



HAL
open science

Risques et perspectives du big data et de l'intelligence artificielle : approche éthique et épistémologique

Relwende Aristide Yameogo

► **To cite this version:**

Relwende Aristide Yameogo. Risques et perspectives du big data et de l'intelligence artificielle : approche éthique et épistémologique. Intelligence artificielle [cs.AI]. Normandie Université, 2020. Français. NNT : 2020NORMLH10 . tel-03357698

HAL Id: tel-03357698

<https://theses.hal.science/tel-03357698>

Submitted on 29 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Normandie Université

THÈSE

Pour obtenir le diplôme de doctorat

Spécialité Informatique

Préparée au sein de l'Université du Havre Normandie

Risques et perspectives du *big data* et de l'intelligence artificielle : approche éthique et épistémologique

Présentée et soutenue par
Relwendé Aristide YAMEOGO

Thèse soutenue publiquement le 28 septembre 2020
devant le jury composé de

| | | |
|-------------------------------|---|------------------------|
| M. Stéfan DARMONI | Professeur des Universités – Praticien Hospitalier Université de Rouen Normandie | Examineur |
| Mme Armelle JACQUET-ANDRIEU | Ingénieure de recherche – HDR Université Paris-Nanterre | Rapporteuse |
| Mme Marie-Christine JAULENT | Directrice de recherche INSERM classe 1 Université Pierre et Marie Curie | Rapporteuse |
| Mme Sylvie LELEU-MERVIEL | Professeure des Universités Université Polytechnique Hauts-de-France | Examinatrice |
| M. Alain MILLE | Professeur des Universités Emérite Université Claude Bernard Lyon 1 | Examineur |
| M. Joël COLLOC | Professeur des Universités Université du Havre Normandie | Directeur de thèse |
| Mme. Béatrice GALINON-MELENEC | Professeure des Universités Emérite Université du Havre Normandie | Co-directrice de thèse |

Thèse dirigée par Joël COLLOC et Béatrice GALINON-MELENEC,
laboratoire IDEES Le Havre, UMR 6266 CNRS



Remerciements

À mon directeur de thèse Prof. Joël COLLOC et ma directrice de thèse Prof. Emérite Béatrice GALINON-MELENEC : merci pour la confiance placée en ma modeste personne pour mener ce travail de thèse. Vos conseils et surtout votre patience ont permis l'aboutissement de cette thèse. Votre rigueur scientifique et votre disponibilité constante ainsi que les qualités humaines qui vous caractérisent ont forcé mon admiration. Vous avez été pour moi au-delà de l'encadrement de cette thèse, des parents.

Au Professeur Marie-Christine JAULENT et au Docteur Armelle JACQUET-ANDRIEU : je vous remercie pour avoir accepté d'être rapporteuses de cette thèse. Je tiens à vous témoigner toute ma gratitude.

Aux Professeurs Sylvie LELEU-MERVIEL, Alain MILLE et Stéfan DARMONI : je suis très honoré que vous ayez accepté d'examiner mon travail et d'être membres du jury de ma thèse. Je vous remercie pour votre disponibilité et vous prie de recevoir l'expression de ma reconnaissance.

Aux Professeurs Athanase BOPDA et Damien OLIVIER : merci d'avoir accepté d'être membres de mon comité de suivi de thèse. Vos conseils m'ont été très bénéfiques au cours de ces années. Veuillez recevoir toute ma gratitude.

Au Laboratoire IDEES Le Havre : vous m'avez offert une famille où j'ai trouvé l'amitié et la solidarité. Merci à tous.

À la famille OUEDRAOGO : j'ai toujours senti que vous êtes ma deuxième famille que j'aime et respecte. Je vous remercie pour tout votre soutien.

Au Prof. Patrice ZABSONRE et à Monsieur Léné SEBGO : votre sens du devoir et vos conseils m'ont permis d'arriver au bout de ce travail et je vous en remercie sincèrement.

À Ronan KIBIROU, Clément MERIC : vous m'avez initié aux systèmes d'information géographique (SIG) et réalisé la cartographie des cabinets médicaux du Havre. Je vous en remercie de ce précieux apport.

À mes amis et collègues docteurs et doctorants (Anne-Solène, Emmanuel, Catalina, Aziz, Diogo, Yao, Stéphanie, Alexandre, Marie) : vos contributions ont été inestimables pour l'élaboration de cette thèse. La bonne ambiance qui régnait entre nous m'a profondément marquée.

Aux médecins du Havre : merci d'avoir dégagé du temps pour participer à mon enquête de terrain malgré vos multiples occupations. Sans votre participation, cette thèse ne serait pas ce qu'elle est aujourd'hui. Recevez toute ma gratitude et mon admiration pour le courage dans l'accomplissement de votre mission sanitaire. Merci **au Docteur LEMAIRE** d'avoir accepté lire la partie « *trace numérique et pratique médicale* » de mon travail.

Dédicace

*Je dédie cette thèse à ma famille pour la patience,
l'amour et la confiance dont elle m'a faite preuve à mon égard.
Cette thèse est aussi le fruit de vos efforts.*

À ma fille Annaëlle : tu me donnes la joie de vivre, le courage et la force pour avancer.

À mon épouse Bertille : ta compréhension et ton soutien sans faille m'ont aidé à avancer dans ce travail. Merci pour les sacrifices consentis pour me suivre dans cette aventure qu'est la thèse, ce travail est le tien. Je te suis infiniment reconnaissant.

À ma mère, Véronique Koudougou ZONGO et à mon père Noraogo YAMEGO : courageux et dévoués ; vous nous avez entourés d'une attention particulière et d'une affection qui nous ont toujours apporté réconfort. Vous nous avez soutenus dans tous les projets que nous avons entrepris. Les mots me manquent aujourd'hui pour vous exprimer toute ma gratitude pour tous les sacrifices et l'ardeur que vous avez endurés pour nous éduquer.

À mes frères : merci pour le soutien fraternel, le courage et la persévérance durant toutes les épreuves que nous avons traversées ensemble. Restons toujours unis dans les moments de bonheur et de souffrance.

Table des matières

| | |
|--|------|
| Remerciements | i |
| Dédicace | iii |
| Table des matières | v |
| Liste des abréviations | xiii |
| Liste des figures..... | xv |
| Listes des tableaux..... | xix |
| Liste des encadrés..... | xxi |
| 1 Introduction générale..... | 1 |
| 1.1 Contexte..... | 3 |
| 1.1.1 Rappel historique..... | 3 |
| 1.1.2 Contexte spécifique de la santé | 4 |
| 1.2 Problématique et hypothèses de recherche | 9 |
| 1.2.1 Problématique des traces numériques et de la privacy dans le cadre du big data et de l'intelligence artificielle. | 9 |
| 1.2.2 Place de l'éthique et questions épistémologiques | 14 |
| 1.2.3 Complexité des questions posées | 16 |
| 1.2.4 Intérêt de notre sujet de thèse | 18 |
| 1.2.5 Objectifs de notre thèse | 19 |
| 2 Méthodologique de la recherche | 21 |
| 2.1 Méthodes..... | 21 |
| 2.1.1 Phase exploratoire | 21 |
| 2.1.2 Méthode d'enquête | 25 |
| 2.2 Techniques de recueil de données | 30 |
| 2.2.1 Technique de questionnaire..... | 30 |
| 2.2.2 Technique d'entrevue ou d'entretien..... | 32 |
| 2.3 Echantillonnage | 33 |
| 2.3.1 Population cible..... | 33 |

| | | |
|-------|--|-----|
| 2.3.2 | Population d'étude..... | 33 |
| 2.4 | Techniques d'analyse..... | 39 |
| 2.4.1 | Techniques d'analyse qualitative | 40 |
| 2.4.2 | Techniques d'analyse quantitative | 41 |
| 2.4.3 | Logiciels d'analyse des données | 41 |
| 2.5 | Réalisation des outils de collecte | 42 |
| 2.5.1 | Guide d'entretien des médecins..... | 44 |
| 2.5.2 | Questionnaire des patients | 46 |
| 2.5.3 | Guide d'entretien des personnes ressources | 46 |
| 2.6 | Déroulement des entretiens et remplissage des questionnaires | 47 |
| 2.7 | Considérations éthiques et juridiques | 47 |
| 3 | Etat de l'art | 49 |
| 3.1 | <i>Big data</i> | 49 |
| 3.1.1 | Définitions | 49 |
| 3.1.2 | Description des technologies du big data | 53 |
| 3.1.3 | Big data et cloud computing..... | 55 |
| 3.2 | Intelligence artificielle | 63 |
| 3.2.1 | Généralités..... | 63 |
| 3.2.2 | Méthodes utilisées dans le développement de l'IA | 77 |
| 3.2.3 | Données de l'IA..... | 115 |
| 3.2.4 | Utilisation de l'IA dans le domaine de la santé..... | 117 |
| 3.3 | Textes juridiques de protection des données de santé | 119 |
| 3.3.1 | Définitions | 121 |
| 3.3.2 | Principes | 122 |
| 3.3.3 | Droits des personnes..... | 126 |
| 3.3.4 | Sécurité et confidentialité des données..... | 130 |
| 3.3.5 | Transfert des données hors de l'UE | 132 |
| 3.4 | Ethique..... | 137 |

| | | |
|-------|---|-----|
| 3.4.1 | Ethique médicale ou théorie du principisme | 140 |
| 3.4.2 | Déontologisme..... | 150 |
| 3.4.3 | Utilitarisme | 152 |
| 3.4.4 | Ethique de la vertu..... | 154 |
| 3.4.5 | Prise de décision en médecine..... | 157 |
| 4 | Système d'aide à la décision médicale à partir d'un espace vectoriel flou : cas de l'embolie pulmonaire | 163 |
| 4.1 | Diagnostic de l'embolie pulmonaire..... | 164 |
| 4.1.1 | Facteurs de risque thromboemboliques | 164 |
| 4.1.2 | Stratégie diagnostique | 165 |
| 4.2 | Espace vectoriel flou : une extension de la logique floue..... | 168 |
| 4.2.1 | Proposition d'un espace vectoriel flou | 169 |
| 4.2.2 | Calcul de la norme de la résultante des trois vecteurs sur E | 171 |
| 4.2.3 | Généralisations à n paramètres..... | 172 |
| 4.2.4 | Représentation du temps | 173 |
| 4.3 | Modèle de la démarche clinique | 174 |
| 4.4 | Modèle temporel flou du diagnostic clinique | 177 |
| 4.5 | Méthodologie de la conception de notre score diagnostique..... | 178 |
| 4.5.1 | Choix des variables | 178 |
| 4.5.2 | Fonction caractéristique des variables..... | 180 |
| 4.6 | Intérêts de l'espace vectoriel flou temporel pour le diagnostic de l'EP | 192 |
| 4.7 | Méthodologie pour la conception d'un EVFT dans un contexte clinique | 193 |
| 4.8 | Perspectives | 197 |
| 5 | Traces numériques et pratique médicale | 199 |
| 5.1 | De la notion de trace | 201 |
| 5.1.1 | Définitions proposées dans le cadre de l'Homme-trace | 201 |
| 5.1.2 | Leur intérêt dans le cadre médical..... | 202 |
| 5.1.3 | Déclinaisons des définitions | 207 |

| | | |
|-------|--|-----|
| 5.2 | Déclinaisons de la notion de trace numérique | 209 |
| 5.2.1 | Enjeu de l'usage statistique des traces numériques..... | 210 |
| 5.2.2 | Nécessité d'une définition scientifique | 215 |
| 5.2.3 | Ce que nous retenons pour notre pratique | 216 |
| 5.2.4 | Prise en compte de l'environnement/ contexte..... | 217 |
| 5.2.5 | Biais interprétatifs nés de la méconnaissance de l'environnement | 218 |
| 5.3 | De la donnée « <i>data</i> » | 220 |
| 5.3.1 | En informatique | 220 |
| 5.3.2 | En SIC | 221 |
| 5.3.3 | Médecine et règles de droit | 223 |
| 5.4 | De l'information et de la connaissance..... | 226 |
| 5.4.1 | Information | 226 |
| 5.4.2 | Connaissance | 227 |
| 5.4.3 | Dynamique des connaissances médicales | 232 |
| 5.5 | Passerelles entre expérience et notions théoriques | 234 |
| 5.5.1 | Médecine clinique et diagnostic en coprésence..... | 236 |
| 5.5.2 | Question de la temporalité..... | 237 |
| 5.5.3 | Connaissance du milieu | 241 |
| 5.5.4 | Notre viatique | 242 |
| 5.6 | Utilisation des traces numériques de santé dans le <i>big data</i> | 245 |
| 5.6.1 | Mode de collecte des données de santé | 246 |
| 5.6.2 | Systèmes d'aide à la décision médicale | 252 |
| 5.6.3 | Interprétation des traces numériques de santé dans le <i>big data</i> | 254 |
| 5.6.4 | Enjeux de l'interprétation des traces dans le domaine de la santé à l'ère du big data et de l'IA..... | 258 |
| 5.6.5 | Définition de la maladie | 259 |
| 5.6.6 | Médecine personnalisée | 262 |
| 6 | Evaluation de l'utilisation des technologies en pratique médicale..... | 269 |

| | | |
|-------|---|-----|
| 6.1 | Description du corpus | 269 |
| 6.2 | Statuts des personnes enquêtées | 270 |
| 6.3 | Méthodes utilisées pour l'analyse des données | 270 |
| 6.3.1 | Analyse qualitative | 270 |
| 6.3.2 | Analyse quantitative | 272 |
| 6.4 | Caractéristiques socio – démographiques des médecins | 274 |
| 6.4.1 | Sexe et âge | 274 |
| 6.4.2 | Spécialité | 274 |
| 6.4.3 | Mode et ancienneté dans la pratique médicale | 274 |
| 6.5 | Caractéristiques globales du corpus de texte des entretiens | 275 |
| 6.5.1 | Analyse de similitude | 277 |
| 6.5.2 | Classification hiérarchique descendante (CHD) selon la méthode Reinert. | 279 |
| 6.5.3 | Analyse factorielle des correspondances (AFC) | 284 |
| 6.6 | Utilisation des technologies de l'information et de la communication dans le cadre de la pratique médicale. | 286 |
| 6.6.1 | Utilisation du dossier patient informatisé..... | 286 |
| 6.6.2 | Utilisation des objets connectés | 289 |
| 6.6.3 | Utilisation des applications et sites web | 299 |
| 6.6.4 | Télémédecine..... | 303 |
| 6.6.5 | Dossier médical partagé (DMP) | 312 |
| 6.6.6 | Assistants médicaux | 318 |
| 6.7 | Gouvernance et gestion des données | 321 |
| 6.7.1 | Participation à la recherche | 321 |
| 6.7.2 | Textes juridiques de protection des données personnelles | 326 |
| 6.7.3 | Confiance des médecins aux institutions pour la protection des données de santé | 328 |
| 6.8 | Big data et intelligence artificielle | 331 |

| | | |
|-------|--|-----|
| 6.8.1 | Modifications du big data et de l'intelligence artificielle dans la pratique médicale..... | 331 |
| 6.8.2 | Promesses du big data et de l'intelligence artificielle dans le domaine de la santé | 340 |
| 6.9 | Finalités éthiques des données dans le cadre du <i>big data</i> et de l'intelligence artificielle..... | 352 |
| 6.9.1 | Bienfaisance | 352 |
| 6.9.2 | Justice | 358 |
| 6.9.3 | Non-malfaisance..... | 362 |
| 6.9.4 | Autonomie | 366 |
| 7 | Vers une intelligence artificielle de confiance | 375 |
| 7.1 | <i>Design thinking</i> | 376 |
| 7.1.1 | Inspiration..... | 378 |
| 7.1.2 | Idéation : imagination et conception | 379 |
| 7.1.3 | Implémentation..... | 380 |
| 7.2 | Application du <i>design thinking</i> dans le domaine de la santé | 381 |
| 7.2.1 | Ecosystème de la santé | 381 |
| 7.2.2 | Différentes étapes de la conception..... | 389 |
| 7.2.3 | Avantages et inconvénients | 395 |
| 8 | Conclusion..... | 399 |
| | Références | 407 |
| | Annexes | i |
| | Annexe 1 : listes des publications et des communications..... | i |
| | Annexe 2 : lettre de demande de terrain à l'ordre des médecins Seine-Maritime..... | i |
| | Annexe 3 : réponse de l'ordre des médecins de la Seine-Maritime | ii |
| | Annexe 4 : lettre de demande de participation des médecins..... | iii |
| | Annexe 5 : formulaire de consentement des médecins | iv |
| | Annexe 6 : fiche d'entrevue des médecins | v |

| | |
|--|-----|
| Annexe 7 : lettre d'information des patients | x |
| Annexe 8 : questionnaire adressée aux patients | xi |
| Annexe 9 : serment d'Hippocrate..... | xv |
| Annexe 10 : participation aux ateliers, séminaires et colloques..... | xvi |
| Résumé | xix |
| Abstract | xx |

Liste des abréviations

| | |
|--------|---|
| ANSM | Agence Nationale de Sécurité du Médicament et des produits de santé |
| ARS | Agence Régionale de la Santé |
| AFC | Analyse Factorielle des Correspondances |
| AP-HP | Assistance Publique-Hôpitaux de Paris |
| APSSIS | Association Pour la Sécurité des Systèmes d'Information de Santé |
| BATX | Baidu, Alibaba, Tencent et Xiaomi |
| BDOA | Bases de Données Orientées Agrégats |
| BDOG | Bases de Données Orientées Graphes |
| CHD | Classification Hiérarchique Descendante |
| CNIL | Commission Nationale de l'Informatique et des Libertés |
| W3C | <i>World Wide Web Consortium</i> |
| DMP | Dossier Médical Partagé |
| EP | Embolie Pulmonaire |
| EVFT | Espace Vectoriel Flou Temporel |
| ESC | <i>European Society of Cardiology</i> |
| GAFAM | Google, Apple, Facebook, Amazon et Microsoft |
| GIP | Groupement d'Intérêt Public |
| HDFS | <i>Hadoop Distributed File System</i> |
| HAS | Haute Autorité de Santé |
| IDEES | Identité et Différenciation des Espaces, de l'Environnement et des Sociétés |
| IA | Intelligence Artificielle |
| MTEV | Maladie Thromboembolique Veineuse |
| NSA | <i>National Security Agency</i> |
| NATU | Netflix, Air BNB, Tesla et Uber |
| NoSQL | <i>Not Only Structured Query Language</i> |
| OLAP | <i>On Line Analytical Processing</i> |
| OLTP | <i>On Line Transaction Processing</i> |
| OMS | Organisation Mondiale de la Santé |
| PMSI | Programme de Médicalisation des Systèmes d'Information |
| RGPD | Règlement Général sur la Protection des Données |
| SIC | Sciences de l'Information et de la Communication |
| SADM | Système d'Aide à la décision Médicale |

| | |
|---------|---|
| SIA | Système d'Information Autonome |
| SI-VIC | Système d'Information d'Identification Unique des Victimes |
| SNDS | Système National des Données de Santé |
| SNIIRAM | Système National d'Information Inter-régimes de l'Assurance Maladie |
| TIC | Technologies de l'Information et de la Communication |
| UML | <i>Unified Modeling Language</i> |
| UE | Union Européenne |

Liste des figures

| | |
|---|-----|
| Figure 1 : pourcentage de saturation en fonction du nombre de personnes enquêtés (GRIFFIN, HAUSER, 1993) | 34 |
| Figure 2 : méthode d'échantillonnage des cabinets et des médecins | 35 |
| Figure 3 : répartition spatiale des cabinets médicaux de la commune du Havre | 36 |
| Figure 4 : répartition spatiale des structures de soins enquêtées dans la commune du Havre | 37 |
| Figure 5 : carte des données santé d'un patient | 52 |
| Figure 6 : modèles de service de <i>cloud computing</i> (FIGER, 2012) | 58 |
| Figure 7 : approches de la définition de l'IA en quatre catégories..... | 67 |
| Figure 8 : contenu d'un système expert | 79 |
| Figure 9 : comparaison entre un ensemble classique et un ensemble flou (CHEVRIE, GUËLY, 1998)..... | 83 |
| Figure 10 : fonction d'appartenance d'un ensemble flou et d'un ensemble classique..... | 83 |
| Figure 11: caractéristiques d'un ensemble flou | 85 |
| Figure 12: variable linguistique « Fréquence cardiaque »..... | 87 |
| Figure 13: variable linguistique « Variation de la tension artérielle systolique » | 87 |
| Figure 14: variable linguistique « Risque d'embolie pulmonaire »..... | 88 |
| Figure 15: représentation des valeurs de notre exemple | 90 |
| Figure 16: inférence de Mamdani: étapes d'agrégation et d'activation | 91 |
| Figure 17: inférence de Mamdani : étape d'agrégation des conclusions | 92 |
| Figure 18: risque d'embolie pulmonaire selon la méthode de la moyenne des maximas.... | 93 |
| Figure 19: risque d'embolie pulmonaire selon la méthode du barycentre..... | 94 |
| Figure 20 : schématisation de l'apprentissage supervisé (CLERGUE, 2009) | 100 |
| Figure 21 : schématisation de l'apprentissage non supervisé (CLERGUE, 2009) | 100 |
| Figure 22 : modèle de régression linéaire..... | 101 |
| Figure 23 : modèle d'arbre de décision pour le diagnostic de l'embolie pulmonaire sans signe de choc ou d'hypotension artérielle | 103 |
| Figure 24 : modèle de SVM (VANNIEUWENHUYZE, 2019)..... | 105 |
| Figure 25 : représentation d'un neurone | 108 |
| Figure 26 : principe d'activation d'un neurone (VANNIEUWENHUYZE, 2019)..... | 108 |
| Figure 27 : principe de fonctionnement du réseau de neurone..... | 109 |
| Figure 28 : réseau de neurone à une couche..... | 110 |
| Figure 29 : modèle de réseau de neurone avec couche cachée..... | 112 |

| | |
|--|-----|
| Figure 30: principe général des algorithmes génétiques (ALLIOT, DURAND, 2005)..... | 115 |
| Figure 31 : cheminement des repères ou « <i>boite à outils</i> » (LERY, COLLOC, 2008)..... | 161 |
| Figure 32 : algorithme diagnostic en cas de suspicion d'une EP avec signe de choc ou d'hypotension artérielle (DELAHAYE, 2015) | 167 |
| Figure 33 : algorithme diagnostic en cas de suspicion d'une EP sans signe de choc ou d'hypotension artérielle (DELAHAYE, 2015) | 167 |
| Figure 34 : fonction sigmoïde définie de $0,1 \rightarrow 0,1$ | 170 |
| Figure 35 : fonction sigmoïde définie de $[-1,1] \rightarrow [-1,1]$ | 171 |
| Figure 36 : vecteur de la fonction $F(x): -1, 1 \rightarrow -1, 1$ | 171 |
| Figure 37 : schéma UML général de l'évolution d'une maladie | 175 |
| Figure 38 : évolution de la maladie (M) dans le temps en fonction des tableaux cliniques (TC), des syndromes (Sd) et des signes (S) (COLLOC, SUMMONS, 2006) | 176 |
| Figure 39 : espace vectoriel flou avec les vecteurs composant les variables et leurs attributs (COLLOC, SUMMONS, 2015)..... | 178 |
| Figure 40: fonction caractéristique $\mu_{\text{RisqueAge}}(x)$ | 182 |
| Figure 41 : fonction caractéristique du risque thromboembolique $\mu_{\text{ThE}}(H(x))$ | 183 |
| Figure 42 : fonction caractéristique floue μ_{kcEP} attribut discret risque d'embolie pulmonaire selon le stade de cancer $x \in [-1,4]$ | 184 |
| Figure 43 : fonction caractéristique floue du risque d'embolie pulmonaire lié au délai écoulé après une fracture ou une intervention chirurgicale..... | 185 |
| Figure 44 : fonction caractéristique de l'augmentation unilatérale du volume du mollet $\mu_{\text{RdiffTailleMolletEP}}(x,t)$ | 187 |
| Figure 45 : fonction caractéristique du Risque d'EP lié à une fréquence cardiaque élevée $\mu_{\text{RFréquenceCard}}(x,t)$ | 188 |
| Figure 46 : fonction caractéristique du risque lié à la tension artérielle $\mu_{\text{RHypoTASystolique}}(x,t)$ | 188 |
| Figure 47 : fonction caractéristique $\mu_{\text{R}\Delta t\text{TA}}$ variation de TAS en 15 minutes..... | 189 |
| Figure 48: fonction résultante de l'EP | 191 |
| Figure 49 : évolution des variables de l'EP de la suspicion au diagnostic | 191 |
| Figure 50: modèle EPICE..... | 194 |
| Figure 51: méthodologie pour la conception d'un EVFT | 195 |
| Figure 52: représentation thématique des discours des acteurs de la lutte contre la Covid-19 dans la presse..... | 203 |
| Figure 53 : statistiques sur l'utilisation du numérique en France (CREDOC, 2019) | 211 |

| | |
|---|-----|
| Figure 54: tracking des personnes contact du Covid-19 en Corée du Sud : sources, catégories et distribution des données collectées (PARK ET AL., 2020)..... | 214 |
| Figure 55 : structuration des données de santé lors d'un épisode de maladie (LELONG, 2019) | 225 |
| Figure 56 : épistémologie possible de la connaissance (ERMINE, 2000)..... | 228 |
| Figure 57: représentation de la production de la connaissance | 230 |
| Figure 58 : cycle de création de la connaissance (NONAKA, 1994)..... | 231 |
| Figure 59 : approches fondamentales de la médecine : médecine clinique et santé publique (JOËL COLLOC, 2015)..... | 235 |
| Figure 60: enjeux et risques de la circulation des traces numériques en matière de santé | 236 |
| Figure 61: processus d'analyse et d'interprétation des signes-traces | 244 |
| Figure 62 : différents types de données figurants dans la base du SNDS | 247 |
| Figure 63 : différentes bases de données alimentant le SNDS..... | 249 |
| Figure 64 : circonstances de partage des informations confidentielles par les employés (DELL, 2017)..... | 250 |
| Figure 65 : comportements à risque des employés (DELL, 2017) | 250 |
| Figure 66 : proportion des attaques de programmes malveillants dans le monde (DELL, 2016)..... | 251 |
| Figure 67 : processus de la science de la donnée (O'NEIL, SCHUTT, 2013) | 255 |
| Figure 68 : graphique logarithmique des rangs et des fréquences..... | 276 |
| Figure 69 : fréquence et type des mots utilisés sous forme active dans le corpus de texte | 277 |
| Figure 70 : graphe de similitude - arbre maximum - du corpus | 278 |
| Figure 71 : dendrogramme montrant les différentes classes (CHD) | 280 |
| Figure 72 : CHD avec les mots associés aux différentes classes | 280 |
| Figure 73 : CHD avec les différentes thématiques des classes | 282 |
| Figure 74 : classification thématique du corpus à partir de la CHD | 283 |
| Figure 75 : pourcentage des différents facteurs de l'AFC | 284 |
| Figure 76 : AFC des classes de la CHD | 285 |
| Figure 77 : AFC des mots associés aux classes de la CHD..... | 286 |
| Figure 78 : utilisation des TIC par les médecins | 289 |
| Figure 79 : utilisation des objets connectés en fonction des indications (n = 07)..... | 291 |
| Figure 80 : utilisation des applications et sites web en fonction des indications (n = 12) | 299 |
| Figure 81 : note d'information sur le site de la SNDS sur les accès aux données..... | 314 |

| | |
|--|-----|
| Figure 82: degré de confiance des médecins pour les capacités des textes juridiques dans la protection des données | 327 |
| Figure 83: opinions des médecins sur la capacité des textes juridiques à protéger les données de santé..... | 328 |
| Figure 84: degré de confiance des médecins aux différentes institutions pour la protection des données de santé..... | 329 |
| Figure 85: degré de confiance des médecins envers les hôpitaux et les cabinets privés pour la protection des données | 330 |
| Figure 86: modifications du big data et de l'intelligence artificielle dans la pratique médicale..... | 332 |
| Figure 87 : promesses du big data et de l'IA dans le domaine de la santé | 340 |
| Figure 88 : finalités éthiques de l'utilisation des données de santé dans le cadre de la bienveillance..... | 356 |
| Figure 89 : détails des réponses des médecins par item de questionnaire relatif à la bienfaisance | 357 |
| Figure 90 : finalités éthiques de l'utilisation des données de santé dans le cadre de la justice | 360 |
| Figure 91 : détails des réponses des médecins par item de questionnaire relatif à la justice | 361 |
| Figure 92 : finalités éthiques de l'utilisation des données de santé dans le cadre de la non-malfaisance..... | 364 |
| Figure 93 : détails des réponses des médecins par item du questionnaire relatif à la non-malfaisance..... | 365 |
| Figure 94 : finalités éthiques de l'utilisation des données de santé dans le cadre de l'autonomie..... | 369 |
| Figure 95 : détails des réponses des médecins par item du questionnaire relatif à l'autonomie..... | 370 |
| Figure 96 : principes du design thinking à la recherche de solution idéale..... | 378 |
| Figure 97 : acteurs de l'écosystème des soins de santé..... | 386 |
| Figure 98 : interactions entre les différents acteurs de l'écosystème de soins de santé..... | 393 |
| Figure 99 : différentes étapes de conception d'un outil d'IA en santé | 394 |

Listes des tableaux

| | |
|---|-----|
| Tableau 1: différentes règles d'inférence possibles..... | 89 |
| Tableau 2 : facteurs prédisposant à une MTEV (KONSTANTINIDES ET AL., 2019)..... | 164 |
| Tableau 3 : score de Wells (DELAHAYE, 2015)..... | 165 |
| Tableau 4: score de Genève (DELAHAYE, 2015)..... | 166 |
| Tableau 5 : dictionnaire des attributs des objets maladie et de leurs composants..... | 175 |
| Tableau 6 : différentes variables prises en compte dans notre modèle de score..... | 179 |
| Tableau 7 : fonction caractéristique floue $\mu_{\text{RisqueAge}}(H(x))$ | 181 |
| Tableau 8 : fonction caractéristique $\mu_{\text{cancer}}(H(x))$ selon le stade du cancer..... | 184 |
| Tableau 9 : paramétrage des valeurs de comparaison de la taille des mollets gauche et droite en cm..... | 186 |
| Tableau 10 : paramétrage de la variable de l'EP..... | 190 |
| Tableau 11 : répartition des médecins en fonction de leur spécialité..... | 274 |
| Tableau 12 : répartition des médecins en fonction de leur ancienneté de pratique..... | 275 |
| Tableau 13 : médiane du score unitaire éthique en fonction de l'ancienneté dans la pratique médicale..... | 352 |

Liste des encadrés

| | |
|--|-----|
| Encadré 1: production des données de santé en France | 51 |
| Encadré 2: instruments juridiques européens | 119 |
| Encadré 3 : éthique selon Paul Ricoeur (RICOEUR, 2015)..... | 158 |
| Encadré 4: coupures de presse chronologiques sur la maladie à coronavirus..... | 203 |
| Encadré 5 : capacités et compétences numériques (INSEE, 2019)..... | 212 |
| Encadré 6 : coût de la location d'un serveur de données de santé. | 288 |
| Encadré 7 : définition d'un dispositif médical | 290 |
| Encadré 8: sécurité des objets connectés..... | 294 |
| Encadré 9: mise en œuvre de la télémédecine..... | 304 |
| Encadré 10: sécurité des données dans les hôpitaux (GALLOIS, 2016) | 337 |
| Encadré 11 : conditions matérielles et techniques de la mise en place du <i>big data</i> | 347 |
| Encadré 12: analyse de l'étude sur l'hydroxychloroquine (MEHRA ET AL., 2020)..... | 371 |

1 Introduction générale

Les avancées technologiques ont démultiplié les circuits d'accès aux données¹, leur traitement et leur diffusion. Les données sont devenues la matière première des systèmes d'information et sont indispensables à la vie et au développement des organisations ; surtout avec la croissance rapide des capacités de calcul des processeurs ces dernières années. Nous assistons, ainsi, à une explosion du stockage et du traitement des données avec les technologies de l'information et de la communication (TIC). Ces technologies ont atteint un niveau de développement très important avec pour aboutissement l'apogée de technologies connexes que sont : le *big data*², le *cloud computing*³, les objets connectés⁴ et l'intelligence artificielle (IA).

Le *big data*, terme très utilisé de nos jours mais méconnu des Français⁵, s'est développé dans un premier temps dans les domaines de l'économie, de la gestion et du marketing pour la gestion de la relation client, l'amélioration des processus opérationnels, la création de

¹ Le terme « donnée » est polysémique et sa définition dépend du domaine dans lequel il est employé. Dans notre contexte, il sera utilisé dans le domaine de l'informatique et des sciences de l'information et de la communication (SIC).

En Informatique, Selon Le Petit Robert 2014, la donnée est une « représentation conventionnelle d'une information (fait, notion, ordre d'exécution) sous une forme (analogique ou digitale) permettant d'en faire le traitement automatique ».

En SIC, nous avons retenu la définition de Jacques Mélése: « une donnée est un enregistrement dans un code convenu par un groupe social, de certains attributs d'un objet ou d'un événement ». (LELEU-MERVIEL, 2008).

² Il n'existe pas de définition unifiée et précise du *big data* qui est un objet complexe polymorphe et qui comporte des définitions différentes selon les communautés et les domaines. Le terme « *big data* » signifie « méga données » en français faisant référence à la quantité des données. Mais la définition va au-delà de la quantité.

Défini au départ par Douglas Laney en 2001 (LANEY, 2001) selon les trois mots « volume – variété – vitesse », les 3V, ce terme est défini de nos jours surtout par ses caractéristiques en 5V (volume – variété – vitesse – véracité – valeur).

³ « Cloud computing ou « informatique en nuage » est un ensemble de processus qui consiste à utiliser la puissance de calcul et de stockage des serveurs informatiques distants à travers un réseau, généralement Internet ». (MONINO, SEDKAOUI, 2016).

⁴ « Les objets connectés ou Internet des objets se définissent comme un réseau de réseaux qui permet, via des systèmes d'identification électronique normalisés et unifiés, et des dispositifs mobiles sans fil, d'identifier directement et sans ambiguïté des entités numériques et des objets physiques et ainsi de pouvoir récupérer, stocker, transférer et traiter, sans discontinuité entre les mondes physiques et virtuels, les données s'y rattachant. » (BENGHOZI ET AL., 2012).

Il n'existe pas une définition consensuelle et acceptée de tous pour ce concept. Cette définition a l'avantage d'être technique et centrée sur les usages.

⁵ Lors d'une enquête réalisée en ligne du 29 février au 2 mars 2016. Échantillon de 1 487 personnes, représentatif des Français âgés de 18 ans et plus, 59% des Français disent ne pas connaître la signification du terme « *big data* » et 87% sont mal informés de l'utilisation du *big data* par les entreprises et les organismes publics. Source : http://harris-interactive.fr/opinion_polls/big-data-quen-pensent-les-francais/ consulté le 20 avril 2017.

nouvelles valeurs et la prise de décision efficiente. Mais ces dernières années, les envies d'utilisation dans le domaine de la santé connaissent un essor important ; envies liées notamment aux promesses d'économies énormes sur les dépenses de la santé, à leur impact sur la recherche dans les domaines médicaux et pharmaceutiques et à l'amélioration de la prise en charge des individus.

Le *big data* a pour objectif de transformer les données brutes en connaissance directement exploitable en lien direct avec les consommateurs. Dans le domaine de la santé, les concepteurs promettent des réductions de dépenses de santé, l'amélioration de la qualité de soin et des politiques de santé. En effet, l'étude McKinsey (MANYIKA ET AL., 2011) sur le système de santé aux Etats Unis a montré qu'en analysant les données de santé⁶, on notait une amélioration considérable de l'efficacité, de l'efficience et de la qualité du système de santé. Cette utilisation des données de santé pourrait créer une valeur de 300 milliards de dollars américains dont deux tiers sous forme de baisse des dépenses de santé. La protection de la santé est non seulement un réflexe individuel mais aussi une obligation pour l'Etat vis-à-vis de la population ; Ainsi, du fait de la rareté des ressources financières et de la recherche permanente d'efficience et de qualité, l'Etat s'est saisi de l'opportunité qu'offre le numérique pour mettre en place une politique de santé fondée sur les TIC.

⁶ La définition de la « donnée de santé » n'existe pas pour le moment dans le droit français. L'Union Européenne à travers le Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données la définit comme suit « *Les données à caractère personnel concernant la santé devraient comprendre l'ensemble des données se rapportant à l'état de santé d'une personne concernée qui révèlent des informations sur l'état de santé physique ou mentale passé, présent ou futur de la personne concernée. Cela comprend des informations sur la personne physique collectées lors de l'inscription de cette personne physique en vue de bénéficier de services de soins de santé ou lors de la prestation de ces services au sens de la directive 2011/24/UE du Parlement européen et du Conseil au bénéfice de cette personne physique; un numéro, un symbole ou un élément spécifique attribué à une personne physique pour l'identifier de manière unique à des fins de santé; des informations obtenues lors du test ou de l'examen d'une partie du corps ou d'une substance corporelle, y compris à partir de données génétiques et d'échantillons biologiques; et toute information concernant, par exemple, une maladie, un handicap, un risque de maladie, les antécédents médicaux, un traitement clinique ou l'état physiologique ou biomédical de la personne concernée, indépendamment de sa source, qu'elle provienne par exemple d'un médecin ou d'un autre professionnel de la santé, d'un hôpital, d'un dispositif médical ou d'un test de diagnostic in vitro.* »

1.1 Contexte

1.1.1 *Rappel historique*

Dès 1998, l'Etat s'est lancé dans la mise en place de la télémédecine⁷ afin d'améliorer l'accès aux soins à un plus grand nombre d'individus. En 2000, le premier arrêté pour la mise en œuvre de la télémédecine⁸ dans un hôpital a été signé. En avril 2005, l'Etat a mis en place un groupement d'intérêt public (GIP-DMP) qui devait mettre en œuvre le dossier médical personnel aujourd'hui devenu le dossier médical partagé⁹ (DMP). Le but du DMP était d'une part, de fournir au médecin traitant l'information la plus complète pour qu'il puisse proposer le traitement ou les examens les plus adaptés et, d'autre part, d'éviter des redondances inutiles d'examens ou de prescriptions.

L'État a encouragé le numérique dans tous les domaines avec la promulgation de la loi n° 2016-1321 du 7 octobre 2016 pour une République numérique¹⁰. L'objectif de l'utilisation du numérique par l'Etat est de réduire les dépenses de santé et d'améliorer la qualité de la prise en charge des patients. Cette loi a été suivie d'actions telles que la relance du DMP et la création d'un « *Data Health Hub* » pour améliorer l'accès aux données de santé à la recherche.

L'utilisation des TIC dans le domaine de la santé a donné naissance à l'e-santé. Pour l'Organisation Mondiale de la Santé (OMS), l'e-santé se définit comme « *les services du numérique au service du bien-être de la personne* » (WHO, 2016). Elle se définit également comme « *l'utilisation des outils de production, de transmission, de gestion et de partage d'informations numérisées au bénéfice des pratiques tant médicales que médico-sociales* » (WHO, 2016). Elle ne se résume pas à la télémédecine et ne doit pas être assimilée au vaste

⁷ Décret no 98-305 du 17 avril 1998 portant publication de l'entente en matière de santé entre le Gouvernement de la République française et le Gouvernement du Québec, signée à Paris le 25 mars 1997.

⁸ Arrêté du 30 mars 2000 relatif à la mise en œuvre par le service de radiologie de l'hôpital d'instruction des armées Sainte-Anne à Toulon d'un traitement automatisé d'informations nominatives dans le cadre d'un réseau de télémédecine en urgences neurochirurgicales.

⁹ Le dossier médical partagé est un carnet de santé numérique accessible de manière sécurisée en ligne à tout moment par le patient et les professionnels de santé autorisés. Il contient les informations médicales suivantes : les compte-rendu hospitaliers et radiologiques, résultats d'analyses de biologie, antécédents et allergies, actes importants réalisés, médicaments qui ont été prescrits et délivrés. Source : <http://www.dmp.gouv.fr/> consulté le 22 avril 2017.

¹⁰ Loi n° 2016-1321 du 7 octobre 2016 pour une République numérique. 7 octobre 2016.

ensemble des TIC-santé qui englobent l'ensemble des applications numériques au service de l'offre de soins.

Mais le domaine de la santé a une particularité spécifique - l'humain comme objet d'étude - qui mérite d'être prise en compte dans l'utilisation des TIC ; car une négligence de cette spécificité peut aboutir à des conséquences importantes pour le système de santé, les individus et la société.

1.1.2 Contexte spécifique de la santé

Le domaine de la santé est un système complexe¹¹ très dynamique qui repose sur la médecine et l'Homme qui sont également des systèmes complexes. Aujourd'hui, la combinaison du *big data*, de l'intelligence artificielle¹² (IA) et des objets connectés conduit à une médecine préventive et individuelle qui prend en compte les vulnérabilités de chaque personne en vue d'assurer un meilleur suivi (un analyseur électrocardiographique surveille en temps réel les alertes cardiovasculaires, la surveillance des patients chroniques à distance ou de l'activité sportive et des paramètres vitaux des patients, etc...). Toutes ces technologies aboutissent à une production massive de données qui nécessitera des moyens pour les stocker, les gérer et les utiliser. Toutefois, la sécurité des données devrait être prise en compte afin de protéger le secret médical - fondement de la pratique médicale -, la vie privée, l'intimité des individus et garantir l'accessibilité des données à tout moment aux soignants afin de leur permettre d'assurer la continuité des soins.

Le *big data* prend essentiellement naissance dans et par l'informatique ; dépassant les solutions classiques existantes pour la collecte, le stockage et le traitement des données. Le changement d'échelle, qui a fait passer des bases de données aux masses de données, réinterroge ces problématiques. Concernant la collecte des données tout d'abord, les

¹¹ « Un système complexe est par définition un système que l'on tient pour irréductible à un modèle fini quelle que soit sa taille, le nombre de ses composants, l'intensité de leur interaction. Pour un observateur, il est complexe parce qu'il tient pour certain l'imprévisibilité potentielle des comportements. » (LE MOIGNE, 1999).

¹² Le terme « intelligence artificielle », créé par John McCarthy, est souvent abrégé par le sigle « IA » (ou « AI » en anglais, pour *Artificial Intelligence*). Marvin Lee Minsky la définit, comme « la construction de programmes informatiques qui s'adonnent à des tâches qui sont, pour l'instant, accomplies de façon plus satisfaisante par des êtres humains car elles demandent des processus mentaux de haut niveau tels que : l'apprentissage perceptuel, l'organisation de la mémoire et le raisonnement critique ». Source : https://www.larousse.fr/encyclopedie/divers/intelligence_artificielle/187257 consulté le 2 novembre 2017.

méthodes de fouille des réseaux sociaux comme l'intégration de données provenant de multiples capteurs (objets connectés, dispositifs de surveillance médicale, puces, smartphones, etc...) et leur croisement est riche de potentialités applicatives tout en posant différemment les problématiques liées à la vie publique / vie privée et à la visibilité plus ou moins contrôlée et plus ou moins contrôlable via les algorithmes, des individus. Le stockage et le traitement calculatoire de ces données, la croissance des capacités de stockage et la structuration, des *data center* (centres de données ou ferme de données), ouvrent des problématiques aussi bien techniques (pour l'organisation et l'optimisation de ces systèmes), qu'énergétiques ou économiques. Pour l'exploitation de ces données, le *big data* a imposé la découverte de nouvelles méthodes dans des domaines pourtant assez anciens, comme le *data mining*¹³ (fouille de données) pour l'exploitation efficace de ces données par l'IA en intégrant à la fois des méthodes de *machine learning*¹⁴ et de *deep learning*¹⁵. Le traitement des données massives pose de sérieux problèmes¹⁶ juridiques - notamment en ce qui concerne leur contrôle, leur protection, leur sécurité et leur propriété intellectuelle - et des questions sur la commensurabilité du parcours de vie et du parcours de soins à travers le numérique.

L'innovation technologique du *big data*, dont l'Internet des objets (objets connectés), et les stratégies de ciblage, de marketing et de visibilité auxquelles elles donnent lieu, forcent à interroger le cadre juridique qui leur est applicable. La visibilité sur le web implique également un questionnement sur la portée des droits fondamentaux à caractère informationnel. S'il existe une volonté de renforcer les droits des personnes sur leurs

¹³ Le *data mining* ou extraction de connaissances à partir des données, a pour objet l'extraction d'un savoir ou d'une connaissance à partir de grandes quantités de données, par des méthodes automatiques ou semi-automatiques. Il utilise un ensemble d'algorithmes issus de disciplines diverses telles que les statistiques, l'intelligence artificielle ou l'informatique pour construire des modèles à partir des données, c'est-à-dire trouver des structures intéressantes ou des motifs selon des critères fixés au préalable, et d'en extraire un maximum de connaissances utiles à l'entreprise. (MONINO, SEDKAOUI, 2016).

¹⁴ Le *machine learning* est une méthode d'analyse de données permettant d'automatiser le développement de modèle analytique. Par le biais d'algorithmes capables d'apprendre de manière itérative, le *machine learning* permet aux ordinateurs de découvrir des *insights* cachés sans être programmés pour savoir où les chercher.

¹⁵ Le *deep learning*, ou apprentissage profond, fait référence à un type d'intelligence artificielle, qui lors d'analyses complexes, les caractéristiques essentielles du traitement des données ne seront plus identifiées par un traitement humain dans l'algorithme préalable, mais directement par l'algorithme de *deep learning*.

¹⁶ Le géant du web Facebook est accusé aux Etats Unis pour avoir influencé sur les résultats de la campagne électorale présidentielle aux Etats Unis ayant abouti à l'élection de Donald Trump. Dans le cadre des procédures judiciaires, il a transmis des données des comptes utilisateurs, des messages publicitaires et des métadonnées à la Commission du Congrès Américain. Source : <http://www.lefigaro.fr/secteur/high-tech/2017/09/22/32001-20170922ARTFIG00157-election-de-trump-facebook-va-cooperer-avec-le-congres-sur-l-enquete-russe.php> consulté le 02 novembre 2017.

données, ceux-ci doivent subir de nécessaires limites, de façon à ce que certaines données puissent être traitées légitimement.

L'utilisation de l'Internet dans le domaine de la santé et des objets connectés laissent des traces¹⁷ dont l'impact social pour les individus sur l'utilisation de leurs données de santé est considérable. Les technologies du *big data* sont fondées sur le web sémantique¹⁸ et l'interrogation transversale (*data mining*) des bases de données permet le profilage qui conduit à exclure tacitement des personnes (COLLOC, 2015). Ces risques d'exclusion ont été rappelés dans le discours de Edith Ramirez (RAMIREZ, 2014) sous la notion de « *discrimination par algorithme* »¹⁹ consécutif au *big data*.

La pratique de la médecine est inconcevable sans l'utilisation des données personnelles. L'exploitation informatisée des données personnelles dans le domaine de la santé est très sensible en ce qu'elles font intimement partie de la vie des individus et de leur famille. Pour protéger la vie privée des individus, le législateur a prévu la loi relative à l'informatique, aux fichiers et aux libertés²⁰ depuis 1978 en France complétée récemment par le règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016²¹ relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant ainsi la directive 95/46/CE²². Ainsi, la question de la protection de la vie privée des individus était déjà abordée dans les années 1970 avant même le développement important des TIC, témoignant de l'importance de celle-ci au sein de la population mais également au niveau de l'Etat. Elle a abouti à la création de la

¹⁷ Le Petit Robert 2014 définit la trace comme « *empreinte ou suite d'empreintes, de marques que laisse le passage d'un être ou d'un objet* », « *marque laissée par une action quelconque* », « *ce à quoi on reconnaît que quelque chose a existé, ce qui subsiste d'une chose passée* », « *très petite quantité perceptible* ». La trace est polysémique et fonction du contexte et du domaine de spécificité.

¹⁸ « *Le web sémantique est un ensemble de technologies visant à rendre l'ensemble des ressources du web accessibles, intelligibles et utilisables par des programmes et agents logiciels, grâce à un système de métadonnées. Le web sémantique représente un ensemble de standards développés et promus par le W3C (World Wide Web Consortium) pour permettre la représentation, la manipulation des connaissances par les outils du web* ». (MONINO, SEDKAOUI, 2016).

¹⁹ L'assureur Generali adapte la prime d'assurance de ses conducteurs en fonction des comportements de conduites de ses assurés à travers la collecte des données par l'intermédiaire des objets connectés.

²⁰ Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés.

²¹ Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données.

²² Directive 95/46/CE du Parlement européen et du Conseil, du 24 octobre 1995, relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données.

Commission Nationale de l'Informatique et des Libertés (CNIL)²³. Toutefois, sur le plan individuel, le partage des données de santé des personnes ne peut se concevoir qu'entre les soignants dans le respect du secret professionnel afin de préserver la confiance dans la relation de soin. Le DMP doit assurer la sécurité (disponibilité, intégrité, traçabilité, confidentialité) des données. Mais le secret médical est actuellement remis en question par l'utilisation des TIC comme les *clouds* administrés le plus souvent à l'étranger ou des hébergeurs²⁴ accrédités par le ministère de la santé pour le stockage des données de santé et l'entrée en jeu d'autres acteurs ne faisant pas partie de l'équipe de soins. Le « *privacy shield*²⁵ » a donc été mis en place pour permettre la protection des données des ressortissants de l'Union Européenne transférées²⁶ vers les Etats Unis en remplacement du « *Safe Harbor* ». L'anonymat est devenu pratiquement « *impossible* » à l'heure du tout numérique. Le concept de « *privacy by design*²⁷ » est l'une des solutions proposées par Ann Cavoukian (CAVOUKIAN, 2006) pour assurer un minimum de protection des données personnelles.

D'une manière générale, comment préserver une part d'intimité dans l'usage d'Internet, du téléphone portable et des réseaux sociaux au niveau mondial lorsque les appareils collectent nos informations sans que nous n'en soyons informés ou quand nous sommes obligés d'accepter les conditions d'utilisation d'un service ou d'une application sans savoir comment les données seront traitées (finalités des données). La sécurité de la population et de la nation

²³ En 1974, le ministère de l'intérieur avait bâti un fichier informatisé au nom évocateur : SAFARI, acronyme de système automatisé pour les fichiers administratifs et le répertoire des individus. Ce système prévoyait de créer une base de données centralisée de la population, en utilisant le fichier de sécurité sociale comme identifiant commun à tous les fichiers administratifs. Devant le tollé généralisé provoqué par ce projet ; le journal Le Monde allant jusqu'à titrer « *SAFARI, ou la chasse aux Français* », le Premier ministre de l'époque, Pierre Messmer, ne put que le retirer et créer dans la foulée une commission dite Informatique et liberté, chargée de proposer une réglementation sur l'utilisation des moyens informatiques. Son travail débouchera sur la publication, dès 1975, du rapport TRICOT, dont le projet de loi relatif à l'informatique et aux libertés est le fils naturel, pour reprendre l'expression d'Alain PEYREFITTE, alors Garde des Sceaux. C'est dans ce contexte encore tendu que s'engage, au Sénat, le 17 novembre 1977, la discussion en première lecture du projet de loi, adopté par l'Assemblée nationale, relatif à l'informatique et aux libertés. Source : <http://www.senat.fr/evenement/archives/D45/context.html>, consulté le 25 juillet 2017

²⁴ Ordonnance n° 2017-27 du 12 janvier 2017 relative à l'hébergement de données de santé à caractère personnel.

²⁵ Le *Privacy Shield* organise et encadre le transfert des données personnelles européennes vers les *Etats Unis* en remplacement du *Safe Harbor* dont les contrôles, particulièrement faibles, liés aux mécanismes d'auto-certification des entreprises américaines. Le *privacy shield* a pour but de combler le déséquilibre existant entre les entreprises américaines et les entreprises européennes en matière de protection des données personnelles.

²⁶ La notion de transfert de données vers un pays tiers a été définie par la CNIL comme « *toute communication, toute copie ou déplacement de données par l'intermédiaire d'un réseau, ou toute communication, toute copie ou déplacement de ces données d'un support à un autre, quel que soit ce support, dans la mesure où ces données ont vocation à faire l'objet d'un traitement dans le pays destinataire* ».

²⁷ Le *privacy by design* consiste à intégrer les principes de protection de la vie privée dès la conception des outils, dans leur architecture. C'est un principe de loi qui est intégré dans le règlement Européen de protection des données personnelles. Source : http://data.europa.eu/eli/dec_impl/2016/1250/oj, consulté le 25 juillet 2017.

rend nécessaire pour les autorités, sous certaines conditions, de pouvoir accéder aux données personnelles pour lutter contre le terrorisme, certaines crises sanitaires²⁸, les conduites déviantes criminelles ou délictuelles ; mais la question est de définir les limites entre les besoins sécuritaires et la protection de la vie privée des individus posant alors les problèmes d'éthique de l'utilisation du profilage des individus et des risques d'une dérive totalitaire conduisant à la dilution de l'individu dans la société et d'une atteinte aux valeurs de la République dans le contexte de la mondialisation.

Ces craintes ont été confirmées par l'affaire Edwards Snowden par la divulgation des secrets du programme *Prism*²⁹ (LE MONDE, 2013) de la NSA (*National Security Agency*) aux Etats Unis et l'affaire d'identification des gilets jaunes blessés à travers la base de données « système d'information d'identification unique des victimes » (*SI-VIC*), dont l'objectif est l'identification et le suivi des victimes de situations sanitaires exceptionnelles, comme les attentats³⁰. Des efforts d'adaptation des textes législatifs ont été faits aussi bien en France qu'en Europe pour la protection et le traitement des données, mais ils restent insuffisants face à l'évolution du *big data*, de l'IA et des objets connectés.

Le *big data* offre des nouveaux moyens d'investigation, de stockage, d'indexation et de comparaison fondés initialement sur les technologies de base de données. Les systèmes multi-agents d'aide à la décision médicale assurent la coopération des connaissances et des ontologies médicales apportant ainsi une aide précieuse aux médecins pour la prise en charge efficace des patients (COLLOC, SYBORD, 2003). Le *big data* présente également un intérêt en épidémiologie et en santé publique (détection des conduites inadéquates, problèmes de pharmacovigilance et des comportements des populations).

²⁸ L'épidémie de maladie à coronavirus est un cas d'école qui vient nous rappeler les capacités de traçage des personnes mais surtout la faiblesse de la protection de la vie privée à l'ère du numérique (Facebook et Google transmettent des données de géolocalisation aux chercheurs pour prédire l'évolution de la maladie), mais également la manière dont les Etats utilisent les lois disponibles sans prise en compte des contextes réelles et de l'impact de l'utilisation des données sur les personnes.

²⁹ Le programme *Prism* est un programme de surveillance mis en place par les Etats Unis pour suivre de manière étendue l'activité en ligne de nombreux individus. Il permet à la NSA de collecter des informations (courriels, photos, fichiers, audio, vidéos et réseaux sociaux) auprès des entreprises américaines dont la plupart sont les géants du web (Microsoft, Apple, Yahoo, Skype, YouTube, Google).

³⁰ Identification des gilets jaunes à l'hôpital : l'ordre des médecins saisit la CNIL, l'AP-HP et l'ARS se défendent. Source : https://www.sfm.org/fr/actualites/actualites-de-l-urgences/identification-des-gilets-jaunes-a-l-hopital-l-ordre-des-medecins-saisit-la-cnil-l-ap-hp-et-l-ars-se-defendent/new_id/61713 consulté le 01 juin 2019.

Les progrès récents des sciences cognitives, des neurosciences, de l'informatique conduisent à développer des systèmes d'information autonomes (SIA)³¹ dotés d'une forme de pensée et de capacités de décision sans intervention humaine. Ces SIA naturellement connectés à Internet et au *big data* disposent d'un accès potentiel à l'ensemble des données de l'humanité et en particulier aux données personnelles³² de chacun d'entre nous. Représentent-ils une chance ou une menace pour l'humanité, en particulier dans le domaine de la santé ?

La pratique médicale, que ce soit dans le domaine de la recherche ou dans la prise en charge des individus, constitue un long processus de gestion avec une part importante sur le plan humain et une confiance en son médecin. La numérisation ou du moins la pratique médicale par les données ne comporte-t-elle pas des risques avec des conséquences que cela pourrait engendrer quand nous savons que le parcours de soins ou de vie des individus n'est pas commensurable.

1.2 Problématique et hypothèses de recherche

1.2.1 *Problématique des traces numériques et de la privacy dans le cadre du big data et de l'intelligence artificielle.*

La visibilité du *big data* est le résultat de l'application de modèles, de choix stratégiques et de tactiques opératoires spécifiques au web qui se doivent d'être interrogés. Mais cette visibilité interroge également le web invisible, c'est-à-dire celui qui opère à travers les algorithmes³³ lors de requêtes sur les moteurs de recherche, et à travers les outils qui autorisent les traitements qualitatifs et quantitatifs des données. Le traitement se fait à partir

³¹ « Le système d'information autonome est un modèle de système de décision dont les opérations consistent à coordonner de nombreuses décisions d'actions que le système doit prendre à chaque instant *t*, en tenant compte des informations endogènes (sa propre activité et régulation) et exogènes : les informations provenant de l'environnement (entrées) et ses actions décidées pour tenter d'agir sur l'environnement (sorties) » (LE MOIGNE, 1999).

³² « [...] Constitue une donnée à caractère personnel toute information relative à une personne physique identifiée ou qui peut être identifiée, directement ou indirectement, par référence à un numéro d'identification ou à un ou plusieurs éléments qui lui sont propres. Pour déterminer si une personne est identifiable, il convient de considérer l'ensemble des moyens en vue de permettre son identification dont dispose ou auxquels peut avoir accès le responsable du traitement ou toute autre personne » Article 2 de la Loi Informatique et libertés.

³³ Le Petit Robert 2014 définit le mot « algorithme » comme « suite finie, séquentielle de règles que l'on applique à un nombre fini de données, permettant de résoudre des classes de problèmes semblables » ; « ensemble des règles opératoires propres à un calcul ou à un traitement informatique », « Calcul, enchaînement des actions nécessaires à l'accomplissement d'une tâche ».

des données stockées dans les bases de données mais également à partir des traces visibles ou invisibles laissées par les individus sur le web de façon volontaire ou non. De plus, toute activité numérique génère des traces numériques. Cette dimension est également un enjeu majeur du *big data*. Chacune de ces dimensions implique des choix éthiques d'autant plus grands qu'ils touchent la santé des individus (GALINON-MÉLÉNEC, 2011), mais également l'interprétation³⁴ de ces traces hors de leurs contextes de production.

Dans quelle mesure la visibilité relève-t-elle de la seule volonté du sujet ? Quel équilibre peut-on établir entre le besoin de connaissance des usagers et la protection des intérêts des personnes dont les faits et gestes deviennent de plus en plus transparents ? Il convient d'évaluer la force juridique des instruments de gouvernance d'Internet et leur application au *big data*. Il y a aussi lieu d'examiner l'adéquation des dispositifs de droit international privé, plus précisément les droits et obligations des gestionnaires de données massives, dont les principaux sont Google, Apple, Facebook, Amazon et Microsoft (GAFAM), Netflix, Air BNB, Tesla et Uber (NATU) pour les Américains et Baidu, Alibaba, Tencent et Xiaomi (BATX) pour les chinois³⁵.

Le concept de « *ethics by design* »³⁶ est en réflexion afin d'éviter des problèmes éthiques et juridiques que nous rencontrons aujourd'hui avec le traitement des données ; la résolution des problèmes se fait déjà à la conception des applications et des systèmes. Ce concept pose des questions sur la décision des personnes de partager ou non les données et le paramétrage des accès. Mais toutes ces propositions restent insuffisantes face aux enjeux liés à l'utilisation des données et leurs conséquences sur la vie des individus, la société, voire l'Etat.

Le patient est un individu, une personne vivant dans une famille, un environnement ayant une culture et des valeurs. L'activité de soins s'intègre dans une interaction entre le médecin et son patient (processus de communication et d'information, prise en compte du contexte et

³⁴ Sur la question de l'interprétation des traces Béatrice Galinon-Méléneec dit : « *toute trace est polysémique et est interprétée différemment selon les contextes et les enjeux qu'elle représente dans les systèmes (politiques, économiques, médiatiques, symboliques, etc.) où elle s'insère* » in (GALINON-MÉLÉNEC, 2011).

³⁵ L'*Executive Order* signé par Donald Trump le 25 janvier 2017 permet aux agences gouvernementales américaines de disposer des données personnelles des non américains et non-résidents américains présent sur le territoire américain. Cette nouvelle loi met à rude épreuve le *privacy shield* qui a pour objectif la protection des données de l'UE transférées vers les Etats Unis.

³⁶ « *L'ethics by design consiste à prendre en considération le plus en amont possible de la conception, et dans le suivi d'une innovation technologique, les questions éthiques et les questions d'acceptabilité sociales* » (DIGNUM ET AL., 2018).

du système relationnel). La dynamique de l'interaction repose sur des négociations et des transactions à partir d'un contrat qui est une convention implicite entre les deux personnes (BAGROS ET AL., 2012). D'après Shea « *l'entretien est un dialogue à la fois verbal et non verbal entre deux individus dont les comportements influencent leurs styles de communication respectifs donnant ainsi lieu à des schémas spécifiques d'interaction* » (BAGROS ET AL., 2012 ; GALINON-MÉLÉNEC, 2011) ; leur mode d'interaction pose la question de la commensurabilité des parcours de soins des individus. Le développement des outils (les bases de données, les objets connectés, les *data center*, la bio-informatique, le web sémantique) sur lesquels s'appuie le *big data* pose des questions épistémologiques aux informaticiens, aux médecins et aux personnels de santé en tant qu'utilisateurs (COLLOC, HÉNOCQUE, 2016). Les données constituent une marchandise précieuse commercialisée et l'usage du *big data* constitue une nécessité pour la survie économique des entreprises devenues des « *carnivores des données* »³⁷.

« *L'homme porte les traces de son histoire et de l'environnement dans lequel il s'est construit et ses comportements, ses actions, ses choix et ses jugements laissent des traces dans son environnement* » (GALINON-MÉLÉNEC, 2011). Cela est d'autant plus vrai dans le monde du numérique où toutes nos navigations sont profilées, les appareils que nous utilisons laissent des traces numériques³⁸ exploitables par les fournisseurs de services. Les puissances et l'amélioration des algorithmes ont facilité la sélection et la collecte de ces traces. Cette accélération technologique fait croître le volume de données disponibles de manière exponentielle. Une aubaine pour la recherche en santé pour qui le *big data* est une source presque inépuisable de nouvelles connaissances indispensables à l'innovation et aux progrès médicaux. En effet, l'exploitation des données de qualité donne de bonnes informations³⁹ qui dans leurs exploitations aboutissent à une connaissance. Mais,

« la trace se repère et s'interprète en fonction des processus d'interactions entre un individu en situation d'interprétation et l'environnement (individuels,

³⁷ Terme utilisé par Béatrice Galinon-Méléneq, Professeur des Universités en SIC, pour expliquer l'avidité des grands groupes du web à dévorer la vie privée des individus à travers les données collectées.

³⁸ Selon Alain Mille « *La trace numérique est constituée à partir d'empreintes numériques laissées volontairement (ou non ?) dans l'environnement informatique à l'occasion de processus informatiques* » cité par (GALINON-MÉLÉNEC, 2015).

³⁹ L'information consiste en des données interprétées, porteuse de sens. Elle est contenue dans des descriptions, des réponses à des questions comme qui, quoi, quand et combien. (MONINO, SEDKAOUI, 2016). En SIC, nous avons retenu la définition de Jacques Mélèse: « *est information pour un être vivant (ou pour un automate) tout signal qui, après perception, produit un effet sur son comportement ou sur son état cognitif (par exemple en modifiant la représentation qu'il se fait d'un phénomène* » cité par (LELEU-MERVIEL, 1997).

familiaux, culturels, sociaux et plus généralement humains et non humains) dans lesquels ils s'insèrent » (GALINON-MÉLÉNEC, 2011).

L'exercice de la médecine nécessite un développement de l'esprit critique. Tant que le symptôme n'est pas interprété en tant que signe indexé dans les taxinomies médicales en vigueur - afin de l'intégrer dans un syndrome de maladie pour guider les examens complémentaires et qu'il reste encore à diagnostiquer définitivement la maladie - ; le coût de la santé pour l'assurance maladie ne va faire que croître. Il est nécessaire de mettre l'accent sur la formation sémiologique des étudiants en médecine avec un raisonnement fondé sur la science et non sur la peur de rater le diagnostic (BAGROS ET AL., 2012). Le soignant doit repérer la signification du symptôme pour le sujet lui-même. Cette signification varie d'un sujet à l'autre en fonction de son histoire, son environnement, son identité, son fonctionnement psychologique, de ses valeurs de références et de ses représentations (BAGROS ET AL., 2012 ; GALINON-MÉLÉNEC, 2011).

« Qu'il vise à rétablir la nature ou à la modifier afin d'améliorer la vie des malades, l'acte médical est toujours un acte invasif, y compris dans le silence, la confidentialité et l'intimité de la rencontre médecin-malade. En conséquence, penser la médecine, c'est d'abord penser la violence » (DUPONT, 2001) ;

amenant à se poser des questions sur l'utilisation du numérique dans la pratique médicale.

L'implication et l'utilisation des TIC dans la pratique de la médecine comportent de nombreux enjeux (économiques, techniques, sociaux) et conduisent à poser des questions sur l'évolution de la médecine numérique (médecine personnalisée, personnalisante et médecine connectée) d'autant plus que les processus d'innovations technologiques au sein des structures de soins reposent le plus souvent sur des « *logiques technico-économiques que sur des logiques médico-intégratives*⁴⁰ » (BONNEVILLE, GROSJEAN, 2007). La médecine numérique est basée sur les prédictions⁴¹. Les prédictions sont fondées sur les traces numériques ; mais les traces sous forme de calcul délivrent-elles tout leur sens ? Les calculs se font à partir des algorithmes qui font des tris sélectifs de l'information. Tout choix

⁴⁰ « *La logique médico-intégrative est fondée sur un modèle organisationnel de collaboration entre des sujets aux domaines d'expertise complémentaires et permet d'aller à la rencontre de l'« usager réel » et qui ne se limite pas à penser un « usager modèle ».* » (BONNEVILLE, GROSJEAN, 2007).

⁴¹ « *La prédiction consiste à faire l'hypothèse d'un événement futur en se fiant aux données observables du passé ou du présent* » (GAUTHIER, 2009).

signifiant abandon, il va s'en suivre que l'algorithme analyse et présente les données qui lui semblent pertinentes et pas forcément ceux qui sont pertinentes à nos yeux et dans notre vie. Derrière la machine, il y a ce que nous souhaitons par rapport aux sociales (CARDON, 2015) ; et ce choix dans le cadre de l'utilisation des algorithmes est fait par le concepteur lui-même. Les traces sont interprétées en dehors de leur contexte, sans lien avec l'individu ou l'environnement qui les produit. « *L'individu est dépossédé des sens de ses agissements* » (MERZEAU, 2013b). On assiste à une « *déliasion⁴² des traces* » (MERZEAU, 2013b) avec les producteurs de ces traces et une aliénation importante de la part des *GAFAM, NATU et BATX*, qui se sont spécialisés dans l'exploitation de ces traces à travers leurs propres algorithmes pour des intérêts économiques. Les algorithmes travaillent en fonction des objectifs fournis par son concepteur. Chaque firme dispose de ses propres algorithmes gardés secrets et elle produit des calculs qu'elle seule peut comprendre et reproduire. Les vérités statistiques sont devenues instrumentales : « *ce n'est plus la valeur propre du chiffre qui importe, mais l'évolution de la valeur mesurée entre deux enregistrements* » (CARDON, 2015). Selon la loi de Goodhart (STRATHERN, 1997) « *Dès qu'une mesure devient un objectif, elle cesse d'être une bonne mesure* ».

Au-delà de la gestion des traces numériques, la gouvernance des algorithmes⁴³ en elle-même pose problème. Sa gestion opaque et invisible pose des questions sur la transparence et l'éthique des calculs qu'elle mène. La seule justification de cette opacité est le secret de la recherche et des affaires dans les différentes firmes. Pourtant, les décisions algorithmiques sont utilisées au nom de la société (même s'il s'agit d'une réduction voire une minimisation de l'existence de l'individu) et de la bonne gestion (même si les décisions sont fondées sur des corrélations sans possibilités de relation de cause à effet) mais surtout au profit des acteurs qui n'ont que des intérêts économiques. Nous sommes dans le cas où les algorithmes ont un réel pouvoir, ou plus exactement, dans le cas d'individus, plus ou moins invisibles qui exercent leur pouvoir à travers les algorithmes. Ceux qui proposent les algorithmes sont ceux qui les utilisent ou les font utiliser pour la prise de décision.

⁴² « *La déliaison des traces numériques débute dès leur production, elles sont agrégées à d'autres données, stockées dans des réservoirs distants, disséminées dans quantité de sites et font l'objet de calculs qui produisent à leur tour d'autres données* » (MERZEAU, 2013b).

⁴³ « *La gouvernamentalité algorithmique peut être définie comme l'extraction automatisée d'informations pertinentes depuis des bases de données massives à des fins de prévision ou d'exclusion (consommation, risque, définitions de clientèles nouvelles etc.* » (ROUVROY, BERNIS, 2013).

1.2.2 *Place de l'éthique et questions épistémologiques*

La multiplicité des intervenants dans la prise en charge des patients avec des responsabilités différentes ou souvent mal définies pose un problème sur le colloque singulier médecin – patient (utilisation d'objet connecté, *cloud computing*, hébergeurs de données) et sur le respect du serment d'Hippocrate. Le *big data* est en train de modifier de manière significative la pratique médicale et la production des connaissances dans le domaine de la santé (secret médical partagé, DMP, médecine connectée, médecine personnalisée, prise de décision basée sur les données) qui vont modifier d'une façon profonde les savoirs et la pratique médicale.

Les questions épistémologiques soulevées par le *big data* sont la définition de la donnée et sa collecte. Une donnée se construit à partir de référentiels ou de paramètres acceptés de tous. Le traitement des données pose un problème car toutes les données ne sont pas disponibles sur Internet ou dans des bases de données et ne sont pas toutes accessibles. La qualité des données est un facteur important dans la prise de décision notamment dans le cadre de l'apprentissage des algorithmes et des corrélations ; d'où la notion de « *garbage in, garbage out* »⁴⁴. Les données ne sont pas équivalentes ; quantité importante de données ne signifient pas une meilleure qualité, ni une prise en compte de tous les facteurs associés et encore moins une exhaustivité de la population étudiée. Cela pose la question de la représentativité des données et des informations que l'on peut tirer de ceux-ci. En effet, la collecte des données se fait selon un cadre conceptuel qui permet son interprétation et sa compréhension ; sorties de leur cadre, les données collectées posent un problème de validité et d'inférence à l'ensemble de la population.

Le *big data* ne vient pas du néant sans une régulation philosophique ; « *les algorithmes sont à la base construits à partir d'un travail de théories et de modèles en arrière-plan* » (KITCHIN, 2014) après des longues périodes de réflexion. Dans le cadre de la pratique médicale, nous passons d'une médecine fondée sur les preuves à une médecine fondée sur les données, avec une dépersonnalisation des données et une personnalisation de la décision et une interprétation réductionniste et fonctionnelle tout en oubliant l'effet de la culture, de

⁴⁴ Terme utilisé en informatique pour expliquer le fait que le système fournit des résultats de mauvaise qualité si on le fournit des données de mauvaises qualités.

la politique et de la gouvernance. Au niveau de la recherche, le *big data* se passe des méthodes de recherche reposant sur des hypothèses, mais fait parler les données toutes seules pour faire des prédictions par l'intermédiaire des calculs automatiques. Faut-il prendre les décisions sur la base de corrélation ou sur la base de relation de cause à effet ? Cette question mérite une réflexion car elle pose la question de la responsabilité et de l'éthique dans un domaine où l'humain, l'autre, est le fondement de la pratique médicale. Penser que l'utilisation du *big data* et de l'IA est la solution pour la gestion des problèmes de santé et devient la source de production du savoir tout en écartant les autres méthodes et sources de connaissances constitue une limite dans la réflexion sur la résolution des problèmes.

L'utilisation de l'informatique et des systèmes d'information dans le domaine de la santé génère des droits et des devoirs et une responsabilité liée au secret médical, qui comporte aujourd'hui de nombreuses exceptions. Tout ceux-ci soulèvent la place de l'éthique⁴⁵. De nos jours, « *l'éthique moderne tend à devenir une sorte de régulation vague de ce qui se passe et de ce que nous faisons* » (LOMBARD, 2009). L'omniprésence du *big data* dans notre vécu quotidien est à l'origine de nombreux problèmes sociaux, politiques et éthiques ; problèmes aggravés par le mode de développement et des utilisations incontrôlées entraînant un emballement global dans toutes les sphères de la société. Il y a une nécessité de mener une réflexion sur les problèmes dans leur globalité afin de mieux situer les responsabilités des acteurs et la prise en compte des individualités ; comme le dit Edgar Morin :

« Toute connaissance (et conscience) qui ne peut concevoir l'individualité, la subjectivité, qui ne peut inclure l'observateur dans son observation, est infirme pour penser tous problèmes, surtout les problèmes éthiques. Elle peut être efficace pour la domination des objets matériels, le contrôle des énergies et les manipulations sur le vivant. Mais elle est devenue myope pour appréhender les réalités humaines et elle devient une menace pour l'avenir » (MORIN, 2004).

Le *big data* peut contribuer d'une manière importante dans le développement de la santé en termes d'informations et de connaissances, de la qualité de prise en charge des patients et de

⁴⁵ L'éthique (du grec *ethos* « caractère, coutume, mœurs ») est une discipline philosophique portant sur les jugements de valeur. L'éthique se définit telle une réflexion fondamentale sur laquelle la morale établira ses normes, ses limites et ses devoirs. L'éthique regroupe un ensemble de règles qui se différencient et complètent les règles juridiques. L'éthique vise à répondre à la question « Comment agir au mieux ? ». Source : <https://comitedethique.wordpress.com/about/definition-ethique/> consulté le 2 novembre 2017

Dans le domaine de la santé, l'éthique médicale se base sur quatre grands principes universels : l'autonomie, la bienfaisance, la non malfaisance et la justice. (LE COZ, 2009).

Davis et Patterson ont défini le cadre éthique des données numériques à partir de quatre éléments pour les particuliers et les organisations : l'identité, la confidentialité, la propriété et la réputation. (DAVIS, PATTERSON, 2012).

la santé publique. Mais toutes ces contributions, pour être bénéfiques (pour les chercheurs, les citoyens, les politiques et l'économie), doivent prendre en compte l'éthique dans tout son ensemble sans ignorer l'écologie d'action⁴⁶ (MORIN, 2004).

Nous sommes en face d'une révolution scientifique qui va nécessiter de faire des choix, c'est-à-dire de prendre des décisions. Ces décisions s'imposent notamment dans le contexte de la médecine, où les données de santé relèvent de l'intimité des personnes, la volonté de l'Etat de réduire les dépenses de santé, la prolifération des modèles économiques reposant sur les données et les pratiques médicales. Ces décisions devront tenir compte de l'éthique médicale et des droits fondamentaux. Penser l'éthique, c'est un questionnement au quotidien auquel il faut apporter des réponses. Ce questionnement se fait en tenant compte du contexte et de l'évolution de la société. Et il est d'actualité dans notre contexte actuel avec le développement des technologies du numérique et surtout l'insertion du *big data* et de l'IA dans la vie des personnes et surtout dans la pratique médicale. Cette thèse se propose d'étudier les problèmes éthiques liés à l'utilisation du *big data* et de l'IA dans le domaine de la santé. Il s'agit principalement de déterminer quelle éthique dans l'utilisation du *big data* et de l'IA dans le domaine de la santé ?

1.2.3 *Complexité des questions posées*

Le domaine de la santé est un milieu complexe du fait qu'elle touche à l'intimité des personnes, de son évolution rapide dans le temps et dans l'espace (les connaissances de la santé évoluent de façon rapide avec les découvertes de nouvelles molécules, de nouvelles maladies et de pratiques différentes en tenant compte des contextes) et surtout de par sa relation avec plusieurs domaines notamment le droit, la sociologie, la philosophie, les sciences de l'information et de la communication et l'informatique. Cette complexité réside même dans la difficulté à trouver une définition pour la « *maladie* ». Toutes les définitions proposées partent à partir de son contraire à savoir la santé. Maël Lemoine le résume bien en disant que la définition de « *la maladie*⁴⁷ est fonction de celui qui la définit » (LEMOINE,

⁴⁶ « *L'écologie d'action enseigne que toute action humaine dès qu'elle est entreprise, échappe à son initiateur et entre dans un jeu d'interactions multiples qui la détourne de son but et parfois lui donne une destination contraire à son intention* » (MORIN, 2004).

⁴⁷ « *[...] est « malade », pour le médecin, celui qui doit être traité, pour l'assureur, celui qu'il doit rembourser, pour la société, celui à qui on doit indulgence et secours, et pour les malades eux-mêmes, celui qui est entravé dans son action habituelle et ne peut faire ce qu'il veut ou doit faire* » (LEMOINE, 2017).

2017). Cette définition pose le problème de l'usage du *big data* pour le diagnostic des maladies et leur prise en charge à travers les algorithmes. Sommes-nous en train d'évoluer vers une définition de la maladie et des soins à travers les données ou plutôt vers une santé productrice des données ?

Bien que les données physiologiques et les examens complémentaires soient importants pour le diagnostic de la maladie, ils ne sauraient dispenser du processus communicationnel dans l'élaboration du diagnostic et la prise en charge du patient. La production des données à travers les objets connectés ces dernières années a connu une croissance importante. En dehors de la production et du traitement des données, il se pose la question de la fiabilité des données produites par ces objets connectés surtout dans le cas de la prise en compte dans le diagnostic médical et la prise de décision. En effet, des études ont montré que les applications des smartphones et les objets connectés ne produisent pas toujours des données fiables et ceci dans des proportions inacceptables pour une décision médicale (CASE ET AL., 2015 ; FARINA, LOWRY, 2017 ; ORR ET AL., 2015 ; FARNIA ET AL., 2017). Il se pose alors le problème de la réglementation de ces outils de production de données et les questions juridiques liées à ces données. Pour le moment, il n'existe pas une réglementation sur la conception et le fonctionnement (fiabilité médicale et sécurité des données) des objets connectés en dehors de ceux qui sont validés comme étant des dispositifs médicaux⁴⁸. Par conséquent, les objets connectés ne peuvent pas être considérés comme des dispositifs médicaux entrant dans le diagnostic ou la surveillance des maladies. Mais quelle est la responsabilité du médecin, s'il ne prend pas en compte les résultats d'un objet connecté et que cela a des conséquences néfastes pour le patient ou si au contraire, il est trompé par une valeur erronée fournie par l'objet connecté ? La question de la sécurité est plus prégnante dans le domaine de la santé en ce sens qu'elle nécessite une forte disponibilité des données tout en assurant leur intégrité et leur confidentialité. Une adaptation des textes juridiques voire un nouveau droit⁴⁹ concernant les objets connectés, l'IA et le traitement des données de santé s'avèrent nécessaires. Les technologies du numérique évoluent à une vitesse impressionnante avec d'importants enjeux économiques et sociétaux. Il faudra trouver un

⁴⁸ L'Agence nationale de sécurité du médicament et des produits de santé (ANSM) est chargée de l'évaluation et de l'expertise des dispositifs médicaux (applications, objets de diagnostics et de surveillance) en France. Source : www.ansm.sante.fr consulté le 02 novembre 2017.

⁴⁹ Le roi Salmane a donné la nationalité Saoudienne à Sophia (robot humanoïde). Il va se poser le problème des droits de ce robot envers elle-même et envers la société. Cette naturalisation pose en même temps un problème éthique auquel nous devons répondre. Source : http://www.lepoint.fr/monde/l-arabie-saoudite-honore-une-femme-robot-30-10-2017-2168688_24.php consulté le 02 novembre 2017.

juste milieu pour la protection de la vie privée des personnes et la confidentialité des données tout en adaptant les textes juridiques et en prenant en compte les enjeux économiques et politiques dans le domaine de la santé.

À côté des enjeux juridiques, sécuritaires et éthiques, nous devons évaluer les risques de l'utilisation des données de mauvaises qualités dans la production de nouveaux savoirs et connaissances dans la pratique médicale et la recherche dans le domaine de la santé. Les conséquences économiques et sociétales ne seront-elles pas supérieures aux effets bénéfiques tant plébiscités des technologies du *big data* ?

« Pratiquer la médecine, c'est entrer dans la vie d'un être humain, dans la mort d'un cadavre. C'est l'interroger, le découvrir de mille et une manières, l'ouvrir, le modifier, le réparer, l'exhiber et l'exhumer contre son gré. Même le geste salvateur de quelque chirurgien greffant une main n'est rien à côté des douleurs physiques et des souffrances psychologiques que la personne redevenue normale (c'est-à-dire avec deux mains) endure » (DUPONT, 2001).

Tout manquement ou erreur est un lourd tribut à payer pour l'individu et pour la société (par exemple si nous prenons le cas des piratages des systèmes d'information des hôpitaux avec le blocage des accès des données patients et la récupération de ces données par les pirates). Dans ce domaine complexe, nous examinons les questions suivantes :

- quels sont les impacts de l'utilisation du *big data* et de l'IA dans la pratique médicale ?
- les moyens juridiques de protection de données sont-ils suffisants et efficaces ?
- quels sont les enjeux de l'utilisation des traces numériques dans le domaine de la santé ?

1.2.4 Intérêt de notre sujet de thèse

Nous avons entrepris ce travail de recherche en fonction de la prégnance des enjeux pour la société et la pratique médicale.

Sur le plan scientifique, il s'agit d'un travail interdisciplinaire (médecine, SIC, droit, sociologie et informatique). La thèse concerne, d'une part, le traitement des données de santé

à l'ère des TIC - notamment le *big data* - et de l'IA, son implication dans la vie sociale des individus et la pratique médicale, et, d'autre part elle propose une réflexion sur les aspects éthiques liés à l'utilisation du numérique.

Conjointement à la conception d'une application concrète de l'IA en médecine, destinée à la prise en charge des patients suspectés d'une embolie pulmonaire (YAMEOGO, COLLOC, 2016), nous avons montré l'applicabilité du paradigme de *l'Homme-trace* (GALINON-MÉLÉNEC, 2011) dans le domaine de la santé et l'évaluation de l'utilisation des TIC dans la pratique médicale par les médecins.

Avec le numérique et l'utilisation du *big data*, les frontières sont abolies. Il est possible de retrouver ses données dans n'importe quel pays à tout moment. Ce point mérite des conditions techniques et humaines de protection des données mais surtout une adaptation des textes juridiques pour que tous les acteurs tirent le meilleur profit du *big data* et de l'IA sans mettre en danger la vie privée et personnelle des individus et la pratique médicale. Pour une meilleure utilisation du *big data* et de l'IA dans notre société, la mise en place d'une stratégie globale et d'une pensée complexe s'avère nécessaire et urgente. La question de l'éthique (protection des données, la perte de confiance en l'administration publique, la monopolisation des données par les géants du web (GAFAM, NATU, BATX), les méthodes de production des informations et des connaissances, le terrorisme étatique⁵⁰ en vigueur dans certains Etats) dans l'utilisation de ces techniques dans notre société est plus que d'actualité de nos jours.

1.2.5 Objectifs de notre thèse

Les objectifs de notre thèse étaient les suivantes :

- établir un état des lieux :
 - des différents modes d'utilisation du *big data* et de l'IA dans la pratique médicale ;
 - des moyens juridiques de protection des données de santé des individus ;
- analyser l'impact de l'utilisation des traces numériques dans la pratique médicale ;
- évaluer les enjeux de l'utilisation du *big data* et de l'IA dans la pratique médicale.

⁵⁰ Collecte des données personnelles sans respect des règles juridiques et sociales et surveillance systématique des individus par l'état sans l'autorisation des personnes au nom de la sécurité nationale.

Ces repères nous ont permis d'évaluer l'hypothèse principale : L'utilisation du *big data* et de l'IA dans la pratique médicale comporteraient des risques pour l'éthique médicale.

De cette hypothèse principale découlaient les hypothèses secondaires que sont :

- l'utilisation du *big data* dans la pratique médicale porterait atteintes à la vie privée du patient et au secret médical ;
- les moyens juridiques actuels ne garantiraient pas suffisamment la protection des données de santé face à l'évolution rapide des technologies du *big data* et de l'IA ;
- l'utilisation des traces numériques dans le domaine de la santé aurait des conséquences néfastes à la pratique de la médecine et pour la société.

Ainsi, pour atteindre les objectifs de notre thèse et confirmer ou infirmer nos hypothèses de recherche, nous nous sommes efforcés de suivre une démarche rigoureuse décrite dans la partie suivante.

2 Méthodologique de la recherche

« La méthodologie de la recherche est un ensemble de méthodes⁵¹ et de techniques qui orientent l'élaboration d'une recherche et guident la démarche scientifique » (DÉPELTEAU, 1998). La méthode est un ensemble de démarches raisonnées que suit l'esprit pour découvrir et démontrer la vérité dans les sciences et la technique, un outil de recherche qui facilite la collecte des données ou un ensemble de procédés employés pour produire une œuvre ou obtenir un résultat déterminé (MUKE ZIHISIRE, 2011). Pour mener à bien notre étude, nous avons retenu des méthodes et des techniques adaptées à nos objectifs.

2.1 Méthodes

2.1.1 Phase exploratoire

Dans le but de mieux explorer et analyser la problématique de notre thèse, nous avons fait un état de l'art et mis en place une stratégie pour étudier le sujet sous un nouvel angle. Nous avons réalisé une recherche documentaire, participé à des séminaires, colloques, ateliers et rencontré des personnes ressources.

La phase de recherche exploratoire s'est faite sur le web (bases de données spécialisées ou non) et dans les bibliothèques. Ce choix s'expliquait par les raisons suivantes :

- la mise à jour régulière de nos connaissances sur le sujet dans un domaine où la recherche et les connaissances évoluent à une vitesse exponentielle ;
- l'obtention d'un point de vue des personnes extérieures à notre travail ;
- l'analyse sous des angles différents et par différents chercheurs dans le domaine ;
- l'analyse de l'évolution des recherches dans le domaine afin de dégager des tendances de perspectives.

Nous avons ainsi mis en place une veille documentaire sur les bases de données afin de recevoir des rappels réguliers sur les nouvelles mises en ligne et les publications dans le domaine.

⁵¹ Le mot « méthode » est un emprunt du mot latin *methodus* qui est, à son tour, emprunté du mot grec *methodos* qui signifie « route, voie », « direction qui mène au but ».

Des recherches sur le web ont été effectuées sur :

- *Google* pour la recherche généraliste et *Google Scholar* pour la recherche des travaux scientifiques et universitaires (articles avec ou sans comité de lecture, thèses universitaires, livres scientifiques). Le choix de *Google Scholar* nous a paru judicieux car elle intègre un nombre important de revues scientifiques dans sa base de données. En 2014, *Google Scholar* prétendait couvrir 85 % des publications scientifiques. Elle a une interaction avec les bibliothèques et permet de rechercher un ouvrage grâce au catalogue international *WorldCat*, et en France via le système *SUDOC*. Mais il faut noter que *Google Scholar* comporte de nombreux bruits⁵² et la qualité des recherches n'atteint pas celle des bases de données spécialisées et des bibliothèques. En plus la recherche avec *Google* et *Google Scholar* nécessite un travail fastidieux de tri et de choix des articles qui consomment un temps important pour la revue de la littérature ;
- *Cairn* qui est une plateforme de diffusion de publications francophones en sciences humaines et sociales. Cairn propose l'accès au texte intégral de plus de 300 revues. Une des difficultés de la recherche sur cette plateforme est que les requêtes se vont en texte libre ou *full text* avec un dépouillement qui prend plus de temps ;
- *Medline / Pubmed* est l'un des principaux moteurs de recherche de données bibliographiques de l'ensemble des domaines de spécialisation de la biologie et de la médecine. Elle a une meilleure qualité de recherche à partir des mots-clés *MeSH*⁵³. La qualité des requêtes permet d'avoir une bonne précision sur la recherche et un gain de temps importants. Nous avons utilisé le site web www.hetop.eu⁵⁴ pour la correspondance des mots-clés du français à l'anglais.

⁵² Pour les SIC, « le bruit documentaire est un ensemble de documents qui souffre d'un déficit de pertinence, en résultats d'une recherche d'information sur un outil de recherche ». Source : Bruit documentaire sur le site de l'APDEN http://apden.org/wikinotions/index.php?title=Bruit_documentaire consulté le 24 avril 2018.

⁵³ *MeSH (Medical Subject Headings)* est le thésaurus de référence dans le domaine biomédical. La NLM (*U.S. National Library of Medicine*), qui l'a construit et le met à jour chaque année, l'utilise pour indexer et permettre d'interroger ses bases de données, notamment *Medline / PubMed*.

⁵⁴ *HeTOP (Health Terminology/Ontology Portal)* inclut les principales terminologies et ontologies de santé. Il contient plus de 2 millions de concepts disponibles dans plusieurs langues à travers plus de 70 terminologies ou ontologies. Le site web a été réalisé par le Département d'Informatique et d'Information Médicales du CHU Hôpitaux de Rouen.

- *Embase*⁵⁵ qui est une base de données biomédicale actualisée et polyvalente. Elle contient la plus grande collection de documents biomédicaux de 1947 à nos jours et tous les articles sont indexés en profondeur à l'aide de *Life Science thesaurus Embase Indexing* d'Elsevier® ainsi que *d'Emtree*®.

La recherche en bibliothèque a été faite dans les bibliothèques de l'Université du Havre et les autres bibliothèques françaises par l'intermédiaire du prêt interbibliothèques.

Les mots-clés utilisés au cours de nos différentes recherches étaient « *big data* », « *objets connectés* », « *intelligence artificielle* », « *machine learning* », « *deep learning* », « *fouilles des données* », « *algorithme* », « *traces* », « *traces numériques* », « *données* », « *système d'information autonome* », « *épistémologie* », « *éthique* », « *vie privée* », « *données santé* », « *données personnelles* ».

Les recherches ont été faites à la fois en anglais et en français. A ces mots-clés, il faut ajouter l'usage des qualificatifs en fonction du raffinement voulu et des combinaisons avec les opérateurs booléens « *AND et OR* » en particuliers. Ces différents mots-clés étaient le plus souvent utilisés en combinaison pour avoir un meilleur rendu des résultats de recherche.

Une recherche particulière a été faite sur les documents juridiques disponibles pour la protection des données à caractère personnel, la pratique de la médecine en France et la gestion des données. Les sites web de LégiFrance⁵⁶ et des lois de l'Union Européenne⁵⁷ ont été utilisés à cet effet.

Les documents juridiques que nous avons exploités au cours de cette analyse sont :

- la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés⁵⁸ ;

⁵⁵ La base de données *Embase* est payante et seule l'université de Paris Descartes disposait d'un accès en France. Nous avons eu accès à cette base de données lors d'une formation organisée par la Cochrane France le 16 mars 2019 ; formation à laquelle nous avons participé.

⁵⁶ <https://www.legifrance.gouv.fr/>.

⁵⁷ https://europa.eu/european-union/law_fr.

⁵⁸ Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés.

- le règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016⁵⁹ relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données ;
- le *privacy shield*⁶⁰ qui encadre les transferts des données personnelles de l'Union Européenne vers les États-Unis ;
- la loi n° 2002-303 du 4 mars 2002 relative aux droits des malades et à la qualité du système de santé⁶¹ ;
- la loi n° 2016-1321 du 7 octobre 2016 pour une République numérique⁶² ;
- la loi n° 2016-41 du 26 janvier 2016 de modernisation de notre système de santé⁶³ ;
- le décret n° 2016-1871 du 26 décembre 2016 relatif au traitement de données à caractère personnel dénommé « Système National des Données de Santé⁶⁴ » ;
- l'ordonnance n° 2017-27 du 12 janvier 2017 relative à l'hébergement de données de santé à caractère personnel⁶⁵ ;
- le décret n° 2018-137 du 26 février 2018 relatif à l'hébergement de données de santé à caractère personnel⁶⁶ ;
- le code de déontologie médicale français⁶⁷ ;

L'analyse de ces différents textes juridiques nous a permis de comprendre les différents processus mis en œuvre pour exploiter et protéger les données de santé considérées comme des données sensibles. Elle a permis de dégager les différentes thématiques abordées dans

⁵⁹Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données.

⁶⁰ *The EU-U.S. Privacy Shield*. In : European Union Law [en ligne]. 2016. [Consulté le 25 janvier 2017]. Disponible à l'adresse : http://data.europa.eu/eli/dec_impl/2016/1250/oj.

⁶¹ Loi n° 2002-303 du 4 mars 2002 relative aux droits des malades et à la qualité du système de santé.

⁶² Loi n° 2016-1321 du 7 octobre 2016 pour une république numérique.

⁶³ Loi n° 2016-41 du 26 janvier 2016 de modernisation de notre système de santé.

⁶⁴ Décret n° 2016-1871 du 26 décembre 2016 relatif au traitement de données à caractère personnel dénommé « système national des données de santé ».

⁶⁵ Ordonnance n° 2017-27 du 12 janvier 2017 relative à l'hébergement de données de santé à caractère personnel.

⁶⁶ Décret n° 2018-137 du 26 février 2018 relatif à l'hébergement de données de santé à caractère personnel.

⁶⁷ Articles R4127-1 à R4127-122 du code de santé publique français.

les différents textes pour ce qui concerne la protection, l'accessibilité et le traitement des données de santé.

Les entretiens et les conseils des personnes extérieures nous ont permis de dégager des pistes et d'orienter notre travail.

Cette phase exploratoire nous a permis de :

- mieux comprendre la technologie du *big data* et les différents domaines dans lesquels il est utilisé. Nous avons amélioré notre compréhension des notions de trace et de données aussi bien dans le domaine de l'informatique que dans celui des SIC ;
- préparer notre méthode d'enquête et le choix des outils techniques pour le recueil des données ;
- mettre en place une méthodologie pertinente, rigoureuse et appliquée pour la réalisation de notre travail.

2.1.2 Méthode d'enquête

A l'issue de ces prospections initiales, nous avons choisi une méthode d'enquête qui consiste à interroger les acteurs sur des aspects d'eux-mêmes (expériences, leurs attitudes et leurs opinions) qu'ils sont en mesure de décrire. A partir des résultats obtenus auprès d'un groupe d'individus, les chercheurs tirent des conclusions sur les caractéristiques de l'ensemble de la population de référence (MUKE ZIHISIRE, 2011).

Les objectifs de cette enquête dans notre travail sur le terrain étaient de :

- décrire l'utilisation des objets connectés et la santé numérique par les patients et les médecins ;
- évaluer la perception de la protection et du respect des règles éthiques dans l'exploitation des données personnelles de santé par les médecins ;
- analyser les craintes et les attentes des médecins quant à l'utilisation du *big data* et de l'IA dans la pratique médicale ;

- évaluer l'impact de l'utilisation du numérique sur la confiance qu'ont les patients envers leur médecin à l'ère du numérique ;
- recueillir les avis d'experts sur l'utilisation du *big data* et de l'IA dans la pratique médicale.

Dans cette recherche, nos populations de référence étaient :

- les médecins praticiens de la commune du Havre qu'ils soient dans un cabinet de soins ou dans un hôpital public de soins ;
- les habitants de la commune du Havre ;
- les personnes ressources dans les domaines du *big data* et de l'IA.

Enquête auprès des médecins et des habitants de la commune du Havre

L'enquête sur les médecins et les habitants du Havre a porté sur leurs caractéristiques sociodémographiques, leurs usages des objets connectés et des applications de santé sur Internet et sur leur téléphone mobile, leur avis sur l'utilisation de l'IA dans le domaine de la santé, l'impact de l'utilisation du *big data* sur la pratique médicale, la protection et la sécurité des données.

Le choix de faire une enquête sur les deux populations reposait sur notre hypothèse de départ. Jusqu'à aujourd'hui, la relation médecin – patient reposait sur un lien fort⁶⁸ (GRANOVETTER, 1983) entre le médecin et le patient durant tout le processus de soins. Ce lien complexe est composé en partie de confiance dans le diagnostic des « *signes-traces* » (GALINON-MÉLÉNEC, 2011) émis par le corps du patient, en partie par l'intimité que crée la coprésence des corps dans la même salle (le cabinet médical), la fréquence de la rencontre médecin – patient et la forme prise par l'interaction (GOFFMAN, 1974) ; l'ensemble de ces composants produisant une relation d'attachement entre les parties prenantes de l'interaction. Cette relation de confiance est associée à une augmentation de la satisfaction, à l'adhérence au traitement et à la continuité des soins (ROLFE ET AL., 2014). L'évolution et les changements dans le système de santé peuvent influencer cette relation et affecter la prise en charge des

⁶⁸ Selon la théorie de lien faible, décrite par Granovetter, il faut distinguer deux types de relations pour un individu : « *celles qui le relient à sa famille et ses amis proches, qui constituent des liens forts, et celles formées par un réseau généralement plus étendu et plus distant, qui constituent les liens faibles* » (GRANOVETTER, 1983).

patients. Ces changements sont surtout marqués par l’informatisation de tout le processus de soins et l’implication de plusieurs acteurs dans la relation de soin, qui était constitué du colloque singulier entre le médecin et le patient. Ils ont abouti à la notion de « *secret partagé*⁶⁹ ».

Notre hypothèse était que l’utilisation du numérique et des objets connectés dans la pratique médicale modifiait la nature de ce lien. Notre enquête visait à préciser également les contours de cette modification.

Pendant notre recherche documentaire, les enquêtes menées dans le cadre de l’utilisation du *big data*, des objets connectés ou de l’IA dans le domaine de la santé ont concerné uniquement l’un des deux groupes à savoir soit le patient ou le médecin mais pas les deux à la fois en tenant compte de la relation médecin - patient (LES ECHOS, 2016 ; LÉO, 2016).

Afin d’éviter les biais dans les réponses :

- concernant les patients, notre procédure d’enquête consistait à faire un choix des habitants de la commune du Havre (patients dans le cadre de notre étude) centré sur le médecin traitant afin de mieux établir une relation médecin - patient et une corrélation dans l’utilisation du *big data* et ses applications dérivées dans la prise en charge médicale ;
- concernant les médecins, le choix a été fait de manière stratifiée et aléatoire à partir de la base de données de l’annuaire de santé de l’assurance maladie, les médecins praticiens inscrits au titre de la ville du Havre dans un premier temps. Ensuite, nous avons dû faire un échantillonnage aléatoire systématique de tous les cabinets médicaux du fait d’une faible participation des cabinets lors de notre première technique d’échantillonnage.

Les résultats devaient permettre :

- dans un premier temps, de mettre en évidence les relations entre les besoins et les usages de chaque participant (médecin et patient). Ce qui sur le plan opérationnel

⁶⁹ La circulaire santé/justice du 21 juin 1996 évoque la possible transmission, entre professionnels de santé, d’informations sur un patient en vue d’une meilleure coordination ou qualité des soins. Elle a été reprise dans loi n° 2002-303 du 4 mars 2002 relative aux droits des malades et à la qualité du système de santé.

devrait aboutir à proposer une approche de l'intégration du *big data* dans le domaine de la santé qui prenne en compte les préoccupations des différents acteurs. La protection de la vie privée des individus et l'éthique médicale ont fait l'objet d'un *punctum* spécifique ;

- dans un second temps, d'analyser l'importance de la contextualisation dans la prise en charge des patients et la pratique de la médecine ;
- et enfin, une analyse des risques perçus et vécus par les médecins dans le cadre de l'utilisation du numérique dans la pratique médicale.

Pour l'obtention des résultats, le traitement de l'enquête alliait une approche mixte, qualitative et quantitative, que nous considérons comme complémentaires pour repérer les paramètres de la nature du lien décrite plus haut.

Personnes ressources dans les domaines connexes

Le corpus pluridisciplinaire de ces personnes se composait de :

- juristes dont nous avons souhaité connaître la vision sur les textes de la protection des données de santé et l'implication juridique de l'utilisation du *big data* et de l'IA dans la pratique médicale, en particulier sur la question de la responsabilité médicale ;
- responsables du département d'information médicale qui nous ont informé des pratiques sur le partage des informations entre professionnels et la confidentialité des données, l'accès des données aux assurances maladies dans le cadre des remboursements, le transfert des données vers la PMSI (programme de médicalisation des systèmes d'information) et la sécurité des données ;
- informaticiens qui sont chargés d'assurer la sécurité, l'intégrité, l'accessibilité et la confidentialité des données, et de centraliser les données de santé dans un seul serveur ;
- médecins spécialistes de l'éthique qui ont donné leur avis sur la numérisation de la pratique médicale ;
- spécialistes en SIC dont les recherches portaient sur l'usage des traces numériques et leurs implications dans le domaine de la santé.

Approche qualitative

« *L'approche qualitative s'intègre dans une démarche compréhensive qui privilégie la connaissance intime d'un phénomène, l'explication des cas et la richesse des données. Elle induit un rapport différent à la réalité empirique* » (LELEU-MERVIEL, 2008). Elle implique un contact personnel avec les sujets de la recherche, principalement par le biais d'entretiens et par l'observation des pratiques dans les milieux mêmes où évoluent les acteurs (PAILLÉ, MUCCHIELLI, 2016).

« *Une méthode qualitative est un agencement spécifique de techniques de recueil et de techniques de traitement, appropriée à la résolution d'une problématique de recherche dans le but d'explicitier, en compréhension, un phénomène humain ou social* » (MUCCHIELLI, 1996).

Les avantages d'une collecte de données qualitatives bien menée résident précisément dans la richesse des données collectées et la compréhension plus en profondeur du problème étudié. Elles visent non seulement à décrire, mais aussi à aider à obtenir des explications plus significatives sur un phénomène. La recherche qualitative est également utile pour générer des hypothèses (KOHN, CHRISTIAENS, 2014)

Approche quantitative

« *L'approche quantitative permet de décrire (obtenir des informations sur les caractéristiques d'une population cible), de comparer (lorsque les données sont collectées sur deux ou plusieurs groupes), de repérer (certains faits sociaux) ou d'expliquer (élucider les liens entre certaines variables). Elle permet aussi de vérifier et de quantifier de façon précise des hypothèses préalablement formulées* » (LELEU-MERVIEL, 2008).

Mais, les résultats de la recherche quantitative peuvent être trop simples et superficiels. Il n'est pas toujours possible de décrire le problème en profondeur. Le chercheur peut oublier des qualités importantes du phénomène étudié, car il se consacre trop à un problème concret et il ne pense pas à la multi dimensionnalité du problème ; cela pose le problème du réductionnisme que connaissent ces méthodes.

Chacune des deux approches comporte des limites qu'il faut prendre en compte. L'association de ces deux méthodes permet ainsi d'avoir un meilleur résultat et surtout de

compenser les inconvénients de chaque méthode. Ces deux méthodes ne sont pas opposées ou incompatibles, mais plutôt complémentaires. La méthode d'enquête impose de choisir une technique de recueil des données.

2.2 Techniques de recueil de données

« *La technique se situe au niveau des faits ou des étapes pratiques. Chaque technique choisie doit garantir un certain niveau de fidélité, de validité, de précision et de neutralité* » (LELEU-MERVIEL, 2008) dans l'atteinte des objectifs de l'étude.

Dans le cadre de notre travail, nous avons choisi les techniques de questionnaire et de l'entretien.

2.2.1 Technique de questionnaire

Le questionnaire est une technique de recueil de données qui permet de comprendre et d'expliquer les faits. C'est une approche quantitative qui s'applique à un ensemble d'échantillons qui doit permettre des inférences statistiques (DÉPELTEAU, 1998 ; MUKE ZIHISIRE, 2011).

Dans le cadre de notre travail, les individus et les médecins ont été enquêtés par la technique de questionnaire pour permettre de décrire leur usage du numérique dans le domaine de la santé, leur utilisation des objets connectés et leur avis sur l'IA et le *big data*. Cette collecte de données permettra d'expliquer les différents choix des individus et les craintes que peut susciter l'utilisation du *big data* dans le domaine de la santé. Enfin, les données recueillies permettront de faire une corrélation entre les besoins et pratiques des patients et des médecins dans le cadre de l'utilisation du *big data* pour leur prise en charge.

Nous avons utilisé à la fois des questions ouvertes et fermées. Ce choix permettait aux individus participant de l'étude de s'exprimer librement. Les questions fermées permettaient d'avoir des données comparables entre les individus afin de mettre en évidence des relations entre des facteurs expliquant en partie les faits.

Pour les questions fermées, nous avons utilisé une échelle de Likert (LIKERT, 1932). L'échelle de Likert ne recourt pas à des juges pour établir les valeurs à accorder aux différentes modalités des questions. C'est directement les réponses des individus qui sont utilisées, mais les scores sont fixés a priori par l'auteur de l'échelle qui décide également de l'orientation de l'item (DEMEUSE, 2008). Nous avons utilisé des questions à cinq modalités centrées par une modalité neutre.

L'administration du questionnaire s'est faite par Internet par la méthode du système web intégré (STÉPHANE, JEAN, 2004). Le questionnement sur Internet a ouvert de nouveaux horizons aux intervenants du domaine : on efface les distances, on donne un accès libre, convivial aux répondants avec un recueil mieux suivi et plus rapide pour celui qui l'organise. C'est aussi un moyen plus ciblé et plus attractif pour les professionnels de la plupart des secteurs d'activité, qui peuvent répondre selon leurs rythmes et leurs contraintes. Le choix de cette technique s'explique par sa facilité de mise en place et son accessibilité. Il comporte des limites en raison de l'équipement et de l'utilisation d'Internet chez les particuliers ; mais, une étude menée sur l'utilisation des technologies de l'information et de la communication en 2012 notait que 80 % des Français avaient utilisé un ordinateur au cours des trois derniers mois et 81 % des ménages avaient accès à Internet (DEROIN, 2013). Une des limites de cette étude est qu'elle dépendait de la volonté et de la motivation du patient à participer à l'étude.

A partir du logiciel Sphinx®, nous avons construit notre questionnaire qui a été hébergé sur un serveur de l'Université du Havre. Les individus de l'enquête étaient informés de l'enquête par leur médecin traitant et une note explicative contenant un lien vers l'URL du formulaire en ligne était remise au patient s'ils consentaient à participer à l'étude.

L'enquête en ligne a l'avantage de permettre le suivi de l'évolution des entrées des données en temps réel et de juger de l'état d'avancement afin de faire des relances ou de trouver des solutions palliatives à des problèmes survenant au cours de la collecte des données. Mais dans le cas de notre étude, il nous était impossible de faire des relances aux patients car nous n'étions pas en contact avec eux ; seule une relance auprès du médecin était possible.

La collecte des données en ligne nous permettait de gagner du temps par rapport à une collecte auprès des patients dans les cabinets des médecins et nous évitait également la réticence des médecins à nous accueillir dans leur cabinet en présence des patients.

2.2.2 *Technique d'entrevue ou d'entretien*

L'entretien est une méthode de recherche et d'investigation. Par le biais de cette méthode, l'enquêteur cherche à obtenir des informations sur les attitudes, les comportements, les représentations d'un ou de plusieurs individus dans la société (TOUMERT, 2017).

La technique d'entretien a été utilisée pour interroger les médecins sur l'utilisation des objets connectés, des applications de santé et les sites d'informations dans leur pratique médicale. Ils ont été également interviewés sur leur avis sur l'utilisation de l'IA et des données santé dans le cadre du *big data*. Nous avons recueilli leur point de vue sur les impacts de l'usage des TIC dans leur pratique quotidienne et future.

L'utilisation de la technique d'entretien nous a permis d'atteindre une saturation rapide de la question. En plus, nous avons l'ensemble des données qui nous permet de faire une analyse qualitative ou quantitative en fonction des objectifs de notre étude et tester ainsi nos hypothèses de recherche.

Pour atteindre nos objectifs, nous avons réalisé des entretiens compréhensifs semi directifs en face-à-face, qui donnent une plus grande marge de liberté aux enquêtés et permettent aux chercheurs de faire une analyse comparative des données recueillies. L'entretien compréhensif s'inscrit dans une démarche inductive où le recueil et l'analyse des informations ne sont pas des étapes successives, mais s'opèrent conjointement, au fur et à mesure des entretiens, en même temps que l'interprétation et l'élaboration de la problématique et des hypothèses. L'objectif est de parvenir à une compréhension intime de la pensée et de l'action des sujets (VAN CAMPENHOUDT, QUIVY, 2011).

En effet, l'utilisation de l'entretien compréhensif semi-directif a fait surgir des thématiques dont la télémédecine, les assistants médicaux et la relance du DMP. Toutes ces thématiques étaient en lien avec l'utilisation du *big data* et de l'IA dans le domaine de la santé.

Dans un souci de comparer scientifiquement les données, nous avons posé un certain nombre de questions qui ont été les mêmes pour tous les interviewés. Durant la réalisation des

enquêtes, les réponses des interviewés ont fait ressortir des aspects et de nouvelles considérations que nous avons pris en compte pour les entretiens suivants. Comme le disent Van Campenhoudt et Quivy, l'entretien permet

« l'analyse du sens que les acteurs donnent à leurs pratiques et aux événements auxquels ils sont confrontés : leurs systèmes de valeurs, leurs repères normatifs, leurs interprétations de situations conflictuelles ou non, leurs lectures de leurs propres expériences » (VAN CAMPENHOUDT, QUIVY, 2011).

Nous avons également interviewé des personnes ressources qui de par leurs compétences, leurs expériences et leurs implications dans la prise des décisions dans l'utilisation du *big data* et de l'IA dans le domaine de la santé nous ont fourni des informations très pertinentes pour cette étude.

2.3 Echantillonnage

2.3.1 Population cible

Notre étude s'est déroulée dans la commune du Havre. Elle concernait tous les médecins de la commune et leurs patients. Nous avons considéré que les patients venaient tous de la commune du Havre même si dans la pratique nous pouvons avoir des chevauchements de zone administrative dans la demande de soins par la population. Etant donné que les médecins doivent couvrir tout le territoire administratif de la commune, l'ensemble de la population de la commune constitue notre population cible.

2.3.2 Population d'étude

Échantillonnage des médecins

Nous avons inclus dans notre travail les médecins toujours en activité régulière de soins, consentant à participer aux entretiens et à remettre les fiches d'informations à leurs patients. Le médecin devait avoir dans sa cohorte de suivi des patients ou lors de ses consultations des patients de plus de 15 ans.

Dans le cadre de nos entretiens, nous avons limité la taille de notre échantillon de médecins à 20. Ce choix s'expliquait par le fait que le recrutement des patients était centré sur les médecins, pour éviter ainsi une dispersion de l'échantillon des patients et permettre une meilleure corrélation de la relation de soin dans le cadre de l'utilisation du *big data* dans le domaine de la santé. Le choix de notre méthode d'analyse nous impose également un nombre raisonnable pour être efficace et efficient dans l'analyse des corpus. Ce nombre de médecins permet d'atteindre une saturation de plus de 90 % du sujet dans le cadre des entretiens individuels (GRIFFIN, HAUSER, 1993) (voir figure ci-dessous).

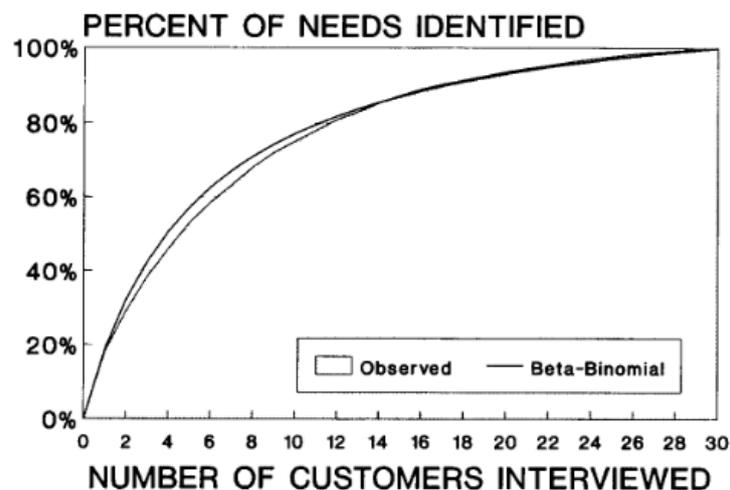


Figure 1 : pourcentage de saturation en fonction du nombre de personnes enquêtés
(GRIFFIN, HAUSER, 1993)

L'apport d'informations nouvelles devient marginal au-delà de 20 personnes. En plus, avec l'aspect asymptotique de la courbe de saturation, la probabilité d'apparition d'idée nouvelle devient très faible ; on peut dire qu'il y a une saturation probable de la question à partir de ce nombre.

Nous avons utilisé une méthode d'échantillonnage à plusieurs degrés (échantillonnage aléatoire stratifié)⁷⁰ pour le choix des médecins. Dans un premier temps, nous avons découpé la commune du Havre selon les cinq quartiers administratifs. Ensuite, nous avons sélectionné

⁷⁰ L'idée de base de l'échantillonnage aléatoire stratifié est de diviser une population hétérogène en sous-populations, habituellement appelées strates, constituées d'éléments homogènes ; on pourra ainsi obtenir une estimation précise d'une moyenne de strate quelconque à partir d'un petit échantillon prélevé dans cette strate, ainsi qu'une estimation précise pour l'ensemble de la population, en combinant ces estimations. L'échantillonnage stratifié donne une meilleure section transversale de la population que la procédure d'échantillonnage aléatoire simple. En général le degré de précision augmente avec le nombre de strates de la population, car plus elles sont nombreuses, plus les unités qu'elles contiennent sont homogènes. Source <http://www.fao.org/3/X6831F/X6831f14.htm> consulté de 20 avril 2018.

de manière aléatoire et proportionnelle les cabinets de médecins dans chaque quartier et nous avons envoyé des lettres d'invitation aux cabinets tirés au sort de manière aléatoire. Chaque tirage a été fait sur la base des cabinets de chaque quartier. Nous avons envoyé trente lettres en tenant compte des possibles refus de participation à l'étude.

La base de l'échantillonnage a été l'annuaire de l'assurance maladie en ligne⁷¹ qui est accessible gratuitement sans identification préalable. L'annuaire permet de faire des requêtes de recherche selon les critères de sélection des médecins et selon les besoins. Pour les requêtes, nous avons pris en compte le type d'acte « **la consultation** » et le lieu de pratique médicale « **Le Havre (76)** ». Cette requête nous permettait de faire une sélection de tous les médecins ayant une activité de consultation sans distinction de la spécialité et ayant comme commune d'installation Le Havre.

Cette requête réalisée le 15 mai 2018, nous retournait un résultat de 284 praticiens. Et comme il existe des pratiques en cabinet individuel ou en groupe, nous avons recensé tous les cabinets où pratiquaient les médecins à partir des adresses fournies par la requête. Ce travail de tri manuel nous a donné un résultat de 89 cabinets.

Nous avons donc constitué notre échantillonnage de médecins sur la base des cabinets. Ce choix permettait de voir une répartition spatiale des cabinets dans les cinq quartiers de la commune de Havre (Voir ci-dessous la carte de répartition des cabinets dans les quartiers de la commune du Havre). Notre choix reposait sur le fait que la pratique en groupe ou la pratique individuelle n'avait pas d'influence sur l'utilisation du numérique dans les cabinets à cause de la mutualisation des ressources par les acteurs.



Figure 2 : méthode d'échantillonnage des cabinets et des médecins

⁷¹ Annuaire des professionnels de santé en ligne. Source : <http://annuairesante.ameli.fr/> consulté le 15 mai 2018.

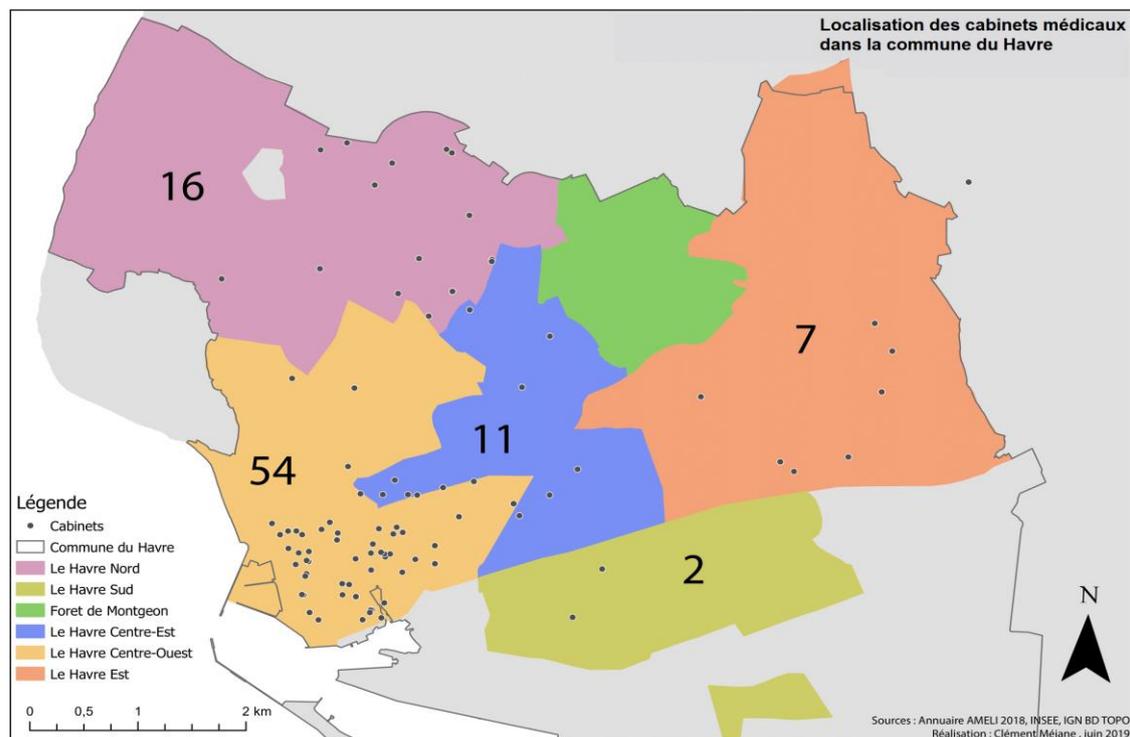


Figure 3 : répartition spatiale des cabinets médicaux de la commune du Havre

Nous avons déposé des lettres (*Annexe 4*) directement auprès des cabinets expliquant les objectifs de l'étude et les différentes étapes en particulier, la phase de recrutement des patients. Lorsque le médecin acceptait de participer à l'étude, un rendez-vous était pris à la convenance du médecin pour l'entretien. Durant cette phase d'échantillonnage et de demande de participation à l'étude qui s'est déroulée du début du mois d'octobre 2018 à la fin du mois de novembre 2018, seulement cinq médecins sur les trente (16 %) ont accepté de participer à l'étude.

Dans un second temps, n'ayant pas atteint le nombre de médecins nécessaires pour notre travail, nous avons envoyé une lettre de demande de participation aux 59 cabinets restants dans notre base d'échantillonnage, passant ainsi à un échantillonnage systématique de tous les cabinets de la commune du Havre.

Cette phase s'est déroulée du début du mois de décembre 2018 à la fin du mois de janvier 2019. Elle nous a permis d'avoir la participation de 16 cabinets sur 59 (27 %).

Nous avons également enquêté deux médecins du groupe hospitalier du Havre à l'Hôpital Jacques Monod et à l'Hôpital Flaubert. Nous avons inclus l'hôpital Jacques Monod, bien

qu'il ne soit pas situé dans la commune du Havre, parce qu'il est le principal centre de référence des cabinets médicaux de la ville du Havre.

Ainsi, nous avons eu un accord de participation de 23 médecins dont deux n'ont pas été enquêtés (soit un taux de participation des cabinets de 23 %). La cartographie ci-dessous montre la répartition spatiale des structures de soins enquêtées dans la commune du Havre.

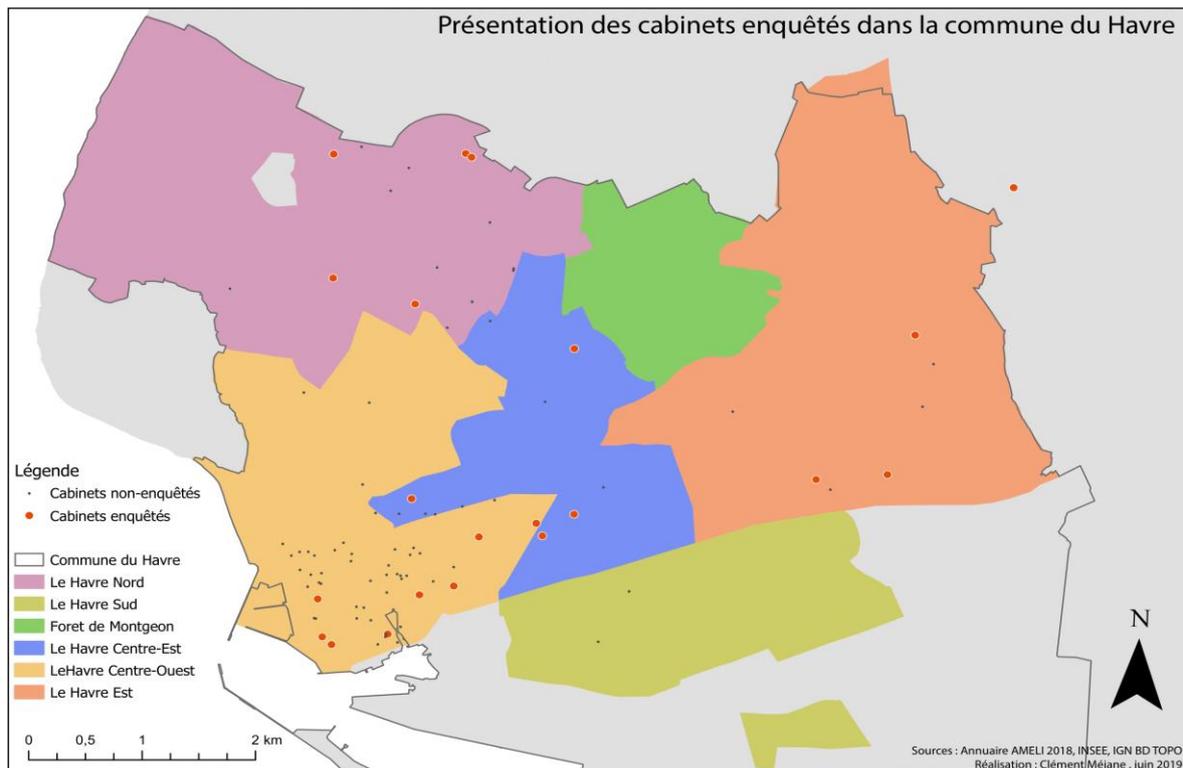


Figure 4 : répartition spatiale des structures de soins enquêtées dans la commune du Havre

Lorsque nous n'avons pas de réponse au bout de deux semaines, nous avons fait des relances pour avoir une confirmation de participation ou non à l'étude.

Echantillonnage des patients

Nous avons inclus dans notre étude des patients ayant un âge supérieur à 15 ans et consentant à participer à notre étude. Nous n'avons donc pas jugé nécessaire d'inclure des patients dont la réponse dépendait des tuteurs du fait que ces derniers pouvaient faire partie de l'échantillon d'étude des patients pour éviter le biais de réponse double.

La taille d'un échantillon d'une étude menée auprès d'une population donnée est un facteur déterminant pour obtenir des données fiables à propos d'une proportion dans une population. La fiabilité des données n'est jamais absolue, mais se situe plutôt dans un intervalle de confiance. Plus l'intervalle de confiance est petit, plus le risque d'erreur est faible.

Nous avons calculé la taille de l'échantillon à partir de la formule de Schwartz (SCHWARTZ, 1993) :

$$n = z^2 \frac{p(1-p)}{d^2}$$

n = taille de l'échantillon

z = le niveau de confiance selon la loi normale. Dans notre cas, nous avons pris un niveau de confiance de 95 %, donc z = 1,96

d = la marge d'erreur tolérée égale à 5 % dans notre cas

p = proportion estimée de la population qui acceptera le partage des données santé dans le cadre de l'utilisation du *big data*. Nous ne disposons pas de données dans ce cadre pour notre population. La valeur de p est alors estimée à 0,5.

Notre taille de l'échantillon à partir de la formule ci-dessus est de donc n = 384.

Nous avons considéré un taux de réponse dans notre étude de 80 % ramenant ainsi la taille de notre échantillon à enquêter à 480. L'échantillon a été réparti entre les médecins de façon équitable. Chaque médecin devait enrôler dans l'étude au moins 24 individus (480/20).

L'enrôlement des patients s'est fait de façon systématique lorsque le patient venait à la consultation et consentait à participer à l'étude. L'enrôlement a été stoppé dès que le nombre de patients requis pour chaque médecin a été atteint. L'enrôlement était également adapté au mode de fonctionnement du médecin afin de ne pas perturber son emploi du temps.

Nous avons mis à la disposition des médecins, des fiches qui expliquaient les objectifs de l'étude, la demande de consentement ainsi que les informations sur le responsable du traitement des données et le lien d'accès au questionnaire en ligne (*Annexe 7*).

Le questionnaire a été mis en ligne le 10 octobre 2018.

Echantillonnage des personnes ressources

Nous avons interviewé des personnes ressources du fait de leurs implications dans le processus d'utilisation du big data dans le domaine de la santé. Ces personnes sont choisies soit pour leurs compétences et leur notoriété scientifique dans le domaine, soit du fait de l'influence de leurs opinions sur la question. Le choix des personnes a été fait à partir des questions durant nos entretiens avec les médecins et de celles que nous nous sommes posées pendant notre phase de revue de la littérature.

Les personnes enquêtées étaient au nombre de quatre :

- une responsable des services informatiques d'un hôpital au Havre pour être plus proche de la réalité de gestion des infrastructures informatiques et des applications dans les hôpitaux du Havre ;
- un responsable de département d'information médicale au Havre pour avoir un point de vue pratique sur la gestion et le traitement des données de santé dans les hôpitaux, en particulier ceux du Havre ;
- un chef d'entreprise diplômé de médecine dans le développement d'applications d'IA en santé à Paris ;
- une juriste spécialiste de l'utilisation des TIC dans le domaine de la santé à Lille ; l'entretien s'est déroulé par téléphone. Ce choix s'explique par le fait que nous ne pouvons pas aborder les questions d'éthique du *big data* et de l'IA dans le domaine de la santé en passant sous silence les questions juridiques de l'utilisation des TIC dans le domaine de la santé.

2.4 Techniques d'analyse

Analyser, c'est répondre à des questions. Pour cela, il faut utiliser les techniques d'analyse adéquates pour avoir les meilleures réponses.

Dans notre travail, nous avons utilisé trois techniques d'analyse : l'analyse thématique de contenu, l'analyse statistique textuelle et l'analyse de données quantitatives. Le choix de ces trois techniques s'explique par les techniques de recueil des données que nous utilisons.

2.4.1 Techniques d'analyse qualitative

Dans le cadre de l'analyse qualitative, nous avons réalisé une analyse de contenu.

C'est « un ensemble de techniques d'analyse des communications visant, par des procédures systématiques et objectives de description du contenu des messages, à obtenir des indications (quantitatives ou non) permettant l'inférence de connaissances relatives aux conditions de production/réception (variables inférées) de ces messages » (BARDIN, 2013).

Nous avons utilisé l'analyse thématique de contenu et l'analyse statistique textuelle.

L'analyse thématique de contenu est une technique d'analyse qualitative qui fait intervenir des procédés de réduction des données.

« L'analyste va faire appel pour résumer et traiter son corpus à des dénominations que l'on appelle les thèmes. Elle a deux fonctions principales : une fonction de repérage et une fonction de documentation. La première fonction concerne le travail de saisie de l'ensemble du corpus. La tâche est de relever les thèmes pertinents en lien avec les objectifs de la recherche, à l'intérieur du matériau à l'étude. La deuxième concerne la capacité de tracer des parallèles ou de documenter les oppositions ou les divergences entre les thèmes » (PAILLÉ, MUCCHIELLI, 2016).

Nous avons utilisé la logique de la thématization continue qui consiste en une démarche ininterrompue d'attribution des thèmes et simultanément de construction de l'arbre thématique (PAILLÉ, MUCCHIELLI, 2016). Cette méthode est très utile car nous procédons après chaque entretien à une analyse de celui-ci afin de permettre une exploration de pistes inconnues suite à l'interaction avec les interviewés. Elle permet une analyse fine et riche du corpus.

Elle peut être mise à contribution seule ou être combinée avec d'autres stratégies d'analyse. L'analyse thématique représente un outil important dans bon nombre de cas car elle permet de poser rapidement un diagnostic, une utilisation dans le cadre d'une recherche-action et une utilisation complémentaire à des méthodes quantitatives.

2.4.2 *Techniques d'analyse quantitative*

L'analyse quantitative désigne l'ensemble des méthodes et des raisonnements utilisés pour analyser des données standardisées (c'est-à-dire des informations dont la nature et les modalités de codage sont strictement identiques d'un individu ou d'une situation à l'autre). Ces données résultent souvent d'une enquête par questionnaire mais peuvent également être produites par le codage de documents d'archives, de dossiers administratifs, de sources sonores ou visuelles (MARTIN, 2012).

« Dans l'analyse quantitative, le chercheur tente de répondre aux questions posées lors d'une enquête et/ou de vérifier des hypothèses émises. La difficulté et l'ampleur de l'analyse dépendent des objectifs de l'étude et de la complexité des variables recherchées » (LELEU-MERVIEL, 2008).

Lors de l'analyse quantitative, nous avons prévu de réaliser une analyse descriptive des caractéristiques des groupes d'individus impliqués dans notre étude. Dans un second temps nous réaliserons des analyses bivariées et multivariées pour mettre en évidence les relations entre les variables à partir des tableaux de contingence et des régressions logistiques. *« La régression logistique a pour but d'isoler les effets de chaque variable, c'est-à-dire d'identifier les effets résiduels d'une variable explicative sur une variable d'intérêt » (LARMARANGE, 2018).* Enfin, nous utiliserons l'analyse factorielle multiple pour réduire le nombre de variables et rendre l'information moins redondante. Elle consiste à transformer des variables liées entre elles en nouvelles variables décorréélées les unes des autres.

2.4.3 *Logiciels d'analyse des données*

Le choix d'un outil d'analyse devrait surtout dépendre de trois éléments : les choix méthodologiques, la constitution du corpus, le moment de l'analyse statistique (FALLERY, RODHAIN, 2007).

Toutes les analyses quantitatives seront réalisées avec le logiciel R (R DEVELOPMENT CORE TEAM, 2008), qui est un logiciel open source avec une forte communauté de développeurs sur les besoins spécifiques de chaque groupe. L'utilisation de R présente plusieurs avantages : c'est un logiciel très puissant :

- développé par ses utilisateurs et modifiable par tout un chacun ;

- dont les fonctionnalités de base peuvent être étendues à l'aide de plusieurs milliers d'extensions ;
- dont le développement est très actif et dont la communauté d'utilisateurs ne cesse de s'élargir ;
- avec des possibilités de manipulation de données qui sont en général largement supérieures à celles des autres logiciels usuels d'analyse statistique ;
- avec d'excellentes capacités graphiques et de nombreuses possibilités d'export ;
- qui est utilisé dans tous les secteurs scientifiques, y compris dans le domaine des analyses d'enquêtes et, plus généralement, des sciences sociales.

Iramuteq est un logiciel d'analyse de textes et de tableaux de données. Il s'appuie sur le logiciel de statistique R (<http://www.r-project.org>) et sur le langage python. Iramuteq propose un ensemble de traitements et d'outils pour l'aide à la description et à l'analyse de corpus de texte et de matrices de type individus/caractères.

La méthode Alceste® ou méthode de Reinert « *consiste à effectuer plusieurs découpages du corpus en « unités de contexte », à réaliser des comparaisons et des regroupements du corpus segmenté selon les lexèmes contenus, puis à chercher les distributions stables, « c'est-à-dire, indépendantes de l'arbitraire de chaque découpage »* (REINERT, 1987).

La classification mise en place dans la méthode Alceste est une classification de type « *hiérarchique descendante* ». Elle procède par itérations successives à partir d'une analyse factorielle des correspondances multiples. Le logiciel Iramuteq propose également des analyses supplémentaires, telles que : analyse prototypique, analyse de similitudes, nuages de mots.

2.5 Réalisation des outils de collecte

Les questionnaires d'enquête des patients (*Annexe 8*) et le questionnaire semi-directifs des médecins (*Annexe 6*) ont été réalisés à partir du modèle d'acceptation de la technologie (TAM) de Davis (VENKATESH, DAVIS, 2000), le modèle éthique du *big data* dans le domaine de la santé proposé par Jérôme Béranger (BÉRANGER, 2016), la « *boîte à outils* » de Nicole et Louis Lery (LERY, COLLOC, 2008) et l'évaluation de la confiance au médecin traitant par

le patient à travers le questionnaire « *Trust in Physician Scale* » (ROLFE ET AL., 2014 ; BIZOUARN, 2008 ; ANDERSON, DEDRICK, 1991).

La *technology acceptance model (TAM)* se fonde sur l'utilité perçue de la technologie, les processus d'influence sociale et ceux de cognition instrumentale. Les processus cognitifs instrumentaux pris en compte, sur la base des théories de l'action, de la motivation du travail et de la décision comportementale considèrent trois composants clefs: la pertinence pour l'emploi; la qualité des résultats fournis par le système ; la démontrabilité des résultats (VENKATESH, DAVIS, 2000).

Le modèle éthique du *big data* dans le domaine de la santé repose sur les quatre principes fondamentaux de l'éthique médicale associés à l'exploitation des données personnelles de santé des patients (BÉRANGER, 2016). Les quatre principes d'éthique médicale sont : la bienfaisance, l'autonomie, la non-malfaisance et la justice (BEAUCHAMP, CHILDRESS, 2009) décrites par Beauchamp et Childress.

La « *boite à outils* » de Nicole et Louis Lery s'appuie sur la compétence professionnelle et les connaissances scientifiques du professionnel. Ce modèle explore le bénéfice/risque/coût de la décision médicale en prenant en compte la fleur relationnelle du patient (famille, entourage, professionnel et institution de santé), les repères juridiques (textes juridiques de protection des données personnelles aussi bien en France que dans l'Union Européenne et les textes relatifs au droit des patients dans le domaine de la santé) existant dans le domaine, la déontologie (code de déontologie médicale dans le cadre de notre travail) et la culture et les croyances des patients (LERY, COLLOC, 2008).

La confiance est la base de la relation médecin - patient. Il s'agit d'un construit psychologique qui peut être mesuré par des questionnaires (ROLFE ET AL., 2014 ; BIZOUARN, 2008 ; ANDERSON, DEDRICK, 1991). Ces questionnaires comportent plusieurs dimensions (compétence, confidentialité, fiabilité, honnêteté, fidélité et capacité communicationnelle) (BIZOUARN, 2008 ; ROLFE ET AL., 2014). Des études ont montré que la confiance aux médecins ou à la profession médicale d'une manière générale est influencée par les mêmes facteurs (BIZOUARN, 2008 ; ROLFE ET AL., 2014). Dans le cadre de notre enquête, nous avons

utilisé le questionnaire « *Trust in Physician Scale* » (ANDERSON, DEDRICK, 1991). Ce questionnaire évalue trois dimensions de la confiance qui sont :

- la prise en compte de l'intérêt du patient dans la décision médicale,
- les compétences et l'habileté technique du médecin ;
- la confidentialité et la fiabilité de l'information entre le médecin et le patient.

Le questionnaire a une bonne corrélation et un coefficient alpha de Cronbach⁷² égal à 0,90. Il est constitué de 11 questions avec cinq modalités de réponse pour chaque question selon une échelle de Likert (ANDERSON, DEDRICK, 1991).

Les questionnaires sont destinés aux patients dans le cadre d'une enquête par Internet, aux médecins et aux différents spécialistes dans le cadre d'un entretien semi-directif.

2.5.1 Guide d'entretien des médecins

Le guide d'entretien des médecins était constitué de quatre parties. Il était composé à la fois de questions ouvertes et des questions fermées. Les parties étaient :

- l'utilisation des objets connectés, sites web et applications : cette partie concernait le but de l'utilisation, le type d'objets connectés, les sites web et les applications de santé, le degré de confiance, les freins à l'utilisation et les actions pour améliorer de l'adhésion des médecins ;
- la protection des données de santé : elle concernait le partage des données et les raisons de ce partage, les conditions dans lesquelles les médecins souhaitaient partager ces données de santé, leur opinion sur les différentes lois de protection des données et le degré de confiance accordé aux institutions pour la protection des données ;
- *le big data* et l'IA : elle concernait les modifications provoquées sur la pratique médicale, les attentes et les promesses de ces technologies dans le système de santé et les risques pour l'éthique médicale ;

⁷² Le coefficient alpha de Cronbach, parfois appelé simplement coefficient α , est une statistique utilisée notamment en psychométrie pour mesurer la cohérence interne (ou la fiabilité) des questions posées lors d'un test (les réponses aux questions portant sur le même sujet devant être corrélées). Sa valeur est inférieure ou égale à 1, étant considérée comme "acceptable" à partir de 0,7.

- l'éthique et la finalité des données : il s'agissait d'évaluer la finalité éthique des données de santé dans le cadre de son utilisation dans le *big data* et l'IA. Nous avons utilisé un questionnaire développé par Jérôme Béranger (BÉRANGER, 2016) qui repose sur les principes de l'éthique médicale (autonomie, bienfaisance, justice, non-malfaisance) (BEAUCHAMP, CHILDRESS, 2009). Ce questionnaire permet de définir un score unitaire éthique variant de 0 à 100.

A côté de ces quatre parties, trois points ont été ajoutés à nos entretiens du fait de l'actualité de la politique de santé. Il s'agissait de :

- la télémédecine : la généralisation et le remboursement de la télémédecine dans toute la France le 15 septembre 2018⁷³. Les entretiens concernaient l'apport de la télémédecine dans la prise en charge des patients, son impact sur la qualité des soins, les modalités pratiques et la responsabilité médicale ;
- le DMP : la relance du DMP le 6 novembre 2018 par la ministre de la santé, Agnès Buzyn⁷⁴. Les entretiens sur cet item concernaient l'apport du DMP dans la prise en charge des patients, la gestion et la gouvernance des données et les responsabilités des différents acteurs ;
- les assistants médicaux : dans le cadre du plan santé, présenté par le président de la République, Emmanuel Macron⁷⁵, il a été décidé la création de ce poste. Les entretiens sur cet item concernaient le rôle des assistants médicaux dans le cabinet médical et les modalités de recrutements des assistants.

⁷³ Téléconsultation : coup d'envoi le 15 septembre. « *Il devient possible en France, à compter du 15 septembre, de bénéficier d'un acte de téléconsultation remboursable par l'Assurance Maladie* ». Source : <https://www.ameli.fr/le-havre/assure/actualites/teleconsultation-coup-denvoi-le-15-septembre> consulté le 12 novembre 2018.

⁷⁴ Agnès Buzyn et Nicolas Revel relancent en grande pompe un DMP dépolvoisiéré. Source : https://www.lequotidiendumedecin.fr/actualites/article/2018/11/06/agnes-buzyn-et-nicolas-revel-relancent-en-grande-pompe-un-dmp-depoussiere_862513 consulté le 12 novembre 2018.

⁷⁵ « ... *Nous devons continuer à décharger les médecins d'actes qui peuvent être faits par d'autres* », a encouragé Emmanuel Macron, mardi 18 septembre, lors de la présentation de son plan santé. Dans cette optique, le président de la République a annoncé la création de 4 000 postes d'assistants médicaux à l'horizon 2022, destinés à aider les médecins pour leur libérer du temps médical. Source : https://www.lemonde.fr/campus/article/2018/09/19/plan-sante-4-000-postes-d-assistants-medicaux-accessibles-apres-une-formation-d-un-an_5357473_4401467.html consulté le 12 novembre 2018.

2.5.2 *Questionnaire des patients*

Nous avons utilisé un questionnaire qui était constitué de quatre parties pour enquêter auprès des patients. Ces parties étaient :

- l'utilisation du numérique en santé : elle concernait la recherche d'informations sur Internet et les réseaux sociaux, la confiance aux informations recueillies et la modification du mode de gestion des maladies des patients ;
- l'utilisation des objets connectés : elle concernait le type d'objets connectés, l'impact des objets connectés sur la relation de soin avec le médecin et les freins à leur utilisation ;
- la protection des données : elle concernait la confiance des patients aux sites et aux réseaux sociaux pour la protection des données, la confiance aux institutions, le partage des données et les craintes par rapport à la numérisation de la santé ;
- la confiance à leur médecin traitant : nous avons utilisé le questionnaire « *Trust in Physician Scale* » (ANDERSON, DEDRICK, 1991).

2.5.3 *Guide d'entretien des personnes ressources*

Le guide d'entretien des personnes ressources était adapté aux domaines de spécialités et comportait les points suivants :

- le rôle du spécialiste dans le traitement des données ;
- la protection juridique des données : l'hébergement des données selon les normes, les traitements et les accès des données par les différents acteurs en fonction de leur grade et de leurs responsabilités en cas de fautes ;
- la qualité des données dans les structures de soins : l'exactitude, la cohérence, l'intégrité et l'exhaustivité ;
- la sécurité des données : la confidentialité, la traçabilité, l'imputabilité, la disponibilité, la gestion des authentifications et la journalisation des activités et leurs auteurs ;
- la sécurité des systèmes d'information de santé : la sécurité matérielle, logicielle et physique.

2.6 Déroulement des entretiens et remplissage des questionnaires

Les entretiens se sont déroulés entre octobre 2018 et mai 2019. Sur les 23 rendez-vous obtenus, un rendez-vous a été annulé pour des raisons de congés de maternité et un autre pour des raisons de calendrier. Les entretiens se sont tous déroulés dans les bureaux des médecins. Les dates et heures des rendez-vous étaient fixées en fonction de l'emploi du temps des médecins enquêtés. Les horaires de rendez-vous étaient soit tôt le matin avant le début des activités, soit à la pause de midi ou soit à « *la descente* » après le dernier patient de la journée.

Les entretiens ont été réalisés en face-à-face sauf pour un entretien pour des raisons de distance et de calendrier (l'entretien a été réalisé dans ce cas par téléphone). Tous les entretiens ont été enregistrés avec le consentement écrit des médecins et des personnes ressources et retranscrits par nous-mêmes. La confidentialité et l'anonymat des participants ont été respectés durant tout le processus de traitement des données.

La durée moyenne des entretiens était de 41 ± 16 minutes avec des extrêmes de 20 et 84.

Les questionnaires ont pâti d'un faible taux de remplissage de la part des patients. Nous n'avons enregistré à la date de mai 2019 que huit questionnaires remplis. Nous étions limité par le temps pour mettre en place une nouvelle stratégie de collecte des données qui aurait nécessité une relance des médecins. Nous avons donc été contraint d'abandonner l'étude sur les patients.

2.7 Considérations éthiques et juridiques

Notre étude rentre dans le cadre d'une recherche n'impliquant pas la personne humaine selon les méthodologies de références⁷⁶ de la Commission Nationale de l'Informatique et des Libertés (CNIL). Il s'agit de la méthodologie de référence 004 (MR-004) mise en place le

⁷⁶ Recherche médicale : comment procéder pour une thèse ou un mémoire ? Source : <https://www.cnil.fr/fr/recherche-medicalement-comment-proceder-pour-une-these-ou-un-memoire> consulté le 15 janvier 2019.

13 juillet 2018 par délibération⁷⁷. Elle vient remplacer les catégories « hors champs » de la loi Jarde dans le cadre de la recherche biomédicale⁷⁸.

Nous avons ainsi déclaré notre étude à la CNIL conformément à la loi sur la protection et le traitement des données à caractère personnel par l'intermédiaire du représentant CNIL du laboratoire IDEES et du référent local de CNIL au niveau de l'Université du Havre.

Le consentement écrit des participants (*Annexe 5*) à l'étude a été demandé au préalable avant le début de l'étude après explication de l'objectif de l'étude, des types de données recueillis, leur protection et leur possibilité d'avoir accès aux résultats de l'étude par l'intermédiaire d'un formulaire dans le cadre des entretiens avec les médecins et les personnes ressources. Pour les patients, il s'agissait d'un formulaire en ligne avec les mêmes explications et chacun devait consentir avant de pouvoir participer à l'étude en cochant une case.

Les entretiens ont été anonymisés afin de préserver la confidentialité des identités des personnes enquêtées comme précisée dans la lettre d'information et la demande de consentement. Ainsi, dans l'optique de préserver la confidentialité, toutes les informations pouvant permettre d'identifier les personnes enquêtées ont été supprimées avant toutes manipulations.

La manipulation des entretiens a été faite uniquement par les personnes impliquées dans l'étude.

⁷⁷ Délibération n° 2018-155 du 3 mai 2018 portant homologation de la méthodologie de référence relative aux traitements de données à caractère personnel mis en œuvre dans le cadre des recherches n'impliquant pas la personne humaine, des études et évaluations dans le domaine de la santé (MR-004).

⁷⁸ Loi n° 2012-300 du 5 mars 2012 relative aux recherches impliquant la personne humaine.

3 Etat de l'art

3.1 *Big data*

Le développement de la technologie a permis une accessibilité facile avec un développement important des outils de productions de données et de collecte. Les capacités des bases de données classiques – bases de données relationnelles – ont été rapidement dépassées demandant ainsi des nouvelles structures de bases pour la collecte, le stockage et le traitement des données. Ces différentes technologies ont donné naissance au *big data*. Le principal objectif initial du développement du *big data* est la valorisation économique des données dans un premier et secondairement par le service rendu à la population. Le *big data* comporte ainsi trois niveaux qui sont :

- la production et la collecte ;
- les infrastructures et le traitement ;
- la restitution et la visualisation des résultats.

Le domaine de la santé n'est pas en marge de cette production importante des données depuis la numérisation du domaine de la santé de l'administration à la gestion quotidienne des soins.

3.1.1 *Définitions*

Le terme de *big data* désigne une nouvelle discipline qui se situe au croisement de plusieurs disciplines tels que les technologies, les statistiques, l'informatique et la santé. Il n'existe pas une définition précise du *big data*. Certaines définitions reposent soit sur la capacité de collecte des données, soit sur la capacité de valorisation économique des données ou sur le service rendu à la population à partir de l'exploitation des données.

La définition du *big data* la plus répandue et la plus utilisée par les gestionnaires et les économistes basée sur les 5V (*Volume, Velocity, Variety, Veracity, Value*) (MARR, 2015) :

- *volume* correspond à la quantité de données produite et stockée dans un temps relativement court. Cette production est augmentée avec l'utilisation des réseaux sociaux, des dispositifs connectés, des capteurs et une explosion du e-commerce. Dans le domaine de la santé cette explosion est due à une numérisation de tous les actes de soins ;

- *velocity* (vitesse) correspond à la rapidité avec laquelle les données sont produites et utilisées en temps réel. Elle est très importante dans la santé avec le développement de la biologie moléculaire et de la numérisation de l'imagerie qui produisent des quantités importantes de données de par leur utilisation ;
- *variety* (variété) : les données sont classées en structurées et non structurées. Les données structurées sont stockées dans des bases de données et ne représentent que 20 % des données. Les données non structurées sont les plus importantes (80 %) des données : textes, images, vidéo, voix. La technologie *big data*, permet d'analyser, comparer, reconnaître, classer des données de différents types comme des messages, des conversations sur les réseaux sociaux, des photos, des données détectées, des vidéos, des enregistrements vocaux ;
- *veracity* (véracité) concerne la crédibilité et la fiabilité des données. La véracité des données est très importante dans la production des informations et des connaissances fiables. Elle sera la base de la production des valeurs. La véracité des données est la base de la qualité de la prise en charge des patients dans le domaine de la santé ;
- *value* (valeur) : l'objectif des entreprises est de tirer un profit des quantités importantes des données qu'elles stockent. Dans le domaine de la santé, elle consiste à une réduction des dépenses et une amélioration de la qualité de la prise en charge des patients.

Le cabinet IDC⁷⁹ (*International Data Corporation*) définit le *big data* comme « *a new generation of technologies and architectures designed to economically extract value from very large volumes of a wide variety of data by enabling high-velocity capture, discovery, and/or analysis* »⁸⁰

⁷⁹ *International Data Corporation (IDC)* est le premier groupe mondial de conseil et d'études sur les marchés des technologies de l'information. IDC fournit aux utilisateurs des technologies informatiques et télécoms, aux cabinets de conseil et d'investissement, des outils de compréhension des offres du marché ainsi que des données économiques précises et fiables sur les marchés et leurs acteurs.

⁸⁰ *Big data Analytics: Future Architectures, Skills and Roadmaps for the CIO, IDC 2011*, Source : <https://triangleinformationmanagement.com/wp-content/uploads/2013/12/bigdata-idc-wp.pdf> consulté le 03 septembre 2019

Encadré 1: production des données de santé en France

Les données de santé sont en pleine explosion, dans tous les pays du monde : leur volume devrait atteindre 2,3 milliards de gigaoctets d'ici à 2020, et le nombre d'objets connectés devrait atteindre 80 milliards en 2020. Le volume de données produites chaque année dans le monde va être multiplié par 44 d'ici 2020. En France, 50 000 applications de santé sont recensées par la Haute Autorité de santé (HAS). Ces données, d'une grande disparité, d'une vaste diversité de formats, proviennent de multiples sources : les essais cliniques, les bases de données médico-administratives comme le SNIIRAM et ses 8,9 milliards de feuilles de soins, 20 milliards de lignes, les cohortes et registres, le dossier médical, les données de patients et leurs objets connectés ou applications, les réseaux sociaux et les sites Internet⁸¹. (Source INSERM).

La figure 5 montre les différentes sources (données pharmaceutiques, données de soins, registres d'essai clinique, réseaux sociaux, données de la vie hors structure de soins) possibles des données d'un patient et les différentes interactions possibles entre les données. Nous avons à la fois des données structurées et non structurées. La particularité des données de santé est que leur structuration dépend de l'utilisation des standards d'interopérabilité (*LOINC*⁸², *CIM*⁸³, *SNOMED*⁸⁴, *DICOM*⁸⁵, *HL7*⁸⁶...). Cette figure montre également la complexité dans le traitement des données afin de leur donner de la valeur.

⁸¹ *Big data* : l'explosion de la production de données par Dr Philippe Massol Source : <https://www.egora.fr/actus-medicales/sante-publique/39040-big-data-l-explosion-de-la-production-de-donnees> consulté le 03 septembre 2019.

⁸² *Logical Observation Identifiers Names and Codes (LOINC)* est une terminologie de référence internationale. Elle permet le codage des analyses élémentaires de biologie pour l'expression des résultats dans un contenu médical.

⁸³ La Classification internationale des maladies (*CIM*) est une classification médicale codifiée classifiant les maladies et une très vaste variété de signes, symptômes, lésions traumatiques, empoisonnements, circonstances sociales et causes externes de blessures ou de maladies.

⁸⁴ *Systematized nomenclature of medicine clinical terms® (SNOMED CT®)*, est un système international de terminologie clinique. Il comprend une liste de termes cliniques contrôlés, détaillés et validés médicalement avec leurs synonymes. Cette terminologie est utilisée dans les soins directs aux patients pour documenter les plaintes physiques et psychologiques, les symptômes, les circonstances, les maladies, les interventions chirurgicales, les diagnostics, les résultats et les décisions thérapeutiques.

⁸⁵ *Digital imaging and communications in medicine (DICOM)*, est une norme standard pour la gestion informatique des données issues de l'imagerie médicale.

⁸⁶ *Health Level 7 (HL7)* est une organisation qui définit un ensemble - auquel il donne son nom - de spécifications techniques pour les échanges informatisés de données cliniques, financières et administratives entre systèmes d'information hospitaliers (SIH). Ces spécifications sont diversement intégrées au corpus des normes formelles américaines et internationales.

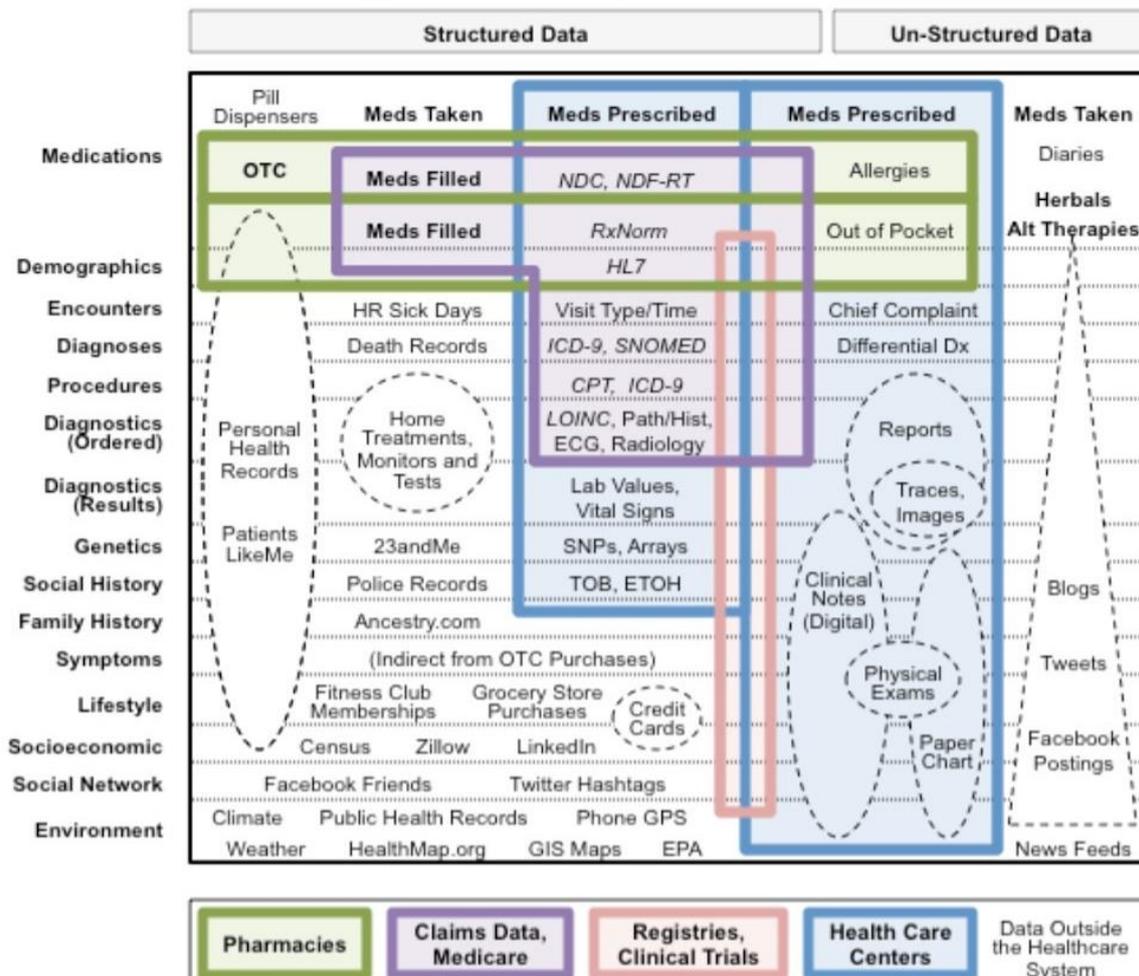


Figure 5 : carte des données santé d'un patient⁸⁷

A partir des définitions, nous constatons que le *big data* ne se limite pas à la désignation de grandes bases de données mais fait référence également aux technologies, aux processus et techniques permettant la gestion, la manipulation de ses données en vue d'extraire des informations et des connaissances pour les transformer en valeur économique.

Les bases de données relationnelles se sont imposées depuis longtemps dans le domaine de stockage des données. Elles servent à stocker des données opérationnelles pour pouvoir insérer, modifier et interroger rapidement la base en toute sécurité. Il utilise de ce fait le OLTP (*On Line Transaction Processing*). Les transactions travaillent sur de faible quantité de données et sur des versions toujours plus récentes. Face à la production de quantité importante de données structurées et non structurées, le nombre de plus en plus croissant des transactions et les besoins analytiques, les bases de données relationnelles ont montré leurs

⁸⁷ Source : <http://numerique.anap.fr/publication/1505-big-data-en-sante-donnees-concernees-usages-entrepot-bio-heterogenes-et-outils-d-exploitation> consulté le 23 avril 2019.

limites (BRASSEUR, 2013). Pour répondre aux besoins de stockage de quantité de données de plus en plus importants et hétérogènes avec une capacité d'historisation, les entrepôts de données (*datawarehouses*) ont vu le jour. Les entrepôts de données reposent sur le système OLAP (*On Line Analytical Processing*). Les programmes consultent des quantités importantes de données pour procéder à des analyses. La mise en place de cette technologie porte sur les bases de données, les plateformes de développement et de traitement des données.

3.1.2 Description des technologies du big data

Bases de données NoSQL (Not only Structured Query Language)

Il est très délicat de donner une définition précise aux bases de données *NoSQL*.

Les bases de données *NoSQL (Not only SQL)* ne s'opposent pas aux bases de données *SQL* ; elles sont une alternative à ces dernières qui n'offrent ni l'évolutivité ni le passage à l'échelle (« *scalability* ») nécessaires aux besoins importants d'augmentation de la quantité de données, de puissance de traitement et des infrastructures qui les hébergent. Elles viennent compléter les fonctionnalités des bases de données afin de les adapter à des usages contextuels.

Les bases de données *NoSQL* ont de ce fait pour objectifs de (LEMBERGER ET AL., 2016) :

- distribuer les traitements et le stockage sur des centaines voire des milliers de nœuds constitués de serveurs ;
- donner la priorité aux performances et à la disponibilité ;
- traiter efficacement les données non structurées ou seulement partiellement structurées.

Les points communs à ces systèmes sont (LEMBERGER ET AL., 2016) :

- la possibilité de *clustérisation* permettant une montée en charge approximativement linéaire ;
- l'absence de schémas car ils ne sont pas adaptés pour les données non structurées ;
- l'absence de transactions au sens habituel du terme, sauf pour des transactions qui garantissent seulement l'intégrité de certains agrégats de données naturels ;
- l'absence de jointure.

Une base de données distribuée devrait posséder trois caractéristiques (*théorème CAP* (*Consistency, Availability, Partition tolerance*)) qui sont nécessaires à son bon fonctionnement (LEMBERGER ET AL., 2016) :

- la cohérence (« *consistency* ») : chaque donnée est présente dans la même version sur tous les nœuds du réseau. Il n’y a pas de retard dans la propagation des mises à jour ;
- la disponibilité (« *availability* ») : l’application est disponible à tout instant et chaque requête reçoit une réponse qui confirme si elle a été traitée avec succès ou non ;
- la résistance au morcellement (« *partition tolerance* ») : le système doit continuer à fonctionner lorsque différents nœuds sont isolés suite à une rupture de réseau.

Ainsi, le mode de fonctionnement des bases de données *NoSQL* permet de distinguer deux grandes catégories - réparties sous quatre modèles - que sont (LAROCHE, 2016 ; LEMBERGER ET AL., 2016 ; BEHMO, TRAVERS, 2019) :

- les bases de données orientées agrégats (BDOA) : on renonce à éclater les données dans une multitude de tables et on les regroupe plutôt au sein d’agrégats qui contiennent les données auxquelles on accède le plus souvent. Cela facilite le déploiement des systèmes sur des clusters. Les BDOA sont de trois modèles : les bases de données clé - valeur ; les bases de données orientées documents et les bases de données orientées colonnes ;
- les bases de données orientées graphes (BDOG) conçues pour naviguer dans un graphe de données.

Hadoop

Hadoop est un *framework* logiciel open source basé sur *Java* permettant de stocker des données, et de lancer des applications sur des grappes de machines standards. Cette solution offre un espace de stockage massif pour tous les types de données, une immense puissance de traitement et la possibilité de prendre en charge une quantité de tâches virtuellement illimitée.

Hadoop repose sur un ensemble de machines formant un cluster *Hadoop*. Chaque machine est appelée nœud. C'est l'addition des capacités de stockage et traitement de ses nœuds qui lui assure un important système de stockage et une puissance de calcul. Le système de stockage est appelé *HDFS (Hadoop Distributed File System)* (EL MALKI, 2016). La puissance de calcul repose sur le paradigme de programmation parallèle (DEAN, GHEMAWAT, 2008).

Parmi les principales entreprises du marché, on compte notamment Amazon Web Services, Teradata Corporation, Cisco Systems, IBM Corporation, Cloudera, Inc., Datameer, Inc., Oracle Corporation, Hortonworks, Inc., VMware, OpenX.

Les entrepôts de données actuelles reposent en majeure partie sur une organisation des stockages en ligne (*cloud computing*) qui permet un traitement plus rapide des transactions et l'enregistrement des événements.

3.1.3 *Big data et cloud computing*

Le développement du *cloud computing* est étroitement lié à celui du *big data*. Des architectures plus agiles et plus puissantes sont requises pour optimiser les ressources et assurer la capacité des infrastructures à tenir la montée en charge sans faire exploser les dépenses d'investissement et de maintenance. L'hébergement et les opérations critiques (migration, maintenance) peuvent être externalisés ; avec le *cloud*, tout devient service.

Le *cloud* offre aujourd'hui un concept opérationnel qui permet d'accéder, via le réseau et à la demande, à des ressources informatiques de stockage et de calcul virtualisées et mutualisées. Ce concept est aujourd'hui de plus en plus en vogue dans le contexte *du big data* notamment pour y déposer des données y compris dans les usages les plus simples comme le stockage de fichiers en ligne pour les particuliers (TANGHE ET AL., 2016).

Définitions

Pour Syntec Numérique⁸⁸, il s'agit « *d'une interconnexion et une coopération de ressources informatiques, situées au sein d'une même entité ou dans diverses structures internes, externes ou mixtes, et dont le mode d'accès est basé sur les protocoles et standards Internet.* » (BRASSEUR, 2013)

La définition opérationnelle proposée par le NIST (*The National Institute of Standards and Technology*)⁸⁹ est :

« *le cloud computing est un modèle informatique qui permet un accès facile et à la demande par le réseau à un ensemble partagé de ressources informatiques configurables (serveurs, stockage, applications et services) qui peuvent être rapidement provisionnées et libérées par un minimum d'efforts de gestion ou d'interaction avec le fournisseur du service* » (SIMMON, 2018).

Le concept de *cloud computing* selon les définitions est la possibilité pour les entreprises et les individus de partager une même infrastructure informatique pour la gestion et le traitement des différents processus. L'accès aux infrastructures se fait à la demande et en fonction des besoins par le réseau. Mais il faut préciser que le principe d'une informatique à la demande n'est pas nouveau. Il est apparu dans les années 1960 la location des lignes téléphoniques, de répondeurs ou de services informatiques (BRASSEUR, 2013).

Dans le *cloud computing*, les données aussi bien que les applications et les traitements peuvent migrer. La disponibilité est la priorité du *cloud computing*. Il est constitué des cinq caractéristiques, de trois modèles de services et de quatre modèles de déploiement.

Caractéristiques essentielles du cloud computing

Le NIST propose cinq caractéristiques essentielles au *cloud computing* (SIMMON, 2018) qui sont :

⁸⁸ Syntec Numérique, créé en 1969 sous le nom Syntec informatique (pour devenir Syntec Numérique en 2010), est l'un des syndicats professionnels français de l'industrie du numérique. Il regroupe des Entreprises de services du numérique (ESN), des éditeurs de logiciels et des sociétés de Conseil en Technologies. Il fournit les outils, les conseils et les ressources nécessaires à la conception, la mise en œuvre et l'exploitation des nouvelles applications.

⁸⁹ Le *National Institute of Standards and Technology (NIST)* (en français « Institut National des Normes et des Technologies ») est une agence du département du commerce des États-Unis. Son but est de promouvoir l'économie en développant des technologies, la métrologie et des standards de concert avec l'industrie.

- *l'accès au service à la demande* : l'utilisateur doit pouvoir modifier les capacités informatiques, telles que le temps de réaction du serveur ou le stockage en réseau, en fonction de ses besoins, sans interaction humaine quel que soit le fournisseur ;
- *l'accès au réseau large bande* : les capacités sont disponibles sur le réseau et accessibles par le biais de mécanismes standard qui facilitent l'utilisation par des plates-formes hétérogènes clients légers ou lourds ;
- *la mise en commun des ressources* : les ressources informatiques sont mises en commun pour servir plusieurs consommateurs à l'aide d'un modèle à locataires multiples, avec différentes ressources physiques et virtuelles attribuées et réaffectées dynamiquement selon la demande des consommateurs. Il y a un sentiment d'indépendance par rapport à l'emplacement dans la mesure où le client n'a généralement aucun contrôle ou connaissance sur l'emplacement exact des ressources fournies, mais peut être en mesure de spécifier l'emplacement à un niveau d'abstraction plus élevé. C'est le cas dans le domaine de la santé en France avec la certification des hébergeurs de santé qui sont obligés lors du processus de fournir la localisation des serveurs ;
- *le redimensionnement rapide* : les capacités peuvent être provisionnées et libérées rapidement, dans certains cas automatiquement, pour s'étendre rapidement vers l'extérieur et l'intérieur en fonction de la demande. Ces mécanismes permettent d'adapter la puissance de calcul en temps réel en fonction des besoins ;
- *la mesure du service utilisé* : le *cloud computing* contrôle et optimise automatiquement l'utilisation des ressources en exploitant une capacité de comptage à un certain niveau d'abstraction approprié. L'utilisation des ressources peut être surveillée, contrôlée et signalée, ce qui assure la transparence tant pour le fournisseur que pour le consommateur du service utilisé.

Modèles de services du cloud computing

On distingue trois modèles de service (LEAVITT, 2009 ; MELL, GRANCE, 2011 ; VAQUERO ET AL., 2008 ; YOUSEFF ET AL., 2008 ; FIGER, 2012) de *cloud computing* qui sont :

- *Software as a Service (SaaS)* : ce modèle de service est caractérisé par l'utilisation d'une application partagée qui fonctionne sur une infrastructure *cloud*. L'utilisateur accède à l'application par le réseau au travers de divers types de terminaux le plus souvent via un navigateur web. L'administrateur de l'application

ne gère pas et ne contrôle pas l'infrastructure sous-jacente (réseaux, serveurs, applications, stockage). Il ne contrôle pas les fonctions de l'application à l'exception d'un paramétrage de quelques fonctions utilisateurs limitées. L'exemple de *SaaS* dans le domaine de la santé est le dossier patient en ligne à travers d'un navigateur où le médecin a accès à la fonction du logiciel en fonction de ses besoins et ses paramétrages ou du dossier médical partagé mis en place par le gouvernement ;

- *Platform as a Service (PaaS)* : l'utilisateur a la possibilité de créer et de déployer sur une infrastructure *cloud PaaS* ses propres applications en utilisant les langages et les outils du fournisseur. L'utilisateur ne gère pas ou ne contrôle pas l'infrastructure *cloud* sous-jacente (réseaux, serveurs, stockage) mais l'utilisateur contrôle l'application déployée et sa configuration ;
- *Infrastructure as a Service IaaS* : l'utilisateur loue des moyens de calcul et de stockage, des capacités réseau et d'autres ressources indispensables (partage de charge, pare-feu, cache). L'utilisateur a la possibilité de déployer n'importe quel type de logiciel incluant les systèmes d'exploitation. L'utilisateur ne gère pas ou ne contrôle pas l'infrastructure *cloud* sous-jacente mais il a le contrôle sur les systèmes d'exploitation, le stockage et les applications.

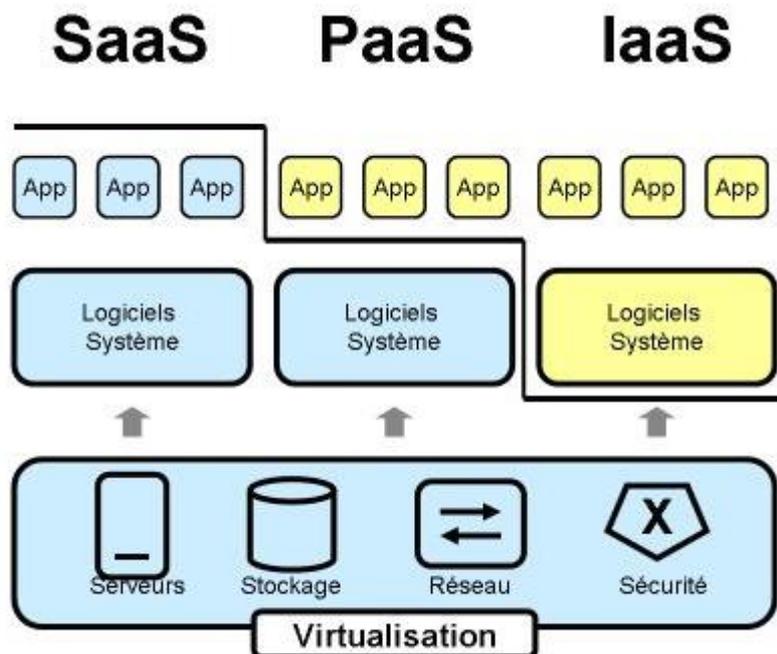


Figure 6 : modèles de service de *cloud computing* (FIGER, 2012)

Modèles de déploiement

Quatre modèles de déploiement sont généralement reconnus (SIMMON, 2018 ; BRASSEUR, 2013). Ces modèles de déploiement ont peu d'influence sur les caractéristiques techniques des systèmes. Ainsi on distingue :

- le *cloud* privé : il est utilisé par une seule organisation en interne pour ses propres besoins. Il peut être géré par l'organisation (*cloud* privé interne) ou par une tierce partie (*cloud* privé externe). L'infrastructure peut être placée dans les locaux de l'organisation ou à l'extérieur ;
- le *cloud* public : il est ouvert au public. Cette infrastructure est possédée par une organisation qui vend des services *cloud* ;
- le *cloud* communautaire : il est partagé par plusieurs organisations pour les besoins d'une communauté qui souhaite mettre en commun des moyens (sécurité, conformité, etc..). Elle peut être gérée par les organisations ou par une tierce partie et peut être placée dans les locaux ou à l'extérieur ;
- le *cloud* hybride : il est composé d'un ou plusieurs modèles qui restent des entités séparées. Ces infrastructures sont liées entre elles par la même technologie qui autorise la portabilité des applications et des données. C'est une excellente solution pour répartir ses moyens en fonction des avantages recherchés.

Avantages et inconvénients du cloud computing

Le *cloud computing* a apporté de nouveaux avantages aux entreprises et aux personnes en termes de réduction des dépenses (pas de dépenses d'immobilisation sur les infrastructures, la maintenance), une gestion simplifiée des infrastructures ou des applications, une grande disponibilité et surtout une montée en puissance à tout moment (BRASSEUR, 2013).

Mais il apporte aussi son lot d'inconvénients qu'est la dépendance du service à la qualité du réseau du fournisseur qui est un élément important dans le domaine de la santé, la perte de la maîtrise de l'implantation des applications et des données (bien que le serveur principal soit situé dans un lieu géographique donné, les systèmes étant distribués, une montée en puissance peut amener à utiliser des serveurs hors du pays d'origine) et la perte du niveau de sécurité car dépendant du fournisseur.

L'absence de standard dans le développement du *cloud computing* (VAQUERO ET AL., 2008) pose un problème non seulement d'interopérabilité et de sécurité. On peut être confronté à une difficulté de transfert de données ou des outils lors d'un changement de prestataire ou d'une plateforme entraînant ainsi un emprisonnement des données ou des outils. Il faut donc gérer la dépendance au prestataire et la pérennité du prestataire.

La question de la propriété des données stockées sur le *cloud* surtout avec l'intervention de prestataires de services pose des problèmes. Il faudra donc s'assurer de la législation en vigueur dans chaque pays avant de s'engager. En France, pour le domaine de la santé, l'hébergement des données de santé est bien encadré par la loi et permet de s'affranchir de ces mesures de sécurité.

La soumission au cadre légal de protection des données dans le pays qui héberge les données est à prendre en compte. C'est le cas par exemple des Etats Unis où le gouvernement peut saisir les données d'une entreprise pour des besoins de la sécurité nationale.

Le *cloud computing* est une réponse pertinente dans le cadre du *big data* dans les entreprises et dans le domaine de la santé. Il permet une montée en puissance en temps réel et une adaptation des besoins de stockage et de puissance. Mais face aux incertitudes sur la sécurité et la confidentialité des données, le choix du transfert vers le *cloud* doit être bien réfléchi.

Utilisation du big data dans le domaine de la santé

Le *big data* pourrait être utile dans le domaine de la santé⁹⁰ pour :

- **la prévention et la prise en charge des maladies**

Les données de santé collectées sur un long temps et stockées dans les bases de données permettront d'identifier des facteurs de risque pour certaines maladies. Ces facteurs serviront ensuite pour construire des messages de prévention, et mettre en place des programmes à destination des populations à risque.

Le *big data* permet le développement de systèmes d'aide au diagnostic et aux traitements. Ces systèmes se fondent sur le traitement de grandes masses de données cliniques individuelles. En effet, les données individuelles sont très importantes pour le test des systèmes mis en place et pour leur entraînement dans le cadre de l'apprentissage machine.

⁹⁰ INSERM. Big data en santé. Source <https://www.inserm.fr/information-en-sante/dossiers-information/big-data-en-sante> consulté le 13 juin 2018

Elles permettent également de surveiller l'efficacité et l'effet des traitements en pratique quotidienne, ce que les essais cliniques ne permettent pas.

- **Prédire des épidémies**

Disposer de nombreuses informations sur l'état de santé des individus dans une région donnée permet de repérer l'augmentation de l'incidence des maladies ou des comportements à risque, et d'alerter les autorités sanitaires.

En France, depuis 1984, le *réseau Sentinelles* suit plusieurs maladies infectieuses. Il alerte sur les épidémies grâce à la contribution de 1 300 médecins généralistes et d'une centaine de pédiatres répartis sur tout le territoire. Ces derniers rapportent au moins une fois par semaine le nombre de cas observés pour sept maladies transmissibles (diarrhée aiguë, maladie de Lyme, oreillons, syndromes grippaux, urétrite masculine, varicelle et zona) ainsi que les actes suicidaires.

- **Améliorer la pharmacovigilance**

L'analyse des données issues de cohortes ou des bases médico-économiques sur le long terme peut donc permettre d'observer beaucoup de phénomènes, et notamment de faire des rapprochements entre des traitements et la survenue d'événements en santé. Cette pratique permet de repérer des événements indésirables graves et d'alerter sur certains risques. En 2013, la base de données du SNIIRAM avait ainsi permis d'étudier le risque d'accident vasculaire cérébral et d'infarctus du myocarde chez les femmes utilisant une pilule contraceptive de 3^e génération (CNAMTS ET AL., 2013).

- **Gestion de flux des patients**

Des applications de gestion de flux de patients comme celles utilisées dans les entreprises sont exploitées dans les hôpitaux. Ces applications permettent d'assurer une utilisation optimale des ressources disponibles. Depuis fin 2016, quatre des hôpitaux qui constituent l'Assistance Publique-Hôpitaux de Paris (AP-HP) ont décidé d'utiliser le *big data* et le *machine learning* pour prédire le nombre de visites par jour ainsi que les jours et les heures d'admission. L'objectif est de pouvoir déployer le nombre d'employés adéquat lorsqu'il y a des pics de fréquentation.

L'utilisation du *big data* dans le domaine de la santé pourrait apporter des avancées importantes dans la qualité des soins et l'organisation des systèmes de santé à condition que les bons moyens soient mis en place pour son exploitation. Ainsi, pour l'exploitation des données dans le cadre du *big data*, l'IA est un atout important pour l'exploitation des masses de données. Mais qu'est que l'IA et quels sont les méthodes et les outils utilisés pour la mettre en œuvre ?

3.2 Intelligence artificielle

L'IA est une discipline jeune d'une soixante d'années, qui est un ensemble de sciences, théories et techniques (notamment logique mathématique, statistiques, probabilités, neurobiologie computationnelle, informatique) qui ambitionne d'imiter les capacités cognitives d'un être humain voire de les dépasser (IA autonome).

Initiés dans le souffle de la Seconde Guerre mondiale, ses développements sont intimement liés à ceux de l'informatique et ont conduit les ordinateurs à réaliser des tâches de plus en plus complexes, qui ne pouvaient être auparavant que déléguées à un humain. L'IA s'est inscrite au carrefour de nombreuses disciplines scientifiques auxquelles elle a emprunté des concepts. Chacune d'elles a pu s'enrichir en retour de cette interaction (CONSEIL DE L'EUROPE, 2014).

3.2.1 Généralités

Le mot Intelligence vient du latin *intellegentia* (faculté de comprendre), dérivé du latin *intellegere* signifiant comprendre, et dont le préfixe *inter* (entre), et le radical *legere* (choisir, cueillir) ou *ligare* (lier) suggèrent principalement l'aptitude à relier des éléments qui sans elle resteraient scindés.

Définir l'intelligence est complexe. L'idée d'intelligence semble présente dans de nombreuses cultures. L'intelligence est multiple, et tous les auteurs s'accordent sur le fait qu'il n'y a pas une mais des intelligences, et que chacun d'entre nous peut présenter des forces et/ou des faiblesses dans les différentes formes d'intelligences (MATHIVET, 2014).

Pour Ibn Khaldoun⁹¹, l'intelligence comporte plusieurs branches (KHALDOUN, 2006) :

- l'intelligence discernante ou tactique ;
- l'intelligence expérimentale ou stratégique ;
- l'intelligence spéculative, scientifique ou politique.

Au moins cinq composantes de l'intelligence, regroupées sous le terme de facultés cognitives sont nécessaires à l'homme pour mémoriser (apprendre), analyser (comprendre) et communiquer (partager), les informations externes⁹² :

⁹¹ Ibn Khaldoun (732-808 de l'hégire, 1332-1406 de J.C.) Historien, philosophe, sociologue, juge, enseignant, poète, aussi bien qu'homme politique.

⁹² Source : <http://www.histophilo.com/intelligence.php> consulté le 25 avril 2019.

- « *l'attention ou distinction est une faculté de faire attention et distinguer des choses ou actes existants dans l'environnement ;*
- *la concentration est une faculté permettant d'évaluer les distances, de se représenter volumes et mouvements par représentation mentale ;*
- *la conscience ou compréhension est une faculté de comprendre les problèmes et les actes le plus souvent ;*
- *le raisonnement ou pensée est la faculté de planifier. Elle est aussi nommée créativité ;*
- *l'humour est une preuve qu'on s'adapte aisément socialement ».*

L'hérédité et l'environnement semblent jouer un rôle dans la détermination des aptitudes d'un individu, il est complexe de cerner dans quelle mesure chacun de ces facteurs joue un rôle, et comment ils s'articulent entre eux.

Dans les formes de l'intelligence (GARDNER, 1997) Howard Gardner, professeur de psychologie cognitive à Harvard a émis l'hypothèse selon laquelle plusieurs types d'intelligence coexisteraient chez chaque être humain. Il distingue neuf types d'intelligence qui sont :

- *l'intelligence logico-mathématique* : elle représente la capacité à travailler à l'aide de nombres et de la géométrie, à analyser les situations, à mettre au point des raisonnements. Elle est mise en avant chez les scientifiques, en particulier en mathématiques et en physique ;
- *l'intelligence linguistique (ou verbale)*, fréquente chez les politiciens, écrivains, poètes est l'aptitude à construire des phrases, à avoir le sens de la rime, convaincre et persuader les autres ;
- *l'intelligence intrapersonnelle*, qui sert à désigner la capacité qu'on a à avoir un regard critique sur soi-même, à juger de ses limites, à comprendre ses réactions, à connaître ses forces et ses faiblesses ;
- *l'intelligence interpersonnelle* est la capacité à comprendre les autres et à réagir de façon adéquate. Elle est liée à la notion d'empathie, à la tolérance, à la sociabilité. Mais elle peut permettre aussi de manipuler les autres ;
- *l'intelligence visio-spatiale*, qui sert à manipuler des objets ou des environnements tridimensionnels dans sa tête, et est utilisée pour suivre une carte, se souvenir d'un

chemin ou imaginer ce que donne une forme dans l'espace à partir du plan. Elle est exploitée surtout par les architectes, géographes, artistes ;

- *l'intelligence naturaliste*, qui sert à trier, organiser, hiérarchiser les objets, et de les différencier en catégories. Particulièrement sollicitée chez les zoologistes, botanistes, archéologues ;
- *l'intelligence musicale*, qui juge la hauteur, la tonalité des sons, le rythme et la mélodie d'une musique. Exploitée en particulier par les musiciens, compositeurs ;
- *l'intelligence kinesthésique*, qui est responsable de tout ce qui a trait aux travaux gestuels. C'est la capacité à avoir une représentation mentale de son corps dans l'espace et à pouvoir mener un mouvement particulier. Développée chez les sportifs, danseurs, mimes, mais également ceux qui pratiquent un travail de minutie (chirurgiens, ...) ;
- *l'intelligence existentialiste ou spirituelle* se définit par l'aptitude à se questionner sur le sens de la vie. Elle se rapproche de la notion de moralité.

L'intelligence n'est pas facile à mesurer, ni à définir car elle couvre un grand nombre de domaines. En fonction des définitions, certaines formes d'intelligence sont sous-estimées voire ignorées.

Pour Jean Piaget, l'intelligence dérive de la capacité d'adaptation du sujet à son milieu. Il distingue deux types d'intelligence (PIAGET, 1998) :

- l'intelligence pratique ou sensori-motrice ;
- l'intelligence intériorisée (verbale ou réfléchie).

« *L'intelligence ne commence ni par la connaissance du moi, ni par celle des choses comme telles, mais par celle de leur interaction.... Elle organise le monde en s'organisant elle-même* » (PIAGET, 1998). Cette définition large de la définition de l'intelligence permet en quelque sorte de prendre en compte toutes les approches de définitions.

Définition de l'intelligence artificielle

On parle de plus en plus d'IA, un ensemble de technologies très évoluées utilisées dans divers domaines et déployées dans de nombreux services, systèmes et applications. L'IA regroupe des sciences et technologies qui permettent d'imiter ou d'étendre l'intelligence humaine à l'aide de machines capables d'apprentissage, de prise de décisions voire de

raisonnement. Toutefois l'IA est très difficile à définir. Il existe plusieurs définitions dont celle des pères fondateurs de l'IA.

Marvin Lee Minsky, l'un des créateurs de l'intelligence artificielle, la définit comme « *la construction de programmes informatiques qui s'adonnent à des tâches qui sont, pour le moment, accomplies de façon plus satisfaisante par des êtres humains car elles demandent des processus mentaux de haut niveau tels que : l'apprentissage perceptuel, l'organisation de la mémoire et le raisonnement critique* »⁹³.

Même si elles respectent globalement la définition de Minsky, il existe un certain nombre de définitions différentes de l'IA qui fluctuent sur deux points fondamentaux (RUSSELL, NORVIG, 2016) :

- les définitions qui lient la définition de l'IA à un aspect *humain* de l'intelligence, et celles qui la lient à un modèle parfait d'intelligence, non nécessairement humaine, appelée *rationalité*.
- les définitions qui insistent sur le fait que l'IA a pour but d'*avoir l'ensemble des apparences* de l'intelligence (humaine ou rationnelle), et celles qui insistent sur le fait que le *fonctionnement interne* du dispositif d'IA doit ressembler aussi à celui de l'être humain ou être rationnel.

L'IA peut être envisagée à partir de deux approches essentielles que sont :

- par le processus de la pensée et du raisonnement ou par le comportement ;
- une évaluation par rapport à la performance humaine ou une évaluation par rapport à un concept idéal de l'intelligence.

Ces deux principales approches permettent de définir plusieurs dimensions de l'IA comme le montre la figure ci-dessous.

⁹³ Intelligence artificielle. Source : http://www.histophil.com/intelligence_artificielle.php#cite_ref-1 consulté 29 avril 2019.

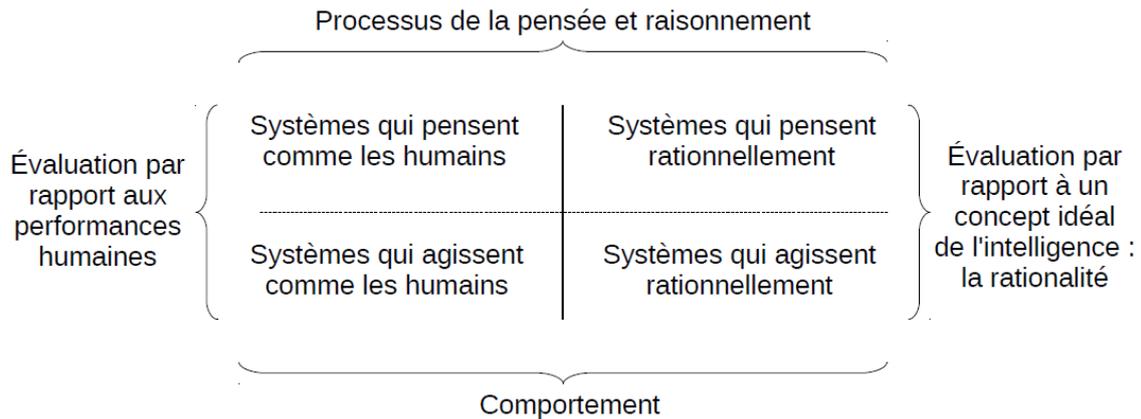


Figure 7 : approches de la définition de l'IA en quatre catégories⁹⁴

Les chercheurs en IA tentent à terme de développer - artificiellement donc - cette aptitude. Ils recherchent ainsi des moyens pouvant permettre à l'ordinateur d'accomplir des tâches réalisées par les humains et d'acquérir des capacités intellectuelles comparables voire supérieures à celles de l'homme. Le but des laboratoires en IA est d'accroître les performances de l'ordinateur dans des activités où l'humain réussit normalement mieux.

Fondements de l'intelligence artificielle

L'IA s'appuie sur de nombreuses disciplines scientifiques et sur des techniques qui s'améliorent au fil du temps. La conceptualisation de l'IA s'appuie également sur plusieurs travaux et inventions antérieurs comme la définition du syllogisme par Aristote en 384 av. JC., la création par Héron d'Alexandrie des premiers automates au Ier siècle après JC ou encore l'invention de l'Ars Magna de Ramon Lulle en 1275. Ces disciplines et travaux ont été les fondements de la mise en place et de l'évolution – idées, perspectives et techniques - de l'IA (RUSSELL, NORVIG, 2010) :

- *« philosophie (rationalisme et logique) : peut-on utiliser des règles formelles pour tirer des conclusions valides ? Quel rapport l'esprit entretient-il avec la nature physique du cerveau ? D'où provient la connaissance ? Comment la connaissance conduit-elle à l'action ?*
- *psychologie (modélisation de la cognition) : comment les hommes et les animaux pensent-ils et agissent-ils ? Quels sont les rapports entre le langage et la pensée ?*

⁹⁴ JÉRÔME CHAMPAVÈRE. Introduction à l'intelligence artificielle. Source : <http://www.grappa.univ-lille3.fr/~champavere/?page=Enseignement> consulté le 27 avril 2019.

- *mathématiques (algorithmique, calculabilité et probabilité) : quelles sont les règles formelles qui permettent de tirer des conclusions valides ? Qu'est ce qui peut être calculé ? Comment raisonne-t-on à partir d'informations incertaines ?*
- *économie (théorie de la décision et recherche opérationnelle) : comment prendre des décisions qui maximisent les gains ? Comment faire quand les autres risquent de ne pas coopérer ? Comment y parvenir alors que les gains sont susceptibles d'être éloignés dans le futur ?*
- *neurosciences (cerveau ordinateur) : quels sont les mécanismes qui permettent au cerveau de traiter l'information ?*
- *linguistique (traitement automatique de la langue, représentations des connaissances) : comment faire pour représenter les connaissances ?*
- *ingénierie informatique : comment construire un ordinateur performant ?*
- *théorie du contrôle et cybernétique : comment faire en sorte que des artefacts opèrent de façon autonome ? »*

Histoire de l'intelligence artificielle

L'histoire de l'IA se déroule parallèlement à celle de l'informatique durant le 19e et le 20e siècle. Elle débute avec les travaux d'Ada Lovelace (1815-1852) qui développa le premier algorithme sur la machine analytique de Charles Babbage (1791-1871). Ce dernier eut l'idée d'incorporer les cartes du métier à tisser de Joseph Marie Jacquard (1801) qui est considéré comme la première machine programmable⁹⁵ (JARRY ET AL., 2018). Mais le terme « *artificial intelligence* » ou « intelligence artificielle » est né durant une école d'été en 1956 à *Dartmouth College* USA qui réunissait des chercheurs célèbres comme John McCarthy, Marvin Minsky, Alan Newell, Herbert Simon (TISSEAU, 1996 ; GANASCIA, 1999).

L'histoire de l'IA a évolué en dents de scie lors des cinquante dernières années. Elle a connu des périodes fastes entre 1957 et 1973, des moments de disette jusqu'au début des années 1980 et de désillusion durant la décennie 1990, le matériel ne suivant pas les besoins. La réduction des investissements destinés au développement de l'IA a suivi les mauvais résultats et les attentes décevantes. Mais portés par la digitalisation de l'économie depuis les

⁹⁵ JOEL COLLOC. De l'Intelligence Artificielle mimétique aux systèmes autonomes : quels objectifs et quels enjeux éthiques dans JARRY, Alban, RAVIX, Valéry, ETIENNE, Loïc, COLLOC, Joël et ZOUARHI, Sajida, 2018. La datasphère et ses applications numériques. In : Vademecum sur le traitement des données numériques : réflexions et bonnes pratiques éthiques. Saint-Herblain : ADEL. p. 42-73.

années quatre-vingt-dix, les recherches ont repris de plus belle et les programmes intelligents sont aujourd'hui incontournables dans de nombreux domaines.

Aperçu de l'évolution de l'intelligence artificielle

L'année 1950 est un tournant décisif dans l'histoire de l'IA, avec la publication de l'article « *Computing Machinery and Intelligence* » d'Alan Turing (TURING, 1950). Dans son article, l'auteur pose les fondements de l'IA moderne et y décrit le test de Turing qui met au défi des testeurs (hommes et femmes) de reconnaître si leur interlocuteur est un humain ou une machine, grâce à l'analyse des réponses à une série de questions posées au travers d'une interface masquant l'interlocuteur.

Warren McCulloch et Walter Pitts en 1943 (MCCULLOCH, PITTS, 1943) menèrent les premiers travaux de l'IA à partir de trois sources : l'état du savoir sur la physiologie de base et la fonction des neurones dans le cerveau, l'analyse formelle de la logique propositionnelle de Russell et Whitehead, et la théorie de calcul de Turing (RUSSELL, NORVIG, 2010). Ils proposèrent un modèle de neurones artificiels dans lequel chaque neurone est caractérisé par un état « *marche* » (excité stimulant un potentiel d'action) ou « *arrêt* » ; le passage à l'état « *marche* » se produisant en réponse à une stimulation émise par un nombre suffisant de neurones voisins. McCulloch et Pitts ont également suggéré que des réseaux de neurones définis de manière appropriée sont capables d'apprentissage.

Donald Hebb, en 1949, a découvert la plasticité neuronale (la faculté du cerveau à se réorganiser, à former de nouvelles connexions entre les neurones, et/ou d'activer de nouveaux neurones, de nouvelles fonctions pour les neurones) (HEBB, 2005 ; SPATZ, 1996), appelée « *apprentissage hebbien* » dans lequel l'efficacité synaptique augmente lors d'une stimulation présynaptique répétée et persistante de la cellule postsynaptique et qui demeure un modèle très influent aujourd'hui (RUSSELL, NORVIG, 2010).

Vers la fin des années 1950, John McCarthy et Fernando Corbató créent la technique du temps partagé et le langage LISP⁹⁶ (TURNER, 2013) qui permettent l'utilisation partagée et

⁹⁶Le langage LISP « *list processing* » ou « traitement de listes » fut inventé par John McCarthy en 1958 alors qu'il était au *Massachusetts Institute of Technology* (MIT). LISP est la plus ancienne famille de langages de programmation à la fois impératifs et fonctionnels. Développé initialement en tant que modèle pratique pour représenter des programmes (par contraste avec la notion théorique de machine de Turing), il est devenu dans les années 1970 et 80 le langage de choix pour la recherche en intelligence artificielle.

simultanée des performances d'un ordinateur grâce à l'usage des temps morts des calculs d'un utilisateur au profit d'un second. John McCarthy, un des organisateurs de la conférence de Dartmouth en 1956, est considéré avec Marvin Lee Minsky comme le pionnier de l'IA. Il représente le courant qui privilégie la logique symbolique. Le premier programme de l'IA développé par Newell, Shaw et Simon est sans doute le *Logic Theorist* dont le but est la démonstration automatique des théorèmes. Ils développent conjointement un système de résolution de problème général (*GPS : General Problem Solver*) fondé sur des techniques dites d'analyse moyen - fin (ALLIOT ET AL., 2003).

Après les années 1950, l'essor de l'IA va aboutir à l'invention des systèmes experts⁹⁷. Dans les années 1960-1970, des programmes pouvant simuler une tâche précise d'un expert humain, programmes que l'on retrouve dans de très nombreuses applications industrielles quotidiennes. La programmation de telles connaissances demandait en réalité beaucoup d'efforts (problèmes de cohérence de règles contradictoires et/ou de l'incomplétude de la base des règles) et à partir de 200 à 300 règles, il y avait un effet « *boîte noire* » où l'on ne savait plus bien comment la machine raisonnait. La mise au point et la maintenance devenaient ainsi extrêmement problématiques et, surtout, on arrivait à faire plus vite et aussi bien avec d'autres manières moins complexes, moins chères. Durant cette période, l'IA va se spécialiser et se théoriser. JA Robinson (ROBINSON, 1965) développe le principe de résolution qui sera à la base de la réalisation du langage PROLOG, des systèmes de résolutions automatiques.

Après les années 1970, les investisseurs, notamment la *Defense Advanced Research Projects Agency (DARPA)*⁹⁸ aux Etats Unis, avaient commencé à réduire leurs budgets, parce que les résultats n'étaient pas à la hauteur de leurs attentes. Le Japon⁹⁹, les Etats Unis¹⁰⁰ et la Grande Bretagne se lancèrent dans de vastes projets d'IA sur les recherches sur les architectures des

⁹⁷ Le premier système expert fut *Dendral* en 1965, créé par les informaticiens Edward Feigenbaum, Bruce Buchanan, le médecin Joshua Lederberg et le chimiste Carl Djerassi. Il permettait d'identifier les constituants chimiques d'un matériau à partir de spectrométrie de masse et de résonance magnétique nucléaire, mais ses règles étaient mélangées au moteur.

⁹⁸ La *DARPA* a été à l'origine du développement de nombreuses technologies qui ont eu des conséquences considérables dans le monde entier dont les réseaux informatiques notamment l'ARPANET qui est devenu Internet.

⁹⁹ En 1981, les Japonais ont lancé le projet Cinquième génération, plan décennal qui prévoyait de construire des ordinateurs intelligents programmés en Prolog.

¹⁰⁰ Les Etats Unis ont créé un consortium de recherche MCC (*Microelectronics and Computer Technology Corporation*) pour préserver leur compétitivité dans le domaine de l'IA.

puces et sur les interactions homme - machine ; mais les projets n'ont jamais atteint leurs ambitieux objectifs dans ces trois pays.

L'IA devient une industrie dans les années 1980. Le premier système expert commercial réussi, R1¹⁰¹, est né chez Digital Equipment Corporation. Presque toutes les grandes entreprises des Etats Unis possédaient un département IA et utilisaient des systèmes experts ou envisageaient d'en utiliser pendant cette période. L'industrie de l'IA s'est développée avec des centaines d'entreprises qui vendaient des systèmes experts, des systèmes de vision artificielle et des robots (RUSSELL, NORVIG, 2010).

En France, à cette époque, l'IA devient une nouvelle branche de l'informatique, les pionniers furent Jacques Pitrat¹⁰², Jean-Louis Laurière¹⁰³, Suzanne Pinson¹⁰⁴, Alain Colmerauer¹⁰⁵, Henri Farreny¹⁰⁶, Olivier Gascuel¹⁰⁷, Henri Prade¹⁰⁸, Alain Bonnet, Jean-Paul Haton et Jean-Michel Truong Ngoc¹⁰⁹ et bien d'autres.

Dans les années 1990, le terme d'IA était presque devenu tabou et des déclinaisons plus pudiques étaient même entrées dans le langage universitaire, comme « *informatique avancée* » (CONSEIL DE L'EUROPE, 2014). L'IA va se scinder en plusieurs branches que l'on peut schématiser ainsi (ALLIOT ET AL., 2003) :

- le traitement du langage naturel ;
- la démonstration automatique des théorèmes ;

¹⁰¹ Le programme R1 configurait des ordinateurs en fonction des commandes des clients.

¹⁰² PITRAT JACQUES, la genèse des systèmes experts, Bull INRIA, 1984, n°97, 2-8. Et, Pitrat Jacques, métaconnaissance, futur de l'intelligence artificielle, Hermès, 1990.

¹⁰³ LAURIÈRE JEAN-LOUIS, un langage et un programme pour énoncer et résoudre les problèmes de combinatoire, Thèse d'état Paris VI, 1976.
LAURIÈRE JEAN-LOUIS, intelligence artificielle, résolution de problèmes par l'homme et la machine, ed. Eyrolles, 1986, 473p.

¹⁰⁴ PINSON SUZANNE, représentation des connaissances dans les systèmes experts, RAIRO, Informatique, 1981, 15 :4, 343-367.

¹⁰⁵ COLMERAUER ALAIN, précédences, analyse syntaxique et langages de programmation, Thèse d'état, Grenoble, 1967.

¹⁰⁶ FARRENY HENRI, les systèmes experts principes et exemples, ed. Cepadues, Toulouse, 1985, 254 p.

¹⁰⁷ GASCUEL OLIVIER, SAM : un système expert dans le domaine médical, Thèse 3e cycle, Paris VI, Nov 1981.

¹⁰⁸ PRADE HENRI, modèles mathématiques de l'imprécis et de l'incertain en vue d'applications au raisonnement naturel, Thèse d'état, Toulouse, 1982.

¹⁰⁹ BONNET ALAIN, HATON JEAN-PAUL, TRUONG NGOC JEAN-MICHEL, systèmes experts vers la maîtrise technique, Inter Editions, Paris, 1986, 288 p.

- les jeux ;
- la résolution des problèmes généraux ;
- la résolution des problèmes experts ;
- la représentation des connaissances ;
- la perception ;
- l'apprentissage ;
- de façon générale tout problème que l'on ne sait pas résoudre informatiquement de façon classique.

La fin des années 1990 fait apparaître des avancées notables de l'IA, avec le développement d'Internet, des moteurs de recherche et surtout des technologies hardware, à présent, suffisamment performantes et disposant de capacités de stockage sans précédent. L'IA entre ensuite dans une ère de perfection, avec des machines de plus en plus performantes dans la résolution des problèmes. En 1997, le champion du monde Russe Garry Kasparov est battu par un supercalculateur d'*IBM* appelé *Deep Blue*, après deux matchs de six parties. Un événement qui marquera les esprits et donnera ses lettres de noblesse à la renaissance de l'IA moderne. Le développement des technologies intelligentes a amorcé une accélération fulgurante à partir des années 2000, grâce à la croissance exponentielle de la puissance de calcul et de stockage des machines.

On assiste aujourd'hui à l'explosion des technologies et des applications révolutionnaires concrètes telles que la reconnaissance d'images, la colorisation automatique d'images, le sous-titrage automatique de centaines de millions de vidéos *YouTube* ou encore le développement de voitures autonomes ou de la robotique. Les avancées sont accélérées du fait d'un plus grand usage de la méthode scientifique dans l'expérimentation et de la comparaison des approches. La *DARPA* dédiée à la recherche et développement militaire a depuis repris ses investissements dans l'IA, ce qui prouve le nouvel intérêt qu'y voient certains gouvernements ne serait-ce qu'en matière de défense (cybersécurité, combattants artificiels, armes autonomes) (IN PRINCIPIO, 2017). Le regain d'intérêt actuel pour l'IA se justifie sans conteste par des symboles forts de succès qu'elle enregistre régulièrement (*IBM*

Watson 2011, DeepMind¹¹⁰ au jeu de Go 2016), par l'avènement du web, du *cloud*, du *big data* et par la démultiplication des applications business concrètes qui fleurissent à présent.

Les facteurs qui expliquent le nouvel élan de l'IA :

- l'accès à des volumes importants de données aussi variées pour entraîner et tester les algorithmes ;
- l'amélioration de la puissance de calcul des ordinateurs pour accélérer le calcul des algorithmes d'apprentissage associée à une réduction importante des coûts d'acquisition ;
- la capacité de distribution des calculs sur un grand nombre de machines connectées en réseaux ;
- une amélioration des méthodes scientifiques et des algorithmes de programmation.

Histoire de l'intelligence artificielle dans le domaine de la santé

L'IA en médecine est née dans les années soixante avec les travaux de Ledley et Lusted en 1959 (LEDLEY, LUSTED, 1959) et Warner en 1964 (WARNER ET AL., 1964) ; les travaux portaient essentiellement sur la manière d'utiliser l'informatique pour élaborer un diagnostic. Les approches impliquant la théorie des ensembles flous, les réseaux bayésiens et les réseaux de neurones artificiels ont été appliqués aux systèmes informatiques intelligents d'aide à la décision en santé (IN PRINCIPIO, 2017).

Dans les années 1970, à la suite du premier système expert, *Dendral* en chimie organique cité précédemment, une équipe a créé son successeur immédiat : le système *MYCIN* (système spécialisé dans le diagnostic des maladies du sang et la prescription de médicaments notamment les antibiotiques) (BUCHANAN, SHORTLIFFE, 1984), considéré comme l'un des plus importants usages pionniers de l'IA en médecine. A partir de *MYCIN* a été proposé le générateur *EMYCIN* qui a permis la réalisation de toutes une génération de systèmes qui n'ont cependant pas intégré le quotidien des professionnels de santé (COLLOC, 2000).

¹¹⁰ Google DeepMind est la division de Google dédiée à l'intelligence artificielle. Elle développe des algorithmes de *Deep Learning* permettant à des agents IA d'apprendre de manière autonome sans avoir été programmés au préalable. L'objectif de DeepMind est de créer des machines intelligentes capables d'apprendre et d'acquérir de nouvelles connaissances à la manière des êtres humains. Pour ce faire, DeepMind développe de puissants algorithmes d'apprentissage général pouvant être combinés pour créer un système IA, aussi appelé agent IA.

La médecine a constitué un domaine d'investigation particulièrement riche pour l'expérimentation de nouveaux systèmes d'apprentissage. Elle a ainsi contribué à l'élaboration de nouveaux modèles s'appuyant sur la maîtrise de la complexité de l'organisme humain et, de manière plus générale, sur la nature et le fonctionnement des êtres vivants. Ces modèles ont ensuite largement été utilisés pour réaliser des applications dans d'autres domaines.

La complexité du corps humain, l'évolution considérable des moyens de diagnostic, d'investigation et de soins font de la médecine un immense champ de recherche pour l'informatique et l'IA. En effet, les techniques utilisées proviennent des découvertes les plus récentes de toutes les sciences fondamentales. L'informatique est présente dans la plupart des dispositifs d'imagerie ou d'exams paracliniques utilisés actuellement en médecine.

Les années 1980 et 1990 ont entraîné la prolifération du micro-ordinateur et de nouveaux niveaux de connectivité réseau. Pendant cette période, les chercheurs et les développeurs ont reconnu le fait que les systèmes d'IA dans le soin doivent être conçus pour pallier l'insuffisance des données et appuyer l'expertise des médecins.

Les progrès technologiques et médicaux ont permis la croissance des soins de santé (scanner, échographie, automatisation des examens biologiques) liés à des applications de l'IA, comprennent :

- des améliorations dans la puissance de calcul rendent plus rapide la collecte de données et le traitement de données ;
- une augmentation du volume et de la disponibilité des données liées à la santé des particuliers et des soins de santé liés à des dispositifs ;
- une amélioration des capacités du séquençage génomique et du traitement des données génomiques ;
- une mise en œuvre généralisée du dossier de santé électronique ;
- des améliorations dans le traitement du langage naturel et de la vision par ordinateur, permettent à des machines de reproduire le processus humain de perception ;
- des améliorations de la précision de la chirurgie robotisée.

Différentes approches de l'intelligence artificielle

L'importance et l'utilité de l'IA sont aujourd'hui incontestables, son aide est déjà inestimable dans le domaine médical (reconnaissance des troubles du rythme à l'électrocardiogramme ou des images tumorales en imagerie médicale), le secteur automobile (conduite automatique) ou encore l'exploration spatiale avec les robots. Ces technologies ont largement montré leur capacité à rivaliser avec l'intelligence humaine et à dépasser les limites physiques de l'homme, grâce à leur puissance et à leur rapidité de calcul hors norme et surtout inédite dans l'histoire moderne. L'évolution de ces programmes innovants et révolutionnaires a conduit à une distinction entre les méthodes d'IA.

IA symbolique

L'IA symbolique¹¹¹ est fondée sur la modélisation du raisonnement logique, sur la représentation et la manipulation de la connaissance par des symboles formels. Elle utilise une approche top-down, s'appuyant sur l'intersection entre logique philosophique et mathématique. Elle reproduit le plus fidèlement possible un comportement spécifique prévu à l'avance, mais pas son fonctionnement. Le principe de compréhension ne s'applique pas à la machine, le système imite un comportement intelligent dans un domaine précis et peut résoudre des problèmes.

Il s'agit d'une forme d'approche pragmatique d'ingénieur : chercher à construire des dispositifs de plus en plus autonomes (pour diminuer le coût de leur supervision), des algorithmes capables de résoudre des problèmes d'une certaine classe. Mais, cette fois, la machine simule l'intelligence, elle semble agir comme si elle était intelligente. Ces programmes parviennent à imiter de façon grossière le comportement d'humains face à d'autres humains lors d'un dialogue appelé souvent « *test de Turing* » (TURING, 1950).

L'IA symbolique s'inscrit dans la droite succession de ce qu'a été la recherche opérationnelle dans les années 1960, le *process control* dans les années 1970, l'aide à la décision dans les années 1980 et le *data mining* dans les années 1990. Et, qui plus est, avec une certaine continuité. Il s'agit en particulier d'intelligence humaine reconstituée, et de programmation d'un apprentissage. Quoi que la machine donne l'impression d'être intelligente, elle simule l'intelligence. Tout se passe seulement, pour l'observateur extérieur, « *comme si* » elle

¹¹¹ Source : http://www.histophilo.com/intelligence_artificielle.php consulté le 08 mai 2018.

comprenait, mais elle ne fait que « *délivrer* » comme dans la métaphore de la « *chambre chinoise* » (SEARLE, 1980).

L'IA symbolique n'est pas conçue pour évoluer seule. Ce ne sont que des programmes informatiques, certes pratiques, mais seulement prédéterminés dans des tâches précises et spécifiques. Ces programmes nécessitent une mise à jour régulière en fonction de l'évolution des connaissances ; les mises à jour sont très importantes notamment dans la pratique médicale car les connaissances évoluent d'une manière exponentielle (par exemple : l'apparition de nouveaux virus, la découverte de nouvelles molécules avec la mise en place de nouveaux protocoles ou la mise en place de nouveaux procédés de diagnostic).

IA connexionniste

Le connexionnisme est un courant de recherche assez vaste qui constitue une voie originale dans l'étude des phénomènes cognitifs. Les modèles connexionnistes utilisent les réseaux de neurones, dont l'organisation et le fonctionnement rappellent, à un certain niveau d'abstraction, les systèmes neuronaux physiologiques. Ces modèles cherchent à faire le lien entre le fonctionnement du cerveau et celui de l'esprit, en proposant des mécanismes qui pourraient expliquer certains aspects de la cognition humaine (VICTORRI, 2008).

La diversité des techniques connexionnistes et de leurs applications entraîne pas mal de confusion sur les buts, les méthodes, et les limites éventuelles des réseaux neuronaux. L'analogie biologique est délibérée, mais le modèle est manifestement très idéalisé. Les unités de traitement sont très simplifiées par rapport à des neurones réels, et la ressemblance est moins convaincante dans le détail qu'au niveau de l'organisation générale du traitement (MEMMI, 1990).

Quelles que soient les approches utilisées dans le développement des applications de l'IA, ces approches ne sont pas totalement contradictoires, et peuvent être complémentaires pour résoudre des problèmes.

Grands domaines et applications de l'intelligence artificielle

L'IA est un domaine très vaste avec de nombreuses applications. Trois types de domaines peuvent être distingués (FALTINGS, SCHUMACHER, 2017 ; RUSSELL, NORVIG, 2010) :

- les programmes qui imitent des capacités cognitives et reproduisent un raisonnement humain, par exemple dans le cas du diagnostic médical ou la recherche de régularités dans une base de données ;
- des programmes qui imitent des capacités sensorielles et sont capables de reconnaître des formes ou des objets, ou bien de comprendre la parole en langage naturel, par exemple la traduction automatique, la reconnaissance des images pathologiques en médecine ou la reconnaissance vocale ;
- des programmes qui imitent des capacités sensori-motrices et sont capables de réagir de façon autonome à leur environnement, par exemple les robots ou les agents autonomes.

L'IA se trouve pratiquement dans tous les domaines informatisés de nos jours. Elle nous entoure et nous l'utilisons parfois sans nous en rendre compte. Elle est en plein essor et son développement est en pleine croissance du fait de la disponibilité des données, des capacités des ordinateurs et surtout des espoirs d'économies qu'elle nous permettra de réaliser dans les domaines dans lesquels elle sera utilisée. Nous décrivons les mécanismes sous-jacents afin de montrer comment nous les exploitons dans notre travail.

Pour analyser les quantités importantes des données produites dans le cadre des soins (données structurées et non structurées surtout), nous présentons un ensemble de méthodes de traitement utilisées en IA.

3.2.2 Méthodes utilisées dans le développement de l'IA

Notre objectif n'est ni d'être exhaustif, ni de décrire des formules mathématiques. Il s'agit de décrire le fonctionnement et les principales utilisations de ces méthodes.

Systèmes experts

« Les systèmes experts ou systèmes à base de règles (de production) sont des systèmes à base de connaissances dans lesquels les contenus de la base de connaissances se réfèrent au

savoir et au savoir-faire des hommes de métier » (GANASCIA, 2006). Ce contenu ne correspond plus à la spécification d'un problème abstrait, mais à l'expertise des professionnels, à leurs astuces, à leur coup d'œil et à leur habilité.

La tâche principale d'un système à base de connaissance est d'exploiter des connaissances d'un domaine pour (aider à) résoudre un problème donné, d'identification, de classification, de reconnaissance, de diagnostic, de planification,... (LE BER ET AL., 2006). Les systèmes experts constituent un moyen privilégié de coder des connaissances d'experts. Il s'avère qu'une bonne partie des éléments d'expertises courants et utiles en résolution des problèmes peuvent s'exprimer sous la forme de règles de production lorsque la déduction est employée comme mode de raisonnement.

Fondamentalement, les inférences réalisées par les systèmes experts sont régies par les principes de la logique, mais c'est le recueil d'expertise qui fait la différence. Une règle de production se constitue d'une prémisse et d'une conclusion et se présente sous la forme

si condition (s) alors conclusion (s).

Un système expert se compose de trois modules séparés et interdépendants. Cette séparation permet de respecter le caractère déclaratif de la représentation (BUCHANAN, SHORTLIFFE, 1984 ; GIARRATANO, RILEY, 2004) :

- la base des règles contient les connaissances du domaine mémorisées sous forme de règles de production. Les règles sont intelligibles pour et par elles-mêmes, et elles sont indépendantes les unes des autres (aucune règle de production ne fait appel à une autre règle, ce qui est très difficile en pratique) ;
- la base de faits contient les faits relatifs au problème courant à résoudre.
- le moteur d'inférence permet d'enchaîner et de contrôler les cycles d'applications des règles, en partant des faits pour atteindre la solution du problème courant. Le cycle d'un moteur d'inférence est à trois temps : sélection des règles applicables, choix d'une règle et déclenchement ;

La base des règles et la base des faits constituent la base de connaissance du système expert.

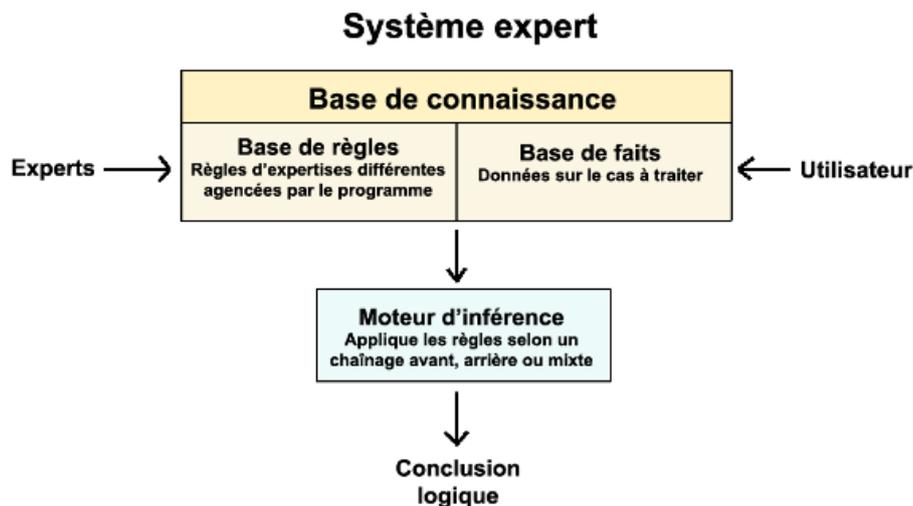


Figure 8 : contenu d'un système expert¹¹²

Base de règles

La base des règles correspond à l'ensemble des règles de production donné sous la forme d'une base de données. Elle représente les connaissances de l'expert sur le domaine (BUCHANAN, SHORTLIFFE, 1984 ; GIARRATANO, RILEY, 2004).

Les règles sont composées de faits liés avec des opérateurs logiques. Elles représentent les connaissances opératoires qui permettent la déduction pour un système expert. Ces règles déductives sont encore appelées règles de production, et permettent de déduire d'autres faits. Les règles sont fournies par l'expert humain suite à son expérience dans le domaine (GIARRATANO, RILEY, 2004).

Ces règles sont sous la forme : *si (ensemble des conditions) alors (nouvelle connaissance)*.

Les conditions d'applications de la règle sont appelées prémisses et il peut y avoir plusieurs prémisses. Les prémisses sont toujours liées entre elles par la coordination « **ET** » et doivent être toutes vraies pour que la règle puisse s'appliquer. Les nouvelles connaissances sont appelées conclusions.

Base de faits

Les prémisses d'une règle peuvent être de deux types (GIARRATANO, RILEY, 2004) :

- des connaissances sur le problème fourni par l'utilisateur du système : les entrées ;
- des connaissances issues de l'application des règles : les faits inférés

¹¹² Source : <http://www.intelligenceartificielle.org/details-progression+fulgurante+de+l+ia+1980+a+aujourd'hui+-8.html> consulté le 25 avril 2019.

Ces connaissances sont enregistrées dans une base de faits qui contient donc toutes les informations sur le problème, quelle que soit leur origine.

Moteur d'inférence

Le moteur d'inférence est le cœur du système. Il permet de faire la mise en relation de la base des faits et la base des règles. Il va permettre ainsi de sélectionner et d'appliquer les règles. Le moteur d'inférence est un programme qui parcourt la base de connaissance en appliquant les règles aux faits connus donnés. Il est l'équivalent de l'esprit logique de l'expert humain qui raisonne sur des faits connus pour déduire de nouvelles connaissances (GIARRATANO, RILEY, 2004 ; BUCHANAN, SHORTLIFFE, 1984). Le moteur d'inférences est responsable de l'ajout des nouveaux faits dans la base des faits. L'ajout des faits que l'on sait faux est tout aussi intéressant que celui des faits que l'on sait juste.

Les faits et les règles utilisent une arborescence dans leur construction. Il faut tenir compte de la structure et appliquer les stratégies de raisonnement qui sont (GIARRATANO, RILEY, 2004) :

- le chaînage avant : la règle est déclenchée par la partie prémisse et la conclusion en résulte. On part des données disponibles dans la base de faits, et on teste pour chaque règle si elle peut s'appliquer ou non. Si oui, on l'applique et on rajoute à la base de faits. Le chaînage avant est dirigé par les données ;
- le chaînage arrière : la règle est déclenchée par la partie conclusion ; pour valider la règle, il faut que la prémisse soit un fait connu. On part des faits que l'on souhaiterait obtenir et on cherche une règle qui pourrait permettre d'obtenir ce fait. On rajoute alors toutes les nouvelles prémisses de cette règle dans les nouveaux buts à atteindre. Le chaînage arrière est dit diriger par le but ;
- le chaînage mixte : qui est une combinaison des deux stratégies ci-dessus.

Le système expert utilise des interfaces pour communiquer entre l'utilisateur et l'expert.

Cette communication s'effectue à deux niveaux :

- au niveau de l'expert qui instruit le système expert en lui donnant les règles du raisonnement ;
- au niveau de l'utilisateur qui pose des questions au système expert et reçoit en réponse le résultat de raisonnement de celui-ci.

La puissance du système expert réside donc dans :

- la quantité de connaissance traitée ;
- la manière de les représenter : la base de faits et la base de règles ;
- la manière de les exploiter : le moteur d'inférence

Un système expert permet de partir de faits et d'y appliquer des règles pour obtenir de nouveaux faits, dits inférés. Le cœur du système expert est son moteur d'inférences. C'est lui qui va choisir et appliquer les règles, et lancer les interactions avec l'utilisateur. La création de ce moteur n'est pas toujours chose aisée, mais il est possible d'en implémenter dans tous les langages ou à partir des générateurs et des moteurs d'inférences. Il est possible d'ajouter la gestion des incertitudes, autant au niveau des faits entrés par l'utilisateur que des règles, pour les domaines plus difficiles à modéliser ou dans des cas où l'utilisateur ne peut pas répondre avec certitude aux questions posées par le système. Les systèmes experts, en raison de leur facilité de mise en œuvre, leur puissance, et leur facilité d'utilisation, se retrouvent aujourd'hui dans de nombreux domaines.

Des moteurs de règles open source existent. On peut citer par exemple CLIPS : moteur de règles dans le domaine public ; Drools : distribué par RedHat ; DTRules : moteur de règles en Java ; Gandalf : moteur de règles tournant sur PHP.

Logique floue

Le terme d'ensemble flou apparaît pour la première fois en 1965 lorsque le professeur Lotfi A. Zadeh, de l'université de Berkeley aux USA, publie un article intitulé « *Ensembles flous* » (*Fuzzy sets*) (ZADEH, 1965). C'est une technique de l'IA déterministe permettant de prendre des décisions. Elle permet ainsi d'avoir un comportement cohérent et reproductible en fonction des règles fournies.

Les avantages de la logique floue viennent notamment de ses capacités à (CHEVRIE, GUÈLY, 1998) :

- formaliser et simuler l'expertise d'un opérateur ou d'un concepteur dans la conduite et le réglage d'un procédé ;
- donner une réponse simple pour les procédés dont la modélisation est difficile ;
- prendre en compte sans discontinuité des cas ou exceptions de natures différentes, et les intégrer au fur et à mesure dans l'expertise ;

- prendre en compte plusieurs variables et effectuer de la « fusion pondérée » des grandeurs d'influence.

La logique floue trouve ses origines dans un certain nombre de constatations :

- la connaissance que l'être humain a d'une situation quelconque est généralement imparfaite, elle peut être incertaine¹¹³, ou imprécise¹¹⁴ ;
- l'être humain résout souvent des problèmes complexes à l'aide de données approximatives.

Pour que l'utilisation de règles floues puisse être envisagée, il faut impérativement qu'il existe une expertise, un savoir-faire humain. Les bases de règles floues n'apportent pas de solution quand personne ne sait comment fonctionne le système ou ne peut le piloter manuellement.

Ensembles flous et degrés d'appartenance

- **Logique booléenne et logique floue**

En logique booléenne (logique classique), une valeur peut être seulement vraie codée 1 ou fausse codée 0. On définit une valeur seuil qui sert de transition. La logique floue repose sur la notion d'appartenance partielle ou sous ensemble flou. Chaque élément appartient graduellement ou partiellement aux sous-ensembles flous qui ont été définis. Il existe une phase de transition durant laquelle la variable se situe entre vrai et faux comme le montre la figure 9 ci-dessus (CHEVRIE, GUÈLY, 1998 ; BOUCHON-MEUNIER, 2007).

¹¹³ Une connaissance (faits ou règles) est incertaine si elle n'est ni vraie, ni fausse, mais seulement probable.

¹¹⁴ Une connaissance est imprécise si elle est mal définie, mal cernée.

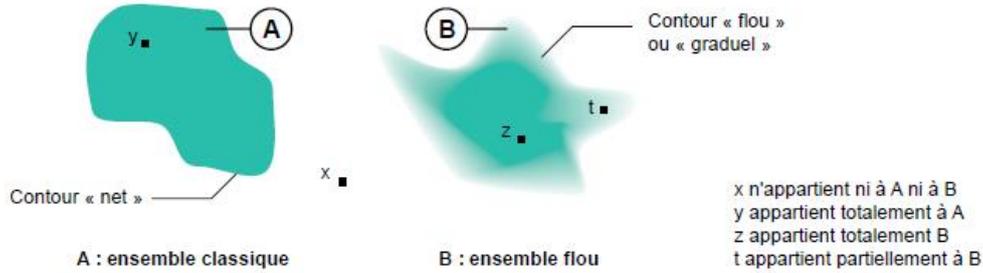


Figure 9 : comparaison entre un ensemble classique et un ensemble flou (CHEVRIE, GUÈLY, 1998)

- **Fonction d'appartenance**

Un sous ensemble flou A de X est défini par une fonction d'appartenance qui associe à chaque élément x de X, le degré, compris entre 0 et 1, avec lequel x appartient à A (BOUCHON-MEUNIER, 2007). Ce degré d'appartenance est déterminé à partir d'une fonction d'appartenance. La figure ci-dessous montre un exemple de fonction d'appartenance à la tachycardie. Sur cette figure, le degré d'appartenance à la tachycardie pour une fréquence cardiaque à 95 battements par minute est de 95% avec la logique floue contre 0 avec la logique booléenne.

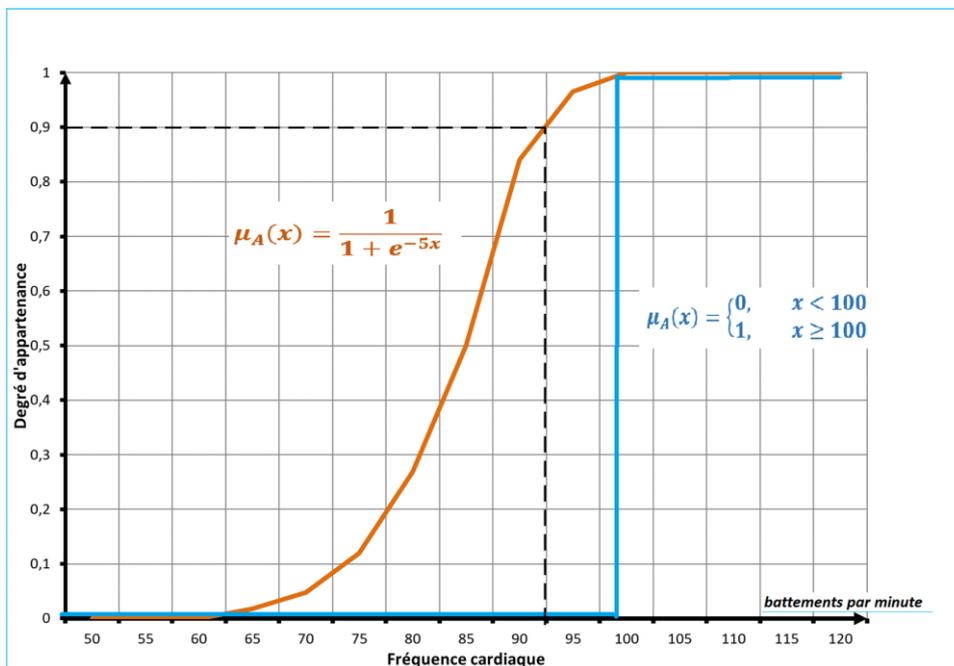


Figure 10 : fonction d'appartenance d'un ensemble flou et d'un ensemble classique

Soit X un ensemble de référence ou univers et x un élément quelconque de X . On dit qu'une partie A de l'ensemble de référence X est un sous-ensemble flou lorsqu'elle est définie de la manière suivante :

$$A = \{(x, \mu_A(x)), x \in X\}, \text{ avec } : \mu_A : X \rightarrow [0, 1].$$

Une partie A d'un ensemble X est usuellement associée à sa fonction d'appartenance. Celle-ci s'applique sur les éléments x de X . Elle prend la valeur 0 si x n'appartient pas à X et 1 si x appartient à X .

En attribuant les éléments x de X un degré d'appartenance, on remarque les trois cas suivants :

- $\mu_A(x) = 0$, lorsque x n'appartient pas à A de façon certaine ;
- $0 < \mu_A(x) < 1$, lorsque x appartient partiellement à A ;
- $\mu_A(x) = 1$, lorsque x appartient à A de façon certaine.

Ces ensembles flous peuvent utiliser différentes fonctions d'appartenance qui associent toutes un degré d'appartenance aux différentes valeurs possibles.

Les ensembles flous diffèrent des ensembles classiques par les principales caractéristiques, que sont :

le support : il correspond à l'ensemble des éléments de X appartenant au moins en partie au sous ensemble A . il est noté $supp(A)$ et se formule ainsi :

$$supp(A) = \{x \in X / \mu_A(x) \neq 0\} ;$$

la hauteur : il correspond au plus haut degré d'appartenance d'un élément x de X à A et s'exprime sous la forme $h(A) = sup_{x \in X} \mu_A(x)$;

le noyau est l'ensemble des éléments de X appartenant de façon absolue au sous-ensemble flou A ; en d'autres termes, tous les éléments pour lesquels la fonction d'appartenance de A vaut 1. Il est noté : $noy(A) = \{x \in X / \mu_A(x) = 1\}$.

La figure ci-dessous présente les caractéristiques d'un sous-ensemble flou.

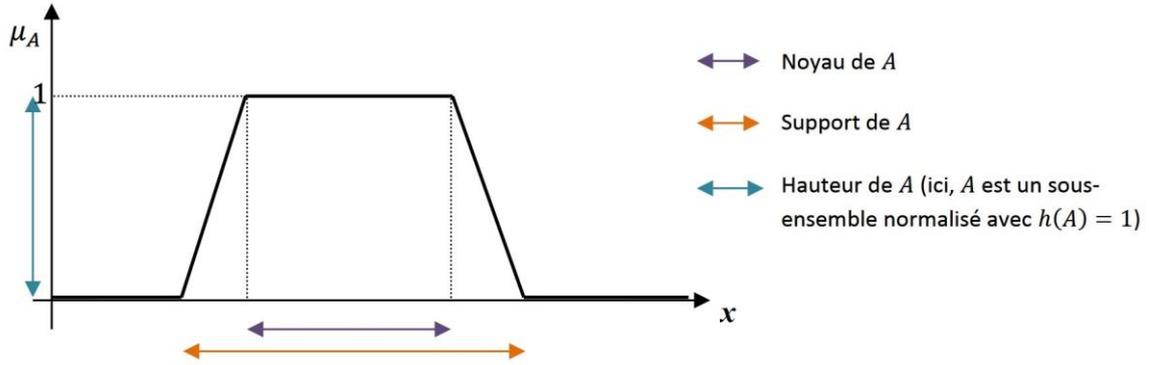


Figure 11: caractéristiques d'un ensemble flou

- **Opérations sur les ensembles**

En logique floue, il n'y a pas de démarcation nette entre ce qui est faux (0) et ce qui est vrai (1). Les opérateurs classiques ne peuvent donc plus s'appliquer. Il faut alors travailler sur les fonctions d'appartenance des ensembles flous notée μ_A pour l'ensemble flou A. Le degré d'appartenance d'une valeur numérique particulière x s'écrit donc $\mu_A(x)$. Pour la négation floue, Non A est défini comme l'ensemble flou ayant pour fonction d'appartenance la fonction définie par : $\mu_{\bar{A}}(x) = 1 - \mu_A(x)$.

Soit A et B, deux sous-ensembles flous de X et $\mu_A(x)$ et $\mu_B(x)$ leurs fonctions d'appartenance respectives (BOUCHON-MEUNIER, 2007) :

l'union de A et B correspond à l'opérateur logique OU, notée :

$$\forall x \in X, \mu_{A \cup B}(x) = \max(\mu_A(x), \mu_B(x)) \text{ où } \max(\) \text{ désigne l'opérateur de maximisation ;}$$

l'intersection de A et B correspond à l'opérateur logique ET, notée :

$$\forall x \in X, \mu_{A \cap B}(x) = \min(\mu_A(x), \mu_B(x)) \text{ où } \min(\) \text{ désigne l'opérateur de minimisation ;}$$

l'égalité de A et B est définie par :

$$A = B \text{ ssi } \forall x \in X, \mu_A(x) = \mu_B(x) ;$$

l'inclusion : A est inclus dans B, si tout les éléments de A appartiennent à B et notée :

$$A \subseteq B \text{ ssi } \forall x \in X, \mu_A(x) \leq \mu_B(x) ;$$

Les opérateurs de minimisation et de maximisation sont les plus utilisés en logique floue.

- **Variables linguistiques et création de règles**

La variable linguistique sert à modéliser les connaissances imprécises ou vagues d'une variable dont la valeur précise est inconnue.

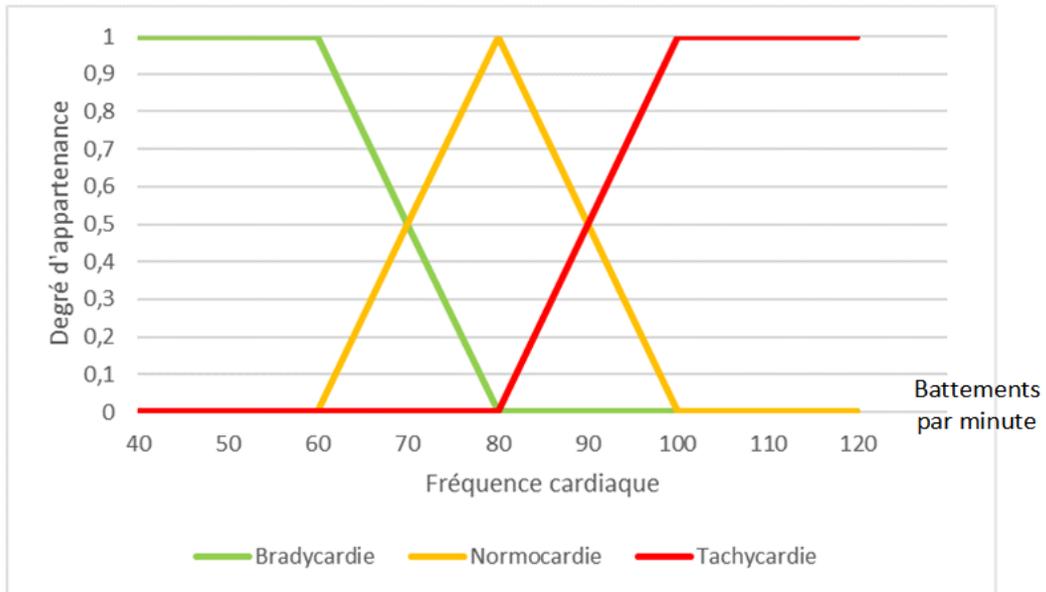
« Une variable linguistique est représentée par un triplet (V, X, T_v) , dans lequel V est une variable (âge, fréquence cardiaque, tension artérielle) définie sur un ensemble de référence X (ensemble des nombres entiers, des réels ...), sa valeur pouvant être n'importe quel élément de X » (BOUCHON-MEUNIER, 2007).

Dans un système flou, les règles utilisent des valeurs floues au lieu des valeurs numériques. *« Une valeur linguistique ou une proposition floue élémentaire est définie à partir d'une variable linguistique (V, X, T_v) par la qualification « V est A », pour une caractérisation floue A que l'on suppose normalisée, appartenant à T_v » (BOUCHON-MEUNIER, 2007).* On note les expressions utilisées dans les règles sous la forme :

« Variable linguistique » EST « valeur linguistique ».

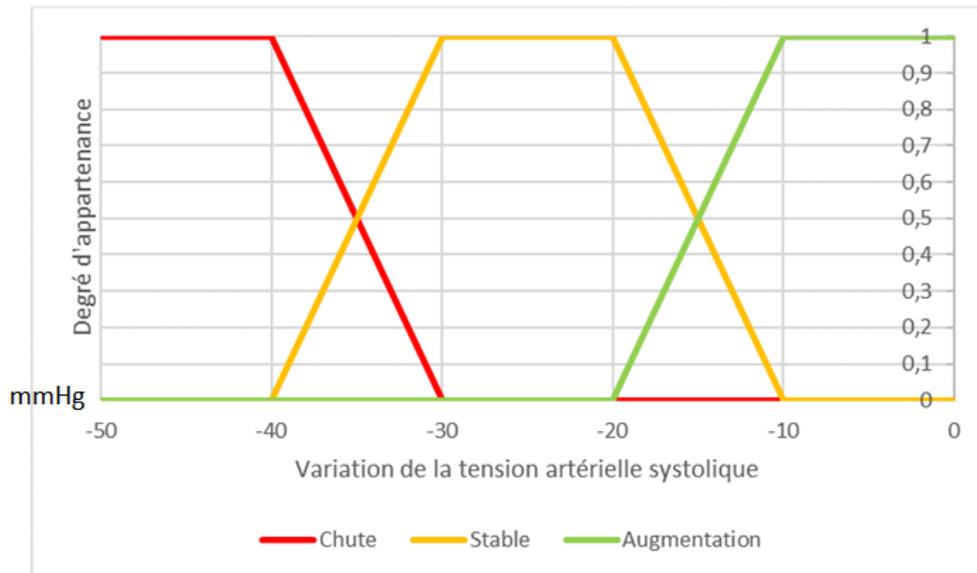
Un système flou peut posséder de nombreuses règles. *« Une proposition floue générale est ainsi obtenue par la composition de propositions floues élémentaires « V est A », « W est B » ... pour des variables V, W ... supposées non interactives » (BOUCHON-MEUNIER, 2007).*

Dans la suite de notre présentation, nous utilisons les variables linguistiques « *Fréquence cardiaque* », « *Variation de la tension artérielle systolique* » et « *Risque d'embolie pulmonaire* » comme fil conducteur. Les figures ci-dessous montrent les fonctions caractéristiques de ses variables linguistiques.



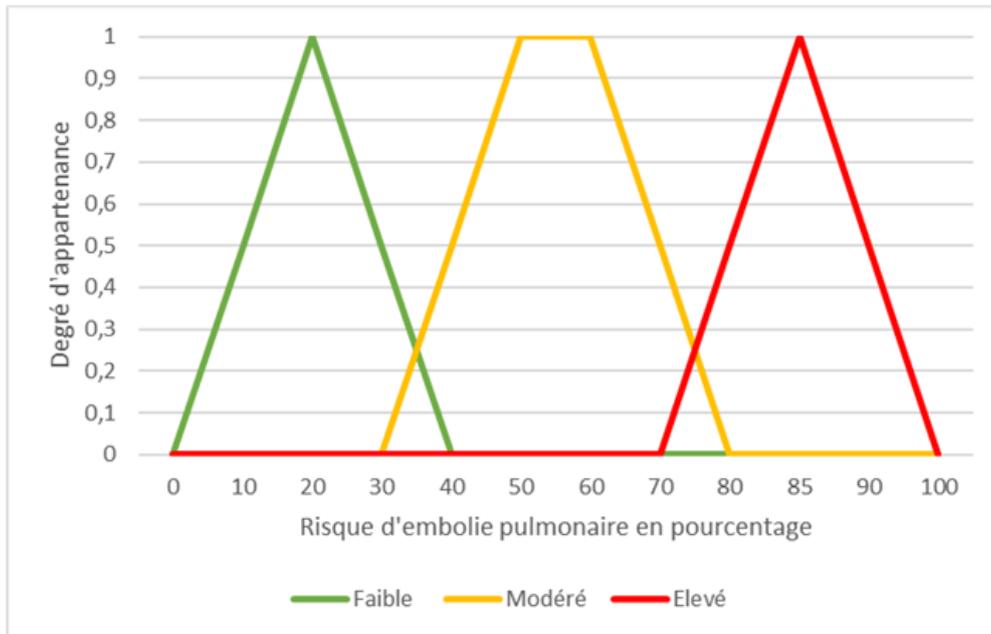
V = Fréquence cardiaque
X = N
Tv = { bradycardie, normocardie, tachycardie }

Figure 12: variable linguistique « *Fréquence cardiaque* »



V = Variation de la tension artérielle systolique
X = N
Tv = { chute, stable, augmentation }

Figure 13: variable linguistique « *Variation de la tension artérielle systolique* »



V = Risque d'embolie pulmonaire
X = N
Tv = { faible, modéré, élevé }

Figure 14: variable linguistique « Risque d'embolie pulmonaire »

- **Fuzzification**

L'étape de fuzzification a pour but de transformer une donnée numérique en entrée en sous-ensemble flou. Pour cela, le concepteur du système flou doit utiliser des fonctions d'appartenances. Les différentes règles possèdent toutes une implication (la clause *alors*). Il va donc falloir exprimer à quel point la règle doit être appliquée en fonction des valeurs numériques mesurées.

On peut utiliser plusieurs variables linguistiques pour caractériser un seul type de données ; il faut alors créer une fonction d'appartenance pour chaque variable. Les base de règles vont être utilisées pour caractériser le fonctionnement désiré du système. Les variables floues sont utilisées dans un mécanisme d'inférence pour créer et déterminer les variables floues de sortie en appliquant les règles développées auparavant. On a alors de nombreux choix pour l'opérateur d'implication afin de déterminer l'ensemble flou résultant.

Dans le cas de notre fil conducteur, nous avons déduit une règle d'inférence qui va conditionner le processus de contrôle :

SI Fréquence cardiaque ET Variation de la tension artérielle systolique ALORS Risque d'embolie pulmonaire.

A partir de cette règle d'inférence nous pouvons déduire les règles suivantes comme le montre le tableau ci-dessous :

Tableau 1: différentes règles d'inférence possibles

| Numéro de la règle | <i>SI</i> | Fréquence cardiaque | <i>ET</i> | Variation de la TAS | <i>Alors</i> | Risque d'embolie pulmonaire |
|--------------------|-----------|---------------------|-----------|---------------------|--------------|-----------------------------|
| 1 | <i>Si</i> | Tachycardie | <i>Et</i> | Chute | <i>Alors</i> | Elevé |
| 2 | <i>Si</i> | Tachycardie | <i>Et</i> | Stable | <i>Alors</i> | Modéré |
| 3 | <i>Si</i> | Tachycardie | <i>Et</i> | Augmentation | <i>Alors</i> | Modéré |
| 4 | <i>Si</i> | Normocardie | <i>Et</i> | Chute | <i>Alors</i> | Elevé |
| 5 | <i>Si</i> | Normocardie | <i>Et</i> | Stable | <i>Alors</i> | Modéré |
| 6 | <i>Si</i> | Normocardie | <i>Et</i> | Augmentation | <i>Alors</i> | Faible |
| 7 | <i>Si</i> | Bradycardie | <i>Et</i> | Chute | <i>Alors</i> | Elevé |
| 8 | <i>Si</i> | Bradycardie | <i>Et</i> | Stable | <i>Alors</i> | Faible |
| 9 | <i>Si</i> | Bradycardie | <i>Et</i> | Augmentation | <i>Alors</i> | Faible |

Nous prenons par exemple un patient qui a une fréquence cardiaque à **90 battements par minute** et une variation de la tension artérielle systolique à **- 37 mmHg**. Pour la variable linguistique « *Fréquence cardiaque* », la valeur appartient à normocardie dans 50% et tachycardie dans 50%. Pour la variable linguistique « *Variation de la tension artérielle systolique* », la valeur appartient à chute dans 70% et stable dans 30%. Ce qui nous permet de déduire quatre règles d'inférence (1, 2, 4 et 5).

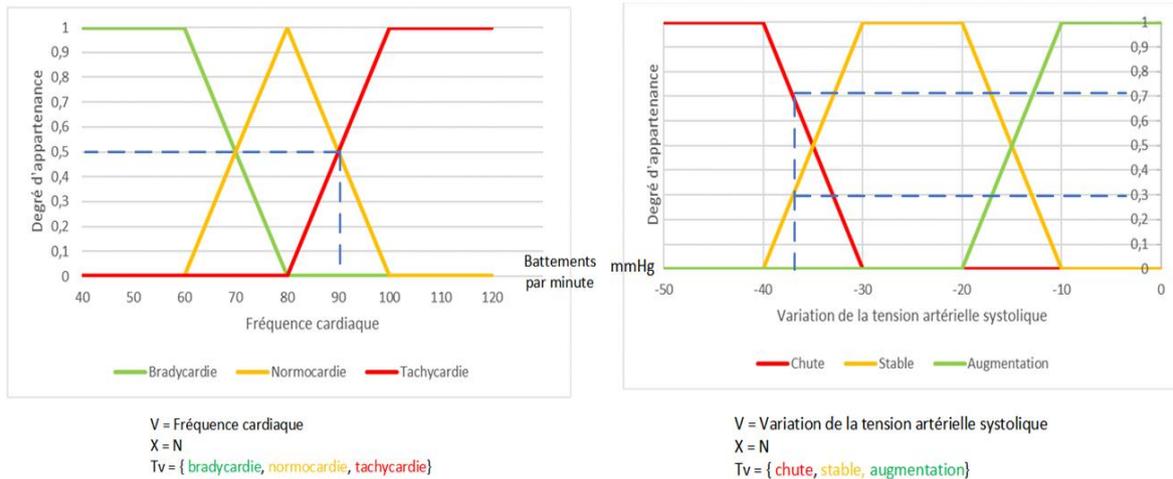


Figure 15: représentation des valeurs de notre exemple

A partir de ces variables floues, nous avons déterminé les variables floues de sortie à partir de l'opérateur d'implication de Mamdani. L'opérateur d'implication de Mamdani (MAMDANI, ASSILIAN, 1975), couramment utilisé peut être décrite en trois phases :

l'agrégation qui est l'évaluation des conditions associées aux règles d'inférence, l'union étant caractérisée par l'opérateur de maximisation et l'intersection par la minimisation ;

l'activation qui est l'évaluation de chaque inférence, et l'application de la relation d'inférence de Mamdani ;

l'agrégation des conclusions qui est l'obtention de la fonction d'appartenance finale à partir du maximum des fonctions d'appartenance.

Un autre opérateur d'implication est la méthode de Larsen, qui, au lieu de tronquer les fonctions d'appartenance, va prendre une réduction de la fonction d'appartenance de conclusion en question proportionnellement au degré d'activation de la règle.

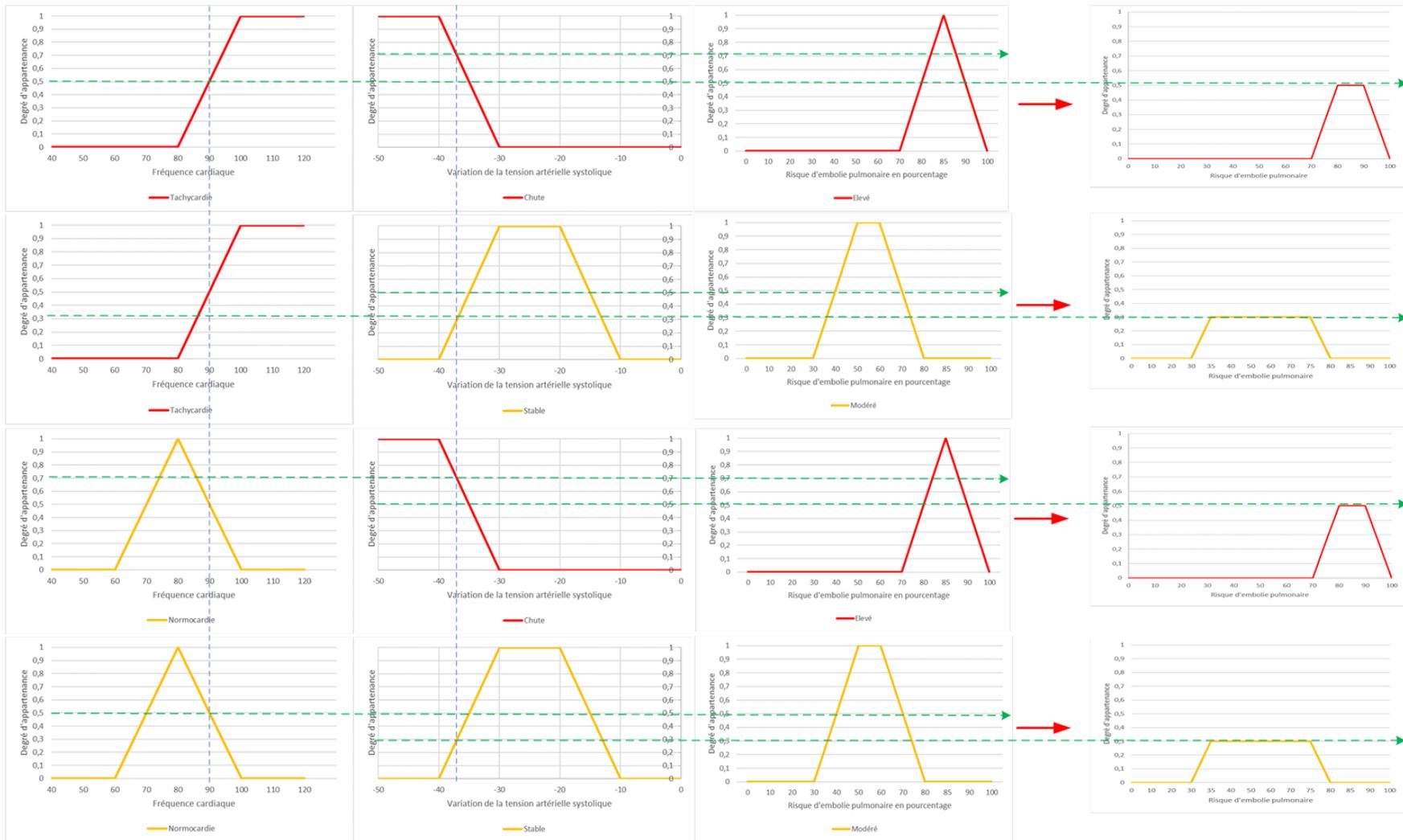


Figure 16: inférence de Mamdani: étapes d'agrégation et d'activation

- **Défuzzification**

Une fois que l'implication (la règle) est évaluée, on obtient un sous-ensemble flou solution, il faut en extraire une valeur réelle pour la prise de décision : c'est l'étape de défuzzification.

L'étape de défuzzification se déroule en deux temps (MARTAJ, MOKHTARI, 2010) :

- d'abord, il faut fusionner les variables linguistiques communes à l'aide d'un opérateur de la logique floue choisi par le concepteur du système. Si on a plusieurs règles d'inférences qui génèrent plusieurs valeurs de la même variable linguistique, on peut choisir un opérateur pour combiner les valeurs de la variable.
- ensuite, on a des variables linguistiques qui caractérisent une seule et même donnée. Ces variables linguistiques possèdent chacune une fonction d'appartenance. Défuzzifier la donnée revient donc à trouver la meilleure valeur quantitative en fonction des fonctions d'appartenances des variables linguistiques.

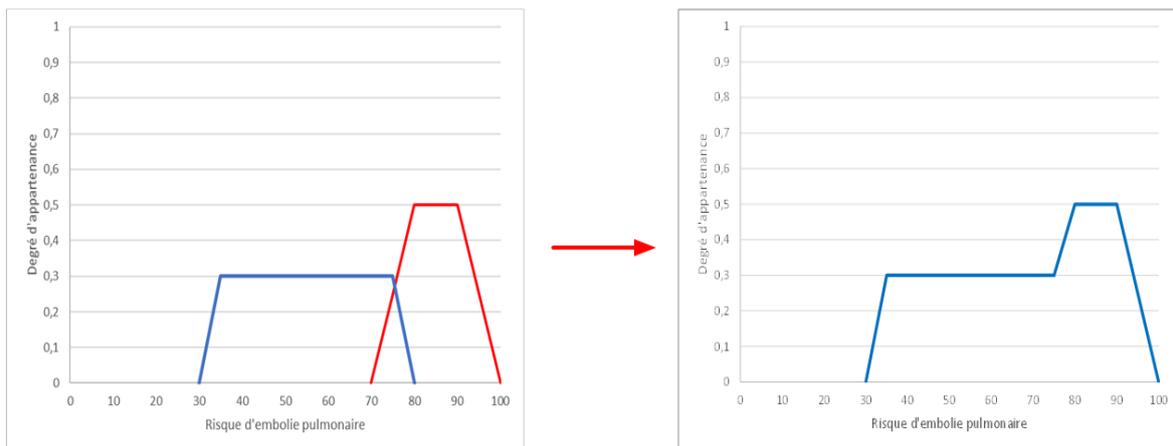


Figure 17: inférence de Mamdani : étape d'agrégation des conclusions

Il existe plusieurs méthodes pour défuzzifier. Parmi les plus utilisés, on peut citer la méthode de la moyenne des maximas et la méthode du centre de gravité.

La défuzzification par la moyenne est la plus simple. La méthode de la moyenne des maximas revient à prendre l'abscisse correspondant à la moyenne des abscisses ayant pour ordonnée la valeur maximale des fonctions d'appartenances.

Le formule de calcul de la moyenne des maximas est :

$$x_{mm} = \frac{\int_S x dx}{\int_S dx} = \frac{\sum_{i=1}^n x_i}{n}$$

$$S = \{x \in A \text{ tel que } \mu_A(x) = \max_{x \in A}(\mu_A(x))\}$$

$n = \text{nombre de points appartenant à } S$

Dans notre exemple, le risque d'embolie pulmonaire est :

$$\frac{(80 + 90)}{2} = 85\%$$

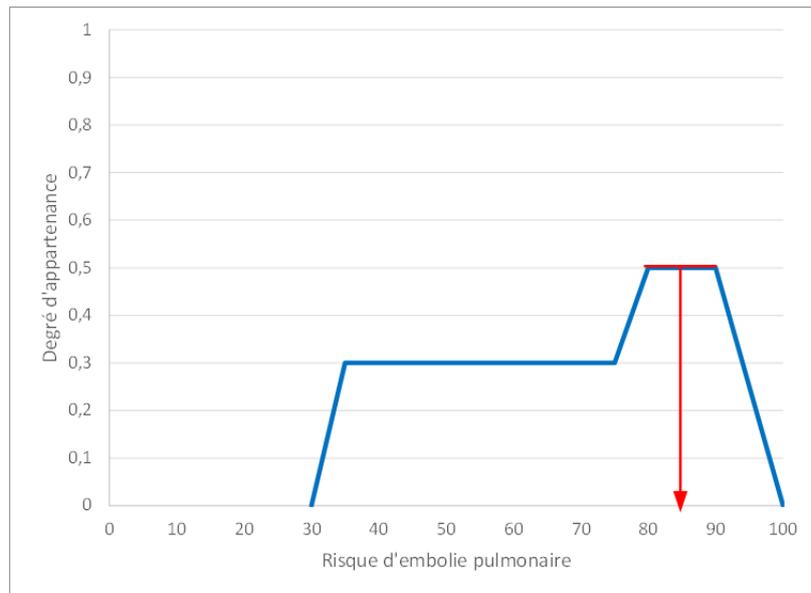


Figure 18: risque d'embolie pulmonaire selon la méthode de la moyenne des maximas

La défuzzification par le barycentre, plus complexe, consiste à chercher le barycentre (aussi appelé centroïde ou, de manière abusive, centre de gravité) de la forme obtenue. Cela permet de prendre en compte l'ensemble des règles, et non uniquement la règle majoritaire. Cette méthode consiste à calculer l'abscisse du centre de gravité de la surface formée par l'intersection de la courbe caractéristique de la fonction obtenue et l'axe des abscisses. Cette valeur retournée est la solution.

La formule du barycentre est :

$$x_g = \frac{\sum_{i=1}^n x_i \mu_A(x_i)}{\sum_{i=1}^n \mu_A(x_i)}$$

$n = \text{nombre de sous-ensembles}$

Le risque d'embolie pulmonaire avec cette méthode dans notre exemple est :

$$\frac{(35 * 0,30) + (75 * 0,30) + (80 * 0,5) + (90 * 0,5)}{0,3 + 0,3 + 0,5 + 0,5} = 73,25\%$$

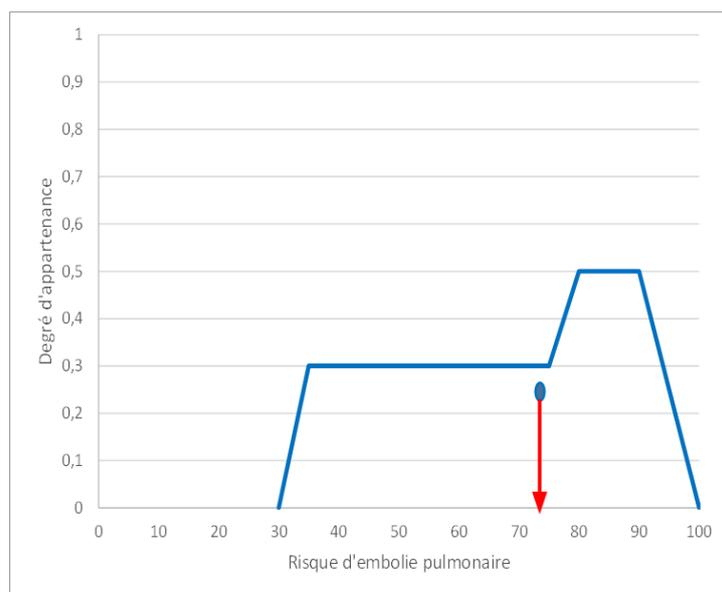


Figure 19: risque d'embolie pulmonaire selon la méthode du barycentre

La logique floue permet de prendre des décisions en fonction de règles imprécises, c'est-à-dire dont l'évaluation est soumise à interprétation. De plus, cela est plus proche du comportement d'un humain, ce qui permet de multiples utilisations. La logique floue est utilisée dans la médecine, les systèmes experts, l'aide à la décision, la reconnaissance des formes et le traitement d'images, la classification, la robotique. Des réalisations industrielles existent par exemple dans l'automobile, l'agro-alimentaire, la chimie, la conception industrielle et la détection de risque.

Systèmes multi-agents (SMA)

Les concepts d'agents et de SMA sont apparus dans les années quatre-vingt. La motivation du domaine des SMA relativement récent est d'essayer de réaliser des entités artificielles (programmes, robots) qui simulent le comportement humain et la façon dont les hommes interagissent entre eux et avec leur environnement.

Le passage de l'agent (l'individu) au multi-agent (le collectif) a nécessité l'extension et la généralisation de divers concepts issus de l'IA tels que l'apprentissage multi-agent (WEIB, 1995), le raisonnement sur les connaissances de groupes (FAGIN, HALPERN, 1994) ou la décision collective.

Le domaine des SMA est à la croisée des chemins entre trois disciplines que sont : l'intelligence artificielle distribuée¹¹⁵, les systèmes distribués¹¹⁶ et le génie logiciel.

Toutes les techniques classées comme systèmes multi-agents ont pour but de mettre en œuvre cette intelligence sociale, qui est appelée en informatique intelligence distribuée. Pour cela, on retrouve :

- un environnement, correspondant aux problèmes que l'on cherche à résoudre ;
- des objets fixes ou non, étant des obstacles ou des points d'intérêt, dont la présence est facultative ;
- des agents, aux comportements simples.

Le but de l'algorithme n'est jamais réellement codé, la solution va émerger de l'interaction de tous ces éléments entre eux.

Agents

Agent vient de « *agri* » : un agent est donc une entité susceptible d'action. Et un agent sera dit rationnel si le choix de ses actions est asservi à la satisfaction de ses besoins, ce qui sous-entend implicitement, qu'un agent rationnel possède des désirs, des souhaits, des buts, ... autrement dit qu'il correspond à ce que le philosophe Daniel Dennett (DENNETT, 1989) appelle un système intentionnel¹¹⁷.

Ferber définit un agent comme « *une entité autonome, réelle ou abstraite, qui est capable d'agir sur elle-même et sur son environnement, qui, dans un univers multiagent, peut communiquer avec d'autres agents, et dont le comportement est la conséquence de ses observations, de ses connaissances et des interactions avec les autres agents* » (FERBER, 1997).

¹¹⁵ « *L'intelligence artificielle distribuée (IAD) se définit comme la branche de l'IA qui s'intéresse à la modélisation de comportement intelligent d'un ensemble d'agents opérant collectivement et de façon décentralisée pour parvenir à un objectif global. Elle propose une distribution de l'expertise sur un groupe d'agents, non soumis à un contrôle centralisé, devant être capable de travailler et d'agir dans un environnement commun et de résoudre des problèmes* » (INPRINCIPO, 2018).

¹¹⁶ « *Un système distribué est un ensemble d'entités autonomes de calcul (ordinateurs, PDA, processeurs, processus, processus léger etc.) interconnectées et qui peuvent communiquer. La technologie distribuée permet au SMA de se focaliser sur les couches supérieures du logiciel et de mettre en évidence leur plus-value-réelle : la prise de décision distribuée et coopérative, l'autonomie, l'adaptabilité, ...* » (SEGHROUCHNI, 2006).

¹¹⁷ La notion de système intentionnel sert à décrire des entités naturelles ou artificielles trop complexes pour que l'on ne maîtrise toutes les déterminations, et à qui l'on prête des « intentions », au sens précédent, c'est-à-dire des affects, des préférences, des buts, ... relatifs aux objets du monde, afin d'en prévoir le comportement.

Une autre définition d'un agent, proposé par Jennings, Sycara et Wooldridge est : « *un agent est un système informatique, situé dans un environnement, et qui agit d'une façon autonome et flexible pour atteindre les objectifs pour lesquels il a été conçu* » (JENNINGS ET AL., 1998).

Les principales caractéristiques d'un agent sont (FERBER, 1997 ; JENNINGS ET AL., 1998) :

- l'autonomie : un agent contrôle son état interne et ses actions. Il est capable d'agir sans l'intervention d'un tiers ;
- la réactivité : un agent perçoit son environnement et réagit à ses changements ;
- la pro activité ; un agent tend par lui-même à atteindre des buts (explicités ou non, plus ou moins élaborés) ;
- la sociabilité : un agent interagit avec d'autres agents du système quand la situation l'exige afin de compléter ses tâches ou aider ces agents à compléter les leurs ;
- la capacité de répondre à temps : l'agent doit être capable d'élaborer une réponse dans les temps requis.

Le comportement d'un agent dépend de sa structure interne (connaissances, croyances, intentions, etc.) et de ses perceptions (communications, interactions). L'autonomie fait référence à l'aspect réflexif d'un agent, c'est-à-dire que chaque agent dispose de son propre contexte d'exécution ; ses propres buts, son propre mécanisme de prise de décision, et est capable de raisonner sur son comportement et d'en changer le cours si nécessaire (FERBER, 1997 ; JENNINGS ET AL., 1998).

Ces caractéristiques des agents permettent de définir une typologie des agents en fonction de leur capacité de raisonnement :

- les agents réactifs, souvent qualifiés de peu intelligents, possèdent une représentation très simplifiée de leur environnement. Leurs capacités consistent à réagir uniquement en mode stimulus/action vu comme une forme de communication (ou perception). Un SMA constitué d'agents réactifs possède généralement un grand nombre d'agents et le pari dans ce type de SMA, porte sur l'émergence d'un comportement collectif intelligent (BROOKS, 1991 ; SEGHROUCHNI, 2006).

- les agents cognitifs possèdent une représentation partielle mais sophistiquée de leur environnement, ont des buts explicites et sont capables de planifier leur comportement, de mémoriser leurs actions passées, de communiquer par envoi de messages ou via des langages d'interactions élaborés, de négocier (FININ, FRITZON, 1994). Un SMA constitué d'agents cognitifs possède généralement un nombre restreint d'agents. Les agents cognitifs peuvent être intentionnels (dotés d'attitudes intentionnelles telles que les croyances, les désirs et les intentions), rationnels (agissant selon une rationalité donnée telle que la rationalité économique), normés (agents évoluant dans un système doté de normes sociales) (SEGHROUCHNI, 2006 ; RUSSELL, NORVIG, 2010).

Un agent rationnel est défini comme un agent cognitif qui agit selon une rationalité donnée en vue d'améliorer son état (par exemple se rapprocher de son but). Cela suppose que l'agent est capable de mesurer sa performance et de savoir à quel moment effectuer cette mesure.

La mesure de la rationalité dépend de quatre éléments (SEGHROUCHNI, 2006) :

- la mesure de performance qui définit le degré de succès d'un agent ;
- l'historique de la perception d'un agent ;
- la connaissance de l'agent à propos de son environnement ;
- les actions qu'un agent peut effectuer.

Un agent rationnel idéal est défini ainsi : « *pour tout historique possible de sa perception, cet agent effectuera l'action dont les connaissances dont il dispose permettent de penser qu'elle maximise sa performance* » (RUSSELL, NORVIG, 2016).

Un agent intentionnel est un agent cognitif doté d'attitudes intentionnelles telles que définies par Dennet (DENNETT, 1989). Dans le domaine des SMA, les agents intentionnels sont incarnés par les agents BDI (*beliefs, desires, intentions*) (RAO, GEORGEFF, 1991) qui sont capables de planifier et d'invoquer des plans selon la théorie du raisonnement pratique (BRATMAN ET AL., 1988). La modélisation d'un système d'agents BDI porte sur le comportement de l'agent ainsi que l'environnement qui est dynamique. Le modèle associé à ce système est un graphe d'états finis où les changements d'états sont provoqués par des

actions. Ces actions sont de deux types : des actions intentionnelles (actions du plan) et des actions non intentionnelles (ou spontanées) produites par l'environnement (appelées aussi événements) (SEGHROUCHNI, 2006 ; FERBER, 1997).

Espace d'un SMA

Les SMA cognitifs constituent un espace qui s'articule autour de trois axes dont la synergie est une condition *sine qua non* pour la conception et la mise en œuvre de systèmes complexes (FERBER, 1997) :

- l'axe cognitif fait référence aux aptitudes cognitives d'un agent telles que l'apprentissage, la prise de décision, la planification autonome qui permettent aux agents de s'adapter aux changements potentiels de leur environnement ;
- l'axe d'interaction fait référence aux mécanismes de communication entre agents qui leur permettent d'échanger de l'information, de partager des objectifs et de coopérer dans le cadre de la résolution des conflits et des négociations ;
- l'axe de concurrence signifie que les agents peuvent s'exécuter en parallèle et mener des activités concurrentes. Ceci permet aux agents de distribuer la résolution des tâches et de l'optimiser, ou enfin de concourir à l'affectation/résolution de tâches.

Principaux algorithmes utilisés dans les SMA

Quatre algorithmes sont utilisés fréquemment dans le cadre des SMA :

- l'algorithme de meutes : à partir de quelques règles simples, il est possible de simuler des comportements de meutes ou de groupes. Craig Reynolds crée ainsi en 1986 les *boids* (REYNOLDS, 1987), des créatures artificielles évoluant en groupe,
- l'optimisation par colonie de fourmis est directement inspirée du fonctionnement des fourmis éclairées. Le but est donc de trouver une solution optimale grâce à la stigmergie¹¹⁸. Cette technique a été créée par Marco Dorigo en 1992 (DORIGO ET AL., 1996) ;

¹¹⁸ En biologie, la stigmergie est un mécanisme de coordination indirecte entre les agents. Le principe est que la trace laissée dans l'environnement par l'action initiale stimule une action suivante, par le même agent ou un agent différent.

- les systèmes immunitaires artificiels s'inspirent des systèmes naturels des vertébrés (comme le nôtre). En effet, plusieurs cellules collaborent pour déterminer le soi et le non soi, et attaquer ce qui a été déterminé comme étranger (virus, bactéries, champignons, voire poisons). Les systèmes immunitaires artificiels font donc évoluer, dans un environnement donné, différents agents de défense. Chacun connaît un ensemble de menaces, qu'il sait détecter (et combattre si nécessaire). Ces agents peuvent être créés aléatoirement au départ (FARMER ET AL., 1986 ; DASGUPTA, 1998 ; TIMMIS ET AL., 2000 ; DE CASTRO, TIMMIS, 2002) ;
- l'automate cellulaire : son environnement est une grille régulière. Chaque agent est situé sur une des cases de la grille et ne peut en bouger. Il possède différents états possibles (par exemple des couleurs). À chaque itération, il va pouvoir changer d'état en suivant quelques règles simples toujours basées sur les états de ses voisins proches. La grille va donc évoluer au cours du temps (VON NEUMANN, BURKS, 1966 ; HEDLUND, 1969 ; WOLFRAM, 1984 ; RENNARD, 2002).

La difficulté réside surtout dans le codage, le choix de l'environnement, des agents et surtout de leurs caractéristiques. En informatique, les systèmes multi-agents contiennent donc un environnement dans lequel se trouvent des objets et des agents. Il n'y a que très peu de règles à suivre sur ceux-ci, et chaque problème pourra donc avoir une ou plusieurs modélisations possibles. Dans tous les cas, c'est la multiplication des agents et les liens qu'ils ont, directement par communication ou indirectement par stigmergie (ou même sans communication entre eux) qui vont faire émerger la solution. On observe alors la puissance de l'intelligence distribuée.

Les SMA sont utilisés dans les domaines de la santé, des finances, des loisirs, du contrôle des processus, des systèmes embarqués et bien d'autres (JENNINGS, WOOLDRIDGE, 1998). Nous avons, par exemple, des applications dans le contrôle du trafic aérien dans les aéroports ; la surveillance continue des patients en soins intensifs ou les SADM.

Machine learning et Deep Learning

Mode d'apprentissage

Dans le cadre du *machine learning*, il existe plusieurs modes d'apprentissages que sont :

- l'apprentissage supervisé qui est un type de *machine learning* dont le but est

d'apprendre à faire des prédictions, à partir d'une liste d'exemples étiquetés, c'est-à-dire accompagnés de la valeur à prédire. Les étiquettes servent de modèles d'enseignement et supervisent l'apprentissage. le système exploite les exemples et acquiert la capacité à les généraliser ensuite sur de nouvelles données de production.

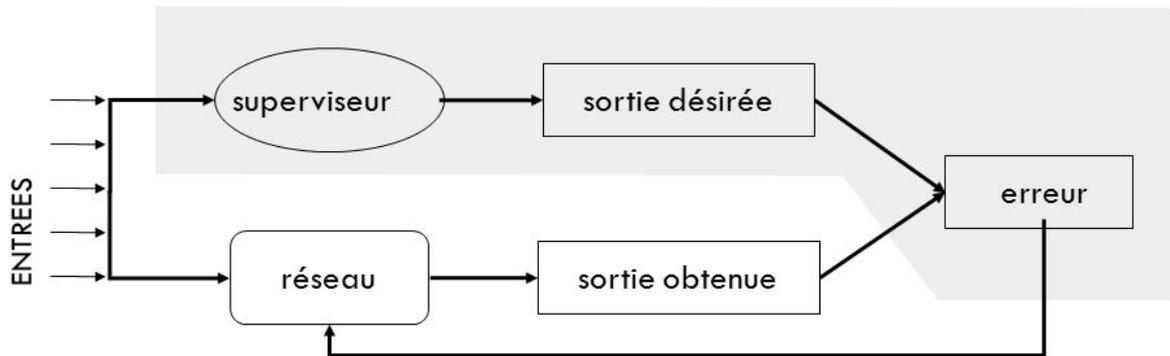


Figure 20 : schématisation de l'apprentissage supervisé (CLERGUE, 2009)

- l'apprentissage non supervisé où les données ne sont pas étiquetées. Il faut alors modéliser les observations pour mieux les comprendre. Le système ne dispose pas de retour explicite sur ses actions.

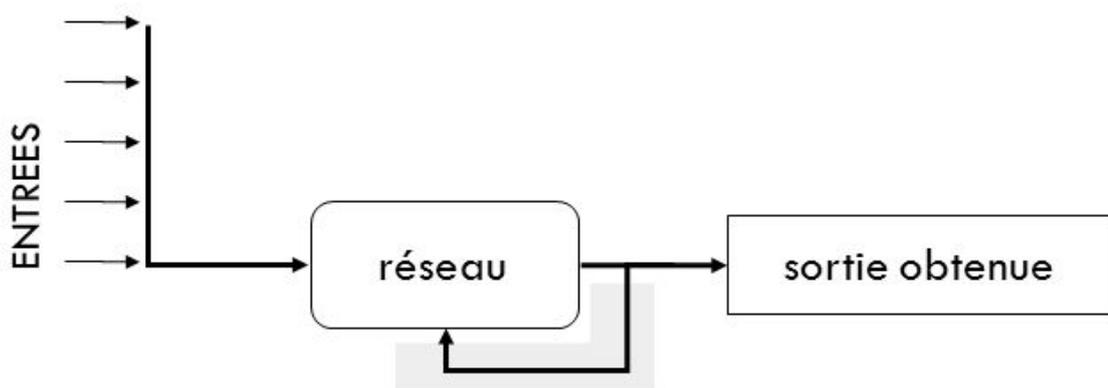


Figure 21 : schématisation de l'apprentissage non supervisé (CLERGUE, 2009)

- l'apprentissage par renforcement est un mode d'apprentissage où le système peut interagir avec son environnement et accomplir des actions. Le système apprend par

des mécanismes de récompenses et de punitions. L'apprentissage consiste dans ce cas à définir une stratégie permettant d'obtenir le meilleur résultat possible. C'est une forme d'apprentissage supervisé incrémental qui utilise des données arrivant au fil de l'eau pour modifier le comportement du système (RUSSELL, NORVIG, 2010).

Description des différentes méthodes

- **Régression linéaire**

Une régression permet de prédire une valeur numérique y en fonction d'une valeur x à partir d'un jeu d'entraînement constitué de paires de données (x, y) . x peut être une seule valeur numérique ou une combinaison de valeurs (vecteur ou matrice de données) (MCCULLAGH, NEDLER, 1989 ; SAPORTA, 2011).

Cet algorithme cherche à établir, sous forme d'une droite, une relation entre une variable expliquée et une variable explicative. Par exemple, prédire la taille d'une personne (variable expliquée) en fonction de son âge (variable explicative).

Là où on utilise une seule variable explicative pour expliquer une autre variable (une taille en fonction d'un âge), on parle de régression linéaire simple ou uni variée ; dans la régression linéaire multivariée on utilise plusieurs variables explicatives.

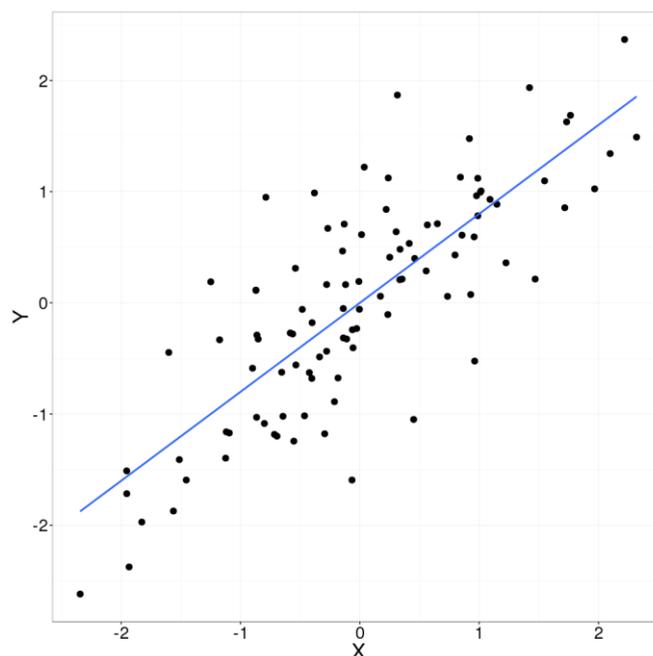


Figure 22 : modèle de régression linéaire

Une étape importante lors de l'utilisation de plusieurs variables explicatives est leur normalisation (mise à l'échelle). La mise à l'échelle (*scaling*) va donc consister à faire en sorte que la moyenne de chaque série d'observations soit égale à 0, que la variance et l'écart-type soient égaux à 1. Cette méthode est également appelée centrage de réduction. Cette méthode permet de faire la prédiction grâce à la méthode de descente de gradient¹¹⁹ ou à la méthode des moindres carrés¹²⁰ (SAPORTA, 2011 ; DRIANCOURT, 1994).

- **Régression polynomiale**

Il est parfois difficile de trouver une droite pouvant passer parmi les points de la série d'observations de façon optimale. Cependant, il est parfois possible de trouver un lien entre les variables à l'aide d'une courbe. C'est ce que permet la régression polynomiale en ajoutant des plis à la courbe à l'aide d'éléments appelés polynômes (AZENCOTT, 2018 ; SAPORTA, 2011 ; MCCULLAGH, NEDLER, 1989).

- **La régression logistique**

Elle utilise, quant à elle, une fonction logistique encore appelée sigmoïde ou courbe en S. Ce type d'algorithme est à appliquer dans des problèmes de classification. La régression logistique a pour but d'isoler les effets de chaque variable, c'est-à-dire d'identifier les effets résiduels d'une variable explicative sur une variable d'intérêt, une fois pris en compte les autres variables explicatives introduites dans le modèle (SAPORTA, 2011 ; SCOTT, SUN, 2013).

Mais l'utilisation des régressions nécessite de faire des compromis qui sont de deux types (FALTINGS, SCHUMACHER, 2017) :

le biais du modèle est la capacité du modèle à représenter la réalité,

la variance de l'apprentissage est l'erreur induite par la sensibilité du modèle aux fluctuations de l'échantillon d'apprentissage.

¹¹⁹ La méthode de descente de gradient : Lorsqu'un système est en phase d'apprentissage, il commet des erreurs. Le taux d'erreur diminue au fur et à mesure de l'apprentissage, mais il se peut qu'à un moment donné l'erreur augmente pour à nouveau diminuer et atteindre un niveau d'erreur plus bas que le précédent qui est le niveau optimal d'apprentissage.

¹²⁰ La méthode des moindres carrés consiste à trouver la droite qui minimise l'écart entre les points théoriques et les points pratiques (mesures réelles). Le point théorique qui correspond au point pratique est celui situé sur la droite de régression à la même abscisse. La méthode consiste à prendre la somme des écarts au carré, et à les minimiser.

- **Arbre de décision**

Cet outil d'aide à la décision ou d'exploration de données permet de représenter un ensemble de choix sous la forme graphique d'un arbre. Il s'agit donc d'une suite de tests réalisés dans le but de prédire un résultat, on retrouve les décisions à chaque extrémité des branches de l'arbre. Les arbres de décision exploitent des critères discriminants, comme dans un moteur de règles. Ils permettent de classer en effectuant une série de tests. Chaque nœud interne correspond à un test sur la valeur de l'un des attributs d'entrée, et les branches partant de ce nœud sont étiquetées avec les valeurs possibles de cet attribut (RUSSELL, NORVIG, 2010 ; SCOTT, SUN, 2013).

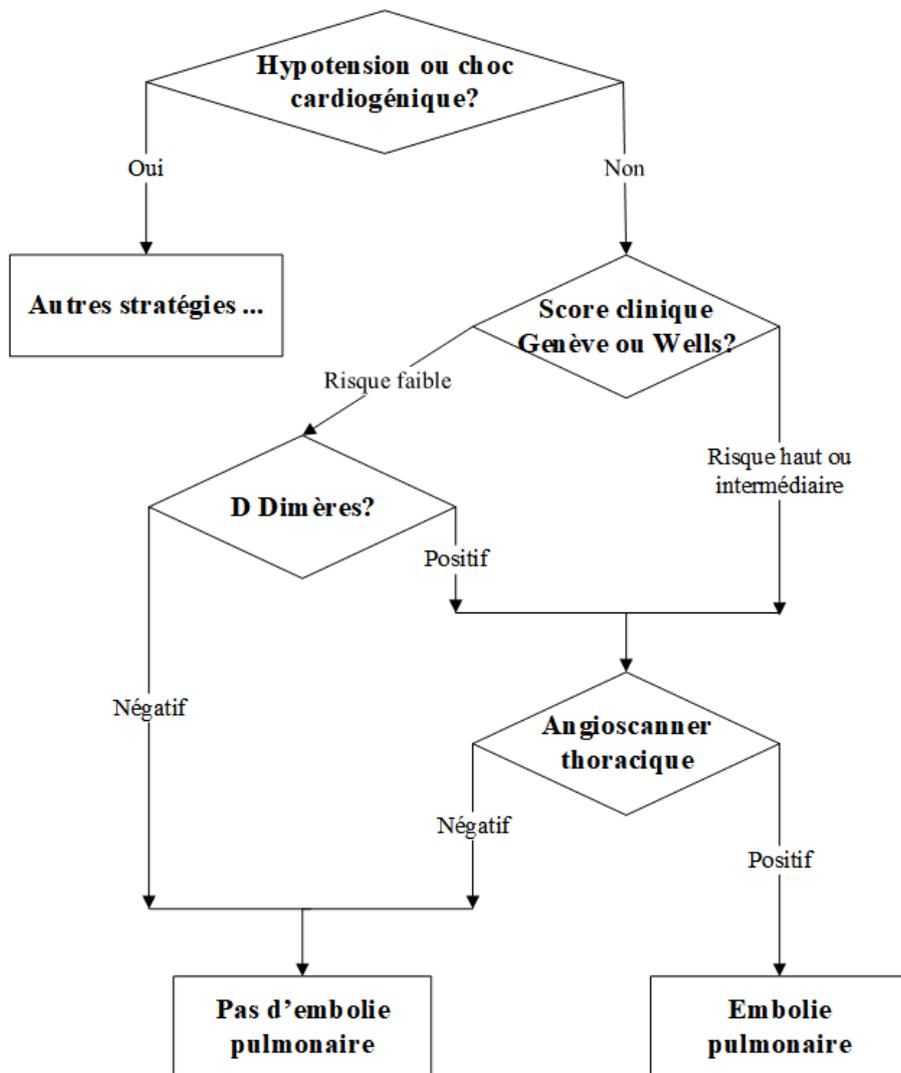


Figure 23 : modèle d'arbre de décision pour le diagnostic de l'embolie pulmonaire sans signe de choc ou d'hypotension artérielle

- **Random forest : (forêt aléatoire)**

Pour chaque test, on partage l'ensemble des données en sous-ensembles et on effectue la moyenne des valeurs de ces sous-ensembles en guise de prédiction. Son fonctionnement consiste à apprendre en parallèle sur plusieurs arbres de décisions construits aléatoirement et entraînés sur des sous-ensembles contenant des données différentes. Chaque arbre propose alors une prédiction et la prédiction finale consiste à réaliser la moyenne de toutes les prédictions. La complexité dans l'utilisation de cet algorithme est de trouver le bon nombre d'arbres à utiliser (BREIMAN, 2001 ; LEMBERGER ET AL., 2016).

- **Machine à vecteurs de support (*Support Vector Machine (SVM)*)**

Le SVM est un algorithme puissant utilisé aussi bien dans les cas de classification que de régression. Son objectif est de déterminer une frontière afin de séparer les observations en groupes distincts tout en maximisant la marge de séparation. Les observations les plus proches de la frontière sont appelées des vecteurs de supports (RUSSELL, NORVIG, 2010 ; BOSER ET AL., 1992).

Les SVM linéaires cherchent à identifier une droite ou un hyperplan dans le cas d'un modèle à plusieurs dimensions qui permette de distinguer les classes d'objets les unes des autres de manière binaire en essayant de les séparer par une marge, représentée par une bande autour de l'hyperplan, aussi large que possible. On les appelle aussi des modèles d'indépendance conditionnelle simples (HEARST, 1998). Les SVM mémorisent tous les exemples d'apprentissages et peuvent avoir besoin de tous les stocker.

La figure suivante illustre le principe du SVM. On peut y observer une frontière séparant deux groupes. Chaque groupe se trouve à une distance (d) de cette frontière. La marge que l'on cherche à maximiser étant alors la somme de ces deux distances. On note également que les observations les plus proches de part et d'autre de la frontière sont alors nommées des vecteurs de supports.

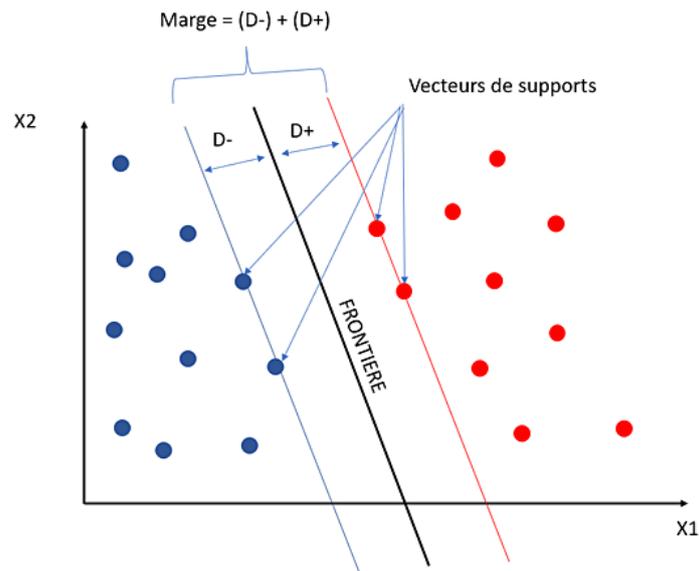


Figure 24 : modèle de SVM (VANNIEUWENHUYZE, 2019)

Ce type d'algorithme est particulièrement efficace lorsque les données sont linéairement séparables (séparables par une droite), ce qui en pratique est rarement le cas. Une technique consiste donc à projeter les données dans un espace vectoriel de plus grande dimension à l'aide d'un élément appelé noyau, permettant alors une création de frontière.

- ***K-Nearest Neighbours (K-NN)***

L'algorithme *K-Nearest Neighbours* est un algorithme de classification visant à déterminer pour une observation donnée son groupe d'appartenance à partir du groupe d'appartenance de ses K plus proches voisins. K étant le nombre de voisins à considérer (COOMANS, MASSART, 1982).

La fonction de décision s'exprime en fonction des données observées et non pas comme une formule analytique fonction des variables. La procédure d'apprentissage consiste à stocker les données du jeu d'entraînement et ne comporte aucun calcul ; mais peut devenir un facteur limitant dans le cadre des grandes bases de données (AZENCOTT, 2018).

- ***Naive Bayes***

Naive Bayes est un algorithme de classification fondé sur les probabilités. Son utilisation est très populaire dans les applications du *machine learning* notamment dans les problématiques de classification de texte. *Naive Bayes* se base sur le Théorème de Bayes fondé sur les

probabilités conditionnelles, c'est-à-dire la détermination de la probabilité qu'un évènement se produise en fonction d'un évènement qui s'est déjà produit. Les classifications naïves bayésiennes utilisent les probabilités pour départager les objets dans des classes en fonction de caractéristiques bien établies. À chaque hypothèse de départ, on associe une probabilité. L'observation d'une ou de plusieurs instances peut modifier cette probabilité. On peut parler de l'hypothèse la plus probable au vu des instances observées. Les probabilités bayésiennes présupposent l'indépendance des attributs utilisés (DOMINGOS, PAZZANI, 1997 ; RISH, 2001).

- **K-Moyennes (*K-Means*)**

L'algorithme des K-Moyennes (*K-Means*) est l'un des algorithmes de *clustering* les plus utilisés. Le *clustering*, le partitionnement ou la segmentation automatique est une méthode d'apprentissage non supervisé qui permet à partir d'un jeu de données non labellisé d'identifier des groupes de données proches les unes des autres, les clusters de données. Les méthodes de *clustering* permettent d'identifier les paramètres discriminants de ces différents segments (ARTHUR, VASSILVITSKII, 2007 ; BAHMANI ET AL., 2012).

Son principe de fonctionnement est relativement simple, car après avoir indiqué à l'algorithme le nombre de clusters à trouver, celui-ci tente par itérations successives de déterminer des centroïdes (un par cluster) autour desquels il est possible de regrouper les données. Ces regroupements s'effectuant en calculant la distance de chaque observation par rapport à un point central de regroupement appelé centroïde et permettant ainsi de classer les observations en plusieurs groupes de façon automatique (DAVIES, BOULDIN, 1979 ; GOODMAN, KRUSKAL, 1954). La méthode des *K-Means* cherche à minimiser la variance intracluster afin de créer des clusters homogènes.

- **Réseaux de neurones**

Les réseaux neuronaux ont connu un fort développement à partir de 2012 et dans leur mise en œuvre d'abord dans la machine *learning* puis avec le *deep learning*, qui exploite des réseaux de neurones avec un grand nombre de couches. Le *deep learning* est un sous-ensemble des techniques de *machine learning* à base de réseaux de neurones qui s'appuient sur des réseaux de neurones à plusieurs couches dites cachées. Les outils de *deep learning* s'appuient sur différentes variantes de réseaux de neurones pour leur mise en œuvre pratique. Leur histoire remonte aux perceptrons de Franck Rosenblatt de 1957 (ROSENBLATT, 1958).

L'histoire du *deep learning* a véritablement démarré plus de 20 ans plus tard, dans les années 1980. Il a cependant fallu attendre 1995 pour que l'on puisse les mettre en œuvre en pratique, sans doute, grâce aux progrès matériels constatés par la loi de Moore mais aussi aux progrès conceptuels, notamment aux travaux d'Alexander Waibel en 1989 (WAIBEL ET AL., 1989), Yann Le Cun en 1985 (LE CUN, 1985 ; LE CUN ET AL., 1998) et à Geoffrey Hinton, particulièrement à partir de 1986 (ACKLEY ET AL., 1985 ; RUMELHART ET AL., 1986).

Le perceptron, réalisé en 1957 par Frank Rosenbalt, est un algorithme d'apprentissage supervisé. Le perceptron sous sa forme la plus simple est également nommé neurone formel et a pour but de séparer des observations en deux classes (ou groupes) distinctes à condition que ces données soient linéairement séparables. Son rôle est donc de classifier.

Neurone formel et perceptron

Un neurone est une cellule, composée :

- d'un corps cellulaire ou péricaryon ;
- d'un noyau ;
- de plusieurs ramifications appelées dendrites ayant pour fonction d'être les points d'entrée de l'information dans le neurone ;
- d'un chemin de sortie de l'information appelé axone, pouvant atteindre une longueur d'un mètre ;
- d'une gaine de myéline protégeant l'axone.

Des terminaisons axonales également appelées synapses connectées aux autres neurones

La communication entre neurones s'opère par l'échange de messages sous forme de variation du potentiel d'action. Un neurone peut recevoir plusieurs messages de la part d'autres neurones auxquels il est connecté.

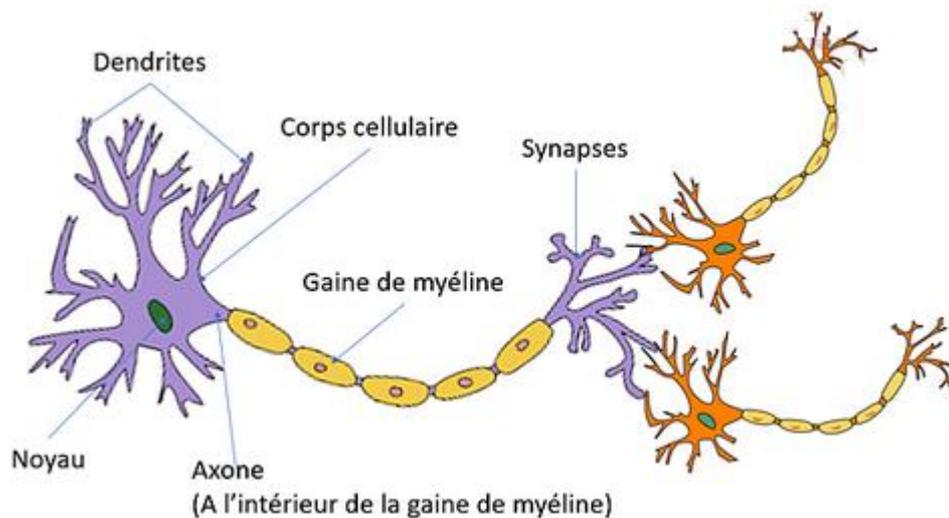


Figure 25 : représentation d'un neurone

Lorsqu'un neurone reçoit plusieurs messages, il effectue la somme des influx nerveux reçus puis, si cette somme dépasse un certain seuil, il s'active et transmet à son tour un message via son axone aux neurones connectés à celui-ci. Cette suite d'activation constitue notre mémoire. Car pour une action donnée, un ensemble de neurones s'active alors que d'autres restent inactifs, car un chemin s'est créé entre l'action et l'activation neuronale.

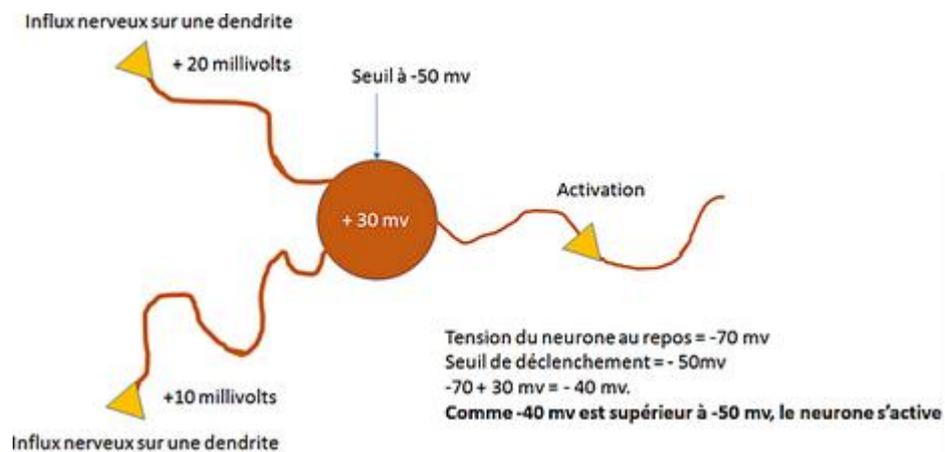


Figure 26 : principe d'activation d'un neurone (VANNIEUWENHUYZE, 2019)

Le neurone artificiel, aussi appelé neurone formel, reprend le fonctionnement du neurone biologique. Le perceptron va se caractériser par une couche de neurones en entrée et un

neurone en sortie. Comme nous l'avons déjà évoqué auparavant, sous cette forme, le perceptron est également appelé neurone formel et ne permet que de réaliser des classifications sur des observations linéairement séparables.

Ces approches sont incomplètes pour représenter une approche neuromimétique pertinente du fonctionnement cérébral. La plupart des approches oublie les approches neurohormonales et l'intégration du système nerveux central avec le système endocrinien, sympathiques et parasympathiques (COLLOC, 2000).

Principe

Un neurone reçoit des entrées et fournit une sortie, grâce à différentes caractéristiques (RUSSELL, NORVIG, 2010 ; MCCULLOCH, PITTS, 1943) :

des poids accordés à chacune des entrées, permettant de modifier l'importance de certaines entrées par rapport aux autres ;

une fonction d'agrégation, qui permet de calculer une unique valeur à partir des entrées et des poids correspondants ;

un seuil (ou biais), permettant d'indiquer quand le neurone doit agir ;

une fonction d'activation, qui associe à chaque valeur agrégée une unique valeur de sortie dépendant du seuil.

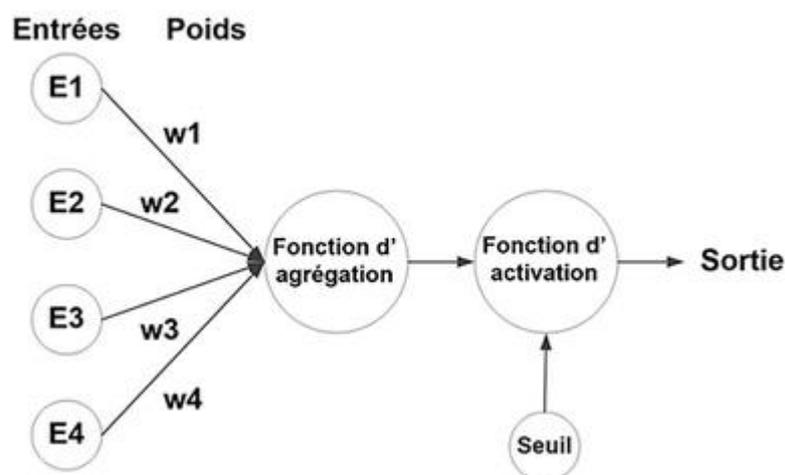


Figure 27 : principe de fonctionnement du réseau de neurone

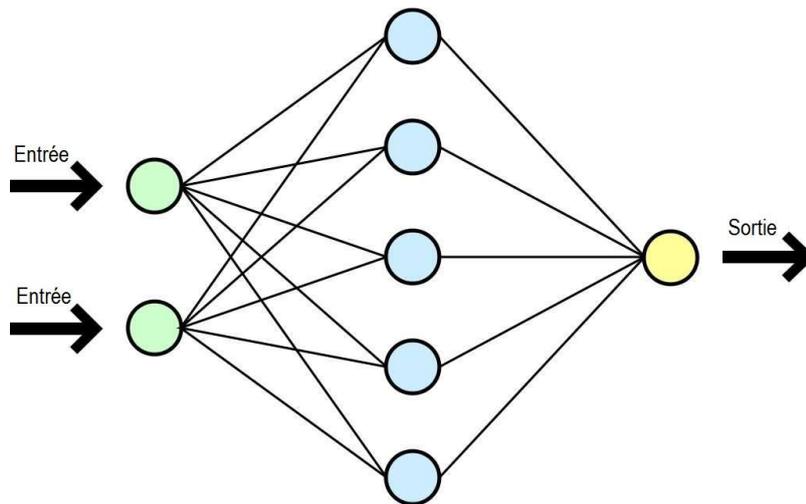


Figure 28 : réseau de neurone à une couche

Dans le cas des réseaux multicouches, les neurones sont classés en trois catégories (CORNUÉJOLS ET AL., 2018) :

neurone d'entrée : unité chargée de transmettre une composante du vecteur x des données (les données d'apprentissage pendant la phase d'apprentissage) ;

neurone de sortie : unité qui fournit l'hypothèse d'apprentissage ;

neurone caché : qui n'est ni une unité d'entrée ni de sortie, qui réalise des traitements intermédiaires.

D'autres modèles existent comme la machine de Boltzmann où tous les neurones formels quel que soit le type sont connectés les uns aux autres (ACKLEY ET AL., 1985).

Fonction d'agrégation

Le but de cette fonction est de transformer l'ensemble des entrées et des poids en une seule valeur, qui sera utilisée pour la fonction d'activation. Il est possible d'imaginer plusieurs fonctions d'agrégation. Les deux plus courantes sont : la somme pondérée¹²¹ et le calcul de distance¹²².

¹²¹ Dans le cas de la somme pondérée, on va simplement faire la somme de toutes les entrées (E) multipliées par leur poids (w).

¹²² Dans le cas du calcul des distances, on va comparer les entrées aux poids (qui sont les entrées attendues par le neurone), et calculer la distance entre les deux. La distance est la racine carrée de la somme des différences au carré.

Fonction d'activation

L'activation du neurone artificiel est déclenchée en fonction d'un seuil comme le neurone biologique. La première étape dans la détermination de l'activation ou non d'un neurone artificiel est de réaliser la somme des poids synaptiques avec les valeurs d'entrées. Cette étape peut être considérée comme étant une pré-activation du neurone. Une fois cette somme réalisée, on utilise une fonction d'activation qui va permettre de déterminer si le neurone doit s'activer ou non. Cette activation sera en fait la prédiction du neurone. Pour calculer cette prédiction, il existe plusieurs fonctions utilisables (MCCULLOCH, PITTS, 1943 ; CORNUÉJOLS ET AL., 2018 ; LESHNO ET AL., 1993) :

la fonction de seuil binaire : cette fonction retourne une valeur égale à 0 quand le seuil d'activation défini n'est pas atteint ou 1 quand le seuil est atteint.

la fonction sigmoïde dont les changements de valeurs entre 0 et 1 sont plus progressifs. La formule mathématique de cette fonction donne naissance à la courbe en S.

la fonction tangente hyperbolique est une fonction sigmoïde. Elle présente cependant l'avantage que pour toute valeur négative celle-ci reste ou devient fortement négative :

la fonction *ReLU* (*Rectified Linear Unit*, unité de rectification linéaire) : La fonction *ReLU* pallie ce problème de saturation¹²³ des fonctions sigmoïdes et hyperboliques via un fonctionnement assez simple. Si la valeur issue de la somme pondérée est inférieure à 0, celle-ci prend la valeur 0, sinon elle prend la valeur de la somme calculée. L'un de ses avantages est son faible coût en termes de calcul, car il s'agit de prendre le maximum entre la valeur 0 et la valeur de la somme pondérée. Son second avantage est qu'elle ne souffre pas de la saturation, car elle n'a pas de valeur limite dans la zone positive. La fonction *ReLU* est la fonction la plus utilisée de nos jours en tant que fonction d'activation dans un réseau de neurones.

la fonction *softMax* attribue des probabilités décimales à chaque classe d'un problème à plusieurs classes. La somme de ces probabilités décimales doit être égale à 1. Cette contrainte supplémentaire permet de faire converger l'apprentissage plus rapidement. Ce type de fonction d'activation est généralement utilisé dans un réseau de neurones multicouche et dans le cas de classifications multi classes.

¹²³L'apprentissage se fait en ajustant les différents poids du réseau de neurones en fonction de l'erreur qu'il a commise lors de sa classification. Dans le cas d'une saturation, la mise à jour de ces poids devient difficile et l'apprentissage s'en trouve impacté.

Réseaux *feed-forward*

Les réseaux de type *feed-forward* ou à couches permettent de dépasser les limites des perceptrons. En effet, ceux-ci ne sont plus limités aux problèmes linéairement séparables.

Les réseaux avec couches cachées sont composés d'une ou plusieurs couches cachées de neurones, reliées aux entrées ou aux couches précédentes, et d'une couche de sortie, reliée aux neurones cachés. On les appelle *feed-forward* car l'information ne peut aller que des entrées aux sorties, sans revenir en arrière (RUMELHART ET AL., 1986 ; BISHOP, 1995 ; LE CUN ET AL., 1998).

Le nombre de couches cachées et le nombre de neurones de chaque couche sont à choisir par l'utilisateur (généralement en faisant des tests sur plusieurs valeurs). On parle de *deep learning* quand le nombre de couches est important (à partir de trois ou quatre couches cachées). Plus le réseau possède de couches et plus il est complexe, permettant un meilleur apprentissage plus rapidement qu'avec une ou deux couches cachées. En effet, chaque couche peut être vue comme étant une étape supplémentaire dans le processus d'apprentissage.

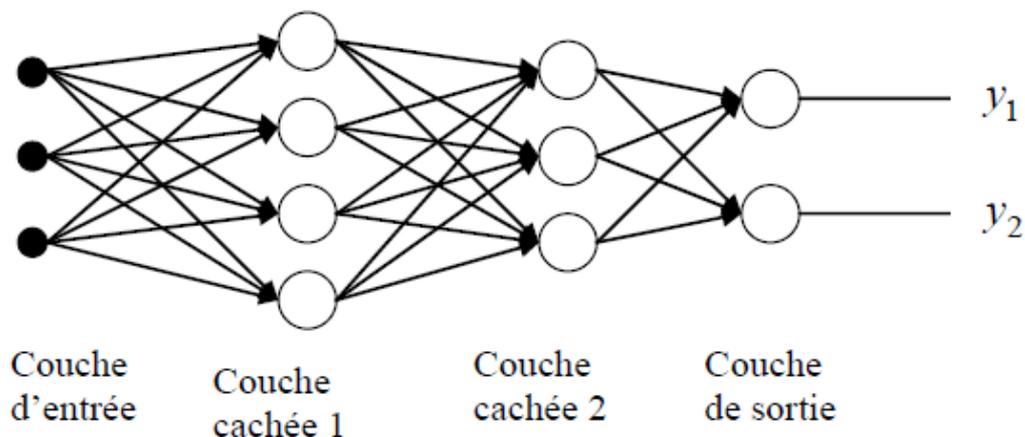


Figure 29 : modèle de réseau de neurone avec couche cachée

Dans le cas des réseaux *feed-forward*, il existe un apprentissage possible par rétro propagation du gradient ; c'est une extension de l'algorithme de descente de gradient. Le principe de l'apprentissage par rétro propagation est de corriger tout d'abord les poids entre les neurones de sortie et les neurones cachés, puis propager l'erreur en arrière et corriger les poids entre les neurones cachés et les entrées. Cette correction se fera exemple après

exemple, et il faudra plusieurs passes (avec des ordres différents si possible) pour converger vers un optimum (DRIANCOURT, 1994 ; LE CUN ET AL., 1998 ; BISHOP, 1995).

Les réseaux *feed-forward* sont fortement utilisés, mais ce ne sont pas les seuls réseaux, et il existe de multiples architectures permettant de travailler sur des problèmes différents.

Réseaux de neurones à convolution

Les réseaux de neurones convolutifs ou *Convolution Network Neurone* (CNN) sont très adaptés pour travailler sur des images et des vidéos, le traitement de langage naturel et les systèmes de recommandations.

Ces réseaux sont composés de multiples couches, pouvant avoir des rôles différents. En particulier, on distingue les couches de convolution, qui traitent des parties de l'image (extracteurs de caractéristiques), les couches de *pooling* (qui combinent les sorties des couches de convolution pour en détecter des fonctionnalités de plus haut niveau) et les couches plus classiques de classification (souvent avec la fonction *ReLU* comme fonction d'activation) (HINTON ET AL., 2006 ; 2012 ; LE CUN ET AL., 1998 ; BOSER ET AL., 1991).

Un traitement correctif non linéaire est appliqué entre chaque couche pour améliorer la pertinence du résultat. L'ensemble des sorties d'une couche permet de reconstituer une image intermédiaire dite de « *carte de caractéristique* » (*feature map*), qui sert de base à la couche suivante (LE CUN ET AL., 1998 ; HINTON ET AL., 2012).

Ces réseaux sont cependant très durs à entraîner en partant de 0. Généralement, on utilise un réseau connu pré entraîné, et on ne fera que de l'entraînement sur notre jeu de données sur quelques générations. Cette utilisation d'un réseau existant s'appelle *transfert learning* ou apprentissage par transfert.

On arrive ainsi à des résultats supérieurs à 95 % de classification correcte avec un temps d'apprentissage qui reste court.

Réseaux de Hopfield

Les réseaux de Hopfield sont des réseaux complètement connectés : chaque neurone est relié à tous les autres. Lorsqu'on soumet une entrée au réseau, on ne modifie l'état que d'un neurone à la fois, jusqu'à la stabilisation du réseau. L'état stable est donc la « *signature* » de l'entrée. L'apprentissage consiste à déterminer les poids de manière à ce que des entrées différentes produisent des états stables différents, mais que des entrées presque identiques conduisent au même état (HOPFIELD, 1982).

De cette façon, si des erreurs entachent légèrement une entrée, elle sera quand même reconnue par le réseau. On peut ainsi imaginer un système permettant la reconnaissance des lettres, même si celles-ci sont abîmées ou moins lisibles.

Algorithmes génétiques

Les algorithmes génétiques appartiennent à la famille des algorithmes évolutionnistes. Ils sont utilisés pour effectuer des optimisations sur des problèmes complexes afin d'obtenir une solution proche de l'optimal (ALLIOT, DURAND, 2005). Les algorithmes génétiques utilisent la notion de sélection naturelle et l'appliquent à une population de solutions potentielles au problème donné. Les algorithmes génétiques ont déjà une histoire relativement ancienne puisque les premiers travaux de John Holland sur les systèmes adaptatifs remontent à 1962 (HOLLAND, 1962). En entrée, un algorithme génétique dispose (FALTINGS, SCHUMACHER, 2017 ; HOLLAND, 1962) :

d'une population initiale à n solutions ; chaque solution est représentée par un chromosome qui est une version codée de toutes ses caractéristiques sous la forme d'une chaîne de symbole ;

d'une fonction f d'évaluation des solutions ; f n'est pas forcément une fonction mathématique de chromosomes. Elle peut aussi consister à une simulation de la performance des solutions dans le problème dans lequel elle doit s'appliquer ; cela permet de modéliser le principe de sélection naturelle.

A chaque itération, l'algorithme génère une nouvelle génération de solutions en appliquant des opérateurs de mutations et de croisements sur des membres de la population courante. Une nouvelle population est ainsi constituée à partir de la population courante en éliminant certaines solutions peu performantes et en créant des nouvelles par des mutations et des croisements. L'application itérative de la procédure conduit à l'optimisation de la population selon les critères d'évaluation choisis au départ.

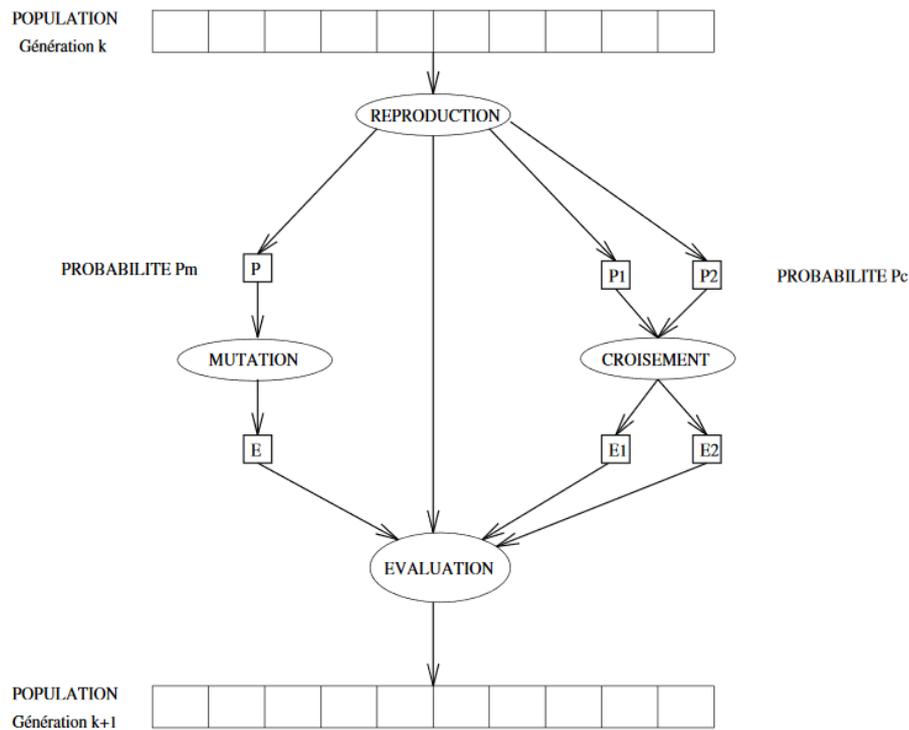


Figure 30: principe général des algorithmes génétiques (ALLIOT, DURAND, 2005)

Les algorithmes génétiques requièrent de nombreuses itérations, ainsi que l'utilisation excessive de la fonction d'évaluation entraînant ainsi un temps long de calcul. En plus, la définition de la fonction d'évaluation est critique, elle doit être optimale car elle constitue la base de l'algorithme.

3.2.3 Données de l'IA

Le point commun du *machine learning* et du *deep learning* est d'exploiter des données pour l'entraînement de modèles probabilistes. Avec les algorithmes et/ou logiciels et le matériel, les données sont la troisième composante clé de la plupart des IA de nos jours. On distingue généralement trois types de données pour entraîner un système de *machine learning* et de *deep learning* : les données d'entraînement, les données de test et les données de production.

Données d'entraînement

Ce sont les jeux de données qui vont servir à entraîner un modèle de *machine learning* ou de *deep learning* pour en ajuster les paramètres. Plus la base est grande, meilleure sera l'entraînement du système, mais plus il sera long. Ces données sont étiquetées à l'avance

pour faciliter l'apprentissage. Il est évidemment nécessaire de disposer de données d'entraînement de qualité ; ce qui nécessite souvent un gros travail de filtrage, de nettoyage et d'élimination des doublons préalable à l'ingestion des données, d'où l'intérêt de construction des systèmes de *big data* fiables pour le traitement des données.

Données de test

Ce sont les données, également étiquetées, qui serviront à vérifier la qualité de l'entraînement d'un système. Ces données doivent avoir une distribution statistique voisine des données d'entraînement, au sens où elles doivent être bien représentatives de la diversité des données que l'on trouve dans la base d'entraînement et que l'on aura dans les données de production.

Les données de tests sont un sous-ensemble d'un jeu de départ dont une partie sert à l'entraînement et une autre partie, plus limitée, sert aux tests. Elles seront injectées dans le système entraîné et on en comparera les étiquettes résultantes avec les étiquettes de la base. Cela permettra d'identifier le taux d'erreur du système. On passera à l'étape suivante lorsque le taux d'erreur sera considéré comme acceptable pour la mise en production de la solution. Le niveau de taux d'erreur acceptable dépend de l'application. Son maximum généralement accepté est le taux d'erreur de la reconnaissance humaine. Mais comme on est généralement plus exigeant avec les machines, le taux véritablement accepté est très inférieur à celui de l'homme.

Données de production

Il s'agit des données non étiquetées qui alimenteront le système lors de son utilisation en production pour faire une prévision des étiquettes manquantes. Alors que les données d'entraînement sont normalement anonymisées pour l'entraînement du système, les données de production peuvent être nominatives ainsi que les prévisions associées générées par la solution.

3.2.4 Utilisation de l'IA dans le domaine de la santé

L'IA est notamment utilisée dans trois principaux domaines de la santé :

- le diagnostic dont l'imagerie médicale ou la génomique et les systèmes d'aide à la décision ;
- la thérapeutique avec les outils pour les biotechnologies ou les robots chirurgiens ;
- la gestion des systèmes de santé et des soins.

L'aide au diagnostic est le domaine de la santé où l'IA a le plus prospéré ces dernières années. L'IA est utilisée pour l'analyse du fond de l'œil afin de détecter diverses pathologies comme les micro rétinopathies diabétiques ou les débuts de glaucome ou de dégénérescence maculaire.

La cardiologie fait appel à plusieurs outils de diagnostic : les échographies, les radios, les IRM (imagerie par résonance magnétique), et les ECG (électrocardiogrammes). Chacun joue son rôle dans le diagnostic ou la prévoyance.

La cancérologie est l'un des plus gros marchés de la santé. Il est fragmenté en de nombreux types de cancers différents qui ont leurs propres techniques de diagnostics, s'appuyant sur différentes formes d'imagerie médicale et d'analyses biologiques. Presque tous les cancers ont leur solution spécifique d'analyse d'imagerie médicale à base d'IA.

Les solutions de systèmes de santé couvrent des besoins divers : le suivi de l'observance des traitements, l'évitement d'erreurs de prises de médicaments, la surveillance des effets secondaires des médicaments, le suivi des dépenses de santé ou l'optimisation des ressources des hôpitaux et praticiens. Ils génèrent de gros volumes de données, d'où les nombreux cas d'usage potentiels de l'IA.

L'univers français des outils dans le domaine de l'IA en santé sont nombreux ; nous pouvons citer entre autres :

- *Cardiologs*¹²⁴ (accrédité CE et FDA) dans l'analyse de l'électrocardiogramme ;
- *I-nside*¹²⁵ avec le smart scope dans l'analyse de l'oreille ;
- *IDx-DR*¹²⁶ dans la recherche de rétinopathie ;

¹²⁴ Source : <https://cardiologs.com/> consulté le 25 janvier 2019.

¹²⁵ Source : <http://i-nside.com/> consulté le 25 janvier 2019.

¹²⁶ Source : <https://www.eyedagnosis.co/> consulté le 25 janvier 2019.

- *Therapixel*¹²⁷ pour le diagnostic du cancer du sein ;
- *VitaDx*¹²⁸ pour la détection des cellules du cancer de la vessie dans les urines ;
- *FreeStyle*¹²⁹ de Abbott dans le suivi continu de la glycémie chez les patients diabétiques ;
- *Owkin*¹³⁰ dans l'analyse et le traitement des données de grands volumes en partenariat avec l'INSERM sur certains projets.

Les principales difficultés de la mise en œuvre de l'IA sont d'ordre légales et concernent l'accès aux données de santé, le partage de ces données et le consentement des patients pour l'utilisation de leurs données. Ces dernières années, le droit a été très prolifique en matière de textes pour la protection des données personnelles dans le cadre de l'IA.

¹²⁷Source : <https://www.therapixel.com/> consulté le 25 janvier 2019.

¹²⁸Source : <http://vitadx.com/> consulté le 25 janvier 2019.

¹²⁹Source : <https://www.freestylelibre.fr/libre/> consulté le 25 janvier 2019.

¹³⁰Source : <https://owkin.com/> consulté le 25 janvier 2019.

3.3 Textes juridiques de protection des données de santé

Ces dernières années, une accélération de la production de textes législatives à la fois au niveau national et supra national relative à l'exploitation des données personnelles a été observée. Cet état de fait s'explique par les enjeux (économiques, juridiques, politiques, sociétaux, éthiques) importants dus aux capacités de traitements de ces données personnelles de nos jours. Toutes ces adaptations et évolutions ont pour but de réduire les risques de non-respect des droits fondamentaux des individus.

La France a été pionnière en la matière avec la création de la loi informatique et libertés (LIL) en 1978 à la suite de la crise sur le projet SAFARI que nous avons expliqué plus haut. En 1995, l'Union Européenne rédigeait une directive sur la protection des données personnelles, mais cette directive était appliquée de manière disparate par les États membres de l'UE. Face à cette disparité d'application, le parlement européen a adopté le Règlement Général sur la Protection des Données (RGPD¹³¹) en avril 2016. Celui-ci est entré en vigueur le 25 mai 2016 dont l'application est placée, en France sous l'autorité de la Commission nationale de l'informatique et des libertés (CNIL).

Encadré 2: instruments juridiques européens¹³²

Les instruments juridiques européens désignent les instruments dont disposent les institutions européennes pour l'accomplissement de leur mission. Ces instruments énumérés par l'article 288 du traité sur le fonctionnement de l'Union européenne sont les suivants :

- le règlement : obligatoire dans tous ses éléments, il est directement applicable dans tous les États membres ;
- la directive : elle lie les pays de l'UE quant aux résultats à atteindre. La directive nécessite une transposition dans le cadre juridique national et laisse une marge de manœuvre quant à la forme et aux moyens de la mise en œuvre ;
- la décision : obligatoire dans tous ses éléments, elle lie les destinataires qu'elle désigne expressément ;

¹³¹ Règlement (UE) 2016/679 du parlement européen et du conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données).

¹³² Source : https://eur-lex.europa.eu/summary/glossary/community_legal_instruments.html?locale=fr consulté le 23 mai 2019.

- la recommandation et l'avis : non obligatoires, ils possèdent un caractère déclaratoire.

La loi n° 2018-493 du 20 juin 2018¹³³ est venue modifier la loi Informatique et Libertés afin de mettre en conformité le droit français avec le cadre juridique européen. Elle permet la mise en œuvre concrète du RGPD.

Le RGPD apporte plusieurs avancées majeures dans la protection des données personnelles. Il crée un cadre unifié pour l'ensemble des États membres de l'UE et offre ainsi une même protection aux Européens quant au traitement des données. Elle renvoie la responsabilité de la protection des données personnelles aux acteurs économiques ; chaque entreprise concernée doit prouver sa mise en conformité avec le RGPD. Il n'y a pas de différence notable entre le RGPD et la loi Informatique et Libertés, le premier complétant désormais la seconde.

Le RGPD laisse aux États membres des articles pouvant faire l'objet d'une adaptation interne. La France a adapté le RGPD sur certains points précis :

- la régulation de l'utilisation des algorithmes dans le secteur public. Tout ce qui est lié à la sécurité nationale n'est pas concerné par le RGPD car est une compétence nationale ;
- l'âge du consentement numérique est porté à 15 ans contre 16 ans dans le RGPD afin d'homogénéiser la majorité numérique avec la majorité dans la vie courante ;
- la loi renforce les compétences de la CNIL dans la mise en œuvre et le contrôle de la loi.

Les deux textes juridiques définissent les principes à respecter lors de la collecte, du traitement et de la conservation de ces données, garantissent un certain nombre de droits pour les personnes, donnent des directives sur la sécurité des données et les conditions de transfert des données hors de l'UE.

Selon ces textes juridiques, la donnée est une émanation de la personnalité juridique d'une personne physique. A l'instar, de l'identité de l'individu, elle bénéficie de la même protection juridique relevant de la vie privée des individus. Toute collecte, stockage,

¹³³ Loi n° 2018-493 du 20 juin 2018 relative à la protection des données personnelles.

traitement, exploitation et circulation des données personnelles sont soumis à l'information préalable et au consentement. Dans le cadre des données de santé, qui sont classées comme des données sensibles, il faut des accords des autorités compétentes avant toutes collectes, le consentement de l'individu ne suffit pas à elle seule.

3.3.1 Définitions

« La donnée à caractère personnel est considérée comme toute information se rapportant à une personne physique identifiée ou identifiable (ci-après dénommée « personne concernée ») ; est réputée être une « personne physique identifiable » une personne physique qui peut être identifiée, directement ou indirectement, notamment par référence à un identifiant, tel qu'un nom, un numéro d'identification, des données de localisation, un identifiant en ligne, ou à un ou plusieurs éléments spécifiques propres à son identité physique, physiologique, génétique, psychique, économique, culturelle ou sociale »¹³⁴.

« Les données à caractère personnel concernant la santé devraient comprendre l'ensemble des données se rapportant à l'état de santé d'une personne concernée qui révèle des informations sur l'état de santé physique ou mentale passé, présent ou futur de la personne concernée. Cela comprend des informations sur la personne physique collectées lors de l'inscription de cette personne physique en vue de bénéficier de services de soins de santé ou lors de la prestation de ces services au sens de la directive 2011/24/UE du Parlement européen et du Conseil au bénéfice de cette personne physique ; un numéro, un symbole ou un élément spécifique attribué à une personne physique pour l'identifier de manière unique à des fins de santé ; des informations obtenues lors du test ou de l'examen d'une partie du corps ou d'une substance corporelle, y compris à partir de données génétiques et d'échantillons biologiques ; et toute information concernant, par exemple, une maladie, un handicap, un risque de maladie, les antécédents médicaux, un traitement clinique ou l'état physiologique ou biomédical de la personne concernée, indépendamment de sa source, qu'elle

¹³⁴ Article 4 considérant 1 du RGPD.

proviennent par exemple d'un médecin ou d'un autre professionnel de la santé, d'un hôpital, d'un dispositif médical ou d'un test de diagnostic in vitro »¹³⁵.

3.3.2 Principes

Trois principes sont clairement identifiés dans le RGPD : « *Les données à caractère personnel doivent être [...] traitées de manière licite, loyale et transparente au regard de la personne concernée* »¹³⁶

Principe de la collecte loyale

La loyauté de la collecte est un pilier fondamental de la nouvelle loi sur les données personnelles. Ce principe définit la légalité de la collecte des données avant la question même du traitement des données.

« Le principe de traitement loyal et transparent exige que la personne concernée soit informée de l'existence de l'opération de traitement et de ses finalités. Le responsable du traitement devrait fournir à la personne concernée toute autre information nécessaire pour garantir un traitement équitable et transparent, compte tenu des circonstances particulières et du contexte dans lesquels les données à caractère personnel sont traitées. [...] Lorsque les données à caractère personnel sont collectées auprès de la personne concernée, il importe que celle-ci sache également si elle est obligée de fournir ces données à caractère personnel et soit informée des conséquences auxquelles elle s'expose si elle ne les fournit pas »¹³⁷.

Le principe de loyauté est ainsi lié à deux obligations que sont :

- l'obligation de finalités : « *... collectées pour des finalités déterminées, explicites et légitimes, et ne pas être traitées ultérieurement d'une manière incompatible avec ces finalités ; le traitement ultérieur à des fins archivistiques dans l'intérêt public, à des fins de recherche scientifique ou historique ou à des fins statistiques n'est*

¹³⁵ Considérant 35 du RGPD.

¹³⁶ Article 5 du RGPD.

¹³⁷ RGPD considérant 60.

pas considéré, conformément à l'article 89, paragraphe 1, comme incompatible avec les finalités initiales (limitation des finalités) »¹³⁸ ;

- *l'obligation de traitement adéquat, transparent et limité « ... adéquates, pertinentes et limitées à ce qui est nécessaire au regard des finalités pour lesquelles elles sont traitées (minimisation des données) »¹³⁹;*

Les principales difficultés de la mise en œuvre de ce principe dans le cadre du *big data* sont le fait qu'il se base sur une collecte de quantité importante de données sans finalités précises et sans connaissance des traitements qui seront appliqués aux données au moment de la collecte. Les données de santé étant définies comme des données sensibles par le RGPD, en conséquence, toute collecte est interdite sauf pour motif légitime défini par le RGPD et complété par la loi informatique et libertés. La manière dont s'opère la collecte dans le cadre du *big data* ne permet pas aux promoteurs de respecter ce principe dès la mise en œuvre. En effet, l'exploitation des données se base sur des modèles exploratoires qui nécessitent la capture de toutes les traces possibles laissées par les individus de manière volontaire ou non.

Principe de licéité

« Le traitement n'est licite que si, et dans la mesure où, au moins une des conditions suivantes est remplie :

- *la personne concernée a consenti au traitement de ses données à caractère personnel pour une ou plusieurs finalités spécifiques ;*
- *le traitement est nécessaire à l'exécution d'un contrat auquel la personne concernée est partie ou à l'exécution de mesures précontractuelles prises à la demande de celle-ci ;*
- *le traitement est nécessaire au respect d'une obligation légale à laquelle le responsable du traitement est soumis ;*
- *le traitement est nécessaire à la sauvegarde des intérêts vitaux de la personne concernée ou d'une autre personne physique ;*
- *le traitement est nécessaire à l'exécution d'une mission d'intérêt public ou relevant*

¹³⁸ Article 5 du RGPD considérant b.

¹³⁹ Article 5 considérant c.

de l'exercice de l'autorité publique dont est investi le responsable du traitement »¹⁴⁰.

La notion de consentement de la personne apparaît dans cet article comme le point central de ce principe. Il faut un consentement libre comme défini dans l'article 4 du RGPD¹⁴¹. Pour chaque donnée et chaque finalité, la personne doit préciser son accord à l'utilisation des données. Néanmoins, le choix de la personne ne doit en aucun cas altérer ou le priver de l'utilisation des services¹⁴². Mais force est de constater que les individus sont obligés d'accepter l'utilisation de leurs données pour avoir accès à des services ; même pour les services facturés aux personnes. En plus, dans le cadre des relations commerciales entre les entreprises, les individus se retrouvent sans le savoir à partager leurs données avec des entreprises dont ils n'ont jamais été en contact.

Le législateur prévoit des exceptions dans le cadre du consentement défini à l'article 6 du RGPD. Parmi ces exceptions, il y a la mission d'intérêt public qui pose un problème puisqu'elle n'a pas été définie dans le RGPD et la loi informatiques et libertés.

« Il n'y a pas de définitions d'intérêt public dans le droit il s'agit de notion standard. En France, on le comprend sous la forme de service public, il faut que ça soit une mission de service public. Il va être défini au cas par cas et laisse une grande liberté aux législateurs et aux services concernés. [...] Il devient très difficile à contrôler »¹⁴³.

Principe de l'obligation de transparence

Ce principe incombe au responsable du traitement ; cette obligation porte sur le droit à l'information de la personne (informations sur les données collectées et les finalités), la gestion de la sécurité et de la confidentialité des données afin d'éviter la divulgation ou l'accès à des personnes non autorisées.

¹⁴⁰ Article 6 du RGPD.

¹⁴¹ On entend par « consentement de la personne concernée, toute manifestation de volonté, libre, spécifique, éclairée et univoque par laquelle la personne concernée accepte, par une déclaration ou par un acte positif clair, que des données à caractère personnel la concernant fassent l'objet d'un traitement ». Article 4 considérant 11 du RGPD.

¹⁴² Article 7 du RGPD.

¹⁴³ Entretien Juriste.

A cet effet, un ensemble d'informations doit être obligatoirement communiqué aux personnes concernées en fonction des cas (collecte directe auprès de la personne ou indirecte)¹⁴⁴ :

- l'identité et les coordonnées du responsable du traitement et du délégué à la protection des données (DPD) ;
- si les données sont transmises à un tiers, l'identité de ce tiers ;
- la finalité du traitement et sa base juridique ;
- l'existence d'un transfert hors UE dans certains cas ;
- les catégories de données à caractère personnel concernées ;
- la durée de conservation des données ;
- un rappel des différents droits des personnes ;

Concernant le principe d'obligation de transparence, « *la personne concernée a le droit de ne pas faire l'objet d'une décision fondée exclusivement sur un traitement automatisé, y compris le profilage, produisant des effets juridiques la concernant ou l'affectant de manière significative de façon similaire* »¹⁴⁵.

Le problème avec le principe de transparence se pose sur la question du comment informer une personne qui n'est pas au courant de l'utilisation de ses données. La collecte des données dans le cadre du *big data* se fait de manière systématique à travers Internet, les smartphones et les objets connectés. Dans le cas où les données sont conservées dans les bases de données médico-administratives, l'information des patients n'est pas faite directement ; l'utilisation des données entre dans le cadre de « *l'intérêt public* » dont la définition dépend de plusieurs facteurs.

¹⁴⁴ Article 13 et 14 du RGPD.

¹⁴⁵ Article 22 du RGPD.

3.3.3 *Droits des personnes*

Le RGPD consacre des droits aux personnes, deux nouveaux droits ont été ajoutés à savoir le droit à la limitation du traitement et le droit à la portabilité.

Droit à l'information

Le droit à l'information des personnes a un lien fort avec le principe d'obligation de transparence. Nous estimons qu'informer une personne de ses données collectées est une forme de transparence. Les personnes ont le droit de demander des informations sur les données collectées comme stipulé à l'article 15 considérant 1 du RGPD :

« La personne concernée a le droit d'obtenir du responsable du traitement la confirmation que des données à caractère personnel la concernant sont ou ne sont pas traitées et, lorsqu'elles le sont, l'accès auxdites données à caractère personnel ainsi que les informations suivantes :

- *les finalités du traitement ;*
- *les catégories de données à caractère personnel concernées ;*
- *les destinataires ou catégories de destinataires auxquels les données à caractère personnel ont été ou seront communiquées, en particulier les destinataires qui sont établis dans des pays tiers ou les organisations internationales ;*
- *lorsque cela est possible, la durée de conservation des données à caractère personnel envisagée ou, lorsque ce n'est pas possible, les critères utilisés pour déterminer cette durée ;*
- *l'existence du droit de demander au responsable du traitement la rectification ou l'effacement de données à caractère personnel, ou une limitation du traitement des données à caractère personnel relatives à la personne concernée, ou du droit de s'opposer à ce traitement ;*
- *le droit d'introduire une réclamation auprès d'une autorité de contrôle ;*
- *lorsque les données à caractère personnel ne sont pas collectées auprès de la personne concernée, toute information disponible quant à leur source;*
- *l'existence d'une prise de décision automatisée, y compris un profilage, visée à l'article 22, paragraphes 1 et 4, et, au moins en pareils cas, des informations utiles concernant la logique sous-jacente, ainsi que l'importance et les conséquences*

prévues de ce traitement pour la personne concernée »¹⁴⁶.

Le responsable du traitement fournit une copie des données à caractère personnel faisant l'objet d'un traitement aux personnes à leur demande. Il peut exiger le paiement de frais raisonnables fondés sur les coûts administratifs pour toute copie supplémentaire demandée par la personne concernée¹⁴⁷.

Le droit à l'information est très utile dans les cas où vous voulez vérifier si les informations vous concernant ont été supprimées ou pas, ou dans les cas où vous voulez vérifier si vos données ont été traitées ou pas sans votre consentement.

Droit d'accès

En plus du droit à l'information, la personne a droit d'accès aux données personnelles le concernant détenues par un organisme. Elle peut avoir accès à ces données sous réserve de faire la demande et de justifier de son identité. Elle peut mandater une tierce personne avec un écrit présentant l'objet du mandat.

Le responsable de traitement doit s'assurer de la communication des informations dans un délai d'un mois avec possibilité de prolongation d'un mois en fonction de la complexité de la demande.

Droit de rectification et droit à l'effacement

Le droit à la rectification est un droit qui permet à la personne de modifier les informations le concernant pour les mettre à jour ;

« La personne concernée a le droit d'obtenir du responsable du traitement, dans les meilleurs délais, la rectification des données à caractère personnel la concernant qui sont inexactes. Compte tenu des finalités du traitement, la personne concernée a le droit d'obtenir que les données à caractère personnel

¹⁴⁶ Article 15 considérant 1 du RGPD.

¹⁴⁷ Article 15 considérant 3 du RGPD.

incomplètes soient complétées, y compris en fournissant une déclaration complémentaire »¹⁴⁸.

Le droit à l’effacement ou « *droit à l’oubli* » devient un peu plus complexe en ce sens que la personne doit motiver les raisons de la demande de son effacement. L’article 17 du RGPD précise les motifs pouvant faire valider ce droit¹⁴⁹ :

- les données à caractère personnel ne sont plus nécessaires au regard des finalités pour lesquelles elles ont été collectées ou traitées d'une autre manière ;
- la personne concernée retire le consentement sur lequel est fondé le traitement ;
- la personne concernée s'oppose au traitement et il n'existe pas de motif légitime impérieux pour le traitement ;
- les données à caractère personnel ont fait l'objet d'un traitement illicite ;
- les données à caractère personnel doivent être effacées pour respecter une obligation légale qui est prévue par le droit de l'Union ou par le droit de l'État membre auquel le responsable du traitement est soumis ;
- les données à caractère personnel ont été collectées dans le cadre de l'offre de services.

Le droit à l’oubli se limite dans le cadre de l’UE. Or, de nos jours, les données voyagent plus que les propriétaires des données dans le *cloud computing* et l’informatique distribuée. Par conséquent, nous pouvons être oubliés dans l’UE mais toujours visible dans le reste du monde. En plus, l’utilisation des *VPN (Virtual Personal Network)* permet aux individus de changer leur situation géographique de manière virtuelle tout en restant physiquement présent sur le territoire de l’UE. Elle reste ainsi un droit difficilement applicable dans la pratique. Pour une effectivité franche, il faut une collaboration internationale, ce qui ne sera pas mis en place tout de suite au regard de la guerre entre les grandes puissances et les géants de l’Internet dont le modèle économique est basé sur les données.

¹⁴⁸ Article 16 du RGPD.

¹⁴⁹ Article 17 considérant 1 du RGPD.

Droit à la limitation du traitement

Ce nouveau droit est créé pour évoluer vers la pratique du traitement des données. Normalement, une fois que les données sont effacées, elles doivent l'être dans toutes les bases de données et les organismes auxquels la personne a donné son consentement pour le traitement des données. Mais dans la réalité, cela est très difficile voire impossible dans certains cas. Ce droit permet de restreindre le traitement des données au strict minimum.

Il existe néanmoins des conditions pour la limitation du traitement des données par le responsable de traitement que sont¹⁵⁰ :

- l'exactitude des données à caractère personnel est contestée par la personne concernée, pendant une durée permettant au responsable du traitement de vérifier l'exactitude des données à caractère personnel ;
- le traitement est illicite et la personne concernée s'oppose à leur effacement et exige à la place la limitation de leur utilisation ;
- le responsable du traitement n'a plus besoin des données à caractère personnel aux fins du traitement mais celles-ci sont encore nécessaires à la personne concernée pour la constatation, l'exercice ou la défense de droits en justice.

Droit à la portabilité

Le droit à la portabilité¹⁵¹ permet d'obtenir toutes les données que la personne a communiquées à un organisme dans un format ouvert et interopérable. Il peut transmettre ces données à un autre organisme dans le cadre de demande de service. La personne peut récupérer directement ses données ou demander que les données soient directement transmises à l'organisme de son choix, si cela est techniquement possible.

¹⁵⁰ Article 18 du RGPD.

¹⁵¹ Article 20 du RGPD.

3.3.4 *Sécurité et confidentialité des données*

Le responsable du traitement met en œuvre des mesures techniques et organisationnelles appropriées pour s'assurer et être en mesure de démontrer que le traitement est effectué conformément au RGPD ; à savoir garantir la confidentialité et la sécurité des données. Ces mesures sont réexaminées et actualisées si nécessaire.

Chaque traitement de données nécessite de mettre en œuvre toutes les mesures de protection possibles en fonction de l'état de l'art afin d'assurer la confidentialité, l'intégrité et la sécurité des données. Le RGPD propose quatre moyens que sont¹⁵² :

- la « *pseudonymisation*¹⁵³ » et le chiffrement des données à caractère personnel ;
- des moyens permettant de garantir la confidentialité, l'intégrité, la disponibilité et la résilience constantes des systèmes et des services de traitement ;
- des moyens permettant de rétablir la disponibilité des données à caractère personnel et l'accès à celles-ci dans des délais appropriés en cas d'incident physique ou technique ;
- une procédure visant à tester, à analyser et à évaluer régulièrement l'efficacité des mesures techniques et organisationnelles pour assurer la sécurité du traitement.

Le dernier point montre la liberté laissée aux organismes pour la gestion de la sécurité des données. Mais, toutes les activités doivent être notées dans un registre afin d'assurer la traçabilité et de permettre les vérifications lors des contrôles de la CNIL.

Le responsable au traitement a une obligation de notifier les failles de sécurité à la CNIL dans le cadre de la France dans les meilleurs délais (trois jours après avoir pris connaissance de la faille). La notification doit comporter au moins les informations suivantes¹⁵⁴ :

- décrire la nature de la violation des données à caractère personnel y compris, si possible, les catégories et le nombre approximatif de personnes concernées par la

¹⁵² Article 32 du RGPD.

¹⁵³ « La « *pseudonymisation* » est le traitement de données à caractère personnel de telle façon que celles-ci ne puissent plus être attribuées à une personne concernée précise sans avoir recours à des informations supplémentaires, pour autant que ces informations supplémentaires soient conservées séparément et soumises à des mesures techniques et organisationnelles afin de garantir que les données à caractère personnel ne sont pas attribuées à une personne physique identifiée ou identifiable ». Considérant 5 du RGPD.

¹⁵⁴ Article 34 du RGPD.

violation et les catégories et le nombre approximatif d'enregistrements de données à caractère personnel concerné ;

- communiquer le nom et les coordonnées du délégué à la protection des données ou d'un autre point de contact auprès duquel des informations supplémentaires peuvent être obtenues ;
- décrire les conséquences probables de la violation de données à caractère personnel ;
- décrire les mesures prises ou que le responsable du traitement propose de prendre pour remédier à la violation de données à caractère personnel, y compris, le cas échéant, les mesures pour en atténuer les éventuelles conséquences négatives.

Lorsqu'une violation de données à caractère personnel¹⁵⁵ est susceptible d'engendrer un risque élevé pour les droits et libertés d'une personne physique, le responsable du traitement communique la violation de données à caractère personnel à la personne concernée dans les meilleurs délais.

Face à ces risques de plus en plus importants et le niveau de responsabilité des responsables de traitement, il est important de faire une analyse d'impact (*Privacy Impact Assessment PIA*). Cette analyse d'impact relative à la protection des données devrait être effectuée par le responsable du traitement, préalablement au traitement, en vue d'évaluer la probabilité et la gravité particulières du risque élevé, compte tenu de la nature, de la portée, du contexte et des finalités du traitement et des sources du risque. Cette analyse d'impact devrait comprendre, notamment, les mesures, les garanties et les mécanismes envisagés pour atténuer ce risque, assurer la protection des données à caractère personnel et démontrer le respect du RGPD¹⁵⁶.

¹⁵⁵ On entend par « violation de données à caractère personnel, une violation de la sécurité entraînant de manière accidentelle ou illicite, la destruction, la perte, l'altération, la divulgation non autorisée de données à caractère personnel transmises, conservées ou traitées d'une autre manière, ou l'accès non autorisé à de telles données ». Article 4 considérant 12 du RGPD.

¹⁵⁶ Considérant 90 du RGPD.

« L'analyse d'impact relative à la protection des données est, en particulier, requise dans les cas suivants :

- *l'évaluation systématique et approfondie d'aspects personnels concernant des personnes physiques, qui est fondée sur un traitement automatisé, y compris le profilage, et sur la base de laquelle sont prises des décisions produisant des effets juridiques à l'égard d'une personne physique ou l'affectant de manière significative de façon similaire ;*
- *le traitement à grande échelle de catégories particulières de données visées à l'article 9, paragraphe 1, ou de données à caractère personnel relatives à des condamnations pénales et à des infractions visées à l'article 10 ;*
- *la surveillance systématique à grande échelle d'une zone accessible au public »¹⁵⁷*

Dans le domaine de la santé, pour assurer une meilleure confidentialité et sécurité des données, l'état a mis en place l'agrément des hébergeurs de données de santé qui a évolué vers la certification des hébergeurs de données de santé. Le but recherché est d'amener les acteurs de l'hébergement au plus haut niveau de sécurité afin d'offrir un espace de confiance aux patients et aux professionnels. L'activité d'hébergement consiste dans l'organisation du dépôt et de la conservation des données personnelles de santé, afin d'assurer leur pérennité et leur confidentialité. Un contrat est passé entre l'hébergeur et la personne ou l'organisme à l'initiative du dépôt de ces données¹⁵⁸. Les hébergeurs doivent journaliser les accès, réussis ou en échec, et les actions effectuées par tous les intervenants sur les systèmes et les données de santé¹⁵⁹.

3.3.5 Transfert des données hors de l'UE

La protection des données dans le cadre du RGPD concerne tous les citoyens de l'UE ; par conséquent les règles de sortie des données hors de l'UE et de leur traitement ont été renforcées.

¹⁵⁷ Article 35 considérant 3.

¹⁵⁸ Article R.1111-13 du code de la santé publique.

¹⁵⁹ Article R.1111-14 du code de la santé publique.

Le consentement de la personne est nécessaire avant le transfert des données hors de l'UE. Le transfert des données fait l'objet de communication à la personne lors de l'exercice de son droit d'accès ou de son droit d'information.

Tous les pays n'assurent pas un niveau de protection égal à celui du RGPD et par conséquent la CNIL doit donner son avis sur le transfert des données hors de l'UE ;

« Un transfert de données à caractère personnel vers un pays tiers ou à une organisation internationale peut avoir lieu lorsque la Commission a constaté par voie de décision que le pays tiers, un territoire ou un ou plusieurs secteurs déterminés dans ce pays tiers, ou l'organisation internationale en question assure un niveau de protection adéquat. Un tel transfert ne nécessite pas d'autorisation spécifique »¹⁶⁰

Les bases de la technologie du *big data* sont une collecte systématique et massive de toutes les données accessibles sans finalités de départ afin de faire émerger des corrélations. Cela a pour conséquence de ne pas faciliter le respect du principe de loyauté et de la licéité des données. En plus ; certains droits (droit à l'effacement ou à la limitation du traitement) sont difficiles à mettre en œuvre quand on sait qu'il existe un marché des données et que les entreprises qui les collectent, les partagent avec d'autres entreprises moyennant des services ou de l'argent. La question de la traçabilité des données après le partage avec des entreprises tierces reste un enjeu à prendre en compte dans les attributions des autorisations des accès. La sécurité des données incombe au responsable du traitement et il doit mettre en place les moyens nécessaires pour assurer leurs protections en tenant compte de la valeur des données et de l'état de l'art ; mais comment évalue-t-on la valeur des données personnelles ? Que se passe-t-il si le responsable du traitement crée une faille de sécurité volontaire dans son système ? Autant de questions que l'on peut se poser.

Des sanctions financières importantes, des peines d'emprisonnement et une augmentation de la responsabilité des organismes en matière de respect des traitements des données ont été adoptées par le RGPD. Ce sont des éléments importants dans la régulation, mais ces sanctions¹⁶¹ comparées à la valeur que peut apporter les données sont dérisoires. Cette

¹⁶⁰ Article 48 du RGPD.

¹⁶¹ Cinq milliards de dollars, c'est le montant de l'amende votée aux Etats-Unis par la *Federal Trade Commission* (FTC) à l'encontre de Facebook. Ce montant représente environ 10% du chiffre d'affaires de l'entreprise sur 2018. [...] Il est effectivement légitime de s'interroger sur l'impact de cette sanction pour le géant américain. Avant le prononcé définitif

faiblesse encourage certaines entreprises à courir le risque de violer les lois sur la protection des données D'autant plus, que la déclaration n'est plus obligatoire et que la preuve de la défaillance doit être faite lors d'un contrôle de la CNIL ou lorsqu'une plainte a été déposée

Bien qu'il y ait eu des avancées en matière de lois sur la protection des données personnelles, elles ne prennent pas en compte certaines caractéristiques des données et en particulier les données de santé :

- les lois ne sont pas écrites dans le cadre d'une protection globale des données individuelles sur le plan sociétal ; mais dans le cadre d'une protection des individus avec le consentement comme base ; il se pose la question de la limite entre la vie privée et l'intimité. Les individus n'appréhendent pas les données sur la vie privée de la même manière et n'accordent pas également la même valeur. Par exemple dans le cas de l'information sur les maladies, certaines personnes divulguent les données de leur maladie et en parlent sans problème et d'autres personnes préfèrent les garder secrètes.
- Les lois ne prennent pas en compte le fait qu'il est impossible d'anonymiser certaines données de santé notamment les données génomiques. Elles constituent un identifiant fiable de l'individu si on fait le lien avec son état civil. Quand elles sont anonymisées, elles perdent toutes leurs valeurs scientifiques et ne permettent pas de développer des applications d'IA ou de faire de la recherche. Comment respecter les lois en ce qui concerne la protection des données de santé ? Ceux d'autant plus qu'il existe des multitudes de bases de données génomiques disponibles en accès libre en ligne et/ou auprès des entreprises de séquençage d'ADN (acide désoxyribonucléique) – la majorité sont américaines – dont le modèle économique est fondé sur la valeur marchande des données.

« *L'intérêt public* » est l'un des arguments avancés pour assouplir les conditions d'accès des données dans le cadre de la recherche aux entreprises et aux chercheurs. Il est laissé à l'appréciation des juges et des comités pour l'appréciation de l'intérêt public des projets. Il est le « *mot-clef* » qui fait sauter tous les verrous de protection des données personnelles et

de cette amende, Facebook avait déjà provisionné plus de 3 milliards de dollars. Bien que considérable, le montant de l'amende semble bien faible rapporté aux bénéfices colossaux générés par une décennie d'utilisation abusive des données personnelles. Source : <https://www.linkedin.com/pulse/confirmation-de-lamende-prononc%C3%A9e-contre-facebook-pierre-hervieu/> consulté le 30 novembre 2019.

en particulier ceux concernant les données de santé. En effet, quand on analyse le nombre de services public et d'acteurs^{162,163} ayant accès aux bases de données de santé au nom de l'intérêt public, on se pose des questions sur l'intérêt de la confidentialité des données et du secret médical, d'autant plus que ces services et acteurs peuvent faire de croisement des bases de données.

Au moment de la finalisation de notre thèse en avril 2020, l'épidémie du Covid19 sévit en France et l'heure est au confinement. En vue du déconfinement progressif prévu à partir du 11 mai, l'État envisage, au nom de l'intérêt public et national, de mettre en place l'application « *StopCovid* »¹⁶⁴. Ce dispositif aurait pour objectif *le tracking* des patients atteints de la maladie et de leurs contacts.

Notre recherche nous conduit à les inviter avant toute décision à examiner les points suivants :

- est-il vraiment justifié d'utiliser une application dans lequel le consentement des personnes n'est plus requis pour la collecte et le traitement des données ?
- ne faut-il pas auparavant mieux circonscrire et mesurer la menace que fait peser la pandémie sur la nation ? et comparer les deux risques (épidémie vs atteinte à la liberté individuelle et démocratique) ?
- n'est-il pas nécessaire que les politiques comprennent quels paramètres seront captés, comment les données seront traitées et les conséquences éthiques de leurs choix ?
- comment s'engager avec certitude au fait que les captations et la durée du traitement des données seront limitées ?
- saisir l'opportunité d'instaurer une surveillance numérique généralisée sous le prétexte de limiter la propagation d'un virus ne constitue-t-il pas une véritable rupture sociale dans le traitement passé des épidémies ?

¹⁶² Liste des services publics et acteurs ayant accès au DMP dans le cadre de l'intérêt public : Article L. 312-1 du code de l'action sociale et des familles.

¹⁶³ Liste des services publics et acteurs ayant accès aux SNDS dans le cadre de l'intérêt public. Article R 1461-12 du décret n° 2016-1871 du 26 décembre 2016 relatif au traitement de données à caractère personnel dénommé « système national des données de santé ».

¹⁶⁴Source : <https://www.linternaute.com/actualite/guide-vie-quotidienne/2492203-stop-covid-macron-confirme-que-le-gouvernement-travaille-sur-une-appli-de-tracage-numerique/> 28 avril 2020.

Toutes ces questions mériteraient d'autant plus d'être posées que, au moment où les politiques envisagent cette mise en place, il n'existe aucune preuve de l'efficacité du *tracking* de la population pour des raisons de prévention de risques sanitaires. En contrepartie, cette mesure fait courir des risques de sécurité liés à la centralisation des données.

Les lois permettent une régulation de la vie en société, mais elle se montrent insuffisantes à certains moments ; c'est le cas notamment de la protection des données personnelles dans le cadre de leur utilisation dans le *big data* et l'IA. Une éthique dans l'utilisation des données est nécessaire aussi bien pour les personnes que pour les entreprises afin d'assurer la protection des données personnelles. Dans le domaine de la santé, l'éthique médicale permet une identification des enjeux éthiques en fonction de l'évolution de la société, l'évaluation des bénéfices et des risques et une proposition de solutions.

3.4 Ethique

*« Celui qui ne peut pas vivre en société,
ou qui n'a besoin de rien parce qu'il se suffit à lui-même,
ne fait pas partie de l'Etat ; c'est une brute ou un Dieu »*

Aristote, éthique à Nicomaque

Étymologiquement, « morale » vient du latin (*philosophia*) *moralis*, traduction par Cicéron, du grec « *ta êthica* » ; les deux termes désignent ce qui a trait aux mœurs, au caractère, aux attitudes humaines en général et, en particulier, aux règles de conduite et à leur justification (WEIL, 2018).

Le Petit Robert définit la morale comme : « *Science du bien et du mal ; théorie de l'action humaine en tant qu'elle est soumise au devoir et a pour but le bien* » ; « *ensemble des règles de conduite considérées comme bonnes de façon absolue* » (COLLECTIF, 2013).

Pour Paul Ricoeur, « *la morale a une double fonction, celle de désigner, d'une part, la région des normes, autrement dit des principes du permis et du défendu, d'autre part, le sentiment d'obligation en tant que face subjective du rapport d'un sujet à des normes* » (RICOEUR, 2000a).

La morale est une obligation pour l'individu de respecter des normes, des valeurs dans une société, mais en même temps un processus d'intériorisation de ses normes en lui dans le cadre de ses actions. Ce processus d'intériorisation des normes se fait tout au long de la vie de l'individu dans la société soit par apprentissage ou par observation des faits et gestes de la communauté dans laquelle il grandit. La société a une importance capitale, en ce sens que, les normes seront différentes d'une société à une autre ; entraînant parfois des conflits dans le processus de décision et d'action de l'individu (LE COZ, KACHANER, 2010). Certaines sociétés se focalisent sur les intentions qui président les actions, d'autres sur les conséquences de ces actions. La morale renvoie ainsi au rapport avec l'autre (l'altérité). Nos actions doivent tenir compte de cette altérité, des règles morales et du droit de la République.

Le mot « éthique » prend son origine dans le terme grec *ethos* signifiant « *les mœurs* » (Cicéron), « *les habitudes* » (Platon et Aristote).

L'éthique est définie dans Le Petit Robert comme étant l'« *ensemble des valeurs, des règles morales propres à un milieu, une culture, un groupe* » (COLLECTIF, 2013).

Cette définition du dictionnaire nous ramène à la définition de l'éthique dans le cadre de la pratique d'une profession.

Dans les années 1990, les nouvelles pratiques de communication publique, le développement des relations publiques et de la publicité, l'émergence de la communication organisationnelle et institutionnelle ont induit un questionnement continu d'ordre moral. Les colloques sur le sujet se multipliaient parmi lesquels on peut citer, en SIC, le colloque *Ethique et communication* de l'Institut des Sciences de l'Information et de la Communication (ISIC) de l'université de Bordeaux III dont les actes ont été publiés en 1990¹⁶⁵. En parallèle, aux États unis, Pettit, Vaught et Pulley (PETTIT ET AL., 1990) proposaient un article offrant « *un cadre théorique d'éthique, de pouvoir et de communication au travail. (...) Un modèle d'éthique conceptuelle développé pour fournir une toile de fond pour visualiser les décisions éthiques (...)* ». L'éthique en milieu de travail y est discutée en montrant « *comment les gestionnaires sont souvent pris dans un dilemme entre les pressions du travail et leur propre code d'éthique personnel (...)* » et comment « *la communication est proposée comme moyen de favoriser un comportement plus éthique dans les organisations* ». Apparaît alors la nécessité de l'acquisition d'une compétence spécifique en matière de communication (GAUTHIER, 1992). Ce qui conduit Gilles Gauthier à instaurer « *la notion d'inhérence* », idée selon laquelle « *la question éthique est intrinsèque à la communication des organisations* » (GAUTHIER, 1992) et de préciser :

« *L'un des grands points de tension de l'éthique est qu'elle est une entreprise par nature théorique qui, par ailleurs, prétend avoir un champ d'application pratique. L'ambition de l'éthique, c'est de fournir une connaissance objective de la conduite idéale. Elle comporte donc une visée cognitive relative à l'action concrète. Le rapport entre le caractère théorique de l'éthique et sa prétendue portée pratique a souvent, dans le passé, été pensé davantage comme un écart que comme un véritable lien. (...) Par l'éthique appliquée, on veut évidemment désigner un certain type de démarche éthique qui au lieu de consister en des spéculations désincarnées porterait immédiatement sur des séquences ou des pans de conduite bien identifiés* » (GAUTHIER, 1992).

¹⁶⁵ HOTIER H. (dir.). *Ethique et communication*. Actes du 5^{ème} colloque Entreprise et communication, Bordeaux, ISIC, 1990.

Dans notre cas - l'éthique médicale -, il s'agit donc bien d'une éthique appliquée en vue de mettre en œuvre la réglementation d'une pratique médicale vertueuse vis-à-vis des patients et de la communauté. Elle fait appel à des principes et des valeurs qui sont acceptés par l'ensemble de la communauté de la profession et qui prend en compte les règles morales de la société dans laquelle s'opère la pratique médicale. La valeur correspond à ce à quoi nous accordons de l'importance, de l'estime. Des valeurs naissent des principes¹⁶⁶ et les principes seront traduits en règle pour guider les actions.

L'éthique est ici une réflexion sur les comportements à adopter et les actions à mener. « *Le but de la réflexion éthique n'est pas de déterminer les valeurs les plus motivantes, sur le plan subjectif, mais celles qui peuvent justifier nos actions de façon rationnelle* »^{167,168}. La réflexion éthique permet de déterminer les valeurs qui constituent des raisons d'agir acceptables par l'ensemble de la société, par les personnes qui partagent l'idéal de pratique et, au niveau particulier, par les personnes et les groupes touchés par une décision.

Cependant, il faut intégrer l'idée que la réflexion éthique peut se positionner sur trois niveaux (PAYEN, 2012 ; BAUMANN, 2004) :

- Le niveau singulier : la situation concrète ;
- Le niveau particulier : celui du groupe et de ses règles ;
- Le niveau universel : celui de la promotion de l'humain en chacun.

L'éthique théorique se situe en amont de la déontologie et du droit dont elle a engendré certains principes fondamentaux. L'éthique appliquée invite le professionnel à réfléchir sur les valeurs qui motivent son action et à choisir, sur cette base, la conduite la plus appropriée. La réflexion éthique d'un acteur en situation de pratique professionnelle fait donc appel à l'autonomie, au jugement et au sens des responsabilités.

Dans le cadre de la médecine, ces actions se mènent dans le long terme et dans le temps avec des actions irréversibles. L'éthique est ainsi une réflexion et une analyse attentive et

¹⁶⁶ « Les principes sont des règles d'action s'appuyant sur un jugement de valeur et constituant un modèle, une règle et un but » (COLLECTIF, 2013). Il désigne également ce qui est à l'origine de quelque chose, qui en est la cause première.

¹⁶⁷ Source : http://gpp.oig.qc.ca/qu_est-ce_que_l_ethique.htm consulté le 24 mai 2019.

¹⁶⁸ Le problème est de savoir quelle rationalité ou plus exactement la rationalité de qui sera prise en compte. C'est là qu'intervient la notion d'altérité, cet autre soi-même et pourtant si différent.

systématique des décisions et comportements moraux, passés, présents ou futurs. Elle est ainsi le fruit d'un débat et d'une argumentation contradictoire, qui évolue dans le temps et en fonction de l'évolution de la société.

3.4.1 *Ethique médicale ou théorie du principisme*

Hippocrate - médecin grec au V^e siècle av. JC - a promu une conception de la responsabilité en médecine à travers le serment qui porte son nom depuis des siècles. Cette responsabilité prône le respect d'autrui comme une personne, la confidentialité des informations reçues, le secret médical et la recherche du bien pour la personne que l'on soigne quel que soit son statut social. L'exercice de la médecine doit s'appuyer sur des compétences techniques et scientifiques, mais aussi sur des vertus telles que l'honnêteté, la compassion, l'habileté, la fidélité, la confiance et l'altruisme. Il mettait en place les bases de l'éthique médicale dans la pratique de la médecine.

« L'éthique médicale est étroitement liée à la bioéthique (éthique biomédicale), sans toutefois lui être identique. Alors que l'éthique médicale s'intéresse principalement aux problèmes soulevés par l'exercice de la médecine, la bioéthique est un vaste sujet qui concerne les questions morales liées au développement des sciences biologiques de manière plus générale » (WILLIAMS, 2015).

L'éthique médicale est fondée sur l'étude des cas ; les questionnements éthiques sont de ce fait interdisciplinaires et nécessitent une collaboration étroite et franche de tous les acteurs intervenant dans la vie de l'homme.

La pratique de la médecine a toujours été une démarche éthique. L'action médicale se caractérise ainsi : *« son objet c'est autrui, son mobile c'est la pitié devant la souffrance d'autrui ; sa finalité c'est la santé d'autrui »* (JOLY, JOLY, 2010 ; RAMEIX, 1998 ; HOERNI, 2003). Cette démarche doit prendre en compte le patient dans sa singularité et non sa maladie dans le cadre de la prise de décision. La définition de la maladie, elle-même, étant variable en fonction de la personne et de la société dans laquelle se situe le malade.

Dans le cas classique de la morale, il s'agit de choisir entre le bien et le mal. Il est d'autant plus simple puisque l'homme reconnaît ce qui est mal et ne souhaite pas qu'on lui fasse cela ; même si dans la réalité cela est difficile. Dans le cas de la médecine, l'éthique pose un dilemme de choisir le bien parmi plusieurs biens. C'est là toute la difficulté du choix à faire et ceux d'autant plus que les effets de ce choix sur l'homme sont irréversibles. Claude Huriet a coutume de dire « *la morale, c'est la réponse avant la question et l'éthique, un questionnement sans réponse* » (SICARD, 2017).

Les raisons de l'interrogation sur l'éthique médicale (RAMEIX, 1998 ; BAUMANN, 2004 ; BENAROYO, 2014) remontent à plusieurs années :

- la pensée occidentale à la fin du XXe siècle marquée par un recul de la religion qui était les bases de la valeur morale. L'apparition de l'individualisme et une société pluraliste dans laquelle les valeurs ne sont plus partagées ;
- l'agir occidental en fin du XXe siècle : marqué par un développement important de la technique qui devient un prolongement de l'organe humain. Nous vivons dans un monde où tout est possible par la technique sans tenir compte de l'éthique. L'homme n'est plus le maître de la technique mais est manipulé par la technique ;
- l'histoire après Auschwitz marquée par l'euthanasie et l'expérimentation sur les hommes « *inférieurs* » ayant abouti au procès de Nuremberg et le code de Nuremberg. Selon le philosophe Allemand T Adorno, 1966 « *Hitler a imposé aux hommes un nouvel impératif catégorique : penser et agir en sorte qu'Auschwitz ne se répète pas* »¹⁶⁹ ; Cette citation devrait guider nos pensées et nos actions de sorte que certaines barbaries du passé ne puissent se répéter. La question que l'on doit se poser aujourd'hui est « *est ce que l'homme n'a-t-il pas oublié Auschwitz* » ? Cette question mérite d'être posée au regard de l'utilisation de la technologie de nos jours ;
- l'évolution de la médecine marquée par les découvertes scientifiques et techniques qui aboutissent à un oubli de l'homme malade (autrui) et se concentrent sur la technique et le savoir pour agir sur la maladie. On aboutit à un refus de la mort et du vieillissement entraînant des nouvelles revendications et des nouveaux droits. L'homme refuse la maladie et la mort comme faisant partie de son histoire. La

¹⁶⁹ ADORNO (1973) *Gesammelte Schriften*, 6, 358 in RAMEIX, SUZANNE 1998. *Fondements Philosophiques de l'Éthique Médicale*. Paris: Ellipses Marketing.

liberté et l'absence du colloque singulier entre le malade et son médecin font que les problèmes sont résolus dans l'espace public. La santé étant une affaire d'État ; il s'en suit des enjeux économiques, sociaux, juridiques et politiques ;

- la spécificité des problèmes posée par la médecine et la biologie contemporaine.

Le développement de la biotechnologie, l'évolution des responsabilités dans la pratique médicale, l'évolution des droits humains et le pluralisme social ont poussé les chercheurs dans le domaine de la santé à mettre en place un noyau consensuel de normes morales dans le cadre de la pratique médicale et de la recherche biomédicale (BENAROYO, 2004 ; WILLIAMS, 2015). Ces normes consensuelles ont été déclinées sous forme de principes au nombre de quatre (BEAUCHAMP, CHILDRESS, 2009 ; WILLIAMS, 2015) :

- le respect de l'autonomie : une norme qui prescrit de respecter les capacités de prise de décision des personnes autonomes ;
- la non-malfaisance : la norme qui prescrit d'éviter de causer du mal ;
- la bienfaisance : la norme qui vise à procurer des bénéfices et qui évalue les bénéfices par rapport aux risques et aux coûts ;
- la justice : la norme qui prescrit la répartition équitable des bénéfices, des risques et des coûts.

Bienfaisance

Le principe de bienfaisance est un principe exigeant car les personnes doivent prendre des mesures pour aider les autres et pas simplement s'abstenir de mal agir. La bienfaisance positive requiert que la personne apporte des bienfaits et l'utilité requiert que la personne pèse les avantages et les inconvénients de son action pour parvenir dans l'ensemble aux meilleurs résultats possibles (BEAUCHAMP, CHILDRESS, 2009 ; DONALDSON, 2012). Les formes de bienfaisance incluent habituellement la notion d'altruisme, d'amour et d'humanité. Dans son sens le plus large, toute action accomplie pour le bien d'autrui ; la bienfaisance se réfère au trait de caractère ou à la vertu qui consiste à être disposé à agir pour le bien d'autrui (HOERNI, 2003 ; TOURNEUR-BAGOT, 2019).

Les obligations d'apporter les bienfaits, d'empêcher et d'éliminer les méfaits et de peser les pour et les contre des avantages possibles, des coûts et des effets néfastes possibles sont des éléments essentiels de l'éthique médicale.

La morale commune ne contient pas de principe de bienfaisance qui exige des gros sacrifices et un altruisme extrême au sein de la vie morale. Nous ne sommes pas moralement contraints d'accomplir un acte de générosité ou de charité pour le bien d'autres personnes. Mais dans le cadre de la médecine, la bienfaisance exhorte le médecin à agir dans l'intérêt supérieur du patient et l'altruisme est l'essence même de la médecine. Dans la définition ancienne de la médecine, *health* signifie « *se mettre à l'abri de* » ; « *se confier* ». Si nous dissociions « *l'autre* », le patient qui se confie au médecin, de la pratique médicale, elle perd tout son sens.

La bienfaisance peut être considérée comme une obligation morale en fonction de laquelle les cliniciens peuvent être jugés, au lieu d'être considérée comme un idéal moral auquel les cliniciens doivent aspirer (HÆRNI, AVRIL, 2009). Cet idéal devrait être un puissant facteur de motivation pour tous les cliniciens. Tout acte médical devant être réalisé sur un patient doit apporter le plus de bénéfice possible et le clinicien doit avoir la conviction qu'il le fait pour le bien de celui-ci.

David Hume (HUME, MILLER, 1987) affirmait que l'obligation d'agir pour le bien d'autrui résulte des interactions sociales : « *toutes nos obligations d'agir pour le bien de la société semblent impliquer quelque chose de réciproque. Je reçois les bienfaits de la société, donc je dois promouvoir ses intérêts* ». La réciprocité est présente dans la vie sociale, sans pour cela réduire la vie morale dans son ensemble à une série d'obligations de réciprocité. Les médecins sont redevables à la société pour l'éducation et aux patients pour la recherche clinique et la pratique clinique. Le médecin a une obligation de réciprocité envers la société et le patient. Mais cette relation de réciprocité n'échappe pas à l'obligation d'agir pour le bien d'autrui. Promouvoir le bien-être des patients est le but, la raison d'être et la justification de la médecine ; le serment d'Hippocrate rédigé depuis des siècles faisait de l'engagement de produire des soins utiles pour le patient et à son avantage, son pilier.

Un exemple de bienfaisance est le principe de solidarité de prise en charge sanitaire en France : des personnes bien portantes cotisent pour la prise en charge des personnes qui ont des maladies nécessitant des soins continus. Ce même principe est à la base de la mise en place de la complémentaire maladie universelle permettant à des personnes indigentes d'avoir accès à des soins de meilleure qualité sans discrimination.

La bienfaisance représente le but premier et la raison d'être de la médecine et des soins. Ni le patient, ni le médecin n'ont une autorité première et prépondérante et aucun principe prééminent n'existe en éthique médicale, pas même l'obligation d'agir dans le meilleur intérêt du patient.

Respect de l'autonomie

Le respect d'autonomie est utilisé pour analyser la prise de décision de l'individu dans les domaines de la santé et de la recherche, et particulièrement les notions de consentement et de refus de soins (CONTANTINIDES, 2014). Le respect de l'autonomie est une obligation professionnelle pour le praticien de la santé et un droit pour le patient.

Le mot autonomie, dérivé du grec *autos* « auto » et *nomos* « règle » « gouvernement » « gouvernance » ou « loi », se réfère à l'origine à l'autorégulation ou à l'autogouvernement des cités indépendantes.

Les théories sur l'autonomie s'accordent sur le fait que deux conditions sont essentielles à l'autonomie (BEAUCHAMP, CHILDRESS, 2009) :

- la liberté, comme indépendance vis-à-vis des influences extérieures ;
- l'action possible comme capacité à agir intentionnellement.

Le respect de l'autonomie est le principe qui reconnaît le droit d'un patient d'avoir ses propres opinions, de faire des choix et d'agir en fonction de ses opinions et croyances particulières. Il implique l'obligation pour le clinicien de ne pas contrôler ou contraindre les patients, et l'importance de l'autodétermination autonome et de l'intégrité corporelle exigent que le consentement du patient soit obtenu par un clinicien proposant une intervention médicale (HIRSCH, MANN, 2014 ; SICARD, 2017 ; QUINCHE, 2005). Le respect de l'autonomie est un droit corrélatif de choisir et non pas un devoir impératif de choisir.

Le choix de déléguer est une forme d'autonomie. Il y a une obligation fondamentale qui est l'assurance pour le patient du droit à choisir, ainsi que du droit à accepter ou refuser d'être informé. Respecter un individu autonome, c'est, au minimum reconnaître le droit de cette personne à avoir des opinions, à faire des choix et à agir en fonction de ses valeurs et de ses croyances. Un tel respect implique une action respectueuse et pas uniquement une attitude

respectueuse. Violer l'autonomie d'une personne revient à la traiter uniquement comme un moyen, en accord avec les buts des autres, sans prendre en compte les propres buts de cette personne (SICARD, 2017 ; HIRSCH, MANN, 2014).

Le respect de l'autonomie des personnes passe par le respect :

- du droit et obligations de liberté ;
- de l'intimité ;
- de la confidentialité ;
- de vérité et de consentement éclairé.

Le respect de cette autonomie peut être outrepassé lorsque le choix de l'individu est une menace pour la population – santé publique - ou dans un contexte de rareté des ressources pour la prise en charge de l'individu. Le médecin dans son devoir de protection de la population peut dans ce cas décider à la place de l'individu mais en prenant en compte les trois autres principes d'éthiques médicales.

Tous les patients ne sont pas en mesure de prendre des décisions autonomes (les enfants, les personnes inconscientes) ; dans ces cas, le concept d'agir au mieux de leurs intérêts (conformément au principe de bienfaisance) devient d'une importance capitale.

Le clinicien avant de pratiquer un acte médical, devrait avoir le consentement du patient afin de respecter l'autonomie de la personne. Ce consentement peut se présenter sous plusieurs formes (BEAUCHAMP, CHILDRESS, 2009 ; SICARD, 2017) :

- Le paradigme de base de l'autonomie dans les soins médicaux, la recherche, la politique et autres contextes est le consentement exprès. Le consentement exprès n'est pas limité dans le temps, à moins que l'individu ne le retire.
- Le consentement tacite est exprimé de façon silencieuse ou passive par omission. Il est limité dans le temps et se termine à la fin de l'acte dans le cas d'un acte médical ou d'une procédure particulière.
- Le consentement implicite ou sous-entendu est déduit par les actions. Le consentement à un acte médical est souvent implicite car il découle d'un consentement spécifique à un autre acte, et le fait d'aller volontairement dans une structure de soins peut impliquer le consentement aux divers actes des médecins et des infirmiers.
- Le consentement présumé est considéré comme étant un autre genre ; si le consentement est présumé à partir de ce que nous savons sur les choix et les valeurs

particulières d'une personne, il se réduit soit au consentement sous-entendu, ou soit à un consentement formel. Si le consentement est présumé à partir des théories sur la volonté rationnelle ou les biens humains, la situation morale devient plus problématique. Le consentement doit se référer aux choix propres d'un individu, et non pas à des suppositions quant aux choix que tout individu ferait ou devrait faire.

Mais le consentement est le plus souvent une formalité qu'un acte de respect de l'autonomie de la personne pour les raisons suivantes (SICARD, 2017) :

- l'absence de temps de réflexion ; or c'est la réflexion qui donne de la valeur au consentement ;
- l'impératif médico-légale qui conduit à présenter le consentement comme une obligation et non comme une approche respectueuse ;
- la préférence de l'écrit ou de la matérialité qui est utilisée comme preuve au détriment de l'oral.

Non-malfaisance

Ce principe affirme l'obligation de ne pas infliger du mal à autrui. Dans l'éthique médicale, il a toujours été associé à la maxime d'Hippocrate « *primum non nocere* » : « *d'abord ne pas nuire* ». Le serment d'Hippocrate exprime clairement l'obligation de non-malfaisance et l'obligation de bienfaisance : « *je dirigerai le régime des malades à leur avantage, suivant mes forces et mes jugements, et je les protégerai de tout mal et de toute injustice* ».

Les obligations de ne pas faire du mal à autrui sont distinctes des obligations d'aider autrui (apport de bienfaits, protection des intérêts et la promotion du bien-être).

La non-malfaisance requiert de s'abstenir intentionnellement d'agir en causant du tort. Elle inclut également des obligations de ne pas imposer de risques de causer du mal (HOERNI, 2003 ; HERNI, AVRIL, 2009 ; BENAROYO, 2014).

Dans le cas des risques imposés, la loi et la morale reconnaissent une norme de soins dus.

« Les soins dus impliquent de veiller à la sécurité de façon suffisante et appropriée en fonction des circonstances, afin d'empêcher de causer du mal ; et cela comme il se doit pour une personne raisonnable et prudente. Cette norme

exige que le but poursuivi justifie les risques qu'on impose pour l'atteindre »

(BEAUCHAMP, CHILDRESS, 2009).

La négligence signifie l'absence de soins dus. Dans le domaine professionnel, elle implique un écart par rapport aux normes qui définissent les soins dus dans un contexte particulier. On peut avoir deux cas de négligence : la négligence intentionnelle ou imprudence et la négligence par inadvertance (BEAUCHAMP, CHILDRESS, 2009).

Les acteurs impliqués de loin ou de près dans le processus de soins (professionnels de la santé, informaticiens, assistants médicaux, secrétaires, gestionnaires) doivent avoir une norme de soins dus aux personnes qui utilisent le système de santé ; ceci est aussi valable pour les hébergeurs des données et les responsables de plateforme). Par exemple, l'utilisation des données personnelles des individus dans le cadre du *big data*, le partage des données avec des entreprises n'apporte aucun bénéfice direct à l'utilisateur d'une plateforme ou d'une application. Le partage ici est intentionnel ; et si toutefois, les informations des personnes sont divulguées, il s'agit d'une négligence intentionnelle de la part des responsables de la plateforme. Les responsables de plateforme doivent assurer la sécurité des données des personnes par tous les moyens possibles afin d'éviter de créer du tort à ces personnes ou de leur faire courir le risque de la divulgation des données. Dans les deux cas, nous sommes en face d'une forme de malfeasance et d'absence de norme de soins dus.

La faute professionnelle est une forme de négligence qui implique le non-respect des normes professionnelles de soins. Les patients ont à faire face à la limitation des soins, à cause des pressions destinées à réduire les coûts des soins médicaux qui montent de manière exponentielle. Les discussions portent plus sur les intérêts collectifs que sur le patient lui-même. L'obligation de soins aux patients implique un choix de traitements qui correspondent à ses préférences et à ses intérêts – dans les limites fixées par les politiques d'allocations équitables -, et non un choix de traitements faits pour ce qu'ils symbolisent dans la société. Les demandes irraisonnables de la part des patients ou des familles ne doivent pas interdire des politiques raisonnables dans les institutions de soins médicaux. Dans ces conditions, le respect de l'autonomie du patient ou de son substitut n'est pas un atout pour bénéficier de soins adéquats, surtout dans des contextes de ressources limités.

La non-malfeasance constitue un équilibre important par rapport au principe de bienfaisance, car aucun acte médical n'est sans risques ni effets secondaires, et toute prise de décision médicale implique un équilibre entre les avantages et les risques. La pratique de la médecine

ne doit pas être écrasée par le désir de faire le bien à n'importe quel prix et la considération incontrôlée de traitements extraordinaires ou héroïques pour les patients pourrait les exposer à des préjudices et des risques importants pour un bénéfice potentiel faible (DONALDSON, 2012 ; LE COZ, 2009 ; LE COZ, KACHANER, 2010).

Justice

« La justice distributive se réfère à une distribution équitable juste et appropriée déterminée par des normes justifiées qui structurent les conditions de la coopération sociale, englobant les politiques qui attribuent les bénéfices et les désavantages divers comme les biens, les ressources, les impôts, les privilèges et les opportunités » (BEAUCHAMP, CHILDRESS, 2009).

Les situations de pénurie peuvent pousser les sociétés à faire des choix tragiques. La médecine est pratiquée dans un monde où les besoins sont importants et augmentent de façon exponentielle dans le temps et les ressources sont limitées ; il est donc important que le médecin prenne en compte le fait que son patient vit dans une société où les ressources médicales doivent être utilisées pour les autres patients.

Les principes valides de justice distributive sont (BEAUCHAMP, CHILDRESS, 2009) :

- chacun doit recevoir une part égale ;
- chacun doit recevoir en fonction de ses besoins ;
- chacun reçoit en fonction de ses efforts ;
- chacun reçoit en fonction de sa contribution ;
- chacun reçoit en fonction de son mérite ;
- chacun reçoit en fonction d'échanges dans le cadre de la libre concurrence

Les théories concernant la justice distributive tentent de faire le lien entre le bien des individus et une répartition moralement justifiable des avantages et contraintes.

Les théories utilitaristes mettent l'accent sur un mélange de critères dans le but de maximiser l'utilité publique. Les décisions sociales doivent maximiser l'utilité sociale. La maximisation de l'utilité entraîne une précarisation de l'accès aux soins et les droits fondamentaux. Elles ne prennent pas en compte la manière dont les biens sont distribués. Par contre, elles jouent

un rôle important dans l'élaboration des politiques de santé (MASSÉ, SAINT-ARNAUD, 2003 ; HOERNI, 2003 ; LE COZ, KACHANER, 2010).

Les théories libertariennes mettent l'accent sur les droits à la liberté sociale et économique. Une société juste protège les droits de propriété et de liberté qui permettent aux individus d'améliorer leur situation et de protéger leur santé de par leurs propres initiatives. Les soins médicaux ne sont pas un droit, et le système de santé idéal est privatisé (MASSÉ, SAINT-ARNAUD, 2003). La solidarité des soins n'existe pas dans ce système et entraîne des inégalités d'accès aux soins.

Les théories communautaristes donnent plus d'importances aux principes et aux pratiques se rapportant à la justice que configurent les pratiques traditionnelles d'une communauté. Les principes de justice sont pluralistes, découlant d'autant de conceptions du bien qu'il y a de communautés morales différentes. La communauté a une responsabilité envers l'individu et l'individu envers la communauté (MASSÉ, SAINT-ARNAUD, 2003).

Les théories égalitaristes soulignent l'accès égal aux biens auxquels toute personne rationnelle aspire (critère matériel de besoin et de nécessité). Certaines égalités de base entre les individus. Une théorie de la justice qui correspond au jugement couramment accepté que nous avons sur l'équité, associé à nos principes généraux. Chaque membre de la société, quelle que soit sa richesse ou sa position, aurait un accès égal à un bon niveau de soins médicaux – le niveau exact d'accès dépendrait en fait des ressources sociales disponibles et des procédures publiques de décision (MASSÉ, SAINT-ARNAUD, 2003).

L'absence de justice sociale dans l'accès aux soins sont à la base de la création d'une inégalité plus prononcée. Cette inégalité peut se faire par l'intermédiaire du pouvoir d'achat des individus qui ne peuvent pas se payer une assurance décente mais aussi par les assurances, qui décident de ne pas prendre en charge les personnes qui souffrent de maladies chroniques ou des antécédents familiaux de maladie. Cette iniquité d'accès aux soins peut être minimisée par la mise en place d'un service de soins minimum décent. Les soins médicaux sont à la charge des individus et la protection collective est à la charge de l'Etat. La thèse d'un socle minimum de soins implique l'acceptation d'un système à double vitesse : une couverture sociale assurée pour les besoins en soins médicaux élémentaires et les soins

urgents ; ainsi qu'une couverture médicale volontaire et privée pour les autres besoins et pour les soins personnels.

Dans le cas d'une justice sociale d'accès aux soins, le risque est la discrimination de certaines personnes qui s'exposent à des risques et la perte de leurs droits fondamentaux telle que l'autonomie. Bien que leur comportement soit contraire aux règles édictées par les politiques de santé, il faut néanmoins prendre en compte les facteurs sociétaux et le choix de la société par rapport à ces personnes à risque. Nous savons tous qu'il est très difficile d'avoir des consensus pour certaines actions, et l'Etat dans sa responsabilité paternaliste peut prendre certaines décisions s'il prouve les relations de causalité et met en place des systèmes pour ne pas discriminer certaines personnes qui ne maîtrisent pas les facteurs.

Les quatre principes de l'éthique médicale ont été adoptés par la majorité des professionnels de la santé. Ils sont simples et faciles à retenir, tout en fournissant au clinicien des outils de raisonnement éthique qui découlent d'une analyse minutieuse et de la fusion de nombreuses théories et traditions de la philosophie morale.

L'approche par principe de l'éthique médicale comporte des limites ; cette méthode est loin d'être complète et ne répond pas au problème de préciser à quel moment un principe l'emporte sur un autre, de sorte que le clinicien peut avoir besoin de se tourner vers d'autres ressources éthiques pour obtenir des conseils lorsqu'il doit prendre une décision éthique et peser entre les principes. (WILLIAMS, 2015 ; SICARD, 2017 ; BYK, 2008 ; BAUMANN, 2004 ; LE GALL, 2016).

3.4.2 *Déontologisme*

L'éthique déontologique décrit un ensemble de théories éthiques qui jugent les actions comme bonnes ou mauvaises sur la base de règles et de devoirs (WILLIAMS, 2015). La déontologie est toujours rattachée explicitement aux devoirs d'une profession. Les règles sont précisées dans un code de déontologie qui s'applique aux membres de cette profession. C'est le cas par exemple du code de déontologie médicale en France.

Le mot déontologie est dérivé du mot grec « *deon* », qui signifie « *devoir* ». Selon le déontologisme, ce ne sont pas les résultats d'une action, mais plutôt quelque chose d'intrinsèque à l'action elle-même, qui la rend bonne ou mauvaise. « *La nature intrinsèque de l'action est jugée en fonction d'une règle ou d'un ensemble de règles, quels que soient la*

nature de l'action et les résultats de cette action » (DONALDSON, 2012). Pour bien agir, l'agent moral doit faire son devoir conformément aux règles.

Selon Kant, nous devrions poser des actes que nous savons être intrinsèquement bons et éviter les actes que nous savons être intrinsèquement mauvais, et laisser les conséquences se manifester comme elles le seront (KANT, FUSSLER, 2003). Nous ne pouvons jamais connaître toutes les conséquences de nos actions et que, parce que nous ne pouvons pas savoir si nos actions auront de bons ou de mauvais résultats. Pour Kant, nous savons si les actions sont bonnes ou mauvaises à partir de la raison et non de leurs conséquences. En fait, Kant a fondé son système éthique entièrement sur la raison (KANT, FUSSLER, 2003). Il cherchait à utiliser la raison pour élaborer un ensemble cohérent et non prépondérant de règles morales qui seraient universelles et contraignantes pour toutes les créatures rationnelles - une loi morale suprême.

La théorie de Kant comporte un seul principe qui est l'impératif catégorique de Kant. Sa formule générale est « *Agis uniquement d'après la maxime qui fait que tu peux vouloir en même temps qu'elle devienne une loi universelle* » (KANT, FUSSLER, 2003). Kant a trois formulations de son impératif catégorique, qu'il considérait comme trois manières de dire la même chose. La première de ses formulations communique le principe de l'universalité et est comme suit « *Agis comme si la maxime de ton action devait être érigée par ta volonté en loi universelle de la nature* ». La deuxième formulation de l'impératif catégorique exprime la valeur et la dignité qui sont intrinsèques agents rationnels selon Kant. C'est ainsi qu'il est écrit « *Agis de telle sorte que tu traites l'humanité aussi bien dans ta personne que dans la personne de tout autre toujours comme une fin et jamais simplement comme un moyen* ». La troisième formulation de l'impératif catégorique exprime le principe de l'autonomie de Kant, « *Tout être rationnel est capable de se considérer comme un créateur de loi universelle* ». Kant définit une autonomie de la volonté opposée à l'hétéronomie (choix non rationnel, impulsif, passionné). L'autonomie est l'acceptation volontaire des principes moraux. Le principe de l'autonomie est le seul principe de la morale et que seule l'autonomie fonde le respect, la valeur, la vraie motivation des personnes.

Le déontologisme donne aux médecins un aperçu de l'importance des règles en éthique. Les règles fournissent une excellente description des attentes que la plupart des patients ont à

l'égard de leurs médecins ; notamment la protection du secret médical, l'honnêteté dans la prise en charge. Plusieurs de ces règles gouvernent la relation humaine en société et en particulier la relation médecin - patient. La théorie éthique de Kant montre que pour qu'une action soit jugée juste, il faut que l'on ait des bonnes raisons de la mettre en œuvre. Cette théorie met en exergue la valeur humaine comme un agent rationnel doué d'une raison et que l'humain doit être traité comme une fin en soi et non comme un moyen. Ces valeurs sont très développées dans le cadre de la pratique médicale.

Mais le déontologisme comporte des limites dans ce sens que les formulations des impératifs catégoriques sont larges et non spécifiques en termes de principes. Ces formulations sont donc difficilement applicables aux différents modes de pratique médicale. En plus, il est basé sur la raison et le droit et ne laisse pas de place à la bienfaisance, l'empathie, la sympathie ou le désir de la personne qui sont fondamentaux dans l'exercice médicale.

3.4.3 Utilitarisme

Les pères de l'utilitarisme sont Jeremy Bentham (BENTHAM, 1789) et John Stuart Mill (MILL ET AL., 2009), dont les écrits ont donné naissance à la théorie de l'utilitarisme pendant le Siècle des lumières. Ils ont suggéré que la moralité n'était pas une question de fidélité à un code ou à des règles inflexibles.

L'utilitarisme fonde la prise de décision éthique sur l'analyse des conséquences ou des résultats des différents choix ou actes. L'action juste est celle qui produit les meilleurs résultats. Il peut y avoir désaccord sur ce que l'on estime être un bon résultat. L'utilitarisme retient « l'utilité » comme unité de mesure et définit ce principe comme « *le plus grand bien pour le plus grand nombre* » (BENTHAM, 1789). Les mesures utilisées dans la prise de décision relative aux soins de santé sont l'analyse du coût et de l'efficacité et les systèmes de mesure de la qualité de la vie, AVCQ (années de vie corrigées par la qualité) ou AVCI (années de vie corrigées de l'invalidité) (SICARD, 2017).

Le principe éthique de base est l'utilité. Nous devrions toujours produire un résultat maximal de valeurs positives par rapport aux valeurs négatives (ou le moins de valeurs négatives si les résultats obtenus ne peuvent qu'être mauvais) (MASSÉ, SAINT-ARNAUD, 2003). Elle ne définit pas le type de valeurs à produire. Nous devons évaluer les meilleurs biens en termes de valeur intrinsèque totale produite pour une action. D'autres soutiennent que le concept d'utilité ne se réfère pas aux biens intrinsèques, mais aux préférences d'un individu, et que

nous devons maximiser la satisfaction globale des préférences d'un plus grand nombre d'individus (BEAUCHAMP, CHILDRESS, 2009).

Bentham décrit une version hédoniste de l'utilitarisme, soutenant que les actions devraient être décidées par la détermination des effets nets des actions alternatives potentielles en termes de production de bonheur ou de réduction de la souffrance. L'action qui produit le plus de bonheur, ou le moins de souffrance, est la bonne action (BENTHAM, 1789).

Pour Mill, l'utilité n'est pas purement hédoniste, sa vision du bonheur ne provient pas du simple plaisir. Il soutenait que tous les plaisirs ne sont pas comparables et il voyait le bonheur en termes d'*eudaimonia* (l'épanouissement humain, le bonheur, la prospérité). Cependant, pour Mill, le plaisir et la douleur sont toujours fondamentaux à sa compréhension du bonheur et sa conception de l'utilité est toujours l'équilibre entre le plaisir et la douleur (MILL ET AL., 2009).

Le jugement probabiliste joue un rôle important dans l'utilitarisme quant à la décision d'une bonne action. L'utilitarisme propose de considérer la totalité des faits et l'ensemble des conséquences possibles en s'informant au mieux de leur probabilité et de leur ampleur.

Dans le cas de l'utilitarisme de l'acte, les règles morales sont utiles pour guider les actions humaines, mais elles ne sont pas nécessaires si elles n'encouragent aucune utilité dans un contexte.

L'utilitarisme a le potentiel de fournir au médecin un système simple et clair pour aborder les dilemmes éthiques. Cela permettrait au médecin d'éviter la confusion de principes contradictoires. Le principe d'utilité permet, en théorie, de fournir une procédure de prise de décision fiable permettant au médecin de choisir la bonne d'action dans chaque situation. De plus, l'utilité ne repose pas sur l'intuition morale du médecin pour identifier ou équilibrer les principes moraux, ce qui peut parfois produire des résultats variables et peu fiables. Le principe d'utilité est très important dans le cadre de la mise en place d'une politique publique et la promotion du bien-être (bienfaisance). Il vise à rendre les gens heureux et à soulager les souffrances ; ces objectifs s'inscrivent dans la motivation du médecin et résument l'objectif de l'éthique médicale.

Mais l'utilitarisme peine à maintenir la distinction essentielle entre les actions moralement obligatoires et les actions surrogatoires situées au-delà de l'obligation morale et effectuées pour atteindre des idéaux personnels. Il ne définit pas clairement ce qui est bien et le bonheur des gens. Il n'existe aucun paramètre clinique au médecin pour mesurer le bonheur de son patient d'autant plus qu'il ne peut pas prévoir toutes les complications pouvant survenir dans un contexte de prise en charge ; elle ne se limite pas seulement au fait de prodiguer des soins de qualité au patient. Il pose la question de la justice distributive dans des cas cliniques où il existe un conflit entre le maximum de biens pour le maximum de personnes¹.

3.4.4 *Ethique de la vertu*

L'éthique de la vertu est une description d'une façon de penser l'éthique qui, plutôt que de mettre l'accent sur l'éthique des règles ou des résultats produits par des actions spécifiques, met l'accent sur le caractère et la motivation qui conduisent à choisir ces actions. Les actions sont jugées à la lumière du caractère, ce qui signifie que la bonne action dans une situation est définie comme celle qu'une personne vertueuse prendrait dans une telle situation. Une personne vertueuse est décrite selon ses vertus (SIMMENAUER, 2014). Par exemple, la compassion est une vertu particulièrement importante pour le médecin ; comme l'honnêteté, la prudence et le dévouement. Les médecins qui possèdent ces vertus sont mieux à même de prendre les bonnes décisions et de bien les appliquer. Cependant, même les personnes vertueuses ne sont souvent pas sûres de la manière d'agir dans certaines situations et ne sont pas à l'abri d'une mauvaise décision.

Selon l'éthique de la vertu, il n'est pas seulement important de faire ce qui est juste, mais il faut avoir la bonne disposition, la motivation et la maîtrise de ses émotions pour être une bonne personne qui fait ce qu'il faut. Cela rend l'éthique de la vertu bien adaptée à la fois à la vie publique et à la vie privée, en comblant le fossé entre les deux et en exigeant l'intégrité dans les deux (SIMMENAUER, 2014).

L'éthique de la vertu, en tant que théorie éthique inspirante s'éloigne du contexte juridique et englobe à la fois l'importance de la spiritualité et la valeur de la communauté. Elle est une éthique d'aspiration et non une éthique de devoir. L'éthique de la vertu remontent à la Grèce antique et Aristote est considéré comme le père fondateur.

Le concept de « *telos* » (*but ou fin*), qui sous-tend l'éthique d'Aristote, décrit le but vers lequel tend l'humanité. Cet objectif donne le sens et le but à l'être humain et est nécessaire pour fournir l'impératif éthique d'être une personne vertueuse. Aristote définit « *les vertus comme*

un trait de caractère qui se manifeste par l'action habituelle, qu'il est bon pour une personne d'avoir » (ARISTOTE, 1998). Ces vertus sont les qualités requises pour avoir une bonne vie et sont le bon moyen entre les deux vices négatifs, représentés par la carence et l'excès. La personne vertueuse, tout en se concentrant toujours sur le *telos* de l'épanouissement humain, est constamment engagée dans des situations et des circonstances de la vie réelle. Le *phronesis* (sagesse pratique) est comment Aristote décrit la capacité de la personne vertueuse à répondre de façon réaliste à des circonstances spécifiques., tout en restant dirigé et guidé par le *telos* de l'épanouissement de l'homme (ARISTOTE, 1998).

L'éthique de la vertu a le grand avantage pour le clinicien de fournir un compte rendu de l'importance de la motivation morale dans la vie éthique. Elle met l'accent sur l'importance de l'éthique morale. L'éthique de la vertu peut traiter certains cas avec une partialité telle dans le cas des amitiés ou avec impartialité telle que dans le cas de la justice ou la bienfaisance ; ainsi le médecin vertueux pourra faire la part des choses au moment de l'action. Par contre, l'éthique de vertu ne donne pas de bonne orientation pour le médecin dans le cadre de la prise de décision clinique. Il est basé sur des théories qui ne sont pas facilement applicables dans la pratique de la médecine. En plus, la pluralité des vertus sans explication claire des choix et du poids de chaque vertu rend difficile la prise de décision.

Toutes les théories éthiques comportent des limites et ne se suffisent pas à elles seules dans la gestion des problèmes éthiques soulevés dans la pratique médicale. L'éthique pratique est une construction singulière adaptée à chaque cas unique. Le reste ne peut être qu'une charte, un modèle, une méthode de guidelines qui ne sont que des indicateurs utiles mais toujours réducteurs. Une combinaison de plusieurs approches est un moyen excellent de prendre des décisions éthiques rationnelles.

« Il s'agira alors de bien tenir compte des règles (déontologie) et des principes (éthique des principes) lorsque l'on identifiera les plus appropriés à une situation donnée et tentera de les appliquer le plus largement possible. Il conviendra aussi d'examiner les conséquences possibles (conséquentialisme) des autres décisions possibles et de déterminer quelles conséquences seraient préférables. Enfin, il faudra s'efforcer de s'assurer que le comportement du décideur, à la fois au regard de la prise de décision et de sa mise en application, est admirable (éthique vertueuse) » (WILLIAMS, 2015).

Si plusieurs approches éthiques ont été utilisées dans le cadre de la décision éthique en médecine, qu'en est-il de l'éthique de l'IA dans le cadre de son utilisation en médecine ? Faut-il écrire une nouvelle éthique pour l'IA ? Ces dernières années nous avons assisté à une rédaction de multitudes de textes et guides dans le cadre de l'éthique médicale ; environ 70 textes (ALGORITHM WATCH, 2019). Un fait important, tous ces textes et guides sont basés sur les quatre principes d'éthique médicale (bienfaisance, autonomie, justice et non malfaisance) en plus d'un principe qui est l'explicabilité (FLORIDI ET AL., 2018). Mais, nous pensons que le principe de l'explicabilité entre dans le processus d'obtention du consentement libre et éclairé du patient avant sa prise de décision. Ce principe peut donc faire partie intégralement du principe d'autonomie.

En plus, tous ces textes et guides ne sont que des recommandations de bonnes pratiques sans engagement obligatoire, ni valeur juridique ; il s'agit pour chaque entreprise ou groupe de personnes de proposer leur modèle de conception d'une IA éthique. Leur non-utilisation n'engage la responsabilité d'aucune structure, il s'agit pour les entreprises de « *montrer patte blanche* » quant au développement de leur système d'IA. Ainsi, l'un des risques est que les différents groupes essaient d'imposer leur manière de faire sans prendre en compte les besoins de chaque groupe et des différents acteurs (FLORIDI, 2019). Les textes et guidelines sur l'éthique de l'IA sont des formes de légaliser les pratiques faites dans les différentes entreprises sans un vrai débat éthique ; par contre l'éthique médicale est une obligation, un impératif avec des possibilités de sanctions par la société et un rejet par ses pairs en cas de manquement.

« *Plus les techniques auxquelles nous recourons sont innovantes et audacieuses, plus s'exerce une liberté qui apparaît sans limite, et plus la démarche éthique semble recevoir de justification et d'aliment* » (LOMBARD, 2009). Car la fonction de l'éthique est fondamentalement la mise en cause d'un pouvoir et son interpellation au nom de valeurs qu'on juge à un moment donné supérieures ; surtout que le pouvoir va s'appliquer et même s'imposer comme porteur d'une décision concernant une personne, le plus souvent vulnérable, avec des conséquences sur sa vie, son intimité voire sa liberté (souvent pesé par rapport à l'intérêt général de la population). L'éthique repose sur l'idée que le possible ne s'identifie pas au souhaitable et qu'il existe une grande différence entre être en mesure de faire (ceci ou cela) et de pouvoir en répondre.

Au fil des découvertes techniques et leurs utilisations dans la pratique médicale, l'éthique s'est toujours adaptée avec une base qui est toujours celle de l'éthique de la médecine. Les technologies de l'IA dans le cadre de leur utilisation dans la pratique médicale peuvent bénéficier également de cette réflexion éthique sans pour autant créer une multitude d'éthiques ; cela est d'autant plus vrai que l'objet de la médecine reste identique avec pour seul apport des nouveaux outils de prise de décision et de prise en charge. Le processus de prise de décision éthique dans la pratique médicale doit prendre en compte l'apport de ces outils. Faut-il pour autant changer la manière dont la décision de soin est élaborée ?

3.4.5 *Prise de décision en médecine*

La prise de décision médicale est un long et périlleux processus qui se construit tout au long du raisonnement médical. Elle doit prendre en compte plusieurs faits importants à savoir le patient et ses désirs, son environnement (famille, contexte culturel, coutumes), la société et les moyens disponibles ; il va sans dire que les conséquences de la prise de décision, même si l'impact direct concerne le patient, impacteront tous les acteurs impliqués. Cette prise de décision met au centre le patient avec pour objectif de faire le maximum de bien et le moins possible de mal pour tous les acteurs en fonction des moyens disponibles et mobilisables. De ce fait, nous pensons que la prise de décision médicale constitue en elle-même une réflexion éthique permanente.

Dans le cadre de la pratique de la médecine, les actions se mènent dans le long terme et dans le temps avec des actions irréversibles. Tous ceux-ci placent la responsabilité au centre de l'éthique. L'éthique est ainsi une réflexion et une analyse attentive et systématique des décisions et comportements moraux, passés, présents ou futurs. Elle est ainsi le fruit d'un débat et d'une argumentation contradictoire, qui évolue dans le temps et en fonction de l'évolution de la société. Elle devra donc prendre en compte le contexte singulier, les groupes de personnes et l'environnement dans lequel le débat a eu lieu.

Encadré 3 : éthique selon Paul Ricoeur (RICOEUR, 2015)

L'éthique vient, dit Paul Ricoeur « *se briser en deux sur la morale* », dont elle forme à la fois l'aval et l'amont. Il y a d'un côté, en amont des normes, *une éthique antérieure*, qui s'intéresse au fondement de l'action, qui s'interroge sur les notions et les principes, puis il y a, d'un autre côté, en aval des normes, *une éthique postérieure* qui vise à insérer ces « *normes dans des situations concrètes* », qui joue le rôle d'une « *sagesse pratique* » permettant de discerner « *la droite règle (orthos logos) dans les circonstances difficiles de l'action* ».

Ainsi, s'interroger sur la valeur de la vie ou sur le sens de l'action de soigner, c'est de l'éthique antérieure, qui précède l'action et qui met jeu les différentes normes. Par contre dans le cadre du choix d'une meilleure attitude diagnostique et thérapeutique dans une situation particulière, avec tels ou tels acteurs, ici et maintenant, c'est de l'éthique postérieure, celle où se confrontent la réalité et les décisions à prendre. L'éthique se trouve donc toujours, d'une manière ou d'une autre, à la jonction des principes et du réel (RICOEUR, 2015).

Les nouvelles techniques et nouveaux savoirs, la pression de la recherche scientifique, les sollicitations du marché (industries pharmaceutiques et de technologies, assurances) et les demandes de la société à l'égard de la médecine les mettent face à des situations confuses. Nous considérons la médecine de nos jours comme la solution miracle à tous les maux. Bien que la pratique de la médecine ait connu des évolutions importantes avec le développement des technologies et la découverte de nouveaux médicaments dans la prise en charge, le statut de malade n'a pas changé ; le malade est toujours resté singulier face à la maladie. Cette singularité est très importante dans les prises de décision dans le cadre des soins et de la recherche dans le domaine de la santé.

Une prise de décision éthique dans le cadre de la pratique médicale doit prendre en compte :

- les connaissances scientifiques des faits ;
- le respect de la dignité de la personne ;
- le respect à la fois de l'identité et de la différence entre les personnes ;
- l'obligation de compétence et d'actualisation des connaissances scientifiques et techniques du médecin ;
- l'altérité ;

- la disponibilité des moyens permettant de prendre en charge la personne en toute sécurité.

Le mode de décision ainsi que le processus de prise de décision sont importants pour une prise de décision éthique.

Décision médicale partagée

Plusieurs modèles de décision médicale existent (LLORCA, 2006 ; EMANUEL, EMANUEL, 1992), mais la décision médicale partagée est celle utilisée en France. Ce choix de modèle de décision s'est vu renforcé avec la loi du 4 mars 2002¹⁷⁰ relative au droit des malades et la qualité du système de santé. En effet l'article 11 de ladite loi dit : « *toute personne a le droit d'être informée sur son état de santé. [...] Toute personne prend, avec le professionnel de santé et compte tenu des informations et des préconisations qu'il lui fournit, les décisions concernant sa santé* ».

La décision médicale partagée décrit deux étapes clés de la relation entre le patient et son médecin que sont l'échange d'informations et la délibération en vue d'une prise de décision acceptée d'un commun accord concernant la santé individuelle d'un patient (LLORCA, 2006 ; EMANUEL, EMANUEL, 1992 ; CHARLES ET AL., 1997 ; POLITI ET AL., 2013) :

- le médecin et le patient partagent de manière bilatérale une information médicale, notamment les éléments de preuve scientifique ;
- le patient reçoit le soutien nécessaire pour envisager les différentes options possibles et exprimer ses préférences ;
- un choix éclairé entre les différentes options est effectué et accepté mutuellement par le patient et les professionnels de santé.

Le fondement moral de ce modèle est le respect de la personne et la solidarité dans la prise de décision. Ainsi, la décision prise au cours de ce consensus doit satisfaire le patient, le médecin et la société. Ce modèle permet de donner une vue simplifiée du processus de décision ; mais il ne doit pas occulter toute la complexité de la relation médecin – patient évoluant dans le temps (HERNI, 2002). Au cours de ce processus de décision, l'avis de tous les acteurs (patient, médecin, société, financiers) doit être pris en compte dans la prise de décision.

¹⁷⁰ Loi n° 2002-303 du 4 mars 2002 relative aux droits des malades et à la qualité du système de santé.

« Boîte à outils » de Nicole et Louis Lery

Dans le cadre du processus de décision dans le soin au quotidien, Nicole et Louis Lery¹⁷¹ après des réunions avec les soignants a mis en place un objet d'analyse qu'il nomme « *boîte à outils* » (LERY, COLLOC, 2008 ; COLLOC, LERY, 1997).

La « *boîte à outils* » est un objet qui permet lors de son cheminement de respecter les étapes préalables à un processus de décision partagée (CHARLES ET AL., 1997) :

- *« établir, par le médecin, un environnement propice à l'échange de sorte que le patient perçoive que, s'il le souhaite, son opinion sur les différentes options thérapeutiques est nécessaire et sera valorisée ;*
- *rechercher les préférences du patient concernant les différentes options du traitement, afin de s'assurer que celles-ci sont compatibles avec son mode de vie et ses convictions ;*
- *pour le médecin, de transmettre des informations techniques sur les différentes options possibles, dont celle de ne pas être traité, sur leurs risques et bénéfices probables, d'une manière claire, objective et compréhensible ;*
- *pour le patient, de faire part des informations qu'il a pu recueillir avant la consultation et de ses valeurs, ce qui est important pour lui au regard du choix du traitement ;*
- *un accord commun entre les parties (au moins le médecin et le patient) est trouvé et une décision prise, que les deux parties acceptent. Cet accord mutuel est une caractéristique essentielle du processus de décision partagée ».*

En effet, Dans le cadre de la « *boîte à outils* », il s'agit d'un « *cheminement jalonné de repères techniques, juridiques, déontologiques, de recommandations et de normes morales ou religieuses, des données psychosociologiques et culturelles aboutissant à une délibération pluridisciplinaire avec et pour le*

¹⁷¹ Nicole et Louis Léry ont fondé une association nommée SEL (Santé, Ethique et Liberté) qui regroupait de nombreux praticiens et qui ont créé une consultation d'éthique médicale à l'Hôpital Edouard Herriot puis à l'Hôpital du Vinatier à Lyon. Les activités de recherche ont conduit à l'élaboration d'une méthode de décision éthique qui a donné lieu à un système multi-agent d'aide à la décision en éthique médicale en 1997, notamment concernant les vaccinations (COLLOC, LERY, 1997).

patient dans la situation concrète dans laquelle, il se trouve à ce moment de son histoire » (LERY, COLLOC, 2008) comme le montre la figure 31.

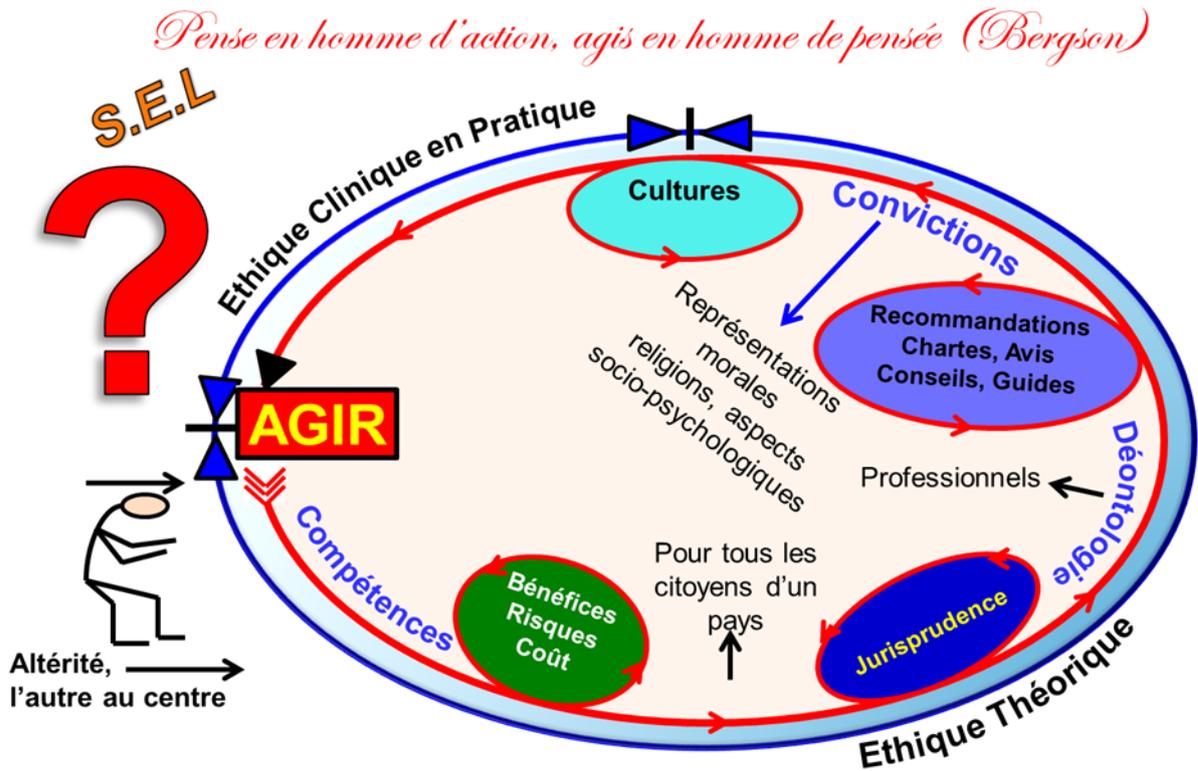


Figure 31 : cheminement des repères ou « boîte à outils » (LERY, COLLOC, 2008)

Dans ce cheminement, il s'agit d'agencer, au cas par cas, une série de procédures évaluables dans le but d'obtenir le résultat le plus acceptable pour et avec le malade. Il permet de respecter la liberté des acteurs que ce soit sur le plan des compétences ou de l'autonomie, de mettre en place des conditions équitables (égalité en droit) et de marquer notre solidarité et sollicitude envers les autres.

Les commentaires dans la presse et par des personnalités sur la place du *big data* et de l'intelligence artificielle dans notre société et plus particulièrement dans le cadre de la médecine ont suscité diverses réactions en fonction des rôles joués par les différents acteurs. Les questions qui se posent aujourd'hui sont : pouvons-nous développer des outils de l'IA respectueux de la pratique médicale ? Quels sont les enjeux de l'utilisation des traces numériques dans le cadre de la prise de la décision en santé ? Et quels sont leurs impacts sur la pratique médicale ?

La mise en place d'un système d'aide à la décision médicale doit intégrer en son sein la dynamique de la décision médicale afin d'apporter une valeur ajoutée à la pratique médicale.

4 Système d'aide à la décision médicale à partir d'un espace vectoriel flou : cas de l'embolie pulmonaire

L'embolie pulmonaire (EP) est la manifestation la plus grave de la pathologie thromboembolique veineuse. Elle est définie comme l'obstruction des artères pulmonaires ou de leurs branches par des emboles, le plus souvent à partir de thromboses cruoriques des veines profondes des membres inférieurs (TAPSON, 2008). Sur la base d'une étude épidémiologique réalisée dans six pays européens avec une population de 454,4 millions, 370 000 décès étaient dus à une maladie thromboembolique veineuse (MTEV). Un décès brutal ou quelques heures après l'épisode aigu était noté dans 34 % avant la mise en place d'un traitement. Dans 59 % des cas, le diagnostic était fait après le décès et ce n'est que dans 7 % des cas que le diagnostic a lieu avant le décès (COHEN ET AL., 2007). Cependant, les études autoptiques montrent que l'EP reste aujourd'hui une des premières causes de décès dans la population générale, plus de 70 % des EP létales survenant sans que le diagnostic ne soit évoqué ou porté du vivant du patient (PULIDO ET AL., 2006). Faire un diagnostic précoce représente ainsi un véritable enjeu médical et économique, mais les circonstances et les facteurs de risque de la MTEV sont multiples (GEERTS ET AL., 2008).

Plusieurs algorithmes d'aide au diagnostic ont été validés et assistent le clinicien dans sa démarche décisionnelle. Ils sont cependant encore peu ou mal utilisés en pratique courante, probablement parce qu'ils font référence à des données complexes ou parce que, dans certaines situations, les tests auxquels ils font appel sont contre-indiqués ou non disponibles (DAMBRINE ET AL., 2008). L'émergence d'outils informatisés d'aide à la décision pourrait permettre de diminuer ces entraves à l'application d'une médecine fondée sur des preuves.

Les SADM sont « *des applications informatiques dont le but est de fournir aux cliniciens en temps et lieu utiles les informations décrivant la situation clinique d'un patient ainsi que les connaissances appropriées à cette situation, correctement filtrées et présentées afin d'améliorer la qualité des soins et la santé des patients* » (DEGOULET, FIESCHI, 1991).

Mais le diagnostic de l'embolie pulmonaire repose sur des scores qui ont l'inconvénient de ne donner qu'une indication statique du risque d'EP à l'admission du patient. Nous proposons ainsi un nouveau score diagnostique basé sur un espace vectoriel flou temporel qui combine les critères traditionnels utilisés dans les scores de Wells et de Genève mais qui prend en compte l'évolution dynamique du diagnostic de la maladie du patient composée des tableaux et des signes cliniques. Ce modèle a déjà été éprouvé dans différents domaines cliniques (COLLOC ET AL., 2017 ; YAMEOGO, COLLOC, 2016 ; COLLOC, SUMMONS, 2006 ;

COLLOC, SYBORD, 2003). Ce nouveau score nous permettra de prendre en compte des incertitudes et des imprécisions liées au recueil de données à un instant initial et de déterminer l'évolution clinique de manière plus proche de la réalité du terrain. Nous utilisons une nouvelle approche basée la logique floue, le modèle orienté objet et le temps.

4.1 Diagnostic de l'embolie pulmonaire

Dans le cadre du diagnostic de l'EP, nous nous fondons sur les recommandations 2019 de la Société Européenne de Cardiologie (ESC) pour la prise en charge de l'EP (KONSTANTINIDES ET AL., 2019). Elles ont été rédigées par plusieurs experts du domaine selon les connaissances actuelles de la médecine ; elles serviront ainsi de données expertes dans le cadre de la construction de notre modèle de score.

4.1.1 Facteurs de risque thromboemboliques

La MTEV est la conséquence d'une interaction de facteurs de risque liés au patient et de facteurs de risques liés à l'environnement. Ces facteurs de risque sont classés en fonction de leur lien avec la survenue de la MTEV ; mais ils peuvent être absents dans certains cas. Le tableau 2 montre les différents facteurs de risques thromboemboliques.

Tableau 2 : facteurs prédisposant à une MTEV (KONSTANTINIDES ET AL., 2019)

| | |
|---|---|
| <p>Facteurs de risque fort (OR* > 10)</p> <ul style="list-style-type: none"> ● Fracture au membre inférieur ● Hospitalisation pour insuffisance cardiaque ou fibrillation/flutter atrial (dans les 3 mois précédents) ● Prothèse de hanche ou de genou ● Traumatisme majeur ● Infarctus du myocarde (dans les 3 mois précédents) ● Antécédent d'accident thromboembolique veineux ● Lésion de la moelle épinière <p>Facteurs de risque modéré (OR entre 2 et 9)</p> <ul style="list-style-type: none"> ● Chirurgie du genou par arthroscopie ● Maladie auto-immune ● Transfusion sanguine ● Cathéter veineux central ● Chimiothérapie ● Insuffisance cardiaque ou respiratoire ● Stimulant de l'érythropoïèse ● Traitement hormonal substitutif (selon la formulation) ● Fécondation <i>in vitro</i> | <ul style="list-style-type: none"> ● Infection (spécifiquement : pneumonie, infection urinaire et VIH) ● Maladie intestinale inflammatoire ● Cancer (risque plus élevé en cas de cancer métastatique) ● Contraception orale ● AVC avec paralysie ● Post-partum ● Thrombose veineuse superficielle ● Thrombophilie <p>Facteurs de risque faible (OR < 2)</p> <ul style="list-style-type: none"> ● Alitement > 3 jours ● Diabète ● Hypertension artérielle ● Immobilité en position assise (par exemple voyage : prolongé en voiture ou avion) ● Âge avancé ● Chirurgie laparoscopique (par exemple : cholécystectomie) ● Obésité ● Grossesse ● Varices <p>* OR : odds ratio.</p> |
|---|---|

4.1.2 Stratégie diagnostique

Être attentif aux symptômes d'une EP est le premier élément d'un diagnostic rapide. Mais il n'existe pas de symptômes spécifiques au diagnostic de l'EP. Il peut s'agir de signes respiratoires (dyspnée brutale), de signes cardiaques (hypotension artérielle voire état de choc, syncope) ou de douleurs basi-thoraciques rendant difficile le diagnostic. Les bilans paracliniques (électrocardiogramme, D dimères, radiographie pulmonaire, échographie cardiaque Doppler) ne sont pas spécifiques pour le diagnostic (GRÉGOIRE LE GAL ET AL., 2005 ; COROLLER-BEC ET AL., 2008 ; G. LE GAL ET AL., 2005 ; FEDULLO, TAPSON, 2003 ; ROY ET AL., 2005 ; STEIN ET AL., 2004).

L'angioscanner thoracique permet le diagnostic, mais sa disponibilité reste limitée dans certains contextes. Cette pathologie a une évolution imprévisible et nécessite un diagnostic rapide, d'où l'importance de la stratégie diagnostique de la prise en charge de l'embolie pulmonaire aux urgences. Elle dépendra en premier de la présence ou non de signe de choc ou d'une hypotension artérielle puis du niveau de probabilité clinique évalué par le score de Wells ou le score de Genève (KONSTANTINIDES ET AL., 2019).

Tableau 3 : score de Wells (DELAHAYE, 2015)

| Algorithme de Wells | Version originale | Version simplifiée |
|--|-------------------|--------------------|
| Antécédent de MVTE | 1,5 | 1 |
| Fréquence cardiaque \geq 100 bpm | 1,5 | 1 |
| Intervention chirurgicale ou immobilisation dans le mois précédent | 1,5 | 1 |
| Hémoptysie | 1 | 1 |
| Cancer en évolution | 1 | 1 |
| Signes cliniques de TVP | 3 | 1 |
| Diagnostic différentiel moins probable que celui d'EP | 3 | 1 |
| Probabilité clinique de "Wells" = | | |
| <i>Score à trois niveaux</i> | | |
| Basse | 0-1 | Non disponible |
| Intermédiaire | 2-6 | Non disponible |
| Haute | \geq 7 | Non disponible |
| <i>Score à deux niveaux</i> | | |
| EP improbable | 0-4 | 0-1 |
| EP probable | \geq 5 | \geq 2 |

Tableau 4: score de Genève (DELAHAYE, 2015)

| Score de Genève révisé | Version originale | Version simplifiée |
|---|-------------------|--------------------|
| Antécédent de MVTE | 3 | 1 |
| Fréquence cardiaque 75-94 bpm | 3 | 1 |
| Fréquence cardiaque \geq 95 bpm | 5 | 2 |
| Intervention chirurgicale ou fracture dans le mois précédent | 2 | 1 |
| Hémoptysie | 2 | 1 |
| Cancer en évolution | 2 | 1 |
| Douleur unilatérale à un membre inférieur | 3 | 1 |
| Douleur à la palpation profonde d'un membre inférieur et œdème unilatéral | 4 | 1 |
| Âge > 65 ans | 1 | 1 |
| Probabilité clinique de "Genève" = | | |
| <i>Score à trois niveaux</i> | | |
| Basse | 0-3 | 0-1 |
| Intermédiaire | 4-10 | 2-4 |
| Haute | \geq 11 | \geq 5 |
| <i>Score à deux niveaux</i> | | |
| EP improbable | 0-5 | 0-2 |
| EP probable | \geq 6 | \geq 3 |

Dans les deux cas, un score bas traduit une faible probabilité d'EP (10 %). Elle augmente à 30 % si le score est intermédiaire et à 65 % s'il est haut.

- **En cas de suspicion d'une EP avec signe de choc ou hypotension artérielle**

Un angiographe ou une échocardiographie transthoracique au lit du patient, selon les disponibilités et les circonstances, est recommandé pour faire le diagnostic. La présence d'une dysfonction ventriculaire droite chez un patient instable nécessite de débiter un traitement même s'il n'existe pas de possibilité de confirmer le diagnostic. Un algorithme décisionnel (figure 32) permet de guider le médecin dans ses décisions.

- **En cas de suspicion d'EP sans signe de choc ou d'hypotension artérielle**

L'évaluation clinique est recommandée afin de fonder la stratégie diagnostique sur la probabilité clinique, évaluée soit par le jugement clinique, soit par un algorithme de prédiction validé (figure 33).

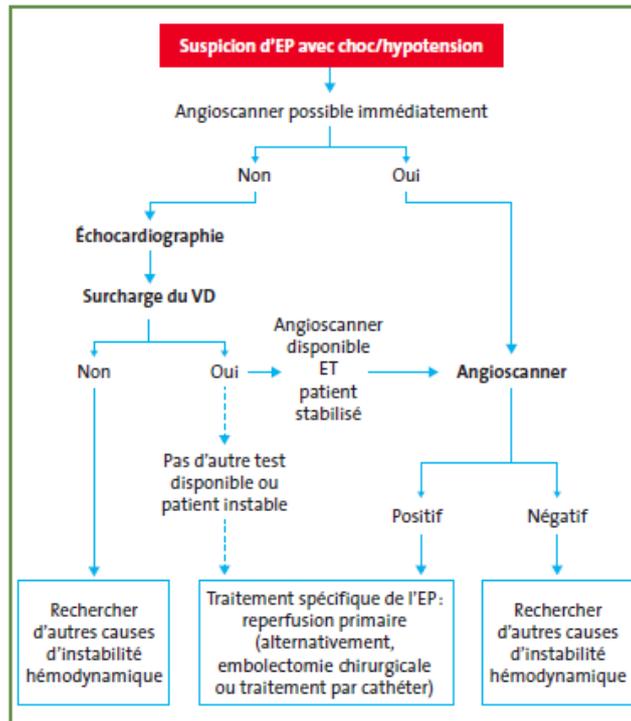


Figure 32 : algorithme diagnostique en cas de suspicion d'une EP avec signe de choc ou d'hypotension artérielle (DELAHAYE, 2015)

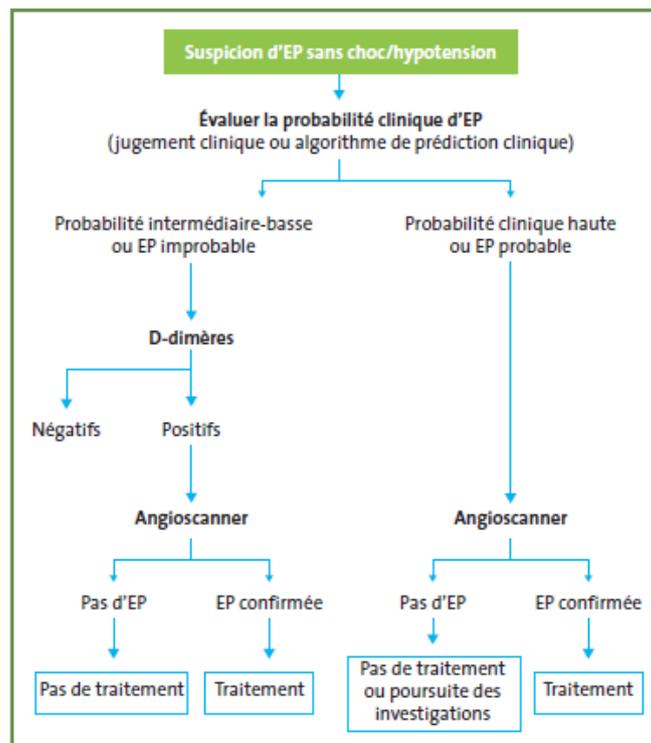


Figure 33 : algorithme diagnostique en cas de suspicion d'une EP sans signe de choc ou d'hypotension artérielle (DELAHAYE, 2015)

4.2 Espace vectoriel flou : une extension de la logique floue

La théorie des sous-ensembles flous (ZADEH, 1965) a permis de développer des applications utiles dans de nombreux domaines en tenant compte de l'incertitude et l'approximation des données nécessaires à la modélisation de connaissances. L'implémentation utilise le plus souvent des règles de production fondées sur l'induction floue dont l'inférence nécessite des méthodes de « *défluzzification* » (MAMDANI, ASSILIAN, 1975 ; SUGENO, 1993). Toutefois, une telle approche devient complexe et impraticable dès que les exceptions à gérer sont nombreuses (MCCARTHY, 1987) et/ou lorsque la décision nécessite d'autres modes de raisonnement que la déduction¹⁷².

Ce travail s'appuie sur des travaux antérieurs réalisés dans le cadre de l'aide à la décision médicale. La modélisation de la décision clinique a concerné la modélisation objet puis agent des maladies, et des différentes étapes de la démarche clinique : le diagnostic Δ , le pronostic Π , la thérapie Θ et le suivi thérapeutique $S\Theta$. Ces recherches multidisciplinaires utilisent des concepts provenant des sciences cognitives, notamment la psychologie cognitive, l'éthique médicale, la sémiologie, la linguistique, les ontologies, l'approche systémique (COLLOC, SYBORD, 2003). Dans ce modèle, Chaque cas clinique est unique et est donc porteur d'une expérience singulière réutilisable. Nous nous appuyons également sur les travaux de Colloc et Summons (COLLOC, SUMMONS, 2006) qui modélisent l'évolution temporelle des signes cliniques à partir de modèle orienté objet.

Le modèle que nous proposons est un espace vectoriel flou qui repose sur les modèles orientés objet et la logique floue. Le modèle entité – association de la « méthode Merise » (TARDIEU ET AL., 1994) et le modèle objet *UML (Unified Modeling Language)* (BOOCH ET AL., 1999) sont les plus utilisés en conception de systèmes d'information. Les opérateurs sémantiques des modèles objet sont utilisés pour modéliser les associations dans les ontologies et les bases de connaissances d'une manière plus simple et plus efficace que l'utilisation des règles de production et l'induction floue. Les opérateurs sémantiques des modèles orientés objets sont précisément dans le méta modèle *de l'ODMG (Object Data Management Group)*. Ils ont été utilisés pour exprimer des ontologies médicales (SHEN ET AL., 2012 ; YAMEOGO, COLLOC, 2016) et donnés lieu à une applications dans les accidents

¹⁷² Par exemple l'analogie qui s'appuie sur la similarité sémantique d'un grand nombre d'objets exprimant des connaissances dans des ontologies de grande taille (plusieurs millions d'objets et relations) (SHEN ET AL., 2018).

vasculaires cérébraux et les aphasies (COLLOC, JACQUET-ANDRIEU, 2013 ; SHEN ET AL., 2015). L'enchaînement et la mise en relation de ces travaux (COLLOC, SYBORD, 2003 ; COLLOC, SUMMONS, 2006 ; SHEN ET AL., 2012 ; COLLOC, JACQUET-ANDRIEU, 2013 ; SHEN ET AL., 2015) ont permis de développer l'espace vectoriel flou temporel (COLLOC ET AL., 2017). L'opérateur de composition « *est partie de* » au sein d'un espace vectoriel flou exprime la dynamique du système¹⁷³. Un objet du système comporte une fonction résultante qui est influencée par les fonctions des objets composants et celles de leurs attributs. Ces fonctions déterminent l'évolution temporelle des propriétés et des structures des objets (dans notre cas l'évolution des signes, des syndromes et des tableaux cliniques qui constituent la maladie), ce qui constitue également une contribution importante de notre approche. Ce modèle est utilisé pour évaluer le diagnostic clinique de l'EP dans notre cas. Ce qui est très important en pratique médicale, car il s'agit d'une maladie très évolutive dont les médecins n'ont souvent pas le temps de poser le diagnostic à l'angioscanner avant le décès du patient dans bon nombre de cas. Une amélioration du diagnostic clinique constitue de ce fait un énorme avantage pour la mise en route du traitement en attendant une confirmation diagnostic.

4.2.1 Proposition d'un espace vectoriel flou

Le modèle des espaces vectoriels flous est une extension de la logique floue de Zadeh appliquée aux approches orientées objets et présenté en détail dans un travail précédent (COLLOC ET AL., 2017).

Une fonction sigmoïde floue classique est définie sur :

$$[0,1], f(x) = \frac{1}{1 + e^{-k(x-s)}}$$

La figure 34 montre la fonction $f(x)$ où x représente la valeur d'une caractéristique ou d'un attribut (par exemple cancer) d'un objet (ex : antécédent) et s le seuil exprimant la valeur médiane pour le paramètre x et k une constante exprimant la précision avec laquelle x est connu.

Pour adapter cette fonction à l'intervalle $[-1,1]$, nous changeons de repère et d'échelle. Nous définissons la fonction :

$$F(x) = 2f(x) - 1.$$

Par conséquent :

¹⁷³ L'évolution des valeurs des attributs, de leurs contraintes et les changements de structure : la composition, les associations, les interactions des objets du système.

$$F : x \rightarrow [-1,1], F(x) = \frac{2}{1+e^{-k(x-s)}} - 1$$

Le graphe de la fonction est représenté par la figure 35.

Comme $F(x)$ est une fonction impaire $F(-x) = -F(x)$ définie et continue sur $[-1,1]$ et donc sur les intervalles $[-1,0]$ d'une part et $[0,1]$ d'autre part qui ont 0 pour intersection. $F(x)$ permet de construire deux fonctions opposées sur les intervalles respectifs :

$$f : x \rightarrow [0,1], f(x) = \frac{2}{1+e^{-k(x-s)}} - 1 \text{ et}$$

$$f' : x \rightarrow [-1,0], f'(x) = \frac{2}{1+e^{-k(x-s)}} - 1.$$

Comme la fonction est impaire, on constate bien que $\forall x \in [-1,1], f(x) = -f'(-x)$.

Le graphe des vecteurs correspondants à f et f' est représenté par la figure 36.

On définit de la même manière deux autres fonctions G et H :

$$G : y \rightarrow [-1,1], y \rightarrow G(y) \text{ et } H : z \rightarrow [-1,1], z \rightarrow H(z)$$

avec leurs parties respectives positives $g(y)$ et $h(z)$ et négatives $g'(y)$ et $h'(z)$. Nous considérons que comme x , les variables y et z correspondent à des caractéristiques, des attributs d'un objet à modéliser.

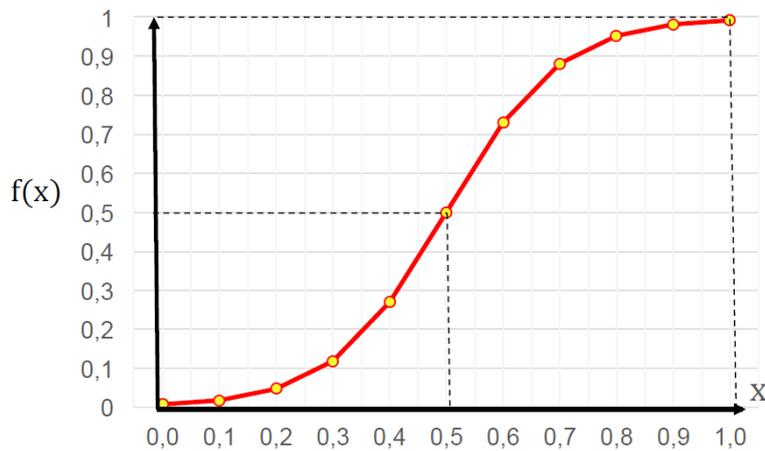


Figure 34 : fonction sigmoïde définie de $[0,1] \rightarrow [0,1]$



Figure 35 : fonction sigmoïde définie de $[-1,1] \rightarrow [-1,1]$

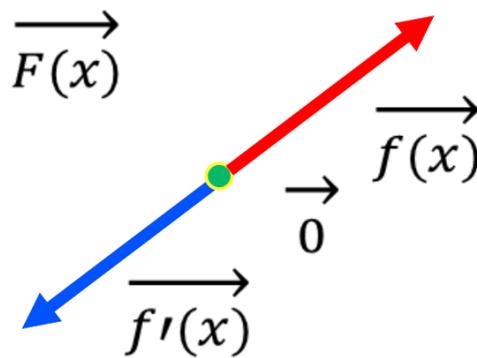


Figure 36 : vecteur de la fonction $\overline{F(x)}$: $[-1, 1] \rightarrow [-1, 1]$

4.2.2 Calcul de la norme de la résultante des trois vecteurs sur E

Dans l'espace vectoriel E sur R^3 , $E(0, \vec{i}, \vec{j}, \vec{k})$ un repère orthonormé, on définit le vecteur résultant $\vec{u} = \vec{f} + \vec{g} + \vec{h}$.

La norme

$$\|\vec{u}\| = \sqrt{f(x)^2 + g(y)^2 + h(z)^2}.$$

Comme $f(x), g(y), h(z) \in [0,1]$, $\max(\|\vec{u}\|) = \sqrt{3}$ et $\|\vec{u}\| \in [0, \sqrt{3}]$.

On définit la fonction résultante r des fonctions d'appartenance positives $f(x)$, $g(y)$ et $h(z)$.

$$\forall (x, y, z) \in [0,1]^3, r(x, y, z) = \frac{1}{\sqrt{3}} \|\vec{u}\| \in [0,1].$$

De même nous avons quand $x < 0$ la fonction d'appartenance négative $f'(x)$ symétrique à $f(x)$.
 $f(x) = -f'(-x)$. Le signe de x exprime si la fonction f est agoniste ou antagoniste f' .

$F(x)$ permet de définir les deux vecteurs opposés \vec{f} et $\overrightarrow{-f}$ à l'aide des fonctions $f(x)$ et $f'(x)$.

$$\forall x \in [-1,1], F(x) = \begin{cases} \vec{f} = f(x), & x \geq 0 \\ \overrightarrow{-f} = f'(x) = -f(-x), & x < 0 \end{cases}$$

Ceci permet de vérifier la propriété V^3 : vecteur opposé

$$\forall \vec{f} \in E, \exists \overrightarrow{-f} \text{ tel que } \vec{f} + \overrightarrow{-f} = \vec{0}.$$

De la même manière sont définis le vecteur \vec{g} avec $g(y)$ et $\overrightarrow{-g}$ avec $g'(y)$ puis le vecteur \vec{h} avec $h(z)$ et $\overrightarrow{-h}$ avec $h'(z)$ selon z avec les mêmes propriétés.

La résultante des valeurs négatives avec

$$f', g' \text{ et } h' : r'(x, y, z) = -\frac{1}{\sqrt{3}} \|\vec{u}\| \in [-1,0]$$

Dans l'espace vectoriel E , on peut affecter chaque vecteur d'un scalaire nécessairement positif et non nul, qui exprime l'importance de la composante dans la combinaison linéaire des vecteurs, $\alpha, \beta, \gamma \in \mathbb{R}^{+*}$, $\vec{u} = \alpha\vec{f} + \beta\vec{g} + \gamma\vec{h}$ et la norme devient

$$\|\vec{u}\| = \sqrt{\frac{\alpha f(x)^2 + \beta g(y)^2 + \gamma h(z)^2}{\alpha + \beta + \gamma}}$$

La fonction résultante r devient

$$\forall (x, y, z) \in [0,1]^3, r(x, y, z) = \frac{1}{\sqrt{3}} \|\vec{u}\| = \frac{1}{\sqrt{3}} \sqrt{\frac{\alpha f(x)^2 + \beta g(y)^2 + \gamma h(z)^2}{\alpha + \beta + \gamma}}, r(x, y, z)$$

$\in [0,1]$.

4.2.3 Généralisations à n paramètres

On généralise le modèle pour prendre en compte n paramètres ou variables $x_i, n \ll \infty$, aux variables $\forall x_i, i \in [1, n], x_i \in [-1,1]$ correspondent n fonctions d'appartenance $F_i(x_i)$ munies d'un scalaire $\alpha_i \in \mathbb{R}^{+*}$. On définit sur \mathbb{R}^n un espace vectoriel E où deux vecteurs $\vec{f} = (x_1, \dots, x_i, \dots, x_n)$ et $\vec{g} = (y_1, \dots, y_i, \dots, y_n)$ ont comme produit scalaire

$$\langle \vec{f}, \vec{g} \rangle = x_1 y_1 + \dots + x_i y_{i1} + \dots + x_n y_n.$$

E est muni d'un repère orthonormé $O(\vec{l}_1, \dots, \vec{l}_i, \dots, \vec{l}_n), \vec{f}_i$ correspond à $\vec{l}_i, \forall i, j \langle \vec{f}_i, \vec{f}_j \rangle = 0$. Le produit scalaire de deux vecteurs orthogonaux est nul. On définit le vecteur \vec{u} dans l'espace vectoriel E sur \mathbb{R}^n comme étant la somme des vecteurs \vec{f}_i munis du scalaire $\alpha_i \in \mathbb{R}^{+*}$. Ainsi vient $\forall \vec{f}_i \in E, \vec{u} = \sum_{i=1}^{i=n} \alpha_i \vec{f}_i$. La norme $\| \vec{u} \|$ est définie ci-dessous.

$$\| \vec{u} \| = \sqrt{\frac{\sum_{i=1}^{i=n} \alpha_i (f_i(x_i))^2}{\sum_{i=1}^{i=n} \alpha_i}}$$

La fonction résultante r de n fonctions caractéristiques est :

$$\forall (x_1, \dots, x_i, \dots, x_n) \in [-1, 1]^n, r(x_1, \dots, x_i, \dots, x_n) = \frac{1}{\sqrt{n}} \sqrt{\frac{\sum_{i=1}^{i=n} \alpha_i (f_i(x_i))^2}{\sum_{i=1}^{i=n} \alpha_i}}$$

Le principal inconvénient est quand n devient grand, la fonction résultante r devient très faible. En effet $\lim_{n \rightarrow +\infty} r(x_1, \dots, x_i, \dots, x_n) = 0$. Toutefois, ce n'est pas très grave, car le nombre d'attributs x_i et donc de fonctions caractéristiques $n \ll +\infty$ pas très grand. Il convient en pratique de traiter d'abord l'ensemble les attributs x_i à valeur positive $r_{x_i \geq 0}(x_i)$ puis ceux x_j à valeur négative $r_{x_j < 0}(x_j)$, puis combiner ensemble les vecteurs $\vec{u} = \overrightarrow{u_{x_i \geq 0}} + \overrightarrow{u_{x_j < 0}}$ ce qui ne pose pas de problème puisque l'addition de vecteurs est associative.

4.2.4 Représentation du temps

A chaque instant t ces paramètres peuvent évoluer plus ou moins rapidement ou être constants. Les échelles temporelles (ex : heures, jours) associées aux différents paramètres sont converties en temps de référence par des fonctions de conversion dans un but de synchronisation. La variable t est utilisée pour définir pour chaque attribut x_i à l'instant t chaque fonction caractéristique $F(x_{i,t})$.

$$F(x_{i,t}): [-1, 1] \rightarrow [-1, 1], F(x_{i,t}) = \frac{2}{1 + e^{-k(x_{i,t}-s)}} - 1, \quad t \in [t_0, t_n]$$

Certaines variables et fonctions caractéristiques sont constantes sur un intervalle de temps $[t_0, t_n]$.

A chaque instant t , le vecteur \vec{u}_t évolue selon les fonctions caractéristiques associées à chaque variable $x_{i,t}$ et les vecteurs correspondants $\vec{f}_{i,t}$.

$$\forall \vec{f}_{i,t} \in E, \vec{u}_t = \sum_{i=1}^{i=n} \alpha_{i,t} \cdot \vec{f}_{i,t}$$

$$\text{La norme } \|\vec{u}_t\| = \sqrt{\frac{\sum_{i=1}^{i=n} \alpha_{i,t} (f_i(x_{i,t}))^2}{\sum_{i=1}^{i=n} \alpha_{i,t}}}$$

Le scalaire $\alpha_{i,t}$ qui représente l'importance de la $i^{\text{ème}}$ variable au temps t permet de définir la fonction caractéristique dans l'ensemble. La fonction résultante devient :

$$\forall (x_{1,t}, \dots, x_{i,t}, \dots, x_{n,t}) \in [-1,1]^n, r(x_{1,t}, \dots, x_{i,t}, \dots, x_{n,t}) = \frac{1}{\sqrt{n}} \sqrt{\frac{\sum_{i=1}^{i=n} \alpha_{i,t} (f_i(x_{i,t}))^2}{\sum_{i=1}^{i=n} \alpha_{i,t}}}, t \in [t_0, t_n].$$

4.3 Modèle de la démarche clinique

Le modèle de la démarche clinique s'appuie sur les travaux de Colloc et col. (SHEN ET AL., 2012; 2015; YAMEOGO, COLLOC, 2016; COLLOC ET AL., 2017; COLLOC, SYBORD, 2003).

Les ontologies réalisent l'axiomatisation d'un domaine de connaissances. Elles sont généralement codées à l'aide d'un langage déclaratif de programmation logique qui permet d'associer des descriptions écrites dans un langage commun aux différents agents avec des instructions formelles. Notre modèle ontologique se fonde sur la richesse sémantique des modèles orientés objet. La notation UML qui permet d'exprimer les relations sémantiques définissant la terminologie d'un domaine de connaissance à l'aide du méta modèle d'UML. L'ontologie (figure 37) décrit la maladie comme une succession de tableaux cliniques comportant des signes cliniques pathognomoniques (SP), obligatoires (SO), évocateurs (SE) ou accessoires (SA) liés à une pathologie. Les signes seront regroupés en syndromes et leurs associations pour définir la forme clinique de la maladie.

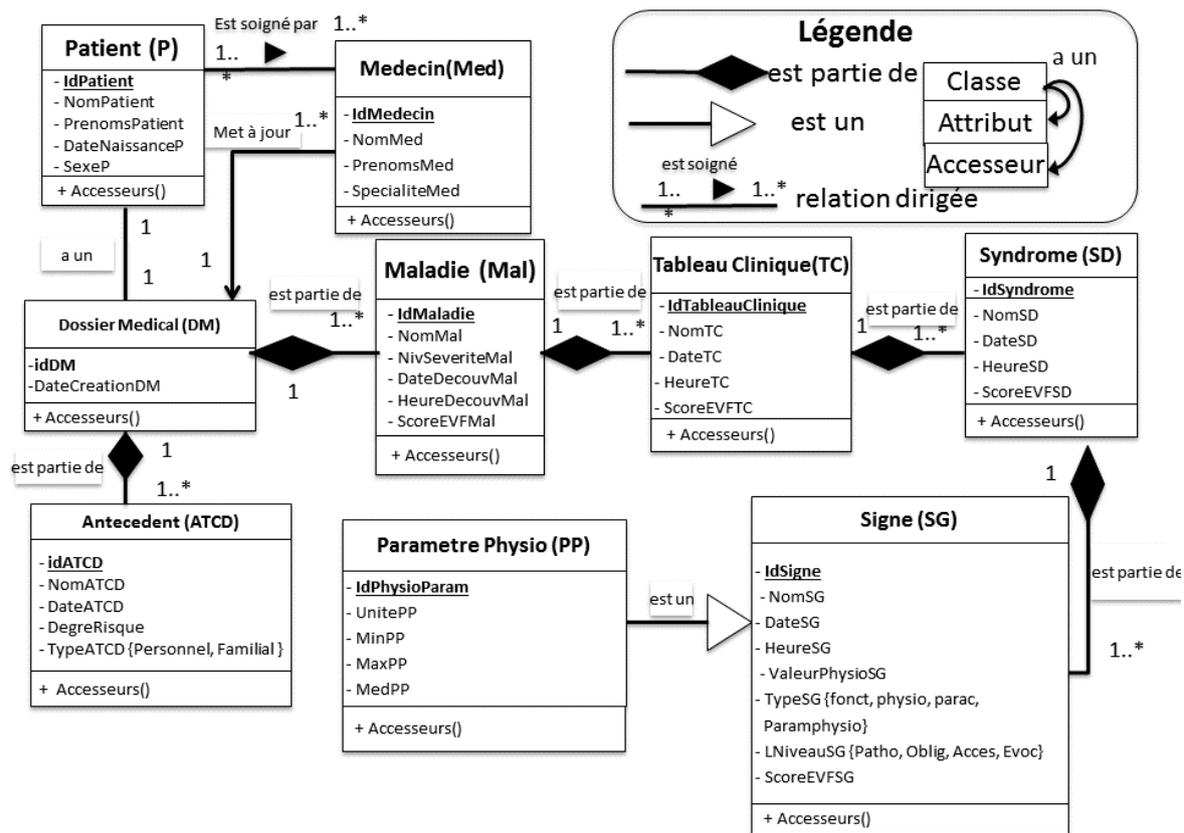


Figure 37 : schéma UML général de l'évolution d'une maladie

Tableau 5 : dictionnaire des attributs des objets maladie et de leurs composants

| Attributs | Libellé des attributs |
|-------------------|---|
| ScoreEVF | Score espace flou vectoriel pour les signes, les syndromes, les formes cliniques et la maladie |
| NivSévérité | Niveau de sévérité |
| NiveauSigne | Niveau de spécificité du signe (Patho = pathognomonique, Oblig = obligatoire, Acces = accessoire et Evoc = évocateur) |
| TypeSigne | Type signe (Phys = signe physique, Fonct = Signe fonctionnel et Parac = Signe paraclinique). |
| ValeurPhysioSigne | Valeur physiologique |
| UnitéPP | Unité du paramètre physiologique et se rapporte à la valeur physiologique |
| MinPP | Valeur minimale du paramètre physiologique |
| MaxPP | Valeur maximale du paramètre physiologique |

| | |
|------|---|
| VCPP | Valeur Centrale PP. Selon le type de distribution et de paramètre on pourra choisir la médiane, la moyenne ou $(MaxPP-MinPP)/2$ du paramètre physiologique. |
| | Les accesseurs publics (+) permettent de construire les objets, de lire, d'écrire, de modifier et de calculer les valeurs des attributs à partir des variables. |

La figure 38 traduit l'évolution des tableaux cliniques dans le temps ainsi que l'évolution de la maladie. Chaque évaluation clinique du patient à un temps t donné correspond à un tableau clinique. Ainsi un patient ayant été évalué n fois correspondrait à lui seul à n tableaux cliniques successifs déterminant des évolutions potentielles ; ce qui est très important pour la base de données de connaissances. Cette prise en compte de l'évolution de la pathologie représente un des points forts du modèle proposé (YAMEOGO, COLLOC, 2016 ; COLLOC ET AL., 2017).

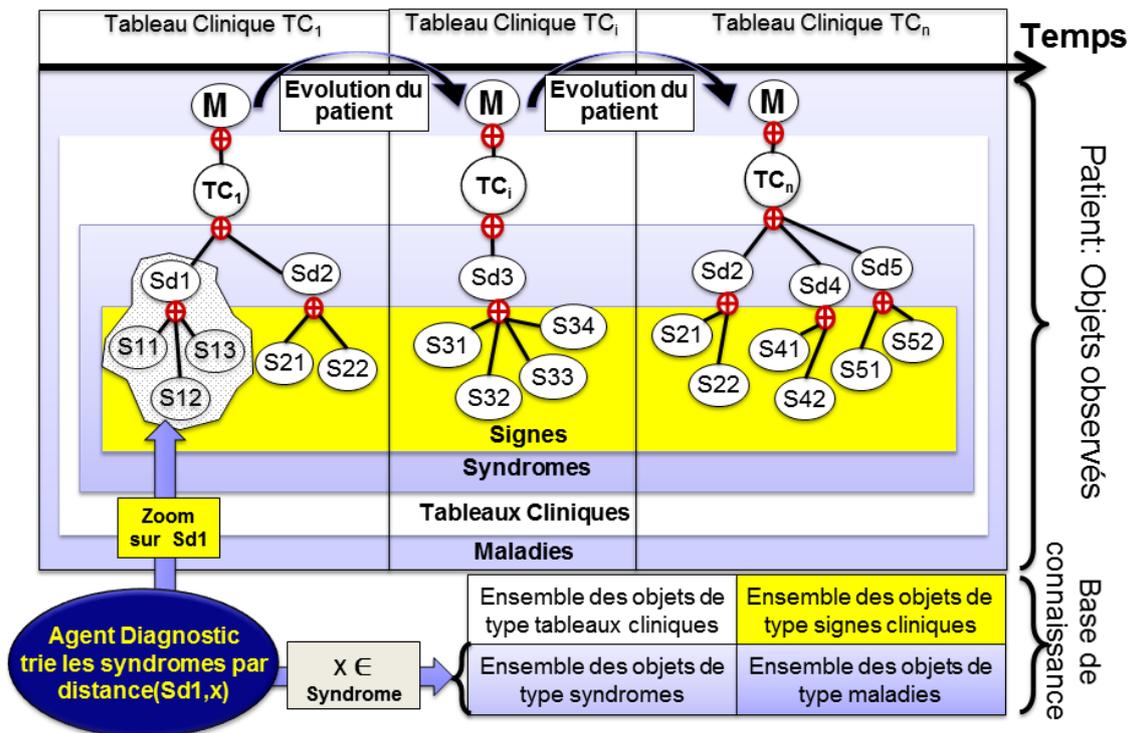


Figure 38 : évolution de la maladie (M) dans le temps en fonction des tableaux cliniques (TC), des syndromes (Sd) et des signes (S) (COLLOC, SUMMONS, 2006)

Cela permet d'améliorer la comparabilité des cas cliniques et de mettre en œuvre un raisonnement analogique non seulement des signes cliniques mais aussi des histoires cliniques dans une approche de raisonnement à partir de cas.

4.4 Modèle temporel flou du diagnostic clinique

Dans le cadre du développement de notre score, nous utilisons un espace vectoriel flou temporel (EVFT) que nous avons déjà utilisé dans des travaux antérieurs (YAMEOGO, COLLOC, 2016 ; COLLOC ET AL., 2017). Il est une extension de la logique floue décrite par Zadeh (ZADEH, 2008 ; COLLOC, 2020). La composition des objets tableaux cliniques est modulée par une fonction caractéristique floue qui évolue dans le temps. L'espace vectoriel flou est représenté par les fonctions caractéristiques booléennes des variables cliniques qui constituent un score exprimant la possibilité d'embolie pulmonaire.

Les attributs des variables sont combinés dans l'espace vectoriel flou qui combine les signes et tableaux cliniques composants qui sont définis par un ensemble de paramètres et d'attributs dont les valeurs sont évaluées par des fonctions caractéristiques floues et des vecteurs associés. Chaque variable est représentée par une fonction sigmoïde définie sur $[-1, 1]$; ce qui permet de les reformuler et de pondérer en fonction de leur importance.

Le score final est lui-même évalué sur un intervalle de $[-1, 1]$ par une fonction sigmoïde dont la valeur indique la « possibilité » (au sens de Zadeh) de diagnostic clinique de l'EP. Ce procédé présente l'avantage d'effacer les incertitudes et les imprécisions au voisinage des seuils des intervalles qui sont fixés par les cliniciens selon leur expérience. Un espace vectoriel flou temporel (EVFT) définit l'évolution des indicateurs cliniques impliqués dans le système. Ainsi, la composition des objets tableaux cliniques est modulée par des fonctions caractéristiques floues dont les valeurs évoluent avec le temps. La figure 31 montre la dynamique des tableaux cliniques, syndromes et signes dans le temps sous forme de vecteurs définis par la fonction d'appartenance (ou fonction caractéristique).

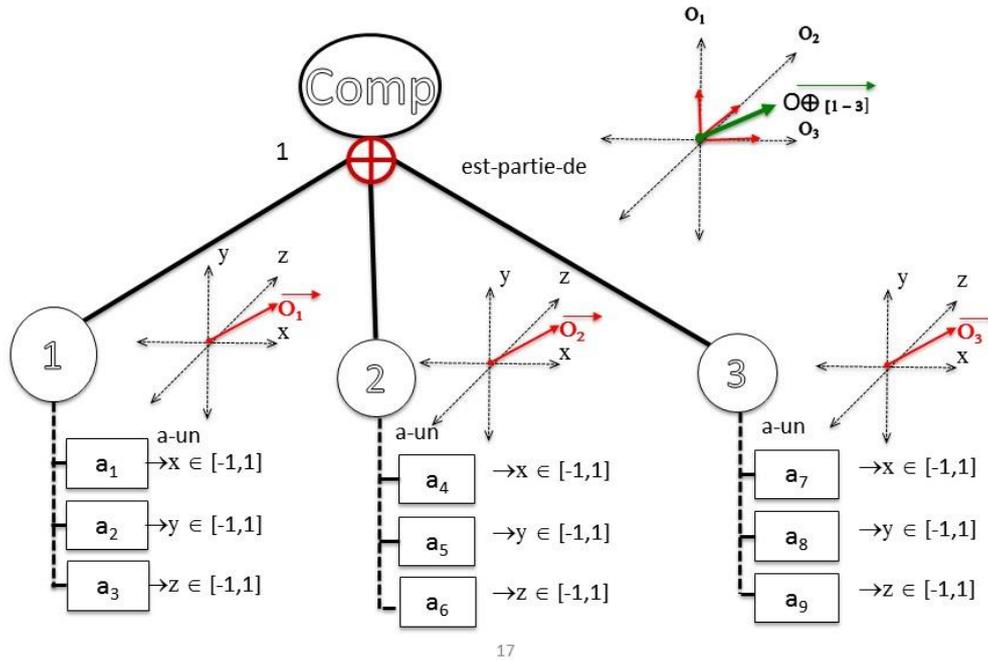


Figure 39 : espace vectoriel flou avec les vecteurs composant les variables et leurs attributs (COLLOC, SUMMONS, 2015)

4.5 Méthodologie de la conception de notre score diagnostic

L'objectif principal de notre travail est la construction d'un score de prédiction clinique de l'EP à partir d'un espace vectoriel flou temporel. Ce score a pour objectif d'améliorer les capacités de diagnostic clinique de l'EP en prenant en compte les incertitudes, les imprécisions et l'évolution clinique de la maladie afin d'améliorer la prise en charge avant la confirmation à l'angioscanner. C'est un enjeu de santé publique et économique au regard des statistiques de mortalité présentées plus haut. L'avantage de l'utilisation de l'espace vectoriel flou est le fait qu'elle se fonde sur des connaissances expertes et les recommandations de la ESC (KONSTANTINIDES ET AL., 2019) sur la prise en charge de l'EP qui ont été validées par les sociétés savantes et les experts du domaine.

4.5.1 Choix des variables

L'analyse des recommandations de prise en charge de l'EP et des scores (score de Wells et score de Genève) nous ont permis de faire le choix de nos variables et de faire une classification des variables. Ainsi, nous distinguons trois types de variables dans le cadre du diagnostic de l'EP :

- les variables statiques et les antécédents qui sont quasi permanents ou permanents ;
- les variables évolutives avec une faible variabilité ;
- les variables dynamiques avec une variation forte et rapide

Tableau 6 : différentes variables prises en compte dans notre modèle de score

| | |
|----------------------|---|
| Variables statiques | Age Facteurs de risque thromboembolique Cancer Intervention chirurgicale ou immobilisation prolongée |
| Variables évolutives | Hémoptysie Douleur unilatérale au membre inférieur Augmentation unilatérale du volume du mollet |
| Variables dynamiques | Fréquence cardiaque Tension artérielle systolique Variation de la tension artérielle systolique |

Nous utilisons la modélisation UML de l'évolution clinique de la maladie (figures 37, 38 et 39) pour définir celle de l'EP. Ainsi, nous disposons des informations sur :

- le patient : l'âge est le paramètre utilisé dans la modélisation de notre score ;
- des antécédents : les facteurs de risque thromboembolique, le cancer et l'antécédent d'hospitalisation ;
- les symptômes et paramètres cliniques : l'hémoptysie, la douleur unilatérale de la jambe, augmentation du volume du mollet, la fréquence cardiaque et la tension artérielle systolique ;
- les syndromes : le syndrome de phlébite constitué de la douleur unilatérale du membre inférieur et de l'augmentation du volume du membre inférieur et le syndrome de l'hypotension artérielle constitué par une tension artérielle systolique < 90 mm Hg et/ou une variation de la tension artérielle > 40 mm Hg. La présence d'un des deux syndromes suffit pour considérer la présence d'une embolie pulmonaire jusqu'à confirmation ou infirmation par l'angioscanner pulmonaire.

Dans le cas de l'EP, il n'existe pas de signes cliniques pathognomoniques ou obligatoires ; mais des signes évocateurs et des signes accessoires rendant difficile le diagnostic. En plus,

il s'agit d'une pathologie très évolutive à court terme, d'où l'intérêt de la prise en compte du temps dans le développement de notre score.

4.5.2 *Fonction caractéristique des variables*

Une fois le choix des variables réalisées, nous avons procédé à la constitution des ensembles flous avec la fonction d'appartenance. Il existe deux grandes catégories de variables, les variables discrètes (qui ne comportent que quelques valeurs possibles) et les variables continues pour lesquelles on peut calculer directement une valeur par l'utilisation d'une fonction caractéristique floue continue. Ce qui présente l'avantage de s'affranchir de la nécessité de définir arbitrairement des seuils et de l'incertitude à leur voisinage. A chaque fois que l'attribut a une valeur continue, nous utiliserons une fonction caractéristique continue. C'est le cas pour la plupart des mesures.

A cet effet, nous utilisons la fonction d'appartenance vectorielle floue définie ci-dessous :

$$\mathbf{x} \in [-1, 1] \rightarrow \overrightarrow{f(\mathbf{x})} \in [-1, 1], \overrightarrow{f(\mathbf{x})} = \frac{2}{(1+e^{-k(x-s)})} - 1.$$

Il s'agit d'une fonction sigmoïde caractéristique des valeurs des attributs sur leur domaine de définition $[min, max] \rightarrow [-1, 1]$. Ce qui permet de normaliser l'espace de définition de chaque attribut et d'obtenir une homogénéité dans l'évaluation de chaque critère de décision clinique.

À partir de cette fonction d'appartenance des valeurs ont été attribuées aux différentes variables afin de pouvoir calculer les vecteurs de chaque variable. Une fois les vecteurs calculés, la résultante de tous les vecteurs permet de calculer le risque de développement de l'embolie pulmonaire. Un espace vectoriel comporte les propriétés de commutativité, d'associativité de l'addition des vecteurs et du produit scalaire qui permettent la composition des attributs des objets composants pour exprimer le vecteur résultant au niveau des objets composés.

Dans le cadre des variables quantitatives, nous utiliserons la fonction ci-dessous pour la normalisation des variables avec des valeurs centrales V_{cent} qui selon le type d'attribut, sa distribution et cardinalité sera : la moyenne, la médiane ou la valeur centrale $(V_{max} + V_{min}) / 2$ ainsi que le minimum (min) et le maximum (max).

$$H: X \in [min, max] \rightarrow [-1, 1], \quad x = H(X) = \begin{cases} X = min, & x = -1 \\ X \geq med, & x = \frac{X - Vcent}{Vmax - Vcent} \\ X < med, & x = \frac{X - Vcent}{Vcent - min} \\ X = max & x = 1 \end{cases}$$

Variables statiques

- Age

L'âge est un facteur de risque dans la survenue de l'EP du fait qu'avec son augmentation les personnes deviennent de plus en plus moins mobiles. L'âge est considéré comme un facteur de risque de MTEV au fur à mesure qu'il augmente avec un âge supérieur à 65 ans comme âge limite. Le risque est fort quand l'âge est supérieur à 65 ans, modéré pour un âge compris entre 50 en 65 ans et faible si l'âge est inférieur à 50 ans. Notre approche permet de nous affranchir de ces seuils en utilisant une fonction caractéristique continue lorsque l'attribut considéré est continu. On peut dire que la fonction $\mu_{RisqueAge}(H(x)) = foH(x)$.

Tableau 7 : fonction caractéristique floue $\mu_{RisqueAge}(H(x))$

| âge | H(x) | $\mu_{age}(H(x))$ |
|-----|-------|-------------------|
| 35 | -1,00 | -1,00 |
| 36 | -0,93 | -0,98 |
| 38 | -0,80 | -0,96 |
| 39 | -0,73 | -0,95 |
| 40 | -0,67 | -0,93 |
| 41 | -0,60 | -0,91 |
| 42 | -0,53 | -0,87 |
| 43 | -0,47 | -0,82 |
| 44 | -0,40 | -0,76 |
| 45 | -0,33 | -0,68 |
| 46 | -0,27 | -0,58 |
| 47 | -0,20 | -0,46 |
| 48 | -0,13 | -0,32 |

| âge | H(x) | $\mu_{age}(H(x))$ |
|-----|------|-------------------|
| 51 | 0,07 | 0,17 |
| 52 | 0,13 | 0,32 |
| 53 | 0,20 | 0,46 |
| 54 | 0,27 | 0,58 |
| 55 | 0,33 | 0,68 |
| 56 | 0,40 | 0,76 |
| 57 | 0,47 | 0,82 |
| 58 | 0,53 | 0,87 |
| 59 | 0,60 | 0,91 |
| 60 | 0,67 | 0,93 |
| 61 | 0,73 | 0,95 |
| 62 | 0,80 | 0,96 |
| 63 | 0,87 | 0,97 |

| | | |
|----|-------|-------|
| 49 | -0,07 | -0,17 |
| 50 | 0,00 | 0,00 |

| | | |
|----|------|------|
| 64 | 0,93 | 0,98 |
| 65 | 1,00 | 1,00 |

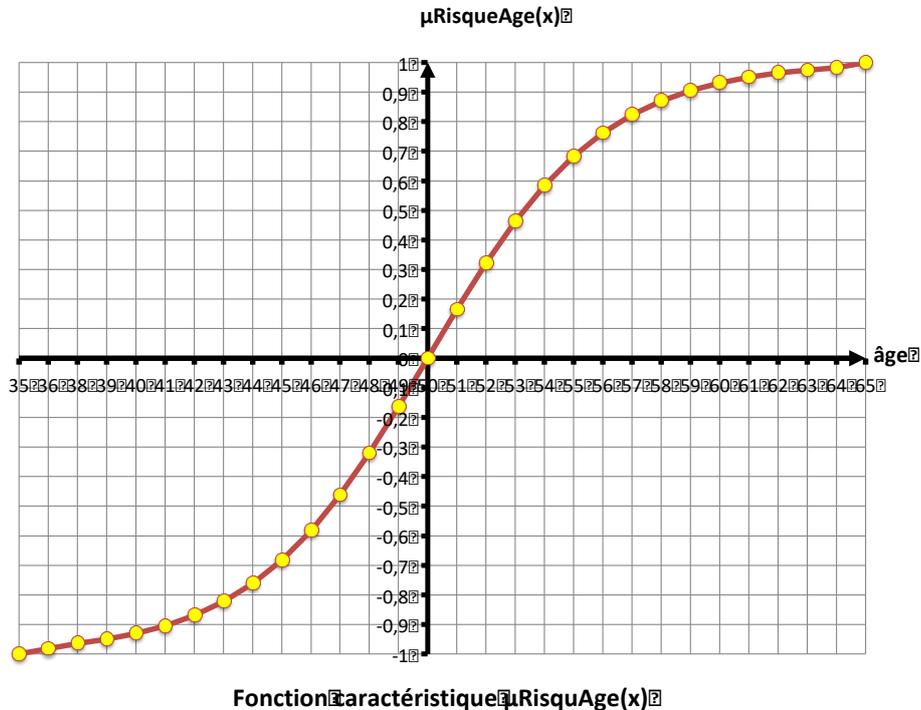


Figure 40: fonction caractéristique $\mu_{RisqueAge}(x)$

- **Facteurs de risque thromboembolique**

Les facteurs de risque sont définis par une présence ou absence en fonction de la liste ci-dessus dans le tableau 1. Quand il est présent, il est classé en risque faible, modéré ou fort. Ces facteurs de risque sont stratifiés en fonction du odd ratio. Ainsi, nous avons associé un facteur multiplicatif au facteur en fonction de son odd ratio. Pour le risque fort, le facteur est égal à $w_i = 8$, pour le risque modéré à $w_j = 4$ et pour le risque faible à $w_k = 2$. Nous avons sept facteurs de risque fort, 18 facteurs de risque modéré et dix facteurs de risque faible. L'équation pour la variable facteur de risque est :

$$Score = w_i \sum_{i=1}^7 RkF + w_j \sum_{i=1}^{18} RkM + w_k \sum_{i=1}^{10} Rkf$$

RkF = risque fort ; RkM = risque modéré ; Rkf = risque faible

La valeur minimale de $y = 0$ et la valeur maximale $y = 20$; la valeur médiane est égale à 7 correspondant à une probabilité égale à 0.

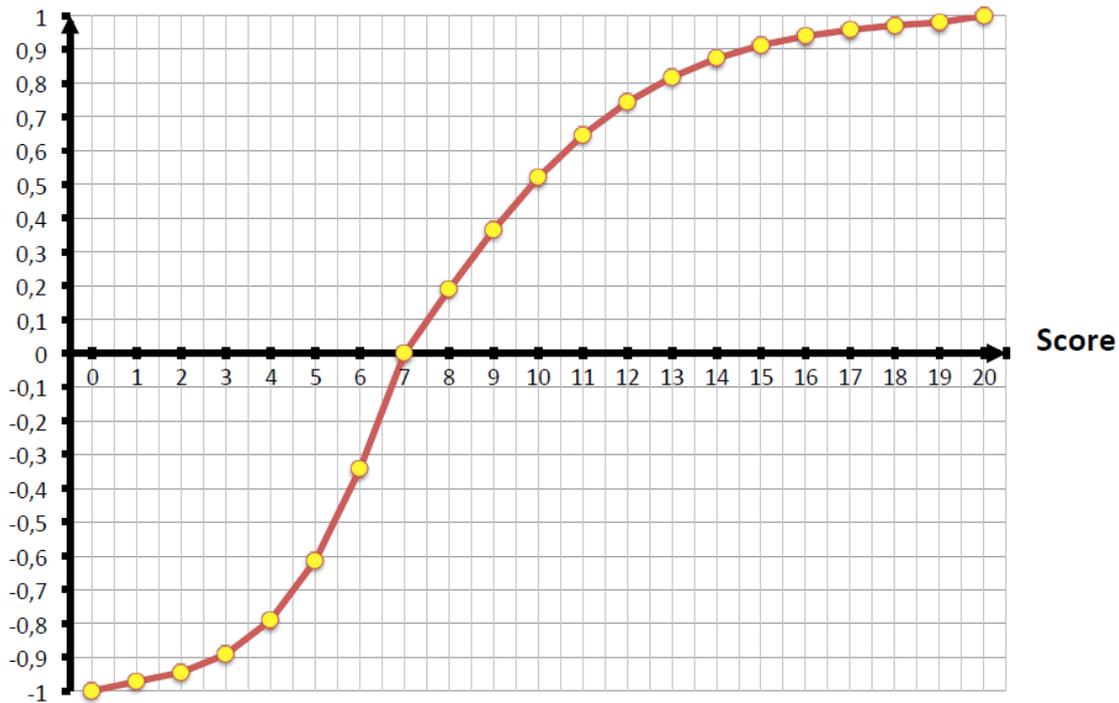


Figure 41 : fonction caractéristique du risque thromboembolique $\mu\text{ThE}(H(x))$

- **Cancer**

Le cancer est classé en fonction de la classification par stade des cancers¹⁷⁴ :

stade 0 – c’est un carcinome in situ ou un changement précancéreux ;

stade 1 – la tumeur est habituellement petite et elle ne s’est pas développée hors de l’organe dans lequel elle a pris naissance ;

stades 2 et 3 – la tumeur est grosse ou s’est développée hors de l’organe dans lequel elle a pris naissance jusque dans les tissus voisins ;

stade 4 – le cancer s’est propagé par le sang ou le système lymphatique jusqu’à un emplacement éloigné (propagation métastatique).

À partir de cette classification, les stades 0 et 1 sont considérés à risque faible, les stades 2 et 3 à risque modéré et le stade 4 à risque fort de MTEV. La valeur -1 désigne l’absence de cancer connu.

Cet attribut stade est discret : voici le tableau des valeurs et la fonction caractéristique correspondante, -1 correspond à l’absence de cancer. La fonction caractéristique est construite de la même manière que celle des facteurs de risque thromboembolique.

¹⁷⁴ Source : <https://fqc.qc.ca/fr/information/le-cancer/classification-cancer> consulté le 20 janvier 2019.

Tableau 8 : fonction caractéristique $\mu_{\text{cancer}}(H(x))$ selon le stade du cancer

| Stade cancer | $H(x)=\exp(\text{stade})$ | $\mu_{\text{cancer}}(H(x))$ |
|--------------|---------------------------|-----------------------------|
| -1 | 0 | -1 |
| 0 | 1 | -0,8 |
| 1 | 3 | -0,4 |
| 2 | 7 | 0,4 |
| 3 | 20 | 0,8 |
| 4 | 55 | 1 |

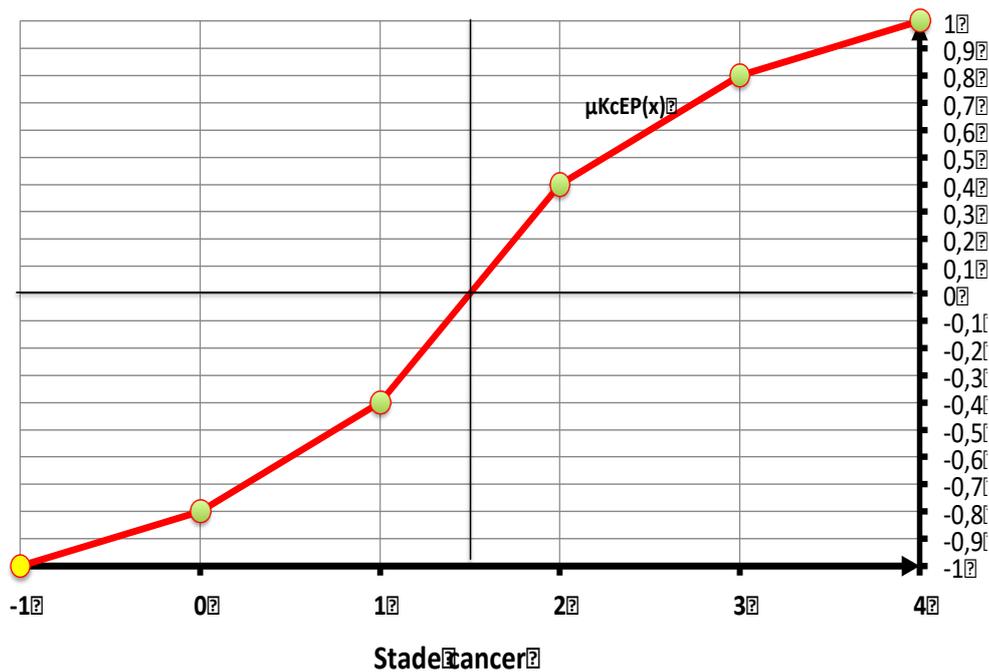


Figure 42 : fonction caractéristique floue μ_{KcEP} attribut discret risque d'embolie pulmonaire selon le stade de cancer $x \in [-1, 4]$

- **Intervention chirurgicale ou immobilisation prolongée**

La durée depuis une intervention chirurgicale ou d'une immobilisation après une fracture est considérée comme un risque d'embolie pulmonaire, si elle est inférieure à un mois.

Comme le temps écoulé est une variable continue, la fonction caractéristique $\mu_{RIntervChir}(t)$ montre que la fonction caractéristique décroît au fur et à mesure que le temps s'écoule (t est exprimé en jours) depuis la date de la fracture ou de l'intervention chirurgicale.

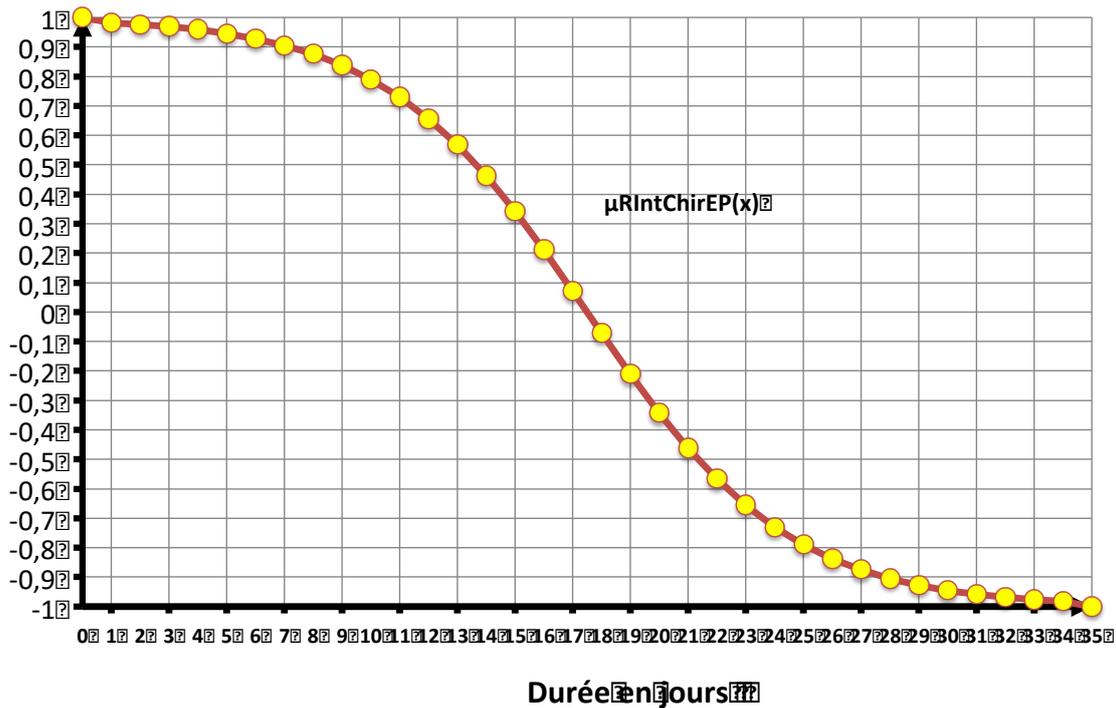


Figure 43 : fonction caractéristique floue du risque d’embolie pulmonaire lié au délai écoulé après une fracture ou une intervention chirurgicale

Variables évolutives

- **Hémoptysie**

L’hémoptysie était définie par la présence ou l’absence. Quand elle était présente, le risque est fort si l’hémoptysie est franche, modéré si ce sont des crachats striés de sang et inexistant en absence d’hémoptysie. L’hémoptysie est un attribut discret, la fonction caractéristique est donc discrète et exprimée par l’équation suivante.

$$\mu RHemoptysie(t) = \begin{cases} \textit{inexistante} : -1 \\ \textit{crachats striés} : 0 \\ \textit{franche} : 1 \end{cases}$$

- **Douleur unilatérale au membre inférieur**

La douleur unilatérale au membre inférieur est définie par une présence ou une absence. Le risque est fort si la douleur survient à la palpation profonde du mollet, modéré si la douleur est spontanée et faible si la douleur est absente. Cet attribut est discret, donc la fonction caractéristique est exprimée de la même manière que celle de l’hémoptysie.

$$\mu_{RDouleurMI}(t) = \begin{cases} \text{absente} : -1 \\ \text{spontanée} : 0 \\ \text{palpation} : 1 \end{cases}$$

- **Augmentation unilatérale du volume du mollet**

L'augmentation unilatérale du membre inférieur est définie par une différence de trois centimètres entre les deux membres au niveau du mollet. Le risque est fort si le membre inférieur est augmenté de volume, modéré si la différence est comprise entre un et deux centimètres et faible si la différence est inférieure à un centimètre.

Tableau 9 : paramétrage des valeurs de comparaison de la taille des mollets gauche et droite en cm

| Diff Taille (cm) | H(x,t) | $\mu_{RDiffTailleMollet}(H(x,t))$ | Vmax | 4 |
|------------------|--------|-----------------------------------|-----------------------|---|
| 0 | -1 | -1,00 | Vmin | 0 |
| 0,2 | -0,9 | -0,98 | (max+min)/2=Vcentrale | 2 |
| 0,4 | -0,8 | -0,96 | Moyenne | 2 |
| 0,6 | -0,7 | -0,94 | Médiane | 2 |
| 0,8 | -0,6 | -0,91 | | |
| 1 | -0,5 | -0,85 | | |
| 1,2 | -0,4 | -0,76 | | |
| 1,4 | -0,3 | -0,64 | | |
| 1,6 | -0,2 | -0,46 | | |
| 1,8 | -0,1 | -0,24 | | |
| 2 | 0 | 0,00 | | |
| 2,2 | 0,1 | 0,24 | | |
| 2,4 | 0,2 | 0,46 | | |
| 2,6 | 0,3 | 0,64 | | |
| 2,8 | 0,4 | 0,76 | | |
| 3 | 0,5 | 0,85 | | |
| 3,2 | 0,6 | 0,91 | | |
| 3,4 | 0,7 | 0,94 | | |
| 3,6 | 0,8 | 0,96 | | |
| 3,8 | 0,9 | 0,98 | | |
| 4 | 1 | 1,00 | | |

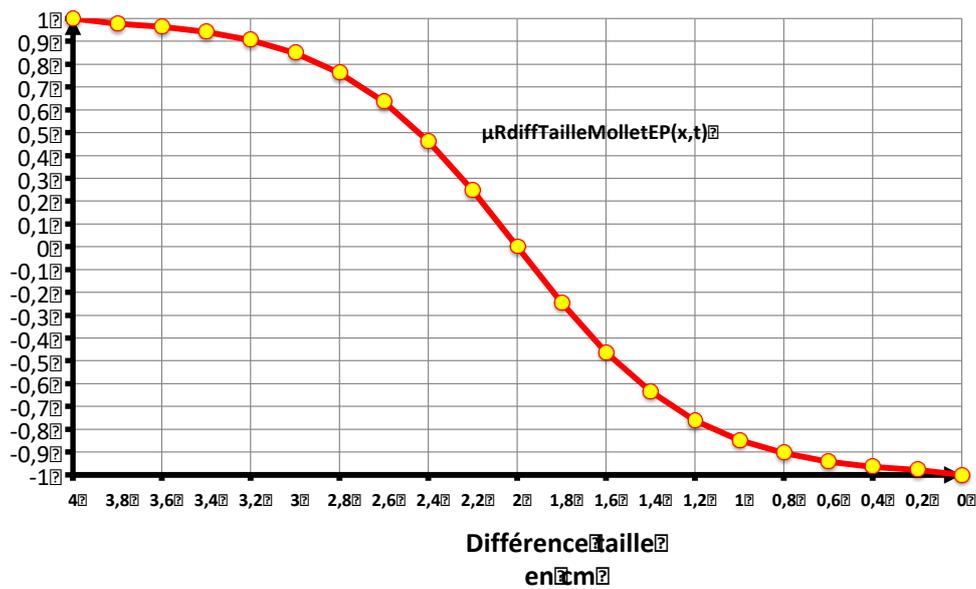


Figure 44 : fonction caractéristique de l'augmentation unilatérale du volume du mollet $\mu_{RdiffTailleMolletEP}(x,t)$

Variables dynamiques

- **Fréquence cardiaque**

La fréquence cardiaque supérieure à 100 battements par minute est un risque fort de MTVE, le risque est modéré quand la fréquence cardiaque est comprise entre 75 et 100 battements par minute et faible si la fréquence cardiaque est inférieure à 75 battements par minute.

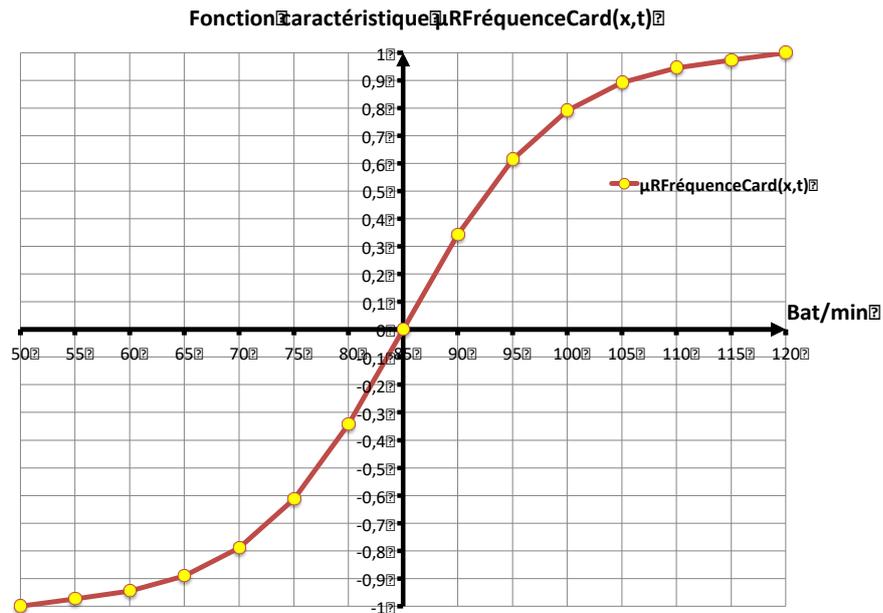


Figure 45 : fonction caractéristique du Risque d'EP lié à une fréquence cardiaque élevée $\mu R_{\text{FréquenceCard}}(x,t)$

- **Tension artérielle systolique**

Lorsque la tension artérielle systolique est inférieure à 90 mm Hg, il existe une hypotension artérielle ou un état de choc et le diagnostic d'embolie pulmonaire est retenu jusqu'à la confirmation ou infirmation par l'angioscanner.

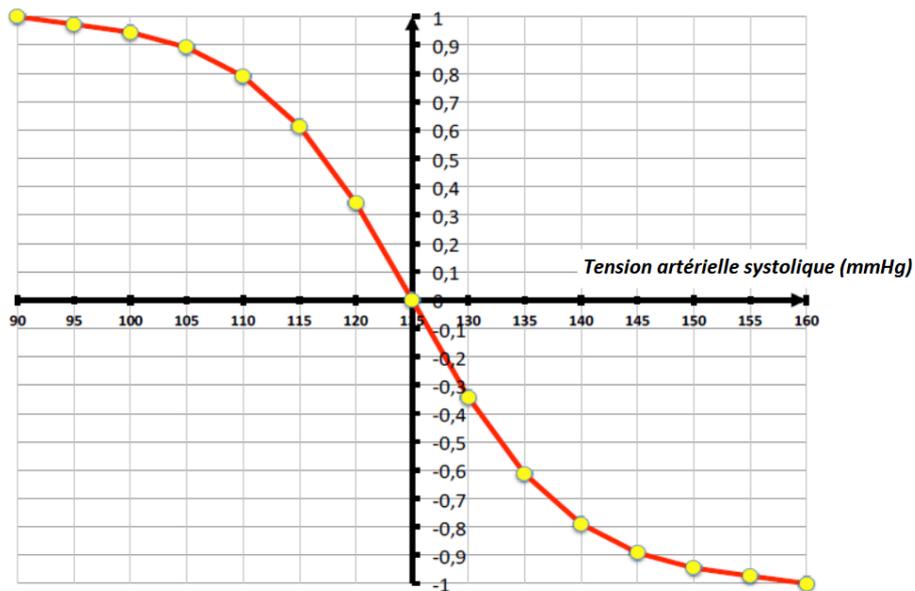


Figure 46 : fonction caractéristique du risque lié à la tension artérielle $\mu R_{\text{HypoTASystolique}}(x,t)$

- **Variation de la tension artérielle**

La variation de la tension artérielle systolique ($\Delta TAS = TAS_{t_{n+1}} - TAS_{t_n}$) ou une chute de la TAS supérieure à 40 mm Hg pendant au moins 15 minutes est considérée comme une hypotension artérielle. Elle a la même valeur diagnostique que la tension artérielle systolique inférieure à 90 mm Hg.

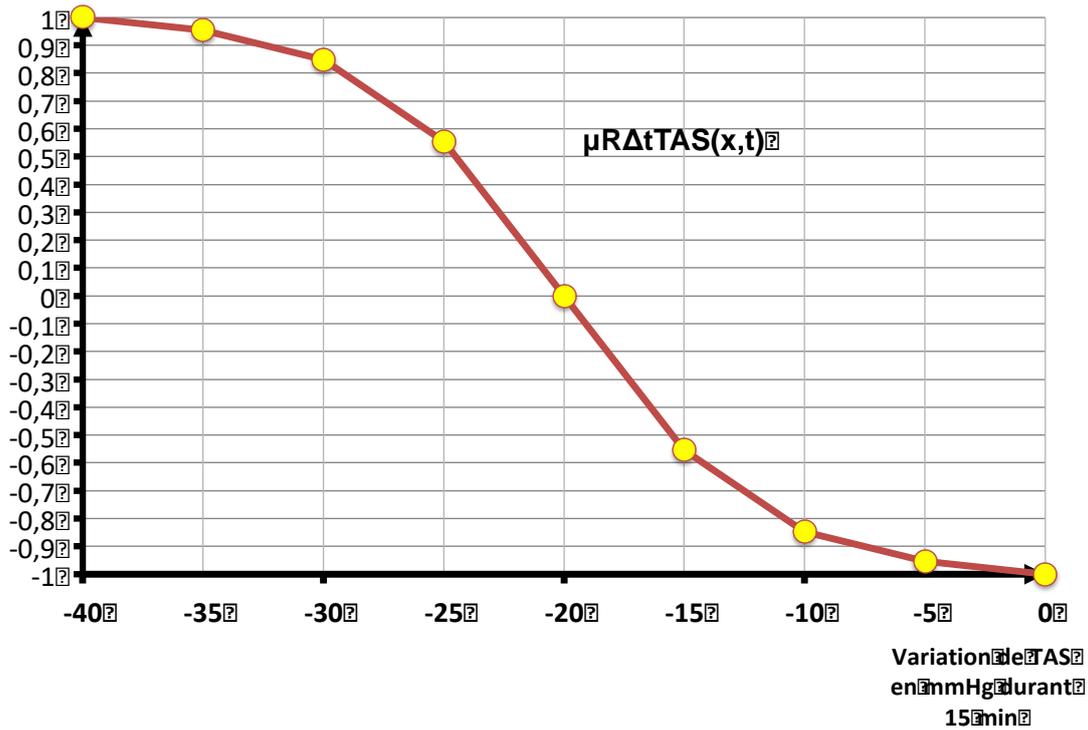


Figure 47 : fonction caractéristique $\mu R \Delta t TAS$ variation de TAS en 15 minutes

Comme l'un ou l'autre critère détermine le risque lié à une chute de la tension artérielle systolique, nous pouvons appliquer l'opérateur OU-inclusif utilisé en logique floue (Max).

$$\mu RTAS(x, t_n) = \text{Max}(\mu RHypoTAS(x, t_n), \mu R \Delta t TAS(x, t_n, t_{n-15}))$$

La maladie correspond à la résultante des valeurs des attributs définissant par compositions successives les objets (signes, syndromes et tableaux cliniques). Pour paramétrer la maladie, il faut paramétrer les différents attributs et définir les vecteurs résultants avant.

La fonction d'évaluation de l'EP est comprise entre [-1, 1] et est définie par la résultante des vecteurs des différentes variables. L'agrégation des variables a été faite en fonction des trois catégories de variables (statiques, évolutives et dynamiques) ; la résultante des trois vecteurs constitue le vecteur de l'EP comme le montre la figure 48 au cours des trois étapes habituelles de la maladie (figure 49) : étape 0 : Antécédents et suspicion de PE ; étape 1 : signes de thrombose veineuse et étape 2 : signes cardiaques pouvant être combiné de manière spécifique dans chaque patient.

Tableau 10 : paramétrage de la variable de l'EP

| Objet | Syndromes et symptômes | α_i | Variable | Valeurs | Fonction |
|-------|------------------------|------------|----------------------------|---|---|
| EP | Age | 1 | Age | D=[35,65], H(x)=[-1 ; 1] | $\mu_{\text{RisqueAge}}(H(x))$ |
| | FRTE | 1 | FRTE | D = [0 ; 26,8] H(x) = [-1 ; 1] | $\mu_{\text{ThE}}(H(x))$ |
| | Cancer | 1 | Cancer | D={-1,0,1,2,3,4} H(x)={-1 ; -0,8 ; -0,4 ; 0,4 ; 0,8 ; 1} | $\mu_{\text{kcEP}}(H(x))$ |
| | IntChir | 1 | IntChir | D=[0,35], H(t)=[-1, 1] | $\mu_{\text{RIntervChir}}(H(t))$ |
| | Hemop | 1 | Hemop | D= {inexistante, strié, franche} H(x)={-1, 0, 1} | $\mu_{\text{RHemoptysie}}(x)$ |
| | SyndPh | 1 | DoulMI | D = {absente, spontanée, palpation} H(x) = {-1, 0, 1} | $\mu_{\text{RDouleurMI}}(t)$ $\mu_{\text{RdiffTailleMolletEP}}(x,t)$ $\text{SyndPh} = \frac{\text{DoulMI} + \text{AugMI}}{2}$ |
| | | 1 | AugMI | D = [0 ; 4] H(x) = [-1 ; 1] | |
| | FC | 1 | FC | D= [50,120] H(x)=[-1 ; 1] | $\mu_{\text{RFréquenceCard}}(H(x), t)$ |
| | HypoTA | 1 | TAS | D= [90,160] H(x)=[-1 ; 1] | $\mu_{\text{RHypoTASystolique}}(H(x), t)$ |
| 1 | | VTAS | D= [-40,0] H(x)=[-1; 1] | $\mu_{\text{R}\Delta t\text{TAS}}(x, t_n, t_{n-15})$ | |

La fonction résultante de chaque catégorie de variable est :

- $\overrightarrow{HIST} = \overrightarrow{Age} + \overrightarrow{FRTE} + \overrightarrow{Cancer} + \overrightarrow{IntChir}$
- $\overrightarrow{STV}(t) = \overrightarrow{Hemop}(t) + \overrightarrow{AugMI}(t) + \overrightarrow{DoulMI}(t)$
- $\overrightarrow{SC}(t) = \overrightarrow{FC}(t) + \overrightarrow{HypoTA}(t)$

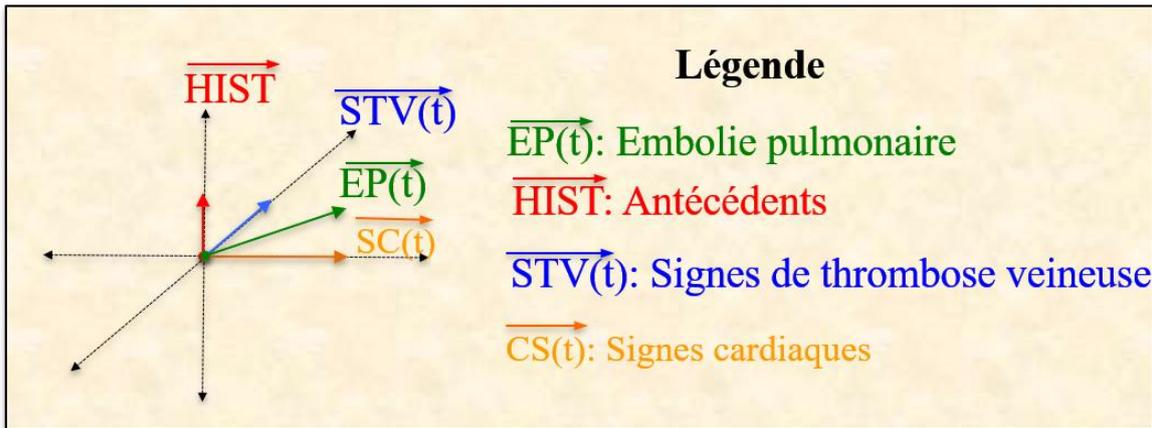


Figure 48: fonction résultante de l'EP

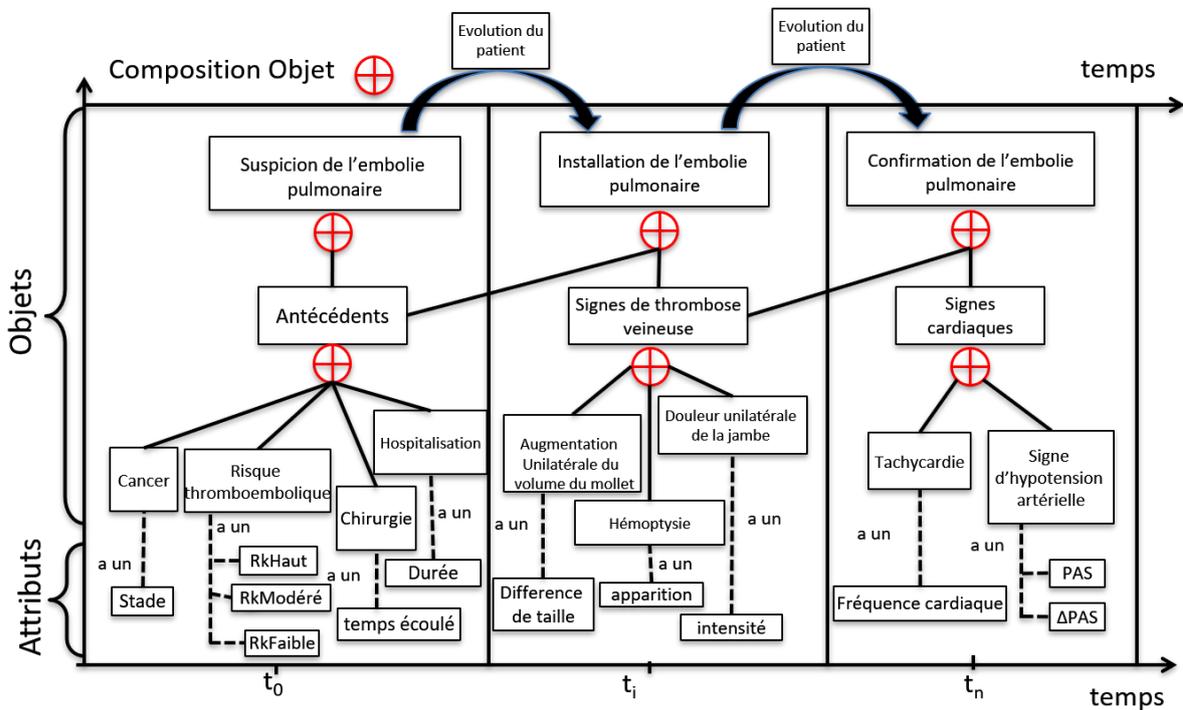


Figure 49 : évolution des variables de l'EP de la suspicion au diagnostic

L'espace vectorielle flou de l'EP est :

$$\overrightarrow{EP}(t) = \overrightarrow{HIST} + \overrightarrow{STV}(t) + \overrightarrow{SC}(t)$$

Ainsi, plus la valeur de la résultante tend vers 1, le risque de présence de l'embolie pulmonaire augmente.

4.6 Intérêts de l'espace vectoriel flou temporel pour le diagnostic de l'EP

Le diagnostic actuel de l'EP à partir des scores cliniques présente l'inconvénient de prendre en compte des indicateurs statiques pour l'évaluation du risque. Cette approche ne prend pas en compte le contexte évolutif de la maladie dans le contexte des urgences.

Ainsi, bien que les scores de Wells et de Genève soient utilisés dans les recommandations de prise en charge de l'EP, ils comportent néanmoins des insuffisances qui sont :

- les scores expriment une photographie de l'état clinique du patient en un temps t donnée sans prise en compte de l'évolution clinique du patient dans le temps. Par exemple, la variation de la fréquence cardiaque n'est pas étudiée dans le temps, pourtant, elle constitue une variable labile ;
- les scores sont fondés sur la logique classique et ne prennent pas ainsi en compte l'imprécision et l'incertitude des variables dont certaines sont constantes alors que d'autres peuvent varier rapidement dans le temps. Par exemple la fréquence cardiaque est une variable très labile, et elle peut faire basculer le patient dans la limite supérieure sous l'effet de facteurs externes ;
- les scores ne prennent pas en compte la variabilité de la tension artérielle qui est pourtant très importante dans le diagnostic de l'hypotension artérielle et l'ensemble des facteurs de risque thromboembolique.

Nous avons utilisé un modèle d'espace vectoriel flou temporel, une nouvelle approche pour combiner la logique flou et le modèle orienté objet et permet la modélisation temporelle des structures et des comportements des objets complexes. Cette approche permet de prendre en compte les propriétés des objets et de les comparer entre eux. Elle permet d'évaluer la dynamique temporelle et cinétique des attributs de certains objets très importante en pratique clinique quand on sait que la dynamique d'évolution d'un signe est plus importante que la valeur de ce signe à un temps t. ce nouveau score de diagnostic clinique de l'embolie pulmonaire peut être très utile dans les zones où l'angioscanner n'est pas disponible.

Approche très importante dans le cas de l'EP qui est une maladie très évolutive à court terme avec une mortalité importante. Cette évolution rapide concerne les variables dynamiques et à moindre mesure les variables évolutives ; car l'évolution de la maladie concerne surtout

les valeurs hémodynamiques du fait de l'obstruction de l'artère. Notre modèle temporel prend en compte également les variables labiles telles que la tension artérielle et la fréquence cardiaque et la définition imprécise de l'hypotension artérielle par une seule mesure avec l'aide de la logique floue. Par exemple, chez un patient hypertendu, une mesure de la tension artérielle peut montrer une absence d'hypotension artérielle, mais la mesure répétée permet de montrer cette chute de la tension artérielle par la variation. Ces facteurs associés rendent très intéressant l'utilisation d'un modèle vectoriel flou temporel en affinant la comparabilité des cas cliniques et de mettre en œuvre un raisonnement analogique non seulement des signes cliniques des pathologies mais des histoires cliniques dans une véritable approche de raisonnement à partir de cas.

Un avantage important dans notre modèle est que nous n'avons pas besoin de données pour le fonctionnement du système et le partage de la base de données n'est pas obligatoire pour son fonctionnement. Le modèle de score se base sur les connaissances actuelles de la maladie avec un respect de la pratique médicale (le médecin peut expliquer tous le processus et les facteurs évolutifs de la maladie et prend la décision finale en fonction du contexte clinique de son patient). Nous réalisons à travers ce modèle une IA responsable respectant la vie privée des patients et l'éthique médicale.

Notre objectif avec ce nouvel algorithme de prédiction clinique est d'augmenter la sensibilité et la spécificité du diagnostic de l'EP par les signes cliniques et symptômes. Cela permettra de réduire le retard du diagnostic et de débiter le traitement le plus tôt possible ; Elle contribuera également à réduire les risques associés à la réalisation de l'angioscanner thoracique pour la confirmation du diagnostic (allergies aux produits de contraste ou insuffisance rénale aiguë).

4.7 Méthodologie pour la conception d'un EVFT dans un contexte clinique

Nous avons fondé notre approche de modélisation sur le modèle EPICE (Emotion, Personnalité, Interactions, Connaissances, Expérience) développé par Colloc et Lery (COLLOC, LÉRY, 2008a ; COLLOC, 2020). Ce modèle permet de prendre en compte les comportements humains, les contraintes du terrain et de l'environnement et les interactions entre les différents acteurs. La figure ci-dessous montre les différentes couches du modèle EPICE.

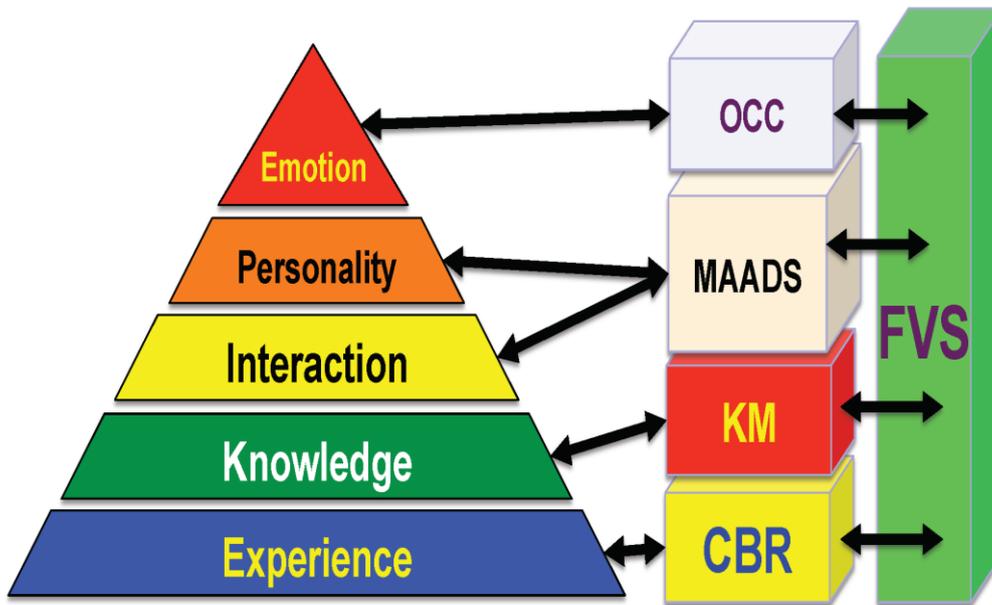


Figure 50: modèle EPICE

OCC : Ortony Clore and Collins model; MAADS : multi-agent aided decision support; KM: knowledge model; CBR: case-based reasoning approach ; FVS: fuzzy vectorial space.

Nous résumons dans cette section les lignes directrices et les tâches nécessaires pour concevoir et implémenter un EVFT à partir d'ontologies dotées de diagramme de classe UML. Notre méthode s'appuie sur notre connaissance orientée objet décrite dans la section 4.3. Les étapes nécessaires à la conception du système et sa mise en œuvre sont décrites dans la figure ci-dessous.

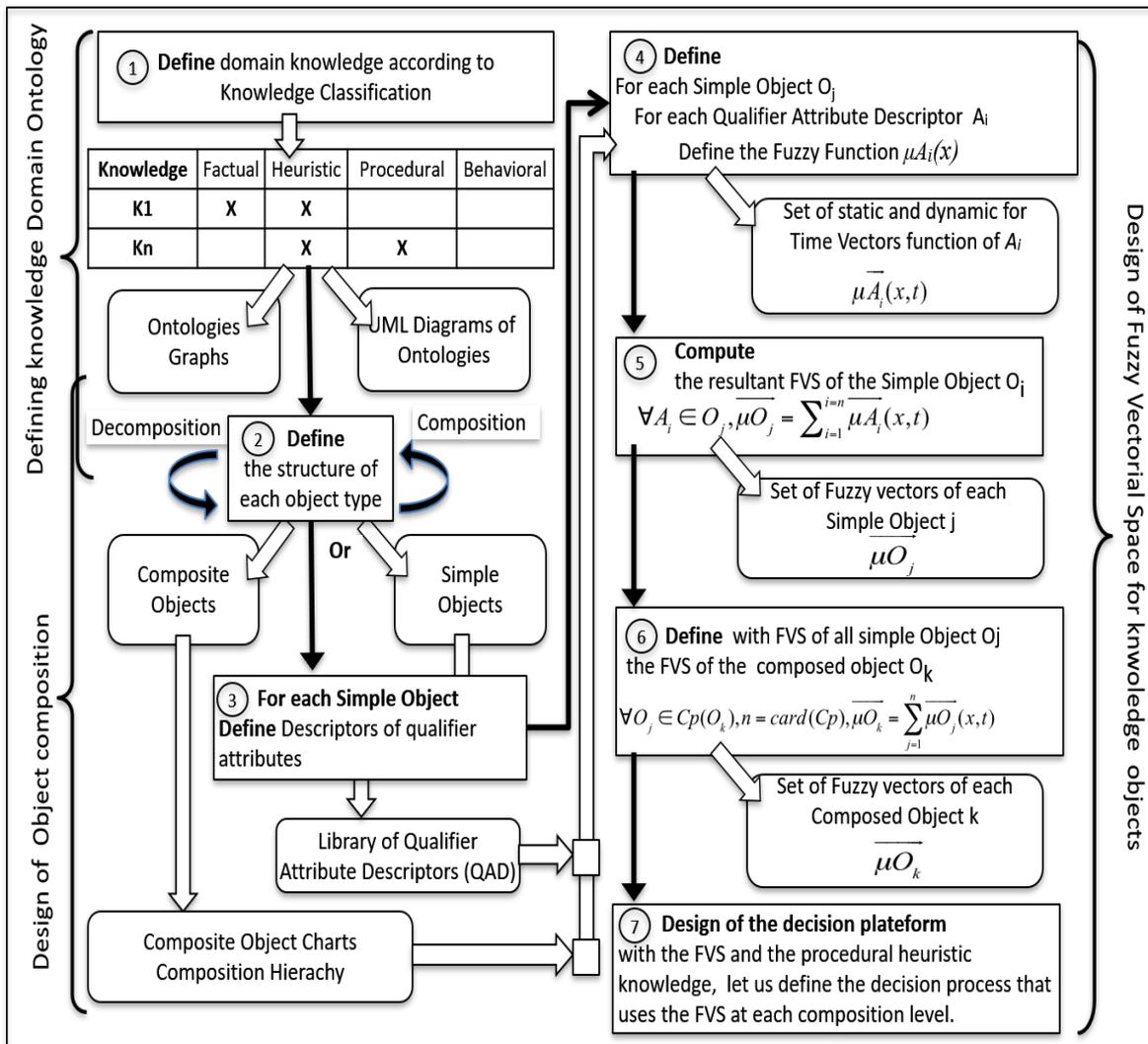


Figure 51: méthodologie pour la conception d'un EVFT

La première étape (1) consiste à concevoir le domaine de connaissance de l'ontologie. Il fournit graphe d'ontologie et son diagramme de classe UML correspondant qui décrit les acteurs, les objets et leurs relations qui sont habituellement utilisés dans le modèle orienté objet : la relation de généralisation/spécialisation (*est un*), la relation de composition (*est une partie de*) et les relations objet-attribut et objet-méthode (*a un*) comme indiqué sur la figure 37, section 4.3. Les connaissances sont évaluées selon la classification des connaissances décrite par Colloc et Lery (COLLOC, LÉRY, 2008a) :

- la connaissance factuelle (*factual knowledge*) : il s'agit de données décrivant un objet du monde réel et généralement accepté par tous comme vrai. Il peut s'agir de classements des faits selon leur degré de certitude et d'exactitude, le plus souvent sur la base d'arguments statistiques.
- la connaissance heuristique (*heuristic knowledge*) : il s'agit d'un état (une conjonction de faits concernant les propriétés d'un objet) ou d'une méthode

habituellement applicable avec succès à un type de situation donnée. Cette connaissance permet de mettre en valeur l'expérience.

- la connaissance procédurale (*procedural knowledge*) : Il s'agit de l'ensemble des procédures ou conduites à suivre : exprimées par une série de tâches à accomplir pour être efficace dans une situation donnée.
- la connaissance comportementale (*behavioral knowledge*): il s'agit d'une connaissance de l'évolution spontanée des faits dans le temps et qui est généralement observée. Elles sont utiles pour la simulation de phénomènes naturels. C'est la perception de différents états de changement des objets du monde réel et leurs interactions prévisibles pendant une période donnée.

La deuxième étape **(2)** vise à définir la structure de chaque type d'objet et, le cas échéant, sa composition (liste des sous-objets composants). Ainsi, il existe deux types d'objets : les objets simples qui n'ont que des attributs et de composition d'objets qui sont composés d'objets simples ou d'autres composition d'objets de manière récursive. Ensuite, pour chaque objet simple, il est nécessaire de définir les descripteurs de chaque qualificatif d'attribut. Un qualificatif d'attribut est une caractéristique obligatoire d'un objet. Le descripteur qualificatif d'attribut (QAD) définit comment l'utiliser comme décrit dans le tableau 5, section 4.3. Le résultat de cette tâche est un ensemble de QAD.

Pour chaque QAD **(3)**, le concepteur doit définir la fonction d'appartenance floue en utilisant les valeurs min, max et centrale et le vecteur correspondant. Cette étape fournit la liste des objets simples, la liste de leur QAD et les graphiques EVF de la composition des objets comme dans la figure 39 section 4.4.

Cette étape **(4)** définit les fonctions d'appartenance floues pour chaque objet simple et pour chaque QADS. Le résultat de cette tâche est un ensemble de fonctions d'appartenance et de fonctions vectorielles appropriées (avec la modélisation temporelle le cas échéant) pour chaque objet simple.

L'étape suivante **(5)** vise à calculer le vecteur résultant de l'EVF de chaque objet simple avec les vecteurs de tous ses QAD. Le résultat de cette étape est une fonction d'appartenance floue résultante qui représente l'état de l'objet simple selon son type d'objet et son rôle dans le

champ d'application du système d'aide à la décision. Cette étape doit prendre en compte les connaissances heuristiques recueillies lors de la première étape de la méthode.

Cette étape (6) permet de définir l'EVF de composition des objets en utilisant les relations de composition et de calculer le vecteur résultant avec les vecteurs résultants déduits lors de l'étape précédente.

Cette étape (7) vise à construire la plateforme de décision à partir des connaissances heuristique et procédurale du processus de décision qu'exploite l'EVF à chaque niveau de composition. Le concepteur doit élaborer deux interfaces principales :

- une interface expert qui permet à l'expert de domaine de mettre à jour les QAD des objets invoqués dans le système et d'éditer la définition des fonctions d'appartenance floue et les vecteurs correspondants.
- une interface d'utilisateur final qui définit les échanges de données nécessaires avec l'utilisateur et le système ou l'interaction avec des objets connectés pour obtenir par exemple le rythme cardiaque, la pression artérielle ou d'autres paramètres de surveillance.

Cette dernière étape (8), qui n'est pas représentée sur la figure 51, concerne le test et l'évaluation de la qualité du système pour chaque cas de demande.

4.8 Perspectives

Nous sommes en train de développer une plateforme pour faciliter la conception du SADM avec le modèle et la méthode. Pour la suite, Nous proposons ainsi de confronter notre modèle sur le terrain. Notre hypothèse est que le nouveau score que nous avons développé permettra de faire un diagnostic clinique efficace et efficient aussi proche des résultats de l'angioscanner sans les inconvénients de l'angioscanner que sont son indisponibilité et la possibilité de complications suite à l'injection des produits de contraste. Si notre hypothèse est validée, elle permettra d'améliorer de façon considérable le diagnostic et la prise en charge de l'embolie pulmonaire dans les pays à ressources limités et de faire des économies en termes de prise en charge de cette maladie.

Les SADM permettent une production des traces numériques dans le cadre de la prise en charge des patients et constituent de ce fait une aide précieuse pour les médecins. Mais, cette numérisation de la trace fait perdre le processus interactionnel de leur production. Il se pose ainsi la question du traitement et de l'interprétation de ces traces.

5 Traces numériques et pratique médicale

Du fait de l'entrée des contemporains dans une société numérique à laquelle l'humain ne peut se soustraire, la question de la trace numérique occupe une place de plus en plus importante dans toutes les disciplines et dans toutes les activités privées comme professionnelles. La médecine n'échappe pas à cette emprise du tout numérique ; ce qui nous préoccupe d'autant plus que, lorsque nous avons décidé d'exercer cette profession, nous avons prêté au serment d'Hippocrate en ces termes :

« Mon premier souci sera de rétablir, de préserver ou de promouvoir la santé dans tous ses éléments, physiques et mentaux, individuels et sociaux » [...] « Admis(e) dans l'intimité des personnes, je tairai les secrets qui me seront confiés ». [...] « Je préserverai l'indépendance nécessaire à l'accomplissement de ma mission. Je n'entreprendrai rien qui dépasse mes compétences. Je les entretiendrai et les perfectionnerai pour assurer au mieux les services qui me seront demandés. J'apporterai mon aide à mes confrères ainsi qu'à leurs familles dans l'adversité » (Extraits du serment d'Hippocrate)¹⁷⁵.

Or, l'obligation numérique qui est faite aujourd'hui dans le domaine médical est susceptible de mettre le médecin en défaut par rapport à ce serment. C'est pourquoi nous avons décidé de nous soustraire pendant quatre ans à notre activité médicale pour cerner ce à quoi notre formation en médecine ne nous avait pas préparé. Nous avons donc commencé à vouloir comprendre la logique informatique et plus précisément les traces dommageables pour l'humain qu'elles pouvaient engendrer.

Les recherches en informatique ont donc retenu plus spécifiquement notre attention. Cependant, à la faveur de l'ouvrage « *De la trace à la connaissance à l'ère du Web* » publié par l'informaticien Alain Mille (MILLE, 2013a) nous avons découvert les paradigmes de *L'Homme-trace* et des *signes-traces* (GALINON-MÉLÉNEC, 2011) et des auteurs en sciences de l'information et de la communication (SIC) tels que Yves Jeanneret, Sylvie Leleu-Merviel, Louise Merzeau. Cette découverte est à l'origine de notre envie d'élargir notre champ de connaissance de la médecine à l'informatique et de l'informatique aux SIC.

¹⁷⁵ Confère le serment d'Hippocrate reporté dans sa totalité en annexe 9.

Cette ouverture disciplinaire s'est trouvée renforcée par un autre article d'Alain Mille intitulé « *Traces numériques et construction de sens* » (MILLE, 2013b) dans la série *L'Homme-trace*¹⁷⁶ publiée chez CNRS éditions. Dans ce chapitre, l'informaticien (MILLE, 2013b) définit la trace comme étant

« *constituée à partir d'empreintes laissées volontairement ou non dans l'environnement informatique à l'occasion d'un processus informatique* ». Il précise que, dans ce contexte, le terme « *environnement* » se réfère à la fois « *à l'espace, aux humains et aux processus interagissant dans le monde informatique* ».

Cette définition est compatible avec celle proposée en SIC par Béatrice Galinon-Méléneq, qui précise : « *l'empreinte est une catégorie de trace qui intègre à la fois à la notion de forme et de durabilité de la trace* » (GALINON-MÉLÉNEC, 2017). Nous la retiendrons donc comme définition passerelle entre les deux disciplines. Dans l'ouvrage « *Traces numériques. De la production à l'interprétation* » (GALINON-MÉLÉNEC, ZLITNI, 2013), informaticiens et chercheurs en SIC s'accordent également sur le fait que, derrière ce que donne à voir l'écran se trouve un travail invisible et complexe, une procédure rationnelle de traitement de données, un chaînage d'opérations nécessaire à la transformation des inscriptions en traces. Notre recherche s'orientera donc également sur la découverte des processus engendrés par les dispositifs numériques en matière de santé et sur les problèmes éthiques qu'ils posent en matière de respect de la vie privée des individus.

Dans cette optique, nous avons lu des auteurs qui, bien avant l'avènement de la « *société numérique* », analysaient déjà la notion de *trace*. Il en ressort des approches qui apparentent la trace à une *empreinte* ou à une *marque psychique* inhérente à la problématique de la *mémoire* (RICOEUR, 2000b), à un *paradigme indiciaire* (GINZBURG, 1980) ou à *l'écriture* (DERRIDA, 1979). Ces auteurs nous ont permis de comprendre que les approches de la notion de trace sont multiples et de faire nôtre la proposition de Béatrice Galinon-Méléneq : « *puisque'il n'existe pas une définition consensuelle de la trace, il convient que tout chercheur qui l'intègre dans sa problématique précise sa perspective épistémologique et l'acception qu'il retient de la notion* » (GALINON-MÉLÉNEC, ZLITNI, LIÉNARD, 2015).

Ce sera l'objet du premier paragraphe, qui sera ensuite contextualisé dans le cadre de l'usage du *big data*, dans le domaine de la santé en particulier.

¹⁷⁶ La série *L'Homme-trace* de CNRS éditions constitue une forme nouvelle d'encyclopédie sur la trace dans la mesure où, depuis 2011, elle publie tous les deux ans un recueil de chapitres scientifiques conduisant le lecteur à s'interroger sur la portée de la trace dans la vie contemporaine.

5.1 De la notion de trace

5.1.1 Définitions proposées dans le cadre de l'Homme-trace

Pour Jeanneret, la trace exprimerait l'idée d'une « *matérialité affectée par un phénomène* » (JEANNERET, 2011)¹⁷⁷. Dans un sens voisin, Galinon-Méléneq indique que

« la trace relèverait d'une terminologie générale qui évoquerait la conséquence, le résultat d'un processus qui relie le présent au passé. En ce sens, tout le réel est trace (conséquence) d'autre chose. La trace serait donc partout et en tous lieux. Elle constituerait toute la réalité » (GALINON-MÉLÉNEC, 2020)¹⁷⁸.

Ces définitions sont à mettre en rapport avec les dictionnaires « Le Robert »¹⁷⁹ qui présentent des définitions de la trace sous différentes versions. S'appuyant sur le Robert intitulé « *Dictionnaire historique de la langue française* », dictionnaire particulièrement référencé quand il s'agit de recherches en Sciences humaines, Béatrice Galinon-Méléneq (GALINON-MÉLÉNEC, 2019)¹⁸⁰ en vient à proposer des définitions où elle positionne la notion de trace et ses déclinaisons au regard de celle de « *conséquence - du passé dans le présent et du présent dans le futur - et de sa définition de signe-trace* ». Les définitions de la notion de trace par Jeanneret et Galinon-Méléneq se rejoignent donc. Cependant, elles diffèrent quand il s'agit de situer l'indice par rapport à la trace. Cette nuance mérite d'autant plus notre attention que, dans le cadre médical, nous nous retrouvons davantage dans les définitions de Galinon-Méléneq.

¹⁷⁷ JEANNERET Y., « La complexité de la notion de trace », dans Galinon-Méléneq B. L'Homme trace. Perspectives anthropologiques des traces contemporaines, Cnrs éditions, Série *L'Homme-trace*, tome 1, 2011, page 59-86.

¹⁷⁸ GALINON-MÉLÉNEC B. *de la trace 1. Voyage au-delà des apparences*, London, ISTE, 2020 à paraître.

¹⁷⁹ Le Robert est une maison d'édition qui publie plusieurs dictionnaires spécialement dédiés à la langue française.

¹⁸⁰ GALINON-MÉLÉNEC B., « Pour une ichtologie générale à l'heure du numérique », in Liénard F et Zlitni S. (édit.) *Communication et traces numériques*, Presses universitaires Rouen Le Havre, 2019/1.

5.1.2 *Leur intérêt dans le cadre médical*

Les définitions proposées par Galinon-Méléneec découlent des axiomes qu'elle définit en introduction de ses articles et ouvrages¹⁸¹. Ils la situent dans une filiation théorique constructiviste où est posé que toute la réalité n'est pas accessible à l'humain.

Pour cette auteure « *la complexité du réel étant voilée l'humain ne peut qu'émettre des hypothèses à son sujet : il le fait à partir de sa dimension anthropologique d'Homme-trace et le résultat relève d'une interaction entre la « réalité-trace » (terme qui signifie que la réalité présente est la conséquence du passé) et chaque « homme-trace »* ».

Dans ce contexte, la représentation de la réalité par chaque individu est dite « *relationnelle* » (*Réalité-trace/Homme-trace*). Dans la mesure où chaque *homme-trace* a incorporé une histoire de vie (sociale, culturelle, affective, émotive, etc.) différente, cette relation est tout à fait spécifique.

L'épidémie de *coronavirus (Covid-19)* que vit l'humanité depuis décembre 2019 fournit la possibilité d'observer l'intérêt pragmatique de ces paradigmes :

- le réel voilé prend ici la figure d'une épidémie dont aucun calcul humain ou robotique n'a pu prédire l'évolution et pour laquelle les chercheurs du monde entier tâtonnent autant pour sa description et que pour sa prise en charge ;
- la *différence* individuelle s'exprime au travers des discours qui semblent exprimer autant de vérités que d'acteurs confrontés à cette réalité, comme en témoignent les extraits de presse dans l'encadré et la figure ci-dessous :

¹⁸¹ GALINON-MÉLÉNEEC (2011, 2013, 2015, 2017, 2019).

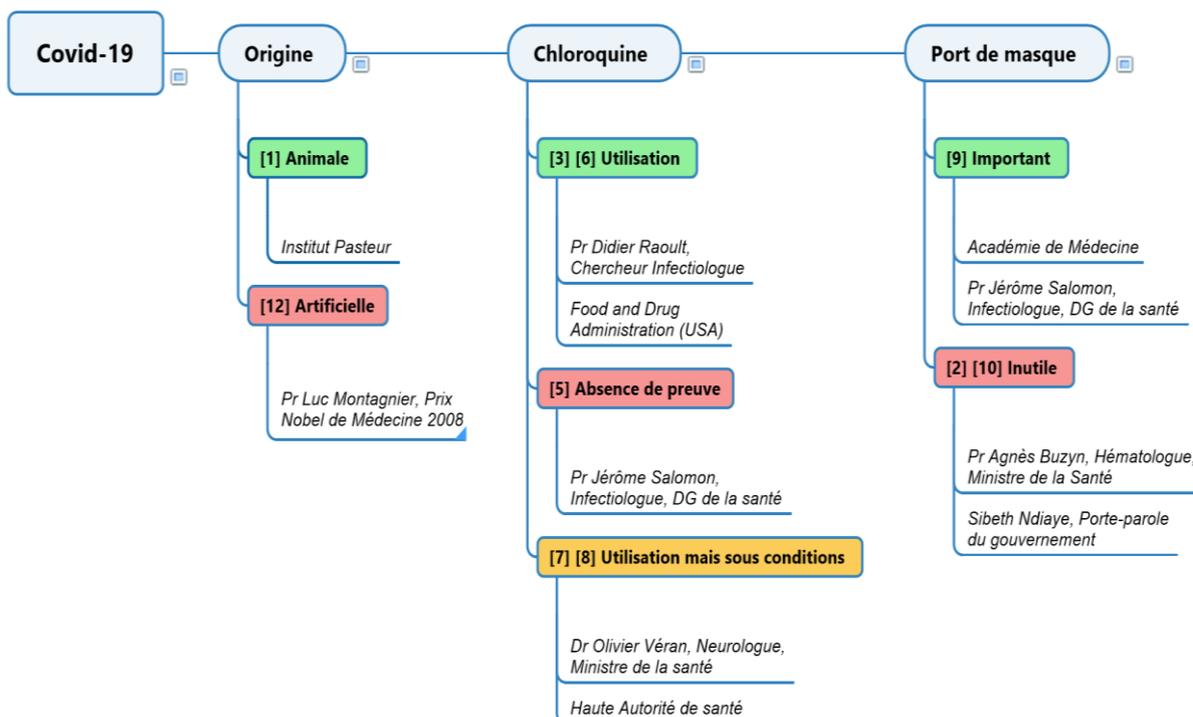


Figure 52: représentation thématique des discours des acteurs de la lutte contre la Covid-19 dans la presse

Encadré 4: coupures de presse chronologiques sur la maladie à coronavirus

[1] « Cette pneumonie est une maladie infectieuse causée par un virus appartenant à la famille des coronavirus, pour le moment identifié sous le nom de SARS-CoV-2. Le réservoir de virus est probablement animal. Même si le SARS-CoV-2 est très proche d'un virus détecté chez une chauve-souris, l'animal à l'origine de la transmission à l'homme n'a pas encore été identifié avec certitude. Plusieurs publications suggèrent que le pangolin, petit mammifère consommé dans le sud de la Chine, pourrait être impliqué comme hôte intermédiaire entre la chauve-souris et l'homme »¹⁸² Institut de Pasteur

[2] « Agnès Buzyn estime "totalement inutile" le port de masque pour les non-contaminés »¹⁸³ 26 janvier 2020

¹⁸² Source : <https://www.pasteur.fr/fr/centre-medical/fiches-maladies/maladie-covid-19-nouveau-coronavirus> consulté le 24 avril 2020.

¹⁸³ Source : <https://www.bfmtv.com/mediaplayer/video/coronavirus-agnes-buzyn-estime-totalement-inutile-le-port-de-masque-pour-les-non-contamines-1217820.html> consulté le 24 avril 2020.

[3] « Dans une vidéo intitulée « Coronavirus : fin de partie ! » mise en ligne mardi 25 février, l'infectiologue Didier Raoult fait état d'une « excellente nouvelle » sur le front de la lutte contre le coronavirus SARS-CoV-2 : la chloroquine, un antipaludéen ancien et peu onéreux, aurait apporté des « améliorations spectaculaires » chez des patients infectés. Le Covid-19 « est probablement l'infection respiratoire la plus facile à traiter » »¹⁸⁴ 27 février 2020

[4] « Je savais que la vague du tsunami était devant nous. » « Le 30 janvier, j'ai averti [le premier ministre] Edouard Philippe que les élections ne pourraient sans doute pas se tenir. » « On aurait dû tout arrêter, c'était une mascarade. »¹⁸⁵ 17 mars 2020

[5] « Mais "aujourd'hui, je n'ai aucune donnée suffisamment validée scientifiquement, médicalement pour tendre vers une recommandation (dans le cadre de l'utilisation de la chloroquine)", conclut Jérôme Salomon, espérant des résultats "dans moins de quinze jours" »¹⁸⁶.

[6] « Didier Raoult continue de défendre l'usage de la chloroquine pour traiter les malades du coronavirus. Il réfute tout d'abord la toxicité du produit : « Le monde devient fou ! », estime-t-il. Il rappelle notamment que la Food and Drug Administration (FDA), l'agence américaine des médicaments, « un gendarme extrêmement sévère » a donné son feu vert pour traiter les New-Yorkais »¹⁸⁷ 24 mars 2020

[7] « Olivier Véran autorise le traitement à la chloroquine sous conditions. Une étude clinique, baptisée Discovery, de grande ampleur (3200 personnes) vient d'être lancée pour tester l'efficacité de la chloroquine »¹⁸⁸ 24 mars 2020

¹⁸⁴ Un antipaludéen pourra-t-il contrer l'épidémie de Covid-19 ? publié dans le journal Le monde le 27 février 2020. Source https://www.lemonde.fr/sciences/article/2020/02/26/un-antipaludeen-pourra-t-il-arreter-l-epidemie-de-covid-19_6030931_1650684.html consulté le 24 avril 2020.

¹⁸⁵ Coronavirus : la confession d'Agnès Buzyn suscite une double polémique ; publié dans le journal Le monde le 18 mars 2020. Source : https://www.lemonde.fr/politique/article/2020/03/18/coronavirus-la-confession-d-agnes-buzyn-suscite-une-double-polemique_6033469_823448.html consulté le 24 avril 2020.

¹⁸⁶ Coronavirus : Didier Raoult teste tout le monde et traite à la chloroquine. Source : <https://www.ledauphine.com/france-monde/2020/03/22/coronavirus-didier-raoult-teste-tout-le-monde-et-traite-a-la-chloroquine> consulté le 24 avril 2020.

¹⁸⁷ Coronavirus : les mises au point de Didier Raoult sur le covid-19. Source : <https://gomet.net/raoult-mises-au-point-covid-19/> consulté le 24 avril 2020.

¹⁸⁸ Traitement à la chloroquine du covid-19 : feu vert du gouvernement sous conditions. Source : <https://gomet.net/chloroquine-traitement-ihu-marseille/> consulté le 24 avril 2020.

[8] « Pour l'heure, si la Haute Autorité de Santé a accepté d'administrer ce traitement (la chloroquine) à certains malades qui présentent des cas graves, c'est au titre que l'on appelle « compassionnel », c'est à dire sans assurance de l'efficacité du traitement mais à défaut de pouvoir en administrer un autre »¹⁸⁹. 24 mars 2020

[9] « Le port d'un masque "grand public" ou "alternatif" aux masques médicaux en pleine pénurie, devrait être rendu obligatoire pour les sorties pendant la période de confinement et lors de sa levée, a recommandé vendredi l'Académie de médecine »¹⁹⁰ 03 avril 2020.

[10] « Comment comprendre le changement de position des autorités françaises vis-à-vis du port du masque ? Alors que Sibeth Ndiaye, la porte-parole du gouvernement, affirmait le 25 mars que les citoyens "n'avaient pas besoin de masques" lorsque les gestes barrières étaient bien respectés, neuf jours plus tard, Jérôme Salomon, le directeur général de la santé, disait "encourager le grand public, s'il le souhaite, à porter des masques et en particulier des masques alternatifs (autres que chirurgicaux ou FFP2) »¹⁹¹ 10 avril 2020

[11] « Vraiment, ce virus est une vacherie. On est même en train de se poser la question de l'immunité. Si quelqu'un est guéri du Covid-19, est-il vraiment si protégé que cela ? On se demande si nous ne sommes pas en train de nous tromper complètement. Il y a une série d'éléments qui suggèrent qu'il y a des phénomènes de réactivation qui peuvent arriver. Donc, nous ne savons pas si le fait d'avoir des anticorps est un élément absolu de protection » a indiqué Jean-François Delfraissy. »¹⁹² 15 avril 2020.

[12] « Pour le Pr Luc Montagnier, prix Nobel de médecine en 2008 pour la découverte du VIH, SARS-CoV-2 serait un virus manipulé par les Chinois avec de l'ADN de VIH, sorti

¹⁸⁹ Chloroquine : le protocole Raoult. Source : <https://www.franceculture.fr/emissions/radiographies-du-coronavirus/chloroquine-le-protocole-raoult> consulté le 24 avril 2020.

¹⁹⁰ Covid: l'Académie de médecine pour le port obligatoire du masque, même « alternatif ». Source : https://www.sciencesetavenir.fr/sante/covid-l-academie-de-medecine-pour-le-port-obligatoire-du-masque-meme-alternatif_143179 consulté le 24 avril 2020..

¹⁹¹ Covid-19 : les données scientifiques en faveur du port du masque. Source : https://www.sciencesetavenir.fr/sante/contre-la-covid-19-le-port-du-masque_143401 consulté le 24 avril 2020.

¹⁹² Source : <https://rnc.bfmtv.com/emission/ce-virus-est-une-vacherie-est-on-vraiment-immunise-apres-avoir-ete-malade-du-coronavirus-1893975.html> consulté le 24 avril 2020.

*accidentellement d'un laboratoire chinois à la recherche d'un vaccin contre le SIDA »¹⁹³.
16 avril 2020.*

Sur le terrain, les divergences interprétatives ont créé une perte de confiance dans les capacités du corps médical à vaincre le virus avec des protocoles adaptés et des autorités à protéger la population. Cette défiance a plusieurs causes. Dans le cadre de cette thèse, nous posons l'hypothèse que l'une d'entre elles trouve sa source dans le fait que les patients ont pris l'habitude de consulter des médecins qui semblent appliquer - sans jamais émettre de doute à leur sujet - les représentations du réel catégorisées en maladies et protocoles de soins incorporées au cours de leurs études en médecine. Le médecin apparaît ainsi comme un expert qui maîtrise l'interprétation des signes du corps.

Pour Galinon-Méléneq (GALINON-MÉLÉNEC, 2011), comme pour nous, il y a là un biais interprétatif qui présente un risque pour la médecine : quand la médecine occidentale contemporaine applique la coupure entre le corps et l'esprit, découpe le corps en morceaux et considère le médecin en tant que cerveau se référant aux taxinomies apprises, on efface de la démarche clinique le rapport médecin-patient interprétée comme la rencontre de deux *corps-traces*. Cette découpe existe déjà du fait du grand nombre de spécialités médicales et chirurgicales.

La posture actuelle ouvre la voie à la possibilité de remplacer le médecin humain par un médecin-robot ou au développement sans nuance de la télémédecine. Ce choix peut s'avérer stratégique dans certains contextes comme, par exemple, pendant la crise sanitaire du *Covid-19* où l'une des pratiques pour éviter la propagation du virus a été la distanciation des corps. C'est ainsi qu'en France, en mars et avril 2020, les salles d'attentes des cabinets des médecins se sont vidées tandis que, en parallèle, la télémédecine effectuait un bon spectacle. Ce qui a permis de pallier certaines des difficultés liées au confinement de la population. Mais quelles « *conséquences-traces* » (GALINON-MÉLÉNEC, 2020) cette expérience de médecine à distance va-t-elle produire sur la formation des médecins ? Notre thèse repose sur l'hypothèse qu'elles vont faire moins appel à la mémoire humaine - largement moins performantes que celle fournie par l'IA - qu'à des questions qui relèvent plus des sciences humaines et de l'éthique.

¹⁹³ Source : <https://www.pourquoidoctor.fr/Articles/Question-d-actu/32184-EXCLUSIF-Pour-Pr-Montagnier-SARS-CoV-2-serait-virus-manipule-Chinois-l-ADN-de-VIH-podcast> consulté le 24 avril 2020.

5.1.3 *Déclinaisons des définitions*

La posture adoptée par la faculté de médecine étant à l'opposé de celle présentée par Galinon-Méléneec (GALINON-MÉLÉNEC, 2011) et de la nôtre, nous exposons ici les définitions proposées par cette auteure :

- *« la trace et le signe-trace : la découverte de la conséquence-trace est en fait une trace anthropocentrée car le réel ne parle pas : c'est l'humain qui le fait « parler ». La découverte de ce qu'il nomme une trace est le résultat de la rencontre entre la trace ontologique (la matière) et la trace incorporée (inscrite dans le corps) du vécu de l'interprétant. Cette rencontre de traces fait émerger ce qu'il nomme « le signe » (le réel interprété). C'est pour signifier ce processus que nous avons créé l'expression « signe-trace ».*
- *la marque¹⁹⁴ : un signe-trace associé à l'identification¹⁹⁵*
La marque est une catégorie de signes-traces, une conséquence visible d'un événement produit, volontairement ou non, reconnaissable par l'humain, identifiée de manière distinctive et signifiant¹⁹⁶ dans la représentation sociale. Exemple : la cicatrice laissée par l'impact d'une balle d'arme à feu.
- *l'indice¹⁹⁷ : un signe-trace représentatif d'une histoire envisagée sous la forme d'un scénario supposé probable¹⁹⁸. L'indice est une catégorie de signe-trace à laquelle un interprétant associe un lien direct lié à une inférence causale indexée à un scénario. Exemple : l'enquêteur qui tente de reconstituer ce qui s'est passé sur la scène d'un crime)¹⁹⁹.*

¹⁹⁴ D'après le dictionnaire Le Robert, le mot « marque » désigne dès sa première attestation un signe mis volontairement sur un objet pour le rendre reconnaissable, en indiquant la propriété.

¹⁹⁵ On pourra également se référer aux définitions proposées par Sylvie Leleu-Merviel (LELEU-MERVIEL, 2013) qui sont légèrement différentes. Notamment, pour elle, « la marque est l'inscription physique de la rétention indicielle appelée à faire trace ».

¹⁹⁶ A l'odeur du parfum, un « nez » identifie le n°5 de Chanel.

¹⁹⁷ D'après le dictionnaire Le Robert, le mot « indice » est initialement associé au sens latin de « dénonciation » ; aujourd'hui il évoque aussi « un signe apparent qui indique quelque chose avec probabilité, puis un fait connu qui sert à constituer une preuve par présomption. Il est associé également au terme index qui lui-même veut dire montrer, indiquer avec l'index (le doigt), le catalogue, la liste ».

¹⁹⁸ C'est ainsi que des objets patrimoniaux peuvent être considérés comme des indices du scénario d'origine ((DAVALLON, 2006), cité par JEANNERET Y., *La fabrique de la trace*, London, ISTE, 2018, p. 41.

¹⁹⁹ « Dans la mesure où nous considérons que tout signe est un signe-trace, ce signe même dans une singularité qui donne l'impression d'un lien direct reste le résultat d'une dynamique d'interactions de traces (conséquences d'évènements

- *l’empreinte*²⁰⁰ : un signe-trace associé à une idée de durée ou d’une altération de forme ou de matière. L’empreinte est une catégorie de traces (conséquence d’un événement) inscrites de façon suffisamment durable dans la forme et/ou la matière. Exemple : l’empreinte du vécu émotionnel de l’individu dans son enfance^{201,202} ».

A propos des « traces latentes », Galinon-Méléneq précise :

« la diversité des contextes conduit à des emplois variés qui permettent d’intégrer jusqu’à l’idée que des traces invisibles, non lisibles, ininterprétables, ne sont pas pour autant inexistantes. Elles sont parfois qualifiées de « traces latentes ». Elles sont dues à l’évolution de la trace et/ou liées aux limites de l’humain à les voir, les lire, les interpréter » (GALINON-MÉLÉNEC, 2020).

Et pour illustrer son propos, elle évoque l’actualité médicale qui conduit à faire appel à l’odorat canin pour détecter les traces olfactives conséquences d’un cancer existant mais non encore diagnostiqué par les humains (GALINON-MÉLÉNEC, 2020). Elle rappelle également que dans la mesure où elle a défini toute la réalité en tant que trace, elle retient l’idée que « la trace ne parle pas ». Quand la trace se signale à un humain - ou dans l’exemple antérieur au chien - ce n’est pas du fait de la chose elle-même. « C’est la résultante d’un processus d’interaction entre la matière observée et celui qui observe » (GALINON-MÉLÉNEC, 2015). Pour signifier ce processus, elle utilise la terminologie « signe-trace²⁰³ » (GALINON-MÉLÉNEC, 2011). Cette terminologie qui a pour objectif de souligner le passage de la trace existante en tant que conséquence à une trace repérée, dépend à la fois de l’ouverture à la

antérieurs) qui [...] met dans une dynamique strictement individuelle en activité interprétative le corps-trace de celui qui perçoit le signe-trace en tant qu’indice. Sa contextualisation dans un scénario indique sa relativité (un scénario différent produit un intérêt pour d’autres signes-traces). Le terme indice introduit donc la notion d’erreur possible dans l’interprétation de la trace (du signe-trace). Cette prise de conscience conduit à s’intéresser aux moindres détails, jusqu’aux signes-traces infinitésimaux » (GALINON-MÉLÉNEC, op.cit.).

²⁰⁰ D’après le dictionnaire Le Robert, le terme « empreinte » découle initialement du verbe *empreindre* issu (1213) du latin *impremere* « appuyer sur » qui signifie initialement (1250) « marquer par pression sur une surface ». LE ROBERT, *Le dictionnaire historique de la langue française*, 2006.

²⁰¹ Le qualificatif « empreinte » se rapporte plus généralement à la conséquence d’un processus quand celle-ci prend la forme d’une inscription plus ou moins durable de l’ordre du fixe. Exemple : les empreintes physiques des mains de stars inscrites durablement dans le béton d’Hollywood boulevard.

²⁰² C’est pourquoi les informaticiens nomment « empreintes » ce que d’autres nomment « traces numériques ». « Si on prend la définition générale de la trace en la spécialisant au cas du numérique, la définition correspondante serait : la trace numérique est constituée à partir d’empreintes numériques laissées volontairement (ou non ?) dans l’environnement informatique à l’occasion de processus informatique » (MILLE A., « Des traces à l’ère du Web » in *Intellectica*, 2013/1/n°59, pp. 8-9.

²⁰³ La terminologie « signe-trace » est construite sur une relation transductive entre le signe et la trace, ce qui signifie que non seulement la trace perçue par l’humain ne l’est que parce qu’elle fait signe mais aussi que le signe ne prend sens pour l’humain que parce qu’il fait écho à une trace.

réception et de sa vectorisation par l'interprétant. Selon ce principe, tout repérage de « *signe-trace* » doit aussi être questionné en fonction des traces - sensorielles, culturelles, cognitives, etc. - qui œuvrent chez l'interprétant.

Quand il s'agit d'une relation entre deux individus, la terminologie *signe-trace* évoque les deux faces de la trace : l'une tournée vers l'histoire, le passé et l'autre qui, tournée vers l'interaction présente, devient signal. Le comportement d'un individu est un *signe-trace* particulier qui donne lieu à jugement, à commentaire par autrui. Mais ce jugement est lui-même un *signe-trace* : de l'orientation de l'attention, d'un tri de quelques *signes-traces* parmi tous ceux qui constituent les comportements de l'Autre et des hypothèses sous-jacentes sur le processus qui les a produits (GALINON-MÉLÉNEC ET AL., 2015 ; MONSEIGNE, 2013).

Comme nous le verrons par la suite, ces définitions sont particulièrement appropriées dans le domaine de la consultation médicale et dans le diagnostic, les choix thérapeutiques et le suivi qui en résulte.

5.2 Déclinaisons de la notion de trace numérique

Les activités de l'Homme dans le cadre du numérique laissent des traces considérables, la plupart des temps à leur insu. Cette collecte de trace fait le plus souvent partie des processus de recherche de traçabilité des activités. Elle a connu une ampleur plus importante avec le développement du web sémantique. Le web sémantique a normalisé la manière de produire les traces numériques.

Dans le contexte professionnel informatique le terme « *empreinte* » est utilisé comme synonyme du terme « *trace* » notamment pour désigner les traces numériques. En quoi ces empreintes numériques nommées « *traces numériques* » par les informaticiens diffèrent-elles de la notion de trace vue sous l'angle de la conséquence précédemment évoquée ? En quoi ce choix terminologique joue-t-il dans l'analyse du processus de traçabilité ? Dans les choix qui sont opérés ? En devenant polysémique et naturalisée dans un sens commun, la notion de « *trace numérique* » ne se prive-t-elle pas des apports de la recherche ? Et de là, de sa dimension épistémologique et éthique ?

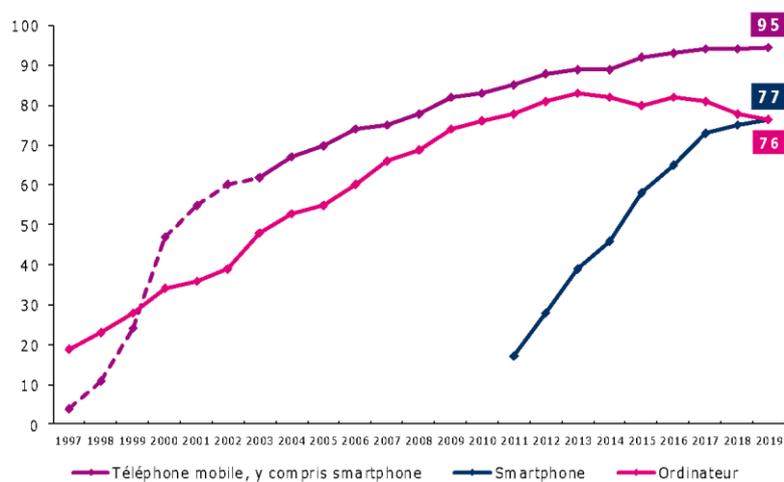
5.2.1 *Enjeu de l'usage statistique des traces numériques*

La traçabilité numérique est un fait social contemporain majeur. Le développement du web sémantique permet de suivre les activités des individus sur le web. Ce suivi des activités est devenu d'autant plus important avec la vulgarisation des téléphones portables et une meilleure accessibilité de la connexion Internet. Le développement de la technologie (portables, tablettes, ordinateurs, objets connectés) et l'amélioration de l'accès de l'Internet rendent difficile la distinction entre la vie *offline* et *online*. De nos jours, ces technologies et ces objets connectés sont utilisés dans le domaine de la santé pour le diagnostic, le traitement et le suivi des patients. Cette production des données s'est associée à un développement des techniques de collecte des traces numériques par les grands groupes de l'Internet et les gouvernements.

La diffusion de l'usage des téléphones portables et une accessibilité de la connexion Internet plus étendue ont permis la mise en place progressive d'une société où le numérique joue un rôle majeur. En France, le baromètre du numérique français en 2019 (Figure 53), étude réalisée par le CREDOC²⁰⁴, indiquait une progression d'un taux d'équipement du smartphone de deux points par rapport à 2017 (77%, +2 points) et dépassant celui de l'ordinateur (76%, -2 points) ; l'équipement en ordinateur s'inscrivant dans une période de déclin (76% en 2019, soit -7 points depuis 2012). Le CREDOC précisait :

« même si le téléphone mobile ou smartphone était déjà l'équipement le plus utilisé quotidiennement par les Français en 2017, cette tendance se poursuit et l'écart se creuse même avec l'utilisation de l'ordinateur : en 2019, 82% des Français indiquent utiliser quotidiennement leur téléphone ou smartphone (+3 points par rapport à 2017) contre 47% pour l'ordinateur (-5 points). Dans cette logique, le smartphone s'impose comme l'équipement privilégié pour se connecter à internet en 2019 (pour 51% des Français, +9 points, contre 31% pour l'ordinateur, -7 points) » (CREDOC, 2019).

²⁰⁴ Par le CREDOC pour le compte du Conseil Général de l'Economie, de l'Industrie, de l'Energie et des Technologies (CGE), l'Autorité de Régulation des Communications Electroniques et des Postes (ARCEP) et l'Agence du numérique CREDOC : Centre de recherche pour l'étude et l'observation des conditions de vie, est un organisme d'études et de recherche au service des acteurs de la vie économique et sociale.



Source : CREDOC, Enquêtes sur les « Conditions de vie et Aspirations ».

Figure 53 : statistiques sur l'utilisation du numérique en France (CREDOC, 2019)

Même si les statistiques montrent une croissance de l'utilisation d'Internet par les Français, il en reste 12% qui ne disposent d'aucun accès à Internet quel que soit le type d'appareil et de connexion (INSEE, 2019) ; cette partie invisible de la population, masquée et subissant les calculs moyennés des statistiques, est rendue visible grâce à cette épidémie de maladie à coronavirus²⁰⁵ où la majorité des activités sont réalisées au travers du numérique.

La question est de savoir si cette évolution s'accompagne de la capacité des individus à comprendre les technologies qu'ils utilisent et les risques qu'elles comportent en termes de traçabilité numérique. Des publications des chercheurs en sociologie (CASTELLS, 2001 ; FLICHY, 2001 ; JAURÉGUIBERRY, 2003 ; LAULAN, 2017 ; PROULX, 2004 ; RIEFFEL, 2014 ; WOLTON, 2009) et des chercheurs en SIC (MERZEAU, 2013b ; PERRIAULT, 2003 ; GALINON-MELENEC, ZLITNI, 2013) accompagnent ces évolutions depuis le début du XXIème siècle. Bien que ces chercheurs fassent l'objet d'invitations de plus en plus fréquentes sur les médias (c'est le cas, par exemple, de Dominique Wolton) et que les consignes en vue de la protection des données personnelles soient relayées par les services publics²⁰⁶, la compétence des utilisateurs reste encore insuffisante. En effet, selon une enquête de l'INSEE en 2019 (INSEE, 2019), 17% de la population adulte de plus de 15 ans n'avait aucune compétence numérique et 38 % des usagers d'Internet apparaissaient manquer d'au moins une compétence dans les quatre domaines de compétences (voir encadré ci-dessous) que sont la

²⁰⁵Le coronavirus risque d'accroître la fracture numérique du pays. Source : <https://www.publicsenat.fr/article/politique/le-coronavirus-risque-d-accroitre-la-fracture-numerique-du-pays-181533> consulté le 25 mars 2020.

²⁰⁶ Source : <https://www.cnil.fr/fr/maitriser-mes-donnees> consulté le 28 avril 2020..

recherche d'information, la communication, l'utilisation de logiciels et la résolution de problèmes.

C'est ainsi, par exemple, que, parmi les utilisateurs, beaucoup ignorent que la majorité des applications²⁰⁷ disponibles sur les smartphones - dont *WhatsApp*²⁰⁸ - ne respectaient pas les obligations du RGPD, alors que leurs usages ne cessent de progresser.

Encadré 5 : capacités et compétences numériques (INSEE, 2019)

Eurostat²⁰⁹ distingue quatre domaines de compétences numériques :

- la recherche d'information (sur des produits et services marchands ou administratifs, etc.) ;
- la communication (envoyer ou recevoir des courriels, etc.) ;
- la résolution de problèmes (accéder à son compte bancaire par Internet, copier des fichiers, etc.) ;
- l'usage de logiciels (traitement de texte, etc.).

Ces compétences sont mesurées à partir des déclarations sur le fait d'effectuer certaines tâches dans l'enquête annuelle auprès des ménages sur les technologies de l'information et de la communication, menée dans tous les pays de l'Union européenne.

Chaque compétence est notée 0 (compétence nulle), 1 (basique) ou 2 (compétence plus que basique). Le non-usage d'Internet au cours de l'année impliquant la note 0 : l'échelle mesure donc une capacité pratique (liée à la possession d'un équipement et à un usage même minimal d'Internet) si l'on considère la population générale, mais une compétence

²⁰⁷ Selon une étude réalisée par les chercheurs de l'Université d'Oxford, après avoir analysé le code de 959 000 applications sur les versions américaines et britanniques du *Google Play Store* des smartphones Android, 88,44% des applications tierces envoient des données à Alphabet, la maison-mère de *Google*. De même, 42,55% transfèrent les données à Facebook, 33,88% à Twitter, 26,27% à Verizon, 22,75% à Microsoft, et 17,91% à Amazon. Parmi les différentes informations partagées par les applications tierces, on compte l'âge, le genre, ou encore la position géographique de l'utilisateur. Ces données permettent ensuite de créer un profil détaillé de chaque individu. Ce profil peut notamment renseigner sur les habitudes d'achat, la classe socio-économique ou les opinions politiques de la personne. Source : <https://www.lebigdata.fr/android-donnees-google-facebook-oxford> consulté le 22 février 2019.

²⁰⁸ « *WhatsApp* envoie chaque contact du carnet d'adresses de ses utilisateurs sur ses serveurs, qui sont bien sûr situés aux États-Unis. Ceci principalement pour synchroniser les données et voir les utilisateurs qui ont déjà un compte, pour que les utilisateurs puissent facilement se connecter les uns aux autres. C'est peut-être très pratique pour vous et moi, mais cela signifie que les données de personnes qui n'ont ni envie ni l'intention d'utiliser l'application de messagerie se retrouveront sur *WhatsApp*. Seulement, en raison du RGPD, cette procédure est à présent interdite. Vous, en tant que propriétaire de votre smartphone, devriez demander l'autorisation à toutes les personnes de votre carnet d'adresses, avant que leurs données ne puissent être légalement transférées aux États-Unis ». Source : <https://blog.avira.com/fr/whatsapp-rgpd-illegale/> consulté le 28 avril 2020.

²⁰⁹ INSEE contribue activement aux travaux européens de Eurostat, qui constitue une des directions générales de la Commission européenne chargé de la conduite de la politique communautaire dans le domaine statistique. L'INSEE y participe directement, soit coordonne les positions des différents services statistiques ministériels.

si l'on se restreint aux usagers d'Internet. Elle sous-estime légèrement les compétences en « logiciels » et « résolution de problèmes » dont les critères ne nécessitent pas tous l'usage d'Internet.

Ces quatre domaines de compétences sont très liés, et en les sommant, on obtient un indicateur global de capacités numériques : une personne n'a ainsi aucune capacité numérique si elle obtient 0 dans chaque domaine (illettrisme) et des capacités plus que basiques si elle obtient 2 dans les quatre domaines. Entre les deux, Eurostat distingue les capacités faibles (au moins une compétence est notée à 0 et au moins une vaut 1) et basiques (aucune des compétences n'est égale à 0 et au moins une est égale à 1).

Cette tendance montre que la progression de l'utilisation des TIC ne signifie pas pour autant la connaissance des processus mis en jeu et la portée de chaque mise en relation et de chaque contact effectué via un écran. Or ces processus produisent des traces numériques qui sont utilisées pour des analyses statistiques et interprétées dans le cadre de prise de décision et de mise en place de projets par des organisations du secteur public ou marchand. L'enjeu invisible pour l'utilisateur est néanmoins majeur pour l'évolution de la société. Il se pose ainsi la question de la définition de la trace numérique, de son mode de production et de son interprétation. Lorsqu'elle est combinée à un domaine aussi sensible que la santé, elle requiert une grande prudence. Les choix qui sont effectués ont une dimension, non seulement éthique²¹⁰ mais aussi sociétale et politique. L'exemple le plus récent est fourni par les différences entre les pays en ce qui concerne la gestion de l'épidémie de coronavirus au printemps 2020. En avril 2020, au moment où se prépare la sortie du confinement en France - prévue le 11 mai - la Chine effectuait un *tracking* strict des déplacements des individus via une application développée par le groupe Chinois Alibaba²¹¹.

La figure 54 montre les différentes sources de données et les catégories de données collectées dans le cadre de la surveillance des personnes atteintes par la *Covid-19* et leurs contacts par l'intermédiaire d'une application en Corée du Sud. Les données sont ensuite croisées et

²¹⁰ MARTA SPRANZI, « Le rôle de l'éthique en médecine : de la décision singulière aux débats sociétaux ». Source : <https://www.youtube.com/watch?v=HU4XyGzpyN8> consulté le 30 avril 2020.

²¹¹ « Pour entrer dans cet immeuble en Chine, les employés doivent présenter un code QR présent dans leur téléphone. Sa couleur est cruciale : vert, l'accès est autorisé ; orange ou rouge, on ne rentre pas. C'est ainsi que les autorités chinoises traitent l'épidémie de Covid-19, en surveillant les individus et leur possible niveau de contagion. Développée par le géant chinois Alibaba, l'application analyse les données de l'utilisateur pour déterminer s'il a voyagé dans des zones à risque. L'application indique même si l'individu a voyagé à proximité de personnes infectées par le Covid-19 ». Source : https://www.francetvinfo.fr/sante/maladie/coronavirus/covid-19-en-chine-des-malades-surveilles-grace-a-numerique_3852249.html consulté le 30 avril 2020.

partagées avec les autorités sanitaires, municipales et le gouvernement local et par la suite divulguées au public. Cette collecte est autorisée par le *Contagious Disease Prevention and Control Act (CDPCA)* qui permet d’outrepasser certaines dispositions des lois sur la protection des données personnelles telles que le consentement des personnes (PARK ET AL., 2020).

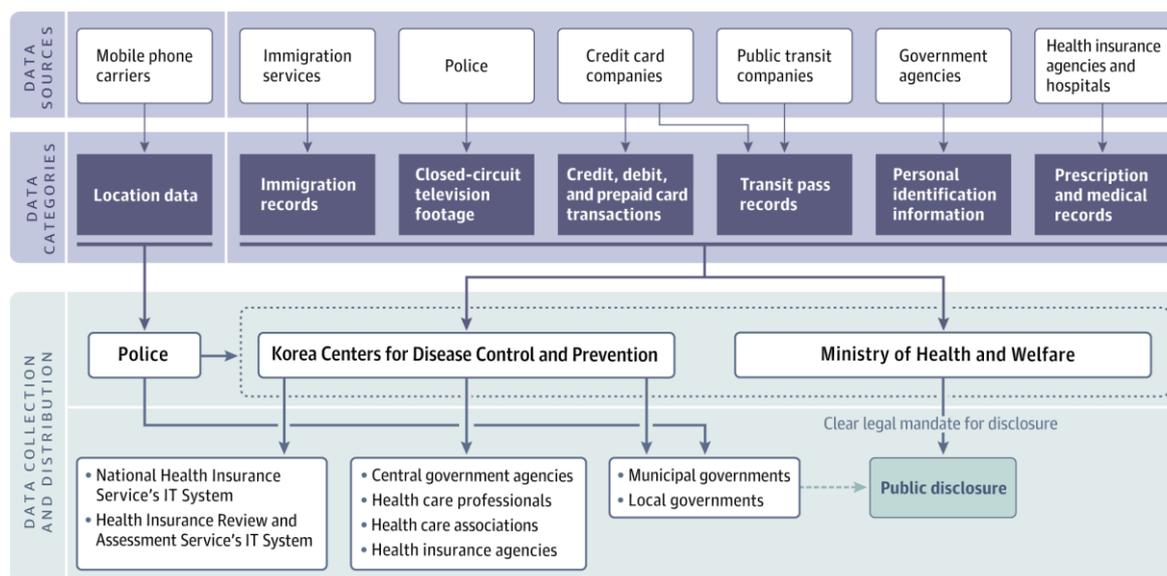


Figure 54: *tracking* des personnes contact du *Covid-19* en Corée du Sud : sources, catégories et distribution des données collectées (PARK ET AL., 2020)

La France considérant les « *conséquences-traces* » (GALINON-MÉLÉNEC, 2020) sur la démocratie pourrait avoir une telle procédure²¹², cherche une méthode de *tracking de contact* permettant de limiter les risques de propagation du virus, mais sans géolocalisation des individus, sans risque de stigmatisation ni mémorisation des données à long terme.

²¹² La mise en place d’un code d’une éventuelle application européenne de *tracking* est alors à l’étude pour assurer la protection des données personnelles. Source : <https://rapportsdeforce.fr/pouvoir-et-contre-pouvoir/tracking-quelle-efficacite-contre-le-covid-19-quel-danger-pour-la-vie-prive-04086659> consulté le 24 avril 2020.

5.2.2 *Nécessité d'une définition scientifique*

Comme nous l'avons vu, en informatique, « *La trace numérique est constituée à partir d'empreintes numériques laissées volontairement (ou non ?) dans l'environnement²¹³ informatique à l'occasion de processus informatique* » (MILLE, 2013a).

Les traces numériques peuvent désigner les informations qu'un dispositif numérique enregistre sur l'activité ou l'identité de ses utilisateurs au moyen de traceurs tels que les cookies²¹⁴, les smartphones ou les objets connectés, soit automatiquement, soit par le biais d'un dépôt intentionnel.

Moteurs de recherche, blogs, sites de réseaux sociaux, sites de e-commerce, mais aussi cartes à puce, titres de transport, téléphones mobiles : tous les systèmes qui requièrent une identification ou une interaction sont susceptibles de capter des données sur l'utilisateur - parcours, requêtes, préférences, achats, connexions, évaluations, coordonnées - qui seront stockées dans des bases de données distribuées à travers le monde entier.

La trace numérique prend forme sur un support matériel qui est dans ce cas la mémoire informatique selon des procédés bien définis avec une possibilité de conservation dans le temps et dans l'espace ; elle prend alors la forme d'une empreinte.

Avec le développement de l'informatique, toutes les activités et les interactions d'un individu peuvent être enregistrées sur des supports numériques de façon volontaires ou non (à son insu). Le développement des programmes et des systèmes se fait selon des normes et des standards de protocoles - le *web sémantique*²¹⁵ - utilisée pour la première fois en 1994 par Tim Berners-Lee (COLLOC, 2015) et qui a été validée par le *World Wide Web Consortium (W3C)*²¹⁶. Le *web sémantique* est un système qui permet aux machines de « *comprendre* »

²¹³ L'environnement est défini comme « *l'ensemble des éléments (biotiques ou abiotiques) qui entourent un individu ou une espèce et dont certains contribuent directement à subvenir à ses besoins* », ou encore comme « *l'ensemble des conditions naturelles (physiques, chimiques, biologiques) et culturelles (sociologiques) susceptibles d'agir sur les organismes vivants et les activités humaines* » Source : Dictionnaire Larousse.

En informatique, un environnement désigne, pour une application, l'ensemble des matériels et des logiciels système, dont le système d'exploitation, sur lesquels sont exécutés les programmes de l'application.

²¹⁴ Le cookie est l'équivalent d'un fichier texte de petite taille, stocké sur le terminal de l'internaute. Existant depuis les années 1990, ils permettent aux développeurs de sites web de conserver des données utilisateur afin de faciliter la navigation et de permettre certaines fonctionnalités.

²¹⁵ Le web sémantique est une toile de données qui peuvent être traitées directement et indirectement par des machines pour aider leurs utilisateurs à créer de nouvelles connaissances.

²¹⁶ Le *World Wide Web Consortium* est un organisme de standardisation à but non lucratif, fondé en octobre 1994 chargé de promouvoir la compatibilité des technologies du *World Wide Web*. Source : <https://www.w3.org/> consulté le 19 novembre 2017.

et de répondre aux demandes complexes de l'homme en fonction du sens de ses demandes. Une telle « *compréhension* » exige que les sources d'information pertinentes aient été sémantiquement structurées au préalable. Cependant, les machines ne peuvent pas accomplir toutes ces tâches sans direction humaine, parce que les pages web sont conçues pour être lues par des personnes et non par des machines. L'utilisation du web sémantique dans la conception des programmes et des applications permet une meilleure interopérabilité et une extraction plus facile des traces numériques.

La notion de traçabilité numérique est de plus en plus présente dans nos sociétés, cela est dû au contexte actuel du *big data*, où toute information est enregistrée et stockée par défaut.

5.2.3 *Ce que nous retenons pour notre pratique*

A l'issue de nos lectures de Jeanneret, Mille, Serres et Galinon-Méléneq (MILLE, 2013a ; SERRES, 2012 ; JEANNERET, 2011 ; GALINON-MELENEC, ZLITNI, 2013), nous proposons sept critères associables à la notion de trace numérique :

- *la matérialité* : contrairement aux idées courantes sur le « virtuel » et la dématérialisation, le numérique est un univers matériel extrêmement complexe, reposant sur de lourdes infrastructures technologiques. Et une trace numérique n'est visible, décodable et lisible que sur un écran ou un dispositif technique ;
- *l'inférence causale* : on considère l'élément observé comme résultant d'une relation cause - conséquence ;
- *la temporalité* : elle est au centre des approches considérant les traces numériques comme une suite d'événements dans le temps ;
- *la calculabilité* : elle apparaît comme la propriété particulière de la trace numérique. Elle est aussi au fondement du phénomène de la « *redocumentarisation du monde* » (SALAÜN, 2007). La finalité pour les géants du web est le recueil des traces des usagers avec des technologies toujours plus sophistiquées, de les intégrer dans de gigantesques bases de données, afin de pouvoir calculer et prédire les comportements ;
- *le différentiel interprétatif* : à l'instar de tout type de trace, cette interprétation se fait à partir des outils d'analyse. Elle ne prend pas en compte les autres aspects de la trace qui ne peuvent pas être numérisés - tels que les odeurs par exemple - en d'autres termes les aspects qualitatifs de la trace ; l'interprétation de la trace

numérique se base sur sa transformation en donnée quantifiable, qui perd contact avec le lieu, l'origine et le producteur de la trace.

- *la capacité mémorielle* : si la mémorisation, ou plus exactement la mise en mémoire, n'a pas attendu le numérique pour caractériser les traces, là encore le numérique a profondément changé la donne et bouleverse de fond en comble les problématiques de la mémoire et de l'archive.
- *le caractère historicisant* : il est au centre d'approches considérant les traces numériques comme une suite d'actions successives que l'on peut reconstituer ;
- *les capacités réflexives* : le potentiel de réflexivité sur l'action est incontestablement plus intense, plus complet avec les traces numériques. Toute interaction homme - machine donne lieu à une abondante production de traces numériques, pouvant ensuite se lire comme autant de signes réflexifs sur l'action.

Puisque toute trace - numérique ou non - est contextualisée dans un environnement bien précis, en ce qui concerne les traces numériques, il convient d'interroger ce que les informaticiens nomment « *l'environnement* ».

5.2.4 *Prise en compte de l'environnement/ contexte*

Les termes environnement, contexte, milieu ne portent pas la même signification selon les disciplines. Dans le contexte de notre thèse, nous retiendrons :

En informatique : Alain Mille (MILLE, 2013a) distingue deux formes d'environnement :

- l'environnement informatique qui est constitué par un ensemble d'artefacts informatiques capables de supporter les processus de mémoire et de calcul numérique, de communiquer entre eux par des canaux numériques et constituant des systèmes artéfactuels, finis et homogènes ;
- l'environnement informatisé qui est l'environnement défini dans l'acception habituelle du terme considéré comme instrumenté par des environnements informatiques.

En SIC : Pour Galinon-Méléneq (GALINON-MÉLÉNEC, 2011) avant toute interprétation de la trace - numérique ou non -, il faut que l'individu qui doit l'interpréter ait une connaissance du contexte et des circonstances ayant abouti à la trace. Ainsi, le médecin effectue-t-il les

choix thérapeutiques en prenant en compte non seulement le patient pris individuellement mais aussi son contexte de vie.

Or, « *la trace se repère et s'interprète en fonction des processus d'interactions entre un individu en situation d'interprétation et les environnements (individuels, familiaux, culturels, sociaux et plus généralement humains et non humains) dans lesquels elle s'insère* » (GALINON-MÉLÉNEC, 2011).

En médecine, quand les pathologies engagent le pronostic vital, les praticiens sortent de l'interprétation individuelle pour confronter les points de vue. Nous y voyons l'illustration du fait que dans les cas les plus délicats, les médecins intègrent - sans en connaître le substrat théorique en anthropologie de la communication - l'idée que chacun d'entre eux est un *homme-trace* inscrit dans sa différence et qu'il convient de partager les représentations de la situation dans laquelle le patient se trouve.

5.2.5 *Biais interprétatifs nés de la méconnaissance de l'environnement*

Les chercheurs auxquels nous nous référons considèrent que les traces de l'interaction entre l'utilisateur informatique et l'environnement s'insèrent dès la collecte et le stockage des données dans l'empreinte matérielle existant dans les mémoires des ordinateurs. Ils prennent en compte a) l'expérience, support premier de construction de la connaissance individuelle (GALINON-MÉLÉNEC, ZLITNI, 2013 ; MILLE, 2006 ; 2013a) ; b) une formalisation structurée de l'expérience sans laquelle les données ne peuvent devenir sources de production de connaissances numérisées (MILLE, 2013a ; 2006 ; LELEU-MERVIEL ET AL., 2018) et c) la prise en compte de la façon dont l'activité et les interactions de l'utilisateur dans l'environnement témoignent de cette expérience (GALINON-MÉLÉNEC, 2020 ; MILLE, 2006 ; CORDIER ET AL., 2013).

Le passage de l'expérience à la capacité de formalisation de cette expérience n'est pas automatique ; il suppose une observation intentionnelle et spécifique et un observateur capable d'attribuer une signification (MILLE, 2006). Or, cette attribution quand elle est immédiate va dépendre de chaque « *homme-trace* », de son histoire de vie ; elle en est le « *signe-trace* » (GALINON-MÉLÉNEC, 2011). L'informaticien ayant conscience qu'il est un « *homme-trace* » comme un autre, avec un « *corps-trace* » inscrit dans la différence (GALINON-MÉLÉNEC, 2020), et a un *punctum* (LELEU-MERVIEL, 2013) qui produit des biais adopte donc une formalisation, des outils et des méthodes appropriés en vue de les canaliser.

Les déliaisons (MERZEAU, 2013a) effectuées lors de la sélection des *captas*²¹⁷ (GALINON-MÉLÉNEC, 2020 ; LELEU-MERVIEL, 2017) et des transformations en *data* doivent être repérées. Tout au long des processus, il faut garder à l'esprit que ces « *données* » sont donc construites et non « *données* » comme le mot le laisserait penser (GALINON-MÉLÉNEC, 2020) : elles sont agrégées à d'autres données, stockées dans des réservoirs distants, disséminées dans quantité de sites, et font l'objet de calculs qui produisent à leur tour d'autres données (MERZEAU, 2013a).

Quand il s'agit d'un environnement médical, les déliaisons - expérience, *capta*, *data*, calculs moyennés, interprétations successives (par l'humain et par la machine) - non maîtrisées conduisent à des risques interprétatifs. En effet, il n'est pas évident de décrire le processus de raisonnement médical ayant abouti au diagnostic. Par conséquent, une inférence du diagnostic à partir des symptômes et résultats notés sur la mémoire de l'ordinateur est source d'erreur. Nous voyons donc ici la limite d'une sémiologie médicale où le signe²¹⁸, considéré *ceteris paribus* coupé de la dynamique de « *signe-trace* », est pris comme *capta*, transformé en *data* et fait ensuite l'objet de regroupements en vue de produire les « signes cliniques » d'une maladie.

Quand les tris des données médicales sont réalisés à partir d'algorithmes préalablement établis par des personnes travaillant pour des organisations - publiques ou privées - dont le calcul est avant tout basé sur le rapport coût/bénéfice, le risque pour la santé du patient s'accroît. Les traces numériques subissent alors une « modélisation de traitement » effectuée au prorata du poids économique.

Ainsi, si toute interprétation de traces numériques relative à une personne pose la question de l'adéquation avec la réalité du vécu des individus (GALINON-MÉLÉNEC, 2020 ; MERZEAU, 2013a), devons-nous dans cette thèse nous poser la question de savoir si l'utilisation des traces numériques et des traces dans le domaine de la santé échappe aux propriétés précédemment décrites ou s'il existe des particularités ? Nous ferons cette analyse dans la partie consacrée à l'utilisation des traces dans le domaine de la santé (confère la section 5.6).

²¹⁷ « *Un captum est un fragment de saisie aspectuelle. Les captas sont brutes au sens où les saisies sont non qualifiées. Ce sont des infra-data : leur conversion dans une moule sémantique et syntaxique préformé les traduit en data* ». LELEU-MERVIEL S. « *La traque informationnelle* ». Londres, ISTE, 2017. p. 123.

²¹⁸ « *Le signe est la partie visible de la trace. Tout signe est signe-trace, toute trace n'est pas signe* » (GALINON-MÉLÉNEC, 2011).

5.3 De la donnée « *data* »

La « donnée » (*data*) est également un terme polysémique dont la définition varie en fonction des domaines et des usages.

Jean Louis Ermine²¹⁹ enseigne que « *la donnée peut être définie comme la perception de la réalité par les sens (qui peut être étendue par des observations faites par des machines). Les données sont ainsi le résultat d'un processus de perception, à travers un système de signes* » (ERMINE ET AL., 2012).

Nous comparerons ci-après cette définition à partir des trois disciplines de notre thèse : l'informatique, la médecine et les SIC. Cependant, en référence au rapport du CNE (Comité National d'Évaluation) sur les SIC (CNE, 1993)²²⁰ et à l'historique la revue *Communication & Organisation*²²¹ en SIC, nous ferons appel, de point en point, aux sciences de gestion.

5.3.1 En informatique

Selon Le Petit Robert, la donnée est une « *représentation conventionnelle d'une information (fait, notion, ordre d'exécution) sous une forme (analogique ou digitale) permettant d'en faire le traitement automatique* ».

Si l'on suit l'informaticien Serge Abiteboul : « *une donnée est une description élémentaire, typiquement numérique pour nous, d'une réalité* » (ABITEBOUL, 2013).

Le qualificatif « typiquement numérique » fait référence au *bit*, une variable qui peut prendre la valeur 0 ou 1. La donnée est alors représentée par une séquence de *bits*. Par exemple : *100110010000*.

Dans ce cadre, on peut distinguer la donnée en fonction de :

²¹⁹ Jean-Louis Ermine est professeur émérite à l'Institut Mines-Télécom (École de management) et expert à l'AIEA. Il est délégué auprès de la commission ISO relative au *Knowledge Management*. Il a fait ses recherches dans le domaine des mathématiques, informatique et management.

²²⁰ SIRINELLI J (DIR.), BRUNET J (DIR.), MASSI-FOLLEA F, CARPENTIER JB, GALINON-MÉLÉNEC B., JOLY J., MARTINET B., MEYRIAT J., MISSIKA JL., SALOMON JC., TETU JF. Réunions de travail au CNE et enquêtes suivies du rapport « *Etat des lieux des Sciences de l'Information et de La communication* », Rapport du Comité National d'Évaluation au Président de la République, 1993,127 pages.

²²¹ CARAYOL V., GRAMACCIA G., (DIR.). *Abécédaire des 20 ans de la revue Communication & Organisation*, Presses Universitaires de Bordeaux, 2013, 241 pages.

- son caractère structuré ou non ;
- sa place dans la hiérarchie : on peut trouver des données élémentaires bits, propriétés d'entités (objet (informatique) ou relations) et des métadonnées.

La structuration des données joue un rôle clé dans la sécurité informatique.

5.3.2 *En SIC*

Le *Publictionnaire* des SIC²²² distingue la trace numérique de la donnée « *par le fait qu'elle n'est pas préformatée et qu'elle nécessite d'être mise en relation avec d'autres attributs pour acquérir un sens* » (BOULLIER, 2015). Ces attributs correspondraient aux qualificateurs de Mioara Mugur-Schächter (LELEU-MERVIEL, 2008) :

« les qualificateurs sont les moyens d'observation et de mesure, biologiques ou instrumentaux. Il n'y a qu'une qualification par mesure et celle-ci n'est pas répétable car généralement l'entité-objet a changé. Mais la multiplication des qualifications donne des vues-aspects proposant des grilles de qualifications effectives et intersubjectives. L'opération peut conduire à la constatation de l'inexistence relative de l'entité-objet créée aux fins d'observation » (MUGUR-SCHÄCHTER, 2006)²²³.

L'entité-objet est générée par un mécanisme permettant au fonctionnement de la conscience, dans le cadre de ses stratégies téléonomiques, de créer quelque chose (un observable) à partir de quoi il pourra procéder à des mesures. Nous procédons de cette façon en permanence dans la vie courante. Nous construisons des « *objets* » d'étude, qui n'existaient pas avant notre intervention²²⁴. Ainsi, la donnée apparaît comme un fait objectif a nécessité l'action d'un agent cognitif²²⁵. Cela induit un questionnement épistémologique qui vise à préciser « *qui*

²²² *Publictionnaire*, notice « Traces numériques ». Source : <http://publictionnaire.huma-num.fr/notice/traces-numeriques/> consulté le 11 mars 2018.

²²³ MUGUR-SCHÄCHTER, Mioara, *Sur le tissage des connaissances* [en ligne]. Paris, France, Hermès, 2006.

²²⁴ Présentation et commentaire du livre « Sur le tissage des connaissances » de Mioara Mugur-Schächter par Jean-Paul Baquiast : <http://philoscience.over-blog.com/article-3410988.html> consulté le 20 mars 2018.

²²⁵ Nous avons préféré le terme d'« agent cognitif » car notre agent peut être un individu ou un algorithme de traitement de données dans le cadre des traces numériques. Le terme cognitif n'est pas utilisé dans le sens propre, mais surtout pour mettre en évidence la manipulation des traces.

produit ces traces numériques et les enregistre [...], qui les traite et comment, pour leur donner du sens » (BOULLIER, 2015).

Ainsi, dans le cadre de la production de la trace numérique, il faut distinguer deux processus :

- le *front-end* qui correspond à l'interface ou à la partie visible par l'utilisateur final. Il interagit dans cette interface par des modalités préfinies par le concepteur du programme ou du système. Dans certains cas - rare avec le développement du web - l'utilisateur ne peut avoir une interaction directe avec le système modifiant son comportement. L'utilisateur ne connaît pas forcément le processus mis en arrière-plan pour le permettre d'avoir une interaction ;
- le *back-end* correspond à l'arrière-boutique où toutes les productions et les interactions possibles sont imaginées par les concepteurs du programme ou du système pour le présenter à l'utilisateur. L'utilisateur ne peut pas y intervenir pour apporter des modifications à ce niveau. C'est dans cette arrière-boutique invisible pour l'utilisateur que se développent toutes les procédures rationnelles de traitement de données, de chaînage d'opérations nécessaire à la transformation des inscriptions en traces (GALINON-MELENEC, ZLITNI, 2013).

Ainsi, dans le cas de la production de la trace numérique nous distinguons deux acteurs :

- l'utilisateur : il interagit avec le programme ou le système et laisse des traces numériques de façon volontaire ou involontaire dans l'environnement informatique comme le définit Alain Mille (MILLE, 2013a). L'utilisateur n'intervient pas dans le choix des traces (au sens commun) à collecter (*les capta*) et ne sait pas à l'avance celles qui le seront, ni comment elles seront mises en relation pour les contextualiser ou mises en relation pour acquérir un sens ;
- Le concepteur²²⁶ à l'origine du programme ou du système qui utilise le web sémantique pour la conception, le mode de stockage et la définition des traces à collecter par le programme ou le système. Dans ce cas, la collecte des traces est volontaire et prévue par le concepteur. Les différentes traces à collecter sont définies à l'avance par un groupe ou un individu avant sa collecte. Par ailleurs, il existe des opérations qui se produisent dans l'environnement informatique durant

²²⁶ Le concepteur peut être celui qui commande - maître d'ouvrage - le programme ou le système en définissant ses besoins ou celui qui réalise la conception technique - maître d'œuvre - ou les deux à la fois.

la phase de collecte des traces que le concepteur ne connaît pas ou ne maîtrise pas.

5.3.3 Médecine et règles de droit

La définition de la « donnée de santé » n'existe pas dans le droit français ; mais elle est considérée comme une donnée sensible²²⁷. Jusqu'à la proposition de la définition par l'Union Européenne²²⁸, seules les informations sur l'état de santé issues du milieu médical avaient valeur de « données de santé ». Il était ainsi difficile de faire une différence nette entre les données de « bien-être » et de santé. Il fallait donc se reporter aux jurisprudences françaises et européennes pour déterminer ce qui relevait du champ des données de santé.

L'Union Européenne à travers le Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données définit :

« Les données à caractère personnel concernant la santé devraient comprendre l'ensemble des données se rapportant à l'état de santé d'une personne concernée qui révèlent des informations sur l'état de santé physique ou mentale passé, présent ou futur de la personne concernée. Cela comprend des informations sur la personne physique collectées lors de l'inscription de cette personne physique en vue de bénéficier de services de soins de santé ou lors de la prestation de ces services au sens de la directive 2011/24/UE du Parlement européen et du Conseil au bénéfice de cette personne physique ; un numéro, un symbole ou un élément spécifique attribué à une personne physique pour l'identifier de manière unique à des fins de santé ; des informations obtenues lors du test ou de l'examen d'une partie du corps ou d'une substance corporelle, y compris à partir de données génétiques et d'échantillons biologiques ; et toute information concernant, par exemple, une maladie, un handicap, un risque de maladie, les antécédents médicaux, un traitement clinique ou l'état physiologique ou biomédical de la personne concernée, indépendamment de sa source, qu'elle provienne par exemple d'un médecin ou d'un autre professionnel de la santé,

²²⁷ Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés.

²²⁸ Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE.

d'un hôpital, d'un dispositif médical ou d'un test de diagnostic in vitro »

Considérant 35 du RGPD.

Cette nouvelle définition des données de santé par l'Union Européenne prend en compte toutes les données produites par les objets connectés reconnus comme dispositif médical. Elle souligne également l'absence d'ancrage temporel des données de santé, dans la mesure où elles peuvent se rapporter à l'état de santé passé, présent ou futur d'une personne.

Les données de santé proviennent de base de données et principalement de trois sources différentes :

- les données issues du « dossier patient » : ce sont celles détenues par les médecins, les laboratoires, les pharmacies, les hôpitaux, les acteurs de la recherche et l'administration (assurance maladie, système national des données santé (SNDS)) ;
- les données issues des objets connectés, des commentaires sur les réseaux sociaux : ce sont les commentaires individuels ou les pages d'association de patients effectués dans l'environnement informatique ;
- les données issues du séquençage génomique humain et de l'épigénétique²²⁹.

Dans ce lot de données, il existe deux types de données :

- les données qui sont capables d'être anonymisées (données des dossiers patients, des assurances et du SNDS) ;
- les données qui sont impossibles d'être anonymisées telles que les données génomiques ; ou du moins quand elles sont anonymisées perdent toutes leurs valeurs dans le cadre de la recherche.

Deux questions qui restent en suspens pour nous sont : qu'en est-il : a) des données provenant des objets connectés non reconnus comme dispositif médical ? b) des données provenant des patients eux-mêmes par l'intermédiaire des réseaux sociaux ?

²²⁹ L'épigénétique correspond à l'étude des changements dans l'activité des gènes, n'impliquant pas de modification de la séquence d'ADN et pouvant être transmis lors des divisions cellulaires. Contrairement aux mutations qui affectent la séquence d'ADN, les modifications épigénétiques sont réversibles. Source : <https://www.inserm.fr/information-en-sante/dossiers-information/epigenetique> consulté le 21 mai 2020.

La loi prévoit de ne pas prendre en compte seulement le caractère intrinsèque des données. Les finalités de traitement des données doivent être prises en compte, seule ou en association avec d'autres données. Cela permet de vérifier et de déterminer la capacité de faire des projections sur la santé future de l'individu. Mais une des particularités des données de santé provient du fait que les données correspondent à un épisode précis, à un moment t de l'état de santé et qu'elles ne sont donc pas forcément cumulables pour être moyennées ou pour faire des classifications. Ainsi, quand des données de santé sont fournies dans le cadre d'une maladie, elles correspondent à un état de crise de l'organisme et sont donc très différentes de celles produites avant la maladie. La non prise en compte de cette variation du contexte dans la captation des données aboutit à leur mauvaise utilisation dans le cadre de la production des informations et des connaissances. Les décisions qui sont prises à partir de résultats ainsi biaisés sont susceptibles d'entraîner des conséquences néfastes pour les individus et la société.

La figure ci-dessous montre l'organisation des données de santé lors d'un épisode de maladie.

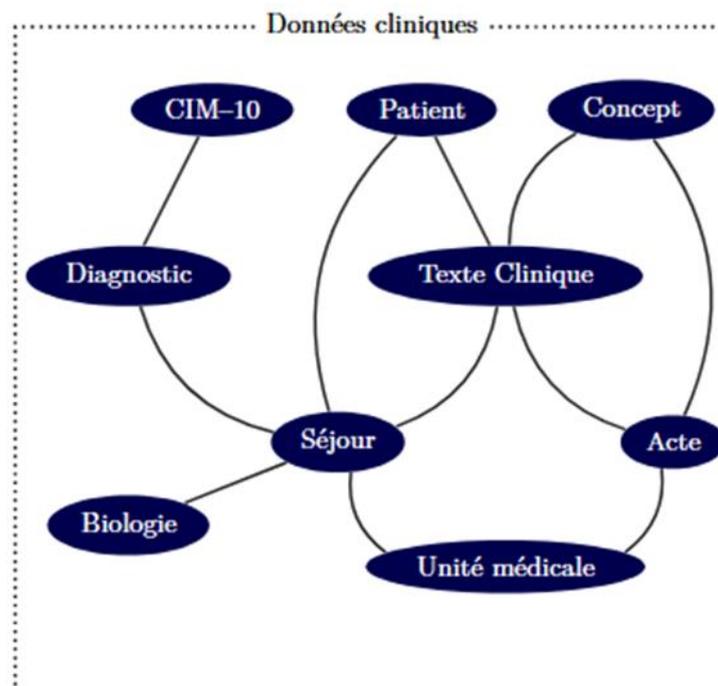


Figure 55 : structuration des données de santé lors d'un épisode de maladie (LELONG, 2019)

5.4 De l'information et de la connaissance

5.4.1 *Information*

Nous avons précédemment exploré les différents usages des termes en fonction des disciplines. Reprenons ici quelques-unes de ces définitions pour mieux situer notre positionnement.

Pour Gilles Balmissse, spécialiste des systèmes d'information et de l'application des outils et applications numérique dans le *knowledge management* (BALMISSSE, 2015), une information est une donnée à laquelle un sens et une interprétation ont été donnés. Une donnée devient une information lorsqu'elle est contextualisée. Le plus souvent l'information naît de la combinaison de plusieurs données entre elles, aboutissant à un fait.

Pour l'économiste Français, Jean Louis Monino, l'information correspond à des données qui ont été consignées, classées, organisées, raccordées et interprétées dans le cadre d'une étude spécifique (MONINO, 2013).

Pour l'économiste Japonais, Ikujiro Nonaka (NONAKA, 1994), l'information peut être vue selon deux perspectives : syntaxique (volume d'information) et sémantique (signification de l'information). La perspective syntaxique relève de la théorie de Shannon, mais la perspective sémantique est plus importante pour la création de connaissances, car elle se focalise sur le transport de sens.

Pour la suite, nous retiendrons particulièrement les travaux de Nonaka car son analyse sémantique de l'information prend toute sa valeur dans le cadre du traitement des données issues de plusieurs sources notamment dans le cas du *big data*. Une analyse hors du contexte et de l'environnement des données pourrait aboutir à une mauvaise interprétation des données avec des conséquences importantes. Tout l'enjeu dans le contexte de la pratique médicale est d'avoir des données de qualité prenant en compte le contexte et le producteur de données tout en protégeant l'intimité et la vie privée. Ce n'est qu'une fois cette étape franchit, que les informations tirées à partir des données pourront évoluer vers la production de connaissance très utile.

5.4.2 Connaissance

La question de la connaissance a été traitée depuis des années par les philosophes (Platon, Aristote, Descartes, Kant, Hume) et continue de l'être de nos jours. Les questions philosophiques de la connaissance portent sur sa nature, ses sources et ses limites.

La définition la plus courante de la connaissance est une « *croyance vraie et justifiée* » (JTB : « *Justified True Belief* », (CHISHOLM, 1982). Ceci signifie : « *je connais quelque chose si je le crois, si j'ai une preuve que c'est vrai, et si c'est vrai* » (ERMINE ET AL., 2012). Cette définition de la connaissance est celle défendue par Platon. Selon cet auteur, « *la connaissance est l'ensemble des conditions auxquelles la formule verbale de la pensée raisonnée - raisonnement qui épouse la confrontation du réel – pourra devenir un instrument sûr de la connaissance vraie* » (DAUDIN, 1951). Cette forme de connaissance ne prend pas en compte l'environnement. Et pour Aristote, disciple de Platon, l'environnement ne peut être ignoré dans la constitution de la connaissance et de la science (DAUDIN, 1951)

Pour Mickaël Polanyi, chimiste, philosophe et économiste, toutes les connaissances ne peuvent pas être décrites ou explicitées : « *we can know than we can tell* » (POLANYI, SEN, 2009). Il distingue deux types de connaissances :

- la connaissance tacite que l'on peut exprimer aisément ou décrire «*tacit knowledge that cannot be easily formalized and put into exact words*» (POLANYI, SEN, 2009);
- la connaissance explicite que l'on peut transmettre ou décrire aisément. Elle correspond aux connaissances scientifiques.

La connaissance tacite est pour cet auteur la base de la connaissance, le moteur de l'innovation. C'est cette connaissance qui permet au chercheur de réfléchir à un problème et de faire des hypothèses de recherche pour ensuite les valider lors des expérimentations : « *Tacit knowing is shown to account (1) for a valid knowledge of a problem, (2) for the scientist's capacity to pursue it, guided by his sense of approaching its solution, and (3) for a valid anticipation of the yet indeterminate implications of the discovery arrived at in the end*» (POLANYI, SEN, 2009). Les deux types de connaissance sont essentielles pour la création de la connaissance. La connaissance explicite qui exclut la connaissance tacite perd rapidement de son sens. La connaissance est créée par les interactions entre les connaissances tacites et les connaissances explicites et non pas de la connaissance tacite seule ou de la connaissance explicite seule (POLANYI, SEN, 2009). Cette théorie de la connaissance correspond bien à notre position épistémologique dans le cadre de notre travail

de thèse. En effet, la création de la connaissance dépend de l'environnement, des inspirations, du sens de l'observation, des interactions et des intuitions des individus.

Ermine (ERMINE, 2000) définit « *la connaissance comme de l'information (un ensemble de messages produits par un système), prenant un sens spécifique dans un contexte spécifique* ». Dans cette définition, il considère que la connaissance a trois composantes imbriquées : information, sens et contexte (LE BLANC, ERMINE, 2006). L'information est régie par la théorie de Shannon, le sens est régi par la théorie sémiotique, et le contexte est régi par la théorie des graphes connexes. Ces trois composants permettent de délimiter un ensemble de sciences et de méthodes qui relèvent de l'étude d'un système de connaissances. Chacune d'elles s'organise plus particulièrement autour d'un axe donné, de sorte à couvrir, en une sorte de continuum, le cercle du microscope (figure 56).

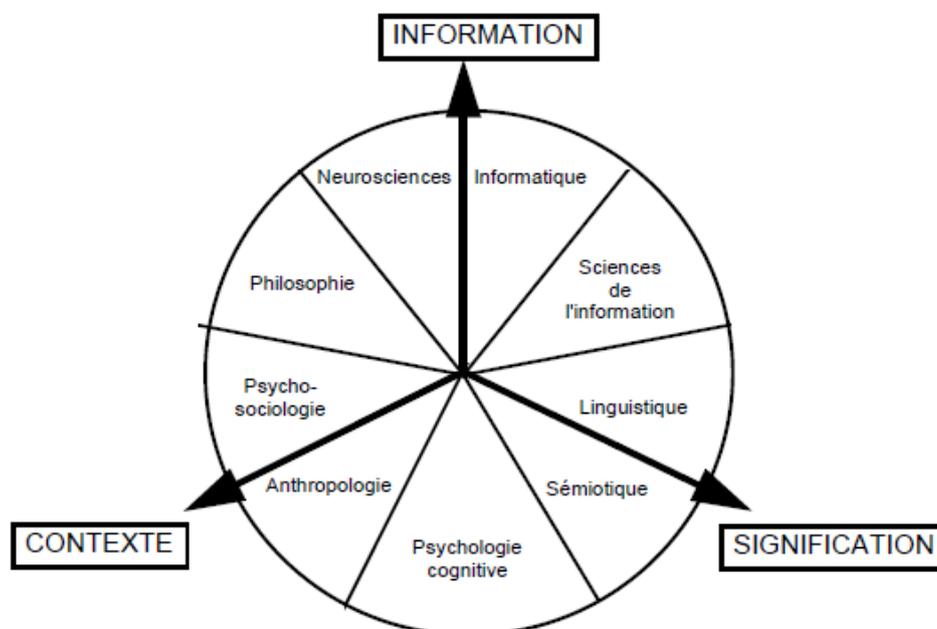


Figure 56 : épistémologie possible de la connaissance (ERMINE, 2000)

Pour Minsky Lee Minsky, spécialiste en sciences cognitives et en IA, « *la connaissance est constituée de données brutes perçues et mémorisées, mais aussi du raisonnement utilisé pour établir des relations entre elles, pour construire des schémas qui seront comparés à d'autres, analogues, déjà rencontrés* » (MINSKY, 1997 ; 1988).

La connaissance est ainsi le résultat d'une réflexion sur les informations analysées en se basant sur : ses expériences, ses idées, ses valeurs, les avis d'autres personnes consultées pour l'occasion ; sa propre expertise et celle de ses pairs.

Cette définition prend en compte la connaissance de l'environnement par l'individu, ses compétences et surtout l'interprétation que se fait l'individu en fonction de sa compréhension de l'information. La connaissance permet de confronter les informations au contexte afin d'avoir une meilleure connaissance et une interprétation élargie des phénomènes mis en lumière par ces informations.

Pour Nonaka (NONAKA ET AL., 2000), « *la connaissance est un processus dynamique créé à travers une interaction sociale entre individus et organisations. La connaissance est spécifique à un contexte* ».

La connaissance naît de la compréhension et de l'assimilation des règles qui régissent les modèles ou les schémas mentaux sous-jacents à ces relations, permettant ainsi de comprendre comment la situation évoluera si les données se modifient. La connaissance permet d'aboutir à une action. Mais cette connaissance ne peut être considérée comme une vérité universelle et indiscutable. Elle est fortement dépendante de l'individu qui la porte, empreinte de ses croyances et de son système de valeur (PAQUET, 2006 ; BLUMENTRITT, JOHNSTON, 1999). La notion d'interprétation est très importante dans la production de connaissance.

Ainsi, Nonaka (NONAKA, 1994) définit la création de connaissance comme « *un processus humain dynamique de justification de croyances personnelles tendant vers la vérité* ».

A partir des travaux des différents auteurs (ERMINE, 2000 ; NONAKA ET AL., 2000 ; NONAKA, 1994 ; MILLE, 2006 ; GALINON-MÉLÉNEC, 2011 ; LELEU-MERVIEL, 2017), nous proposons la figure ci-dessus pour une représentation du processus de production de la connaissance.

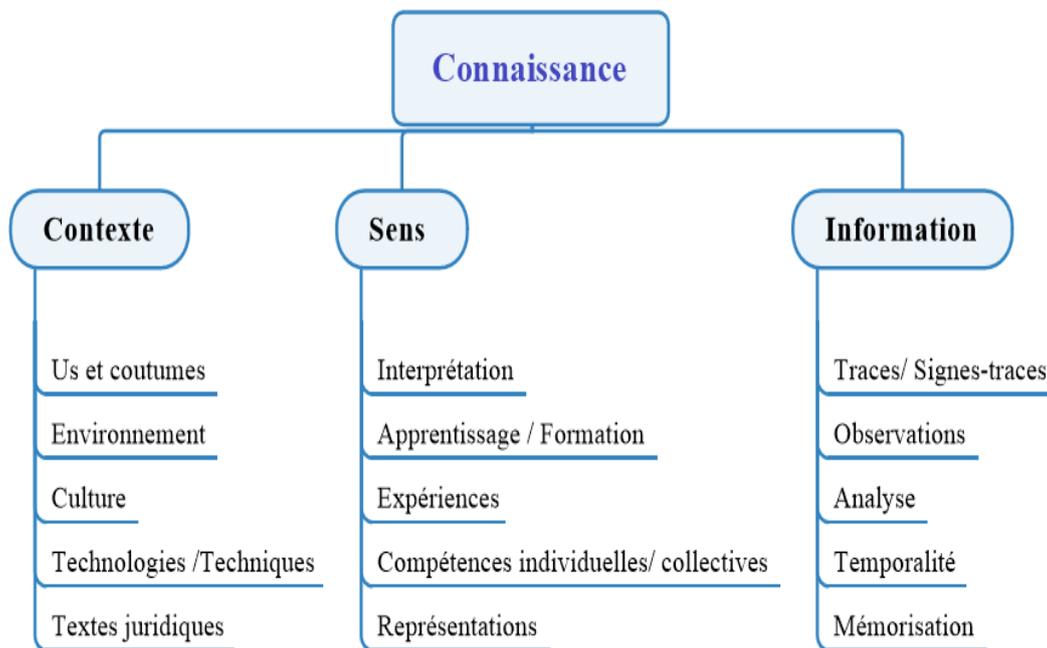


Figure 57: représentation de la production de la connaissance

Plusieurs formes de connaissance coexistent. Deux grandes formes de connaissance sont communément admises : la connaissance explicite et la connaissance tacite.

La connaissance explicite peut être codifiée, exprimée dans un langage formel et partagée sous forme de formules, de spécifications, de schémas, de manuels de procédures ou de livres. Une fois consignée sous l'une de ces formes, elle peut facilement se stocker et se transmettre (PAQUET, 2006 ; BLUMENTRITT, JOHNSTON, 1999).

La connaissance tacite au contraire, incorpore des éléments cognitifs et techniques qui ne peuvent être codifiés. Les éléments techniques de la connaissance tacite sont liés à l'habileté mise en œuvre dans un contexte spécifique (PAQUET, 2006 ; BLUMENTRITT, JOHNSTON, 1999). Ces connaissances contiennent une part de subjectivité liée à la personnalité de celui qui la détient, ce qui rend difficile leur formalisation et leur communication.

A partir des deux formes de connaissance, Nonaka (NONAKA, 1994) a mis en place le « modèle SECI » qui définit quatre modes de création des connaissances permettant de passer de la connaissance tacite à la connaissance explicite et inversement (figure 58) :

- la socialisation : passage d'un état de connaissance tacite à un autre état de connaissance tacite. Cela se produit notamment lors d'un partage d'expérience entre employés via l'observation et la reproduction d'une pratique ;
- l'extériorisation (explication) : cette étape complexe permet d'explicitier des connaissances tacites en utilisant des concepts abstraits (analogie, métaphore, hypothèse, modèle...). Ce mode de conversion est notamment à l'œuvre lors d'échanges informels entre plusieurs salariés ou lors de réflexions collectives ;
- l'intériorisation (appropriation) : la connaissance explicite est assimilée par l'individu au point de devenir un automatisme ;
- la combinaison (partage) : passage d'une connaissance explicite à une connaissance tacite. Lorsque plusieurs individus possédant un langage commun échangent des connaissances explicites, cela crée de nouvelles connaissances.

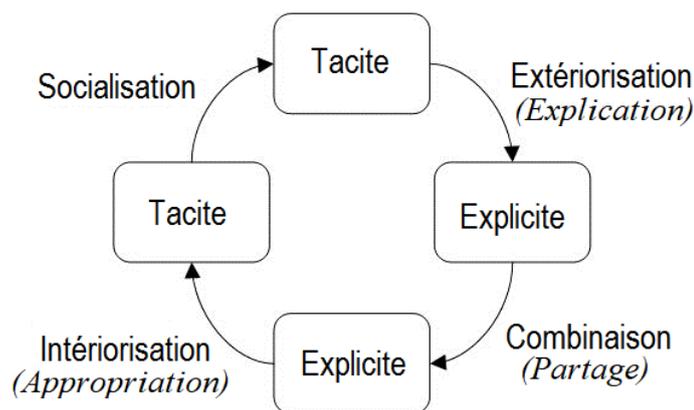


Figure 58 : cycle de création de la connaissance (NONAKA, 1994)

Le processus de la création de la connaissance est une spirale continue entre l'individu et le collectif ; un mouvement de dedans en dehors et inversement et le plus souvent une combinaison des différents modes de création. Il ne s'agit pas d'un processus linéaire mais plutôt d'un processus complexe qui demande une construction et une mise en œuvre minutieuse.

5.4.3 *Dynamique des connaissances médicales*

De l'expérience

En tant que médecin, nous posons des diagnostics qui combinent diverses dynamiques de traces (individuelles et sociales) dont nous souhaitons rendre compte ici.

Prenons un exemple comme point de départ : lors de la consultation, mon patient présente une hypertension artérielle pulmonaire avec une pression artérielle pulmonaire systolique à 70 mm Hg. Si, et seulement si, je rapproche cette information d'une autre où il m'est indiqué que ce même patient a présenté une embolie pulmonaire massive il y a un an, je pourrai dire que l'hypertension artérielle pulmonaire est une complication de son embolie pulmonaire. La confrontation des informations me permet de passer à la connaissance de l'état du patient. En plus, les données de la pression artérielle pulmonaire de ce patient, ne peuvent pas être moyennées avec les valeurs antérieures à sa crise d'embolie pulmonaire ; ces données ayant des contextes et des référentiels différents chez le même patient.

La compréhension et l'assimilation des règles qui régissent les modèles ou les schémas mentaux qui leurs sont sous-jacents me permettent de comprendre comment la situation évoluera si les résultats du patient se modifient. Nous utilisons ici une *connaissance explicite* mais aussi une *connaissance tacite*. Les connaissances tacites contiennent une part de subjectivité liée à la personnalité de celui qui la détient. Elles sont fortement dépendantes de l'individu et portent les *conséquences-traces* (GALINON-MÉLÉNEC, 2011) de ses croyances et de son système de valeur (PAQUET, 2006 ; BLUMENTRITT, JOHNSTON, 1999).

Ainsi, dans l'exemple ci-dessus, les résultats d'hypertension pulmonaire artérielle resitués en connaissance explicite d'embolie pulmonaire nécessitent des actions et une prise de décision adaptée au contexte, à l'environnement du patient, à la pratique médicale du médecin. Cette décision combine connaissance explicite et connaissance tacite.

L'importance de la subjectivité a été mise en évidence une fois de plus lors de la déclaration de la *Covid-19* par les médecins Chinois. A partir de leurs expériences et de leurs connaissances tacites, ils avaient compris que les symptômes et les signes cliniques ne correspondaient pas à une maladie connue aboutissant ainsi à la découverte d'une nouvelle maladie connue. Ils ont donc abouti ainsi à la découverte d'une nouvelle maladie. Cependant,

l'information qu'a tenté de faire circuler l'un d'entre eux - Li Wenliang²³⁰ - sur le sujet a fait l'objet de la censure immédiate.

Cet exemple illustre une pratique médicale générale qui tend à écarter toutes connaissances tacites dans le champ de la recherche²³¹. Ce combat s'illustre de nouveau dans la violente controverse qui oppose dans le cadre des propositions thérapeutiques contre la *Covid-19*. Le médecin qui livre aux médias son expérience personnelle pour montrer l'efficacité de son protocole thérapeutique entraînant ainsi des nombreux débats (voir encadré 5 dans la section 5.2.1). Ces débats se justifient dans la mesure où pour valider des nouvelles connaissances, il faut des preuves scientifiques probantes, confirmées à travers des méthodes statistiques appliquées sur un grand nombre de patients.

Nous venons donc à poser des questions sur les méthodes de construction des connaissances reconnues par les sociétés modernes :

- les expériences individuelles ou collectives n'ont-elles pas de valeurs tant qu'elles n'ont pas été démontrées mathématiquement ?
- toutes les connaissances sont-elles quantifiables ?
- n'existe-il pas plusieurs méthodes de productions des connaissances ?

Ces questions nous paraissent d'autant plus importantes à poser que notre origine culturelle nous a permis d'observer des traitements et guérisons dues à l'application de méthodes ancestrales n'ayant jamais été soumises au filtre des statistiques.

Passage de l'expérience individuelle au social (et inversement)

Quand les médecins ont la volonté et la capacité langagière de passer de la connaissance tacite à la connaissance explicite, ils peuvent la partager et créer ainsi de nouvelles connaissances utilisables de façon plus générale. Le processus d'une intériorisation de l'expérience transformée en connaissance tacite puis en connaissance explicite est présenté dans la figure 48.

²³⁰Source : https://www.huffingtonpost.fr/entry/coronavirus-mort-de-li-wenliang-medecin-chinois-lanceur-dalerte_fr_5e3c8d81c5b6f1f57f0df316 consulté le 18 mai 2020.

²³¹ Ce site web présente les résultats des études faites dans le cadre la recherche sur le traitement de la maladie à coronavirus. Les preuves d'efficacité sont basées sur les essais cliniques et les méta-analyses. Source : <http://www.metaevidence.org/covid19.aspx> consulté le 20 avril 2020.

Le processus de la création de la connaissance médicale suit un circuit circulaire continu entre chaque médecin et le collectif : après les phases de collecte, d'analyse et d'interprétation, les résultats sont soumis à une vérification par les pairs avec une mise à disposition de tous les éléments de l'étude et une critique des différentes étapes. Les modifications et les recommandations de mise en pratique en fonction des contextes et de l'environnement sont ensuite validées par les sociétés savantes médicales.

Formation des médecins

Les connaissances ainsi produites sont ensuite intégrées dans la formation des étudiants en médecine qui, dans un premier temps, recevront la formation de leurs enseignants (phase d'internalisation), suivi d'une phase d'apprentissage sur le terrain de stage en milieu hospitalier avec les collègues et les enseignants (phase de socialisation et externalisation). Cette phase d'apprentissage aboutira à une assimilation similaire de tous les étudiants en formation des connaissances de la médecine (phase de combinaison). Dans le domaine de la santé, cette boucle est fermée et le médecin passe d'une phase à une autre en fonction des avancées dans le domaine de la médecine et du contexte dans lequel se déroule sa pratique médicale.

Ce qui nous conduit à proposer de faire évoluer la formation des médecins en y intégrant la dynamique conceptuelle portée par les sciences humaines et en particulier par celles rendues explicites dans les publications de la série *L'Homme-trace*. Par conséquent, la qualité de mise en place des différentes étapes est très primordiale pour la production et la validation des connaissances. Dans le cadre du *big data* et de l'IA, l'usage des traces numériques pour la production de valeur nous amène à définir la notion de traces et les implications dans son utilisation et son interprétation dans le cadre de la pratique médicale.

5.5 Passerelles entre expérience et notions théoriques

Selon Hippocrate, « *il est recommandé aux médecins d'observer l'homme dans son environnement pour lui venir en aide, lui porter les secours de la médecine, restaurer sa santé* » (CONSEIL NATIONAL DE L'ORDRE DES MÉDECINS, 2018). A cette fin, il doit être en mesure d'effectuer un diagnostic pertinent, adapté à la spécificité d'un patient ; ce qui induit non seulement la mise en place éventuelle d'un protocole associé à telle ou telle pathologie mais aussi une compréhension de son contexte de vie.

Ainsi, deux approches de la médecine coexistent (figure 59) :

- une approche individuelle basée sur les données cliniques du patient, les dossiers médicaux du patient et son environnement (hérédité, prédispositions, facteurs de risque, coutumes, entourage) ;
- une approche de santé publique basée sur l'agrégation des données individuelles, l'approche systémique à partir des données épidémiologiques pour la surveillance et l'évaluation de l'émergence des pathologies avec une simulation des crises.

Ces deux approches fonctionnent en boucle et dépendent l'un de l'autre.

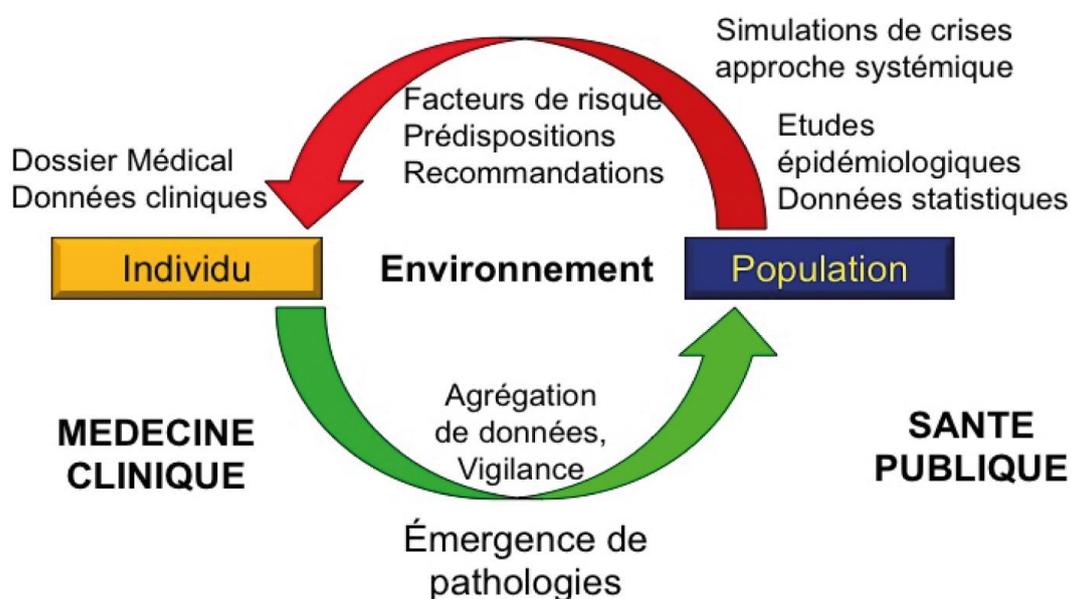


Figure 59 : approches fondamentales de la médecine : médecine clinique et santé publique (COLLOC, 2015)

Cette complémentarité s'est trouvée interrogée au moment de la crise sanitaire de 2020, la Covid-19 transmise par le contact d'individu à individu étant devenu un problème de santé publique. A l'interface entre les deux, la question des traces numériques s'est posée avec d'autant plus d'acuité que la gestion par la Chine posait en parallèle la question du respect des libertés individuelles et de la *privacy*. Cela nous conduit à approfondir chacune des deux approches dans cette thèse. Nous synthétisons la façon dont nous abordons la question dans la figure (ci-dessous) : « enjeux et risques de la circulation des traces numériques en matière de santé ».

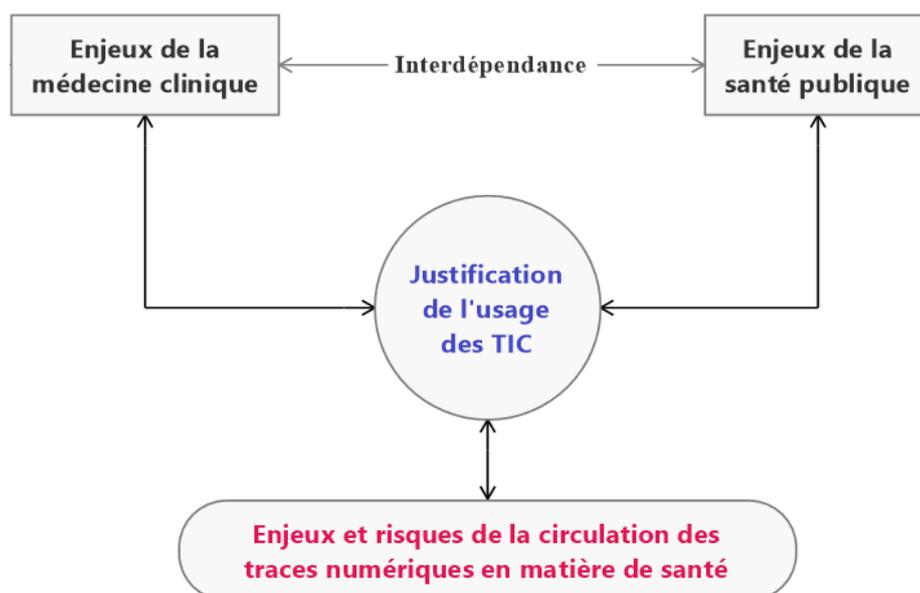


Figure 60: enjeux et risques de la circulation des traces numériques en matière de santé

5.5.1 Médecine clinique et diagnostic en coprésence

Pour le médecin - également docteur en philosophie -, Georges Torris (TORRIS, 2018), le diagnostic est un acte individuel du médecin résultant d'une consultation qui vise à déterminer l'origine et la nature de la souffrance présentée par le patient. Il est indispensable à l'établissement du pronostic et de la thérapeutique. Il est moins une phase de l'examen médical ou paramédical qu'une conclusion de celui-ci. Facile et même évident dans certains cas, il peut être très difficile dans d'autres ; devant l'urgence thérapeutique, le médecin devra alors décider si et quand les documents qui s'accumulent permettent l'arrêt de l'investigation et le passage à une conclusion, qui reste forcément révisable, mais n'en est pas moins nécessaire. Ce que Galinon-Méléneq nomme « les *signes-traces* du *corps-trace* du patient » (GALINON-MÉLÉNEC, 2011) pouvant ainsi devenir l'objet d'une enquête interminable²³², il faut savoir y mettre fin. La question qui se pose alors est celle du choix du moment.

²³² C'est en ce sens qu'on comprendra que le même mot grec (*diagnôsis*) ait signifié à la fois discernement, décision et diagnostic médical.

5.5.2 *Question de la temporalité*

Michel Foucault, philosophe issu d'une famille de médecins, s'intéresse à l'épistémologie de la médecine et publie en 1963 « *Naissance de la clinique : une archéologie du regard médical* ». Pour lui, la médecine clinique accorde une grande importance à la question du temps. Elle s'emploie à retracer l'enchaînement naturel, successif des symptômes, reconstruisant l'évolution chronologique des maladies (FOUCAULT, 1963 ; 2015). L'examen clinique correspond donc à un repérage de la vie. Les différents temps (les trois M : le temps du médecin, le temps du malade, le temps de la maladie) décrits par Hippocrate (DUPONT, 2001)²³³ permettent de mieux comprendre le processus de l'examen clinique dans le cadre du diagnostic et de la prise en charge :

- le temps du malade permet : a) d'écouter les plaintes et la description des symptômes et la signification que la malade accorde à chaque symptôme, b) d'informer et d'éclairer le malade, c) de prendre en compte son environnement et d) de respecter l'autonomie du patient ;
- le temps de la maladie permet : a) de percevoir et interpréter les symptômes, d'analyser les moyens d'accès au diagnostic, b) décrire l'évolution de la maladie et c) d'analyser l'effet des thérapeutiques entreprises dans le cadre de la prise en charge de cette maladie ;
- le temps du médecin permet : a) de faire un diagnostic objectif avec un choix de traitement adéquat et b) de prendre en compte les contradictions entre les médecins et/ou les équipes de soins et les incertitudes dans le raisonnement médical.

Ces différents temps permettent aux médecins de faire une recherche des *signes-traces* de la maladie et l'effet des différentes actions du médecin sur la maladie. Ils permettent également de prendre en compte l'état antérieur du patient pour apprécier l'évolution de ces *signes-traces*. La prise en compte de ces trois temps est très importante pour l'interprétation des *signes-traces* dans le cadre du diagnostic médical et social de la maladie par le médecin.

La prise en compte de cette temporalité dans le cadre de la *Covid-19* a permis de comprendre qu'il existait une phase - de 14 jours en moyenne - à la fois asymptomatique et contagieuse d'une période. Au début de la *Covid-19*, la détresse respiratoire a été rapidement liée à

²³³ Bernard-Marie Dupont est un exemple de l'interdisciplinarité en médecine : il est docteur 1) en médecine, 2) en génétique, 3) en philosophie 4) en droit privé et sciences criminelles.

l'atteinte des alvéoles par la maladie ; mais avec les observations autopsiques des cas de décès, les médecins ont fait l'hypothèse que la détresse respiratoire était liée à la vascularité due au tropisme du virus pour les parois vasculaires. Il en est de même pour les symptômes, qui étaient initialement constitués de la fièvre et de la dyspnée, se sont progressivement étoffés au fil du temps avec d'autres symptômes (diarrhée, agueusie, perte de l'odorat, troubles du rythme, ...). L'accumulation des observations au cours du temps a permis d'interpréter des symptômes en tant que *signes-traces* de la présence du coronavirus.

Le diagnostic médical est fondé sur un raisonnement analogique et déductif à partir d'un ensemble de signes et de symptômes pour déterminer la cause d'une affection du corps ou de l'esprit²³⁴. Dans cette thèse, nous les avons reliés aux recherches en SIC et les avons nommés *signes-traces* en prenant en compte le contexte de l'anthropo-sémiotique développée par Galinon-Méléneq.

Cependant, il convient de se rappeler que les origines de la sémiologie et de la sémiotique développées par Saussure, Barthes ou Peirce - pour ne citer que quelques auteurs de cette discipline relativement récente - ont largement été inspirées par la pratique médicale. En effet, dès 1855, *Le Dictionnaire de médecine* utilisait le terme *séméiologie* pour désigner la « *partie de la médecine qui traite des signes des maladies* », définition reprise récemment dans le tome IV de la série *L'Homme-trace* en ces termes : « *est une branche de la médecine qui a pour objectif de déduire les dysfonctionnements physiques et/ou psychiques du corps à partir des signes extérieurs ou intérieurs du corps et de l'esprit* » (COLLOC, BONENFANT, 2017).

Aujourd'hui la lecture des *signes-traces* ne se limite plus au dehors du corps. Des techniques de plus en plus sophistiquées permettent de rendre visible ce qui se passe dans le dedans du corps. De telles investigations permettent de gagner du temps dans la compréhension de ce qui se passe chez le patient et cette évolution technique participe grandement à la possibilité de diagnostics précoces. Le diagnostic est ainsi ancré dans une temporalité. Il est d'ailleurs perpétuellement revu en fonction de l'évolution de la maladie et de l'interaction du médicament avec le traitement.

²³⁴La distinction corps esprit relève ici de l'approche de la médecine qui cloisonne corps et esprit.

La précocité d'un diagnostic ne relève pas que de l'évolution technique. L'expérience individuelle et collective des médecins une grande importance²³⁵ (COLLOC & BONENFANT, 2017).

La question de la temporalité concerne aussi la consultation médicale en ce qu'elle comporte deux phases :

- l'*anamnèse*²³⁶ ou l'interrogatoire ;
- l'examen physique²³⁷.

L'interrogatoire va porter sur les antécédents personnels et familiaux²³⁸ du patient, son mode de vie et son entourage. Elle porte également sur la description des symptômes ressentis par le patient avec leur évolution dans le temps que le médecin devra interpréter. Dans ce cas, le médecin utilise la mémoire du patient à la recherche de trace cognitive : « [...] *mon livre est un plaidoyer pour la mémoire comme matrice d'histoire, dans la mesure où elle reste la gardienne de la problématique du rapport représentatif du présent au passé* » (RICOEUR, 2000b).

Durant cette phase de remémoration du patient, le médecin doit traduire les récits du patient en termes médicaux, les interpréter ensuite de manière à construire un scénario probabiliste (GALINON-MÉLÉNEC, 2020) guidant son examen clinique vers la recherche d'indices (GINZBURG, 1980 ; GALINON-MÉLÉNEC, 2020) évoquant ce scénario.

Dans cet examen, le médecin est totalement investi : il utilise tous ses sens pour mieux rechercher tous les *signes-traces* pouvant être liés à la description des symptômes du malade.

²³⁵ On retrouve ici l'idée avancée plus haut que la circulation et le partage des informations et les représentations participent à faire évoluer la connaissance médicale.

²³⁶ Anamnèse vient du grec. *aná-*: «de bas en haut» -*mnêsis*, «mémoire» c'est-à-dire «faire remonter les souvenirs». En médecine, l'*anamnèse* retrace les antécédents et l'historique de la plainte actuelle du patient pour reconstituer l'histoire de la maladie.

²³⁷ L'examen physique est l'examen du corps du patient par l'intermédiaire de quatre moyens d'observations utilisant les sens du médecin : inspection (la vue et l'odorat), palpation (le toucher), percussion (le toucher et l'ouïe) et auscultation (l'ouïe).

²³⁸ La prise en compte des antécédents familiaux dans le processus de diagnostic témoigne de la transmission intergénérationnelle des signes-traces (THOMAS-VASLIN, 2017).

Sa perspicacité est elle-même le *signe-trace* de son savoir et de son expérience. Ainsi, certains commencent leur enquête dès la salle d'attente du médecin à la recherche de certaines positions ou attitudes du patient, autres *signes-traces* pouvant orienter ou faire suspecter le diagnostic. Par exemple, un patient qui se plaint de douleurs thoraciques et se met en position mahométane (position avec le thorax antéfléchi) fait suspecter une péricardite²³⁹.

L'examen clinique est l'occasion pour le médecin de rechercher des *signes-traces* de la maladie sur le corps du malade. Ainsi :

- *l'absence de « signe-trace » de la maladie crainte par le patient* est très importante car elle permet d'éliminer ou retenir systématiquement certains diagnostics. Cependant, l'absence de traces visibles pouvant être le *signe-trace* d'autres phénomènes, il convient aussi de s'interroger sur cette invisibilité ;
- *les signes-traces trouvés mais non évoqués par le patient* : les *signes-traces* retrouvés à l'examen physique peuvent amener le médecin à reprendre l'anamnèse pour s'assurer que le patient n'a pas oublié²⁴⁰ de donner certains détails. Il s'agit d'un perpétuel mouvement de va-et-vient dans la singularité et l'individualité du patient.

A partir des *signes-traces* retrouvés - présents ou absents en première lecture -, le médecin va poser des hypothèses diagnostiques à partir d'un raisonnement déductif.

Avec le volume des données produites lors des différents examens, le médecin peut recourir aux systèmes d'aide à la décision médicale (SADM) pour faire le tri des données et formuler des hypothèses diagnostiques. Mais, « *les statistiques et les algorithmes n'ont pas de sens moral, ni de tact* » (CARDON, 2015) ; cette phrase rappelle l'importance de la prise en compte des limites des statistiques et des algorithmes par les médecins dans le cadre de leur prise de décision.

²³⁹ La péricardite est une inflammation des parois du cœur.

²⁴⁰ L'oubli comme le décrit Paul Ricœur peut être volontaire ou involontaire (RICOEUR, 2000b). Dans le cadre de l'anamnèse médicale, nous agissons le plus souvent pour réduire l'oubli involontaire.

5.5.3 *Connaissance du milieu*

Même si le patient vient seul en consultation, il est un être social et son milieu de vie est important à connaître car rien n'est cloisonné. Selon le paradigme de l'Homme-trace, tout ce qui est le patient est inscrit dans son « *corps-trace* », lequel évolue en portant les *conséquences-traces* de ses interactions avec son milieu. Il s'agit « *d'un processus complexe et continu d'intériorisation de l'extériorité et d'extériorisation de l'intériorité* » (GALINON-MÉLÉNEC, 2011).

Même si le médecin utilise l'expression individuelle du corps, la subjectivité du patient et son environnement dans le processus de soins, le contexte socioculturel du malade, ses interactions avec son entourage et sa relation avec le médecin sont très importants.

Ainsi, l'interprétation des symptômes et des résultats des examens complémentaires prend en compte les conditions de l'apparition des *signes-traces* pour aller ensuite vers l'hypothèse diagnostic après un processus d'interaction systémique dans le cadre du paradigme des *signes-traces*. Dans ce cadre,

« l'interprétation des signes-traces émises par le corps apparaît comme inscrite dans une dynamique systémique liant le corps-trace de l'interprété et le corps-trace de l'interprétant au sein d'un complexe où des corps en coprésence sont animés par des interactions permanentes de traces » (GALINON-MÉLÉNEC, 2017).

Par exemple, la présence et l'interprétation des images en verres dépolies à la radiographie pulmonaire ou au scanner pulmonaire dépendent du contexte d'apparition et de l'environnement du patient. En effet, avec l'épidémie de la *Covid-19*, ces images sont considérées comme des signes indirects de la maladie et permettent d'évoluer vers le diagnostic. Pourtant, ces images étaient les signes indirects de la pneumocystose (maladie parasitaire survenant chez les personnes immunodéprimées) dans les années 2000 avant l'avènement de la trithérapie antirétrovirale.

5.5.4 *Notre viatique*

Pour résumer notre conception du diagnostic médical, nous dirons que le médecin n'analyse pas le symptôme dans son unité ; il l'analyse dans un système, un ensemble complexe avec des interactions importantes. Cet ensemble est constitué de la cellule jusqu'à l'organe et aux systèmes organiques, en tenant compte de l'interaction de ces différents éléments entre eux et avec l'environnement.

De notre point de vue, le cheminement de l'analyse des *signes-traces* de Galinon-Méléneec appliqué à un scénario médical comporte trois caractéristiques communes au paradigme indiciaire de Ginzburg, tel qu'il est présenté par Serres (SERRES, 2012) :

- *l'individualité de l'objet d'analyse* : dans le cadre de la médecine, elle se base sur les données qualitatives et les mesures individuelles à partir d'appareils technologiques, avec une prise en compte de la singularité²⁴¹ de l'individu dans le diagnostic de la maladie et dans le processus de soin. Par exemple dans le cadre du diagnostic d'une même maladie, l'individu ne présentera pas forcément les mêmes symptômes dans l'évolution de la maladie, ni la même progression des signes ; c'est en cela que la médecine décrit des formes cliniques²⁴² de la maladie ;
- *le caractère indirect du déchiffrement* : il n'existe pas une cause à effet direct dans le cadre de l'analyse des *signes-traces* retrouvés par le médecin. L'observation directe des *signes-traces* ne suffit pas, il faut un processus d'interprétation avec une base de connaissance importante aussi bien de la discipline et de l'environnement ;
- *le caractère conjectural* désigne la marge irréductible de l'aléatoire, d'incertitude dans le déchiffrement des *signes-traces*. Nous constatons des erreurs ou des retards de diagnostic en médecine dus à l'interprétation des *signes-traces*.

Nous proposons ainsi à partir des travaux de Galinon-Méléneec, Mille, Ermine, Nonaka, Colloc et Lery (GALINON-MÉLÉNEEC, 2011 ; GALINON-MELENEEC, ZLITNI, 2013 ; MILLE, 2013a ; 2006 ; ERMINE, 2000 ; NONAKA, 1994 ; LERY, COLLOC, 2008 ; COLLOC, LÉRY,

²⁴¹ La singularité de l'individu peut être d'origine anatomique (organe manquant, surnuméraire, dextrocardie ou asymétrie), génétique ou liée à l'environnement.

²⁴² Les formes cliniques sont des sous catégories de la même maladie en fonction de l'évolution ou du mode l'apparition des symptômes ou de la réponse à un traitement donné ou en fonction des caractéristiques propres à l'individu.

2008b ; LELEU-MERVIEL, 2013), la figure ci-dessous pour la description du processus d'analyse, d'interprétation des *signes-traces* dans le cadre de la pratique médicale.

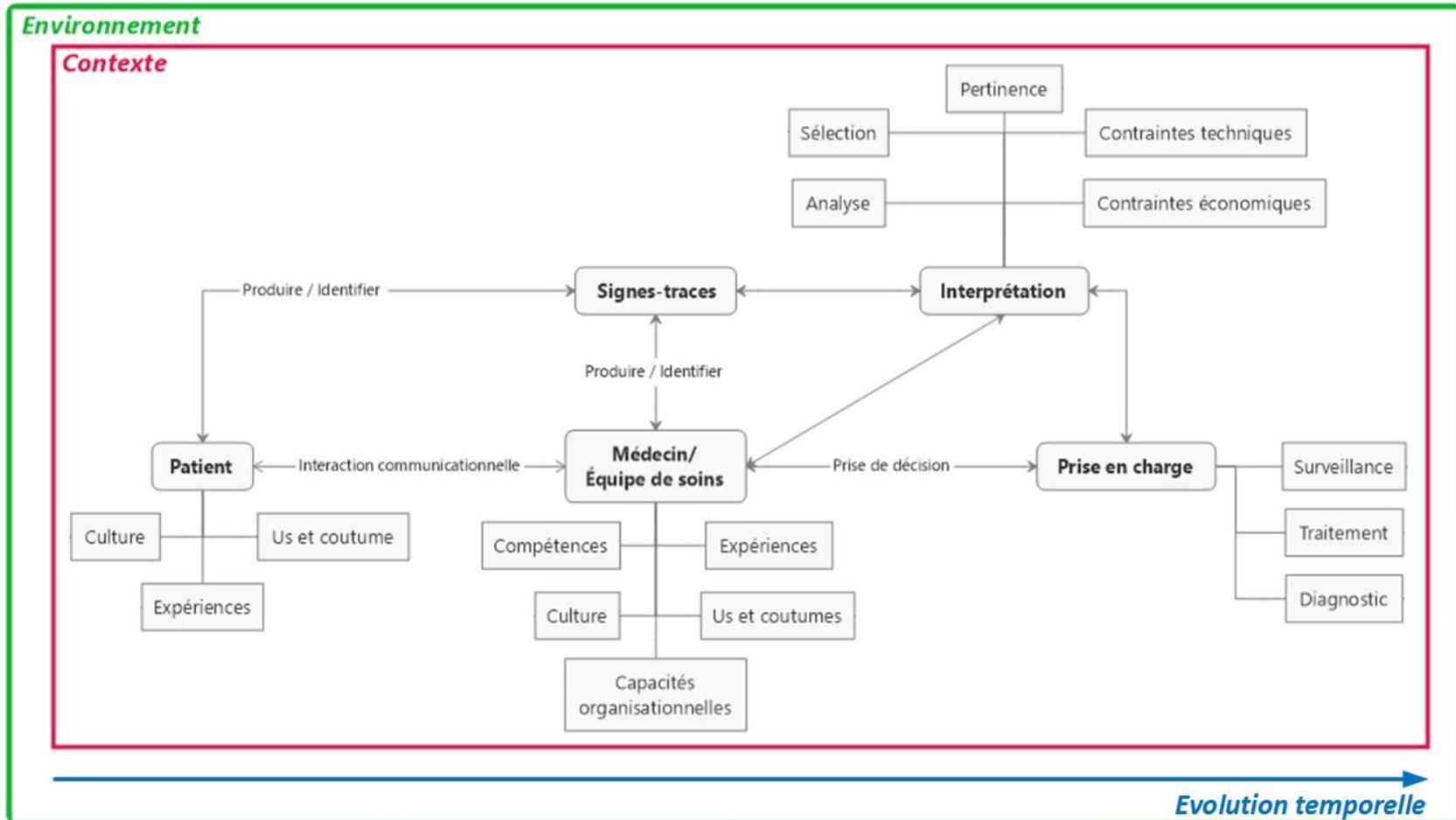


Figure 61: processus d'analyse et d'interprétation des *signes-traces*

Lors de l'examen clinique du patient dans le cadre du diagnostic, le médecin utilise plusieurs types de *signes-traces* - seuls ou en combinaison - afin d'aboutir à un diagnostic correct à partir d'un raisonnement déductif très bien codifié ; connaissance acquise au cours de plusieurs années de formation et d'expériences dans la pratique de l'art médical.

A l'issue de l'examen clinique, le médecin inscrit toutes les informations recueillies dans le dossier médical du patient. Avec la dématérialisation du dossier médical du patient, cette inscription écrite se fait dans un environnement informatique. Nous pouvons considérer cette phase d'écriture comme le « *moment de l'archive* » défini par Paul Ricœur comme celui de « *l'entrée en écriture de l'opération historiographique* » (RICOEUR, 2000b). Ce processus est une transcription de tous les actes du médecin dans le processus de soin en écrit la transformant ainsi en trace écrite et en trace numérique.

Le processus d'écriture²⁴³ digitale du médecin se fait à partir des informations d'interprétation des *signes-traces* réalisées par celui-ci. Nous pouvons ainsi dire que la trace numérique dans le domaine de la santé est une « *trace de la trace* » - selon l'expression de R. Barthes²⁴⁴ à propos de la photographie - une « *conséquence-trace scripturale* » (GALINON-MÉLÉNEC, 2020) de l'interprétation de *signes-traces* du *corps-trace* du patient par le médecin.

5.6 Utilisation des traces numériques de santé dans le *big data*

L'accumulation des traces numériques produites par le médecin dans les environnements numériques, couplées à celles produites par le patient et l'administration de santé seront à l'origine de la production de grande base de données et du *big data* dans le domaine de la santé. *Le big data* dans le domaine de la santé constitue des espoirs pour les uns et des craintes pour les autres, l'analyse et l'interprétation des traces doivent être ainsi discutées.

Avec le développement des TIC et l'amélioration de l'accès à Internet, nous avons assisté à une informatisation croissante du système de santé. Cette informatisation du système de santé s'est faite progressivement avec l'informatisation des processus de soins, la création de la carte vitale, la mise en place du dossier médical patient, le développement et

²⁴³ « *L'écriture est un système de notation attaché à l'activité humaine* » (DERRIDA, 1979). Dans le cadre de l'activité du médecin, elle va correspondre à tous les actes de soins. Par ailleurs tous les actes ne peuvent pas être traduits en trace écrite.

²⁴⁴ Cité par GALINON-MÉLÉNEC, 2020.

l'amélioration de l'accessibilité de la bio-informatique et l'utilisation de plus en plus croissante des objets connectés. Cette informatisation a abouti à une production importante des données de santé au fil des années aussi bien dans le public que dans le privé (SALGUES, 2016 ; MONINO, SEDKAOUI, 2016).

Avec la production croissante de volume de données, nous assistons à une utilisation du *big data* pour le stockage et traitement des données. La rareté des ressources et l'augmentation des dépenses de santé dans la plupart des pays et particulièrement en France - du fait de la solidarité sociale dans la prise en charge et des innovations thérapeutiques²⁴⁵ dans le domaine de la santé - ; la valorisation des données collectées a évolué dans le domaine de la santé avec pour objectif une amélioration considérable de l'efficacité, de l'efficience et de la qualité du système de santé (MANYIKA ET AL., 2011 ; MONINO, SEDKAOUI, 2016).

5.6.1 Mode de collecte des données de santé

La collecte et l'analyse des données de santé sont devenues une priorité des gouvernements et des industries de l'informatique pour les raisons suivantes :

- la qualité des données : les données de la santé font partie des données les plus fiables de par leurs processus de collecte et de leurs impacts sur la vie des individus. De ce fait, elles nécessitent moins de traitements pour assurer leur qualité, donc moins coûteuses pour les institutions. En effet, les données de santé sont utilisées dans le cadre de la planification du budget de la Sécurité sociale, les actions de santé publique et le financement des hôpitaux par l'intermédiaire de la tarification à l'acte. Mais la qualité de l'information dépend de l'annotation faite lors du stockage des données. En médecine, des systèmes de classification sont utilisés pour uniformiser le stockage des données, chacun avec ses forces et ses faiblesses. On peut citer des outils comme la classification internationale des maladies (CIM), le *Systematized Nomenclature Of MEDicine (SNOMED)*, l'*Unified Medical Language System (UMLS)* ou le CCAM et le PMSI²⁴⁶ dans le

²⁴⁵ Les innovations d'une manière générale et en particulier dans le cas de la médecine comportent des coûts importants de la découverte au début de commercialisation. Leur accessibilité coûte chère dès la mise sur le marché. Dans le cas de la prise en charge des cancers, une immunothérapie coûte en moyenne 4 000 euros par mois. Il faut les situations de monopoles ou de quasi monopoles dans le domaine du *big data* et de l'IA.

²⁴⁶ Le PMSI utilise des stratégies de codage en vigueur faussent les statistiques du fait que ces données permettent d'obtenir des financements selon les classes de pathologies codées (COLLOC, VINOT, 2000).

contexte français ;

- la valeur des données : les données de santé décrivent l’histoire de vie des individus depuis la naissance. Elles comportent les informations sur les différents épisodes de maladies des individus, les prédispositions héréditaires, les consommations médicamenteuses et les handicaps. Ces données sont en général couplées aux données administratives dans le cadre de remboursement des soins par la Sécurité sociale ou les complémentaires santé. En France, le Système National des Données de Santé (SNDS)²⁴⁷ - créé par la loi de modernisation du système de santé du 26 janvier 2016²⁴⁸ - rassemble et met à disposition, pour des finalités définies, des données individuelles de santé « *pseudonymisées* » et des données agrégées issues des principales sources médico-administratives. La figure ci-dessous montre les différents types de données stockées par le SNDS ;

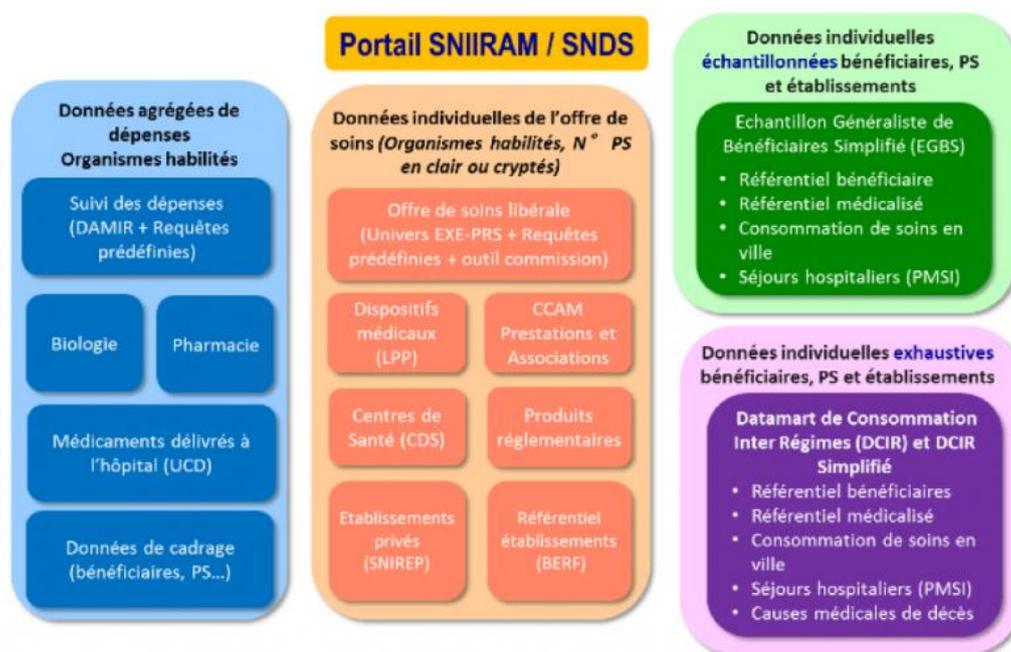


Figure 62 : différents types de données figurants dans la base du SNDS²⁴⁹

- le volume des données : les données de santé, en particulier en France, constituent

²⁴⁷ Le SNDS est géré par la CNAMTS (Caisse Nationale de l'Assurance Maladie des Travailleurs Salariés). Les données qui le composent sont conservées pendant une durée de dix-neuf ans en plus de l'année au cours de laquelle elles ont été recueillies. Passé ce délai, ces données sont archivées pour une durée de dix ans. Il concerne 57 millions de personnes (soit 86 % de la population) et enregistre tous les remboursements de chaque bénéficiaire tout au long de sa vie.

²⁴⁸ Loi n° 2016-41 du 26 janvier 2016 de modernisation de notre système de santé. 26 janvier 2016.

²⁴⁹ Source : <https://www.indsante.fr/fr/les-composantes-du-snds> consulté le 18 novembre 2017.

une base de données « *presque* » exhaustive de la population Française. En effet, avec le système de Sécurité sociale et des complémentaires de santé, toute personne vivant sur le territoire français a droit aux soins et bénéficie des avantages de la Sécurité sociale. Elle constitue de ce fait une base de données très importante car prenant en compte les personnes non « *connectées* ». Ce volume de données est également lié à l'intensité de l'activité médicale²⁵⁰ et de recherche en France. La figure 62 montre les différentes bases de données qui alimentent le SNDS ;

- la variété des données : les données sont disponibles en fonction de leur nature (génomique²⁵¹, physiologique, biologique, clinique, sociale), leur format (texte, valeurs numériques, signaux, images 2D et 3D), leur dispersion au sein de plusieurs systèmes d'information (groupes hospitaliers, laboratoires de recherche, bases publiques...). La variété des données dans les SIH et les entrepôts de données constituent un défi majeur notamment dans le cadre de l'interopérabilité sémantique et de l'extraction. Avec le *big data*, qui comporte à la fois des données structurées et des données non structurées, cette question est devenue cruciale pour les différents acteurs du système pour une valorisation des données. Ainsi, les méthodes classiques de recherche d'informations et d'extraction des données peuvent s'avérer inefficaces ou inadaptées dans le contexte du *big data*. Des outils comme le *Semantic Search Engine No SQL* et le *Semantic Search Engine SQL* développé par l'équipe D2IM de Rouen permettent de faire de la recherche d'information documentaire, bibliographique ou clinique dans le dossier clinique du patient (LELONG, 2019).

²⁵⁰ Le nombre de consultation par habitant et par an était de 6,1 et douze millions de personnes hospitalisées par an, 204 scanners et 114 IRM pour 1000 habitants en France. Source : https://www.oecd-ilibrary.org/social-issues-migration-health/data/statistiques-de-l-ocde-sur-la-sante/eco-sante-ocde-utilisation-des-ressources-en-sante_data-00542-fr consulté le 18 avril 2019.

²⁵¹ La liste des plateformes génomiques en France. Source : <https://www.france-genomique.org/plateformes-et-equipements/liste-plateformes/> consulté le 20 janvier 2020.

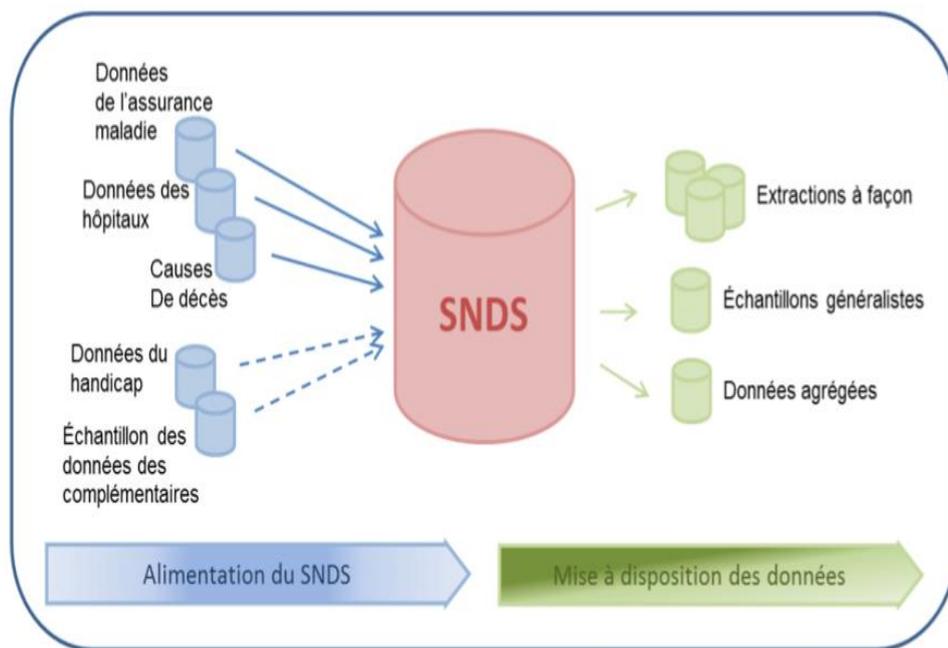


Figure 63 : différentes bases de données alimentant le SNDS²⁵²

Mais cette course effrénée pour l’informatisation de la santé et le traitement des données de santé comporte des enjeux au niveau de la sécurité des données, de la protection de la vie privée et de l’interprétation des données et des traces numériques.

En effet, selon l’enquête Dell *End User*²⁵³ en 2017, présentée par les figures 64 et 65, près de trois employés sur quatre (72 %) se déclaraient prêts à partager des informations d’entreprise sensibles, confidentielles ou soumises à des réglementations dans certaines circonstances et 45 % des employés d’organisations admettaient adopter des comportements à risque pendant leur journée de travail (DELL, 2017).

²⁵² Source : <https://www.indsante.fr/fr/les-composantes-du-snds> consulté le 18 novembre 2017.

²⁵³ Dimensional Research a mené une enquête en ligne commandée par Dell Data Security auprès de 2608 professionnels ayant personnellement accès à des données et informations confidentielles, sensibles ou soumises à des réglementations et les utilisant chaque jour dans des entreprises comptant plus de 250 employés. Les participants ont été interrogés dans huit pays : Allemagne, Australie, Canada, États-Unis, France, Inde, Japon et Royaume-Uni. Les recherches se sont déroulées entre le 24 février et le 9 mars 2017. Source : <https://france.emc.com/collateral/analyst-reports/dell-end-user-security-survey-2017.pdf> consulté le 22 novembre 2018.

72 %

des employés se disent prêts à partager des informations d'entreprise sensibles, confidentielles ou soumises à des réglementations dans certaines circonstances



obéissent à la demande de leur supérieur



partagent les données avec une personne autorisée à les recevoir



pensent que le risque est très faible et les bénéfices élevés



ont l'impression que cela les aidera à faire leur travail plus



ont l'impression que le destinataire fera son travail plus efficacement

Figure 64 : circonstances de partage des informations confidentielles par les employés (DELL, 2017)

45%

des employés d'organisations admettent adopter des comportements à risque pendant leur journée de travail

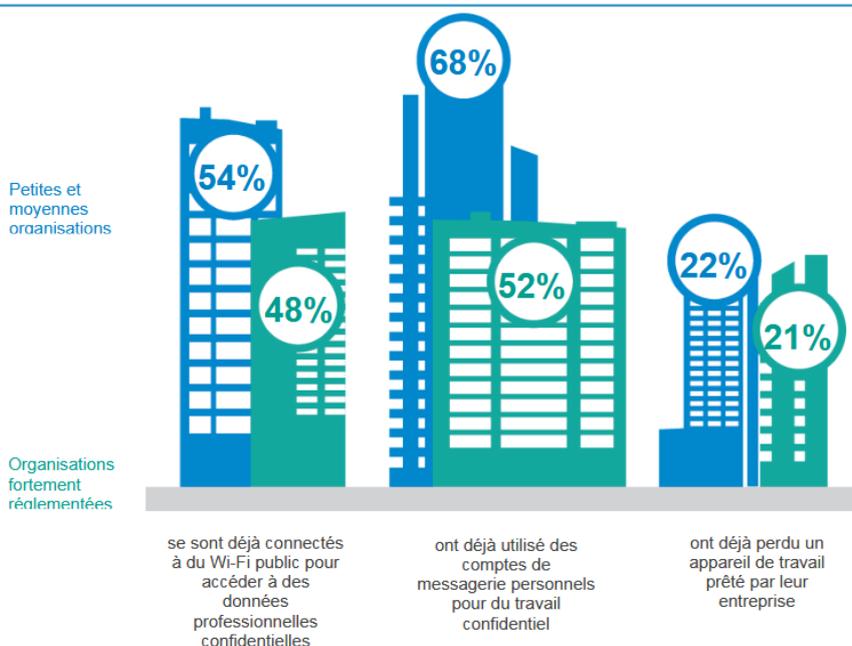


Figure 65 : comportements à risque des employés (DELL, 2017)

En 2016 déjà, l'enquête Dell (DELL, 2016) sur la sécurité des données²⁵⁴ montrait que :

- près de trois décideurs sur quatre admettaient que la sécurité des données était l'une des priorités des dirigeants de leur organisation, mais s'inquiétaient du fait que les hauts dirigeants ne s'intéressaient pas assez aux problématiques et aux outils de sécurité des données, et ne soient pas bien informés sur le sujet ;
- près des trois quarts (73 %) des décideurs étaient « *assez très préoccupés* » par les logiciels malveillants et les menaces avancées persistantes, alors que la majorité d'entre eux utilisait déjà des logiciels contre les programmes malveillants pour assurer la sécurité des données de leur organisation ;
- environ 82 % des décideurs essayaient de limiter les points d'accès aux données pour renforcer la sécurité et seuls 18 % d'entre eux étaient absolument certains que les données étaient sécurisées lorsque les employés y accédaient à distance.

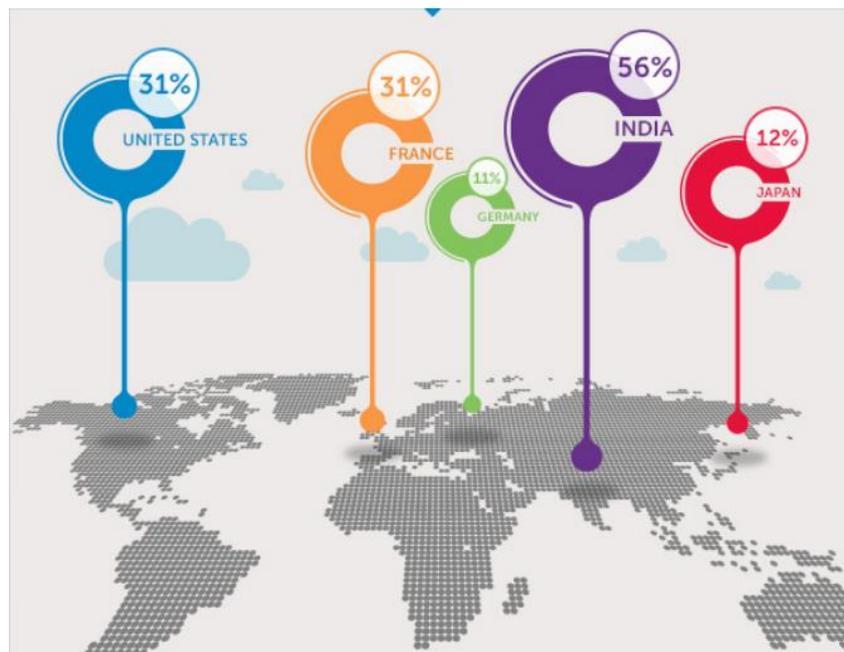


Figure 66 : proportion des attaques de programmes malveillants dans le monde (DELL, 2016)

²⁵⁴ Le cabinet Penn Schoen Berland a mené une enquête en ligne commandée par Dell auprès de 1 302 professionnels travaillant aux États-Unis, en Europe et dans la zone Asie-Pacifique, Japon inclus. Les personnes interrogées étaient des décideurs informatiques et métier. L'enquête a eu lieu entre le 19 novembre et le 8 décembre 2015. Cette enquête avait pour objectif de décrire la manière dont les décideurs informatiques des entreprises de taille intermédiaire (et les cadres dirigeants) perçoivent les évolutions en matière de sécurité des données et l'impact de ces dernières sur leur activité. http://i.dell.com/sites/doccontent/shared-content/data-sheets/fr/Documents/Enquete_securite_des_donnees_2016_FR.pdf consulté le 22 novembre 2018.

La sécurité dans le domaine de la santé n'est pas au point comme dans les institutions bancaires. En plus, la sécurité des systèmes n'est pas le premier souci des hôpitaux, d'autant plus qu'ils doivent faire face aux restrictions budgétaires pour assurer les services de soins. Ainsi, le domaine de la santé est sous tension du fait des attaques incessantes²⁵⁵ par les cybercriminels et la propagation des objets connectés (*pacemakers*, pompe à insuline, tensiomètre, glucomètre) qui constituent des failles potentielles dans les systèmes d'information de santé. Ces attaques peuvent être des vols de données²⁵⁶, des blocages des systèmes pour des rançons ou la modification de l'intégrité des données²⁵⁷. Ces failles de sécurité peuvent avoir des conséquences néfastes pour les patients et l'hôpital : perte de confidentialité des données, escroqueries des patients, procès pour l'hôpital et blocage des activités médicales.

5.6.2 *Systèmes d'aide à la décision médicale*

L'informatique est un outil utile pour le médecin, mais il doit avoir à l'esprit que c'est lui seul qui est apte à faire la synthèse entre les signes-*traces du corps* du patient et les informations qu'il reçoit via les « *systèmes d'aide à la décision médicale* » (SADM). Le médecin peut y recourir à condition de se souvenir qu'ils sont basés sur des calculs probabilistes – pour faire le tri des données et formuler des hypothèses diagnostiques. En effet, le but des SADM est de

« fournir aux cliniciens en temps et en lieux utiles les informations décrivant la situation clinique d'un patient ainsi que les connaissances appropriées à cette

²⁵⁵ Depuis le 1er octobre 2017, 478 incidents ont été déclarés par les structures de santé sur le portail de signalements (<https://www.cyberveille-sante.gouv.fr/>), soit une moyenne de 27 signalements par mois. Dans le détail, 88% des déclarations d'incidents proviennent des établissements de santé, 6% d'autres structures de type EHPAD (établissements d'hébergement pour personnes âgées dépendantes), 4% des laboratoires de biologie médicale et 2% des centres de radiothérapie. Dans la plupart des cas (46%), les problèmes informatiques ont conduit la structure à mettre en place un fonctionnement dégradé du système de prise en charge des patients. Sur les 12% de cas ayant présenté une mise en danger potentielle des patients, cinq ont entraîné une "mise en danger patient avérée". Source : <https://www.ticsante.com/story/4617/cybersecurite-478-incidents-declares-par-les-structures-de-sante-depuis-18-mois.html> consulté le 05 septembre 2019.

²⁵⁶ Le dossier médical se vend entre 22 et 150 € sur le marché noir, selon Vincent Trely, le président de l'Association Pour la Sécurité des Systèmes d'Information de Santé (APSSIS). Ce prix a évolué selon une enquête de Kaspersky pour atteindre 300 € en 2019.

²⁵⁷ Des chercheurs en sécurité informatique ont décrit une attaque inédite : s'introduire dans le système radiologique d'un hôpital pour en modifier les clichés d'imagerie, afin d'induire de faux diagnostics de cancer du poumon (MIRSKY ET AL., 2019).

situation, correctement filtrées et présentées afin d'améliorer la qualité des soins et la santé des patients » (BERNER, 2009).

Le système donne ainsi des recommandations et des suggestions au médecin à partir des données cliniques qu'il a introduites sans pour autant prendre la décision à sa place. Quel que soit le type de système, il doit aider le médecin dans l'immédiat dans le cadre de sa pratique quotidienne et dans la prise de décision. Les médecins ont en général besoin d'un SADM coopératif qui les complète et non d'un SADM capable de les remplacer. En effet, un SADM ne peut pas intégrer l'expérience du médecin qui est fondamentale dans la prise en charge des patients. Les SADM doivent prendre également en compte les choix et les préférences du patient. Les SADM ne doivent pas être conçus pour prendre une décision de manière autonome dans le domaine de la santé.

L'évolution des connaissances médicales nécessite que les bases de connaissances des SADM soient régulièrement mises à jour. Ainsi, les SADM fournissent un apport considérable à la synthèse des connaissances pour le médecin, mais aussi pour les étudiants en formation.

Les systèmes experts apparus dans les années soixante-dix, sont les premiers SADM avec notamment *eMycin* et *Mycin* dans le domaine de la santé ; mais leur déploiement en France est resté modeste car les médecins les perçoivent comme des outils portant atteinte à leur indépendance professionnelle et à leur pouvoir de prise de décision (SEROUSSET AL., 2011). Au fil des années, les SADM se sont développés avec l'utilisation des techniques de plus en plus avancées, passant des modèles déterministes (systèmes experts, logique floue, système multi agent, raisonnements à partir de cas) aux modèles autonomes (*machine learning*, réseaux de neurones, *deep learning*) ; ces derniers temps, on assiste aux développements de modèles hybrides. Mais force est de constater que cette évolution n'a pas permis une intégration des SADM dans la pratique médicale du fait de l'évolution rapide et constante des connaissances dans le domaine de la santé, l'interopérabilité des systèmes et la résistance des professionnels de la santé (DARMONI ET AL., 2014).

Mireille Cléret (CLÉRET ET AL., 2001) explique cela par le fait que « *les développeurs de ces systèmes ont souvent des ambitions de recherche en matière d'intelligence artificielle et de processus de raisonnement et appliquent de nouvelles méthodes à des connaissances médicales plutôt que d'adopter la démarche inverse qui consiste à résoudre un problème médical donné en*

appliquant une méthode adaptée. [...] les médecins n'imaginent pas toujours les bénéfices potentiels ou même les limites des systèmes d'aide à la décision ».

Il existe un problème de communication entre les médecins et les développeurs mais également la méthodologie de conception des SADM doit être revue pour prendre en compte les besoins réels des utilisateurs et les caractéristiques des problèmes à résoudre.

Avec le *big data* et l'IA, l'utilisation des SADM, en particulier les systèmes autonomes, pose des questions juridiques, déontologiques et éthiques. Avec ces technologies, « nous agrégeons des données d'après ce que l'on sait déjà, c'est-à-dire d'après la somme de connaissances déjà disponibles ; le risque est donc grand de conforter ce qui est déjà connu, jusqu'à la tautologie » J-C Lambert (CACHIA ET AL., 2015). Ces méthodes d'agrégation des données et d'analyses automatiques ont montré leurs limites et les conséquences (erreur d'interprétation, mauvais diagnostic et thérapeutique, expansion d'une épidémie) qu'elles pourraient engendrer avec la *Covid-19*. Dans ce contexte, l'absence de données et les incertitudes montrent les limites des SADM. En plus, ces méthodes posent le problème de la subjectivité de la médecine et l'objectivité face à une personne en fonction de ses données. La question de la responsabilité, de l'explicabilité, de l'exactitude, de l'auditabilité et de la justiciabilité des algorithmes utilisés dans les SADM constitue également des freins à son utilisation dans un domaine aussi complexe et sensible qu'est la santé.

Une analyse de ces questions s'avère importante pour intégrer les interactions entre les différents acteurs du système de santé dès la conception des outils, mais surtout prendre en compte le contexte particulier de la France en ce qui concerne l'organisation des soins. Nous proposons à la section 7 de notre thèse une méthodologie de conception des SADM.

5.6.3 Interprétation des traces numériques de santé dans le big data

L'analyse des traces dans le cadre du *big data* se base sur l'analyse exploratoire des données²⁵⁸ plutôt que sur l'analyse confirmatoire²⁵⁹. Or comme le précise Tukey JW (TUKEY, 1980), « aucun catalogue de techniques ne peut transmettre une disposition à

²⁵⁸ « L'analyse exploratoire des données (*exploratory data analysis*) vise plutôt à laisser ouverte la question sur laquelle portera la recherche et à manipuler un ensemble d'outils de représentation des données pour y déceler progressivement la question pertinente et les types d'analyses possibles » (DAGIRAL, PARASIE, 2017).

²⁵⁹ « L'analyse confirmatoire des données (*confirmatory data analysis*) consiste, à partir d'une question bien définie, à mettre en place un protocole d'enquête et à collecter des données que l'on analyse ensuite pour en tirer une conclusion » (DAGIRAL, PARASIE, 2017).

chercher ce qui peut être observé, que cela ait été anticipé ou non ». C’est pourtant là qu’est le cœur de l’analyse exploratoire de données.

« *Toute trace est polysémique et est interprétée différemment en fonction des contextes et des enjeux* » (GALINON-MÉLÉNEC, 2011). De ce fait, en fonction des disciplines, le processus d’interprétation de la trace ne sera pas la même. Mais avant toute interprétation de la trace, il faut que celui qui doit l’interpréter ait une connaissance du contexte et des circonstances ayant abouti à la trace.

La collecte des données, la gestion et le traitement des bases de données et la visualisation sont réalisés en majeure partie par des acteurs étrangers aux statistiques et du domaine dans lequel les données doivent s’appliquer. La collecte des données repose alors sur des choix et ce processus se fait de manière subjectif en fonction de la sensibilité des acteurs et non de manière objective comme le montre la figure 67 ci-dessous (DAGIRAL, PARASIE, 2017).

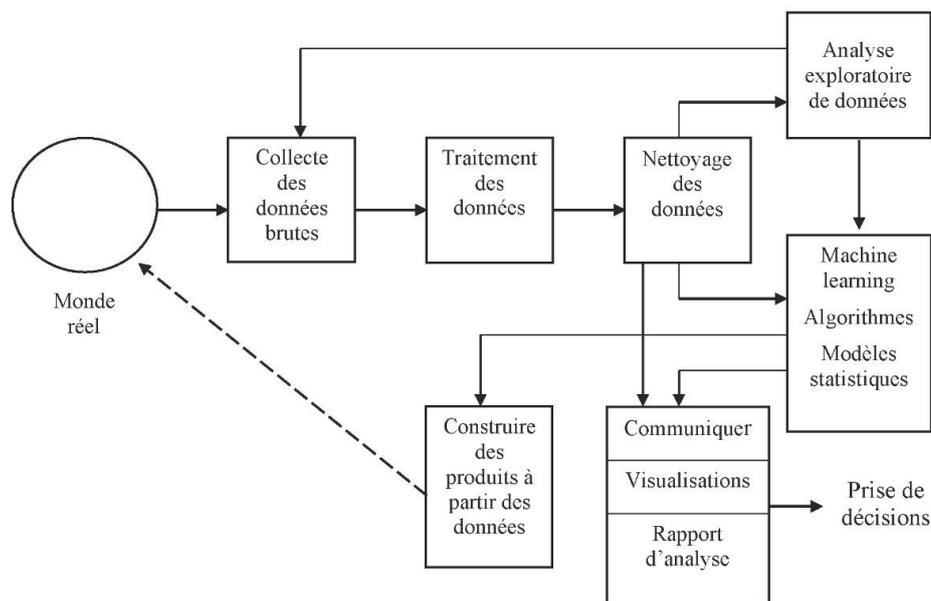


Figure 67 : processus de la science de la donnée (O’NEIL, SCHUTT, 2013)

« *La trace se repère et s’interprète en fonction des processus d’interactions entre un individu en situation d’interprétation et les environnements (individuels, familiaux, culturels, sociaux et plus généralement humains et non humains) dans lesquels elle s’insère* » (GALINON-MÉLÉNEC, 2011). Ainsi, la trace ne peut être interprétée par un seul spécialiste. Le résultat

de l'interprétation doit être l'intersection de plusieurs disciplines entrant en jeu dans le cadre d'un contexte donné pour un individu avec la prise en compte des normes²⁶⁰ s'appliquant dans chaque domaine.

La déliaison de la trace numérique de son contexte de production et de son auteur, en particulier dans le domaine de la santé, est sujette à des erreurs d'interprétation. Non seulement les analyses s'opèrent dans une méthode exploratoire et elles sont faites par des algorithmes dont le mécanisme de production des résultats est difficilement explicable. Les *signes-traces* perdent ainsi leurs caractéristiques propres (individualité de l'objet d'analyse, caractère indirect du déchiffrement et caractère conjectural).

Dans le cadre du *big data*, les données issues des pratiques médicales sont couplées aux données des objets connectés (dont la fiabilité et les contextes mesures ne sont pas toujours définis) et aux données des réseaux sociaux. La qualité des données, leur fiabilité et le mode d'accessibilité avant le processus d'analyse et d'interprétation pose des problèmes juridiques (protection des données personnelles, consentement explicite, anonymisation des données...) et surtout éthiques (confidentialité des données, respect du secret médical...).

Comme nous l'avons précisé plus haut, l'écriture qu'elle soit digitale ou non, est un système de notation attaché à l'activité humaine. De ce fait, il existe des traces écrites et des traces non écrites (DERRIDA, 1979). Toutes les traces n'étant pas écrites, l'interprétation des données de santé à partir des traces écrites du médecin est parcellaire et ne permet pas dans ce cas une prise de décision dans le cas de l'individu d'autant plus que le médecin n'est pas tenu de noter toutes les informations dans le dossier médical du patient²⁶¹. Il se pose alors la question de la complétude des données.

En effet, le médecin, dans le cadre du processus de diagnostic, utilise non seulement les différentes catégories de *signes-traces* du corps du patient mais aussi les traces numériques qu'il met en jeu dans un mécanisme de raisonnement et d'interprétation. Le médecin effectue

²⁶⁰ Une norme, du latin *norma* « équerre, règle », désigne un état habituellement répandu, moyen, considéré le plus souvent comme une règle à suivre. Ce terme générique désigne un ensemble de caractéristiques décrivant un objet, un être, qui peut être virtuel ou non. Dans le domaine médical, des auteurs (Georges Canguilhem, Michel Foucault) se sont posés la question sur la norme.

²⁶¹ « Indépendamment du dossier médical prévu par la loi, le médecin tient pour chaque patient une fiche d'observation qui lui est personnelle ; cette fiche est confidentielle et comporte les éléments actualisés, nécessaires aux décisions diagnostiques et thérapeutiques. Les notes personnelles du médecin ne sont ni transmissibles ni accessibles au patient et aux tiers. Dans tous les cas, ces documents sont conservés sous la responsabilité du médecin. Article 45 du code de déontologie médicale » (article R.4127-45 du code de la santé publique).

cette démarche sans nécessairement se rendre compte qu'elle est impactée par sa différence subjective (*signes-traces* cognitifs, ouverture sensorielle et empathique, etc.). Quand c'est le cas, il est - dans les cas difficiles - plus enclin à demander l'avis de ses confrères.

L'interprétation de la trace numérique telle que conçue ramène la notion de trace à la notion de donnée sur sa base technique. Prise isolément, elle n'a guère de sens. Mais regroupée, traitée et combinée dans d'importantes bases de données, elle peut révéler des informations significatives, stratégiques ou sensibles ; mais « *la donnée est elle-même le résultat d'analyses massives en quelque sorte. L'objet est un résultat et non une évidence* » Arnaud Cachia (CACHIA ET AL., 2015). Son mode d'interprétation occulte ainsi la complexité de la notion de trace (JEANNERET, 2011). De ce fait, nous pouvons dire que l'interprétation de la trace numérique est très réductionniste et nous ramène aux raisonnements cartésiens²⁶². On assiste également à une dépersonnalisation des données par le regroupement vers une personnalisation de l'interprétation après les processus de traitement imposant ainsi les caractéristiques moyennées d'un ensemble de population à l'individu.

L'utilisation traces numériques de santé dans le *big data* et les outils de l'IA est confrontée à des problèmes de fiabilité, de complétude, d'accessibilité et le mode d'accessibilité qui seront à l'origine de biais et d'erreurs lors du processus d'interprétation et d'analyse. La modélisation des données de la réalité est réductionniste et entraîne un appauvrissement voire une déformation de cette réalité ; par ailleurs, elle ne prend pas en compte l'hétérogénéité irréductible et la complexité du domaine de la santé. Elle pose alors le problème de la validité et de la liaison des données à la réalité des individus. Ainsi, l'interprétation traces numériques de santé dans le cadre du *big data* et son utilisation en IA comporte de nombreux enjeux dont il est important d'évoquer.

²⁶² Descartes ,dans « *le discours de la méthode* », définit quatre préceptes pour le développement de connaissance basée sur la raison que sont : ne recevoir aucune chose pour vraie tant que son esprit ne l'aura clairement et distinctement assimilée préalablement ; diviser chacune des difficultés afin de mieux les examiner et les résoudre ; établir un ordre de pensées, en commençant par les objets les plus simples jusqu'aux plus complexes et divers, et ainsi de les retenir toutes et en ordre, passer toutes les choses en revue afin de ne rien omettre. La pensée cartésienne ne prend pas en compte le caractère holistique des problèmes dans la recherche de solution et la complexité de l'humanité avec ses différentes interactions.

5.6.4 *Enjeux de l'interprétation des traces dans le domaine de la santé à l'ère du big data et de l'IA.*

L'analyse des données répond à des questions posées ou des problèmes à résoudre. Par conséquent, l'analyse est adaptée à la spécificité du problème et est guidée par une personne. Les décisions de prédiction et d'interprétation dépendront donc de la compréhension de l'individu sur la question. En plus, dans le cadre de l'analyse exploratoire des données comme le montre la figure 51, il n'y a pas de question de départ, les données sont directement utilisées avec leurs limites constitutives pour créer des produits ou prendre des décisions :

« avec le big data, on se passe du schéma rationaliste classique de l'hypothèse aux données. En un sens, on peut parler de science de la découverte par comparaison à une science de l'hypothèse » Arnaud Cachia (CACHIA ET AL., 2015). L'hypothèse de recherche n'est plus la source de l'investigation, elle est l'ingrédient d'ajustement théorique à la suite de l'exploration des données.

Dans le domaine de la santé, l'interprétation des signes cliniques prend en compte l'évolution des signes cliniques dans le temps, l'effet des thérapeutiques antérieures et présentes administrées et l'interaction avec l'environnement du patient. Cette prise en compte de la temporalité dans l'interprétation permet de poser le diagnostic de la maladie, définir le pronostic évolutif et proposer une thérapeutique en fonction du patient (COLLOC, BONENFANT, 2017). La médecine compose avec un objet très complexe - l'Homme - dans un environnement tout aussi complexe que les modèles et les probabilités mathématiques ont du mal à mettre en œuvre. En effet, la maladie est un fait biologique et un événement biographique. *« Si la personne est la finalité, elle incite à penser la personne comme un objet, une chose »*²⁶³. Nous devons aborder la question de la santé dans une pensée holistique²⁶⁴ et complexe (LE MOIGNE, 1999 ; MORIN, 2015).

L'interprétation des données de santé issues des traces numériques ne comporte pas les trois caractéristiques - l'individualité de l'objet, le caractère indirect du déchiffrement et le

²⁶³ Propos de Jean François Mattei lors d'une émission intitulée « qu'est-ce que l'éthique » organisée par l'espace éthique méditerranéen de Marseille en octobre 2005. Source : <https://www.youtube.com/watch?v=siYk3pNINiM> consulté le 21 décembre 2016.

²⁶⁴Holistique vient du grec, « *holos* » qui signifie « *totalité, l'entier* ». L'approche holistique consiste donc à prendre en compte la personne dans sa globalité (avec une prise en compte des facteurs physiques, psychologiques, sociaux, environnementaux, économiques etc...) plutôt que de la considérer de manière morcelée dans une approche centrée sur un organe, une fonction ou un symptôme d'une maladie.

caractère conjectural - du paradigme indiciaire décrites plus haut. Elle procède par une agrégation des données et un déchiffrement à partir des valeurs moyennées.

La médecine est à la fois une science et un art. La science s'occupe de ce qui peut être observé et mesuré et un médecin compétent reconnaît les signes d'une maladie et sait comment rétablir une bonne santé. Mais la médecine scientifique a ses limites (WILLIAMS, 2015). L'art de la médecine suppose l'application de la science et de la technologie médicale aux patients, aux familles et aux communautés.

« La plupart des différences entre les individus, les familles et les communautés ne sont pas, pour l'essentiel, physiologiques, et c'est dans la reconnaissance et la prise en compte de ces différences que les arts, les humanités, les sciences sociales jouent, avec l'éthique, un rôle majeur » (WILLIAMS, 2015).

La consultation médicale est un jeu de rôle entre le médecin et son patient ; elle se déroule par l'intermédiaire d'une communication multicanale (modulations de la voix, la mimique, les gestes, le regard) (HOUTAUD, 1999). Le diagnostic médical se construit à partir de la lecture des *signes-traces* par le médecin et à l'aide des outils de diagnostic technologiques. Cette vision a été réaffirmée par le Conseil National de l'Ordre des Médecins (CNOM) :

« Nous pensons que la médecine comportera toujours une part essentielle de relations humaines, quelle que soit la spécialité, et ne pourra jamais s'en remettre aveuglément à des « décisions » prises par des algorithmes dénués de nuances, de compassion et d'empathie. Les algorithmes et l'intelligence artificielle seront nos alliés, comme un apport essentiel pour l'aide à la décision et à la stratégie thérapeutique, ce qu'aucun médecin ne peut rejeter » (CONSEIL NATIONAL DE L'ORDRE DES MÉDECINS, 2018).

5.6.5 Définition de la maladie

Pour le moment, il n'existe pas une définition de la maladie à notre connaissance. Sa définition se fait en fonction de son opposé qu'est la santé. La santé est définie par l'Organisation Mondiale de la Santé (OMS) en 1946 comme : *« la santé est un état de complet bien-être physique, mental et social, et ne consiste pas seulement en une absence de maladie ou d'infirmité »*. En se référant à cette définition de la santé, on se rend compte de la complexité de la notion de santé. Elle est à la fois subjective, objective et globale, contextuelle et multidimensionnelle. En effet, la santé est le produit d'une interaction

complexe d'un individu avec son environnement, son patrimoine génétique, son milieu social et culturel. La santé n'est pas l'affaire des médecins seulement, il implique l'action de plusieurs acteurs à tous les niveaux y compris le patient ; elle ne se résume pas non plus à des signes cliniques et des valeurs objectives, mais prend en compte une part importante de subjectivité et enfin elle doit concilier une vie sociale et économique de qualité.

La maladie est socialement construite avant qu'elle ne soit une maladie en tant que telle dans le domaine médical. Si le patient ou la société ne définit pas un état comme pathologique, il ne consultera pas un médecin pour sa prise en charge et même si le médecin intervient, les recommandations ne seront pas suivies. Cet état de fait explique l'échec de nombreux programmes de santé. Ainsi, la prise en compte de cette construction sociale est très importante pour le diagnostic, la prise en charge et surtout le respect des recommandations par le patient et la société.

« Est « malade », pour le médecin, celui qui doit être traité, pour l'assureur, celui qu'il doit rembourser, pour la société, celui à qui on doit indulgence et secours, et pour les malades eux-mêmes, celui qui est entravé dans son action habituelle et ne peut faire ce qu'il veut ou doit faire » (LEMOINE, 2017).

Cette définition de Lemoine de la maladie montre une fois de plus l'importance de la prise en compte de tous les acteurs pour une meilleure prise en charge de la maladie nous ramenant ainsi à la problématique de la définition de la maladie à l'heure du *big data* et de l'IA.

La définition de la maladie à l'heure du *big data* et de l'IA se fera en fonction de valeurs agrégées et moyennées qui seront utilisées comme des normes²⁶⁵ en ne prenant pas en compte la complexité de la maladie dans son ensemble mais de manière réductionniste sur des aspects simplifiés. Ainsi, ce n'est ni le malade, ni la société, ni le médecin qui définira la maladie, mais les organisations dirigeantes (financiers, géants de l'Internet²⁶⁶ et Etats). En France, les médecins dans les hôpitaux le ressentent déjà avec la gestion entrepreneuriale des hôpitaux et de la santé. Cette vision normalisante de la maladie mettra en cause l'indépendance du médecin, l'autonomie du patient et imposera des normes à la société qui

²⁶⁵ La norme est à l'origine d'exigences et de coercitions qui s'exercent dans les domaines dans lesquels elle s'applique. Elle est imposée par un rapport de force en fonction des intérêts en jeu dans les différentes catégories (FOUCAULT, 2015).

²⁶⁶ C'est le cas de l'application de surveillance des cas de maladie à coronavirus développée par Alibaba. Le traitement des données pour les autorisations d'accès sont faites par les algorithmes développés par le groupe.

ne correspondent pas à leur vision de la maladie et aboutira ainsi à des injustices et à une dégradation de la qualité des soins. Ces craintes sont plus marquées d'autant plus que l'on sait que cette mise en norme de la maladie se fera dans une vision plus économique qu'humaniste²⁶⁷ ; il nous suffit de voir l'écosystème des financements de l'IA et du *big data* dans le domaine de la santé²⁶⁸ pour comprendre que les enjeux sont très importants et nombreux.

Dans la pratique clinique, la maladie n'est pas un fait universel, mais représente un problème clinique particulier par un individu à un temps donné (LEMOINE, 2017). La maladie doit être abordée sur le plan de la singularité du malade. C'est une expérience subjective vécue ou spontanée. Par exemple, un patient hypertendu n'aura pas le même comportement ou la même appréhension qu'un patient normotendu face à l'accident vasculaire cérébral. C'est tout son mode de vie et son interaction avec l'environnement qui change. La maladie affecte non seulement son organisme, mais également son milieu de vie. Pour Canguilhem (CANGUILHEM, 2013), l'individu se définit par sa capacité à créer des normes qui l'individualise ; « *être normal, c'est pouvoir s'adapter au changement du milieu* ». Les aspects sociaux et environnementaux de la normalité ne doivent pas être occultés dans l'appréhension de la maladie ; les maladies doivent être adaptées à la société, il s'agit là d'une des principales limites de l'IA et du *big data*. Le début de la maladie constitue une nouvelle vie avec des nouvelles normes pour le malade. Même s'il est complètement guéri de cette maladie, il garde en lui les *signes-traces* et *conséquences-traces* de cette maladie. Même s'il lui semble que la maladie n'a laisser aucune trace, le corps ne peut retrouver exactement dans l'état d'avant la maladie. Dans sa vie quotidienne et en société, l'effet de cette maladie se ressent dans ses activités et ses interactions. C'est une des raisons principales d'aider le malade à se réintégrer dans son environnement sans chercher à le ramener à des normes antérieures impossibles à retrouver pour lui mais en adaptant l'environnement et les normes à son nouvel état.

²⁶⁷ La gestion économique de la santé en France a débuté avec la mise en place du PMSI et la tarification à l'acte dans les structures de soins.

²⁶⁸ Le rapport de *Startup Health* montrait que les startups étaient parvenues à lever 13 Md€ en 2018, soit une augmentation de 25 % par rapport à l'année précédente. Les investissements dans les entreprises d'e-santé représentent deux tiers du total des sommes investies dans l'ensemble des *startups*. Source : <https://www.bpifrance.fr/A-la-une/Actualites/Le-marche-de-l-e-sante-porte-par-l-investissement-et-l-IA-46282> consulté le 10 décembre 2019.

En, plus la temporalité est un élément important dans la maladie. En effet les concepts médicaux et leur contexte évoluent, la conception des phénomènes pathologiques et les normes sociales aussi ; or cette évolution n'est pas toujours visible dans toute la société en même temps, mais dans la singularité des individus et des sociétés. Cela pose ainsi le problème de la généralisation des techniques de l'IA et du *big data* dans la prise en charge des maladies.

5.6.6 Médecine personnalisée

La pratique de la médecine depuis Hippocrate a toujours été personnalisée et individualisée. En effet, Hippocrate considérait que, pour prendre en charge de manière optimale son patient, le médecin devait prendre en compte toutes les informations le concernant dans le cadre du soin (son environnement, son mode de vie et ses préférences). En France, la prise de décision médicale est partagée entre l'équipe de soin et le patient en prenant en compte ses préférences, sa culture, son environnement et surtout les moyens mis à disposition pour la prise en charge. Cette prise en compte de la singularité du patient est mise en avant dans la relation médecin - patient. Il se pose alors la question du « *concept de médecine personnalisée* ». Est-ce une nouvelle révolution de la prise en charge du patient ?

Définition de la médecine personnalisée

« *La médecine personnalisée, ou médecine de précision, vise à traiter chaque personne de façon individualisée en fonction de ses spécificités génétiques et physiologiques ainsi que des caractères particuliers de sa maladie* » (CHAN, GINSBURG, 2011). Elle repose sur les caractéristiques moléculaires de l'individu établies à partir des techniques de cartographie à grande échelle avec l'aide du séquençage génomique. La personne humaine est limitée à ses données moléculaires et physiologiques qui sont analysées pour déterminer sa spécificité pour le prendre en charge. Dans ce contexte, il s'agit plutôt d'une médecine technique et scientifique basée sur la production de grande quantité de données, d'analyses statistiques et bio-informatiques avec l'aide du *big data* et de l'IA.

Ce terme relègue les pratiques médicales antérieures dans le domaine de l'imprécision où la décision thérapeutique est tout un art plus qu'une science (BILLAUD, GUCHET, 2015). Le problème est que la communauté scientifique ne s'accorde pas sur la définition de la *médecine personnalisée* ; elle ne s'entend pas davantage à propos de son appellation,

puisqu'on la nomme aussi *médecine de précision, ou santé personnalisée* (BILLAUD, GUCHET, 2015) ; en effet, nous avons d'une part ceux dont la définition de la *médecine personnalisée* est basée sur la biologie moléculaire, et, d'autre part, ceux dont la définition de la *médecine personnalisée* est centrée sur le patient.

La *médecine personnalisée* a été utilisée dans un premier temps par la pharmacologie génétique par l'intermédiaire du séquençage du génome humain ; De nos jours, elle inclut les capteurs, les objets connectés, les appareils d'auto-surveillance (*quantified self*) (LANGRETH, WALDHOLZ, 1999 ; PAULINE ET AL., 2009).

La définition et la délimitation de son contenu posent des problèmes aux acteurs, rendant ambigu leur positionnement. Ayant déjà abordé les avantages et les limites de l'interprétation des traces plus haut, la suite de nos propos sur la *médecine personnalisée* sera axée sur la définition qui s'appuie sur la biologie moléculaire.

Pratique de la médecine personnalisée

La médecine personnalisée fonde ses approches diagnostiques et thérapeutiques sur le profil moléculaire de la maladie à traiter ou du patient dans le cadre de l'évaluation du risque. Cette évolution de la médecine est basée sur les avancées importantes dans le domaine de la biologie et de la biomédecine. Elle a pris un tournant décisif avec la capacité de séquençage moléculaire et la réduction importante des coûts d'acquisition des appareils. « *La médecine personnalisée ne cherche plus à identifier seulement les perturbations moléculaires des cellules dues à la maladie, mais cherche à anticiper et prévenir l'apparition de la maladie* » (BILLAUD, GUCHET, 2015).

La médecine personnalisée pour nous s'inscrit dans la continuité de la pratique médicale et ne constitue ni un concept nouveau, ni une révolution dans la pratique médicale. L'anticipation et la prévention de la maladie ont toujours été considérées dans la prise en charge des patients pour empêcher l'apparition de la maladie elle-même ou des complications. Elle a comme base la biologie moléculaire ; or l'utilisation de la biologie moléculaire dans le diagnostic où le traitement des maladies n'est pas nouvelle dans la pratique de la médecine (MORANGE, 2003 ; TAMBOURIN, 2005 ; NELSON ET AL., 2013).

L'histoire de la médecine est jalonnée de ruptures intellectuelles remettant chaque fois en cause la manière de penser ou de pratiquer la médecine. Dans la naissance de la clinique

(FOUCAULT, 2015), Michel Foucault identifie trois grandes classes de la pratique médicale qui ont évolué dans le temps et toujours d'actualité ; il s'agit de :

- *la médecine des espèces ou médecine classificatrice* : les maladies sont classées par famille. Elle décrit la maladie en se fondant sur les signes cliniques, et non sur les causes ; elle s'intéresse au visible. La maladie s'exprime à travers le corps et permet de la classer. La médecine des espèces ne s'intéresse pas au processus évolutif de la maladie ; elle s'intéresse à la description et à l'analogie de la maladie. Dans cette médecine, le patient est l'élément par lequel la maladie s'exprime ; la prise en compte de la description du patient brouille la maladie ;
- *la médecine des épidémies* : il n'y a pas de différence entre la maladie individuelle et la maladie collective ; il s'agit d'une analyse de santé publique (une vision collective) avec une quantification et une définition de la distribution géographique de la maladie. La prise en charge de la maladie est basée sur des contraintes physiques et sociales pour maîtriser la maladie. Cette médecine est à l'origine de l'institutionnalisation de la médecine avec l'Etat comme régulateur et une extension du rôle de la médecine au mode de vie des individus.
- *la médecine clinique et anatomopathologique* : elle se développe avec la formation théorique et pratique au sein de l'hôpital et partage d'expérience avec un médecin expérimenté. Ainsi, « *la médecine ne se définit pas comme clinique sans se définir comme savoir multiple de la nature et connaissance de l'homme en société* » (FOUCAULT, 2015). La médecine clinique a pour principe la recherche de la cause de la maladie en tenant compte de tous les aspects à savoir l'homme dans son environnement, le caractère évolutif et variable de la maladie et la diversité de ses manifestations. Elle accorde une grande importance à la question du temps dans l'évolution de la maladie.

L'anatomopathologie a contribué à redéfinir la médecine clinique en rendant l'invisible visible. Les symptômes de la maladie sont expliqués par les lésions des tissus et des cellules au niveau de l'organisme grâce au développement du microscope. L'anatomopathologie a permis un ancrage de la maladie dans le corps.

La biologie moléculaire, sur laquelle se fonde la médecine personnalisée, va plus loin ; elle va au-delà de la cellule et du tissu pour rechercher le siège de la lésion observé au niveau de la molécule au cours de la maladie ou dans le cadre de la prévention. Ainsi, pour nous, elle

n'est autre qu'un prolongement de la médecine clinique et de l'anatomopathologie à travers la biologie moléculaire. Comme l'anatomopathologie, elle renforce l'ancrage de la maladie dans le corps à travers l'étude des modifications moléculaires. Elle doit être intégrée de ce fait dans le cadre du processus de soins dans la pratique clinique pour faire avancer la prise en charge des patients.

Considérée comme une révolution scientifique dans la prise en charge des patients et se suffisant à elle-même, la médecine personnalisée devient alors une médecine de classification ne s'intéressant qu'à la lésion moléculaire pour classifier la maladie tout en oubliant le patient, ou une médecine des épidémies servant à classifier la population dans des cadres définis afin de prendre des décisions pour la collectivité dans le cadre de la santé publique ; dans les deux cas, elle ignore non seulement l'individu mais lui enlève tout pouvoir de décision (non-respect de l'autonomie du patient) au profit des institutions (Etat, entreprises privées). Nous sommes face à une utilisation de la santé des individus comme levier de pouvoir sur eux, le bio pouvoir de Foucault (GENEL, 2004). Une utilisation de la médecine personnalisée hors de la médecine clinique aura un impact important sur le système de santé avec un risque de discrimination, d'inégalité d'accès aux soins et un contrôle de la population au travers de leurs données par l'imposition de normes de comportement et de vie, les privant ainsi de leur liberté de choisir d'autant plus que « *la pratique de la médecine personnalisée est de séquencer, stratifier et cibler* » (TOPOL, 2014 ; ARNEDOS ET AL., 2014 ; DE BONO, ASHWORTH, 2010 ; CHAN, GINSBURG, 2011).

« Dans nos travaux, nous sommes par exemple amenés à considérer un échantillon de 10000 individus pour 600000 variables (SNPs). Pour ces variants²⁶⁹, les données brutes représentent 50 Go (gigaoctets) et les données traitées après filtrage de l'information sont de l'ordre de 2 Go » (GRENIER-BOLEY, 2015).

Ainsi en ne travaillant pas sur l'intégralité des données mais sur des données traitées, on perd forcément des informations et on choisit volontairement d'ignorer certaines informations. Ces craintes sont d'autant plus importantes que les financements aussi bien privés que publique dans le domaine sont considérables. Des nouveaux acteurs jusque-là loin du domaine de la santé investissent des sommes importantes pour avoir accès aux données

²⁶⁹ Le variant génétique est une modification apparue spontanément dans un gène. Elle peut avoir, ou pas, une incidence sur le fonctionnement cellulaire et l'apparence (phénotype) de la personne. Elle rentre donc dans la définition générale d'une mutation mais sans notion de maladie ou de trouble induits. Source: <https://www.federationdesdiabetiques.org/information/diabete/dico-du-diabete/variant-genetique> consulté le 21 mai 2020.

de santé. L'objectif des laboratoires pharmaceutiques est d'augmenter la cible des patients non malades qui sont plus rentables que la prise en charge des vrais malades ; l'enjeu ultime pour eux est leur intérêt économique plutôt que le bien-être des peuples. Ainsi, le prix Nobel britannique de médecine, Richard J. Roberts déclarait :

« les médicaments qui guérissent ne sont pas rentables et ne sont donc pas développés par les laboratoires pharmaceutiques, qui parallèlement développent des médicaments qui chronicisent les maladies et qui doivent être consommés de manière ininterrompue »²⁷⁰.

Ainsi, l'un des problèmes de la médecine personnalisée est la gestion des données et des algorithmes servant à la prise de décision.

Gestion des données et des algorithmes

La protection des données médicales dans le cadre du séquençage génomique est un enjeu crucial. Au-delà de l'individu, les données génomiques concernent également la famille de la personne. Ce sont des données qui comportent les traits particuliers d'une famille et propre à chaque individu ; elles se suffisent à elles seules pour l'identification des personnes. Si elles sont anonymisées, elles n'ont aucune valeur dans le cadre de son utilisation pour le diagnostic et le traitement. Ainsi, protéger les données de séquençage génomique revient à renoncer à l'apport de cette technique dans la prise en charge des patients. Il se pose alors la question éthique de la confidentialité des données et du secret médical. Cette question est d'autant prégnante qu'il existe des bases de données génomiques *open access* ou payante. La question de la gouvernance des données de génomique se pose avec acuité et nécessite une réaction des Etats pour protéger la population et trouver des moyens surs de partage des données dans le cadre de la recherche et la mise en place de solutions innovantes dans le domaine de la santé.

Mais, renoncer aux avantages de la médecine personnalisée (thérapie ciblée des cancers, diagnostic précoce des maladies génétiques) ne serait pas raisonnable non plus, d'où la nécessité de la positionner de façon claire dans le processus de soin en tant qu'examen complémentaire permettant aux médecins de l'utiliser de manière précise en fonction des

²⁷⁰Un prix Nobel de médecine accuse: « guérir des maladies n'est pas rentable pour les laboratoires ». Source : <https://reseauinternational.net/un-prix-nobel-de-medecine-accuse-guerir-des-maladies-nest-pas-rentable-pour-les-laboratoires/> consulté le 27 novembre 2019.

connaissances scientifiques actuelles pour le diagnostic et le traitement des patients. Il devra s'aider de ce fait des algorithmes pour l'interprétation des grandes quantités de données.

Les algorithmes prédictifs s'appuient sur des études scientifiques en utilisant les résultats de ces études pour tirer des probabilités et des règles de calcul que les algorithmes prendront en compte dans les équations. Ces algorithmes ne déterminent pas si le patient aura nécessairement la maladie. Il indique uniquement les risques de développer cette pathologie. Or, pour qu'il soit valide, il faut que la population étudiée soit représentative et importante. Mais une des difficultés des statistiques est de passer du niveau de la population au niveau individuel. Car les probabilités sont constituées à partir de moyennes et que la moyenne ne fonctionne pas forcément à l'échelle individuelle. Dans le cas d'une maladie où le facteur génétique ne joue qu'un rôle mineur, la médecine prédictive n'aura pas d'intérêt. Il faudra donc définir les maladies concernées et faire une gradation de la fiabilité de ces résultats (SCHUTZ, 2018).

Dans certains cas, les maladies sont multicausales et / ou multifactorielles : soit génétiques, soit environnementales, soit une combinaison des deux. L'algorithme doit prendre en compte ces deux types de facteurs, génétiques et environnementaux, et les signes cliniques (avec la prise en compte de leur évolution dans le temps) pour établir son résultat, qui sera une probabilité, donc un risque de développer cette maladie. Ces facteurs évoluent énormément à travers le monde et à travers le temps et influencent de manière importante l'évolution de la maladie sans compter les capacités adaptatives de l'humain en fonction de l'environnement. Dans le cas de la génétique, seule une petite partie des mutations causales de maladies monogéniques sont connues et validées sur les milliers voire les millions d'autres variants, actuellement méconnues (SCHUTZ, 2018).

La technologie a été et sera d'un apport dans la pratique médicale. Elle a été à l'origine d'une amélioration nette de la prise en charge des maladies, mais elle ne devra pas oublier que la médecine est un art qui s'aide de la science et de la technologie pour atteindre ses objectifs. Ainsi, la médecine personnalisée basée sur la biologie moléculaire permettra d'étayer la classification causale ou multifactorielle des maladies et d'apporter une amélioration importante dans la prise en charge des maladies et de constituer de ce fait un progrès dans la pratique médicale. Elle doit donc être intégrée dans le processus de raisonnement médical

en fonction de sa contribution. Son utilisation ne devrait pas occulter la relation de soin entre le patient et son médecin.

6 Evaluation de l'utilisation des technologies en pratique médicale

6.1 Description du corpus

Nous avons enquêté les médecins de la ville du Havre. Nous avons envoyé une lettre d'information à l'Ordre des Médecins (*Annexe 2*) de la région Seine-Maritime pour expliquer les objectifs de l'étude et utiliser le canal d'information de l'ordre pour informer les médecins. Nous avons reçu leur réponse avec des recommandations en matière de protection des données (*Annexe 3*).

Pour Stéphane Beaud et Florence Weber : « ... *faire du terrain, c'est avoir envie de se connecter avec les faits, de discuter avec les enquêtées, de mieux comprendre les individus et les processus sociaux. Sans cette soif de découvrir, sans cette envie de savoir, presque « d'en découdre », le terrain devient une formalité, un exercice scolaire, plat, sans intérêt »* (BEAUD, WEBER, 2015).

Nous avons choisi l'entretien, une méthode de recherche et d'investigation à travers lequel, l'enquêteur cherche à obtenir des informations sur les attitudes, les comportements, les représentations d'un ou de plusieurs individus dans la société (TOUMERT, 2017).

Les entretiens compréhensifs semi directifs en face-à-face nous ont donné une plus grande marge de liberté aux enquêtés et nous ont permis de faire une analyse comparative des données recueillies. Il s'inscrit dans une démarche inductive où le recueil et l'analyse des informations ne sont pas des étapes successives, mais s'opèrent conjointement, au fur et à mesure des entretiens, en même temps de l'interprétation que l'élaboration de la problématique et des hypothèses. « *L'objectif est de parvenir à une compréhension intime de la pensée et de l'action des sujets »* (VAN CAMPENHOUDT, QUIVY, 2011).

En effet, Quivy et L. Van Campenhoudt, insistent sur le fait que l'entretien permet « *l'analyse du sens que les acteurs donnent à leurs pratiques et aux événements auxquels ils sont confrontés : leurs systèmes de valeurs, leurs repères normatifs, leurs interprétations de situations conflictuelles ou non, leurs lectures de leurs propres expériences »* (VAN CAMPENHOUDT, QUIVY, 2011)

6.2 Statuts des personnes enquêtées

Notre échantillon était constitué de 25 personnes réparties comme suit :

- 21 médecins praticiens ;
- un agent de la direction de services informatiques d'un hôpital ;
- un agent (médecin) du département d'informatique médicale ;
- un chef d'entreprise dans le développement d'applications informatiques en santé et médecin ;
- un juriste, spécialiste du droit de la robotique et des TIC dans le domaine de la santé.

L'analyse quantitative et qualitative a concerné les 21 médecins. Une analyse de contenu a été réalisée dans les autres cas.

6.3 Méthodes utilisées pour l'analyse des données

6.3.1 *Analyse qualitative*

Nous avons utilisé le logiciel Iramuteq pour l'analyse statistique textuelle des données. C'est un logiciel qui permet une analyse statistique textuelle. Mais il faut noter que l'analyse qualitative est un exercice intellectuel et concret qui doit s'adapter à la réalité du terrain. Bien que l'analyse automatisée permette un gain de temps et une classification des données, il doit être complété par une analyse classique de contenu pour faire émerger le sens des discours. Nous avons complété l'analyse statistique textuelle par une analyse de contenu.

« L'analyse de contenu est un ensemble d'instruments méthodologiques de plus en plus raffinés et en constante amélioration s'appliquant à des « discours » extrêmement diversifiés et fondé sur la déduction ainsi que l'inférence. Il s'agit d'un effort d'interprétation qui se balance entre deux pôles, d'une part, la rigueur de l'objectivité, et, d'autre part, la fécondité de la subjectivité »
(BARDIN, 2013).

Le logiciel Iramuteq est un logiciel open source qui se base sur la lexicométrie²⁷¹ pour la statistique textuelle. Les outils numériques et le développement des capacités de calcul des ordinateurs ont permis une diffusion de cette méthode. Il permet de faire de l'analyse textuelle simple, l'analyse factorielle des correspondances (AFC), l'analyse de similitude et la classification hiérarchique descendante (CHD) ou la *méthode Reinert*.

Le document est sous forme de texte au format UTF8. Iramuteq fait une segmentation²⁷² du texte tous les 40 mots tout en respectant la ponctuation et une lemmatisation²⁷³ lors du traitement statistique à l'aide des dictionnaires de mots intégrés dans le logiciel.

L'analyse textuelle simple propose des statistiques simples sur les corpus de texte : effectif de toutes les formes, effectif des formes actives et supplémentaires, liste des hapax²⁷⁴.

L'analyse de similitude est une technique, reposant sur la théorie des graphes, classiquement utilisée pour décrire des représentations sociales, sur la base de questionnaires d'enquête (VERGÈS, BOURICHE, 2001 ; MARCHAND, RATINAUD, 2012). L'objectif de l'analyse de similitude est d'étudier la proximité et les relations entre les éléments d'un ensemble, sous forme d'arbres maximum : le nombre de liens entre deux items évoluant « *comme le carré du nombre de sommets* » ; elle cherche à réduire le nombre de ces liens pour aboutir à « *un graphe connexe et sans cycle* » (MARCHAND, RATINAUD, 2012).

L'AFC va transformer statistiquement les données et les organiser sous forme de graphique à deux dimensions, montrant les relations lexicales entre elles. Elle va permettre de voir la différence entre les classes ou les groupes de mots. Cette analyse nous a permis de lier et de hiérarchiser les informations des textes. Ce type d'analyse peut être effectué sur des classes de la CHD, soit sur une variable ou sur certaines modalités de variables.

²⁷¹ « *La lexicométrie peut être définie comme un « ensemble de méthodes permettant d'opérer des réorganisations formelles de la séquence textuelle et des analyses statistiques sur le vocabulaire d'un corpus* » (LEBART, SALEM, 1994).

²⁷² Les segments de texte sont construits à partir d'un critère de taille et de ponctuation. Iramuteq cherche le meilleur ratio taille/ponctuation. L'objectif est d'avoir des segments de tailles homogènes en respectant le plus possible la structure du langage.

²⁷³ La lemmatisation est l'obtention des racines des mots. Les occurrences sont réduites à leur racines ; les verbes sont ramenés à l'infinitif, les noms au singulier et les adjectifs au masculin singulier.

²⁷⁴ Les hapax sont des mots qui apparaissent une seule fois dans le texte.

La CHD ou la *méthode Reinert* (REINERT, 1987) se présente sous forme de dendrogramme. L'objectif est de regrouper des « mondes lexicaux » et de mettre en évidence les thématiques générales du corpus. Cette méthode a été mise au point par Max Reinert, elle cherche à rendre compte de l'ordre interne d'un discours et à mettre en évidence ses mondes lexicaux. Elle découle de l'AFC. La *méthode Reinert* va découper le texte en segment, puis observer la distribution des formes pleines dans chaque segment. Elle va ensuite regrouper les segments en classes en fonction des mots qui les composent par un processus d'itération.

6.3.2 *Analyse quantitative*

En plus de l'analyse qualitative, nous avons réalisé une analyse quantitative pour déterminer des fréquences et des moyennes dans notre travail. Cette analyse quantitative s'est faite à partir des données quantitatives recueillies pendant les entretiens et à partir de la catégorisation lors de l'analyse de contenu. Cette analyse catégorielle a permis de quantifier les données qualitatives recueillies pendant les entretiens. L'analyse quantitative a été réalisée avec le logiciel R. Les variables continues ont été exprimées sous forme de moyenne \pm écart type et les variables catégorielles sous forme de pourcentage.

Les variables catégorielles qui ont été cotées à partir d'une échelle de Likert ont été regroupées en deux catégories par la suite afin de réaliser les graphes. Les pourcentages des deux valeurs positives ont été additionnés pour avoir la part de ces différentes réponses.

Pour l'analyse du score unitaire éthique, nous avons attribué un nombre de points aux différents attributs de l'échelle (*4 points pour tout à fait d'accord, 3 points pour d'accord, 2 points pour neutre, 1 point pour pas d'accord et 0 point pour pas du tout d'accord*). Le questionnaire pour l'évaluation de la finalité éthique des données dans le cadre du *big data* était constitué de 25 questions réparties comme suit :

- neuf questions pour la bienfaisance ;
- cinq questions pour la justice ;
- six questions pour la non-malfaisance ;
- cinq questions pour l'autonomie.

Le score éthique unitaire a été calculé selon la formule suivante :

$$\text{Score éthique} = \frac{\text{Total de nombre de points des questions} * 100}{4 * \text{nombre de questions}}$$

Ainsi, nous avons calculé le score unitaire éthique global et le score unitaire éthique spécifique aux quatre principes de bioéthique. Les scores vont de 0 à 100 points ; zéro pour un désaccord pour l'atteinte des finalités éthiques lors de l'utilisation des données dans le cadre du *big data* et 100 pour un accord.

Les graphes ont été réalisés à l'aide du logiciel Calc de la suite Open Office.

6.4 Caractéristiques socio – démographiques des médecins

6.4.1 Sexe et âge

Nous avons enquêté 21 médecins dont cinq de sexe féminin et 16 de sexe masculin.

L'âge moyen des médecins était de 52 ± 10 ans avec des extrêmes de 33 et 72 ans.

L'âge moyen des femmes était de 46 ± 10 ans avec des extrêmes de 35 et 61 ans et celui des hommes était de 53 ± 10 ans avec des extrêmes de 33 et 72 ans.

6.4.2 Spécialité

Les médecins étaient des spécialistes de la médecine générale dans 14 cas. Deux médecins généralistes avaient une double spécialité (expertise judiciaire et médecine du sport).

Le tableau ci-dessous montre la répartition des médecins en fonction de leur spécialité.

Tableau 11 : répartition des médecins en fonction de leur spécialité

| | Fréquence |
|---|-----------|
| Médecine générale | 14 |
| Neurologie | 01 |
| Anesthésie - Réanimation | 01 |
| Médecine du travail | 01 |
| Médecine d'urgence | 01 |
| Endocrinologie – Diabétologie – Nutrition | 01 |
| Chirurgie | 01 |
| Dermatologie | 01 |
| Total | 21 |

6.4.3 Mode et ancienneté dans la pratique médicale

La pratique médicale était faite sous trois modes dans notre échantillon de médecins et répartie comme suit :

- la pratique de groupe : 12 médecins ;
- la pratique individuelle : sept médecins ;

- la pratique hospitalière : deux médecins.

Le nombre moyen d'année de pratique était de 22 ± 11 ans avec des extrêmes de 1 et 42 ans.

Nous avons subdivisé nos praticiens en trois groupes en fonction de l'histoire de l'informatisation du système de santé en France ces 50 dernières années²⁷⁵. Treize médecins avaient un nombre d'année de pratique médicale supérieur à 20 ans.

Tableau 12 : répartition des médecins en fonction de leur ancienneté de pratique

| | Fréquence | Pourcentage |
|------------------------|-----------|-------------|
| < = 10 ans | 03 | 14 |
| > 10 ans et < = 20 ans | 05 | 24 |
| > 20 ans | 13 | 62 |
| Total | 21 | 100 |

6.5 Caractéristiques globales du corpus de texte des entretiens

Notre corpus de texte était constitué de 21 textes d'entretiens, 761 segments de textes, 27995 occurrences²⁷⁶, 973 hapax (3,5 % des occurrences) et 2123 formes.

Le nombre moyen d'occurrence par texte était de 1333.

L'analyse statistique de notre corpus montre un graphique dont la courbe est globalement sous la forme d'une droite descendante selon la loi de Zipf²⁷⁷.

²⁷⁵ Des années 60 aux années 90 : l'informatisation des fonctions administratives par des structures publiques. Il s'agissait dans un premier temps de système centralisé avec la création dite d'applications dites « nationales » puis la création de structures régionales d'informatiques hospitalières. Des années 90 aux années 2000 : l'informatisation des plateaux médicotechniques et l'ouverture du marché avec une structuration et un développement de l'offre logicielle privée. Des années 2000 à aujourd'hui : l'informatisation de la production de soins et les plans nationaux avec la complexité que l'on connaît avec la multiplicité des acteurs. (ANAP, 2014).

²⁷⁶Le nombre d'occurrences correspond au nombre de mots.

²⁷⁷ la loi de Zipf prévoit que dans un texte donné, la fréquence d'occurrence $f(n)$ d'un mot est liée à son rang n dans l'ordre des fréquences par une loi de la forme $f(n) = \frac{K}{n}$ où K est une constante.

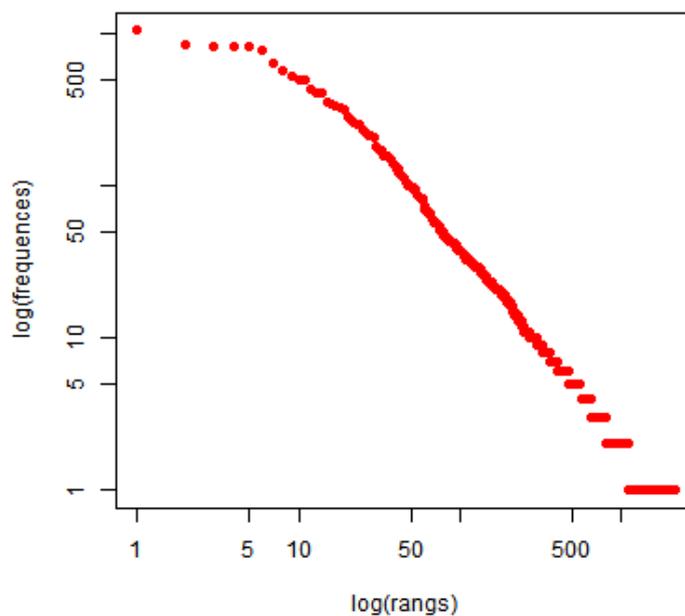


Figure 68 : graphique logarithmique des rangs et des fréquences

La forme active la plus utilisée était le mot « *patient* », utilisé 409 fois, suivi des mots « *donnée* », utilisé 213 fois et « *médecin* », utilisé 179 fois. La figure 69 montre la répartition des mots utilisés sous forme active dans le corpus de texte en fonction de leur fréquence et leur type.

| Forme | Freq. ↓ | Types |
|------------------|---------|-------|
| patient | 409 | nom |
| donnée | 213 | nom |
| médecin | 179 | nom |
| médical | 158 | adj |
| aller | 150 | ver |
| utilisation | 117 | nom |
| prendre | 116 | ver |
| utiliser | 107 | ver |
| santé | 104 | nom |
| information | 100 | nom |
| cadre | 97 | nom |
| objets_connectes | 85 | nr |
| big_data | 83 | nr |
| pratique | 79 | adj |
| voir | 73 | ver |
| charger | 68 | ver |
| mettre | 66 | ver |
| exemple | 65 | nom |
| permettre | 63 | ver |
| temps | 58 | nom |
| maladie | 57 | nom |
| site | 55 | nom |
| artificiel | 51 | adj |
| intelligence | 51 | nom |
| cas | 47 | nom |
| problème | 47 | nom |
| web | 45 | nom |
| médecine | 44 | nom |
| secret | 43 | nom |
| améliorer | 42 | ver |
| partager | 42 | ver |
| recherche | 42 | nom |
| risque | 42 | nom |
| dépense | 41 | nom |

Figure 69 : fréquence et type des mots utilisés sous forme active dans le corpus de texte

6.5.1 *Analyse de similitude*

Pour notre analyse de similitude, nous avons choisi les mots qui avaient une fréquence d'occurrence supérieure ou égale à 20. Ce choix donnait une liste de 91 mots inclus dans le graphe ci-dessous.

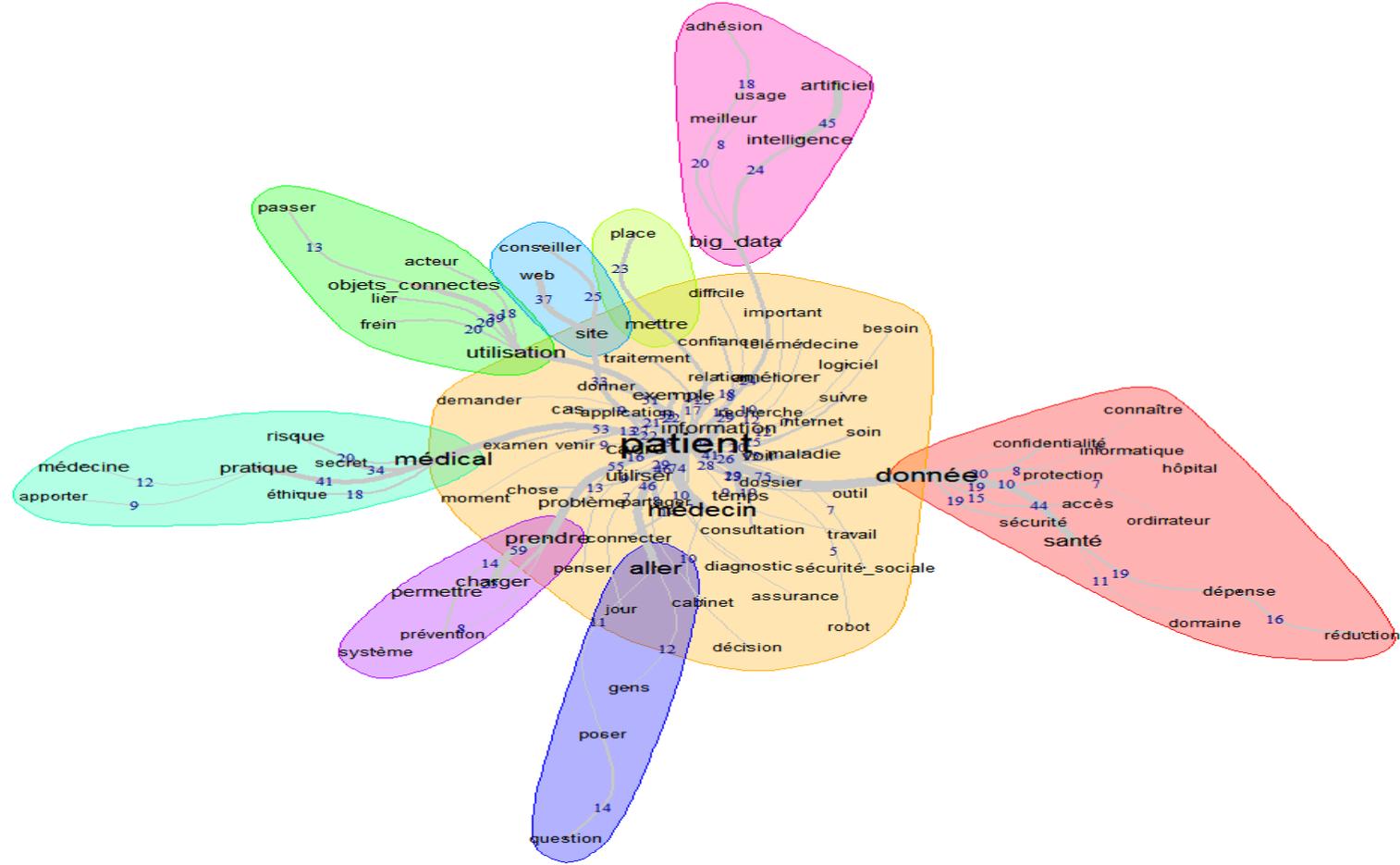


Figure 70 : graphe de similitude - arbre maximum - du corpus

Nous avons obtenu un graphe étoilé avec le mot « patient » au centre du graphe. Il existait une forte relation entre les mots « médecin », « donnée » et « patient ».

Le mot « donnée » avait une relation avec les mots « confidentialité », « protection », « sécurité », « accès » et « santé ». Cette liaison portait sur la gestion des données de santé d'une manière générale. Cette relation entre le mot « donnée » et « patient » s'explique par l'importance du respect du secret professionnel et la protection des données de santé qui sont la pierre angulaire de l'éthique médicale.

Le mot « patient » était directement lié au mot « big data », « objets connectés » et « site web ». Cependant, on notait une absence relation directe de ces mots avec le mot « médecin ». Le choix des outils était de la responsabilité du médecin ; mais la décision d'utilisation dépendait du patient. La relation médecin-patient reste importante dans l'utilisation des TIC dans la pratique médicale dans cette analyse.

Nous pouvons à partir de cette analyse de similitude que le graphe montre des îlots de groupe qu'étaient :

- la gestion des données de santé ;
- l'utilisation des objets connectés, site web et applications dans la pratique médicale ;
- l'utilisation du *big data* et de l'IA dans la pratique médicale ;
- l'impact des TIC sur la pratique médicale.

Le patient est au centre de ces îlots. Le médecin jouait le rôle de médiation entre le patient et les TIC dans cette analyse de similitude.

6.5.2 Classification hiérarchique descendante (CHD) selon la méthode Reinert

La CHD selon la *méthode Reinert* permet de classer les formes dans des classes de formes regroupées selon leur indépendance mesurée par un test au Chi². L'algorithme de la CHD selon la *méthode Reinert* a permis de mettre en évidence cinq classes et avec la classification de 82 % des segments de texte ; démontrant d'une bonne qualité de l'analyse.

La CHD divisait les classes en deux sous-ensembles :

- un sous-ensemble 1 avec la classe 5 et représentant 19,8 % des segments classés ;

- un sous-ensemble 2 avec les classes 1, 2, 3, 4 et représentant 80,1 % des segments classés. Le sous-ensemble 2 est subdivisé en deux avec un regroupement des classes 1 et 4 et un regroupement des classes 2 et 3.

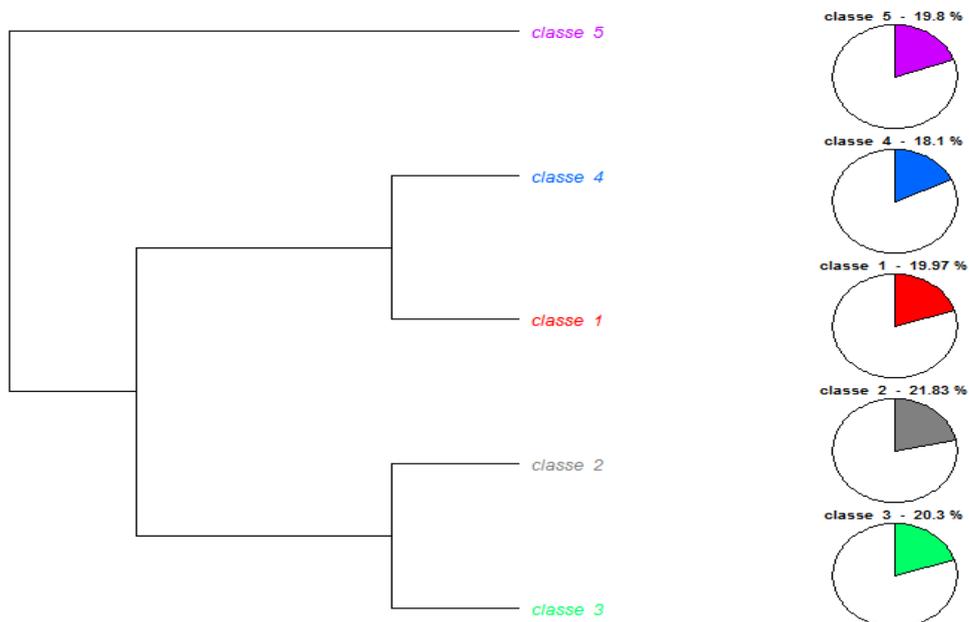


Figure 71 : dendrogramme montrant les différentes classes (CHD)

Ces cinq classes ont été classées à partir des segments de mots du corpus de texte qui permettaient de catégoriser chaque classe de façon indépendante.

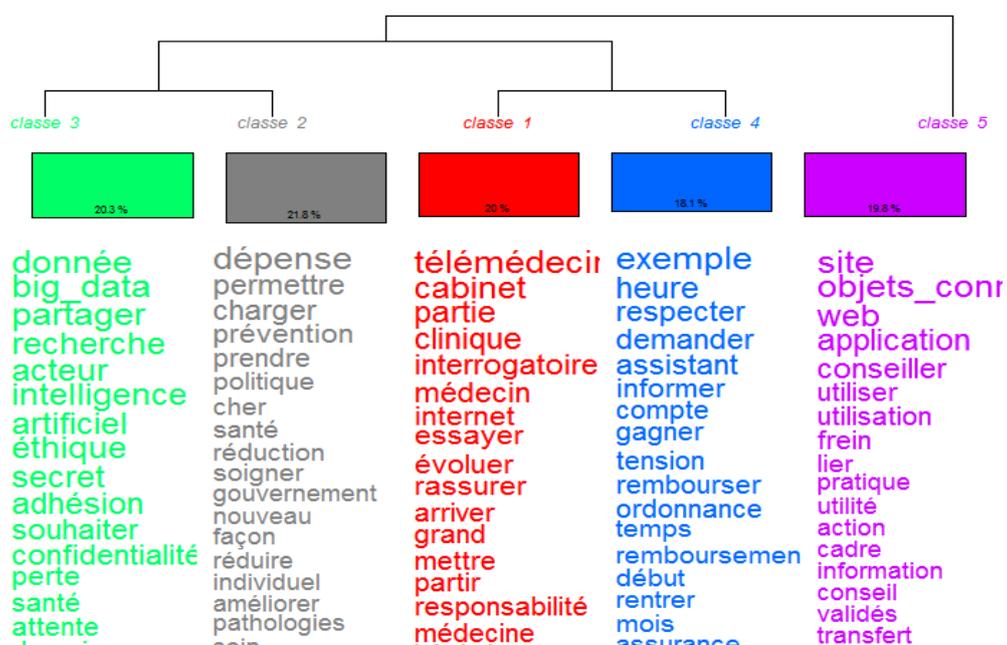


Figure 72 : CHD avec les mots associés aux différentes classes

Ainsi l'analyse des segments de mots et du corpus de texte des différentes classes nous a permis de proposer un thème aux différents sous corpus :

- la classe 1 était associée aux mots « *télémedecine* » ($chi2 = 37$), « *cabinet* » ($chi2 = 32$), « *clinique* » ($chi2 = 27$), « *interrogatoire* » ($chi2 = 24$), « *médecin* » ($chi2 = 24$), « *Internet* » ($chi2 = 22$), « *responsabilité* » ($chi2 = 15$). La classe des mots était liée à « *l'impact des TIC sur la pratique médicale* ». Elle concernait la place du médecin dans la relation des soins à l'ère des TIC, l'impact des TIC lors de la consultation médicale en particulier le cas de la télémedecine, la relation du médecin avec les autres prestataires de soins et la responsabilité médicale ;
- la classe 2 : les mots étaient liés à « *l'impact du big data et de l'IA sur la politique de santé* ». Les mots associés à cette classe étaient « *dépense* » ($chi2 = 114$), « *prévention* » ($chi2 = 65$), « *politique* » ($chi2 = 41$), « *santé* » ($chi2 = 35$), « *gouvernement* » ($chi2 = 25$), « *améliorer* » ($chi2 = 20$), « *réduction* » ($chi2 = 31$). Cette classe concernait l'intérêt du *big data* et de l'IA en termes d'attente et de promesses d'amélioration dans la pratique médicale, la gestion des dépenses de santé et les coûts liés à l'utilisation du *big data* et de l'IA sur le système de santé et le rôle du gouvernement dans l'élaboration des politiques de santé ;
- la classe 3 concernait les risques d'éthique médicale liés à l'utilisation du *big data* et de l'IA dans la pratique médicale, la gestion et la gouvernance des données de santé, du DMP et les conditions pour une meilleure adhésion des acteurs à l'utilisation du *big data* et de l'IA. Les mots associés à cette classe étaient « *donnée* » ($chi2 = 65$), « *big data* » ($chi2 = 65$), « *partager* » ($chi2 = 61$), « *secret* » ($chi2 = 57$), « *éthique* » ($chi2 = 53$), « *confidentialité* » ($chi2 = 41$), « *consentement* » ($chi2 = 20$). Cette classe des mots était liée à « *l'utilisation du big data et de l'IA dans la pratique médicale* » ;
- la classe 4 était associée aux mots « *exemple* » ($chi2 = 51$), « *assistant* » ($chi2 = 31$), « *heure* » ($chi2 = 37$), « *respecter* » ($chi2 = 32$), « *temps* » ($chi2 = 24$), « *gagner* » ($chi2 = 20$), « *responsabilité* » ($chi2 = 15$). La classe des mots était liée aux « *assistants médicaux et exemples* ». Elle concernait le rôle que pourrait jouer les assistants médicaux dans la pratique médicale avec les contraintes associées pour les médecins et les exemples utilisés pour étayer les propos des médecins lors des entretiens ;
- la classe 5 : les mots étaient liés aux « *objets connectés, sites web et applications* ».

Les mots associés à cette classe étaient « *site web* » ($chi2 = 189$), « *objets connectés* » ($chi2 = 175$), « *applications* » ($chi2 = 139$), « *utilisation* » ($chi2 = 85$), « *frein* » ($chi2 = 72$), « *utilité* » ($chi2 = 41$), « *action* » ($chi2 = 35$). Cette classe concernait l'utilisation, le degré de confiance, les freins à l'utilisation des objets connectés, sites web et applications et les actions pour une meilleure adhésion des acteurs de la santé.

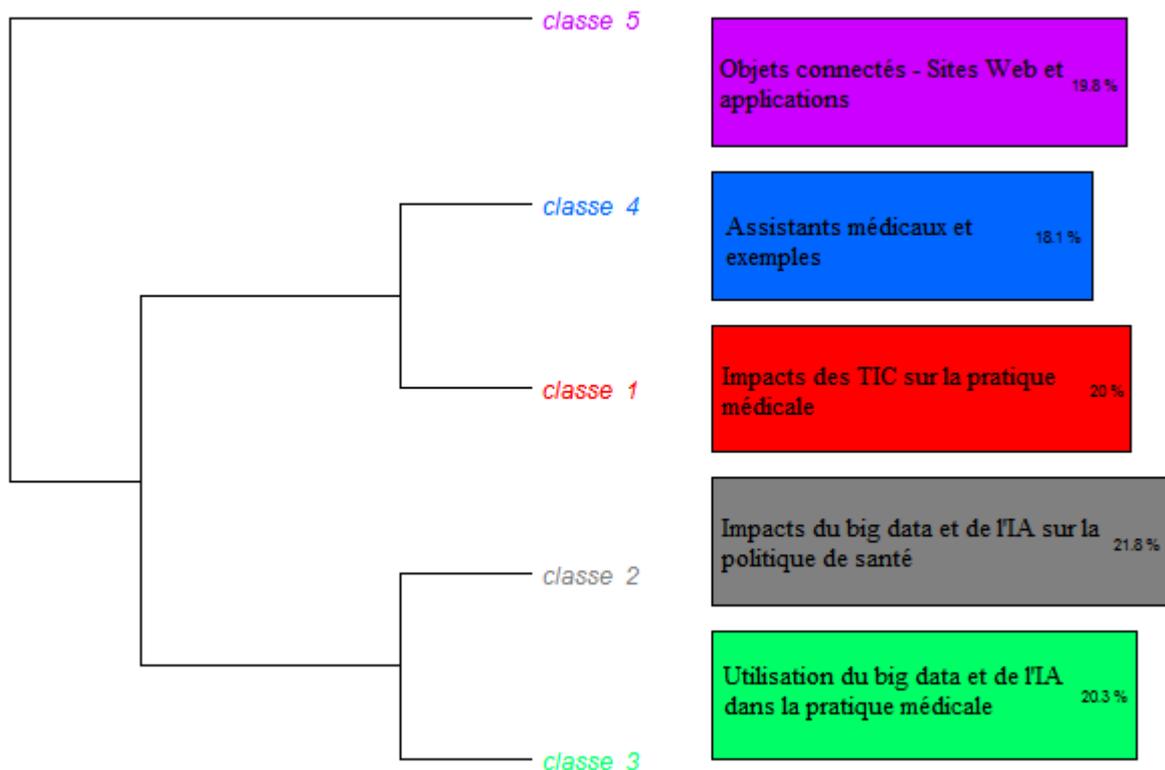


Figure 73 : CHD avec les différentes thématiques des classes

Les thèmes 1 et 2 (figure 74) correspondaient respectivement aux classes 1 et 4 de la CHD que nous avons nommé « *pratique médicale à l'ère de l'informatique* » (sous ensemble 1 figure 74). Les thèmes 3 et 4 (figure 74) correspondant aux classes 2 et 3 ont été nommés « *big data et IA en santé* » (sous ensemble 2 figure 74). Les deux sous-ensembles formaient l'ensemble 1 correspondant à « *pratique médicale, big data et intelligence artificielle* ».

L'ensemble 2 correspondait à la classe 5 et nommé « *objets connectés, sites et applications* ».

Les ensembles 1 et 2 correspondaient à l'ensemble de notre corpus de texte que nous avons nommé « *TIC et santé* ».

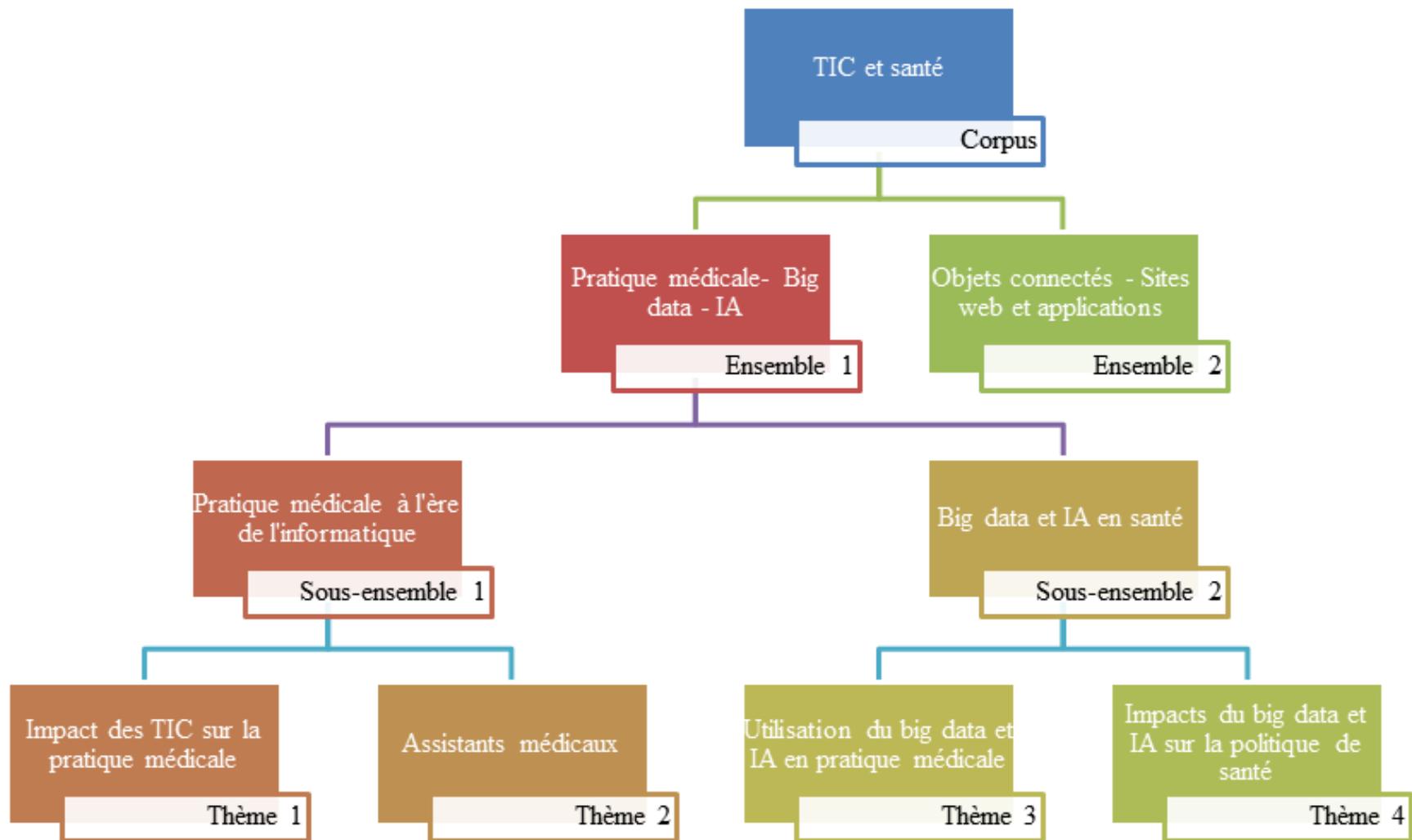
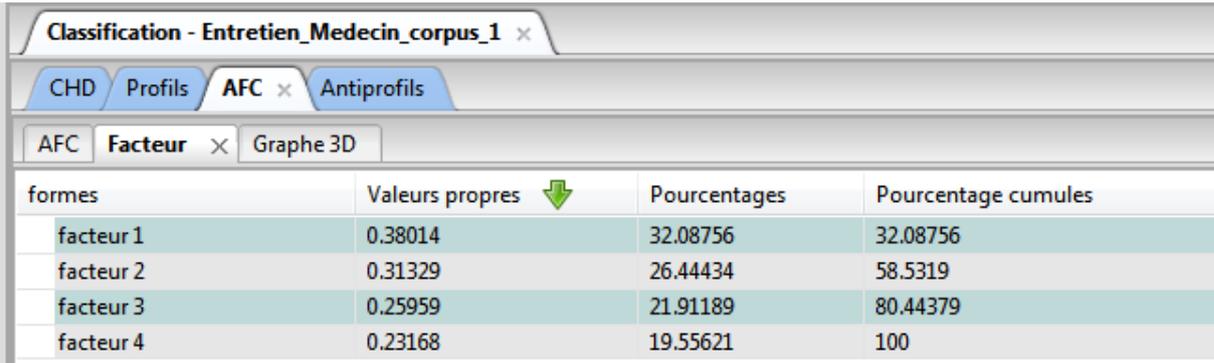


Figure 74 : classification thématique du corpus à partir de la CHD

6.5.3 Analyse factorielle des correspondances (AFC)

L'AFC de notre corpus de texte a permis de mettre en évidence quatre facteurs. Ces facteurs pouvaient être couplés en fonction de deux axes pour expliquer les résultats. Dans notre analyse, les facteurs 1 et 2 expliquaient 58,5 % des résultats.



| formes | Valeurs propres | Pourcentages | Pourcentage cumules |
|-----------|-----------------|--------------|---------------------|
| facteur 1 | 0.38014 | 32.08756 | 32.08756 |
| facteur 2 | 0.31329 | 26.44434 | 58.5319 |
| facteur 3 | 0.25959 | 21.91189 | 80.44379 |
| facteur 4 | 0.23168 | 19.55621 | 100 |

Figure 75 : pourcentage des différents facteurs de l'AFC

Par rapport à l'axe 1 (facteur 1), nous avons un regroupement des classes 2 et 3 qui s'expliquait par leur point commun qui est « *le big data et l'IA dans le domaine de la santé* » comme le soulignait le titre du sous-ensemble 2 de la figure 74. Ces deux classes étaient opposées à la classe 5 « *objets connectés, sites et applications* » (ensemble 2 de la figure 74). Cette opposition s'expliquait par le mode d'acquisition et de traitement des données. Nous avons choisi de nommer cet axe « *Gestion et gouvernance des données de santé numérique* ».

Au niveau de l'axe 2 (facteur 2), nous avons les classes 1 et 4 qui étaient groupées et traitaient de la « *pratique médicale à l'ère des TIC* ». Elles étaient opposées aux classes 2, 3 et 5 qui portaient sur les « *TIC utilisés dans le domaine de la santé* ». Nous avons choisi de nommer cet axe « *TIC et pratique médicale* ».

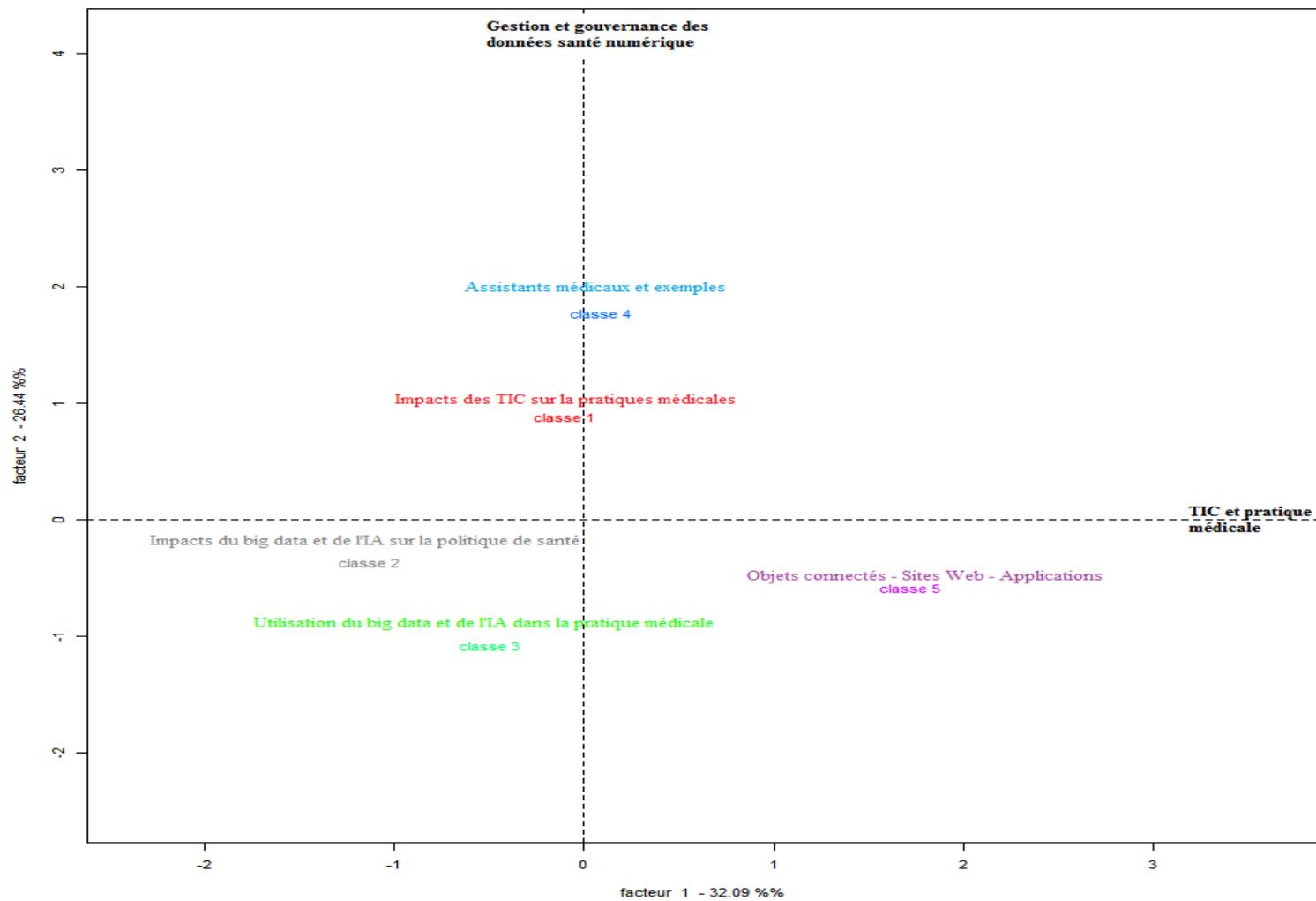


Figure 76 : AFC des classes de la CHD

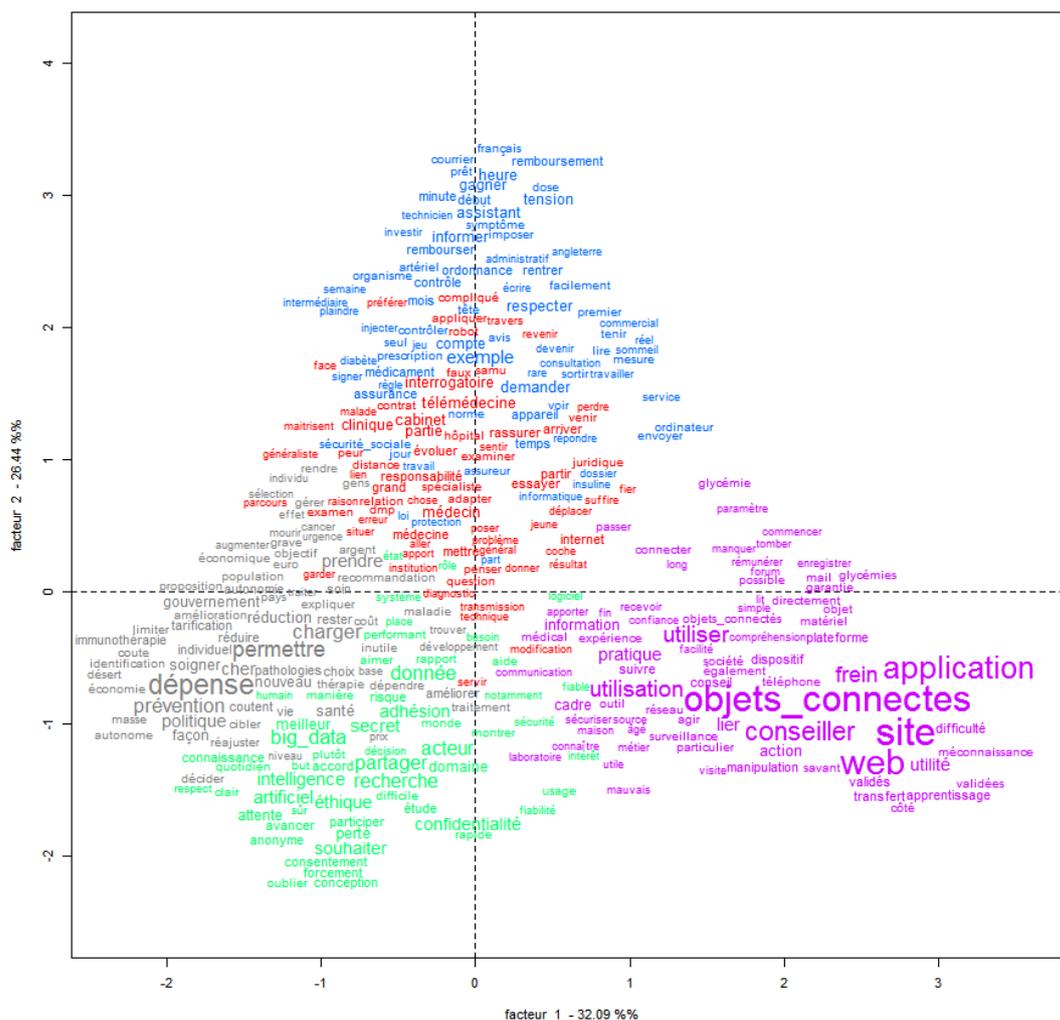


Figure 77 : AFC des mots associés aux classes de la CHD

6.6 Utilisation des technologies de l'information et de la communication dans le cadre de la pratique médicale.

6.6.1 *Utilisation du dossier patient informatisé*

Vingt médecins sur les 21 enquêtés utilisaient un dossier patient informatisé (DPI). Le seul médecin qui n'utilisait pas de DPI, le faisait pour des raisons personnelles, départ de la retraite proche et manque de confiance à la technologie.

L'utilisation du DPI est une réalité dans la pratique médicale quel que soit l'âge du médecin. Cette utilisation s'expliquerait par plusieurs raisons : la facilité de recherche de l'information

et du suivi pour un patient donné, les fonctions de rappels et de détection des interactions médicamenteuses dans les prescriptions et l'amélioration de la qualité des soins.

L'informatisation des soins s'est accélérée avec la publication des ordonnances Juppé²⁷⁸ qui ont imposé aux professionnels de santé d'être en mesure de télétransmettre les feuilles de soins avant le 31 décembre 1998 et la mise en place par le comité de gestion d'un fonds de réorientation et de modernisation de la médecine libérale (FORMMEL) a permis de mettre en place des mesures d'aide à l'informatisation des médecins. A côté des contraintes législatives, la dématérialisation des processus de soins a contribué également à l'utilisation du DPI : « *J'ai exercé la médecine pendant quarante-deux et je suis informatisé que depuis treize ans. Je me suis mis en informatique parce que la tendance m'obligeait* »²⁷⁹ disait un médecin lors de nos entretiens.

L'utilisation du DPI permet également d'avoir toutes les informations du patient avec l'évolution dans le temps et un archivage facile des données des patients avec moins d'encombrement spatial ; les médecins ont l'obligation de conserver les dossiers des patients vingt ans après la dernière consultation²⁸⁰. Mais les médecins pensaient que l'utilisation du DPI parasitait la relation avec leur patient du fait de l'écran et la saisie des notes qui ne leurs permettait pas de se concentrer et d'écouter avec attention le patient.

La dématérialisation du dossier patient a entraîné avec elle, la question du stockage et de la conservation des données de santé. Ainsi, L'Etat a alors mis en place l'agrément des hébergeurs de données de santé qui a évolué vers une certification²⁸¹ en 2016 afin de permettre un stockage sécurisé mais surtout un partage facilité des données entre les médecins dans le cadre de la continuité des soins des patients. La certification n'est délivrée que si des conditions très strictes de confidentialité, de sécurisation, de disponibilité, sont remplies par les prestataires. Mais ces mesures semblent être peu utilisées par les médecins. En effet, seulement trois médecins hébergeaient leurs données de santé chez un hébergeur agréé de données de santé. Dans les autres cas (n = 18), les données étaient conservées au

²⁷⁸ Ordonnance no 96-346 du 24 avril 1996 portant réforme de l'hospitalisation publique et privée. JORF n°98 du 25 avril 1996 page 6324.

²⁷⁹ Entretien 3.

²⁸⁰ Source : <https://www.service-public.fr/particuliers/vosdroits/F12210> consulté le 25 avril 2019.

²⁸¹ L.1111-8 du code de la santé publique, modifié par la loi n° 2016-41 du 26 janvier 2016.

niveau local avec des sauvegardes régulières, sur des disques durs externes, faites par les médecins.

Les coûts d'hébergement des données de santé constituaient un frein à l'utilisation de ce mode d'hébergement surtout pour les praticiens libéraux. Dans les hôpitaux, en plus des coûts liés à l'hébergement, il faut tenir compte du personnel nécessaire pour assurer l'entretien et la sécurité des installations. Les services étant payés en fonction de l'utilisation et de l'espace occupée par les données, les coûts ne font qu'augmenter avec le temps du fait de la quantité des données produites chaque jour qui ne fait que croître et de l'obligation de conservation des données pendant une période relativement longue.

Encadré 6 : coût de la location d'un serveur de données de santé²⁸².

Pour un seul serveur, les tarifs peuvent varier de 300 à 1 000 € par mois, selon la complexité de la machine. S'il s'agit d'héberger les données de tout un hôpital, ça peut monter à plusieurs dizaines de milliers d'euros par mois : pour un hôpital qui a besoin de 40 machines héberger, ça fera 40 000 € et on ajoute tous les services associés. Les services associés, c'est-à-dire l'entretien, l'antivirus, la surveillance 24 heures/24, et le conseil adapté au client, peuvent faire doubler le tarif, ce qui peut nous amener à près de 100 000 € mensuels pour l'hébergement des données d'un grand hôpital. Certains médecins ne peuvent pas payer une telle somme, et le regrettent amèrement car cela freine leurs projets de recherche.

Le coût de l'hébergement sécurisé est donc un obstacle, mais dans les hôpitaux il y a d'autres problèmes. Même si les données sont placées dans des coffres-forts chez des hébergeurs agréés, il reste que l'hôpital souffre d'une carence de personnel qualifié en matière de sécurité informatique.

Par contre la mise en place de la messagerie sécurisée est très appréciée par les médecins dans le cadre de la transmission des informations entre professionnels, en témoigne son utilisation. En effet, 16 médecins l'utilisaient dans le cadre de leur échange avec les collègues et la transmission des informations et 15 médecins recevaient directement les résultats de laboratoires de leurs patients par télétransmission. Cette utilisation des moyens sécurisés de communication était associée à une confiance qu'accorder les médecins aux outils déployés à cet effet.

²⁸² Source : <https://www.techniques-hospitalieres.fr/blog/sante-ce-que-preparent-les-geants-du-numerique-n2216> consulté le 25 octobre 2018.

Néanmoins, certains médecins préféraient toujours utiliser les lettres d’information (cinq médecins qui n’utilisaient pas de messagerie sécurisée d’e-mails dans le cadre de la transmission des informations ou des échanges avec leurs collègues). Ils expliquaient cet état par le fait de l’habitude mais surtout une envie de maintenir le contact humain, la matérialité de la relation avec leurs collègues ; Pour eux, cela permettait de porter plus d’attention à leur message.

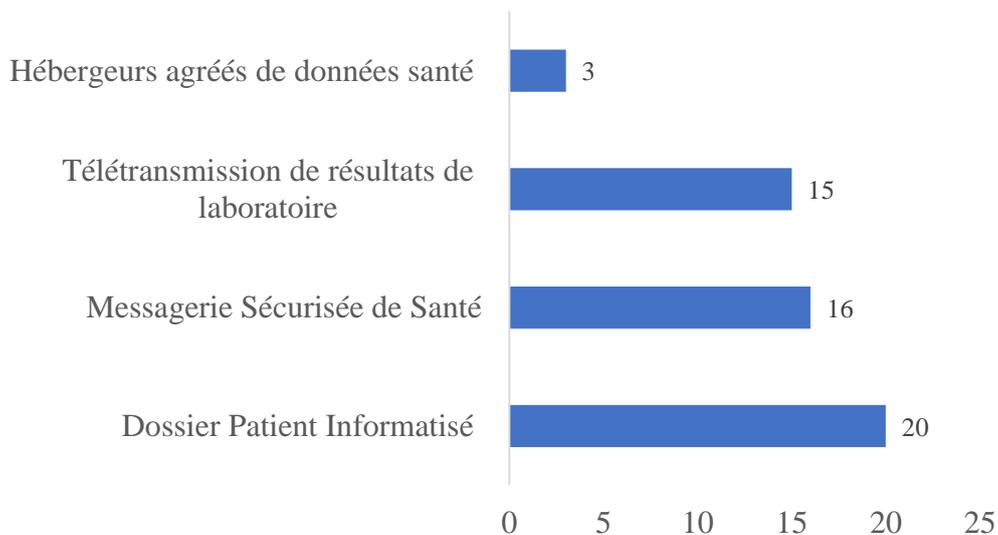


Figure 78 : utilisation des TIC par les médecins

6.6.2 Utilisation des objets connectés

Malgré la relative disponibilité des objets connectés dans la vie courante des citoyens, leur utilisation semble peu fréquente dans le domaine de la santé. Seulement sept médecins utilisaient des objets connectés dans le cadre de leur pratique médicale et quatre médecins parmi ceux-ci utilisaient des objets connectés validés dispositifs médicaux. En effet, il faut distinguer les objets connectés validés dispositifs médicaux des objets connectés dits de « *bien être* » ; mais le processus de labélisation n’est pas obligatoire, il dépend de la volonté du fabricant de faire la déclaration.

Le règlement UE 2017/745²⁸³ du Parlement européen et du conseil du 5 avril 2017 relatif aux dispositifs médicaux rend obligatoire la certification des dispositifs par les fabricants à

²⁸³ Règlement (UE) 2017/745 du Parlement européen et du Conseil du 5 avril 2017 relatif aux dispositifs médicaux, modifiant la directive 2001/83/CE, le règlement (CE) no 178/2002 et le règlement (CE) no 1223/2009 et abrogeant les directives du Conseil 90/385/CEE et 93/42/CEE (JO L 117 du 5.5.2017, p. 1-175).

partir du 26 mai 2020 ; il constitue de ce fait une avancée majeure dans le processus de labélisation des objets connectés.

Encadré 7 : définition d'un dispositif médical²⁸⁴

Selon la Directive Européenne 93/42CEE :

est considéré comme dispositif médical tout instrument, appareil, équipement, logiciel, matière ou autre article, utilisé seul ou en association, y compris le logiciel destiné par le fabricant à être utilisé spécifiquement à des fins diagnostique et/ou thérapeutique, et nécessaire au bon fonctionnement de celui-ci. Le dispositif médical est destiné par le fabricant à être utilisé chez l'homme à des fins de :

- diagnostic, prévention, contrôle, traitement ou d'atténuation d'une maladie ;
- diagnostic, contrôle, traitement, d'atténuation ou de compensation d'une blessure ou d'un handicap ;
- d'étude ou de remplacement ou modification de l'anatomie ou d'un processus physiologique ;
- maîtrise de la conception,

et dont l'action principale voulue dans ou sur le corps humain n'est pas obtenue par des moyens pharmacologiques ou immunologiques ni par métabolisme, mais dont la fonction peut être assistée par de tels moyens.

Les objets connectés dans le domaine du bien-être collectent des données qui concernent des données personnelles telles que l'âge, le sexe, la localisation et des paramètres médicaux tels que la tension artérielle, le poids, le nombre de pas effectués par jour qui sont indéniablement des données de santé. Il s'agit d'une manière détournée de collecter et de stocker des données sensibles sans être hors la loi. Ces objets connectés posent également le problème du lieu de stockage des données et leur gouvernance. Pourtant, la nouvelle définition des données de santé considère les données issues de ces objets comme des données sensibles et doivent bénéficier de la même garantie de protection que les données issues des dispositifs médicaux.

Dans le cadre de la pratique médicale, quand ils sont validés dispositifs médicaux, ils sont d'un apport important dans la prise en charge des patients de par le respect des normes

²⁸⁴ Directive 93/42/CEE du conseil du 14 juin 1993 relative aux dispositifs médicaux. JO L 169 du 12.7.1993, p. 1-43.

métiers et d'une grande fiabilité. En plus, leur utilisation n'engage pas la responsabilité du médecin tant que celui-ci respecte les indications décrites par le fabricant. Le « *marquage CE* » qui est la certification de l'objet comme dispositif médical et d'autorisation de mise sur le marché est un gage de qualité et d'assurance pour les médecins ; c'est un facteur qui favoriserait l'utilisation de ces objets par les médecins. Les objets connectés étaient utilisés dans le cadre de la surveillance et du suivi de la maladie par les médecins. La figure 79 présente la répartition de l'utilisation des objets connectés en fonction des indications.

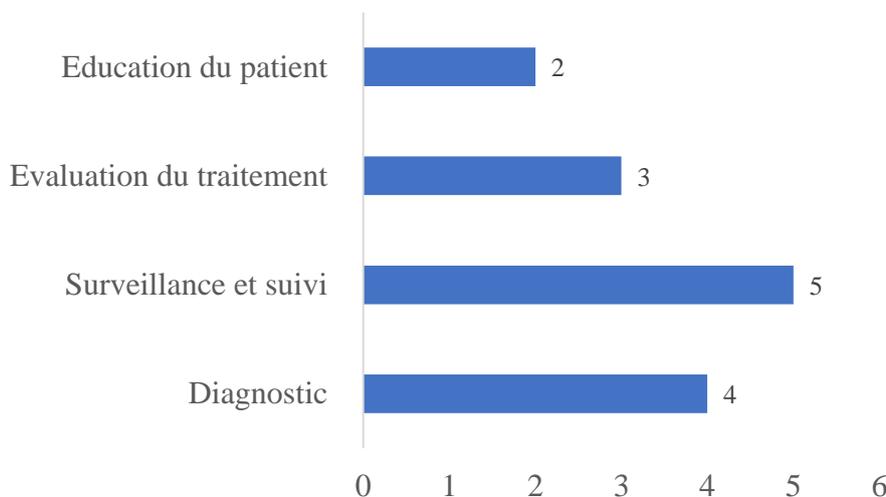


Figure 79 : utilisation des objets connectés en fonction des indications (n = 07)

Les médecins, dans le cadre de l'utilisation pour les soins de ces objets ont obligation de prescrire des objets avec le « *marquage CE* ». S'ils décident d'utiliser des objets sans le « *marquage CE* », leur responsabilité est engagée en cas de préjudice causé au patient. Mais au-delà de cette responsabilité, il s'agit d'une question éthique pour le médecin de ne pas nuire à son patient (non-malfaisance) par l'utilisation de procédés ou de techniques dont il n'est pas sûr de l'efficacité et de la validité et rappelée à l'article 39 du code de déontologie médicale « *les médecins ne peuvent proposer aux malades ou à leur entourage comme salutaire ou sans danger un remède ou un procédé illusoire ou insuffisamment éprouvé. Toute pratique de charlatanisme est interdite* ».

Dans une étude réalisée par une équipe du LIMICS²⁸⁵ (FARNIA ET AL., 2017), une évaluation d'objets connectés (pèse-personne, podomètre, oxymètre de pouls, tensiomètre) dans le cadre de la télésurveillance des patients insuffisants cardiaques a été réalisée selon six critères (validité médicale, fiabilité technique, facilité d'utilisation, ergonomie, respect des normes juridiques, la précision des mesures). Une note finale sur 20 était attribuée à

²⁸⁵ Le LIMICS (Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances en e-Santé) est une unité de recherche mono-équipe et interdisciplinaire en informatique et en informatique médicale.

chaque objet connecté. L'objet connecté devait avoir au moins une note de 15 pour être validé. A l'issue des analyses, aucun objet ne respectait les critères définis par les auteurs. En, effet, l'absence de validation scientifique et de « *marquage CE* » constituaient des freins à l'utilisation des objets connectés par les médecins lors de notre enquête de terrain.

Freins à l'utilisation des objets connectés

L'analyse de contenu des entretiens avec les médecins mettait en évidence trois principaux freins à l'utilisation des objets connectés dans la pratique médicale qui étaient :

- **le processus de conception des objets connectés**

Le respect des normes métiers est un élément important dans la conception des outils. Il permet une fiabilisation des outils et surtout une validation des paramètres mesurer par l'outil. Les médecins pensaient que l'association des spécialistes de la santé dans la conception permettrait de garantir la fiabilité et la qualité des objets.

« [...] la diffusion de l'information sur les objets connectés disponibles et leurs utilités auprès des médecins est très importante et en particulier les objets connectés et les applications validées par des sociétés savantes. Il y a beaucoup d'applications qui sont créées, mais on ne connaît pas les critères de qualité, qui a créé, les références et les recommandations utilisées par les concepteurs »²⁸⁶.

Cette procédure permet d'obtenir des données validées et de qualité à partir des objets connectés ; c'est une phase très importante du fait que les données doivent être utilisées au niveau individuel et au niveau de la population pour des prises de décision. Les conséquences seront importantes aussi bien pour les individus que le système de santé si des données erronées étaient amenées à être utilisées. L'absence de validation comme dispositifs médicaux met en doute la fiabilité des outils dans la prise en charge médicale. Mais surtout, aucun médecin ne voulait engager sa responsabilité pour des outils dont il ne connaissait pas les concepteurs.

La majorité des objets connectés ne sont pas interopérables avec les DPI. Les données devaient dans un premier temps être transférées sur l'ordinateur du médecin et dans un second temps être introduites dans le DPI. Il s'agit d'une perte de temps pour les médecins avec des risques d'erreurs sur les valeurs lors de la manipulation :

²⁸⁶ Entretien 1

« Il y'a également le problème de la compatibilité avec les logiciels médicaux (il faut prendre du temps pour saisir les informations) et le manque de temps à l'éducation du patient. Les logiciels de gestion médicaux constituent déjà un investissement et il est très difficile d'ajouter d'autres contraintes »²⁸⁷.

Avec le développement de l'informatisation de la santé, chaque acteur ne tient pas compte des fonctionnalités disponibles sur le marché et des recommandations d'interopérabilité²⁸⁸. Le manque d'interopérabilité rend difficile le partage des données. Mais, l'interopérabilité n'est pas uniquement technique, elle aussi juridique et métier. En effet, l'interopérabilité doit respecter les règles de protection juridiques des données personnelles, mais également, la confidentialité, la sécurité et le secret médical dans cadre de la pratique médicale.

- **la gestion des données**

Le consentement obligatoire des patients à céder les données avant l'utilisation des objets connectés met à mal l'autonomie du patient à faire un choix. Le consentement doit être libre, éclairé et explicite avec un droit de refus. Mais dans le cadre des objets connectés et des applications, l'utilisateur n'a pas le choix, soit il accepte les conditions d'utilisation précisées par le fabricant, il ne peut pas utiliser l'objet qu'il a pourtant acheté. Il s'agit d'un consentement par contrainte, posant la question sur l'autonomie des patients à gérer leurs données comme le stipule la loi.

A côté du consentement, il se pose la question du stockage, des accès, de la sécurité et de la confidentialité des données. Que ce soit le médecin ou le patient, aucun ne connaît le lieu de stockage réel des données, les utilisations qui sont faites et les différents organismes et /ou personnes qui ont accès aux données. Le risque dans ces cas est surtout la marchandisation des données de santé par les fabricants, la perte de la confidentialité et l'atteinte à la vie privée. Cela concerne l'utilisation des objets connectés comme moyens de surveillance des utilisateurs. Le partage de données des objets connectés se fait également de manière « invisible » et à l'insu des utilisateurs entre les concepteurs et des sociétés tierces. Ces risques sont d'autant plus importants que les objets connectés sont reliés à Internet,

²⁸⁷ Ibid.

²⁸⁸ L'interopérabilité est la capacité pour un système informatique d'utiliser les informations produites par un autre système informatique comme les informations produites par lui-même et de mettre à disposition des autres systèmes les informations qu'il a produites.

Bluetooth, wifi ou des interfaces *web client*. Ce sont des enjeux juridiques qui doivent être réglés afin de garantir une utilisation efficace et efficiente des outils.

Encadré 8: sécurité des objets connectés

Une étude de HP Fortify²⁸⁹, la division sécurité de Hewlett Packard, s'est intéressée à l'écosystème de l'Internet des objets au niveau de la sécurité. Dans cette étude, au moins de 250 vulnérabilités ont été découvertes dans dix des objets connectés les plus populaires à l'heure actuelle. Selon HP, sept d'entre eux ne chiffraient même pas les données transférées aux réseaux, huit ne demandaient même pas de mots de passe ou alors n'imposent aucune restriction quant à sa complexité et sa longueur. Neuf objets sur dix stockaient au moins une partie des données personnelles via le dispositif, le *cloud* ou l'application mobile. Soixante pour cent des objets connectés ne sécurisaient pas le téléchargement des mises à jour logicielles et étaient donc vulnérables aux attaques dites de « *cross-site scripting* » (type de faille de sécurité permettant d'injecter du contenu dans une page, et provoquer ainsi des actions sur les navigateurs web visitant la page).

Le défaut de prise en compte des exigences de la sécurité informatique dans la conception des systèmes existe dans la majorité des objets connectés et des applications. Les concepteurs privilégient la mise en place rapide sur le marché des produits qui n'intègrent pas toutes les mesures de sécurité pour des raisons de performance, de convivialité ou de coût. « *La sécurité coûte cher et personne, à moins qu'elle n'y soit contrainte, ne souhaite supporter les coûts de la sécurité informatique* » (BALAGUÉ ET AL., 2018).

Le secret médical et la confidentialité des données de santé sont un enjeu majeur pour le médecin et pour le patient dans le cadre de la relation de soin. L'intégrité, la disponibilité, la fiabilité, la confidentialité des données sont des éléments importants que ce soit dans le diagnostic, le suivi et la surveillance des patients ; ils sont non négociables dans le cadre de la sécurité des objets connectés utilisés dans le domaine de la santé. Les médecins préfèrent ne pas prendre de risque pour leurs patients tant qu'ils ne sont pas sûrs des outils : « *Je n'ai aucune confiance à l'informatique et dans les données qui circulent dans le web. Les*

²⁸⁹ Source : <https://www.developpez.com/actu/73797/Une-etude-de-HP-Fortify-fait-etat-d-un-nombre-important-de-failles-dans-la-securite-des-objets-connectes-et-interpelle-les-fabricants/> consulté le 05 mars 2019.

données d'une manière ou d'une autre vont se retrouver dans le circuit du marché »²⁹⁰, disait un médecin lors de nos entretiens. Il s'agit d'une crainte partagée par l'ensemble des médecins que nous avons interrogés.

- **les conditions d'utilisation :**

La perte de temps dans la manipulation des outils, l'absence d'utilité dans leur pratique et le parasitage de la relation médecin – patient ont été évoqués par les médecins comme des freins à l'utilisation des objets connectés. En effet, les objets connectés constituent un budget important pour son acquisition ; et acheter un objet jugé inutile n'intéressait pas les médecins.

« Les freins à l'utilisation des objets connectés sont liés au temps pris pour les manipuler ; la perte du temps, la question de la gestion des pannes et des bugs et c'est difficile d'avoir un dépanneur. Il nous faut un vrai service après-vente. A chaque fois que l'on paie les outils, on se débrouille nous-même pour les installer et quand il y a les pannes c'est très difficile d'avoir quelqu'un pour nous dépanner immédiatement. Déjà l'ordinateur me prend énormément de temps. Je n'ai pas besoin des objets connectés »²⁹¹

Les outils peuvent être des sources d'angoisse pour les patients à cause des fausses alertes (faux positifs) avec pour conséquence des consultations fréquentes sans intérêt et une perte de temps des médecins pour les rassurer. Cette perte de temps pour les médecins a été mise en évidence dans le travail de Nachoua Guizani (GUIZANI, 2016) sur la gestion de l'acheminement des messages dans le cadre de la e-santé.

L'accessibilité des objets connectés est un facteur important pour son utilisation dans le cadre du soin. Dans un premier temps, lorsque l'objet n'a pas « *marquage CE* », il ne peut prétendre à un processus de remboursement par la Sécurité sociale. En France, c'est la Haute Autorité de Santé (HAS) qui est chargée de l'évaluation de l'opportunité de remboursement selon des critères bien définis : évaluation du service attendu (intérêt pour le patient et pour la santé publique) et l'évaluation de l'amélioration du service attendu²⁹².

²⁹⁰ Entretien 19.

²⁹¹ Entretien 14.

²⁹²Source : https://www.has-sante.fr/portail/jcms/c_2845863/fr/specificites-methodologiques-d-evaluation-clinique-des-dispositifs-medicaux-connectes?xtmc=&xtcr=1 consulté le 25 mai 2019.

La rapidité d'évolution de la technologie et de la médecine est un facteur inquiétant pour les médecins. Les sociétés qui développent les outils sont des petites ou moyennes entreprises (PME) qui n'inspirent pas confiance car elles disparaissent ou se font racheter par des grandes entreprises. Ces PME n'ont tout simplement pas la capacité d'effectuer les mises à jour techniques en termes de sécurité ou de connaissances nécessaires à la pratique médicale. Ces objets connectés deviennent rapidement obsolètes et doivent être remplacés. Ainsi, les médecins proches de la retraite ne trouvaient aucun intérêt à investir dans du matériel informatique ou des objets connectés.

La réticence des médecins à utiliser les objets connectés était liée à la protection des intérêts des patients (utilité des outils, fiabilité, protection des données), au maintien d'une relation médecin-patient humaine qui n'est pas fondée sur les données transmises par les objets connectés. Le domaine de la santé est un secteur où les nouvelles idées et les essais de solutions pour améliorer les conditions de vie et de prise en charge des patients évoluent rapidement. Seule une infime partie des solutions arrivent à révolutionner la prise en charge des patients.

L'utilisation des objets connectés dans le domaine de la santé n'est pas nouvelle. Les pacemakers, les holters ECG ont été utilisés depuis des années dans le cadre de la surveillance des maladies. Mais la réduction importante des coûts de développement des applications, la disponibilité des données en grande quantité et la puissance de calcul des machines ont entraîné une prolifération des entreprises de conception d'outils destinés au soin et une modification importante du rapport aux données de santé.

Les problèmes que l'on peut rencontrer dans le cadre de l'utilisation des objets connectés sont très variés et ne présentent pas tous le même degré de gravité. Le risque peut porter sur des informations incomplètes ou d'absence de certaines fonctionnalités, des erreurs de calcul ou d'orientations diagnostiques. On peut classer les risques selon le fait qu'ils portent sur :

- la protection des données de santé et la confidentialité qui conduit à une atteinte au secret médical ;
- l'absence de validation scientifique des données cliniques ou le non-respect des finalités de l'application ;
- le dysfonctionnement des objets et des logiciels ou un manque de fiabilité des capteurs ;

- la vulnérabilité avec des failles de sécurité des produits.

Propositions d'action pour une meilleure adhésion à l'utilisation des objets connectés

Les médecins ont proposé la mise en place d'actions par les différents acteurs participants à la conception des objets connectés pour une meilleure adhésion. Les actions ont été catégorisées comme suit :

- **une implication de tous les acteurs (patients, médecins, sociétés savantes) dès la conception des objets connectés afin de mieux évaluer les besoins :**

La participation de tous les acteurs dans le processus de développement permettrait de créer un climat de confiance dans la mise en place des objets. Ce climat de confiance est très important pour faciliter l'appropriation des objets connectés par les utilisateurs finaux.

Par exemple, l'entreprise pourrait définir un projet d'objet à créer. Les patients et les médecins pourront définir leur besoin réel et leurs attentes. Ils définissent ensemble les variables nécessaires, pertinentes et utiles pour atteindre l'objectif. Ce processus de transparence mis en place facilite plusieurs aspects de la conception pour l'entreprise, les médecins et les patients :

une validation de l'utilité de l'application par les différents acteurs avec une définition claire du service à améliorer ;

une collecte des données utiles pour le développement de l'objet tout en respectant les principes de protection et de l'utilisation des données de santé ;

une participation des sociétés savantes tout au long du processus de conception permet un respect des normes métiers, une certification accélérée pour le marquage CE et le remboursement par la Sécurité sociale.

- **une amélioration de la gestion des données en termes de sécurité, de confidentialité et de préservation du secret médical ;**

Un des grands problèmes des objets connectés est la gestion et la gouvernance des données. Si ces questions ne sont pas résolues, le risque éthique dans le cadre de la pratique médicale est très important et n'aura pas l'assentiment des médecins. Une des solutions est d'assurer la sécurité des objets (matérielles et logicielles) par des entreprises pérennes et la mise en place des équipes capables de faire les mises à jour. Il faudra avoir une transparence dans le

transfert des données lors de l'utilisation. Mais le gros problème réside dans le modèle économique des entreprises qui repose en général sur les traitements des données et la valeur marchandes que l'on peut en tirer.

- **une valorisation de l'utilisation des objets connectés dans les actes médicaux ;**

L'utilisation des objets connectés nécessite du temps pour la transmission des données et l'analyse de ses données par les médecins. Sauf que les actes médicaux liés aux objets connectés ne sont pas remboursés officiellement par la Sécurité Sociale créant ainsi un manque à gagner pour les médecins libéraux du fait du mode de tarification à l'acte appliqué en France. Ainsi, les médecins préféreraient prendre des rendez-vous de consultation plutôt que d'interpréter des résultats sans rentabilité économique. Mais ce mode de réflexion de la part des médecins est-il déontologique ? Ne met-il pas en avant le gain par rapport au bien-être du patient ? Mais, faut-il que le médecin mette en péril le fonctionnement de son cabinet pour le bien-être des patients ? Nous avons la dimension économique de la prise en charge qui rentre en jeu. Elle mérite une réflexion sérieuse si nous voulons tirer profit des TIC dans le domaine de la santé. Si la santé n'a pas de prix, elle a un coût qui doit être pris en charge.

- **une simplification des outils avec une interopérabilité avec les logiciels de dossiers patients informatisés ;**

Le nombre des objets connectés augmentent chaque année et ne sont pas compatibles avec les outils existants. Les concepteurs développent en fonction de leurs propres normes ; les médecins se retrouvent avec un nombre élevé d'interface à manipuler. La gestion de l'interopérabilité serait une grande avancée dans la gestion des données de santé.

Pour aider les médecins dans le choix et l'intégration des applications et des objets connectés dans leur pratique médicale, un projet en cours (*ApiAppS*) en collaboration avec l'équipe D2IM (Département d'Informatique et d'Information Médicales) de Rouen et SESSTIM (Sciences Economiques et Sociales de la Santé & Traitement de l'Information Médicale) d'Aix Marseille évalue la qualité des outils afin de mettre en place un outil d'aide au choix. Ce projet permettrait de vérifier la qualité et l'utilité des outils et le respect des normes de protection des données de santé (DUFOUR ET AL., 2019). Le développement d'outils permettant de faire des choix fondés sur des preuves est très important pour l'intégration de la *m-health* et leur appropriation par les médecins.

6.6.3 *Utilisation des applications et sites web*

Les médecins conseillaient l'utilisation des applications et sites web à leurs patients dans 13 cas. Ils utilisaient ces technologies dans le cadre de conseils (n = 12) et de l'information (n = 11) aux patients.

La figure 80 montre la répartition de l'utilisation des applications et sites web en fonction des indications.

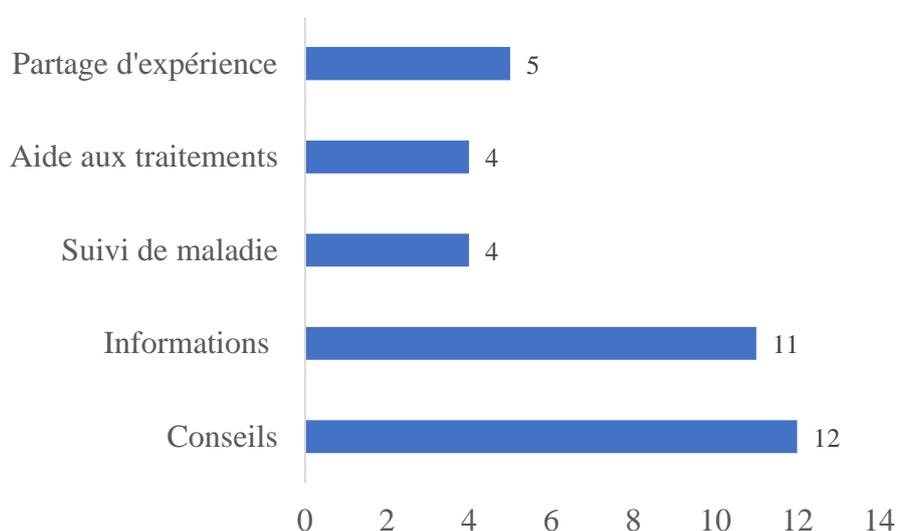


Figure 80 : utilisation des applications et sites web en fonction des indications (n = 12)

Mode d'utilisation des sites web et des applications

Les médecins utilisaient les sites web et les applications pour les conseils et l'information aux patients dans le but de les aider à être acteur du soin. L'analyse de contenu des entretiens montrait qu'ils avaient un processus de sélection des sites web destinés aux patients. Ces processus étaient :

- une vérification des sites web et des applications avant de faire la proposition aux patients ;
- une utilisation des sites web et des applications dans le cadre de pathologies bien définies, notamment dans les cas où ils existaient des patients experts dans le domaine :

« J'ai pris conscience de l'importance des sites web lors de la formation sur le baclofène. Il y a une patiente qui faisait des témoignages sur son usage du baclofène à partir du réseau mis en place sur ce produit. J'ai essayé le site avant de le proposer à mes patients »²⁹³.

- une utilisation des sites web institutionnels (HAS, assurance maladie) et des sociétés savantes ;
- une recherche de l'information avec le patient en ligne avec une explication adaptée à son cas.

Le choix de l'utilisation se faisait toujours avec le consentement du patient.

Freins à l'utilisation des sites web et des applications

Dans les cas où les sites web et les applications n'étaient pas utilisés, les raisons avaient été :

- la méconnaissance de l'existence des sites web et des applications ;
- la méconnaissance des acteurs concepteurs du site web ou des applications
- la capacité d'analyse et d'adaptation du message par le patient telles que limites dans la compréhension des messages, la fiabilité des informations sur les sites web et applications, les quantités importantes d'informations et l'angoisse causée au patient à la vue des informations ;
- le manque d'utilité des sites web et des applications pour les médecins ;
- les questions techniques telles que la collecte des traces à l'insu des patients, la sécurité et la confidentialité des données et la publicité ciblée ;
- la préférence pour des conseils verbaux avec une adaptation au cas du patient.

Avec le développement des technologies et la facilité de mise en ligne des informations, la recherche d'informations aussi bien pour les professionnels de santé que pour les patients devient de plus en plus compliquée, exige du temps pour le tri des informations et une adaptation de l'information pour les patients.

Les médecins utilisaient les sites web des sociétés savantes, des institutions étatiques comme la HAS ou des outils de recherche bibliographiques tels que Lissa (Littérature scientifique

²⁹³ Entretien 6.

de santé)²⁹⁴, Pubmed/Medline et CISMef²⁹⁵ pour la recherche des informations ; ces sites web et outils leur font gagner du temps car ils comportent des informations scientifiques validées. Par contre pour les outils de recherche bibliographiques, il faut une formation pour mieux exploiter les outils et avoir des meilleurs résultats.

Pour les patients, la recherche des informations sur Internet est encore plus compliquée. Il existe une multitude de sites web proposant des informations sur les maladies sans que l'on ne sache exactement qui est l'auteur et quelles sont ses sources. Des méthodes ont été mises en place pour évaluer la qualité de l'information des sites web, mais elles ne font pas l'unanimité pour les différents acteurs ; néanmoins, elles ont l'avantage de guider les utilisateurs.

Les critères de qualité d'une information sur Internet selon l'Institut de recherche et documentation en économie de la santé (IRDES)²⁹⁶ sont :

- une information ayant une référence vérifiée (auteurs ou responsable du site web) ;
- une information pertinente ;
- une information récente et actualisée ;
- une information accessible.

La certification *HONcode*²⁹⁷ est un engagement moral pour divulguer les informations éditoriales. Le *HONcode* est un code de déontologie guidant les responsables de sites web dans la mise en place de mécanismes fondamentaux permettant de mettre à disposition une information médicale de qualité, objective et transparente adaptée à la mission et à l'audience

²⁹⁴ C'est une base de données qui comprend aujourd'hui plus de 800 000 références issues de PubMed mais aussi de 72 journaux scientifiques qui n'y figurent pas et de 53 qui y figurent de façon incomplète. Si l'on considère uniquement les années supérieures à 2000, la moitié du contenu de LiSSa est inédite. Plus de 80 000 résumés en français sont également proposés. Source : www.lissa.fr consulté le 05 janvier 2020.

²⁹⁵ CISMef est une réalisation du D2IM – Département d'Informatique et d'Information Médicales du CHU Hôpitaux de Rouen. Ce projet a pour objectif le Catalogage et l'Indexation des Sites Médicaux de langue Française (CISMef) accessibles sur l'Internet. Ce catalogue indexe les principaux sites et documents francophones. Il a dépassé les 43.000 ressources indexées en mai 2008 avec une moyenne de 55 nouvelles ressources par semaine. CISMef utilise deux outils standards pour organiser l'information : le thesaurus MeSH (Medical Subject Headings) utilisé notamment pour la base de données bibliographique Medline et le format de métadonnées du Dublin Core. Chaque site indexé comprend une notice descriptive. CISMef est un outil utile pour la communauté francophone, professionnel de santé ou cyber-citoyen : chaque jour ouvré, 30.000 ordinateurs se connectent sur ce site représentant environ 40.000 personnes. Source : <http://www.chu-rouen.fr/cismef/projet-cismef/a-propos/> consulté le 05 janvier 2020.

²⁹⁶ Source : <https://www.irdes.fr/documentation/documents/sources-d-information-et-methodologie-de-recherche-documentaire.pdf> consulté le 05 janvier 2020.

²⁹⁷ Source : https://www.hon.ch/HONcode/Patients/Visitor/visitor_f.html consulté le 05 janvier 2020.

du site. La transparence d'un site améliore l'utilité et l'objectivité de l'information et la publication de données correctes. Les sites demandant la certification et les sites certifiés s'engagent à respecter les bonnes pratiques éditoriales qui sont au nombre de huit :

- une autorité : indiquer la qualification des rédacteurs ;
- une complémentarité : compléter et non remplacer la relation patient – médecin ;
- une confidentialité : préserver la confidentialité des informations personnelles soumises par les visiteurs du site web ;
- une attribution : citer la/les source(s) des informations publiées et dater les pages du site web ;
- une justification : justifier toute affirmation sur les bienfaits ou les inconvénients de produits ou traitements ;
- un professionnalisme : rendre l'information la plus accessible possible, identifier le webmaster, et fournir une adresse de contact ;
- une transparence du financement : présenter les sources de financements ;
- une honnêteté dans la publicité et la politique éditoriale : séparer la politique publicitaire de la politique éditoriale.

Cependant, *HONCode* n'évalue pas la qualité intrinsèque des informations présentes sur le site web, mais demande le respect d'un ensemble de règles qui oblige les rédacteurs de site web d'être transparent. Il permet aux visiteurs de sites d'avoir toutes les informations nécessaires pour évaluer le site web eux-mêmes. En plus, la certification est volontaire et ce n'est pas tous les sites qui en font la demande ; ce n'est pas parce que le site ne possède pas aussi la certification qu'elle n'est pas fiable.

Le score *NetScoring* (PERRIN, 2016) a été élaboré à l'initiative de Centrale Santé qui a constitué un groupe de travail multidisciplinaire comprenant des ingénieurs, des professionnels de santé, un juriste, un bibliothécaire médical et des organisations professionnelles. Ce score compte 49 critères de qualité, répartis en huit catégories principales : crédibilité, contenu, liens, design, interactivité, aspects quantitatifs, aspects déontologiques, accessibilité.

Cette grille d'évaluation est essentiellement destinée aux décideurs de santé, aux administrateurs de sites web. Elle peut être utilisée par le grand public.

La difficulté d'évaluation de l'information sur Internet dans le domaine de la santé est telle que la nécessité de la médiation du médecin pour expliquer et aider le patient à faire le choix des informations adaptées à son contexte est très importante. « *Je reçois des patients qui consultent Internet avant de venir nous voir. Il ne s'agit pas d'un phénomène chez les jeunes seulement ; mais tous les patients sont concernés* »²⁹⁸.

En effet, une des craintes des médecins était de devoir expliquer aux patients que les informations trouvées sur Internet ne s'appliquaient pas à leur cas et de les contextualiser.

*« Il est très fréquent que des patients viennent très angoissés après des recherches sur Internet sur leurs symptômes. Par exemple Docteur, j'ai fait du sang dans les selles, j'ai un cancer alors que c'est juste une crise d'hémorroïdes. On est obligé de rappeler que c'est un tout qui fait le diagnostic et le patient n'a pas cette objectivité »*²⁹⁹.

Néanmoins, les applications et les sites web peuvent constituer des outils complémentaires utiles à la prise en charge des patients. Ils peuvent soutenir et renforcer la relation de soins et améliorer l'adhésion des patients aux conseils de prévention et aux traitements ; « *Par exemple, concernant les maladies rares, certains patients qui en sont atteints en connaissent plus que nous les médecins (qui n'ont jamais vu de cas). Lorsque les patients ont fait des consultations sur Internet, j'essaie de savoir ce qu'ils ont retenu et compris et je leur apporte mon aide dans la compréhension* »³⁰⁰.

6.6.4 Télémédecine

La télémédecine était utilisée par un seul médecin dans le cadre de suivi du diabète gestationnel³⁰¹.

« La télémédecine est une forme de pratique médicale à distance utilisant les technologies de l'information et de la communication. Elle met en rapport, entre eux ou avec un patient, un ou plusieurs professionnels de santé, parmi lesquels

²⁹⁸ Entretien 5.

²⁹⁹ Entretien 10.

³⁰⁰ Entretien 5.

³⁰¹ Le diabète gestationnel ou diabète de grossesse est une augmentation de la glycémie. Il apparaît pendant la grossesse et disparaît après l'accouchement.

figure nécessairement un professionnel médical et, le cas échéant, d'autres professionnels apportant leurs soins au patient. Elle permet d'établir un diagnostic, d'assurer, pour un patient à risque, un suivi à visée préventive ou un suivi post-thérapeutique, de requérir un avis spécialisé, de préparer une décision thérapeutique, de prescrire des produits, de prescrire ou de réaliser des prestations ou des actes, ou d'effectuer une surveillance de l'état des patients »³⁰².

Encadré 9: mise en œuvre de la télémédecine

Le décret n° 2010-1229 du 19 octobre 2010 donne les conditions de mise en œuvre de la télémédecine clinique en précisant les différentes pratiques médicales (art. R. 6316-1) :

- la téléconsultation qui a pour objet de permettre à un professionnel médical de donner une consultation à distance à un patient. Un professionnel de santé peut être présent auprès du patient et, le cas échéant, assister le professionnel médical au cours de la téléconsultation ;
- la télé expertise qui a pour objet de permettre à un professionnel médical de solliciter à distance l'avis d'un ou de plusieurs professionnels médicaux en raison de leurs formations ou de leurs compétences particulières, sur la base des informations médicales liées à la prise en charge d'un patient ;
- la télésurveillance médicale qui a pour objet de permettre à un professionnel médical d'interpréter à distance les données nécessaires au suivi médical d'un patient et, le cas échéant, de prendre des décisions relatives à la prise en charge de ce patient ;
- la téléassistance médicale qui a pour objet de permettre à un professionnel médical d'assister à distance un autre professionnel de santé dans la réalisation d'un acte ;
- la réponse médicale apportée dans le cadre de la régulation médicale : elle se décline en plusieurs prestations médicales, dont l'aide médicale urgente et le télé conseil médical personnalisé

³⁰² Loi n° 2009-879 du 21 juillet 2009 portant réforme de l'hôpital et relative aux patients, à la santé et aux territoires.

La télémédecine est un des outils de régulation de l'offre de soins mis en place par la loi Hôpital Patient Santé Territoire (HPST)³⁰³ du 21 juillet 2009. L'enjeu de cette loi était de réduire les inégalités de répartition démographique de l'offre médicale et de lutter contre un accès aux soins qui serait inégal entre les habitants du territoire français. La télémédecine est donc indissociablement liée à la désertification médicale. Dans le cadre de cette loi, la télémédecine devait se pratiquer entre deux médecins : un médecin requérant qui demande l'acte de télémédecine et un médecin requis qui pratique l'acte. Depuis le 15 septembre 2018, l'acte de téléconsultation³⁰⁴ est remboursable par la Sécurité sociale ; mais fait nouveau, il peut se pratiquer directement entre le médecin et le patient mais à condition de respecter le parcours de soins.

Dans ce contexte, la réduction des déserts médicaux reste difficile car dans ces vastes territoires, les patients n'ont pas de médecins traitants pour leur suivi et doivent se déplacer dans la ville la plus proche. En plus, il faudra que le médecin dégage du temps pour les actes de téléconsultation, un temps qui n'est déjà pas suffisant pour les patients présents en cabinet. A côté de ces difficultés, la télémédecine nécessite l'aménagement des cabinets médicaux, l'acquisition des plateformes de téléconsultation et la capacité du patient à manipuler les applications.

Un rapport de la HAS en 2013 (efficience de la télémédecine : état des lieux de la littérature internationale et cadre d'évaluation) (HAS, 2013) montrait que la télésurveillance médicale a un impact certain sur les événements morbides de certaines maladies chroniques comme l'insuffisance cardiaque ou l'insuffisance respiratoire en prévenant les hospitalisations et de ce fait les accidents de décompensation ; *« par contre la démonstration d'un impact sur la santé peut demander plusieurs années dans les maladies chroniques comme l'hypertension artérielle ou le diabète. Un résultat coût/efficacité favorable n'a pas été démontré sur une période de 12 mois »* (HAS, 2013).

Une revue systématique de la littérature réalisée par une équipe du LIMICS (FARNIA ET AL., 2018) montrait que l'évaluation de la télésurveillance devraient prendre en compte six dimensions (économique, clinique, éducationnelle, technique, perception des utilisateurs, organisationnelle). Dans cette revue, 95% des études prenaient en compte au moins trois

³⁰³ Ibid.

³⁰⁴ L'avenant n°6 à la convention médicale de 2016 a été signé le 14 juin 2018 par l'Union nationale des caisses d'assurance maladie (Uncam) et les cinq syndicats représentatifs des médecins libéraux (FMF, CSMF, MG France, Le Bloc, SML). Le texte a été approuvé par un arrêté du 1^{er} août 2018, paru au Journal officiel le 10 août 2018 (les annexes de l'avenant ont été publiées au Journal officiel le 23 août 2018).

dimensions et aucune étude ne prenait en compte toutes les dimensions. Les dimensions clinique et économique étaient plus étudiées durant les phases d'implémentation et d'intégration et moins durant la phase et de développement. Les dimensions organisationnelle et technique, très importantes dans le déploiement des solutions, n'ont été étudiées que respectivement dans 5% et 4% des études. La mise en place des outils de e-santé d'une manière générale est très complexe et doit prendre en compte plusieurs aspects. En effet, dans son étude, Solène Houdard-Brunet (HOUDARD-BRUNET, 2014) a montré que les patients greffés du rein préféraient le télé suivi à leur domicile par téléconsultation au suivi habituel dans le service de transplantation. La qualité de prise en charge médicale était aussi meilleure que dans une prise en charge en face-à-face et cela leur permettait de faire moins de déplacements pour aller à l'hôpital. Les patients étaient prêts à payer eux-mêmes le maintien de cette prestation à distance tant elle leur apportait une meilleure qualité de vie (moins de déplacements, moins de perte de journée de travail, un contact plus personnalisé avec le médecin).

Les médecins pensaient que le choix des indications était très important pour que la télémédecine ait un impact réel sur la vie des patients sans pour autant entraîner des modifications importantes aussi bien pour la pratique médicale et la vie du patient. Par exemple dans le cadre du suivi des maladies chroniques chez les sujets jeunes, la télésurveillance serait d'un grand apport en termes de qualité de prise en charge et de confort de vie. Ainsi, ces patients ne seront pas obligés de venir à l'hôpital tout le temps pour leur suivi. « *Dans le cadre de la relation médecin – patient, on a beaucoup de jeunes gens qui n'aimeraient pas être tout le temps à l'hôpital, et la télémédecine pourra être un grand apport pour ces jeunes diabétiques* »³⁰⁵.

Des sociétés savantes proposent des solutions de télémédecine à leurs adhérents en fonction d'indications claires pour améliorer le confort des patients :

*« Par exemple dans le cas de l'anesthésie, la SFAR (Société Française d'Anesthésie Réanimation) est en train de réfléchir à l'utilisation de la télémédecine pour les patients qui doivent subir une deuxième intervention dans un délai d'un mois par exemple. Au lieu que le patient se déplace, il envoie ses résultats à distance et se connecte en ligne pour avoir une consultation avec l'anesthésiste pour faire une évaluation de son état générale »*³⁰⁶.

³⁰⁵ Entretien 16.

³⁰⁶ Entretien 15

Bien que la télémédecine semble être à la mode ces dernières années, elle n'est pas pour autant une technique récente dans la pratique médicale. Dans les années 1920, (a été délivrée la première licence pour radio de service médical aux bateaux à New York), on avait recours à la téléassistance par une radio consultation pour aider les médecins dans le domaine de l'aide médicale au navire (AMENTA, 2000). Elle se fait également par l'intermédiaire du SAMU et des conseils téléphoniques. Les moyens technologiques et le mode de la pratique qui sont en train de changer. Néanmoins, la pratique de la télémédecine exige le respect des règles juridiques et déontologiques qui encadrent la profession médicale car elle est considérée comme un acte médical. Il se pose ainsi la question de la responsabilité des différents acteurs intervenants dans le cadre de la télémédecine.

Plusieurs acteurs interviennent dans le cadre de la télémédecine avec des responsabilités différentes :

- le médecin (requérant ou requis) en fonction du type d'acte doit informer le patient de l'acte de télémédecine en lui expliquant les raisons pour lesquelles l'acte est demandé, les avantages et les inconvénients. Ces explications doivent être claires, précises et faciles à comprendre pour le patient. Il devra obtenir le consentement éclairé et écrit si possible du patient après l'avoir donné toutes les informations concernant le déroulement de l'acte de télémédecine, les informations personnelles et médicales qui seront partagées avec les autres professionnels de santé. En ce sens, il doit recueillir toutes les informations nécessaires pour la médecin requis pour que celui-ci puisse mener à bien son acte (ORDRE NATIONAL DES MÉDECINS, 2014 ; DGOS, 2012). Tous les actes réalisés et les décisions prises lors de l'acte de la télémédecine doivent être clairement notifiés pour chaque intervenant sous la supervision du médecin requérant afin d'assurer la traçabilité des données et de situer les responsabilités en cas de faute ;
- le médecin requis s'assurera d'exercer sa profession dans des conditions qui ne puisse pas compromettre la qualité des soins et des actes médicaux ou la sécurité des patients conformément à l'article 71 du code de déontologie médicale. Il devra avoir les compétences nécessaires pour le recours à l'acte de la télémédecine. Il définit toutes les informations nécessaires dont il aura besoin pour mener à bien sa mission. Tout comme le médecin requérant, le médecin requis est tenu au secret

professionnel et à la confidentialité des données. Il est responsable de tous ceux qui interviennent dans le cas de l'acte de télémédecine sous sa responsabilité.

Le médecin requis devra s'assurer du bon fonctionnement du système et avoir une maîtrise parfaite de l'utilisation de tous les outils nécessaires à l'acte de la télémédecine dans le cadre des obligations de moyens (DGOS, 2012 ; WILLIATTE-PELLITTERI, 2013).

- l'organisateur : les activités de télémédecine se mènent dans le cadre d'un programme national, soit sous forme d'un programme régional de télémédecine ou soit sous forme d'un contrat conclu avec un particulier. Ce contrat permettait d'identifier les obligations ainsi que les responsabilités des acteurs organisationnels de l'activité de la télémédecine. L'organisateur est également responsable des professionnels de santé qu'il intègre dans l'exercice de l'activité de télémédecine au sens où il doit garantir au patient que le professionnel qui le prend en charge a légalement les compétences nécessaires. Dans le cas contraire, il risque d'être susceptible de poursuite pour complicité d'exercice illégal de la médecine. L'organisateur doit veiller à ce que les professionnels de santé intervenants dans l'activité de la télémédecine soient formés à l'utilisation des appareillages qu'ils utilisent. L'organisateur garantir la fiabilité de l'hébergeur des données et de celle du fournisseur d'accès. Il répondra par conséquent des défaillances du système sans préjudice pour lui d'un recours subrogatoire contre les principaux concernés (WILLIATTE-PELLITTERI, 2013). L'organisateur est en charge de la responsabilité consécutive à la défaillance du matériel utilisé dans le cadre de l'activité de télémédecine. En application des principes de la responsabilité civile du fait des choses, il appartient à celui qui a la garde des choses de répondre aux préjudices dont elles sont la cause.
- les tiers technologiques sont les différents prestataires techniques des réseaux de télémédecine. Ils sont indépendants les uns vis-à-vis des autres et ne sont responsables que des prestations mises à leur disposition.

Dans le cadre de la télémédecine, le matériel utilisé est considéré comme un outil de diagnostic, de prévention, de contrôle et de traitement (article R 5211-1 Code de Santé Publique (CSP)). Le matériel utilisé pour réaliser les actes de télémédecine répond à la définition du dispositif médical (article L 5211-1 CSP). Il est soumis à une exigence de conformité (article L 5211-3 CSP) et de déclaration des incidents ou des risques d'incidents (article L 5211-2 CSP). Par conséquent,

les fabricants de matériels doivent installer du matériel conforme aux dispositions légales avec une garantie de la qualité et une planification de la maintenance préventive et curative du système et la mise en place d'un système de notification des incidents. En plus, ils devront assurer la formation des acteurs de la télémédecine à l'utilisation du matériel et mettre à disposition une description technique du matériel.

- l'hébergeur doit être habilité par le ministère de la santé à héberger des données de santé. L'hébergeur veille à la sécurité de l'information notamment en ce qui concerne les données personnelles de santé. Cette sécurité concerne la confidentialité, l'intégrité, la pérennité et la disponibilité des données. Elle concerne également la vérification des personnes habilitées à accéder, le contrôle des droits d'accès et la traçabilité des données. L'hébergeur devra permettre l'authentification de l'émetteur de l'information et l'impossibilité pour le récepteur de nier une réception réussie.

Les fournisseurs d'accès Internet doivent fournir un débit suffisant pour la pratique de la télémédecine. Ils doivent respecter les normes en ce qui concerne cette connexion Internet et veiller à ce que les informations soient accessibles et transmises à temps.

En d'autres termes, la pratique de la télémédecine doit garantir :

- la disponibilité, l'intégrité, la confidentialité et la sécurité des données de santé ;
- l'identification sécurisée des professionnels de santé ;
- l'identification sécurisée des patients ;
- la traçabilité des actes ;
- l'interopérabilité avec l'infrastructure nationale et/ou régionale.

Les médecins pensaient que la télémédecine n'est pas adaptée à la pratique libérale pour plusieurs raisons :

- la mise en place des infrastructures technologiques est trop onéreuse ;
- la question de la responsabilité des différents acteurs reste en suspens et difficile à mettre en évidence. C'est le médecin qui est accusé dans un premier temps et c'est à lui de prouver la responsabilité des autres acteurs ;

- les risques d'erreurs médicales sont plus importants du fait de l'absence de l'examen clinique et une difficulté d'appréciation du degré d'urgence avec à la clé des poursuites judiciaires ;
- la télémédecine ne permet pas de concilier qualité de la consultation avec une optimisation des soins ;
- la sécurité des réseaux pour la transmission des flux de données est insuffisante ;

En effet, les contraintes et les obligations de la mise en œuvre de la télémédecine constituaient des freins ; les médecins pensaient que les SAMU étaient plus aptes à son utilisation :

« Pour ce qui concerne la télémédecine, nous les urgentistes nous la pratiquons depuis longtemps à travers la régulation des appels SAMU. Cela est très complexe et compliqué car nous devons nous fier aux descriptions du patient et faire l'interrogatoire et mobiliser les équipes pour y intervenir. Nous sommes protégés par la loi et la responsabilité ne nous incombe pas directement en cas de faute. Ce n'est pas le cas dans le cadre de la télémédecine, la responsabilité du médecin sera directement engagée surtout dans la pratique libérale »³⁰⁷.

Dans le cadre de l'épidémie à coronavirus, le ministre français de la santé et de la solidarité, Olivier Véran, a levé les conditions de l'avenant 6 dans le cadre des conditions de la pratique de la télémédecine et les contraintes techniques et juridiques pour éviter la propagation de l'épidémie et permettre la continuité de soins des Français jusqu'au 30 avril 2020. Mais, les normes techniques ont été allégées pour permettre d'augmenter l'offre des plateformes et la principale condition était le respect du RGPD. Ainsi des plateformes telles que *Skype*, *WhatsApp* ou *Facetime* ont été acceptées pour la pratique de la télémédecine. L'utilisation de la télémédecine dans des situations particulières ou des indications précises pourrait être d'un grand intérêt, mais elle doit prendre en compte la gestion et la gouvernance des données sur le long terme. Que deviendront toutes ces plateformes qui ont obtenu l'autorisation exceptionnelle après l'épidémie ? Comment les données de santé seront-elles gérées ?

³⁰⁷ Entretien 19.

Pour les médecins, le risque est que la télémédecine rentre dans le cadre d'une logique structurelle ou économique plutôt que conjoncturelle avec un risque « *d'ubérisation* » de la pratique médicale³⁰⁸ avec la multitude de plateformes de télémédecine qui est en train d'être créée à tous les niveaux.

En plus, le déploiement de la télémédecine et des technologies de l'information de la communication d'une manière générale dans le domaine de la santé, nécessite le changement des équipements dans les cabinets ou dans les hôpitaux pour s'adapter. Ce déploiement est à l'origine d'une augmentation importante des coûts des projets technologiques. Il faudra faire une analyse complète du système de santé et proposer des solutions prenant en compte le fonctionnement global du système dans sa complexité et non pas gérer les situations au cas par cas.

Les besoins fonctionnels liés à la réalisation d'un acte de télémédecine mobilisent différents composants applicatifs et systèmes d'information qui doivent ainsi être mis en œuvre en respectant les recommandations suivantes :

- harmoniser les infrastructures de télémédecine avec les solutions déjà en place ou en cours de déploiement ;
- utiliser les référentiels fonctionnels applicables en matière d'identification des différents acteurs ;
- veiller à l'intégration des logiciels et systèmes d'information métiers existants, dans le dispositif de télémédecine, selon leurs caractéristiques et leurs fonctionnalités (Système d'Information Hospitalier, Système d'Information de Laboratoires, etc.) ;
- favoriser le recours à des composants et infrastructures « standards » pour éviter la multiplication de solutions qui aboutit à la « non-interopérabilité » et qui pourrait s'avérer très onéreuse à mettre en œuvre et surtout à maintenir ;
- privilégier la mutualisation des services, gage de cohérence technique et d'économies d'échelle en matière d'investissement et de fonctionnement ;
- privilégier le recours aux outils disponibles sur le marché aux dépens de développements spécifiques.

³⁰⁸ La start-up Medadom, qui gère une plateforme de prise de rendez-vous médicaux à domicile, se lance dans la téléconsultation et teste actuellement ses services "dans une dizaine d'officines à Paris et en Seine-Saint-Denis. Source : <https://www.ticsante.com/story/4630/telemedecine-medadom-prend-ses-quartiers-dans-les-pharmacies-d-officine.html> consulté le 20 novembre 2019.

6.6.5 *Dossier médical partagé (DMP)*

Le Dossier Médical Partagé (DMP) a été mis en place pour favoriser la prévention ainsi que la coordination, la qualité et la continuité des soins³⁰⁹. En effet, lors de nos entretiens, les médecins sont unanimes que le partage des informations entre professionnels de la santé est très important dans le cadre de la prise en charge des patients en termes de sécurité, de continuité de soins et de qualité : « *le DMP va clairement aider les médecins dans la prise en charge. Les médecins des hôpitaux ont du mal à communiquer les gardes et les examens complémentaires, mais à condition que ça fonctionne bien* »³¹⁰ ; le partage est également important en termes de santé publique notamment dans le cadre de la réduction des redondances des examens complémentaires, qui constituent une source de dépenses importantes pour la Sécurité sociale même si elles restent marginales. Mais les médecins ont des craintes et des appréhensions par rapport à son utilisation à savoir :

la gestion du consentement

*« Une fois que le bénéficiaire de l'assurance maladie a consenti à la création de son dossier médical partagé, il ne peut, sauf motif légitime, s'opposer à ce que les professionnels de santé qui le prennent en charge versent dans son dossier médical partagé les informations utiles à la prévention, la continuité et la coordination des soins qui lui sont ou seront délivrés »*³¹¹.

La question du consentement est très importante dans le cadre de la pratique médicale. Il constitue la base du respect de l'autonomie de la personne dans le cadre de l'éthique médicale et du respect de la dignité de la personne. L'autonomie est un engagement à faire participer le patient au processus décisionnel, de ce fait, à respecter son autodétermination. Le concept est bien balisé en droit avec la loi n° 2002-303 du 4 mars 2002 relative aux droits des malades et à la qualité du système de santé « *aucun acte médical ni aucun traitement ne peut être pratiqué sans le consentement libre et éclairé de la personne et ce consentement peut être retiré à tout moment* »³¹², en médecine avec l'article 45 du code de déontologie

³⁰⁹ Article R1111-26 du code de santé publique.

³¹⁰ Entretien 10.

³¹¹ Article R1111-36 du code de santé publique.

³¹² Article L1111-4 du code de santé publique.

« ... à la demande du patient ou avec son consentement, le médecin transmet aux médecins qui participent à la prise en charge ou à ceux qu'il entend consulter les informations et documents utiles à la continuité des soins »³¹³ et sur le fait que le consentement en elle-même a une valeur morale intrinsèque et constitue un privilège.

Ainsi dans l'article Article R1111-36 du CSP et Article L1111-4 du CSP, nous constatons une contradiction sur le consentement qui ne peut être retiré dans le premier cas sauf en cas de motif légitime – qui d'ailleurs n'est pas défini dans la loi – et n'est plus nécessaire une fois le DMP créé.

« Le motif légitime est également une notion standard. Il a été évoqué plein de fois dans la loi européenne et il n'y a pas de définitions. C'est le tribunal qui va apprécier si vous avez un motif légitime ou pas. La CNIL donne des exemples de motifs légitimes. C'est à la personne de prouver qu'elle a un motif légitime »³¹⁴.

Le consentement est nécessaire pour chaque acte médical ou traitement. Est-ce à dire que le processus de versement des données dans le DMP ne constitue pas un acte médical ? Si la constitution du dossier professionnel et la mise à jour à chaque consultation font partie de l'acte médical, le processus de versement constitue également un acte médical et doit nécessiter le consentement du patient à chaque consultation. Le consentement doit être libre, spécifique à un traitement (précise ceux à quoi le patient est consentant, la durée dans les limites légales), éclairé (explications claires, simples et compréhensibles pour le patient), transparent avec un droit de retrait ou de correction des données pour le patient.

Pour chaque acte que doit effectuer le médecin, il doit avoir l'avis du patient aussi bien dans la gestion des informations. Cette gestion de l'information fait partie de la confidentialité des données et du respect du secret professionnel. Dans le cadre du DMP, le consentement va au-delà de la création du dossier qui est un acte différent du versement des données dans le DMP. L'article R1111-36 du CSP dépossède le patient et le médecin de l'autorité sur la gestion des informations et du secret médical. Elle constitue une intrusion dans la relation médecin-patient. L'articulation avec le DMP suppose que le patient ait donné son accord pour le transfert de données médicales qui le concernent dans une base informatique où elles seront hébergées. Ceci à toutes les étapes de son « *parcours de soins* ».

³¹³ Article R4127-45 du code de santé publique.

³¹⁴ Entretien Juriste.

En d'autres termes, le médecin ne peut pas inscrire dans le DMP de données médicales sans l'accord du patient, mais il pourrait inscrire des codes d'actes comme il le fait actuellement sur des feuilles de soins, qu'elles soient électroniques ou non, afin que la prestation soit remboursée par l'assurance maladie dans le respect de l'autonomie du patient.

la gestion de la confidentialité des données et de la sécurité des données

Au vu du nombre de personnes ayant accès aux données des patients dans le cadre de soins et surtout dans le cadre du service public³¹⁵, les médecins ont des craintes sur la confidentialité des données et de ce fait au respect du secret professionnel. Cette inquiétude réside dans le fait que l'intérêt public n'est pas défini dans les textes juridiques et est à l'appréciation du juge où des comités mis en place pour l'évaluation des projets afin de décider de l'accès des données.

Généralités

Rappel à destination des organismes bénéficiant d'un accès permanent au SNDS

Les services de l'État, établissements publics, et organismes chargés d'une mission de service public bénéficiant d'un accès permanent n'ont pas le pouvoir de donner librement un accès au SNDS aux personnes de leur choix. Leurs possibilités d'habilitation se limitent à leurs personnels (ou assimilés, par exemple, un chercheur intégré dans une équipe, mobilisé pour le compte de l'organisme et sous l'autorité hiérarchique du directeur) et aux sous-traitants, agissant au nom des organismes précités et pour leur compte dans le strict cadre de leurs missions de service public. Les accès par les personnes habilitées :

- doivent être délivrés par le directeur général de l'organisme (ou son délégué),
- doivent être ponctuels et limités aux seuls besoins du projet concerné,
- ne peuvent être utilisés qu'à cette fin (articles L.1461-3-III du code de la santé publique et R1461-16 du code de la santé publique issu du décret n°2016-1871 du 26 décembre 2016 relatif au traitement de données à caractère personnel dénommé «système national des données de santé »).

Les personnes (personnels et sous-traitants) habilitées à accéder au SNDS sont responsables nominativement. Elles ont ainsi l'interdiction formelle :

- de faire bénéficier d'autres personnes de ces accès sous leur identité,
- d'utiliser ces accès pour mener des études déconnectées de la mission de service public qui a justifié l'octroi des accès permanents, qu'il s'agisse de manipulations menées pour le compte de tiers ou déclenchées sous couvert de conventions de partenariat.

Figure 81 : note d'information sur le site de la SNDS³¹⁶ sur les accès aux données

Cette note sur le site de la SNDS témoigne de la conscience du risque d'utilisation non autorisée des accès aux bases de données pour d'autres fins.

³¹⁵ Liste des services publics et acteurs ayant accès aux SNDS dans le cadre de l'intérêt public. Article R 1461-12 du décret n° 2016-1871 du 26 décembre 2016 relatif au traitement de données à caractère personnel dénommé « système national des données de santé ».

³¹⁶ Source : <https://www.snds.gouv.fr/SNDS/Actualites/Actu-13> consulté le 21 octobre 2019.

La mise en place du DMP ne prend pas en compte les compétences techniques du patient en matière de TIC ; il aura besoin dans ce cas d'un tiers, pas forcément médecin, qui aura accès aux données du patient. La divulgation ultérieure pourrait dans ce cas être imputée à l'un des médecins précédemment consultés, à plus forte raison s'ils ont été nombreux. D'où l'intérêt sans doute d'une traçabilité et d'un catalogue chronologique des accès. Il est probable que le secret médical sera alors du niveau de celui que l'on apprécie dans les salles d'attente ou les jours de foire où chacun raconte ses histoires, et même celles des voisins ; « *l'utilisation des données est un frein à la mise en place du « serpent de mer » qu'on appelle le dossier médical partagé (DMP). Il se pose la question de la sécurité des données, l'accès des données et la gouvernance des données* »³¹⁷ disait un médecin lors de nos entretiens.

Le secret professionnel devient de ce fait le secret du patient dont les discussions concernant sa santé, s'il ne prête pas attention jusqu'à présent aux conséquences, vont devenir, preuves à l'appui, une arme redoutable contre lui, par rapport à son employeur, aux assurances ou à la justice. Le patient, devenu responsable de par la loi, devra assumer les risques encourus sans aucune formation de base et sans connaissance des enjeux ; c'est pourquoi l'information du patient, dans le cadre du DMP, prend une dimension capitale. Il est donc nécessaire pour le patient d'être maître de ses données mais il ne faut pas que cela engendre des responsabilités pour auxquelles le patient n'est pas préparé et ne peut prévenir les conséquences à long terme. En outre, dans une enquête Odoxa, « *les données de santé sur internet et les réseaux sociaux* », réalisée pour le compte de Healthcare Data Institute³¹⁸, 29% des Français ont déjà parlé de leur santé sur le web et dans les réseaux sociaux (Facebook, Twitter, Instagram) dans 73% des cas et dans 29% des cas dans des forums de communautés de patients.

Les médecins avaient une faible confiance à l'hébergement des données en ligne. Pour eux, une fois que les données étaient en ligne, elles étaient accessibles à tout moment et par tout le monde. Malgré la certification des hébergements des données de santé, une centralisation des données de santé dans un serveur serait un risque important et faciliterait le travail des

³¹⁷ Entretien 18.

³¹⁸ Les français et les données de santé sur internet et les réseaux sociaux (Communiqué) Healthcare Data Institute. Source : <https://www.dsih.fr/article/3142/les-francais-et-les-donnees-de-sante-sur-internet-et-les-reseaux-sociaux.html>, consulté le 25 avril 2019.

pirates en cas d'attaques. En effet, ces derniers temps, les données de santé sont l'objet d'attaques croissantes.

la relation médecin – patient

Une des craintes des médecins est que l'utilisation du DMP avec les prérogatives qu'il donne au patient n'influence leur relation. La relation médecin-patient est une relation qui se construit sur la base de la confiance mais surtout par l'intermédiaire d'un jeu d'interaction avec une influence réciproque des acteurs. C'est une relation qui se construit sur le long terme. Ainsi, pour chaque annonce, le médecin met en place des stratégies de communication dont la durée dépend de l'individualité du patient. Par contre dans le cadre du DMP, l'article R 1111-42 précise :

« Lorsqu'un professionnel de santé estime qu'une information sur l'état de santé versée dans le dossier médical partagé ne doit pas être portée à la connaissance du patient sans accompagnement, il peut rendre cette information provisoirement inaccessible au titulaire du dossier en attendant qu'elle soit délivrée à ce dernier par un professionnel de santé lors d'une consultation d'annonce. Dans un délai de deux semaines suivant le versement d'une information inaccessible, et en l'absence de la consultation d'annonce, le patient est informé par tout moyen y compris dématérialisé d'une mise à jour de son dossier médical partagé, l'invitant à consulter un professionnel de santé, notamment son médecin traitant, pour en prendre connaissance. Si la consultation d'annonce n'a pas eu lieu un mois après le versement de l'information dans le dossier médical partagé du patient, elle devient automatiquement accessible »

Cet article vient influencer la relation médecin-patient en ce sens qu'il impose des délais pour informer le patient sans connaissance réelle de l'interaction communicationnelle existante entre les deux acteurs.

L'annonce du diagnostic ou d'un état d'aggravation significatif de sa maladie est un choc³¹⁹ pour le patient. Le vécu de l'annonce peut être différent en fonction de la personne, en fonction de la personnalité, du milieu socioculturel, de la gravité effective et ressentie de la

³¹⁹ Par humanité, beaucoup de médecins n'appliquent pas la surtarification à laquelle ils ont droit dans le cas où ils doivent annoncer une très grave maladie. Ce tact ne pourrait exister dans le cas d'un médecin robot dont les comportements suivent les logiques de programmes informatiques.

maladie et de la représentation que se fait le patient de la maladie. Le médecin s'adresse au patient dans son individualité, sa singularité et l'annonce doit être adaptée à sa prédisposition psychique et cognitive. Il doit s'efforcer de réaliser cette annonce de façon à atténuer autant que possible le ressenti des informations. La durée de l'acceptation de la maladie peut varier en fonction du patient et le médecin doit être à l'écoute de manière attentive pour aider le patient durant cette période. La question de la responsabilité va se poser lorsque le patient est informé par une accessibilité directe de l'information à partir du DMP. Le médecin peut-il être incriminé pour le tort causé par ce mode d'information s'il survenait un incident du fait de l'annonce de l'information ? Si le DMP doit servir pour la coordination des soins, il est important de permettre aux médecins et aux patients de gérer les informations en fonction de la relation qui les lie et non pas en fonction des besoins techniques de l'information.

Le DMP n'est pas exhaustif ; il est destiné à rassembler les éléments de prévention, de diagnostic et thérapeutiques nécessaires à la coordination des soins.

la valeur juridique du DMP en cas de litige ;

« Le DMP ne se substitue pas au dossier que tient chaque établissement de santé ou chaque professionnel de santé, quel que soit son mode d'exercice, dans le cadre de la prise en charge d'un patient »³²⁰.

De ce fait, les médecins préféreraient avoir un dossier patient bien tenu et complet dans leurs cabinets que de perdre du temps à remplir un DMP qui n'a pas de valeur juridique, ni de valeur dans la prise en charge des patients,

« Par exemple on nous pousse à mettre en place le DMP, je préfère avoir le dossier patient de mon cabinet bien fait et complet et accessible de façon pour mes collègues de cabinet que de perdre mon temps à remplir le DMP ou garder les informations dans la mémoire du médecin »³²¹.

Si le DMP ne peut pas se substituer aux dossiers patients du cabinet, quelle est donc sa place dans le cadre de la coordination des soins ? Le médecin peut-il se fier aux données du DMP pour prendre en charge son patient ? Mais, du fait de la nature des données qu'il contient, il devient tout de même une pièce médico-légale.

³²⁰ Article R 1111 – 28 du CSP.

³²¹ Entretien 20.

La réponse se trouve dans les conditions générales d'utilisation du DMP³²² qui dit :

« le DMP n'est pas exhaustif. Il est destiné à rassembler les éléments de prévention, de diagnostic et thérapeutiques nécessaires à la coordination des soins. Toute interprétation, diagnostic ou toute décision, de quelle que nature que ce soit, se basant sur ces données relève de la seule responsabilité de la personne qui les consulte ».

Il se pose alors la question de la mise en place du DMP dans le cadre du soin. Est-ce un outil de collecte de données de santé dans le cadre de la recherche en intelligence artificielle comme l'a préconisé le rapport Villani le rapport Villani (VILLANI ET AL., 2018) : *« lancer un nouveau chantier spécifique de production d'informations et de données de santé afin d'entraîner les technologies de l'IA ».*

6.6.6 Assistants médicaux

Au moment de la réalisation de notre enquête auprès des médecins, le rôle des assistants médicaux n'était pas encore défini de manière claire et la question de la rémunération de ces assistants médicaux était toujours en discussion.

Mais les médecins, quand ils étaient favorables à cette idée, leur préféraient un rôle purement administratif dans la gestion des courriers, des télétransmissions, de l'intégration des lettres dans les dossiers des patients, en d'autres termes le rôle de secrétaire médicaux. Certains médecins souhaitaient leur implication dans la préparation de la consultation à savoir la prise de constante, la préparation des dossiers afin d'augmenter le nombre de patients en consultations. Et d'autres ne voulaient pas de ce rôle car pour eux réduire le temps de contact médical avec le patient pourrait altérer la relation médecin – patient et réduire la relation de confiance qui est très importante dans la pratique médicale.

Le déploiement des assistants médicaux répond à un triple enjeu³²³ :

- *« favoriser un meilleur accès aux soins des patients : épaulé par un assistant médical, le médecin sera accompagné dans sa pratique quotidienne, ce qui lui*

³²²Source : <https://media.dmp.gouv.fr/MPA/ConditionsUtilisation.html> consulté le 21 octobre 2019.

³²³ Arrêté du 14 août 2019 portant approbation de l'avenant n° 7 à la convention nationale organisant les rapports entre les médecins libéraux et l'assurance maladie signée le 25 août 2016.

permettra de se dégager de certaines tâches, par exemple administratives, et de libérer du temps médical. Il pourra donc recevoir davantage de patients ; les assurés auront de leur côté moins de difficultés à trouver un médecin traitant et à obtenir un rendez-vous, avec un spécialiste notamment, dans un délai raisonnable ;

- *assurer de meilleures conditions d'exercice : le médecin pourra consacrer plus de temps aux soins et au suivi médical de qualité des patients ;*
- *rechercher davantage d'efficacité, une meilleure prise en charge et un suivi amélioré : engagé dans une démarche de coordination des soins, le médecin pourra plus facilement assurer la coordination et la continuité des soins avec l'ensemble des autres acteurs de la prise en charge de ses patients ».*

L'arrêté³²⁴ précise que « *les missions que les médecins confient à l'assistant médical sont laissées à leur appréciation en fonction de leurs besoins et de leurs modes d'organisation, et selon le profil soignant et/ ou administratif des personnes recrutées dans le cadre du référentiel métier* ».

Le rôle de l'assistant médical se situe à la fois entre le rôle de secrétaire médical, de l'infirmier ou des deux à la fois. La liberté est laissée aux médecins de décider comment l'assistant médical sera employé sans pour autant remplacer le secrétaire médical ou l'infirmier. Plusieurs questions se posent : pourquoi ne pas permettre aux médecins de recruter des infirmiers directement pour les aider à accomplir cette triple mission ? Si les médecins veulent de l'aide dans les tâches administratives, pourquoi ne pas permettre de recruter des secrétaires médicaux ? La solution de la création des assistants médicaux pose la question de la gestion des dépenses dans le système de santé. Même si le dispositif permet la création d'emploi, l'efficacité et l'efficience quant à l'amélioration de la prise en charge des patients restent à démontrer. Cela risque de créer des tensions avec le corps des infirmiers et celui des secrétaires médicaux.

L'accès à l'aide à l'embauche d'assistants médicaux est conditionné par le territoire, la spécialité et le mode d'exercice de la profession. Mais une fois les conditions d'accès

³²⁴ Ibid.

remplies, les médecins ont un contrat d'engagement qui conditionne la poursuite du financement à savoir :

- une augmentation de la patientèle adulte du médecin traitant : cet indicateur mesure le nombre total de patients de plus de 16 ans ayant choisi le médecin généraliste comme médecin traitant,
- une augmentation de la file active : la notion de file active correspond au nombre de patients différents tout âge confondu vus dans l'année par le médecin.

Le financement est indexé à une obligation de résultat stipulé dans un contrat. N'est-il pas contraire au code de déontologie en son article 97 qui dit :

« un médecin salarié ne peut, en aucun cas, accepter une rémunération fondée sur des normes de productivité, de rendement horaire ou toute autre disposition qui auraient pour conséquence une limitation ou un abandon de son indépendance ou une atteinte à la qualité des soins ».

Le fait de vouloir respecter les objectifs quantitatifs fixés pourrait porter atteinte à l'indépendance du médecin avec des préjudices sur la qualité des soins offerts à long terme (réduction du temps de contact avec les patients, altération de la relation médecin-patient).

Par exemple dans le cas des maladies chroniques, un médecin après un certain nombre d'années de pratique est saturé en termes de patientèle. Les patients doivent être revus régulièrement selon une fréquence donnée en dehors des urgences. Sauf en cas de décès, il ne peut pas intégrer d'autres patients sous peine de réduire la qualité de la prise en charge et les temps de la consultation. Cela est d'autant plus vrai dans les zones sous tension où l'aide à l'embauche sera octroyée.

Les médecins généralistes constituent la cheville ouvrière du système de santé ; ils assurent le tri des patients avant de les diriger chez les spécialistes. La proximité et la relation de confiance créées pendant des années avec les familles font que le temps passé avec son médecin traitant en consultation est plus long et peut entraîner une réticence des patients vis-à-vis des assistants médicaux. Ce constat est évident par exemple dans le cadre de la formation des internes qui assistent les titulaires lors de la consultation : *« J'ai des internes et je vois très bien qu'avoir une troisième personne dans une salle influence*

de façon importante la relation médecin patient ; on ne parle plus de la même chose »³²⁵.

Nous craignons que le temps médical gagné par les assistants médicaux soit vite perdu à moyen et long terme du fait des demandes croissantes de soins et des départs à la retraite de médecins qui ne sont pas remplacés. Les effectifs de médecins généralistes nouvellement formés ne compensent pas ceux partant à la retraite et la liberté d'installation des médecins ne permet pas de contrôler la gestion des zones d'installation (LUCAS-GABRIELLI, CHEVILLARD, 2018). La formation des médecins en nombre reste une des meilleures options pour la satisfaction des besoins de soins de la population.

6.7 Gouvernance et gestion des données

6.7.1 Participation à la recherche

Les médecins étaient en majorité d'accord pour participer à des projets de recherches scientifiques (18 médecins) ; mais à la condition que les patients consentent à participer à l'étude et qu'ils jouent un rôle actif dans la réalisation de l'étude et non un rôle de fournisseur de données.

Les motifs de partage de données de santé dans le cadre de la recherche étaient de permettre selon les médecins à :

- une amélioration des connaissances sur les maladies et la découverte de nouveaux traitements ;
- un contrôle de la pratique médicale (surveillance des prescriptions pour l'identification des abus, le contrôle des dépenses de soins, la pharmacovigilance) ;
- une surveillance épidémiologique des maladies afin d'être alerte ;
- un partage des expériences de la pratique médicale avec une implication des patients et des institutions de recherche.

³²⁵ Entretien 1.

Parmi les médecins qui refusaient de participer aux projets de recherche, les raisons étaient liées à leur spécialisation (médecin expert judiciaire et médecin du travail) chez deux médecins. Pour le troisième médecin, il s'agissait de convictions personnelles.

Le partage des données de santé constitue un enjeu majeur pour les systèmes de santé. Ce partage permet un accès aux données médicales des patients, avec une fiabilité et une qualité des données exceptionnelles. Les professionnels de santé disposent dans le cadre des soins et du suivi les antécédents, les historiques de consultation, les résultats des examens complémentaires dans le cadre de l'accomplissement de leur mission. Les données répondent donc à un besoin d'amélioration continu de la qualité de soin. La disponibilité et l'accès à des informations médicales relatives à un patient permettent d'éviter les examens redondants et de faire ainsi des économies.

En plus, avec l'informatisation des systèmes de santé, il existe un partage important entre les professionnels de santé (médecins, biologistes, pharmaciens) dans le cadre de transmission des résultats des examens complémentaires et des demandes d'avis auprès de leurs confrères médecins.

Au-delà de ces avantages de partage de données (réduction des dépenses, qualité de soins, la continuité des soins), le partage des données de santé avec d'autres acteurs permettra de mieux exploiter les données et de faire un apport important dans l'évolution de la pratique médicale comme le témoignait nos entretiens avec les médecins ; « *Souvent la recherche se limite aux laboratoires en milieu clos. Nous avons souvent des informations et des patients qui pourront faire avancer la connaissance médicale, même en médecine générale* »³²⁶.

En effet, les données en vie réelle³²⁷ constituent un atout important dans la recherche à la suite des essais cliniques qui se limitent à une population plus restreinte. Les données en vie réelle pourraient permettre de détecter des événements qui n'avaient pas été observés lors des essais cliniques (effets secondaires et indésirables des médicaments, efficacité secondaire sur une indication non prévue par l'essai clinique) et de transposer les résultats à une plus large population. Ainsi, plusieurs structures de soins peuvent cumuler leurs données pour une pathologie donnée pour comprendre l'évolution d'une maladie dans la « vraie » vie.

³²⁶ Entretien 1.

³²⁷ Les données de vie réelle sont les données collectées lors de l'activité de soins habituelle sans intervention sur les modalités de prise en charge, elles reflètent la pratique courante des agents de santé.

Mais le désir de partage des données se heurte à de nombreuses difficultés que sont : l'interopérabilité³²⁸ des systèmes, la loi sur la protection des données de santé qui sont considérées comme des données sensibles, le consentement du patient qui reste tout de même le propriétaire des données et la protection du secret professionnel.

Dans le domaine de la santé, le premier principe du partage de données est l'identification des patients. Il faut une identification sans ambiguïté des patients et des professionnels de santé dans le cadre du partage des données entre les professionnels de la santé. Il faut de ce fait créer un identifiant unique afin d'éviter les doublons et les confusions. Des initiatives ont été prises au niveau de l'Etat avec la création de l'identifiant national de santé (INS) qui permettra de faire l'identification des utilisateurs du système de santé et d'améliorer le parcours de soins. « *On parle du parcours patient dans le GHT (groupement hospitalier de territoire), alors qu'aujourd'hui on n'a pas d'identifiant GHT. On a un projet c'est l'INS pour 2020, mais une étape pouvait être les numéros de GHT* »³²⁹ déplorait un médecin DIM. Pour les professionnels de santé ; le système de CPS (carte professionnelle de santé) permet cette identification et est utilisé dans le cadre de la transmission des feuilles de soins électroniques (FSE). Mais si la carte est obligatoire en pratique libérale pour la transmission des FSE, ce n'est pas le cas dans les structures publiques, ce mode de fonctionnement n'incite donc pas son déploiement systématique avec des risques de failles dans l'identification des professionnels et la surveillance des habilitations.

*« Les niveaux d'habilitation ne sont pas les mêmes pour chaque praticien. Le cadre médico-administratif a un rôle de contrôle et de gestion. Il y a une balance entre les DIM et la facturation, qui travaille au DIM. Nous avons intégré DIM et facturation. Le contrôle devrait être sur les coûts et le contrôle médico-administratif et ne devait pas avoir accès aux informations nominatives. Mais dans les hôpitaux, il y a des passoires et les gens ont des accès qui dépassent leurs attributions et c'est difficile de reprendre la main »*³³⁰.

³²⁹ Entretien DIM.

³³⁰ Ibid.

Le DMP a été relancé afin de faciliter le partage des données entre professionnels mais surtout la disponibilité des données dans le cadre de la recherche et le soin, mais il est confronté à de nombreuses critiques et difficultés que nous avons évoquées plus haut (*section 6.6.5*).

Une meilleure exploitation des données dans le cadre du partage pour son utilisation dans le *big data* et l'IA est sa structuration. Cette structuration doit se faire sur le plan informatique à partir des référentiels d'interopérabilité et sur le plan médical avec des définitions précises des données transmises et des contextes de recueil des informations.

« Beaucoup de données médicales sont de textes libres et il faut une codification des données. Il faut structurer les données en les codifiant en CIM10 ou SNOMED. Ce sont ces genres de données qui sont très intéressants pour le big data, le deep learning ou l'IA »³³¹.

En effet, la plupart des données stockées dans les bases de données sont en texte libres ou en images sans respects des normes standard. Dans le cadre du projet *EHR4CR*³³² (*Electronic Health Record for Clinical Research*), les auteurs notaient qu'il existait des difficultés de mise en place de systèmes interopérables au niveau des pays entre hôpitaux et structures de recherche, mais également dans le cadre de la collaboration entre les différents pays (DANIEL ET AL., 2016).

L'agence des systèmes d'information partagés en santé (ASIP santé) a mis en place des normes et des référentiels dans le cadre de la conception des logiciels médicaux mais leur application reste toujours insuffisante sur le terrain. Aussi bien dans la production de soins que dans la gestion des dossiers de santé, le choix des outils se fait en fonction des préférences des acteurs pouvant être à l'origine d'une multitude d'outils différents non interopérables dans la même structure ;

« On a des dossiers patients informatisés contenant des données médicaux et médico-administratifs et on a des logiciels métiers contenant des données un peu éparpillées en réanimation, en cardiologie ou en réanimation néonatale et qui ne sont pas à chaque fois interopérables »³³³.

³³¹ Entretien Médecin entrepreneur en IA.

³³² Le projet *EHR4CR* (*Electronic Health Record for Clinical Research*) est un projet européen qui avait pour objectif de réduire le coût des essais cliniques à travers une meilleure utilisation des données cliniques contenues dans les DPI.

³³³ Entretien DIM

L'interopérabilité logicielle est déjà difficile à réaliser. Mais ce qui est plus complexe à réaliser, c'est l'interopérabilité sémantique. Dans le cadre de la normalisation des données médicale, des systèmes de codification existent. Il s'agit entre autres de la *CIM 10*, le *SNOMED*, l'*UMLS*, *LOINC*, *HL7* qui permettent d'avoir des définitions des termes utilisées dans le domaine médical et comparable dans le monde entier. Ils sont tous intégrables dans les bases de données des logiciels de santé. Cette normalisation des données rend plus efficace le partage et l'utilisation des données de santé quelle que soit la provenance.

Le mode de partage des données de santé se fait de plusieurs manières selon les acteurs ; les partages de données à travers :

- les objets connectés par les patients avec d'autres patients ou avec les entreprises de conception de ces objets (de manière volontaire ou contraignante à travers l'acceptation des conditions d'utilisation) ;
- les réseaux sociaux lors des échanges ;
- les entrepôts de données de santé tels que le SNDS, SNIIRAM ou les systèmes d'information des hôpitaux.

Le partage des données fait par les patients sur les réseaux sociaux et à travers les objets connectés n'est pas de la responsabilité du médecin tant qu'il n'intervient pas dans ce processus de partage. Mais, il est important de discuter de cette question ; car pour ne pas être paternaliste, faut-il donner autant de responsabilité aux patients de décider des informations à partager surtout en ces périodes de vulnérabilité qu'est la maladie ? Il ne faudra pas que l'*empowerment* des patients ne se transforme en déresponsabilisation de certains acteurs dans le domaine de la santé. « *La santé n'est pas un bien marchand* » affirme le code de déontologie médicale dans son article 19. Le patient est considéré comme une personne vulnérable, influençable, plus encline à consommer ce qui lui est conseillé. On est face à une forme de dépendance et dans ces conditions, le médecin doit défendre les intérêts de son patient considéré comme vulnérable. L'auto - dispensation de soins a des limites que les TIC ne permettent pas encore de neutraliser. En un sens, la relation de confiance entre le médecin et le patient reste un élément fondamental dans la relation de soin, quel que soit l'outil ou le vecteur d'information utilisé.

Le partage des données personnelles est régi par des textes juridiques en France et dans l'Union Européenne. Ces lois sont mises en place pour garantir la protection des individus et éviter les abus dans l'utilisation de leurs données.

6.7.2 Textes juridiques de protection des données personnelles

La loi informatique et libertés et le décret sur l'hébergement des données de santé n'étaient pas connus par cinq médecins.

Douze médecins sur les vingt-un faisaient confiance à la loi informatique et libertés et le décret sur l'hébergement des données de santé pour une meilleure protection des données de santé. Le RGPD et le *Privacy Shield* n'étaient pas connus par les médecins dans la moitié des cas (n =10). Huit médecins pensaient que le RGPD permettait de protéger les données de santé et trois médecins pour le *Privacy Shield*.

Nous avons noté une insuffisance de connaissances des textes juridiques par les médecins interrogés au cours de notre enquête. Cette insuffisance de connaissance concernait surtout le RGPD et le *Privacy shield* et était constatée pour près de la moitié des médecins. « *Je ne connais pas très bien les procédures de protection des données de santé. Je n'ai pas cherché à les connaître et personne ne nous a parlé de ça* »³³⁴. Il est nécessaire de mettre en place des enseignements post-universitaires sur la loi auprès de ces acteurs car elle peut être à l'origine de plusieurs effets néfastes pour la protection des données ou pour l'utilisation des technologies dans le domaine de la santé. Par exemple, les médecins disaient qu'ils faisaient confiance à leurs informaticiens dans le cadre de la mise en place de leur système informatique pour assurer le respect des normes de protection des données ; mais ce contrat était surtout limité au choix des logiciels des DPI.

³³⁴ Entretien 3.

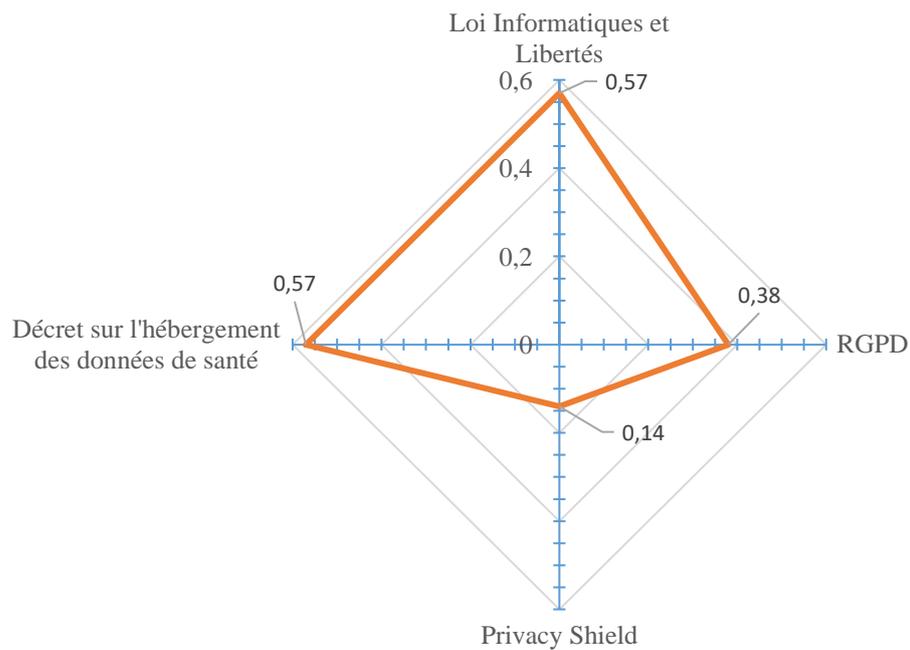


Figure 82: degré de confiance des médecins pour les capacités des textes juridiques dans la protection des données

Les figures 82 et 83 montrent la répartition des opinions des médecins sur la capacité de la loi informatique et libertés, le décret sur l'hébergement des données de santé, RGPD et du *Privacy Shield* à protéger les données de santé.

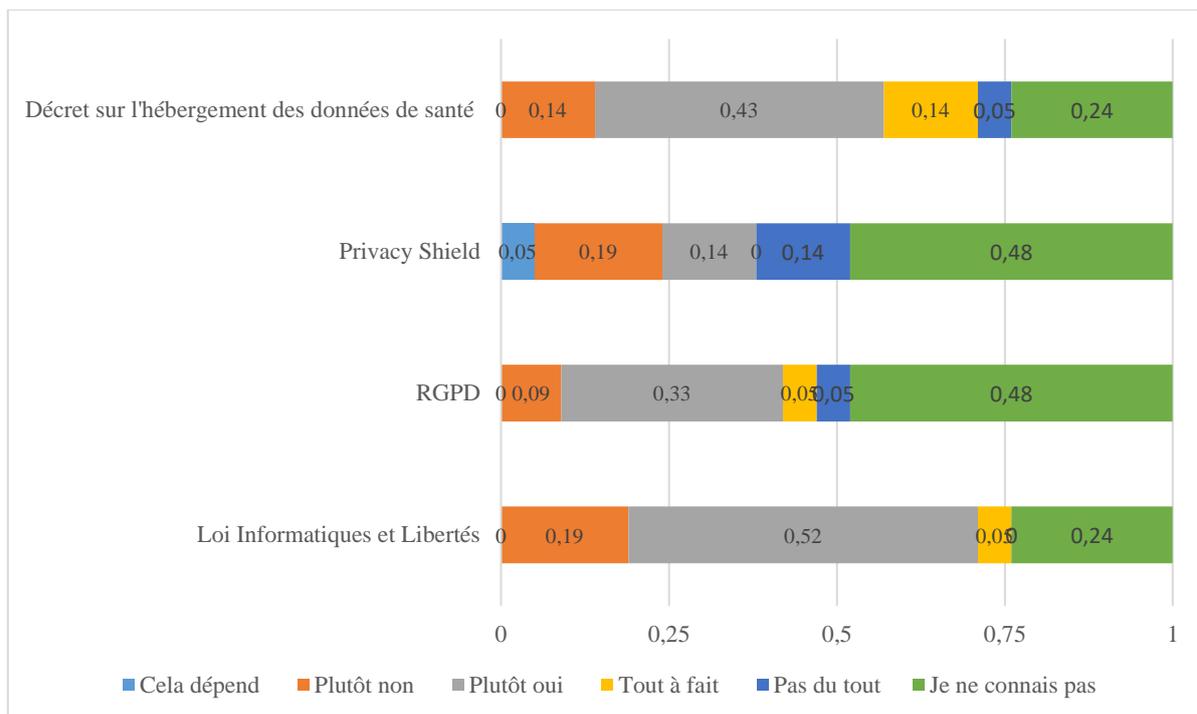


Figure 83: opinions des médecins sur la capacité des textes juridiques à protéger les données de santé

Pour les médecins, les textes juridiques servaient surtout à encadrer une manière de faire et prévoir des sanctions ; mais dans le cadre de la protection des données de santé, il s'agissait surtout d'un engagement personnel des différents acteurs à respecter la loi et protéger les personnes qui leur confiaient leurs données. Le rapport risque/bénéfice pour ceux qui enfreignaient la loi est à prendre en compte car le marché des données est en pleine croissance avec de gros enjeux financiers. Ils pensaient que lorsque les données sont mises en ligne sur Internet, personne ne maîtrisait quoi que ce soit et la meilleure solution était de ne pas divulguer des données de santé sur Internet sans une anonymisation forte.

6.7.3 *Confiance des médecins aux institutions pour la protection des données de santé*

Les médecins faisaient confiance dans la majeure partie des cas aux hôpitaux/cabinets privés (76%) et aux institutions de l'état (sécurité sociale (81%) – institut national des données de santé (76%)) pour la protection des données santé. Par contre, les médecins avaient une méfiance importante vis-à-vis des assurances, des entreprises privées (entreprises pharmaceutiques et entreprises de conception des objets connectés et fournisseurs

d'applications d'intelligence artificielle). La figure 84 montre le degré de confiance qu'accordent les médecins aux différentes institutions.

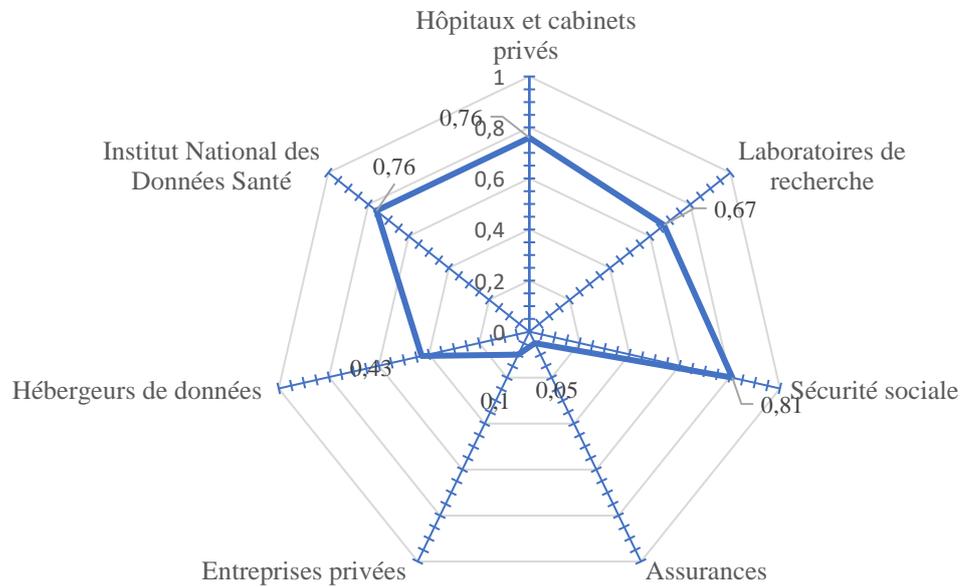


Figure 84: degré de confiance des médecins aux différentes institutions pour la protection des données de santé

La figures 85 montre les détails des résultats du degré de confiance par institutions pour la protection des données de santé. Nous avons noté une indifférence des médecins (40%) envers les hébergeurs des données de santé pour la protection de ces données.

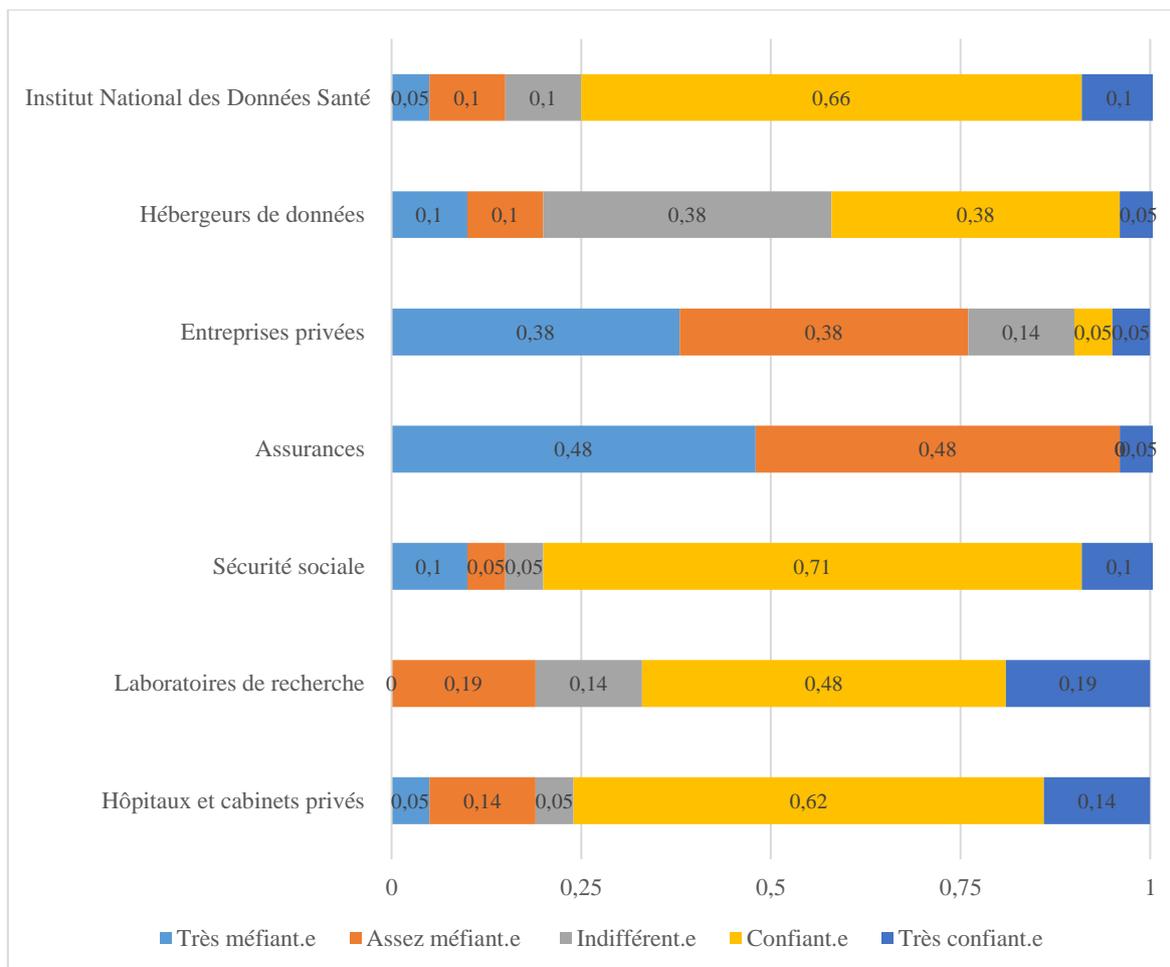


Figure 85: degré de confiance des médecins envers les hôpitaux et les cabinets privés pour la protection des données

Malgré l'existence des textes juridiques et leurs améliorations ces dernières années, les médecins ne faisaient pas confiance à tous les acteurs intervenant dans le processus de traitement des données de santé. En effet, les plateformes de partage de données (réseaux sociaux, forum), les concepteurs d'objets connectés fondent leurs modèles économiques sur la valeur marchande des données de santé avec des entreprises partenaires telles que les assurances ou les laboratoires de recherche. Les entreprises telles que les GAFAM, les NATU ou les BATX se positionnent à travers une médiation technique pour se désengager de tout pouvoir et de toutes responsabilités. Or, l'économie des données aujourd'hui est basée sur la captation d'un maximum de données. Les entreprises ont compris la valeur des données de santé et investissent des milliards de dollars dans le développement du *big data* et des outils d'IA appliqués au domaine de la santé. En plus, les fusions-acquisitions³³⁵ des entreprises de développement des objets connectés ou des TIC par les grands groupes

³³⁵ L'acquisition de Fitbit, entreprise de fabrication des objets connectés, par Google en novembre 2019 avec une base de données comportant les données de plus de 100 millions d'utilisateurs.

rendent très compliqués la gestion des données et les différentes transactions correspondantes.

Les médecins n'avaient pas confiance aux assureurs quant au partage et la gestion des données par ces derniers. En effet, le risque pour le patient est de voir leurs données de santé être exploitées par les assurances pour mettre en place des scores de bonne santé et d'indexer les primes d'assurances sur ces indicateurs. L'assurance santé qui jusqu'à présent reposait sur la mutualisation du risque, pourrait être fondée sur le risque individuel, avec pour conséquence la création d'une inégalité d'accès à la santé selon un mode de vie « *idéalisé* » par des individus qui sera utilisé comme des normes sociales.

6.8 Big data et intelligence artificielle

Les médecins (n=17) avaient déjà entendu parler du *big data* et son principe de fonctionnement. Trois médecins n'avaient jamais entendu parler du *big data*. Dans ces trois cas, une explication du principe du *big data* a été faite aux médecins. Le principe est connu des médecins, mais le terme *big data* pour le définir n'était pas connu.

6.8.1 *Modifications du big data et de l'intelligence artificielle dans la pratique médicale*

Les médecins pensaient que l'utilisation du *big data* et l'IA dans la pratique médicale impacterait le secret médical, la confidentialité des données (n = 15) et la relation médecin-patient (n = 11).

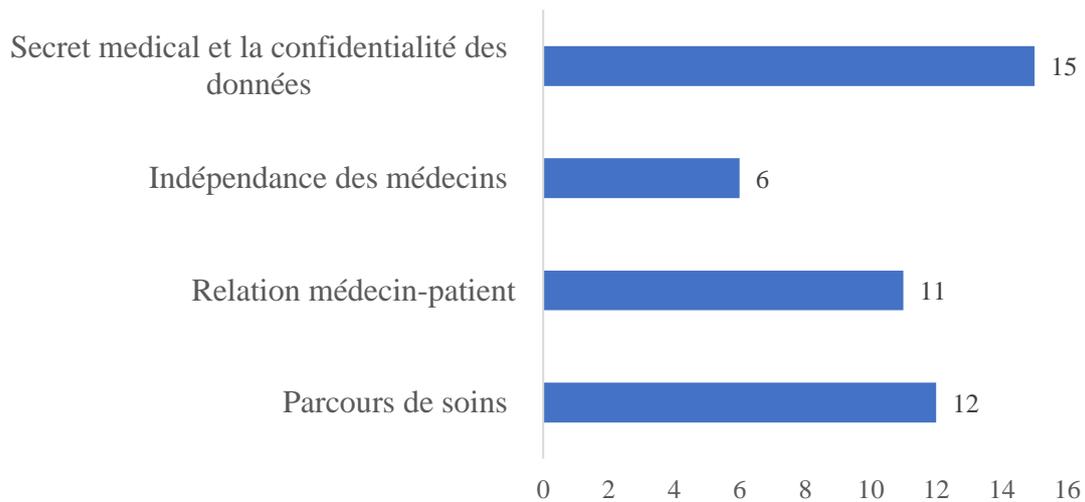


Figure 86: modifications du *big data* et de l'intelligence artificielle dans la pratique médicale

Ces modifications sur ces quatre aspects concernaient :

le parcours de soins

Le parcours de soins permet de s'appuyer sur les bonnes pratiques de chaque professionnel mais aussi de développer les actions d'anticipation, de coordination et d'échanges d'informations entre tous les acteurs impliqués afin d'offrir les meilleurs soins possibles aux patients. Une coordination de qualité nécessite la participation de tous les professionnels concernés qui varie selon la complexité de la situation. La démarche permet au final une meilleure intégration des différentes dimensions de la qualité des soins : la pertinence, la sécurité, l'efficacité clinique mais aussi l'accessibilité, la continuité de soins et le point de vue du patient.

L'utilisation des technologies dans le cadre du parcours de soins va permettre aux médecins d'anticiper les différentes étapes de la prise en charge du patient à l'aide des alertes. Elle permettra également une meilleure communication et le partage des données entre les professionnels de santé dans le cadre de la continuité des soins. Les techniques de l'IA peuvent être d'un apport considérable à travers les SADM par exemple sur la génération des alertes, l'assistance diagnostique, les choix thérapeutiques ou la gestion des flux des patients.

La virtualisation du parcours de soins permet également une meilleure gestion des rendez-vous et une meilleure interaction entre le patient et son médecin. Mais l'une des difficultés majeures de cette virtualisation est la capacité du patient à pouvoir utiliser les outils technologiques mis à sa disposition. Une formation de la population s'avère nécessaire afin de ne pas créer une marginalisation de certains groupes de personnes du système de santé.

Le patient doit être impliqué dans l'évaluation des parcours soins qui le concernent et faire des propositions d'amélioration, en particulier lorsque ses données personnelles sont traitées par l'IA. Il doit pouvoir être un acteur et connaître comment ses données sont traitées par les systèmes de l'IA. Il doit participer à la construction des algorithmes dans le cadre de la transparence et de l'explicabilité. Le consentement du patient est capital quel que soit le bénéfice apporté par les technologies dans le parcours de soins.

la relation médecin-patient

L'accessibilité et la disponibilité de la technologie peuvent entraîner une perte du climat de confiance entre le médecin et son patient instauré par un long processus de relation de soin. Le *big data* et l'IA fonctionnent à partir des données agrégées pour évoluer vers des décisions sur l'individu. Or, la pratique médicale clinique est centrée l'individu pour adapter la prise de décision en tenant compte de son environnement, sa culture, les règles juridiques et les moyens disponibles. Ce mode de pratique nécessite une relation de proximité intense et approfondie qui va permettre au médecin de peaufiner la prise en charge de son patient avec empathie, compassion, altruisme et bienveillance. Certaines technologies comme les applications, les objets connectés, la téléassistance pourront améliorer cette relation de proximité, mais il faudra les intégrer dans le processus de soins existant de sorte qu'ils n'entraînent pas des bouleversements majeurs dans l'organisation des soins.

Le remplacement du colloque singulier entre le médecin par les codes informatiques et les chiffres constituent une crainte profonde des médecins. Laissez les patients seuls face à la technologie dans un moment de vulnérabilité (la peur, la solitude, la dépression, la souffrance) qu'est la maladie aura des conséquences néfastes pour leur santé. L'utilisation des données produites par les patients dans le cadre du soin sans la prise en compte de l'épisode, du contexte et de l'environnement dans laquelle les données ont été produites

pourrait aboutir à une déshumanisation voire « *inhumanisation* » de la pratique médicale considérant

*« l'homme comme une machine à réparer » ; « J'ai peur que la pratique de la médecine n'évolue dans les années à venir. Les gens vont être soigner à distance, ça me faire peur ; il faut manier l'informatique. Moi, je préfère prendre ma retraite avant que ça ne vienne. Dans ce type de procédé, il n'y a plus la relation médecin patient directe comme on le connaissait, on devient des petites machines que l'on répare, là ça me fait peur »*³³⁶ déclarait un médecin.

Par contre, une médiation par le médecin permettrait de tirer le meilleur profit des avancées technologiques sans mettre en danger le patient et sans interférer dans la relation de soin. Cette médiation est plus d'actualité de nos jours car les technologies produisent des données de plus en plus complexes à analyser et interpréter. Le patient a besoin des compétences du médecin pour lui expliquer les données et l'aider dans le processus de décision dans le soin. Mais cette manière de pensée nécessite de revoir tout le processus de conception des outils de l'IA et du *big data* en intégrant tous les acteurs et surtout en assurant la transparence.

L'indépendance du médecin

« Le médecin ne peut aliéner son indépendance professionnelle sous quelque forme que ce soit » Article 5 du code de déontologie médical (article R.4127-5 du code de la santé publique).

L'indépendance professionnelle fonde la confiance du patient. Son corollaire est la responsabilité personnelle du médecin. L'indépendance, la confiance et la responsabilité constituent les éléments essentiels du contrat tacite qui lie le patient et son médecin. Le contrat ne serait pas loyal si le médecin se disposait à agir sous d'autres influences que l'intérêt du patient. Cette indépendance est acquise quand chacun de ses actes professionnels est déterminé seulement par le jugement de sa conscience et les références à ses connaissances scientifiques, avec, comme seul objectif, l'intérêt du patient.

³³⁶ Entretien 3.

Dans le cadre du *big data* et de l'IA, les algorithmes deviennent de plus en plus puissants avec une capacité de détection des maladies souvent supérieure³³⁷ à celle des professionnels (le cas du diagnostic en imagerie médicale). La crainte est un risque d'asservissement et d'accoutumance du médecin au décision du système. Il aura tendance à valider tout ce que propose le système sans chercher à comprendre le processus utilisé pour prendre la décision, perdant ainsi son indépendance.

Dans le système de santé français, les dépenses de santé sont prises en charge dans la majorité des cas par l'Etat. Dans le contexte de rareté des ressources, il cherche par tous les moyens à les maîtriser. Ainsi, la crainte des médecins est de voir une normalisation de la pratique médicale par des canevas imposés. C'est par exemple, le cas des alertes sur la prescription des médicaments de spécialité par la Sécurité sociale. Cette normalisation pourrait aboutir à une obligation non suivie des outils technologiques par les médecins et également une utilisation abusive de la technologie comme moyens de contrôle et de surveillance de leur exercice en lieu et place d'un outil d'amélioration de la pratique médicale. Dans un contexte économique très contraignant pour les hôpitaux, les décisions prises par les médecins peuvent être mises en cause par les gestionnaires des hôpitaux et ces derniers jouent un rôle important dans le choix des outils technologiques dans les hôpitaux. En effet, les médecins estimaient lors de nos entretiens que le déploiement des outils informatiques dans la pratique médicale était décidé plutôt selon des arguments économiques que dans le but d'améliorer la qualité de soins. Ils ne s'opposaient pas aux contrôles de leur pratique, mais il faudrait que ces contrôles respectassent l'éthique médicale et le code de déontologie.

« La Sécurité sociale fonctionne avec de l'argent et quand quelqu'un coûte cher, elle regarde. Mais je trouve, qu'elle ne regarde pas assez, notamment dans les cas d'arrêt maladie, les transports inutiles et les soins inutiles. On utilise l'argent des contribuables et il faut rendre compte quand on les utilise et il faut accepter les contrôles. Les bases de données vont permettre de faire les contrôles des dépenses de soins et d'identifier les abus »³³⁸.

³³⁷ En général, les algorithmes sont entraînés à partir de données étiquetées et vérifiées par plusieurs spécialistes. Par conséquent, sa capacité de détection des cas correspond au cumul des expériences de tous les experts ayant participé à l'étiquetage et ce gap est très important dans les cas rares.

³³⁸ Entretien 3.

Le respect de l'indépendance du médecin impose qu'il ait la mainmise sur les outils. Les outils ne doivent ni remplacer le médecin, ni participer directement à la prise de décision. Ils ne doivent pas contraindre la décision du médecin mais au contraire améliorer la prise de décision et le raisonnement clinique, en d'autres termes ne fournir qu'un avis indicatif. En plus, le médecin doit s'approprier les outils qu'il utilise par une capacité d'expliquer les résultats, être capable de comprendre leur fonctionnement, le processus de raisonnement, qui sont les concepteurs et quelles données sont utilisées pour l'apprentissage.

le secret médical

« Le secret professionnel, institué dans l'intérêt des patients, s'impose à tout médecin dans les conditions établies par la loi. Le secret couvre tout ce qui est venu à la connaissance du médecin dans l'exercice de sa profession, c'est-à-dire non seulement ce qui lui a été confié, mais aussi ce qu'il a vu, entendu ou compris » Article 4 du code de déontologie médicale (article R.4127-4 du code de la santé publique).

Le secret médical est une tradition très ancienne dans la pratique médicale quel que soit son statut. En effet, *« il n'y a pas de soins sans confidences, de confidences sans confiance, de confiance sans secret »* (HOERNI, 2000). Le médecin est responsable de la protection des informations qu'il aurait obtenues dans le cadre des soins. Dans le cadre de l'informatisation des dossiers patients, le médecin doit respecter la loi informatique et libertés et le RGPD, quel que soit son mode de pratique.

La sécurité de ces données constitue donc une obligation juridique, professionnelle et éthique pour les médecins, mais les hôpitaux et les médecins ne sont pas préparés pour faire face aux cyberattaques qui portent atteinte à la confidentialité des données et au secret médical. La sécurité est une composante cruciale dans l'activité humaine et elle joue un rôle important dans la confiance que portent les citoyens à l'Etat et aux institutions. La confiance constitue un enjeu de souveraineté nationale. Les données des personnes et en particulier les données de santé n'échappent pas à cette règle. Cette confiance (THÉVENOT, 2006 ; MANGEMATIN ET AL., 1999 ; MARTIN, QUIVIGER, 2007) dépend de la transparence de mise en œuvre des solutions proposées par les plateformes ou les sociétés de développement informatiques. De

cette confiance et transparence dépendra le niveau d'utilisation et d'appropriation des solutions proposées.

Encadré 10: sécurité des données dans les hôpitaux (GALLOIS, 2016)

Les hôpitaux sont des proies faciles pour les hackers. Contrairement aux banques qui sont devenues des forteresses numériques, les ordinateurs des professionnels de la santé restent très vulnérables. La sécurité informatique est le dernier des soucis des hôpitaux et de leurs personnels, qui ont déjà d'énormes difficultés à offrir des soins de qualité dans un contexte marqué par les restrictions budgétaires. Pour un pirate expérimenté, pénétrer les serveurs d'un hôpital est très facile comme l'a démontré un expert en sécurité :

« en avril 2016, à l'occasion du congrès national de l'APSSIS, il n'a pas fallu plus de cinq minutes à ce gentil hacker pour prendre le contrôle d'un hôpital, puis de mettre la main sur des milliers de dossiers médicaux devant les yeux ébahis des 150 responsables informatiques présents dans la salle selon Vincent Trely, Président de l'APSSIS dans « les pirates à l'abordage des hôpitaux » » (GALLOIS, 2016).

Au-delà des accès aux données, il y a l'altération de l'intégrité des données dans les systèmes d'information, qui peuvent avoir des conséquences néfastes pour la prise en charge des patients et la prise de décision. Des chercheurs d'une université israélienne ont démontré comment l'IA pouvait être utilisée pour créer de faux cancers. Elle pouvait aussi être utilisée pour effacer des données réelles sur des clichés. Les images, présentées ensuite à des radiologues, les ont trompés de manière efficace. Lorsqu'une image de faux cancer leur était montrée, les spécialistes le diagnostiquaient dans 99 % des cas. Et lorsqu'une tumeur était effacée par l'IA, les radiologues se trompaient dans 94 % des cas (MIRSKY ET AL., 2019). Cette possibilité témoigne de l'enjeu de la sécurité des données dans les hôpitaux.

Mais les failles de la sécurité ne sont pas seulement liées à la technologie utilisée, il y a aussi les comportements des acteurs de santé tels que le partage de mot de passe entre collègues, l'utilisation de matériels non vérifiés par les informaticiens et l'utilisation des messageries non sécurisées. *« On a également une gestion des droits qui est mal gérée. Par exemple des*

médecins qui passent leurs codes à des infirmières ou à la secrétaire. On n'a aucune traçabilité là-dessus. Il s'agit des erreurs dues à la personne »³³⁹.

Ces comportements constituent des points de faiblesses pour les systèmes même les plus robustes et ils mettent à mal la confidentialité des données et le secret médical.

« Le secret médical est violé et potentiellement des cas existent. Le problème ici s'est multiplié à cause du dossier patient informatique. Avec le dossier papier, il fallait s'assurer des procédures d'accès au dossier. La difficulté était d'assurer la protection des données dans les services ou les dossiers traînaient sur les chariots et que tout le monde pouvait y avoir accès. Aujourd'hui, nous avons pas mal de plaintes, la plupart des temps, c'est exclusivement des agents de santé qui nous disent qu'on est allé voir mon dossier ou on m'a parlé d'informations qui ne se trouvent que dans mon dossier médical »³⁴⁰.

La sous-traitance et la multiplicité des acteurs dans le processus de soins constituent également des failles de sécurité. En effet :

« Pour les hôpitaux, ils sous-traitent avec des cabinets privés les vérifications comptables avec les identifications et le diagnostic des patients dans le cadre de la tarification à l'acte pour des vérifications. Le secret médical n'est pas préservé. Je peux me connecter sur n'importe quel patient et regarder. Quand le pilote, Michael Schumacher a fait son accident, il y avait 9 000 vues sur son scanner en plein hôpital »³⁴¹.

La sécurité dans les hôpitaux doit tout faire pour atteindre le niveau de protection des données au niveau des banques ; mais il faut également enseigner la culture de la sécurité aux personnels des hôpitaux qui constituent eux-mêmes des failles très importantes de sécurité. Une éducation à la protection des données et surtout aux respects des éthiques est très importante et constitue le premier socle de la sécurité des données.

Avec la centralisation des données par l'intermédiaire du DMP et la possibilité d'accès par les patients à partir de leur ordinateur, les besoins en matière de sécurité seront de plus en

³³⁹ Entretien DIM.

³⁴⁰ Ibid.

³⁴¹ Entretien 20.

plus importants avec les difficultés qu'ont les hôpitaux pour recruter des spécialistes dans un domaine très compétitif. Il va falloir donner les moyens aux hôpitaux pour faire face à la concurrence dans le recrutement des personnes compétentes dans le domaine. En effet, sur le *dark web*, Internet souterrain, un dossier médical rapporte mieux plus qu'une carte de crédit ; « *un dossier médical avec des données personnelles, adresses mails et mot de passe se vend au bas mot 20 euros, alors qu'un code de carte bancaire se monnaie à moins de deux euros* » (GALLOIS, 2016). Cette valeur a atteint 350 euros selon une enquête menée par Kaspersky France en 2019³⁴².

Pour être très efficace, l'utilisation du *big data* et de l'IA nécessite de croiser des données très détaillées en matière de données dans les domaines de la santé, de la culture, de l'environnement. Or, même si elles sont indépendamment anonymisées, ces données, une fois croisées, permettent une réidentification des personnes (DOMINGO-FERRER ET AL., 2016 ; TANGHE, GIBERT, 2017). Ce qui pose la question de savoir si l'on peut faire du *big data* ou de l'IA tout en préservant la vie privée des patients ? A cette question, Cédric VILLANI répond négativement³⁴³. Il s'avère donc nécessaire de trouver un compromis entre protection des données et la vie privée et les avantages de l'utilisation de ces technologies dans la pratique médicale. Cet équilibre très fragile et précaire dans le domaine de la santé reste très difficile, mais peut être trouvé avec la contribution de tous les acteurs concernés sans exception à travers une co-construction des outils basée sur leurs besoins et non sur les intérêts économiques.

³⁴² Source : <https://media.kasperskydaily.com/wp-content/uploads/sites/93/2019/10/25084649/Manifesto-Kaspersky-Protection-des-donn%C3%A9es-de-sant%C3%A9-du-curatif-au-pr%C3%A9ventif.pdf> consulté le 03 mars 2020.

³⁴³ Lors d'un colloque auquel nous avons participé, il s'exprime en ces termes : « *Nous ne pouvons pas faire du big data et de l'IA efficace en protégeant la vie privée des individus* ». Colloque cycle bioéthique 2016 : Numérique et santé (intelligence artificielle, données de santé, *big data*) organisé le 16 mai 2019 à la Cour de cassation de Paris sous la direction scientifique de Mmes Bénédicte BOYER-BÉVIÈRE, Astrid MARAIS et Dorothée DIBIE.

6.8.2 Promesses du big data et de l'intelligence artificielle dans le domaine de la santé

Pour les médecins, le *big data* et l'IA permettraient d'améliorer la prise en charge et la prévention des maladies (n = 16) et une amélioration de l'autonomie du patient (n = 5).

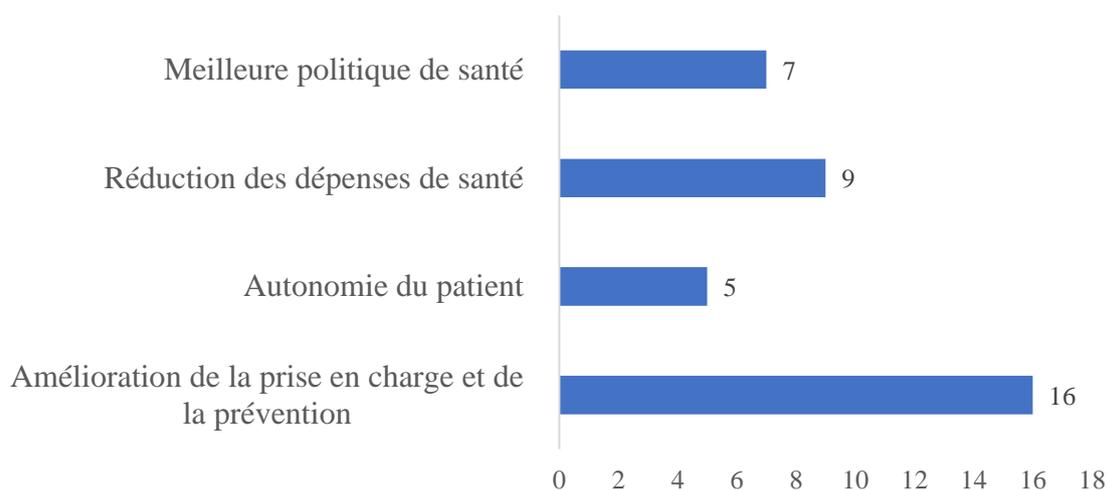


Figure 87 : promesses du *big data* et de l'IA dans le domaine de la santé

Les raisons qui motivaient les médecins à croire aux promesses du *big data* et l'IA étaient :

l'amélioration des soins

Les données présentes dans les systèmes d'information ou les entrepôts de données des hôpitaux sont d'une valeur importante. En effet, ces données peuvent contribuer à la mise en place des systèmes d'aide à la décision médicale, la constitution de cohorte dans le cadre de la recherche clinique et la mise en place d'outils de planification décisionnelle à condition d'utiliser des outils de qualité pour l'extraction des données.

Le *big data* et l'IA effectuent les reconnaissances des formes nécessaires à la classification des images annotées par des spécialistes du domaine. Ainsi, ils vont permettre aux médecins de gagner plus de temps dans l'interprétation des données sans pour autant remettre en cause leur rôle. Les outils d'IA ont besoin de données annotées pour s'entraîner et le pouvoir de diagnostic se limite aux données déjà rencontrées. Ils ont un besoin perpétuel du médecin pour apprendre sur les nouveaux cas ou les nouvelles maladies. C'est le cas par exemple de

Thérapixel³⁴⁴ et de Cardiologs³⁴⁵ deux entreprises françaises, qui ont développé des algorithmes capables de détecter respectivement les lésions cancéreuses du sein à la mammographie et des troubles du rythme cardiaque à l'électrocardiogramme. L'algorithme de Cardiologs a reçu le *marquage CE* et FDA (*Food and Drug Administration*), il s'agit du premier système d'IA à être accrédité dans l'UE. Dans le cas du cancer, des algorithmes d'apprentissage peuvent traiter en un temps record des milliards de données d'un séquençage d'ADN provenant de la tumeur d'un patient, une tâche impossible même pour les médecins les plus expérimentés ; il s'agit d'un gain de temps important associé à une prise en charge rapide du patient. Les logiciels d'aide à la prescription et à la dispensation des médicaments sont déjà utilisés dans les DPI et sont même obligatoires dans le cadre de l'obtention de la certification HAS pour les dossiers informatiques.

Les algorithmes contribuent à la lutte contre la propagation des épidémies. Par exemple, les données de géolocalisation des téléphones mobiles permettent de suivre les mouvements de la population à travers les bornes de transmission et ainsi anticiper le développement de maladies dans un pays. C'est le système qui a été utilisé dans le cadre de la lutte contre la maladie Ebola en Afrique³⁴⁶. Mais, elle met en évidence le problème de la protection de la vie privée des individus à travers ce mode de collecte et les limites de son utilisation. Doit-on la rendre systématique dans le cas des questions de santé publique ou doit-on respecter la vie privée des individus ? Il s'agit d'une question éthique qui mérite une discussion avec tous les acteurs car dans certains cas, l'intérêt public doit l'emporter sur l'intérêt individuel mais l'information des acteurs est toujours nécessaire. Le risque de détournement de la finalité première du dispositif est également un enjeu capital à prendre en compte, d'où l'intérêt d'un tiers de confiance pour la surveillance des usages.

Dans le cadre de la santé individuelle, la collecte des données à partir d'application permettra de faire des suivis à distance et de prédire par exemple des risques d'hospitalisation, de récurrence d'une maladie ou de décompensation d'une maladie chronique en fonction des informations renseignées par le patient. Mais pour un meilleur résultat, il faut que les patients

³⁴⁴ Source : <https://www.therapixel.com/> consulté le 25 janvier 2019.

³⁴⁵ Source : <https://cardiologs.com/> consulté le 25 janvier 2019.

³⁴⁶ Source : <https://www.nouvelobs.com/rue89/rue89-tech/20140825.RUE5400/les-portables-pourraient-predire-les-epidemies-d-ebola.html> consulté le 03 avril 2020.

soient éduqués à l'utilisation des applications et une éducation thérapeutique pour la compréhension de la maladie qui n'est pas aussi aisée. Dans le cas de *Moovcare*³⁴⁷ par exemple, le patient répond à un certain nombre de questions dans le cadre du suivi post thérapeutique du cancer du poumon et l'algorithme permet d'évaluer le risque de rechute. Le praticien peut alors convoquer le patient rapidement pour évaluer son état. L'application rentre dans le cadre d'un processus de soin et le médecin décide de la conduite à tenir en fonction des recommandations. Cette application a été validée par la HAS.

Pour la robotique en chirurgie, c'est un outil qui permet au chirurgien d'opérer de manière plus précise avec peu de complications ; mais elle nécessite une longue période de formation pour une meilleure utilisation. Par ailleurs, son utilisation n'exempte pas du besoin d'une équipe au bloc ;

« On opère sur place dans la même salle avec le moniteur, le robot et la console de manipulation. Aujourd'hui, ça nous apporte un grand confort dans les interventions chirurgicales : une vision en 3D, une capacité de manipulation des instruments à 560 degrés, une réduction des tremblements et un grossissement en 10 fois des images. Seul inconvénient, on n'a pas le palper. Les avantages sont : peu de douleurs, peu de saignements, une morbidité et une mortalité faible, un raccourcissement de la durée d'hospitalisation, une rapidité de mobilité des patients. En général, nous utilisons le robot pour la chirurgie élective »³⁴⁸.

Pour le moment c'est le chirurgien qui manipule le robot et la présence de toute l'équipe en salle opératoire est nécessaire pour la bonne marche de l'intervention.

L'utilisation du *big data* et de l'IA dans le cas des maladies rares à travers la recherche de cas similaires permet de poser des diagnostics précoces et une prise en charge de ces cas. Il s'agit d'une interpénétration de la recherche et des soins pour permettre une utilisation directe des résultats de la recherche à travers le *big data*. C'est le cas par exemple de l'hôpital Necker de Paris à travers l'entrepôt de données Dr Warehouse, qui leur permet de recueillir toutes les données du patient et de les comparer à des cas antérieurs.

³⁴⁷ Source : <https://www.moovcare.com/> consulté le 03 mars 2020.

³⁴⁸ Entretien 20.

La pharmacovigilance est un domaine où l'application du *big data* sont très bénéfique dans le cadre de la surveillance des médicaments après leur mise sur le marché. L'enregistrement des feuilles de soins et des effets secondaires dans les bases de données est une source importante pour l'étude des effets indésirables des médicaments et des vaccins de faible fréquence ou des médicaments à faibles prescriptions³⁴⁹. L'analyse de ces données permettra de retirer certains médicaments du marché avant que des drames sociétaux ne surviennent et de réduire les coûts de la prise en charge des conséquences des effets secondaires de ces médicaments.

Le *big data* auront un apport indéniable dans le domaine de la santé, mais il faut prendre en compte leurs limites et leurs impacts sur la pratique médicale pour tirer le meilleur profit pour les patients. Cela concerne notamment :

- la responsabilité liée à l'utilisation des outils : en effet, l'utilisation du *big data* et des outils de l'IA engage la responsabilité du médecin ; il devra faire attention dans le choix des outils pour ne pas causer de tort à son patient. On pourrait assimiler par exemple la conception des outils comme un avis d'experts ; dans ce cas la responsabilité de l'éditeur pourra être engagée dans le cadre des biais d'édition (faute dans la programmation) ou dans le cadre du dysfonctionnement des outils ayant causé des incohérences dans la prise de décision. Il serait important d'appliquer le régime des dispositifs médicaux aux logiciels d'aide à la décision médicale pour améliorer la qualité et le niveau de confiance.
- l'accès aux outils : la mise en place des outils constitue un investissement important qui dépasse la capacité financière de la plupart des hôpitaux surtout dans un contexte de restrictions budgétaires. Elle entraîne des inégalités d'accès aux technologies de santé aussi bien au niveau des hôpitaux et des individus. Une mutualisation des projets développés au niveau national peut être une solution pour les hôpitaux.
- la mise à jour des plateformes et des outils doit être réalisée dans le cadre de l'évolution de la pratique médicale et des technologies. En effet, l'investissement initial étant important, il faut éviter que les outils ne tombent rapidement en désuétude du fait de l'absence de mise à jour des connaissances du système. La

³⁴⁹ Etudes pharmaco-épidémiologiques publiées sur le site de l'assurance maladie et réalisées à partir de la base de données de la SNIIRAM. Source : <https://www.ameli.fr/l-assurance-maladie/statistiques-et-publications/etudes-en-sante-publique/etudes-pharmaco-epidemiologiques/index.php> consulté le 20 octobre 2019.

mise à niveau du plateau technique peut faire courir un risque important d'atteinte à la sécurité des données et d'erreurs durant la prise en charge des patients en cas de défaillance. Les algorithmes étant construits dans la majorité des cas par des start-up, ceux-ci font faillites ou sont rachetées par des grands groupes qui décident de ne plus assurer la mise à jour ou d'arrêter la production des outils.

- l'utilisation des outils du *big data* et de l'IA peut avoir un impact négatif sur la relation médecin – patient. Il est important de maintenir la participation du patient dans le processus de décision et de renforcer l'obligation d'information et du recueil du consentement du patient dans le cadre de l'utilisation d'un outil de prédiction. Le médecin doit pouvoir expliquer le fonctionnement de l'outil et son choix au patient.
- la généralisation des résultats : les outils de l'IA sont développés dans un contexte particulier pour un groupe d'individus particuliers ; pour généraliser les résultats afin de les appliquer à une large population. Il faut s'assurer que les outils et les résultats obtenus sont toujours valables. Sinon, il y a un risque d'obtenir des résultats erronés avec des conséquences pouvant être néfastes pour les patients et la société.
- les outils du *big data* et de l'IA dans le domaine de la santé concernent en général un domaine bien circonscrit et bien balisé ; or l'humain dans son fonctionnement n'est pas linéaire et est imprévisible. Les médecins devraient par conséquent intégrer les autres aspects de la vie et de la prise en charge du patient dans une approche systémique.

la réduction des dépenses de santé

Le rapport des économies de santé réalisées avec les coûts liés au déploiement technologies de *big data* et de l'IA serait-il favorable ?

On peut espérer : une diminution de la redondance des examens cliniques, un meilleur contrôle des activités cliniques des médecins par la Sécurité sociale et surtout une réduction importante des coûts de prise en charge des effets indésirables des médicaments.

« Par exemple dans le cas du partage des données, l'intelligence artificielle permettra d'éviter les examens redondants et permettra de limiter les dépenses. »

Par contre, les dépenses sont liées aux nouvelles thérapies et aux technologies. L'augmentation des coûts de la santé sera quelque chose de permanente. On pourra faire des économies de dépenses à la marge mais ça ne sera pas des réductions révolutionnaires avec des sommes extraordinaires »³⁵⁰.

Cependant, les médecins prescrivaient les examens complémentaires non seulement pour les besoins de soins des patients mais aussi pour se protéger en cas d'éventuel procès.

« Dans le cadre de la judiciarisation de la société, le nombre des examens complémentaire va augmenter même si ces examens sont normaux. Les médecins vont ouvrir les parapluies pour se protéger. Le big data ne servira à rien car on aura toujours les avocats pour dire qu'on n'a pas fait tel ou tel examen »³⁵¹.

La mise en place des technologies du *big data* et de l'IA nécessite des investissements tels que l'achat ou la location des *data centers*, le développement des algorithmes d'IA et le recrutement de personnels qualifiés, ce qui est onéreux. A cela, s'ajoutent : les frais d'entretien, de mise à jour et surtout de la gestion de la sécurité (confidentialité, disponibilité, intégrité, accessibilité) des systèmes d'information de santé qui sont devenus les cibles privilégiées des pirates.

Ces technologies sont brevetées par des entreprises privées (surtout pour le séquençage génomique ou la robotique) qui fixent librement les tarifs des appareils, des consommables et des logiciels. Les Etats exercent une influence faible sur les fixations des prix entraînant ainsi des spéculations. Par exemple dans le cas de la chirurgie robotique, les pinces du *robot Da Vinci* doivent être changées toutes les dix interventions, entraînant un surcoût de 2000 euros pour chaque patient. La thérapie ciblée engendre des coûts supplémentaires.

Par conséquent, le rapport coût de mise en œuvre et économie des dépenses est donc défavorable. On assiste à une monopolisation des évolutions technologiques par quelques grands groupes ; une augmentation des coûts individuels de prise en charge liées à de nouvelles thérapies coûteuses.

« Pour la réduction des dépenses de santé, je ne me fais pas d'illusion, ça diminuera certaines choses, mais ça augmentera d'autres. La preuve, on nous demande de nous

³⁵⁰ Entretien 8.

³⁵¹ Entretien 10.

informatiser, de payer des logiciels. On change juste les priorités »³⁵².

Pour les médecins, la réduction des dépenses est un numéro d'équilibriste et de redéfinitions des priorités des politiques de santé des gouvernements ;

« En quelque sorte, c'est la course entre les nouvelles dépenses et l'encadrement des dépenses précédentes. On a une course vers la modernisation de la santé et la réduction des dépenses. Les deux doivent être jumelés, l'argent attribué à la santé est limité et il faut trouver l'équilibre entre les deux. Après c'est le souci de savoir ce qui doit être pris en charge par la société et par les individus. C'est une des raisons pour lesquelles certaines maladies ne sont pas bien remboursées. Et quand c'est le cas, les gens se soignent mal »³⁵³.

Le risque de la réduction des dépenses de santé avec le *big data* et l'IA est de considérer l'humain comme un objet à partir duquel on doit tirer le maximum de profit tout en oubliant ce qu'est l'humain, à savoir : les sentiments, l'empathie, la compassion, la tristesse et l'amour. En effet, pour les médecins, l'informatisation du système de santé avait beaucoup plus une logique technico-économique plutôt que médico-intégrative ;

« Le PMSI était au départ un assez bel outil d'épidémiologie avec une bonne description des patients, des trajets et des actes. Depuis que nous sommes passés à la tarification à l'acte, il est devenu un outil médico-économique pour la valorisation financière de séjours.

[...] Par exemple, les codeurs regardent plusieurs sources de données qui vont permettre de retrouver des données qui vont valoriser les actes et les diagnostics. En quelque sorte, nous cherchons tout ce qui peut rapporter de l'argent aux hôpitaux dans la description »³⁵⁴.

³⁵² Entretien 13.

³⁵³ Entretien 3.

³⁵⁴ Entretien DIM.

Encadré 11 : conditions matérielles et techniques de la mise en place du *big data*

« [...] Si nous voulons séquencer les exons complets de 1 000 individus, on doit composer avec 30 To (téraoctets) de données brutes pour des données effectives traitées de 10 Go. Quant au génome complet de 1 000 individus, son étude nécessite un espace de travail de 500 To, pour un volume de données effectivement traitées de 200 Go. Sur le plan du transfert des données, il faudrait un an et demi à un réseau de débit ordinaire (1 Mo/s) pour acheminer 1 000 génomes complets. Il ne faut que 56 jours avec un réseau universitaire, par exemple celui de l'Institut Pasteur à Lille dont le débit est de 10 Mo/s. [...] Le traitement des données est également très exigeant. Si on devait traiter les génomes complets de 1 000 individus, sur un processeur, il faudrait 240 années afin de terminer le processus. Bien évidemment, le calcul s'appuie sur des clusters comprenant de nombreux processeurs pour conduire des tâches en parallèle. [...] Les robots de sauvegarde et le recours à des bandes magnétiques coûtent cher, d'autant plus qu'à l'image des disques durs, les bandes magnétiques ont une durée de vie limitée. On doit recopier les données de bande à bande continuellement, ce qui est également fastidieux. Le cloud a pu être évoqué comme la solution ultime aux problèmes de stockage. Or, sur les volumes de données que nous manipulons, son coût est prohibitif. [...] Une infrastructure de stockage matériel des données est garantie 7 ans. Autrement dit, tous les 7 ans son renouvellement doit être financé. Il va de soi qu'il n'existe aucune sécurité absolue. En cas d'incendie au lieu de stockage, il subsiste toujours un risque de tout perdre »(GRENIER-BOLEY, 2015).

L'autonomie du patient

Dans le cadre de la participation des patients dans les soins, l'utilisation des objets connectés et ses outils de l'IA permettront aux patients de participer de façon active dans le processus de production des données de santé qui serviront pour l'analyse de leur situation. Ces outils pourraient permettre aux médecins et aux patients d'avoir :

- une meilleure connaissance de la maladie ;
- une participation au processus de décision des soins ;
- une meilleure gestion du suivi de la maladie.

Mais, l'*empowerment* des patients pose à l'évidence la question de leur motivation à consacrer du temps, de l'énergie et des efforts à la gestion ou à la valorisation de leurs propres données. Cet *empowerment* est un transfert de responsabilité auquel le patient n'a pas été préparé. La manipulation et la compréhension des données personnelles nécessitent un apprentissage et la mise à disposition d'interfaces très simples et intuitives pour la gestion des données. La place de médecin en tant que médiateur entre ces outils sera encore plus importante pour guider le patient et l'aider à faire les choix dans une quantité importante de données et d'informations ; et éviter ainsi les mauvaises compréhensions et interprétations des informations. Le principal obstacle sera le temps nécessaire pour gérer cette tâche.

« La masse de données disponibles et les technologies vont au contraire rendre le patient plus dépendant de son médecin dans le tri et le choix des décisions. De toutes les façons, l'Homme face à une vulnérabilité comme la maladie a besoin d'un médiateur dans ses choix et le médecin est la personne indiquée dans ces moments »³⁵⁵.

En effet, une étude réalisée dans la population française sur l'utilisation des objets connectés et des outils d'IA montrait des résultats très intéressants à prendre en compte dans le processus de mise en place et d'utilisation des outils dans les soins (TRAN ET AL., 2019). Ces résultats se présentaient comme suit :

- sur 1 183 patients seulement 50 % utilisaient les outils de la *m-health* ;
- seulement 20 % des participants ont estimé que les avantages potentiels de la technologie l'emportaient largement sur ses dangers potentiels ;
- environ 20 % des patients atteints de maladies chroniques s'opposaient à l'utilisation des objets connectés et d'outils fondés sur l'IA dans leurs soins ;
- seule une minorité de participants étaient prêts à intégrer des outils d'IA entièrement automatiques dans leurs soins.

Les auteurs signalaient néanmoins que l'étude a été réalisée avec une population de patients engagés dans une initiative scientifique citoyenne visant à accélérer la recherche sur leurs maladies chroniques (TRAN ET AL., 2019).

³⁵⁵ Entretien 17.

L'utilisation des technologies dans les soins peut être un atout important, mais elles doivent prendre en compte à la fois les besoins et les attentes des médecins et des patients.

la politique de santé

« *La politique de santé a pour but de faire un meilleur usage de l'ensemble des ressources disponibles pour la santé de sorte que toute la population, où qu'elle soit, ait accès à des soins sanitaires de qualité et puisse, en conséquence, vivre plus longtemps, en meilleure santé* »³⁵⁶.

Sa mise en place dépend des déterminants de la santé dans un pays donné. Ainsi, Il est difficile pour les Etats de définir des priorités et de faire des choix faisant consensus face à des enjeux multiples : financement de nouveaux traitements coûteux, vieillissement de la population, risques épidémiques ou environnementaux émergents, exigences accrues en matière de sécurité et de qualité des soins.

Pour les médecins, l'utilisation du *big data* et de l'IA permettrait :

- une meilleure répartition de la prise en charge des pathologies ;
- un meilleur ciblage des maladies à prendre en charge ;
- une modernisation de la santé en fonction des besoins.

Mais, la vraie problématique dans la politique de santé est qu'il s'agit d'un concept qui englobe les demandes de soins, de bien être, de confort et de proximité de soins pour les individus d'une part et d'autre part, une demande de rationalisation, d'économies pour les Etats et les acteurs privés de la santé. Le problème d'indicateurs objectifs pour évaluer les différents besoins des individus et permettre aux Etats de prendre une décision est une question cruciale et difficile à définir. En effet, « *Dès qu'une mesure devient un objectif, elle cesse d'être une bonne mesure* » (STRATHERN, 1997).

Bien que les outils du *big data* et de l'IA permettent d'importantes collectes de données, ils n'éclipsent pas le problème de leur interprétation déjà soulevé plus haut dans notre travail (*Section 5.2.1*).

³⁵⁶ Source : <https://www.who.int/nationalpolicies/about/fr/> consulté le 23 mai 2019.

De l'avis des médecins, la politique de santé est plutôt une volonté du gouvernement de mettre en œuvre les moyens et les outils nécessaires pour une qualité de soins pour les individus. Toutefois, une de leurs craintes est le risque de donner la priorité aux aspects économiques de la santé au lieu de la qualité de la prise en charge. Par exemple, un médecin disait :

« Je pense que les politiques jouent sur les déserts médicaux pour réduire les dépenses. Les endroits où il existe des déserts médicaux vivent les gens les plus vulnérables avec une absence de moyens de transport. Les politiques pensent que les déserts médicaux vont réduire le recours aux soins et faire des économies de manières indirectes tout en faisant semblant de lutter contre le phénomène. Les personnes qui sont dans des déserts médicaux finissent par avoir marre de lutter pour avoir le moindre soin et abandonnent pour mourir par la suite »³⁵⁷.

En effet, une enquête réalisée par France Assos Santé³⁵⁸ sur les difficultés d'accès aux soins en novembre 2019 montrait que deux Français sur trois ont déjà dû reporter ou renoncer à des soins, pour des raisons financières ou faute de médecin disponible. Les difficultés d'accès aux soins touchaient tous les Français, quels que soient leur âge, leur situation économique ou leur lieu de vie. Les répondants déclaraient avoir dû renoncer ou reporter des soins en raison du manque de médecins disponibles dans 49% des cas et pour des difficultés financières dans 45% des cas. Les renoncements touchaient en particulier les malades chroniques (67%) et les personnes en situation de handicap (77%). Soixante-sept pour cent des Français étaient confrontés à des dépassements d'honoraires dans les zones de désert médicales.

L'Etat doit s'assurer de la maîtrise de la gouvernance des données, qui est un enjeu de souveraineté aujourd'hui. Les géants Chinois et Américains sont les champions de la collecte des données et investissent des fonds importants dans le développement des outils de l'IA et du *big data*. Une maîtrise de la gouvernance des données par les Etats permet de maintenir la confiance des acteurs et un accès équitable aux données.

³⁵⁷ Entretien 19

³⁵⁸ Enquête BVA pour France Assos Santé : notre système de santé de plus en plus malade ! Source : https://www.france-assos-sante.org/communiqu%C3%A9_presse/enquete-bva-pour-france-assos-sante-notre-systeme-de-sante-de-plus-en-plus-malade/ consulté le 24 avril 2020.

Dans le cadre de la protection de la vie privée et de l'intimité des personnes, il y a un enjeu géostratégique qu'il faut prendre en compte :

- la valeur marchande des données de santé ne fait qu'augmenter dans un monde où les individus produisent des données sans se rendre compte et qu'il est considéré de nos jours comme « *un produit (une dérivée) de ses données* » plutôt qu'un « *individu produisant des données* » ;
- Malgré la protection des données en France et dans l'UE, des firmes étrangères les exploitent sans respecter la loi. L'utilisation des « *marchés bifaces* »³⁵⁹ (STOEKLÉ ET AL., 2018) permettent d'appâter les utilisateurs en offrant des services que les Etats n'offrent pas;
- l'intérêt individuel, de l'intérêt collectif et l'intérêt général dans la protection des données constituent un enjeu démocratique.

En France, l'Etat a mis en place l'Agence Nationale de la Sécurité des Systèmes d'Information (ANSSI)³⁶⁰ en 2015 et qui répond aux nouveaux enjeux nés des évolutions des usages numériques et des menaces qui y sont liées avec cinq objectifs :

- garantir la souveraineté nationale ;
- apporter une réponse forte contre les actes de cyber malveillance ;
- informer le grand public ;
- faire de la sécurité numérique un avantage concurrentiel pour les entreprises françaises ;
- renforcer la voix de la France à l'international.

L'intégration du numérique, qui apporte de nombreux avantages à la prise en charge des individus, doit prendre en compte la dimension sociale, économique, et culturelle des individus. Cette intégration ne pourra pas se faire sans l'implication et la participation des acteurs concernés et c'est le pacte social qui devra être le ciment de cette politique de santé. L'État doit pouvoir proposer une « *assurance santé* » et non une « *assurance maladie* » qui prend en charge le patient dans sa globalité.

³⁵⁹ Offre gratuite de services aux utilisateurs dans une face en termes d'échange et d'usage et une seconde face auprès des entreprises et institutions où les services sont offerts avec des contreparties, rattrapant ainsi les pertes faites dans la première face.

³⁶⁰ Source : <https://www.ssi.gouv.fr/> consulté le 21 février 2020.

6.9 Finalités éthiques des données dans le cadre du *big data* et de l'intelligence artificielle

Nous avons utilisé le questionnaire mis en place par Jérôme Béranger (BÉRANGER, 2016) pour l'évaluation des finalités éthiques du *big data*.

La transformation des modalités de réponses en score nous a permis de calculer les scores unitaires d'éthique. Le score était en faveur de l'atteinte de la finalité éthique du *big data* lorsqu'il était supérieur ou égal à 75.

Le score unitaire d'éthique globale était égal à 67 ± 14 avec des extrêmes de 43 et 99. Le score éthique était supérieur ou égal à 75 dans seulement cinq cas.

Le tableau 13 montre la médiane du score unitaire d'éthique en fonction de l'ancienneté des médecins dans la pratique médicale. La différence entre les différents groupes n'était pas statistiquement significative ($\text{Chi}^2 = 3,48$, $p = 0,17$).

Tableau 13 : médiane du score unitaire éthique en fonction de l'ancienneté dans la pratique médicale

| Ancienneté dans la pratique médicale | Médiane | 1 ^{er} quartile – 3 ^e quartile |
|--------------------------------------|---------|--|
| < = 10 ans | 77 | 69 – 81 |
| > 10 ans et < = 20 ans | 63 | 63 – 71 |
| > 20 ans | 57 | 53,5 – 70 |

6.9.1 *Bienfaisance*

Le principe de bienfaisance se réfère à l'obligation d'agir pour le bien d'autrui. Dans le cadre de la pratique médicale, il s'agit pour le médecin d'apporter le bienfait des techniques médicales et de leurs utilités pour le patient. Il doit évaluer ainsi les avantages et les inconvénients dans l'intérêt supérieur de son patient (BEAUCHAMP, CHILDRESS, 2009). Ce principe engage ainsi la responsabilité du médecin dans le choix des techniques et des outils. Tout acte médical sur le patient doit chercher le maximum de bénéfice possible pour le patient.

La disponibilité des données dans le cadre de processus de soins est très importante pour la prise en charge des patients et la continuité des soins. Ainsi, la continuité de soins est un gage de qualité d'un système de santé et nécessite un système de partage d'informations entre les différents acteurs du système. Avant le développement du numérique, le partage d'information se faisait à partir de lettre de référence ou de demande d'avis avec les informations nécessaires pour prendre en charge des patients ; le médecin traitant ne donnait que les informations utiles pour la prise en charge du patient.

Avec le numérique des entrepôts de données ont été développés au sein des hôpitaux pour assurer une meilleure accessibilité des données et des demandes de soins à travers les systèmes d'information des hôpitaux. Ils permettent ainsi une meilleure coordination des soins entre les professionnels de santé et une planification efficace des services (par exemple l'occupation des lits, la gestion des blocs opératoires ou des appareils d'imagerie). Cette organisation permet de fluidifier la gestion des patients pour leur bien-être avec un gain de temps dans la gestion pluridisciplinaire du patient à l'hôpital. Mais de façon pratique, il se pose la gestion des habilitations au sein des hôpitaux qui entraîne ainsi la perte de la confidentialité des données et du secret médical. Ces lacunes ne sont pas seulement d'ordre technique, mais aussi liés aux comportements des personnes concernant le partage des mots de passe ou l'utilisation d'outils informatiques extérieurs à la structure.

Mais cette numérisation est confrontée à des problèmes d'interopérabilité technique. Il n'y a pas un modèle national unique de système d'information hospitalier. Chaque hôpital développe son système en fonction des moyens disponibles en son sein avec une entreprise de son choix. De ce fait, les données sont cloisonnées au sein d'un même hôpital et ne peuvent pas suivre les mouvements du patient qui doit disposer d'une copie pour la présenter au médecin traitant. En plus, il existe une absence de connexion entre la pratique libérale et la pratique hospitalière sur le même territoire réduisant ainsi les possibilités de partage d'information entre professionnels de soins et entravant ainsi la continuité des soins en ville. Le DMP constitue une alternative mais les difficultés expliquées plus haut (*section 6.6.5*) font que son utilisation dans le cadre de la continuité des soins reste problématique.

Le seul point commun entre les différents systèmes est la transmission des feuilles de soins et les codages PMSI dans le cadre des remboursements. En effet, le PMSI est de nos jours un outil plutôt économique pour le remboursement des soins des hôpitaux. Ainsi la qualité des données de santé ne sont pas adaptées à une utilisation en IA et le *big data* nécessite des

outils spécifiques dans le domaine de la santé. En effet, des équipes en particulier en France travaillent sur des systèmes pour une meilleure extraction de données de santé cohérentes dans le cadre du *big data* (LELONG, 2019). Il n'a jamais eu une vocation épidémiologique. C'est un grief qui lui a été fait dès le début par les cliniciens qui ne se voyait pas comme des artisans d'une comptabilité analytique et qui sous-traitaient à des petites mains le codage des RSS (résumé standardisé de sortie). Il en résulte la piètre qualité des données et leur inutilité en épidémiologie (COLLOC, VINOT, 2000). Elles ne sont donc pas utilisables en IA ou pour déployer des études à l'aide du *big data*. En effet, des équipes, en particulier en France, travaillent sur des méthodes et des systèmes pour une meilleure extraction de données de santé cohérentes dans le cadre du *big data* (LELONG, 2019).

L'utilisation des données de mauvaises qualités dans l'apprentissage des algorithmes expose à des risques de biais, pour la santé du patient, de responsabilités pour le médecin et de conséquences sur la santé publique pour toute la société. Il s'agit donc de construire des systèmes de gestion de données qui sont fiables, intègres, vérifiables et accessibles par tous les acteurs. Ainsi, le médecin doit s'assurer de la validité de l'algorithme et de l'apport d'un bénéfice avant son utilisation. Le respect des recommandations et des protocoles de bonne pratique clinique sont très importants dans ce cas et doivent être intégrés dans le développement des algorithmes afin d'assurer de meilleures garanties de résultats et d'utilisation.

Les algorithmes ont besoin des données annotées (métadonnées) pour s'entraîner quel que soit le mode d'apprentissage supervisé ou non. L'harmonisation des annotations des données de santé constitue une priorité absolue pour une amélioration de la qualité des données et une optimisation des algorithmes surtout quand ils sont non supervisés. Dans le domaine de la médecine, il existe plusieurs ontologies. Chacune comporte des forces et des faiblesses. Les plus utilisées sont la CIM 10 développée par l'OMS et le SNOMED.

Le score unitaire d'éthique de bienfaisance était égal à 66 ± 16 avec des extrêmes de 37 et 97. Le score éthique était supérieur ou égal à 75 dans seulement six cas.

Les médecins pensaient que l'utilisation des données dans le cadre du *big data* permettrait d'améliorer la continuité des soins (85 %) et d'œuvrer pour le bien des patients (75 %). Par contre, seulement 40 % des médecins pensaient que l'utilisation des données permettrait de

promouvoir la qualité, l'organisation, la gestion et la planification de la prise en charge des patients.

Les figures 88 et 89 montrent les détails des réponses des médecins par item du questionnaire relatif à l'utilisation des données de santé dans le cadre de la bienfaisance.

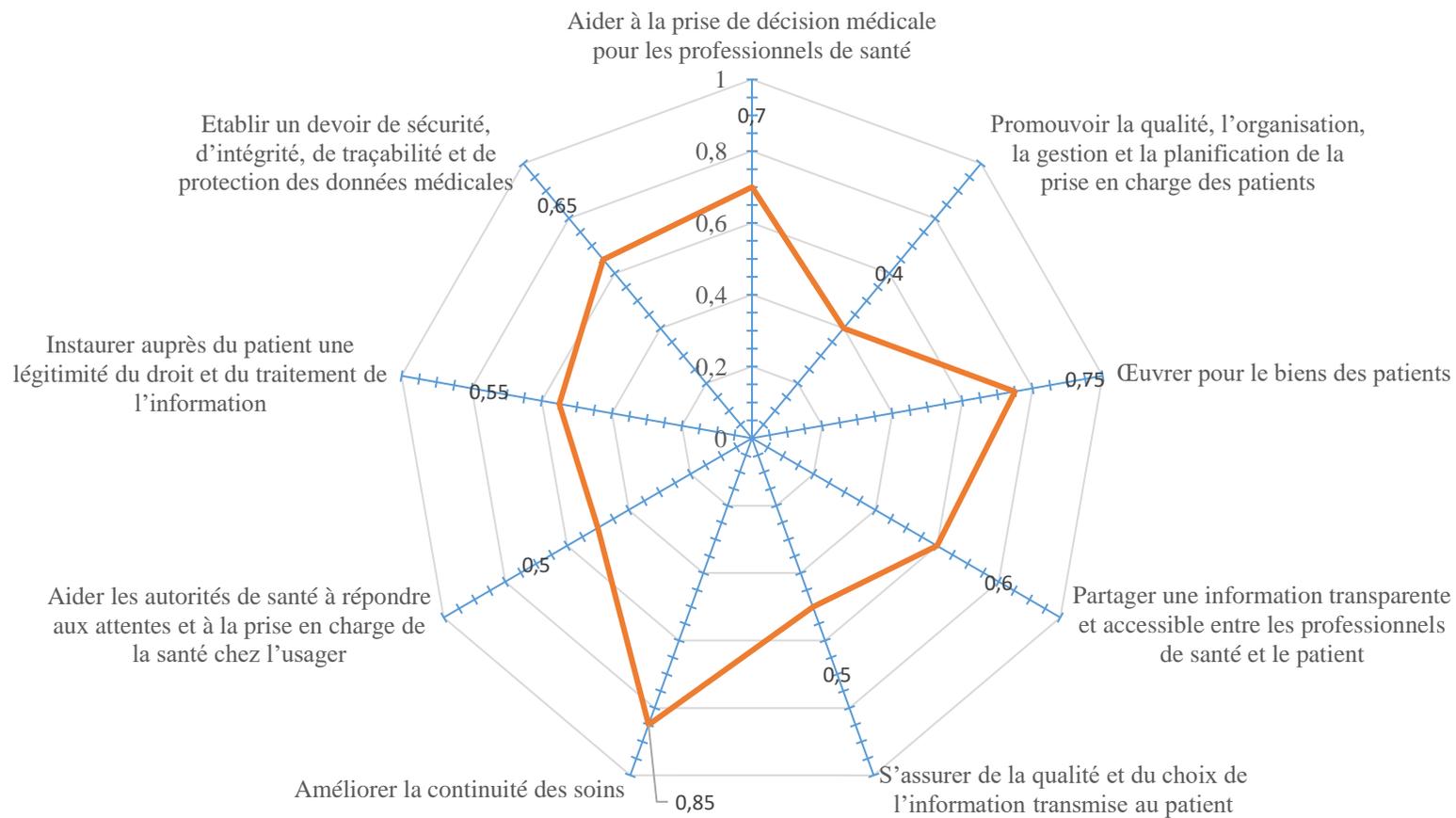


Figure 88 : finalités éthiques de l'utilisation des données de santé dans le cadre de la bienveillance

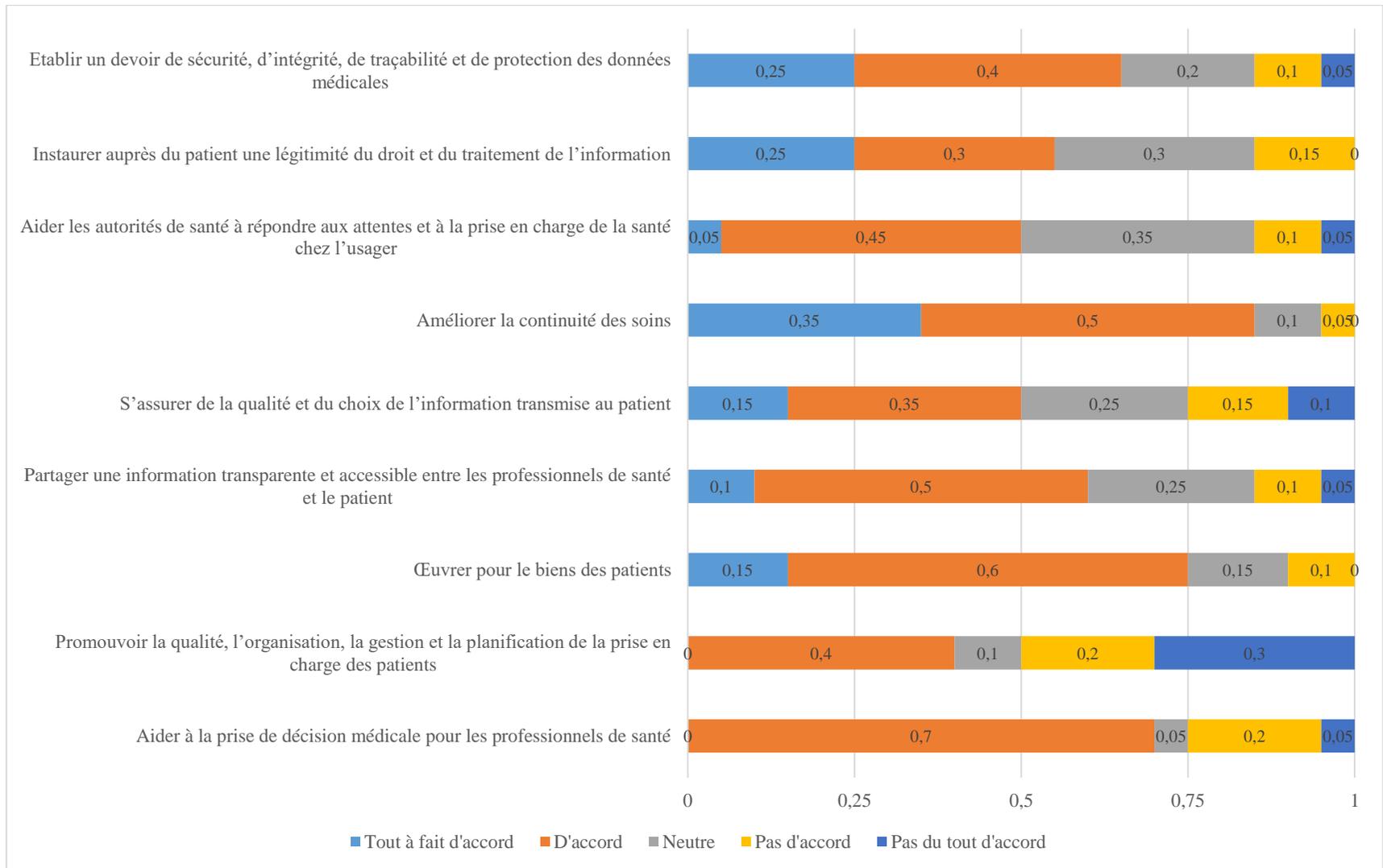


Figure 89 : détails des réponses des médecins par item de questionnaire relatif à la bienfaisance

6.9.2 Justice

Le principe de justice a pour objectif de trouver le juste équilibre pour améliorer l'accès aux soins et aux technologies en prenant en compte le patient, la société dans laquelle il vit et les ressources disponibles pour les soins.

L'analyse des données de soins au niveau national permet d'analyser la demande de soins dans les hôpitaux et les villes et mieux cibler les besoins des patients pour des actions plus efficaces en termes de disponibilité de soins et de qualité. Elle permet également d'évaluer les interventions de l'Etat en matière de politique de santé. Le choix et l'utilisation des données pertinentes disponibles dans les entrepôts de données permettront une meilleure allocation des ressources dans le système de prise en charge des patients et améliorer ainsi leur efficacité et leur efficience.

Mais la crainte des médecins est que les outils puissent être utilisés comme des outils de surveillance et de coercition contre les médecins ou de surveillance des patients. Par exemple, ce sont les répressions et les sanctions faites aux médecins depuis l'informatisation de la rédaction des arrêts maladie ou de travail³⁶¹ ou l'utilisation de la base de données des victimes d'attentats pour identifier les blessés lors de la crise des gilets jaunes à Paris³⁶². Ces constats créent des méfiances auprès des différents acteurs et rendent difficile une exploitation des avantages du *big data* et de l'IA. La médecine ne peut se pratiquer sans la confiance entre les acteurs et ces dérives dans un contexte de numérisation de tout le processus de soins ne sont pas faites pour améliorer leur appropriation. Les médecins sont disponibles pour que les données de soins et de prescriptions soient utilisées pour améliorer les performances du système de santé et offrir une meilleure qualité de soins mais pas constituer un outil de répression.

La mise en place des groupements hospitaliers de territoire (GHT) devrait permettre le décroisement des hôpitaux et l'interconnexion de tous les acteurs des soins (pratique libérale, pratique hospitalière, pratique médico-sociale). Mais, il faudra une vision systémique du parcours et une mise à contribution des acteurs.

³⁶¹ Source : <https://www.mgfrance.org/actualite/profession/1976-les-medecins-generalistes-et-les-arrets-de-travail-encore-et-toujours-seuls-et-responsables> consulté le 22 octobre 2019.

³⁶² Source : <https://www.capital.fr/economie-politique/lintrigant-fichage-des-gilets-jaunes-par-les-hopitaux-de-paris-1335543> consulté le 22 octobre 2019.

Le *big data* et l'IA permettent de développer des nouvelles techniques de prise en charge telles que la robotique ou la thérapie ciblée avec le séquençage des génomes. Les coûts de ces techniques limitent également l'accès à des hôpitaux qui ne disposent de moyens pour les acquérir ou à des personnes ne disposant pas de ressources nécessaires pour se prendre en charge. Cela crée ainsi une inégalité sociale d'accès aux soins sur le plan territorial et individuel. Le vrai problème est de trouver un équilibre entre les besoins et les ressources disponibles ; entre les allocations de ressources et les besoins de rationnement. La prise en charge de plus en plus chères et qui concernent un faible nombre de pathologies, notamment les cancers, pose aux médecins la question du choix des priorités de la politique de santé du gouvernement.

« En France, on fait de la chimiothérapie de compassion. On ne rallonge pas la vie du patient, mais les semaines qui restent à vivre sont malheureuses. Il faut gérer la fin de vie. Il s'agit de décision éthique et des décisions politiques. Il faut un bon usage des ressources. La santé n'a pas un prix mais à un coût, que tout le monde va supporter ou alors seuls les riches pourront se soigner »³⁶³.

L'utilisation de la thérapie fondée sur le génome est une avancée mais risque de créer une discrimination des individus et des groupes ethniques. Le génome étant une carte d'identité unique pour chaque individu avec des traits particuliers à chaque groupe ethnique, la génomique pourrait entraîner une discrimination d'accès des personnes ou des groupes ethniques les moins nantis. Il faudra trouver des solutions pour rendre disponible et accessible les soins à ces groupes et cela va nécessiter un investissement et une prise en charge de la majorité des coûts par l'Etat.

Le score unitaire d'éthique de justice était égal à 68 ± 12 avec des extrêmes de 50 et 100. Le score éthique était supérieur ou égal à 75 dans seulement six cas

L'utilisation des données de santé dans le cadre du *big data* permettrait de faire des analyses épidémiologiques ou statistiques (100 %), d'évaluer les performances et cerner les domaines où l'action s'impose (85 %) et d'améliorer et renforcer l'interactivité avec les acteurs extérieurs à la structure de soins (80 %) selon les médecins. Par contre, seulement 30 % des médecins pensaient qu'elle permettrait de renforcer la disponibilité des médecins.

Les figures 90 et 91 montrent les détails des réponses des médecins par item du questionnaire relatif à l'utilisation des données de santé dans le cadre de la justice

³⁶³ Entretien 11

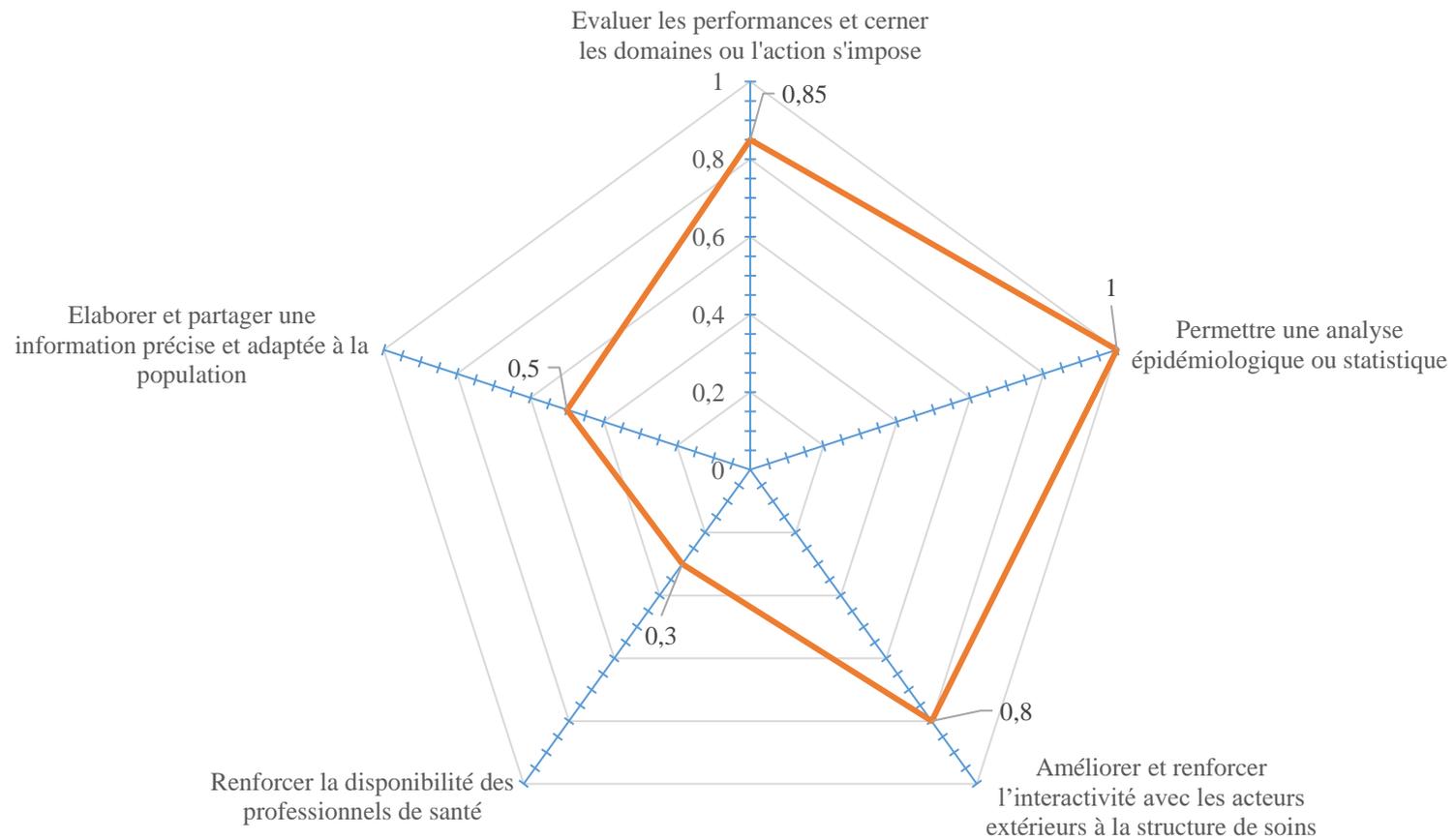


Figure 90 : finalités éthiques de l'utilisation des données de santé dans le cadre de la justice

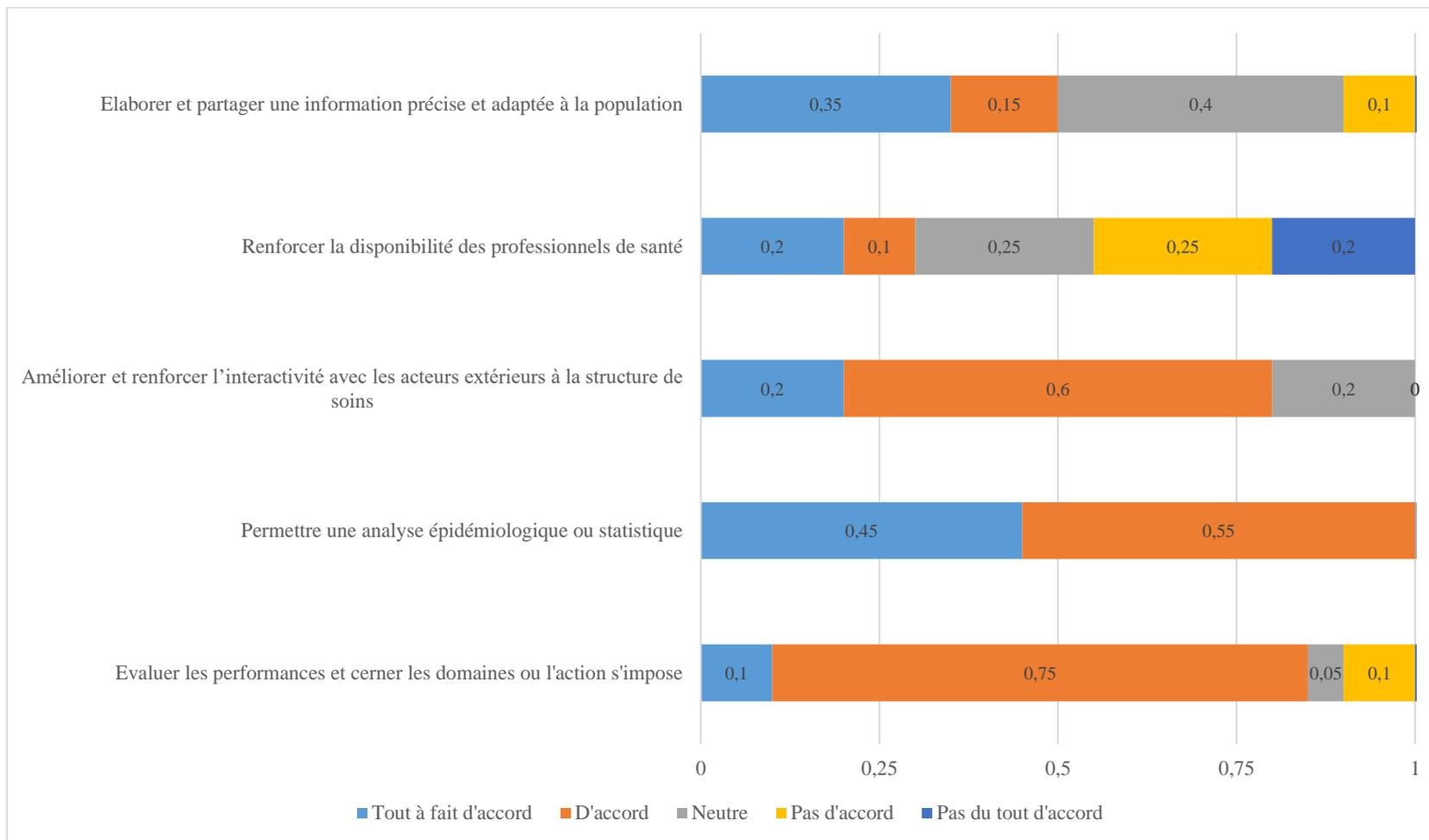


Figure 91 : détails des réponses des médecins par item de questionnaire relatif à la justice

6.9.3 *Non-malfaisance*

Le principe de non-malfaisance a pour objet principal de s'abstenir de nuire. Le médecin dans le cadre de la pratique de la médecine a une obligation éthique de ne pas nuire à son patient notamment par les outils et les techniques qu'il utilise. La question se pose effectivement dans le cas des techniques du *big data* et de l'IA. Dans le cadre des médicaments et des dispositifs médicaux, des essais cliniques randomisés sont réalisés pour valider leurs apports dans la prise en charge des patients et l'innocuité lors de l'utilisation ; la médecine fondée sur les preuves. Ainsi, les techniques du *big data* et des outils de l'IA devront être validés comme dispositifs médicaux pour que les médecins les utilisent sans engager leur responsabilité médicale et surtout respecter article 14 du code de déontologie médicale :

« Les médecins ne doivent pas divulguer dans les milieux médicaux un procédé nouveau de diagnostic ou de traitement insuffisamment éprouvé sans accompagner leur communication des réserves qui s'imposent. Ils ne doivent pas faire une telle divulgation dans le public non médical ».

Mais les systèmes de *deep learning* sont confrontés à la question de l'explicabilité et l'interprétabilité des algorithmes. En effet, avec les algorithmes de *deep learning*, il est impossible de mettre en relation des données ou des caractéristiques avec le résultat de la décision. Les chercheurs connaissent les données en entrée et les résultats en sortie mais ne peuvent pas expliquer les méthodes et les processus utilisés par l'algorithme pour aboutir aux résultats. Or, dans la pratique médicale, le médecin doit pouvoir expliquer toutes les méthodes utilisées et le processus d'obtention des résultats aux patients avant d'obtenir leur consentement. Des études sont en cours pour essayer de résoudre ce problème qui constitue un frein de l'utilisation de l'IA dans le domaine médical. Par contre, d'autres techniques d'IA, dont celle que nous avons utilisées, sont fondées sur les objets cliniques de la sémiologie médicale et la chronologie des événements que l'on peut expliquer au patient.

Le score unitaire d'éthique de non-malfaisance était égal à 69 ± 16 avec des extrêmes de 38 et 100. Le score éthique était supérieur ou égal à 75 dans dix cas.

Les médecins pensaient qu'il faudrait s'assurer de la pertinence technique et du bien-fondé humain des outils du *big data* et de l'IA (70 %) et que l'utilisation des données dans le cadre du *big data* et de l'IA permettrait d'assurer la fiabilité des données médicales et leurs

permanences (80 %). Par contre, seulement 45 % des médecins pensaient qu'elle permettrait de minimiser ou éliminer des torts faits aux patients du fait d'une mauvaise information.

Les figures 92 et 93 montrent les détails des réponses des médecins par item du questionnaire relatif à l'utilisation des données de santé dans le cadre de la non-malfaisance.

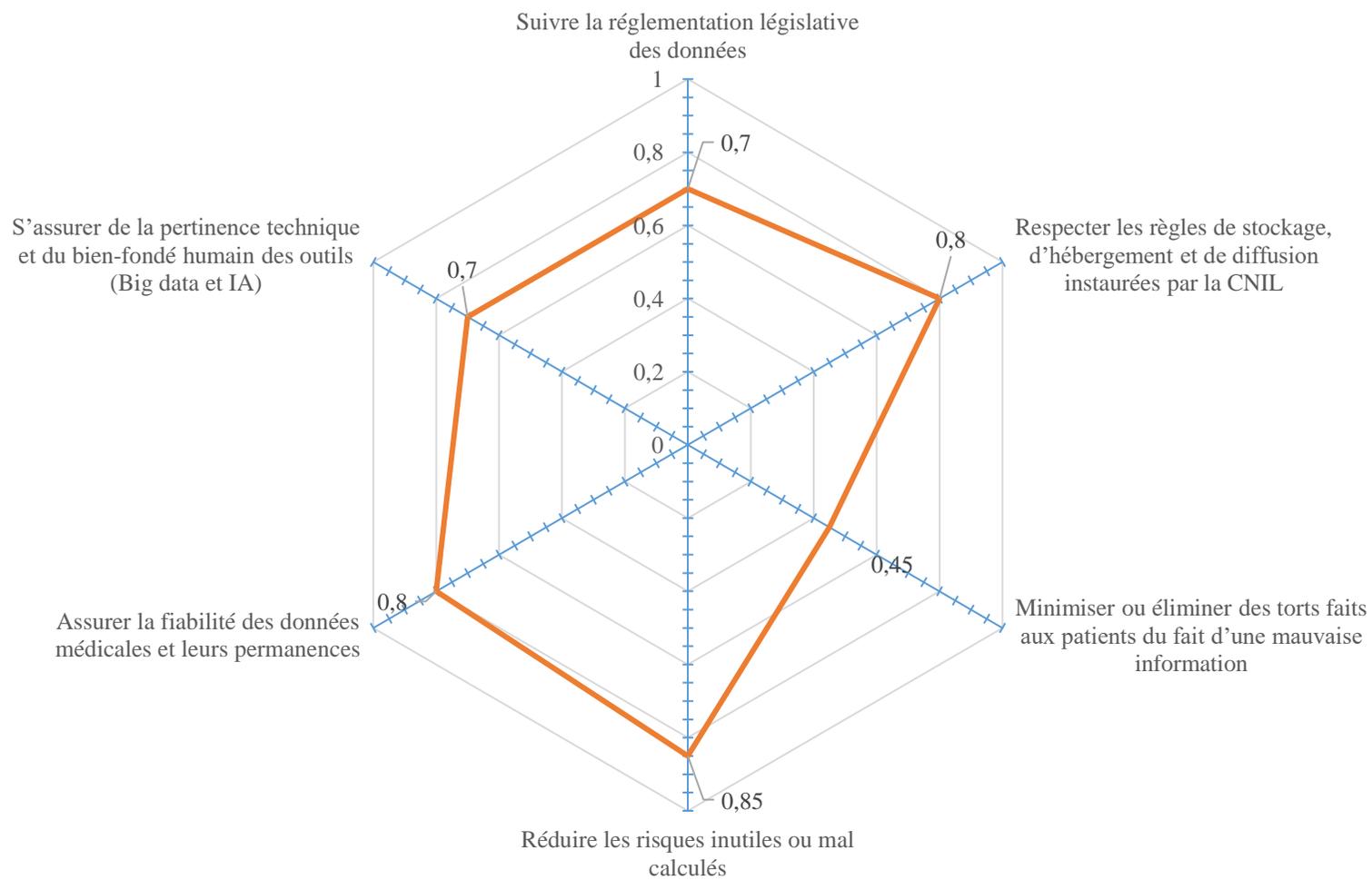


Figure 92 : finalités éthiques de l'utilisation des données de santé dans le cadre de la non-malfaisance

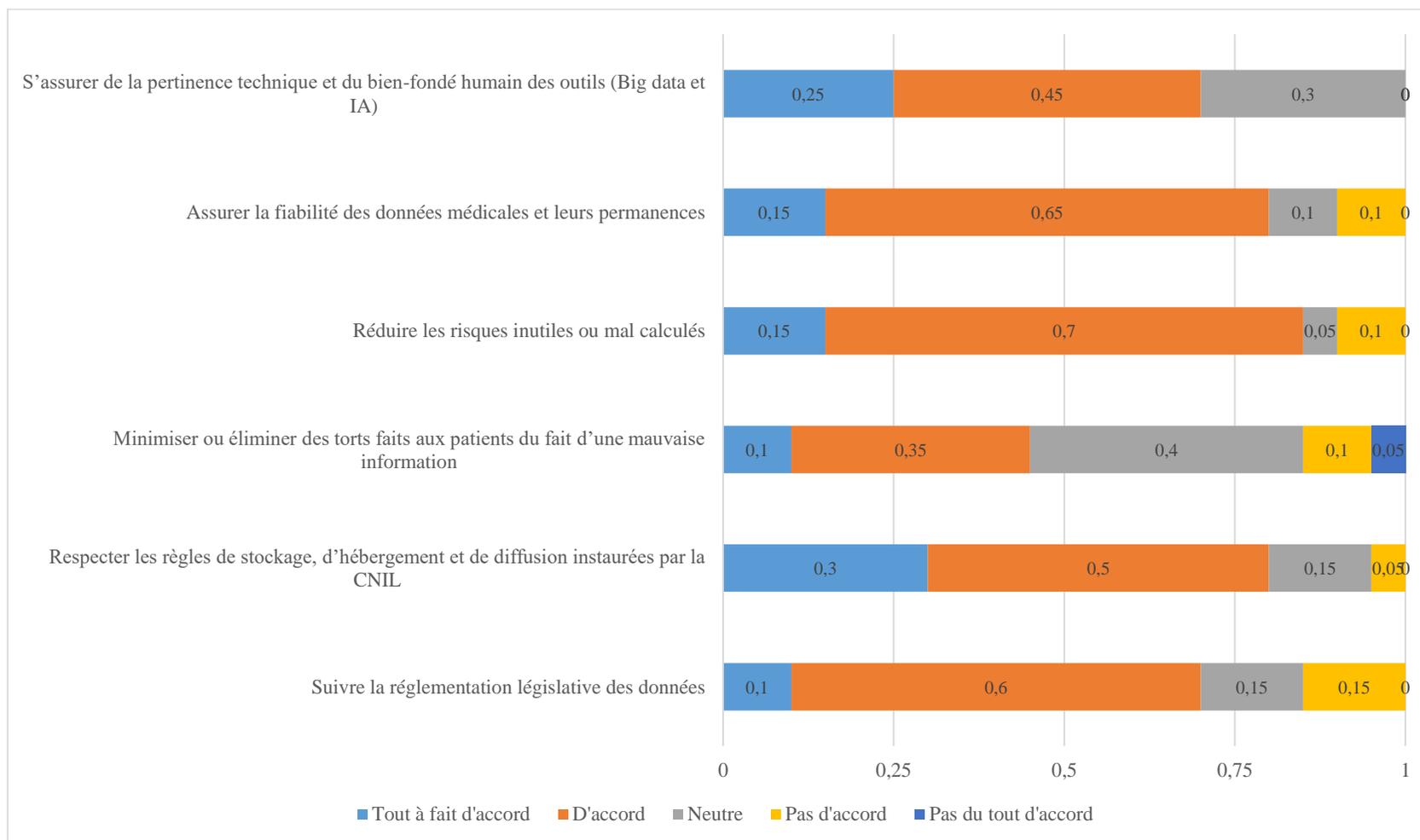


Figure 93 : détails des réponses des médecins par item du questionnaire relatif à la non-malfaisance

6.9.4 *Autonomie*

« *Nul ne sera l'objet d'immixtions arbitraires dans sa vie privée, sa famille, son domicile ou sa correspondance, ni d'atteintes à son honneur et sa réputation. Toute personne a droit à la protection de la loi contre de telles immixtions ou de telles atteintes* » Article 12 de la déclaration universelle des droits de l'Homme, Nations Unies, 10 décembre 1948.

Cette déclaration est plus que d'actualité de nos jours avec le développement du *big data* et de l'IA dans notre société. Le respect de la dignité de la personne humaine dépasse le cadre de la loi ; il pose des questions éthiques sur l'autonomie de la personne humaine.

Le respect de l'autonomie est l'un des principes fondamentaux de l'éthique médicale. L'autonomie du patient est la base du respect de la dignité de l'Homme. Quelles que soient les conditions, cette dignité doit être respectée. Le consentement du patient à utiliser ses données est la base du respect de l'autonomie. Le consentement doit être libre et éclairé pour une action bien précise. Le recueil du consentement nécessite une information préalable du patient (pas toujours évidente dans le cadre du *big data* et de l'IA) avec une possibilité de retirer son consentement à tout moment. Le problème du *big data* est qu'il s'agit d'une collecte systématique des données des patients dans les bases de données et sur les réseaux sociaux sans finalité préalable de traitement. La collecte n'est pas de ce fait transparente et loyale d'autant plus que le patient ne sait même pas que ses données sont collectées. Respecter l'autonomie, c'est également demander le consentement des patients même si les données sont en accès libre à partir du moment où ce sont des données personnelles. Ce n'est pas parce que les données sont disponibles que leur utilisation est libre.

Le respect de l'autonomie des patients concerne également la vente des données personnelles ou le partage entre entreprises dans le cadre de partenariat commercial. En effet, il est très difficile pour le patient et même pour les autorités de protection de loi de suivre la traçabilité des données au sein des entreprises ou des institutions de recherches. Il faut donc compter sur l'éthique individuelle des chercheurs ou des entreprises de respecter leur engagement. Mais la valeur monétaire des données et la rareté des ressources peuvent faire que les entreprises ne respectent pas leur engagement initial.

Comme la vente des données, l'achat des données pour entraîner des algorithmes constitue également une atteinte à l'autonomie et à la dignité des personnes. Les dispositifs conçus avec ces données ne respectent pas non plus la valeur et les droits humains même si elles

offrent des résultats meilleurs. Faut-il les accepter du fait de leurs résultats dans le cadre de la pratique médicale ou faut-il les rejeter pour le respect des valeurs et des droits humains ? Les lois sur la protection des données ne disent rien sur l'achat des données par les entreprises européennes. Les lois ne pourront pas réguler tous les cas possibles d'utilisation ; mais la réflexion éthique nous permettra de faire des avancées majeures dans l'utilisation des données.

Le respect de la protection de la vie privée, de la confidentialité des données et du secret médical sera de plus en plus difficile dans un monde où tout est connecté. Des études ont montré que les données peuvent être réidentifiées par des recoupements avec des bases de données et par des méthodes d'IA, même si elles sont anonymisées (KONDOR ET AL., 2018 ; NARAYANAN, SHMATIKOV, 2008). D'autres études ont montré également que si le processus d'anonymisation des données est correctement fait, les données perdent en qualité lors du processus d'analyse (DOMINGO-FERRER ET AL., 2016 ; BERGEAT, 2014 ; TANGHE, GIBERT, 2017). Toute la difficulté réside dans le fait qu'en modifiant la véracité des données, l'exploitation des données est faussée, donnant lieu à des résultats d'analyse moins fiables. Appliquées à la recherche médicale, ces techniques conduisent donc à rendre des jeux de données anonymes inutilisables, car potentiellement dangereux pour la production des connaissances et dans le cadre des décisions relatives à la santé des patients.

Aucune méthode d'anonymisation des données satisfaisante ne permet de combiner à la fois la richesse des données anonymisées, un faible coût de prétraitement et une protection infaillible de la vie privée surtout dans le cas du *big data* (BERGEAT, 2014 ; DOMINGO-FERRER ET AL., 2016 ; TANGHE, GIBERT, 2017). La protection des données personnelles ne peut se faire qu'entre un compromis entre l'utilité et la protection de la vie privée ; d'où l'intérêt de l'implication des acteurs dans la prise de décision.

Le mode de recueil du consentement dans le cadre de la pratique médicale reste très difficile et coûteux pour les acteurs et n'est pas adapté dans le cadre du *big data* et de l'IA. Des auteurs ont proposé des nouveaux modes de consentement à savoir le consentement large ou le consentement global qui consiste à autoriser toutes les utilisations possibles des données dès la collecte sans définition claire des objectifs de l'étude dès le départ (HOFMANN, 2009 ; GRADY ET AL., 2015). Mais, ce mode de consentement pose problème sur le plan juridique car tout traitement de données de santé doit être fait dans le cadre d'un objectif clair avec une durée de conservation bien déterminée. Ce consentement ne respecte aucun principe du

RGPD. Il n'est pas adapté dans le cadre des changements de situation tels que la modification de la loi ou la rétractation du patient après avoir donné son avis.

D'autres auteurs ont proposé le consentement dynamique qui consiste à recueillir le consentement des patients en ligne à l'aide d'une plateforme de communication avec le choix pour le patient de définir les usages possibles de ses données (BUDIN-LJØSNE ET AL., 2017 ; KAYE ET AL., 2015). Ce type de consentement pose non seulement le problème de la compréhension et de la responsabilité du patient ; mais aussi la suppression de toute médiation humaine directe dans le cadre du consentement et la capacité à manipuler la technologie. Il entraîne non seulement une inégalité dans la participation à la recherche et porte une atteinte à la relation entre le médecin et son patient. Une éducation de la population aussi bien sur la responsabilité de leur acte et sur l'utilisation des technologies s'avère nécessaire afin de mettre en place un système de consentement en ligne.

Il faut reconnaître que l'utilisation des données de santé dans le cadre du *big data* et de l'IA pose d'énormes difficultés, mais il faudra trouver le juste milieu pour profiter des avantages de ces outils pour l'amélioration de la prise en charge des patients. La limitation de la collecte des données est utile pour l'atteinte des objectifs et respecter les règles de partage est très important pour le maintien et la confiance entre les acteurs. En plus, les rôles des différents acteurs de la santé doivent être clairement définis.

Le score unitaire d'éthique d'autonomie était égal à 66 ± 20 avec des extrêmes de 40 et 100. Le score éthique était supérieur ou égal à 75 dans seulement cinq cas.

Seulement 35 % des médecins pensaient que l'utilisation des données de santé dans le cadre du *big data* et de l'IA permettrait de respecter la vie privée et le droit au secret médical et à la confidentialité et 45 % pensaient qu'elle permettrait de replacer le patient au centre de la décision en lui apportant une information médicale plus complète et rapide.

Les figures 94 et 95 montrent les détails des réponses des médecins par item du questionnaire relatif à l'utilisation des données de santé dans le cadre de l'autonomie.

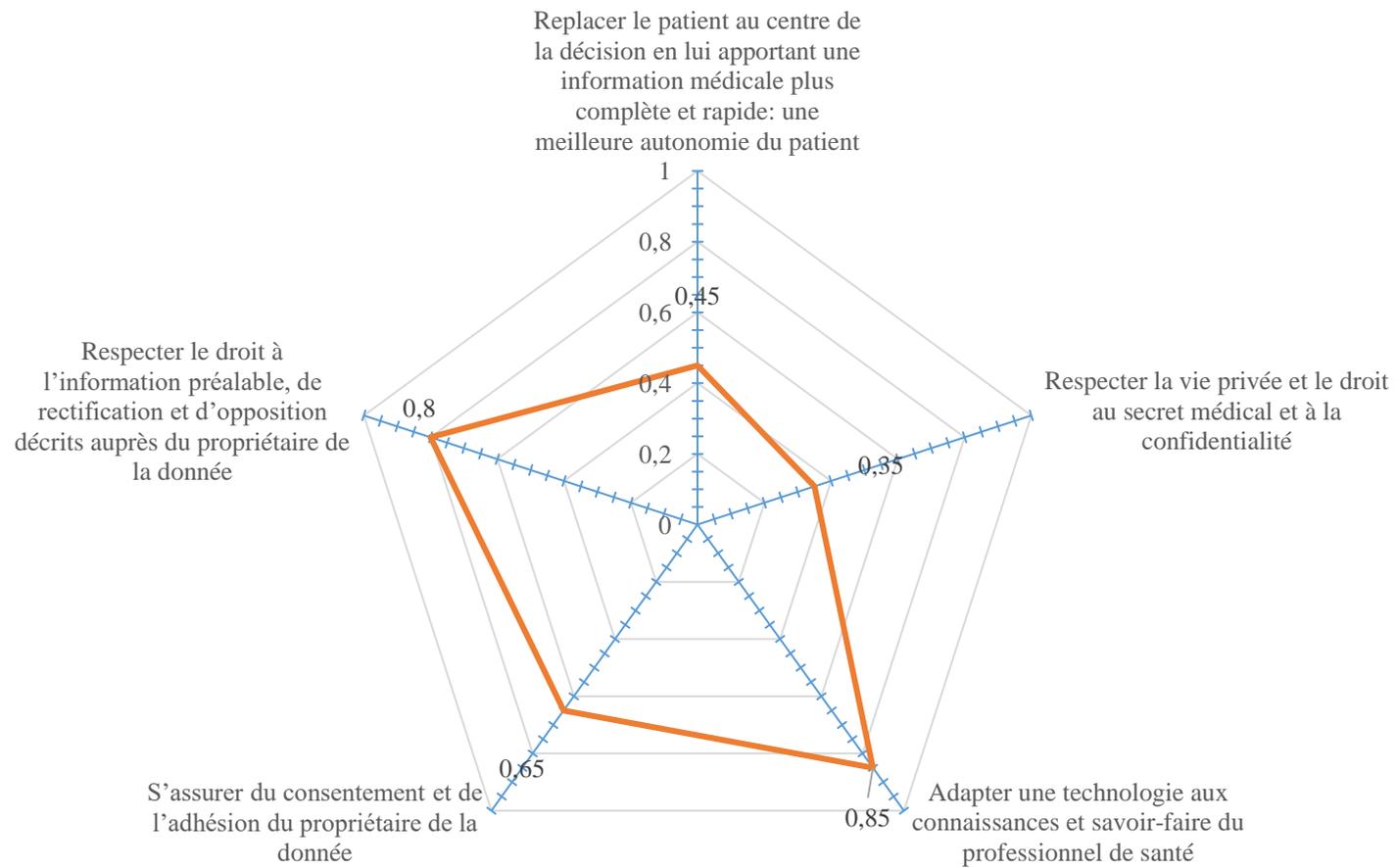


Figure 94 : finalités éthiques de l'utilisation des données de santé dans le cadre de l'autonomie

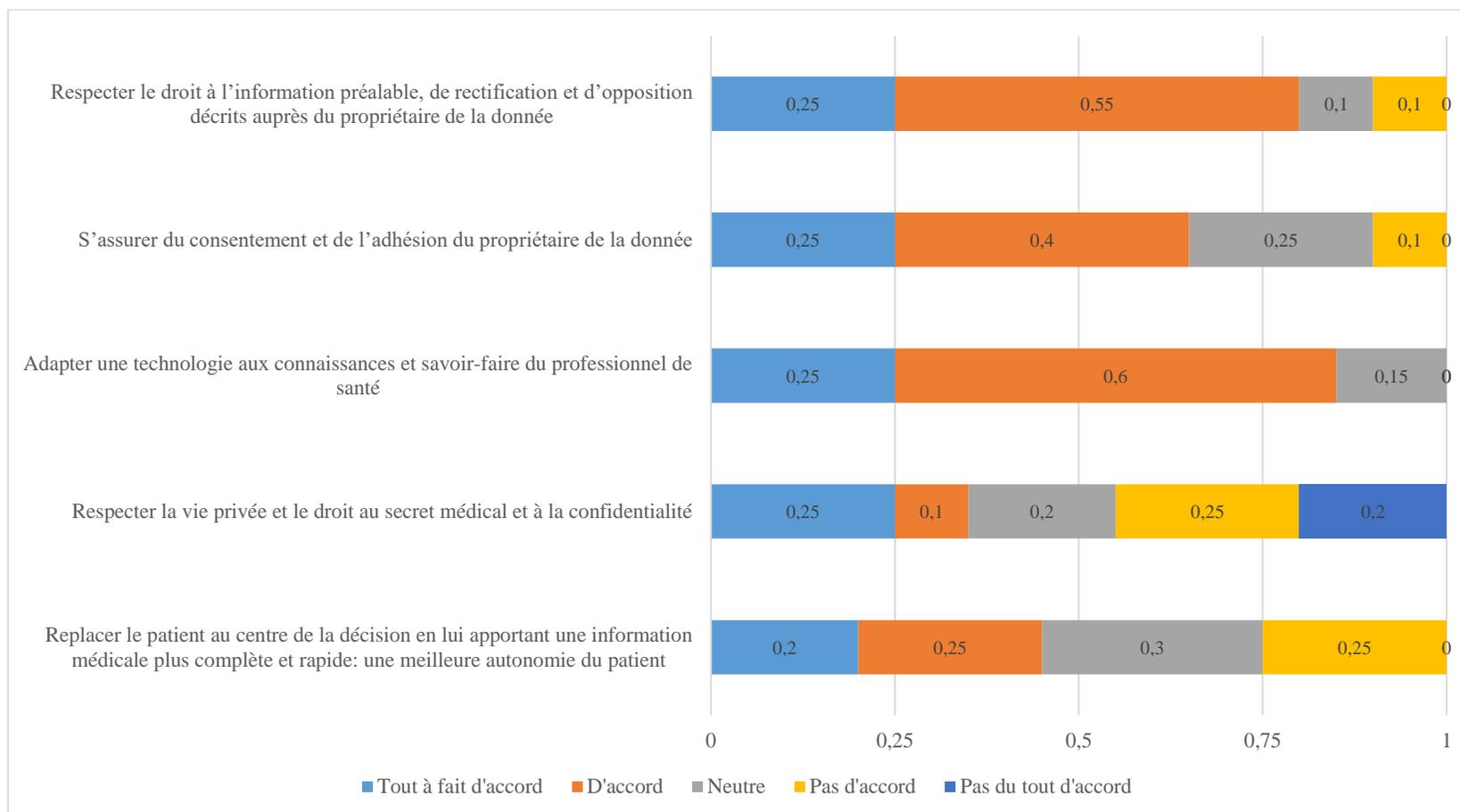


Figure 95 : détails des réponses des médecins par item du questionnaire relatif à l'autonomie

L'utilisation du *big data* et de l'IA dans la pratique médicale comporte des risques pour l'éthique médicale tels que :

- la disparition du consentement des patients dans l'utilisation des données de santé avec une perte de leur autonomie et de leurs droits fondamentaux ;
- la perte de la confidentialité des données par la multiplicité des acteurs et des accès aux données avec une atteinte à l'intimité et à la vie privée des individus et une disparition totale du secret médical ;
- La dépersonnalisation voire la déshumanisation de la pratique médicale avec une mise en second plan de la relation humaine au profit de l'interaction homme – machine ;
- le risque d'asservissement des médecins à la technologie avec une obéissance aveugle à ces outils qui deviennent de plus en plus performants entraînant ainsi la perte de l'indépendance du médecin ;
- le renforcement des inégalités d'accès aux soins du fait de niveau d'accès aux technologies par les individus et les hôpitaux ;
- la discrimination de certains groupes d'individus par l'utilisation de la thérapie génomique et de l'échantillonnage des données dans le cadre de l'entraînement des algorithmes ;
- un détournement de l'usage des technologies dans le cadre de la pratique médicale comme moyen de surveillance et de répression.

Encadré 12: analyse de l'étude sur l'hydroxychloroquine (MEHRA ET AL., 2020)

L'étude publiée dans *The Lancet* sur l'utilisation de l'hydroxychloroquine (MEHRA ET AL., 2020) qui défraie l'actualité constitue pour nous un exemple pratique de ce que nous développons dans le cadre de notre thèse. Dans cette étude, les données provenaient de 671 hôpitaux répartis dans 45 pays et sur 6 pays. Elles ont été fournies par une entreprise américaine « *Surgisphère* ». Les auteurs concluaient qu'ils n'étaient pas à mesure de préciser si « *l'utilisation de l'hydroxychloroquine seul ou en association avec l'azithromycine n'apportait aucun bénéfice aux patients. Par contre, ils indiquaient que cette utilisation augmentait le taux de mortalité intra hospitalière et le risque de troubles de rythme graves* » (MEHRA ET AL., 2020).

Cette étude comporte plusieurs problèmes d'ordre éthico-méthodologique :

- la qualité et les sources des données³⁶⁴ ne sont pas fournies³⁶⁵ ;
- le processus ayant abouti au choix des hôpitaux et des pays n'a été ni défini, ni expliqué ;
- aucune information concernant l'approbation des comités d'éthiques et le consentement des patients n'est fournie ; les auteurs n'ont pas fourni de documents concernant les autorisations des commissions informatiques et libertés des différents pays ;
- le transfert des données européennes vers les États Unis ne correspond pas aux normes : il nécessite une approbation de toutes les commissions informatiques et libertés au niveau de chaque pays de l'UE et l'inscription sur la liste des certifications du *privacy shield*³⁶⁶ ; ce qui n'est pas le cas pour l'entreprise *Surgisphere* ;
- sont comparées des données provenant de pays avec des populations ayant des profils, des environnements et des contextes de prise en charge différents³⁶⁷.

Des auteurs du monde entier ont écrit une lettre ouverte aux auteurs et à l'éditeur de la revue (WATSON, 2020) pour dénoncer la qualité insuffisante de cette étude. Mais les autorités françaises ont malgré tout pris des décisions sur la base de cette étude et ont interdit la prescription dérogatoire de la chloroquine³⁶⁸. De même l'arrêt des essais cliniques à base de chloroquine a été demandée par l'OMS le 24 mai 2020 avant de demander la reprise des études le 03 juin 2020³⁶⁹.

³⁶⁴ Les autorités australiennes ont signalé que les données leur concernant n'étant pas conformes aux données nationales et qu'elles n'avaient donné aucune autorisation d'accès ou de transfert de leurs données vers la société « *Surgisphere* ». Source : <https://www.theguardian.com/science/2020/may/28/questions-raised-over-hydroxychloroquine-study-which-caused-who-to-halt-trials-for-covid-19> consulté le 04 juin 2020.

³⁶⁵ Pour ne pas diffuser leurs sources, les auteurs avancent la protection des travaux des chercheurs. Le grand public s'en étonne. Source : https://www.sciencesetavenir.fr/sante/interview-le-coauteur-de-l-etude-du-lancet-sur-la-chloroquine-repond-aux-questions_144780 consulté le 04 juin 2020.

³⁶⁶ Source : <https://www.privacyshield.gov/list> consulté le 04 juin 2020.

³⁶⁷ Comment par exemple comparer les données de mortalité entre les pays africains - dont beaucoup ne disposent pas de moyens de réanimation - et les États Unis où la mortalité est liée aux inégalités d'accès à des soins de réanimation existants.

³⁶⁸ Source : <https://france3-regions.francetvinfo.fr/provence-alpes-cote-d-azur/bouches-du-rhone/marseille/coronavirus-haut-conseil-sante-publique-siffle-fin-partie-hydroxychloroquine-interdite-france-1833556.html> consulté le 04 juin 2020.

³⁶⁹ Source : https://www.sciencesetavenir.fr/sante/covid-19-et-chloroquine-l-etude-du-lancet-est-mise-en-doute-par-le-journal_144875 consulté le 04 juin 2020.

On voit ici comment le défaut dans l'application de règles méthodologiques et juridiques et le manque d'éthique produit des risques. Cet exemple montre également l'impact que peuvent avoir un mauvais traitement des *data* a) sur le processus de création des connaissances et sur la prise de décision et b) sur la confiance que les individus accordent à la recherche scientifique en général et à la pratique médicale en particulier.

La conséquence principale sur la pratique médicale sera la perte totale de confiance aux médecins ; confiance qui est l'essence même de la pratique médicale. La rupture de la confiance entre le médecin et son patient aura des répercussions importantes sur la société avec des enjeux économiques et sociétaux très importants à savoir :

- un faible recours à la consultation médicale avec pour conséquence le diagnostic tardif des maladies et un risque important de propagation des épidémies ;
- une rupture de confiance entre les patients, le médecin et les autorités sanitaires ;
- une dérégulation du processus d'informatisation mis en place par un refus d'utilisation des TIC par les médecins en vue de protéger leur profession et le secret médical ;
- une mauvaise protection sanitaire des patients du fait de refus de transmission des feuilles de soins au niveau de la Sécurité sociale ;
- une perte économique importante pour les entreprises qui travaillent dans le domaine des TIC de la santé.

L'utilisation du *big data* et de l'IA comporte certes des risques pour l'éthique médicale et des enjeux sociétaux, économiques et politiques. Mais ces risques et enjeux ne doivent pas occulter les avantages que peuvent apporter leurs utilisations dans le domaine de la santé. C'est surtout le mode de développement des outils et de l'utilisation des données qui doivent être questionnés afin de trouver un ajustement acceptable entre les bénéfices et les risques pour le patient comme cela a toujours été le cas en pratique médicale.

7 Vers une intelligence artificielle de confiance

« Tout ce qui est fait pour moi, sans moi, est fait contre moi ».

Nelson MANDELA.

Le processus de développement des outils de l'IA et du *big data* dans le cadre de la pratique médicale est une étape importante pour son acceptabilité et son appropriation par les acteurs ; surtout avec la participation de nouveaux acteurs dans le processus de soin. Ce processus se doit ainsi de respecter des contraintes et des normes aussi bien pour les utilisateurs que sont les patients et les médecins et surtout pour la profession médicale. Ainsi, le processus de développement doit respecter :

- les procédures métiers dans le domaine de la santé ;
- le respect des connaissances médicales dans le domaine ;
- le respect des droits des patients ;
- le respect des textes juridiques d'accès des données de santé ;
- le respect de l'éthique médicale

Le respect des normes et des règles métiers permettent de créer un climat de confiance, très important dans le cadre de la pratique médicale.

Pour la satisfaction de leurs attentes, les médecins pensaient qu'il fallait revoir le processus de mise en place des outils avec des actions telles que :

- l'implication de tous les acteurs dans le processus de conception et de développement des solutions afin de mieux cerner les besoins réels des utilisateurs, les règles à respecter ; l'intérêt et la fiabilité des outils dans la pratique médicale ;
- la définition de la gestion des données pour s'assurer du respect du secret professionnel et des normes de pratique médicale ; pour assurer la confidentialité ; la sécurité des données et mettre en place une gouvernance transparente des données de santé ;
- la validation des outils par un processus d'essai clinique afin de prouver l'efficacité et l'efficacité des outils avant la mise sur le marché ;
- la valorisation de l'utilisation des TIC dans la pratique médicale par le gain de temps avec des outils simples à mettre en place, adapter aux connaissances des

utilisateurs et la transformation de certains usages en acte médicale ;

- la définition de la place de l'Etat dans le processus de mise en place des technologies ; l'Etat doit être un facilitateur, un créateur d'environnement propice et non un acteur à temps plein dans le processus.

Pour notre part, nous pensons que toutes les conditions peuvent être mise en place pour un développement « *vertueux* » du *big data* et de l'IA sans pour autant perdre en compétitivité, en qualité et en rapidité de mise sur le marché des outils. Pour ce faire, le projet doit être pensé dans le cadre de la recherche-action.

« La recherche-action vise à résoudre des problèmes pertinents dans un contexte donné au travers d'une recherche démocratique où les chercheurs professionnels collaborent avec les participants dans un effort pour chercher et mettre en action des solutions à des problèmes d'importance majeure pour les acteurs locaux » (GREENWOOD, LEVIN, 1998).

La recherche-action s'inscrit dans un ensemble large de démarches de recherche participantes parfois qualifiées de méthodes interactives à visée transformatrice. La méthodologie de la recherche-action s'apparente à l'expérimentation de laboratoire (hypothèse – vérification) transposée en contexte social. La phase de diagnostic dans une recherche – action est essentielle. *« Une action réussie repose sur l'analyse correcte par le chercheur de la situation, sur l'identification de toutes les solutions possibles et sur le choix de la plus appropriée » (JOUISON-LAFFITTE, 2009).*

Pour la mise en œuvre d'une recherche-action au cours du processus de développement du *big data* et de l'IA, nous proposons d'utiliser des outils que sont :

- le *design thinking* ;
- la relation de confiance entre les différents acteurs ;
- la « *boîte à outils* » de Nicole et Louis Lery déjà décrite plus haut (section 3.4.5) dans le cadre de la prise de décision partagée.

7.1 *Design thinking*

Le *design thinking* est un processus utilisé par les designers pour innover à partir d'usages existants ou à inventer. Il s'agit d'une méthode d'intelligence collective qui place l'humain, ses usages et ses besoins au centre de la réflexion. Concrètement, l'idée est de mettre en

place une équipe pluridisciplinaire et volontaire, composée de communicants, d'ingénieurs, de dirigeants, de consommateurs ou usagers, afin de résoudre une problématique donnée par une marque ou une entreprise (BROWN, 2008).

Le *design thinking* fait partie des nouvelles approches d'innovation axées sur l'utilisateur et nommées « *design centré utilisateur* ». Adopter la perspective des utilisateurs dès la conception est au cœur des valeurs du *design thinking*.

« Plutôt que de partir du développement des nouvelles technologies ou d'analyses quantitatives liées au marché, il est nécessaire d'aller sur le terrain pour comprendre les gens. Les appréhender dans leur contexte réel est une voie plus prometteuse et moins risquée pour identifier de nouvelles opportunités et créer des innovations pertinentes et génératrices de croissance, qu'elles soient incrémentales ou de rupture » (MATHIEU, HILLEN, 2016).

Le *design thinking* permet de définir une expérience idéale, à la croisée de ce qui est désirable pour ceux pour qui on veut innover, ce qui est techniquement réalisable et ce qui est viable économiquement par le porteur de projet. « *C'est une philosophie avec un manifeste de valeurs liées à l'empathie et l'action, ainsi qu'une méthode qui allie processus et outils, et dont l'objectif est de guider la pratique pour innover d'une façon pragmatique, collective et efficace* » (MATHIEU, HILLEN, 2016).

Le *design thinking* permet ainsi de résoudre des problèmes en partant des besoins des utilisateurs et en développant des solutions pour y répondre. La technologie est ainsi vue comme un outil pour répondre à des besoins et non comme une fin en soi. Il est centré sur l'utilisateur, les expériences et l'exploration. Il permet d'apporter à une problématique donnée, une solution innovante qui répond à ces trois principes fondamentaux (BROWN, 2008) :

- la désirabilité qui correspond à l'expression des besoins et de l'utilité centrées sur l'utilisateur ;
- la faisabilité technique et organisationnelle ;
- la viabilité en termes de performance, d'efficience, d'efficacité et surtout de viabilité économique.

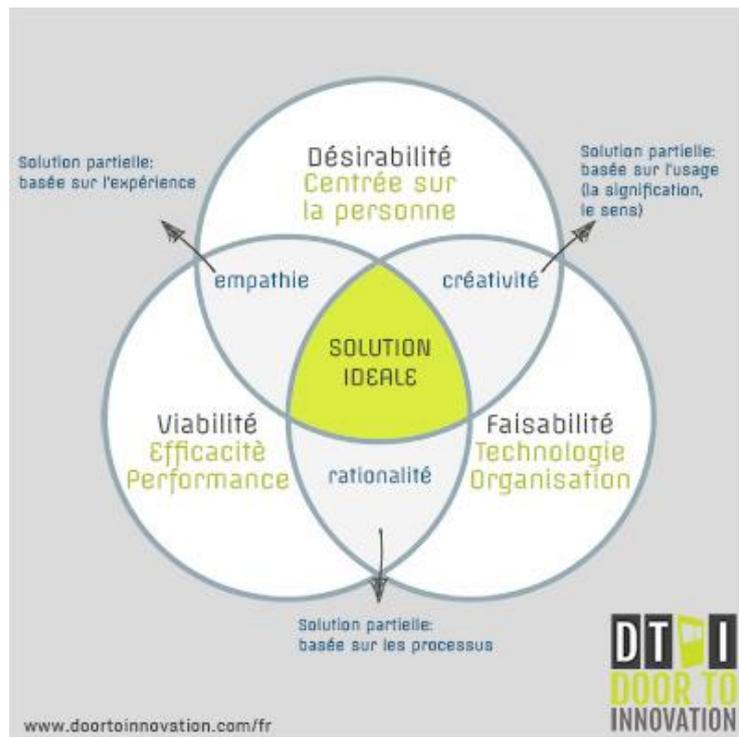


Figure 96 : principes du design thinking à la recherche de solution idéale³⁷⁰

Le travail en mode *design thinking* s'organise autour de trois logiques (MATHIEU, HILLEN, 2016) :

- une logique de co-création où les différents acteurs travaillent ensemble et non pas chacun dans son domaine de spécialisation ;
- une réflexion intellectuelle alternant des phases d'intuition et d'analyse, dans une logique d'ouverture/fermeture ;
- une importance majeure accordée à l'étude de terrain (observation ethnographique) qui offre une compréhension pleine et entière des expériences, contrairement aux classiques études quantitatives et qualitatives.

Le design thinking peut être décrit en trois principales étapes que sont (MATHIEU, HILLEN, 2016 ; BROWN, 2008) :

7.1.1 *Inspiration*

Durant cette étape, il s'agit d'identifier le problème à résoudre. Cette identification du problème se fait à partir du terrain avec les utilisateurs concernés. Elle permet de faire une

³⁷⁰ Jean-Pierre LEAC. Qu'est-ce-que le *design thinking* ?. Source : <https://www.lescahiersdelinnovation.com/2016/02/qu-est-ce-que-le-design-thinking/> consulté le 20 octobre 2019.

description plus détaillée du problème avec les expériences et les vécus de chaque utilisateur.

Cette phase de terrain exige (MATHIEU, HILLEN, 2016) :

- une combinaison entre empathie, méthode et intuition ;
- un état d'esprit, discipline et pratique ;
- de construire une expertise spécifique pour type chaque d'utilisateur et de contexte.

L'objectif est de gagner en pertinence par rapport à ceux pour qui on veut innover, afin de « frapper juste » : juste en pertinence dans l'identification des besoins réels, et juste en quantité dans la compréhension de leur hiérarchie.

Elle consiste à reformuler la problématique en quittant le traditionnel questionnement « *comment faire ?* » pour se demander « *pourquoi faire ?* » (BROWN, 2008).

7.1.2 Idéation : imagination et conception

L'idéation est le processus de générer, développer et tester des idées qui peuvent aboutir en des solutions. Durant cette phase, il s'agit de générer autant d'idées que possible pour répondre à ces besoins sans les juger. Pour explorer et créer de nouvelles possibilités, on utilise la pensée divergente, « *qui consiste à multiplier les options pour créer des choix, autrement dit, générer un maximum d'idées afin d'augmenter la possibilité de trouver la bonne solution* » (MATHIEU, HILLEN, 2016 ; BROWN, 2008).

Cette étape inclut également le prototypage des solutions et les tests avec les utilisateurs en contexte réel. Le prototypage est une étape clé du process du *design thinking*. Il possède un pouvoir fédérateur très puissant : c'est la démonstration par la preuve. La réalisation des prototypes permet d'accélérer l'avancée du projet et d'explorer plusieurs idées en parallèles (MATHIEU, HILLEN, 2016).

La matérialisation de l'idée permet une évaluation facile des solutions, de les perfectionner et de faire le choix de la méthode la plus adaptée sur le plan humain, technique et économique. Elle permet de gagner du temps et éviter de dépenser sur des prototypes industriels qui ne répondent pas aux besoins des utilisateurs.

« *La mise en scène des prototypes (produits ou services) nécessite une attention très particulière pour recueillir un retour pertinent et constructif de la part des utilisateurs. C'est un exercice très délicat qui peut être totalement biaisé par la façon de présenter. L'état d'esprit est à nouveau l'empathie, avec une volonté d'écouter pour améliorer les solutions proposées* » (MATHIEU, HILLEN, 2016).

7.1.3 Implémentation

La phase d'implémentation commence au moment où émergent une ou plusieurs solutions matures, résultantes de l'interaction entre les phases d'inspiration et d'idéation. C'est la transition vers l'exploitation, alors que les deux premières phases sont exploratoires. L'implémentation regroupe toutes les activités nécessaires pour transformer le projet de l'état de conception à une réalité (MATHIEU, HILLEN, 2016).

La solution retenue – celle qui répond à la désirabilité, la faisabilité et la viabilité - est testée sur le terrain pour vérifier si elle répond à l'objectif de départ des utilisateurs et pour identifier les sources d'amélioration. Le retour et les commentaires des utilisateurs sont très importants à cette étape et peuvent déboucher à de nouveaux marchés. Ce retour permettra aux concepteurs également de définir un modèle économique.

La force du *design thinking* réside dans son principe « *d'échouer vite pour réussir encore plus vite* » (BROWN, 2008). Ce principe de valorisation de l'erreur est très puissant en termes de conception rapide de solutions pertinentes. Il permet également de réduire le taux d'échec de lancement des solutions d'IA.

Le *design thinking* est une démarche qui permet de concilier des contraintes variées. Même si c'est dur, même si c'est long, la création d'expériences client en vaut la peine. De très bonnes expériences construisent la fidélité et certaines études affirment que si l'on améliore la fidélité des consommateurs de 5 %, on peut s'attendre à une augmentation des bénéfices de 25 % à 50 %³⁷¹. En effet, une étude réalisée par l'International Data Corporation (IDC) en 2019 notait un taux d'échec de 50 % des projets d'IA dans les entreprises. Les principales

³⁷¹ Op. Cit. Jean Pierre LEAC p 334.

causes étaient le manque de compétences, les attentes irréalistes et l'absence des données de qualité³⁷².

Ces trois phases qui semblent distinctes se recoupent en réalité grâce à une démarche itérative qui met l'utilisateur et son expérience au centre du processus : « *nous sommes toujours dans un état d'esprit empathique vis-à-vis de ceux pour qui on veut innover, qu'ils soient clients ou futurs clients, et ce souvent dans un système complexe de parties prenantes aux enjeux différents* » (MATHIEU, HILLEN, 2016).

Le *design thinking* est une méthode agile centrée sur l'humain et l'intelligence collective. Il peut s'appliquer à tout processus d'innovation. Il permet à des équipes pluridisciplinaires de travailler ensemble sur des projets innovants en mettant en valeur les attentes et l'expérience des utilisateurs, la faisabilité technologique et fonctionnelle et la viabilité économique. Le *design thinking* permet de mettre en avant l'intérêt du public plutôt que les considérations de profit et de rentabilité. Le processus de gestion du projet est vu comme un système dans son ensemble comme faisant partie d'un écosystème sociétal, organisationnel et fonctionnel.

7.2 Application du *design thinking* dans le domaine de la santé

7.2.1 *Ecosystème de la santé*

L'écosystème de la santé est constitué de plusieurs acteurs et cela s'est complexifié avec l'informatisation de la santé. L'utilisation du *big data* et de l'IA rend encore cet écosystème encore plus complexe. En plus des acteurs traditionnels, il faut compter avec les concepteurs de solutions d'informatique.

Nous pouvons définir les acteurs de la santé en quatre groupes dans le cadre du *big data* et de l'IA. Ce sont :

Equipe de soins

L'équipe de soins est constituée des différents acteurs entrant dans le processus de soin du patient sur le plan médical. Il est constitué du médecin dans le cadre de la pratique libérale ou en consultation et d'une équipe dans le cadre d'une prise en charge pluridisciplinaire.

³⁷² Source : <https://venturebeat.com/2019/07/08/idc-for-1-in-4-companies-half-of-all-ai-projects-fail/> consulté le 28 octobre 2019.

L'équipe de soins est membre d'un ordre professionnel qui définit les règles de pratique du métier et veille au respect de ces règles par les membres de l'équipe de soins. Les membres de l'équipe sont également membres de sociétés savantes qui mettent à leur disposition des recommandations de bonnes pratiques cliniques fondées sur les preuves de leur efficacité. Les différentes organisations collaborent dans le cadre de la pratique médicale d'une manière générale.

Patients

Le patient est en contact avec l'équipe de soin dans une relation de médecin-patient. Mais il fait partie d'un environnement avec une famille et des amis, avec ses besoins propres et ses désirs ; bien que celui-ci puisse être influencé par les avis de la famille et des amis. Il peut être membre d'une association de patients. Le développement de la participation des patients dans les soins en tant qu'expert est très important pour le partage de son expérience et de son vécu (HERVÉ ET AL., 2017). Ainsi le patient peut être d'une grande contribution dans le processus de mise en place des outils car étant l'un des utilisateurs finaux. Par ailleurs, tous les acteurs peuvent être à la fois des patients.

Etat

L'Etat offre des opportunités aux acteurs du système de soins de pouvoir exercer dans leur domaine respectif. Il régule également la relation qui existe entre les différents acteurs à travers des lois et veille aux respects des contrats entre eux.

Dans le cadre du processus de développement du *big data* et de l'IA dans le domaine de la santé, certaines structures de l'Etat sont impliquées. Il s'agit :

- la HAS³⁷³ : la HAS a pour mission d'évaluer d'un point de vue médical et économique les produits de santé. Ses avis rendus aux pouvoirs publics permettent d'éclairer leur décision en vue d'une prise en charge par la solidarité nationale et de la fixation de leur prix. Elle évalue le service rendu des médicaments, des dispositifs médicaux, des actes professionnels et des technologies de santé et apprécie leur progrès au regard des stratégies thérapeutiques existantes. Lorsqu'ils se revendiquent comme innovants et qu'ils sont susceptibles d'avoir un impact significatif sur les dépenses de santé, les produits de santé font l'objet d'une évaluation médico-économique par la HAS.

Elle élabore également des recommandations pour les professionnels de la santé,

³⁷³ Source : https://www.has-sante.fr/jcms/c_1002212/fr/missions-de-la-has consulté le 29 octobre 2019.

du social et du médico-social, afin d'optimiser et d'harmoniser leurs pratiques et leurs organisations.

La HAS a pour rôle de mesurer et d'améliorer la qualité des soins et la sécurité des patients dans les établissements de santé et en médecine de ville, et la qualité des accompagnements dans les établissements sociaux et services médico-sociaux ;

- l'Agence nationale de sécurité du médicament et des produits de santé (ANSM)³⁷⁴ a été créée par la loi du 29 décembre 2011 relative au renforcement de la sécurité sanitaire des médicaments et des produits de santé. Elle a deux missions principales que sont : offrir un accès équitable à l'innovation pour tous les patients et garantir la sécurité des produits de santé tout au long de leur cycle de vie, depuis les essais initiaux jusqu'à la surveillance après autorisation de mise sur le marché.

Sa compétence s'applique aux médicaments, aux produits biologiques, aux dispositifs médicaux et aux produits cosmétiques et de tatouage.

Pour tous les dispositifs médicaux, exceptés ceux de la classe 1³⁷⁵, l'apposition du marquage est subordonnée à l'obtention d'un certificat CE, délivré par un organisme appelé organisme notifié qui est habilité par les autorités compétentes. En France, l'autorité compétente est l'ANSM. A ce jour, le seul organisme notifié français pour les dispositifs médicaux est le GMED.

- le Comité économique des produits de santé (CEPS)³⁷⁶, organisme interministériel placé sous l'autorité conjointe des ministres chargés de la santé, de la Sécurité sociale et de l'économie, est principalement chargé par la loi de fixer les prix des médicaments et les tarifs des dispositifs médicaux à usage individuel pris en charge par l'assurance maladie obligatoire. Le CEPS est composé de 2 sections, la section du médicament et la section des dispositifs médicaux.

Dans le cadre des orientations qu'il reçoit des ministres, le CEPS est chargé de proposer les tarifs de remboursement des produits et prestations de l'article L. 165-1 du code de Sécurité sociale et, le cas échéant leur prix. Il peut pour la fixation de ces tarifs et de ces prix conclure avec les fabricants et/ou les distributeurs concernés

³⁷⁴ Source : [https://www.ansm.sante.fr/Activites/Mise-sur-le-marche-des-dispositifs-medicaux-et-dispositifs-medicaux-de-diagnostic-in-vitro-DM-DMIA-DMDIV/DM-classe-I-DM-sur-mesure-assemblage-Declaration/\(offset\)/4](https://www.ansm.sante.fr/Activites/Mise-sur-le-marche-des-dispositifs-medicaux-et-dispositifs-medicaux-de-diagnostic-in-vitro-DM-DMIA-DMDIV/DM-classe-I-DM-sur-mesure-assemblage-Declaration/(offset)/4) consulté le 29 octobre 2019.

³⁷⁵ Le dispositif médical de classe 1 sont des dispositifs à risque faible comme les seringues ou les pèse-personnes.

³⁷⁶ Source : <https://solidarites-sante.gouv.fr/ministere/acteurs/instances-rattachees/article/ceps-comite-economique-des-produits-de-sante> site consulté le 29 octobre 2019.

des conventions pouvant notamment porter sur les volumes de ventes. Le CEPS assure un suivi périodique de ces dépenses.

- la CNIL³⁷⁷ a pour mission essentielle de protéger la vie privée et les libertés individuelles ou publiques. Elle est chargée de veiller au respect de la loi "informatique et libertés" qui lui confie quatre missions principales : informer et protéger les droits des usagers, particuliers comme professionnels ; accompagner et conseiller la mise en conformité ; anticiper et innover à la recherche de signaux faibles et aux sujets émergents ; contrôler les organismes et les sanctionner ou mettre en demeure en cas de manquements constatés.

Tous ces organismes de l'Etat travaillent en étroite collaboration entre eux mais également avec les différents acteurs du système de santé.

Promoteurs

De nouveaux acteurs apparaissent dans l'écosystème de santé et participent de manière directe ou indirecte dans le soin au travers des outils qu'ils développent. Ces acteurs peuvent être membre d'une association professionnelle. Il s'agit :

- le *data scientist* : son rôle est de gérer le traitement des données et de produire des algorithmes sur les données pour anticiper leur comportement, recommander des actions et catégoriser les données ;
- l'informaticien : son rôle est de mettre en place l'infrastructure nécessaire pour le traitement des données et assurer leur sécurité. Il peut jouer le rôle de *data scientist* à la fois ;
- les ingénieurs : dans le cadre de développement de dispositifs médicaux, ils participent à la mise en œuvre du projet
- le bailleur de fonds : il apporte des ressources financières pour le bon déroulement du projet. Il peut être un hôpital, une entreprise, une personne ou l'Etat.

Au vu de l'analyse des acteurs de l'écosystème, il faut non seulement une collaboration forte entre les différents acteurs et surtout une pluridisciplinarité pour le développement des

³⁷⁷ Source : <https://www.cnil.fr/fr/les-missions-de-la-cnil> consulté le 29 octobre 2019.

projets d'IA et de *big data*. De ce fait, la création d'un climat de confiance est une condition *sine qua non* pour la réussite des projets dans le domaine de la santé.

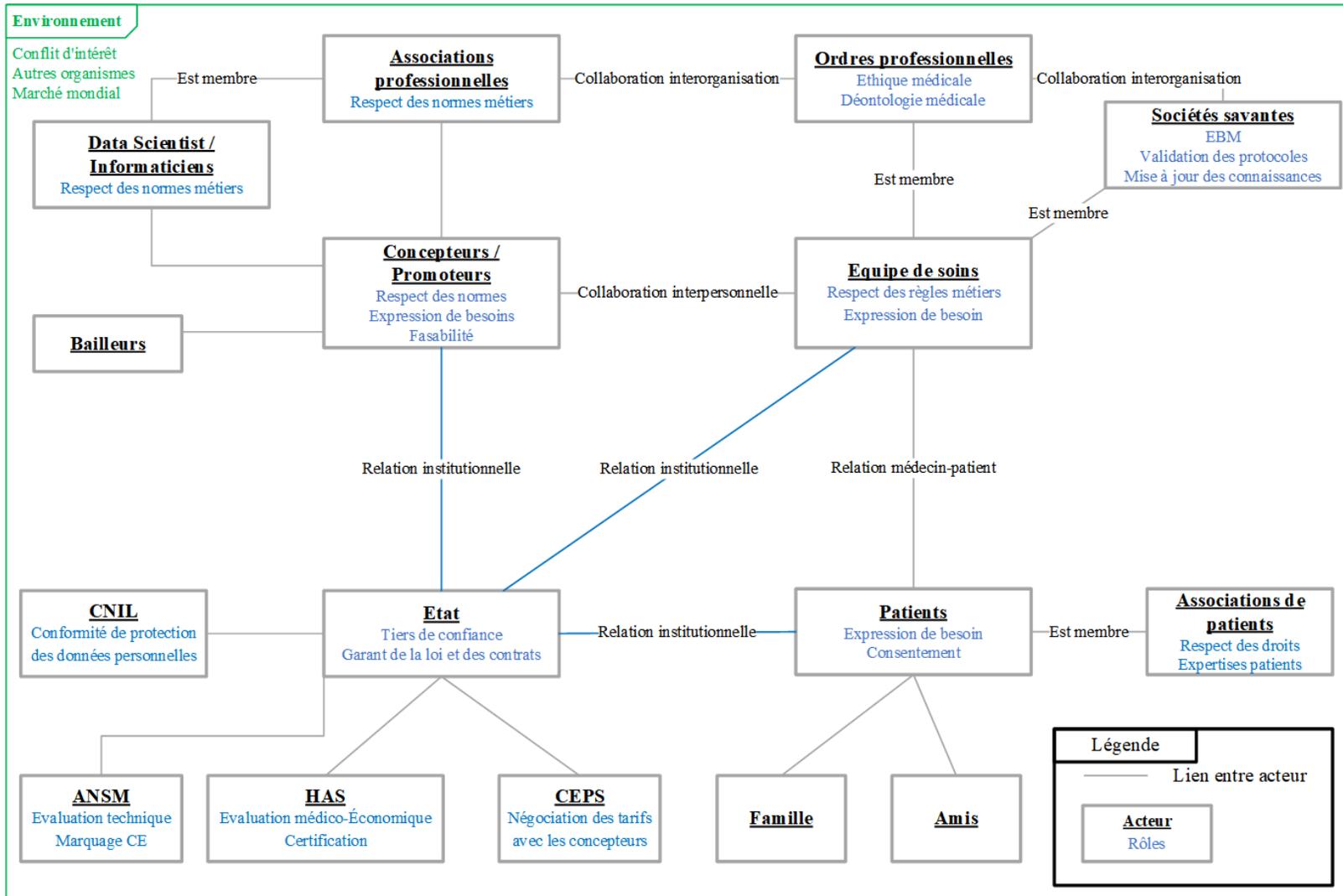


Figure 97 : acteurs de l'écosystème des soins de santé

Confiance

On ne peut imaginer l'acte de soin sans la confiance ; elle est l'essence même de l'acte médical. Dans le cadre de la relation de soins, il s'agit d'être digne de confiance plutôt que d'un accord de confiance. Tout au long de la relation de soins, il existe une confiance interpersonnelle, entre des personnes et des organismes et entre des organismes, sans lesquelles il serait difficile de pratiquer la médecine. La confiance est l'un des fondements du serment d'Hippocrate :

« [...] Je ne tromperai jamais leur confiance et n'exploiterai pas le pouvoir hérité des circonstances pour forcer les consciences » et est renforcée par le secret médical en ces termes *« [...] Admis(e) dans l'intimité des personnes, je tairai les secrets qui me seront confiés. Reçu(e) à l'intérieur des maisons, je respecterai les secrets des foyers et ma conduite ne servira pas à corrompre les mœurs »* Serment d'Hippocrate.

En l'absence de confiance dans le système de santé, les conséquences sont catastrophiques pour la société : refus de participation à la vaccination, non-respect des règles de santé, refus de participation aux études et dans notre cas le refus d'utilisation du *big data* et de l'IA par manque de transparence et de la fiabilité des outils, de risque de perte de confidentialité et du secret médical et la question de la responsabilité des acteurs.

« La confiance est une attitude, un état psychologique survenant dans un contexte d'incertitude, de risque et de vulnérabilité » (BIZOUARN, 2008). Les définitions du terme sont multiples, selon les points de vue adoptés par les auteurs et le contexte dans lequel est insérée la notion de confiance (sociologie, psychologie, économie) (SIMON, 2007 ; ORLÉAN, 2000).

Zucker (ZUCKER, 1986) définit *« la confiance comme une série d'attentes sociales partagées par les personnes impliquées dans un échange économique. L'hypothèse sous-jacente à cette définition est la présence de valeurs communes pour les parties engagées dans l'échange »*. Elle est le *« lubrifiant »* à la base du fonctionnement du système social et de son efficacité (SIMON, 2007).

« Faire confiance, c'est d'abord se fier à quelqu'un, s'en remettre à quelqu'un, en éprouvant sa fiabilité (en anglais, to rely, et son substantif reliance). Une

personne est fiable dès lors qu'elle est compétente, efficace, soigneuse pour réaliser une tâche ou un bien demandé. La relation de confiance est une relation à trois : A se fie à B pour réaliser X » (HARDIN, 2002).

Zucker (ZUCKER, 1986) décrit trois typologies de la confiance à savoir :

- La confiance interpersonnelle basée sur une relation entre les individus. Elle est perçue comme une importante ressource sociale qui facilite la coopération, elle permet une meilleure coordination des interactions (MAYER ET AL., 1995). Elle comporte une composante intentionnelle (qu'un individu fait confiance à un autre parce qu'il croit que ce dernier respectera ses engagements sans faire preuve d'opportunisme) qui est la bienfaisance et le respect de l'autonomie dans le cadre de la relation médecin – patient et une composante de compétence (qu'un individu fait confiance à un autre parce que ce dernier détient les capacités et les compétences requises) correspondant à la non-malfaisance ;

La confiance interpersonnelle se développe lors de la poursuite d'objectifs communs qui nécessitent des communications fréquentes et une certaine proximité.

- La confiance institutionnelle ou systémique accordée à une autorité supérieure dans une société (MANGEMATIN ET AL., 1999) ; son rôle est la protection des individus contre les aléas moraux, il s'agit d'une confiance partagée entre les individus grâce à des construits normatifs et sociaux. Elle a recours à des lois, des cadres réglementaires et des contrats avec l'acceptation des règles qui en découlent. *« La construction de la confiance institutionnelle s'apparente davantage à un contrat social dans lequel chacun accepte d'abandonner une partie de sa liberté pour fixer des règles de jeu et accepte de déléguer l'autorité et les moyens de coercition à une tierce partie comme l'État, un ordre professionnel » (MANGEMATIN ET AL., 1999).*
- La confiance inter-organisationnelle est la confiance accordée par les membres d'une organisation à l'organisation associée (ZAHEER ET AL., 1998) ; elle implique au moins deux organisations en tant que personnes morales. Elle permet de tisser des liens sur le long terme et bâtir un réseau de relations efficace. La confiance inter-organisationnelle dépasse la relation interpersonnelle pour embrasser l'organisation partenaire avec laquelle s'effectue un échange. Cette confiance

conduit à des comportements coopératifs et à l'apprentissage collectif.

La pluridisciplinarité du développement du *big data* et de l'IA impose une collaboration franche entre les acteurs avec un niveau de confiance élevé et une personne garant des contrats. L'Etat dans ce cas-ci en partenariat avec des organismes de régulation peut jouer le rôle de tiers de confiance. Durant le processus de développement des outils, la relation de confiance entre les médecins, les promoteurs et les patients va se développer dans l'exercice de co-conception et se renforcer dans le temps. A travers eux, la relation de confiance entre leurs différentes organisations se trouvera renforcer.

Tout n'est pas à refaire, puisque la relation de confiance entre le médecin et le patient existe déjà ; aussi bien que la relation de confiance entre les médecins, l'ordre des médecins, les associations de patients et l'Etat. Le lien à enrichir et à renforcer est le lien entre les concepteurs et les autres acteurs. Pour cela, il va falloir que les concepteurs acceptent se mettre en position de vulnérabilité et travaillent en toute transparence avec les autres acteurs. La construction de cette relation de confiance entre tous les acteurs de la santé est très importante en ce sens qu'elle peut devenir un moyen de gouvernance (DYER, 1997) et palier là où la loi ne peut être appliquée (par exemple dans le cas de l'explicabilité des algorithmes si un processus de co-construction des outils est mise en place de manière transparente).

7.2.2 *Différentes étapes de la conception*

Dans le cadre de l'application du modèle de *design thinking* dans le développement du *big data* et de l'IA dans le domaine de la santé, nous utilisons les trois étapes décrites plus haut et associons une étape d'évaluation du fait de la spécificité du domaine médicale. Cette étape d'évaluation est très importante pour la relation de confiance et une meilleure utilisation des outils par les utilisateurs finaux.

Inspiration

L'identification du problème ou de l'expression de besoin peut émaner de tous les acteurs du système de santé. Mais une fois le problème identifié, il doit être confronté au terrain pour s'enquérir de l'expérience des futurs utilisateurs. Durant cette phase, le problème sera discuté et bien cerné par tous les acteurs avec une analyse des solutions existantes, des avantages et des inconvénients et le retour des utilisateurs sur ces solutions.

La confrontation du terrain permet dans le cadre de soins de définir le problème en tenant compte des impacts et des contraintes sur la pratique médicale et les besoins d'organisations possibles. Elle permet de prendre en compte les besoins de tous les acteurs sans exception pour proposer des solutions et d'adapter les solutions à chaque contexte de pratique. Par exemple, un outil basé sur le web et dépendant d'Internet pourra bien fonctionner en France, mais ne sera pas adapté pour les pays Africains.

Idéation

Une fois que le problème a été correctement décrit, tous les acteurs du système de santé proposent des solutions qui sont toutes retenues dans un premier temps sans exception. Toutes les idées de solutions sont ensuite discutées une à une pour un meilleur choix sans prépondérance de voix d'un acteur sur un autre acteur. Lors du choix des solutions, les acteurs décident ensemble des données utiles et nécessaires pour l'atteinte des objectifs de départ.

Cette phase est très importante car elle permet de respecter les principes fondamentaux du RGPD (collecte loyale, licéité et obligation de transparence et d'informations), les règles de métiers des différents corps (médecins, ingénieurs et informaticiens) et les droits et les devoirs des utilisateurs. Elle donne l'occasion aux ingénieurs et aux informaticiens d'expliquer le principe de fonctionnement et le mode d'obtention des résultats surtout pour les médecins et les patients dans le cadre du soin. Ce processus de co-construction permet de respecter les textes juridiques pendant toute la phase de conception et surtout de permettre aux concepteurs d'avoir accès à un terrain d'expérimentation dans la vie réelle avec une garantie de sécurité pour tous les acteurs.

Les solutions retenues sont prototypées dans un premier temps et présentées aux acteurs pour un test. Au cours de ce test, les acteurs font des commentaires et des critiques pour améliorer les différentes solutions.

Implémentation

Une fois les commentaires et critiques des utilisateurs pris en compte pour les différentes solutions, les acteurs doivent procéder au choix de la solution finale. Ce choix se fait en fonction de la désirabilité (résolution du problème identifié), la faisabilité technique et organisationnelle (le niveau de difficulté technique et l'impact sur l'organisation des

différents acteurs) et la viabilité économique (le coût de production, d'exploitation et d'entretien et de mise à jour, important pour l'accessibilité de la technologie). L'idéal est que ces trois fonctions soient respectées ; mais certes difficile dans la pratique.

L'intérêt du *design thinking* est que les acteurs étant autour de la même table peuvent faire des compromis en connaissance des causes et des conséquences. En plus, il permet de définir le mode d'utilisation et la place des outils développés dans le processus de soins comme outils d'aide à la décision. Les forces et les faiblesses des outils sont connues d'avance par les différents acteurs et les compromis *que chacun d'eux sont* prêts à accepter sont connus.

Mais dans le domaine de la santé, il est important de faire l'évaluation des outils entrant dans le cadre de soins et avant son utilisation à large échelle pour s'assurer de son innocuité, sa fiabilité, sa qualité et l'apport d'un bénéfice pour le patient, la société et le médecin. Nous avons donc intégré le processus d'évaluation dans notre modèle.

Evaluation

L'évaluation des outils va comporter plusieurs étapes :

- l'essai clinique : le but de l'essai clinique est de vérifier les apports en termes de bénéfice et d'innocuité de l'utilisation de l'outil dans la pratique médicale. On élargit le cadre d'évaluation à d'autres utilisateurs finaux. Avant la réalisation de l'essai clinique, l'outil est soumis au comité de protection des personnes qui rend son avis sur la validité de la recherche et la CNIL pour la validation des procédures de protection des données personnelles. En plus, il faut une autorisation de l'ANSM pour mener l'essai clinique. Tous les acteurs participent à ce processus. Cet essai clinique servira non seulement pour l'évaluation médicale des outils par les médecins et les patients, l'évaluation technique et le *marquage CE* par l'ANSM et l'évaluation médico-technique et la certification par la HAS.
- l'évaluation technique : l'évaluation technique a pour objectif de vérifier la qualité, l'efficacité et la sécurité d'emploi des outils. Cette évaluation technique comporte une évaluation clinique qui a pour but de démontrer la sécurité et les performances pour chaque indication revendiquée par les concepteurs pour le dispositif concerné. Cette évaluation se fait avant la mise sur le marché et tout au long de la vie de l'outil et permet d'anticiper la commercialisation. Cette évaluation est réalisée par l'ANSM en collaboration avec tous les acteurs du système de santé.

- l'évaluation médico-économique : cette évaluation intervient après le marquage CE obtenu pour le dispositif médical. L'évaluation est réalisée par la HAS en collaboration avec tous les acteurs du système de santé. L'évaluation médico-économique évalue le service attendu/rendu du dispositif en fonction de³⁷⁸ :
l'intérêt du produit : son effet thérapeutique, diagnostique ou de compensation d'une situation de handicap, ainsi que des effets indésirables ou des risques liés à son utilisation, et sa place dans la stratégie thérapeutique, diagnostique ou de compensation d'une situation de handicap compte tenu des autres thérapies ou moyens disponibles.
- l'intérêt de santé publique : son impact sur la santé de la population, sa capacité à répondre à un besoin non couvert eu égard à la gravité de la pathologie ou de la situation de handicap, son impact sur le système de soins et sur les politiques ou les programmes de santé publique.
Ce critère permet d'apprécier le bien-fondé du remboursement. Cette évaluation est appréciée à une date donnée, dans un environnement évolutif. C'est la CEPS qui discute des prix et des modalités de remboursement avec le concepteur.
- la certification : elle est optionnelle et non obligatoire pour les logiciels aide à la décision dans le processus de soin³⁷⁹. Mais elle est importante pour donner plus de crédibilité aux logiciels développés.

Une fois l'évaluation réalisée, le dispositif est prêt pour être commercialisé et surtout à être utilisé dans la pratique médicale. L'évaluation est très importante car elle facilite l'accès au marché mondial en toute assurance et permet un gain de temps pour les concepteurs.

³⁷⁸ Source : https://www.has-sante.fr/jcms/c_928541/fr/comprendre-l-evaluation-des-dispositifs-medicaux consulté le 31 octobre 2019.

³⁷⁹ Source : https://www.has-sante.fr/jcms/c_989142/fr/certification-des-logiciels-des-professionnels-de-sante consulté le 31 octobre 2019.

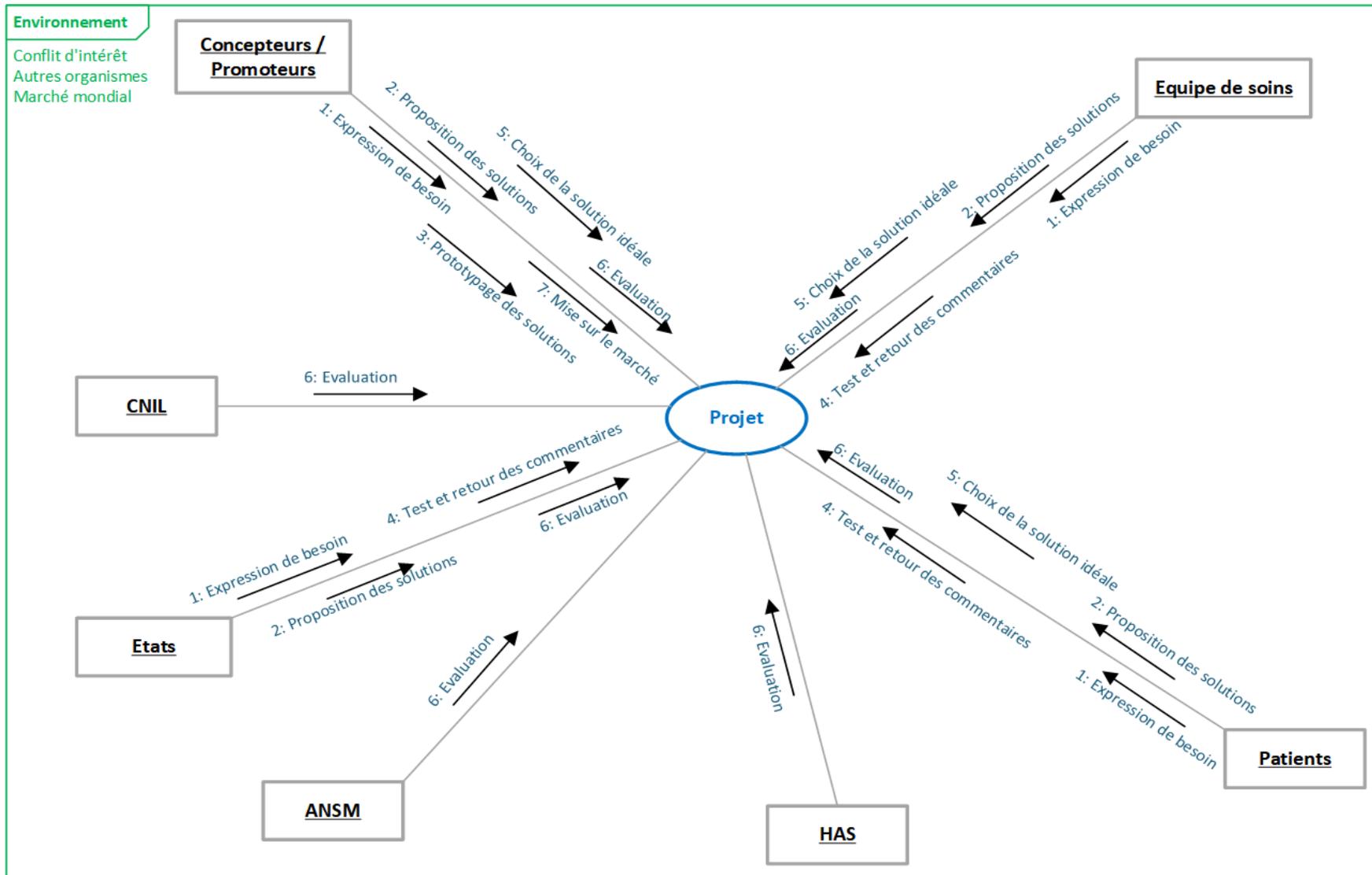


Figure 98 : interactions entre les différents acteurs de l'écosystème de soins de santé

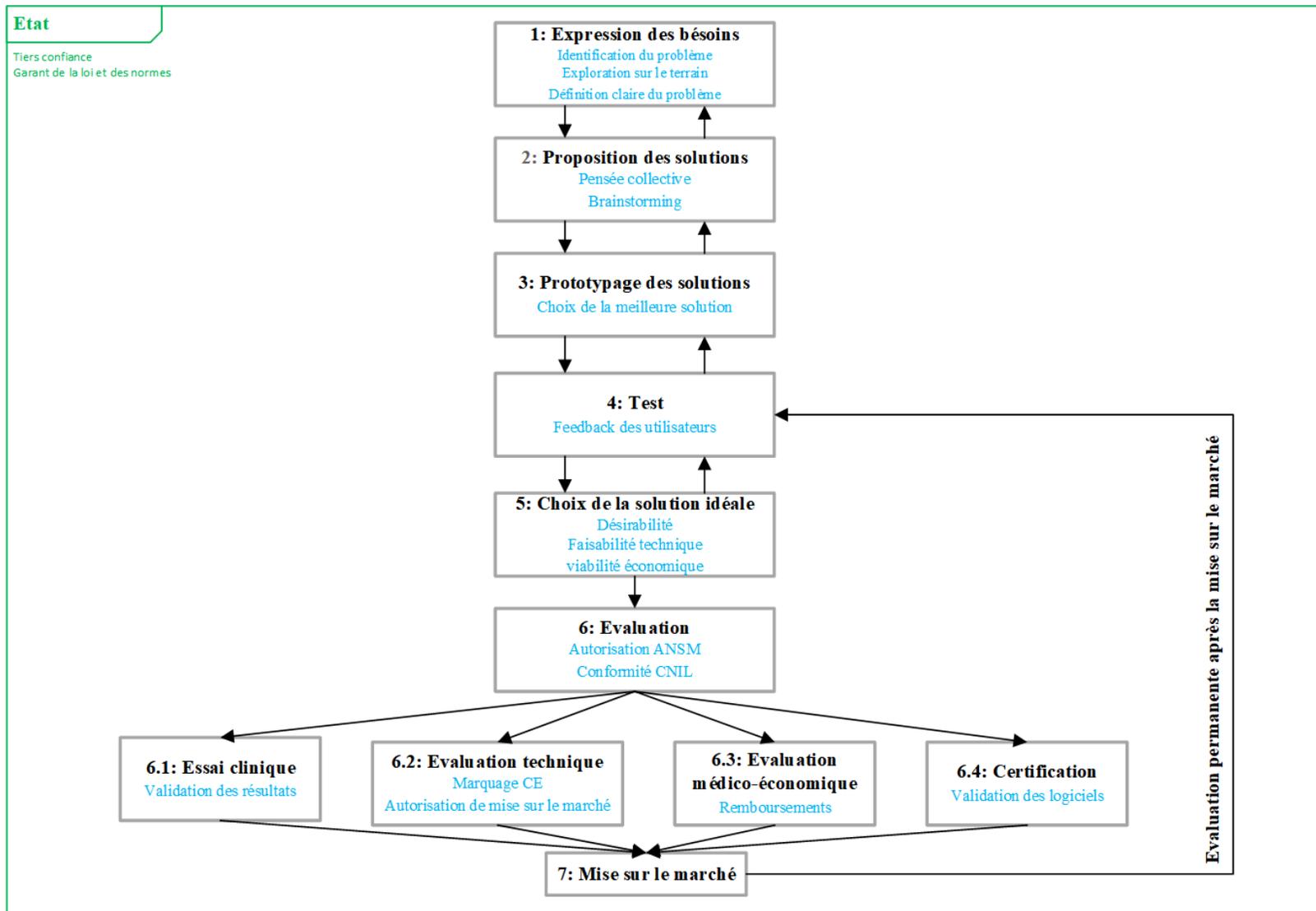


Figure 99 : différentes étapes de conception d'un outil d'IA en santé

7.2.3 *Avantages et inconvénients*

L'utilisation du *design thinking* dans le développement des outils dans le cadre du *big data* et de l'IA dans le domaine de la santé comporte de nombreux avantages pour les acteurs.

Ces avantages sont :

- l'acceptation et l'appropriation rapide des dispositifs par les utilisateurs finaux. En effet, la participation de tous les acteurs dans le processus de conception crée un climat de confiance qui permet son expansion rapide.
- le processus de développement intègre la procédure de validation du dispositif par les sociétés savantes en fonction de l'état de l'art permettant ainsi aux médecins d'utiliser de manière sereine les outils. En plus, le processus prend en compte les règles métiers à savoir la protection du secret médical et le positionnement des outils comme des SADM et sous la responsabilité du médecin.
- tous les acteurs sont informés du mode de conception des outils et surtout du fonctionnement des algorithmes, de la qualité des données utilisées pour l'entraînement des algorithmes et du niveau d'explicabilité des algorithmes. Même si tous les algorithmes ne sont pas explicables pour le moment, le mode de co construction des outils avec la connaissance des données en entrée et en sortie par tous les acteurs réduit de manière significative la méfiance vis-à-vis des algorithmes. La possibilité de biais est ainsi réduite de manière importante et avec la possibilité de faire recenser les risques de biais par les différents acteurs ;
- le respect des principes fondamentaux de protection des données personnelles (collecte loyale en fonction de la finalité et d'un traitement limité, licéité par le consentement éclairé des utilisateurs et l'obligation de transparence et d'information en assurant la confidentialité et la sécurité des données) et des droits fondamentaux des utilisateurs (droit à l'information, droit d'accès et de rectification et le droit à la limitation du traitement).
- la responsabilisation et la valorisation des différents acteurs participant à la mise en œuvre du projet. Les actions de chaque acteur sont définies de manière consensuelle avec tous les acteurs. L'influence de chaque acteur sur l'autre et sur le processus de développement permet de créer un climat de confiance entre tous les acteurs.
- une économie importante dans le développement des outils avec un taux d'échec

considérablement réduit et une accélération de mise sur le marché mondial.

Une telle organisation du développement des outils avec plusieurs acteurs comporte des inconvénients qui sont :

- la gestion de plusieurs groupes de compétences et de personnes dans le développement pouvant être une source de ralentissement et de blocage dans le développement des outils. Mais un bon management des équipes permet rapidement de surmonter ce problème ;
- la possible perte du secret dans la conception des projets et le sentiment de prise de risque et de changement continuels au cours de la mise en œuvre du projet qui peuvent déstabiliser les différents acteurs. Mais pour la création d'une relation de confiance, il faudra que pour une fois les concepteurs se mettent en position de vulnérabilité face aux utilisateurs finaux, qui s'avèrera payante à la fin du processus ;
- le *design thinking* pose le problème de la gestion des propriétés intellectuelles et des brevets surtout dans le cadre de l'utilisation à but commercial. Mais cela ne doit pas empêcher le travail collaboratif des acteurs. Une attente entre les acteurs peut toujours être négociée facilement. Le rôle de chaque acteur et les bénéfices pourront être définis par un contrat.

Notre modèle de développement permet de répondre aux besoins et aux attentes de tous les acteurs sans pour autant mettre en péril les droits et les principes fondamentaux d'un acteur. A travers ce modèle, nous proposons un processus de mise en place du *big data* et de l'IA dans le domaine de la santé respectueuse de :

- l'éthique médicale ;
- des normes et des compétences techniques ;
- la prise en compte du cycle de vie des SADM et leur méthodologie de conception (figure 51) ;
- la protection des données personnelles et surtout des données de santé ;
- la vie privée et de l'intimité des patients ;
- la prise en compte du contexte, de l'environnement et de la singularité des utilisateurs ;

- la pratique médicale.

Les innovations dans le domaine de la santé ne doivent pas être considérées de nos jours comme « *la solution* ». Elles ne doivent pas non plus être vues que sous l'angle des risques qu'elles induisent. La santé est l'un des domaines qui a beaucoup profité du développement de la technologie pour améliorer la prise en charge des maladies. Mais, la technologie a toujours été adaptée et intégrée dans la pratique médicale ; c'est ce qui doit se faire dans le cadre du *big data* et de l'IA et mettre le patient au centre des usages.

Le *big data* et l'IA, réactualisent le mythe de feu de Prométhée³⁸⁰, « *toute innovation technique est à double face selon le changement qu'elle opère dans la distribution et l'exercice du pouvoir, elle enlève le pouvoir à certains pour le donner à d'autres, en changeant la réalité pour tous* » (MALAFOSSE ET AL., 2018).

³⁸⁰ Le mythe de Prométhée reflète la notion universelle de l'ambiguïté du feu « dangereux ami », utile et dangereux.

8 Conclusion

Les progrès de la science mis au service des soins au malade bousculent la relation patient-médecin construite depuis des siècles sur le colloque singulier et le secret médical. Ils produisent l'évolution du système de santé et appellent des logiques pluridisciplinaires et multidimensionnelles. L'avènement de la communication numérique supporte l'ensemble et fournit une puissance dans le traitement des données qui, si elle présente un potentiel énorme, pose aussi de nouvelles questions juridiques, éthiques et sociétales. Parmi elles, celle qui est à la base de notre recherche : l'intérêt et les risques de l'utilisation du *big data* et de l'IA dans la pratique médicale. Une exploration de l'état de l'art dans différentes disciplines nous a conduit à : a) proposer un nouveau système d'aide à la décision médicale (SADM) qui vise à concilier progrès technique et éthique médicale ; b) interroger les risques liés la protection des données ; c) poser la question de l'engagement des acteurs, ce qui nous induit à intégrer l'apport des sciences humaines dans l'analyse de la complexité du système de santé ; d) étudier l'évolution de la pratique médicale face au *big data* et à l'IA ; e) proposer des remédiations aux problèmes posés et f) nous interroger sur l'apport de cette thèse interdisciplinaire à notre pratique personnelle de médecin.

Un SADM reposant sur une extension floue d'un modèle orienté objet

Dans ce travail de thèse, nous avons développé un espace vectoriel flou temporel pour la conception des SADM en médecine clinique. Il s'agit d'une méthodologie de conception des SADM qui repose sur le modèle orientée objet et la logique floue. Nous avons, à cet effet, développé à partir du diagramme de classe d'UML, une ontologie de description des maladies. Nous avons ensuite intégré la gestion de la dynamique temporelle de la maladie, paramètre important pour la comparaison du patient avec lui-même ; chaque évaluation à un instant t correspondant à un cas clinique.

Les avantages de cette méthode sont : a) toutes les étapes de développement des SADM sont explicables et compréhensibles par les médecins ; b) la facilité de mise à jour du SADM qui se fait l'intermédiaire d'une interface expert ; c) le SADM est facilement adaptable au contexte et à l'environnement de travail du médecin ; d) la base de données d'entraînement n'est pas nécessaire ; e) la modélisation de l'évolution temporelle des événements cliniques à différents niveaux de granularité offrant la perspective du raisonnement à partir de cas temporelle des histoires cliniques (RàPCT) ; f) la défuzzification est absente dans le

processus d'analyse et de traitements afin d'obtenir les résultats et g) l'EVFT peut être utilisé pour le développement d'un SADM pour toutes les pathologies.

A partir de l'EVFT, nous avons développé un score de diagnostic clinique de l'EP à partir des recommandations de la ESC. L'EP est une maladie d'évolution rapide, nécessitant une prise en charge rapide. Mais cette pathologie n'a aucun signe clinique spécifique et le diagnostic de confirmation se fait à partir de l'angioscanner. Nous nous sommes appuyés sur les signes cliniques de l'EP pour les classés en trois catégories (statiques, évolutives et dynamiques). Cette classification nous a permis de prendre en compte la dynamique temporelle des signes cliniques qui est l'élément important en cas de suspicion de l'EP en attendant la confirmation diagnostic. Nous espérons à travers ce nouveau score améliorer la prise en charge de cette pathologie grâce à la clinique et sans recourir à un plateau technique insuffisant ou absent. Cette contribution a été précieuse pour compléter et renforcer notre compétence dans la pratique de l'IA.

Mais pour des contraintes de temps, nous n'avons pas pu terminer le développement de la plateforme de conception des SADM qui est actuellement en cours. Par conséquent, nous n'avons pas pu évaluer notre nouveau score (capacité diagnostic, acceptabilité par les médecins et son utilité dans le cadre de la pratique médicale).

Des interrogations sur la protection des données

Nous avons analysé les moyens juridiques de protection via l'état de l'art et notre enquête de terrain auprès des médecins. Notre analyse a montré que les moyens juridiques comportent des insuffisances quant à la protection des données et des difficultés pour leur mise en application. Nous avons noté une évolution importante des moyens juridiques de protection des données et avec surtout l'harmonisation du RGPD dans tout l'espace de l'UE. Mais la limite territoriale de cette loi rend difficile l'application des principes et le respect des droits fondamentaux des personnes. Avec le RGPD, la responsabilité de la sécurité des données incombe au responsable du traitement des données et qu'il doit mettre les moyens nécessaires pour les sécuriser en fonction de la valeur des données et de l'état de l'art. Mais des questions demeurent. Comment évalue-t-on la valeur des données personnelles ? Qui évalue cette valeur ? Et que se passe-t-il si le responsable crée une brèche de sécurité volontaire dans son système ? Les sanctions adoptées par le RGPD en cas de manquement sont des éléments importants dans la régulation. Mais ces sanctions comparées à la valeur que peuvent apporter les données sont dérisoires. Cette faiblesse encourage certaines entreprises à courir le risque de violer les lois. D'autant plus, que la déclaration n'est plus

obligatoire et que la preuve de la défaillance doit être faite lors d'un contrôle de la CNIL ou lorsqu'une plainte a été déposée. La loi ne prend pas en compte le fait qu'il est impossible d'anonymiser certaines données de santé, notamment les données génomiques. Nous avons conclu que les moyens juridiques de protection des données ne sont pas suffisants face à l'évolution rapide de la technologie du *big data* et de l'IA mais constituent néanmoins une avancée majeure en ce sens qu'elles donnent les moyens à la CNIL de prendre des sanctions lorsque la loi n'est pas respectée. Ainsi, ce qui se joue à travers les contraintes juridiques qui servent à encadrer des manières de faire ; ce n'est pas la protection des données mais aussi l'évaluation des comportements d'engagement sécuritaire par les différents acteurs.

L'engagement des acteurs : l'apport des sciences humaines dans l'analyse de la complexité

A travers le paradigme *l'homme-trace* appliqué à la médecine, nous énonçons que les signes (au sens commun), interprétés comme des *signes-traces* d'une maladie diagnostiquée par le médecin ne relèvent pas seulement de l'approche clinique en cabinet médical ou des examens complémentaires réalisés à un instant t. la biographie du patient joue un rôle important, de même que, l'épigénétique et l'histoire familiale. Dans ce travail, nous avons montré que la trace numérique est une partie de la trace en médecine. Elle est issue de l'interprétation des *signes-traces* du corps mise en mémoire sur des supports numériques. Selon le paradigme des *signes-traces*, cette interprétation devrait être faite en tenant compte de l'environnement et des interactions complexes et multiformes qui régissent la vie des acteurs de cette émergence interprétative. Or, la prise en compte de cette complexité reste difficile voire impossible pour les modèles mathématiques et probabilistes à l'origine des algorithmes. En effet, d'une part, le corps est un exemple de système complexe dont toutes les interactions ne sont encore ni « scriptables » ni, pour celles qui le seraient, inscrites sur des supports numériques. D'autre part, le mode d'interprétation de la trace numérique (agrégation de valeur, calcul de moyenne) fait perdre leurs caractéristiques propres (l'individualité de l'objet, le caractère indirect du déchiffrement et le caractère conjectural) et incite à penser l'humain comme un objet, une chose. Or, dans la mesure où « l'objet » de la médecine est l'humain, elle ne constitue pas une science comme un autre : c'est aussi un art, au sens où la finesse d'un diagnostic relève aussi de l'indicible, de la sensibilité sensorielle et émotionnelle. Ainsi, même si le *big data* couplée à l'IA, joue un rôle important à la prise en charge des patients, ils modifient également une relation médecin-patient qui intègre l'interaction des *signes-traces*. C'est pourquoi dans notre analyse du système de

santé, nous avons souhaité intégrer non seulement les recherches en information (et donc les traces au sens informatique) mais aussi les recherches en communication (sur la relation entre *homme-trace*).

Notre enquête sur la pratique médicale face aux *big data* et l'IA

Nous avons réalisé une enquête terrain auprès des médecins pour évaluer l'impact de l'utilisation des TIC dans la pratique médicale. Elle montre une utilisation effective du *big data* et de l'IA se matérialisant par l'utilisation des dossiers patients informatiques : stockage de toutes des données liées aux soins, utilisation des logiciels d'aide à la prescription médicale pour éviter les erreurs et par l'utilisation des objets connectés dans le cadre de la surveillance des maladies. Les médecins sont conscients que ces pratiques comportent des risques d'atteinte à la vie privée des patients via la dégradation de la confidentialité des données dues non seulement à l'évolution interne du système santé mais aussi aux attaques de hackers à l'affût d'informations susceptibles d'être revendues à des tiers, par exemple aux assurances. La capacité de réidentification des données ou la perte de leur qualité en cas d'anonymisation forte augmente encore ce risque. A cela s'ajoutent : a) la présence dans le processus de soins de nouveaux partenaires dont la responsabilité n'est pas clairement définie ; b) un modèle économique des outils fondé sur la valeur marchande des données personnelles ; c) le manque de visibilité dans la gouvernance des données et d) l'absence de validation scientifique des outils de l'IA et des objets connectés. Notre enquête montre que ces paramètres entraînent un manque de confiance des médecins à l'utilisation du *big data* et de l'IA. Mais ce résultat reste à compléter dans la mesure où l'une des principales limites de notre enquête est l'absence de prise en compte de la parole des patients. Cependant, dans l'état actuel de nos connaissances, nous posons comme fortement probable qu'existe de leur part une résistance à l'introduction des *big data* et de l'IA du fait de : a) la disparition du consentement des patients dans l'utilisation de leurs données de santé est associée à une perte de leur autonomie et de leurs droits fondamentaux ; b) la perte de la confidentialité des données par la multiplicité des acteurs et des accès aux données produit une atteinte à l'intimité et à la vie privée des individus et une disparition totale du secret médical ; c) la dépersonnalisation voire la déshumanisation de la pratique médicale met en second plan de la relation humaine au profit de l'interaction homme-machine ; d) l'asservissement des médecins à la technologie risque d'aboutir à une obéissance aveugle à ces outils qui, en devenant, de plus en plus performants, sont susceptibles d'entraîner la perte de l'indépendance du médecin ; e) Les inégalités d'accès aux soins se renforcent du fait du

niveau d'accès inégal aux technologies par les individus et les structures de soins ; f) l'utilisation de la thérapie génomique et de l'échantillonnage des données dans le cadre de l'entraînement des algorithmes peut générer la discrimination des certains groupes d'individus et g) l'usage des technologies dans le cadre de la pratique médicale peut faire l'objet de détournements et devenir un moyen de surveillance et de répression envers les médecins et les patients. La conséquence principale de cet ensemble étant susceptible de provoquer la perte de la confiance dans les médecins, essence même de la pratique médicale, nous avons cherché à élaborer un nouveau processus de développement des systèmes.

La prise en compte d'un écosystème intégrant l'humain, les TIC et les organisations

La robustesse d'un système informatique n'est pas la seule variable à prendre en compte lors du processus de développement d'un meilleur système de santé. La mise en place d'un processus qui assure la confiance des acteurs aux systèmes est essentielle. Nous proposons un processus de développement qui prend en compte tous les acteurs de l'écosystème de la santé. Dans ce processus, la conception du système se fonde sur les besoins des utilisateurs et leurs expériences individuelles et collectives. Les différentes phases de la conception se font de manières itératives et tous les acteurs sont impliqués à chaque étape. Cette méthode permet : a) une appropriation rapide des systèmes par les utilisateurs ; b) un respect des normes métiers et de l'éthique et c) une économie importante en matière de coût. Nous observons actuellement son application au Burkina Faso où le processus est utilisé dans le processus de développement de notre score diagnostique de l'EP avec l'implication de cardiologues et de médecins généralistes et cela bien que le processus comporte encore des inconvénients pour certains acteurs et pour la gestion de la propriété intellectuelle. A la suite de notre thèse, nous comptons donc poursuivre notre recherche : a) d'abord, en terminant le développement de la plateforme de conception des SADM afin d'implémenter notre nouveau score de diagnostic de l'EP et de faire les simulations ; b) en évaluant, après la phase d'implémentation, le score sur le terrain afin de le valider pour un usage en pratique médicale et c) en complétant ensuite notre enquête de terrain auprès des patients avec une nouvelle approche qui permettra d'appréhender les mécanismes d'adaptation du patient. Ainsi notre recherche pourra mieux contribuer à intégrer les TIC dans un système de santé conçu en fonction des besoins réels du patient.

Apport de cette thèse interdisciplinaire à ma pratique

Ma conception de la relation médecin-patient porte les traces d'une culture ancestrale qui n'opère une coupure ni entre le corps et l'esprit ni entre l'humain et le milieu dans lequel il vit. Cette thèse m'a conduit à suspendre pendant quatre ans ma pratique de cardiologue au sein du système de santé au Burkina Faso et à m'intégrer comme doctorant dans un laboratoire de recherche français en sciences humaines et sociales. Nous y avons trouvé l'opportunité de rapprocher notre culture incorporée et les SIC françaises via le paradigme anthropologique de *l'homme-trace*. Nous l'avons donc pris en compte dans une recherche a priori centrée sur la partie la plus contemporaine de l'évolution de la pratique médicale : la question des avantages et inconvénients du *big data* et de l'IA au regard des fondements humains et éthiques de la médecine. Notre approche qui se veut non seulement multidisciplinaire mais aussi interdisciplinaire nous a permis de confronter nos hypothèses en prenant en compte des notions issues de disciplines aussi variées que la médecine, l'informatique, les sciences de l'information et de la communication, la sociologie, le droit, et la philosophie. Cette pluridisciplinarité nous a permis de construire l'objet de notre recherche dans une interdisciplinarité prenant en compte la complexité du système de santé et de proposer une approche systémique pour une aide à la décision médicale. Au moment de la relecture de ma thèse, s'est produit un événement inattendu : la France entrait dans une vague aigue du Covid-19. Chaque citoyen y prenait conscience de l'interaction de tous les systèmes concernés par la gestion de la santé : pas seulement le rôle des agents de santé encore moins celui du numérique ou de l'économie mais celui de tous les acteurs. La participation de tous les individus est apparue comme nécessaire pour que la politique de santé soit efficace. Elle a également montré que la santé est intimement liée à des contextes différents (société, économie, politique) qui sont tous complexes et incertains. L'expérience de vie de chacun lui a montré que, comme nous l'explicitons dans la thèse : a) tout changement entraîne des dynamiques qui peuvent être à l'origine d'instabilité dans le système avec des bouleversements importants et b) l'utilisation de la communication à distance, si elle est d'un recours utile malgré les risques de détournement de traces numériques qu'elle implique, ne peut remplacer la communication en co-présence. Autrement dit, que nulle part, la médiation technologique qui produit une relation humaine biaisée ne peut remplacer la complexité des échanges de *signes-traces* réalisée par des *corps-traces* situés dans une relation de face à face. Cette compréhension a imprégné le plus grand nombre de façon plus ou moins explicite. Mais elle n'est pas pour autant intégrée dans les études médicales occidentales. Aussi, nous envisageons via les séminaires de formation

médicale continue que suivent les médecins en exercice de partager avec les confrères de plusieurs pays les connaissances que nous avons acquises lors de cette thèse soutenue dans le cadre du laboratoire IDEES (UMR 6266) de l'école doctorale Homme, Sociétés, Risques, Territoire (ED HSRT) de Normandie université.

Références

ABITEBOUL, SERGE, 2013. Sciences des données : de la logique du premier ordre à la Toile : Leçon inaugurale prononcée le jeudi 8 mars 2012. Chaire d'Informatique et sciences numériques. In : *Sciences des données : de la logique du premier ordre à la Toile : Leçon inaugurale prononcée le jeudi 8 mars 2012* [en ligne]. Paris : Collège de France. Leçons inaugurales. [Consulté le 31 octobre 2018]. ISBN 978-2-7226-0171-0. Disponible à l'adresse : <http://books.openedition.org/cdf/529>.

ACKLEY, DAVID H., HINTON, GEOFFREY E. ET SEJNOWSKI, TERRENCE J., 1985. A learning algorithm for boltzmann machines. In : *Cognitive Science*. 1 janvier 1985. Vol. 9, n° 1, p. 147-169. DOI 10.1016/S0364-0213(85)80012-4.

ALGORITHM WATCH, 2019. AI Ethics Guidelines Global Inventory. In : *Algorithm Watch* [en ligne]. 21 juin 2019. [Consulté le 17 juillet 2019]. Disponible à l'adresse : <https://algorithmwatch.org/en/project/ai-ethics-guidelines-global-inventory/>.

ALLIOT, JEAN-MARC ET DURAND, NICOLAS, 2005. *Algorithmes génétiques*. Centre d'Etudes de la Navigation Aérienne.

ALLIOT, JEAN-MARC, SCHIEX, THOMAS, BRISSET, PASCAL ET GARCIA, FRÉDÉRIK, 2003. *Intelligence artificielle et informatique théorique*. Toulouse : Cépaduès. ISBN 978-2-85428-578-9.

AMENTA, F., 2000. The International Radio Medical Centre (C.I.R.M.): an organization providing free medical assistance to seafarers of any nationality world wide. In : *International Maritime Health*. 2000. Vol. 51, n° 1-4, p. 85-91.

ANAP, 2014. *Audit des systèmes d'information hospitaliers auprès d'établissements représentatifs*. Paris. ANAP.

ANDERSON, LA ET DEDRICK, ROBERT, 1991. Development of the Trust in Physician Scale: A Measure to Assess Interpersonal Trust in Patient-Physician Relationships. In : *Psychological reports*. 1 janvier 1991. Vol. 67, p. 1091-100. DOI 10.2466/pr0.1990.67.3f.1091.

ARISTOTE, 1998. *Éthique de Nicomaque*. Paris : Flammarion. ISBN 978-2-08-070043-8.

ARNEDOS, MONICA, VIELH, PHILIPPE, SORIA, JEAN-CHARLES ET ANDRE, FABRICE, 2014. The genetic complexity of common cancers and the promise of personalized medicine: is there any hope? In : *The Journal of Pathology*. janvier 2014. Vol. 232, n° 2, p. 274-282. DOI 10.1002/path.4276.

ARTHUR, DAVID ET VASSILVITSKII, SERGEI, 2007. k-means++: the advantages of careful seeding. In : *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*. New Orleans, Louisiana : Society for Industrial and Applied Mathematics. 7 janvier 2007. p. 1027–1035.

AZENCOTT, CHLOÉ-AGATHE, 2018. *Introduction au Machine Learning*. Paris : Dunod. ISBN 978-2-10-078080-8.

BAGROS, PHILIPPE, BENSÄÏAH, CLAUDE, CAMUS, VINCENT, RUSCH, EMMANUEL, TOFFOL, BERTRAND DE ET LEMOINE, MAËL, 2012. *Manuel de sciences humaines en médecine*. Paris : Ellipses. ISBN 978-2-7298-7865-8.

BAHMANI, BAHMAN, MOSELEY, BENJAMIN, VATTANI, ANDREA, KUMAR, RAVI ET VASSILVITSKII, SERGEI, 2012. Scalable k-means++. In : *Proceedings of the VLDB Endowment*. 1 mars 2012. Vol. 5, n° 7, p. 622–633. DOI 10.14778/2180912.2180915.

BALAGUÉ, CHRISTINE, PICARD, FLORENCE, DOS, BENOIT, VAYRE, JEAN-SÉBASTIEN, BÉRIOT, QUENTIN, PON, DOMINIQUE ET MAÎTREJEAN, SERGE, 2018. Les pistes de réflexions éthiques pour tendre vers une confiance numérique. In : *Vademecum sur le traitement des données numériques: Réflexions et bonnes pratiques éthiques*. Saint-Herblain : ADEL. p. 110-147.

BALMISSE, GILLES, 2015. *Digital Workplace et gestion des connaissances: concepts et mise en oeuvre*. Saint-Herblain : Éd. ENI. ISBN 978-2-7460-9574-8.

BARDIN, LAURENCE, 2013. *L'analyse de contenu*. Paris : Presses universitaires de France. ISBN 978-2-13-062790-6.

BAUMANN, F., 2004. L'éthique médicale au quotidien : un concept efficace d'éthique clinique. In : *Éthique & Santé*. 1 septembre 2004. Vol. 1, n° 3, p. 131-133. DOI 10.1016/S1765-4629(04)94684-3.

- BEAUCHAMP, TOM L. ET CHILDRESS, JAMES F., 2009. *Principles of biomedical ethics*. 6th ed. New York : Oxford University Press. ISBN 978-0-19-533570-5. R724 .B36 2009
- BEAUD, STÉPHANE ET WEBER, FLORENCE, 2015. *Guide de l'enquête de terrain: produire et analyser des données ethnographiques*. Paris : La Découverte. ISBN 978-2-7071-6008-9.
- BEHMO, REGIS ET TRAVERS, NICOLAS, 2019. Maîtrisez les bases de données NoSQL. In : *OpenClassrooms* [en ligne]. 5 janvier 2019. [Consulté le 10 septembre 2019]. Disponible à l'adresse : <https://openclassrooms.com/fr/courses/4462426-maitrisez-les-bases-de-donnees-nosql>.
- BENAROYO, LAZARE, 2004. Méthodologie en éthique clinique: une approche intégrant les diverses dimensions éthiques du soin. In : *Médecine et Hygiène*. 2004. Vol. 62, p. 1304- 07.
- BENAROYO, LAZARE, 2014. Emergence de la bioéthique. In : *Traité de bioéthique: Tome I, Fondements, principes, repères*. Toulouse : Eres. p. 95-106. ISBN 978-2-7492-1305-7.
- BENGHOZI, PIERRE-JEAN, BUREAU, SYLVAIN ET MASSIT-FOLLÉA, FRANÇOISE, 2012. Définir l'internet des objets. In : *L'Internet des objets: Quels enjeux pour l'Europe* [en ligne]. Paris : Éditions de la Maison des sciences de l'homme. praTICs. p. 15-23. [Consulté le 23 octobre 2017]. ISBN 978-2-7351-1587-7. Disponible à l'adresse : <http://books.openedition.org/editionsmsh/84>.
- BENTHAM, JEREMY, 1789. *An introduction to the principles of morals and legislation* [en ligne]. London (UK) : Oxford: Clarendon Press,. [Consulté le 12 mai 2020]. Disponible à l'adresse : <https://gallica.bnf.fr/ark:/12148/bpt6k93974k>.
- BÉRANGER, JÉRÔME, 2016. *Les Big Data et l'éthique*. Industrialisation de la santé. London (UK) : ISTE éditions. ISBN 978-1-78405-129-7.
- BERGEAT, MAXIME, 2014. Données individuelles : bien les protéger pour mieux les diffuser. In : *46èmes journées de Statistique de la SFdS*. Rennes : SFdS. 2014. p. 7.
- BERNER, ETA S, 2009. Clinical Decision Support Systems: State of the Art. In : *Rockville, Maryland: Agency for Healthcare Research and Quality*. 2009. n° 09-0069-EF., p. 26.

BILLAUD, MARC ET GUCHET, XAVIER, 2015. L'invention de la médecine personnalisée - Entre mutations technologiques et utopie. In : *médecine/sciences*. 1 août 2015. Vol. 31, n° 8-9, p. 797-803. DOI 10.1051/medsci/20153108020.

BISHOP, CHRISTOPHER M., 1995. *Neural Networks for Pattern Recognition*. New York, NY, USA : Oxford University Press, Inc. ISBN 978-0-19-853864-6.

BIZOUARN, P., 2008. Le médecin, le malade et la confiance. In : *Éthique & Santé*. 1 septembre 2008. Vol. 5, n° 3, p. 165-172. DOI 10.1016/j.etiqe.2008.04.001.

BLUMENTRITT, ROLF ET JOHNSTON, RON, 1999. Towards a Strategy for Knowledge Management. In : *Technology Analysis & Strategic Management*. 1 septembre 1999. Vol. 11, n° 3, p. 287-300. DOI 10.1080/095373299107366.

BONNEVILLE, LUC ET GROSJEAN, SYLVIE, 2007. Les défis que soulève l'informatisation de la pratique médicale sur le plan de l'innovation technologique. In : *Canadian Journal of Communication*. 12 novembre 2007. Vol. 32, n° 3, p. 435-56. DOI 10.22230/cjc.2007v32n3a1914.

BOOCH, GRADY, RUMBAUGH, JAMES ET JACOBSON, IVAR, 1999. *The Unified Modeling Language User Guide*. Redwood City, CA, USA : Addison Wesley Longman Publishing Co., Inc. ISBN 978-0-201-57168-4.

BOSER, B.E., SACKINGER, E., BROMLEY, J., LE CUN, Y. ET JACKEL, L.D., 1991. An analog neural network processor with programmable topology. In : *IEEE Journal of Solid-State Circuits*. décembre 1991. Vol. 26, n° 12, p. 2017-2025. DOI 10.1109/4.104196.

BOSER, BERNHARD E., GUYON, ISABELLE M. ET VAPNIK, VLADIMIR N., 1992. A training algorithm for optimal margin classifiers. In : *Proceedings of the fifth annual workshop on Computational learning theory* [en ligne]. Pittsburgh, Pennsylvania, USA : Association for Computing Machinery. 1 juillet 1992. p. 144–152. [Consulté le 30 avril 2020]. Disponible à l'adresse : <https://doi.org/10.1145/130385.130401>.

BOUCHON-MEUNIER, BERNADETTE, 2007. *La logique floue* [en ligne]. 4e. Paris : Presses Universitaires de France. [Consulté le 4 mai 2020]. Que sais-je? ISBN 978-2-13-056260-3. Disponible à l'adresse : <https://www.cairn.info/la-logique-floue--9782130562603.htm>.

BOULLIER, DOMINIQUE, 2015. Les sciences sociales face aux traces du big data. In : *Revue française de science politique*. 2015. Vol. 65, n° 5, p. 805-828. DOI 10.3917/rfsp.655.0805.

BRASSEUR, CHRISTOPHE, 2013. *Enjeux et usages du big data : Technologies, méthodes et mise en oeuvre*. Paris : Hermes Science Publications. ISBN 978-2-7462-4520-4.

BRATMAN, MICHAEL E., ISRAEL, DAVID J. ET POLLACK, MARTHA E., 1988. Plans and resource-bounded practical reasoning. In : *Computational Intelligence*. 1988. Vol. 4, n° 3, p. 349-355. DOI 10.1111/j.1467-8640.1988.tb00284.x.

BREIMAN, LEO, 2001. Random Forests. In : *Machine Learning*. 2001. Vol. 45, n° 1, p. 5-32. DOI 10.1023/A:1010933404324.

BROOKS, RODNEY A, 1991. How to Build Complete Creatures Rather than Isolated Cognitive Simulators. In : *Architectures for Intelligence*. Carnegie Mellon Symposia on Cognition Series. Laurence Erlbaum Ass., Hillsdale, New Jersey : Kurt Van Lehn, Kurt Van Lehn. p. 225-239.

BROWN, TIM, 2008. Design Thinking. In : *Harvard Business Review*. 1 juin 2008. n° June 2008, p. 10.

BUCHANAN, BG ET SHORTLIFFE, EDWARD H., 1984. *Rule-based Expert Systems: MYCIN Experiments of the Stanford Heuristic Programming by Bruce G. Buchanan*. First Edition edition. Massachusetts, USA : Addison Wesley Longman Publishing Co.

BUDIN-LJØSNE, ISABELLE, TEARE, HARRIET J. A., KAYE, JANE, BECK, STEPHAN, BENTZEN, HEIDI BEATE, CAENAZZO, LUCIANA, COLLETT, CLIVE, D'ABRAMO, FLAVIO, FELZMANN, HEIKE, FINLAY, TERESA, JAVAID, MUHAMMAD KASSIM, JONES, ERICA, KATIĆ, VIŠNJA, SIMPSON, AMY ET MASCALZONI, DEBORAH, 2017. Dynamic Consent: a potential solution to some of the challenges of modern biomedical research. In : *BMC Medical Ethics*. 25 janvier 2017. Vol. 18, n° 1, p. 4. DOI 10.1186/s12910-016-0162-9.

BYK, CHRISTIAN, 2008. Pour une bioéthique fondée sur le mariage de la réflexion et de l'action. In : *Médecine & Droit*. 1 novembre 2008. Vol. 2008, n° 93, p. 163-164. DOI 10.1016/j.meddro.2008.10.005.

CACHIA, ARNAUD, LAMBERT, JEAN-CHARLES, NABI, HERMANN, COUTELEC, LÉO ET CHOURAKI, VINCENT, 2015. Emettre des hypothèses ou multiplier les découvertes à partir des données? In : *Cahier de l'Espace Ethique*. septembre 2015. n° 02, p. 35-37.

CANGUILHEM, GEORGES, 2013. *Le normal et le pathologique*. 12e édition. Paris : Presses Universitaires de France - PUF. ISBN 978-2-13-061950-5.

CARDON, DOMINIQUE, 2015. *A quoi rêvent les algorithmes: nos vies à l'heure des big data*. Paris : La République des idées : Seuil. La République des idées. ISBN 978-2-02-127996-2. QA76.9.C66 C37 2015

CASE, MEREDITH A., BURWICK, HOLLAND A., VOLPP, KEVIN G. ET PATEL, MITESH S., 2015. Accuracy of smartphone applications and wearable devices for tracking physical activity data. In : *JAMA*. 10 février 2015. Vol. 313, n° 6, p. 625-626. DOI 10.1001/jama.2014.17841.

CASTELLS, MANUEL, 2001. *The Internet Galaxy: Reflections on the Internet, Business, and Society*. USA : Oxford University Press, Inc. ISBN 978-0-19-924153-8.

CAVOUKIAN, ANN, 2006. *Creation of a Global Privacy Standard* [en ligne]. novembre 2006. S.l. : Information & Privacy Commissioner. [Consulté le 24 avril 2018]. Disponible à l'adresse : www.ipc.on.ca/images/Resources/gps.pdf.

CHAN, ISAAC S. ET GINSBURG, GEOFFREY S., 2011. Personalized medicine: progress and promise. In : *Annual Review of Genomics and Human Genetics*. 2011. Vol. 12, p. 217-244. DOI 10.1146/annurev-genom-082410-101446.

CHARLES, C., GAFNI, A. ET WHELAN, T., 1997. Shared decision-making in the medical encounter: what does it mean? (or it takes at least two to tango). In : *Social Science & Medicine (1982)*. mars 1997. Vol. 44, n° 5, p. 681-692. DOI 10.1016/s0277-9536(96)00221-3.

CHEVRIE, FRANÇOIS ET GUÈLY, FRANÇOIS, 1998. La logique floue. In : *Cahier Technique Schneider*. mars 1998. n° 191, p. 1-30.

CHISHOLM, RODERICK, 1982. Knowledge as justified true belief. In : *The Foundations of Knowing*. NED, New Edition. Minnesota (USA) : University of Minnesota Press. p. 43-49.

CLÉRET, MIREILLE, BEUX, PIERRE LE ET DUFF, FRANCK LE, 2001. Les systèmes d'aide à la décision médicale. In : *Les Cahiers du numérique*. 2001. Vol. 2, n° 2, p. 125-154.

CLERGUE, MANUEL, 2009. Réseaux de neurones artificiels. In : [en ligne]. Université Antilles Guyane. 2009. [Consulté le 18 novembre 2019]. Disponible à l'adresse : <http://www-lisic.univ-littoral.fr/~verel/TEACHING/08-09/sac-M1/cRdNV9.pdf>.

CNAMTS, DSES ET DSEP, 2013. *Risque d'embolie pulmonaire, d'accident vasculaire cérébral ischémique et d'infarctus du myocarde chez les femmes sous contraceptif oral combiné en France: une étude de cohorte sur 4 millions de femmes de 15 à 49 ans à partir des données du SNIIRAM et du PMSI*. Paris. CNAMTS.

CNE, 1993. *Rapport du Comité Nationale d'Evaluation sur les Sciences d'Informations et de la Communications*. Paris. CNE.

COHEN, ALEXANDER T., AGNELLI, GIANCARLO, ANDERSON, FREDERICK A., ARCELUS, JUAN I., BERGQVIST, DAVID, BRECHT, JOSEF G., GREER, IAN A., HEIT, JOHN A., HUTCHINSON, JULIA L., KAKKAR, AJAY K., MOTTIER, DOMINIQUE, OGER, EMMANUEL, SAMAMA, MEYER-MICHEL, SPANNAGL, MICHAEL ET VTE IMPACT ASSESSMENT GROUP IN EUROPE (VITAE), 2007. Venous thromboembolism (VTE) in Europe. The number of VTE events and associated morbidity and mortality. In : *Thrombosis and Haemostasis*. octobre 2007. Vol. 98, n° 4, p. 756-764.

COLLECTIF, 2013. *Petit robert 2014: dictionnaire de la langue française*. Edition 2014. Paris : Le Robert. ISBN 978-2-321-00344-1.

COLLOC, J ET VINOT, D, 2000. A Decision Support System to encode the diseases in hospital. In : *Proceedings of ICSSHC 2000*. Budapest : s.n. 2000. p. 55-61.

COLLOC, JOEL, 2000. *Les apports réciproques de la modélisation objet et de la connaissance médicale*. HDR. Lyon - France : Université Claude Bernard - Lyon 1.

COLLOC, JOËL, 2015. Santé et Big Data : l'Etat et les individus, impuissants face aux pouvoirs des réseaux. In : *L'Espace Politique* [en ligne]. 22 juillet 2015. n° 2. [Consulté le 4 janvier 2017]. DOI 10.4000/espacepolitique.3493. Disponible à l'adresse : <http://espacepolitique.revues.org/3493>.

COLLOC, JOËL, 2020. Fvsoomm a Fuzzy Vectorial Space Model and Method of Personality, Cognitive Dissonance and Emotion in Decision Making. In : *Information*. avril 2020. Vol. 11, n° 4, p. 229. DOI 10.3390/info11040229.

COLLOC, JOEL ET BONENFANT, MAUDE, 2017. Des traces humaines dans le big data : du diagnostic médical à la transduction machinique du corps. In : *L'Homme-Trace: des traces du corps au corps-trace*. Paris : CNRS Editions. Homme Trace. p. 85-110.

COLLOC, JOËL ET HÉNOCQUE, BRUNO, 2016. Introduction - Enjeux du big data et identifications des données médicales. In : *Les Cahiers du numérique*. 25 mai 2016. Vol. 12, n° 1, p. 9-12.

COLLOC, JOËL ET JACQUET-ANDRIEU, ARMELLE, 2013. Système multi-agents d'aide à la décision clinique, appliqué à la prise en charge neuropsychologique des aphasiques par AVC. In : *Les systèmes informatisés complexe en santé*. Paris : IREB, Dalloz. 2013. p. 183-207.

COLLOC, JOËL ET LÉRY, LOUIS, 2008a. Un métamodèle d'aide à la décision en éthique médicale. In : *Santé Décision Management*. 30 mars 2008. Vol. 11, n° 1-2, p. 255-274. DOI 10.3166/sdm.11.1-2.255-274.

COLLOC, JOËL ET LÉRY, LOUIS, 2008b. Un métamodèle d'aide à la décision en éthique médicale. In : *Santé Décision Management*. 30 mars 2008. Vol. 11, n° 1-2, p. 255-274. DOI 10.3166/sdm.11.1-2.255-274.

COLLOC, JOËL ET LÉRY, NICOLE, 1997. A multi-expert decision support system in medical ethics, Health and System Science, Paris : Hermès, Vol 1, N°1, pp. 39-55. In : *Health and System Science*. 1997. Vol. 1, n° 1, p. 39-55.

COLLOC, JOEL ET SUMMONS, PETER, 2006. An object oriented time model of a decision support system for intoxication diagnoses. In : *4th International Industrial Simulation Conference 2006: ISC'2006* [en ligne]. Palerma, Italie : EUROSIS- ETI. juin 2006. p. 387-392. [Consulté le 5 novembre 2019]. Disponible à l'adresse : <https://ogma.newcastle.edu.au/vital/access/manager/Repository/uon:2776>.

COLLOC, JOEL ET SUMMONS, PETER, 2015. An analogical model to design time in clinical objects. In : *RITS 2015* [en ligne]. Dourdan, France : French Society of BioMedical

Engineering (SFGBM). 2015. p. 124-125. [Consulté le 5 novembre 2019]. Disponible à l'adresse : <https://ogma.newcastle.edu.au/vital/access/manager/Repository/uon:2776>.

COLLOC, JOEL ET SYBORD, CHRISTINE, 2003. A multi-agent approach to involve multiple knowledge models and the case base reasoning approach in decision support systems. In : *Proceedings of the 35th Southeastern Symposium on System Theory, 2003*. Morgantown, USA : IEEE. mars 2003. p. 247-251.

COLLOC, JOËL, YAMEOGO, RELWENDE A., SUMMONS, PETER F., SHEN, YING, PARK, MIRA ET ARONSON, JANINE E., 2017. EPICE an emotion fuzzy vectorial space for time modeling in medical decision. In : *Proceedings of the 1st International Conference on Internet of Things and Machine Learning*. Liverpool : ACM. 2017. p. 1-7.

CONSEIL DE L'EUROPE, 2014. Histoire de l'intelligence artificielle. In : *Conseil de l'Europe* [en ligne]. 2014. [Consulté le 9 avril 2019]. Disponible à l'adresse : <https://www.coe.int/fr/web/artificial-intelligence/history-of-ai>.

CONSEIL NATIONAL DE L'ORDRE DES MÉDECINS, 2018. *Médecins et patients dans le monde des data, des algorithmes et de l'intelligence artificielle* [en ligne]. France : Citizen Press. Disponible à l'adresse : https://www.conseil-national.medecin.fr/sites/default/files/cnomdata_algorithmes_ia.pdf.

CONTANTINIDES, YANNIS, 2014. Limites du principes d'autonomie. In : *Traité de bioéthique: Tome I, Fondements, principes, repères*. Toulouse : Eres. p. 158-173. ISBN 978-2-7492-1305-7.

COOMANS, D. ET MASSART, D.L., 1982. Alternative k-nearest neighbour rules in supervised pattern recognition. In : *Analytica Chimica Acta*. 1982. Vol. 136, p. 15-27. DOI 10.1016/S0003-2670(01)95359-0.

CORDIER, AMÉLIE, LEFEVRE, MARIE, CHAMPIN, PIERRE-ANTOINE ET MILLE, ALAIN, 2013. Modéliser les traces d'interaction pour raisonne a partir de l'expérience tracée ? In : *IC - 24èmes Journées francophones d'Ingénierie des Connaissances* [en ligne]. Lille, France : s.n. juillet 2013. [Consulté le 5 juin 2020]. Disponible à l'adresse : <https://hal.inria.fr/hal-01103671>.

CORNUÉJOLS, ANTOINE, MICLET, LAURENT ET BARRA, VINCENT, 2018. *Apprentissage artificiel: deep learning, concepts et algorithmes*. 3ème édition. Paris : Eyrolles. ISBN 978-2-212-67522-1.

COROLLER-BEC, C., DAMBRINE, S., CHOUKROUN, J., KERLIR, G., DELORI, M. ET ROY, P.-M., 2008. Comparaison de différentes méthodes d'évaluation de la probabilité clinique lors d'une suspicion d'embolie pulmonaire. In : *Journal Européen des Urgences*. mars 2008. Vol. 21, p. A87-A88. DOI 10.1016/j.jeur.2008.03.299.

CREDOC, 2019. 19 : *Enquête sur la diffusion des technologies de l'information et de la communication dans la société française en 2019* [en ligne]. Paris. CREDOC. [Consulté le 31 mai 2018]. Disponible à l'adresse : <http://www.credoc.fr/pdf/Rapp/R337.pdf>.

DAGIRAL, ÉRIC ET PARASIE, SYLVAIN, 2017. La « science des données » à la conquête des mondes sociaux : ce que le « Big Data » doit aux épistémologies locales. In : MENGER, Pierre-Michel et PAYE, Simon (éd.), *Big data et traçabilité numérique : Les sciences sociales face à la quantification massive des individus* [en ligne]. Paris : Collège de France. Conférences. p. 85-104. [Consulté le 3 octobre 2018]. ISBN 978-2-7226-0467-4. Disponible à l'adresse : <http://books.openedition.org/cdf/4999>.

DAMBRINE, S., BORDOT, A. -S., PATTIER, K. ET ROY, P. -M., 2008. Facteurs prédictifs et éléments diagnostiques de l'embolie pulmonaire aux urgences. In : *Réanimation*. décembre 2008. Vol. 17, n° 8, p. 745-752. DOI 10.1016/j.reaurg.2008.09.001.

DANIEL, CHRISTEL, OUAGNE, DAVID, SADOU, ERIC, FORSBERG, KERSTIN, GILCHRIST, MARK MC, ZAPLETAL, ERIC, PARIS, NICOLAS, HUSSAIN, SAJJAD, JAULENT, MARIE-CHRISTINE ET MD, DIPKA KALRA, 2016. Cross border semantic interoperability for clinical research: the EHR4CR semantic resources and services. In : *AMIA Summits on Translational Science Proceedings*. 20 juillet 2016. Vol. 2016, p. 51-59.

DARMONI, STEFAN, GRIFFON, NICOLAS ET MASSARI, PHILIPPE, 2014. Les systèmes d'aide à la décision médicale. In : *Annales des Mines - Realites industrielles*. 19 décembre 2014. Vol. Novembre 2014, n° 4, p. 47-50.

DASGUPTA, D. (éd.), 1998. *Artificial immune systems and their applications*. Berlin ; New York : Springer. ISBN 978-3-540-64390-6. QR182.2.C65 A78 1998

DAUDIN, H., 1951. Les rapports de l'être et de la connaissance chez Platon et chez Aristote. In : *Revue des Études Anciennes*. 1951. Vol. 53, n° 1, p. 26-41. DOI 10.3406/rea.1951.3439.

DAVALLON, JEAN, 2006. *Le don du patrimoine: une approche communicationnelle de la patrimonialisation*. Paris : Lavoisier : Hermès science. Collection Communication, médiation et construits sociaux. ISBN 978-2-7462-1436-1. CC135 .D38 2006

DAVIES, DAVID L. ET BOULDIN, DONALD W., 1979. A Cluster Separation Measure. In : *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1 février 1979. Vol. 1, n° 2, p. 224–227. DOI 10.1109/TPAMI.1979.4766909.

DAVIS, KORD ET PATTERSON, DOUG, 2012. *Ethics of big data*. Sebastopol, CA : O'Reilly. ISBN 978-1-4493-1179-7. HF5387 .D36 2012

DE BONO, J. S. ET ASHWORTH, ALAN, 2010. Translating cancer research into targeted therapeutics. In : *Nature*. 30 septembre 2010. Vol. 467, n° 7315, p. 543-549. DOI 10.1038/nature09339.

DE CASTRO, LEANDRO N. ET TIMMIS, JONATHAN, 2002. *Artificial immune systems: a new computational intelligence approach*. London ; New York : Springer. ISBN 978-1-85233-594-6. QA76.87 .D43 2002

DEAN, JEFFREY ET GHEMAWAT, SANJAY, 2008. MapReduce: simplified data processing on large clusters. In : *Communications of the ACM*. 1 janvier 2008. Vol. 51, n° 1, p. 107. DOI 10.1145/1327452.1327492.

DEGOULET, PATRICE ET FIESCHI, MARIUS, 1991. *Traitement de l'information médicale: méthodes et applications hospitalières*. Paris : Masson. ISBN 978-2-225-82514-9.

DELAHAYE, F, 2015. Recommandations de la Société européenne de Cardiologie sur l'embolie pulmonaire. In : *Réalités Cardiologiques* [en ligne]. 6 mai 2015. [Consulté le 7 octobre 2019]. Disponible à l'adresse : <https://www.realites-cardiologiques.com/2015/05/06/recommandations-de-la-societe-europeenne-de-cardiologie-sur-lembolie-pulmonaire/>.

DELL, 2016. *Enquête Dell 2016 sur la sécurité des données* [en ligne]. Enquête. New York, NY, USA. Dell. Disponible à l'adresse : http://i.dell.com/sites/doccontent/shared-content/data-sheets/fr/Documents/Enquete_securite_des_donnees_2016_FR.pdf.

DELL, 2017. *Enquête sur la sécurité Dell End-User 2017* [en ligne]. New York, NY, USA. Dell. Disponible à l'adresse : <https://france.emc.com/collateral/analyst-reports/dell-end-user-security-survey-2017.pdf>.

DEMEUSE, MARC, 2008. Echelles de Likert ou méthode des classements additionnés. In : *Introduction aux théories et aux méthodes de la mesure en sciences psychologiques et en sciences de l'éducation* [en ligne]. 2008. [Consulté le 8 mars 2018]. Disponible à l'adresse : http://ired.u-bourgogne.fr/images/stories/Documents/Cours_disponibles/Demeuse/Cours/p5.3.pdf.

DENNETT, DANIEL CLEMENT, 1989. *The Intentional Stance*. Massachusetts, USA : MIT Press. ISBN 978-0-262-54053-7.

DÉPELTEAU, FRANÇOIS, 1998. *La démarche d'une recherche en sciences humaines: de la question de départ à la communication des résultats*. Sainte-Foy, Québec : Presses de l'Université Laval. ISBN 978-2-7637-7604-0.

DEROIN, VALÉRIE, 2013. Les ménages et les technologies de l'information et de la communication (tic) en France et en Europe en 2012. In : *Culture chiffres*. 31 mai 2013. n° 2, p. 1-8. DOI 10.3917/culc.132.0001.

DERRIDA, JACQUES, 1979. *L'écriture et la différence*. Paris : Seuil. ISBN 978-2-02-005182-8.

DGOS, 2012. *Télémedecine et responsabilités juridiques engagées*. Paris. DGOS.

DIGNUM, VIRGINIA, LOPEZ-SANCHEZ, MAITE, MICALIZIO, ROBERTO, PAVÓN, JUAN, SLAVKOVIK, MARIJA, SMAKMAN, MATTHIJS, VAN STEENBERGEN, MARLIES, TEDESCHI, STEFANO, VAN DER TOREE, LEON, VILLATA, SERENA, DE WILDT, TRISTAN, BALDONI, MATTEO, BAROGLIO, CHRISTINA, CAON, MAURIZIO, CHATILA, RAJA, DENNIS, LOUISE, GÉNOVA, GONZALO, HAIM, GALIT ET KLIESS, MALTE S., 2018. Ethics by Design: Necessity or Curse? In : *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society - AIES '18* [en ligne]. New Orleans, LA, USA : ACM Press. 2018. p. 60-66. [Consulté le 13 mars 2020]. Disponible à l'adresse : <http://dl.acm.org/citation.cfm?doid=3278721.3278745>.

DOMINGO-FERRER, JOSEP, SÁNCHEZ, DAVID ET SORIA-COMAS, JORDI, 2016. Database Anonymization: Privacy Models, Data Utility, and Microaggregation-based Inter-

model Connections. In : *Synthesis Lectures on Information Security, Privacy, and Trust*. 11 janvier 2016. Vol. 8, n° 1, p. 1-136. DOI 10.2200/S00690ED1V01Y201512SPT015.

DOMINGOS, PEDRO ET PAZZANI, MICHAEL, 1997. On the Optimality of the Simple Bayesian Classifier under Zero-One Loss. In : *Machine Language*. 1 novembre 1997. Vol. 29, n° 2-3, p. 103–130. DOI 10.1023/A:1007413511361.

DONALDSON, THOMAS M., 2012. Ethical Resources for the Clinician: Principles, Values and Other Theories. In : CLARK, Peter A. (éd.), *Contemporary Issues in Bioethics* [en ligne]. London (UK) : InTech. p. 15-38. [Consulté le 28 juin 2019]. ISBN 978-953-51-0169-7. Disponible à l'adresse : <http://www.intechopen.com/books/contemporary-issues-in-bioethics/ethical-resources-for-the-clinician-principles-values-and-other-theories->.

DORIGO, M., MANIEZZO, V. ET COLORNI, A., 1996. Ant system: optimization by a colony of cooperating agents. In : *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*. février 1996. Vol. 26, n° 1, p. 29-41. DOI 10.1109/3477.484436.

DRIANCOURT, XAVIER, 1994. *Optimisation par descente de gradient stochastique de systemes modulaires combinant reseaux de neurones et programmation dynamique. Application a la reconnaissance de la parole* [en ligne]. thesis. Paris : Université Paris 11. [Consulté le 30 avril 2020]. Disponible à l'adresse : <http://www.theses.fr/1994PA112203>.

DUFOUR, JEAN-CHARLES, GROSJEAN, JULIEN, DARMONI, STEFAN, YASINI, MOBIN, MARCHAND, GUILLAUME, SIMON, CHRISTIAN, SARRADON-ECK, ALINE, PRÉAU, MARIE, DARMON, DAVID, SCHUERS, MATTHIEU, HASSANALY, PARINA ET GIORGI, ROCH, 2019. ApiAppS: A Project to Study and Help Practitioners in Recommending mHealth Apps and Devices to Their Patients. In : *Studies in Health Technology and Informatics*. 21 août 2019. Vol. 264, p. 1919-1920. DOI 10.3233/SHTI190713.

DUPONT, BERNARD-MARIE, 2001. Evidence-Based Medicine : un développement à risques. In : *Actes des XVIes journées d'éthique de Lyon*. Lyon : SEL. 2001. p. 87-94.

DYER, JEFFREY H., 1997. Effective interim collaboration: how firms minimize transaction costs and maximise transaction value. In : *Strategic Management Journal*. 1997. Vol. 18, n° 7, p. 535-556. DOI 10.1002/(SICI)1097-0266(199708)18:7<535::AID-SMJ885>3.0.CO;2-Z.

EL MALKI, MOHAMMED, 2016. *Modélisation NoSQL des entrepôts de données multidimensionnelles massives* [en ligne]. Theses. Toulouse : Université Toulouse le Mirail - Toulouse II. [Consulté le 10 septembre 2019]. Disponible à l'adresse : <https://tel.archives-ouvertes.fr/tel-02057102>.

EMANUEL, EZECKIEL J. ET EMANUEL, LINDA L., 1992. Four models of the physician-patient relationship. In : *JAMA*. 22 avril 1992. Vol. 267, n° 16, p. 2221-2226.

ERMINE, JEAN-LOUIS, 2000. *Les systèmes de connaissances*. Paris : Hermes Science Publication. ISBN 978-2-7462-0159-0.

ERMINE, JEAN-LOUIS, MORADI, MAHMOUD ET BRUNEL, STÉPHANE, 2012. Une chaîne de valeur de la connaissance. In : *Management international*. 2012. Vol. 16, p. 29. DOI 10.7202/1012391ar.

FAGIN, RONALD ET HALPERN, JOSEPH Y., 1994. Reasoning About Knowledge and Probability. In : *J. ACM*. mars 1994. Vol. 41, n° 2, p. 340–367. DOI 10.1145/174652.174658.

FALLERY, BERNARD ET RODHAIN, FLORENCE, 2007. Quatre approches pour l'analyse de données textuelles: lexicale, linguistique, cognitive, thématique. In : *XVI^{ème} Conférence de l'Association Internationale de Management Stratégique AIMS* [en ligne]. Montréal, Canada : AIMS. 2007. p. pp 1-16. [Consulté le 23 avril 2018]. Disponible à l'adresse : <https://hal.archives-ouvertes.fr/hal-00821448>.

FALTINGS, BOI ET SCHUMACHER, MICHAEL, 2017. *L'intelligence artificielle par la pratique*. 2^{ème} édition. Lausanne : Presses polytechniques et universitaires romandes. ISBN 978-2-88915-075-5.

FARINA, NICOLAS ET LOWRY, RUTH G., 2017. The Validity of Consumer-Level Activity Monitors in Healthy Older Adults in Free-Living Conditions. In : *Journal of Aging and Physical Activity*. 22 mai 2017. p. 1-23. DOI 10.1123/japa.2016-0344.

FARMER, J. DOYNE, PACKARD, NORMAN H ET PERELSON, ALAN S, 1986. The immune system, adaptation, and machine learning. In : *Physica D: Nonlinear Phenomena*. 1 octobre 1986. Vol. 22, n° 1, p. 187-204. DOI 10.1016/0167-2789(86)90240-X.

FARNIA, TROSKAH, JAULENT, MARIE CHRISTINE, MARCHAND, GUILLAUME ET YASINI, MOBIN, 2017. A Collaborative Evaluation Framework for Biometric Connected Devices in Healthcare. In : *Studies in Health Technology and Informatics*. 2017. Vol. 245, p. 69-73.

FARNIA, TROSKAH, JAULENT, MARIE-CHRISTINE ET STEICHEN, OLIVIER, 2018. Evaluation Criteria of Noninvasive Telemonitoring for Patients With Heart Failure: Systematic Review. In : *Journal of Medical Internet Research* [en ligne]. janvier 2018. Vol. 20, n° 1. [Consulté le 6 mai 2020]. DOI 10.2196/jmir.7873. Disponible à l'adresse : <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6257336/>.

FEDULLO, PETER F. ET TAPSON, VICTOR F., 2003. Clinical practice. The evaluation of suspected pulmonary embolism. In : *The New England Journal of Medicine*. 25 septembre 2003. Vol. 349, n° 13, p. 1247-1256. DOI 10.1056/NEJMcp035442.

FERBER, JACQUES, 1997. *Les systèmes multi agents. Vers une intelligence collective*. Paris : Dunod. ISBN 978-2-7296-0665-7.

FIGER, JEAN-PAUL, 2012. L'informatique en nuage [Cloud Computing]. In : www.figer.com [en ligne]. 21 avril 2012. [Consulté le 12 septembre 2019]. Disponible à l'adresse : <https://www.figer.com/Publications/nuage.htm>.

FININ, TIM ET FRITZON, RICK, 1994. KQML--A Language and Protocol for Knowledge and Information Exchange. In : *Proceedings of the Thirteenth International Workshop on Distributed Artificial Intelligence*. Lake Quinalt, WA : s.n. 1994. p. 126-136.

FLICHY, PATRICE, 2001. *L'imaginaire d'Internet* [en ligne]. Paris : La Découverte. [Consulté le 17 avril 2020]. ISBN 978-2-7071-3537-7. Disponible à l'adresse : <https://www.cairn.info/l-imaginaire-d-internet--9782707135377.htm>.

FLORIDI, LUCIANO, 2019. Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical. In : *Philosophy & Technology*. 1 juin 2019. Vol. 32, n° 2, p. 185-193. DOI 10.1007/s13347-019-00354-x.

FLORIDI, LUCIANO, COWLS, JOSH, BELTRAMETTI, MONICA, CHATILA, RAJA, CHAZERAND, PATRICE, DIGNUM, VIRGINIA, LUETGE, CHRISTOPH, MADELIN, ROBERT, PAGALLO, UGO, ROSSI, FRANCESCA, SCHAFER, BURKHARD, VALCKE, PEGGY ET VAYENA, EFFY, 2018. AI4People—An Ethical Framework for a Good AI Society:

Opportunities, Risks, Principles, and Recommendations. In : *Minds and Machines*. 1 décembre 2018. Vol. 28, n° 4, p. 689-707. DOI 10.1007/s11023-018-9482-5.

FOUCAULT, MICHEL, 1963. *Naissance de la clinique*. 1ere ed. Vendôme : Presses universitaires de France.

FOUCAULT, MICHEL, 2015. *Naissance de la clinique*. 9e ed. Paris : Presses universitaires de France. ISBN 978-2-13-065197-0.

GALINON-MÉLÉNEC, B, ZLITNI, S ET LIÉNARD, F, 2015. *L'Homme-trace. Inscriptions corporelles et techniques*. Paris : CNRS Editions. ISBN 978-2-271-08324-1.

GALINON-MÉLÉNEC, BÉATRICE, 2011. *L'Homme trace. Perspectives anthropologiques des traces contemporaines*. Paris : CNRS Editions. ISBN 978-2-271-07139-2.

GALINON-MÉLÉNEC, BÉATRICE, 2015. À la recherche de la trace. In : *Communication & Organisation*. 14 octobre 2015. n° 47, p. 31-50.

GALINON-MÉLÉNEC, BÉATRICE, 2017. *L'homme-trace: des traces du corps au corps-trace*. S.l. : CNRS Editions. Homme Trace. ISBN 978-2-271-11417-4.

GALINON-MÉLÉNEC, BÉATRICE, 2019. Pour une ichtnologie générale à l'heure du numérique. In : LIENARD, Fabien et ZLITNI, Sami (éd.), *Communication et traces numériques*. Rouen : Presses universitaires Rouen Le Havre.

GALINON-MÉLÉNEC, BÉATRICE, 2020. *L'odyssée de la trace 1. Voyage au-delà des apparences*. London (UK) : ISTE Éditions.

GALINON-MELENEC, BEATRICE ET ZLITNI, SAMI, 2013. *Traces numériques - De la production à l'interprétation*. Paris : CNRS Editions. ISBN 978-2-271-07239-9.

GALLOIS, JEAN-BERNARD, 2016. Les pirates à l'abordage des hôpitaux. In : *OInet*. 2 novembre 2016. Vol. 852, p. p.34-36.

GANASCIA, JEAN GABRIEL, 2006. Connaissance. In : *Encyclopédie de l'informatique et des systèmes d'information*. Paris : Vuibert. p. 1165-75. ISBN 978-2-7117-4846-4.

GANASCIA, JEAN-GABRIEL, 1999. *Les sciences cognitives: un exposé pour comprendre, un essai pour réfléchir*. Paris : Flammarion. ISBN 978-2-08-035418-1.

GARDNER, HOWARD, 1997. *Les formes de l'intelligence*. Paris : Odile Jacob. ISBN 978-2-7381-9762-7.

GAUTHIER, Benoît (éd.), 2009. *Recherche sociale: de la problématique à la collecte des données*. 5th ed. Québec : Presses de l'Université du Québec. ISBN 978-2-7605-1600-7. H62 .R363 2009

GAUTHIER, GILLES, 1992. L'éthique de la communication des organisations : de la dénonciation à la reconnaissance. In : *Communication et organisation* [en ligne]. 1 novembre 1992. n° 2. [Consulté le 7 février 2017]. DOI 10.4000/communicationorganisation.1593. Disponible à l'adresse : <https://communicationorganisation.revues.org/1593?lang=en>.

GEERTS, WILLIAM H., BERGQVIST, DAVID, PINEO, GRAHAM F., HEIT, JOHN A., SAMAMA, CHARLES M., LASSEN, MICHAEL R., COLWELL, CLIFFORD W. ET AMERICAN COLLEGE OF CHEST PHYSICIANS, 2008. Prevention of venous thromboembolism: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). In : *Chest*. juin 2008. Vol. 133, n° 6 Suppl, p. 381S-453S. DOI 10.1378/chest.08-0656.

GENEL, KATIA, 2004. Le biopouvoir chez Foucault et Agamben. In : *Methodos. Savoirs et textes*. 2 mai 2004. n° 4, p. 57. DOI 10.4000/methodos.131.

GIARRATANO, JOSEPH C. ET RILEY, GARY D., 2004. *Expert Systems: Principles and Programming*. 4th Revised edition. Boston, Mass. : Course Technology Inc. ISBN 978-0-534-38447-0.

GINZBURG, CARLO, 1980. Signes, traces, pistes. In : *Le Débat*. 1980. n° 6, p. 3-44. DOI 10.3917/deba.006.0003.

GOFFMAN, ERVING, 1974. *Les Rites d'interaction*. Paris : Les Editions de Minuit. Le sens commun. ISBN 978-2-7073-0022-5.

GOODMAN, LEO A. ET KRUSKAL, WILLIAM H., 1954. Measures of Association for Cross Classifications. In : *Journal of the American Statistical Association*. 1 décembre 1954. Vol. 49, n° 268, p. 732-764. DOI 10.1080/01621459.1954.10501231.

GRADY, CHRISTINE, ECKSTEIN, LISA, BERKMAN, BEN, BROCK, DAN, COOK-DEEGAN, ROBERT, FULLERTON, STEPHANIE M., GREELY, HANK, HANSSON, MATS G., HULL, SARA, KIM, SCOTT, LO, BERNIE, PENTZ, REBECCA, RODRIGUEZ, LAURA, WEIL, CAROL, WILFOND, BENJAMIN S. ET WENDLER, DAVID, 2015. Broad Consent for Research With Biological Samples: Workshop Conclusions. In : *The American journal of bioethics: AJOB*. 2015. Vol. 15, n° 9, p. 34-42. DOI 10.1080/15265161.2015.1062162.

GRANOVETTER, MARK, 1983. The Strength of Weak Ties: A Network Theory Revisited. In : *Sociological Theory*. 1983. Vol. 1, p. 201-233. DOI 10.2307/202051.

GREENWOOD, DAVYDD J. ET LEVIN, MORTEN, 1998. Action research, science, and the co-optation of social research. In : *Studies in Cultures, Organizations and Societies*. 1 septembre 1998. Vol. 4, n° 2, p. 237-261. DOI 10.1080/10245289808523514.

GRENIER-BOLEY, BENJAMIN, 2015. Big data: conditions matérielles et limites techniques. In : *Cahier de l'Espace Ethique*. septembre 2015. n° 02, p. 23.

GRIFFIN, ABBIE ET HAUSER, JOHN R., 1993. The Voice of the Customer. In : *Marketing Science*. février 1993. Vol. 12, n° 1, p. 1–27. DOI 10.1287/mksc.12.1.1.

GUIZANI, NACHOUA, 2016. *Conception d'une architecture de services d'intelligence ambiante pour l'optimisation de la qualité de service de transmission de messages en e-santé* [en ligne]. phdthesis. Lyon : Université de Lyon. [Consulté le 1 juin 2020]. Disponible à l'adresse : <https://tel.archives-ouvertes.fr/tel-01412597>.

HARDIN, RUSSELL, 2002. *Trust and trustworthiness*. New York, NY, US : Russell Sage Foundation. Trust and trustworthiness. ISBN 978-0-87154-342-4.

HAS, 2013. *Efficience de la télémédecine : état des lieux de la littérature internationale et cadre d'évaluation* [en ligne]. Paris. HAS. Disponible à l'adresse : https://www.has-sante.fr/upload/docs/application/pdf/2013-07/efficience_tlm_vf_2013-07-18_14-48-38_743.pdf.

HEARST, MARTI A., 1998. Support Vector Machines. In : *IEEE Intelligent Systems*. 1 juillet 1998. Vol. 13, n° 4, p. 18–28. DOI 10.1109/5254.708428.

HEBB, D. O., 2005. *The Organization of Behavior: A Neuropsychological Theory*. London (UK) : Psychology Press. ISBN 978-1-135-63191-8.

HEDLUND, G. A., 1969. Endomorphisms and automorphisms of the shift dynamical system. In : *Mathematical systems theory*. 1 décembre 1969. Vol. 3, n° 4, p. 320-375. DOI 10.1007/BF01691062.

HERVÉ, Christian, STANTON-JEAN, Michèle et MAMZER, Marie-France (éd.), 2017. *La participation des patients*. Paris : Dalloz. ISBN 978-2-247-17034-0.

HINTON, GEOFFREY E., OSINDERO, SIMON ET TEH, YEE-WHYE, 2006. A Fast Learning Algorithm for Deep Belief Nets. In : *Neural Computation*. juillet 2006. Vol. 18, n° 7, p. 1527-1554. DOI 10.1162/neco.2006.18.7.1527.

HINTON, GEOFFREY E., SRIVASTAVA, NITISH, KRIZHEVSKY, ALEX, SUTSKEVER, ILYA ET SALAKHUTDINOV, RUSLAN R., 2012. Improving neural networks by preventing co-adaptation of feature detectors. In : *arXiv:1207.0580 [cs]* [en ligne]. 3 juillet 2012. [Consulté le 30 avril 2020]. Disponible à l'adresse : <http://arxiv.org/abs/1207.0580>.

HIRSCH, EMMANUEL ET MANN, JONATHAN M, 2014. Droit de l'Homme et respect de la personne. In : *Traité de bioéthique: Tome I, Fondements, principes, repères*. Toulouse : Eres. p. 249-261. ISBN 978-2-7492-1305-7.

HOERNI, BERNARD, 2000. *Ethique et déontologie médicale : Permanence et progrès*. Paris : Editions Masson. ISBN 978-2-294-00313-4.

HOERNI, BERNARD, 2002. Participation des patients à la décision en cancérologie. In : *Bulletin du Cancer*. 19 novembre 2002. Vol. 89, n° 10, p. 904-7.

HOERNI, BERNARD, 2003. L'éthique médicale: principales évolutions au fil des siècles. In : *Histoire des Sciences Médicales*. 2003. Vol. XXXVII, n° 3, p. 331-338.

HOERNI, BERNARD ET AVRIL, ANTOINE, 2009. Quel bien ? Du principe de bienveillance au principe d'autonomie. In : *Bulletin du Cancer*. 1 juillet 2009. Vol. 96, n° 7, p. 769-775. DOI 10.1684/bdc.2009.0909.

HOFMANN, B., 2009. Broadening consent--and diluting ethics? In : *Journal of Medical Ethics*. février 2009. Vol. 35, n° 2, p. 125-129. DOI 10.1136/jme.2008.024851.

HOLLAND, JOHN H., 1962. Outline for a Logical Theory of Adaptive Systems. In : *Journal of the ACM*. 1 juillet 1962. Vol. 9, n° 3, p. 297-314. DOI 10.1145/321127.321128.

HOPFIELD, J. J., 1982. Neural networks and physical systems with emergent collective computational abilities. In : *Proceedings of the National Academy of Sciences*. 1 avril 1982. Vol. 79, n° 8, p. 2554-2558. DOI 10.1073/pnas.79.8.2554.

HOUDARD-BRUNET, SOLÈNE, 2014. *Le suivi des patients greffés rénaux par télé-médecine: étude de leurs préférences individuelles par la méthode des choix discrets*. Thèse de médecine. France : Université de Nantes.

HOUTAUD, ALPHONSE D', 1999. *La santé à travers les sciences humaines et sociales: approche linguistique et sociologique*. Paris : Masson. ISBN 978-2-225-85699-0. RA418 .H68 1999

HUME, DAVID ET MILLER, EUGENE G., 1987. *Essays, Moral, Political, and Literary*. 2nd Revised edition. Indianapolis : Liberty Fund Inc. ISBN 978-0-86597-055-7.

IN PRINCIPIO, 2017. L'intelligence artificielle de 1950 à aujourd'hui. In : *L'intelligence Artificielle* [en ligne]. 2017. [Consulté le 9 avril 2020]. Disponible à l'adresse : <https://www.inprincipio.xyz/annees-1950/>.

INPRINCIPO, 2018. Systèmes multi agents et IA distribuée : Qu'est ce que c'est ? In : *L'intelligence Artificielle* [en ligne]. 2018. [Consulté le 13 juin 2018]. Disponible à l'adresse : <https://www.inprincipio.xyz/ia-distribuee/>.

INSEE, 2019. Enquête annuelle auprès des ménages sur les technologies de l'information et de la communication. In : [en ligne]. octobre 2019. [Consulté le 16 avril 2020]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/4241397>.

JARRY, ALBAN, RAVIX, VALÉRY, ETIENNE, LOIC, COLLOC, JOËL ET ZOUARHI, SAJIDA, 2018. La datasphère et ses applications numériques. In : *Vademecum sur le traitement des données numériques: Réflexions et bonnes pratiques éthiques*. Saint-Herblain : ADEL. p. 42-73.

JAURÉGUIBERRY, FRANCIS, 2003. *Les branchés du portable*. Paris : Presses Universitaires de France - PUF. ISBN 978-2-13-053670-3.

JEANNERET, YVES, 2011. Complexité de la notion de trace. De la traque au tracé. In : *L'Homme trace: Perspectives anthropologiques des traces contemporaines*. Paris : CNRS éditions. Homme Trace. p. 59-86.

JENNINGS, N. R. ET WOOLDRIDGE, M., 1998. Applications of intelligent agents. In : *Agent technology: foundations, applications, and markets*. Berlin, Heidelberg : Springer-Verlag. p. 3–28. ISBN 978-3-540-63591-8.

JENNINGS, NICHOLAS R., SYCARA, KATIA ET WOOLDRIDGE, MICHAEL, 1998. A Roadmap of Agent Research and Development. In : *Autonomous Agents and Multi-Agent Systems*. 1 janvier 1998. Vol. 1, n° 1, p. 7–38. DOI 10.1023/A:1010090405266.

JOLY, C. ET JOLY, L.-M., 2010. Éthique et déontologie médicale: droits des malades, problèmes liés au diagnostic, au respect de la personne et à la mort. In : LEHOT, Jean-Jacques et ARVIEUX, Charles-Christian (éd.), *Réanimation et urgences* [en ligne]. Paris : Springer Paris. Le point sur ... p. 475-481. [Consulté le 25 septembre 2018]. ISBN 978-2-287-99129-5. Disponible à l'adresse : https://doi.org/10.1007/978-2-287-99129-5_35.

JOUISSON-LAFFITTE, ESTÈLE, 2009. La recherche action : oubliée de la recherche dans le domaine de l'entrepreneuriat. In : *Revue de l'Entrepreneuriat*. 2009. Vol. Vol. 8, n° 1, p. 1-35.

KANT, EMMANUEL ET FUSSLER, JEAN-PIERRE, 2003. *Critique de la raison pratique*. Paris : Flammarion. ISBN 978-2-08-071090-1.

KAYE, JANE, WHITLEY, EDGAR A., LUND, DAVID, MORRISON, MICHAEL, TEARE, HARRIET ET MELHAM, KAREN, 2015. Dynamic consent: a patient interface for twenty-first century research networks. In : *European Journal of Human Genetics*. février 2015. Vol. 23, n° 2, p. 141-146. DOI 10.1038/ejhg.2014.71.

KHALDOUN, IBN, 2006. *Les prolégomènes*. Quebec : Bibliothèque Paul -Émile Boulet de l'Université. Les classiques des sciences sociales.

KITCHIN, ROB, 2014. Big Data, new epistemologies and paradigm shifts. In : *Big Data & Society*. 10 juillet 2014. Vol. 1, n° 1, p. 2053951714528481. DOI 10.1177/2053951714528481.

KOHN, LAURENCE ET CHRISTIAENS, WENDY, 2014. Les méthodes de recherches qualitatives dans la recherche en soins de santé : apports et croyances. In : *Reflets et perspectives de la vie économique*. 2014. Vol. Tome LIII, n° 4, p. 67-82.

KONDOR, DANIEL, HASHEMIAN, BEHROOZ, DE MONTJOYE, YVES-ALEXANDRE ET RATTI, CARLO, 2018. Towards matching user mobility traces in large-scale datasets. In : *IEEE Transactions on Big Data*. 2018. p. 1-1. DOI 10.1109/TBDATA.2018.2871693.

KONSTANTINIDES, STAVROS V., MEYER, GUY, BECATTINI, CECILIA, BUENO, HÉCTOR, GEERSING, GEERT-JAN, HARJOLA, VELI-PEKKA, HUISMAN, MENNO V., HUMBERT, MARC, JENNINGS, CATRIONA SIAN, JIMÉNEZ, DAVID, KUCHER, NILS, LANG, IRENE MARTHE, LANKEIT, MAREIKE, LORUSSO, ROBERTO, MAZZOLAI, LUCIA, MENEVEAU, NICOLAS, NÍ ÁINLE, FIONNUALA, PRANDONI, PAOLO, PRUSZCZYK, PIOTR, RIGHINI, MARC, TORBICKI, ADAM, VAN BELLE, ERIC, ZAMORANO, JOSÉ LUIS, GALIÉ, NAZZARENO, GIBBS, J. SIMON R., ABOYANS, VICTOR, AGENO, WALTER, AGEWALL, STEFAN, ALMEIDA, ANA G., ANDREOTTI, FELICITA, BARBATO, EMANUELE, BAUERSACHS, JOHANN, BAUMBACH, ANDREAS, BEYGUI, FARZIN, CARLSEN, JØRN, DE CARLO, MARCO, DELCROIX, MARION, DELGADO, VICTORIA, SUBIAS, PILAR ESCRIBANO, FITZSIMONS, DONNA, GAINÉ, SEAN, GOLDBERGER, SAMUEL Z., GOPALAN, DEEPA, HABIB, GILBERT, HALVORSEN, SIGRUN, JENKINS, DAVID, KATUS, HUGO A., KJELLSTRÖM, BARBRO, LAINSCAK, MITJA, LANCELLOTTI, PATRIZIO, LEE, GERALDINE, GAL, GRÉGOIRE LE, MESSAS, EMMANUEL, MORAIS, JOAO, PETERSEN, STEFFEN E., PETRONIO, ANNA SONIA, PIEPOLI, MASSIMO FRANCESCO, PRICE, SUSANNA, ROFFI, MARCO, SALVI, ALDO, SANCHEZ, OLIVIER, SHLYAKHTO, EVGENY, SIMPSON, IAIN A., STORTECKY, STEFAN, THIELMANN, MATTHIAS, NOORDEGRAAF, ANTON VONK, BECATTINI, CECILIA, BUENO, HÉCTOR, GEERSING, GEERT-JAN, HARJOLA, VELI-PEKKA, HUISMAN, MENNO V., HUMBERT, MARC, JENNINGS, CATRIONA SIAN, JIMÉNEZ, DAVID, KUCHER, NILS, LANG, IRENE MARTHE, LANKEIT, MAREIKE, LORUSSO, ROBERTO, MAZZOLAI, LUCIA, MENEVEAU, NICOLAS, NÍ ÁINLE, FIONNUALA, PRANDONI, PAOLO, PRUSZCZYK, PIOTR, RIGHINI, MARC, TORBICKI, ADAM, VANBELLE, ERIC, LUISZAMORANO, JOSÉ, WINDECKER, STEPHAN, ABOYANS, VICTOR, BAIGENT, COLIN, COLLET, JEAN-PHILIPPE, DEAN, VERONICA, DELGADO, VICTORIA, FITZSIMONS, DONNA, GALE, CHRIS P., GROBBEE, DIEDERICK, HALVORSEN, SIGRUN, HINDRICKS, GERHARD, IUNG, BERNARD, JÜNI, PETER, KATUS, HUGO A., LANDMESSER, ULF, LECLERCQ, CHRISTOPHE, LETTINO, MADDALENA, LEWIS, BASIL S., MERKELY, BELA, MUELLER, CHRISTIAN, PETERSEN, STEFFEN E., SONIA PETRONIO, ANNA, RICHTER, DIMITRIOS J., ROFFI, MARCO, SHLYAKHTO, EVGENY, SIMPSON, IAIN A., SOUSA-UVA, MIGUEL, TOUYZ, RHIAN M., HAMMOUDI, NAIMA, HAYRAPETYAN, HAMLET, MASCHERBAUER, JULIA,

IBRAHIMOV, FIRDOVSI, POLONETSKY, OLEG, LANCELLOTTI, PATRIZIO, TOKMAKOVA, MARIYA, SKORIC, BOSKO, MICHALOLIAKOS, IOANNIS, HUTYRA, MARTIN, MELLEMKJAER, SØREN, MOSTAFA, MANSOUR, REINMETS, JULIA, JÄÄSKELÄINEN, PERTTI, ANGOULVANT, DENIS, BAUERSACHS, JOHANN, GIANNAKOULAS, GEORGE, ZIMA, ENDRE, VIZZA, CARMINE DARIO, SUGRALIYEV, AKHMETZHAN, BYTYÇI, IBADETE, MACA, AIJA, EREMINIENE, EGLE, HUIJNEN, STEVE, XUEREB, ROBERT, DIACONU, NADEJDA, BULATOVIC, NEBOJSA, ASFALOU, ILYASSE, BOSEVSKI, MARIJAN, HALVORSEN, SIGRUN, SOBKOWICZ, BOŽENA, FERREIRA, DANIEL, PETRIS, ANTONIU OCTAVIAN, MOISEEVA, OLGA, ZAVATTA, MARCO, OBRADOVIC, SLOBODAN, ŠIMKOVA, IVETA, RADSEL, PETER, IBANEZ, BORJA, WIKSTRÖM, GERHARD, AUJESKY, DRAHOMIR, KAYMAZ, CIHANGIR, PARKHOMENKO, ALEXANDER ET PEPKE-ZABA, JOANNA, 2019. ESC Guidelines for the diagnosis and management of acute pulmonary embolism developed in collaboration with the European Respiratory Society (ERS)The Task Force for the diagnosis and management of acute pulmonary embolism of the European Society of Cardiology (ESC). In : *European Heart Journal*. 2019. p. 61. DOI 10.1093/eurheartj/ehz405.

LANEY, DOUGLAS, 2001. *3D Data Management: Controlling Data Volume, Velocity, and Variety* [en ligne]. Stamford. META Group. Disponible à l'adresse : <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>.

LANGRETH, ROBERT ET WALDHOLZ, MICHAEL, 1999. New era of personalized medicine: targeting drugs for each unique genetic profile. In : *The Oncologist*. 1999. Vol. 4, n° 5, p. 426-427.

LARMARANGE, JOSEPH, 2018. Régression logistique. In : *Analyse-R* [en ligne]. 13 mars 2018. [Consulté le 13 mars 2018]. Disponible à l'adresse : <http://larmarange.github.io/analyse-R/regression-logistique.html>.

LAROCHE, NICOLAS, 2016. Les bases de données NoSQL clé-valeur et la base Cassandra open source. In : *DataFactory par JEMS group* [en ligne]. 23 septembre 2016. [Consulté le 9 septembre 2019]. Disponible à l'adresse : <http://blog.jemsdatafactory.com/2016/09/23/les-bases-de-donnees-nosql-cle-valeur-et-la-base-cassandra-open-source/>.

LAULAN, Anne-Marie (éd.), 2017. *La coopération à l'ère du numérique*. Paris : L'Harmattan. Communication et civilisation. ISBN 978-2-343-12547-3.

LE BER, FLORENCE, LIEBER, JEAN ET NAPOLI, AMEDEO, 2006. Systèmes à base de connaissances. In : *Encyclopédie de l'informatique et des systèmes d'information*. Paris : Vuibert. p. 1097-1208. ISBN 978-2-7117-4846-4.

LE BLANC, BENOÎT ET ERMINE, JEAN-LOUIS, 2006. A Shannon's theory of knowledge. In : *Creating collaborative advantage through knowledge and innovation* [en ligne]. Singapour : World Scientific Publishing Company. Innovation and Knowledge Management. p. 51-68. [Consulté le 1 octobre 2018]. Disponible à l'adresse : <https://hal.archives-ouvertes.fr/hal-00470413>.

LE COZ, P. ET KACHANER, J., 2010. Justice et équité dans la décision médicale. In : *Archives de Pédiatrie*. 1 février 2010. Vol. 17, p. S32-S38. DOI 10.1016/S0929-693X(10)70006-3.

LE COZ, PIERRE, 2009. Place de la réflexion philosophique dans la décision médicale. In : *Bulletin de l'Académie Nationale de Médecine*. 1 février 2009. Vol. 193, n° 2, p. 499-510. DOI 10.1016/S0001-4079(19)32601-9.

LE CUN, Y., BOTTOU, L., BENGIO, Y. ET HAFFNER, P., 1998. Gradient-based learning applied to document recognition. In : *Proceedings of the IEEE*. novembre 1998. Vol. 86, n° 11, p. 2278-2324. DOI 10.1109/5.726791.

LE CUN, YANN, 1985. Une procedure d'apprentissage pour reseau a seuil asymmetrique. In : *Proceedings of Cognitiva 85, Paris, France*. 1985. p. 599-604.

LE GAL, G., RIGHINI, M., ROY, P.-M., MEYER, G., AUJESKY, D., PERRIER, A. ET BOUNAMEAUX, H., 2005. Differential value of risk factors and clinical signs for diagnosing pulmonary embolism according to age. In : *Journal of thrombosis and haemostasis: JTH*. novembre 2005. Vol. 3, n° 11, p. 2457-2464. DOI 10.1111/j.1538-7836.2005.01598.x.

LE GAL, GRÉGOIRE, TESTUZ, ARIANE, RIGHINI, MARC, BOUNAMEAUX, HENRI ET PERRIER, ARNAUD, 2005. Reproduction of chest pain by palpation: diagnostic accuracy in suspected pulmonary embolism. In : *BMJ (Clinical research ed.)*. 26 février 2005. Vol. 330, n° 7489, p. 452-453. DOI 10.1136/bmj.38331.602384.8F.

LE GALL, JEAN-ROGER, 2016. Bernard Hoerni. Éthique et déontologie en médecine, d'Hippocrate à nos jours, Éditions Glyphe, 2015, 290 p. In : *Bulletin de l'Académie Nationale de Médecine*. 1 avril 2016. Vol. 200, n° 4, p. 1033-1034. DOI 10.1016/S0001-4079(19)30681-8.

LE MOIGNE, JEAN LOUIS, 1999. *La modélisation des systèmes complexes* [en ligne]. Paris : Dunod. [Consulté le 25 novembre 2019]. ISBN 978-2-10-004382-8. Disponible à l'adresse : <https://www.dunod.com/sciences-humaines-et-sociales/modelisation-systemes-complexes>.

LE MONDE, 2013. Comprendre le programme « Prism ». In : *Le Monde.fr* [en ligne]. 2013. [Consulté le 25 janvier 2017]. Disponible à l'adresse : http://www.lemonde.fr/international/infographie/2013/06/11/le-programme-prism-en-une-infographie_3427774_3210.html.

LEAVITT, N., 2009. Is Cloud Computing Really Ready for Prime Time? In : *Computer*. janvier 2009. Vol. 42, n° 1, p. 15-20. DOI 10.1109/MC.2009.20.

LEBART, LUDOVIC ET SALEM, ANDRÉ, 1994. *Statistique textuelle*. Paris : Dunod. ISBN 978-2-10-002239-7.

LE COZ, PIERRE, 2009. Les principes éthiques et les émotions dans la décision médicale. In : *Médecine thérapeutique / Pédiatrie*. 1 novembre 2009. Vol. 12, n° 6, p. 383-390. DOI 10.1684/mtp.2009.0260.

LEDLEY, R. S. ET LUSTED, L. B., 1959. Reasoning foundations of medical diagnosis; symbolic logic, probability, and value theory aid our understanding of how physicians reason. In : *Science (New York, N.Y.)*. 3 juillet 1959. Vol. 130, n° 3366, p. 9-21. DOI 10.1126/science.130.3366.9.

LELEU-MERVIEL, SYLVIE, 1997. *La Conception en communication: méthodologie qualité*. Paris : Hermès. ISBN 978-2-86601-631-9.

LELEU-MERVIEL, SYLVIE, 2008. *Objectiver l'humain?* Paris : Hermès Science publications : Lavoisier. ISBN 978-2-7462-1552-8.

LELEU-MERVIEL, SYLVIE, 2013. Traces, information et construits de sens. Déploiement de la trace visuelle de la rétention indicielle à l'écriture. In : *Intellectica*. 2013. Vol. 59, n° 1, p. 65-88. DOI 10.3406/intel.2013.1086.

LELEU-MERVIEL, SYLVIE, 2017. *La traque informationnelle*. Londres : ISTE Éditions. Sciences, société et nouvelles technologies. ISBN 978-1-78405-292-8.

LELEU-MERVIEL, SYLVIE, SCHMITT, DANIEL ET USEILLE, PHILIPPE, 2018. *De l'UXD au LivXD*. London (UK) : ISTE Editions. ISBN 978-1-78405-524-0.

LELONG, ROMAIN, 2019. *Accès sémantique aux données massives et hétérogènes en santé* [en ligne]. thesis. S.l. : Normandie. [Consulté le 25 octobre 2019]. Disponible à l'adresse : <http://www.theses.fr/2019NORMR030>.

LEMBERGER, PIRMIN, BATTY, MARC, MOREL, MÉDÉRIC ET RAFFAËLLI, JEAN-LUC, 2016. *Big Data et Machine Learning - 2e éd. - Les concepts et les outils de la data science*. 2e édition. Malakoff : Dunod. ISBN 978-2-10-075463-2.

LEMOINE, MAËL, 2017. *Introduction à la philosophie des sciences médicales*. Paris : Hermann. Med phi. ISBN 978-2-7056-9384-8. R723 .L+

LÉO, MAGALI, 2016. Patient connecté et données de santé : les vrais risques. In : *I2D – Information, données & documents*. 4 octobre 2016. Vol. me 53, n° 3, p. 65-66.

LERY, LOUIS ET COLLOC, JOËL, 2008. Prise de décision dans l'éthique au quotidien - Comment décider dans le soin ? In : *Santé Décision Management*. 2008. Vol. 11, n° 1-2, p. 243-254.

LES ECHOS, LES ECHOS, 2016. Le Big data dans la santé : quelles réalités et perspectives en France ? In : *lesechos.fr* [en ligne]. 18 novembre 2016. [Consulté le 18 octobre 2017]. Disponible à l'adresse : <https://www.lesechos.fr/idees-debats/cercle/cercle-162778-le-big-data-dans-la-sante-queelles-realites-et-perspectives-en-france-2043994.php#Xtor=AD-6000>.

LESHNO, MOSHE, LIN, VLADIMIR YA., PINKUS, ALLAN ET SCHOCKEN, SHIMON, 1993. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. In : *Neural Networks*. 1 janvier 1993. Vol. 6, n° 6, p. 861-867. DOI 10.1016/S0893-6080(05)80131-5.

LIKERT, R., 1932. A technique for the measurement of attitudes. In : *Archives of Psychology*. 1932. Vol. 22 140, p. 55-55.

LLORCA, GUY, 2006. L'accord mutuel librement consenti dans la décision médicale. In : *Médecine*. 1 septembre 2006. Vol. 2, n° 7, p. 330-333.

LOMBARD, JEAN, 2009. *Ethique médicale et philosophie : L'apport de l'Antiquité*. Paris : Editions L'Harmattan. ISBN 978-2-296-09866-4.

LUCAS-GABRIELLI, VÉRONIQUE ET CHEVILLARD, GUILLAUME, 2018. « Déserts médicaux » et accessibilité aux soins : de quoi parle-t-on ? In : *médecine/sciences*. 1 juin 2018. Vol. 34, n° 6-7, p. 599-603. DOI 10.1051/medsci/20183406022.

MALAFOSSE, JEANNE BOSSI, WILLIATTE, LINA, RIAL-SEBBAG, EMMANUELLE, LE BARS, HUGUES ET GHERNAOUTI, SOLANGE, 2018. Le droit et la réglementation autour du traitement des données numériques. In : *Vademecum sur le traitement des données numériques: Réflexions et bonnes pratiques éthiques*. Saint-Herblain : ADEL. p. 74-109.

MAMDANI, E. H. ET ASSILIAN, S., 1975. An experiment in linguistic synthesis with a fuzzy logic controller. In : *International Journal of Man-Machine Studies*. 1 janvier 1975. Vol. 7, n° 1, p. 1-13. DOI 10.1016/S0020-7373(75)80002-2.

MANGEMATIN, V, THUDEROZ, C, MANGEMATIN, V ET HARRISSON, D, 1999. La confiance: un mode de coordination dont l'utilisation dépend de ses conditions de production. In : *La confiance: approche économique et sociologiques*. Boucherville, Quebec : Gaëtan Morin. p. 322.

MANYIKA, JAMES, CHUI, MICHAEL, BROWN, BRAD, BUGHIN, JACQUES, DOBBS, RICHARD, ROXBURGH, CHARLES ET BYERS, ANGELA HUNG, 2011. *Big data: The next frontier for innovation, competition, and productivity | McKinsey & Company* [en ligne]. New Jersey (USA). The McKinsey Global Institute. [Consulté le 25 janvier 2017]. Disponible à l'adresse : <http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation>.

MARCHAND, PASCAL ET RATINAUD, PIERRE, 2012. L'analyse de similitude appliquée aux corpus textuels : les primaires socialistes pour l'élection présidentielle française (septembre-octobre 2011). In : *JADT 2012*. Liège-Belgique : Lexicométrica. juin 2012. p. 687-699.

MARR, BERNARD, 2015. *Big Data: Using SMART Big Data, Analytics and Metrics To Make Better Decisions and Improve Performance*. 1 édition. Chichester, West Sussex, United Kingdom ; Hoboken, New Jersey : Wiley. ISBN 978-1-118-96583-2.

MARTAJ, DR. NADIA ET MOKHTARI, DR. MOHAND, 2010. Contrôle par logique floue. In : MARTAJ, Nadia et MOKHTARI, Mohand (éd.), *MATLAB R2009, SIMULINK et STATEFLOW pour Ingénieurs, Chercheurs et Etudiants* [en ligne]. Berlin, Heidelberg : Springer. p. 747-805. [Consulté le 14 avril 2020]. ISBN 978-3-642-11764-0. Disponible à l'adresse : https://doi.org/10.1007/978-3-642-11764-0_16.

MARTIN, OLIVIER, 2012. Analyse quantitative. In : *Sociologie, Les 100 mots de la sociologie* [en ligne]. 1 avril 2012. [Consulté le 24 mai 2019]. Disponible à l'adresse : <http://journals.openedition.org/sociologie/1204>.

MARTIN, THIERRY ET QUIVIGER, PIERRE-YVES, 2007. *Action médicale et confiance*. Besançon : Presses universitaires de Franche-Comté. Annales littéraires de l'université de Franche-Comté, 828. ISBN 978-2-84867-204-5. R727.3 .A28 2007

MASSÉ, RAYMOND ET SAINT-ARNAUD, JOCELYNE, 2003. *Ethique et santé publique : Enjeux, valeurs et normativité*. Sainte-Foy, Québec : Presses Université Laval. ISBN 978-2-7637-7970-6.

MATHIEU, FLORENCE ET HILLEN, VÉRONIQUE, 2016. *Le design thinking par la pratique: de la rencontre avec l'utilisateur à la commercialisation d'un produit innovant pour les seniors*. Paris : Eyrolles. ISBN 978-2-212-14385-0.

MATHIVET, VIRGINIE, 2014. *L'intelligence artificielle pour les développeurs concepts et implémentations en C#*. Saint-Herblain : Éditions ENI. ISBN 978-2-7460-9215-0.

MAYER, ROGER C., DAVIS, JAMES H. ET SCHOORMAN, F. DAVID, 1995. An Integrative Model of Organizational Trust. In : *The Academy of Management Review*. 1995. Vol. 20, n° 3, p. 709-734. DOI 10.2307/258792.

MCCARTHY, JOHN, 1987. Generality in artificial intelligence. In : *Communications of the ACM*. 1 décembre 1987. Vol. 30, n° 12, p. 1030–1035. DOI 10.1145/33447.33448.

MCCULLAGH, PETER ET NEDLER, JA, 1989. *Generalized Linear Models, Second Edition*. 2. London, New York : Chapman and Hall/CRC. ISBN 978-0-412-31760-6.

MCCULLOCH, WARREN S. ET PITTS, WALTER, 1943. A logical calculus of the ideas immanent in nervous activity. In : *The bulletin of mathematical biophysics*. 1 décembre 1943. Vol. 5, n° 4, p. 115-133. DOI 10.1007/BF02478259.

MEHRA, MANDEEP R, DESAI, SAPAN S, RUSCHITZKA, FRANK ET PATEL, AMIT N, 2020. Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis. In : *The Lancet* [en ligne]. 22 mai 2020. [Consulté le 30 mai 2020]. DOI 10.1016/S0140-6736(20)31180-6. Disponible à l'adresse : <http://www.sciencedirect.com/science/article/pii/S0140673620311806>.

MELL, PETER ET GRANCE, TIM, 2011. NIST Special Publication (SP) 800-145 : *The NIST Definition of Cloud Computing* [en ligne]. Maryland (USA). National Institute of Standards and Technology. [Consulté le 13 septembre 2019]. Disponible à l'adresse : <https://csrc.nist.gov/publications/detail/sp/800-145/final>.

MEMMI, DANIEL, 1990. Connexionnisme, intelligence artificielle, et modélisation cognitive. In : *Intellectica*. 1990. Vol. 9, n° 1, p. 41-79. DOI 10.3406/intel.1990.880.

MERZEAU, LOUISE, 2013a. Du signe à la trace : l'information sur mesure, From Signs to Traces - Reflections on Customized Information. In : *Hermès, La Revue*. 26 octobre 2013. n° 53, p. 21-29.

MERZEAU, LOUISE, 2013b. L'intelligence des traces. In : *Intellectica - La revue de l'Association pour la Recherche sur les sciences de la Cognition (ARCo)*. 2013. Vol. 1, n° 59, p. p.115-135.

MILL, JOHN STUART, AUDARD, CATHERINE ET THIERRY, PATRICK, 2009. *L'utilitarisme. Essai sur Bentham*. 2. Paris : Presses Universitaires de France - PUF. ISBN 978-2-13-057480-4.

MILLE, ALAIN, 2006. From case-based reasoning to traces-based reasoning. In : *Annual Reviews in Control*. octobre 2006. Vol. 30, p. 223-232.

MILLE, ALAIN, 2013a. De la trace à la connaissance à l'ère du Web. In : *Intellectica*. 2013. Vol. 59, n° 1, p. 7-28. DOI 10.3406/intel.2013.1083.

MILLE, ALAIN, 2013b. Traces numériques et construction de sens. In : GALINON-MELENEC, Béatrice et ZLITNI, Sami (éd.), *Traces numériques : De la production à*

l'interprétation [en ligne]. Paris : CNRS Éditions. Sociologie. p. 111-128. [Consulté le 22 juillet 2019]. ISBN 978-2-271-13013-6. Disponible à l'adresse : <http://books.openedition.org/editionscnrs/21750>.

MINSKY, MARVIN, 1988. *Society Of Mind*. S.l. : Simon and Schuster. ISBN 978-0-671-65713-0.

MINSKY, MARVIN, 1997. *La société de l'esprit*. 1re. Paris : InterEditions. ISBN 978-2-7296-0188-1.

MIRSKY, YISROEL, MAHLER, TOM, SHELEF, ILAN ET ELOVICI, YUVAL, 2019. CT-GAN: Malicious Tampering of 3D Medical Imagery using Deep Learning. In : *USENIX Security 2019* [en ligne]. Santa Clara, CA, USA : USENIX. 11 janvier 2019. p. 19. [Consulté le 6 septembre 2019]. Disponible à l'adresse : <http://arxiv.org/abs/1901.03597>.

MONINO, JEAN-LOUIS, 2013. L'information au cœur de l'intelligence économique stratégique. In : *Marché et organisations*. 17 juillet 2013. n° 18, p. 25-39.

MONINO, JEAN-LOUIS ET SEDKAOUI, SORAYA, 2016. *Big Data, Open Data et valorisation des données*. London : ISTE Éditions. ISBN 978-1-78405-122-8.

MONSEIGNE, ANNICK, 2013. Recension de Galinon-Méléne B. Des signes-traces à l'Homme-trace. Perspectives anthropologiques, in « De la trace à la connaissance à l'ère du Web » (Mille, dir.) *Intellectica*, 2013/1, n° 59. In : *Communication et organisation*. 1 décembre 2013. n° 44, p. 229-232.

MORANGE, MICHEL, 2003. *Histoire de la biologie moléculaire*. Paris : La Découverte. ISBN 978-2-7071-4057-9.

MORIN, EDGAR, 2004. *Ethique*. Paris : Seuil. La méthode, 6. ISBN 978-2-02-078638-6. BJ12 .M67 2004

MORIN, EDGAR, 2015. *Introduction à la pensée complexe*. Paris : Le Seuil. ISBN 978-2-02-066837-8.

MUCCHIELLI, ALEX, 1996. Pour des recherches en communication. In : *Communication et organisation* [en ligne]. 1 novembre 1996. n° 10. [Consulté le 27 février 2018].

DOI 10.4000/communicationorganisation.1877. Disponible à l'adresse : <http://journals.openedition.org/communicationorganisation/1877>.

MUGUR-SCHÄCHTER, MIOARA, 2006. *Sur le tissage des connaissances*. Paris : Hermes Science Publications. ISBN 978-2-7462-1438-5.

MUKE ZIHISIRE, MODESTE, 2011. *La recherche en sciences sociales et humaines: guide pratique, méthodologie et cas concrets*. Paris : L'Harmattan. ISBN 978-2-296-56175-5.

NARAYANAN, ARVIND ET SHMATIKOV, VITALY, 2008. Robust De-anonymization of Large Sparse Datasets. In : *Proceedings of the 2008 IEEE Symposium on Security and Privacy* [en ligne]. USA : IEEE Computer Society. 18 mai 2008. p.111–125. [Consulté le 17 avril 2020]. Disponible à l'adresse : <https://doi.org/10.1109/SP.2008.33>.

NELSON, NICOLE C., KEATING, PETER ET CAMBROSIO, ALBERTO, 2013. On being “actionable”: clinical sequencing and the emerging contours of a regime of genomic medicine in oncology. In : *New Genetics and Society*. 1 décembre 2013. Vol. 32, n° 4, p. 405-428. DOI 10.1080/14636778.2013.852010.

NONAKA, IKUJIRO, 1994. A Dynamic Theory of Organizational Knowledge Creation. In : *Organization Science*. 1994. Vol. 5, n° 1, p. 14-37.

NONAKA, IKUJIRO, TOYAMA, RYOKO ET KONNO, NOBORU, 2000. SECI, Ba and Leadership: a Unified Model of Dynamic Knowledge Creation. In : *Long Range Planning*. 1 février 2000. Vol. 33, n° 1, p. 5-34. DOI 10.1016/S0024-6301(99)00115-6.

O'NEIL, CATHY ET SCHUTT, RACHEL, 2013. *Doing Data Science: Straight Talk from the Frontline*. 1er. Californie (USA) : O'Reilly Media.

ORDRE NATIONAL DES MÉDECINS, 2014. *Vadémecum télémédecine*. Paris : Ordre National des médecins.

ORLÉAN, ANDRÉ, 2000. La théorie économique de la confiance et ses limites. In : *La confiance en question*. Cahiers de Socio-Économie, intitulés. Paris : L'Harmattan. Logiques sociales. p. 59-77.

ORR, KRYSSTN, HOWE, HOLLY S., OMRAN, JANINE, SMITH, KRISTINA A., PALMATEER, TESS M., MA, ALVIN E. ET FAULKNER, GUY, 2015. Validity of smartphone pedometer

applications. In : *BMC Research Notes* [en ligne]. 30 novembre 2015. Vol. 8, n° 733. [Consulté le 2 novembre 2017]. DOI 10.1186/s13104-015-1705-8. Disponible à l'adresse : <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4666074/>.

PAILLÉ, PIERRE ET MUCCHIELLI, ALEX, 2016. *L'analyse qualitative en sciences humaines et sociales*. Paris : Armand Colin. ISBN 978-2-200-61470-6.

PAQUET, PHILIPPE, 2006. *De l'information à la connaissance*. Écully, France : IAE Orléans.

PARK, SANGCHUL, CHOI, GINA JEEHYUN ET KO, HAKSOO, 2020. Information Technology–Based Tracing Strategy in Response to COVID-19 in South Korea—Privacy Controversies. In : *JAMA* [en ligne]. 23 avril 2020. [Consulté le 26 avril 2020]. DOI 10.1001/jama.2020.6602. Disponible à l'adresse : <https://jamanetwork.com/journals/jama/fullarticle/2765252>.

PAULINE, NG C, SARAH, S. MURRAY, SAMUEL, LEVY ET CRAIG, J VENTER, 2009. An agenda for personalized medicine. In : *Nature*. octobre 2009. Vol. 461, n° 7265, p. 724-726. DOI 10.1038/461724a.

PAYEN, MARIE CHRISTINE, 2012. L'éthique au quotidien. In : *Kairos*. 2012. Vol. 32, p. 15-18.

PERRIAULT, JACQUES, 2003. *Éducation et nouvelles technologies*. Paris : Nathan université. ISBN 978-2-09-191164-9.

PERRIN, SOPHIE CLÉMENT, 2016. *Quels sites internet utilisent les médecins généralistes en consultation ? Enquête de pratique auprès de 100 médecins généralistes libéraux installés en Gironde* [en ligne]. Thèse de médecine. Bordeaux : Université de Bordeaux. [Consulté le 12 novembre 2019]. Disponible à l'adresse : <https://dumas.ccsd.cnrs.fr/dumas-01278965>.

PETTIT, JOHN D., VAUGHT, BOBBY ET PULLEY, KATHY J., 1990. The Role of Communication in Organizations: Ethical Considerations. In : *The Journal of Business Communication* (1973). 1 juin 1990. Vol. 27, n° 3, p. 233-249. DOI 10.1177/002194369002700302.

PIAGET, JEAN, 1998. *La construction du réel chez l'enfant*. Lausanne : Delachaux et Niestlé. ISBN 978-2-603-00741-9.

POLANYI, MICHAEL ET SEN, AMARTYA, 2009. *The Tacit Dimension*. Revised ed. Chicago ; London : University of Chicago Press. ISBN 978-0-226-67298-4.

POLITI, MARY C., WOLIN, KATHLEEN Y. ET LÉGARÉ, FRANCE, 2013. Implementing clinical practice guidelines about health promotion and disease prevention through shared decision making. In : *Journal of General Internal Medicine*. juin 2013. Vol. 28, n° 6, p. 838-844. DOI 10.1007/s11606-012-2321-0.

PROULX, SERGE, 2004. *La révolution internet en question*. Montréal : Québec Amérique. ISBN 978-2-7644-0270-2.

PULIDO, TOMÁS, ARANDA, ALBERTO, ZEVALLOS, MARCO ANTONIO, BAUTISTA, EDGAR, MARTÍNEZ-GUERRA, MARIA LUISA, SANTOS, LUIS EFRÉN ET SANDOVAL, JULIO, 2006. Pulmonary embolism as a cause of death in patients with heart disease: an autopsy study. In : *Chest*. mai 2006. Vol. 129, n° 5, p. 1282-1287. DOI 10.1378/chest.129.5.1282.

QUINCHE, F., 2005. Respect du droit et de l'autonomie ou bienfaisance ? In : *Éthique & Santé*. 1 février 2005. Vol. 2, n° 1, p. 41-45. DOI 10.1016/S1765-4629(05)80507-0.

R DEVELOPMENT CORE TEAM, 2008. *R: A Language and Environment for Statistical Computing* [en ligne]. Vienna, Austria : R Foundation for Statistical Computing. ISBN 3-900051-07-0. Disponible à l'adresse : <http://www.R-project.org>.

RAMEIX, SUZANNE, 1998. *Fondements Philosophiques de l'Éthique Médicale*. Paris : Ellipses Marketing. ISBN 978-2-7298-9641-6.

RAMIREZ, EDITH, 2014. Big Data: A Tool for Inclusion or Exclusion? Opening Remarks of Chairwoman Edith Ramirez. In : *FTC Big Data Workshop* [en ligne]. Washington DC (USA). 15 septembre 2014. [Consulté le 24 avril 2018]. Disponible à l'adresse : <https://www.ftc.gov/public-statements/2014/09/big-data-tool-inclusion-or-exclusion-opening-remarks-chairwoman-edith>.

RAO, AS ET GEORGEFF, MP, 1991. Modeling rational agents within a BDI-architecture. In : *Principles of Knowledge Representation and Reasoning: Proceedings of the Second*

International Conference (KR91). Cambridge, Massachusetts : Morgan Kaufmann Publishers. p. 317-338. ISBN 978-1-55860-165-9.

REINERT, PAR MAX, 1987. Classification Descendante Hierarchique et Analyse Lexicale par Contexte - Application au Corpus des Poesies D'A. Rihbaud. In : *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*. 1 janvier 1987. Vol. 13, n° 1, p. 53-90. DOI 10.1177/075910638701300107.

RENNARD, JEAN-PHILIPPE, 2002. *Vie artificielle: où la biologie rencontre l'informatique : illustré avec Java*. Paris : Vuibert informatique. ISBN 978-2-7117-8694-7.

REYNOLDS, CRAIG W., 1987. Flocks, herds and schools: A distributed behavioral model. In : *Proceedings of the 14th annual conference on Computer graphics and interactive techniques* [en ligne]. New York, NY, USA : Association for Computing Machinery. 1 août 1987. p. 25–34. [Consulté le 29 avril 2020]. Disponible à l'adresse : <https://doi.org/10.1145/37401.37406>.

RICOEUR, PAUL, 2000a. De la morale à l'éthique et aux éthiques. In : *Divinatio*. 2000. n° 11, p. 31-42.

RICOEUR, PAUL, 2000b. *La Mémoire, l'histoire, l'oubli*. Paris : Editions du Seuil. ISBN 978-2-02-034917-8.

RICOEUR, PAUL, 2015. *Soi-même comme un autre*. Paris : Le Seuil. ISBN 978-2-02-114486-4.

RIEFFEL, RÉMY, 2014. *Révolution numérique, révolution culturelle?* Paris : Gallimard. ISBN 978-2-07-045172-2.

RISH, IRINA, 2001. An empirical study of the naive Bayes classifier. In : *IJCAI 2001 workshop on empirical methods in artificial intelligence*. New York, NY, USA : IBM. août 2001. p. 41-46.

ROBINSON, J. A., 1965. A Machine-Oriented Logic Based on the Resolution Principle. In : *J. ACM*. janvier 1965. Vol. 12, n° 1, p. 23–41. DOI 10.1145/321250.321253.

ROLFE, ALIX, CASH-GIBSON, LUCINDA, CAR, JOSIP, SHEIKH, AZIZ ET MCKINSTRY, BRIAN, 2014. Interventions for improving patients' trust in doctors and groups of doctors.

In : *The Cochrane Database of Systematic Reviews*. 4 mars 2014. n° 3, p. CD004134. DOI 10.1002/14651858.CD004134.pub3.

ROSENBLATT, F., 1958. The perceptron: a probabilistic model for information storage and organization in the brain. In : *Psychological Review*. novembre 1958. Vol. 65, n° 6, p. 386-408. DOI 10.1037/h0042519.

ROUVROY, ANTOINETTE ET BERNS, THOMAS, 2013. Gouvernamentalité algorithmique et perspectives d'émancipation, Faced with algorithmic governmentality. In : *Réseaux*. 2 mai 2013. n° 177, p. 163-196.

ROY, PIERRE-MARIE, COLOMBET, ISABELLE, DURIEUX, PIERRE, CHATELLIER, GILLES, SORS, HERVÉ ET MEYER, GUY, 2005. Systematic review and meta-analysis of strategies for the diagnosis of suspected pulmonary embolism. In : *BMJ (Clinical research ed.)*. 30 juillet 2005. Vol. 331, n° 7511, p. 259. DOI 10.1136/bmj.331.7511.259.

RUMELHART, D. E., HINTON, G. E. ET WILLIAMS, R. J., 1986. Learning internal representations by error propagation. In : *Parallel distributed processing: explorations in the microstructure of cognition, vol. 1: foundations*. Cambridge, MA, USA : MIT Press. p. 318–362. ISBN 978-0-262-68053-0.

RUSSELL, STUART J ET NORVIG, PETER, 2010. *Intelligence artificielle: avec plus de 500 exercices*. 3ème édition. Paris : Pearson Education. ISBN 978-2-7440-7150-8.

RUSSELL, STUART ET NORVIG, PETER, 2016. *Artificial Intelligence: A Modern Approach, Global Edition*. 3. Boston Columbus Indianapolis New York San Francisco : Pearson. ISBN 978-1-292-15396-4.

SALAÜN, JEAN-MICHEL, 2007. La redocumentarisation, un défi pour les sciences de l'information. In : *Études de communication. langages, information, médiations*. 1 octobre 2007. n° 30, p. 13-23. DOI 10.4000/edc.428.

SALGUES, BRUNO, 2016. *Industrialisation de la santé*. ISTE éditions. London : ISTE éditions. ISBN 978-1-78405-131-0.

SAPORTA, GILBERT, 2011. *Probabilités, analyse des données et statistique*. 3e. Paris : Editions TECHNIP. ISBN 978-2-7108-0814-5.

SCHUTZ, FRÉDÉRIC, 2018. *Algorithme prédictif*: « Il y a une gradation dans la fiabilité de [en ligne]. Santé personnalisée et société. 2018. S.l. : s.n. [Consulté le 11 janvier 2019]. Disponible à l'adresse : <https://www.santeperso.ch/Pour-comprendre/Algorithme-predictif-Il-y-a-une-gradation-dans-la-fiabilite-des-resultats>.

SCHWARTZ, DANIEL, 1993. *Méthodes statistiques à l'usage des médecins et des biologistes*. Paris : Médecine Sciences Publications. ISBN 978-2-257-12326-8.

SCOTT, JAMES G. ET SUN, LIANG, 2013. Expectation-maximization for logistic regression. In : *arXiv preprint:1306.0040 [math, stat]* [en ligne]. 31 mai 2013. [Consulté le 30 avril 2020]. Disponible à l'adresse : <http://arxiv.org/abs/1306.0040>.

SEARLE, JOHN R., 1980. Minds, brains, and programs. In : *Behavioral and Brain Sciences*. septembre 1980. Vol. 3, n° 3, p. 417-424. DOI 10.1017/S0140525X00005756.

SEGHROUCHNI, AMAL EL FARAH, 2006. Systèmes multi-agents. In : *Encyclopédie de l'informatique et des systèmes d'information*. Paris : Vuibert. p. 1232-42. ISBN 978-2-7117-4846-4.

SEROUSSI, BRIGITTE, SAUQUET, DOMINIQUE, FALCOFF, HECTOR, JULIEN, JACQUES ET BOUAUD, JACQUES, 2011. Formalisation de l'attitude des médecins vis à vis des propositions d'un système d'aide à la décision : évaluation de l' ' e-iatrogénie ' sur un cas d'hypertension avec ASTI mode guidé. In : *Actes de IC2011* [en ligne]. Chambéry (France) : IC 2011. 16 mai 2011. p. 673-688. [Consulté le 18 février 2014]. Disponible à l'adresse : <http://hal.archives-ouvertes.fr/hal-00746731>.

SERRES, ALEXANDRE, 2012. Problématiques de la trace à l'heure du numérique. In : *Sens-Dessous*. 2012. n° 10, p. 84-94. DOI 10.3917/sdes.010.0084.

SHEN, Y, JACQUET-ANDRIEU, A ET COLLOC, J, 2012. Un système multi-agents d'aide à la décision clinique fondé sur les ontologies,. In : *Applications Médicales de l'Informatique : Nouvelles Approches*. Madhia, Tunisie : AMINA. 2012. p. 136-145.

SHEN, YING, COLLOC, JOËL, JACQUET-ANDRIEU, ARMELLE ET LEI, KAI, 2015. Emerging medical informatics with case-based reasoning for aiding clinical decision in multi-agent system. In : *Journal of Biomedical Informatics*. août 2015. Vol. 56, p. 307-317. DOI 10.1016/j.jbi.2015.06.012.

SHEN, YING, YUAN, KAIQI, CHEN, DAOYUAN, COLLOC, JOËL, YANG, MIN, LI, YALIANG ET LEI, KAI, 2018. An ontology-driven clinical decision support system (IDDAP) for infectious disease diagnosis and antibiotic prescription. In : *Artificial Intelligence in Medicine*. 1 mars 2018. Vol. 86, p. 20-32. DOI 10.1016/j.artmed.2018.01.003.

SICARD, DIDIER, 2017. *L'éthique médicale et la bioéthique* [en ligne]. 5e éd. Paris : Presses Universitaires de France - PUF. [Consulté le 11 mai 2020]. Que sais-je? ISBN 978-2-13-079970-2. Disponible à l'adresse : <https://www.cairn.info/l-ethique-medicale-et-la-bioethique--9782130799702.htm>.

SIMMENAUER, BENJAMIN, 2014. Qu'est ce qu'une valeur morale? In : *Traité de bioéthique: Tome I, Fondements, principes, repères*. Toulouse : Eres. p. 121-131. ISBN 978-2-7492-1305-7.

SIMMON, ERIC, 2018. NIST SP 500-322 : *Evaluation of cloud computing services based on NIST SP 800-145* [en ligne]. Gaithersburg, MD. National Institute of Standards and Technology. [Consulté le 13 septembre 2019]. Disponible à l'adresse : <http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.500-322.pdf>.

SIMON, ÉRIC, 2007. La confiance dans tous ses états. In : *Revue française de gestion*. 1 juillet 2007. n° 175, p. 83-94. DOI 10.3166/rfg.175.83-94.

SPATZ, HANNS-CHRISTOF, 1996. Hebb's concept of synaptic plasticity and neuronal cell assemblies. In : *Behavioural Brain Research*. 1 juin 1996. Vol. 78, n° 1, p. 3-7. DOI 10.1016/0166-4328(95)00221-9.

STEIN, PAUL D., HULL, RUSSELL D., PATEL, KALPESH C., OLSON, RONALD E., GHALI, WILLIAM A., BRANT, ROLLIN, BIEL, RITA K., BHARADIA, VINAY ET KALRA, NEERAJ K., 2004. D-dimer for the exclusion of acute venous thrombosis and pulmonary embolism: a systematic review. In : *Annals of Internal Medicine*. 20 avril 2004. Vol. 140, n° 8, p. 589-602.

STÉPHANE, GANASSALI ET JEAN, MOSCAROLA, 2004. Protocoles d'enquête et efficacité des sondages par Internet. In : *Décisions marketing (Bruxelles)*. 2004. n° 33, p. 63-75.

STOEKLÉ, HENRI-CORTO, FORSTER, NINON, CHARLIER, PHILIPPE, BLOCH, OUDY C., HERVÉ, CHRISTIAN, TURRINI, MAURO ET VOGT, GUILLAUME, 2018. Le partage des

données génétiques : un nouveau capital. In : *médecine/sciences*. 1 août 2018. Vol. 34, n° 8-9, p. 735-739. DOI 10.1051/medsci/20183408022.

STRATHERN, MARILYN, 1997. 'Improving ratings': audit in the British University system. In : *European Review*. juillet 1997. Vol. 5, n° 3, p. 305-321. DOI 10.1002/(SICI)1234-981X(199707)5:3<305::AID-EURO184>3.0.CO;2-4.

SUGENO, M., 1993. Fuzzy measures and fuzzy integrals- a survey. In : DUBOIS, Didier, PRADE, Henri et YAGER, Ronald R. (éd.), *Readings in Fuzzy Sets for Intelligent Systems* [en ligne]. S.l. : Morgan Kaufmann. p. 251-257. [Consulté le 5 novembre 2019]. ISBN 978-1-4832-1450-4. Disponible à l'adresse : <http://www.sciencedirect.com/science/article/pii/B9781483214504500274>.

TAMBOURIN, PIERRE, 2005. Les grands instruments de la biologie moléculaire, prémices de la médecine de demain. In : *La revue pour l'histoire du CNRS* [en ligne]. 3 mai 2005. n° 12. [Consulté le 15 novembre 2019]. DOI 10.4000/histoire-cnrs.1308. Disponible à l'adresse : <http://journals.openedition.org/histoire-cnrs/1308>.

TANGHE, HÉLÈNE, DAOUD, SANAA AIT, GIBERT, PAUL-OLIVIER, SAHRI, SOROR, BENBERNOU, SALIMA, TOTEL, JÉRÔME, BERTIN, LOÏC, DIACONU, CRISTINEL, GHORBEL, ITEBEDDINE, GELDWERTH-FENIGER, DANIELLE, ABIDI, LEILA, CÉRIN, CHRISTOPHE, WERLE, PHILIPPE ET LAFAILLE, MARIE, 2016. *approches contemporaines en hébergement et gestion de données*. Paris : Digital et Ethics.

TANGHE, HÉLÈNE ET GIBERT, PAUL-OLIVIER, 2017. L'enjeu de l'anonymisation à l'heure du big data. In : *Revue française des affaires sociales*. 2017. n° 4, p. 79-93.

TAPSON, VICTOR F., 2008. Acute Pulmonary Embolism. In : *New England Journal of Medicine*. 6 mars 2008. Vol. 358, n° 10, p. 1037-1052. DOI 10.1056/NEJMra072753.

TARDIEU, HUBERT, ROCHFELD, ARNOLD ET COLLETTI, RENÉ, 1994. *La méthode Merise*. Paris : Éditions d'Organisation. ISBN 978-2-7081-1699-3.

THÉVENOT, LAURENT, 2006. *L'action au pluriel* [en ligne]. Paris : La Découverte. [Consulté le 2 juin 2020]. ISBN 978-2-7071-4640-3. Disponible à l'adresse : <https://www.cairn.info/l-action-au-pluriel--9782707146403.htm>.

THOMAS-VASLIN, VÉRONIQUE, 2017. Le rôle des traces dans le système immunitaire : des anticorps au corps. In : B, Galinon-Méléneec (éd.), *L'Homme-trace: Des traces du corps au corps-trace* [en ligne]. S.l. : CNRS édition. L'Homme-trace. p. 255-294. [Consulté le 31 octobre 2018]. Disponible à l'adresse : <https://hal.archives-ouvertes.fr/hal-01565303>.

TIMMIS, JON, NEAL, MARK ET HUNT, JOHN, 2000. An artificial immune system for data analysis. In : *Biosystems*. février 2000. Vol. 55, n° 1-3, p. 143-150. DOI 10.1016/S0303-2647(99)00092-1.

TISSEAU, GÉRARD, 1996. *Intelligence artificielle: problèmes et méthodes*. 1. éd. Paris : Presses Univ. de France. Psychologie et sciences de la pensée. ISBN 978-2-13-047429-6.

TOPOL, ERIC J., 2014. Individualized medicine from prewomb to tomb. In : *Cell*. 27 mars 2014. Vol. 157, n° 1, p. 241-253. DOI 10.1016/j.cell.2014.02.012.

TORRIS, GEORGES, 2018. *Diagnostic* [en ligne]. Paris : Encyclopædia Universalis. [Consulté le 13 septembre 2018]. Disponible à l'adresse : <http://www.universalis-edu.com/encyclopedie/diagnostic/>.

TOUMERT, T., 2017. Entretien ou questionnaire : quelle méthode de collecte de données pour son mémoire ? In : *art, langage, apprentissage* [en ligne]. 2017. [Consulté le 12 mars 2018]. Disponible à l'adresse : <https://arlap.hypotheses.org/8170>.

TOURNEUR-BAGOT, O., 2019. Éthique et téléconsultation, des grands principes à une approche humaniste ancrée dans la pratique. In : *Éthique & Santé*. 1 septembre 2019. Vol. 16, n° 3, p. 120-124. DOI 10.1016/j.etiqe.2019.07.001.

TRAN, VIET-THI, RIVEROS, CAROLINA ET RAVAUD, PHILIPPE, 2019. Patients' views of wearable devices and AI in healthcare: findings from the ComPaRe e-cohort. In : *npj Digital Medicine*. 14 juin 2019. Vol. 2, n° 1, p. 1-8. DOI 10.1038/s41746-019-0132-y.

TUKEY, JOHN W., 1980. We Need Both Exploratory and Confirmatory. In : *The American Statistician*. 1 février 1980. Vol. 34, n° 1, p. 23-25. DOI 10.1080/00031305.1980.10482706.

TURING, AM, 1950. Computing machinery and intelligence. In : *Mind*. 1 octobre 1950. Vol. LIX, n° 236, p. 433-460. DOI 10.1093/mind/LIX.236.433.

TURNER, D. A., 2013. Some History of Functional Programming Languages. In : LOIDL, Hans-Wolfgang et PEÑA, Ricardo (éd.), *Trends in Functional Programming*. S.l. : Springer Berlin Heidelberg. 2013. p. 1-20.

VAN CAMPENHOUDT, LUC ET QUIVY, RAYMOND, 2011. *Manuel de recherche en sciences sociales*. 4e éd. entièrement revue et augmentée. Paris : Dunod. ISBN 978-2-10-056301-2.

VANNIEUWENHUYZE, AURÉLIEN, 2019. *Intelligence artificielle vulgarisée - Le Machine Learning et le Deep Learning par la pratique*. Paris : Editions ENI. ISBN 978-2-409-02073-5.

VAQUERO, LUIS M., RODERO-MERINO, LUIS, CACERES, JUAN ET LINDNER, MAIK, 2008. A break in the clouds: towards a cloud definition. In : *ACM SIGCOMM Computer Communication Review*. 31 décembre 2008. Vol. 39, n° 1, p. 50-55. DOI 10.1145/1496091.1496100.

VENKATESH, VISWANATH ET DAVIS, FRED D., 2000. A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. In : *Management Science*. 1 février 2000. Vol. 46, n° 2, p. 186-204. DOI 10.1287/mnsc.46.2.186.11926.

VERGÈS, P ET BOURICHE, B, 2001. *L'analyse des données par les graphes de similitude* [en ligne]. Auxerre : Sciences Humaines. Disponible à l'adresse : <https://www.scienceshumaines.com/textesInedits/Bouriche.pdf>.

VICTORRI, BERNARD, 2008. Le connexionnisme. In : *Traité de neuropsychologie clinique* [en ligne]. 1ere. Paris : De Boeck Supérieur. Neurosciences & cognition. p. 53-64. [Consulté le 23 mars 2020]. ISBN 978-2-8041-5678-7. Disponible à l'adresse : <https://www.cairn.info/traite-de-neuropsychologie-clinique--9782804156787-page-53.htm>.

VILLANI, CÉDRIC, SCHOENAUER, MARC, BONNET, YANN, BERTHET, CHARLY, CORNUT, ANNE-CHARLOTTE, LEVIN, FRANÇOIS ET RONDEPIERRE, BERTRAND, 2018. *Donner un sens à l'intelligence artificielle*. Mission parlementaire. Paris. Premier Ministère.

VON NEUMANN, JOHN ET BURKS, ARTHUR W, 1966. *Theory of Self-Reproducing Automata*. Urbana, London : University of Illinois Press.

WAIBEL, A., HANAZAWA, T., HINTON, G., SHIKANO, K. ET LANG, K.J., 1989. Phoneme recognition using time-delay neural networks. In : *IEEE Transactions on Acoustics, Speech, and Signal Processing*. mars 1989. Vol. 37, n° 3, p. 328-339. DOI 10.1109/29.21701.

WARNER, HOMER R., TORONTO, ALAN F. ET VEASY, L. GEORGE, 1964. Experience with Baye's Theorem for Computer Diagnosis of Congenital Heart Disease*. In : *Annals of the New York Academy of Sciences*. 1964. Vol. 115, n° 2, p. 558-567. DOI 10.1111/j.1749-6632.1964.tb50648.x.

WATSON, JAMES, 2020. An open letter to Mehra et al and The Lancet. In : [en ligne]. 28 mai 2020. [Consulté le 31 mai 2020]. DOI 10.5281/zenodo.3862789. Disponible à l'adresse : <https://zenodo.org/record/3862789>.

WEIL, ÉRIC, 2018. *Morale* [en ligne]. Paris : Encyclopædia Universalis. [Consulté le 25 septembre 2018]. Disponible à l'adresse : <http://www.universalis-edu.com/encyclopedie/morale/>.

WEISS, GERHARD, 1995. Adaptation and learning in multi-agent systems: Some remarks and a bibliography. In : *Adaption and Learning in Multi-Agent Systems* [en ligne]. Berlin : Springer, Berlin, Heidelberg. Lecture Notes in Computer Science. p. 1-21. [Consulté le 13 juin 2018]. ISBN 978-3-540-60923-0. Disponible à l'adresse : http://link.springer.com/chapter/10.1007/3-540-60923-7_16.

WHO, 2016. WHO | eHealth. In : *WHO* [en ligne]. 2016. [Consulté le 25 janvier 2017]. Disponible à l'adresse : <http://www.who.int/topics/ehealth/en/>.

WILLIAMS, JOHN R., 2015. *Manuel d'Éthique Médicale*. 3. Ferney-Voltaire : Association Médicale Mondiale.

WILLIATTE-PELLITTERI, L., 2013. Télémedecine et responsabilités juridiques. In : *European Research in Telemedicine / La Recherche Européenne en Télémedecine*. mars 2013. Vol. 2, n° 1, p. 17-22. DOI 10.1016/j.eurtel.2013.01.003.

WOLFRAM, STEPHEN, 1984. Universality and complexity in cellular automata. In : *Physica D: Nonlinear Phenomena*. 1 janvier 1984. Vol. 10, n° 1, p. 1-35. DOI 10.1016/0167-2789(84)90245-8.

WOLTON, DOMINIQUE, 2009. *Informers n'est pas communiquer*. Paris : CNRS éditions. ISBN 978-2-271-06820-0. HM258 .W64 2009

YAMEOGO, RELWENDÉ ARISTIDE ET COLLOC, JOËL, 2016. Système multi-agents d'aide à la décision clinique, appliqué à la prise en charge de l'embolie pulmonaire. In : *Applications Médicales de l'Informatique : Nouvelles Approches*. Monastir, Tunisie : AMINA. 2016. p. 92-98.

YOUSEFF, LAMIA, BUTRICO, MARIA ET SILVA, DELSON DA, 2008. Toward a Unified Ontology of Cloud Computing. In : *2008 Grid Computing Environments Workshop*. 2008. p. 1-10. DOI 10.1109/gce.2008.4738443.

ZADEH, L. A., 1965. Fuzzy sets. In : *Information and Control*. 1 juin 1965. Vol. 8, n° 3, p. 338-353. DOI 10.1016/S0019-9958(65)90241-X.

ZADEH, LOTFI A., 2008. Is there a need for fuzzy logic? In : *Information Sciences*. 1 juillet 2008. Vol. 178, n° 13, p. 2751-2779. DOI 10.1016/j.ins.2008.02.012.

ZAHEER, AKBAR, MCEVILY, BILL ET PERRONE, VINCENZO, 1998. Does Trust Matter? Exploring the Effects of Interorganizational and Interpersonal Trust on Performance. In : *Organization Science*. 1 avril 1998. Vol. 9, n° 2, p. 141-159. DOI 10.1287/orsc.9.2.141.

ZUCKER, LYNNE G., 1986. Production of trust: Institutional sources of economic structure, 1840–1920. In : *Research in Organizational Behavior*. 1986. Vol. 8, p. 53-111.

Annexes

Annexe 1 : listes des publications et des communications

Articles dans des conférences internationales avec comité de lecture :

- **Yameogo, RA**, Colloc, J. Système multi-agents d'aide à la décision clinique, appliqué à la prise en charge de l'embolie pulmonaire AMINA'2016 "Applications Médicales de l'Informatique : Nouvelles Approches", Bedoui, M. H. (Ed.) Monastir, Tunisie, ISBN : 978-9938-9531-2-1, 2016, 92-98
- Colloc, J. ; **Yameogo, RA** ; Summons, P. ; Shen, Y. ; Park, M. & Aronson, J. E. EPICE an Emotion Fuzzy Vectorial Space for Time Modeling in Medical Decision International Conference on Internet of Things and Machine Learning Proceedings IML'2017, 2017 ACM doi : doi.acm.org/10.1145/3109761.3158409

Articles dans des revues :

- **Yameogo, RA**, Colloc, J. A temporal fuzzy vector space indicator for real time clinical follow-up of pulmonary embolism in Artificial Intelligence In Medecine (AIIM) (en cours de soumission).

Chapitre d'ouvrage

- **Yameogo RA**. Enjeux de l'interprétation des traces dans le domaine de la santé. Tome 5 de la série L'Homme-trace (Béatrice Galinon-Méléne dir.). CNRS Editions 2019. (À paraître).

Communication dans des conférences sans actes

- Table ronde Jeunes chercheur(e)s et une communication aux 3es journées éthique et intelligence artificielle à la plate-forme de l'AFIA (Association Française d'Intelligence Artificielle) : dans le cadre de la conférence PFIA'2018, Université de Nancy le 2 juillet 2018 :

Yameogo RA, Colloc J. Le consentement et l'usage de l'intelligence artificielle en médecine. Journée « Ethique & IA », 11^e Plate-forme d'Intelligence Artificielle, PFIA 2018, Nancy, 2 juillet 2018.

Colloc J., **Yameogo RA**. De l'IA mimétique à l'IA autonome en santé : Une boîte à outils éthique pour évaluer le statut des utilisateurs humains dans la relation de soin, Journée

« Ethique & IA », 11^e Plate-forme d'Intelligence Artificielle, PFIA 2018, Nancy, 2 juillet 2018.

- Table Ronde sur l'apport des TIC et la problématique de la gestion des données dans la prise en charge des maladies chroniques au Burkina Faso lors des 6es journées de la Société de Cardiologie du Burkina (Ouagadougou, Burkina Faso) le 20 juillet 2018) :

Yameogo RA. Santé numérique et maladies chroniques au Burkina Faso : état des lieux

Séminaire :

- Colloc J., **Yameogo RA.**, Big-data, cloud, objets connectés appliqués à la santé : avantages et risques éthiques. Campus Logistique PIL, demi-journée avec le pôle TES Axe Thématique e-Santé. 23 novembre 2016, Le Havre.

Annexe 2 : lettre de demande de terrain à l'ordre des médecins Seine-Maritime

ides



UMR 6266 CNRS IDEES

Docteur Joël COLLOC
Professeur d'Informatique à l'université du Havre
Laboratoire UMR CNRS 6266 IDEES
25, Rue Philippe Lebon
Université du Havre, 76600 Le Havre
Tel : 06 19 19 30 41
joel.colloc@univ-lehavre.fr

Le Havre, le 11 juin 2018

Docteur Patrick DAIMÉ
Président du Conseil Départemental
de la Seine-Maritime de l'Ordre des médecins

Objet : Recherche de terrain concernant l'impact de l'utilisation du numérique, des outils d'aide à la décision sur la relation médecin-patient par le Docteur Relwendé Aristide Yaméogo

Monsieur le Président, mon cher confrère,

Nous sollicitons l'agrément du Conseil Départemental de la Seine Maritime de l'Ordre des Médecins pour une étude scientifique que nous effectuons auprès de nos confrères et de leurs patients.

Contexte : Le Docteur Relwendé Aristide Yaméogo est médecin cardiologue du Burkina-Faso, titulaire d'un Master 2 de Santé Publique de l'université d'Aix Marseille. Il effectue actuellement une thèse transdisciplinaire en informatique médicale sous ma direction et en Sciences d'information et de la communication sous la direction du Professeur Béatrice Galinon-Méléneq. Durant sa thèse le Docteur Yaméogo est employé à l'université du Havre à l'aide d'une allocation de recherche du MENSUR. Le titre est « Risques et perspectives du Big Data en santé et des systèmes d'information autonomes : approche éthique et épistémologique ». Dans le cadre de sa thèse, le Docteur Yaméogo doit effectuer une recherche de terrain auprès des médecins du département de Seine Maritime et de leurs patients afin d'étudier l'impact de l'utilisation du numérique sur la relation de soin médecin-patient, en particulier de celui des nouveaux outils informatiques : le dossier médical informatisé, le « Big data », les « Cloud », Internet, les objets connectés, les systèmes d'aide à la décision, les robots, le « dossier médical partagé » (DMP). L'étude concerne l'acceptabilité de ces outils par les médecins, les aspects éthiques liés à cette utilisation. Les questions auxquelles nous souhaitons répondre sont les suivantes : Quels usages les médecins font effectivement de l'informatique au sein de leur cabinet. Quelle perception ont les patients lors de l'usage de ces outils par leur médecin ? Quels sont les impacts sur la relation de soin ?

Méthode : Cette étude comprend un entretien avec les médecins et un questionnaire destiné à leurs patients. Il ne contient aucune donnée clinique des patients. Le questionnaire patient ne comporte que des questions concernant l'usage de l'informatique au sein du cabinet du médecin. Les données recueillies seront anonymes. Elles seront traitées statistiquement et les résultats globaux de l'étude seront communiqués aux participants.

Conformément au Règlement Général de la Protection des Données (RGPD), les données personnelles collectées durant cette étude seront stockées par notre université sous la responsabilité du référent RGPD de notre université Romain Baudry et de moi-même en tant que référent pour le laboratoire IDEES-Le Havre.

Ainsi, je demande au Conseil de nous donner son agrément et son soutien en faisant la publicité de notre recherche au sein de son prochain bulletin d'information afin de sensibiliser nos confrères à l'intérêt scientifique de cette étude et de réserver le meilleur accueil à notre confrère le Dr Yaméogo pour mener à bien cette recherche.

Je reste à votre disposition pour toute information complémentaire.

Veuillez agréer Monsieur le Président, Mon cher Confrère l'expression de mes sentiments confraternels et je vous remercie de transmettre à l'ensemble des membres du Conseil mes meilleures salutations.

Docteur Joël COLLOC
Professeur d'informatique
Réfèrent RGPD IDEES-Le Havre

Annexe 3 : réponse de l'ordre des médecins de la Seine-Maritime



ORDRE NATIONAL DES MÉDECINS
Conseil Départemental de l'Ordre

Docteur Joël COLLOC
Joel.colloc@univ-lehavre.fr

ROUEN, le 5 juillet 2018

PD/ID

Monsieur et Cher Confrère,

Notre Conseil a pris connaissance de votre courrier du 11 juin dernier, avec toute l'attention requise.

Il ne nous appartient pas de donner un agrément mais plutôt de rappeler les règles déontologiques qui s'appliquent. Citons notamment le respect de la confidentialité du secret professionnel, l'accord du patient, le respect des nouvelles obligations du RGPD. A ce sujet, un document est disponible sur le site du Conseil National de l'Ordre des Médecins, document élaboré de concert avec la CNIL.

Cette recherche du Docteur YAMÉOGO pourrait faire l'objet d'une insertion dans notre prochain bulletin à la rubrique « communiqué », sous réserve de la validation par l'équipe éditoriale.

Recevez, Monsieur et Cher Confrère, l'expression de nos sentiments confraternels et dévoués.



Docteur Patrick DAIMÉ
Président

PS : date de sortie du prochain bulletin : automne 2018

Annexe 4 : lettre de demande de participation des médecins



Le Havre, le 29 septembre 2018

Dr Relwendé Aristide YAMEOGO
Laboratoire IDEES-Le Havre
UMR CNRS 6266
25, Rue Philippe Lebon BP 1123
76063 Le Havre Cedex
07 53 05 24 33
relwende-aristide.yameogo@univ-lehavre.fr

Objet : Enquête « Impact de l'utilisation de l'informatique et de l'intelligence artificielle sur la pratique médicale et la relation médecin-patient »

Chère Consœur, Cher Confrère,

Je suis médecin cardiologue du Burkina-Faso et je prépare une nouvelle thèse, à l'université du Havre dont le titre est « Risques et perspectives des « Big data » en santé et des systèmes d'information autonomes : approche éthique et épistémologique », sous la direction du docteur Joël Colloc, professeur d'informatique et de Béatrice Galinon-Méléneq professeur en sciences de l'information et de la communication.

Dans le cadre de cette thèse je réalise une enquête afin d'évaluer l'impact de l'utilisation des outils informatiques et des systèmes d'information sur la relation de soin médecin-patient (dossier médical informatisé, « big data », « cloud », Internet, objets connectés, systèmes d'aide à la décision, robots, «dossier médical partagé»). Il s'agit d'étudier l'acceptabilité, l'utilisation effective de ces outils au sein du cabinet, les questions éthiques qui se posent, la perception des patients de cet usage et les impacts sur la relation de soin.

Déroulement de l'enquête

Dans le cadre de cette enquête, je propose un entretien avec les médecins et un questionnaire en ligne destiné à leurs patients qui ne contient aucune donnée clinique. Les questions concernent uniquement l'usage des outils informatiques rappelés ci-dessus dans le cadre des soins au patient. Les données recueillies seront anonymes. Je traiterai les données recueillies à l'aide de modèles statistiques usuels et les résultats globaux de l'enquête seront communiqués aux participants.

Conformément au Règlement Général de la Protection des Données (RGPD), les données personnelles collectées durant cette étude seront stockées par notre université sous la responsabilité du référent RGPD de notre université Romain Baudry et de Joël Colloc en tant que référent du laboratoire IDEES-Le Havre.

Je reste à votre disposition pour toutes vos questions et je vous serai très reconnaissant, si vous consentiez à participer à cette enquête, de bien vouloir prendre contact avec moi par téléphone ou par mail afin que nous puissions prendre un rendez-vous.

Veuillez agréer, chère consœur, cher confrère, l'expression de mes confraternelles salutations.

Dr Relwendé Aristide Yameogo

relwende-aristide.yameogo@univ-lehavre.fr

Annexe 5 : formulaire de consentement des médecins



FORMULAIRE DE CONSENTEMENT

Titre de la recherche :

**Risques et perspectives des big data en santé et des systèmes d'information autonomes :
approche éthique et épistémologique**

Fait à, Le

Je soussigné(e)

accepte de participer à l'enquête « Impact de l'utilisation de l'informatique et de l'intelligence artificielle sur la pratique médicale et la relation médecin-patient » dans le cadre de la thèse dont le titre est rappelé ci-dessus.

Les objectifs et modalités de l'enquête m'ont été expliqués dans lettre d'information transmise par Docteur YAMEOGO Relwendé Aristide.

J'ai pris connaissance des objectifs et du déroulement de cette enquête.

J'accepte de réaliser l'interview dans le cadre de l'enquête. A l'exception de ces personnes, qui traiteront les informations dans le plus strict respect de la confidentialité, mon anonymat sera préservé.

J'ai bien compris que ma participation à l'étude est volontaire. Je suis libre d'accepter ou de refuser de participer, et je suis libre d'arrêter à tout moment ma participation à cette enquête.

Mon consentement ne décharge pas les organisateurs de cette étude de leurs responsabilités. Je conserve tous mes droits garantis par la loi.

Après en avoir discuté et avoir obtenu la réponse à toutes mes questions, j'accepte librement et volontairement de participer à l'enquête qui m'est proposée.

Nom et signature de l'investigateur

Signature du médecin

Annexe 6 : fiche d'entrevue des médecins

1. Etat Civil

1.1. Date de naissance

1.2. Sexe Homme Femme

1.3. Spécialité médicale

1.4. Date de soutenance

1.5. Date de début de la pratique médicale

← Précédent

Suivant →

2. Utilisation des objets connectés

2.1. Utilisez-vous les objets connectés dans le cadre de la pratique médicale? Oui Non

2.2. Dans quel(s) but(s): Diagnostic Evaluation du traitement
 Suivi et surveillance Education du patient

2.3. Utilisez-vous des objets connectés validés dispositifs médicaux? Oui Non

2.4. Si non, quel est le degré de confiance accordé à ces objets connectés

2.5. Conseillez-vous l'usage des applications/site web à vos patients? Oui Non

2.6. Si non, pourquoi?

2.7. Pour quel(s) usages: Informations Partage d'expérience
 Conseils Aide au traitement
 Suivi de leur maladie

← Précédent

Suivant →

2.8. Quels sont selon votre opinion, les freins à l'utilisation des objets connectés

2.9. Selon votre opinion, quelles sont les actions pouvant améliorer l'utilisation des objets connectés

3. Protection des données

3.1. Aimeriez-vous partager les données de santé de vos patients dans le Oui Non
cadre de la recherche

3.2. Pour quelles raisons ?

3.3. Selon votre opinion, les données sont-elles protégées par les différentes procédures mise en place:

| | Pas du tout | Plutôt non | Cela dépend | Plutôt oui | Tout à fait |
|---|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Loi informatique et libertés | <input type="radio"/> |
| Agrément hébergeurs données de santé | <input type="radio"/> |
| RGPD | <input type="radio"/> |
| Traité de transfert de données UE- USA (Privacy Shield) | <input type="radio"/> |

3.4. Quel degré de confiance accordez-vous aux institutions pour la protection des données?

| | Très méfiant | Assez méfiant | Indifférent | Confiant | Très confiant |
|--|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Hôpitaux et cabinets privés | <input type="radio"/> |
| Laboratoire de recherche (chercheurs) | <input type="radio"/> |
| Sécurité sociale | <input type="radio"/> |
| Assurance | <input type="radio"/> |
| Entreprise privée | <input type="radio"/> |
| hébergeurs de données | <input type="radio"/> |
| Institut National des Données de Santé | <input type="radio"/> |

3.5. Quelles sont vos attentes par rapport à l'usage des big data et l'intelligence artificielle dans le domaine de la santé

← Précédent Suivant →

4. Big data et intelligence artificielle

4.1. Selon votre opinion, quelles sont les modifications que les big data et l'IA pourront apporter dans la pratique médicale sur :

- | | |
|--|--|
| <input type="checkbox"/> le parcours de soins | <input type="checkbox"/> l'indépendance du médecin |
| <input type="checkbox"/> la relation médecin-patient | <input type="checkbox"/> le secret médical et la confidentialité des données |

4.2. Quelles sont ces modifications

4.3. Le big data et l'IA pourront-elles tenir leurs promesses?

- | | |
|---|--|
| <input type="checkbox"/> Amélioration de la prise en charge et de la prévention | <input type="checkbox"/> Réduction des dépenses de santé |
| <input type="checkbox"/> Autonomie du patient | <input type="checkbox"/> Meilleure politique de santé |

4.4. Pouvez-vous commenter les promesses qui seront tenues?

← Précédent Suivant →

4.5. Quels sont les risques pour l'éthique médicale quant à l'utilisation du big data et de l'IA

4.6. Selon votre opinion, quelles sont les conditions à réunir pour une meilleure adhésion des acteurs de la santé à l'usage du big data et de l'IA

← Précédent Suivant →

5. Ethique et finalités des données

5.1. Bienfaisance

| | Tout à fait d'accord | D'accord | Neutre | Pas d'accord | Pas du tout d'accord |
|---|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Aider à la prise de décision médicale pour les professionnels de santé | <input type="radio"/> |
| Promouvoir la qualité, l'organisation, la gestion et la planification de la prise en charge des patients | <input type="radio"/> |
| Œuvrer pour le bien des patients | <input type="radio"/> |
| Partager une information transparente et accessible entre les professionnels de santé et le patient | <input type="radio"/> |
| S'assurer de la qualité et du choix de l'information transmise au patient | <input type="radio"/> |
| Améliorer la continuité des soins | <input type="radio"/> |
| Aider les autorités de santé à répondre aux attentes et à la prise en charge de la santé chez l'utilisateur | <input type="radio"/> |
| Instaurer auprès du patient une légitimité du droit et du traitement de l'information | <input type="radio"/> |
| Etablir un devoir de sécurité, d'intégrité, de traçabilité et de protection des données médicales | <input type="radio"/> |

← Précédent Suivant →

5.2. Justice

| | Tout à fait d'accord | D'accord | Neutre | Pas d'accord | Pas du tout d'accord |
|--|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Evaluer les performances et cerner les domaines où l'action s'impose | <input type="radio"/> |
| Permettre une analyse épidémiologique ou statistique | <input type="radio"/> |
| Améliorer et renforcer l'interactivité avec les acteurs extérieurs à la structure de soins | <input type="radio"/> |
| Renforcer la disponibilité des professionnels de santé | <input type="radio"/> |
| Elaborer et partager une information précise et adaptée à la population | <input type="radio"/> |

5.3. Non Malfaisance

| | Tout à fait d'accord | D'accord | Neutre | Pas d'accord | Pas du tout d'accord |
|--|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Suivre la réglementation législative des données | <input type="radio"/> |
| Respecter les règles de stockage, d'hébergement et de diffusion instaurées par la CNIL | <input type="radio"/> |
| Minimiser ou éliminer des torts faits aux patients du fait d'une mauvaise information | <input type="radio"/> |
| Réduire les risques inutiles ou mal calculés | <input type="radio"/> |
| Assurer la fiabilité des données médicales et leurs permanences | <input type="radio"/> |
| S'assurer de la pertinence technique et du bien-fondé humain des outils (Big data et IA) | <input type="radio"/> |

5.4. Autonomie

| | Tout à fait d'accord | D'accord | Neutre | Pas d'accord | Pas du tout d'accord |
|--|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Replacer le patient au centre de la décision en lui apportant une information médicale plus complète et rapide: une meilleure autonomie du patient | <input type="radio"/> |
| Respecter la vie privée et le droit au secret médical et à la confidentialité | <input type="radio"/> |
| Adapter une technologie aux connaissances et savoir-faire du professionnel de santé | <input type="radio"/> |
| S'assurer du consentement et de l'adhésion du propriétaire de la donnée | <input type="radio"/> |
| Respecter le droit à l'information préalable, de rectification et d'opposition décrits auprès du propriétaire de la donnée | <input type="radio"/> |

Annexe 7 : lettre d'information des patients



Le Havre, le 29 septembre 2018

Dr Relwendé Aristide YAMEOGO
Laboratoire IDEES-Le Havre
25, Rue Philippe Lebon BP 1123
76063 Le Havre Cedex
07 53 05 24 33
relwende-aristide.yameogo@univ-lehavre.fr

Objet : Enquête « Impact de l'utilisation de l'informatique et de l'intelligence artificielle sur la pratique médicale et la relation médecin-patient »

Chère Madame, Cher Monsieur,

Je suis médecin cardiologue du Burkina-Faso et je prépare une nouvelle thèse, à l'université du Havre dont le titre est « Risques et perspectives des « Big data » en santé et des systèmes d'information autonomes : approche éthique et épistémologique », sous la direction du docteur Joël Colloc, professeur d'informatique et de Béatrice Galinon-Méléneq professeur en sciences de l'information et de la communication.

Dans le cadre de cette thèse je réalise une enquête afin d'évaluer l'impact de l'utilisation des outils informatiques et des systèmes d'information sur la relation de soin médecin-patient (dossier médical informatisé, « big data », « cloud », Internet, objets connectés, systèmes d'aide à la décision, robots, « dossier médical partagé»). Il s'agit d'étudier l'acceptabilité, l'utilisation effective de ces outils au sein du cabinet, les questions éthiques qui se posent, la perception des patients de cet usage et les impacts sur la relation de soin.

Dans le cadre de cette enquête, je propose un entretien avec votre médecin et un questionnaire en ligne pour vous qui ne contient aucune donnée clinique. Les questions concernent uniquement votre usage des outils informatiques rappelés ci-dessus dans le cadre de votre santé. Les données recueillies seront anonymes. Je traiterai les données recueillies à l'aide de modèles statistiques usuels et les résultats globaux de l'enquête vous seront communiqués.

Conformément au Règlement Général de la Protection des Données (RGPD), les données personnelles collectées durant cette étude seront stockées par notre université sous la responsabilité du référent RGPD de notre université Romain Baudry et de Joël Colloc en tant que référent du laboratoire IDEES-Le Havre.

Je reste à votre disposition pour toutes vos questions et je vous serai très reconnaissant, si vous consentiez à participer à cette enquête, de bien vouloir vous connectez à l'adresse :

<https://goo.gl/G6x56B>
Votre code médecin : 20

Veuillez agréer, chère Madame, cher Monsieur, l'expression de mes confraternelles salutations.

Dr Relwendé Aristide Yameogo

07 53 05 24 33

relwende-aristide.yameogo@univ-lehavre.fr

Annexe 8 : questionnaire adressée aux patients

Enquête sur l'utilisation de la santé numérique

Nous souhaitons recueillir votre avis **sur l'utilisation du numérique dans le domaine de la santé.**

Le questionnaire comportent des questions à caractère obligatoire et facultative.

YAMEOGO Relwendé Aristide, doctorant au laboratoire IDEES Le Havre dans le cadre de sa thèse , est **responsable du traitement des données.**

Vous pouvez avoir accès à vos données, les rectifier ou vous opposer à leurs utilisations en envoyant **vos données** à l'adresse

relwende-aristide.yameogo@univ-lehavre.fr

Code Médecin

Numéro de compteur

Votre consentement est nécessaire pour la participation à cette enquête : Oui Non

1. Utilisation de la santé numérique

1.1. Faites-vous des recherches de santé sur internet Oui Non

1.2. Pour qui faites-vous ces recherches? pour vous même Par curiosité pour vous informer
 Pour un parent, un ami Autre

Si 'Autre' précisez :

1.3. Vous réalisez vos recherches sur internet Avant la consultation Des moments sans lien avec la consultation
 Après la consultation Autre

Si 'Autre' précisez :

1.4. Les recherches sur internet modifient-ils votre comportement?

- | | |
|--|---|
| <input type="checkbox"/> Non | <input type="checkbox"/> Arrêt du traitement |
| <input type="checkbox"/> Automédication | <input type="checkbox"/> Modification du traitement |
| <input type="checkbox"/> Changement de style de vie | <input type="checkbox"/> Autre |
| <input type="checkbox"/> Réduction des visites chez le médecin | |

Si 'Autre' précisez :

1.5. Avez-vous confiance aux informations sur internet?

Très méfiant Assez méfiant Indifférent Confiant Très confiant

1.6. Utilisez-vous les réseaux sociaux dans le cadre de votre santé?

Non Twitter
 Facebook Forum de discussion
 Whatsapp Autre

Si 'Autre' précisez :

1.7. Etes-vous membre d'un groupe de réseau de patient en ligne? Oui

Non

1.8. Si oui, quel est l'intérêt du groupe social pour vous

Echanger avec d'autres patients Conseils d'autres patients
 Partager mon expérience Combler la difficile relation avec mon médecin
 M'informer sur les médicaments Autre

Si 'Autre' précisez :

1.9. Quel degré de confiance accordez-vous aux réseaux sociaux pour ces informations de santé

Très méfiant Assez méfiant Indifférent Confiant Très confiant

2. Utilisation des objets connectés

2.1. Utilisez-vous des objets connectés? Oui

Non

2.2. Quels sont outils que vous utilisez?

Glucomètre ECG
 Tensiomètre Balance
 Montre connectée Autre

Si 'Autre' précisez :

2.3. Les objets connectés ont-ils été recommandés par votre médecin? Oui

Non

2.4. L'utilisation des objets connectés améliore-t-elle la relation avec votre médecin

Beaucoup moins Moins Egal Plus Beaucoup Plus

2.5. Quels sont les freins de l'usage des objets connectés

Aucun intérêt J'ai peur des fuites de données
 Peur de ne pas savoir les utiliser Le prix des objets connectés est élevé
 mon médecin n'en parle pas Je préfère le contact avec mon médecin
 Je ne suis pas convaincu de l'efficacité Mon médecin me déconseille

3. Protection des données

3.1. Faites-vous confiance aux site web et réseaux sociaux pour la protection de vos données

Très méfiant Assez méfiant Indifférent Confiant Très confiant

3.2. Aimerez-vous partager vos données de santé dans le cadre de la recherche

Oui Non

3.3. Faites-vous confiance aux institutions suivantes pour la protection de vos données?

| | Très méfiant | Assez méfiant | Indifférent | Confiant | Très confiant |
|----------------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Professionnels de santé | <input type="radio"/> |
| Chercheurs | <input type="radio"/> |
| Sécurité sociale | <input type="radio"/> |
| Firme pharmaceutique | <input type="radio"/> |
| Assurances | <input type="radio"/> |
| Entreprise privée | <input type="radio"/> |
| Les institutions étatiques | <input type="radio"/> |

3.5. Quelles sont vos craintes par rapport à la numérisation de la santé?

- Surveillance à tout moment
- Manque d'intimité et de vie privée
- Perte du secret médical
- Divulgence des données personnelles
- Diminution de votre pouvoir de décision sur votre santé
- Discrimination (embauche, assurance, prêt bancaire, réputation)
- Déshumanisation de la pratique médicale
- Divulgence de vos données santé
- Autre

Si 'Autre' précisez :

4. Confiance en votre médecin

Quelle degré de confiance accordez-vous à votre médecin?

| | Tout à fait d'accord | D'accord | Neutre | Pas d'accord | Pas du tout d'accord |
|---|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Je ne doute pas que mon médecin me soigne en tant que personne | <input type="radio"/> |
| Mon médecin tient compte de mes besoins qu'il considère en premier | <input type="radio"/> |
| J'ai tellement confiance en mon médecin que j'essaie toujours de suivre ses conseils | <input type="radio"/> |
| Si mon médecin me dit quelque chose, alors je considère qu'elle est vraie | <input type="radio"/> |
| J'ai toujours confiance en l'opinion de mon médecin et je juge inutile de demander un deuxième avis | <input type="radio"/> |
| J'ai confiance au jugement de mon médecin concernant les soins médicaux prodigués | <input type="radio"/> |
| Je pense que mon médecin fait tout ce qu'il devrait | <input type="radio"/> |
| Je fais confiance en mon médecin pour mettre mes besoins au-dessus de toutes autres considérations quand il traite mes problèmes médicaux | <input type="radio"/> |
| Mon médecin est réellement un expert pour prendre soin de mes problèmes de santé | <input type="radio"/> |
| Je fais confiance en mon médecin pour me dire si une erreur a été commise au sujet de mon traitement | <input type="radio"/> |
| Je regrette parfois que mon médecin ne garde pas confidentielles les informations dont nous avons discutées | <input type="radio"/> |

5. Etat Civil

5.1. Date de naissance

Date type jour/mois/année

5.2. Quel est votre niveau d'études ?

- | | | | |
|--|---------------------------------|---|--|
| <input type="radio"/> DEA, doctorat, diplôme d'ingénieur | <input type="radio"/> BTS, DEUG | <input type="radio"/> Première, terminale | <input type="radio"/> BE, BEPC, classe de collège et seconde |
| <input type="radio"/> Licence, maîtrise (bac + 3 ou 4) | <input type="radio"/> Bac | <input type="radio"/> CAP, BT, BET, BEP | <input type="radio"/> Sans diplôme, CEP |

5.3. Catégorie socio-professionnelle

- | | |
|--|--|
| <input type="radio"/> Agriculteurs exploitants | <input type="radio"/> Employés |
| <input type="radio"/> Artisans, commerçants, chefs entreprise | <input type="radio"/> Ouvriers |
| <input type="radio"/> Cadres, professions intellectuelles sup. | <input type="radio"/> Retraités |
| <input type="radio"/> Professions intermédiaires | <input type="radio"/> Autres sans activité professionnelle |

5.4. Précisez votre catégorie professionnelle antérieure

5.5. Êtes-vous... Un homme Une femme

5.6. Depuis combien d'années êtes-vous suivi par votre médecin?

Annexe 9 : serment d’Hippocrate

Texte revu par l’Ordre des médecins en 2012

« Au moment d’être admis(e) à exercer la médecine, je promets et je jure d’être fidèle aux lois de l’honneur et de la probité.

Mon premier souci sera de rétablir, de préserver ou de promouvoir la santé dans tous ses éléments, physiques et mentaux, individuels et sociaux.

Je respecterai toutes les personnes, leur autonomie et leur volonté, sans aucune discrimination selon leur état ou leurs convictions. J’interviendrai pour les protéger si elles sont affaiblies, vulnérables ou menacées dans leur intégrité ou leur dignité. Même sous la contrainte, je ne ferai pas usage de mes connaissances contre les lois de l’humanité.

J’informerai les patients des décisions envisagées, de leurs raisons et de leurs conséquences.

Je ne tromperai jamais leur confiance et n’exploiterai pas le pouvoir hérité des circonstances pour forcer les consciences.

Je donnerai mes soins à l’indigent et à quiconque me les demandera. Je ne me laisserai pas influencer par la soif du gain ou la recherche de la gloire.

Admis(e) dans l’intimité des personnes, je tairai les secrets qui me seront confiés. Reçu(e) à l’intérieur des maisons, je respecterai les secrets des foyers et ma conduite ne servira pas à corrompre les mœurs.

Je ferai tout pour soulager les souffrances. Je ne prolongerai pas abusivement les agonies. Je ne provoquerai jamais la mort délibérément.

Je préserverai l’indépendance nécessaire à l’accomplissement de ma mission. Je n’entreprendrai rien qui dépasse mes compétences. Je les entretiendrai et les perfectionnerai pour assurer au mieux les services qui me seront demandés.

J’apporterai mon aide à mes confrères ainsi qu’à leurs familles dans l’adversité.

Que les hommes et mes confrères m’accordent leur estime si je suis fidèle à mes promesses ; que je sois déshonoré(e) et méprisé(e) si j’y manque. »

Annexe 10 : participation aux ateliers, séminaires et colloques

- 18 au 20 octobre 2017 : Conférence internationale de Hypertextes et Hypermédias, Produits – Outils – Méthodes sur le thème « le numérique à l'ère des designs, de l'hypertexte et de l'hyperexpérience » (Valenciennes)
- 07 décembre 2017 : Journée Unitwin etrace « Mon robot, ma famille et moi » (Université du Havre)
- 13 décembre 2017 : Séminaire de l'Académie Internationale d'éthique, médecine et politique publique sur le thème « la médecine au défi de l'intelligence artificielle » (Paris)
- 14 décembre 2017 : Journée organisée par Etudes Digitales sur le thème « la donnée n'est pas la pensée » (Paris)
- 19-20 mars 2018 : Séminaire de l'Académie Internationale d'éthique, médecine et politique publique sur le thème « données massives : enjeux éthiques, juridiques en recherche et en clinique » (Paris)
- 16 au 20 avril 2018 : Doctoriales 2018 : séminaire de formation pour l'insertion professionnelle des doctorants (Saint Valery-en-caux)
- 29 mai 2018 : Séminaire AI-LOG sur le thème « Intelligence artificielle : historique et enjeux » (Université du Havre).
- 6 au 8 juin 2018 : Colloque international « réseaux sociaux, traces numériques et communication électronique » (Université du Havre)
- 12 décembre 2018 : Colloque « droits de l'Homme au défi de nouveaux territoires de la bioéthique » organisé par l'Espace Ethique IDF
- 13 décembre 2018 : Co-organisateur de la journée des doctorants du Havre 2018 sous le thème « Centres et centralités en sciences humaines et sociales »
- 7 au 9 janvier 2019 : Colloque « La personne humaine au 21ème siècle : interculturalités et progrès des science et techniques appliqués au corps » organisée par l'Académie Internationale d'Ethique, de Médecine et de Politique Publique.
- 15 mai 2019 : Colloque Poléthis : « données massives en recherche : reconfiguration du paysage de la recherche : enjeux éthiques et scientifiques »

organisé par l'Espace Ethique IDF

- 16 mai 2019 : Colloque numérique et santé (IA, big data, données de santé)
organisé par la Cour de Cassation de Paris

Résumé

Au XXIème siècle, l'utilisation du *big data* et de l'IA dans le domaine de la santé s'est progressivement étendue bien qu'elle soit accompagnée par de nombreux problèmes liés à l'émergence de pratiques basées sur un usage des traces numériques. Cet usage ne prend pas toujours en compte les règles éthiques de la pratique médicale. Notre pratique de médecin nous a conduit à produire une recherche qui s'inscrit dans ce contexte. Elle a pris la forme d'une thèse en informatique qui interroge la part de l'humain et a pour but d'évaluer l'utilisation du *big data* et de l'IA dans la pratique médicale, de découvrir les processus engendrés par les outils numériques en matière de santé et de souligner les problèmes d'éthique qu'ils posent en matière de respect de la vie privée et de secret professionnel.

Au moment de notre recherche, l'utilisation des TIC dans la pratique médicale passe essentiellement par l'utilisation des dossiers patients informatisés, de logiciels d'aide à la prescription et d'objets connectés. Ces usages soulèvent de nombreux problèmes pour les médecins conscients du risque encouru quant à la protection des données de santé des patients. L'absence de validation scientifique de la plupart des outils d'IA et des objets connectés constituent donc pour eux des freins à leur utilisation. Dans ce travail, nous mettons en place une méthode de conception de système d'aide à la pratique médicale, l'espace vectoriel flou temporel (EVFT). Cette méthode s'appuie sur le modèle orientée objet, la logique floue et la dynamique temporelle. Elle ne nécessite pas un processus de défuzzification et permet de prendre en compte l'évolution clinique du patient et de comparer les cas cliniques. Cette méthode nous permet de modéliser un nouveau score diagnostique clinique de l'embolie pulmonaire. A travers le paradigme anthroposémiotique l'« homme-trace », notre recherche permet, non seulement de prendre la mesure de la limitation dans l'usage des TIC, mais aussi de mettre en évidence les biais interprétatifs dus à la déliaison effectuée entre l'individu pris dans sa complexité d'« homme-trace » et les *data* circulant à son sujet *via* les traces numériques. Si le *big data*, couplé à l'IA peut jouer un grand rôle dans la mise en place de système d'aide à la décision médicale, il ne peut pas se limiter pas à ce champ. En effet, en bouleversant les fondements de la pratique médicale, ils sont susceptibles de priver le patient de la relation humaine avec son médecin, cet autre humain dont le « corps-trace » a été conduit à expérimenter pour lui-même les difficultés inhérentes au corps vivant. Nous étudions donc également comment la situation de communication entre le médecin et le patient est susceptible d'évoluer dans une société où la communication numérique permet de relier tous les acteurs du « care ». L'IA non seulement ne peut remplacer cette correspondance entre « corps-traces » vivants mais aussi peut provoquer des problèmes éthiques. C'est pourquoi nous étudions aussi comment mettre en place des processus de développement du *big data* et de IA respectant les droits des personnes, des règles déontologiques et d'éthique médicale associées à l'appropriation des TIC par les acteurs du système de santé.

Mots-clés : *E-santé - Systèmes d'information - Big data - Intelligence artificielle - Ethique - Traces - Anthroposémiotique - Homme-trace - Corps-trace - Systèmes complexes.*

Abstract

In the 21st century, the use of big data and artificial intelligence (AI) in the field of health has gradually expanded, although it is accompanied by many problems linked to the emergence of practices based on the use of digital traces. This use does not always take into account the ethical rules of medical practice. Our practice as doctor has led us to produce research that is part of this context. It took the form of a thesis in informatics that questions the human part and aims to evaluate the use of big data and AI in medical practice, to discover the processes generated by digital tools in health and to highlight the ethical problems they pose in terms of privacy and professional secrecy.

At the time of our research, the use of ICTs in medical practice is mainly through the use of electronics health records, prescription support software and internet of things. These uses raise many issues for physicians who are aware of the risk to the protection of patients' health data. The lack of scientific validation of most AI tools and internet of things is therefore a barrier to their use. In this work, we are implementing a method for designing a decision support system, the temporal fuzzy vector space (TFVS). This method is based on the object-oriented model, fuzzy logic and temporal dynamics. It does not require a defuzzification process and allows to take into account the clinical evolution of the patient and to compare clinical cases. This method allows us to model a new clinical diagnosis score of pulmonary embolisms. Through the "*human-trace*" anthroposemiotic paradigm, our research, not only allows us to take the measure of this limitation in the use of ICTs, but also to highlight the interpretative biases due to the delinking of the individual in his complexity as a "human-trace" and the data flowing about him via digital traces. If big data, coupled with AI can play a major role in the implementation of decision support systems, it cannot be limited to this field. Indeed, by upsetting the foundations of medical practice, they are likely to deprive the patient of the human relationship with his doctor, this other human whose "body-trace" has been led to experience for himself the difficulties inherent in the living body. We are therefore also studying how the situation of communication between doctor and patient is likely to evolve in a society where digital communication makes it possible to link all those involved in "care". AI cannot replace this correspondence between living "body-trace" but it can cause ethical problems. This is why we are also studying how to set up processes for the development of big data and AI that respect the rights of individuals, deontological rules and medical ethics associated with the appropriation of ICTs by the actors of the health system.

Keywords: E-health - Information systems - Big data - Artificial intelligence - Ethics - Traces - Anthroposemiotic – Human-trace – body-trace - Complex systems.

Relwende Aristide YAMEOGO
Risques et perspectives du *big data* et de l'intelligence artificielle : approche éthique et épistémologique

Au XXIème siècle, l'utilisation du *big data* et de l'IA dans le domaine de la santé s'est progressivement étendue bien qu'elle soit accompagnée par de nombreux problèmes liés à l'émergence de pratiques basées sur un usage des traces numériques. Cet usage ne prend pas toujours en compte les règles éthiques de la pratique médicale. Notre pratique de médecin nous a conduit à produire une recherche qui s'inscrit dans ce contexte. Elle a pris la forme d'une thèse en informatique qui interroge la part de l'humain et a pour but d'évaluer l'utilisation du *big data* et de l'IA dans la pratique médicale, de découvrir les processus engendrés par les outils numériques en matière de santé et de souligner les problèmes d'éthique qu'ils posent en matière de respect de la vie privée et de secret professionnel.

Au moment de notre recherche, l'utilisation des TIC dans la pratique médicale passe essentiellement par l'utilisation des dossiers patients informatisés, de logiciels d'aide à la prescription et d'objets connectés. Ces usages soulèvent de nombreux problèmes pour les médecins conscients du risque encouru quant à la protection des données de santé des patients. L'absence de validation scientifique de la plupart des outils d'IA et des objets connectés constituent donc pour eux des freins à leur utilisation. Dans ce travail, nous mettons en place une méthode de conception de système d'aide à la pratique médicale, l'espace vectoriel flou temporel (EVFT). Cette méthode s'appuie sur le modèle orientée objet, la logique floue et la dynamique temporelle. Elle ne nécessite pas un processus de défuzzification et permet de prendre en compte l'évolution clinique du patient et de comparer les cas cliniques. Cette méthode nous permet de modéliser un nouveau score diagnostique clinique de l'embolie pulmonaire. A travers le paradigme anthroposémiotique l'« *homme-trace* », notre recherche permet, non seulement de prendre la mesure de la limitation dans l'usage des TIC, mais aussi de mettre en évidence les biais interprétatifs dus à la déliaison effectuée entre l'individu pris dans sa complexité d'« *homme-trace* » et les *data* circulant à son sujet *via* les traces numériques. Si le *big data*, couplé à l'IA peut jouer un grand rôle dans la mise en place de système d'aide à la décision médicale, il ne peut pas se limiter pas à ce champ. En effet, en bouleversant les fondements de la pratique médicale, ils sont susceptibles de priver le patient de la relation humaine avec son médecin, cet autre humain dont le « *corps-trace* » a été conduit à expérimenter pour lui-même les difficultés inhérentes au corps vivant. Nous étudions donc également comment la situation de communication entre le médecin et le patient est susceptible d'évoluer dans une société où la communication numérique permet de relier tous les acteurs du « *care* ». L'IA non seulement ne peut remplacer cette correspondance entre « *corps-traces* » vivants mais aussi peut provoquer des problèmes éthiques. C'est pourquoi nous étudions aussi comment mettre en place des processus de développement du *big data* et de IA respectant les droits des personnes, des règles déontologiques et d'éthique médicale associées à l'appropriation des TIC par les acteurs du système de santé.

Mots-clés : E-santé - Systèmes d'information - Big data - Intelligence artificielle - Ethique - Traces - Anthroposémiotique - Homme-trace - Corps-trace - Systèmes complexes.