



HAL
open science

Gérer la collaboration entre l'expert métier et l'Intelligence Artificielle : Deux études de cas dans le système de soins.

Joël Perez-Torrents

► **To cite this version:**

Joël Perez-Torrents. Gérer la collaboration entre l'expert métier et l'Intelligence Artificielle : Deux études de cas dans le système de soins.. Gestion et management. Institut Polytechnique de Paris, 2024. Français. ⟨NNT : 2024IPPAX054⟩. ⟨tel-04780193⟩

HAL Id: tel-04780193

<https://theses.hal.science/tel-04780193v1>

Submitted on 13 Nov 2024

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.



HAL Authorization

Gérer la collaboration entre l'expert métier et l'Intelligence Artificielle. Deux études de cas dans le système de soins.

Thèse de doctorat de l'Institut Polytechnique de Paris
préparée à l'École polytechnique

École doctorale n°626
Institut Polytechnique de Paris (ED IP Paris)

Spécialité de doctorat : Sciences de gestion

Thèse présentée et soutenue à Palaiseau, le 6 septembre 2024, par

Joël Perez Torrents

Composition du Jury :

Hervé DUMEZ Directeur de recherche CNRS, École polytechnique (i3-CRG)	Président
Irène GEORGESCU Professeur des universités, université de Montpellier (Institut Montpellier Management)	Rapporteuse
Philippe LORINO Professeur éminent émérite, ESSEC Business School	Rapporteur
John R. KIMBERLY Professeur émérite, The Wharton School, The University of Pennsylvania (Management Department)	Examineur
Marjolaine ROSTAIN Assistant professor, Warwick Business School, The University of Warwick (ISM-Analytics Group)	Examinatrice
Étienne MINVIELLE Directeur de recherche CNRS, École polytechnique (i3-CRG)	Directeur de thèse



À mes proches



Remerciements

Cette thèse est le résultat d'un parcours de plusieurs années, et nous tenons à exprimer notre gratitude à toutes les personnes qui ont contribué à sa réalisation, qui nous ont soutenu et aidé tout au long de ce travail exigeant.

Nos premiers remerciements vont évidemment à Etienne Minvielle, notre directeur de thèse. Son accompagnement bienveillant a fait de cette aventure doctorale un chemin riche en apprentissages, nous transformant irréversiblement en une personne meilleure.

Nous exprimons ensuite notre gratitude envers Hervé Dumez, Irène Georgescu, Philippe Lorino, John R. Kimberly et Marjolaine Rostain pour le grand honneur qu'ils nous font en composant le jury de cette thèse. Nous sommes également reconnaissants envers Philippe Lorino et Simon Mathias Waelli pour leur écoute et leurs conseils avisés lors de la mini-soutenance.

Ce doctorat a été réalisé au sein du Centre de Recherche en Gestion (i3, CNRS) de l'Institut Polytechnique de Paris, un laboratoire de recherche unique et un lieu de développement privilégié pour les doctorants. Nous souhaitons d'abord exprimer notre gratitude envers toutes les personnes qui ont contribué à faciliter le déroulement de cette thèse, notamment Michèle Breton, Marie-Claude Cléon, Catherine Gayda, Elodie Gigout, Franck Guimonet, Jacky Mpoy et Carine Robert.

Cette thèse n'aurait pas été possible sans le financement de la chaire Sanofi « Numérisation des Données en Santé » à l'École polytechnique, tenue par Etienne Minvielle et Marc Lavielle.

Nos remerciements s'adressent naturellement à tous ceux qui nous ont aidés dans la collecte des données. Bien que nous ne puissions nommer individuellement chaque acteur, nous apprécions profondément leur contribution, s'étant prêtés au jeu des entretiens et des observations. En particulier, les Dr G et Dr S nous ont offert un accueil chaleureux, partageant leur quotidien, leurs connaissances et leurs réflexions. De même, tous les membres derrière les équipes des projets WILSON, KALI et KOVAK, ainsi que les acteurs rencontrés à l'Université de Pennsylvanie, ont apporté énormément au niveau du matériau empirique et de nos réflexions personnelles. Enfin, ces rencontres n'auraient pas été possibles sans l'aide d'Etienne Minvielle, de John R. Kimberly, du club Datacraft avec Isabelle Hilali et Xavier Lioneton de l'AX (l'association des anciens élèves polytechniciens), et de Gaétan Casanova. Nous espérons que ces travaux honorent la confiance qu'ils nous ont témoignée.

De plus, cette thèse a bénéficié de nombreux échanges avec nos pairs, dont nous souhaitons évoquer les principaux, à défaut de pouvoir être exhaustifs. Nous avons énormément apprécié l'apport de Philippe Lorino, que ce soient ses retours durant la 3ème conférence de PHILOS qui nous ont apporté un regard neuf et plus percutant sur notre travail,

ou à travers son livre *Pragmatism and Organisational Studies* dont l'influence sur la thèse et sur notre façon de voir le monde est évidente.

Marjolaine Rostain nous a honoré de discussions très enrichissantes, nous aidant à mieux saisir notre terrain, l'approche ethnographique et à nous guider dans notre exploration de la littérature par des recommandations judicieuses.

John R. Kimberly a contribué de manière significative à notre parcours, car outre des échanges toujours pertinents, il nous a offert un séjour à la Wharton School afin de rencontrer des médecins et des data scientists déployant des outils d'IA dans le système de soins. Cette expérience dans le Nouveau Monde nous a ouvert les yeux à bien des égards. Pour ce séjour, nous remercions également l'accueil chaleureux et le soutien financier de Valery Yakubovich et de l'équipe du Mack Institute for Innovation Management de la Wharton School.

Nous avons grandement bénéficié des retours du 6ème atelier SIG sur la théorie de la conception par le Centre Scientifique de Gestion des Mines ParisTech, l'atelier « PDW in OMS » par l'European Group for Organizational Studies, Organization Studies et l'Université de LUMSA, ainsi que de la 3ème conférence de PHILOS. Ces événements ont été l'occasion de rencontrer et d'échanger avec des esprits brillants comme Benoit Weil, Mollie Painter, Simon Altmejd et Jen Sloan.

La formation du programme du Centre Européen de Formation Approfondie à la Gestion (CEFAG) de la Fondation Nationale pour l'Enseignement de la Gestion des Entreprises nous a également beaucoup apporté et nous tenons à remercier l'ensemble des organisateurs, dont les directeurs du programme Hervé Laroche et Géraldine Schmitt. De plus, nous avons pu bénéficier des précieux conseils d'Hervé Laroche, Assaad El Akremi, Thomas Roulet, Calypso Coudeville et Emmanuel Bazzucchi, ainsi que ceux des autres participants du programme.

Enfin, les séminaires et ateliers organisés par le Centre de Recherche en Gestion se sont avérés des étapes clés pour faire progresser notre recherche. Un grand merci à l'ensemble des chercheurs et doctorants de ce laboratoire pour contribuer à un écosystème aussi riche. En particulier, Hervé Dumez et Camille Toussaint y contribuent avec l'organisation des ateliers d'écriture et des ateliers Munitantes. Merci également à Olfa Chelib, Laure Colin et Ludivine Dupont pour leur participation à ces derniers.

Merci également à ceux qui ont fait des lectures intermédiaires et des retours sur le manuscrit, dont Laure Colin, Hervé Dumez, Beinset V. Hounwanou, Allan Lubart, Pierre-Louis Meuric, Etienne Minvielle, Marjolaine Rostain, Christophe Sakuiwongsa, Véronique Steyer, Montserrat Torrents Moglia, Camille Toussaint, Léa Turpin, Estelle Rémondeau et Philippe Lorino.

Outre l'aide sur la thèse, nous souhaitons également rendre compte du soutien reçu pendant cette aventure. Ces remerciements vont d'abord à notre famille et tout

particulièrement à notre mère, mais également à notre directeur de thèse (encore !), et à nos amis, parmi lesquels nous n'avons pas encore cité Samuel Bernold, Etienne Bonnafoux, Martin Buron, Valentin Chomel, Amadeo Cervone, Léna Chouchan, Leeza DiGriz, Lorenzo Fioni, Aurore Fierobe, Arthur Gâteau, Klaudia Glogowska, Guillaume Goujard, Jonathan Langlois, Carolyn LaGreca, Rodirgue Lelotte, Lucie Liversain, Jeanne Redaud, Simon Queyrut, Sara Ray, Jules Sanchez, Alain Sinou, Guillaume Solignac, Frédéric Vialade, Floyd van Bolderen, Zigfrid Zvezdin, Matthew Jauregui, Lucas Minéo, Delphine Mion, Matthew Myers, Anna Miller, Romain Fouilland et Germain Poullot.

À l'heure d'écrire ces remerciements, nous réalisons davantage à quel point la thèse, bien que faite de moments de solitude, est remplie de rencontres stimulantes, rendant l'expérience transformatrice et enrichissante. Merci à tous ceux qui y ont contribué !



Sommaire

REMERCIEMENTS	5
SOMMAIRE	9
INTRODUCTION GENERALE	13
CHAPITRE 1 : COMPRENDRE LES USAGES DES OUTILS D'IA DANS LE SYSTEME DE SOINS	21
1. UNE HISTOIRE DE L'INTELLIGENCE ARTIFICIELLE.....	21
1.1. <i>Évolution de la perception de l'intelligence</i>	24
1.1.1. Des machines et des humains	24
1.1.2. L'intelligence comme la capacité à résoudre des problèmes	25
1.1.3. L'intelligence comme la capacité à s'adapter.....	25
1.1.4. L'intelligence par la construction de machines.....	26
1.2. <i>Approches symboliques</i>	27
1.2.1. Premières approches symboliques	27
1.2.2. Jusqu'aux systèmes experts	30
1.2.3. Limites des approches symboliques	32
1.3. <i>Approches d'apprentissage machine</i>	34
1.3.1. Début des approches d'apprentissage machine	34
1.3.2. Différentes formes d'apprentissage machine	35
1.3.3. Produire des outils d'IA, un besoin dévorant de données.....	37
1.3.4. Quelle est la valeur des outils d'IA ?	38
2. APPORTS DES OUTILS D'IA DANS LE CONTEXTE DU SYSTEME DE SOINS	41
2.1. <i>Des outils d'IA pour révolutionner le système de soins ?</i>	41
2.1.1. Des outils d'IA capables de changer la manière de produire ?.....	41
2.1.2. Spécificités de la production du système de soins	43
2.1.3. La place des outils d'IA dans la production de soins	45
2.2. <i>Quatre types d'usage des outils d'IA dans le système de soins</i>	46
2.2.1. Les outils d'IA pour des tâches administratives.....	47
2.2.2. Les outils d'IA au cœur des soins	50
2.2.3. Les outils d'IA dans la recherche médicale	56
2.2.4. Les outils d'IA et les patients	58
CHAPITRE 2 : QUELS MODES DE GESTION POUR LA COLLABORATION ENTRE HUMAINS ET OUTILS D'IA ? 61	61
1. UN MOMENT CHARNIERE POUR COMPRENDRE LES COLLABORATIONS ENTRE HUMAINS ET OUTILS D'IA	61
1.1. <i>Évolution des connaissances sur les outils d'IA</i>	61
1.1.1. Des développements techniques à l'utilisation des outils d'IA	61
1.1.2. Positionnement en sciences de gestion.....	62
1.2. <i>Une première perspective sur la gestion algorithmique</i>	64
1.2.1. La politique de la gestion automatisée	65
1.2.2. Des applications contestées.....	68
1.2.3. La gouvernementalité ou l'illustration de l'apport des études sur les outils d'IA	69
1.3. <i>Le passage vers une perspective sur la collaboration humain-machine</i>	71
1.3.1. Le débat à l'origine de la perspective de la collaboration humain-outil d'IA	71
1.3.2. Le besoin de recherches spécifiques à la collaboration entre humains et outils d'IA.....	74
2. DEUX LITTÉRATURES CONVERGEANT VERS LES MODES DE GESTION DE LA COLLABORATION HUMAINS-OUTILS D'IA.....	77
2.1. <i>Les modèles conceptuels de collaboration humain-machine</i>	78

2.1.1.	Les Plateformes Numériques Hybrides Humain-IA.....	79
2.1.2.	La Collaboration Humain-IA sous l'angle de l'Intelligence	80
2.1.3.	Systèmes d'Intelligence Hybrides Humain-IA ou Intelligence Collaborative	81
2.1.4.	Coordonner l'Apprentissage Humain et l'Apprentissage Machine	82
2.1.5.	Symbiose Humain-IA dans la Prise de Décisions	83
2.1.6.	Humain dans la Boucle.....	84
2.1.7.	Épistémologie Hybride.....	85
2.1.8.	Augmentation de l'Intelligence.....	86
2.1.9.	Un Tournant post cognitiviste dans l'Augmentation de l'Intelligence	87
2.1.10.	Besoin d'observer les collaborations humain-machine en action	90
2.2.	<i>Les usages des outils d'IA.....</i>	<i>91</i>
2.2.1.	Caractéristiques des outils d'IA.....	92
2.2.2.	Usages développés en réponse aux caractéristiques des outils d'IA.....	99
2.2.3.	Prismes conceptuels pour étudier les usages des outils d'IA	107
3.	DEVELOPPER EMPIRIQUEMENT DES MODES DE GESTION POUR UNE COLLABORATION ENTRE L'EXPERT METIER ET L'OUTIL D'IA	117
CHAPITRE 3 : DEMARCHE DE NOTRE RECHERCHE		121
1.	CHOIX DE L'OBJET DE RECHERCHE ET DES PRINCIPES METHODOLOGIQUES.....	121
1.1.	<i>Choix et présentation de l'objet de recherche.....</i>	<i>121</i>
1.1.1.	Position du chercheur.....	121
1.1.2.	Position du laboratoire.....	122
1.2.	<i>Le défi méthodologique de la thèse.....</i>	<i>124</i>
1.2.1.	Le besoin d'une démarche compréhensive pour étudier les usages des outils d'IA.....	124
1.2.2.	Choisir les études de cas pour théoriser un phénomène	127
1.2.3.	Les tactiques de triangulation.....	130
2.	CONSTRUCTION DES ETUDES DE CAS	133
2.1.	<i>Construction de l'étude de cas RADO : usage des outils d'IA dans les soins courants.....</i>	<i>133</i>
2.1.1.	Accès au terrain	133
2.1.2.	Observation par shadowing.....	134
2.1.3.	Données complémentaires.....	137
2.1.4.	Analyse des données	139
2.2.	<i>Construction de l'étude de cas sur le projet KOVAK : usage des outils d'IA dans la recherche médicale</i>	<i>145</i>
2.2.1.	Les projets WILSON et CALI et leurs enseignements.....	145
2.2.2.	Collecte des données du projet KOVAK	146
2.2.3.	Analyse des données	148
2.3.	<i>Synthèse générale des études de cas.....</i>	<i>150</i>
2.3.1.	De quoi mes cas sont-ils le cas ?	150
2.3.2.	De quoi mes cas sont-ils composés ?.....	151
2.3.3.	Que peuvent produire nos cas ?	152
CHAPITRE 4 : USAGES DES OUTILS D'IA DANS LES SOINS COURANTS		155
1.	USAGES DES OUTILS D'IA DANS L'ANALYSE MAMMOGRAPHIQUE	155
1.1.	<i>Gestion de la prise en charge au niveau du service</i>	<i>155</i>
1.1.1.	Parcours de soins dans le dépistage du cancer du sein	155
1.1.2.	Organisation du service pour la prise en charge	157
1.2.	<i>Collaboration entre le radiologue et l'outil d'IA.....</i>	<i>159</i>
1.2.1.	Prise en main de l'outil d'IA	159

1.2.2.	Diagnostic assisté par l'outil d'IA	162
1.3.	<i>Pratiques interrogatives mises en place</i>	164
1.3.1.	Pratiques interrogatives professionnelles.....	164
1.3.2.	Pratiques interrogatives sur l'outil.....	166
1.3.3.	Limites à comprendre l'augmentation de l'expert métier	167
2.	COMPRENDRE L'OUTIL D'IA COMME UN OUTIL DE GESTION	170
2.1.	<i>Outil d'IA en tant qu'outil de gestion</i>	170
2.1.1.	Le concept d'outil de gestion	170
2.1.2.	Application du concept d'outil de gestion à l'outil d'IA	172
2.2.	<i>Repenser l'usage de l'outil d'IA</i>	174
2.2.1.	Le courant de l'instrumentation située	174
2.2.2.	L'augmentation de l'expert à travers le cadre de conformation/exploration	175
2.2.3.	Articuler les pratiques interrogatives avec le rôle exploratoire de l'outil d'IA.....	180
CHAPITRE 5 : USAGES DES OUTILS D'IA DANS LA RECHERCHE MEDICALE		183
1.	DEVELOPPEMENT D'UN OUTIL D'IA POUR REPONDRE A UNE QUESTION DE RECHERCHE MEDICALE	183
1.1.	<i>Narration du projet KOVAK</i>	183
1.1.1.	Analepse : origines du projet KOVAK	184
1.1.2.	Compréhension des données	186
1.1.4.	Préparation des données	188
1.1.5.	Modélisation	190
1.1.6.	Évaluation	194
1.1.7.	Déploiement.....	196
1.2.	<i>Usage de l'outil d'IA dans la construction de connaissances</i>	198
1.2.1.	Mise en place des pratiques interrogatives	198
1.2.2.	Type d'augmentation procuré par l'usage de l'outil d'IA	203
1.2.3.	Articulation entre pratiques interrogatives et conformation/exploration	206
2.	ROLE DE L'OUTIL D'IA DANS LA CONSTRUCTION DE CONNAISSANCES	210
2.1.	<i>Le processus d'enquête selon Charles S. Peirce</i>	210
2.1.1.	Épistémologie peircienne : inférences abductives, inductives et déductives	213
2.1.2.	Les outils d'IA comme des quasi-esprits.....	216
2.2.	<i>La place de l'outil d'IA dans l'enquête pragmatiste</i>	220
2.2.1.	L'outil d'IA comme expérimentateur	220
2.2.2.	Vérification de l'expérience par le Data Scientist.....	221
2.2.3.	Validation de l'expérience par l'expert métier	223
2.2.4.	Relecture du projet KOVAK selon l'enquête pragmatiste	226
CHAPITRE 6 : L'EXPERT-MACHINE COLLABORATION COMMUNITY POUR DEPASSER LES MODELES DE COLLABORATION HUMAIN-IA.....		229
1.	SYNTHESE DES RESULTATS DES ETUDES DE CAS.....	229
1.1.	<i>Discussion entre les résultats des deux études de cas</i>	229
1.1.1.	Apports et réorientations du cadre conformation/exploration.....	230
1.1.2.	Les études de cas à l'aune des inférences peirciennes	231
1.2.	<i>Typologie des pratiques interrogatives</i>	234
1.2.1.	Les pratiques interrogatives professionnelles	234
1.2.2.	Les pratiques interrogatives sur l'outil.....	237
1.3.	<i>Les configurations des pratiques interrogatives, modes de gestion de la collaboration</i>	241
1.3.1.	Mode de gestion par normalisation	242
1.3.2.	Mode de gestion par approfondissement	245

1.3.3.	Mode de gestion par opposition.....	246
2.	L'EXPERT MACHINE COLLABORATION COMMUNITY, UNE COLLABORATION AI NIVEAU DE LA COMMUNAUTE D'ENQUETE 249	
2.1.	<i>Considérer l'outil d'IA comme un acteur</i>	249
2.1.1.	Extension de la notion de pratiques interrogatives.....	249
2.1.2.	Voies de recherche pour de nouvelles pratiques interrogatives.....	250
2.2.	<i>Construction de l'EMC2</i>	252
2.2.1.	Modes de gestion composant le modèle de l'EMC2.....	252
2.2.2.	L'EMC2 dans la communauté d'enquête.....	254
2.3.	<i>Positionnement théorique de l'EMC2</i>	255
2.3.1.	L'EMC2 comme modèle de collaboration humain-machine.....	256
2.3.2.	L'enquête pragmatiste de Peirce et la compréhension des usages des outils d'IA.....	257
2.4.	<i>Deux leviers pour mettre en place de l'EMC2</i>	259
2.4.1.	Promouvoir les modes de gestion de l'EMC2 auprès des experts métier.....	259
2.4.2.	Concevoir autrement les outils d'IA.....	261
	CONCLUSION	263
	GLOSSAIRE	267
	LISTE DES FIGURES	270
	LISTE DES TABLEAUX	271
	ANNEXES	272
	ANNEXE 1 : PRESENTATION DE L'ALGORITHME MYCIN.....	272
	ANNEXE 2 : ÉVOLUTION DES PRISMES THEORIQUES POUR COMPRENDRE LES USAGES DE LA TECHNOLOGIE.....	275
	<i>De la théorie de la contingence à des approches plus constructivistes en sciences de gestion</i>	275
	<i>L'émergence des études des sciences et des techniques</i>	279
	ANNEXE 3 : ENTRETIENS DE RADO.....	281
	<i>Entretiens avec les radiologues</i>	281
	<i>Entretiens avec les concepteurs d'outils d'IA</i>	282
	ANNEXE 4 : CODAGE DU MATERIAU DE RADO.....	283
	ANNEXE 5 : ENTRETIENS LIES AU PROJET KOVAK.....	284
	ANNEXE 6 : CODAGE DU MATERIAU DU PROJET KOVAK.....	285
	ANNEXE 7 : MODELE CRISP-DM.....	286
	ANNEXE 8 : LA SEMIOTIQUE TRIADIQUE ET L'INTERACTION AVEC LE MONDE.....	287
	ANNEXE 9 : ÉLÉMENTS BIOGRAPHIQUES DE CHARLES S. PEIRCE.....	289
	BIBLIOGRAPHIE	293

Introduction générale

En 2020, au début de cette thèse, l'Intelligence Artificielle (IA) était déjà un sujet brûlant, touchant l'ensemble de la société. Le champ scientifique perfectionnait des méthodes d'apprentissage machine*, c'est-à-dire des méthodes d'analyse de données qui automatisent la création de modèles statistiques, dont l'apprentissage profond* (un type d'apprentissage machine fondé sur des réseaux très denses de « neurones » artificiels, permettant de saisir des structures complexes dans les données).

La particularité de cette technologie réside dans le paradoxe suivant : les outils d'IA mis au point rivalisent avec l'expertise humaine sur de nombreuses tâches, comme la détection d'objets dans des images ; pourtant, la manière exacte dont les résultats sont obtenus reste opaque, aussi bien pour les concepteurs que pour les utilisateurs (Burrell 2016; Pearl et Mackenzie 2018, Kellogg, Valentine, et Christin 2019; Stohl et al. 2016).

L'industrie lance des projets tous azimuts avec des investissements colossaux. L'enseignement forme des data scientists* en masse. La réglementation s'efforce d'encadrer les usages et limiter les débordements. Pendant ce temps, la presse traque chaque nouvelle évolution, et l'opinion publique oscille entre attentes et craintes.

Plus récemment, une application de traitement du langage accessible au grand public, ChatGPT, est apparue et a renforcé ce tumulte. ChatGPT fait partie de l'IA dite « générative », et plus spécifiquement des modèles LLM pour *Large Language Models*. Les outils d'IA génératifs sont entraînés dans le but de produire de nouvelles données, comme des textes, des images ou de la musique ayant une vraisemblance avec les données d'entraînement. En particulier, les LLM sont spécialisés dans le traitement du langage naturel, destinés par exemple à la traduction automatique, la génération de texte, la réponse aux questions ou au résumé de textes. Ils utilisent des méthodes d'apprentissage profond pour apprendre à partir de grands ensembles de données textuelles.

Notre recherche ne traite pas directement de ChatGPT et des outils d'IA générative. Elle porte sur des outils d'IA d'apprentissage machine plus « classiques », qui ont eu davantage de temps pour s'intégrer dans les organisations et les activités des professionnels.

Ces outils d'IA utilisent de l'apprentissage machine, et notamment de l'apprentissage profond, pour trouver une relation entre des entrées et des sorties spécifiques. Les tâches réalisées comme la classification, la régression, et la segmentation s'appliquent alors à la reconnaissance d'images, la recommandation de produits ou la mesure de scores de risque. L'évaluation de la qualité de l'outil se fait généralement en mesurant la précision, le rappel, et d'autres métriques statistiques sur des données de test.

Nous pouvons cependant généraliser les résultats de notre thèse aux outils d'IA fondés sur les modèles LLM tel que Chat GPT. En effet, au-delà de la qualité intrinsèque de l'outil, se pose une question d'usage, à savoir comment l'outil IA s'intègre dans l'organisation des activités. Au cœur de cette question d'usage se trouve une question encore plus centrale : celle de la collaboration entre l'humain et la machine.

L'usage des outils d'IA peut se caractériser comme une collaboration en raison des hautes performances de cette technologie qui provoquent un changement de perspective : de simples outils, ils gagnent le statut de partenaires participant à la création de valeur (Brachten et al., 2020; Seeber et al., 2020; Mirbabaie et al. 2021; Anthony, Bechky, et Fayard 2023). L'humain et la machine travaillent donc ensemble pour atteindre des objectifs, communs ou complémentaires, dans le déroulé de l'activité.

Les considérations à prendre en compte sont alors nombreuses. Certaines portent sur la manière de répartir les rôles et la coordination à mettre en place. Selon le rôle envisagé de l'outil d'IA, une division des tâches peut s'établir, affirmant un remplacement plus ou moins marqué du travail des acteurs humains, ou à l'inverse, une assistance augmentant leurs capacités d'agir. D'autres considérations portent sur l'éthique et la réglementation avec l'attribution des responsabilités des décisions et actions prises. Selon le type de collaboration, cette responsabilité varie. De même, nous pouvons considérer les moyens par lesquels le processus de collaboration est évalué et amélioré. Nous réunissons l'ensemble de ces considérations sous le terme de « modes de gestion » de la collaboration humain-machine.

Le point cardinal de ces modes de gestion concerne le paradoxe de l'usage des outils d'IA, à savoir l'opacité qui entoure les résultats de cette technologie dans un contexte de collaboration avec des experts métier. La complexité du fonctionnement des outils d'IA rend en effet les résultats difficiles à saisir, tant pour les experts métier que pour les concepteurs (Burrell 2016; Pearl et Mackenzie 2018, Kellogg, Valentine, et Christin 2019; Stohl et al. 2016). L'adoption de ces nouveaux outils peut dérouter les acteurs (Boyd et Crawford 2012), les conduire à ignorer les résultats (Pachidi et al. 2021; Lebovitz, Lifshitz-Assaf, et Levina 2022; Jussupow, Spohrer, et Heinzl 2022) ou encore à inventer de fausses explications (Dourish, 2016, Anthony 2021; Borch et Min 2022; Elish et Watkins 2020).

La collaboration humain-machine pose donc une multitude de problèmes dans l'organisation de l'action collective. Cette thèse s'interroge sur les modes de gestion mis en œuvre pour faire face à une collaboration avec l'expertise métier, un type de collaboration particulièrement saillant dans le contexte du système de soins.

Le cas spécifique du système de soins fait ressortir la forte responsabilité professionnelle, pour des raisons à la fois légales et morales (Heimer 1999; Kellogg 2009). La prise de décision médicale affecte en effet directement la vie des patients. Or, cette décision doit être prise

malgré un fort niveau d'incertitude, inhérent à la pratique médicale. L'évolution des besoins, en raison notamment de la part grandissante des maladies chroniques, incite par ailleurs les professionnels à multiplier ces décisions.

Dans ce contexte, les outils d'IA suscitent encore de nombreuses peurs, comme la déshumanisation de la prise en charge, et de nombreux espoirs, comme une plus grande personnalisation des soins à moindre coût. La collaboration humain-machine semble devoir se faire au bénéfice des experts métier en augmentant leurs capacités d'agir.

Précisément, des études récentes montrent que cette collaboration au bénéfice des experts métier existe déjà. Notamment, certains travaux de recherche décrivent comment des médecins arrivent à développer des « métacognitions », des processus cognitifs de second ordre pour surveiller et contrôler le raisonnement, afin d'intégrer les résultats des outils d'IA dans leur réflexion (Jussupow, et al. 2021; Lebovitz, Lifshitz-Assaf, et Levina 2022). Ils évaluent ainsi la pertinence d'incorporer les résultats des outils d'IA à leurs décisions (Jussupow, et al. 2021). Par exemple, dans l'ethnographie de Lebovitz, Lifshitz-Assaf, et Levina (2022), certains radiologues mettent en place des « pratiques interrogatives », c'est-à-dire des pratiques qui permettent de confronter les résultats des outils d'IA à leurs connaissances. La présence de telles pratiques traduit une « collaboration engagée », car les acteurs entrent en dialogue avec l'outil d'IA pour rendre moins tacite son expertise et arriver à une décision plus informée (Lebovitz, Lifshitz-Assaf, et Levina 2022).

Ces travaux apportent des premiers éléments de connaissance sur les modes de gestion mis en œuvre pour organiser cette collaboration. Néanmoins, ils laissent plusieurs questions ouvertes. Par exemple, il reste à saisir la manière dont les pratiques interrogatives contribuent à l'augmentation des capacités des experts métier. Nous souhaitons donc prolonger ces travaux en développant un mode de gestion de la collaboration humain-machine au service du professionnel. Dans ce cadre, le contexte du système de soins est un cas exemplaire nous aidant à aborder la question de recherche suivante : quels sont les modes de gestion appropriés pour une collaboration entre des experts métier et les outils d'IA ?

Afin de dépasser les discours prospectifs, nous avons mené deux études observationnelles de situations dans lesquelles les professionnels de santé utilisent des outils d'IA. Nos investigations de terrain ont produit deux études de cas. La première, dénommée RADO, concerne l'usage d'un outil d'IA pour l'analyse mammographique : des radiologues s'emparent d'un outil détectant les zones sur une image susceptible de représenter un cancer. La seconde étude de cas concerne le projet KOVAK : une équipe de recherche médicale utilise des outils d'IA dans le but de traiter les données d'une cohorte de patients.

Dans nos deux études de cas, les acteurs s'engagent dans une collaboration avec l'outil d'IA en exerçant des pratiques interrogatives. L'analyse de la première étude de cas nous amène à

considérer les outils d'IA observés comme des outils de gestion, en utilisant le cadre d'analyse conformation/exploration (Moisdon 2005a). Issu du courant de l'instrumentation dans la littérature sur les outils de gestion, ce cadre théorique explique comment les outils de gestion ont une double dimension : d'une part, ils normalisent et optimisent une activité, et d'autre part, ils ouvrent d'autres possibilités d'action en étant source d'apprentissage.

Avec ce cadre, nous constatons qu'en pratique, l'usage de l'outil d'IA augmente les capacités des experts métier dans le registre de la conformation. Il les aide à améliorer leur activité, mais les apprentissages apparaissent limités. Au niveau théorique, ces résultats contribuent à caractériser la manière dont les outils d'IA augmentent les capacités d'agir des experts métier. Plus précisément, nous identifions que les augmentations dans le registre de la conformation et de l'exploration dépendent des modes de gestion établis sur les configurations des pratiques interrogatives. C'est la manière d'assembler différentes pratiques interrogatives qui pousse à la conformation de l'activité ou à des apprentissages.

Grâce à cette analyse et aux données du second cas, KOVAK, nous caractérisons alors trois types de modes de gestion majeurs. Cependant, dans ce contexte de recherche médicale, nous observons aussi que l'usage de l'outil d'IA rapproche davantage ce dernier d'un collaborateur que d'un simple outil (Anthony, Bechky, et Fayard 2023). Pour éclairer cette observation, nous mobilisons le concept d'enquête pragmatiste de Peirce (1877; 1878a; 1878b; 1878c; 1878d; 1878e; 1887).

Nous remarquons que la puissance explicative de la vision peircienne englobe le cadre d'analyse conformation/exploration. Elle l'explique tout en proposant une interprétation plus large, qui nous amène à étendre la notion de pratiques interrogatives en considérant l'outil d'IA comme un partenaire du processus de construction de connaissances. Nous développons ainsi plus en profondeur les modes de gestion dans la collaboration entre experts métier et outils d'IA pour aboutir au modèle de l'*Expert Machine Collaborative Community* (EMC2). La vision peircienne nous offre aussi l'opportunité de questionner l'intégration des acteurs dans une approche collective de la collaboration humain-IA.

Ce modèle s'applique également à des outils d'IA comme ChatGPT. Nous avons nous-mêmes bénéficié des résultats de cette thèse au moment de la rédiger en développant une collaboration avec cet outil d'IA. En l'interrogeant, notre écriture a pu être rendue plus fluide. Concernant le fond, lorsque cet outil est apparu, nous étions sceptique, et quelques tentatives infructueuses ont renforcé cet avis. Par construction, l'IA générative ne fait que produire du vraisemblable en fonction de choses existantes. Sur le sujet de notre thèse, l'outil tend à se faire l'écho de propos maintes fois entendus, incluant au passage les plus ineptes, mais il est incapable de développer un sujet précis, et encore moins de produire une nouvelle idée.

En revanche, armé du modèle EMC2, nous avons su mettre en place des usages sur la dimension éditoriale. La compréhension technique de l'IA générative nous sert de correcteur perfectionné. Il corrige des erreurs évidentes comme les mots manquants, mal orthographiés ou incorrects, ce qui allège la charge mentale de relecture. Des pratiques interrogatives sont nécessaires pour écarter les suggestions inappropriées et prendre en compte les suggestions de suppression de certains passages. Il offre une amélioration incrémentale du texte, optimisant notre travail sans en changer la nature.

Grâce à nos réflexions théoriques, nous avons su développer un usage plus exploratoire de ChatGPT. L'idée principale est de bénéficier des nombreux textes ayant servi à son entraînement : ces textes sont vus comme des structures narratives pouvant enrichir notre récit. Cet usage est venu à une phase tardive de la rédaction. Il nous a simplement aidés à trouver une solution pour rendre notre terrain plus accessible aux lecteurs. En effet, des commentaires sur des versions intermédiaires ont souligné la difficulté à comprendre certains termes techniques issus du domaine médical ou des sciences des données. À l'issue des échanges avec ChatGPT, nous avons choisi d'ajouter un glossaire contenant ces termes (identifiés par un astérisque en début de chapitre). Bien qu'il ne s'agisse que d'une amélioration ponctuelle, cet apport d'une autre expertise enrichit notre travail. Nous avons appris et transformé notre pratique d'écriture. Dans notre conclusion, nous revenons sur cet exemple à la lumière du modèle EMC2, en espérant qu'il puisse inspirer d'autres lecteurs à se saisir et faire bon usage de ces technologies.

Structuration du manuscrit

Le chapitre 1 introduit les technologies des outils d'IA en adoptant une approche historique de leur développement. Cette description aboutit au constat de l'enjeu des usages des outils d'IA. L'examen des usages actuels des outils d'IA dans le système de soins, notamment dans les soins courants et la recherche médicale, s'inscrit naturellement comme la seconde partie de ce chapitre.

Le chapitre 2 fait l'état de l'art des connaissances sur la collaboration entre humains et outils d'IA. Nous commençons par expliquer pourquoi ce sujet est d'actualité, et comment les sciences de gestion sont appropriées pour y répondre. Nous décrivons ensuite une première littérature interdisciplinaire, présentant une variété de modèles de collaboration humain-IA, parfois trop abstraits, ce qui nous conduit à une littérature plus empirique sur les usages des outils d'IA. En combinant les connaissances de ces deux corpus, nous constatons le besoin de poursuivre les efforts pour établir des modes de gestion de la collaboration entre outils d'IA et experts métier.

Le chapitre 3 décrit la démarche compréhensive adoptée et comment cette méthodologie répond aux spécificités du sujet traité. Nous présentons d'abord l'objet de recherche, notre

position de chercheur, et les principes méthodologiques choisis pour aborder notre problème. Ensuite, nous décrivons notre démarche et plus particulièrement comment nos deux études de cas ont été construites.

Le chapitre 4 explore les usages des outils d'IA dans les soins courants (étude de cas RADO). Après une description du terrain – deux services de radiologie diagnostiquant le cancer du sein – nous concentrons notre analyse sur la collaboration entre les radiologues et l'outil d'IA dans l'interprétation des mammographies. En mobilisant la notion de pratiques interrogatives, nous examinons l'outil d'IA comme un outil de gestion pour caractériser les améliorations de son usage avec le cadre de conformation/exploration (Moisdon 2005a).

Le chapitre 5 illustre les usages des outils d'IA dans la recherche médicale (étude de cas KOVAK). A travers l'observation de l'intégration des outils d'apprentissage machine pour l'exploitation des données d'une cohorte, nous observons l'usage de l'outil d'IA par une équipe de médecins et de data scientists afin de construire des connaissances médicales. Le concept d'enquête pragmatiste de Peirce (1877; 1878a; 1878b; 1878c; 1878d; 1878e; 1887) éclaire particulièrement bien ce phénomène et caractérise l'usage de l'outil d'IA comme un processus d'inférence inductive qui s'articule et se complète avec les efforts d'autres acteurs.

À partir de nos résultats, le chapitre 6 propose une synthèse des deux études de cas, où nous faisons une typologie des pratiques interrogatives identifiées, et montrons comment elles peuvent s'articuler pour former des modes de gestion améliorant l'efficacité de la collaboration. De plus, la vision peircienne apparaît comme particulièrement pertinente pour interpréter l'ensemble des résultats et développer le concept de pratiques interrogatives en considérant l'outil d'IA comme un participant à part entière.

Ces déductions se concrétisent dans l'EMC2, notre modèle théorique final, que nous situons dans la littérature sur les collaborations humain-IA. Cela nous permet de démontrer l'intérêt de la vision peircienne pour l'étude des usages des outils d'IA. Nous concluons le chapitre par des considérations managériales pour implémenter l'EMC2 sur le terrain.

En raison de la nature technique du phénomène étudié, au croisement de l'Intelligence Artificielle et du système de soins, les termes marqués d'un astérisque (au début de chaque chapitre) sont répertoriés dans un glossaire en fin de document.

De même, plusieurs annexes contextualisent des points complexes, comme l'illustration des systèmes experts par l'algorithme MYCIN (annexe 1), le contexte historique ayant abouti aux prismes théoriques mobilisés dans les analyses des usages des technologies présentés dans le chapitre 2 (annexe 2), une illustration du modèle CRISP-DM mobilisé dans la narration du chapitre 5 (annexe 7) ainsi que la sémiotique triadique qui sous-tend l'ensemble de l'œuvre de Peirce (annexe 8) et une courte biographie de ce penseur pour mieux contextualiser son œuvre (annexe 9).

CHAPITRE 1 : COMPRENDRE LES USAGES DES OUTILS D'IA DANS LE SYSTEME DE SOINS

1. UNE HISTOIRE DE L'INTELLIGENCE ARTIFICIELLE

- 1.1. Évolution de la perception de l'intelligence
- 1.2. Approches symboliques
- 1.3. Approches d'apprentissage machine

2. APPORTS DES OUTILS D'IA DANS LE CONTEXTE DU SYSTEME DE SOINS

- 2.1. Des outils d'IA pour révolutionner le système de soins ?
- 2.2. Quatre types d'usage des outils d'IA dans le système de soins

CHAPITRE 2 : QUELS MODES DE GESTION POUR LA COLLABORATION ENTRE HUMAINS ET OUTILS D'IA ?

1. UN MOMENT CHARNIERE POUR COMPRENDRE LES COLLABORATIONS ENTRE HUMAINS ET OUTILS D'IA

- 1.1. Évolution des connaissances sur les outils d'IA
- 1.2. Une première perspective sur la gestion algorithmique
- 1.3. Le passage vers une perspective sur la collaboration humain-machine

2. DEUX LITTERATURES CONVERGEANT VERS LES MODES DE GESTION DE LA COLLABORATION HUMAINS-OUTILS D'IA

- 2.1. Les modèles de collaboration humain-machine
- 2.2. Les usages des outils d'IA

3. DEVELOPPER EMPIRIQUEMENT DES MODES DE GESTION POUR UNE COLLABORATION ENTRE L'EXPERT METIER ET L'OUTIL D'IA Quels sont les modes de gestion appropriés pour une collaboration entre des experts métier et les outils d'IA ?

CHAPITRE 3 : DÉMARCHÉ DE NOTRE RECHERCHE

1. CHOIX DE L'OBJET DE RECHERCHE ET DES PRINCIPES METHODOLOGIQUES

- 1.1. Choix et présentation de l'objet de recherche
- 1.2. Le défi méthodologique de la thèse

2. CONSTRUCTION DES ETUDES DE CAS

- 2.1. Construction de l'étude de cas RADO
- 2.2. Construction de l'étude de cas sur le projet KOVAK
- 2.3. Synthèse générale des études de cas

CHAPITRE 4 : USAGES DES OUTILS D'IA DANS LES SOINS COURANTS

1. USAGES DES OUTILS D'IA DANS L'ANALYSE MAMMOGRAPHIQUE

- 1.1. Gestion de la prise en charge au niveau du service
- 1.2. Collaboration entre le radiologue et l'outil d'IA
- 1.3. Pratiques interrogatives mises en place

2. COMPRENDRE L'OUTIL D'IA COMME UN OUTIL DE GESTION

- 2.1. Outil d'IA en tant qu'outil de gestion
- 2.2. Repenser l'usage de l'outil d'IA

CHAPITRE 5 : USAGES DES OUTILS D'IA DANS LA RECHERCHE MEDICALE

1. DEVELOPPEMENT D'UN OUTIL D'IA POUR REpondre A UNE QUESTION DE RECHERCHE MEDICALE

- 1.1. Narration du projet KOVAK
- 1.2. Usage de l'outil d'IA dans la construction de connaissances

2. ROLE DE L'OUTIL D'IA DANS LA CONSTRUCTION DE CONNAISSANCES

- 2.1. Le processus d'enquête selon Charles S. Peirce
- 2.2. La place de l'outil d'IA dans l'enquête pragmatiste

CHAPITRE 6 : L'EXPERT-MACHINE COLLABORATION COMMUNITY

1. SYNTHESE DES RESULTATS DES ETUDES DE CAS

- 1.1. Discussion entre les résultats des deux études de cas
- 1.2. Typologie des pratiques interrogatives
- 1.3. Les configurations des pratiques interrogatives

2. L'EXPERT MACHINE COLLABORATION COMMUNITY

- 2.1. Considérer l'outil d'IA comme un acteur
- 2.2. Construction de l'EMC2
- 2.3. Positionnement théorique de l'EMC2
- 2.4. Deux leviers pour mettre en place de l'EMC2

Figure 1 : Plan de thèse



Chapitre 1 : Comprendre les usages des outils d'IA dans le système de soins

L'Intelligence Artificielle (IA) est considérée comme une technologie à usage général, car elle s'adapte à un large éventail de tâches et de domaines. Au travers d'une rétrospective de son développement (1), nous mettons en avant son évolution constante et la diversité de techniques et de paradigmes qui se cachent derrière ce terme.

Ensuite, nous illustrons comment les outils d'IA (l'apprentissage machine* et l'apprentissage profond*) sont adaptés pour améliorer des processus existants ou développer de nouvelles opportunités dans le système de soins (2). Ce contexte spécial génère des questions quant à l'usage des outils d'IA, en particulier l'apport de ces outils pour des experts métier.

1. Une histoire de l'Intelligence Artificielle

Nous présentons ici *une* histoire de l'Intelligence Artificielle. Il est important de souligner que de diverses narrations sont possibles, selon le message que l'on souhaite véhiculer (Dick 2019). Par exemple, Stephanie Dick (2019) retrace les changements de perception sur ce qu'est l'intelligence et les conséquences pour les travailleurs humains lorsque les outils d'IA participent à l'action collective. Elle illustre comment l'IA « *sert de puissant levier pour révéler des hypothèses sur l'être humain* »¹(Suchman 2007, 266).

D'un autre côté, avec *Apocalypse Managériale* (2022), Vaujany se penche sur la rencontre entre le mouvement cybernétique* et la montée du management aux États-Unis, montrant comment l'histoire de l'IA est liée à notre façon d'organiser le monde. Il explique comment la numérisation est un processus par lequel le monde est rendu traitable par un algorithme qui fragmente et réduit la réalité. Des entités sont divisées en tranches spatio-temporelles en faisant abstraction des éléments jugés peu importants (Babushkina et Votsis 2022).

Notre récit s'appuie sur quatre sources principales : *Artificial Intelligence, A Modern Approach* (Russell et Norvig 2021), *The Myth of Artificial Intelligence* (Larson 2021) et les deux susmentionnées. De nombreux événements historiques sont documentés dans la quatrième édition de l'ouvrage de Russell et Norvig, une référence pour les étudiants en informatique et

¹ [Artificial Intelligence] works as a powerful disclosing agent for assumptions about the human (Suchman, 2006, p. 266).

en data science. Larson (2021) propose une analyse de la recherche en IA établie sur les travaux de Charles S. Peirce. Cela le conduit à distinguer les « approches symboliques », principalement développées entre les années 1950 et 1990, et les « approches d'apprentissage machine », qui ont gagné en importance à partir des années 1990.

Les **approches symboliques** reposent sur la manipulation explicite de symboles et de règles formelles pour représenter et traiter les connaissances. Elles sont limitées par leur incapacité à gérer efficacement des tâches complexes nécessitant de prendre en compte une grande quantité de données ou de s'adapter à des environnements changeants.

Les **approches d'apprentissage machine** se concentrent sur des bases de données pour trouver des modèles statistiques (sans nécessiter une programmation explicite de règles) à partir desquels des décisions peuvent être prises. Elles permettent des avancées significatives dans des tâches variées comme la reconnaissance de formes et la traduction automatique.

Pour comprendre la pertinence de cette distinction, nous pouvons effectuer une rapide recherche sur Ngram Viewer². Elle nous montre l'évolution de la fréquence de l'expression « *Artificial Intelligence* » dans un corpus de textes anglais numérisés par Alphabet, couvrant la période de 1945 à 2019 (voir la figure 2).

Cette analyse confirme que le terme « Intelligence Artificielle » fait son apparition dans les écrits à partir de 1956, année considérée comme le point de départ officiel de la recherche en IA, avec l'atelier d'été de l'Université de Dartmouth. Toutefois, ce moment est le fruit de recherches antérieures dont nous retraçons les principaux courants (1.1).

La figure 2 révèle que l'IA connaît deux grandes vagues de popularité. La première, liée aux approches symboliques (1.2), fait référence à une famille de techniques qui s'appuient sur la manipulation explicite de symboles pour représenter des connaissances et réaliser des raisonnements. Les outils d'IA utilisent alors le raisonnement logique pour tirer de nouvelles connaissances à partir des connaissances existantes et résoudre des problèmes. Étant donné que les connaissances sont représentées sous forme de règles et de symboles explicites, il est facile de comprendre et d'expliquer comment l'outil est parvenu à une conclusion particulière.

L'apogée de cette approche se conclut avec le développement des systèmes experts*, suscitant un grand intérêt industriel dans les années 1980. Toutefois, des résultats mitigés ont

² Ngram Viewer est un outil en ligne proposé par Google pour explorer les tendances des mots ou phrases dans un corpus de livres numérisés dans Google Books. En entrant un mot ou une phrase, il génère un graphique montrant la fréquence de son utilisation au fil du temps.

entraîné un désenchantement général, provoquant un « hiver » pour l'IA, c'est-à-dire une période de ralentissement marqué de la recherche.

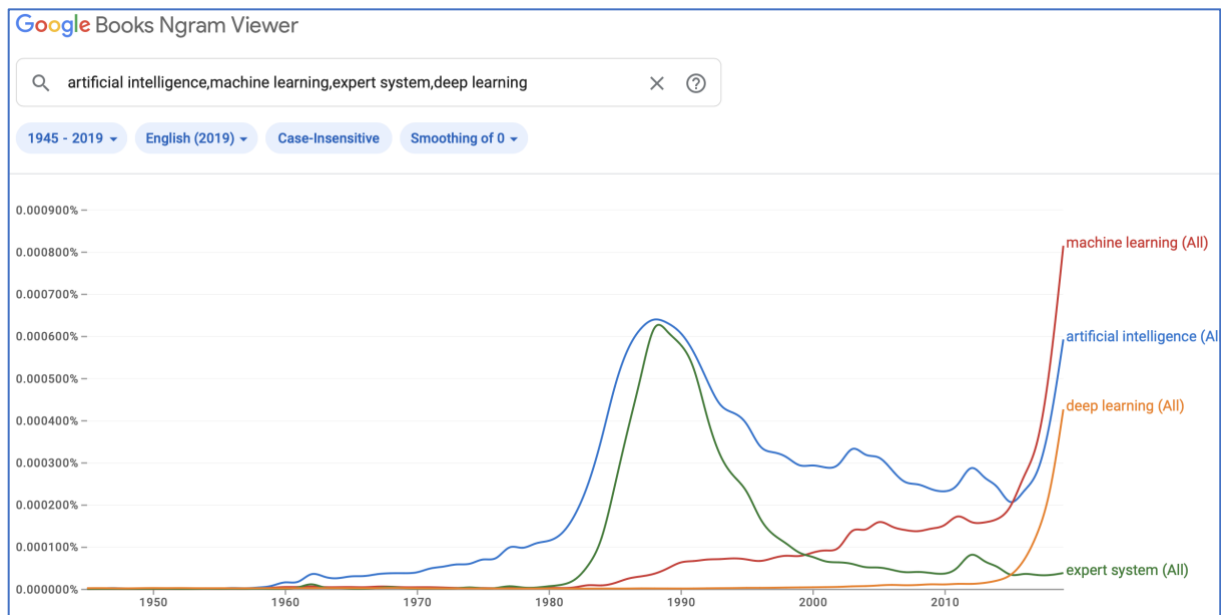


Figure 2 : Ngram Viewer sur les vagues liées à l'IA

La seconde grande vague en IA émerge dans les années 1990, avec les approches d'apprentissage machine ou connexionniste (1.3). Ces outils d'IA « apprennent » en identifiant des motifs, des structures ou des relations dans les données, ce qui leur permet de faire des prédictions ou de prendre des décisions.

Ces approches sont à l'origine des progrès récents dans de nombreux domaines de l'IA comme la reconnaissance vocale, la vision par ordinateur, la traduction automatique et le jeu. Ainsi, elles génèrent un nouvel engouement industriel, avec des investissements massifs³. De nouveaux outils d'IA se développent pour une multitude de cas d'usage, et nous commençons à prendre conscience de l'ampleur de leurs conséquences aujourd'hui. En effet, malgré leur grande flexibilité leur et puissance, pouvant faire émerger des modèles cachés dans les données, il n'en reste pas moins que ces outils présentent un manque de transparence et de compréhensibilité.

³ Par exemple, en 2020, le gouvernement des États-Unis investit environ 1,5 milliard de dollars dans la recherche et le développement de technologies non classifiées liées à l'IA (National Security Commission on Artificial Intelligence, 2021)

1.1.Évolution de la perception de l'intelligence

Avant de préciser la constitution du champ de recherche en IA (1.1.4), il est intéressant de retracer ses origines. Nous exposons les initiatives pour construire des machines et leur relation avec le travail humain (1.1.1), ainsi que les concepts fondamentaux de l'intelligence qui ont contribué à façonner la discipline, de la résolution de problèmes (1.1.2) à la capacité d'apprentissage (1.1.3).

1.1.1. Des machines et des humains

Selon Russell et Norvig (2021), l'histoire de l'IA trouve ses racines dans les contributions de savants ayant développé le concept d'intelligence et des machines capables d'accomplir des opérations cognitives. Parmi les premiers exemples figurent les travaux d'Aristote sur la démonstration et les syllogismes au IV^{ème} siècle av. J.C., l'*Ars Generalis Ultima* de Ramon Lulle qualifiée de « machine logique » au XIII^{ème} siècle, ou encore la « machine d'arithmétique » de Blaise Pascal, capable d'effectuer des calculs simples au XVII^{ème} siècle. Ainsi, dès ses débuts, l'histoire de l'Intelligence Artificielle est liée aux machines et repose sur des preuves empiriques visant à créer une forme d'intelligence semblable à celle des humains.

Cependant, la création de ces machines a un effet paradoxal sur la conception de l'intelligence elle-même. L'historienne des sciences Stephanie Dick (2019, 3) souligne comment « *ce qui est considéré comme de l'intelligence est une cible mouvante dans l'histoire de l'Intelligence Artificielle.* »⁴. Autrefois, les calculs complexes étaient considérés comme l'apanage des génies, mais Gaspard de Prony démocratise leur résolution en appliquant les principes de la division du travail d'Adam Smith. Ainsi, au XIX^e siècle, les calculs deviennent une simple mécanique réalisable par une main d'œuvre peu qualifiée. Un siècle plus tard, Alan M. Turing propose de remplacer les calculateurs, des ouvriers et ouvrières dédiés à faire des calculs, par des « machines de Turing » (Turing 1936) – les précurseurs des ordinateurs.

Selon Dick (2019), cet exemple montre que la reproduction mécanique d'une faculté mentale diminue la valeur sociale qui lui est attribuée. En extrapolant l'évolution des calculs complexes aux nombreuses applications que promettent les outils (reconnaissance vocale, identification d'objets, etc.), les craintes suscitées par ces outils semblent légitimes.

Les prémices du développement de l'IA se sont concentrées sur la compréhension de la nature de l'intelligence en construisant des machines capables de reproduire les capacités humaines. Cette mécanisation a conduit à dévaloriser les facultés qu'elle cherchait à

⁴ *what counts as intelligence is a moving target in the history of artificial intelligence*

reproduire. Malgré cela, la délégation de certaines tâches humaines à des machines offre des opportunités trop intéressantes pour que cette recherche s'arrête. Turing a joué un rôle crucial dans ces développements, en particulier pendant la Seconde Guerre mondiale, où ses travaux ont eu un impact significatif sur l'issue du conflit (Hodges 2014).

1.1.2. *L'intelligence comme la capacité à résoudre des problèmes*

En 1940, l'armée allemande avait perfectionné son système de communication grâce à la machine Enigma pour crypter ses transmissions radio. Pour décoder ces messages, les forces Alliées ont mis en place un important effort, y compris l'espionnage et la collecte de fragments d'Enigma parmi les débris des bâtiments ennemis. Cependant, même avec une compréhension du fonctionnement de la machine, le nombre de combinaisons possibles était astronomique, rendant le déchiffrement impossible pour que l'information puisse être utilisée à temps.

Les choses changent dans la base militaire de Bletchley Park. Les cryptologues reprennent les travaux du mathématicien polonais Rejewski sur une machine de décodage rebaptisée « Bombe ». Utilisée avec les « Turingismus », une adaptation de Turing des principes de la logique déductive pour réduire les combinaisons à tester, la Bombe trouve des résultats en moins d'un jour. Cela donne un avantage considérable aux forces Alliées et les aide à gagner la guerre.

Cette expérience démontre que les machines sont capables non seulement de remplacer, mais également de surpasser les êtres humains dans la réalisation de certaines tâches intellectuelles. Cela a conduit Turing à envisager que des machines suffisamment puissantes pourraient dépasser l'intuition humaine, créant ainsi une forme d'intelligence artificielle.

Turing a simplifié la notion d'intelligence en la réduisant à la capacité de résoudre des problèmes (Turing 1950), éliminant d'autres aspects tels que l'intelligence situationnelle ou sociale (Hodges 2014). Cette idée fondatrice façonne la recherche en IA, en se concentrant sur le déploiement d'outils d'IA « faibles » conçus pour résoudre des tâches spécifiques. De plus, le test de Turing permet aux chercheurs de se concentrer sur la manière de reproduire les résultats des capacités humaines, sans imiter leur processus de réflexion.

1.1.3. *L'intelligence comme la capacité à s'adapter*

Le mouvement cybernétique façonne une partie du champ de recherche en IA. Il émerge grâce aux conférences de Macy, tenues de 1942 à 1953. Ce mouvement interdisciplinaire se centre sur le concept clé de rétroaction* (ou feedback), formalisé dans *Cybernetics* (N. Wiener 1948).

Les recherches issues de ce mouvement ont influencé la recherche en intelligence artificielle en introduisant deux idées clés : l'intelligence en tant que capacité d'apprentissage et l'équivalence entre les systèmes organiques et les machines. En effet, la théorie du contrôle cybernétique s'applique aussi bien aux organismes vivants qu'aux machines.

Cette théorie a inspiré le connexionnisme*, une approche visant à imiter le fonctionnement mécanique du cerveau humain, en particulier celui des neurones. Les pionniers du connexionnisme, McCulloch et Pitts, ont proposé un modèle de neurone artificiel* fondé sur la physiologie : la logique propositionnelle et les machines de Turing (McCulloch et Pitts 1943). Ils ont théorisé que tout calcul pouvait être reproduit par un assemblage de ces neurones et ont émis l'hypothèse que ces réseaux de neurones étaient capables d'apprendre. Ces neurones artificiels peuvent être « allumés » ou « éteints ». Ils passent à l'état « allumé » lorsque suffisamment de neurones voisins sont déjà « allumés ». Quelques travaux ultérieurs poursuivent l'approche connexionniste, mais elle est délaissée, avant d'être reprise dans les années 1990.

Enfin, le mouvement cybernétique a une grande influence sur l'ensemble de la société et sur nos expériences quotidiennes (de Vaujany 2022). Ces transformations reposent sur l'application des principes des systèmes auto-organisés aux institutions humaines (N. Wiener 1948) grâce à l'informatique. Comme le décrit Vaujany, la *digitalisation* et la *managérialisation* ont engendré une instrumentation de gestion à une échelle inédite. Elle provoque alors une « sémiologie digitale », c'est-à-dire une transformation numérique qui affecte la manière dont nous créons, interagissons avec, et donnons un sens aux signes et aux symboles. Les outils d'IA participent ainsi à transformer la façon d'éprouver le monde.

1.1.4. *L'intelligence par la construction de machines*

L'idée d'intelligence comme capacité à résoudre des problèmes, ou comme capacité d'apprentissage, le besoin d'évaluer les résultats produits ou d'envisager les équivalences entre les systèmes biologiques et les machines, procurent des perspectives de recherche qui se cristallisent lors de l'émergence d'un champ scientifique dédié à la reproduction de l'intelligence humaine. Ce champ débute en 1956 à l'Université de Dartmouth, autour d'un atelier d'été de deux mois.

Malgré les différentes perspectives qui la composent, la discipline scientifique fait converger ces tentatives vers l'objectif de donner corps à l'idée de l'Intelligence Artificielle, c'est-à-dire de prouver par l'expérience que cette notion pouvait être matérialisée. Cette proposition est énoncée dans le projet de l'atelier (McCarthy et al. 2006, 12) : « *chaque aspect*

de l'apprentissage, ou de toute autre caractéristique de l'intelligence, peut en principe être décrit avec une telle précision qu'il est possible de le simuler à l'aide d'une machine »⁵.

Aucun résultat probant ne résulte de ces deux mois de travail intensif. Le travail réalisé sert cependant à établir les fondements pour les recherches ultérieures. Les années suivantes permettent d'accumuler des preuves empiriques destinées à répondre aux critiques à l'encontre de cette discipline. Les chercheurs démontrent que l'IA peut résoudre, bien que de manière encore rudimentaire, des énigmes, des jeux comme les échecs, des tests de QI, etc. Sa capacité est alors résumée par l'expression « *Regarde maman, sans les mains !* »⁶. Une tendance de fond apparaît avec l'exploration d'approches symboliques.

1.2. Approches symboliques

Les approches symboliques en IA modélisent des concepts du monde réel et leurs relations par un système de symboles et de règles. Par exemple, un outil d'IA médical peut contenir des informations sur les symptômes, les pathologies, les traitements et les protocoles médicaux. Grâce à une interface utilisateur, le médecin peut alors saisir les informations du patient pour que l'outil d'IA pose un diagnostic, une évaluation des risques ou des recommandations de traitement. Ce cas d'usage a inspiré le premier système expert : l'algorithme MYCIN (annexe 1).

Les premiers outils d'IA symboliques résolvaient des problèmes mathématiques (1.2.1). La mise en place d'heuristiques*, c'est-à-dire des règles adaptées des connaissances empiriques, a conduit aux systèmes experts, capables de répondre à des problèmes plus complexes (1.2.2). Cependant, les approches symboliques présentent des limites. L'ensemble de connaissances est fermé, sans une compréhension profonde des concepts, ce qui limite leur intérêt dans des domaines où les connaissances sont incertaines (1.2.3).

1.2.1. Premières approches symboliques

Un an avant l'atelier de Dartmouth, Newell et Simon entreprennent de travailler sur un algorithme pour effectuer un raisonnement automatisé, nommé le « Logic Theorist ». Simon a déjà publié *Administrative Behavior* (1947) où il a étudié le fonctionnement des bureaucraties et formulé sa théorie sur la rationalité limitée. Ses recherches l'amènent à s'intéresser à la résolution de problèmes et à la prise de décisions. À l'époque, il travaille à la

⁵ Every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it.

⁶ Look, Ma, no hands!

RAND Corporation, un laboratoire de réflexion au service de l'armée américaine, dont l'objectif est d'améliorer la politique et le processus décisionnel. Dans le cadre de ses travaux, Simon commence à considérer qu'une machine est capable de simuler la prise de décision en manipulant des symboles et par extension la pensée humaine.

Il rencontre alors Newell, un chercheur spécialisé dans la logistique et la théorie de l'organisation, qui vient d'assister à une présentation d'un nouvel algorithme, le « filtrage par motifs », préfigurant les algorithmes d'exploration de bases de données. Cette présentation éclaire Newell sur l'importance d'assembler des actions simples pour produire un effet global complexe. Enfin, Shaw, informaticien à la RAND Corporation, se joint à ce duo. Ensemble, ils entreprennent le développement du Logic Theorist.

Le Logic Theorist est un algorithme capable de prouver des théorèmes mathématiques. Pour traduire les problèmes dans un langage compréhensible par l'algorithme, Newell et Simon se réfèrent aux *Principia Mathematica* (1910-1913) de Bertrand Russell et Alfred North Whitehead. Ces trois volumes reprennent les connaissances mathématiques, exposant les théorèmes fondamentaux de la discipline avec une nouvelle notation appelée « logique symbolique ».

Ainsi, le Logic Theorist utilise un raisonnement en arborescence pour prouver un théorème. La racine de son raisonnement est fondée sur des cinq axiomes des *Principia Mathematica*, desquels découlent de nombreuses déductions. L'algorithme en emploie trois types : la *modus ponens* (ou détachement), la substitution de variables* et le remplacement de formules par leur définition. Le programme crée trois branches à partir de la racine, qui se subdivisent à leur tour, et ainsi de suite. Chaque branche est explorée par l'algorithme jusqu'à ce qu'il aboutisse au théorème à prouver. Le chemin permettant d'aller de la racine à la branche finale constitue une séquence logique d'énoncés servant de preuve.

Cependant, l'utilisation de ces trois types de déductions entraîne une croissance exponentielle des branches, rendant le processus de plus en plus long. Pour résoudre ce problème, Newell et Simon introduisent des « règles empiriques » ou « heuristiques » pour guider l'algorithme vers des chemins plus susceptibles de mener à la solution. Pour illustrer ce point : l'ouverture d'un coffre-fort à cadrans pourrait se faire en essayant toutes les combinaisons une par une, mais en écoutant le cliquetis des gâchettes se mettant en place, on peut trouver une solution bien plus rapide et qualifiée d'intelligente.

Le Logic Theorist est présenté à l'atelier de Dartmouth, suscitant un intérêt limité, même s'il s'agit du seul projet fonctionnel de l'atelier. L'algorithme est capable de prouver 38 des 52 théorèmes du chapitre 2 des *Principia Mathematica*. Notamment, la démonstration du théorème 2.85 est plus élégante que celle présentée dans l'ouvrage original. Les chercheurs tentent de la publier dans le *Journal of Symbolic Logic* sans succès, car cette amélioration n'est

pas jugée significative. Néanmoins, le Logic Theorist a une influence considérable sur la recherche en IA et les sciences cognitives, en inspirant le « computationnalisme* », une théorie de l'esprit qui le conçoit comme un processus de traitement de l'information (McCorduck 2004, 127). Qui plus est, les heuristiques deviennent un principe de fonctionnement essentiel pour faire avancer les approches symboliques en IA.

Encouragés par ce succès, Newell, Simon et Shaw poursuivent leurs recherches avec un projet plus général, d'emblée sur des méthodes heuristiques. En 1959, ils proposent le « *General Problem Solver* » (GPS), une version améliorée du Logic Theorist, capable de résoudre des problèmes géométriques, des casse-têtes logiques et des parties d'échecs simples, en plus des théorèmes des *Principia Mathematica*.

L'innovation conceptuelle majeure de cet algorithme est l'application de l'heuristique de « l'analyse des moyens et des fins ». Cette approche sépare la connaissance du problème de ses stratégies de résolution. Elle commence par définir un état actuel (équivalent de l'axiome racine pour le Logic Theorist) et un état cible (le théorème à prouver). Ensuite, le GPS choisit une action qui rapproche ces deux états. Après chaque action, il redéfinit l'état actuel (qui devient l'équivalent d'une branche pour le Logic Theorist) et répète l'opération jusqu'à ce que l'état actuel concorde avec l'état cible.

Le succès de ce nouveau programme conduit Newell et Simon à formuler l'hypothèse suivante : « *Un système de symboles physiques possède les moyens nécessaires et suffisants pour une action intelligente générale.* » (Newell et Simon 1972, 116). Cette déclaration, en accord avec le computationnalisme, suggère que l'IA peut être réalisée à l'aide d'un système de symboles, une idée qui s'applique également à l'esprit humain, considéré comme une « *espèce du même genre* » (Newell et Simon, 1972, 116). Par conséquent, le computationnalisme modélise l'esprit comme un processus de traitement de l'information, tandis que l'approche symbolique précise que les informations manipulées sont des symboles. Ces théories sous-jacentes se développent et se raffinent au fil du temps, et elles servent de justification conceptuelle à l'approche symbolique.

Le GPS inspire de nombreux algorithmes fonctionnant sur les principes de la logique déductive et des heuristiques. Toutefois, les chercheurs se heurtent à des limites conceptuelles dès qu'ils s'aventurent hors des domaines familiers. Les heuristiques fonctionnent efficacement tant que les problèmes sont bien compris, comme en mathématiques, mais aborder des domaines d'expertise différents exige une analyse plus approfondie du problème, nécessitant la capacité d'imaginer le type de solution que l'algorithme peut apporter, puis d'inventer les moyens pour y parvenir.

Une autre limite est plus conceptuelle, car de nombreuses situations ne se prêtent pas à un formalisme symbolique, malgré leur apparence. Par exemple, les conversations informelles et

des textes littéraires demandent de comprendre des références et des allusions qu'un ensemble de règles et de symboles formel ne saurait modéliser. De même, le diagnostic médical est souvent réalisé dans un contexte très spécifique et imprécis, difficilement généralisable et ayant des données manquantes qui mettent à mal le formalisme symbolique.

1.2.2. Jusqu'aux systèmes experts

Les approches symboliques en IA atteignent leur apogée avec les systèmes experts. Dans les années 1960, Feigenbaum (1977) remarque que les outils d'IA sont souvent limités par la connaissance du concepteur plutôt que par la puissance de la méthode utilisée. Il se met à développer des heuristiques plus complexes, ce qui résulte dans la création de l'algorithme DENDRAL, développé par Feigenbaum, Buchanan, Lederberg et Djerassi. DENDRAL est le premier système « à connaissance extensive ».

Dans le projet DENDRAL, les chercheurs mettent l'accent sur l'importance des connaissances spécialisées par rapport aux méthodes génériques de résolution de problèmes, comme le GPS. Ils choisissent d'aborder un problème en chimie analytique : déterminer la structure chimique d'une molécule à partir des signaux de ses éléments chimiques obtenus par spectrographie de masse (voir figure 3).

DENDRAL construit des explications à partir des données empiriques, c'est-à-dire des signaux obtenus par la spectrographie (les bâtons dans la figure 3). Chaque signal correspond à un type de liaison entre des atomes, identifiable par sa valeur. De nombreuses combinaisons sont possibles, mais au lieu d'explorer toutes les possibilités une par une, l'algorithme applique des règles sur la structure moléculaire d'un composé chimique, à l'instar des Turingismus de Bombe. La difficulté de conception de DENDRAL réside dans le développement d'un système de représentation des connaissances et d'heuristiques qui traduisent les « règles de cuisine » (« *cookbook recipes* ») des chimistes analytiques plutôt que des principes théoriques (Buchanan, Feigenbaum, et Lederberg 1971).

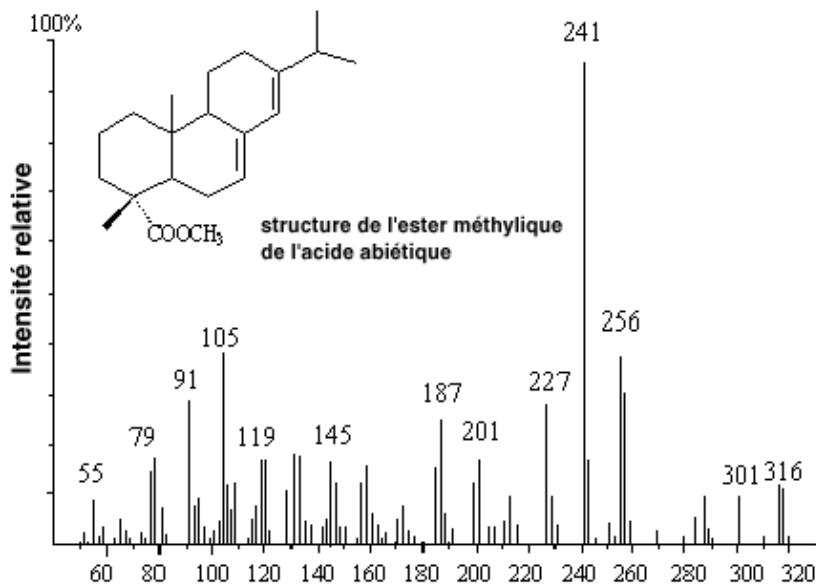


Figure 3 : Résultat d'une spectrographie de masse de l'ester méthylique de l'acide abiétique

La réussite de DENDRAL incite les chercheurs à poursuivre cette voie. Ils développent alors MYCIN, un algorithme conçu dans les années 1970 à l'Université de Stanford par Buchanan et Shortliffe pour aider au diagnostic des infections liées au sang et prodiguer des conseils thérapeutiques sur les antibiotiques associés. MYCIN s'appuie sur des bases de données médicales, la connaissance métier et des règles élaborées à partir d'entretiens avec des médecins experts (voir détails dans l'annexe 1 : présentation de l'algorithme MYCIN).

Cet outil d'IA marque le début des « systèmes experts », qui visent à fournir des solutions équivalentes à celles d'un expert dans des domaines complexes tout en restant compréhensibles. Dans leur forme canonique, ces systèmes commencent par une requête de l'utilisateur liée à son domaine d'expertise. Par exemple, dans le cas de DENDRAL, un chimiste analytique demande la structure d'une molécule à partir des données de spectrographie de masse, tandis que dans le cas de MYCIN, un médecin sollicite un diagnostic et des options thérapeutiques pour une infection sanguine.

Le système expert utilise des connaissances spécifiques au domaine pour construire une série de déductions, de manière semblable au Logic Theorist ou au GPS, mais avec des connaissances adaptées au domaine. Ces connaissances sont stockées dans deux bases distinctes : une « base de faits » et une « base de règles » qui s'appliquent aux faits. Le système expert utilise un « moteur d'inférences » pour automatiser les déductions en combinant des faits et des règles jusqu'à trouver une réponse, en suivant trois principes de fonctionnement différents. Quand le système expert fournit une solution, il est capable de reconstituer le cheminement depuis les informations initiales fournies par l'utilisateur, y compris les règles et les faits requis. De plus, pour être aussi proche que possible des besoins des utilisateurs, ces

derniers peuvent actualiser la base de faits et de règles sans nécessiter d'expertise informatique.

1.2.3. *Limites des approches symboliques*

Les systèmes experts représentent la quintessence des approches symboliques en IA. Leur fonctionnement algorithmique reste globalement le même, avec quelques ajustements. Nous pouvons considérer qu'il s'agit d'un paradigme, selon la définition de Kuhn (1962) dans *La Structure des Révolutions Scientifiques*, car il procure une perspective unique sur la résolution des problèmes et les méthodes à utiliser.

Ainsi, après des débuts rudimentaires, le domaine de l'IA s'est structuré autour du computationnalisme et d'une modélisation de la pensée humaine grâce à des symboles et des règles d'application. En raison des limitations des ordinateurs et des particularités empiriques de chaque problème, l'approche a évolué, à partir des déductions abstraites issues du formalisme mathématique vers des heuristiques établies sur l'expérience des experts du domaine.

Les systèmes experts ont poussé cette approche en construisant des bases de faits et de règles spécifiques à chaque situation. Pour ce faire, des méthodes d'acquisition des connaissances ont été développées, notamment des entretiens avec des experts pour ajouter à la théorie générale des éléments de raisonnement pratiques des professionnels. Cette collecte de données a été à l'origine d'un travail conceptuel pour créer une ontologie qui traduit ces connaissances en un langage symbolique et une classification appropriée pour les algorithmes.

L'engouement de l'industrie pour les systèmes experts incite à de nombreuses expérimentations. Comme dans la notion de paradigme, les chercheurs consolident leur approche en multipliant les exemples empiriques. Cependant, aucune théorie ne correspondant parfaitement à la réalité, les difficultés rencontrées lors de ces expérimentations permettent de mieux comprendre les limites des approches symboliques en IA. D'abord, un système expert repose sur un ensemble de connaissances fermé : il ne crée jamais de connaissance mais se contente d'utiliser celles qui préexistent. De plus, il ne possède pas de compréhension approfondie des concepts, se limitant à des calculs. En cas de programmation incorrecte, il peut choisir une méthode inappropriée pour résoudre un cas particulier, ce qui peut entraîner des pertes de temps. Les systèmes experts ne sont pas en mesure d'apprendre et doivent être reprogrammés pour s'améliorer.

En somme, les connaissances des systèmes experts sont assez superficielles, ce qui pose deux grands défis. D'une part, leur conception nécessite une équipe d'ingénieurs pour traduire les connaissances dans l'algorithme, ce qui n'est pas toujours justifié en termes de gains produits. Ensuite, comme illustré par MYCIN, les utilisateurs restent sceptiques quant à

l'intervention de la machine dans des problèmes éthiques. Malgré les explications fournies par les systèmes experts, la méfiance des utilisateurs persiste, en particulier dans des domaines où les connaissances sont incertaines.

Les domaines ayant des connaissances incertaines sont ceux dans lesquels les informations disponibles ne sont pas précises ou déterministes, comme en médecine, en finance, en météorologie ou dans la jurisprudence

- Les diagnostics et les pronostics peuvent être incertains en raison de la variabilité des symptômes, des réponses individuelles aux traitements et de la complexité des maladies.

- Les prévisions de marché, les évaluations des risques et les décisions d'investissement sont souvent fondés sur des données incertaines en raison de la volatilité des marchés et de l'incertitude économique.

- Les prévisions météorologiques et les modèles climatiques doivent composer avec l'incertitude liée à la complexité des systèmes atmosphériques et à la variabilité des conditions environnementales.

- Les décisions judiciaires peuvent être fondées sur des preuves et des témoignages incertains, ce qui rend difficile de déterminer la culpabilité ou l'innocence d'une personne.

Dans les années 1980, l'industrie de l'IA connaît une phase d'investissements massifs, avec des attentes exponentielles quant aux capacités des systèmes experts. Leurs capacités à modéliser et à reproduire l'expertise humaine dans des domaines tels que la médecine, la finance, l'ingénierie, et bien d'autres ont suscité de nombreux projets pour des tâches de diagnostic, de conseil, de planification et de prise de décision. Ils offrent des avantages significatifs, notamment la rapidité et la cohérence de leurs réponses, la disponibilité 24h/24 et la capacité à stocker et à mettre à jour des connaissances spécialisées. Ils sont aussi une source de préoccupation concernant la déprofessionnalisation (Baer 1986), étant donné leur potentiel à rendre le travail professionnel « *de plus en plus standardisé et routinisé, laissant moins de jugements discrectionnaires* » (Stoeckle 1988, 82).

Cependant, ils présentent des limites (Gill 1995), notamment la difficulté d'acquérir et de maintenir des connaissances spécialisées, ainsi que des questions éthiques et juridiques liées à leur utilisation en pratique clinique et dans d'autres domaines. Les approches symboliques ne parviennent pas à produire des résultats durables. En rétrospective, cet ensemble de technologies est parfois qualifié de « GOFAI » pour « *Good Old-Fashioned AI* », par opposition au nouveau paradigme des approches d'apprentissage machine qui se développe par la suite.

Par ailleurs, en sciences de gestion, nous pouvons voir les systèmes de gestion de connaissances (« *Knowledge management systems* ») (Taskin et Van Bunnem 2015), comme un des héritages des approches symboliques.

1.3. Approches d'apprentissage machine

Après l'exploration des débuts des approches d'apprentissage machine (1.3.1), nous passons en revue les principales techniques (1.3.2) et expliquons l'importance des données (1.3.3). Enfin, nous questionnons la valeur ajoutée de ces outils (1.3.4).

1.3.1. Début des approches d'apprentissage machine

Les approches d'apprentissage machine trouvent leurs racines dans le connexionnisme. Cette perspective considère que les phénomènes mentaux résultent de l'interconnexion entre les neurones. Ainsi, les outils d'IA composés de réseaux de neurones artificiels interconnectés peuvent apprendre à effectuer des tâches spécifiques en ajustant les poids des connexions en réponse à des données d'entrée. Plus généralement, les outils d'IA d'apprentissage machine utilisent une variété de modèles et d'algorithmes, allant des plus simples, comme la régression linéaire, aux plus complexes, comme les réseaux de neurones profonds. Le choix du modèle dépend de la tâche à accomplir et de la nature des données disponibles. Par exemple, des modèles d'apprentissage profond peuvent être entraînés sur de nombreuses images de mammographies annotées pour apprendre à identifier les caractéristiques visuelles subtiles associées aux tissus cancéreux, afin d'évaluer si une patiente est atteinte de cancer.

Ainsi, l'apprentissage machine ne se concentre pas sur la formalisation de règles pour manipuler des symboles. Au contraire, il vise à reproduire le processus qui mène à la création de ces symboles et règles en se concentrant sur l'apprentissage. Comme nous l'avons vu précédemment (1.1.3), l'apparition du neurone artificiel, inventé par McCulloch et Pitts en 1943, marque le début du mouvement cybernétique. Ces chercheurs se positionnent à la croisée du computationnalisme et du connexionnisme, en considérant l'intelligence humaine comme un processus de calcul de l'information. Ils proposent de connecter plusieurs neurones pour modéliser le fonctionnement du cerveau. Cette idée évolue grâce aux travaux de Hebb en 1949, mettant en avant comment la force de connexion entre les neurones artificiels peut engendrer un apprentissage au sein du réseau qu'ils forment.

Cependant, au début, le connexionnisme peine à produire des résultats convaincants en raison de limitations techniques. En effet, Minsky et Edmonds construisent le SNARC, la première machine à réseau neuronal, en 1950, mais les résultats pratiques s'avèrent insuffisants. Selon les travaux de Winograd et Cowan en 1963, le réseau de 40 neurones artificiels est trop petit pour permettre un apprentissage significatif (Winograd et Cowan 1963). Plus précisément, le livre *Perceptrons* de Minsky et Papert en 1969 souligne le besoin d'un réseau neuronal complexe, impossible à réaliser avec les technologies existantes (Minsky et Papert 1969). Ainsi, malgré quelques études sur les méthodes d'apprentissage des réseaux, l'engouement initial pour l'approche connexionniste retombe.

En marge et limitrophe au connexionnisme, l'apprentissage machine se développe dès le début de la recherche en IA. Turing avait évoqué la possibilité que des machines puissent apprendre, un moyen à terme plus efficace que les règles formelles (Turing, 1950). Quelques années plus tard, en 1959, Samuel développe un algorithme d'apprentissage automatique pour jouer aux dames et donne ainsi naissance à l'expression – et l'approche – de « l'apprentissage machine » (« *machine learning* »). Cependant, les travaux sur les approches symboliques, dont les systèmes experts, éclipsent cette voie de recherche jusqu'à la fin des années 1980.

L'échec des systèmes experts remet en question le paradigme dominant des approches symboliques, poussant ainsi la recherche en IA à explorer de nouvelles directions. Les travaux de McAllester en 1998 sont représentatifs de cette transition, démontrant une ouverture de la recherche en IA vers des approches multidisciplinaires. McAllester a souligné que, au début de l'IA, il semblait plausible que de nouvelles méthodes de calcul symbolique, comme les cadres et les réseaux sémantiques, puissent rendre désuète une grande partie de la théorie classique.

Cette perspective a entraîné une forme d'isolationnisme, détachant l'IA du reste des sciences informatiques. Cependant, il est désormais reconnu que l'apprentissage machine ne doit pas être séparé de la théorie de l'information, que le raisonnement incertain doit être intégré avec la modélisation stochastique, que la recherche en IA doit rester en lien avec l'optimisation et le contrôle classiques, et que le raisonnement automatisé doit être connecté aux méthodes formelles et à l'analyse statique (Russell et Norvig 2021, 75).

1.3.2. Différentes formes d'apprentissage machine

Les diverses théories sous-jacentes à l'apprentissage machine trouvent leur origine dans le domaine de la statistique, qui examine les phénomènes en collectant des données, les traitant, les analysant, puis interprétant les résultats pour les présenter. L'apprentissage machine opérationnalise ce processus en l'automatisant grâce à des algorithmes. Dans cette optique, la communauté autour de la découverte de connaissances dans les bases de données (*Knowledge in Databases Discovery*) émerge à partir des années 1980 pour unifier les différentes techniques (Fayyad, Piatetsky-Shapiro, et Smyth 1996). Ces techniques spécifiques se développent en fonction de la disponibilité des données et des résultats souhaités. Bien qu'elles puissent être regroupées en types d'apprentissage relativement stables, la prolifération et la complexification des techniques rendent une classification parfaite difficile, et il est toujours possible de trouver des sous-catégories plus précises ou des techniques à la frontière entre deux catégories.

L'apprentissage non supervisé* s'applique lorsqu'une base de données non étiquetée est disponible. Par exemple, dans le contexte d'une cohorte prospective* de patients en cours de

traitement, il peut être impossible de déterminer des facteurs pronostiques de survie. L'apprentissage non supervisé examine l'ensemble des données collectées pour identifier des groupes, d'où son nom de « clustering » en anglais. Ces groupes résultent de similitudes statistiques entre les données, et peuvent aider à identifier des facteurs distinctifs. Généralement, la base de données est subdivisée en plusieurs groupes de manière itérative pour déterminer le nombre optimal sur la base de critères statistiques. De plus, plusieurs méthodes statistiques pour calculer la « similarité » entre les données sont employées pour créer ces groupes. La partition des données est ensuite interprétée par un expert, tel qu'un épidémiologiste, afin de donner un sens aux groupes identifiés.

L'**apprentissage supervisé*** est un archétype de l'apprentissage machine, utilisé lorsque les catégories des données sont connues. Par exemple, dans le domaine des algorithmes de diagnostic, lors de la phase d'entraînement, un algorithme est construit à partir d'une base de données médicales dans laquelle on sait, a posteriori, si un patient est malade ou non. Ces données servent à entraîner l'algorithme, qui modélise les données pour identifier des motifs potentiels, c'est-à-dire des combinaisons de facteurs qui déterminent si le patient est malade ou non. Lors de l'utilisation de l'algorithme, il attribue une étiquette (encore inconnue) aux données, c'est-à-dire un diagnostic établi sur les informations fournies. D'un point de vue mathématique, l'apprentissage supervisé peut être formulé comme une fonction $Y=f(X)$, où Y est le résultat souhaité (par exemple, si les patients sont malades ou non), X représente les informations, et la fonction f est le modèle appris par l'algorithme pour prédire la bonne réponse.

L'**apprentissage par renforcement*** vise à trouver une solution optimale dans un environnement donné. L'algorithme est programmé avec un ensemble de règles qui décrivent cet environnement, ainsi qu'une fonction de coût qui récompense ou pénalise les résultats trouvés par l'algorithme pour l'inciter à s'améliorer. Par exemple, dans le domaine des soins de santé, pour déterminer le meilleur traitement pour un patient, l'algorithme doit comprendre les règles physiologiques et cliniques liées au patient et aux médicaments, puis effectuer des simulations pour trouver le traitement optimal en minimisant la fonction de coût. Bien que cette méthode rappelle les approches symboliques grâce à un ensemble de règles explicites pour modéliser l'environnement, elle va au-delà, car elle expérimente dans le cadre défini par ces règles pour découvrir la relation $Y=f(X)$. Dans l'exemple précédent, Y indique si le patient a réagi positivement au traitement, X représente l'état initial du patient et f détermine les stratégies thérapeutiques en fonction des situations X .

L'**apprentissage profond*** est un type d'apprentissage qui traverse les autres catégories. Il peut être appliqué à l'apprentissage par renforcement, à l'apprentissage non supervisé et à l'apprentissage supervisé. La spécificité de cette approche réside dans l'utilisation de réseaux de neurones artificiels*. Le terme « profond » se réfère au grand nombre d'opérations

statistiques appliqué aux données. Chaque neurone effectue une de ces opérations. Les neurones sont organisés selon une architecture de réseau complexe, généralement composée d'une première couche prenant les données initiales et d'une dernière couche produisant le résultat. Par exemple, pour détecter une image suspecte de cancer sur une radiographie, la première couche de neurones traite l'information provenant de chaque pixel de l'image, tandis que la dernière couche marque les zones suspectes.

Remarquons que les concepteurs déterminent les conditions sous lesquelles un outil d'IA apprend, mais que l'apprentissage réel (trouver des patterns dans les données d'entraînement et les combiner dans un modèle unifié) est effectué par des règles statistiques (Jordan et Mitchell 2015).

1.3.3. Produire des outils d'IA, un besoin dévorant de données

À partir des années 2000, une course aux données massives*, connue sous le nom de « Big Data », s'engage pour améliorer les performances des techniques d'apprentissage machine (Pantanowitz et al. 2021). Les chercheurs démontrent que plus l'algorithme peut s'entraîner sur un grand volume de données, plus le modèle peut être complexe et précis, indépendamment de la technique d'apprentissage employée (Hays et Efron 2007).

Parmi les initiatives de construction de bases de données massives pour l'entraînement, ImageNet est la plus connue, car elle contribue à structurer la recherche en IA depuis la dernière décennie. Créée en 2006 par puis Fellbaum, cette base d'images recense plus de 14 millions d'images annotées à la main en 2021, avec plus de 20 000 catégories (par exemple, plus de 120 races de chiens identifiées). De 2010 à 2017, ImageNet propose un concours annuel d'algorithmes de classification et de détection d'objets. Lors de la compétition de 2012, un algorithme d'apprentissage profond utilisant un nouveau processeur graphique pour plus de puissance de calcul dans le traitement d'images fait sensation avec un taux d'erreur inférieur de 10,8 points de pourcentage au second meilleur. Selon *The Economist*, cette démonstration crée une prise de conscience collective : « soudain, les gens ont commencé à prêter attention, non seulement au sein de la communauté de l'IA, mais aussi dans l'ensemble du secteur technologique. »⁷.

Ainsi, avec l'augmentation des capacités de calcul et de la masse de données, les outils d'IA ne cessent d'être développés pour des tâches de plus en plus variées et avec des performances de plus en plus proches des capacités humaines.

⁷ "From not working to neural networking". *The Economist*, 25 Juin 2016

Cependant, la capacité d'apprentissage de l'IA contemporaine présente également certaines limites. Les chercheurs en IA prennent conscience que le volume de données ne suffit pas, la qualité des données est également cruciale. Les outils d'IA d'apprentissage machine sont caractérisés par leur focus limité (Larson 2021), apprenant à partir de données avec une fonction objective clairement définie à optimiser. Étant donné que l'outil d'IA dépend entièrement des données d'entraînement et des mécanismes d'apprentissage prédéfinis, sa performance est limitée à l'ensemble de données d'entraînement. Toute information contextuelle ou environnementale est invisible pour l'outil d'IA si elle n'est pas déjà présente dans les données pré-sélectionnées ou, dans le cas de l'apprentissage par renforcement, dans l'environnement choisi.

De plus, la base de données d'entraînement sur laquelle les outils d'IA se fondent, appelée « *ground truth* », peut reposer sur des expertises incertaines, limitant ainsi les performances des outils, notamment en médecine (Lebovitz, Levina, et Lifshitz-Assa 2021). En effet, les données ne sont pas neutres (Gitelman 2013; Desrosières et Kott 2005), et elles peuvent induire un biais moral dans le fonctionnement des outils d'IA (O'Neil 2016), que ce soit pour prendre des décisions de justice, de distribution des ressources de santé ou pour recruter (voir Chapitre 2, 1.2.2). Il est aujourd'hui évident que la collecte de données doit être effectuée avec attention et de manière itérative pour déterminer les données pertinentes.

En dehors des conséquences liées à la qualité des données sur le fonctionnement des outils d'IA, la collecte, le stockage et la préparation des données massives posent d'autres problèmes. Pour répondre au besoin de données massives et à leur annotation, de nombreux services en ligne se développent, comme Mechanical Turk d'Amazon. Or, ces activités sont assurées par des « travailleurs du clic », une main d'œuvre précaire (Casilli 2019). En raison de leurs besoins en ressources, les données massives et l'industrie de l'IA ont des conséquences négatives sur l'environnement, allant de la destruction de territoires, jusqu'aux conflits géopolitiques dus aux besoins de lithium et d'eau pour faire fonctionner les centres de données (Crawford 2021).

1.3.4. *Quelle est la valeur des outils d'IA ?*

Aujourd'hui, le paradigme des approches d'apprentissage machine domine la recherche en IA et les applications qui en résultent. Grâce à leur nature statistique, les méthodes d'apprentissage machine excellent dans la découverte d'associations ou de corrélations entre les données. L'objectif n'est donc plus tant lié à l'automatisation d'un raisonnement supposé humain, mais davantage à l'efficacité, quelles que soient les similarités avec l'intelligence humaine (Russell et Norvig 2021). Les concepteurs n'ont plus besoin de coder chaque étape de l'algorithme, mais de collecter des données pertinentes et d'évaluer les résultats des outils d'IA. De plus, avec l'augmentation de la capacité de calcul et l'accès à un volume massif de

données, les performances des outils d'IA ont progressé dans une gamme de tâches de plus en plus variées.

Désormais, l'attention se tourne vers les usages des outils d'IA. De nombreux rapports à travers diverses professions, comme le droit (Whitehouse 2020), la médecine (Olthof, Van Ooijen, et Rezazade Mehrizi 2020; Topol 2019a; 2019b), la comptabilité (Sutton, Holt, et Arnold 2016) et l'architecture (Geertman et al. 2019) montrent que ces experts métier cherchent à comprendre le potentiel et la menace de ces outils. Chaque application est unique, mais les préoccupations communes se concentrent sur des questions d'ordre éthique, juridique et organisationnel. En fin de compte, il s'agit de déterminer la valeur ajoutée des outils d'IA, d'établir le rapport bénéfice/risque.

Les questions relatives aux usages des outils d'IA dans des contextes de haute expertise revêtent une importance particulière, en raison de l'opacité de ces outils. Dans des domaines comme la médecine, ces outils participent à la prise de décisions cruciales ayant un impact majeur sur la vie des individus, la société ou l'économie. Toutefois, les outils d'IA, en particulier les modèles d'apprentissage profond, sont souvent d'une complexité vertigineuse, avec des milliards de paramètres. Bien qu'ils soient capables de déceler des schémas dans les données qui échappent souvent à la compréhension humaine, leur complexité rend difficile l'interprétation et l'explication de leurs résultats.

Ce phénomène est couramment désigné sous le terme « **d'opacité** » des outils d'IA ou encore « d'effet boîte noire ». Contrairement aux approches symboliques, ces outils d'IA fournissent des réponses sans expliciter leur raisonnement. Cette opacité est souvent observée dans les modèles d'apprentissage profond puisque la relation entre les entrées et les sorties est complexe et non linéaire. Dans une moindre mesure, même les modèles linaires peuvent aboutir à des résultats opaques pour les utilisateurs car les données en entrée et les liens établis sont eux-mêmes peu interprétables. Par conséquent, même si un outil d'IA aboutit à des résultats précis, il peut être difficile de comprendre comment il est parvenu à ces résultats.

Face à ces enjeux, le domaine de l'*Explainable Artificial Intelligence** connaît une intense activité, bien que cette recherche demeure insuffisante. Cette discipline établit des directives de conception visant à réduire l'opacité des outils d'IA, tout en développant des méthodes d'évaluation pour détecter les risques de biais dans ces outils (Meske et al. 2022). Toutefois, ces efforts se concentrent sur la phase de conception et délaissent les aspects liés à l'utilisation et les répercussions organisationnelles (Hafermalz et Huysman 2021). À mesure que les outils d'IA se multiplient, il devient impératif de mettre en place des approches complémentaires pour étudier leurs utilisations, notamment en observant ce qui se passe sur le terrain (Kostick-Quenet et Gerke 2022; Meske et al. 2022). Par rapport à l'opacité, l'objectif est de comprendre

comment les utilisateurs interprètent les résultats des outils d'IA. Cette question est essentielle pour évaluer la valeur de ces outils.

La recherche en IA commence par explorer des approches symboliques. L'accent est alors mis sur la formalisation d'ensembles de règles établies par les concepteurs d'outils d'IA, en s'appuyant sur l'expertise et les connaissances pratiques des spécialistes du domaine concerné. Cependant, l'enthousiasme quant à la portée des outils d'IA retombe rapidement face aux difficultés de formalisation des connaissances dans des environnements incertains et aux réticences juridiques et morales de déléguer des tâches importantes à des machines.

Face à la double déception, tant sur le plan théorique qu'industriel, la recherche en IA se réoriente vers des approches d'apprentissage machine, pour permettre aux machines d'apprendre les règles préalablement formalisées. Dans ce nouveau paradigme, ce qui importe davantage, c'est d'acquérir des données d'entraînement et de créer des outils d'IA capables de produire des résultats pertinents. Néanmoins, l'opacité de fonctionnement inhérente à ces outils d'IA complique, voire rend impossible, la compréhension de leur processus décisionnel.

Les usages des outils d'IA dans le contexte du système de soins représentent un cas extrême. Contrairement aux sciences dites « dures », la médecine est une science humaine et donc imprécise, qui est donc sujette à une grande variabilité voire une imprécision des données, qui sont sujettes à interprétation contextuelle. De plus, le risque de biais dans les données d'entraînement peut mener à des recommandations inévitables ou inappropriées par les outils d'IA, exacerbant potentiellement les inégalités existantes. Des outils d'IA fiables et d'une grande précision sont donc requis dans ce contexte où toute erreur peut avoir des conséquences directes sur la santé et la vie des patients. Ainsi, on attend transparence et intelligibilité du cheminement aboutissant aux propositions de décisions médicales fondées sur les outils d'IA. Pourtant, cela est difficile en raison de la nature des approches d'apprentissage machine ce qui limite l'acceptation par les professionnels de santé. En outre, la réglementation, qui doit garantir la sécurité des patients, ralentit souvent l'adoption de ces outils. L'ensemble de ces problèmes met en exergue la question fondamentale que rencontrent les outils d'IA aujourd'hui : quelle est leur valeur ?

2. Apports des outils d'IA dans le contexte du système de soins

Grâce à une technologie polyvalente (McAfee and Brynjolfsson 2017), les outils d'IA sont maintenant applicables dans divers domaines, comme la médecine (Esteva et al. 2017; Jha and Topol 2016), en effectuant de nombreuses tâches, comme l'analyse d'images et de textes (Hinton et al. 2012; Deng and Yu 2014).

En adoptant une perspective organisationnelle, nous situons l'intérêt suscité par les usages des outils d'IA en fonction de leur capacité à changer les moyens de production, en particulier dans le système de soins (2.1). Ensuite, nous distinguons quatre grands cas d'usage des outils d'IA dans ce système (2.2).

2.1. Des outils d'IA pour révolutionner le système de soins ?

L'usage des outils d'IA génère de nombreuses spéculations quant à leur valeur, notamment leur capacité à engendrer une révolution industrielle. Le système de soins semble être l'un des secteurs les plus concernés, avec un enthousiasme de la part de la majorité des acteurs du système de soins, y compris les agences gouvernementales (Talley et al. 2011), les prestataires de soins (Krittanawong et al. 2017), les compagnies d'assurance (Kose, Gokturk, et Kilic 2015) et les fabricants de produits pharmaceutiques (Ekins 2016). Comme en témoigne le rapport de Villani et al. (2018, 197) : « *Le développement de l'IA est appelé à transformer en profondeur les pratiques des professionnels de santé [...] l'enjeu est bien d'organiser des interactions vertueuses entre l'expertise humaine et les apports de l'IA dans l'exercice quotidien de la médecine* ».

Pour comprendre cette position, nous examinons comment les usages des outils d'IA peuvent influencer les moyens de production en général (2.1.1). Cela nous amène à considérer quelles sont les spécificités de la production dans le système de soins (2.1.2), puis la place des outils d'IA dans ce contexte particulier (2.1.3).

2.1.1. Des outils d'IA capables de changer la manière de produire ?

Les outils d'IA semblent capables de changer radicalement la façon de produire. Certains voient les moteurs de la révolution des outils d'IA comme étant économiques : ces systèmes rendent plus productifs, plus coordonnés, et capables de développer des avantages concurrentiels (Rouse et Spohrer 2018; Wang et al. 2017; Weking et al. 2020). D'autres ajoutent que ces outils permettent de nouvelles approches pour l'engagement des clients, l'augmentation des compétences des employés, de nouvelles offres de produits et de services, ainsi que l'exploration de nouveaux modèles d'affaires (Mittal et al. 2017).

Pour comprendre ces nombreux changements, Wang et al. (2017) retracent l'évolution des paradigmes de production et positionnent les outils d'IA comme la prochaine étape de cette

progression. Ils identifient quatre grandes phases de ce développement. La première, avant le XXe siècle, est celle de la production artisanale, caractérisée par la création individualisée de biens en fonction des besoins de chaque client, offrant ainsi une grande variété de produits, au détriment de la productivité en raison des coûts élevés des matières premières et de la main-d'œuvre. Cette production décline avec la première révolution industrielle qui mécanise de nombreuses tâches et rend l'énergie plus abordable.

Au début du XXe siècle, la production de masse prend le dessus, marquée par les principes de rationalisation de Taylor, considérés comme la deuxième révolution industrielle (Wang et al. 2017, 2). La citation de Ford (1926) : « *Tout client peut avoir une voiture peinte de la couleur qu'il veut, pourvu qu'elle soit noire* » résume cet état d'esprit. Bien que la variété des produits soit limitée, ils deviennent très accessibles. En conséquence, les ménages s'équipent de nombreux biens, dont les appareils électroménagers.

Le concept de personnalisation de masse émerge à la fin des années 1980 en réponse à une demande sociale croissante de produits adaptés aux besoins individuels. Cette notion se consolide grâce au développement de l'informatique et des technologies d'automatisation liées aux ordinateurs. Les machines à commande numérique, telles que les robots industriels, et les progiciels de gestion intégrée (*Enterprise Resource Planning System* ou ERP en anglais) assurent une production plus flexible, générant une variété de produits tout en maintenant une productivité élevée à un coût relativement bas.

Sur un plan plus global, les auteurs évoquent qu'à partir du début du XIXe siècle, les entreprises envisagent un nouveau changement de paradigme de production appelé l'Industrie 4.0. Cette évolution repose sur les avancées des technologies de l'information et de la communication, y compris les outils d'IA, ainsi que sur la convergence entre le monde virtuel et le monde physique. En particulier, l'Industrie 4.0 est censée intégrer une participation accrue des clients à la phase de conception, entraînant une plus grande personnalisation et une distribution de la valeur améliorée.

Cependant, Martinelli et al. (2021) relativisent le concept d'Industrie 4.0. Ils soulignent que les révolutions technologiques profondes, celles qui transforment la production et la société, sont caractérisées par des « technologies à usage général », telles que la machine à vapeur. Ces technologies se distinguent par leur omniprésence et leur complémentarité avec l'ensemble du système technique. En suivant cette perspective, ils examinent le paradigme de l'Industrie 4.0 en comparant la fabrication additive, l'Internet des objets, la robotique, le cloud, les données massives et les outils d'IA.

Ils concluent que, par rapport à la personnalisation de masse et à la troisième révolution industrielle, ces technologies ont tendance à renforcer les dynamiques existantes plutôt que de proposer un paradigme différent. Ils évoquent le potentiel des données massives et des

outils d'IA à devenir des technologies à usage général. En effet, elles sont déjà répandues et s'intègrent dans l'ensemble du tissu productif à mesure que la numérisation progresse.

Les avancées technologiques ont remodelé la façon de produire des biens, suscitant des spéculations sur la transformation radicale de la production du système de soins par les outils d'IA. Ils suscitent des opportunités pour une personnalisation accrue, une plus grande efficacité et des innovations médicales de pointe. Cependant, il est essentiel d'analyser les spécificités de la production du système de soins pour comprendre l'importance que peuvent avoir les outils d'IA dans cette organisation.

2.1.2. Spécificités de la production du système de soins

Aujourd'hui, le système de soins est une organisation complexe qui peut se définir par sa mission : la prise en charge des patients. En effet, le système de soins représente une infrastructure organisée, complexe et interconnectée visant à fournir des services de santé et de soins médicaux à une population donnée. Il englobe l'ensemble des institutions, des professionnels de la santé, des ressources financières et des installations médicales qui contribuent à la prise en charge des patients. Parce que le patient est à la fois l'objet du travail (le soin) et le sujet (il se soigne lui-même à l'aide des professionnels) de cette organisation, l'activité du système de soins a un impact direct sur les conditions de vie et de mort, justifiant ainsi de le caractériser comme « la plus existentielle des organisations » (Dumond 2022, 169).

Pour remplir sa mission, le système de soins est devenu une bureaucratie professionnelle (Mintzberg 1993). Son évolution reste difficile à décrire avec précision, mais de nombreux auteurs constatent une « industrialisation » croissante de ses activités, en particulier à partir de la seconde moitié du XX^e siècle (Pfaff et al. 2010; Mannion et Exworthy 2017; Dumond 2022). Plusieurs facteurs ont motivé cette évolution, notamment des contraintes économiques, des changements démographiques, des avancées médicales et des décisions politiques.

Cependant, dans une première approximation, nous considérons que le haut niveau d'expertise professionnelle confère aux professionnels de la santé une large autonomie, rarement remise en cause par les instances de direction, malgré une certaine standardisation des compétences. Une asymétrie d'information existe entre les professionnels de la santé sur le terrain et les gestionnaires en coulisses, ces derniers ayant souvent du mal à comprendre pleinement l'activité, et donc à la diriger (Moisdon 2012).

De nos jours, le système de soins est soumis à une tension entre la standardisation et la personnalisation de la prise en charge. La standardisation est davantage portée par les « gestionnaires » lors de l'évaluation de l'activité (Mannion et Exworthy 2017). La pratique médicale est concernée, notamment avec la médecine fondée sur les preuves* (« *evidence based medicine* » ou EBM) visant à standardiser la démarche diagnostique et thérapeutique

afin de guider les décisions des professionnels vers des solutions garantissant le meilleur rapport bénéfices/risques pour le patient. Plus précisément, elle établit ainsi des guides de bonnes pratiques dans l'espoir que l'homogénéité des actions réduira les variations dans les diagnostics et les traitements (Timmermans et Almeling 2009; Gawande 2011). Cette démarche de standardisation est définie comme la recherche d'un consensus sur les pratiques, les critères, les méthodes et les conceptions afin d'améliorer la qualité, la sécurité, la compatibilité, l'interopérabilité et la reproductibilité des interventions (Leotsakos et al. 2014).

Cependant, cette logique de standardisation, bien qu'elle repose sur des données scientifiques, éprouve des difficultés à être mise en œuvre en raison de problèmes de diffusion de l'information (Coiera 2015) et des résistances identitaires de la part des professionnels de santé, qui considèrent leur métier comme un « art » et doivent s'adapter au cas par cas (Waring et al. 2016). La prise en charge au cas par cas de chaque patient exige nécessairement la personnalisation des recommandations globales si l'on souhaite prendre en considération les particularités de chaque individu.

Trois approches distinctes de la personnalisation des soins se dégagent : la médecine personnalisée* (ou de précision), qui s'appuie, entre autres, sur la génomique pour potentialiser les réponses cliniques plus adaptées (Hamburg et Collins 2010) ; la prise en charge centrée sur le patient* (ou « *patient-centered care* »), qui répond aux besoins et aux demandes liés à la vie et au bien-être du patient (Oshima Lee et Emanuel 2013) ; enfin, des dispositifs mis en place par les pouvoirs publics pour améliorer l'équité des soins pour l'ensemble de la population, notamment en proposant des aides à des populations vulnérables (Minvielle et al. 2014).

En raison du caractère existentiel du système de soins, la gestion de cette tension entre standardisation et personnalisation ne semble pas pouvoir être résolue par le concept de personnalisation de masse, tel que défini dans la littérature, c'est-à-dire une adaptation de la production de masse aux besoins individuels tout en maintenant des coûts bas (S. M. Davis 1987). En effet, la littérature suggère que l'adaptation se fait progressivement en commençant par personnaliser les phases finales de la production : distribution, assemblage, fabrication, puis conception (Lampel et Mintzberg 1996). La spécificité du système de soins exige une approche inverse, à savoir « gérer la singularité à grande échelle », en commençant par évaluer l'unicité des besoins de chaque patient, puis en adaptant les stratégies thérapeutiques en conséquence (Minvielle 2018).

Pour résumer, nous pouvons considérer que le système de soins adopte un paradigme de production visant à rationaliser la qualité de la prise en charge. En effet, la qualité des soins est essentielle pour prévenir des complications résultant d'une prise en charge inadaptée. Cela implique un ajustement aux besoins spécifiques de chaque patient. Toutefois, il est crucial de gérer de manière rationnelle les ressources disponibles pour que chaque personne nécessitant

des soins puisse en bénéficier. De plus, la recherche médicale doit continuer à produire des connaissances pour améliorer la prise en charge.

2.1.3. La place des outils d'IA dans la production de soins

Les innovations dans le domaine de la santé sont à l'origine de certaines des améliorations les plus spectaculaires des résultats en matière de santé publique dans les pays développés (Rye et Kimberly 2007). En particulier, il existe un large consensus dans la communauté scientifique selon lequel les outils d'IA détiennent un grand potentiel pour soutenir les flux de travail dans pratiquement chaque branche de la médecine, (Das, Nair, et Peter 2020), avec par exemple la rationalisation des flux de travail et la réduction des erreurs (Topol 2019b). Les outils d'IA suscitent ainsi de grands espoirs grâce à leur capacité à reproduire certaines tâches à un niveau expert, contribuant ainsi à rationaliser la qualité des soins.

D'autres chercheurs y voient également des opportunités pour améliorer la prise de décision clinique personnalisée (Obermeyer et Emanuel 2016) et poser les fondements d'un nouveau paradigme dans la médecine de précision personnalisée (Acs, Rantalainen, et Hartman 2020). Dans son rapport, Bollier (2017) adopte une position plus forte en considérant que le rôle principal de ces outils est l'augmentation des compétences des médecins.

Au niveau des médecins, la plupart des radiologues s'accordent à dire qu'il est nécessaire d'adopter des outils d'IA pour faire face à la charge de travail croissante, à la complexité des cas et à la précision diagnostique requise, mais croient également que ces systèmes peuvent profondément changer leur travail de manière encore indéterminée (Miller et Brown 2018).

Nous constatons également qu'en pratique, la numérisation du système de soins progresse lentement. La littérature identifie des facteurs organisationnels à cette lenteur (Béjean et al. 2015). Mais d'autres indiquent les limites de ces technologies, perçues parfois comme une nouvelle source de problèmes. Par exemple, les logiciels en place sont peu adaptés à la complexité des situations, manquent de compatibilité entre eux et nécessitent des mises à jour fréquentes pour suivre les avancées médicales, ce que les professionnels de santé n'ont pas toujours les moyens de faire (Koppel et Gordon 2017)

Dans ce sens, il existe une forte résistance de la part des médecins à les adopter en raison du risque de négliger l'unicité des patients. Or, les médecins jouent souvent le rôle de gardiens dans les questions d'adoption technologique dans le système de soins (Lapointe et Rivard 2005), avec des conséquences massives pour le reste du système de soins (Romanow, Rai, et Keil 2018).

Plus généralement, les technologies numériques en place ne peuvent pas remplacer les professionnels de santé. Elles posent leurs propres défis, qui s'aggravent avec l'introduction de nouvelles technologies, en raison de la complexité des dispositifs superposés. L'accumulation de dispositifs créent des systèmes d'une complexité qui dépasse la compréhension des

acteurs, piégés entre la nécessité de conserver des dossiers papier et la transition vers le numérique. Malgré ces défis, les promesses technologiques demeurent, surtout lorsqu'il s'agit des outils d'IA. Par exemple, *Deep Medicine* (Topol 2019a) fait un éloge aux usages potentiels des outils d'IA, tout en critiquant le système technologique actuel. Face à ce paradoxe, la question se pose de savoir quelle est la place des outils d'IA dans le système de soins ?

2.2. Quatre types d'usage des outils d'IA dans le système de soins

Les usages des outils d'IA sont difficiles à cartographier en raison de leur nature émergente et exploratoire. La diversité des tentatives illustre cette complexité. Par exemple, en France, l'Agence Nationale de la Performance Sanitaire et Médico-Sociale (ANAP) a mis en place une plateforme répertoriant des projets utilisant l'IA dans le système de soins⁸. Le site recense sept catégories⁹ qui se superposent. Ainsi, deux projets d'analyse des prescriptions sont considérés à la fois comme un suivi du patient et comme de la médecine prédictive.

Dans un autre effort de catégorisation, Bpifrance a publié un « *panorama [non-exhaustif] des startups de santé françaises utilisant de l'IA* »¹⁰. 191 entreprises sont répertoriées dans sept catégories voisines de celles de l'ANAP¹¹.

Enfin, le Parlement européen a publié un rapport sur les applications, les risques et les impacts éthiques et sociétaux de l'IA en santé (European Parliament 2022). Il évoque les possibles classifications en fonction des acteurs susceptibles de les utiliser (patients et citoyens, professionnels de santé, administrateurs et gestionnaires ou régulateurs et décideurs), ou les lieux où ils peuvent être déployés (établissements de soins, laboratoires ou lieux administratifs). Cependant, il choisit de classer les usages des outils d'IA selon quatre pratiques : la clinique, la recherche, la santé publique et l'administration.

Dans notre thèse, nous identifions quatre types d'usages des outils d'IA dans le système de soins : les tâches administratives (2.2.1), les soins courants (2.2.2), la recherche médicale (2.2.3) et les usages destinés aux patients (2.2.4). D'une part, nous considérons que les usages dans la santé publique révèlent davantage du système de santé dans son ensemble. D'autre

⁸ <https://ia.anap.fr>

⁹ Diagnostic, prise de décision, médecine prédictive et préventive, organisation, automatisation, suivi du patient et gestion des données.

¹⁰ Publié le 7 juillet 2020 : <https://lehub.bpifrance.fr/panorama-startups-sante-francaises-ia/>

¹¹ Prévention de la santé, développement de nouvelles thérapies, diagnostic du patient, traitement du patient, parcours du patient, suivi du traitement et autres.

part, nous considérons que les usages des outils d'IA pour la recherche et la santé publique sont semblables en pratique, même s'ils peuvent avoir des finalités différentes. Enfin, nous ajoutons les usages des outils d'IA à destination des patients.

2.2.1. Les outils d'IA pour des tâches administratives

L'expertise sur l'IA est détenue par des ingénieurs en dehors du système de soins. Les difficultés qu'ils rencontrent pour appréhender les spécificités de cette organisation révèlent certaines dimensions à estimer. Une brève histoire de la jeune entreprise française Nabla Technologies illustre ces propos. Elle est construite à partir d'un entretien exploratoire avec l'un des anciens collaborateurs au début de l'aventure entrepreneuriale, que nous appelons Éric¹², de deux articles publiés par les co-fondateurs¹³, ainsi que d'une annonce publiée par le PDG sur les réseaux sociaux en 2021¹⁴.

Au début, l'entreprise réunissait une dizaine de jeunes data scientists* portés par l'IA et l'entrepreneuriat. Ils ont commencé par mettre en place un système de gestion des clients (*Customer Relationship Management*), car « tous ceux qui commencent dans l'IA passent par le marketing et on voulait vendre le produit qu'on faisait pour nous aussi. » (Éric).

Cependant, l'entreprise a cherché à se réorienter. Après des mois de discussions, ils ont décidé de se tourner vers le secteur de la santé, car « c'est sexy et personne ne va dire 'non, moi je ne veux pas aider les gens à aller mieux', donc ça mettait tout le monde d'accord » (Éric).

À ce stade, l'idée était de créer une clinique privée qui servirait de lieu d'expérimentation pour développer leurs outils d'IA. L'objectif était double : optimiser leurs processus pour gagner en compétitivité et démontrer leurs preuves de concept pour les vendre. Ils ont envisagé plusieurs tentatives, notamment la mise en place d'un système de saisie semi-automatique pour augmenter l'efficacité des médecins, ainsi qu'une infrastructure de données capable de fournir des informations aux futurs modèles d'IA.

Cependant, le projet ambitieux peine à se concrétiser. Il s'est transformé dans un premier temps en une application médicale pour le suivi des femmes. En 2021, après des mois de travail, le PDG a annoncé dans un post LinkedIn le lancement de l'application (qui n'est plus disponible aujourd'hui) :

¹² Nom anonymisé. Entretien exploratoire de 93 minutes conduit le 14 octobre 2021.

¹³ Delphine Groll, co-fondatrice et COO, 21 novembre 2022, <https://www.nabla.com/blog/operations-nabla/>
Alexandre Lebrun, co-fondateur et CEO, 2 juin 2022), <https://www.nabla.com/blog/platform-launch/>

¹⁴ <https://www.linkedin.com/feed/update/urn:li:activity:6785791305374085120/>

« Nous sommes très heureux de partager deux nouvelles aujourd’hui : nous lançons l’application de santé que toutes les femmes méritent et nous avons levé 17 millions d’euros auprès d’excellents investisseurs. Après avoir mené un bêta-test approfondi et interrogé plus de 2000 femmes, il nous est apparu clairement qu’il y avait des lacunes critiques dans les parcours de santé des femmes à tous les stades de la vie. Notre recherche a montré que plus de 2/3 des femmes souhaitent être mieux soutenues et comprises. Nous avons créé la première application gratuite de soins de santé dédiée aux femmes, qui donne accès à des conseils médicaux sans tabou et permet d’identifier des tendances personnalisées et exploitables pour améliorer la santé » (Traduit de l’anglais.)

Le choix de cette application témoigne d’une approche assez classique parmi les concepteurs d’IA. La volonté première était de développer un outil d’IA, puis de trouver un cas d’utilisation ayant un grand impact. Les arguments invoqués sont souvent établis sur des chiffres, comme la proportion de personnes concernées, les gains de temps, ou d’autres dimensions quantifiables. L’application proposait de dialoguer avec une équipe de médecins employés par l’entreprise. Cependant, la confidentialité des informations échangées entre la patiente et le médecin, ainsi que d’autres aspects législatifs, limitaient les possibilités d’action des ingénieurs pour améliorer le dispositif.

Un autre obstacle majeur était le modèle de revenus vis-à-vis des médecins employés, qui n’était pas viable, car il n’était pas aligné sur les demandes fluctuantes des patientes. L’entreprise a fait face à un « choc des cultures » entre le monde des start-ups et le monde médical. Plus précisément, dans les start-ups, la communication asynchrone se fait via des messages électroniques, et le processus décisionnel est assez linéaire. En revanche, les médecins ont des temps de réunion présentielle qui se déroulent de manière plus informelle, à l’occasion desquels ils échangent des informations sans aboutir systématiquement à une décision. Cela leur permet de surveiller l’évolution clinique des patientes et de s’adapter à la singularité de chaque situation.

Aujourd’hui, l’entreprise s’est réorientée vers une application destinée aux médecins pour les aider à rédiger leurs comptes rendus. Ils proposent de réduire le temps que les médecins consacrent à la rédaction de leurs notes et à la gestion de ces informations de manière plus efficace. Les informations restent stockées sur l’ordinateur du médecin, ce qui évite de se heurter à de lourdes réglementations concernant l’externalisation et la protection des données patients. Cette solution rejoint d’autres entreprises en IA dans le domaine de la santé qui fournissent des services « administratifs », par exemple Lifem ou Doctolib. Les témoignages de l’un des dirigeants, ainsi que d’un « Officier Data » d’un service du Centre Hospitalier Perelman (IMP1, voir annexe 5) convergent sur les difficultés d’entrer dans le système de soins par le biais de la « porte clinique ». En revanche, le système informatique est encore peu

mature en termes de gestion des données, ce qui représente une opportunité de développer des outils d'IA.

Pour ces raisons, de nombreux outils d'IA dans le système de soins se consacrent à des tâches administratives. Ils manipulent des informations à des fins d'organisation, de contrôle, etc. Les produits actuels de Nabla et Lifen, pour ne citer que quelques-unes des entreprises françaises, s'inscrivent dans cette catégorie.

Cependant, il convient de nuancer deux points. Certaines offres de valeur se situent entre deux eaux, par exemple, des systèmes de recommandations qui fournissent aux patients des informations pertinentes en fonction de leur condition clinique. De plus, comme le révèlent les témoignages des fondateurs de Nabla, même les outils d'IA liés à des tâches administratives doivent s'adapter aux particularités du milieu. Ces particularités découlent du caractère existentiel du système de soins, notamment l'importance accordée à la confidentialité des données des patients.

En conclusion, le développement des infrastructures numériques et la maturité croissante dans la gestion et le traitement des données renferment la promesse d'outils d'IA administratifs capables d'atténuer les tensions organisationnelles. Ces outils, liés au travail bureaucratique, justifient leur valeur par le biais de la quantification de variables* telles que le temps gagné, le nombre de personnes concernées, etc. Ils ne s'aventurent pas dans les spécificités cliniques, ce qui rend l'expertise des ingénieurs suffisante pour développer ces solutions (l'expertise des professionnels de santé est beaucoup moins nécessaire), et permet d'éviter une conformité réglementaire lourde. De plus, ces outils d'IA peuvent s'insérer dans des modèles d'affaires classiques pour les solutions numériques.

Cependant, le mouvement de numérisation des hôpitaux portait déjà cette promesse sans qu'elle ne se soit concrétisée. Au contraire, les outils numériques (et leur adoption) exacerbent souvent certains problèmes, voire en créent de nouveaux. De plus, nous constatons que, en dehors des outils d'IA, l'informatisation du système de soins suit une tendance similaire : plus de quarante ans de développements numériques dans les hôpitaux sont consacrés aux fonctions administratives (par exemple, la facturation, la gestion des dossiers des patients...) ou techniques (la numérisation des résultats, avec l'envoi informatisé aux demandeurs, la planification des blocs opératoires...) (Béjean et al. 2015).

Malgré l'intérêt présenté par les outils d'IA administratifs, la thèse n'approfondit pas davantage ces questions, au profit des deux autres catégories présentées par la suite. En effet, si ces outils d'IA possèdent des particularités uniques dans le système de soins, ils sont semblables à des applications dans d'autres industries. En revanche, les deux types d'outils d'IA qui nous intéressent sont profondément liés au caractère existentiel du contexte de soins.

2.2.2. Les outils d'IA au cœur des soins

La deuxième catégorie proposée concerne le cœur de l'organisation du système de soins, à savoir les services cliniques, « là où se fabriquent les trajectoires des patients par la décision médicale, et où donc se structure le fonctionnement quotidien du système de production des soins » (Béjean et al. 2015). Cette catégorie regroupe les outils d'IA qui interviennent dans la pratique médicale courante, que ce soit à des fins de diagnostic, de traitement, d'assemblage ou de délivrance des soins (Minvielle 2018).

Les outils d'IA interviennent pour une grande variété d'actes de soins, dont en mammographie (Mayo et al. 2019 ; Wu et al. 2020 ; Badré et al. 2021; Sheth et Giger 2020) et en pneumologie (Kaplan et al. 2021; Almaslukh 2021; Almalki et al. 2021), mais également en ophtalmologie avec le dépistage du glaucome, pour la rétinopathie diabétique et la rétinopathie hypertensive (Gulshan et al. 2016; Poplin et al. 2018), ainsi qu'en dermatologie avec le cancer de la peau et les soins cosmétiques (Elder et al. 2021; Rundle, Hollingsworth, et Dellavalle 2021).

Ces outils d'IA sont généralement catégorisés comme des dispositifs médicaux embarquant de l'IA. Bien que la définition de dispositif médical varie selon les législations, le terme est utilisé dans de nombreux documents règlementaires. Par exemple, la Haute Autorité de Santé a mis à jour ses grilles d'évaluation pour les demandes d'accès au remboursement en septembre 2020 (avec une date de validation en février 2021 et une récente mise à jour le 17 février 2022). L'annexe répertorie 42 questions spécifiques aux outils d'IA (Haute Autorité de Santé 2022). Elle demande de préciser la finalité de l'usage, les échantillons de la population à partir desquels les données d'entraînement sont obtenues, les données d'entrée utilisées et le prétraitement effectué, le type d'apprentissage algorithmique utilisé et les étapes réalisées pour l'entraînement, les performances et la robustesse du modèle, ainsi que les éléments d'explicabilité et d'interprétabilité.

En somme, le document demande de détailler les étapes de développement de l'outil d'IA, comme cela peut être trouvé dans le modèle CRISP-DM* (voir annexe 7), tout en incluant la place des êtres humains dans ce processus. Par exemple, la question 24 demande de décrire « la manière dont les personnes intervenant dans le développement du système sont référencées », en précisant qui intervient à chaque étape. Farah et al. (2023) montrent, dans la revue de littérature de sept guides d'évaluation d'agences de régulation (dont la HAS avec le document précédent), que les principaux critères d'évaluation reposent sur les performances, l'interprétation et l'explicabilité des outils d'IA, en raison de l'opacité de leur fonctionnement qui suscite de la méfiance et entrave leur adoption.

De plus en plus de dispositifs médicaux sont autorisés sur le marché. Par exemple, à la date du 5 octobre 2022, la FDA¹⁵ répertorie 521 dispositifs médicaux embarquant de l'IA légalisés aux États-Unis. Les outils d'IA pour la quantification des images en radiologie sont l'un des cas d'application les plus matures de la technologie pour les soins courants (392 sur les 521). En particulier, l'apprentissage profond est capable de segmenter des motifs de lésions ou d'anomalies structurelles dans les images cliniques (scanners CT, radiographies, mammographies, etc.) (Jiang et al. 2017; Peng et Wang 2021). Cependant, ces outils d'IA sont généralement des boîtes noires avec peu d'informations pour que les radiologues comprennent à la démarche aboutissant aux résultats (Fazal et al. 2018).

Cette maturité s'explique en partie aussi par l'abondance de données. Tout d'abord, la radiologie est une pratique très numérisée même si des spécialités, comme l'anatomopathologie¹⁶, utilisent encore des équipements analogiques (comme des microscopes). Des projets de numérisation sont en cours d'étude, notamment à l'institut Gustave Roussy, un des centres de lutte contre le cancer Parisien. Ensuite, les données en radiologie sont relativement faciles d'accès. Pour certaines pathologies, elles peuvent s'acheter (légalement) déjà labellisées auprès de centres au Brésil. Ainsi, le grand volume de données et de variables* rend la précision des outils suffisamment intéressante pour espérer rivaliser avec les expertises humaines.

Les questions d'usage des outils d'IA apparaissent clairement le domaine de la radiologie car la transformation est en cours depuis plusieurs années (Wulczyn et al. 2021). Plus précisément, les outils d'IA sont prometteurs pour aider les radiologues à quantifier les images. Par exemple, les techniques de segmentation localisent et délimitent automatiquement les frontières des structures anatomiques ou des lésions (Peng et Wang 2021). En mammographie de dépistage, McKinney et al. (2020) montrent par une étude menée auprès de 6 radiologues que l'analyse de ces examens par l'outil d'IA surpasse celle réalisée par les experts (selon la courbe ROC*) et peut réduire la charge de travail d'un second lecteur de 88%. Ainsi, ces outils d'IA sont aussi envisagés comme des moyens de triage pour permettre aux radiologues de se concentrer sur les images les plus susceptibles d'être anormales (Lee et al. 2018; Peng et Wang 2021).

¹⁵ <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices>

¹⁶ L'anatomopathologie est l'étude morphologique des anomalies macroscopiques et microscopiques des tissus biologiques et des cellules pathologiques. Par exemple, pour le cancer du sein, c'est elle qui analyse les biopsies des tissus suspects et qui détermine la présence ou non de cancer.

Une méta-analyse a récemment comparé les performances des radiologues et des outils d'IA établis sur l'apprentissage profond pour le diagnostic établi à partir de l'imagerie (Liu et al. 2019). Si les performances diagnostiques des outils ont une précision comparable à celles des professionnels de santé, l'étude montre que les évaluations sont souvent très limitées en raison du manque de comparaisons solides entre l'humain et la machine. Par exemple, les données servant à faire les comparaisons entre radiologies et outils d'IA peuvent être différents. De plus, très peu de ces analyses sont effectuées dans le cadre d'études prospectives de la vie réelle. Ainsi, bien que les outils d'IA en radiologie semblent être techniquement au point, leur mise en œuvre réelle est sujette à caution, ce qui soulève de nombreuses questions liées à leur utilisation.

D'une part, malgré la haute performance des outils d'IA, ceux-ci commettent des erreurs (Challen et al. 2019; Ellahham, Ellahham, et Simsekler 2020). Les causes identifiées sont multiples. Par exemple, les données d'entraînement peuvent différer des données sur lesquelles l'outil d'IA s'applique, que ce soit en raison de la différence entre les populations de patients (Subbaswamy et Saria 2019) ou celle des appareils de mesure utilisés (Campello et al. 2021).

Les données d'entraînement peuvent être erronées, ce qui est d'autant plus fréquent si le contexte médical dans lequel les images ont été recueillies est incertain et que les experts qui les classifient ont eux-mêmes des opinions divergentes, ou que des informations complémentaires à celles sur l'image sont nécessaires (Lebovitz, Levina, et Lifshitz-Assa 2021). Par exemple, pour le diagnostic du cancer du sein, l'image radiologique aide à détecter des anomalies ou des caractéristiques qui suscitent des soupçons quant à la présence d'une condition médicale particulière. Cependant, cette observation seule ne constitue pas un diagnostic définitif, mais plutôt une indication de la nécessité d'une évaluation plus approfondie. Le diagnostic final et définitif est souvent obtenu par l'anatomopathologie, avec l'analyse du type de cellules présentes dans le tissu suspect.

D'autre part, les outils d'IA connaissent aussi des biais dans leurs résultats. Par conception, les outils d'IA doivent trouver un équilibre entre laisser passer plus de faux positifs (par exemple, annoncer qu'un individu est malade alors qu'il est sain) ou de faux négatifs (l'inverse). Cependant, une autre source de biais plus néfaste provient des jeux de données eux-mêmes. Les inégalités dans l'accès aux soins se retrouvent intégrées aux données d'entraînement. Les biais constatés sont souvent liés au genre, à l'âge, à l'ethnicité, aux revenus, à l'éducation et à la zone géographique. Par exemple, des études nord-américaines sur la gestion de la douleur des patients ont montré comment les professionnels de santé tendent à minimiser les plaintes des personnes non caucasiennes par rapport à celles des caucasiens (Hoffman et al. 2016). De même, pour les femmes comparées aux humains (Samulowitz et al. 2018). Ces différences de

traitement ont été constatées dans d'autres décisions algorithmiques (Obermeyer et al. 2019; Seyyed-Kalantari et al. 2020; Gichoya et al. 2022).

Les différences peuvent être liées à des facteurs externes à l'outil d'IA. Par exemple, Dai et Tayur (2022) démontrent comment le diagnostic avec l'outil d'IA peut varier selon le degré d'ambiguïté de chaque cas ainsi que le coût des visites pour les patients (le niveau de la prise en charge pouvant varier selon les systèmes des pays et des mutuelles des individus). Li, Zhou, Zhan, Xu, Zhang, Shan et Liu (2022) montrent comment la pression temporelle influence l'amélioration des performances diagnostiques des médecins grâce à l'usage de l'outil d'IA. Sans celle-ci, la sensibilité diagnostique augmente alors que la spécificité reste identique, alors lorsque s'exerce une pression temporelle sur les professionnels, leur sensibilité diagnostique et leur sont toutes deux améliorées. Cependant, ces différents facteurs et leurs conséquences sont encore mal identifiés (Dai et Tayur 2022).

En raison de ces problèmes et du contexte dans lequel les outils opèrent, le consensus tend à privilégier des usages de des outils de l'IA soumis à une garantie humaine (Hosny et al. 2018). Leurs usages dans le milieu du soin sont ainsi limités à des rôles d'aide à la décision. Même cet usage est complexe en raison du manque de transparence du fonctionnement des outils, de leurs potentiels mauvais usages par les professionnels de santé et des questions de responsabilité en suspens. Le manque de transparence des outils d'IA est en partie souhaité par les concepteurs qui cherchent à protéger leur propriété intellectuelle, malgré la demande de traçabilité constatée de la part des organismes régulateurs (European Commission 2019; Harvey et Gowda 2020).

Par exemple, l'outil d'IA pour la mammographie évoqué précédemment (McKinney et al. 2020) a généré de l'attention médiatique en raison de ses très bonnes performances, à la fois en vitesse et en qualité du diagnostic dans de nombreux contextes. Cependant l'entreprise a essuyé des critiques en raison du manque de clarté quant au manque de détails techniques, car cela diminue la sécurité et l'utilité d'un tel outil (Wiggers 2020). L'exemple a même servi à un groupe de chercheurs pour lancer un appel dans la revue *Nature* en faveur de plus de transparence dans les outils d'IA pour les soins (Haibe-Kains et al. 2020).

Le manque de transparence est aussi lié à la technologie elle-même, dont le fonctionnement reste opaque, surtout avec l'apprentissage profond. Ce phénomène est souvent nommé l'effet « boîte noire » (ou black-box) des outils d'IA (Fazal et al. 2018) : le fonctionnement et les mécanismes des décisions prises ne sont pas visibles ni compréhensibles pour les utilisateurs. Cette opacité entraîne un manque de confiance envers les résultats des outils d'IA, ce qui en limite l'adoption ou la bonne utilisation (Cui et Zhang 2021).

La mauvaise utilisation des outils d'IA provient aussi du manque d'experts médicaux dans le développement de ces outils. Ceux-ci sont souvent conçus par des ingénieurs et des data scientists, qui entretiennent très peu d'interactions avec les utilisateurs finaux. Les professionnels de soins doivent alors s'adapter à des outils dont le fonctionnement peut leur échapper, ce qui accroît les risques d'utilisation incorrecte. Le problème est exacerbé par le manque de formations liées à l'utilisation de ces technologies. Par exemple, un questionnaire outre-Manche mené auprès de près de 500 étudiants provenant de 19 écoles médicales montre qu'aucun n'a reçu d'enseignement spécifique sur l'IA dans leurs cours obligatoires (Sit et al. 2020).

Actuellement, il existe plusieurs incitations (D. Li, Kulasegaram, et Hodges 2019; Paranjape et al. 2019) qui visent à améliorer la littératie des professionnels de santé sur les outils d'IA. Elles soulignent la nécessité de remplacer la variation idiosyncrasique par un raisonnement éclairé sur le rôle que ces outils devraient jouer dans la prise de décision clinique. Cependant, il est difficile de savoir quel type de directives les professionnels de santé devraient recevoir alors que si peu de conclusions empiriques ont été tirées sur la façon dont les outils d'IA devraient être utilisés dans la prise de décision, dont la décision clinique (Kostick-Quenet et Gerke 2022).

Certains considèrent que les outils d'IA devraient être testés de manière plus prospective, à travers une variété de contextes reflétant les usages prévus et les interactions humain-IA possibles (Gerke et al. 2020; E. Wu et al. 2021). De même, les régulateurs devraient, au moins dans certains cas, souligner l'importance de bien concevoir des essais cliniques pour tester les facteurs humains et autres résultats de l'usage des outils d'IA dans des contextes réels (Babic et al. 2021).

Le manque de transparence rend plus difficile l'identification des sources d'erreurs, ce qui empêche les experts de les contourner et d'attribuer les responsabilités lors de mauvaises décisions. La régulation concernant les erreurs liées aux outils d'IA n'est pas encore établie, ce qui génère de nombreux doutes et inquiétudes parmi les utilisateurs, en particulier les médecins qui sont les premiers responsables en cas d'erreur médicale.

Ce manque réglementaire pose également des défis à long terme. Certains considèrent qu'à mesure que les outils d'IA se perfectionnent, les médecins devraient être tenus responsables des dommages aux patients s'ils font des suggestions qui s'écartent de ce que propose la machine (Price, Gerke, et Cohen 2019; Russell et Norvig 2021). Cependant, les médecins pourraient intentionnellement éviter d'utiliser les outils d'IA — même lorsque ces outils pourraient aider à atténuer l'incertitude en fournissant un deuxième avis — car leur usage augmenterait leur responsabilité (Price, Gerke, et Cohen 2019; Dai et Tayur 2022).

D'un point de vue commercial, nous avons remarqué que certains cas d'usage sont très saturés. Par exemple, trois startups francophones, Milvue, AZmed et Gleamer, commercialisent des outils d'IA pour le diagnostic des fractures osseuses. Les trois entreprises mentionnées, ainsi que la majorité des entreprises qui développent des outils d'IA pour la pratique médicale, sont fondées à la fois par des médecins et des data scientists, ou, dans le cas des grands groupes, possèdent des équipes aux expertises mixtes dans des rôles de décision. Le volet médical manque souvent aux entreprises qui se sont (re)orientées vers les outils d'IA administratifs.

Cette tendance peut s'expliquer par la propension des ingénieurs (et de ceux qui ne connaissent pas le système de soins) à « simplifier les problèmes », comme le fait remarquer Steve, un expert en IA à l'Université de Pennsylvanie (IMP4, voir annexe 5). Il poursuit en disant : « *L'opinion commune est de croire qu'il faut laisser l'expertise à des entreprises tierces et que l'hôpital doit ensuite les mettre en place. Ma position, biaisée par mon parcours hybride entre sciences des données et expertise clinique, me pousse à défendre l'idée que nous [à l'hôpital] devons développer une expertise en interne. Nous observons des détails que des personnes extérieures sont incapables de remarquer.* »

De plus, la mise en place réelle de ces applications suscite de nombreuses interrogations éthiques, comme en témoigne l'avis commun « Diagnostic Médical et Intelligence Artificielle : Enjeux Éthiques » du Comité national pilote d'éthique du numérique (CNPEN) et du Comité consultatif national d'éthique pour les sciences de la vie et de la santé (CCNE) de novembre 2022. Saisi par le Premier ministre en juillet 2019, le CNPEN a inclus le CCNE dans sa réflexion. Ils ont abouti à une présentation holistique des questions en cours avec 16 recommandations et 7 points de vigilance concernant « les enjeux d'éthique des systèmes d'intelligence artificielle appliqués au diagnostic médical visant l'amélioration des performances techniques des praticiens » (CCNE et CNPEN 2022, 5). En particulier, leur bénéfice pourrait être le plus important dans les centres moyens, c'est-à-dire ceux qui manquent d'une expertise pointue pour pallier le manque de compétences spécialisées.

En somme, s'il est nécessaire de prendre du recul par rapport aux résultats fournis par l'outil d'IA, il est recommandé de ne pas négliger les avantages apportés par ces outils. Ils insistent sur la nécessité de les mobiliser en priorité dans une optique d'amélioration du soin, et non dans un but organisationnel, économique ou managérial. À cet égard, les évaluations réglementaires doivent évoluer pour démontrer à la fois l'innocuité de l'outil et son efficacité clinique. De plus, la garantie humaine semble un élément qui reste indispensable et qui nécessite des éléments explicatifs suffisants quant à son fonctionnement.

L'ensemble de ces constats montre le besoin d'établir un modèle de collaboration entre les professionnels de soins et les outils d'IA. Les outils d'IA se multiplient dans le domaine des soins, apportant une forme d'expertise professionnelle, par exemple avec une forte précision

dans le diagnostic. Cependant, ils restent faillibles, notamment face à l'unicité des situations rencontrées, et leurs résultats sont difficilement interprétables par les utilisateurs. Même si le cadre réglementaire reste à préciser, la responsabilité principale repose sur le médecin. Cela demande au professionnel d'établir un rapport avec les résultats des outils d'IA qui soit bénéfique à son activité. Il a besoin de gérer cette nouvelle technologie qui apporte un nouveau type de connaissances. Par conséquent, la perspective de la collaboration humain-machine est pertinente à l'analyse.

2.2.3. Les outils d'IA dans la recherche médicale

La troisième catégorie des usages des outils d'IA concerne la recherche médicale. Par exemple, la Dr. Delalogue¹⁷ témoigne des avancées en oncologie grâce à des études menées sur des cohortes de patients. Le programme Interception qu'elle dirige démontre comment les outils d'IA créent des modèles statistiques complexes pour identifier les risques individuels de développer un cancer. Ce projet de recherche translationnelle va au-delà de l'objectif habituel de la recherche médicale utilisant l'IA. En plus de générer des connaissances et des publications scientifiques, il propose un moyen d'améliorer les soins courants. Cet exemple illustre la connexion entre ces deux types d'utilisation : la recherche médicale alimente les outils d'IA pour trouver des réponses, et à la fin de ce processus, les outils d'IA résultants sont utiles dans les soins courants.

En d'autres termes, la recherche médicale soutient les soins courants en actualisant les pratiques et en ouvrant la voie à de nouvelles possibilités d'action. Les outils d'IA font partie intégrante de ce processus en fournissant aux acteurs de nouvelles capacités de traitement des données et, en fin de compte, une meilleure compréhension de la situation (Cui et Zhang 2021).

La forme de ces outils diffère de celle des deux usages précédents. Pour les tâches administratives et les soins courants, les outils d'IA sont déjà stabilisés, c'est-à-dire qu'ils sont entraînés et validés. Dans la recherche médicale, ils sont en cours de développement. Les outils sont créés pendant le processus de recherche. Au départ, ce sont des fonctions algorithmiques « vierges », disponibles dans des bibliothèques (*packages*) où le code générique est déjà écrit. Ils doivent ensuite être entraînés avec des données collectées dans le cadre de cohortes de recherche, du système d'information ou d'autres sources telles que les réseaux sociaux ou des dispositifs de collecte de données spécialisés. Ces données sont ensuite traitées et modélisées par de nombreux algorithmes, dont les data scientists ajustent les

¹⁷ Présentation lors du séminaires Polytechnique Santé le 25 mai 2023

hyperparamètres* pour adapter le code générique à la situation spécifique. Dans le meilleur des cas, des modèles statistiquement valides sont obtenus, puis ils sont évalués sur le plan clinique.

Ces usages sont de plus en plus courants dans la recherche médicale académique. En plus des statistiques descriptives et de la recherche médicale traditionnellement menées par des médecins ayant des compétences en statistiques ou des biostatisticiens, l'utilisation de techniques d'apprentissage automatique plus avancées nécessite une collaboration entre des médecins et des data scientists. Les modèles ainsi générés sont plus complexes et qualifiés de « prédictifs ».

De même, le domaine pharmaceutique adopte les outils d'IA pour accélérer le processus de découverte de médicaments, c'est-à-dire pour identifier des molécules ayant un effet bénéfique sur l'organisme humain. Les récents progrès dans la matière font la une des journaux. Par exemple, depuis juin 2023¹⁸, la molécule INS018_055, développée par Insilico Medicine, représente une première historique en entamant les essais cliniques sur des sujets humains.

La découverte de nouveaux médicaments est une entreprise complexe, marquée par d'intenses activités de recherche et des interactions sophistiquées de cause à effet (Vamathevan et al. 2019; Webel et al. 2020). Dans cet environnement, les outils d'IA jouent un rôle crucial en évaluant les molécules candidates pour déterminer les plus prometteuses. Ils accomplissent cela, entre autres, en identifiant des sous-structures chimiques qui peuvent indiquer la toxicité d'une molécule, nécessitant une formation sur des ensembles de données comprenant d'importantes quantités de représentations moléculaires et leurs effets pharmacologiques connexes (Webel et al. 2020).

Si les outils d'IA peuvent identifier des sous-structures moléculaires précédemment ignorées (Preuer et al. 2019; Webel et al. 2020), l'intervention humaine demeure indispensable. D'abord, les chimistes et les data scientists jouent un rôle crucial en configurant les données et les algorithmes, sélectionnant des ensembles de données pertinents en fonction de leur expertise (Schneider et al. 2020).

À un niveau plus exploratoire, situé à la frontière entre la recherche et les usages administratifs, les outils d'IA sont également envisagés pour faciliter la sélection des patients éligibles aux essais cliniques. Par exemple, un médecin et chercheur spécialisé dans les outils d'IA (IMP8, voir annexe 5) nous explique que pour identifier des patients répondant aux

¹⁸ <https://www.cnn.com/2023/06/29/ai-generated-drug-begins-clinical-trials-in-human-patients.html>

critères d'éligibilité aux essais cliniques, il est nécessaire de consulter de nombreux documents, y compris des comptes rendus médicaux contenant des données non structurées. Ce travail prend en moyenne une heure par patient, ce qui limite la portée de la recherche en raison des ressources limitées pour effectuer cette sélection. En partenariat avec l'industrie, des outils d'IA sont développés pour simplifier cette étape.

Plus généralement, l'un des avantages les plus convaincants des outils d'IA dans la recherche est de réduire l'incertitude des situations et de transférer ces connaissances, voire les outils d'IA eux-mêmes dans les soins courants. Par exemple, en calculant une estimation personnalisée selon laquelle l'état d'un patient va se détériorer après X temps ou bénéficiera d'une survie prolongée de Y années après une intervention. Cependant, le succès des outils d'IA dans la réduction de l'incertitude dépend encore largement de la manière dont les estimations sont interprétées et mises en œuvre (Kostick-Quenet et Gerke 2022). La question des usages de ces outils demeure donc centrale.

A l'instar des usages des outils d'IA dans les soins courants, nous retrouvons ici le besoin d'établir un mode de collaboration entre les experts et les outils d'IA. Ces outils prennent une place indispensable au processus d'émergence de connaissances, en complément des chercheurs. En particulier, l'incertitude des résultats des outils d'IA et le besoin d'interprétation et d'explication par les professionnels de santé demande d'établir un cadre de travail entre la machine et l'humain.

2.2.4. Les outils d'IA et les patients

L'objectif du système de soins est d'améliorer les résultats et les expériences des patients. Dans ce sens, une approche des soins centrée sur le patient (*Patient-Centered Care*) s'est imposée ces dernières années, prenant davantage en compte leurs expériences (Gluyas 2015) et les considérant de manière croissante comme des acteurs dans la prise en charge de leur santé. Par exemple, les plateformes numériques de feedback des patients peuvent servir non seulement comme des moyens d'échanger des connaissances entre les hôpitaux et leurs patients, mais également entre les hôpitaux et les patients et/ou gestionnaires des hôpitaux concurrents (Bez, Georgescu, et Farazi 2022).

Certains considèrent alors qu'il faudrait concevoir des usages d'outil d'IA en gardant à l'esprit l'expérience des patients (Dai et Tayur 2022), voire de faire des patients leurs utilisateurs principaux.

Dans cette perspective, Auerswald (2017) soutient que « *les plus grandes avancées dans la production de services de soins viendront d'une combinaison de technologies portables, de diagnostics soutenus par des applications des Données Massives, d'opérations de pair à pair et d'autres innovations dans le code algorithmique qui distribuent la prestation de soins loin des modèles hautement centralisés qui ont dominé au vingtième siècle.* » (Auerswald 2017, 160–

161). Par exemple, l'outil d'IA DBLG1 développé par Diabeloop est connecté à une pompe à insuline et à un capteur de glucose via Bluetooth, permettant un traitement du diabète plus personnalisé.

Les expérimentations dans la matière restent cependant encore très balbutiantes. Dumez et Minvielle (2017), explorent la question au niveau de l'e-santé, en analysant comment elle peut influencer la démocratie sanitaire. Ils identifient deux effets principaux : une co-production améliorant la relation patient-professionnel et une expertise patient accrue grâce à l'information en ligne.

Toutefois, il existe également des risques, notamment la mauvaise utilisation des informations trouvées par les patients qui pourrait être source de troubles au sein de la relation médecin-patient (Dumez et Minvielle 2017). Ces craintes sont également exprimées pour les usages des outils d'IA, qui peuvent modifier les manières de prendre des décisions médicales de nombreuses manières, notamment en excluant le professionnel de santé (Triberti, Durosini, et Pravettoni 2020).

Ainsi, les usages des outils d'IA pour les patients présentent un potentiel considérable pour transformer la manière dont les soins sont dispensés et reçus. En mettant l'accent sur la personnalisation, l'efficacité et l'accessibilité, les outils d'IA peuvent contribuer significativement à améliorer les résultats pour les patients et à rendre le système de soins plus résilient et réactif.

Cependant, il est crucial que cette intégration se fasse en portant une attention particulière aux implications éthiques, à la sécurité des données, et à la préservation de la relation de confiance entre les patients et les professionnels de santé, ce qui demande un dialogue continu entre les différentes parties prenantes.

Parmi les nombreux défis des usages des outils d'IA, leur introduction dans le système de soins met en exergue la tension entre la standardisation (la production de masse à bas coût) et la personnalisation (l'adaptation de la production aux besoins particuliers). Cette tension de la production se retrouve plus largement dans l'idée d'une nouvelle Révolution Industrielle et réussir à la gérer pourrait répondre à des questionnements transversaux sur l'apport des usages des outils d'IA, dont la mitigation des risques de sous-utilisation, surutilisation ou mauvaise utilisation (Fihn et al. 2019). Prendre en compte les spécificités du contexte du système de soins pour développer des usages des outils d'IA adaptés pourrait aider à développer une collaboration humain-machine efficace et responsable.

Parmi les quatre grands types d'usages des outils d'IA que nous identifions, les tâches administratives sont éloignées du cœur de métier. Elles développent des moyens pour améliorer l'efficacité et la qualité de la production grâce à la gestion de l'information. Ces solutions sont plus accessibles pour les concepteurs, car elles requièrent moins de

connaissances médicales et sont soumises à moins d'exigences réglementaires. De même, les usages des outils d'IA pour les patients sont intéressants, mais soit ils s'assimilent à des dispositifs médicaux classiques, soit ils rentrent dans une logique très différente du système de soins actuel et sont alors à des stades trop embryonnaires pour les analyser.

En revanche, les usages des outils d'IA pour les soins courants et la recherche médicale mettent en avant les spécificités du système de soins et, en particulier, ils soulèvent une question de la collaboration entre l'humain et la machine particulière : comment les usages des outils d'IA peuvent-ils potentialiser les capacités des experts métier dans un contexte de haute professionnalisation, d'incertitude et de prises de décisions critiques ?

Chapitre 2 : Quels modes de gestion pour la collaboration entre humains et outils d'IA ?

Les progrès techniques conduisent à sortir les outils d'IA fondés sur l'apprentissage machine des laboratoires de recherche pour les intégrer dans les organisations. Si les processus d'adoption touchent déjà à leur fin pour bon nombre d'entre eux, différentes questions d'usage restent à traiter, dont les modes de gestion de la collaboration humain-machine.

Dans ce chapitre, nous explorons pourquoi nous sommes à un moment charnière pour comprendre et établir des modes de collaboration entre humains et outils d'IA (1). Nous présentons alors les connaissances dont nous disposons sur le sujet (2). La revue de littérature se conclut sur la nécessité de poursuivre les recherches sur la collaboration entre l'expert métier et l'outil d'IA afin d'identifier comment elle bénéficie à l'utilisateur (3).

1. Un moment charnière pour comprendre les collaborations entre humains et outils d'IA

Après une vue d'ensemble sur l'évolution des connaissances des outils d'IA (1.1), nous présentons une première perspective des sciences sociales des années 2010 centrée sur la gestion algorithmique (1.2). Ensuite, nous explorons l'approche qui émerge depuis les années 2020. Avec le débat sur les apports des outils d'IA, l'attention se centre davantage sur les questions de collaboration humain-machine (1.3).

1.1. Évolution des connaissances sur les outils d'IA

Après une présentation succincte sur la progression des connaissances sur les outils d'IA (1.1.1), nous expliquons pourquoi les sciences de gestion sont en mesure d'apporter des contributions essentielles à ce sujet (1.1.2).

1.1.1. Des développements techniques à l'utilisation des outils d'IA

Le chapitre 1 illustre l'évolution de la recherche en IA, depuis les techniques fondamentales à des applications qui affectent les organisations et la société. Cette recherche en IA a surtout été axée sur le développement et le raffinement d'algorithmes, jetant les bases pour les applications futures. Elle a connu différentes ères marquées par des avancées et des revers, pouvant se réduire à l'approche symbolique dont les systèmes experts en ont fait l'apogée et sa fin, puis la période actuelle, dominée par les techniques d'apprentissage machine (Salehi et Burgueño 2018).

À mesure que les technologies d'IA mûrissent, l'attention de la recherche s'est déplacée vers les différents domaines d'application et la compréhension de leurs implications sociétales. Ce changement appelle à des approches interdisciplinaires, combinant l'informatique avec les sciences sociales pour relever des défis sociétaux complexes.

Or, nous remarquons que la production de connaissances sur l'IA est avant tout issue des sciences dures. En réalisant une série de requêtes à propos de la littérature sur l'IA sur la base de données Scopus, nous observons que les mots-clés « artificial intelligence » ou « AI » dans le titre seulement, rédigés en anglais, publiés dans des revues à comité de lecture, entre 2011 et le 1^{er} août 2023, renvoient à : 8 440 articles dans le domaine des sciences de l'informatique, 6 351 articles dans les domaines de l'ingénierie, 5 674 articles en médecine, 3 948 articles en sciences sociales et 2 082 en « Business, Management and Accounting ».

En incluant des termes plus techniques, comme « Machine Learning » ou les noms d'approches et d'algorithmes plus précis, comme « Random Forest », les sciences sociales, dont les sciences de gestion sont encore moins représentées. La littérature regorge de présentations d'algorithmes optimisés pour des contextes d'utilisation et des domaines particuliers, tels que l'ingénierie et la médecine.

Bien que les sciences dures disposent de ressources financières et humaines différentes, ainsi que de processus de publication distincts, ces résultats mettent en évidence que la majeure partie de la production académique sur le sujet est d'ordre « technique ».

Ainsi, au cours des deux dernières décennies, l'IA a ainsi connu une croissance exponentielle, passant de cadres théoriques à des applications pratiques affectant divers secteurs, y compris la santé, la finance et l'éducation (Dwivedi et al. 2021). C'est alors au tour des sciences sociales, notamment les sciences de gestion, d'étudier la mise en place, les usages des outils d'IA et leurs conséquences.

1.1.2. Positionnement en sciences de gestion

Les effets des outils d'IA sont un sujet en plein essor en sciences de gestion, en particulier en Théorie des Organisations et en Gestion des Systèmes d'Information. Ce domaine permet d'éclairer comment les usages de l'IA modifient les comportements organisationnels, les processus de prise de décision et les systèmes d'information.

Parmi les ouvrages les plus anciens en sciences de gestion sur les outils d'IA, nous trouvons *The Second Machine Age* de Brynjolfsson et McAfee (2014). Les auteurs s'appuient sur les effets qu'ils provoquent au moment de la publication pour discuter des changements à venir dans de nombreux secteurs, dans la société et dans l'économie en général. Ils fondent leur argumentation sur l'idée d'une accélération de la technologie qui finira par surpasser les capacités humaines.

Cette idée est reprise par Huang et Rust (2018). Ils observent également que les outils d'IA automatisent déjà des tâches mécaniques, et commencent à s'attaquer aux tâches analytiques. Ils estiment que si les outils d'IA assistent les travailleurs, cette phase est transitoire, avant que l'évolution technologique ne finisse par remplacer totalement l'individu sur une tâche donnée. En poursuivant cette tendance, ils suggèrent que les outils d'IA pourraient finir par suppléer à l'humain pour les tâches impliquant l'intuition et l'empathie.

D'autres travaux d'ordre prospectif ont suivi cette idée. Par exemple, l'article de Kaplan et Haenlein (2019) propose une définition de l'IA fondée sur trois étapes de développement technologique : l'IA faible, qui effectue des tâches très spécifiques, correspondant aux outils d'IA actuels ; l'IA générale, comparable à l'intelligence humaine dans sa capacité à s'adapter à différents environnements ; et la super IA, ayant une conscience et nettement plus performante pour tout. Les auteurs élaborent des taxonomies sur les usages des outils d'IA en fonction de ces trois niveaux technologiques. Dans un autre article, Haenlein et Kaplan (2019) évoquent également l'importance que prendra la régulation sur l'IA.

Ainsi, les prédictions suggèrent que les outils d'IA pourraient considérablement modifier les organisations et le travail, notamment celui des experts métier (Davenport et Kirby 2016; Wirtz et al. 2018; Jaiswal, Arun, et Varma 2022; Tschang et Almirall 2021).

Au fur et à mesure que ces outils ont été introduits dans les organisations, la littérature en sciences de gestion a évolué pour s'intéresser aux défis posés par leur intégration (Chen et al. 2023). L'une des problématiques soulevées est la régulation à partir des considérations éthiques, puisque ces outils d'IA ont des conséquences sur la vie des personnes. Les débats éthiques émergent ainsi, dans des domaines d'application comme le recrutement (Hunkenschroer et Luetge 2022) ou la santé (Morley et al. 2020), voire de manière transversale à tous les secteurs (Ashok et al. 2022).

De même, de nombreux articles s'intéressent aux conséquences de l'adoption des outils d'IA dans des domaines plus ou moins précis allant de l'agriculture (Spanaki et al. 2022) aux chemins de fer (Tang et al. 2022). À cet égard, les revues de littérature sur les effets de l'adoption des outils d'IA dans les domaines de la santé sont particulièrement nombreuses (Becker 2019; Carlin et al. 2021; Khanijahani et al. 2022; Galetsi, Katsaliaki, et Kumar 2019; Sukums et al. 2023; Mahdi et al. 2023). Par exemple, la revue de littérature de Morley et al. (2020) identifie trois types de défis éthiques spécifiques aux outils d'IA dans le système de soins, allant de l'individuel au sectoriel.

Certains défis sont épistémiques. Ces problèmes concernent la fiabilité des données et des résultats algorithmiques. Les auteurs soulignent que les résultats des outils d'IA sont fondés sur des données disponibles et sont de nature probabiliste, ce qui les rend faillibles et non causaux. De plus, leur fonctionnement est souvent difficile à comprendre pour les utilisateurs.

Il convient d'anticiper les conséquences sociétales introduites par l'utilisation des outils d'IA. Par exemple, ils pourraient accroître certaines inégalités sociales et affecter profondément certains groupes de personnes. En outre, les outils d'IA pourraient induire des transformations imprévues de notre compréhension de la réalité, remettant en question les normes et valeurs préétablies.

Il existe également des enjeux sur la responsabilité et la réparation des préjudices causés par les outils d'IA. Compte tenu des difficultés de tracer les effets causés par l'utilisation des outils d'IA, il semble difficile d'identifier les parties responsables.

Outre les considérations éthiques, nous retrouvons souvent des préoccupations relatives à la confidentialité des données et au besoin d'une compréhension plus approfondie de l'interaction humain-IA.

Tout comme les connaissances sur les outils d'IA suivent l'évolution des avancées empiriques, la recherche en sciences de gestion sur les usages des outils d'IA co-évolue avec les expérimentations sur le terrain. Ainsi, nous remarquons que la littérature s'est concentrée d'abord sur la perspective de la gestion algorithmique, en lien avec un discours des acteurs centré sur l'automatisation et le remplacement de l'humain. Après avoir présenté cette perspective, nous évoquerons comment ce discours change dans les années 2020 pour accentuer les aspects de la collaboration humain-machine.

1.2. Une première perspective sur la gestion algorithmique

La « gestion algorithmique » ou « gouvernance algorithmique » renvoie à l'utilisation des outils d'IA pour automatiser et optimiser la prise de décisions et la coordination au sein des organisations et des systèmes socio-économiques. La gestion algorithmique se décompose en deux fonctions qui étaient auparavant du ressort des gestionnaires : la coordination algorithmique et le contrôle algorithmique (Cram et Wiener 2020; Möhlmann et al. 2021; Benlian et al. 2022).

Cette gestion algorithmique provient de la capacité des outils d'IA à collecter et à analyser des données à une échelle inédite pour prendre des décisions de coordination et de contrôle. Selon Kitchin (2017, 15), cette tendance marque l'avènement d'une « ère de gouvernance algorithmique généralisée », où le contrôle et la coordination algorithmiques remplacent progressivement l'intervention humaine dans de nombreux processus décisionnels. Par conséquent, la gestion algorithmique met en exergue la substitution des tâches par les outils d'IA. L'évolution vers la gestion algorithmique s'accompagne alors de questions autour de la transparence, de la responsabilité et de l'éthique de la prise de décision automatisée.

Cette perspective trouve ses sources dans les études des sciences et des techniques depuis la fin des années 1980, notamment les travaux de la sociologue Zuboff (1988; 2019). À cette

époque, Zuboff publie un ouvrage clé qui dresse un état des lieux de la numérisation des métiers dans divers lieux de travail (bureaux, usines, ateliers d'artisans, etc.) sur plusieurs années (Zuboff 1988). Elle explique comment la technologie informatique joue un rôle clé en raison de sa capacité à reconfigurer le champ des possibles. Les nombreux thèmes abordés conservent leur pertinence pour analyser la situation actuelle (Burton-Jones 2014).

Malgré ces possibles, quarante ans après, Zuboff (2019) expose un usage plus critique, mettant en lumière le phénomène de contrôle algorithmique (une des deux composantes de la gestion algorithmique). Elle explique ainsi qu'au XX^{ème} siècle la tension principale portait sur le capital et le travail, mais que la ligne de front s'est déplacée au XXI^{ème} siècle vers la surveillance de masse rendue possible par les algorithmes et les outils d'IA en particulier.

La perspective de Zuboff s'inscrit dans une littérature qui prend forme au milieu des années 2010. Elle s'appuie sur des études dans le secteur de la finance qui connaît à cette époque une adoption massive des outils d'IA. Par exemple, Lenglet (2011), MacKenzie (2014), Arnoldi (2016) et Pasquale (2015) montrent comment les usages des outils d'IA ont profondément restructuré tous les aspects du fonctionnement du secteur financier par des formes de coordination algorithmique et de contrôle algorithmique, dont la manière de classer les clients et d'attribuer des crédits.

Plus généralement, les outils d'IA deviennent performatifs (Muniesa 2014), c'est-à-dire qu'ils produisent les effets qu'ils annoncent. En d'autres termes, la gestion algorithmique est une forme de gestion fondée sur une automatisation des tâches par des machines, imposant leur réalité sur les humains.

Face au déploiement massif de ces usages, une communauté académique interdisciplinaire (études des sciences et des techniques, études sur les médias, Gestion des Systèmes d'Information, etc.) propose une étude critique des algorithmes, dont les outils d'IA d'apprentissage machine. Le but est de mieux comprendre la nature des algorithmes et leur pouvoir d'action pour les déconstruire.

Nous présentons les résultats de ces recherches en abordant la politique de la gestion automatisée (2.1.1). Nous évoquons ensuite les controverses en réponse à ces usages (2.1.2). Nous concluons en illustrant par le concept de la gouvernamentalité comment ces études peuvent enrichir la théorie (2.1.3).

1.2.1. La politique de la gestion automatisée

Selon les concepteurs, l'outil d'IA serait un moyen impartial offrant un résultat objectif. En outre, ils estiment que les échanges économiques nécessitent des « systèmes de contrôle » (Granovetter 2005; Wood 2018). Or, comme le soutiennent Rosenblat et Stark (2016), « *les espaces numériques facilitent et soutiennent de nouveaux systèmes de surveillance et*

opportunités de contrôle à distance sur les travailleurs. »¹⁹ Ainsi, des systèmes de notation et de réputation fondés sur les plateformes (M. K. Lee et al. 2015; Rosenblat et Stark 2016) engendrent une gestion algorithmique.

Leur impartialité supposée est argumentée par l'absence de décideurs humains : les algorithmes prennent des décisions de manière autonome et automatisée. Les outils d'IA offrent une « méthode systématique ou automatique de calcul »²⁰ (Steiner 2012, 55), une gestion automatisée (Dodge et Kitchin 2007). Ces outils d'IA apportent une nouvelle façon de gouverner (Introna 2016; Kitchin 2017). Par exemple, des outils d'IA utilisant l'apprentissage profond sont testés pour la gestion de portefeuilles bancaires, de façon automatique et sans intervention humaine. Il en est attendu que la prise de décisions soit plus rapide et plus fiable (Heaton, Polson, et Witte 2017).

Allant dans le sens de la vision des concepteurs, une grande partie de la recherche s'est efforcée de montrer que ce type d'usage des outils d'IA capte de la valeur économique de nombreuses façons, comme une meilleure prise de décision, de coordination et d'apprentissage organisationnel (Kellogg, Valentine, et Christin 2020). Par exemple, les nombreuses données sur des clients permettent aux entreprises de développer des outils d'IA pour améliorer leur marketing (Hollebeek et al. 2016; Gomez-Uribe et Hunt 2016). De même, des outils d'IA servent à automatiser une grande partie des tâches de coordination entre employés (Little et al. 2010). Par exemple, en agissant comme une interface entre les vendeurs et le personnel en production, les outils d'IA s'assurent des stocks et permettent de prendre en compte des demandes plus personnalisées à des prix de production moindres (G. F. Davis 2015; 2016). Enfin, des systèmes sont mis en place pour identifier les informations importantes à travers les comportements des utilisateurs afin d'améliorer les produits et services en temps réel (Y.-E. Liu et al. 2014) grâce à des boucles de rétroaction* (Nikolaidis et Shah 2013; Sachon et Boquet 2017).

Les nombreuses études sur la plateforme Uber, une application visant à mettre en contact des utilisateurs et des voitures de transport avec chauffeur, illustrent la complexité de saisir toutes les facettes de la gestion algorithmique. Cette technologie semble souvent bien accueillie par les travailleurs qui se sentent protégés du non-paiement et du vol de salaire grâce à la garantie de paiement offerte par la plateforme (D'Cruz et Noronha 2016). Rosenblat et Stark (2016) démontrent également que la gestion algorithmique permet à Uber d'adopter

¹⁹ *digital spaces facilitate and scaffold new systems of monitoring and opportunities for remote control over workers*

²⁰ *method of systematic or automatic calculation*

un contrôle « doux ». Pourtant plus récemment, Möhlmann et Henfridsson (2019) font remonter trois types de plaintes des chauffeurs Uber contre ces systèmes de gestion algorithmique : la surveillance constante, la déshumanisation et le manque de transparence. Au total, les études convergent vers un bénéfice de la gestion algorithmique pour les entreprises de plateformes, tandis qu'elle crée des tensions dans l'environnement de travail avec de possibles impacts négatifs sur le bien-être des travailleurs (Gal, Jensen, et Stein 2020; Jarrahi et al. 2021; M. Wiener, Cram, et Benlian 2023; Möhlmann et al. 2023).

Ainsi, les critiques du discours pro-gestion algorithmique soulignent que les avantages attendus sont liés à la préservation du pouvoir (Zuboff 2019). La gestion algorithmique sert à mieux capter la valeur, qu'elle soit symbolique, économique, etc. (Introna 2016; Kitchin, 2017; Zuboff 2019). Plus précisément, les usages des outils d'IA servent de moyen pour accroître l'efficacité de l'accumulation de la valeur grâce aux possibilités d'automatisation qu'ils offrent à l'exercice du pouvoir (Musiani 2013), au détriment de ceux sur qui ces outils agissent (Kellogg, Valentine, et Christin 2020).

En somme, la politique de la gestion algorithmique propose une automatisation pour que les acteurs en position de pouvoir maintiennent leurs privilèges. La coordination algorithmique et le contrôle algorithmique sous-jacents agissent de façon performative, en rendant vrai ce qu'ils annoncent, en imposant aux travailleurs une réalité formulée par les concepteurs au service du capital. Cet instrument d'exercice du pouvoir est d'autant plus fort qu'il repose sur l'opacité des outils d'IA.

L'opacité renforce le fonctionnement des outils d'IA dans la perspective de la gestion algorithmique (Diakopoulos 2015; Lash 2007; Zuboff 2019; Kellogg, Valentine, et Christin 2020). Selon Lash (2007), les algorithmes introduisent un nouvel ensemble de normes qui complètent les règles constitutives et régulatrices. Ces règles sont inscrites dans le code algorithmique. Elles sont d'autant plus efficaces qu'elles restent dans la boîte noire de l'algorithme, qui les applique de manière automatique. Le résultat produit est alors difficilement contestable, car difficile à comprendre.

En réaction, la communauté scientifique appelle à ouvrir la boîte noire des outils d'IA et à déconstruire les mécanismes par lesquels ils exercent leur pouvoir (Diakopoulos 2015; Zuboff 2019). Par exemple, Thrift et French (2002, 331) proposent de « *comprendre les nouvelles formes de politique technologique et les nouvelles pratiques d'invention politique, une*

légitimité et une intervention que nous commençons à peine à comprendre comme politiques : des politiques de normes, de classifications, de mesures et d'analyses... »²¹.

Une des motivations clé de cette recherche est de rendre la gestion algorithmique plus éthique. Au lieu de s'y opposer frontalement, Möhlmann (2021, 1) et Wiener et al. (2023) expliquent que « *les incitations algorithmiques ne doivent pas être non éthiques* », et que la gestion algorithmique ne doit pas être considérée comme une fatalité pour les travailleurs. La compréhension de ces usages pourrait amener à développer des formes plus éthiques de gestion algorithmique.

1.2.2. Des applications contestées

Malgré la possibilité de construire une gestion algorithmique éthique, de nombreuses applications se font au détriment des travailleurs. Par exemple, ces outils sont utilisés dans le contexte de la « *gig economy*²² » via des services numériques médiatisés par des plateformes de travail en ligne. Ces plateformes de travail sont devenues centrales aux pratiques d'emploi moderne (Taylor et al. 2017). Ces outils proposent deux types d'emploi : un « *travail à distance* » livré directement via des plateformes, et un « *travail local* », transigé via des plateformes, mais nécessitant la présence physique du travailleur (Huws, Spencer, et Joyce 2016).

Concernant l'effet sur les travailleurs, Wood et al. (2019) constatent que la gestion algorithmique dans la « *gig economy* » offre des niveaux élevés de flexibilité, d'autonomie, de variété et de complexité des tâches. Cependant, les salaires sont bas, avec des horaires de travail irréguliers et ils marginalisent les travailleurs, qui peuvent éprouver le surmenage, le manque de sommeil et l'épuisement.

En s'intéressant de plus près à la composante du contrôle algorithmique au-delà de la *gig economy*, Kellogg, Valentine, et Christin (2020) montrent qu'il combine et transcende le contrôle technique (supervision par la technique) et le contrôle bureaucratique (généralisation des règles) du siècle dernier. Les employeurs ont à leur disposition différents moyens pour contrôler le travail de leur main d'œuvre. Ils dirigent les travailleurs en recommandant les

²¹ to understand new forms of technological politics and new practices of political invention, legibility and intervention that we are only just beginning to comprehend as political at all: politics of standards, classifications, metrics, and readings...

²² L'économie des petits boulots, à la tâche ou à la demande, est un système dans lequel les emplois sont flexibles, temporaires ou indépendants. Elle repose sur le concept du travail à la tâche, également appelé travail à la pièce, et peut être perçue comme un retour au travail de tâcheron du capitalisme originel par opposition au travail salarié. Son essor au XXI^e siècle implique souvent une connexion à des plateformes en ligne.

actions ou décisions à prendre, ou en restreignant l'accès à l'information et aux possibilités d'action. De plus, ils évaluent les travailleurs en enregistrant leurs activités et en effectuant des comparaisons (notation fondée sur de nombreuses variables de productivité), ou ils les disciplinent en remplaçant les employés moins performants, parfois en masquant cette violence par la ludification du travail.

Cependant, les acteurs ne sont pas complètement démunis face à la gestion algorithmique. Dans la partie suivante, nous détaillons l'étude de cas d'Introna (2016) qui montre la capacité d'adaptation des étudiants dont le travail est surveillé par un outil d'anti-plagiat. En plus des possibilités d'adaptations, Kellogg, Valentine, et Christin (2020) constatent que les employés soumis à des algorithmes de gestion algorithmique développent des tactiques de contestation et de résistance qu'elles appellent « algoactivisme ». Ils jouent avec l'opacité de l'outil pour créer de nouvelles zones d'incertitude, et donc des espaces de liberté. Par exemple, les forums de partage de connaissances améliorent l'autonomie des travailleurs en enseignant des pratiques de contournement de mesures disciplinaires.

1.2.3. La gouvernementalité ou l'illustration de l'apport des études sur les outils d'IA

Nous concluons notre présentation de la perspective de la gestion algorithmique en détaillant le concept de gouvernementalité proposé par Introna (2016). Celui-ci illustre comment l'étude des outils d'IA peut produire de nouvelles théories et actualiser notre compréhension de la technologie face aux progrès du numérique (Faraj, Pachidi, et Sayegh 2018; von Krogh 2018; Bailey et al. 2022; Faraj et Leonardi 2022).

Introna (2016) offre une description du fonctionnement de l'outil Turnitin pour détecter le plagiat dans le milieu universitaire. Il présente la motivation de son usage comme le fruit de l'introduction du néolibéralisme dans l'éducation, qui met en exergue la valeur marchande et la compétitivité de la production du savoir. Une série de transactions de valeur s'opère à l'université : la rédaction académique de l'étudiant se transforme en un crédit d'enseignement, ce crédit est échangé contre un diplôme, qui à son tour est un passeport pour l'emploi, etc. (Introna 2016, 33). Sous ce prisme, la détection du plagiat devient une préoccupation majeure pour s'assurer de la valeur du diplôme (Olssen 2004). La logique institutionnelle capitaliste pousse donc à la création de moyens, dont des algorithmes, pour contrôler la rédaction académique et garantir sa valeur de manière plus sécurisée. Turnitin propose donc de légitimer l'originalité de l'écrit de l'étudiant par le biais d'un calcul complexe et opaque de « similarité » entre le texte évalué et une base de données textuelle.

Pour comprendre ce qui se passe, Introna (2016) part de la notion de gouvernance, qui est étroitement liée à la gestion algorithmique (Dodge and Kitchin 2007; Kitchin 2017). Cette notion met en avant une pluralité d'acteurs, aucun d'entre eux n'ayant suffisamment de pouvoir pour imposer une décision unilatérale. Le processus de contrôle est le résultat d'un

mélange complexe d'acteurs, humains et non humains, qui identifient et résolvent ensemble des problèmes (Kooiman 2008). L'ordre social résulte donc de multiples acteurs, plus ou moins autonomes, mais interdépendants, ayant des objectifs liés, voire divergents (Chhotray and Stoker 2009). De même, le pouvoir d'action des outils d'IA n'est pas suffisant en soi, comme le montrent les adaptations ou les résistances des acteurs. L'enjeu est donc de comprendre comment ces agents de contrôle et de coordination interagissent avec les autres acteurs.

Pour aller plus loin, Introna (2016) propose de mobiliser le concept de gouvernementalité pour examiner le « comment » plutôt que le « quoi » de la gouvernance. Ce concept met en évidence la performativité de l'algorithme. Il explique comment l'algorithme agit de manière procédurale dans la dimension temporelle de l'action : « *Le flux temporel de l'action [est] une chaîne ou un flux continu d'actions interpénétrées – antérieures et postérieures – qui comparent, échangent, trient, répartissent, administrent, etc.* »²³ (Introna 2016, 21). Dans ce flux temporel, les actions de l'algorithme se déroulent en fonction des actions produites par l'environnement, puis en suivant son propre code, ce qui génère de nouvelles actions qui affectent le monde. La gouvernementalité vise à relier ces actions, à comprendre les relations qu'elles forment dans l'assemblage d'objets techniques, d'humains, de règles et d'institutions.

Ainsi, la gouvernementalité de l'algorithme est liée à des régimes de connaissance et crée de nouvelles subjectivités, comme l'adoption d'autres pratiques d'écriture par les étudiants (Introna 2016). En poursuivant cette analyse, Cellard (2022) constate que les algorithmes entraînent des médiations vers différents modes de gouvernance liés à des contextes hétérogènes. De ces possibles, les visions conflictuelles offrent aux acteurs humains des moyens de dévier des utilisations initialement prévues.

En résumé, les acteurs en position de pouvoir cherchent à mobiliser les outils d'IA dans des usages qui leur permettent de capturer de la valeur. L'automatisation et l'opacité de ces outils créent alors une nouvelle forme de gestion. Cependant, la vision d'une machine impitoyable, appliquant systématiquement des règles, sans possibilité d'échapper à son autorité, est remise en question dans les pratiques qui en découlent. Les effets des usages des outils d'IA dépassent les intentions des concepteurs. Les motivations données aux outils d'IA ne forment qu'une partie des pratiques produites, car les acteurs exercent leur intentionnalité en adaptant leurs comportements ou en résistant aux décisions des algorithmes. Les médiations offertes par l'algorithme sont alors orientées par la volonté des acteurs pour négocier de nouvelles réalités.

²³ Temporal flow of action [is] a continuous string or stream of interpenetrating —prior and subsequent— actions that compare, swap, sort, allocate, administer, and so forth.

En effet, l'automatisation de l'algorithme permet aux acteurs de prévoir et de contourner ses effets.

Le concept de gouvernementalité déconstruit la vision d'un outil d'IA avec un pouvoir direct et incontestable, pour mettre en avant les possibilités d'action. Il explique comment la technologie est mobilisée dans un contexte, en mettant en lumière le pouvoir d'action de chaque acteur (humain et non humain). S'il reste pertinent dans la perspective de la gestion algorithmique, c'est-à-dire d'outils d'IA conçus pour une automatisation du contrôle et de la coordination, il s'adapte moins bien aux nouveaux usages des outils d'IA, davantage tournés vers une forme de collaboration humain-machine.

1.3. Le passage vers une perspective sur la collaboration humain-machine

Le postulat derrière la perspective sur le contrôle algorithmique n'est pas spécifique aux outils d'IA ; chaque nouvelle technologie soulève la question de ce qu'elle remplace. Dans les années 2010, les concepteurs ont mis en avant la capacité des outils d'IA d'apprentissage machine à automatiser des tâches spécialisées, générant ainsi un courant de recherche sur cette thématique et dont la littérature présentée précédemment a révélé les particularités de ces usages.

Dans les années 2020, le débat évolue : les concepteurs se concentrent davantage sur l'augmentation apportée par les outils d'IA, ouvrant la voie à la perspective de collaboration homme-machine (1.3.1). Bien que cette question concerne également d'autres technologies, il est crucial de comprendre les spécificités propres à la collaboration avec les outils d'IA (1.3.2).

1.3.1. Le débat à l'origine de la perspective de la collaboration humain-outil d'IA

La perspective sur la gestion algorithmique repose sur les possibilités d'automatisation de tâches spécialisées offertes par les outils d'IA. Plus globalement, dans les années 2010, les propos tenus par les concepteurs de ces outils tendaient à présenter cette technologie comme un moyen de rendre les processus plus efficaces en remplaçant l'humain par la machine. Par exemple, en 2016 le gourou de l'IA Geoff Hinton déclarait que former des radiologues était devenu inutile : « *people should stop training radiologists now. It is just completely obvious within five years deep learning is going to do better than radiologists* » (Creative Destruction Lab 2016).

Les perspectives d'usage suscitées par l'automatisation des tâches dépassent la gestion algorithmique. Auerswald (2017), Beyer (2016), et Brynjolfsson et McAfee (2014) parlent de la transition vers « l'automatisation » comme un phénomène inévitable qui remplace progressivement les emplois routiniers et l'intelligence humaine par les machines. Ce scénario suscite une attention considérable, car il signifierait un changement dans la nature du travail touchant des millions d'emplois, voire les structures sociales elles-mêmes.

De plus, Brynjolfsson (2022) argumente que la réplication et l'automatisation des tâches humaines menacent de marginaliser la main-d'œuvre, conduisant à une perte de leur pouvoir de négociation et à une concentration accrue de richesse et de pouvoir entre les mains de ceux qui contrôlent les technologies.

De même, des chercheurs mettent en évidence une tendance alarmante : la part du travail humain dans l'économie globale est en déclin (Karabarbounis et Neiman 2014; Frank, Dalenogare, et Ayala 2019). Plus préoccupant encore, cet ajustement ne se produit pas de manière uniforme à travers toutes les catégories de travailleurs actifs. Les écarts entre différents groupes de travailleurs se creusent, certains se retrouvant nettement plus désavantagés que d'autres.

Cette menace s'étend au travail des experts métier. Si les professionnels revendiquent leur expertise par rapport à d'autres groupes d'acteurs pour résoudre certains problèmes (Freidson 1970) et les inscrivent dans des institutions (comme des corps professionnels) pour s'auto-gouverner (Muzio, Brock, et Suddaby 2013), de nombreux efforts sont également faits pour inscrire cette expertise dans les technologies et des processus organisationnels (Abbott 1991; Barley et Tolbert 1991). Alors que les technologies tendaient à aider les experts métier (Autor, Katz, et Kearney 2006; Barrett et al. 2012), les outils d'IA seraient en mesure de remplacer, au moins partiellement, leurs activités analytiques selon certains (Brynjolfsson et McAfee 2014; Huang, Rust, et Maksimovic 2019).

L'effet de l'automatisation va au-delà du marché du travail, touchant également les systèmes de services traditionnels. Spohrer, Giuiusa, Demirkan et Ing (2013) discutent de la réduction rapide des coûts des « travailleurs numériques », c'est-à-dire les outils d'IA et autres algorithmes effectuant un travail comparable à celui des humains, et de son impact sur les structures familiales, éducatives, entrepreneuriales et gouvernementales. Cette dynamique met en question la capacité des systèmes existants à s'adapter à un rythme de progrès sans précédent.

Face à ces discours alarmants, Norman (2014) offre un aperçu plus nuancé de l'automatisation, soulignant les limites inhérentes des outils d'IA. Il observe que, si l'automatisation fonctionne bien dans des conditions normales et prévisibles, elle devient moins fiable et peut même échouer lorsque confrontée à des situations difficiles, complexes ou anormales. Ainsi, dans ces scénarios, les usages des outils d'IA pour la gestion algorithmique sont limités.

Cependant, de plus en plus de chercheurs remettent en question la focalisation sur l'automatisation (Markoff 2016; Bollier 2017; Rouse et Spohrer 2018; Raisch et Krakowski 2021). Ils proposent de regarder les outils d'IA comme des moyens certes d'automatiser

certaines tâches, mais en complément des travailleurs humains, procurant ainsi une augmentation, que ce soit de la production directement, ou des capacités des humains.

Cette perspective est renforcée par les travaux d'Acemoglu et Restrepo (2018), qui mettent en évidence les bénéfices potentiels de l'augmentation des capacités humaines par les technologies. Ils suggèrent que lorsque les technologies servent à augmenter plutôt qu'à remplacer les compétences humaines, les gains économiques tendent à bénéficier davantage aux travailleurs humains, influençant positivement la croissance économique, la répartition des revenus et l'emploi.

Par exemple, Brynjolfsson (2022) propose le concept d'augmentation, qu'il comprend comme le but d'augmenter les capacités humaines. Il envisage ainsi de créer une synergie dans laquelle les humains et les machines travaillent de concert, ouvrant la voie à des innovations et des capacités inédites. Cette complémentarité entre l'humain et la machine signifie que les travailleurs ne sont pas seulement nécessaires, mais qu'ils restent au cœur de la création de valeur, conservant ainsi leur importance et leur influence sur le marché du travail et dans les décisions politiques.

Bollier (2017) illustre cette approche dans le domaine de la médecine, en mettant en avant le rôle des outils d'IA dans l'augmentation des capacités des médecins plutôt que dans leur remplacement. Il rappelle les ambitions initiales du projet MYCIN (annexe 1) et les objectifs du système Watson d'IBM, qui visent à renforcer les compétences diagnostiques et thérapeutiques des professionnels de la santé. Cependant, ils reconnaissent que certaines tâches, en particulier celles impliquant la reconnaissance de motifs, peuvent être considérablement améliorées par l'apprentissage profond, illustrant un mariage réussi entre l'automatisation et l'augmentation.

Rouse et Spohrer (2018) mettent en évidence que l'automatisation et l'augmentation ne sont pas des alternatives distinctes, mais plutôt deux extrémités d'un spectre. Observant que de nombreux emplois combinent des tâches routinières et non routinières, ils soutiennent que les usages des outils d'IA vont automatiser des aspects routiniers tout en améliorant la capacité des humains à exécuter les aspects non routiniers de leur travail, créant ainsi une synergie entre l'humain et la machine.

De même, Raisch et Krakowski (2021) plaident pour une approche qui reconnaît l'entrelacement et la complémentarité de l'automatisation et l'augmentation des usages des outils d'IA. Selon cette perspective, l'automatisation et l'augmentation ne sont pas mutuellement exclusives, mais coexistent en pratique pour maximiser les avantages tout en minimisant les inconvénients potentiels.

Nous voyons ainsi émerger la perspective de la collaboration humain-machine, dans laquelle l'enjeu est d'établir un rapport entre ces deux entités, c'est-à-dire de gérer cette

relation entre l'humain d'une part et l'outil d'IA de l'autre (Brachten et al. 2020; Seeber et al. 2020; Mirbabaie et al. 2021; Anthony, Bechky, et Fayard 2023). D'où la motivation de trouver des modes de gestion dans la collaboration humain-machine pour les outils d'IA.

1.3.2. Le besoin de recherches spécifiques à la collaboration entre humains et outils d'IA

Certains auteurs considèrent qu'il y aurait « *plus de similitudes entre les interactions humain-humain et humain-machine que de différences* »²⁴ (Krämer, Von Der Pütten, et Eimler 2012, 233). Ainsi, les travaux sur la collaboration entre humains pourraient servir d'inspiration à celle entre les humains et les outils d'IA.

La littérature sur la collaboration entre humains aborde de nombreux sujets, comme les rôles et tâches nécessaires à un groupe pour parvenir à une collaboration efficace (Bunderson et Sutcliffe 2002; Aritzeta, Senior, et Swailes 2005; Van De Water, Ahaus, et Rozier 2008; Belbin 2010). De même, un courant de recherche s'est intéressé aux mécanismes structurels (Okhuysen et Bechky 2009) et aux processus dialogiques (Tsoukas 2009; Fayard et Metiu 2014) permettant l'incorporation de multiples perspectives dans les solutions (Hardy, Lawrence, et Grant 2005) et aidant à développer des compréhensions communes du travail (Bechky 2003).

Plus précisément, concernant la collaboration entre experts de différents domaines, les études montrent comment ces acteurs apprennent à travailler ensemble pour exploiter et combiner leurs capacités complémentaires (Maguire, Hardy, et Lawrence 2004; Hardy, Lawrence, et Grant 2005). Spécifiquement, dans le cas des métiers liés à la connaissance (« *knowledge workers* »), la littérature souligne le besoin de mettre en relation les savoirs entre différents acteurs pour une collaboration efficace (Carlile 2004; Levina 2005; Levina et Vaast 2005). Par conséquent, chacun doit être capable de comprendre l'apport de l'autre, de le contester ou de modifier sa propre réflexion (Carlile 2004; Levina 2005). D'ailleurs, dans ces travaux la technologie joue un rôle de « médium » servant à partager les connaissances. La technologie facilite ainsi la communication entre des acteurs ayant des expertises, des intérêts et des pouvoirs divers (Bechky 2003; Carlile 2002).

Cependant, un nombre croissant d'auteurs considèrent que les outils d'IA forment une collaboration spécifique avec les humains (Daugherty et Wilson 2018; Boyacı, Canyakmaz, et De Véricourt 2023; Seeber et al. 2020; Khadpe et al. 2020; Gao et al. 2021; Puranam 2021; Lebovitz, Lifshitz-Assaf, et Levina 2022; Siemon 2022; Anthony, Bechky, et Fayard 2023).

Par conséquent, les travaux sur la collaboration entre humains ne suffisent pas, et il faut poser à nouveaux de nombreuses questions, comme le sujet de la confiance envers l'outil d'IA (Elson,

²⁴ *more similarities between human-human and human-machine interactions than differences*

Derrick, et Ligon 2020; Jessup, Afb, et Gibson 2020), des formes de réciprocité (Goodman et al. 2016), ou de l'anthropomorphisme de l'outil d'IA (Qiu et Benbasat 2009; Araujo 2018; Watson 2019).

Par exemple, Lebovitz, Lifshitz-Assaf, et Levina (2022) argumentent le besoin de penser la gestion de l'opacité des résultats de l'outil d'IA, car malgré des études sur « *la manière dont les travailleurs du savoir traitent les connaissances tacites au cours des trois dernières décennies de recherche organisationnelle (voir Kogut et Zander 1992), nous en savons relativement peu sur la façon de traiter l'opacité des technologies modernes.* »²⁵ (Lebovitz, Lifshitz-Assaf, and Levina 2022, 3).

Nous avons présenté l'évolution des connaissances sur les outils d'IA, en faisant un zoom sur les sciences sociales et les sciences de gestion, qui commencent à s'emparer du sujet. Nous identifions un premier courant de recherche sur la gestion algorithmique, une perspective clé des années 2010, qui illustre comment les outils d'IA ont été utilisés pour automatiser et optimiser la prise de décision et la coordination au sein des organisations. Cette approche met en avant les défis, dont les enjeux de pouvoir et les reconfigurations de l'action collective que suscitent ces usages.

En entrant dans les années 2020, la perspective évolue vers la compréhension des spécificités d'une collaboration entre humains et outils d'IA. Il est essentiel de comprendre et d'établir les modes de gestion spécifiques aux collaborations entre humains et outils d'IA.

À l'aune des travaux précédents, ces modes de gestion de la collaboration se caractérisent comme les moyens pour orchestrer la collaboration entre les êtres humains et les outils d'IA, en incluant la définition des rôles et responsabilités (Siemon 2022), la coordination des tâches (Ren, Chen, et Qiu 2023), la communication et l'interaction entre acteurs (Stowers et al. 2021; Ren, Chen, et Qiu 2023) ainsi que le contrôle et l'amélioration des processus collaboratifs (Siemon 2022).

Nous présentons maintenant deux courants de littérature qui s'approchent de cet objectif. Le premier, trop normatif, conceptualise des modèles de collaboration humain-IA, mais il souffre d'un manque d'ancrage empirique. Le second, trop descriptif, se focalise sur

²⁵ *how knowledge workers deal with tacit knowledge over the last three decades of organizational scholarship (e.g., Kogut and Zander 1992), we know relatively little about dealing with the opacity of modern technologies.*

les usages des outils d'IA, sans intégrer ses résultats dans des modèles permettant de gérer la collaboration.

2. Deux littératures convergeant vers les modes de gestion de la collaboration humains-outils d'IA

Dans les années 2010, les usages des outils d'IA étaient principalement tournés vers la gestion algorithmique, grâce notamment à une emphase sur les capacités d'automatisation des outils d'IA et de leur potentiel à remplacer les humains sur bon nombre de tâches.

Dans les années 2020, on reconnaît l'intérêt à développer des usages d'outils d'IA qui viennent en support aux humains, plutôt que de chercher à les remplacer. Kline (2020) appelle la synergie entre humains et technologie le « facteur de techno-extension » (*techno-extension factor*), soulignant que les utilisateurs de la technologie doivent être conscients du rôle des machines et de la mesure dans laquelle les humains devraient compter sur elles.

Alors que dans le contrôle algorithmique, l'enjeu était de comprendre comment les outils d'IA captaient du pouvoir et de saisir le lien « vertical » qui lie les outils d'IA aux humains, avec la collaboration humain-machine, le but se tourne vers la relation « horizontale », ou les modes de gestion de cette collaboration, c'est-à-dire la manière dont les outils d'IA et les humains interagissent et se complètent mutuellement, ou non.

Cette collaboration a lieu dès la conception. Spohrer et Banavar (2015) soulignent que l'augmentation de l'expertise humaine grâce à des technologies cognitives peut entraîner une amélioration significative de la productivité et de l'innovation. Ils soutiennent que par l'intégration des connaissances à travers différentes disciplines, ces systèmes peuvent co-crée de la valeur sociale et économique, envisageant un futur où l'expertise humaine est amplifiée par l'assistance intelligente. Cette vision nécessite une collaboration multidisciplinaire pour peaufiner la boucle de conception des systèmes sociotechniques, d'où l'importance de l'interaction humaine dans l'amélioration de l'expertise via les outils d'IA.

La collaboration continue ensuite dans les usages. Bonetti et al. (2023), dans une étude ethnographique sur le marketing, observent que l'interaction avec les outils d'IA est un processus dynamique de co-évolution, nécessitant une conception, une adaptation et un (ré)alignement collectifs. Cette approche collaborative est essentielle pour intégrer les usages des outils d'IA dans les pratiques de travail. Elle souligne que la technologie et les utilisateurs humains doivent évoluer ensemble pour maximiser le potentiel des outils d'IA.

Comme le remarque Markoff (2016) ces synergies diffèrent selon les contextes. Les outils d'IA peuvent étendre les capacités humaines dans certains domaines tout en les automatisant et les remplaçant dans d'autres. Pour explorer le sujet, de nombreux modèles de collaboration humain-machine cherchent à définir cette synergie (2.1). Cependant, ces modèles sont conceptuels et manquent d'une empreinte empirique, ce qui nous amène à étudier la littérature sur les usages des outils d'IA (2.2). Cette littérature, riche d'enseignements et qui

fait émerger du terrain les prémisses de modes de gestion de la collaboration, n'aboutit pas encore sur des modèles de collaboration humain-machine.

2.1. Les modèles conceptuels de collaboration humain-machine

Les artefacts cognitifs sont des dispositifs créés par l'homme qui augmentent les capacités cognitives humaines en modifiant la manière dont les tâches sont accomplies (D. A. Norman 1991). Ces artefacts ne changent pas directement les capacités d'un individu, mais transforment la nature des tâches exécutées. Du point de vue du système, l'artefact est vu comme une partie intégrante d'un système comprenant la personne et la tâche, augmentant ainsi la capacité cognitive globale du système. En revanche, pour l'utilisateur individuel, l'artefact change la tâche elle-même plutôt que d'améliorer directement ses capacités cognitives.

Par conséquent, Norman (1991) souligne l'importance de comprendre les artefacts cognitifs pour améliorer la performance humaine et concevoir de meilleurs outils. Il appelle donc à des études approfondies des interactions entre les humains. Ces efforts sont aujourd'hui appliqués à la mise en place de collaborations entre humains et outils d'IA, mais la littérature sur le sujet est encore naissante. En particulier, les modes de gestion de cette collaboration restent une question ouverte.

Par exemple, Fügner et al. (2022) et Tong et al. (2021) indiquent que pour certaines tâches, en particulier celles de classification, la combinaison des outils d'IA et de l'intervention humaine conduit aux meilleurs résultats. Or, cette efficacité est surtout évidente lorsque les outils d'IA délèguent des tâches aux humains, ou lorsque les outils d'IA sont employés différemment selon le niveau d'expérience des employés qui les utilisent. La personnalisation et la compréhension des rôles dans la collaboration IA-humain semblent donc cruciales.

De même, Hang et Chen (2022) soulignent que les outils d'IA peuvent augmenter les revenus en améliorant la productivité des employés, en augmentant la satisfaction des consommateurs, en fixant des prix compétitifs et en créant des ressources uniques. Toutefois, ils notent également que l'adoption des usages des outils d'IA, la nature spécifique des tâches et la gestion des outils d'IA constituent des obstacles majeurs à la réalisation de leur plein potentiel. Ils appellent à une recherche future axée sur l'amélioration des compétences interpersonnelles des outils d'IA, ce qui implique une collaboration étroite entre les humains et ces outils, fondée sur une compréhension des modes de gestion appropriés.

Nous identifions des modèles de collaboration entre humains et outils d'IA issus de différents horizons disciplinaires. Ces modèles sont difficiles à classifier, car les concepts peuvent être voisins, se chevaucher ou constituer des cas particuliers d'autres modèles plus généraux. Nous avons cependant essayé de les présenter en commençant par ceux dont

l'objectif de la collaboration est directement lié à l'amélioration de la tâche à effectuer, pour présenter ensuite ceux dont l'objectif inclut également l'amélioration des capacités des humains, et enfin ceux qui sont focalisés uniquement sur l'augmentation de l'intellect humain.

2.1.1. Les Plateformes Numériques Hybrides Humain-IA

Un premier courant de la recherche sur la collaboration humain-outil IA s'intéresse principalement au résultat final, à la production de biens ou de services. Au sein de ce courant, des études se centrent sur l'émergence de plateformes numériques hybrides humain-IA. Rai, Constantinides, et Sarker (2019) décrivent les plateformes de travail numérique comme des environnements en ligne dans lesquels les services numériques sont fournis en échange d'une rémunération, avec des tâches spécifiques exécutées et coordonnées par des outils d'IA, issus de la période de gouvernance algorithmique. À partir de ce constat, ils théorisent l'évolution de ces outils vers des « hybrides humain-IA » et leur potentiel pour générer de la valeur ajoutée.

Ces plateformes hybrides Humain-IA offrent des opportunités uniques pour la substitution, l'augmentation et l'assemblage dans le cadre du travail. En combinant les capacités humaines et artificielles de manière dynamique, elles cherchent à créer une intelligence collective sensible au contexte, dans laquelle les forces de chaque agent (humain ou outil d'IA) sont employées de manière optimale pour accomplir des tâches complexes.

L'adoption massive des outils d'IA dans la gestion des ressources humaines (GRH) a inspiré le type de collaboration des plateformes numériques hybrides humain-IA. Prikshat et al. (2023) définissent la GRH augmentée par les outils d'IA comme l'intégration technologique dans les systèmes pour améliorer le traitement et l'analyse des données, ce qui assiste dans la résolution de problèmes et la prise de décision pour des résultats opérationnels, relationnels et transformationnels.

Ce secteur a fait évoluer les outils d'IA pour impliquer activement les employés, notamment dans le partage de connaissances (Malik, Nguyen, et Budhwar 2024; Nguyen et Malik 2022; Malik et al. 2022). Des études complémentaires montrent comment l'adoption des outils d'IA dans les pratiques RH mène à une amélioration significative de la tâche à accomplir, dans ce contexte celle-ci étant la satisfaction au travail et de l'expérience des employés (Nguyen et Malik 2022; Malik et al. 2023).

Inspirés par ces résultats, Malik, Nguyen, et Budhwar (2024) mobilisent la littérature théorique sur le partage des connaissances et la GRH pour élaborer un modèle théorique de collaboration centré sur l'amélioration de l'efficacité de la GRH, ainsi que les effets de personnalisation qu'ils peuvent entraîner. En collaborant avec les outils d'IA pour analyser et comprendre les besoins et les comportements des employés, les entreprises peuvent créer des environnements de travail plus engageants et plus productifs. Adoptant une perspective

des sciences de service, Spohrer (2016) argumente même qu'avec les outils d'IA, nous pourrions concevoir le travail optimal pour chaque travailleur.

En résumé, l'incorporation des outils d'IA dans des plateformes numériques hybrides représente une continuation de la perspective de la gestion algorithmique. Le succès de la collaboration se mesure seulement en fonction de l'augmentation la production de biens et de service. Le rôle principal de l'outil d'IA est d'effectuer des tâches routinières, tandis que le travailleur humain comble les déficits rencontrés lors d'évènements anormaux. Ce type de collaboration reste cependant très conceptuel et les études citées ne rendent pas compte de son application concrète, ni des moyens de communication entre humains et outils d'IA ou des dispositifs de contrôle de la collaboration.

2.1.2. La Collaboration Humain-IA sous l'angle de l'Intelligence

La majorité de la littérature choisit d'aborder la collaboration entre l'humain et la machine sous l'angle de l'intelligence afin d'établir les complémentarités à développer. Comme l'ont montré Legg et Hutter (2007) en compilant 70 définitions différentes de l'intelligence, les perspectives possibles sont nombreuses.

Si les formes d'intelligence sont principalement attribués aux humains, il existe un consensus général sur le fait que les machines, et les outils d'IA en particulier, manifestent des formes d'intelligence spécifiques, y compris une intelligence émotionnelle émergente (Huang et Rust 2018).

Plus généralement, la recherche attribue aux outils d'IA une forme d'intelligence analytique, tandis que l'humain, par opposition, possède une intelligence intuitive. Hung (2003) et Martin (2009) soulignent l'importance de cette dualité, arguant que les performances commerciales optimales nécessitent un mélange de maîtrise analytique et d'originalité intuitive.

Par exemple Jaiswal, Arun, et Varma (2022) fournissent une analyse approfondie de la manière dont les compétences associées aux quatre types d'intelligence identifiés par Huang et Rust (2018) – mécanique, analytique, intuitive et empathique – se rapportent aux outils d'IA. Leurs découvertes suggèrent que, tandis que les outils d'IA peuvent exceller dans des tâches nécessitant de l'intelligence mécanique, les tâches nécessitant des compétences analytiques, intuitives et empathiques restent encore un défi. Cela indique que l'avenir du travail, en particulier dans le secteur des technologies de l'information, exigera que les employés se concentrent davantage sur les compétences analytiques, intuitives et empathiques pour rester pertinents et apporter de la valeur à leurs organisations.

Aussi, la recherche met en avant les limites notables de l'intelligence des outils d'IA, notamment en termes de diversité et de biais. Comme le mettent en avant Hassani et al. (2020), les outils d'IA risquent de limiter la diversité en privilégiant constamment les mêmes

sources d'information, et chaque outil d'IA peut intégrer des biais implicites (voir les politiques cachées, sous-partie 1.2.2).

Des chercheurs tels que D. Norman (2014) et Dane, Rockmann, et Pratt (2012) soulignent au contraire l'importance de la pensée réflexive, intuitive et abstraite dans la prise de décision humaine, et donc les capacités uniques de l'intelligence humaine dans des situations non conventionnelles et créatives. Cela indique que, bien que les outils d'IA puissent exceller dans des tâches bien définies, l'intuition et la créativité restent des domaines où l'intelligence humaine conserve un avantage distinct.

Ainsi, pour une collaboration optimale entre l'intelligence humaine et les outils d'IA, ces travaux tendent à recommander d'attribuer aux outils d'IA des tâches appelant à l'analyse et aux humains des rôles mobilisant leur intuition. Alors que les outils d'IA peuvent transformer de nombreux aspects de notre travail et de notre vie, la pensée réflexive, intuitive et empathique humaine reste indispensable à la prise de décision et la créativité. La coordination des tâches doit alors se faire dans l'idée de complémentarité, en fonction des forces et limites respectives. Mais au-delà de ces propos généraux, nous manquons de précisions. De plus, nous remarquons que dans ces travaux, la communication et l'interaction entre acteurs ainsi que le contrôle et l'amélioration des processus restent des dimensions des modes de gestion de la collaboration qui ne sont pas encore abordés.

2.1.3. Systèmes d'Intelligence Hybrides Humain-IA ou Intelligence Collaborative

Complémentaires au concept de plateformes hybrides, les systèmes d'intelligence hybrides humain-IA mettent en œuvre une collaboration entre l'intelligence humaine et celle de la machine pour un meilleur développement et usage des outils d'IA. Selon Dellerman et al. (2019) et H. Liu, Lai, et Tan (2021), les systèmes d'intelligence hybride fusionnent les capacités complémentaires des humains et des outils d'IA, encourageant une collaboration qui mène à des résultats supérieurs par rapport à des entités fonctionnant de manière isolée.

De manière similaire aux systèmes hybrides humain-IA, l'intelligence collaborative explore comment les humains et les outils d'IA peuvent œuvrer de concert pour atteindre des objectifs communs (Johnson et Vera 2019; Bogert, Schecter, et Watson 2021). Par exemple, Pakdemirli (2019) et Topol (2019a) proposent une vision des professionnels de santé augmentés par l'automatisation des outils d'IA sur des tâches spécifiques afin que les travailleurs expérimentés se concentrent sur des aspects plus complexes et tacites de leur travail, rendant ainsi leur approche globale plus productive.

Dans cette dynamique, humains et outils d'IA sont généralement vus comme des partenaires égaux dans la résolution de tâches, comme le démontre Siemon (2022). Cette approche de la complémentarité humain-IA se focalise sur la synergie entre les compétences humaines et machiniques : tandis que les humains excellent grâce à leurs sens, perceptions,

intelligence émotionnelle et compétences sociales, les outils d'IA se distinguent dans la détection de motifs et le calcul de probabilités (Braga et Logan 2017; Dellermann et al. 2019).

Ces aptitudes complémentaires favorisent une performance accrue dans des tâches spécifiques par le biais de la collaboration. Par exemple, les managers peuvent appliquer leur intelligence émotionnelle pour établir des relations et motiver les employés, tandis que des tâches répétitives et monotones peuvent être déléguées aux outils d'IA (Davenport et Kirby 2016). Les discussions sur les types de tâches qui devraient être laissées aux outils d'IA sont encore d'actualité (Fügener et al. 2022).

Pour mettre en place des systèmes d'intelligence hybrides, les recherches mentionnent avant tout le développement de l'étape de conception et les manières d'améliorer les outils d'IA, notamment en prenant en compte l'intelligence humaine (Dellermann et al. 2019; Guo et al. 2020; H. Liu, Lai, et Tan 2021; Siemon 2022). Par exemple, Carroll et al. (2019) ont développé un outil d'IA adaptatif capable de s'ajuster aux comportements humains, tandis que Sadigh et al. (2018) ont permis aux outils d'IA embarqués dans les voitures autonomes d'intégrer les réactions potentielles d'autres conducteurs humains pour une coordination plus efficace. Ces innovations montrent comment les outils d'IA peuvent être ajustés pour mieux interagir et collaborer avec les humains.

Cette approche est renforcée par le cadre de « cross-training humain-robot » proposé par Nikolaidis et Shah (2013), permettant aux outils d'IA et aux travailleurs humains d'apprendre de manière itérative les uns des autres pour améliorer la collaboration. Ainsi, des méthodologies sont élaborées pour la formation continue ou l'actualisation des outils d'IA existants afin d'améliorer le travail d'équipe avec les humains (G. Bansal et al. 2019). Ces études soulignent l'importance d'une mise à jour constante des outils d'IA pour mieux comprendre et aligner leurs objectifs avec ceux des agents humains. Par exemple, Fügener et al. (2021) et Sturm et al. (2021) abordent respectivement le rôle de la connaissance humaine en complément de celle des outils d'IA, et comment ces outils d'IA affectent les systèmes d'apprentissage organisationnel.

En somme, les approches des systèmes hybrides humain-IA et de l'intelligence collaborative proposent des outils d'IA qui viennent en support en remplaçant des tâches mécaniques et analytiques, afin d'augmenter les capacités émotionnelles et créatives des agents humains. Si des modèles de communication et d'amélioration des processus existent (Nikolaidis et Shah 2013), nous pouvons émettre un doute sur la portée réelle d'un apprentissage mutuel lorsque la collaboration est fondée sur une telle division des tâches.

2.1.4. Coordonner l'Apprentissage Humain et l'Apprentissage Machine

Face à la critique précédente, nous pouvons évoquer un courant qui s'intéresse à la coordination entre l'apprentissage humain et l'apprentissage des outils d'IA au sein des

organisations. Cette démarche vise à maximiser l'efficacité globale du processus d'apprentissage organisationnel, permettant ainsi à l'ensemble du système d'accumuler les connaissances (Sturm et al. 2021).

A ce niveau organisationnel, l'apprentissage humain a été décrit comme un processus essentiellement social, impliquant l'interaction, le partage et la construction collective de la connaissance (March 1991). À l'opposé, l'apprentissage des outils d'IA repose sur le traitement des données, sans les composants sociaux et émotionnels qui caractérisent l'apprentissage humain.

Cette distinction souligne le besoin d'une nouvelle théorie ou d'une approche révisée pour coordonner efficacement un système intégrant à la fois des apprenants humains et des outils d'IA (Seidel et al. 2018). En d'autres termes, si l'idée de coordonner les apprentissages entre humains et outils d'IA a été évoquée, les modes de gestion associés à un tel type de collaboration restent encore à développer et concrétiser sur le terrain.

2.1.5. Symbiose Humain-IA dans la Prise de Décisions

La prise de décision fait face aux défis de l'incertitude, la complexité et l'équivoque (Choo 1991; Simon 1997), la rendant difficile même pour les individus les plus compétents. En particulier, la prise de décision clinique, comme pour le traitement de la COVID-19 (Vickery et al. 2022), la stimulation cérébrale profonde (Muñoz et al. 2021), ou les thérapies géniques (Hampson et al. 2018), illustre l'impact de ces facteurs. C'est pourquoi, un courant de la recherche sur la collaboration entre l'humain et l'outil d'IA s'intéresse à la symbiose entre l'intelligence humaine et les outils d'IA dans la prise de décision organisationnelle.

Par exemple, face à l'équivocité, c'est-à-dire la présence de multiples interprétations possibles d'une situation (Weick et Roberts 1993), Luo et al. (2021) et Fügener et al. (2021) reconnaissent que les usages des outils d'IA présentent des avantages significatifs en matière de compétences computationnelles. Ils peuvent ainsi entrer en symbiose avec les capacités humaines fondées sur leur expérience et leurs connaissances générales.

De plus, Jarrahi (2018) souligne que les interactions entre humains et les outils d'IA peuvent rendre les deux parties plus intelligentes au fil du temps. Les outils d'IA peuvent étendre la cognition humaine en abordant des problématiques complexes, tandis que les humains offrent une perspective plus holistique et intuitive, essentielle dans le traitement de l'incertitude et de l'équivoque. Cette collaboration reflète l'idée d'augmentation de l'intelligence humaine, développée ci-dessous.

Kostick-Quenet et Gerke (2022) mettent en évidence un avantage convaincant des outils d'IA dans le domaine de la santé : la réduction de l'incertitude, en fournissant par exemple des estimations personnalisées sur l'évolution de la condition d'un patient. Toutefois, la contribution effective des outils d'IA à la réduction de l'incertitude dépend grandement de

l'interprétation et de la mise en action des estimations par les humains. Ainsi, les recherches devraient également s'intéresser à la manière dont les utilisateurs humains font usage de ces outils d'IA pour la prise de décision.

Paschen, Wilson, et Ferreira (2020) opèrent avec une prémisse similaire, en considérant que les outils d'IA sont capables de traiter de vastes quantités de données en informations pour une meilleure création et gestion des connaissances. Dans leur analyse théorique du processus de vente d'entreprise à entreprise, ils évoquent les transformations que les outils d'IA pourraient apporter. S'ils venaient à assister les décideurs humains, comme suggéré dans les recommandations managériales, les auteurs soulignent que ces outils changeraient le processus de vente traditionnel. Il est donc également intéressant de remarquer que la collaboration humain-machine pour la prise de décision risque de changer en profondeur les processus existants et l'organisation elle-même.

En somme, la symbiose humain-IA dans la prise de décisions met l'emphase sur le besoin de contrôler et d'améliorer les processus collaboratifs, en définissant le rôle de l'outil d'IA au traitement de grands volumes d'information sur une situation spécifique, tandis que l'humain prend appui sur son expérience et des connaissances générales. Nous avons là une inversion curieuse par rapport aux types de collaboration précédents, dans lesquels les aspects généraux et routiniers étaient délégués aux machines, pour que les humains se concentrent sur les cas spécifiques. Mais, à l'instar des autres modèles, les modes de gestion spécifiques restent élusifs.

2.1.6. Humain dans la Boucle

Le concept de « humain dans la boucle » souligne pour sa part l'importance cruciale de l'intégration humaine dans les processus décisionnels assistés par les outils d'IA, en particulier dans des contextes à haut risque tels que la santé et la justice. Le concept, annexe à celui de la symbiose humain-IA dans la prise de décision organisationnelle se centre davantage sur le rôle et la responsabilité de l'humain. L'approche est motivée par les inquiétudes concernant les biais sociétaux négatifs qui peuvent être reproduits par des outils d'IA (Obermeyer et Emanuel 2016; Goyal et al. 2020), et plus généralement de l'opacité des résultats des outils d'IA (sous-partie 1.2.1) et de leurs politiques cachées (sous-partie 1.2.2).

Face à ces enjeux, des chercheurs tels que Loftus et al. (2020) et Rudin (2019) argumentent que les outils d'IA ne devraient pas remplacer entièrement les décisions considérées comme « à haut risque » en raison des impacts significatifs qu'elles peuvent avoir. Dans le domaine de la santé, il est recommandé que même les outils d'IA capables de diagnostiquer de manière autonome les maladies soient toujours soumis à la confirmation par des examens dirigés par des humains (Kelly et al. 2019; Yu, Beam, et Kohane 2018). D'autres chercheurs suggèrent d'aller plus loin et d'éviter les outils d'IA pour certaines décisions (Gillespie 2014; Burrell 2016).

D'autres chercheurs font la distinction entre garder les humains « dans » la boucle versus « sur » la boucle, impliquant respectivement que les humains jouent un rôle actif dans la prise de décision ou qu'ils puissent intervenir ou remettre en question les conclusions de l'outil d'IA lorsque nécessaire, même s'ils n'agissent pas activement dans le processus décisionnel initial.

Dans ce contexte, Kostick-Quenet et Gerke (2022) reconnaissent les risques d'une dépendance excessive aux outils d'IA dans les décisions à haut risque. Ils plaident pour un maintien des humains « dans la boucle » et recommandent des actions pour améliorer la sécurité des outils d'IA, notamment en augmentant la précision computationnelle.

Zanzotto (2019) introduit la notion de « *human-in-the-loop artificial intelligence* » mettant en avant la nécessité de valoriser les contributions humaines dans le processus de décision des outils d'IA. Ce modèle propose une synthèse entre l'apprentissage supervisé et non supervisé, où l'expertise humaine guide la formation des outils d'IA à travers des données étiquetées, tout en permettant à ces outils de révéler des structures dans de larges ensembles de données de manière autonome. De même, Zheng et al. (2017) avancent le concept de « *human-in-the-loop augmented intelligence* », qui promeut une collaboration effective entre les humains et les outils d'IA.

En somme, l'approche de l'humain dans la boucle favorise une collaboration étroite et réfléchie entre l'humain et les outils d'IA, en reconnaissant que les compétences humaines en matière d'intuition, de jugement moral et de compréhension contextuelle sont indispensables pour guider, superviser et améliorer les processus décisionnels assistés par les outils d'IA. La responsabilité est donc mise sur l'humain, avec un rôle de l'outil d'IA très auxiliaire. De plus, quelques exemples de moyens de contrôle et l'amélioration sont présentés. Il reste à voir comment ce principe se traduit dans la pratique.

2.1.7. *Épistémologie Hybride*

Pour aller plus loin dans l'analyse de la collaboration entre les différents types d'intelligence le concept d'épistémologie hybride représente une fusion entre les processus cognitifs humains et les procédures algorithmiques des outils d'IA, créant ainsi un nouveau paradigme dans la manière dont les connaissances sont générées, manipulées et interprétées (Babushkina et Votsis 2022). Cette approche reconnaît que la cognition humaine et les capacités des machines traitent et transforment l'information de manières fondamentalement différentes. Alors que la cognition humaine est profondément enracinée dans des processus sociaux, culturels et émotionnels, les outils d'IA, guidés par l'apprentissage automatique, fonctionnent à partir de modèles de données et d'algorithmes.

L'appel de Babushkina et Votsis (2022) à une épistémologie hybride cherche à remettre en question et à aller au-delà de la tendance actuelle qui consiste à répliquer purement et simplement la cognition humaine dans les outils d'IA. Ils critiquent la tendance à utiliser des

métaphores pour décrire les capacités des outils d'IA, argumentant que de telles analogies sont inadéquates et mènent à des conclusions erronées sur ce que les outils d'IA peuvent réellement accomplir. Au lieu de cela, ils préconisent une approche plus nuancée et réaliste qui reconnaît les outils d'IA comme produisant un type unique de produit épistémologique qui ne remplace pas, mais plutôt complémente, la cognition humaine.

Dans le contexte de l'épistémologie hybride, cette interaction est étendue pour inclure les contributions uniques des outils d'IA, soulignant qu'une compréhension complète des phénomènes (comme une pathologie ou le cas d'un patient) ne peut se faire sans considérer à la fois les perspectives humaines et algorithmiques.

Dans ce sens, Jussupow et al. (2021) mettent l'accent sur l'importance des compétences méta-cognitives dans les usages des outils d'IA pour le processus diagnostique dans le domaine médical, suggérant que la décision augmentée nécessite une compréhension claire de la manière dont les résultats des outils d'IA sont intégrés dans les processus cognitifs humains. Cette perspective souligne la nécessité d'une distinction nette entre les rôles et les tâches épistémiques attribués respectivement aux outils d'IA et à l'expert humain, une notion également soutenue par (Van Baalen, Boon, et Verhoef (2021) et Boon (2020).

Ces travaux montrent donc un niveau de précision plus important sur les modes de gestion de la collaboration qu'ils proposent. Toutefois, il reste à traduire la distinction des rôles et des tâches épistémiques en propositions concrètes. De même, il reste à comprendre comment ce type de modèle gère le contrôle et l'amélioration des processus collaboratifs.

2.1.8. Augmentation de l'Intelligence

L'augmentation de l'intelligence est un concept qui met l'accent sur l'amélioration des capacités humaines grâce aux usages des technologies. D. Norman (2014) critique la tendance actuelle à privilégier une orientation centrée sur la machine, réduisant les humains à un rôle de soutien, au lieu de concevoir des technologies augmentant les capacités humaines.

Cette perspective hérite des idées de Licklider (1960), qui envisageait une symbiose humain-ordinateur dans laquelle les machines et les humains travailleraient ensemble pour résoudre des problèmes et prendre des décisions complexes. Engelbart (1962) a approfondi cette idée en introduisant le terme d'augmentation de l'intelligence, définissant le concept comme l'augmentation de la capacité humaine à aborder des situations problématiques complexes pour obtenir des solutions adaptées à leurs besoins. Selon Engelbart, l'objectif est d'exploiter la technologie pour étendre les capacités cognitives et créatives humaines.

Zhou et al. (2021) définissent l'augmentation de l'intelligence comme l'amélioration de la capacité, de l'intelligence et de la performance humaines avec l'aide des technologies de l'information. Cette définition souligne que l'intention initiale des outils d'IA était de

surmonter les limitations cognitives humaines, comme le mentionnent également Jain et al. (2021).

La distinction entre l'augmentation de l'intelligence et l'IA est importante. Bien que les technologies sous-jacentes soient les mêmes, l'augmentation de l'intelligence place les humains au centre du système de prise de décision, contrairement à l'IA qui met l'accent sur les capacités de la machine. Cette distinction est cruciale pour comprendre les objectifs différents de l'augmentation de l'intelligence par rapport à l'IA : l'un cherche à rendre les gens plus intelligents, tandis que l'autre vise à rendre les machines plus intelligentes. Dans l'optique de l'augmentation de l'intelligence, Zhou et al. (2021) soulignent la nécessité d'accorder plus d'attention aux facteurs humains dans cette relation collaborative pour améliorer tant les systèmes de l'augmentation de l'intelligence que de ceux d'IA.

Barile et al. (2019) proposent de repenser le concept de l'augmentation de l'intelligence comme une interaction entre outils d'IA et humains qui dépasse l'amplification des capacités humaines en transformant notre cognition. Cette mutation change la structure même de la pensée humaine et ses processus cognitifs. Par conséquent, Barile et al. (2019) appellent à formuler de nouvelles règles pour créer des collaborations entre les humains et les outils d'IA, soulignant la nécessité d'une interaction plus intégrée.

Cependant, au-delà de principes très généraux pour définir la collaboration humain-machine, l'Augmentation de l'Intelligence ne propose pas encore des modes de gestion permettant de traduire cette approche en des moyens d'action pour les acteurs.

2.1.9. Un Tournant post cognitiviste dans l'Augmentation de l'Intelligence

Des travaux en philosophie post cognitiviste explorent davantage l'idée de l'augmentation de l'intelligence. Par exemple, van Breemen, Farkas, et Sarbo (2011) analysent la formation de concepts dans la résolution de problèmes en s'appuyant sur la théorie cognitive du traitement de l'information et sur la théorie sémiotique de Peirce, modélisant la conceptualisation comme un processus dynamique plutôt que statique. Ils concluent que les systèmes d'augmentation de l'intelligence humaine nécessitent une représentation des connaissances compatible avec l'humain, mais aussi flexible et transférable d'un domaine à un autre.

Ces travaux s'appuient sur ceux de Skagestad (1993), qui explore les implications philosophiques des technologies en tant qu'extensions de l'intellect humain. Inspiré par Engelbart (1962), il intègre la sémiotique peircienne pour discuter de la manière dont les ordinateurs et autres technologies de l'information transforment nos activités intellectuelles et notre perception. Il souligne une boucle de rétroaction entre la pensée humaine et l'innovation technologique, suggérant une co-évolution de notre intellect et de nos usages des technologies. Cette relation transformatrice remet en question les anciennes pratiques et

façon de nouvelles façons de penser, de résoudre des problèmes et de créer des connaissances.

En continuant sur cette lancée, Skagestad (1996) explore les effets profonds de l'innovation technologique sur notre culture et nos pratiques intellectuelles. Il se penche sur les implications culturelles et sémiotiques de l'ordinateur personnel, mettant en évidence comment il a transformé nos habitudes de travail intellectuel, de communication et d'interaction sociale. L'avènement des technologies informatiques personnelles a non seulement facilité de nouvelles formes de travail intellectuel, mais a également remis en question les notions traditionnelles de logique et de raisonnement. Skagestad appelle à une compréhension approfondie de ces changements d'un point de vue sémiotique pour appréhender les transformations culturelles plus larges induites par la technologie.

Cet appel est repris par plusieurs travaux qui explorent le lien entre les théories de Peirce et les outils d'IA. Par exemple, la théorie de la sémiotique triadique s'applique à l'analyse des outils d'IA symboliques (Sowa 1984; Tiercelin 1984; Fetzer 1990; Bolter 1991). Plus récemment, Skagestad (2004) et P. Steiner (2013) s'intéressent à ses travaux dans la vague des approches post-cognitivistes sur l'intelligence, notamment avec les concepts de cognition étendue (« *extended cognition* ») et de cognition intégrée (« *embedded cognition* ») (Haugeland 1993; Clark 2008; Clark et Chalmers 1998).

Cognition étendue (Clark et Chalmers 1998) : Les processus cognitifs peuvent s'étendre au-delà des frontières du cerveau pour inclure des objets externes et l'environnement. Ainsi les outils et technologies que nous utilisons peuvent devenir des parties intégrantes de nos systèmes cognitifs.

Cognition intégrée (Haugeland 1993) : les processus cognitifs ne peuvent pas être pleinement compris sans tenir compte de leur contexte environnemental. Ils sont influencés, voir façonnés, par les interactions entre l'organisme et son environnement comme les contextes physiques, sociaux et culturels.

En résumé, ces travaux en philosophie mettent en lumière la nécessité d'une approche dans l'augmentation de l'intelligence dans laquelle la technologie ne remplace pas simplement l'intellect humain, mais travaille en synergie avec celui-ci pour favoriser une évolution cognitive et culturelle. Cette synergie, enrichie par une compréhension sémiotique, permet d'améliorer l'efficacité humaine et de redéfinir nos méthodes de pensée et d'interaction avec le monde. À l'instar de l'Augmentation de l'Intelligence, nous n'avons pas pu identifier des modes de gestion qui concrétisent ce type de collaboration.

Ce type de collaboration conclut notre synthèse en se situant à l'extrême des modèles conceptuels visant à étendre l'intellect humain. Il clôture ainsi le continuum de formes de

collaboration que nous avons présenté, débutant avec le modèle des Plateformes Numériques Hybrides Humain-IA, dont l'objectif était de maximiser la tâche à accomplir (figure 4).

Collaboration humain-machine focalisée sur la tâche à accomplir

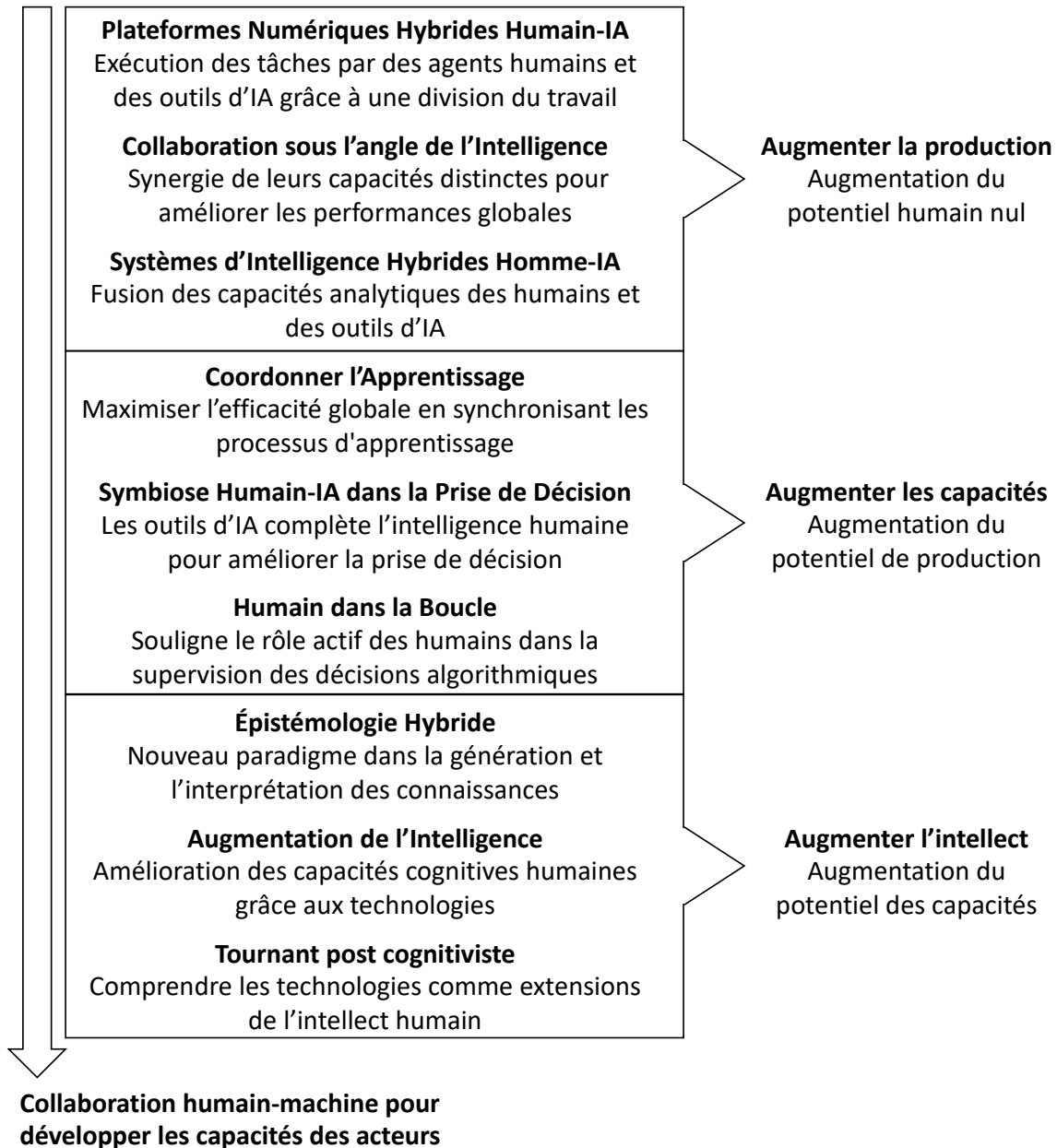


Figure 4 : Graduation des modèles de collaboration humain-IA

Dans cet ordre de présentation, ces modèles évoluent vers l'enrichissement des capacités des humains, que nous percevons comme une progression vers l'augmentation du potentiel humain.

Les premiers modèles agissent principalement sur la tâche à exécuter. Bien qu'ils puissent concrétiser les capacités d'action en optimisant la production, l'augmentation du potentiel est nulle car le travailleur humain reste contraint dans un rôle fixe, défini par le rôle de la machine.

En tant qu'étape intermédiaire, nous identifions les modèles de collaboration axés sur la complémentarité des capacités, qui accroissent le potentiel de production en offrant aux humains de nouvelles méthodes pour réaliser une action.

Enfin, nous trouvons des modèles conceptuels se concentrant sur l'extension de l'intellect humain, procurant ainsi une augmentation du potentiel des capacités d'action.

Chaque type de collaboration cherche à atteindre son objectif d'augmentation (production, capacités ou intellect) en augmentant le potentiel humain de l'ordre inférieur (aucun, production ou capacités). Ces modèles restent cependant très conceptuels et ils n'explicitent pas, ou très peu, les modes de gestion de la collaboration qui permettraient de mettre en place l'augmentation de ces potentiels.

2.1.10. Besoin d'observer les collaborations humain-machine en action

Les discussions conceptuelles présentées précédemment nous mènent à un point crucial : la nécessité d'observer comment ces outils d'IA fonctionnent réellement dans des contextes pratiques.

Maedche et al. (2019) ainsi que Sturm et al. (2021) préconisent d'adopter une perspective davantage socio-technique et structurationnelle (DeSanctis et Poole 1994; Orlikowski et Robey 1991) pour étudier les usages des outils d'IA, et donc analyser leur impact sur les acteurs humains, l'influence de ces derniers sur les outils d'IA, et les contextes dans lesquels ces interactions ont lieu.

Ce besoin s'illustre très facilement par les défis de l'intégration des outils d'IA dans le système de soins. Ces outils soulèvent des questions concernant le rôle des professionnels de la santé et la nécessité de maintenir une approche centrée sur l'humain. Par exemple, le guide de la *Food and Drug Administration* (FDA) de 2022 souligne les risques liés au biais d'automatisation et travaille sur un nouveau cadre réglementaire pour les outils d'IA considérés comme des dispositifs médicaux (Harvey et Gowda 2020). Cette approche reflète une prise de conscience croissante de la nécessité d'équilibrer les avantages des outils d'IA avec les impératifs de sûreté et d'efficacité pour les patients.

Dans cette perspective, Babic et al. (2021), Gerke (2021) et Gerke et al. (2020) mettent en avant la nécessité de tests cliniques et de vérifications prospectives pour évaluer la performance des outils d'IA dans des contextes réels, dépassant les simples évaluations d'utilisabilité pour comprendre comment ces outils interagissent avec les processus décisionnels humains.

Par exemple, D. Li, Kulasegaram, et Hodges (2019) et Paranjape et al. (2019) prouvent l'importance d'améliorer la compréhension des résultats des outils d'IA par les professionnels de la santé pour permettre une intégration plus éclairée et efficace de ces technologies dans la pratique clinique. Cette nécessité est d'autant plus cruciale que l'expérience des travailleurs peut influencer significativement leur interaction avec les outils d'IA, notamment dans des domaines de connaissances spécialisées (W. Wang, Gao, et Agarwal 2023; Allen et Choudhury 2022). Par exemple, les médecins expérimentés sont plus susceptibles de rejeter les suggestions des outils d'IA (Jussupow et al. 2021), ce qui souligne l'importance de comprendre comment l'expérience influence la collaboration humain-IA.

En allant plus loin, Jussupow, Spohrer, et Heinzl (2022) expliquent que les radiologues interprètent la confirmation et la réfutation des outils d'IA de manière à former des schémas d'usage particuliers, et qu'il faut nécessairement passer par un examen dans les pratiques réelles pour les observer et comprendre.

Plus généralement, Zhang et al. (2021) ajoutent que ce besoin est renforcé par le caractère opaque des outils d'IA. Ils sont capables d'apprendre et d'exécuter de nouvelles actions indépendamment, sans que les relations entrée-sortie de ces outils soient connues des agents humains, à la fois *ex ante* et *ex post*.

De même, E. Wu et al. (2021) et Price (2019) soulignent l'importance de tests utilisateurs approfondis pour identifier et atténuer les biais émergents et contextuels pouvant survenir en raison de discordances entre la conception du produit et ses utilisateurs ou contextes d'utilisation. En complément, Kostick-Quenet et Gerke (2022) ainsi que Ansell et McDonald (2015) plaident pour une approche systématique visant à identifier et à réduire les biais utilisateurs lors de l'application des outils d'IA, suggérant la nécessité d'une prise de décision plus critique et réfléchie de la part des utilisateurs, car ils sont des imparfaitement rationnels.

La complexité de la collaboration entre les humains et les outils d'IA nécessite d'adopter une approche plus empirique avant de faire émerger un type de collaboration et les modes de gestion associés. Par conséquent, nous explorons un courant de littérature complémentaire : les travaux qui portent sur les usages des outils d'IA.

2.2. Les usages des outils d'IA

La littérature sur les usages des outils d'IA met en avant différentes caractéristiques de cette technologie (2.2.1), et les réponses que les acteurs expriment en retour (2.2.2). Cette littérature adopte un ensemble de prismes conceptuels pour théoriser ces usages (2.2.3), dont les concepts ont le potentiel, sans encore l'exprimer, de penser les modes de gestion de la collaboration humain-machine.

2.2.1. Caractéristiques des outils d'IA

Les chercheurs ont identifié plusieurs caractéristiques importantes des outils d'IA, telles que leur capacité à fonctionner à l'échelle mondiale, à fournir des résultats instantanés, à permettre une interaction dynamique avec les utilisateurs et à maintenir certaines parties de leurs processus inaccessibles ou opaques (Kellogg, Valentine, and Christin 2020). D'autres études définissent cet effet dit de « boîte noire », qui décrit la difficulté à comprendre comment ces outils prennent des décisions, la numérisation à une échelle globale, l'aptitude à prédire ou quantifier des événements avant qu'ils ne se produisent, et les orientations stratégiques non explicites qui guident le développement et l'application des outils d'IA (Faraj, Pachidi, and Sayegh 2018), ainsi que la capacité de ces technologies à évoluer et à s'adapter au fil du temps (Burrell 2016; Michalski, Carbonell, and Mitchell 2013).

Dans cette sous-partie, nous détaillons les caractéristiques les plus saillantes dans l'usage des outils d'IA. Nous abordons ainsi l'opacité des résultats des outils d'IA, la politique cachée de cette technologie, ainsi que la quantification anticipatrice, la valeur symbolique et l'émergence de travailleurs du clic.

Opacité des outils d'IA

Aujourd'hui, la plupart de la littérature existante sur l'opacité considère cette notion comme une propriété des outils d'IA, en particulier lorsqu'ils utilisent des méthodes d'apprentissage profond* (Burrell 2016; Pearl et Mackenzie 2018). Cette opacité renvoie à la difficulté pour les utilisateurs, voire pour les concepteurs, de comprendre les raisonnements derrière une décision lorsque ce raisonnement est dissimulé (Stohl, Stohl, et Leonardi 2016).

Elle provient de multiples sources (Burrell 2016). Tout d'abord, l'outil d'IA peut fonctionner en tâche de fond et donc son fonctionnement peut passer inaperçu pour les utilisateurs (Gillespie 2012). Pourtant, cet outil influence notre expérience quotidienne (de Vaujany 2022).

Même conscient, l'utilisateur n'a souvent pas accès au code qui est bloqué en raison du secret industriel. Pour de nombreux algorithmes, les choix de conception et le code restent internes à l'entreprise. Par exemple, des algorithmes tels que le moteur de recherche de Google et les recommandations vidéo de TikTok sont cruciaux pour leur succès commercial. Ils valent des milliards de dollars et sont étroitement protégés en tant que propriété intellectuelle (O'Neil 2016).

Cependant, même si l'information était disponible, les outils d'IA restent « inscrutables » car ils sont fondamentalement difficiles à comprendre et à examiner (Leonardi and Treem 2020). La compréhension du fonctionnement des algorithmes nécessite une expertise technique élevée. Selon Dourish (2016), même dans le cas d'un audit, seuls quelques

professionnels hautement qualifiés seraient capables de comprendre le code en raison de sa taille et de sa complexité.

De plus, ces outils d'IA sont fondés sur les techniques d'apprentissage machine* (chapitre 1, sous-section 1.3). Ils établissent des associations statistiques qui peuvent résulter de phénomènes purement probabilistes. Ce processus diffère des logiques de programmation traditionnelles et peut conduire à des résultats inattendus, souvent inexplicables même pour les concepteurs eux-mêmes (Dourish 2016).

Il peut être extrêmement difficile à retracer et à interpréter le processus de fonctionnement exacte (Pasquale 2015; Dourish 2016; Diakopoulos 2020), voire impossible (Gillespie 2014; Rudin 2019). Les chercheurs, y compris certains informaticiens, discutent désormais de « l'opacité fondamentale » des outils d'IA, arguant d'une impossibilité technique à produire de la transparence (Ananny et Crawford 2018).

Certains auteurs proposent deux autres façons de considérer l'opacité. D'une part, Lebovitz, Lifshitz-Assaf, et Levina (2022) déplacent le lieu auquel se rattache l'opacité. Traditionnellement, l'opacité est située dans la phase de conception, ce qui demande d'évaluer les algorithmes d'apprentissage et les données d'entraînement (Lebovitz, Levina, et Lifshitz-Assa 2021). Cependant, en adoptant l'approche de la sociomatérialité (Orlikowski 2000; Leonardi 2012), l'opacité devient le résultat d'un assemblage, un phénomène situé, produit à la fois par l'outil d'IA et par l'utilisateur (Lebovitz, Lifshitz-Assaf, et Levina 2022). Cette vision permet de questionner les leviers d'action susceptibles d'être mis en place par les utilisateurs pour gérer l'opacité de l'outil d'IA, au lieu de se limiter à des solutions techniques lors de la phase de conception.

D'autres auteurs préconisent de dépasser la distinction entre l'utilisation et la conception de l'outil d'IA pour comprendre l'opacité comme un phénomène issu du processus de conception/usage (Anthony, Bechky, et Fayard 2023). Ils s'appuient sur l'approche des réseaux sociotechniques (Callon 1986; Latour 1996) pour voir la création de connaissances comme des réseaux impliquant des acteurs humains et non humains. Par conséquent, ils analysent la dynamique sociale sous-jacente à la construction de l'outil d'IA, avant même son arrivée sur le lieu de travail. En somme, cette vision propose de chercher également des solutions organisationnelles (en plus des solutions techniques) dans la phase de conception de l'outil d'IA.

En quoi l'opacité des outils d'IA pose-t-elle problème ?

Certains considèrent que l'opacité des outils d'IA est positive. Moins les gens comprennent comment l'outil d'IA aboutit à un certain résultat, moins ils essaieront de « manipuler » le système (O'Neil 2016). Cela pourrait entraîner une diminution de la corruption dans divers cas tels que le scoring de crédit dans les processus de demande de prêt, les appels d'offres

concurrentiels ou l'évaluation de la performance au travail. Dans cette perspective, les outils d'IA peuvent fonctionner comme des panoptiques, imposant sur le lieu de travail une discipline souvent souhaitée par les managers, même si le procédé est contestable (Zuboff 1988; Burton-Jones 2014).

D'autres, ont une position neutre vis-à-vis de l'opacité des outils d'IA. Certaines littératures convergent avec l'intuition profane selon laquelle les outils n'ont pas besoin d'être compris pour être utilisés (Simon 1947). D'habitude, même si les utilisateurs ne comprennent pas le fonctionnement interne des technologies, ils peuvent les considérer comme fiables (W. C. Barley, Treem, et Kuhn 2018). Par exemple, il n'est pas nécessaire de comprendre comment les pistons s'enflamment dans un moteur à combustion interne pour démarrer et conduire une voiture ; nous faisons confiance à ceux qui ont conçu, construit et réglementé la voiture. En conséquence, le fonctionnement de la technologie peut être considéré comme acquis, fonctionnant d'une certaine manière sans compréhension ni observations directes. Les chercheurs ont qualifié cela de boîte noire de la technologie (Anthony 2018; 2021).

Cependant, l'opacité des outils d'IA limite les prises de décision critiques, au point de considérer d'éviter de les utiliser dans ces contextes (Burrell 2016; Cui et Zhang 2021). Outre des exemples encore hypothétiques, comme l'accident de la voiture autonome (Knight 2017), le cas de Medicaid illustre concrètement ce type de problèmes. Ce programme nord-américain garantissant une assurance maladie aux plus démunis avait changé son algorithme visant à financer l'assistance à domicile des soins. Tout ce qu'il transmettait alors était le refus de la décision (Hao 2020). La logique de l'outil d'IA était opaque au point où ni les patients, ni les prestataires de soins pouvaient comprendre la décision, provoquant de nombreux problèmes de coordination pour l'organisation et de mal-être pour les patients. De même, en 2020, le centre médical de Stanford avait commencé à utiliser un outil d'IA pour décider de la priorisation des vaccins contre la Covid. Or, le système s'est avéré dangereux, laissant de côté les professionnels de santé, pourtant en première ligne et endossant un rôle clé contre la pandémie (E. Guo et Hao 2022).

L'opacité des outils d'IA est également problématique dans les contextes d'expertise élevée, créant un malaise dans l'acceptation des décisions de l'outil d'IA (Burt 2019) et affectant négativement la relation entre l'humain et la machine. Des études sur la construction des connaissances et de l'expertise montrent que les experts ont un fort désir de comprendre et de décortiquer leurs outils (Bailey et Barley 2011; Anthony 2021). Les novices peuvent trouver les technologies obscures ou difficiles à comprendre, mais l'expertise dans une technologie se développe en parallèle de son exploration (Bailey, Leonardi, et Barley 2012). Plus un utilisateur est expert, plus il est capable de remettre en question ses outils.

En pratique, malgré les défis persistants liés à leur opacité, les outils d'IA sont de plus en plus mis en œuvre pour formuler des jugements dans des contextes critiques. Les

professionnels sont censés intégrer leurs propres connaissances aux résultats de l'outil d'IA lorsqu'ils formulent (Nunn 2018; Razorthink Inc 2019). Alors que le fonctionnement de l'outil d'IA reste opaque, son usage peut changer la distribution des responsabilités et de la redevabilité (*accountability*) des experts métier (Ananny et Crawford 2018; Donovan et al. 2018). Par exemple, dans la finance, les traders ne contrôlent plus leurs dispositifs financiers, ce sont les outils d'IA qui les soumettent à leur logique interne (Pasquale 2015; MacKenzie 2019; Anthony 2021).

Enfin, l'opacité des outils d'IA modifie les dynamiques de travail en équipe. La technologie, dont les technologies de l'information (Levina et Vaast 2005), peut servir d'objet frontière²⁶ (Star et Griesemer 1989), c'est-à-dire qu'elle peut permettre un partage des connaissances, faciliter la communication et rendre plus facile la collaboration entre des groupes avec des expertises, intérêts et pouvoirs différents (Carlile 2002; Bechky 2003).

Cependant, l'opacité du résultat de l'outil d'IA empêche les groupes de confronter différentes hypothèses et de comprendre les nuances de la situation. En fin de compte, cela pourrait éluder des conversations, sources de partages et de négociations entre ces groupes. Par exemple, l'outil d'IA pour le recrutement, vendu comme moyen d'améliorer l'efficacité et l'équité (Raghavan et al. 2020; Harwell 2022), peut renforcer les biais et réduire la diversité (Ajunwa et Greene 2019; Kellogg, Valentine, et Christin 2020). Ils peuvent classer des candidats sans qu'aucune personne ne voie les candidatures refusées, alors que les décisions de recrutement collégiales créent une meilleure adéquation des employés à l'entreprise (Cohen et Mahabadi 2022).

Politique cachée de la technologie

Benjamin (2019; 2020) montre comment les outils d'IA ont le potentiel de cacher, d'accélérer et d'approfondir la discrimination malgré une apparente neutralité, voire un fonctionnement bienveillant et équitable proclamés par leurs concepteurs.

La technologie n'est jamais « neutre ». À l'instar d'autres technologies, les outils d'IA sont politiques dès leur conception, comme le montre l'approche des réseaux sociotechniques (Akrich 1992; Winner 2007). Les concepteurs qui créent les technologies ont une certaine vision de la société et des futurs utilisateurs de la technologie. Cela influence donc implicitement ou explicitement, leurs décisions concernant le développement de l'outil

²⁶ Artefact ou un concept construit socialement qui joue un rôle essentiel dans la délimitation et la négociation des frontières entre différentes communautés de pratique et différents domaines de connaissance.

(Suchman 2007; Introna 2016). Ces choix sont souvent pris de manière informelle, intuitive et idiosyncrasique par l'individu qui conçoit l'algorithme (Ananny 2016; Kitchin 2014).

Cette vision idéalisée peut être liée à des préjugés. Par exemple, des chercheurs ont montré que des concepteurs ont produit des outils d'IA d'analyse faciale intégrant des biais qui pénalisent certaines minorités telles que les femmes à la peau foncée (Buolamwini et Gebru 2018). Les choix de développement sont également guidés par leurs idées politiques, reflétant des schémas d'évaluation, des croyances et des normes éthiques. Par exemple, un projet interne d'Apple visait à réécrire l'assistant vocal Siri pour ne pas « s'engager », « dévier » et enfin « informer » sur des « sujets sensibles » comme le féminisme et le mouvement #MeToo (Hern 2019). Intentionnellement ou non, les outils d'IA intègrent des préjugés sociaux et des idées politiques, accordant à certaines personnes, objets, idées et événements plus de pouvoir, de crédit, de visibilité ou d'importance que d'autres (Ananny 2016).

De même, les concepteurs ont certaines volontés et une vision propre du monde qui les motivent à entreprendre le développement d'un outil d'IA. Selon C. Steiner (2012), les concepteurs élaborent des algorithmes en tenant compte de deux dimensions : dans quelle mesure ils peuvent provoquer des changements conséquents dans le domaine visé et quels sont les avantages attendus pour l'usage associé. « *Déterminer le prochain domaine à envahir par les [algorithmes] est la somme de deux fonctions simples : le potentiel de perturbation ainsi que la récompense pour cette perturbation.* »²⁷ (C. Steiner 2012, 199). Que la motivation soit altruiste ou mercantile, l'ambition de développer un outil d'IA est portée par une envie de créer une nouvelle action collective, de changer le monde.

Enfin, au niveau des concepteurs, le manque d'expertise peut également introduire des idées fausses ou biaisées dans l'outil d'IA (Diakopoulos 2015; Schuetz et Venkatesh 2020; Choudhury, Allen, et Endres 2021).

Au-delà des intentions des concepteurs, le jeu de données servant à entraîner l'outil d'IA provoque également des effets politiques. Les données ne sont jamais « données », elles ne sont pas « brutes » (Gitelman 2013). La constitution de bases de données fait l'objet d'un processus (Aaltonen, Alaimo, et Kallinikos 2021). Comme l'argumentent Desrosières et Knott (2005, 2), ces données sont souvent issues d'une « quantification », c'est-à-dire une façon d'« *exprimer et de faire exister sous une forme numérique ce qui, auparavant, était exprimé seulement par des mots et non par des nombres* », provenant d'une convention et d'une

²⁷ *Determining the next field to be invaded by bots [automated algorithms] is the sum of two simple functions: the potential to disrupt plus the reward for disruption.*

mesure. Loin d'un reflet du monde, le jeu de données implique sa classification, avec des inclusions et des exclusions, qui s'accroissent lors de la collecte de données. De plus, Loukissas (2019) défend l'idée que « toute donnée est locale » : un jeu de données est attaché à un lieu, il reproduit des connaissances localisées. Par conséquent, les résultats des outils d'IA n'ont pas la valeur universelle qui leur est prêtée. Ils sont difficilement applicables en dehors du référentiel où les données d'entraînement ont été produites.

Cette vision du monde et cette localité du savoir se transfèrent ensuite à l'outil d'IA, encodant un ensemble particulier de valeurs dans un modèle formel réifié (Lindebaum, Vesa, et Den Hond 2020). Par exemple, pour un outil d'IA déterminant l'individu qui peut avoir un crédit, le jeu de données d'entraînement peut introduire un biais en associant les minorités à faible revenu aux personnes avec une faible probabilité de remboursement, discriminant davantage un groupe défavorisé (O'Neil 2016).

Ensuite, les biais des outils d'IA influencent les décisions auxquelles ils participent. Les valeurs portées par l'outil d'IA en raison des concepteurs et du jeu de données d'entraînement, orientent vers une vision du monde. Le résultat de l'outil d'IA médiatise un champ de possibles, et les utilisateurs agissent en fonction de ces résultats (Murray, Rhymer, et Sirmon 2021; Balasubramanian, Ye, et Xu 2022).

Par exemple, dans le système de justice pénale aux États-Unis, l'outil d'IA COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) évalue la probabilité qu'un prévenu criminel récidive, notamment en fonction de l'historique criminel, des facteurs sociaux et des facteurs psychologiques. Il sert à prendre des décisions concernant la caution, la peine et la libération conditionnelle. Or, il est l'objet de controverses et de critiques car il amplifie les biais présents dans le système de justice pénale, notamment le biais racial (Israni 2017; Angwin et al. 2022).

Malgré les arguments à son encontre, COMPAS demeure largement utilisé, opaque pour les juges et impénétrable pour un audit public. De nombreux exemples abondent dans ce sens, comme les outils d'IA qui prédisent quels patients bénéficieront d'une attention médicale supplémentaire. Souvent, ils sous-estiment les besoins de santé des patients noirs les plus malades, amplifiant les disparités raciales historiques dans la médecine (C. Y. Johnson 2022). Ces exemples illustrent l'ensemble des politiques cachées des outils d'IA, leur fondement étant intimement lié à celui de l'opacité.

Autres caractéristiques

Quantification anticipatrice

En lien avec la politique cachée des outils d'IA, Faraj et al. (2018, 68) évoquent également l'effet de « quantification anticipatrice ». Les outils d'IA transforment de vastes quantités de données en prédictions ou probabilités, une tendance soutenue par la logique de la

quantification numérique (Espeland and Stevens 2008; Wagner-Pacifici, Mohr, and Breiger 2015). Ils peuvent prédire avec précision des aspects personnels tels que l'orientation sexuelle ou les préférences politiques à travers les traces que nous laissons sur Internet (Grassegger and Krogerus 2017; Youyou, Kosinski, and Stillwell 2015). Cette approche prétend pouvoir représenter le monde grâce à l'accumulation des données, mais les modèles prédictifs produits sont réducteurs et peinent à saisir la complexité et l'évolution des individus (Ananny 2016). La représentation numérique est imparfaite et incomplète. Les individus sont réduits à des catégories algorithmiques sans prendre en compte l'intégralité de leurs caractéristiques (Boyd and Crawford 2012).

Valeur symbolique

L'outil d'IA, porteur d'une forte valeur symbolique (Anthony, Bechky, et Fayard 2023), captive l'imagination publique au-delà des capacités techniques envisageables (Larson 2021). Cet imaginaire, nourri par la science-fiction et des visionnaires tels qu'Irving John Good (1966) et le philosophe Bostrom (2014), dépasse celui d'autres technologies comme le profilage ADN (Bechky 2020) et l'ingénierie (Elsbach 2003) par son ampleur. À la différence de ces dernières, souvent confinées à des contextes spécialisés, les récits autour de l'IA influencent l'économie entière, des travailleurs aux industries, générant un buzz important sur la collaboration humain-IA. Or, les technologies en tant que source d'autorité symbolique influencent les hiérarchies sociales au sein d'une organisation (S. R. Barley 1986; Bechky 2003).

Travailleurs du clic

Cette valeur symbolique fabrique aussi des mythes qui masquent le travail humain derrière les outils d'IA. Le développement des outils d'IA influence la division du travail en créant un nouvel espace de sous-traitance avec les plateformes de crowdsourcing et les travailleurs du clic, pour générer, annoter, et enrichir les données d'entraînement (Roberts 2019; Le Ludec, Cornet, et Casilli 2023). Il s'agit de l'exécution de tâches simples et répétitives pour lesquelles il n'est pas nécessaire d'être spécifiquement qualifié. Ce travail représente environ 80 % du temps de développement d'un outil d'IA.

Ces nouvelles formes de travail induites sont parfois perçues comme une forme de colonialisme moderne du fait de l'équilibre de pouvoir entre ceux qui en bénéficient, les gestionnaires et les investisseurs, et les travailleurs du clic dont l'activité est dévaluée (Shestakofsky 2017; Le Ludec, Cornet, et Casilli 2023).

L'exemple canonique est le service d'Amazon Mechanical Turk²⁸, dont la technologie reposait non pas sur une réelle automatisation, mais sur une succession d'actions manuelles réalisées par des opérateurs humains, souvent dans des conditions de travail pénibles (Aytes 2012).

L'ensemble des caractéristiques ci-dessus – l'opacité des résultats des outils d'IA, la politique cachée de cette technologie, ainsi que la quantification anticipatrice, la valeur symbolique et l'émergence de travailleurs du clic – contient les éléments les plus saillants et les plus discutés dans la littérature.

De même, bien que nous ayons présenté ces caractéristiques de manière distincte, il est crucial de tenir compte de leurs interactions et combinaisons, qui peuvent conduire à l'émergence de comportements et d'effets imprévisibles (Anthony, Bechky, et Fayard 2023). Par exemple, l'étude d'Elish et Watkins (2020) sur Sepsis Watch, un outil d'IA destiné à prédire le risque de septicémie chez les patients, a révélé que les infirmières devaient déployer des efforts considérables pour expliquer aux médecins traitants le score de risque fourni par l'outil, en raison de l'opacité de ses résultats. De plus, le fonctionnement de l'outil changeait sans préavis, compliquant encore davantage la tâche de compréhension et d'interprétation pour les infirmières.

Prendre en compte ces caractéristiques est essentiel pour établir des modes de gestion efficaces de la collaboration humain-machine, un aspect souvent négligé dans la littérature antérieure. À présent, nous explorons comment les études sur les usages des outils d'IA reflètent la manière dont les acteurs s'approprient cette technologie.

2.2.2. Usages développés en réponse aux caractéristiques des outils d'IA

Au regard de ces caractéristiques, les acteurs expriment des réponses variées. Nous détaillons ci-dessous les quatre grands types de réponses que nous avons identifié. À un extrême, Kostick-Quenet et Gerke (2022) évoquent que la méfiance envers les outils d'IA consiste à rejeter les résultats de manière systématique, ce que nous explorons dans un premier temps. À l'opposé, certains utilisateurs accordent une confiance aveugle et inventent des récits factices sur leur fonctionnement, ce que nous regardons ensuite.

Ces deux types d'attitudes sont limitantes et peuvent avoir des impacts négatifs, par exemple pour des diagnostics cliniques, avec une sous-dépendance pouvant entraîner des

²⁸ En référence au nom d'une machine de la fin du 18^{ème} siècle qui prétendait être un automate joueur d'échecs alors qu'un opérateur caché manipulait les pièces grâce à un système ingénieux.

erreurs d'omission et une sur dépendance, des erreurs de commission (Kostick-Quenet et Gerke 2022).

Entre ces deux extrêmes, des études ont montré les moyens mis en place par des équipes pour adopter un outil d'IA, ce que nous détaillons dans un troisième temps.

Enfin, d'autres études s'intéressent à des acteurs seuls face à l'outil d'IA, comme les experts métier, pour voir comment ils développent des pratiques pour intégrer les résultats des outils d'IA dans leur réflexion.

Ignorer les résultats

Si certains concepteurs font semblant que l'outil d'IA fonctionne de manière autonome, alors qu'ils sous-traitent les calculs aux travailleurs du clic (Aytes 2012), il en va de même pour les utilisateurs. Certains font semblant d'adopter les résultats des outils d'IA (Pachidi et al. 2021; Lebovitz, Lifshitz-Assaf, et Levina 2022). Par exemple, Valentine et Hinds (2022) ont réalisé une ethnographie dans une entreprise du commerce de détail et elles ont montré que les employés du service achat de mode préféraient leur propre compréhension du goût des clients aux suggestions de l'outil d'IA que l'équipe de data scientists avait mise au point.

La combinaison de l'opacité de l'outil d'IA et de sa valeur symbolique semble favoriser ce type d'usage. Les connaissances d'un outil d'IA diffèrent aussi de celles d'un être humain, ce qui implique de remettre en question les savoirs existants en cas d'adoption, notamment ce qui constitue des informations utiles et la manière dont elles sont produites (Boyd et Crawford 2012).

Par exemple, Jussupow, Spohrer, et Heinzl (2022) observent que certains radiologues abandonnent les outils d'IA pour l'aide à la prise de décision clinique : ils ne s'appuient pas sur ses évaluations binaires pour prendre des décisions précoces. De même, ils n'enregistrent pas les résultats de l'outil d'IA dans leurs rapports médicaux. Ces médecins vont jusqu'à demander à retirer ces outils d'IA de la pratique clinique. En contrepartie, ils soulignent l'importance des compétences personnelles, de l'expérience et de la compétence comme décisives pour un diagnostic correct.

En effet, une collaboration engagée avec l'outil d'IA peut nécessiter des changements substantiels dans la façon de travailler et dans la structure organisationnelle (Faraj, Pachidi, et Sayegh 2018; Zuboff 2019). Lorsque les outils d'IA sont imposés par la hiérarchie, faire semblant d'adopter l'outil est une forme de résistance à ces changements.

Par exemple, Pachidi et al. (2021) décrivent comment l'équipe de vente d'une entreprise fait semblant d'utiliser un outil d'IA développé par une équipe de data scientists, tout en conservant ses pratiques traditionnelles. L'outil d'IA propose une autre lecture du monde, de la façon de considérer les clients en fonction de métriques et de données comportementales,

alors que l'équipe de vente préfère garder une connaissance des clients en fonction de leurs échanges réguliers et de leur intuition. Les concepteurs étaient au courant de la duperie, mais ils préféraient laisser faire pour maintenir l'apparence que l'outil d'IA était adopté auprès de la direction de l'entreprise.

De plus, comme le résultat de l'outil d'IA est opaque, les experts métier ne font pas nécessairement confiance à l'outil, surtout s'il entre en conflit avec leur expertise. Par exemple, Lebovitz, Lifshitz-Assaf, et Levina (2022) constatent que lorsque les radiologues ne peuvent pas discerner comment l'outil d'IA parvient au diagnostic qu'il propose, de nombreux médecins décident de l'ignorer et de se fier à leur propre jugement professionnel. Plus précisément, dans leur ethnographie, les chercheuses montrent comment les outils d'IA pour l'aide à l'analyse des mammographies génèrent des résultats opaques et augmentent l'incertitude de la situation. Ainsi, face à des résultats souvent divergents de l'outil d'IA, les médecins ont tendance à rejeter les zones signalées comme non pertinentes en raison de leur habitude à considérer les signalements d'anomalies comme des faux positifs. Par conséquent, les résultats sont jugés superflus, ajoutant une charge de travail sans valeur ajoutée significative à leur processus décisionnel. La non-intégration des résultats de l'outil d'IA s'explique en partie par l'absence de suivi complet de l'évolution de la santé des patients, ce qui rend difficile l'évaluation de l'impact réel de ces décisions sur les traitements.

Le fait que les experts puissent choisir de faire semblant d'adopter une technologie opaque soulève de nombreuses questions, surtout en considérant que l'outil d'IA est un « outil » en relation avec un expert métier (Anthony, Bechky, et Fayard 2023). Par exemple, quelles conditions favorisent ce type d'usage ? Quels sont les bénéfices et les dangers de faire semblant ? Par exemple, dans le cas étudié par Pachidi et al. (2021), la direction finit par valider l'adoption de l'outil d'IA.

Inventer des récits

Lorsque les utilisateurs s'engagent dans l'usage de l'outil d'IA, ils mettent en place des récits pour répondre à l'opacité des résultats. Pour Dourish (2016, 7), « *chaque acte de catégorisation – ou plus précisément de narration – est un mouvement du grand vers le petit, une réduction de la masse de points à un élément narratif ou à une caractéristique inhérente, issue généralement du domaine que l'on cherche à connaître* »²⁹. En d'autres termes, les récits aident les acteurs à donner du sens, à interpréter en fonction du contexte et de la situation

²⁹ *Each act of categorization – or more accurately, of narration – is a move from big to small, a reduction of a mass of data points to a narrative element or a defining characteristic, drawn generally from the domain of which we want to know.*

dans laquelle ils se trouvent le résultat – par nature abstrait – de l’outil d’IA. Pour le comprendre, nous pouvons également évoquer le processus de sensemaking (Weick, Sutcliffe, et Obstfeld 2005) dans lequel la construction de sens implique de transformer les circonstances en une situation explicitement comprise en mots qui servent de tremplin à l’action. Ainsi, ces récits jouent un rôle central dans la manière dont les acteurs interagissent avec les outils d’IA, en évitant de les réduire à de simples boîtes noires.

Les récits sont construits par le collectif. Les utilisateurs s’engagent dans des interactions sociales variées, allant de conversations informelles à des réunions structurées, pour explorer et comprendre les outils et leurs résultats (S. R. Barley 1986; Lave et Wenger 1991; Orr 1996; Bechky et Chung 2018; Anthony 2021). Par exemple, Borch et Min (2022) étudient le cas d’une entreprise de trading qui adopte des outils d’IA. Ils décrivent comment l’anthropomorphisation de l’outil d’IA, c’est-à-dire le fait de parler de l’outil comme s’il était humain, ainsi que la « technomorphisation », où les utilisateurs tentent de comprendre la machine en adoptant son point de vue, contribuent à mieux saisir la dynamique humain-machine. Pour clarifier son fonctionnement, l’outil d’IA devient un compagnon, et à l’instar d’autres relations humaines, sa compréhension varie en intensité.

Cependant, ces récits peuvent reposer sur un raisonnement « magique » (Anthony, Bechky, et Fayard 2023). Ils ne sont pas toujours fondés sur des preuves tangibles du fonctionnement des outils d’IA. Les récits qui viennent éclairer l’effet boîte noire des technologies s’appuient souvent sur des expériences antérieures avec d’autres technologies, avec le risque qu’elles n’aient que peu de pertinence.

Par exemple, S. R. Barley (1986) a observé que des techniciens utilisaient leur expérience avec des technologies antérieures pour diagnostiquer des scanners CT, ce qui conduisait à des suppositions incorrectes.

De même, Elish et Watkins (2020) ont remarqué que des infirmières s’appuyaient sur des récits imprécis pour expliquer les scores de risque de sepsis générés par un outil d’IA, ce qui, par chance, n’a pas entravé la mise en place du bon traitement. Ainsi, ces récits, même inexacts, influencent les actions.

Un autre problème des récits sur un outil d’IA émerge de son évolution (Anthony, Bechky, et Fayard 2023). Le décalage entre la rhétorique et la réalité s’intensifie lorsque les logiques sous-jacentes de l’outil d’IA changent à l’insu des utilisateurs en raison de son opacité. C’est pourquoi, les récits incorrects pourraient conduire à des actions exacerbant les problèmes et aboutissant à des résultats inattendus. Les conséquences des récits non vérifiés et obsolètes sur l’outil d’IA méritent d’être davantage étudiées, tout comme la manière dont ces récits changent avec la technologie, et comment les utilisateurs forment, partagent et convainquent les autres de la validité de leurs perceptions.

Intégrer les outils d'IA dans des équipes

Les outils d'IA au sein d'équipes suscitent un grand nombre de questions, souvent inspirées des recherches traditionnelles sur les dynamiques de collaboration (Bunderson et Sutcliffe 2002; Carlile 2002; Bechky 2003; Carlile 2004; Maguire, Hardy, et Lawrence 2004; Hardy, Lawrence, et Grant 2005; Levina et Vaast 2005; Levina 2005; Belbin 2010; Fayard et Metiu 2014; Hardy, Lawrence, et Grant 2005; Van De Water, Ahaus, et Rozier 2008; Okhuysen et Bechky 2009; Tsoukas 2009). Par exemple, l'outil d'IA au sein d'une équipe suscite aide-t-il les acteurs à résoudre des problèmes et à accomplir des tâches conjointes dans les organisations ? Quelles sont les qualités de l'outil d'IA en tant qu'objet-frontière³⁰ ? Comment modifie-t-il les pratiques par lesquelles les équipes développent une compréhension commune pour travailler ensemble ? De quelle manière pourrait-il modifier les hiérarchies ? Comment l'usage de l'outil d'IA pourrait-il faciliter ou au contraire restreindre la négociation et la création de sens du groupe ?

Deux grands types d'usage apparaissent lors de l'intégration des outils d'IA dans une équipe. Tout d'abord, plusieurs équipes peuvent œuvrer à l'intégration d'un outil d'IA, à corriger et à expliquer les résultats pour les rendre compatibles avec le travail de l'organisation (Elish et Watkins 2020). Cependant, ce type d'intégration de l'outil d'IA n'est pas une forme de collaboration dans laquelle les équipes résolvent des problèmes et apprennent davantage sur le travail des autres, mais plutôt une manière d'intégrer efficacement le résultat dans les processus organisationnels. De plus, même s'il est réalisé par des négociations entre plusieurs groupes, l'interprétation des résultats de l'outil d'IA peut propager des récits magiques.

Par exemple, Sachs (2020) parle d'un « travail de réparation ». Dans son ethnographie, il observe comment des groupes d'experts en art identifient et résolvent des « effondrements sociotechniques », c'est-à-dire des résultats de l'outil d'IA qui ne correspondent pas ou ne répondent pas aux attentes, aux normes ou aux objectifs des experts métier ou des utilisateurs. Ce travail itératif de maintenance, d'interprétation et de réparation des outils d'IA souligne l'interaction continue et les négociations entre les experts métier et l'outil pour s'assurer que les résultats soient culturellement crédibles et adaptés à la situation. Sachs (2020) montre également que ce travail des experts, introduisant leurs jugements subjectifs, peut devenir invisible dans le produit final, perpétuant l'illusion de l'objectivité de l'outil d'IA.

³⁰ Un objet-frontière, ou objet-limite, fait référence à un artefact ou un concept qui se situe à l'intersection de diverses disciplines et qui ont la capacité à transcender les limites disciplinaires conventionnelles pour intégrer des connaissances multidisciplinaires ou transdisciplinaires.

Un autre moyen d'intégrer l'outil d'IA dans une équipe est de mettre en place un nouveau type d'acteur qui fasse le lien entre les résultats de l'outil d'IA et le reste de l'organisation. Ce type de professionnels dédiés est appelé Knowledge Brokers. Ils agissent comme des traducteurs entre les résultats de l'outil d'IA et le contexte organisationnel dans lequel il s'inscrit. Cette profession émergente fait l'objet d'une littérature spécialisée en sciences de gestion (Haas 2015), même si des questions comme les conditions d'émergence du Knowledge Broker et le codage de ses activités sont encore ouvertes (Kwon et al. 2020).

L'étude de Waardenburg, Huysman, et Sergeeva (2022), une ethnographie de 31 mois dans une unité de police intégrant un outil d'IA pour prédire la récurrence des crimes, montre que les Knowledge Brokers prennent une forme de pouvoir qui dépasse la simple « traduction » des résultats. Au début, ils jouent le rôle de messagers pour rendre signifiants les résultats de l'outil aux yeux de leurs collègues sur le terrain. Face aux nombreuses lacunes des résultats de l'outil à cause de la décontextualisation, les chefs et les data scientists leur demandent de prendre un rôle d'interprète, c'est-à-dire de prendre appui sur leurs propres connaissances. Pour répondre à cette nouvelle exigence, les Knowledge Brokers doivent alors apprendre davantage, sur la communauté policière et sur celle des concepteurs. Ces nouveaux apprentissages acquis, les Knowledge Brokers se mettent à filtrer l'information qu'ils jugent pertinente.

Ainsi, les relations de statut entre les groupes peuvent changer lorsque l'outil d'IA est introduit. Cela renvoie à la valeur symbolique de l'outil d'IA en générant une autorité symbolique et une légitimité au groupe qui s'en fait l'interprète. L'outil d'IA, comme toute technologie, peut façonner les hiérarchies de statut (S. R. Barley 1986; Bechky 2003).

Développer des pratiques interrogatives

Dans un contexte où des experts métier doivent prendre des décisions critiques, par exemple en médecine, en droit et en finance, l'usage des outils d'IA demande de vérifier, interroger et expliquer les résultats de cette technologie parce que les professionnels restent responsables du diagnostic et des conseils (Kluttz et Mulligan 2019; Anthony 2021; Lebovitz, Lifshitz-Assaf, et Levina 2022). Sans pouvoir déléguer ce travail à des Knowledge Brokers et devant éviter d'inventer des récits magiques, les experts métier qui s'engagent dans la collaboration avec les outils d'IA doivent ancrer leur compréhension des résultats algorithmiques dans leur contexte.

Ici, la question centrale est celle de la gestion de l'opacité des résultats de l'outil d'IA en fonction de l'expertise de l'humain. Comment l'utilisateur travaille-t-il avec une technologie opaque ? Et dans quelles conditions les avantages d'utiliser une « boîte noire » l'emportent sur les risques qu'elle génère ? De même, où situer la responsabilité lorsque l'utilisateur prend une décision en fonction de résultats dont il ne peut pas vérifier la fiabilité ? Quand et

comment les utilisateurs peuvent-ils choisir de ne pas utiliser l'outil d'IA ? Et, plus fondamentalement, comment les experts développent-ils des compétences avec des technologies opaques ?

En menant des expériences en laboratoire avec des radiologues novices et expérimentés pour diagnostiquer des cas de patients à l'aide d'un outil d'IA, Jussupow, et al. (2021) montrent que les médecins utilisent des processus cognitifs de second ordre, qu'ils appellent des « métacognitions », pour surveiller et contrôler leur raisonnement tout en évaluant les conseils de l'outil. Ces métacognitions déterminent si les médecins sont capables de tirer pleinement parti de l'outil d'IA ou non. Parmi ces processus cognitifs, ils distinguent ceux déclenchés par des résultats de l'outil d'IA confirmant leur évaluation et ceux qui la contredisent. Pour naviguer avec succès dans ces conditions, les experts métier doivent concevoir un raisonnement plus élaboré que celui amenant à produire des conseils confirmant l'outil d'IA.

En poursuivant, Jussupow, Spohrer, et Heinzl (2022) identifient certains radiologues ayant un usage intensifié. Ces experts métier interagissent avec le système pour compenser les erreurs. Ils ont accumulé des connaissances et des pratiques visant à garantir qu'aucune évaluation erronée n'affecte le traitement de leurs patients, ni les erreurs produites par l'outil, ni les erreurs produites par eux-mêmes. Dans ce cadre, aucun des radiologues ne se repose sur l'outil d'IA comme moyen principal pour diagnostiquer le cas d'un patient. Ils ont tous les compétences et connaissances nécessaires pour évaluer les cas de patients indépendamment et le font selon la routine clinique. Par exemple, pour un traitement rapide d'AVC aigu, l'outil d'IA sert à prendre des décisions cliniques plus rapides lorsqu'il confirme les diagnostics du radiologue. Plus tard, les médecins reviennent dessus pour créer un rapport détaillé. Les vues détaillées et quantifiées sur différents segments du cerveau du patient sont ainsi particulièrement utiles.

Dans ce contexte, Lebovitz, Lifshitz-Assaf, et Levina (2022) développent la notion de « pratiques interrogatives » pour éclairer la manière dont les experts questionnent les résultats de l'outil d'IA, offrant ainsi une approche où l'expert métier interagit avec l'outil de manière à gérer son opacité et à intégrer les résultats dans la tâche à accomplir.

L'ethnographie de Lebovitz, Lifshitz-Assaf, et Levina (2022) explore le travail des radiologues dans trois services d'un hôpital américain, couvrant l'imagerie pédiatrique, thoracique et mammaire. Chaque service utilise un outil d'IA spécifique pour le diagnostic. Le diagnostic des patients, un processus critique pouvant mettre leur vie en danger, nécessite une expertise significative des radiologues, acquise au fil d'années d'études et de pratique. La qualité des soins est cruciale non seulement pour la vie des patients, mais aussi parce que les complications médicales représentent un fardeau pour l'hôpital. Ignorer une tumeur maligne pourrait avoir des conséquences désastreuses. Cependant, dans un environnement soumis à

des contraintes fortes, le temps consacré aux patients tend à diminuer pour augmenter le volume des examens médicaux. Ainsi, les données empiriques suggèrent la perspective d'une collaboration humain-machine, en mettant particulièrement l'accent sur la relation entre l'expert métier et l'outil d'IA.

L'analyse préliminaire de Lebovitz, Lifshitz-Assaf, et Levina (2022) révèle une frustration générale chez les radiologues lorsque les résultats de l'outil d'IA divergent de leur diagnostic, souvent en raison d'une compréhension insuffisante du fonctionnement de l'outil. Bien que les trois outils d'IA partagent des caractéristiques similaires, comme la technique d'apprentissage profond, seuls les radiologues en imagerie thoracique intègrent les résultats de l'outil d'IA dans leur raisonnement, contrairement aux deux autres départements où les radiologues finissent par ignorer l'outil.

Lebovitz, Lifshitz-Assaf, et Levina (2022) apportent diverses explications à cette divergence. La différence fondamentale identifiée chez les radiologues thoraciques (qui incorporent les résultats de l'outil d'IA) réside dans leur approche consistant à interroger celui-ci avec leur savoir-faire médical. Ces pratiques, appelées « pratiques interrogatives », représentent une contribution significative à la gestion de l'opacité des outils d'IA. Elles modifient la perspective en soulignant que l'opacité n'est pas exclusivement une caractéristique intrinsèque de l'outil d'IA, mais plutôt liée aux usages qu'en font les acteurs. Ces pratiques complètent les solutions traditionnelles telles que la limitation des outils d'IA dans des contextes critiques ou la recherche d'une plus grande explicabilité. Bien que les pratiques interrogatives ne remplacent pas les explications détaillées des raisonnements des outils d'IA, elles offrent aux acteurs des moyens de contextualiser la situation présentée par ces outils.

De plus, Lebovitz, Lifshitz-Assaf, et Levina (2022) apportent une précision au concept d'augmentation dans la collaboration humain-machine. Contrairement à l'idée courante selon laquelle l'objectif est d'augmenter la performance des tâches, en particulier dans des contextes critiques, cette augmentation devrait bénéficier aux experts métier. Cependant, cette amélioration n'est pas toujours observée, notamment lorsque les experts métier ne s'engagent pas dans l'utilisation des outils d'IA. Ainsi, bien que la notion de pratiques interrogatives permette d'identifier les cas où l'usage augmente l'expert métier, elle ne précise pas le type d'augmentation en question.

Enfin, en adoptant une perspective relationnelle sur l'expertise métier, Pakarinen et Huisling (2023) suggèrent que les outils d'IA deviendront partie intégrante du réseau d'acteurs et de connaissances en fonction de la manière dont les groupes de professionnels développeront des usages. Par conséquent, la manière dont les experts métier mettent en place des métacognitions (Jussupow, Spohrer, et Heinzl 2022), dont les pratiques interrogatives (Lebovitz, Lifshitz-Assaf, et Levina 2022), semble être la clé du type de collaboration humain-machine adopté.

Les résultats obtenus sur les caractéristiques des outils d'IA et les réponses des acteurs fournissent une description détaillée des usages des outils d'IA, et même une perspective initiale sur ce qui pourrait constituer des modes de gestion de la collaboration humain-machine, notamment à travers des pratiques interrogatives (Lebovitz, Lifshitz-Assaf, et Levina 2022). Toutefois, nous constatons qu'une étape supplémentaire est nécessaire pour élaborer des modèles de collaboration accompagnés de leurs modes de gestion correspondants. Pour mieux cerner cette lacune, nous approfondissons notre compréhension de cette littérature en examinant les prismes théoriques sur lesquels elle développe ces concepts. Nous recherchons si ces prismes disposent des outils conceptuels nécessaires pour établir des modes de gestion de la collaboration entre l'humain et les outils d'IA.

2.2.3. Prismes conceptuels pour étudier les usages des outils d'IA

Nous décrivons ici brièvement les prismes conceptuels en Théorie des Organisations et en Gestion des Systèmes d'Information utilisés dans les études antérieures : l'approche de l'affordance, l'approche des réseaux sociotechniques et l'approche de la sociomatérialité. La synthèse de ces approches montre l'importance d'analyser le pouvoir d'agir, ou « agency », des acteurs humains et des outils d'IA. Étudier ces nouvelles dynamiques de pouvoir et les relations sociales qui les façonnent permet de discerner les modes de gestion d'une collaboration humain-machine.

Approche de l'affordance

Hutchby (2001) critique les approches constructivistes en études des sciences et des techniques, qui considèrent que les objets technologiques soient des *tabula rasa* et dont la signification est uniquement donnée par l'interprétation de leurs utilisateurs (Bijker 1997; Bijker, Hughes, et Pinch 2012).

Pour réconcilier les perspectives réalistes et constructivistes, il réintroduit le concept d'affordance de Gibson (Gibson 2014), selon lequel les possibilités d'action offertes par un objet technologique ne changent pas en fonction des besoins de l'observateur : « *l'affordance de quelque chose est supposée ne pas changer en fonction des besoins de l'observateur* »³¹ (Gibson 1982, 409).

Issu de la psychologie de la perception, Hutchby place ce concept dans un cadre sociologique, soulignant que les objets technologiques offrent des possibilités d'action

³¹ The *affordance* of something is assumed *not* to change as the need of the observer changes.

inhérentes, ou « affordances », qui ne dépendent pas des interprétations subjectives mais guident et limitent les actions des utilisateurs selon leur matérialité intrinsèque.

Hutchby élargit cette notion pour qu'elle soit applicable au-delà des caractéristiques tangibles des objets, incluant également des facteurs tels que le contexte social et les dynamiques d'interaction (Hutchby 2001). Par exemple, les systèmes experts* partagent une matérialité au niveau de leur interface, affectant la navigation et restreignant les interactions à ce qui a été rendu possible par le concepteur (Hutchby 2013).

Zammuto et al. (2007) explorent les implications des affordances dans le cadre des technologies de l'information, observant comment elles transforment les modes de communication et de décision au sein des organisations.

Burton-Jones et Volkoff (2017) ainsi que Bansal et Shukla (2021) prolongent cette réflexion en plaidant pour des théories spécifiques qui prennent en compte la nature dynamique et contextuelle des affordances.

Cette approche met en avant la nature dynamique et relationnelle des affordances, en lien avec les performances des outils d'IA et la perception des capacités des utilisateurs. Par exemple, Fügener et al. (2021) montrent que la décision de déléguer une tâche à un outil d'IA est motivée par la perception qu'ont les utilisateurs de leur capacité à effectuer la tâche. Cependant, cette perception est souvent biaisée, car les humains ont tendance à surestimer leurs capacités. De plus, l'aversion envers les conseils algorithmiques est motivée par la confiance relative perçue des utilisateurs dans leurs propres capacités par rapport à la précision perçue de l'outil d'IA (Dietvorst, Simmons, et Massey 2015; Dietvorst et Bharti 2020).

En somme, l'approche de l'affordance se centre sur la « matérialité ». Les affordances perçues par les utilisateurs résultent de la matérialité de l'objet technologique, de l'organisation et de l'environnement. Cette approche souligne deux aspects importants pour la collaboration humain-machine : les collaborations dépendent du couplage entre l'outil d'IA et le contexte d'usage, et il est crucial d'identifier l'ensemble des affordances dans ce couplage pour établir les modes de gestion de la collaboration les plus efficaces.

Approche des réseaux sociotechniques

L'approche des réseaux sociotechniques constitue une avancée majeure dans les études des sciences et des techniques, contrastant avec la tradition anglo-saxonne de la « construction sociale de la technologie » (Bijker 1997; Bijker, Hughes, et Pinch 2012). Établie par Akrich, Callon et Latour du Centre de Sociologie de l'Innovation de l'École des Mines de Paris (Muniesa 2015), cette perspective se concentre sur l'intégration des éléments techniques et sociaux dans ce qu'ils appellent la « sociologie de la traduction », le « réseau socio-matériel » ou la « théorie de l'acteur-réseau » (ANT).

Initialement employée pour comprendre la formation de faits scientifiques (Latour et Woolgar 1979), cette approche a été étendue pour analyser comment les éléments scientifiques, techniques et sociétaux s'entremêlent et s'influencent mutuellement dans des « réseaux sociotechniques » (Latour 1996). Ces réseaux sont des ensembles hybrides qui dépassent les dichotomies traditionnelles entre nature et culture, en soulignant comment les faits scientifiques, bien que socialement construits, s'appuient sur des objets physiques pour se matérialiser.

Cresswell, Worth et Sheikh (2010) ont démontré l'utilité de cette approche pour analyser l'introduction des technologies informatiques dans les systèmes de soins, en tenant compte de la complexité organisationnelle et du rôle actif de la technologie. Concernant les outils d'IA, Kim et Kwon (2022) la mobilisent afin d'analyser le système de recommandations de Facebook, critiquant comment les utilisateurs sont limités par des choix et des actions autonomes qui servent les intérêts de l'opérateur.

Cette vision reconnaît que les artefacts technologiques, les institutions et les discours politiques agissent au sein de réseaux qui les définissent et les constituent (Latour et Woolgar 1979; Latour 1993; 1996). La notion « d'actant », centrale dans cette approche, désigne à la fois les acteurs humains et non-humains qui participent à la dynamique du réseau (Akrich et Latour 1992). De même, la notion de « script » décrit comment un objet technologique favorise certains comportements des utilisateurs, à l'image du script de film (Akrich 1992; Latour 1992). Ces scripts technologiques révèlent les politiques cachées des outils d'IA, soulignant la nature performative de la technologie.

Dans le prolongement de cette approche, la théorie de la médiation technologique de Verbeek (2006) intègre également des éléments de la phénoménologie, proposant une analyse des interactions entre les humains et la technologie. Elle met en lumière comment les objets technologiques façonnent nos actions et notre perception du monde à travers « l'intentionnalité technologique ». Cette perspective montre que les objets technologiques influencent notre relation sensorielle avec la réalité et participent à la construction même de cette réalité en amplifiant ou réduisant certaines informations (Ihde 1990).

En somme, l'approche des réseaux sociotechniques permet de comprendre la co-construction de la réalité par les humains et la technologie. Elle offre une vision plus nuancée de l'influence réciproque entre la technologie et la société. Pour les concepteurs et les utilisateurs, cette double vision permet de voir à la fois la transparence de l'usage et du contexte, redéfinissant ainsi les responsabilités liées aux conséquences des outils d'IA. Par conséquent, cette approche montre que la mise en place de la collaboration humain-machine est une cocréation entre plusieurs acteurs, y compris les outils d'IA, les concepteurs et d'autres actants de l'environnement.

Approche de la sociomatérialité

L'approche de la sociomatérialité, développée par Orlikowski (2007; 2000), considère l'interdépendance inextricable entre les artefacts matériels et les aspects sociaux, comme les normes et les structures sociales, en un concept unifié qui reflète une co-construction de la réalité sociale (Leonardi 2013).

Cette perspective est devenue très influente dans les études sur les usages des technologies numériques. Par exemple, elle a été utilisée pour étudier les usages organisationnels des plateformes d'intégration d'entreprise (Wagner, Moll, et Newell 2011), des outils de médias sociaux (Scott et Orlikowski 2014), des plateformes communautaires en ligne (Barrett, Oborn, et Orlikowski 2016), et des outils robotiques (Barrett et al. 2012; Beane et Orlikowski 2015).

Sous ce prisme, Lebovitz, Lifshitz-Assaf, et Levina (2022) considèrent « l'opacité » de l'outil d'IA (la difficulté pour les utilisateurs de comprendre les résultats) non pas comme une simple caractéristique de l'outil, mais comme une entité sociomatérielle qui dépend de la situation de l'acteur (Haraway 2006).

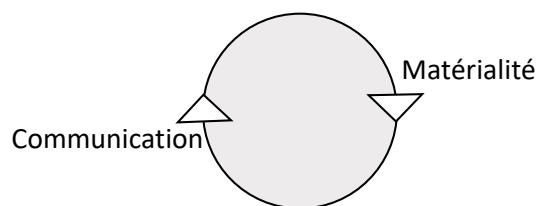
Cependant, l'approche de la sociomatérialité connaît un schisme entre deux versions : celle inspirée du réalisme agential (*agential realism*) de Barad (2003; 2007) et celle du réalisme critique (*critical realism*) proposée par Mutch (2013). Leonardi (2013) offre une analyse comparative entre ces deux versions.

Le réalisme agential propose une vision radicale qui transcende les dichotomies traditionnelles du matériel et du social, considérant ces entités comme des catégorisations arbitraires, de « coupures agentielles » (Barad 2003, 816). Au contraire, ce qu'on appelle des éléments matériels et sociaux forme un processus continu d'interactions sans frontières claires (Orlikowski 2007; 2010). Par conséquent, seules les pratiques sociomatérielles sont observables. La recherche vise donc à identifier ces pratiques dans le processus organisationnel et à démontrer comment elles façonnent ce processus (Scott et Orlikowski 2013). (Voir la partie A de la figure 5.)

En revanche, Mutch (2013) plaide pour une reconnaissance des caractéristiques persistantes des objets et des contextes sociaux distincts, soulignant leur interrelation ou leur « enchevêtrement », plutôt que leur fusion (voir la partie B de la figure 5). Selon le réalisme critique, les pratiques sont l'espace où les agences matérielles et sociales se combinent.

De même, Mutch distingue l'agency humaine, qui est la capacité des acteurs humains à atteindre des objectifs, de l'agency matérielle, qui est la capacité des objets à agir sans intervention humaine (Leonardi 2012). À l'inverse, selon Barad (2003), l'agency n'est ni sociale ni matérielle, leur distinction étant dépourvue de sens.

A. Sociomatérialité fondée sur le réalisme agentiel



Domaine de l'action (Pratiques)

B. Sociomatérialité fondée sur le réalisme critique

Domaine de la structure (matérialité)

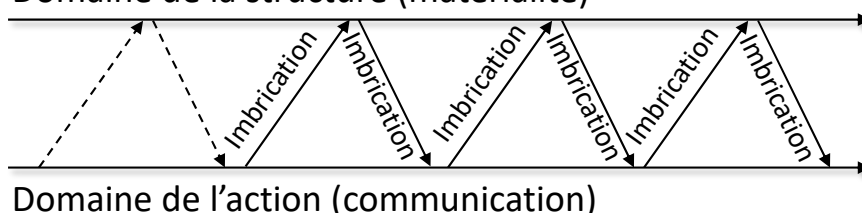


Figure 5 : Représentation des concepts de la sociomatérialité (Leonardi, 2013)

L'approche de la sociomatérialité demande d'analyser deux éléments conjoints (Dourish 2016) : la présence et le rôle de l'algorithme, ainsi que ce qui est produit par le processus collectif d'identification de ces éléments.

Introna (2016) suggère que la sociomatérialité selon le réalisme agentiel est utile pour comprendre les assemblages numériques dynamiques, mais peut être limitée lorsqu'il s'agit d'analyser en détail des éléments spécifiques d'un système.

Le réalisme critique, avec son concept d'agency, facilite l'analyse du rôle des outils d'IA en mettant en avant leur capacité à influencer les environnements organisationnels indépendamment de l'action humaine directe. Par exemple, Zhang et al. (2021) montrent que l'expérience des concepteurs utilisant des outils d'IA devient liminale, marquée par l'ambiguïté et des trajectoires de conception multiples.

De même, Murray, Rhymer, et Sirmon (2021) explorent quatre formes technologiques définies par leur agency conjointe sur les routines organisationnelles : les technologies d'assistance, de limitation, d'augmentation et d'automatisation. Les technologies d'assistance nécessitent une manipulation humaine directe, comme les machines de chirurgie cardiaque. Les technologies de limitation sélectionnent des actions sans développer de protocoles, comme certains outils d'IA. Les technologies d'augmentation développent des protocoles mais

ne sélectionnent pas des actions de manière autonome, tandis que les technologies d'automatisation ressemblent à des Intelligences Artificielles Générales, capables d'apprendre et de s'optimiser sans intervention humaine.

Par conséquent, pour établir des modes de gestion de la collaboration, l'approche de la sociomatérialité envisage d'étudier les pratiques sociomatérielles qui se produisent, éventuellement en analysant les agences qui apparaissent.

Synthèse des approches pour théoriser les usages des outils d'IA

Pour saisir les nuances entre ces approches et les différentes manières d'aborder les usages de la technologie, nous les comparons de manière systématique en utilisant des exemples illustratifs et en évoquant quelques-unes de leurs limites. Cette comparaison s'appuie sur le travail de Leonardi (2013).

Tout d'abord, nous situons l'ontologie et l'épistémologie de ces approches (tableau 1) et les positionnons selon leur rapport à la notion de réalité (figure 6), afin de comprendre comment elles répondent à la tension entre le réalisme et le constructivisme. Ensuite, nous examinons comment ces approches traitent des concepts de matérialité, de social, de sociomatérialité, ainsi que de l'agency et des pratiques (tableau 2). Nous les positionnons ensuite selon leur rapport au concept d'agency (figure 7). Enfin, nous évoquons les implications méthodologiques et les contributions possibles de ces approches.

Le tableau 1 compare l'ontologie et l'épistémologie de ces approches. Il montre que l'approche de la sociomatérialité selon le réalisme critique et celle de l'affordance adoptent une position plus proche du réalisme, en considérant qu'il existe une réalité objective (figure 6). Cependant, l'accès à cette réalité ne se fait qu'à travers des interprétations.

En revanche, les approches des réseaux sociotechniques et de la sociomatérialité selon le réalisme agentiel adoptent une position plus constructiviste, considérant que la réalité résulte des interactions entre les entités (figure 6). Elles soulignent également que le rôle des entités matérielles est tout aussi crucial dans la construction de la réalité, et que celle-ci n'est pas uniquement le fruit des interactions humaines.

Pour comprendre comment ces approches peuvent aider à théoriser les usages de la technologie, et des outils d'IA en particulier, nous comparons également des concepts essentiels tels que la matérialité, le social, la sociomatérialité, ainsi que l'agency et les pratiques (tableau 2). Bien que les trois premiers concepts soient liés aux approches sociomatérielles, ils nous servent à distinguer ces différentes approches.

	Ontologie générale	Épistémologie générale
Affordance	Les affordances émergent de l'interaction entre les acteurs et leur environnement. La réalité est perçue à travers les possibilités d'action qu'un environnement offre aux acteurs.	Approche interactionniste, mettant l'accent sur la relation directe entre la perception et l'action. La connaissance émerge de l'engagement actif avec l'environnement.
Réseaux sociotechniques	Les acteurs humains et non-humains sont au même plan, ayant tous les deux une agency. Mais, les entités existent sous la forme des réseaux, de relations entre actants.	Approche constructiviste où la connaissance est vue comme émergeant des interactions au sein des réseaux. Les chercheurs sont considérés comme des acteurs au sein de ces réseaux.
Sociomatérialité (réalisme agentiel)	Il n'y a pas d'interaction sociale séparée qui ne soit distincte de la matérialité ; il n'y a qu'un « sociomatériel » fusionné.	La connaissance est construite à travers des interactions actives avec le monde. Les chercheurs sont partie prenante de l'entrelacement social et matériel qu'ils étudient.
Sociomatérialité (réalisme critique)	Le contexte social et la matérialité sont distincts. Le social et le matériel deviennent « sociomatériels » à mesure que les entités s'imbriquent.	La connaissance est une construction humaine, mais cherche à refléter la réalité objective. Reconnaît une réalité indépendante de nos perceptions et accessible à travers la science et la réflexion.

Tableau 1 : Comparaison de l'ontologie et l'épistémologie des approches théoriques

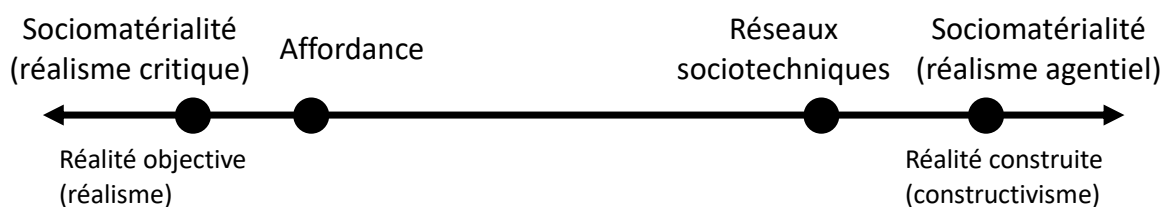


Figure 6 : Schéma situant les approches théoriques selon leur façon de concevoir la réalité

	Socio+matérialité	Agency	Pratiques
Affordance	La matérialité offre des affordances, des possibilités d'action qu'elle permet ou contraint. Elle est liée à la façon dont les individus perçoivent et interagissent avec leur environnement, suggérant une relation dynamique et réciproque entre acteurs et objets. Le social est un ensemble d'interactions entre les acteurs et leur perception des affordances. La sociomatérialité émerge de cette perception et de la mise en œuvre des affordances.	Capacité des acteurs de percevoir et d'agir sur les affordances. L'environnement matériel et social influence cette capacité.	Actions ou comportements émergeant de l'utilisation des affordances disponibles dans un environnement particulier.
Réseaux sociotechniques	La matérialité participe aux réseaux, avec une capacité à agir et à co-former les réalités avec les acteurs humains. Elle porte des inscriptions et médiatise les intentions et les actions. La sociomatérialité est le réseau d'actants humains et non-humains co-construisant la réalité.	Distribuée parmi les acteurs humains et non-humains. Les actions sont le résultat d'interactions complexes au sein des réseaux.	Ensembles d'actions et d'interactions au sein de réseaux d'actants en constante co-construction.
Sociomatérialité (réalisme agentiel)	Aucune entité matérielle ou sociale existe. Il n'y a que la sociomatérialité, l'inhérente inséparabilité entre ce qu'on appelle le social et le matériel.	Résultat de l'entrelacement sociomatériel, l'agency est distribuée et émerge des interactions entre entités sans être localisée dans une entité.	Accomplissements sociomatériels.
Sociomatérialité (réalisme critique)	La matérialité est l'arrangement d'un objet dans des formes particulières qui se retrouvent à travers le temps et les contextes. Le social a ses propriétés et son autonomie relative, tout en influençant et étant influencées par la matérialité. La sociomatérialité est la mise en œuvre d'un ensemble particulier d'activités qui fusionnent la matérialité avec le social.	Généralement attribuée aux acteurs humains dans un contexte où les structures matérielles et sociales ont des pouvoirs causaux.	Espace dans lequel les agences sociales et la matérialité s'enchevêtrent de manière constitutive à travers le processus d'imbrication.

Tableau 2 : Comparaison des notions de matérialité, de social, de sociomatérialité, d'agency et de pratiques

Ces réflexions nous amènent à considérer l'agency comme le concept le plus fondamental dans l'étude des usages des technologies. Dans ce sens, Jatou (2020) souligne les controverses actuelles sur l'agency des algorithmes (Barocas et Selbst 2016; Bechmann et Bowker 2019; Grosman et Reigeluth 2019). De même, Ziewitz (2016) considère que, face au « mythe moderne » qu'ils représentent, les algorithmes doivent être analysés pour mettre en lumière ce qui est en jeu, à quel point ils influencent l'organisation et la société, en d'autres termes, quelle est leur agency. Nous situons donc ces approches en fonction de ce concept (figure 7).

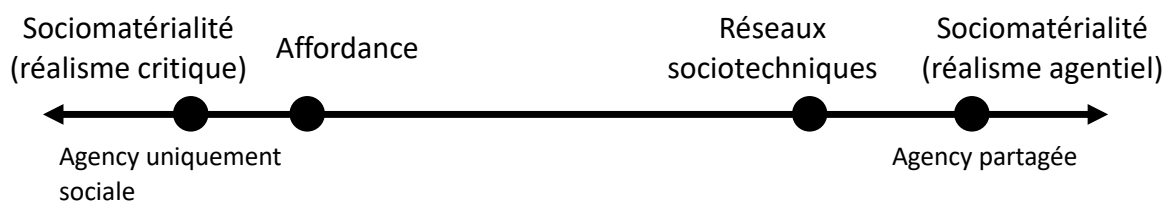


Figure 7 : Schéma situant les approches théoriques selon leur conception de l'agency

Les différences d'ontologie, d'épistémologie et de positionnement sur les concepts vus ci-dessus ont des implications sur la méthodologie à adopter et les contributions potentielles selon l'approche choisie.

L'approche de l'affordance prend comme unité d'analyse les affordances spécifiques offertes par l'environnement et la manière dont les acteurs perçoivent et interagissent avec ces affordances. Ce type de recherche vise à identifier les affordances et la relation entre leur perception et les actions résultantes. Elles peuvent fournir une perspective détaillée sur l'interaction entre les acteurs et la technologie, éclairant la conception de la technologie et les stratégies d'implémentation dans les organisations.

L'approche des réseaux sociotechniques prend comme unité d'analyse le réseau d'actants et leurs interactions, y compris comment ils se forment, se maintiennent et se transforment. Le but est d'identifier les processus d'évolution des réseaux, y compris les processus de traduction, d'inscription et de médiation. Cela peut fournir une compréhension approfondie de la dynamique des réseaux sociotechniques, mettant en évidence sa co-construction par les différents actants.

La sociomatérialité selon le réalisme agentiel observe les pratiques sociomatérielles et les processus performatifs. Elle vise à identifier les implications des pratiques sociomatérielles sur les processus organisationnels (par exemple l'identification, la négociation, etc.) afin de fournir une perspective nuancée sur la manière dont la technologie et l'organisation sont intrinsèquement liées. Elle éclaire ainsi les dynamiques de pouvoir, les résistances et les potentialités pour le changement organisationnel et social.

La sociomatérialité selon le réalisme critique prend comme unité d'analyse l'agency sociale et les influences matérielles. Elle cherche à identifier les imbrications, la manière dont le social et le matériel deviennent le sociomatériel, et quelles implications cela a pour l'organisation. Elle fournit ainsi une compréhension des mécanismes sous-jacents aux phénomènes sociaux et matériels. Cela privilégie un fondement pour critiquer et transformer les structures existantes dans les organisations et la société.

En somme, les approches théoriques évoquées sont une panoplie d'outils conceptuels pour réussir à mieux saisir un phénomène empirique (Mueller, Renken, et Van Den Heuvel 2016; Elbanna 2016). Elles sont à envisager comme des aides à la compréhension, une aide d'autant plus génératrice que le chercheur s'engage à les faire dialoguer.

Elles ont ainsi pu mettre en avant des caractéristiques saillantes des outils d'IA (dont l'opacité et les politiques cachées) et des usages développés par les acteurs (faire semblant de l'utiliser, inventer des récits, mettre en place de nouveaux rôles ou développer des métacognitions pour intégrer les résultats dans leur réflexion).

En outre, la combinaison de ces approches aide à réduire leurs limites. Par exemple, dans leur étude de cas sur la rencontre des étudiants grecs et du système d'anti-plagiat, Introna et Hayes (2011) orchestrent la polyphonie de ces prismes théoriques pour montrer la nature historiquement et culturellement située des imbrications sociomatérielles et l'émergence des différentes agences.

Ces approches offrent ainsi un point de départ pour repenser le concept de technologie dans les organisations (Faraj, Pachidi, et Sayegh 2018; von Krogh 2018; Bailey et al. 2022; Faraj et Leonardi 2022). Plus particulièrement, elles permettent d'établir des modèles de collaboration humain-machine à partir des données empiriques, notamment grâce au concept d'agency. En analysant les pouvoirs d'action de chaque acteur, humain et non humain, et leur assemblage mutuel, une autre façon de voir des formes de collaboration humain-machine émerge.

Mais comme le remarquent Anthony, Bechky, et Fayard (2023), les études empiriques sur les usages des outils d'IA soulèvent encore de nombreuses questions, que ces outils soient considérés comme des « outils », des « médiums » ou des « partenaires ». Répondre à ces questions permettrait de se rapprocher un peu plus d'un modèle de collaboration humain-machine pertinent.

La gestion de la collaboration entre les humains et les outils d'IA est devenue un défi ces dernières années pour l'organisation de l'action collective. La littérature propose soit des modèles théoriques à compléter, notamment par leur mise en place sur le terrain, soit des observations sur les usages qui se développent, mais dont les conclusions n'aboutissent pas encore à des modes de gestion d'une collaboration efficiente et responsable.

3. Développer empiriquement des modes de gestion pour une collaboration entre l'expert métier et l'outil d'IA

Nous avons présenté un ensemble interdisciplinaire de recherches qui propose des modèles conceptuels pour comprendre la collaboration avec les outils d'IA. Diverses notions ont été avancées pour théoriser la collaboration humain-machine, comme les plateformes hybrides humain-IA (Priksat et al. 2023), les systèmes hybrides humain-IA (Dellermann et al. 2019; H. Liu, Lai, et Tan 2021), l'intelligence collaborative (M. Johnson et Vera 2019; Bogert, Schechter, et Watson 2021), l'intégration de l'humain dans la boucle (Zheng et al. 2017; Yu, Beam, et Kohane 2018; Kelly et al. 2019; Zanzotto 2019), ainsi que l'augmentation de l'intelligence (Barile et al. 2019; Zhou et al. 2021).

Ces différentes perspectives explorent la manière dont les outils d'IA et les humains peuvent collaborer, que ce soit dans une symbiose visant à améliorer la productivité et les tâches accomplies, ou dans une démarche d'augmentation des capacités des acteurs humains pour une meilleure réalisation de leurs missions.

Cependant, bien que ces concepts soient variés, ils demeurent souvent trop abstraits, négligeant les spécificités contextuelles et offrant une compréhension limitée des stratégies de gestion requises pour ces collaborations.

En complément, nous avons également examiné un courant de littérature axé sur les usages des outils d'IA. À travers ses diverses approches théoriques, cette littérature met en lumière les caractéristiques distinctives des outils d'IA ainsi que les réactions des utilisateurs. Bien que ces travaux fournissent des perspectives importantes, ils s'aventurent peu dans l'exploration des modes de gestion de la collaboration entre humains et outils d'IA, et ne proposent aucun modèle de collaboration explicite.

Pourtant, les prismes théoriques sous-jacents offrent des outils conceptuels dans cette direction, notamment en ce qui concerne l'agency, c'est-à-dire la capacité d'action des acteurs, qu'ils soient humains ou non humains.

Pour mieux comprendre ce manque, Anthony, Bechky, et Fayard (2023) observent que la majorité de ces travaux considèrent les outils d'IA soit comme des « outils », soit comme des « médiums ». Elles proposent en complément de les approcher comme s'il s'agissait de partenaires à part entière (« *couterpart* »).

Sous cet angle, l'intérêt se pose sur le couplage entre processus de développement et usages. Par exemple, comment la culture et la structure d'un organisme façonnent-elles la conception des outils d'IA ? Comment cela se reflète-t-il dans la manière dont les utilisateurs collaborent avec l'outil d'IA ? Ces questions demandent de démêler les problèmes liés au pouvoir et à la valeur symbolique entourant l'outil. De même, quels récits autour de l'outil d'IA

sont développés et/ou contestés, et par qui ? Qui a développé le récit autour de l'usage de l'outil d'IA, notamment celui concernant la collaboration ?

D'après Anthony, Bechky, et Fayard (2023), il est crucial d'adopter une approche qui intègre cette vision afin de parvenir à une compréhension exhaustive des dynamiques de la collaboration humain-machine. Pour aller plus loin, nous pouvons également souligner le manque « d'organisation » qui apparaît dans la littérature présentée. Le niveau d'analyse reste centré sur la dyade entre l'utilisateur et la machine, ce qui constitue déjà une forme d'action collective, mais qui pourrait s'étendre aux autres acteurs qui influencent, ou devraient participer à cette collaboration.

La démarche initiée par les études sur les usages des outils d'IA présente donc un intérêt considérable, pourvu qu'elle soit poussée jusqu'à l'identification des modes de gestion de cette collaboration, qui se traduisent par la définition des rôles et responsabilités, la coordination des tâches, la communication et l'interaction entre acteurs ainsi que le contrôle et l'amélioration des processus collaboratifs.

Afin de poursuivre ces efforts, nous développons une démarche empirique, s'inspirant des études sur les outils d'IA et de leurs perspectives théoriques. De plus, grâce à l'ensemble de la revue de littérature, nous considérons le système de soins comme un contexte particulièrement pertinent pour illustrer un type de collaboration humain-IA car caractérisé par des niveaux élevés d'expertise, des jugements critiques et une grande incertitude.

Dans le système de soins, la responsabilité finale est attribuée à l'expert métier et l'outil d'IA à un rôle d'assistant ou de soutien aux professionnels de la santé. En effet, alors que la société cherche à rendre le travail professionnel plus transparent et responsable (Kellogg 2009; Huising et Silbey 2018), le système de soins met en lumière le rôle crucial de l'expert métier face à des décisions critiques et à une grande incertitude.

De plus, la responsabilité de l'expert métier s'étend à la gestion de l'opacité des résultats de l'outil d'IA, ce qui oriente les possibilités de communication et d'interaction : les utilisateurs doivent développer des métacognitions (Jussupow et al., 2021), comme les pratiques interrogatives (Lebovitz, Lifshitz-Assaf, et Levina 2022). En effet, les professionnels de santé doivent prendre des décisions critiques, mais également évaluer les résultats des outils d'IA, car ils demeurent responsables du diagnostic et des recommandations faites (Kluttz et Mulligan 2019; Anthony 2021; Lebovitz, Lifshitz-Assaf, et Levina 2022).

Notre objectif est alors de comprendre comment la collaboration entre l'humain et la machine peut être gérée de manière efficiente et responsable afin d'améliorer les processus de collaboration, par exemple avec l'augmentation des capacités à prendre des décisions. Cet objectif inclut la recherche de l'organisation sous-jacente à cette collaboration, faisant ressortir

si possible les acteurs participant (ou devant participer) à la mise en place de cette action collective.

Ainsi, notre question de recherche peut être formulée comme suit : **quels sont les modes de gestion appropriés pour une collaboration entre des experts métier et les outils d'IA ?**

Notre thèse revêt ainsi une double dimension, à la fois descriptive et normative, s'inscrivant dans l'objectif des sciences de gestion de produire des connaissances pour une gestion efficaces et responsables de l'action collective. Le prochain chapitre détaille notre approche pour répondre à cette question.



Chapitre 3 : Démarche de notre recherche

Dans ce chapitre, nous explorons les implications méthodologiques de l'étude des outils d'IA. D'abord, nous présentons le choix de l'objet de recherche et des principes méthodologiques qui en découlent (1). Ensuite, nous illustrons leur application à travers la construction de nos deux études de cas (2).

1. Choix de l'objet de recherche et des principes méthodologiques

Nous présentons d'abord le choix de l'objet de recherche (1.1) avant d'aborder la question de ses implications méthodologiques (1.2).

1.1. Choix et présentation de l'objet de recherche

Nous expliquons comment notre position de chercheur (1.1.1) et la position du laboratoire (1.1.2) ont motivé le sujet de recherche et la façon de l'aborder.

1.1.1. Position du chercheur

Au cours de la décennie 2010, l'Intelligence Artificielle a suscité un intérêt croissant dans les communautés de data scientists*, puis au sein de l'ensemble de l'industrie et, enfin, de la société en général (chapitre I, partie 1). C'est ainsi qu'en septembre 2017, le gouvernement français a commandé un rapport sur l'importance de l'IA à Villani, député et mathématicien. Paru mars 2018, ce rapport a souligné le rôle stratégique de ces technologies, en particulier dans le secteur de la santé.

Notre position en tant que chercheur est influencée par nos études d'ingénieur, en partie liées à l'évolution de la recherche en IA. En 2016, alors que la data science connaissait une croissance vertigineuse, nous avons entamé nos études d'ingénieur généraliste. Parmi les fondements de notre formation, nous avons acquis des connaissances en informatique et en IA, ce qui nous a conduit à développer des projets utilisant ces technologies. Cependant, nous avons été frustré par le manque d'aboutissement de ces projets, qui s'arrêtaient au mieux à la validation technologique. Alors que ces outils d'IA semblaient promettre une révolution, nous nous sommes demandé quelle était leur réelle portée. Ainsi, en septembre 2018, nous avons choisi de nous spécialiser dans la gestion de l'innovation, avec l'espoir de trouver des réponses à nos interrogations.

Notre intuition était la suivante : à mesure que les avancées techniques se poursuivraient, il deviendrait essentiel d'examiner en détail les résultats pratiques des outils d'IA. Par conséquent, en 2019, nous avons poursuivi nos études avec un master de recherche en gestion de l'innovation en alternance au sein d'un cabinet de conseil spécialisé en data science. Ce

master nous a offert l'opportunité de développer une approche qualitative, jugée plus adéquate pour aborder ces phénomènes complexes et émergents.

Cependant, le travail des consultants que nous avons observé se limitait à l'élaboration de conseils établis sur les résultats des outils d'IA qu'ils avaient développés, sans véritablement étudier leurs conséquences au sein des organisations. Si nous n'avons pas trouvé de réponse satisfaisante à notre question, notre intuition s'est renforcée en interagissant avec des concepteurs d'outils d'IA et en examinant la littérature existante (Marchand et Peppard 2013).

Il nous est apparu que ces technologies exigeaient une approche différente, une nouvelle manière de les appréhender. Plus précisément, après avoir constaté autant de preuves de concept sans lendemain, nous avons cherché à comprendre ce qui se passe du côté des utilisateurs pour espérer dévoiler une partie du mystère que représente l'IA.

Nous avons entrepris une thèse en sciences de gestion pour répondre à notre question. Notre rencontre avec notre directeur de thèse, Etienne Minvielle, nous a conduit à traiter la question suivante : « *quelles sont les conséquences des données massives* sur la gestion de la personnalisation à grande échelle du parcours patient ?* » Notre profil était adapté par notre familiarité avec les outils d'IA, bien que le système de soins nous fût étranger.

Cette thèse a également été l'occasion d'approfondir notre expertise en recherche qualitative, ou devrions-nous dire compréhensive (Dumez 2021). Elle s'est renforcée en partie grâce à un dépassement de la rationalité cartésienne, en adoptant une épistémologie plus empirique, et en explorant des courants philosophiques comme le pragmatisme.

Ainsi, le choix de notre sujet de recherche a découlé en partie de notre attrait et de notre familiarité avec l'objet empirique. Ce choix, et notre position, qui a évolué tout au long de la thèse, ont également été influencés par le laboratoire de recherche et son approche de la démarche scientifique.

1.1.2. *Position du laboratoire*

Au fil de son évolution, le Centre de Recherche en Gestion (CRG) a forgé son identité en se distinguant par une approche scientifique étroitement liée au terrain, axée sur des questions interdisciplinaires autour de l'innovation, ce qui a permis de produire des recherches pertinentes à la fois au niveau empirique et théorique. Fondé en 1972 à l'École polytechnique par Bertrand Collomb, le CRG s'est démarqué en se concentrant sur l'étude des pratiques réelles, participant activement avec les organisations, par exemple, à travers la recherche-action ou la recherche-intervention.

Initialement axés sur l'étude des outils de gestion, les travaux pionniers du CRG, notamment synthétisés dans le rapport *Une technologie invisible* de Berry (1983), ont posé les bases d'une approche empirique et orientée vers les préoccupations des acteurs étudiés. Cette approche

empirique est également illustrée par *L'auto qui n'existait pas* de Midler (1993). L'étude de cas décrit avec minutie le terrain et offre une analyse approfondie des méthodes de gestion de projets au sein des entreprises.

Cette démarche empirique s'articule avec la théorie pour produire des contributions qui redéfinissent des concepts existants et en créent de nouveaux face aux phénomènes émergents étudiés (Dumez 2011; 2013; 2021; 2023). Si nous construisons une étude de cas, la question de sa caractérisation se pose dès le début de la recherche. Il faut alors chercher des théories et des concepts qui orientent les premières phases (Whyte 1984).

Cependant, en partant d'une théorie pour choisir le cas, celle-ci ne doit pas structurer l'étude du cas, sous peine de produire une circularité épistémique³² (Ayache et Dumez 2011; Dumez 2012; 2013; 2021). Pour enrichir les discussions académiques, la caractérisation du cas d'étude permet de dévoiler sa nature théorique, insoupçonnée au début de la recherche. C'est ainsi que les travaux du CRG ont pu générer des contributions théoriques qui redéfinissent des concepts existants et en créent de nouveaux face aux phénomènes émergents étudiés.

Cette approche se caractérise par ce que Girin (1989; 1990) a nommé l'opportunisme méthodique. Elle reconnaît que la réalité de la recherche sur le terrain peut être complexe et imprévisible. Elle adopte ainsi une flexibilité méthodologique qui ajuste les méthodes de collecte et d'analyse des données en fonction des situations rencontrées sur le terrain, et l'évolution des questions de recherche au fur et à mesure que de nouvelles informations émergent.

Notre recherche a ainsi commencé avec la gestion de la personnalisation des soins avec le concept de gestion de la singularité à grande échelle (Minvielle 2018) comme *théorie orientante*. En cherchant à répondre aux questions des acteurs sur le terrain, nous nous sommes centrés sur l'usage des outils d'IA. Les échanges fréquents avec notre directeur de thèse, les présentations au sein du laboratoire, les interactions avec d'autres chercheurs lors de conférences et d'échanges, notamment lors de notre séjour à la Wharton School ont tous contribué à enrichir notre dialogue entre le terrain et la théorie. Nous avons caractérisé nos

³² Un problème méthodologique et théorique où les conclusions d'une recherche sont influencées ou déterminées par les présupposés ou les hypothèses initiales. Le raisonnement ou les preuves utilisées pour soutenir une conclusion sont en réalité dépendants de cette même conclusion. Cela crée une boucle de rétroaction où les éléments de preuve sont sélectionnés ou interprétés de manière à confirmer les hypothèses de départ, plutôt que de les tester de manière indépendante et critique. Par exemple, si l'on cherche dans un dictionnaire la définition du mot « A », qui se réfère au mot « B », et que la définition de « B » renvoie à « A », nous sommes dans une boucle circulaire sans fin où la définition de l'un dépend de l'autre.

deux études de cas comme des usages type des outils d'IA dans le contexte du système de soins, dont les caractéristiques mettent en avant la relation de ces outils avec les experts métier.

Le choix du sujet de la thèse a donc été le résultat d'une synergie entre notre position en tant que chercheur et la démarche du CRG, travaillant en tandem avec le terrain pour répondre à des impératifs empiriques et théoriques. Notre objet d'étude empirique sur les usages des outils d'IA dans le contexte du système de soins, abordé à travers deux études de cas, répond à un vide théorique sur la collaboration humain-machine : la théorisation de l'augmentation des experts métier par les outils d'IA. Ce sujet pose également des défis méthodologiques, que nous explorons pour démontrer la robustesse de notre démarche et décrire son application concrète.

1.2. Le défi méthodologique de la thèse

Le choix du sujet, et son angle d'approche, excluent des méthodes quantitatives et des modélisations conceptuelles. Nous avons choisi une démarche compréhensive (1.2.1) et plus particulièrement le travail par études de cas (1.2.2) pour explorer ce nouveau phénomène. Nous avons mis en place des stratégies de triangulation afin de pallier les limites de cette approche (1.2.3).

1.2.1. *Le besoin d'une démarche compréhensive pour étudier les usages des outils d'IA*

Notre recherche nous a amené à défendre l'idée que les approches quantitatives sont limitées pour appréhender la complexité et les nuances des usages des outils d'IA, que ce soit sur le plan empirique ou théorique. Lors de notre RADO, l'étude de cas portant sur les usages des outils d'IA dans le diagnostic mammographique (chapitre 4), nous avons pris connaissance de plusieurs études quantitatives qui concluent à une amélioration de la détection du cancer, parfois associée à des gains de temps (Rodríguez-Ruiz et al. 2019; van Winkel et al. 2021), ainsi qu'une réduction de la charge de travail (Raya-Povedano et al. 2021).

Or, ces analyses se fondent sur un nombre relativement restreint d'observations, ce qui rend les résultats difficilement généralisables étant donné la grande diversité des situations dans le contexte du système de soins. Plus fondamentalement, ces études quantitatives négligent la manière dont les radiologues d'approprient ces outils d'IA et comment ces derniers modifient leurs capacités. Ces études portent essentiellement sur des considérations de coût, de temps et de qualité, sans dévoiler les mécanismes sous-jacents.

Cette limitation des approches quantitatives s'inscrit dans un courant de littérature consacré à l'évaluation des technologies numériques dans le système de soins (B. Kaplan 1997; 2001; B. Kaplan et Shaw 2004). Il en ressort que l'essai randomisé contrôlé, pourtant

actuellement considéré comme la méthodologie la plus robuste pour démontrer l'utilité d'un médicament, peut apporter des évidences de l'usage d'une technologie numérique en matière de qualité des soins, d'expérience du patient et de contrôle des coûts, mais ne permettent pas l'analyse des transformations organisationnelles induites. Les approches qualitatives permettent de mettre en lumière les enjeux et les conséquences de l'introduction d'une technologie numérique (S. R. Barley 1986; B. Kaplan et Shaw 2004).

Pour comprendre les usages des outils d'IA avec des approches qualitatives, la littérature propose diverses stratégies méthodologiques (Kitchin 2017; Seaver 2017; Lange, Lenglet, et Seyfert 2019; Christin 2020; Jonker-Hoffrén 2021). Une perspective particulièrement complète sur le sujet est présentée par Kitchin (2017), qui distingue six méthodes qualitatives distinctes pouvant être appliquées à l'étude critique des algorithmes.

En premier lieu, l'analyse du code, ou du pseudo-code, se présente comme le moyen le plus direct pour comprendre le fonctionnement de l'algorithme (Kitchin 2017). À ce titre, Alcaras et Larribeau (2022) ont introduit un programme de recherche en sciences sociales sur le code, qualifié également de « texte de l'algorithme ». Ils soulignent que « l'écriture informatique » constitue l'un des points de départ pour éclairer le fonctionnement interne des infrastructures numériques, ainsi que les croyances et pratiques associées à ces environnements.

Cette approche vise à comprendre les processus d'inscription et, de manière plus générale, le contexte social qui donne naissance aux algorithmes. Kitchin (2017) évoque la possibilité de déconstruire le code pour en saisir la logique, ainsi que l'analyse de l'évolution des versions de l'algorithme, dont la généalogie illustre comment un problème est abordé au fil du temps. De plus, l'analyse de la traduction d'une même fonction algorithmique en différents langages informatiques permet de mieux comprendre sa matérialité. Cependant, l'accès au code (ou à ses différentes versions) est rarement possible, l'analyse nécessite un niveau d'expertise qui dépasse souvent les chercheurs. Enfin, l'analyse directe du code est décontextualisée de la manière dont il est écrit et dont s'insère dans son contexte social, limitant ainsi la compréhension des effets réels du code.

Une alternative à l'analyse *ex post* du code est le travail réflexif pendant sa production (Kitchin 2017). Cette méthode demande au chercheur d'écrire lui-même l'algorithme et de faire de cette pratique une auto-ethnographie. Le chercheur s'interroge alors sur ses propres expériences et sur la manière dont ses pratiques font partie intégrante des facteurs sociaux, culturels, réglementaires, etc. Cependant, outre les difficultés de mise en œuvre, cette méthode néglige les pratiques inconscientes et l'influence de la réalisation d'une auto-ethnographie sur l'écriture du code.

En troisième lieu, la rétro-ingénierie du code s'avère utile lorsque le code d'un algorithme est gardé secret par le concepteur (Kitchin 2017), ou que l'outil d'IA est opaque, notamment

lorsqu'il est entraîné par apprentissage profond*. Si l'algorithme est une boîte noire, tester ses réactions en fonction de différentes simulations donne un aperçu de son fonctionnement. Pour rendre cette méthode plus crédible, il suggère de systématiser les tests à l'aide de robots capables d'imiter des utilisateurs potentiels dans une large variété de situations. La simulation de conditions réelles d'utilisation suggère une analyse sur l'inscription de l'algorithme dans son contexte d'usage. Cependant, si cette méthode aide les chercheurs à reconstruire le code de l'algorithme, il en va de même pour les concurrents de l'entreprise qui l'a produit. C'est pourquoi, les concepteurs cherchent à empêcher que des robots utilisent leurs algorithmes. La rétro-ingénierie est donc possible dans un nombre limité de situations et ne reflète pas la complexité de ses usages.

Une quatrième méthode consiste à mener des entretiens avec les concepteurs ou une ethnographie des équipes de programmation (Kitchin 2017). L'objectif consiste à comprendre les raisons et les modalités de mise en place d'un algorithme. Par exemple, l'ouvrage de Florian Jaton (2020) offre une vision détaillée des pratiques au sein de cette communauté, mettant en lumière ses contingences et son aspect relationnel, à la manière de *La Vie de Laboratoire* (Latour et Woolgar 1979). Cependant, ces données ne suffisent pas à rendre compte de l'adoption de l'algorithme, de ses usages, et de ses conséquences sur l'organisation et la société.

Pour surmonter ces limites, la cinquième méthode intègre le système socio-technique dans lequel l'algorithme opère (Kitchin 2017). Par exemple, cela implique la collecte de données liées à la promotion de l'algorithme par l'entreprise (documentation interne, discours et pratiques promus, brochures et modes d'emploi destinés aux utilisateurs, présentations lors de conférences professionnelles, etc.). L'analyse discursive met en lumière la manière dont les algorithmes sont imaginés et racontés, ainsi que la compréhension qu'en ont les concepteurs. Face au volume et à la diversité des données collectées par cette méthode, il est recommandé de constituer une équipe de recherche (Kitchin 2017; Jonker-Hoffrén 2021). Le travail en équipe offre également l'opportunité de traiter le matériau sous différents angles disciplinaires, ce qui est crucial pour caractériser l'étendue des ramifications des usages de ces technologies (Jonker-Hoffrén 2021).

Enfin, l'ethnographie complète ces méthodes pour comprendre les usages des outils d'IA (Kitchin 2017; Seaver 2017; Christin 2020). Elle est étroitement liée à l'opacité des outils d'IA, constituant un défi pour les experts métier impliqués dans la collaboration humain-machine et pour les chercheurs étudiant ces usages (Lange, Lenglet, et Seyfert 2019). Elle consiste à observer les acteurs en contact avec ces outils afin de comprendre comment ils les adoptent et s'adaptent en apprenant à les utiliser, parfois en les détournant de leurs objectifs initiaux. Plusieurs options s'offrent alors aux chercheurs ethnographes. Seaver (2017) détaille une tactique ethnographique appelée « fouille » (*scavenge*). Elle consiste à collecter des données

à travers divers lieux (Gusterson 2004). Cette technique ethnographique se rapproche des conditions de notre vie quotidienne qui font émerger la connaissance : la convergence de diverses informations. Ainsi, même si elle peut être critiquée en raison de l'hétérogénéité de ses formes, elle ouvre des perspectives sur de nombreuses interactions entre les outils d'IA et les êtres humains, sur l'émergence locale des représentations, et sur la manière dont la situation acquiert un sens dans divers contextes sociaux et culturels.

Christin (2020) met en avant trois stratégies ethnographiques élaborées à partir des réseaux sociotechniques (Callon 1986). En premier lieu, la « réfraction algorithmique » considère l'algorithme comme un prisme réfractant la lumière. Les outils d'IA reflètent et reconfigurent la structure sociale en mettant en lumière les réorganisations entre les individus et les institutions grâce à leur introduction et leur utilisation, s'inscrivant dans la lignée des approches constructivistes (S. R. Barley 1986; DeSanctis et Poole 1994; Orlikowski 2000). La deuxième stratégie est la « comparaison algorithmique ». En mettant en évidence les similitudes et les différenciations entre les études de cas (ou l'ethnographie multisite), cette approche révèle, par des jeux de contrastes, les caractéristiques uniques des usages des outils d'IA malgré l'opacité de leur fonctionnement interne. Cette stratégie s'aligne sur la notion de matérialité, examinant les formes et les propriétés d'un produit qui perdurent à travers différents contextes spatio-temporels. Enfin, les outils d'IA servent aussi à réaliser des « triangulations algorithmiques » : ce sont des outils permettant au chercheur de collecter davantage de données.

Il en ressort que l'approche quantitative est une méthodologie adaptée à l'étude des usages des outils d'IA. Nous avons choisi d'adopter une approche ethnographique, nous inspirant de la fouille (Seaver 2017), pour collecter une diversité de données sur les usages des outils d'IA. Nous avons structuré ce matériau en deux études de cas afin d'aboutir à une théorisation pertinente.

1.2.2. Choisir les études de cas pour théoriser un phénomène

L'étude de cas constitue un élément central de la recherche en sciences sociales, suscitant de nombreux efforts visant à la caractériser. Il s'agit de définir ce qu'est un cas, quels sont les critères de sélection appropriés, et quelles sont les caractéristiques d'un cas de qualité (H. S. Becker et Ragin 1992; Herreid 1998; Abbott 2004; Gerring 2004; Yin 2011). Pour naviguer au sein de cette abondante littérature, nous nous appuyons sur les travaux d'Hervé Dumez (Dumez 2011; 2013; 2021; 2023).

Saisir l'essence de ce qu'est une étude de cas est complexe dans la mesure où tout phénomène suscitant un intérêt scientifique peut être considéré comme un cas d'étude. La définition classique d'un cas repose sur l'identification de caractéristiques spécifiques et remarquables, c'est-à-dire dignes d'intérêt (Yin 2018; 2011). En fin de compte, cette définition

permet de considérer toute chose comme un cas, pourvu que l'on puisse mettre en évidence ce qui est intéressant dans un contexte donné. Appliquée à la recherche, la question clé est de savoir comment conférer une valeur scientifique à un cas, comment le rendre pertinent du point de vue académique (Dumez 2023). Trois sous-questions méthodologiques émergent alors. La première concerne la collecte de données pour assurer la validité scientifique du cas. La deuxième se penche sur le traitement de ces données. Enfin, la troisième vise à comprendre comment ces données contribuent à alimenter une discussion théorique.

Alors que les approches quantitatives privilégient généralement la quantité de données pour démontrer la force des corrélations, les méthodes qualitatives visent à explorer un phénomène inconnu selon diverses dimensions. Dans cette perspective, semblable à la tactique de la fouille ethnographique (Gusterson 2004; Seaver 2017), Blumer évoque une première phase d'exploration (Blumer 1986; Dumez et Toussaint 2022), c'est-à-dire une démarche adaptable au cours de laquelle le chercheur navigue entre différentes sources de données, élargit les points d'observation et affine sa compréhension des données pertinentes à mesure de l'avancement de son étude de cas. À ce stade, la littérature guide la recherche (Whyte 1984) en tant que ressource, sans toutefois déterminer l'expérience et son interprétation. La diversification des angles d'approche du phénomène étudié enrichit l'étude de cas (Dumez 2023).

Le matériau hétéroclite est traité par deux approches similaires : la description et la narration. La description peut être réalisée selon diverses techniques, produisant des résultats variés et réducteurs pour un même phénomène (Dumez 2021). Son utilité varie, servant à illustration d'une théorie – une démarche problématique en raison du risque de circularité –, mais également à faire voir un phénomène et à l'analyser. Elle peut également offrir une nouvelle perspective : « *le but d'une description ethnographique consiste à présenter les phénomènes de manière nouvelle et révélatrice* » (Hammersley 1990, 599).

Pour ce faire, la description se construit par des itérations de « décrire comme », en débutant par une description minimale qui s'enrichit progressivement, dévoilant ainsi la multiplicité des dimensions et la construction d'autres points de vue (Dumez 2021). Le premier « décrire » est crucial du point de vue théorique : il doit être formulé sous la forme d'une exclusion de l'élément théorique à discuter, qui sera réintroduit par la suite (Dumez 2021). Les ajouts ultérieurs peuvent également inclure des éléments iconographiques, à condition qu'ils ne soient pas des « images pléonastiques » qui n'apportent rien (Dagognet 1973, 109). Au contraire, l'image ou le schéma doit constituer une « codification profondément modificatrice et éclairante » (Dagognet 1973, 47), par exemple en offrant une vue synoptique qui rapproche des éléments et facilite la comparaison.

La narration, bien qu'étant semblable à la description en tant qu'outil liant des données empiriques de natures diverses, s'applique davantage à des processus se déroulant dans le

temps pour expliquer une dynamique (Abell 2004). Elle commence par établir des chronologies multiples, un travail qui doit être entrepris dès le début de la collecte de données afin de détecter les lacunes nécessitant des réponses (Dumez 2021).

Le matériau est ensuite organisé à travers une analyse séquentielle, mettant notamment en lumière les points de basculement qui finissent et ouvrent une nouvelle séquence. La narration bénéficie de l'application du principe de symétrie (Bloor 1976). Il s'agit d'expliquer les événements par des causes similaires, de considérer la possibilité d'équifinalité (plusieurs mécanismes différents peuvent expliquer une série d'événements) et d'intégrer ainsi la diversité des données pour enrichir le récit (Dumez 2021).

En résumé, les deux techniques, description et narration, sont des outils pour traiter un matériau hétérogène et produire de nouvelles connaissances.

Enfin, l'étude de cas offre des contributions théoriques variées. Appliquée aux outils d'IA, Kitchin (2017) identifie trois types de contributions possibles : une étude détaillée d'un outil d'IA (ou d'une catégorie d'outils d'IA) pour examiner leur nature (Mackenzie 2007; Bucher 2012; Geiger 2014) ; une analyse approfondie des usages spécifiques des outils d'IA dans un domaine particulier, comme la finance (Lenglet 2011; MacKenzie 2014; Pasquale 2015; Arnoldi 2016) ; ou une évaluation critique plus générale des outils d'IA, de leur nature et de leur fonctionnement (Introna 2016).

D'une manière générale, l'étude de cas présente un intérêt limité pour la vérification d'une théorie par rapport aux approches qualitatives. De même, son utilisation pour réfuter ou invalider une théorie (Koenig 2009) semble peu pertinente. Cependant, elle évalue le pouvoir explicatif d'une théorie (Dumez 2023), que ce soit à travers des histoires hypothétiques (R. I. Sutton et Staw 1995) ou l'identification de mécanismes sociaux (Hedström et Swedberg 1998; Hedström et Bearman 2009). Elle permet également la construction de typologies exploratoires et théoriques, allant au-delà d'une simple description (Dumez 2013). Enfin, elle peut contribuer à redéfinir un concept en clarifiant son champ d'application et ses contextes d'utilisation (Dumez 2011).

En résumé, l'étude de cas découle d'un processus au cours duquel la littérature guide le chercheur dans l'exploration d'un phénomène qu'il ne comprend pas entièrement au départ. La collecte d'un matériau diversifié facilite la multiplication des perspectives lors du traitement, que ce soit par des techniques de description ou de narration. Tout au long de ce processus, le chercheur effectue des comparaisons systématiques ou « *constant comparative method* » (Glaser, Strauss, et Strutzel 1968), pour caractériser son étude de cas. Ce travail révèle l'essence même de l'étude de cas en dévoilant ce qui la rend scientifiquement intéressante. Il génère ainsi des contributions théoriques qui enrichissent les débats académiques. Par conséquent, la qualité de la collecte et le traitement des données

conditionnent le potentiel de l'étude de cas. Nous avons donc mis en place des tactiques de triangulation pour réaliser ces étapes au mieux.

1.2.3. *Les tactiques de triangulation*

La triangulation, élaborée par Denzin (2009) dans le domaine de la recherche en sciences sociales, englobe un ensemble de tactiques méthodologiques visant à renforcer la validité et la fiabilité des résultats en diversifiant les sources, méthodes ou perspectives pour examiner un phénomène donné. Inspirée par la navigation maritime, où la triangulation implique l'utilisation de plusieurs points de repère pour déterminer la position exacte, cette approche a été introduite dans les années 1970. L'objectif central de la triangulation est de minimiser les biais et d'obtenir une compréhension plus approfondie en combinant différentes approches indépendantes les unes des autres et complémentaires. Dans le cadre de notre recherche, nous mettons en œuvre diverses triangulations : théorique, empirique et méthodologique. Bien que la triangulation des enquêteurs n'ait pas été appliquée, étant donné que nous avons été les seuls à réaliser la collecte de données, nous avons sollicité l'avis de notre directeur de thèse et d'autres chercheurs pour discuter du traitement de celles-ci.

Nous adoptons une forme de triangulation théorique pour enrichir la théorisation des usages des outils d'IA. Cette tactique de triangulation est rarement mise en place (Denzin 2009, 303). La plupart des recherches sont déductives, et elles ne soulèvent qu'un nombre restreint d'hypothèses tout au long de l'étude. Pourtant, elle peut se révéler utile lorsque de nombreuses théories concurrentes existent pour expliquer un phénomène : divers points de vue théoriques pourraient être placés côte à côte pour évaluer leur utilité et leur puissance (Westie 1957). De même, elle est adaptée aux phénomènes émergents en manque de théorisation, par exemple, dans le cadre des usages des outils d'IA dont plusieurs appels ont été faits dans ce sens (Faraj, Pachidi, et Sayegh 2018; von Krogh 2018; Bailey et al. 2022).

Nous avons suivi l'approche du CRG, en réalisant de nombreux aller-retours entre le terrain et la théorie (1.1.2). Ainsi, notre matériau empirique a donné lieu à deux études de cas. Les deux répondent à une question théorique sur l'augmentation de l'expert métier par les outils d'IA. Tout en restant dans les approches de l'hybridation technologique, chaque étude de cas est l'occasion d'explorer un nouveau pan de la littérature pour expliquer le phénomène observé.

Notre première étude de cas, RADO présentée au chapitre 4, porte sur les usages des outils d'IA pour l'aide au diagnostic par des radiologues et fait émerger le cadre d'analyse conformation/exploration (Moisdon 2005a).

Notre seconde étude de cas, sur le projet KOVAK présentée au chapitre 5, porte sur les usages des outils d'IA dans la recherche médicale et fait émerger la vision peircienne sur les inférences (Peirce 1877; 1878a; 1878b; 1878c; 1878d; 1878e; 1887).

Dans notre contribution, ces théories discutent ensemble pour apporter un cadre sur l'augmentation dans la collaboration humain-machine (chapitre 6). Par conséquent, notre tactique de triangulation théorique s'applique au dialogue entre nos deux études de cas pour rendre plus robustes les hypothèses que nous proposons. Elle a été rendue possible grâce aux triangulation empirique et méthodologique.

La triangulation empirique a été mobilisée au sein de chaque étude de cas. Indépendamment des méthodes de collecte, la triangulation des données s'intéresse aux sources, c'est-à-dire aux contextes empiriques. Ceux-ci sont caractérisés par trois dimensions liées : la temporalité, le lieu et les acteurs. Trois niveaux peuvent être distingués pour les acteurs : l'individu est sélectionné, sans pouvoir établir des liens sociaux entre les personnes observées, l'interaction dans laquelle l'unité devient des personnes en interaction, et la collectivité avec comme unité l'organisation ou le groupe.

Pour RADO, l'étude de cas sur les usages des outils d'IA dans les soins courants, nous distinguons quatre grands types de sources. Les deux premiers sont nos principales sources de données. Une première partie de l'observation a eu lieu dans une clinique privée en France où l'outil d'IA commence à être utilisé par les radiologues. Nous y avons mené trois entretiens et nous avons suivi le quotidien d'un médecin représentatif de ce groupe pendant 10 jours (40 heures d'observation).

Une seconde partie du travail s'est déroulé au sein d'un centre de lutte contre le cancer de référence en France, auprès de radiologues déjà familiarisé à l'outil d'IA. Nous y avons suivi une radiologue représentative de ce groupe à chaque moment de la journée pendant 7 jours (28 heures) et nous avons réalisé deux entretiens en complément. Afin de confirmer les données issues de l'étude de ces deux premiers terrains, nous avons mené une première série de cinq entretiens avec des radiologues français qui étudient l'outil d'IA puis une autre série de quatre entretiens avec des radiologues aux États-Unis ayant des degrés divers d'expérience avec les outils d'IA. En complément, nous avons mené des entretiens semi-structurés avec 17 concepteurs d'outils d'IA en radiologie.

Pour l'étude de cas sur les usages des outils d'IA dans la recherche médicale, nous avons initialement sélectionné trois projets, mais nous avons choisi de ne traiter que les données que du dernier projet, KOVAK, en raison de limites méthodologiques présentées ci-dessous. Chaque projet implique des lieux et avec des acteurs semblables, c'est-à-dire des équipes de recherche *ad hoc* rassemblant des experts médicaux et des data scientists. Ces acteurs nous paraissent assez représentatifs du groupe professionnel auquel nous les rattachons, et dans la collecte nous avons étudié leurs interactions. Les projets se sont déroulés entre 3 et 18 mois, pendant lesquels nous avons effectué des collectes de données quotidiennes, hebdomadaires et mensuelles. De même, les supports sont assez variés, avec des observations, des entretiens

et des documents complémentaires comme les différentes versions du code qui a servi à développer l'outil d'IA.

La triangulation méthodologique peut prendre de nombreuses formes, mais sa caractéristique fondamentale est la combinaison de deux ou plusieurs stratégies de recherche différentes dans l'étude des mêmes unités empiriques. Denzin (2009) note deux formes de triangulation méthodologique. La première est la triangulation intra-méthode : au sein d'une méthode de collecte de données, plusieurs moyens sont employés pour examiner ses données. Dans nos études de cas, nous avons effectué plusieurs itérations pour traiter le matériau, principalement avec la technique de description ou de narration, puis un codage théorique. Il ne s'agit pas de générer des variétés distinctes de données triangulées, mais d'explorer au mieux le matériau en fonction d'un processus qui fait émerger ses dimensions distinctives, c'est-à-dire celles qui présentent un intérêt pour l'étude de cas.

Nous avons également réalisé une seconde forme de triangulation méthodologique, que Denzin appelle « la triangulation entre ou à travers les méthodes » (2009, 308). Les défauts d'une méthode sont souvent les points forts d'une autre, et en combinant les méthodes, les observateurs peuvent tirer le meilleur de chacune, tout en surmontant leurs lacunes respectives. Dans nos deux cas, nous avons utilisé plusieurs formes d'observation (observation directe et shadowing), ainsi que plusieurs formes d'entretiens (exploratoires, semi-structurés, et structurés). Ces méthodes sont pertinentes pour rendre compte et comprendre les usages des outils d'IA, puisque nous sommes au plus proche de l'action.

Notre recherche a été conditionnée par nos appétences personnelles et par la démarche caractéristique du CRG, qui privilégie l'exploration du terrain avant d'entamer la théorisation du matériau empirique. Nous avons choisi d'étudier les usages des outils d'IA, et plus particulièrement la dynamique de collaboration humain-machine.

Notre approche méthodologique s'inscrit dans un courant de recherche qui préconise l'utilisation de l'ethnographie. Le but est de recueillir une diversité de données issues de multiples sources afin de mieux appréhender un phénomène complexe. Nous avons utilisé des méthodes de triangulation afin de renforcer cette démarche tant au niveau empirique, méthodologique que théorique.

Notre processus de recherche, caractérisé par des itérations constantes entre le terrain et la théorie, a conduit à deux études de cas permettant de collecter et de traiter le matériau pertinent dans une approche qualitative. Dans la suite de ce travail, nous détaillerons la manière dont ces études de cas ont été élaborées.

2. Construction des études de cas

Notre recherche nous a mené à élaborer deux études de cas sur les usages des outils d'IA. La première étude de cas, RADO, porte sur les usages des outils d'IA dans les soins courants, et le second sur les usages des outils d'IA dans la recherche médicale. Nous commençons par expliquer la façon dont ils ont été construits (2.1 et 2.2) afin de monter les aller-retours et l'émergence des cadres théoriques. Enfin, nous synthétisons ces deux études de cas (2.3), selon les trois axes classiques : de quoi mon cas est-il le cas ? De quoi mon cas est-il fait ? Que peut produire mon cas ? (Dumez 2021).

2.1. Construction de l'étude de cas RADO : usage des outils d'IA dans les soins courants

Afin de bâtir RADO, nous exposons d'abord l'accès au terrain (2.1.1). Narrer cet accès offre une meilleure appréhension des données (Jensen 2010; Seaver, 2017). En outre, cet exercice explore les limites de la collecte, que nous aborderons en deux volets : tout d'abord, la collecte de données par *shadowing*, une technique qui consiste à suivre un acteur pas à pas pour éprouver son quotidien et qui occupe une place centrale dans notre matériau (2.1.2) ; ensuite, les données complémentaires intégrées à nos tactiques de triangulation (2.1.3).

Nous procédons ensuite à une description détaillée de l'analyse de ces données, en mettant en lumière les trois itérations qui ont contribué à l'émergence d'un nouveau cadre théorique (2.1.4).

2.1.1. Accès au terrain

En ce qui concerne l'étude sur l'usage des outils d'IA par les radiologues, l'accès au terrain a été facilité par la méthode de la boule de neige, également connue sous le nom de « *snowball sampling* » (Parker, Scott, et Geddes 2019). Cette approche débute avec un petit nombre de contacts initiaux invités à participer à la recherche, qui sont ensuite sollicités pour recommander d'autres contacts, et ainsi de suite. Elle est généralement clôturée une fois qu'une taille d'échantillon cible ou un point de saturation a été atteint.

Cette méthode s'est révélée particulièrement avantageuse pour entrer en contact avec des médecins, une population souvent difficile d'accès en raison de la charge de travail. De plus, trouver des médecins répondant à notre critère de recherche (faisant usage d'un outil d'IA) s'est révélé difficile en l'absence de registres officiels et en raison de la confidentialité entourant les carnets d'adresses des concepteurs.

Grâce à cette approche, nous avons pu observer le travail d'équipes de radiologues dans deux établissements de santé distincts, accompagnés d'observations en *shadowing* d'un radiologue pour chaque centre (2.1.2). En plus des observations, nous avons mené 14

entretiens avec des radiologues à propos de leur usage des outils d'IA (2.1.3). En complément de l'accès aux médecins par la méthode de la boule de neige, nous avons également réalisé 17 entretiens structurés avec des concepteurs (2.1.3), à l'occasion d'un congrès médical.

Cependant, l'atteinte d'une saturation a constitué un défi en raison de trois difficultés majeures liées à l'accès au terrain.

Tout d'abord, nous avons constaté que peu de professionnels de santé utilisent des outils d'IA dans leur pratique quotidienne. Parmi les neuf centres médicaux contactés, seuls trois utilisaient effectivement ces outils, tandis que les autres se limitaient à des études de faisabilité sans les intégrer pleinement dans leur pratique quotidienne. Leur évaluation porte sur la valeur réelle de ces outils dans leur pratique, les coûts associés, les budgets disponibles, et les protections nécessaires pour les patients.

En outre, certains médecins étaient réticents à participer à des études scientifiques d'observation. Ces réticences semblaient liées à la crainte d'exposer leurs pratiques à un jugement extérieur ou en raison d'un sentiment de manque de temps.

Dans chacun des trois centres observés, un seul médecin a accepté d'être observé de manière prolongée par *shadowing*. Cinq autres médecins ont consenti à des entretiens (trois dans le premier centre et deux dans le deuxième), tandis que les autres entretiens ont été réalisés dans des centres que nous n'avons pas pu visiter en France (cinq entretiens) et aux États-Unis (quatre entretiens).

Enfin, le cadre législatif, en particulier aux États-Unis, était un frein à l'accès au terrain. Même avec un médecin ayant consenti à l'observation, nos visites sur site ont été entravées par des exigences d'accords de confidentialité et des besoins en formation que nous n'avons pas pu satisfaire en raison de contraintes de temps lors de notre séjour à la Wharton School. Ces trois contraintes d'accès au terrain semblent partagées par les concepteurs eux-mêmes, comme ils ont pu nous en témoigner.

Parallèlement à la méthode par boule de neige, nous avons participé à un congrès médical, les Journées Francophones en Radiologie, du 7 au 10 octobre 2022 à Paris. Cet événement réunit annuellement de nombreux médecins partageant leurs expériences, ainsi que des industriels cherchant à promouvoir leurs produits. Nous avons saisi cette opportunité pour assister à des conférences sur les outils d'IA, nous entretenir avec des médecins, et échanger avec des concepteurs. Ces interactions ont été fructueuses, 17 concepteurs acceptant de participer à des interviews structurées et à des discussions informelles.

2.1.2. Observation par *shadowing*

Les données principales de cette étude sont liées à l'observation des interactions entre les radiologues et les outils d'IA utilisés pour l'aide au diagnostic. Nous avons utilisé la méthode

du *shadowing* (McDonald 2005; Czarniawska 2021), couramment employée dans les approches qualitatives axées sur l'observation d'individus dans des contextes organisationnels (Mintzberg 1970; Walker, Guest, et Turner 2017).

Elle implique de suivre physiquement et de près un individu de l'organisation pendant une période prolongée. Le chercheur devient l'ombre de la personne étudiée, la suivant pas à pas, avec la possibilité de poser des questions pour obtenir des précisions ou comprendre leur motivation.

La durée des observations peut varier, allant d'un jour à un mois, et peut se concentrer sur un seul membre de l'organisation, un rôle spécifique, ou un plus grand nombre de personnes et de postes.

Les données collectées sont riches, denses et offrent une image détaillée, directe et multidimensionnelle du rôle, de l'approche, de la philosophie et des tâches de la personne observée.

Nous avons appliqué le *shadowing* auprès de deux radiologues expérimentés. Nous avons délibérément arrêté la collecte de données après avoir atteint la saturation dans ces observations (Lowe et al. 2018). Cela nous a permis de privilégier la collecte de données sur d'autres terrains.

Premièrement, le Dr S. exerce dans une clinique privée en Île-de-France depuis plus de vingt ans. Il interprète en moyenne une centaine d'exams par semaine, principalement des exams de dépistage du cancer mammaire. Les patientes concernées par ces exams ont, la plupart du temps, plus de 50 ans, sauf en cas d'antécédents familiaux ou personnels qui impliquent un dépistage précoce. Certaines sont des habituées de l'établissement que le Dr S. reconnaît par leur nom ou à la lecture de leur dossier médical. Il utilise l'outil d'IA pour le diagnostic des mammographies depuis près d'un an, à raison de deux demi-journées par semaine.

Nous l'avons observé pendant 9 demi-journées entre juin 2022 et novembre 2022, soit 40 heures d'observation. Une attention particulière a été accordée à l'usage de l'outil d'IA, affiché sur une tablette distincte ajoutée à ses consoles d'interprétation. Le Dr S. commente, qualifiant la tablette de « solution du pauvre », expliquant que le concepteur de l'outil d'IA n'est pas autorisé à l'intégrer au système d'archivage lié à l'écran de lecture (voir figure 8). Nous avons également eu accès à la documentation technique des outils d'IA fournie au radiologue par le concepteur.

Deuxièmement, la Dr G., radiologue depuis plus de dix ans, travaille dans un service spécialisé dans un hôpital de pointe, où elle utilise l'outil d'IA depuis plus de huit ans. Le service accueille des patientes aux situations cliniques plus complexes, avec fréquentes suspicions de cancer.

Nous l'avons observée pendant 7 demi-journées entre octobre 2022 et février 2023, soit un total de 30 heures d'observation. L'outil d'IA utilisé par la Dr G. est intégré directement dans l'image numérisée de la mammographie sur la console d'interprétation, ce qui rend son usage moins discernable. Les explications de la Dr. G, obtenues par des questions posées pendant son exercice, ont compensé cette limitation. Son observation a également inclus la présence de collègues et d'internes.

Par rapport à d'autres approches qualitatives, le *shadowing* se distingue par son niveau d'analyse. Les données collectées sont plus diversifiées et détaillées qu'elles ne le sont lors d'un entretien. Le chercheur a ainsi accès à des aspects difficiles à étudier, tels que les événements triviaux et les éléments difficiles à exprimer.

De plus, l'unité d'analyse se concentre sur des actions contextualisées dans leur situation, avec moins de couches interprétatives par rapport à une série d'entretiens. Elle n'est pas conçue comme un simple moyen « quantitatif » d'observation neutre, comme elle a pu l'être dans certaines études, mais comme un moyen de percevoir le monde à travers les yeux de la personne suivie (Mintzberg 1970; Stanley et al. 1998; Snyder et Glueck 2019). Cette technique capture donc la nature brève, fragmentée, variée, orale et interrompue de la vie organisationnelle (Weick 1974). En plus de répondre au quoi et au comment, elle donne une meilleure compréhension du pourquoi.

Les radiologues observés ont spontanément verbalisé de nombreuses actions et expliqué leur raisonnement. De plus, nous avons constaté une forme de pédagogie dans leurs explications, avec de nombreux détails au début, devenant plus succincts pour les cas déjà rencontrés lorsque nous manifestions notre compréhension. Au fur et à mesure que notre compréhension d'améliorait, les explications données étaient de plus en plus pointues, au sujet des cas particuliers.

Cette méthode présente cependant certaines limites. L'un des principaux problèmes du *shadowing* est l'accès aux données. En plus des difficultés à trouver des acteurs prêts à être observés, comme évoqué précédemment, nous avons rencontré quelques limites sur le terrain. S'il nous a été possible de suivre la Dr G. dans la salle dédiée aux biopsies, cet accès ne nous a pas été accordé dans la clinique privée. Un autre problème habituel du *shadowing* concerne la relation avec la personne observée. Snyder et Glueck (2019) ont observé qu'il pouvait y avoir un certain inconfort initial. Dans notre cas nous n'avons pas perçu ce problème, les médecins, experts étant habitués à être suivis et observés par des étudiants (Cydulka, Emerman, et Jouriles 1996; Rancour 1996).

Nous sommes porté à croire que notre présence n'a pas eu une grande influence sur les acteurs observés. Cependant, l'effet « Hawthorne » (Shipman 2014), c'est-à-dire l'altération du

phénomène observé due à la présence de l'observateur, ne peut être ni mesuré ni écarté (Snow et Thomas 1994).

La gestion des données émerge comme l'un des défis majeurs de la méthode de *shadowing*. Les études observationnelles peuvent générer rapidement d'abondantes données difficiles à maîtriser (Forsblad 1984). Nous bénéficions d'un terrain qui présente une répétition dans les actions. Ainsi, pour simplifier le processus et de saisir et capter le maximum d'informations, nous avons adopté des abréviations. Par exemple, l'action d'écrire un compte-rendu médical est représentée par « CR » suivi de mots-clés distinctifs. En l'absence d'informations particulières, le symbole « \emptyset » est utilisé. Ainsi, notre matériel comprend en moyenne 3 000 mots par demi-journée d'observation, par rapport à la norme de la méthode du *shadowing* qui atteint généralement 4 000 à 5 000 mots (McDonald 2005).

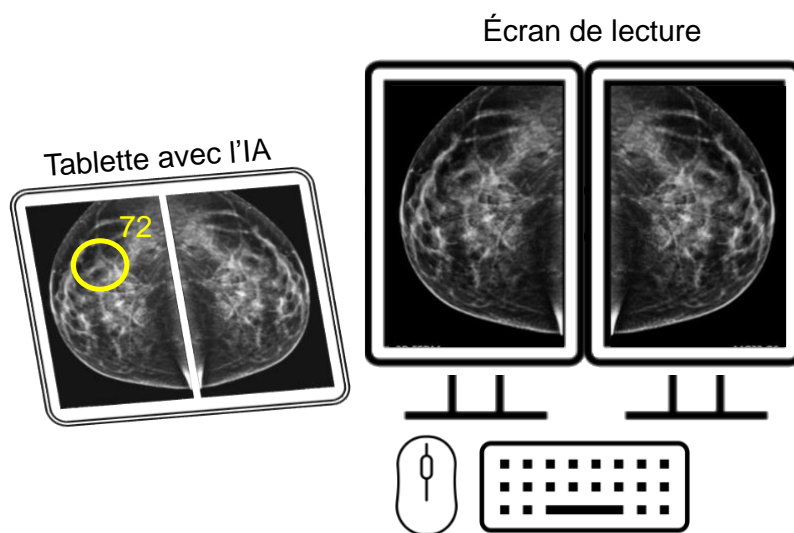


Figure 8 : Description du poste d'analyse des mammographies du Dr G

Les observations ont été consignées dans un journal de terrain en format papier, puis numérisées à la fin de chaque journée ou le jour suivant. L'anonymat des patientes et des professionnels de santé a été préservé lors de la prise de notes sur le terrain, et a été renforcé lors du processus de transcription. Par exemple, à l'écrit, les noms des patientes ont été substitués par des lettres telles que « X », « Y », ou « Z » pour éviter toute confusion dans les références. D'autres détails, tels que le pays d'origine ou l'âge, ont été généralisés ou modifiés. Ensuite, au format électronique, les données ont été classées de manière itérative en relation avec l'analyse.

2.1.3. Données complémentaires

Des données complémentaires à ces observations ont été recueillies. Tout d'abord, avant même les observations par *shadowing*, nous avons effectué des observations plus générales au niveau des services dans lesquels nous avons pu être introduit, totalisant 10 heures

d'observation par établissement de santé. Elles contextualisent les situations, montrant comment les acteurs interagissaient les uns avec les autres, au-delà de la simple interaction entre le radiologue et l'outil d'IA.

L'observation dans ces deux services nous a également permis de mener des entretiens individuels avec cinq radiologues en présentiel. Ces entretiens ont été complétés par cinq autres entretiens en ligne (trois par visioconférence et deux par téléphone) avec des radiologues français d'autres centres que nous n'avons pas pu visiter.

Nous avons également effectué un séjour de deux mois à l'Université de Pennsylvanie, aux États-Unis, du 1er juillet au 10 septembre 2023, à la Wharton School. L'un de nos objectifs était de confirmer les données observées en France. Un mois avant notre arrivée, nous avons trouvé un terrain. Cependant, la réglementation américaine exige une certification pour avoir accès à des données médicales. Nous n'avons pas été en mesure de compléter cette formation dans les délais. Nous avons donc opté pour la réalisation d'entretiens.

Quatre radiologues ont ainsi partagé leurs expériences avec nous. Trois entretiens ont été menés en ligne (deux par visioconférence et un par téléphone), et le quatrième s'est déroulé dans le bureau du médecin. Tous les entretiens ont été réalisés de manière individuelle. Un questionnaire d'entretien avait été préalablement réalisé au moyen des observations et des analyses empiriques effectuées en France. D'autres questions ont émergé pendant l'entretien, notamment parce que trois d'entre eux n'utilisaient pas d'outils d'IA au quotidien. L'entretien était orienté sur leurs usages des outils et leurs points de vue quant à la valeur de ces outils. L'ensemble des entretiens a été enregistré avec un magnétophone avec accord oral préalable des médecins, sauf pour cinq d'entre eux qui ont préféré la prise de notes. Les informations ont ensuite été retranscrites par nos soins après les échanges. (Voir l'annexe 3 : Entretiens de RADO.)

Enfin, lors de l'observation du Dr S., les congrès médicaux nous ont été exposés comme étant des lieux d'échange entre médecins et concepteurs. Nous avons donc décidé d'assister aux Journées Francophones en Radiologie 2022 à Paris les 8 et 9 octobre. Le premier jour, nous nous sommes familiarisés avec l'environnement. En ce qui concerne la collecte de données, nous avons assisté à trois présentations officielles, lors desquelles nous avons pris des notes, transcrites par la suite pour être analysées.

En plus des présentations, le congrès nous a offert l'opportunité de rencontrer des concepteurs. Un espace de 730 mètres carrés était exclusivement réservé aux concepteurs d'outils d'IA, le « Forum Intelligence Artificielle ». Nous avons donc visité de nombreux stands le premier jour. Si aucun événement particulier n'attirait notre attention, nous continuions notre exploration, nous arrêtant environ toutes les deux à trois heures pour rédiger un résumé des observations dans notre journal de terrain papier. Lors d'événements plus intéressants,

nous essayions de mémoriser un maximum d'informations avant de nous isoler pour retranscrire au mieux la situation.

Par exemple, une conversation informelle avec un vendeur a été interrompue par un entretien programmé avec une médecin. Celle-ci souhaitait acquérir leur solution depuis plus d'un an sans parvenir à obtenir le financement nécessaire au sein de son établissement. Avec l'accord des protagonistes, nous avons pu assister à ces échanges. Nous avons ensuite retranscrit de mémoire le discours de vente et les moyens évoqués par l'industriel pour débloquer la situation.

Lors de cette première journée sur le terrain nous avons constaté que les concepteurs passaient la majeure partie du congrès au niveau de leurs stands. Nous avons donc élaboré un guide d'entretiens structurés. Cette démarche empirique opportuniste nous a permis de réaliser 17 entretiens d'une durée moyenne de 14 minutes (allant de 7 minutes à 39 minutes au minimum). Un tableau résume les personnes interrogées (annexe 3, tableau 8). Conformément aux accords passés avec les personnes interrogées, les données ont été anonymisées.

Le tableau 8 comprend un identifiant pour les verbatims utilisés dans les résultats, la fonction de la personne interrogée (divisée entre des rôles techniques ou commerciaux, junior s'il ou elle a moins de 5 ans d'expérience professionnelle, senior sinon, ou faisant partie du groupe dirigeant de l'entreprise). L'entreprise dans laquelle il ou elle travaille est classifiée en fonction de la période de création (période 1 : avant 2012 ; période 2 : de 2012 à 2016 ; période 3 : de 2017 à 2022), de son origine (française, du reste de l'Europe ou des États-Unis) et de sa taille selon la nomenclature française. Le tableau indique également le type d'outils d'IA proposés et leur nombre. Ces entretiens ont été menés en face à face avec un magnétophone, puis retranscrits pour analyse.

2.1.4. Analyse des données

Le terrain empirique a été traité par trois itérations, avec des méthodes différentes (triangulation méthodologique). Ces itérations ont eu un effet entonnoir, allant de considérations générales à la tension particulière que reflète le matériau, pour gérer la diversité des données, identifier les caractéristiques empiriques de l'étude et trouver sa valeur scientifique.

Faire un état des lieux du matériau empirique

La première analyse a été réalisée à mi-chemin de l'observation du Dr S. (après le sixième jour d'observation). Nous avons également scruté la manière de travailler au niveau du service et mené des entretiens avec des radiologues de cet établissement et d'autres centres. À ce stade, notre questionnement demeure principalement empirique : comment les radiologues

intègrent-ils les outils d'IA dans leur activité ? L'objectif dans cette première phase de traitement du matériau était de prendre du recul par rapport au terrain.

Aucune approche théorique avait été privilégiée afin de minimiser le risque de circularité (Ayache et Dumez 2011a; Dumez 2012; 2013; 2021). L'absence d'hypothèses théoriques ne signifie pas pour autant l'absence de cadres directeurs pour orienter les données. Ces cadres restent suffisamment flexibles pour ne pas structurer excessivement le matériau et, par conséquent, les résultats (Ayache et Dumez 2011b).

Nous sommes influencés par deux courants intellectuels. D'une part, l'approche des réseaux sociotechniques, notamment telle qu'explorée par Callon (1986), qui nous amène à considérer l'outil d'IA comme un acteur, un actant, à part entière du système. D'autre part, le pragmatisme, découvert à la lecture de l'ouvrage de Lorino (2020), qui oriente notre regard vers la médiation de l'activité³³, en particulier par le biais de l'outil d'IA. Ensemble, ils nous amènent à regarder l'usage de l'outil d'IA dans l'action et plus particulièrement le type de relation qui se crée entre les acteurs, dont l'outil d'IA.

Nous choisissons la technique descriptive pour comprendre nos données. Nous organisons le matériau empirique en séries de données, englobant les observations générales, le *shadowing*, et les entretiens avec les autres radiologues. La description commence par exclure l'usage de l'outil d'IA par le radiologue. Au contraire, elle met en lumière différents catégories d'acteurs (principalement les radiologues, manipulateurs et patientes) ainsi que leurs actions (relations entre eux, coordination, actes médicaux...) et les dimensions qui influent sur les événements (l'organisation, le système informatique, etc.).

Entre autres, des schémas illustrent la répartition spatiale et temporelle de ces acteurs et de leurs actions (voir les figures 10 et 11). L'usage de l'outil d'IA est introduit par la suite, dans d'autres parties de la description du phénomène, mettant en évidence la médiation produite, la manière dont l'outil participe au sein du réseau d'actants, etc.

Cette analyse nous a donné l'opportunité d'observer de près la relation entre le radiologue et l'outil d'IA. Grâce à la tablette et aux explications fournies par le Dr S., nous avons obtenu des données d'une grande précision. La description a également contribué à une meilleure compréhension de la place de cette relation au sein de l'activité collective globale (la prise en charge de la patiente) et des dimensions qui influent sur cette relation.

³³ C'est un regard sur la manière dont une activité est influencée et structurée par des outils ou des artefacts, en particulier ici par un outil d'IA. Il s'agit de comprendre comment ces outils ne sont pas simplement des aides passives, mais des *actants* qui modifient la dynamique et la nature même de l'activité en question.

Comparaison des examens faits avec l'outil d'IA

La deuxième itération du traitement des données a eu lieu six mois plus tard, après la période d'observation du Dr S., les observations générales du deuxième service, et trois jours d'observations en *shadowing* de la Dr G. Nous avons également réalisé des entretiens supplémentaires avec des radiologues et avons participé au congrès médical dans lequel nous avons rencontré les concepteurs.

L'analyse a débuté en poursuivant les efforts de description précédents. Nous avons enrichi les séries de données, notamment avec de nouvelles sources provenant des entretiens avec les concepteurs. Cette démarche a amélioré la description et les manières de « décrire comme ». Elle a confirmé également les conclusions précédentes.

Cependant, l'observation de la relation entre le radiologue et l'outil d'IA était moins marqué dans le cas de la Dr G. en raison de sa familiarité avec l'outil, développée cours de huit ans d'utilisation, rendant implicite des actions que le Dr S. aurait verbalisé.

Nous avons alors décidé de procéder à une analyse plus approfondie de la démarche diagnostique, dans lequel l'outil d'IA était utilisé. Face à l'hétérogénéité des examens, nous avons structuré les données grâce à un « *template* » (Dumez 2021, 93). Le *template*, qualifié d'outil préthéorique, exclut des éléments jugés peu pertinents pour l'analyse, sans introduire de cadres théoriques. Plus exactement, le *template* facilite la comparaison des données en mettant en évidence les différences entre les situations.

Nous représentons le processus de diagnostic sous forme de schéma (voir la figure 9, issue du diagnostic 96, dont la transcription est en dessous). Les éléments inclus dans le schéma sont choisis en fonction des données rapportées dans le compte-rendu médical (mammographie, échographie, informations sur les antécédents médicaux...) et des éléments identifiés comme importants lors de la description précédente. Notre *template* met également en évidence la chronologie des actions menant au diagnostic. Les flèches rendent compte des liens entre les éléments.

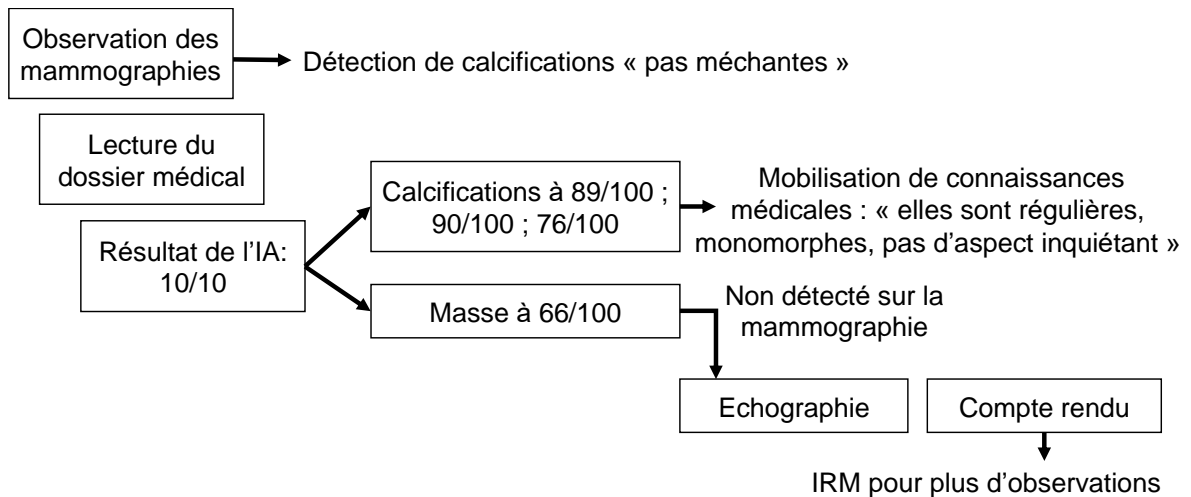


Figure 9 : Exemple de template (diagnostic 96, jour 8)

Diagnostic 96 (jour 8)

Il regarde la mammographie de la patiente et remarque la présence de clips des deux côtés (stigmates d'interventions antérieures). Il reconnaît le chirurgien aux clips ! Il regarde de plus près et juge que les calcifications ne semblent pas méchantes. « *L'IA sera sûrement à 10.* »
[Anticipation score]

Il lit les anciens comptes rendus. Dernière intervention il y a un an, donc la visite est de contrôle post-opératoire...

[Attente]

L'IA est à 10. « *Elle a sûrement remarqué les calcifications.* »
[Anticipation détection] Il regarde, l'IA trouve des calcifications : 76, 89, et 90/100 « *C'est même un 10+ !* ».

Il regarde de nouveau les mammographies. « *Oui, mais bon. Elles sont régulières, monomorphes, pas d'aspect inquiétant. Il faut juste continuer à observer.* » [Connaissances médicales] [Confrontation résultat]

Il revient sur le résultat de l'IA. « *Cependant, il y a une masse à 66/100. Je ne la vois pas dans la radiographie... Faudrait peut-être lui faire une IRM, elle a 50 ans.* »

Part faire l'échographie, puis revient. « *C'est compliqué. Il m'explique la trajectoire et la situation de la patiente. Je vais lui faire une IRM quand-même. Elle est jeune, 50 ans, avec une mère qui a eu un cancer, mais le test génétique est négatif.* » [Relation patiente]

Il écrit le compte-rendu en relisant les anciens documents. Elle a déjà eu une macro-biopsie « *je ne vais pas la faire biopsier à nouveau sur ça.* » [Relation patiente]

Les *templates* de chaque examen sont consignés sur des fiches, dont la comparaison a fait ressortir quatre grandes catégories en fonction du risque perçu par le radiologue et du score attribué par l'outil d'IA. Ces *templates* sont ensuite organisés dans un tableau à deux dimensions (voir tableau 4).

D'une part, le médecin semble réagir différemment selon qu'il perçoit une anomalie ou non. Les colonnes du tableau distinguent les diagnostics au cours desquels le radiologue semble percevoir un risque ou non. Bien que cette distinction binaire soit souvent évidente en fonction de ses réactions et explications du radiologue, certaines situations plus ambiguës ont nécessité l'exercice de notre jugement et de notre expérience, introduisant ainsi une marge d'erreur et d'imprécision.

D'autre part, le score de risque attribué par l'outil d'IA influence également le processus de diagnostic. Les rangées du tableau séparent le score global de l'outil d'IA entre un risque faible ou moyen et un risque élevé. Ces catégories sont définies dans la documentation fournie par le concepteur : un score inférieur à 7 est considéré comme faible ou moyen, tandis qu'un score de 8 à 10/10 est qualifié de fort.

En conséquence, le tableau 4 classe les diagnostics selon quatre configurations. Tout d'abord, dans l'accord d'absence (AA), le radiologue et l'outil d'IA ne perçoivent aucun risque. De manière symétrique, dans l'accord sur le risque (AR), le radiologue et l'outil d'IA tous les deux en alerte. Ensuite, le désaccord de type I (α) correspond à une perception de risque chez le radiologue, mais non chez l'outil d'IA. Enfin, le désaccord de type II (β) correspond aux situations dans lesquelles l'outil d'IA attribue un score de risque élevé, alors que le radiologue ne perçoit aucun risque initial.

Notre deuxième itération dans l'analyse se conclut par la lecture, l'appropriation et la détection des différences empiriques entre ces quatre configurations, contribuant à une compréhension approfondie des usages des outils d'IA par le radiologue.

Codage théorique

La troisième et dernière itération de l'analyse du matériau introduit la dimension théorique. La collecte de l'ensemble de nos données a généré un matériau volumineux et particulièrement varié. Cependant, les itérations précédentes ont apporté de l'ordre dans ce matériau, nous fournissant ainsi une première lecture empirique pour mettre en lumière les caractéristiques saillantes de cette étude de cas. Elle nous a ainsi permis de sélectionner les

cadres théoriques appropriés, un processus amorcé en parallèle de la collecte des données par des lectures, des discussions et des expérimentations avec le matériau.

Plusieurs cadres théoriques ont été mobilisés. D'abord, lors d'échanges avec un chercheur, nous avons découvert l'article de Lebovitz, Lifshitz-Assaf, et Levina (2022) sur les pratiques interrogatives. En plus de la similitude empirique, il éclaire de manière intéressante le matériau en soulignant son importance scientifique pour comprendre l'interaction entre l'outil d'IA et l'expert métier. Plus précisément, il met en avant la manière dont ce dernier gère l'opacité de l'outil, indiquant qu'il s'engage dans une collaboration avec l'outil. Cependant, la notion de pratiques interrogatives ne permet pas de saisir ce que la collaboration apporte à l'expert métier, en quoi il est augmenté.

En parallèle à cette lecture, notre attention a été attirée par la convergence entre cet outil d'IA et les « outil de gestion », dans le sens donné par la littérature éponyme. L'outil devient gestionnaire lorsqu'il participe à la prise de décision. Dans notre étude de cas, la description montre que l'usage de l'outil d'IA intervient dans l'évaluation d'une situation clinique et influe sur son résultat.

Au sein de la littérature sur les outils de gestion, le courant de l'instrumentation située s'intéresse plus précisément au lien entre l'outil de gestion qui médie une situation et la façon dont l'acteur peut agir sur celle-ci. Ce courant puise ses origines dans le pragmatisme (Aggeri et Labatut 2014) et dans l'étude d'anciens outils d'IA : les systèmes experts* (Suchman 1987; Rabardel 1995).

Dans ce contexte, le cadre d'analyse conformation/exploration (Moisdon 2005a) met en évidence la double nature de l'augmentation procurée par les outils de gestion, entre l'augmentation des capacités à faire mieux et à faire différemment. Ce cadre théorique semble apporter une perspective complémentaire à notre matériau et s'articuler avec la notion de pratiques interrogatives pour combler le fossé théorique identifié.

Afin de tester ces deux cadres théoriques, nous avons opté pour un codage multithématique dans notre analyse (Ayache et Dumez 2011b). Ce type de codage implique une démarche systématique visant à organiser un ensemble vaste et souvent chaotique de données hétérogènes afin de comprendre le phénomène empirique et de faire émerger des éléments d'analyse. Ce processus consiste d'abord à identifier des unités de sens (groupes de mots ou de phrases) pour ensuite les associer au sein de différents ensembles. Compte tenu de la possibilité pour une même unité de sens d'appartenir à plusieurs ensembles, le codage adopté est de nature multinomiale. Ainsi, l'approche multithématique a été utilisée, découpant le matériau en plusieurs thèmes, découlant de notre cadrage théorique.

Les thèmes identifiés comprennent les pratiques interrogatives, le rôle de l'outil dans la conformation et dans l'exploration. A ces thèmes théoriques, nous avons ajouté la relation

entre le médecin et le concepteur, ainsi que la relation entre le médecin et la patiente (voir annexe 4 : tableau de codage de RADO).

Cette analyse a été confrontée à la description ainsi qu'aux *templates* issus des diagnostics.

2.2. Construction de l'étude de cas sur le projet KOVAK : usage des outils d'IA dans la recherche médicale

Nous avons commencé des recherches empiriques au sujet du développement des outils d'IA dans la recherche médicale avant même d'observer des outils d'IA utilisés dans les soins courants. Nous retraçons ici les trajectoires des projets WILSON et CALI (2.2.1), des recherche-intervention et recherche-observation qui ont servi de socle pour appréhender la collecte des données du projet KOVAK (2.2.2) et de son analyse (2.2.3). Ce dernier aboutit à l'étude de cas sur l'usage des outils d'IA dans la recherche médicale présentée au chapitre 5.

2.2.1. Les projets WILSON et CALI et leurs enseignements

Le projet WILSON, instigué par OrgExp en juin 2020, a été notre premier terrain d'étude. Ce projet vise à élaborer un cadrage théorique sur les données numériques en vue de personnaliser le parcours du patient. De manière concomitante, il cherche à tester cette conceptualisation en entraînant un outil d'IA pour prédire le risque de complications induites par le traitement du cancer. Sous la tutelle de OrgExp, un groupe d'étudiants ingénieurs (GrpEtudiant) est chargé de mener une recherche scientifique d'une durée d'un an.

Le partenariat établi avec l'équipe de recherche médicale (GrpMedical) dans un centre de lutte contre le cancer permet à GrpEtudiant d'accéder aux données nécessaires pour l'entraînement des outils d'IA. En contrepartie de cet accès et de l'expertise fournie par GrpMedical, la propriété intellectuelle revient à cette dernière, avec l'intention de publier des articles scientifiques résultant de ces travaux collaboratifs.

Le déroulement du projet WILSON se divise en deux phases significatives. Après une année de travail, les résultats obtenus par GrpMedical semblent prometteurs, bien que non concluants. Pour conclure le travail, ORGEXP, nous propose à nous et à un data scientist, DataExp de reprendre le projet. À cette époque, l'idée d'une mission rapide, d'un mois au plus, est mise à mal lorsque DataExp met en lumière des faiblesses dans les constructions algorithmiques de GrpEtudiant. La reprise intégrale du projet devient donc nécessaire. Un an plus tard, le projet aboutit à nouveau à des résultats mitigés. Cependant, il nous fournit des leçons substantielles pour la suite de notre recherche.

Sur le plan empirique, il révèle des vulnérabilités liées à la collecte de données. Le démarrage tardif de la collecte, six mois après le début du projet, et la résiliation prématurée de la collaboration par GrpMedical limitent la possibilité d'une collecte de données

approfondie. Toutefois, l'expérience nous a aidé à nous familiariser avec l'approche ethnographique et avec l'analyse de ce type de données. Nous avons ainsi mieux compris ce type de phénomène et la manière de l'aborder.

D'un point de vue théorique, l'analyse des données collectées a transformé notre question de recherche. Initialement axée sur les conséquences des données massives sur la gestion de la personnalisation à grande échelle du parcours patient, la question évolue vers le rôle du développement d'outils d'IA en tant que moyen de générer des connaissances. Ainsi, nous avons adopté le modèle CRISP-DM* comme cadre théorique, issu des travaux de la communauté Knowledge Discovery in Databases (Fayyad, Piatetsky-Shapiro, et Smyth 1996b; 1996a; Cazacu et Titan 2021).

En plus du projet WILSON, nous avons eu l'opportunité d'observer le début d'un autre développement d'outil d'IA avec le projet CALI. Lancé en novembre 2021 en collaboration entre GrpMedical et le partenaire industriel GrpIndustriel, CALI vise à développer des outils d'IA pour suivre les patients en cours de traitement contre le cancer. Malheureusement, malgré des termes de collaboration similaires à ceux de WILSON, le projet CALI prend fin après quatre mois en raison du refus d'accès aux données par un acteur tiers.

En tant qu'observateurs nous avons plus de recul sur la situation et, forts de notre expérience avec le projet WILSON, nous avons pu mieux caractériser l'étude de cas sur le développement d'outils d'IA dans la recherche médicale. L'accès aux données médicales se révèle souvent complexe, en raison de contraintes réglementaires, mais également de divergences de vision entre data scientists et médecins. Ces constats renforcent l'idée que le développement d'outils d'IA est une entreprise délicate, reposant sur la confiance dans la capacité des outils d'IA et des données à décrire le phénomène en question.

Fort de ces expériences, nous avons entrepris des recherches pour identifier d'autres projets de développement d'outils d'IA. Nous avons constaté que de nombreuses initiatives étant encore à un stade préliminaire et que les entreprises impliquées préféreraient rester discrètes. Parmi ces investigations, le projet KOVAK nous est apparu comme une opportunité prometteuse pour développer nos connaissances, que nous avons saisi dans une démarche d'opportunisme méthodique (Girin 1990).

2.2.2. Collecte des données du projet KOVAK

Le projet KOVAK vise à élucider les facteurs associés aux effets indésirables* chez les patients sous immunothérapie*, un traitement mobilisant les défenses immunitaires pour lutter contre diverses pathologies, notamment des cancers. Cette étude repose sur une cohorte prospective* regroupant plus de 3 000 patients atteints de cancers et mobilise une équipe multidisciplinaire comprenant CoExp (experte de la cohorte), MedExp (experte

médicale), DataExp (expert en données), OrgExp (expert en sciences de gestion), et DatAdvice (experte des données aidant DataExp ponctuellement).

CoExp, épidémiologiste, actualise et met à disposition les données de la cohorte depuis 2020, tandis que MedExp, oncologue engagée dans la cohorte depuis 2017, enregistre les données des patients qu'elle prend en charge. Le projet, lancé par la collaboration antérieure entre CoExp et MedExp sur d'autres projets liés à la cohorte, s'enrichit de l'expertise de DataExp, data scientist, et d'OrgExp, chercheur en santé publique. DatAdvice, data scientist senior, intervient ponctuellement pour conseiller DataExp.

Notre immersion dans le projet commence avec des échanges informels avec DataExp et ORGEXP, en raison de nos intérêts professionnels respectifs. Au fil du temps, nous avons identifié une caractéristique distinctive du projet : la qualité du dialogue entre DataExp et MedExp. Cette dynamique, combinée à notre position en retrait, nous a motivé à faire du projet KOVAK une étude de cas en adoptant une approche ethnographique.

Notre méthodologie s'est inspirée de l'ethnographie appliquée à l'étude critique sur les algorithmes (Kitchin 2017; Seaver 2017), et de nos expériences précédentes. Nous avons choisi de mettre l'accent sur les acteurs clés (CoExp, MedExp, DataExp, DatAdvice, et OrgExp) ainsi que sur l'outil d'IA en cours de développement.

Nous avons collecté des données lors des réunions en ligne, principales occasions de collaboration étant donné la dispersion géographique de l'équipe. Nous avons participé en tant qu'observateur à ces réunions d'une durée de 42 à 93 minutes, abordant les progrès, les obstacles et les discussions approfondies sur le contexte et la signification des données. Nous avons enregistré ces conversations, que nous avons retranscrit en y ajoutant les supports des présentations, et des notes des participants.

En dehors des réunions formelles, les acteurs de l'étude communiquaient par mail pour préciser des informations. Nous avons conservé une trace des échanges courriel et des documents (articles scientifiques, etc.). Des échanges plus informels et directs avaient régulièrement lieu entre les collaborateurs d'un même site. Si nous n'avons pas toujours assisté aux discussions entre CoExp et MedExp, nous avons assisté régulièrement présents lors des échanges entre DataExp et OrgExp. Nous avons retranscrit ces échanges dans un journal de terrain chaque fin de journée.

Nous avons suivi le développement de l'outil d'IA via l'enregistrement des trois versions principales du code totalisant respectivement 2 462, 2 154, et 2 496 lignes de code. Il ne s'agissait pas d'une description détaillée de chaque piste d'exploration ni de tous les allers-retours effectués. Néanmoins, elles ont permis de tracer les principales pistes de recherche, y compris certaines qui n'ont pas abouti à des résultats probants. Notons que le code était dissocié des données cliniques, auxquelles nous n'avions pas accès.

En outre, nous avons mené deux entretiens semi-structurés (42 et 58 minutes) avec DATAEXP. Il s'agissait, en s'appuyant sur les différentes versions du code, de comprendre son point de vue sur le développement de l'outil d'IA et mieux appréhender les interactions qui ont eu lieu.

Par ailleurs, nous avons eu l'occasion de réaliser dix entretiens semi-structurés avec des acteurs travaillant dans l'introduction des outils d'IA dans le système de soins à l'occasion de notre visite à l'Université de Pennsylvanie (voir annexe 5). Leurs témoignages à propos des projets réalisés, ou en cours, nous ont aidé à mieux saisir les enjeux de l'étape de déploiement, et ainsi de réfléchir à la manière dont le projet KOVAK aurait pu continuer après notre collecte.

2.2.3. Analyse des données

Nous avons utilisé la narration (Dumez 2021) pour traiter les données, en raison de leur diversité et de la nature procédurale du phénomène. Pour élaborer cette narration, nous avons organisé l'ensemble des données en chronologies, mettant en lumière l'évolution du code et les actions des intervenants à travers les transcriptions des réunions. Nous avons ensuite superposé ces chronologies et identifié différentes séquences dans le projet. Ce découpage a été effectué de manière aussi indépendante que possible des théories, afin de prévenir tout risque épistémologique de circularité (Ayache et Dumez 2011a; Dumez 2012; 2013; 2021).

Cependant, notre perspective est influencée par le modèle CRISP-DM, que nous finissons par adopter pour décrire ce processus. CRISP-DM propose de concevoir le développement d'outils d'IA comme une série d'étapes interactives et itératives, composées de six éléments (voir l'annexe 7).

La première étape consiste à formaliser les besoins et les objectifs du projet pour pouvoir les traiter de façon algorithmique. Ensuite, la compréhension des données implique la collecte des données incluses, donnant une idée générale de la structure et du contenu des données disponibles à l'aide de statistiques descriptives. La préparation des données s'accompagne de procédures et de techniques visant à transformer les données initiales en une base de données opérationnelle pour les algorithmes d'apprentissage. La phase de modélisation consiste à entraîner des algorithmes d'apprentissage sur les données pour qu'ils identifient des modèles empiriques. L'évaluation teste ces modèles obtenus avec des scores de performance pour vérifier leur pertinence statistique et détermine leur adéquation avec les besoins du projet. Une fois le modèle stabilisé, il est introduit dans son contexte d'utilisation.

Contrairement à la linéarité que le modèle suggère, le projet KOVAK, tout comme d'autres projets de développement d'outils d'IA que nous connaissons, est marqué par une séquence d'étapes « superposées et entremêlées » plutôt « qu'interactives et itératives ». De nombreux allers-retours entre les différentes étapes se produisent, parfois de manière très rapide. La simplicité induite par le modèle nous a poussé à appliquer une forme de rasoir d'Ockham au

processus narratif, ne conservant que l'essentiel, une « manière de décrire » qui commence par exclure les interactions entre les acteurs, mais que nous avons ajouté au fil de nos itérations.

En entamant le processus par la superposition des chronologies, nous avons classé chaque unité de sens en fonction des six éléments du modèle CRISP-DM. Cela a conduit à une fragmentation du matériau selon l'avancement du projet. Nous avons également constaté que certaines données renvoyaient simultanément à plusieurs catégories, les étapes étant « superposées ». Par exemple, bien que la compréhension de la situation soit principalement réalisée au début, elle continue de se compléter tout au long du projet. Un autre phénomène observé est que la prévision des étapes futures influe sur les étapes en cours. Certaines opérations de prétraitement, par exemple, tiennent compte de la modélisation et de l'évaluation ultérieures.

À la suite de cette catégorisation, nous avons synthétisé des blocs d'unités descriptives de plus en plus étendus, tout en nous permettant une certaine imprécision. Par exemple, une série d'allers-retours rapides entre prétraitement, modélisation et évaluation, qui se traduisait initialement par une séquence de dix-neuf étiquetages, était regroupée en seulement trois grands blocs.

Cette première phase de convergence des unités descriptives nous a permis d'obtenir une vue d'ensemble de la situation. Avec cette nouvelle organisation du matériau, trois grandes approches séquentielles ont émergé dans le projet, indépendamment de toute théorie. En premier lieu, le projet suivait une approche fondée sur l'expérience de CoExp et MedExp, mais qui limitait les possibilités d'action de DataExp. Cette approche ne permettait pas de produire des modèles statistiquement performants. Ensuite, DataExp a adopté une nouvelle approche, fondée sur les données, qui générait des modèles performants mais incompréhensibles pour les experts médicaux. Les intervenants ont alors développé une approche mixte, favorisant un échange approfondi entre eux. Cela a permis une intégration plus granulaire de leurs expertises et a abouti à un résultat probant.

Nous avons finalisé la construction de la narration en croiser les étapes du modèle CRISP-DM avec les trois grandes approches du projet. Nous avons ensuite synthétisé les différentes unités descriptives pour produire 14 étapes (voir figure 12), présentées dans l'étude de cas (chapitre 5).

L'analyse de la narration a fait apparaître un processus de construction de sens. L'interaction entre l'expert médical, l'expert en données et les algorithmes appliqués aux données conduisait à la production de connaissance par l'entraînement de l'outil d'IA. Cette façon de comprendre l'usage des outils d'IA dans la recherche médicale nous a conduit à adopter le concept d'enquête pragmatise de Peirce. Ce cadre théorique appréhende la nature de la

collaboration humain-machine en dépassant les dichotomies classiques, complétant la notion de pratiques interrogatives et englobant le cadre d'analyse conformation/exploration.

Nous avons ensuite procédé à un codage sur l'ensemble de notre matériau pour mieux comprendre les liens entre ces trois cadres théoriques. Nous avons donc réalisé à nouveau un codage multithématique (Ayache & Dumez, 2011). Les thèmes abordés étaient l'abduction, l'induction, la déduction, les pratiques interrogatives, la conformation, et l'exploration (voir annexe 6 : tableau de codage sur le projet KOVAK).

2.3. Synthèse générale des études de cas

Nos deux études de cas ont été élaborées au moyen d'un processus de comparaison systématique, instaurant un dialogue constant entre le terrain et la littérature. Dans notre contribution, cette comparaison s'étendait à un échange entre nos deux études de cas, favorisant ainsi le développement du travail conceptuel. Cependant, il nous est apparu essentiel de clarifier en premier lieu pourquoi nos deux études de cas peuvent entretenir un dialogue mutuel, en les synthétisant au travers des trois questions clés des études de cas (Dumez 2021) :

- De quoi mes cas sont-ils le cas ? (2.3.1)
- De quoi mes cas sont-ils composés ? (2.3.2)
- Que peuvent produire mes cas ? (2.3.3)

2.3.1. De quoi mes cas sont-ils le cas ?

L'interrogation « *What is it a case of?* » (H. S. Becker et Ragin 1992) se rapporte à la catégorisation empirique et théorique des cas. À quelles catégories nos cas peuvent-ils être comparés ? Tenter d'identifier la catégorie à laquelle le cas étudié pourrait être ramené implique nécessairement une comparaison de ce cas avec des cas similaires. Le cas est ainsi défini comme l'instance d'une classe de phénomènes (George et Bennett 2005), ou plutôt, il peut être rangé sous plusieurs catégories, correspondant à la multiplicité de ses propriétés complexes (Abbott 2004). Cette catégorisation s'opère au niveau empirique et au niveau théorique (Dumez 2013).

Notre première « théorie orientante », ou « *orienting theory* » d'après Whyte (1984), portait sur la gestion de la personnalisation, et plus particulièrement sur la singularité à grande échelle (Minvielle 2018), une théorie de cette perspective productive adapté au système de soins. Elle nous a conduit à notre premier terrain, le projet WILSON. Les données et leur traitement ont fait ressortir le besoin de comprendre les questions de développement des outils d'IA (Fayyad et al., 1996) et l'approche des réseaux sociotechniques. Ces deux cadres insistent sur l'effacement de la barrière entre développement et usage d'un outil. Pour mieux

le comprendre, nous avons commencé par explorer un terrain permettant d'étudier les usages des outils d'IA.

Ces théories ont permis d'ouvrir les terrains empiriques et ont aidé à caractériser ces derniers. Tout d'abord, nous avons considéré les deux études de cas comme une sous-catégorie des « technologies émergentes » au même titre que la blockchain, l'internet des objets, l'impression 3D, etc. Elles suscitent des réflexions pour penser la technologie différemment (Faraj et Pachidi 2021).

Cependant, notre analyse s'est concentrée davantage sur les usages des outils d'IA, qui ont une place à part dans ces technologies, en raison de leur développement et aussi de leur usage, qui plus est dans le système de soins. En effet, ce contexte se caractérise par une production, la prise en charge, qui demande une grande expertise professionnelle et se déroule dans un contexte d'incertitude et de jugement critique (face aux vies en jeu). Par conséquent, les usages des outils d'IA doivent aider les experts métier, que ce soit dans les soins courants ou pour construire de nouvelles connaissances qui, à terme, donneront de nouvelles possibilités d'action aux experts.

Les caractéristiques terrain donnent leur valeur scientifique aux études de cas. En raison du contexte empirique, les deux études de cas s'inscrivent dans le paradigme de la collaboration humain-machine, dans le cas où l'outil d'IA doit venir augmenter l'expert métier. Dans cette perspective, nous savons que la gestion de l'opacité de l'outil par l'utilisateur est cruciale, sans pour autant connaître quelle est l'augmentation possible. La première étude de cas a jeté une première lumière en considérant l'outil d'IA comme un outil de gestion et en articulant le cadre d'analyse conformation/exploration (Moisdon 2005a) à la notion de pratiques interrogatives (Lebovitz, Lifshitz-Assaf, et Levina 2022). La seconde étude de cas a précisé la nature de la collaboration humain-machine avec l'épistémologie de Peirce (1877; 1878a; 1878b; 1878c; 1878d; 1878e; 1887).

2.3.2. De quoi mes cas sont-ils composés ?

Une étude de cas regorge de manières de la voir et l'exhaustivité de ces possibilités est impossible (Dumez 2021). Pour complexifier davantage l'exercice, ces descriptions évoluent à mesure que de nouveaux éléments sont apportés, que ce soit par la collecte de données ou par leur traitement. Identifier ce qui suscite l'intérêt du chercheur devient donc d'autant plus délicat que la cible est en constante évolution. Nos deux études de cas peuvent être découpées en unités de sens et codées différemment.

Dans notre étude de cas sur les usages des outils d'IA dans les soins courants, nous avons identifié deux grands niveaux d'analyse. Lorsque nous avons effectué notre description, le matériau a été principalement examiné au niveau du service, et plus globalement de l'ensemble des acteurs concernés et des relations qu'ils entretenaient entre eux (approche du

réseau sociotechnique). Ensuite, nous avons approfondi notre analyse sur la relation entre le radiologue et l'outil d'IA. À travers la notion de pratiques interrogatives (Lebovitz, Lifshitz-Assaf, et Levina 2022), nous avons examiné la manière dont l'expert métier établit des liens entre ses connaissances et les résultats de l'outil d'IA.

Le matériau offre un autre niveau de lecture en observant les actions réalisées pour assurer la prise en charge. À cet égard, à l'aide des *templates* créés, nous avons comparé les diagnostics, issus des actions respectives du radiologue et de l'outil d'IA, ainsi que de leurs interactions. L'analyse de ces actions a permis d'étudier les médiations que l'outil rend possibles, comme le suggère le courant de l'instrumentation située avec le cadre d'analyse conformation/exploration (Moisdon 2005a).

Dans notre étude de cas sur les usages des outils d'IA pour la recherche médicale, nous avons identifié trois niveaux d'analyse. Tout d'abord, dans la narration, le matériau a été analysé comme un projet, avec des étapes (CRISP-DM) et différentes approches pour résoudre la question médicale.

Ensuite, notre démarche nous a conduit à examiner les relations entre les acteurs et leurs actions au sein de ce processus. Au niveau des relations, en plus de celles liées à l'outil d'IA, la notion de pratiques interrogatives (Lebovitz, Lifshitz-Assaf, et Levina 2022) peut être étendue à la manière dont les différents experts relient leurs connaissances à celles des autres.

Au niveau des actions, les médiations rendues possibles par l'outil ont été analysées par le cadre d'analyse conformation/exploration (Moisdon 2005a). De même, le matériau peut être examiné à travers l'enquête pragmatiste de Peirce et plus particulièrement des inférences produites par les acteurs, dont les outils d'IA (Peirce 1877; 1878a; 1878b; 1878c; 1878d; 1878e; 1887).

Il est évidemment nécessaire de faire preuve de prudence lors de la comparaison d'un élément provenant d'un cas avec un élément issu d'un autre cas. La comparaison présente des limites. Cependant, nous avons constaté des points de convergence entre nos deux études de cas quant aux interactions entre les experts : l'outil d'IA avec l'expert métier, l'expert métier avec le data scientist, le data scientist avec l'outil d'IA. Les études de cas sont ainsi constituées de la nature, du *modus operandi* et de la portée de ces échanges entre experts.

2.3.3. *Que peuvent produire nos cas ?*

Dumez (2023) souligne les limites inhérentes à une étude de cas. Il exclut la vérification de connaissances car face à une théorie établie, une preuve supplémentaire dans son sens reste une contribution trop faible. Réfuter ou infirmer une théorie (Koenig 2009), bien que possible, semble également exiger un investissement disproportionné par rapport au coût requis par une étude de cas.

En revanche, l'étude de cas peut contribuer à préciser le pouvoir explicatif d'une théorie, à condition que les effets prédits aient été spécifiés, par exemple sous la forme d'histoires hypothétiques (R. I. Sutton and Staw 1995). Il s'agit d'identifier le genre de scénarios qui devraient se produire si la théorie est valide.

Cependant, en raison de l'état de l'art et de notre approche sur le terrain, nous avons utilisé l'étude de cas pour sa capacité à générer des connaissances théoriques (Lijphart 1971; Eckstein 2000). Comme mentionné précédemment, une théorie nous sert de point de départ pour entrer sur le terrain, mais l'expérience du terrain éclaire la littérature par de nouveaux éléments. En contrepartie, nos conclusions et nos perspectives sur le phénomène nécessiteront des examens complémentaires pour les rendre plus valides.

Le dialogue entre nos deux études de cas n'a pas confirmé nos hypothèses, mais il les a approfondies, complétées et structurées pour rendre compte des modes de gestion d'un type de collaboration humain-machine au service de l'expert métier, et du besoin d'ouvrir ce cadre à l'ensemble de l'organisation. Cela a abouti à un modèle de collaboration entre humains et outils d'IA que nous appelons Expert Machine Collaborative Community (EMC2) (Chapitre 6, partie 2).

Dans un souci de transparence méthodologique et afin de fournir une perspective approfondie sur la construction de nos études de cas, nous avons détaillé le processus de leur élaboration. Ces récits ont pour but d'éclairer les conditions empiriques, mettant en évidence à la fois les aspects saillants et les lacunes des données.

Nous avons explicité la démarche qui nous a amené à adopter cette perspective, parmi la multitude d'approches, de cas et de théories qui auraient pu être appliqués au matériau empirique. Nous avons détaillé le cheminement qui a abouti à une théorie sur l'augmentation de l'expert métier au moyen d'une mise en relation étroite et constante entre la littérature existante et nos observations sur le terrain.



Chapitre 4 : Usages des outils d'IA dans les soins courants

RADO, notre première étude de cas, établit une description, puis elle analyse les usages d'un outil d'IA dans l'analyse mammographique (1). Ces résultats nous conduisent à conceptualiser l'outil d'IA comme un instrument de gestion. En nous appuyant sur la littérature existante associée, nous éclairons alors les potentialités d'augmentation de cet outil dans la collaboration humain-machine (2).

1. Usages des outils d'IA dans l'analyse mammographique

Après avoir illustré la prise en charge des patientes (1.1), nous détaillons la collaboration entre le radiologue et l'outil d'IA (1.2), ce qui nous amène à souligner les pratiques interrogatives que l'expert métier développe (1.3).

1.1. Gestion de la prise en charge au niveau du service

Dans le cadre de l'examen mammographique, la patiente peut suivre deux grands types de parcours (1.1.1). L'examen mammographique, réalisé dans le cadre de dépistage du cancer du sein, produit de nombreux résultats négatifs. Le parcours habituel est donc court, avec une seule visite chez le radiologue. En revanche, lorsque l'examen est positif, le parcours de soins se poursuit au-delà du service de radiologie.

Notre description, située au niveau du service de radiologie, concerne le moment critique où le radiologue prend une décision médicale, faisant basculer le parcours d'un côté ou de l'autre. Nous détaillons comment les acteurs s'organisent pour cette prise en charge (1.1.2).

1.1.1. Parcours de soins dans le dépistage du cancer du sein

Le cancer du sein demeure la pathologie la plus fréquemment diagnostiquée et la principale cause de décès par cancer chez les femmes en France. 61 214 diagnostics ont été effectués en 2023 (33 % des cancers diagnostiqués), et 12 100 personnes avaient décédé en 2018 (18 % des décès liés au cancer) (Institut national du cancer 2023).

En raison de l'importance de cette maladie, une réponse organisationnelle axée sur le dépistage précoce a été mise en place. La diminution de la mortalité de 1,6 % entre 2010 et 2018 s'explique en partie par la détection des cancers à un stade initial, qui représente environ 60 % des cas de cancer du sein. Cette approche permet une prise en charge plus efficace, moins coûteuse, et limite les séquelles associées à certains traitements. Ainsi, le dépistage organisé du cancer du sein a été généralisé en France en 2004.

En l'absence de symptômes ou de facteurs de risque autres que l'âge, le dépistage concerne toutes les femmes de 50 à 74 ans, leur offrant un accès biennal à une analyse

mammographique. En cas de symptômes ou de facteurs de risque, des modalités de suivi spécifiques sont recommandées, par exemple un accès annuel.

De plus, la recherche médicale se poursuit afin d'améliorer la stratégie de dépistage, notamment avec une plus grande personnalisation comme le montre l'étude européenne MyPeBS (pour « *My Personal Breast Screening* »)³⁴.

L'importance du diagnostic du cancer du sein a également conduit à une forme de standardisation des résultats des examens. Ainsi, le BI-RADS (pour « *Breast Imaging Reporting and Data System* ») est une classification élaborée par le Collège Américain de Radiologie pour standardiser la description des lésions mammaires visualisées en imagerie médicale en fonction du degré de suspicion de malignité. Selon le score, il existe des recommandations de prise en charge qui favorise une meilleure communication entre les radiologues et les autres professionnels de la santé impliqués, notamment dans le traitement.

Nous distinguons deux types de parcours en fonction du score BI-RADS.

Le parcours de soins le plus fréquent se limite à un diagnostic révélant un BI-RADS 1 (absence de lésion mammaire visualisée) ou 2 (lésion d'aspect bénin). Ce parcours concerne plus de 80 % des patientes de la clinique privée observée (celle avec le Dr S). L'équipe médicale réalise des examens de routine, exécutés de manière très cadencée. Les processus diagnostiques sont homogènes, simples et rapides, durant environ quinze minutes.

Si aucune lésion suspecte de cancer n'est visualisée sur l'examen, la patiente peut revenir dans deux ans. Le délai peut être réduit à un an si elle présente des facteurs de risque. Dans le cadre d'un dépistage, il convient d'évaluer la balance bénéfice-risque. En effet, en cas de faux positif de la mammographie (image décrite comme suspecte sur l'imagerie s'avérant être une lésion bénigne), cela engendre des analyses supplémentaires, comme une biopsie, impliquant le prélèvement d'un échantillon de la zone suspecte, source d'inconfort physique et psychique pour la patiente.

Le parcours de soins devient plus complexe lorsque l'examen d'imagerie signale une lésion de score BI-RADS plus élevé. Le score BI-RAD 3 correspond à une lésion probablement bénigne, mais pour laquelle un contrôle d'imagerie est recommandé pour confirmation. En cas de score BI-RADS à 4 ou 5 indiquant une lésion radiologique suspecte ou hautement suspecte de malignité, une biopsie ou d'autres examens complémentaires peuvent être réalisés afin de confirmer le diagnostic.

³⁴ <https://www.mypebs.eu/fr/>

La plupart de ces parcours plus complexes ont été observés au sein du centre hospitalier de pointe (avec la Dr G). De nombreuses patientes ont été orientées vers ce centre sur les recommandations d'un autre radiologue. L'équipe médicale complète alors l'examen d'imagerie par une biopsie, permettant de prélever un échantillon de la lésion suspecte pour examen anatomopathologique.

1.1.2. Organisation du service pour la prise en charge

Dans les deux types de parcours de soins, les établissements de santé suivent une organisation de la prise en charge semblable. D'après nos observations, la prise en charge se résume à trois acteurs principaux : la patiente, le manipulateur et le radiologue. Nous schématisons les actions prises entre ces acteurs sur le plan temporel (figure 10) et spatial (figure 11).

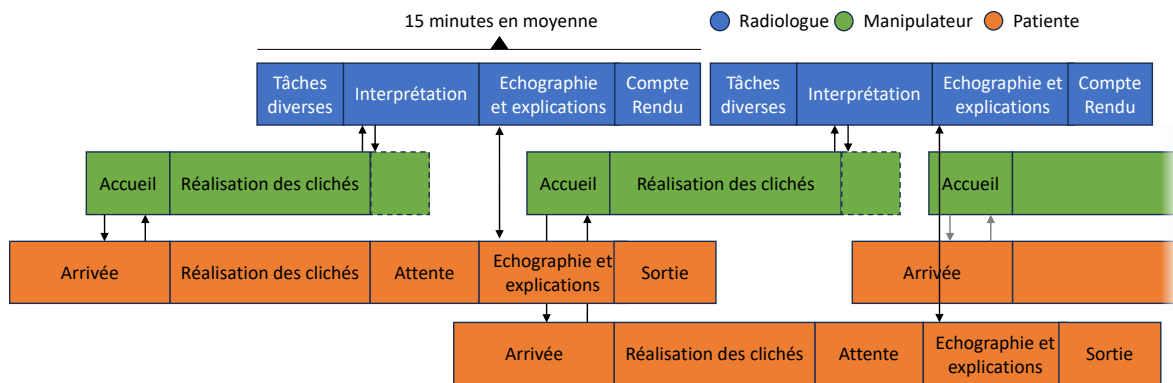


Figure 10 : Schéma temporel des actions prises entre les acteurs dans la prise en charge

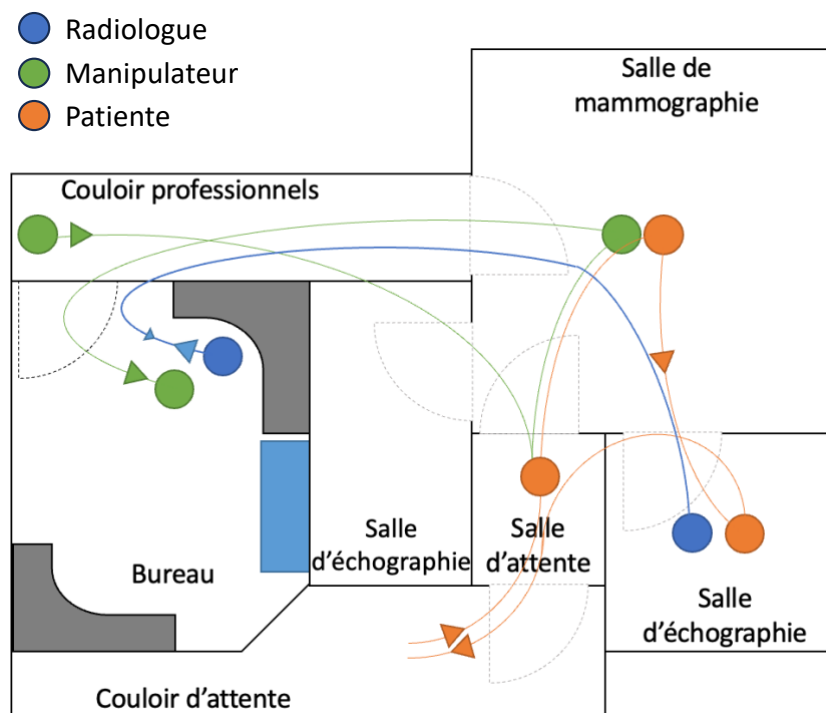


Figure 11 : Schéma spatial des actions prises entre les acteurs dans la prise en charge

La prise en charge de la patiente débute par un passage au secrétariat où elle est accueillie et dépose son dossier médical. Elle attend ensuite dans la salle d'attente où il lui est demandé de remplir un questionnaire médical. Elle est ensuite prise en charge par le manipulateur qui l'installe en salle de mammographie pour réaliser les images de ses seins. Ensuite, elle rentre dans la salle d'échographie où elle patiente le temps que le radiologue vienne réaliser l'examen échographique et lui explique son diagnostic. Une fois l'examen terminé, la patiente récupère le compte-rendu au secrétariat et passe aux caisses afin de régler les éventuels frais.

Au début de la prise en charge, le manipulateur récupère le dossier médical de la patiente et le remet au radiologue dans la salle d'interprétation, avant d'accueillir la patiente en salle d'attente. Il vérifie les réponses au questionnaire médical, puis réalise les images demandées par le radiologue. Par exemple, le protocole standard inclut des clichés de face, de côté et en 3D. Si la visite est motivée par des raisons plus spécifiques, le radiologue donne des indications supplémentaires au manipulateur. Ainsi, pour explorer une image suspecte, il peut demander un cliché centré sur une zone précise.

Dans la salle d'analyse, le radiologue consulte le dossier médical de la patiente avant d'analyser les images mammographiques. Il demande des clichés complémentaires au manipulateur si nécessaire. Lorsqu'il rencontre la patiente en salle d'échographie, il réalise un examen clinique, un interrogatoire, et lui explique les résultats. Il retourne ensuite en salle d'interprétation pour rédiger le compte rendu. Ce document possède une valeur médicale et légale qui permet de tracer l'interprétation du radiologue, en décrivant les éléments qui lui ont permis d'établir ses conclusions.

La schématisation de la prise en charge montre comment celle-ci est organisée autour de l'acte d'interprétation par le radiologue. Cet acte est d'une importance majeure pour la suite de la prise en charge de la patiente. Différents points soulignent cette importance. Le schéma temporel (VI.1) illustre comment le temps du radiologue est optimisé, quitte à faire attendre les patientes. Lorsque l'organisation temporelle du radiologue est perturbée, le manipulateur s'occupe de gérer le flux de patientes et de se coordonner avec les autres services de l'établissement afin d'éviter les temps morts, comme le montre cet échange (jour 2) :

Manipulateur : *Mme Y est en retard.*

Dr S : *Ce n'est pas grave.*

Nous : *Pas grave ?*

Dr S : *Si, c'est chiant. Du coup on la fait patienter, on en choisit une autre... On a l'embarras du choix.*

Manipulateur : *J'ai fait venir Mme Z.*

L'acte d'interprétation se fait principalement lors de l'analyse de la mammographie, là où apparaît un autre « acteur », l'outil d'IA. C'est un acteur qui n'intervient dans la prise en charge que par l'intermédiaire du radiologue, et seulement au moment de l'analyse de la mammographie. En effet, l'usage de l'outil d'IA ne se fait qu'avec le radiologue, au moment où celui-ci reçoit la mammographie dans sa salle d'analyse. Dans de rares occasions, des manipulateurs regardent les résultats de l'outil, par curiosité, mais cela reste et n'altère pas le processus de la prise en charge. De même, les patients ne sont informés de l'usage de l'outil d'IA qu'à l'oral, si le radiologue estime opportun de le mentionner. Le résultat de l'outil d'IA est absent du compte rendu, dont la valeur clinique et légale pose des questions de responsabilité liées à l'interprétation des images et de relation avec la patiente (jour 14, discussion entre la Dr G et un collègue) :

Collègue : Pourquoi tu recommandes de ne pas évoquer le résultat de l'IA dans le compte rendu ?

Dr G : C'est complexe, ça a des implications. Si jamais tu négatives un résultat, et qu'il y a des problèmes, tu peux être embêté.

Mais on est plus intelligents, s'indigne le collègue ! On peut prendre plus de résultats en compte.

Mais la patiente, elle peut se retourner contre toi en demandant pourquoi tu utilises un outil d'IA si ce n'est pas pour l'écouter. Et là, tu es embêté...

Par conséquent, le radiologue établit une forme de collaboration avec l'outil d'IA pour réaliser l'analyse de la mammographie. Cet usage semble circonscrit à la dyade formée entre l'expert métier et la machine, les autres acteurs et l'organisation plus généralement semble ne pas intervenir dans cette relation.

Au vu de l'importance de cet acte dans la prise en charge, nous précisons comment la collaboration, et donc la prise de décision, se déroule.

1.2. Collaboration entre le radiologue et l'outil d'IA

Nous explorons la collaboration entre le radiologue et l'outil d'IA à travers la manière dont l'expert métier prend en main l'outil (1.2.1) et comment il intègre les résultats dans son diagnostic (1.2.2).

1.2.1. Prise en main de l'outil d'IA

Les radiologues se familiarisent avec l'outil d'IA par une série d'étapes. Dans un premier temps, les symposiums offrent un espace pour comprendre pourquoi et comment utiliser

l'outil d'IA. À cet égard, nous avons assisté à celui organisé par des concepteurs et présenté par un radiologue émérite, le Dr E, lors des Journées Francophones de Radiologie (JFR) en octobre 2022. Ce type de présentation est également disponible en dehors des congrès médicaux, les concepteurs les organisent à d'autres moments de l'année.

Ces présentations sont une grande force, mélangeant des études pour prouver la valeur de l'outil d'IA et le témoignage des médecins pour illustrer ses usages.

Le Dr E commence la présentation en notant que dans 30 % des cas où un cancer est détecté lors d'une mammographie, ce cancer était déjà présent dans l'examen précédent. Une détection précoce réduit considérablement les complications et la lourdeur des traitements. L'outil d'IA pourrait ainsi éviter ces détections tardives, s'inscrivant dans la continuité des efforts de dépistage, en se concentrant sur la qualité de l'analyse de la mammographie.

Ensuite, le Dr E expose une étude comparative recensant les avantages de l'usage des outils d'IA en termes statistiques, avec la sensibilité, la spécificité, le taux de rappel et le temps de lecture (voir tableau 3) (Conant et al. 2019). Ces données n'opposent pas le radiologue et l'outil d'IA ; elles comparent les avantages d'une collaboration. Par ailleurs, cette perspective d'un « outil en aide » et non en substitution du radiologue domine le discours des concepteurs interrogés lors des JFR.

	Sans outil d'IA	Avec outil d'IA	Différence
Sensibilité	77,0 %	85 %	+ 8 %
Spécificité	62,7 %	69,6 %	+ 6,9 %
Taux de rappel	38,0 %	30,9 %	- 7,2 %
Temps de lecture	64,1 secondes	30,4 secondes	- 52,7 %

Tableau 3 : Comparaison des performances des radiologues avec et sans les outils d'IA (Conant et al. 2019)

Enfin, le Dr E explique comment il envisage l'usage de l'outil. Il estime qu'il joue un double rôle au niveau de la qualité de l'analyse : corriger les erreurs de détection et celles d'interprétation. Il ajoute que l'outil d'IA lui fait gagner du temps en l'absence d'anomalies, et même en cas d'anomalies, l'évaluation est plus facile réduisant ainsi le risque de manquer des lésions supplémentaires. Il partage son expérience quotidienne et explique comment la manière de lire les images change. Lorsqu'il examine des images en 2D, il commence par regarder la mammographie avant de consulter le résultat de l'outil d'IA. En revanche, lorsqu'il examine des images en 3D, il préfère obtenir d'abord l'avis de l'outil d'IA, car il présente une bonne valeur prédictive négative, évitant ainsi de perdre trop de temps à vérifier les images

lorsque aucune anomalie n'est présente. Nous avons retrouvé des descriptions similaires chez la majorité des radiologues interviewés (dix sur quatorze).

En plus ce type de présentations, les concepteurs offrent des présentations plus individualisées pour les radiologues souhaitant adopter un outil d'IA. Une première option consiste à faire une démonstration en demandant l'aide à des radiologues qui ont déjà adopté ces outils. Ainsi, les potentiels clients peuvent observer comment des médecins en font usage, comme nous le rapporte le Dr S.

Souvent, lorsque l'outil d'IA est encore peu adopté et pour les services plus réticents, une période d'essai est offerte pendant laquelle le service vérifie que l'outil correspond aux besoins de l'équipe médicale. Comme le rapportent une radiologue chef de service en France et une autre aux États-Unis, certaines études statistiques ne sont pas pertinentes dans les conditions de fonctionnement d'un service particulier, notamment en raison du profil clinique des patientes.

Enfin, lorsqu'un service a adopté l'outil d'IA de manière pérenne, les internes apprennent à en faire usage à l'instar des autres outils et savoir-faire à connaître, comme nous avons pu l'observer dans le service de la Dr G.

En revanche, la prise en main de l'outil par les pairs semble plus difficile lorsqu'il s'agit de s'aider mutuellement. Parmi les étapes présentées, une forme d'hierarchie ou de statut existe, avec le médecin qui présente dans une position d'expert, que ce soit dans le cadre d'une conférence ou vis-à-vis d'un interne. Mais, comme nous le relate le Dr S, la culture médicale difficile des modes d'apprentissage horizontaux (Jour 6) :

« C'est dommage qu'on n'ait pas la culture de l'apprentissage de nos erreurs. Je me souviens, mon chef en Belgique, il avait fait une réunion avec son staff pour leur montrer toutes les erreurs qu'il avait commises, alors qu'il était brillant. Mais la réception de ses pairs l'ont fait arrêter. On apprend de nos erreurs [théoriquement], le problème c'est qu'on est jugés de façon sanglante. Il y a une concurrence malsaine qui fait que les gens ne vont pas... »

Nous observons également une évolution dans la prise en main de l'outil d'IA au cours du temps. En comparant les observations du début avec le Dr S à celles réalisées à la fin, et à celles réalisées auprès de la Dr G, qui utilise l'outil depuis plus de huit ans, nous constatons l'apparition de certaines habitudes.

Par exemple, au début de nos observations, le Dr S s'appuyait beaucoup sur la documentation technique, un geste qui a disparu par la suite, tout en conservant les

arguments utilisés. De même, vers la fin de nos observations, il adoptait un comportement semblable à celui de la Dr G, consistant à anticiper dans quelles conditions l'outil d'IA n'allait pas être pertinent (à cause d'une cicatrice de chirurgie, d'implants mammaires, etc.). En somme, ils anticipaient plus facilement les comportements de l'outil.

La relation avec les concepteurs évolue également avec le temps. Initialement, ces derniers sont très présents, organisant des symposiums, des réunions d'information et des démonstrations. Même pendant la période d'essai, ils se montrent réactifs aux besoins des radiologues.

Cependant, après l'achat, la Dr G évoque de manière cynique comment « il ne faut pas se faire d'illusions ». Malgré cela, la relation avec les concepteurs demeure relativement ambiguë. Si pour la Dr G « nos retours sont une goutte d'eau dans un lac », le Dr S fait remonter des dysfonctionnements, ou des cas dans lesquels l'outil d'IA s'est avéré particulièrement utile, et exprime le souhait d'une relation plus approfondie.

Selon les concepteurs, ces informations sont synthétisées puis discutées pour orienter le développement de l'outil d'IA.

1.2.2. Diagnostic assisté par l'outil d'IA

Pour le diagnostic, le radiologue s'appuie sur les informations fournies par le dossier médical, examine l'image, compare le sein gauche et le sein droit, trace l'évolutivité avec les clichés précédents s'ils existent, et prend en compte les indications fournies par l'outil d'IA.

L'outil d'IA, quant à lui, calcule des probabilités sur les distributions de pixels en fonction de ses données d'entraînement. Concrètement, lorsqu'il traite les données dans la mammographie, l'outil affiche initialement un score global de 1 à 10, suivi de scores locaux entourant des zones suspectes de l'image sur une échelle de 1 à 100, différenciant les anomalies liées à des masses et à des calcifications.

Selon la documentation technique, les scores globaux de 1 à 7 visent à « rassurer », car environ seulement 0,3 cas sur 1 000 sont des cancers. Pour les scores de 8 à 9, l'outil « assiste » (environ 6 cas sur 1 000), et à 10, l'outil d'IA « alerte » (environ 1 cas sur 10) le radiologue. En outre, les termes « risque faible », « moyen », et « élevé » remplaceront probablement ces scores pour éviter toute confusion, selon les améliorations prévues. D'autres métriques accompagnent la documentation technique pour relativiser les chiffres fournis par l'outil.

Ces données s'insèrent dans un processus de prise de décision plus complexe que nous avons schématisé dans des templates, en prenant en compte les éléments du compte-rendu tels que les facteurs de risque dans le dossier médical et les antécédents médicaux, ainsi que nos observations, comme les raisonnements mobilisés ou les demandes de clichés complémentaires.

Ce travail n'a pas pu être réalisé avec les données liées à la Dr G, car le résultat de l'outil d'IA est directement intégré à la mammographie, et la verbalisation des actions est moins marquée et effectuée a posteriori.

Sur les 103 diagnostics observés auprès du Dr S, 87 diagnostics ont été modélisés et répertoriés dans le tableau 4, les 16 autres cas n'ont pas été interprétés en raison de l'absence de mammographie (seuls des contrôles échographiques ont été réalisés pour les cas n° 6, 15, 36, 65 et 101), des prothèses entravant l'outil d'IA (cas n° 25, 38, 46, 53, 87, 102) ou une mauvaise documentation (cas ° 21, 37, 43, 67, 70).

	Peu de risque	Risque possible	Total
Score IA ≤ 7	<u>Accord d'absence (AA)</u> N°2, 5, 7, 8, 12, 13, 16, 17, 18, 19, 23, 24, 26, 30, 31, 33, 35, 39, 40, 41, 44, 45, 48, 49, 51, 52, 54, 55, 58, 61, 64, 66, 67, 71, 73, 74, 78, 79, 80, 82, 90, 92, 93, 94, 97, 98, 99 47 diagnostics – 54%	<u>Désaccord type I (α)</u> N°3, 27, 32, 62, 81, 86, 89 7 diagnostics – 9%	54 diagnostics – 62%
	<u>Désaccord type II (β)</u> N°1, 4, 9, 19, 11, 14, 22, 28, 29, 34, 50, 56, 57, 59, 60, 69, 75, 76, 77, 82, 84, 85, 88, 95, 96, 103 26 diagnostics – 29%	<u>Accord de risque (AR)</u> N°20, 42, 47, 62, 72, 91, 100 7 diagnostics – 8%	
Score IA > 7			33 diagnostics – 38%
Total	73 diagnostics – 84%	14 diagnostics – 16%	87 diagnostics – 100%

Tableau 4 : Configurations des diagnostics du Dr S

Lors de la configuration « accord d'absence » (AA), le radiologue ne perçoit aucun risque, et l'outil d'IA affiche un score global faible (inférieur ou égal à 7). Ce cas prédomine dans la clinique privée du Dr S. L'usage de l'outil d'IA semble rassurer et faire gagner du temps. Pour l'analyse des images 2D, le gain de temps est marginal. Mais, pour la tomographie, un radiologue interviewé confirme qu'il parcourt les images « *en gardant l'œil ouvert en cas d'anomalie* », sans s'attarder plus que nécessaire, contrairement à l'absence d'outil d'IA où il peut passer plus de temps par crainte de passer à côté négliger quelque chose. L'usage de l'outil d'IA correspond donc à une forme de « triage », réduisant l'attention lorsque cela n'est pas nécessaire.

Dans la configuration « désaccord type I » (α), l'outil d'IA affiche un score global faible, pourtant le radiologue perçoit un risque possible. Les radiologues préfèrent alors se fier à leur expertise et questionnent la mammographie sans prendre en compte le résultat de l'outil d'IA. Ce comportement souligne la responsabilité professionnelle des radiologues, et qui s'étend même aux actions des manipulateurs. Par exemple, si le sein n'a pas été tiré assez, empêchant de voir une zone critique, les radiologues demandent qu'ils soient refaits.

Dans la configuration « désaccord type II » (β), l'outil d'IA indique un risque élevé (un score entre 8 et 10) alors que le radiologue ne perçoit initialement aucun risque. Cela permet d'améliorer la qualité des examens en détectant des anomalies plus précoces. L'usage de l'outil d'IA correspond alors à une forme de « filet de sécurité », ou de second avis médical. Bien que l'outil d'IA agisse comme « *si 50 radiologues donnaient leur avis* », selon les concepteurs, il existe un risque de faux positif (la patiente n'a rien malgré ce qu'indique l'outil). Ce risque est d'autant plus grand que l'outil d'IA est limité par construction aux informations sur l'image, alors que le radiologue peut vérifier qu'il n'y ait pas d'autres explications, comme une anomalie bénigne qui n'a pas évolué depuis des années. Le radiologue doit donc rester attentif et être capable de « négativer » le résultat de l'outil d'IA pour éviter un surtraitement. C'est ici que les pratiques interrogatives prennent tout leur sens.

Avant d'aborder les pratiques interrogatives, précisons que la configuration « accord de risque » (AR) apporte peu d'informations. Les quelques cas identifiés montrent qu'il est nécessaire de distinguer les situations dans lesquelles le radiologue et l'outil d'IA identifient une même anomalie, de ceux qui sont la somme de la configuration α et β . En effet, dans plusieurs situations, l'outil d'IA attribue un score élevé à une zone jugée bénigne par le radiologue, alors que ce dernier remarque une anomalie ailleurs.

1.3. Pratiques interrogatives mises en place

Dans notre étude de cas, nous identifions deux types de pratiques mobilisées par l'expert métier : des pratiques interrogatives professionnelles (1.3.1) et des pratiques interrogatives sur l'outil (1.3.2). Ces pratiques montrent comment l'expert gère l'opacité de l'outil et s'engage dans une collaboration en intégrant les résultats dans sa prise de décision. Cependant, il reste à voir comment elles s'articulent avec l'augmentation de ses capacités (1.3.3).

1.3.1. Pratiques interrogatives professionnelles

Les exemples de pratiques interrogatives détaillés par Lebovitz, Lifshitz-Assaf, et Levina (2022) présentent des similitudes avec celles employées par les radiologues que nous avons observées, que nous définissons comme des « pratiques interrogatives professionnelles ».

Ces pratiques ne se limitent pas à la compréhension du résultat de l'outil d'IA (dans la configuration β) ; elles sont également employées dans la configuration α pour confirmer la nature maligne de l'anomalie identifiée.

Par exemple, dans le cas 27, malgré un score bas de l'outil d'IA, le radiologue repère une zone qui lui semble suspecte. Il utilise différentes techniques d'analyse, notamment l'échographie qui ne révèle rien. En raison des importants facteurs de risque de la patiente, il choisit de s'assurer de la bénignité de la zone en recommandant une IRM, une méthode d'imagerie moins invasive que la biopsie.

Cas 27 (jour 3, configuration α)

Le Dr S revient de sa consultation : « *C'était une patiente compliquée. Elle fait un suivi annuel car sa mère a eu un cancer du sein. L'année dernière on a fait une biopsie qui s'est avérée bénigne. Mais cette année des formes se dessinent ailleurs. En tomosynthèse, l'IA donne 6, avec la zone à 26. En écho je ne trouve rien. On va faire une IRM pour vérifier.* »

De même, le cas 89 illustre comment le radiologue examine l'évolution de la zone préoccupante à travers l'historique des images.

Cas 89 (jour 8, configuration α)

Dr S : « *Alors Mme X, [le score de l'IA] est à 7.* »

Pendant qu'il regarde la mammographie en 2D, puis la 3D : « *C'est nickel... Ah ! Un petit... On va regarder... Peut-être un petit kyste ?* »

Il compare alors avec les mammographies d'autres années pour vérifier l'évolutivité : « *2015 non, 2017 oui, il y était, ça va.* »

Il va faire une échographie, puis revient écrire le compte rendu. Il nous témoigne : « *Je n'ai pas trouvé l'anomalie, mais le sein est très gras. C'est très complexe.* »

Nous qualifions ces pratiques interrogatives de « professionnelles » car elles existent indépendamment du fonctionnement de l'outil d'IA. La gestion de l'opacité de l'outil se fait en recourant aux connaissances métier, à l'image de ce qu'un expert ferait pour infirmer le

résultat d'un autre médecin. Il mobilise ses connaissances médicales pour caractériser la zone, demande des clichés complémentaires, utilise d'autres techniques d'imagerie et vérifie l'évolution de la zone à risque. Cependant, le radiologue met également en œuvre ces pratiques interrogatives professionnelles lorsqu'il cherche à infirmer le résultat de l'outil d'IA (configuration β).

1.3.2. Pratiques interrogatives sur l'outil

Les « pratiques interrogatives sur l'outil » constituent une seconde typologie que nous mettons en avant, en complément des pratiques interrogatives professionnelles, décrites par Lebovitz, Lifshitz-Assaf, et Levina (2022).

Ces pratiques reposent sur une compréhension du fonctionnement de l'outil d'IA qui se développe au fur et à mesure de son utilisation. Elles se retrouvent dans la configuration β en plus des pratiques interrogatives professionnelles.

Par exemple, dans le cas 56, le radiologue décide de réaliser un cliché supplémentaire car l'outil d'IA affiche un score global de 10 avec une zone à 61/100, alors que pour lui, « ce n'est rien ». Il se réfère également à la documentation pour évaluer l'importance du score attribué à la zone, constatant que malgré un score global de 10/10, le score local demeure « intermédiaire », remettant davantage en question l'anomalie détectée.

Cas 56 (jour 5, configuration β)

Dr S : « *Alors patiente suivante c'est Mme X.* »

Il regarde les clichés, puis le compte rendu précédent : « *Elle a été opérée du sein droit.* » Il regarde à nouveau. Puis il regarde l'outil d'IA. « *Ah ! il est à 10. Des masses. Le cliché gauche est à 61, ici à l'extrémité entre le sein et les côtes.* »

A la manipulatrice : « *Vous pouvez lui faire une 3D de profil en tirant sur son sein. Pour moi ce n'est rien. Sur le cliché oblique, il n'y a pas d'anomalie. Mais l'IA a détecté cette zone donc on va le faire plus précisément.* »

La manipulatrice part refaire le cliché. Dr S, en le recevant : « *Ha, voilà, c'est impeccable ! C'est bénin. Elle avait été classée en intermédiaire [il cherche la documentation technique], oui, c'est ça.* »

Il part faire l'échographie. Il revient et rédige le compte rendu.

Le cas 95 est particulièrement intéressant car il montre que le radiologue anticipe le résultat de l'outil d'IA. Son pronostic se révèle incorrect. Il raisonne ensuite en termes de probabilités liées au résultat, et il complète son analyse par des pratiques interrogatives professionnelles.

Cas 95 (jour 8, configuration β)

Le manipulateur installe les radiographies analogiques. Le Dr S les regarde à la loupe : « *Ah ! Il y a une petite calcification.* » Il lit les anciens comptes-rendus. Les mammographies arrivent, il les regarde.

Le manipulateur entre : « *Alors on ne la connaît pas. Elle a eu un cancer en avril dernier. Depuis elle fait une chimio [...] Elle a 4X ans.*

Elle a eu droit à la totale.

C'est le Dr. Z qui l'envoie. Il y a eu N ganglions enlevés. Elle était en larmes. Elle craint le résultat surtout.

Je vais aller la voir. Je n'ai pas encore eu de résultat de l'IA mais ce n'est pas grave. »

Le Dr S part faire l'échographie. Il revient et voit le score global de l'IA à 10. « *Eh oui ! car il y a une distorsion à cause de l'opération.* » Il regarde en détail. « *Ah ! Non, ce sont les calcifications qui l'embêtent ! Elles sont à 58/100 et 63/100. Ok.* » Il regarde à nouveau, en zoomant sur les zones suspectes. Il conclut le compte rendu en indiquant qu'il y a peu d'évolutivité et qu'un suivi annuel est recommandé.

Les pratiques professionnelles liées à l'outil se manifestent de différentes manières, évoluant en fonction de la familiarité acquise par le radiologue. Initialement, le radiologue peut se référer à la documentation technique pour vérifier le nombre de vrais positifs associés au score de l'outil d'IA. Puis, il cesse de consulter la documentation, ayant intériorisé ces probabilités. Avec l'expérience, d'autres facteurs entrent en jeu, tels que le type d'anomalie détectée (notamment le fait que les calcifications identifiées par l'outil d'IA sont plus souvent des faux positifs) ou les antécédents médicaux des patientes (par exemple, les cicatrices d'opérations). Enfin, nous observons l'émergence de l'anticipation des résultats de l'outil d'IA, suggérant que prévoir ses résultats équivaut à les comprendre.

1.3.3. Limites à comprendre l'augmentation de l'expert métier

La littérature montre que certains médecins, face à des résultats divergents de l'outil d'IA, déterminent que l'effort de comprendre le résultat est trop important par rapport à ce qu'il

apporte dans leur examen clinique. Jussupow et al. (2021) ajoutent que ces experts métier ont tendance à surestimer leurs compétences par rapport à celles qu'ils attribuent à l'outil d'IA.

Par opposition, Lebovitz, Lifshitz-Assaf, et Levina (2022), ainsi que Jussupow, Spohrer, et Heinzl (2022) observent des radiologues qui interagissent avec l'outil d'IA de manière plus soutenue, afin de s'assurer d'aucune négligence, bien qu'ils aient tous les compétences nécessaires pour évaluer les cas de patients sans l'aide de cette technologie.

Pour interagir avec l'outil d'IA, ils mobilisent des pratiques interrogatives (Lebovitz, Lifshitz-Assaf, et Levina 2022), ou d'autres formes de métacognition (Jussupow, Spohrer, et Heinzl 2022). Avec les pratiques interrogatives, l'expert métier s'engage dans la collaboration humain-machine, au lieu d'écarter systématiquement le résultat ou de l'intégrer dans le compte rendu sans reconsidérer la décision (Lebovitz, Lifshitz-Assaf, et Levina 2022).

Cependant, la notion de pratiques interrogatives – et celle plus large de métacognition – ne clarifie pas de quelle manière la gestion de l'opacité des résultats de l'outil d'IA apporte une augmentation de la capacité d'agir des experts métier. Par conséquent, cette littérature ne précise pas comment améliorer le processus collaboratif.

Dans ce type de collaboration, la responsabilité de l'acte médical est attribuée à l'expert métier. Un mode de gestion adapté doit donc permettre au radiologue de renforcer sa capacité à diagnostiquer le cas clinique, par exemple par une amélioration de la qualité, du temps nécessaire ou par une réduction du coût, à en croire les études quantitatives sur le sujet (Rodríguez-Ruiz et al. 2019; van Winkel et al. 2021; Raya-Povedano et al. 2021).

Notre étude de cas identifie deux usages majeurs, le filet de sécurité et le triage, nous analysons quel type d'augmentation ils produisent et leur possible lien avec les pratiques interrogatives.

Dans l'usage de l'outil d'IA comme filet de sécurité, c'est la mobilisation de pratiques interrogatives qui augmente les capacités du radiologue. Celles-ci sont nécessaires en cas de divergence des résultats comme dans la configuration β . Les pratiques interrogatives servent alors à réduire l'opacité du résultat de l'outil d'IA afin de l'incorporer dans la prise de décision. Cela évite des erreurs de détection (si l'outil d'IA a identifié une zone anormale), de négligence (si le radiologue se fixe sur une première tumeur et en laisse passer une seconde, plus discrète), mais également de surtraitement en remettant en question les faux positifs. La décision finale est donc de meilleure qualité, au détriment du temps et d'un coût supplémentaire induit par la mobilisation de techniques supplémentaires ou simplement l'énergie du radiologue.

Ensuite, l'engagement avec l'outil d'IA à travers des pratiques interrogatives n'est pas nécessaire à une forme d'augmentation. Dans la configuration AA, lorsque l'avis du radiologue concorde avec celui de l'outil d'IA, cela rassure le radiologue et lui permet de gagner du temps.

L'usage de l'outil d'IA a alors une fonction de triage de l'attention. L'incertitude propre à la situation clinique est réduite, l'opacité du résultat n'est pas remise en question, donc il n'y a pas lieu d'adopter des pratiques interrogatives.

Par conséquent, pour une prise de décision située, cette augmentation peut avoir lieu en dehors de la mobilisation des pratiques interrogatives (usage de l'outil en triage, comme dans la configuration AA), et lorsque les pratiques interrogatives ont lieu (usage de l'outil en filet de sécurité, comme dans la configuration β), elles entraînent des effets secondaires. Il nous semble donc nécessaire de développer notre analyse à l'aide d'un autre cadre théorique.

Un dépistage précoce du cancer du sein est essentiel pour réduire la mortalité de cette maladie. Les parcours de soins amènent les patientes à faire des examens mammographiques. Parmi les nombreux cas bénins, les radiologues doivent éviter des faux négatifs et détecter les anomalies dangereuses. L'organisation du service de radiologie vient en soutien à cette expertise, tout comme l'introduction récente d'outils d'IA pour l'analyse mammographique.

La collaboration entre le radiologue et l'outil d'IA s'approfondit selon un processus d'approvisionnement dans lequel l'expert métier développe des pratiques interrogatives professionnelles et sur l'outil. Si les pratiques interrogatives professionnelles sont adaptées de ses connaissances médicales, celles sur l'outil nécessitent de nouvelles connaissances, issues d'indicateurs statistiques et de l'expérience face aux réactions des outils d'IA.

La gestion de l'opacité des résultats de l'outil d'IA traduit un engagement dans la collaboration humain-machine de la part des experts métier. Cependant, une collaboration, même engagée, ne se traduit pas nécessairement par une augmentation des capacités des acteurs. Le sens de cette collaboration pourrait ne bénéficier que le résultat de la tâche.

Pour caractériser l'augmentation suscitée, et pour déterminer, s'il existe, un lien avec les pratiques interrogatives, nous continuons à adopter une démarche compréhensive afin de comprendre ce qui se passe dans l'interaction. Qui plus est, les données du terrain semblent indiquer qu'ici, l'outil d'IA se comprend le mieux comme un « outil », au lieu d'un médium ou d'un partenaire (Anthony, Bechky, et Fayard 2023), ce qui n'exclut pas que cet outil puisse avoir un effet dépassant une simple aide neutre, comme le sous-entendent les primes conceptuelles sur les usages des technologies (voir chapitre 2, 2.2.3).

2. Comprendre l'outil d'IA comme un outil de gestion

En considérant l'outil d'IA comme un outil de gestion (2.1), la littérature associée nous aide à mieux saisir ses usages et son potentiel (2.2), et ainsi à articuler les pratiques interrogatives avec le cadre d'analyse conformation/exploration (Moisdon 2005a) issu du courant sur l'instrumentation située (2.3).

2.1. Outil d'IA en tant qu'outil de gestion

Après avoir défini ce qu'est un outil de gestion selon la littérature éponyme (2.1.1), nous explorons comment ce concept s'applique à l'outil d'IA étudié (2.1.2).

2.1.1. Le concept d'outil de gestion

Origines du concept

Le concept d'outil de gestion contraste avec le paradigme décisionnel. Il confère une « épaisseur » à ces outils, leur attribuant une autonomie propre qui dépasse, voire contraint, la volonté des acteurs, au lieu de les voir comme des aides à la décision neutres pour accroître la rationalité des décideurs.

Les racines théoriques de cette approche puisent dans les travaux de Weber et Foucault, suggérant que la gouvernance moderne repose sur une panoplie d'outils de gestion (Deleuze et Foucault 1986; Foucault 2003). Les contributions successives de la littérature sur les outils de gestion ont affiné la définition ce concept, montrant comment ces outils s'articulent avec le reste de l'organisation.

Ce n'est qu'avec le développement des grandes entreprises à partir du milieu du XIXème siècle que les outils de gestion émergent véritablement (Chandler 1993; Lorino et Teulier 2005). Cette prolifération d'outils visant à faciliter l'action collective s'explique par plusieurs facteurs : la généralisation de la grande entreprise industrielle, l'intensification des contraintes concurrentielles, le développement des techniques de calcul et de gestion des données, ainsi que l'émergence de nouvelles compétences (Moisdon 1997, 17).

En effet, les outils de gestion répondent à au besoin organisationnel de rendre des situations simples pour prendre des décisions rapidement et à moindre coût. « *Toute situation de gestion est d'une complexité dépassant les capacités d'analyse des humains. [...] L'encombrement du temps des agents et l'urgence des choix poussent aux outils simples, aux idées simples, aux critères de jugement expéditifs.* » (Berry 1983, 5).

L'intérêt académique anglo-saxon pour les outils de gestion devient manifeste à la fin des années 1960, puis l'école française s'y intéresse à la fin des années 1970. Plus précisément, le Centre de Recherche en Gestion (CRG) et le Centre de Gestion Scientifique (CGS), en

collaboration avec des écoles d'ingénieurs, conjuguent la recherche opérationnelle avec une perspective organisationnelle et adoptent une approche politisée des outils de gestion (Aggeri et Labatut 2014).

La littérature s'intéresse alors aux effets structurants des outils de gestion qui dépassent les acteurs et finissent par entraver le bon fonctionnement de l'organisation. Malgré les intentions et la volonté des acteurs, ceux-ci se trouvent souvent à faire des choix induits par l'usage des outils de gestion. C'est « l'intendance qui commande » (Berry 1983, 7). Par exemple, les tableaux de bord ou les indicateurs exercent un effet structurant sur l'action bien plus puissant que les volontés politiques (Lascoumes et Le Galès 2004).

Le problème survient lorsque « *une contradiction surgit entre la fidélité de la description des phénomènes en cause et la simplicité qui économise du temps.* » (Berry 1983, 5). L'action générée par l'outil se retrouve en décalage avec la situation. Le gain de simplicité n'est bénéfique que sur une période relativement courte. Mais, en raison de son effet structurant, même lorsque sa pertinence est dépassée, il devient très difficile de remplacer l'outil de gestion.

Approfondissement du concept d'outil de gestion

Les fondements de la littérature francophone sur les outils de gestion sont consolidés dans deux ouvrages clés : *Technologie invisible* (Berry 1983) et *Du mode d'existence des outils de gestion* (Moisdon 1997). Ils proposent notamment une syntaxe (Moisdon 1997; de Vaujany 2006) qui nous permet de mieux saisir les concepts principaux.

La règle a une visée explicitement normative et oblige, même si elle peut être informelle (de Vaujany 2006). En revanche, l'outil de gestion (*management tool*) renvoie à l'extension des capacités de la main, comme un ustensile pour l'artisan. Il est conçu pour aider un groupe d'acteurs à réfléchir sur les fonctionnements dans lesquels leur action s'inscrit et à anticiper les évolutions possibles.

Ensuite, l'outil de gestion peut être considéré comme « *une abstraction, un modèle, petit ou gros, qui relie entre elles plusieurs quantités (des productions, des prix, des nombres de défauts, des effectifs de personnel, etc.)* » (Moisdon 2005b, 131).

L'outillage gestionnaire est défini comme « *un ensemble de raisonnements et de connaissances reliant de façon formelle un certain nombre de variables issues de l'organisation [...] et destinés à instruire les divers actes classiques de la gestion [...] : prévoir, décider, contrôler* » (Moisdon 1997, 7).

L'instrument de gestion résulte d'une opération de pensée de niveau supérieur, issue d'un processus d'intervention, générant des connaissances sur de nouveaux objets (Hacking, 1983). Les effets épistémiques qu'il produit modifient la perception de la réalité en construisant et

interprétant de nouveaux phénomènes. Il possède également une dimension politique, véhiculant une intention. Son but est de produire des effets, notamment en contrôlant les comportements. Autrement dit, l'instrument de gestion est « *le produit d'une opération de pensée intellectuelle (doctrine d'usage, schéma abstrait) et [il comporte] une dimension politique, implicite ou explicite, susceptible d'être révélée dans le cadre d'actions organisées et finalisées* » (Aggeri et Labatut 2014, 66).

Les théories, méthodes et modèles divers qui conduisent à la conception d'outils de gestion ne doivent pas être confondus avec ces derniers. Par exemple, la régression linéaire* n'est pas un outil de gestion, mais le devient lorsqu'elle est utilisée dans une situation particulière, comme lorsqu'elle relie les dépenses en marketing au volume des ventes pour une entreprise donnée (Moison 1997, 10).

La technologie de gestion fait référence à une catégorie plus large qui englobe les instruments de gestion. Elle désigne l'état de l'art des connaissances organisées dans un domaine. Par conséquent, la technologie de gestion dépasse l'instrument particulier pour englober un domaine plus vaste (comme les technologies de gestion dans le système de soins), composé de l'ensemble des instruments, techniques et connaissances. À l'instar de l'instrument, elle comporte aussi une intention politique (Moison 2005a).

Enfin, les dispositifs de gestion renvoient, dans une perspective foucauldienne, aux « *arrangements des humains, des objets, des règles et des outils [qui paraissent opportuns à un instant donné]* » (Moison 1997, 10-11). Cette notion désigne un arrangement hétérogène dans le temps d'instruments et d'acteurs en vue d'un objectif ou en réponse à une urgence stratégique (Foucault 2003). En intégrant la question de la temporalité, elle propose une perspective qui plonge l'analyse dans un environnement en mouvement, examinant ainsi l'évolution de la conception et les agencements liés aux outils.

2.1.2. Application du concept d'outil de gestion à l'outil d'IA

Le concept d'outil de gestion s'élargit au-delà des technologies matérielles conçues spécifiquement pour les gestionnaires. Selon Aggeri et Labatut (2014, 67) « *acquérir une dimension gestionnaire dès lors [qu'elle transforme] effectivement les relations et les apprentissages entre acteurs* ».

La distinction entre un outil de gestion et d'autres types d'outils ne réside ni dans la nature de l'organisation concernée, ni dans ses propriétés matérielles, ni dans les objectifs pour lesquels il a été conçu initialement. L'outil devient gestionnaire lorsqu'il participe aux trois grands actes élémentaires de la gestion : déléguer, évaluer, coordonner (Hatchuel et Moison 1993).

Par exemple, le compteur d'eau, mis au point pour mesurer le débit d'eau (observer un phénomène physique), devient l'instrument de calcul de la consommation d'eau par les

services de gestion et les entreprises de distribution (facturation, abonnements, etc.) (Hatchuel 2017).

Dans cette perspective, nous considérons que les outils d'IA observés dans RADO peuvent être vus comme des outils de gestion. Durant leur phase de développement, le concepteur entraîne l'outil avec des mammographies identifiées par des radiologues, créant un modèle empirique. Cet outil évalue ensuite les mammographies et identifie les risques de tumeurs, contribuant ainsi à la prise de décision clinique du radiologue, comme décrit précédemment.

De manière plus générale, les outils d'IA dans les soins courants peuvent être assimilés à des outils de gestion. Certes, chaque situation mérite d'être analysée plus en profondeur, mais nos entretiens avec les concepteurs révèlent que ces outils cherchent à répondre à la nécessité d'améliorer le travail des professionnels de la santé, et les entretiens avec les radiologues corroborent ces propos dans la manière dont ils décrivent leurs usages.

En effet, le système de soins connaît une mise en tension due à sa complexification, avec une spécialisation accrue des tâches, une demande en soins croissante, et des contraintes de ressources humaines et financières.

Dans ce contexte, les usages espérés des outils d'IA sont tournés vers la rationalisation des tâches, aussi bien administratives que propres à l'expertise métier des professionnels de la santé. Les outils d'IA sont donc développés pour être simples à utiliser et intégrés au flux de travail, selon les concepteurs que nous avons interrogés. Cette simplicité est cruciale pour l'adoption par les radiologues, qui témoignent de l'importance de la rapidité et de la facilité d'utilisation.

Cependant, les concepteurs qualifient ces outils d'aides, se dégageant ainsi de la responsabilité de la décision finale qui revient aux radiologues. Par conséquent, les résultats des outils d'IA doivent être compris par les experts métier qui les utilisent.

L'opacité des outils d'IA, surtout dans des contextes de jugement critique, nécessite une gestion consciente. Les exigences réglementaires et éthiques (CCNE et CNPEN 2022) créent une tension avec la simplification promise par ces outils. C'est pourquoi la littérature parle de collaboration humain-machine.

En outre, cette question met en avant la spécificité des outils d'IA en tant que technologie, mais également leur unicité en tant qu'outil de gestion : bien qu'ils soient le fruit d'un raisonnement et qu'ils comportent un modèle formalisé de connaissances produisant des résultats similaires à des experts, les résultats restent opaques pour les acteurs.

L'opacité des résultats des outils d'IA doit être traitée par un mode de gestion approprié, par exemple, avec des pratiques interrogatives, gardant ainsi l'outil d'IA au premier plan et en relation permanente avec l'utilisateur. Il reste cependant à saisir comment ce mode de gestion

participe à l'amélioration de la collaboration humain-machine. En le traduisant dans notre étude ce cas, nous cherchons à mieux saisir comment l'outil d'IA augmente le radiologue.

2.2. Repenser l'usage de l'outil d'IA

Le courant de l'instrumentation située, brièvement introduit (2.2.1), aide à conceptualiser l'augmentation des compétences des experts métier (2.2.2) et à articuler la notion de pratiques interrogatives avec cette augmentation (2.2.3).

2.2.1. *Le courant de l'instrumentation située*

La situation empirique des outils de gestion change dans les années 1980, entraînant de nouveaux courants de recherche. Les nouvelles technologies numériques permettent de nouveaux modes de production (Y. Wang et al. 2017), comme la personnalisation de masse (Pine 1993), et introduisent d'autres modèles organisationnels (Lampel et Mintzberg 1996), accentuant les contraintes concurrentielles et rendant l'innovation essentielle pour rester compétitif (Aggeri et Labatut 2014). L'essor de la production de services, plus difficilement formalisable, s'accompagne également d'une multiplication des critères de performance.

Ces perturbations entraînent un renouvellement massif des outils de gestion et de nouvelles approches théoriques. Leur rôle de contrôle se complète, voire s'efface parfois, en faveur de leur capacité à explorer la réalité. Il ne s'agit plus seulement d'apprendre sur l'organisation, mais aussi sur le contenu des métiers et des activités (Moisdon 1997).

La recherche met alors en avant les outils de pilotage de la transformation (Hatchuel et Molet 1986), leur rôle dans les processus d'innovation managériale (David 1998; de Vaujany 2006), leur capacité à changer la perception du monde (Hacking 1983), et leur rôle dans l'initiation de dialogues entre différents acteurs (Joerges et Shinn 2001).

Dans les années 1990, le courant de l'instrumentation située émerge, remettant en question la perspective cognitiviste des systèmes experts, dont l'échec a conduit à un « hiver de l'IA » (chapitre 1, 1.2.3).

Pionnière de ce courant, Suchman (1987) montre comment l'action humaine est continuellement construite en interaction avec l'environnement social et matériel en étudiant les systèmes experts de Xerox. Elle soutient que notre cognition est située, c'est-à-dire que le savoir et l'action sont inséparables du contexte. L'action est instrumentée par des artefacts. Par conséquent, l'instrumentation située propose une approche micro-analytique des situations de gestion et du rôle des instruments dans l'activité des organisations.

Remarquons que ce courant est interdisciplinaire, car l'émergence des outils informatiques attire l'attention de nombreuses disciplines pour explorer le lien entre les instruments et l'activité. Par exemple, Lorino reprend les travaux d'ergonomie et adopte une approche

pragmatiste et sémiotique en étudiant l'introduction d'un système de planification des ressources de l'entreprise chez EDF (Lorino et Peyrolle 2005; Lorino 2007).

À l'instar des concepts de cognition étendue et de cognition intégrée (Haugeland 1993; Clark 2008; Clark et Chalmers 1998), présentés dans la revue de littérature (chapitre 2, 2.1.9), nous retrouvons d'autres concepts similaires, dont celui de la cognition située (Suchman 1987) et de la cognition distribuée (Hutchins 1994).

Cognition distribuée (Hutchins 1994) : les processus cognitifs ne se limitent pas à l'esprit individuel, mais sont distribués à travers les membres d'un groupe, les artefacts, et l'environnement dans lequel ces membres interagissent, comme l'illustre l'exemple du cockpit d'un avion de ligne.

En effet, l'ensemble de ces travaux s'inspirent implicitement ou explicitement du pragmatisme (Peirce 1931; Dewey 1986). La situation se construit au cours de l'activité, dans la relation entre les sujets et les instruments, plutôt que de manière *ex ante* par les concepteurs (Lorino et Teulier 2005). Ils tendent également à converger autour du concept de l'activité médiatisée (Vygotski 1930).

Activité médiatisée (Vygotski 1930) : les interactions humaines avec le monde sont toujours médiatisées par des outils et des signes, et que ces médiations sont essentielles pour le développement cognitif et l'apprentissage.

Parmi ces analyses, Rabardel (2005) distingue deux types d'activités médiatisées. D'une part, l'activité productive est « *orientée vers l'atteinte des buts en situation ainsi que la configuration des situations* » (Rabardel 2005, 254). Elle opère dans l'espace des possibilités d'action, fait de délibérations, de compromis, d'initiatives et d'interventions par rapport à une situation donnée. Elle agit à court et moyen terme.

D'autre part, l'activité constructive est « *orientée vers l'accroissement, le maintien, la reconfiguration des ressources du sujet pour l'activité productive à venir* » (Rabardel 2005, 254). Elle agit à moyen et long terme. Par exemple, le sportif travaille son potentiel d'action lors de son entraînement (activité constructive), alors qu'il atteint des performances lors de compétitions (activité productive).

En s'inspirant de ces propositions ainsi que de recherches-interventions dans le système de soins, Moisdon (2005a) met en avant une double dimension des outils de gestion avec le cadre d'analyse conformation/exploration.

2.2.2. L'augmentation de l'expert à travers le cadre de conformation/exploration

En se concentrant sur l'activité médiatisée, le courant de l'instrumentation située est pertinent pour comprendre comment l'usage de l'outil d'IA augmente l'expertise métier. L'analyse de l'activité médiatisée est particulièrement intéressante en radiologie, une

spécialité médicale qui repose sur des instruments et des techniques d'imagerie. Lorsqu'une mammographie est insuffisante pour établir un diagnostic, le radiologue peut demander des clichés complémentaires, recourir à l'échographie, ou recommander une IRM voire effectuer une biopsie pour comprendre la situation. Chacune de ces techniques offre au radiologue des moyens supplémentaires pour formuler son diagnostic.

Dans ce système sociotechnique d'instruments et de signes, l'outil d'IA propose une nouvelle forme de lecture de la mammographie. Il ajoute à l'image mammographique un ensemble de scores influençant la manière de voir et de comprendre la situation.

Pour comprendre comment l'outil d'IA médiatise la situation, nous utilisons le cadre d'analyse conformation/exploration (Moisdon 2005a). Celui-ci explique le double rôle que peuvent jouer les outils de gestion, entre « *normer les comportements et [...] créer et propager du savoir* » (Moisdon 1997). En effet, au-delà des outils de gestion qui incarnent des savoirs pour contrôler, diriger et orienter l'activité, Moisdon (2005a) propose une « *doctrine d'usage [qui met] davantage l'accent sur les effets d'apprentissage induits par les outils que sur les effets de jugement et de conformation* » (Moisdon 2005a, 242).

Nous examinons maintenant comment ce cadre éclaire notre étude RADO selon chacune de ces deux dimensions.

L'outil d'IA dans son rôle de conformation

Dans un premier cas, l'outil joue un rôle de conformation en s'intégrant dans une activité médiatisée productive (Rabardel 2005). L'acteur se conforme à l'appareillage gestionnaire qui encadre son activité et l'optimise. Les outils de gestion instaurent une forme singulière de gouvernement, un « *gouvernement à distance par les chiffres* » (Moisdon 2005a, 240).

L'instrument exerce son action à travers un jugement : « *cette entité se comporte-t-elle dans le sens voulu ? Cette action est-elle efficace ?* » L'action se transforme alors en une opération de conformation : « *cette entité doit se corriger dans ce sens ; nous devons faire ce choix* ».

Empiriquement, cet usage est très majoritaire et dominant dans les observations réalisées : l'acteur se conforme à l'appareillage gestionnaire qui encadre son activité. Par conséquent, ce type d'usage augmente ses capacités d'agir mieux, en améliorant la qualité de la production, en réduisant le temps nécessaire et en diminuant les coûts.

Dans RADO, l'outil d'IA aide le radiologue à répondre aux exigences de production. Les deux usages identifiés, le triage et le filet de sécurité, impliquent respectivement un plus grand volume et une meilleure qualité du diagnostic. Lorsque l'outil d'IA et le radiologue ne détectent pas d'anomalie, l'usage de l'outil d'IA conforte (configuration AA), et lorsque l'outil d'IA voit une anomalie inaperçue par le radiologue (configuration β), l'usage de l'outil oblige le radiologue à regarder davantage. Nous observons que cette lecture des outils de gestion est

très répandue, tant dans les évaluations quantitatives des outils d'IA que dans les propos des acteurs interrogés.

Cependant, certains acteurs ressentent les limites du rôle de conformation de l'outil d'IA, axé sur la productivité, et se questionnent sur la véritable valeur qu'apporte l'outil. L'outil d'IA est limité, par construction, aux données mammographiques. Par exemple, lorsque les données d'entraînement sont réduites, l'outil d'IA est beaucoup moins fiable, comme l'a confirmé un radiologue spécialisé en pédiatrie (RAD10). De plus, l'analyse de l'outil d'IA est limitée à l'image mammographique, alors que les radiologues complètent leur analyse avec le dossier médical et, éventuellement, l'échographie. L'absence de contexte d'une mammographie peut rendre son évaluation beaucoup plus incertaine (Lebovitz, Levina, et Lifshitz-Assa 2021). Pire, l'outil ne sait pas comparer des mammographies d'une même patiente prises à différentes dates :

« Si l'outil d'IA comparait [avec les clichés précédents], ce serait bien... S'il faisait une superposition, du type trouver les sept différences, ce serait parfait. L'évolutivité c'est la seule vraie question. » (Jour 2, Dr G)

Les radiologues rapportent également une autre limite liée à la collaboration avec des experts métier. Une radiologue experte américaine nous a expliqué qu'elle employait l'outil parce que les concepteurs le proposaient gratuitement à son service pour évaluer ses performances. Cependant, elle ne serait pas prête à payer pour l'outil, car elle estime que son expertise suffit :

« [L'outil d'IA] ne remplace en aucun cas mon jugement clinique et mon expérience, c'est plus une redite. Je l'utilise seulement parce que les concepteurs nous l'ont proposé gratuitement. »³⁵(RAD14)

De même, le Dr S et la Dr G nous témoignent que pour eux, c'est surtout un outil de confort. Bien qu'il aide à agir mieux, l'outil d'IA reste limité en raison de leur expertise professionnelle.

Une autre radiologue, RAD12, qui évalue les outils d'IA pour son service avant de décider de les acheter, a partagé des observations similaires. Parmi les neuf outils d'IA évalués, quatre avaient des performances moindres qu'annoncées, car le contexte de développement de l'outil différait de la situation clinique de RAD12. Trois d'entre eux étaient jugés trop chers par rapport à ce qu'ils apportaient. Enfin, les deux restants étaient encore en cours d'étude.

³⁵ [The AI tool] doesn't replace my clinical judgment and experience at all; it's more of a redundancy. I only use it because it is free, the developers offered it.

Le jugement sur la valeur de l'outil d'IA est certainement biaisé par notre échantillon composé de radiologues expérimentés. Deux autres interlocuteurs proposent des pistes de réflexion. RAD3, un radiologue français dans une clinique privée, nous a évoqué qu'il avait décidé de payer un tel outil pour « *mettre au niveau tous les radiologues, surtout ceux qui viennent en remplacement* ».

Dans le même esprit, RAD7, un chef de service français, nous évoquait l'idée d'employer de tels outils dans les urgences (plus particulièrement pour les radiographies des os), lorsque des experts radiologistes ne sont pas disponibles. Ainsi, l'outil d'IA, dans son rôle de conformation, pourrait venir compléter l'expertise d'un professionnel de santé moins expérimenté ou qualifié.

L'outil d'IA dans son rôle d'exploration

Envisager uniquement le rôle de conformation des outils de gestion est réducteur face à l'ampleur des transformations qu'ils peuvent induire. Une autre dimension de l'outil de gestion doit être envisagée : celle de l'exploration (Moisdon 2005a). Le rôle d'exploration renvoie à une activité médiatisée constructive (Rabardel 2005). Ce rôle met l'accent sur les apprentissages générés par les outils, car ces connaissances procurent de nouvelles capacités d'action aux acteurs. Ce type d'usage augmente donc les capacités de l'acteur pour agir autrement.

Par exemple, le PMSI (programme de médicalisation des systèmes d'information) collecte de nombreuses données en routine permettant d'ajuster la redistribution des ressources, mais son rôle de conformation est faible car seul un faible pourcentage des ressources est effectivement alloué en fonction de ces données (Engel et al. 2000). L'impact majeur du PMSI se situe au niveau des apprentissages (Moisdon 2007; 2017), levant « *le voile d'opacité [du système de soins] s'est levé, en tout cas partiellement !* » (Moisdon 2005a, 244).

De même, une étude sur le dossier patient informatisé montre que cet outil joue principalement un rôle de confirmation mais qu'il existe des tentatives d'exploration, comme la mise en place de nouveaux indicateurs de performance (Béjean, Kletz, et Moisdon 2018).

Les apprentissages se déclinent en trois catégories (Moisdon 2005) : le niveau stratégique, le niveau organisationnel et le niveau opérationnel. En d'autres termes, le rôle d'exploration offre trois configurations d'usage : les outils de pilotage pour la mutation, les outils d'investigation des fonctionnements organisationnels et les outils d'exploration du nouveau (Moisdon 1997).

Concernant le niveau stratégique, les outils de gestion offrent des connaissances pour le pilotage, participant à « *la construction progressive de représentations partagées* » (Moisdon 1997, 35). Le changement produit est parfois inattendu, les outils pouvant être moins prescripteurs que prévu, générant des discussions pour mieux comprendre la situation (F.

Engel, Kletz, et Moisdon 1997). Par exemple, ils ont modifié la représentation de l'action politique chez les cadres de l'Administration (Jobert 1994).

Concernant le niveau organisationnel, les instruments de gestion mettent en lumière les déterminants essentiels de l'entité. Par exemple, le développement lent du PMSI est attribuable non seulement à des difficultés techniques, mais aussi aux hésitations quant à son utilisation interne (un outil conçu par l'administration centrale, mais dont les établissements disposeraient librement) ou externe (un outil de régulation pour l'allocation des ressources, par les tutelles locales). Les outils de gestion peuvent ainsi révéler des enseignements susceptibles de faire évoluer de nouveaux schémas organisationnels (Hatchuel et Moisdon 1997).

Concernant l'activité opérationnelle, les instruments de gestion permettent d'envisager de nouvelles façons de produire en orientant les métiers vers des transformations de leurs savoirs de base (Hatchuel et Weil 1992). Par exemple les modes opératoires ont été bouleversés par la mise en place du PMSI, avec comme surprise « *des professionnels de santé discutant ensemble des résultats économiques de leur secteur et essayant de comprendre les différences entre leurs établissements, pour finir par repérer des variations au niveau des protocoles de soins eux-mêmes* » (Moisdon, 2005a, 246). Les opérations de construction de nouveaux savoirs liés à l'introduction d'outils de gestion peuvent avoir lieu pour tout type d'acteur, et de façon plus ou moins autonome par rapport aux logiques d'action initiales (François Engel et al. 2000).

Cependant, le rôle exploratoire de l'outil de gestion reste une « doctrine d'usage » peu observée sur le terrain et résultant davantage d'effets inattendus que d'une mise en œuvre consciente (Moisdon 2005a). Par exemple, l'analyse du projet du PMSI révèle que les nouveaux savoirs restent majoritairement inactifs (Moisdon 2005a), en partie parce que les dirigeants comprennent mal l'activité des professionnels de santé que des réformes budgétaires soudaines risqueraient de déstabiliser (Moisdon 2012). En effet, le PMSI ne fournit pas d'indications claires sur les actions à entreprendre.

Plus généralement, « *l'outil de gestion, dans son mode de construction même, rassemble de nombreuses variables pour les condenser en des quantités manipulables [...] mais par là-même, il exclut que l'on puisse facilement retrouver les chaînes de causalité qui ont conduit aux résultats.* » (Moisdon 2005a, 247). Dans ce sens, l'outil d'IA exacerbe ce problème et se fait un cas extrême de l'outil de gestion. Réussir à développer le rôle exploratoire de l'outil d'IA pourrait donc contribuer à trouver des leviers d'action pour opérationnaliser cette « doctrine d'usage ».

Cependant, dans RADO, le rôle exploratoire de l'outil d'IA est difficile à saisir. Pour les dimensions organisationnelle et stratégique, nous ne pouvons que faire des extrapolations,

faute d'entretiens avec les gestionnaires et dirigeants. Les seuls propos entendus, par exemple sur le remplacement des radiologues, restent très spéculatifs.

Nos observations et entretiens avec les radiologues et concepteurs dévoilent des apprentissages sur l'activité opérationnelle, mais ces apprentissages restent très marginaux, sans chamboulement du mode opératoire.

En effet, l'outil d'IA semble peu lié à l'organisation. L'outil d'IA est rattaché au radiologue, mais non au manipulateur ou à la patiente – d'autres acteurs comme les gestionnaires et les dirigeants ne sont pas mentionnés. Nous n'identifions qu'une relation supplémentaire, celle entre le radiologue et le concepteur. Celle-ci est surtout présente au moment de l'adoption de l'outil. Cependant, lorsque vient le moment de comprendre les résultats de l'outil en routine, la relation s'appauvrit, avec des remontées partielles des radiologues intégrées dans un long processus de développement du produit et sans retour précis aux radiologues.

Ainsi, bien que certains soient persuadés que le travail du radiologue puisse être bouleversé (Topol 2019), à ce stade, l'outil d'IA est davantage utilisé comme une technologie additive, plutôt que transformative.

Nous explorons le rôle des pratiques interrogatives, en tant que mode de gestion de la collaboration avec le radiologue, dans l'expression du rôle exploratoire de l'outil d'IA. Autrement dit, nous cherchons un lien entre les pratiques interrogatives et l'augmentation des capacités d'agir autrement de l'expert métier.

2.2.3. Articuler les pratiques interrogatives avec le rôle exploratoire de l'outil d'IA

En explorant la collaboration humain-machine entre le radiologue et l'outil d'IA, nous avons identifié la gestion de l'opacité, réalisée à travers des pratiques interrogatives, comme nécessaire pour que l'usage de l'outil d'IA apporte une valeur ajoutée à l'expert métier.

Nous avons cherché à éclairer l'augmentation procurée par l'outil d'IA. En le comparant à un outil de gestion, nous avons utilisé le cadre de conformation/exploration (Moison 2005a), issu du courant d'instrumentation située pour examiner l'activité médiatisée.

Nous identifions une première dimension de l'augmentation de l'outil d'IA dans le registre de la conformation, qui optimise l'activité du radiologue. Les pratiques interrogatives jouent un rôle crucial en mobilisant les connaissances du radiologue pour comprendre le résultat de l'outil d'IA, assurant ainsi la qualité du diagnostic et limitant les biais potentiels. Plus précisément, l'opacité de l'outil est limitée par des pratiques interrogatives professionnelles, qui garantissent la qualité du diagnostic, et par des pratiques interrogatives sur l'outil, qui limitent le temps passé sur des cas dépassant les capacités de l'outil.

La seconde dimension de l'augmentation, liée au rôle d'exploration de l'outil d'IA, est censée permettre à l'expert métier de transformer sa pratique grâce aux apprentissages

suscités par l'outil. Cette dimension n'est pas fréquemment observée empiriquement (Moisdon 2005a; Béjean, Kletz, et Moisdon 2018), et nous ne l'avons pas identifiée dans notre étude de cas.

Pour expliquer cette absence, nous identifions le manque d'intégration de l'outil d'IA dans l'organisation. Cependant, en faisant un parallèle avec l'analyse du PMSI et des limites des outils de gestion, nous pouvons identifier une autre source, certainement dépendante au manque d'intégration : l'effet de l'opacité des résultats de l'outil d'IA. Par conséquent, les pratiques interrogatives, en tant que moyen pour gérer cette opacité, pourraient concrétiser les apprentissages.

Afin de comprendre leur potentiel, nous revenons aux fondements théoriques de la notion de pratiques interrogatives (Lebovitz, Lifshitz-Assaf, et Levina, 2022). Elles sont issues de la littérature sur la collaboration entre experts, qui identifie l'importance de comprendre et d'intégrer les points de vue des autres experts pour une action collective efficace (Carlile 2004; Maguire, Hardy, et Lawrence 2004; Hardy, Lawrence, et Grant 2005; Levina 2005), et qui transpose ces concepts à la collaboration humain-machine à travers le prisme théorique de la sociomatérialité (Barad 2003; Suchman 2007; Orlikowski 2007; Orlikowski et Scott 2008; Leonardi 2011).

Toutefois, nous remarquons que les pratiques interrogatives, bien qu'observées et décrites dans la littérature (Lebovitz, Lifshitz-Assaf, et Levina 2022), restent unidirectionnelles. Bien qu'elles soient construites sur une notion de collaboration entre experts, elles impliquent uniquement des radiologues face à des outils d'IA.

Nous suggérons donc de réfléchir à établir des pratiques interrogatives collectives, impliquant non seulement les experts métier entre eux, mais aussi des pratiques interrogatives de la part de l'outil d'IA pour comprendre le raisonnement derrière le diagnostic du radiologue, voire des pratiques interrogatives incluant d'autres acteurs comme les manipulateurs, les patientes, les gestionnaires et les dirigeants. Cette forme plus aboutie pourrait alors aboutir à des apprentissages, notamment grâce aux dialogues produits, et donc à exprimer le rôle exploratoire de l'outil de gestion. Nous réservons la poursuite de ces considérations au chapitre 6, consacré à la synthèse des études de cas et à la discussion théorique.

L'étude de cas RADO nous amène à mobiliser le cadre de conformation/exploration (Moisdon 2005a) pour identifier les usages de l'outil d'IA, selon qu'ils normalisent des pratiques, permettant aux acteurs d'augmenter leurs capacités à faire mieux, ou selon qu'ils procurent des apprentissages, permettant aux acteurs d'augmenter leurs capacités à faire autrement.

Cependant, RADO n'illustre que la première dimension : l'usage de l'outil d'IA optimise un processus en améliorant la qualité et parfois le temps du diagnostic, tout en rassurant le radiologue.

Pour aller plus loin et augmenter les capacités d'exploration, il serait intéressant de développer des pratiques interrogatives davantage collaboratives, intégrant ainsi un plus large éventail d'acteurs dans le processus de collaboration entre l'expert métier et l'outil d'IA.

Chapitre 5 : Usages des outils d'IA dans la recherche médicale

Notre seconde étude de cas décrit l'entraînement d'un outil d'IA, un processus visant à répondre à une question de recherche médicale (1). Elle invite ainsi à comprendre le rôle de l'outil d'IA, et des autres acteurs, dans la construction de connaissances (2).

1. Développement d'un outil d'IA pour répondre à une question de recherche médicale

Pour décrire le projet KOVAK, nous commençons par la narration du processus d'entraînement de l'outil d'IA (1.1), mettant en lumière l'usage de l'outil d'IA dans la construction de connaissances (1.2).

1.1. Narration du projet KOVAK

Pour narrer le projet KOVAK, nous procédons à une analepse, un retour en arrière pour établir le contexte et les origines du projet (1.1.1). Ensuite, la lecture peut se faire de manière chronologique, en suivant les étapes de 0 à 14, ou de manière thématique selon le modèle CRISP-DM* (voir annexe 7), comprenant la compréhension de la situation (1.1.2), la compréhension des données (1.1.3), la préparation des données (1.1.4), la modélisation (1.1.5), l'évaluation (1.1.6) et le déploiement (1.1.7).

La figure 12 illustre le processus global et les différentes étapes. Aussi, l'annexe 7 présente un exemple fictif suivant un processus linéaire du modèle CRISP-DM pour comprendre les étapes clé de l'entraînement des outils d'IA.

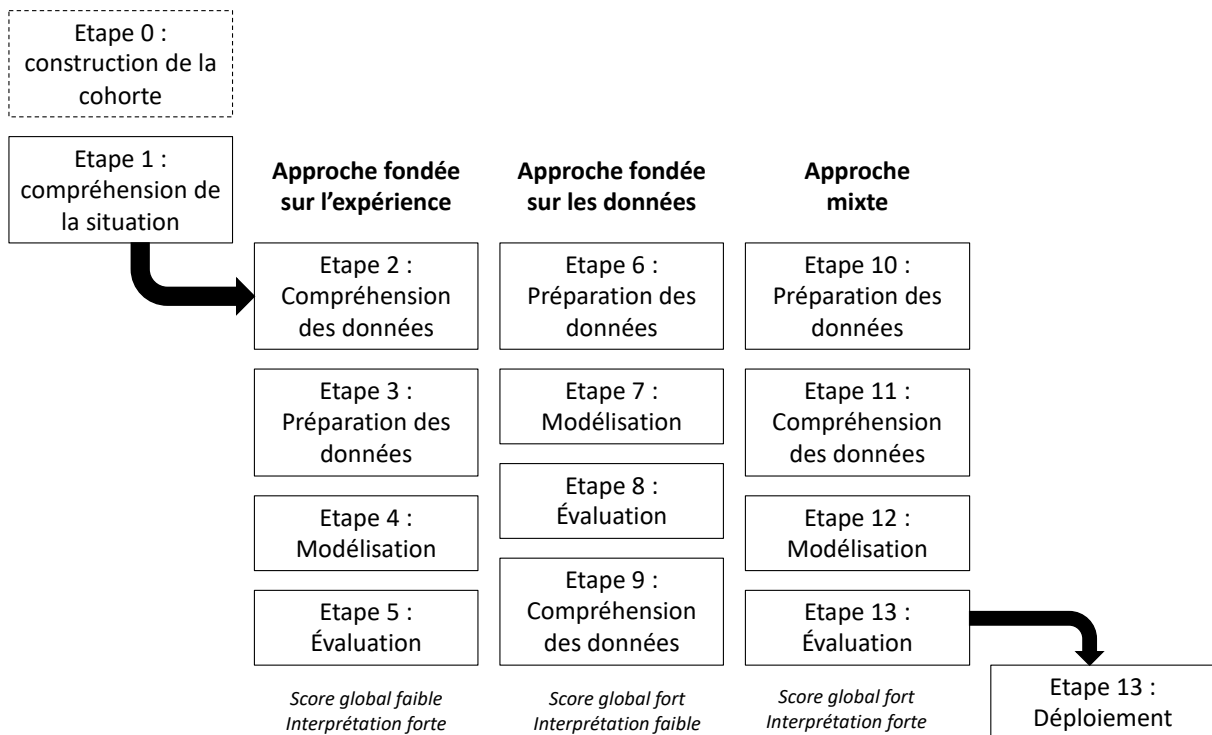


Figure 12 : Schéma de la narration du projet KOVAK selon le modèle CRISP-DM

1.1.1. Analepse : origines du projet KOVAK

Ces dernières décennies, la recherche médicale a révolutionné le traitement du cancer, notamment par le développement de campagnes de prévention généralisées contre le cancer du sein. Les avancées dans le diagnostic et les traitements ont amélioré les taux de survie au point que la recherche se concentre désormais sur les conditions de vie après le traitement.

L'immunothérapie* est l'un des traitements les plus récents, mais environ 30 % des patients développent des effets indésirables* allant de vomissements à des complications graves, voire au décès. Comprendre et prédire ces réactions est un enjeu majeur. L'identification des patients à risque permet aux oncologues de planifier des stratégies thérapeutiques adaptées.

Pour répondre à cette question, les chercheurs ont constitué des cohortes, une méthode qui remonte au XIXe siècle avec l'introduction de la biométrie et des approches statistiques (Phillips 2020). Les progrès dans la manipulation des données numériques (capture, stockage et traitement) ont permis de développer des méthodes statistiques plus sophistiquées. L'usage des outils d'IA a poussé ces avancées encore plus loin en automatisant en partie le traitement des données. Cependant, les outils d'IA de haut niveau nécessitent une collaboration étroite entre les médecins et les experts en données.

Le projet KOVAK est né du besoin de prédire quels patients développent des effets indésirables au sein d'une cohorte prospective* comptant plus de 3 000 patients traités par immunothérapie. Il repose sur la capacité à répondre à cette question à l'aide d'outils d'IA.

L'équipe de recherche est composée de cinq membres : CoExp, MedExp, DataExp, OrgExp et DatAdvice. CoExp est une épidémiologiste qui travaille sur la cohorte depuis 2020, mettant régulièrement à jour les données et les rendant disponibles aux équipes de recherche. MedExp est une oncologue qui collabore avec la cohorte depuis 2017 et enregistre les données des patients sous sa responsabilité. CoExp et MedExp ont déjà travaillé ensemble sur d'autres projets liés à la cohorte et sont à l'origine du projet KOVAK. Elles ont intégré DataExp, un expert en données, ainsi qu'OrgExp, un chercheur en santé publique. DatAdvice, une data scientist* senior, soutient DataExp et intervient dans le projet en cas de difficultés particulières.

Étape 1/14

La compréhension de la situation varie d'un acteur à l'autre et évolue tout au long du projet. Au début, MedExp est la personne la plus informée grâce à son expérience quotidienne et à sa participation dans la construction de la cohorte. En effet, le contact direct de MedExp avec les patients, sa formation de médecin et la lecture de la littérature clinique lui fournissent des connaissances tacites et explicites. Derrière les chiffres abstraits, elle peut « ressentir » intuitivement ce qui se passe. Elle génère des hypothèses explicatives et établit des liens entre les variables* utilisées et ses propres expériences et ses connaissances médicales. Sa motivation dans le projet KOVAK est double. À court terme, elle souhaite contribuer aux connaissances médicales en publiant un article scientifique. À long terme, elle espère que l'outil d'IA produit, ou du moins les connaissances générées, sauront mieux prédire les besoins des patients qu'elle traite.

L'épidémiologiste CoExp a moins d'expérience sur le terrain, car elle n'est pas confrontée à la pratique médicale. En revanche, sa formation et son expérience professionnelle lui procurent une bonne compréhension des enjeux de la cohorte, ainsi que du contexte dans lequel les données sont collectées et de la manière dont elles peuvent être utilisées (ce qui se superpose à l'étape de la compréhension des données).

Par conséquent, ensemble, MedExp et CoExp ont une vision très claire du but poursuivi et des moyens pour l'atteindre. Elles formulent une question claire qui perdure tout au long du projet : quels sont les facteurs associés à la survenue d'un effet indésirable lors d'une immunothérapie ? Elles s'associent à OrgExp et DataExp pour qu'ils traitent les données. Lors de leur première réunion, elles leur présentent la cohorte, la base de données avec ses différentes variables et le problème à résoudre. De plus, DataExp complète sa compréhension de la situation une fois qu'il reçoit la base de données, en effectuant des recherches sur Internet sur les termes médicaux qu'il ne connaît pas encore, en lisant la documentation technique sur la cohorte ainsi que les articles scientifiques publiés à ce sujet, et en interrogeant ponctuellement CoExp et MedExp.

1.1.2. Compréhension des données

Étape 0/14

La base de données de la cohorte est déjà construite avant le début du projet, notamment grâce à la contribution de MedExp et CoExp, ainsi que d'autres collaborateurs. Plus précisément, MedExp fait partie de la commission qui a choisi quelles informations collecter, une décision qui a été prise après une revue de la littérature sur l'immunothérapie. CoExp a régulièrement examiné les nouveaux comptes rendus médicaux et autres informations dispersées dans les systèmes informatiques pour enrichir la base de données de la cohorte.

Étape 2/14

Au départ, DataExp pense travailler sur l'ensemble des données de la cohorte complète. Cependant, MedExp et CoExp décident de lui fournir une base de données déjà organisée avec les variables qu'elles considèrent pertinentes pour la question de recherche. C'est le début de l'approche « fondée sur l'expérience ». En effet, la sélection des variables est faite selon les connaissances du terrain et du contexte dans lequel les données ont été produites. Ainsi, la base de données est composée d'une variable dépendante binaire (oui/non) sur les effets indésirables (grades II et plus) et de 37 variables indépendantes composées de 4 variables numériques, 32 variables qualitatives (différentes catégories) et une variable textuelle. Cette dernière n'est pas censée entrer dans un modèle. Elle a été créée pour se référer au contexte clinique du patient en faisant un résumé des antécédents médicaux indiqués dans les comptes rendus des médecins.

MedExp et CoExp considèrent que cette base de données réduite aidera DataExp dans la compréhension des données et leur prétraitement. En effet, DataExp exprime son étonnement en partageant que, d'après son expérience, rares sont les projets qui demandent si peu d'efforts lors de ces deux étapes. Elles constituent plus de la moitié du temps passé sur un projet.

Outre les recherches sur Internet et la lecture de la documentation technique, DataExp utilise des méthodes de description statistique pour recueillir des informations sur les variables et ainsi comprendre la structure des données. Par exemple, il s'intéresse au degré de corrélation entre les variables indépendantes. Ces informations l'aident à mieux comprendre les effets éventuels que les données pourraient produire et à prévoir un prétraitement adéquat. En suivant l'exemple précédent, si deux variables indépendantes trop corrélées sont introduites dans un même modèle, leurs effets risquent de s'estomper. Les informations qu'il tire des descriptions statistiques lui serviront aussi dans l'évaluation des modèles, en particulier comme des « garde-fous » pour vérifier que le comportement des variables n'est pas erratique.

Dans l'approche fondée sur les données, DataExp contourne le besoin de comprendre la signification médicale des données. Il met en œuvre des techniques de traitement automatique du langage naturel* pour créer des variables alternatives à la base de données fournie précédemment par CoExp et MedExp. Ces variables sont dérivées des mots et des expressions contenus dans la variable textuelle. Nous les nommons « variables-mot ».

Les variables-mot sont uniquement sélectionnées selon des critères statistiques. Un premier tri est effectué en fonction de leur fréquence d'apparition. Après une batterie de tests, les variables-mot gardées sont celles qui optimisent le plus les performances des modèles. Lorsque DataExp présente ce processus de sélection au reste de l'équipe, il précise que cette approche est « très empirique et naïve ». Il l'a seulement mobilisée pour « voir ce qui se passe ». S'il a choisi de présenter ces résultats, c'est parce que les scores globaux sont bien meilleurs (voir étape 8).

Au cours de la discussion subséquente, MedExp cherche à comprendre à quoi ces variables font référence. Elle retrouve alors, grâce aux comptes-rendus de la variable textuelle, à quelles situations cliniques les variables-mot sont associées. En s'imprégnant du contexte médical, elle émet des doutes quant à leur pertinence : certains mots sont employés dans des situations trop variées pour qu'elle puisse les interpréter.

OrgExp : J'ai du mal à comprendre. Si dans la variable textuelle d'origine [avec la synthèse des comptes-rendus], il y a le terme « cancer pulmonaire », c'est le mot « pulmonaire » que tu gardes ? Tu retiens le mot « cancer » indépendamment, mais tu peux ajouter le mot « pulmonaire » en plus de « cancer » ?

DataExp : Oui. Je garde le mot « pulmonaire ». Apparemment, le terme « cancer pulmonaire » n'a pas été retenu. Ça se trouve, il n'a même pas été présent dans la variable textuelle. Enfin, il a sûrement été très présent, je suppose, mais sans être statistiquement significatif, donc il n'a pas été retenu par le modèle linéaire*. C'est pour ça que je pense qu'il y a une sélection de mots à faire avant de lancer les algorithmes.

MedExp : C'est... [Hésitation] Je ne sais pas comment l'interpréter... « Pulmonaire », ça veut dire beaucoup... Je regarde dans le jeu de données à quoi ça correspond... C'est peut-être un emphysème pulmonaire, par exemple. Ça peut être... [Regarde dans la base de données] Ok. Oui, c'est ça. Ça peut-être une métastase pulmonaire, ça peut... Enfin... Attendez ! [Donne d'autres termes médicaux] En gros, il

t'a pris tous les cancers pour lesquels il y ait une [terme médical] de fait, ou une mutation du moins significative.

Étape 11/14

Dans l'approche mixte, DataExp commence par appliquer des techniques de traitement automatique du langage naturel améliorées grâce à des étapes de prétraitement (voir l'étape 10). Contrairement à l'approche précédente, il fournit simultanément les variables-mot et les mots les plus fréquemment associés à chacune de ces variables, ce qui offre une première couche contextuelle.

Ensuite, MedExp et CoExp filtrent les variables-mot qu'elles pourront exploiter ultérieurement. Elles cherchent à les interpréter, à comprendre les mécanismes sous-jacents liés à l'apparition d'effets indésirables. Ce travail interprétatif procède par étapes progressives pour situer la variable-mot dans un contexte qui lui donne du sens. D'abord, grâce à la couche contextuelle introduite par DataExp, la majorité des variables-mots deviennent facilement interprétables. Ces quelques mots supplémentaires permettent à MedExp et CoExp d'identifier les contextes médicaux associés. Lorsque cela n'est pas suffisant, MedExp et CoExp retournent voir la variable textuelle d'origine. Elles observent dans quelles phrases des antécédents la variable-mot apparaît. En dernier recours, MedExp et CoExp examinent le reste des données liées aux patients concernés par la variable-mot. Si MedExp et CoExp n'arrivent toujours pas à relier leurs connaissances à la variable-mot, celle-ci est exclue de la base de données finale.

1.1.4. Préparation des données

Étape 3/14

Dans l'approche fondée sur l'expérience, la base de données créée par MedExp et CoExp n'est pas simplement issue d'une sélection des variables de la cohorte initiale. Elles dérivent également de certaines variables d'origine de nouvelles variables pour mettre en avant ce qu'elles considèrent comme important. Par exemple, dans la cohorte, la variable « tabac » est graduée, avec des nuances apportées entre des patients qui subissent un tabagisme passif de la part de leur entourage, entre ceux qui fument modérément et ceux qui fument plus d'un paquet par jour. Dans la base de données finale, MedExp et CoExp décident de transformer ces nuances en deux valeurs : le patient fume ou non (variable binaire).

En résumé, la base de données transmise à DataExp n'a pratiquement plus besoin de prétraitements. Seules quelques manipulations restent à faire pour préparer les données au format adéquat pour les algorithmes d'apprentissage. Si cette réduction de la charge de travail

lui apparaît d'abord comme une aubaine, DataExp revient dessus après plusieurs modélisations et évaluations. Il exprime alors de la frustration en raison du manque de « marge de manœuvre ».

Étape 6/14

La base de données obtenue par la nouvelle approche comporte un nombre excessif de variables par rapport au nombre de patients, un problème connu sous le nom du « fléau de la dimension ». Le nombre de variables caractérisant les patients est généralement trop élevé par rapport au nombre d'individus enregistrés, ce qui invalide les résultats statistiques.

Pour réduire la dimensionnalité, il est possible d'effectuer une « sélection de caractéristiques », consistant à trouver un sous-ensemble de variables pertinentes, ou une « extraction de caractéristiques », c'est-à-dire de définir des variables plus pertinentes, à l'instar des transformations réalisées par MedExp et CoExp (voir étape 3).

Cependant, là où elles effectuaient ces modifications en s'appuyant sur leur expertise médicale, DataExp combine ces deux stratégies grâce à des moyens statistiques. Les techniques de réduction des dimensions* combinent les variables et créent des composites. Ainsi, les variables deviennent statistiquement pertinentes, mais le reste de l'équipe reste très sceptique quant à leur signification réelle.

MedExp : Donc là si j'ai bien compris, le modèle fait ressortir les variables dans toute la base de données initiale, y compris les mots clés dans les antécédents médicaux.

DataExp : C'est ça, c'est exactement. En considérant ces mots associés aux antécédents, il y a un gain de score de 0.1, ce qui est beaucoup. Il y a vraiment un décalage au niveau des performances.

MedExp : Le seul truc c'est que j'ai du mal à interpréter les mots. Mais après, on peut partir de ce modèle exploratoire en disant, voilà, le fait d'avoir une « lésion pulmonaire » ou une « légion métasternale », ça a un impact. En revanche, le mot « lésion » tout seul, celui-là j'ai un peu du mal à l'intégrer car il est extrêmement large. Pareil, le « nodule » pour moi ce n'est pas assez précis. Et avoir « tumorale », je ne suis pas trop à quoi ça correspond, c'est l'algorithme qui l'a sorti mais... Bon...

Étape 10/14

Pour l'approche mixte, DataExp envisage une préparation plus approfondie des données relatives aux variables-mots. Chaque mot sélectionné est associé à ceux trouvés le plus fréquemment dans la même phrase. Par exemple, « pulmonaire » est subdivisé en « lésion

pulmonaire », « métastases pulmonaires », etc. Comme dans l'approche précédente, ces variables sont soumises à de nombreux processus de réduction des dimensions afin de conserver les plus pertinentes d'un point de vue statistique, avant la sélection de MedExp et de CoExp.

1.1.5. Modélisation

Étape 4/14

La base de données issue de l'approche fondée sur l'expérience est un exemple de cas idéal pour les méthodes en data science. Le problème est très bien formulé, et la variable dépendante est parfaitement adaptée à l'apprentissage supervisé*. Dans ce type d'apprentissage, les algorithmes reçoivent à la fois les données d'entrée et de sortie (la variable dépendante concernant les patients développant des effets indésirables). Au moyen de nombreux tests statistiques, les algorithmes tentent de découvrir les relations qui relient ces données, créant ainsi un modèle empirique fondé sur les corrélations entre les différentes variables.

Tout d'abord, DataExp applique des algorithmes d'apprentissage linéaires, qui produisent des modèles empiriques sous la forme d'une simple équation mathématique : $\mathbf{y} = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \epsilon$, où \mathbf{y} représente la variable dépendante, β correspond aux coefficients de l'équation, \mathbf{x} aux variables indépendantes et ϵ à l'erreur. Cette forme simplifie la lecture et l'interprétation du modèle, permettant de comprendre l'effet de chaque variable et leurs relations. Cependant, DataExp ne parvient pas à obtenir des résultats statistiquement satisfaisants.

Face à ces résultats insatisfaisants, DataExp explore l'application d'algorithmes de régression non linéaire*. Les modèles empiriques générés par ces algorithmes ne peuvent pas être réduits à de simples équations mathématiques, ce qui les rend moins interprétables. Cependant, dans certains cas, ils offrent de meilleures performances. Malheureusement, les premiers essais de DataExp ne sont pas concluants, ce qui le conduit à entreprendre un processus de diagnostic pour améliorer les résultats.

Le code lui-même est complexe, car il est composé d'un ensemble de fonctions algorithmiques développées par différents auteurs. De plus, rétrograder dans le code ne serait que d'une utilité limitée, car il ne montre que le processus d'apprentissage (la série de tests statistiques effectués), sans révéler ce que l'algorithme apprend, c'est-à-dire le modèle qu'il génère.

Ainsi, pour réaliser leur diagnostic, DataExp analyse les résultats produits par les modèles. Il mobilise et renforce ses connaissances théoriques (en lisant des articles scientifiques sur ces modèles, en explorant des forums spécialisés, en interagissant avec DatAdvice, etc.). Il développe des outils graphiques pour éclairer la manière dont les variables se comportent

dans les modèles. Par exemple, grâce à la figure 13, il peut observer les interactions entre la durée d'administration (variables quantitatives) et la toxicité (variable binaire) : la partie en rouge indique une interaction significative lorsque la toxicité est présente et que la durée d'administration dépasse un certain seuil. Ces informations peuvent orienter le prétraitement en supprimant ces interactions significatives pour permettre à d'autres variables de s'exprimer. Malgré tous ses efforts, il dispose de peu de moyens pour améliorer les modèles.

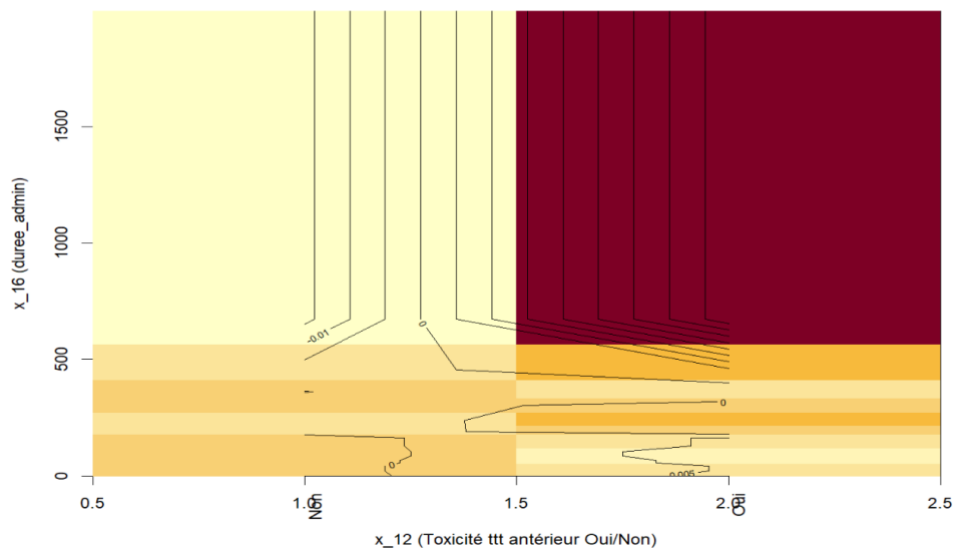


Figure 13 : Interaction durée d'administration et toxicité

Étape 7/14

Dans l'approche fondée sur les données, DataExp travaille avec les mêmes cinq algorithmes d'apprentissage (trois non linéaires et deux linéaires) pour comparer l'impact de la nouvelle base de données. Bien que les algorithmes d'apprentissage restent les mêmes, DataExp continue d'ajuster les « hyper-paramètres* » des modèles. Les hyper-paramètres sont des paramètres dont les valeurs sont utilisées pour réguler le processus d'apprentissage. Par exemple, l'algorithme LASSO*, utilisé par DataExp, comporte un hyper-paramètre qui permet de sélectionner un sous-ensemble de variables afin de prévenir le surapprentissage, un problème courant dans les situations de grande dimensionnalité.

Contrairement aux autres paramètres, dont les valeurs sont déterminées par les algorithmes d'apprentissage, la valeur d'un hyper-paramètre est choisie par un data scientist. Par exemple, les coefficients β dans les modèles linéaires sont calculés uniquement en fonction des opérations effectuées par les algorithmes d'apprentissage. La détermination de la meilleure valeur pour un hyper-paramètre se fait généralement par des essais et erreurs, grâce aux scores de performance obtenus.

L'approche mixte instaure un dialogue entre DataExp et MedExp qui se poursuit jusqu'à la phase de modélisation. DataExp commence par restreindre le choix des algorithmes d'apprentissage aux deux méthodes linéaires. D'une part, dans la phase d'évaluation précédente, les algorithmes d'apprentissage non linéaires n'ont pas montré de meilleures performances, ce qui laisse supposer que le phénomène en question peut être modélisé de manière linéaire. D'autre part, à ce stade du projet, l'objectif est de pouvoir discuter des résultats, ce qui nécessite des équations suffisamment simples.

Ensuite, DataExp présente les résultats des modèles sous forme de graphiques plus accessibles pour le reste de l'équipe (voir figure 14 ci-dessous). Ces graphiques servent de base à la discussion :

MedExp : *[en regardant la figure 14, partie en haut]* Ok, « cortico/baseline », ok ça, c'est parfait. Avoir un mélanome avec des méta-ganglionnaires, c'est associé. Avoir des « adénopathies/pulmonaires », ok, c'est vraiment plus clair pour moi de savoir qu'elles sont dans même phrase, je le comprends mieux. Avoir une « lésion osseuse » ... Ça reste sur des « métastase/osseuse », à voir comment je formule après. Mais, c'est clair pour moi. Avoir des « métastase/pulmonaires » ... Alors là, je comprends moins bien... Oui, par rapport à, « métastase/pulmonaire » et « métastase/ganglionnaire », il faudra voir comment on l'oriente.

DataExp : Je peux aussi les regrouper, enfin, je peux refaire des modifications, pour que ça ressorte mieux.

MedExp : Oui, là, ça me sert un peu moins. Effectivement sur les modèles du bas, j'ai un peu de mal à me représenter ce que ça peut signifier. La « mutation/négative », ça, c'est très large. Je vais avoir du mal à l'interpréter.

DataExp : Les derniers coefficients, c'est de l'aléatoire. C'est-à-dire que si je retourne le modèle, je ne suis pas convaincu que ces variables-là ressortent.

MedExp : Ah, ok. Parce que les trois dernières, je suis moins convaincue. Et tu ne trouves rien dans ce modèle qui réduit les risques de toxicité ?

DataExp : Avec le modèle LASSO*, oui, c'est ce qui va arriver. Alors que les modèles BIC en général, ils ont très peu de variables qui réduisent le risque.

MedExp : D'accord, ce n'est pas fait pour ça, quoi.

DataExp : C'est... j'ai l'impression, enfin, je ne pourrais pas l'expliquer théoriquement parlant, mais ça n'a pas l'air de fonctionner pour ça. En revanche, le modèle LASSO*, là, dans ce cas-là, il y a beaucoup plus de variables qui ressortent et qui réduisent les toxicités [*en montrant la figure 14, partie en bas*].

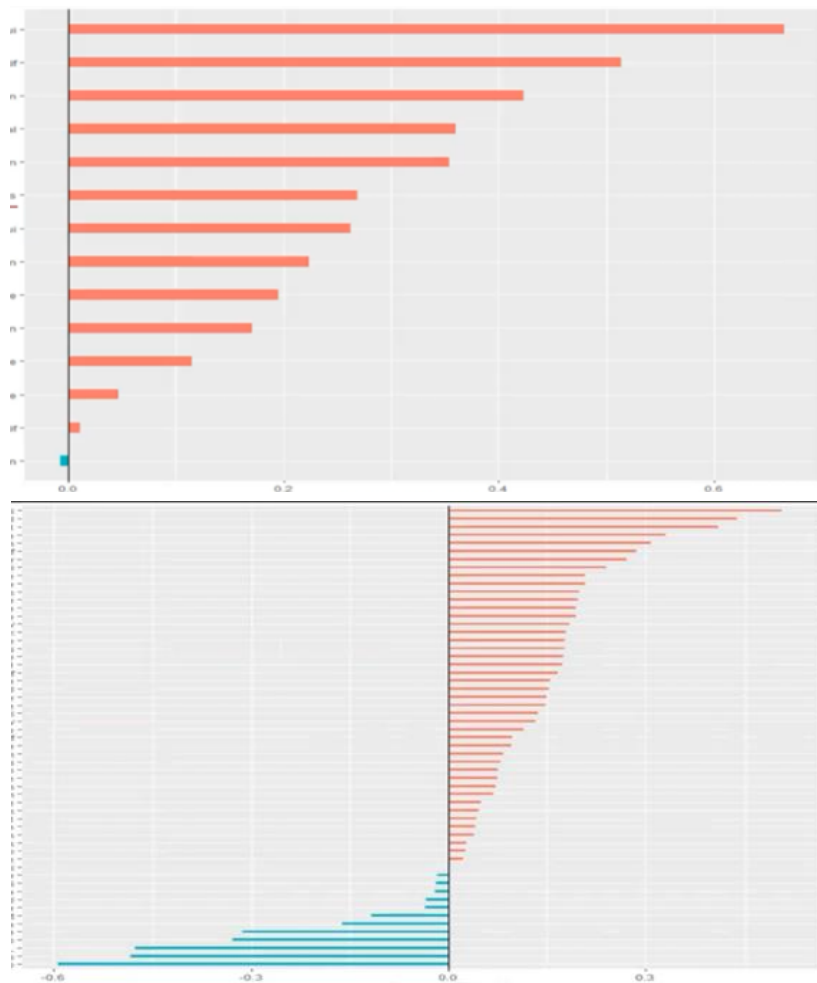


Figure 14 : Comparaison des modèles adaptatifs BIC (en haut) et LASSO (en bas) : influence des variables sur les effets indésirables (en orange les favorisant)

1.1.6. Évaluation

Étape 5/14

Dans l'approche fondée sur l'expérience, CoExp et MedExp souhaitent évaluer les modèles en utilisant le critère le plus courant dans la littérature médicale, le score AUC* (Aire Sous la Courbe) de la courbe ROC* (Receiver Operating Characteristic). De plus, ces indicateurs sont fréquemment employés en statistiques pour évaluer les performances de classificateurs binaires (dans le cadre du projet KOVAK, pour détecter les patients développant des effets indésirables). Plus le score AUC (l'aire sous la courbe ROC) se rapproche de 1, plus le modèle tend vers l'idéal, tandis qu'une valeur proche de 0,5 signifie une performance similaire à un choix aléatoire.

En utilisant ce critère, les scores obtenus par les algorithmes non linéaires sont encourageants (se rapprochant de 0,8), mais insuffisants pour une publication scientifique. DataExp se retrouve limité dans ses possibilités d'amélioration. Il sollicite alors l'aide de DatAdvice. Ensemble, ils réalisent que ces critères d'évaluation ne conviennent pas à la base de données. La catégorie de patients développant des effets indésirables est trop restreinte (environ 20 % du total de patients). Dans cette situation, le score AUC masque un grand nombre de faux positifs, c'est-à-dire de patients incorrectement associés au développement d'effets indésirables.

En effet, derrière ces valeurs, comme pour tout critère d'évaluation statistique classique, la courbe ROC et le score AUC sont liés à la sensibilité (la capacité à détecter un maximum de malades lorsque l'hypothèse est vérifiée) et à la spécificité (la capacité d'un test à détecter uniquement les malades lorsqu'ils sont malades). DatAdvice et DataExp décident donc d'utiliser une métrique d'évaluation plus adaptée à la situation : le score F1. Cependant, les algorithmes d'apprentissage ne parviennent pas à obtenir un score F1* satisfaisant, et leur amélioration reste limitée par la base de données.

Étape 8/14

En comparant les résultats des cinq modèles entre l'approche fondée sur l'expérience et celle fondée sur les données, DataExp souligne « un bond dans le score global ». Les nouveaux modèles affichent une amélioration du score global et une plus grande spécificité (le nombre de faux positifs diminue sensiblement). Face à ces résultats, l'équipe est convaincue de la nécessité de changer d'approche. Cependant, le manque d'interprétabilité des variables (voir étape 9) les conduit vers une méthode mixte pour tirer le meilleur des deux approches précédentes. De plus, ces résultats indiquent que les modèles linéaires sont tout aussi performants que les modèles non linéaires. Ce constat amène DataExp à ne conserver que les premiers dans la prochaine phase de modélisation (étape 12) :

DataExp : [Expliquant au reste de l'équipe ce qu'il a fait] J'ai aussi regardé les méthodes un peu plus orientées boîtes noires, avec moins d'explicabilité, et j'ai pu constater que les scores ont augmenté un peu, mais pas de beaucoup, pas suffisamment pour les prendre en compte

Étape 13/14

L'approche mixte produit des modèles avec des scores F1 suffisamment élevés pour envisager un article scientifique. Cependant, l'évaluation des modèles tient également compte de leur pertinence par rapport à la question initiale. En particulier, dans ce cas, il ne suffit pas d'avoir un bon classificateur ; les médecins souhaitent tirer des enseignements sur les mécanismes de décision de l'outil d'IA afin de les intégrer dans leur pratique médicale de manière indépendante. De plus, pour une publication scientifique, il est préférable de présenter des liens de causalité plutôt que des liens de corrélation. C'est pourquoi MedExp accorde une grande importance à l'interprétation des variables.

Elle analyse la pertinence clinique de chaque variable. Ce travail ressemble à la compréhension des données (voir étape 11), mais il va plus en profondeur. MedExp évalue les variables et les coefficients choisis par les modèles. Elle trouve des explications, en relation avec la présence des variables et leur importance relative, en s'appuyant sur ses connaissances médicales et le contexte des données. En l'absence de preuves de causalité plausible, elle demande à DataExp d'exclure la variable. Certaines variables sont également exclues en fonction de critères statistiques. Par exemple, le poids de certaines variables est jugé trop faible. L'équipe préfère simplifier l'outil d'IA, même au détriment de sa performance globale. Chacun de ces choix est contextuel. Outre les critères explicites, l'intuition et l'expérimentation contribuent à l'évaluation :

MedExp : D'accord. En tout cas ça me parle. [Reprend la liste de variables] Celle-là ouais, c'est compliqué. Ouais, je pense que celle-là, si tu la retrouves [en refaisant des tests], je ne saurai pas trop te dire à quoi elle correspond. Voilà, à la limite, tu peux regarder si ça correspond aussi à des trucs cérébraux. Mais, si tu vois que c'est très aléatoire, je pense qu'il vaut mieux l'enlever.

DataExp : Ok !

MedExp : Ensuite, [reprend la liste de variables], ça, c'est trop large, si jamais elle retombe trop souvent, tu nous diras. J'ai l'impression que la variable est assez faible en termes de coefficient, mais je ne sais pas si elle a un impact. [Reprend la liste de variables]. Là, c'est pareil. Si ça ressort beaucoup dans le modèle, j'essaierai de l'esquiver si possible.

Cela étant, ça n'a pas l'air d'être quelque chose de puissant... [*Reprend la liste de variables*]. Ok. Non, mais c'est vraiment top ! Je trouve que ça fait une jolie figure. Après, ce sera toujours sujet à discussion. Mais là, tu as un truc super avec un score vraiment top. Donc là, on arrive vraiment à un truc vraiment très sympa.

1.1.7. Déploiement

Étape 14/14

L'évaluation de l'outil d'IA s'est conclue de manière positive (voir étape 13). Avec quelques dernières retouches, les modèles se stabilisent. L'équipe entre maintenant dans la phase de déploiement. Une répartition des tâches est effectuée pour répondre au premier objectif (contribuer aux connaissances médicales en rédigeant un article scientifique). MedExp est chargée de rédiger la revue de littérature et la discussion, tandis que DataExp complète la partie méthodologique et les résultats. La collecte des données s'arrête à ce stade du projet. Cependant, il est envisageable que, en fonction des retours du comité de relecture, le projet entre dans de nouvelles itérations plus ou moins importantes.

Bien que le projet n'ait pas achevé l'étape de déploiement d'un outil d'IA pour les soins courants, les échanges au sein de l'équipe mettent en lumière l'usage envisagé. Lors de l'évaluation, le taux de faux positifs s'avère conséquent. Pourtant, les experts métier n'y voient qu'un inconvénient mineur, pourvu que les patients à risque soient identifiés, puisque le but est de limiter les faux négatifs. Pour les médecins, ce qui a de la valeur dans ce contexte est la capacité d'anticiper l'apparition d'effets indésirables avec un nombre de variables réduit et facile à percevoir en consultation.

MedExp : Tu as combien de pourcentage de faux positifs là avec ces modèles ?

DataExp : Un instant je vais revoir ça... Pourcentage de faux positifs... Je vais juste calculer... Parmi ceux que le prédicteur considère comme malades, il y a 7/10 qui sont des faux positifs.

MedExp : D'accord, ouais. Ok, ok. [*Réfléchit*] Dans un modèle parfait, si je comprends bien, le F1 score est à 1 et le modèle ne se trompe jamais, c'est ça ? Et quand il est à 0, il se trompe tout le temps ?

DataExp : C'est ça. Oui. Quand il est à 0, il se trompe tout le temps. Plus le F1-score est bon, plus le nombre de faux positifs diminue et le nombre de faux négatifs diminue aussi. Souvent, quand un algorithme admet un F1 score assez faible, c'est que qu'il a soit un défaut au niveau

de la gestion des faux positifs, soit au niveau des faux négatifs. Dans notre cas, ce sont les faux positifs.

MedExp : Mais ça à la limite c'est moins grave pour le clinicien. Par exemple, se dire que tu as identifié trop de patients à risque, ce n'est pas très grave. En revanche, si on a loupé des patients à risque, c'est plus embêtant sur le modèle.

[...]

OrgExp : La question c'est, l'utilité pour le clinicien. Finalement, ces résultats serviraient en routine ?

MedExp : Oui complètement. Parce qu'il identifie des facteurs hyper simples, hyper clean. Par exemple, le fait d'avoir un mélanome, d'avoir déjà fait une toxicité avant, d'avoir reçu plusieurs lignes de traitement auparavant, d'avoir ou non de la chimio... ce sont des variables hyper simples en consultation. Et donc en fait si on arrive avec ça à identifier un pourcentage de risques d'effets indésirables sévères, ça peut être sympa.

Je voulais te demander, DataExp, si c'est faisable avec ce genre de modèle... Je pose une question bête, est-ce que, pour un patient donné, tu peux arriver à prédire, en rentrant les données de ce modèle ?

DataExp : Oui, oui, bien sûr.

MedExp : C'est ça qui est intéressant. Il faut le pondérer, bien sûr, dans la discussion de l'article en disant « *bah ouais, on a un F1 score qui est de temps avec un risque de faux positif, donc ce n'est pas le graal.* »

De plus, les entretiens menés à l'Université de Pennsylvanie auprès d'acteurs déployant des outils d'IA dans le système de soins nous conduisent à évoquer trois défis principaux pour la suite du projet KOVKA. Premièrement, le cadre réglementaire complexe demande du temps pour répondre aux différentes exigences.

Ensuite, l'intégration avec les systèmes informatiques existants, souvent obsolètes ou fonctionnant sur des plateformes hétérogènes, peut nécessiter une personnalisation au niveau de chaque établissement de santé. De même, le manque de données normalisées en fonction des centres et des professionnels de soins peut entraver le bon fonctionnement de l'outil d'IA.

Enfin, son adoption dans les pratiques de soins peut être délicate et variable en fonction des habitudes de chaque médecin. Réussir à intégrer l'outil d'IA dans un flux de travail efficace demande encore une fois des adaptations au cas par cas. Autrement, il est nécessaire d'investir du temps dans la formation et de trouver des moyens pour dépasser les résistances au changement. L'ensemble de ces défis peut évidemment remettre en question certains aspects du fonctionnement de l'outil d'IA, ce qui entraîne de nouvelles itérations.

1.2. Usage de l'outil d'IA dans la construction de connaissances

L'apprentissage machine utilisant des modèles linéaires aboutit à des résultats moins opaques que l'apprentissage profond, mais les corrélations trouvées demeurent des abstractions que les acteurs humains doivent comprendre et interpréter. C'est pourquoi ils mettent en place des pratiques interrogatives (1.2.1).

Nous observons les augmentations apportées par l'usage de l'outil d'IA dans ce contexte (1.2.2) et comment elles s'articulent avec les pratiques interrogatives identifiées (1.2.3).

1.2.1. Mise en place des pratiques interrogatives

Dans le projet KOVAK, nous observons des pratiques interrogatives professionnelles et sur l'outil d'IA, avec les experts médicaux privilégiant les premières et le data scientist les secondes. Les experts médicaux gèrent l'opacité des résultats de l'outil d'IA en s'appuyant sur leur connaissance clinique, mettant en œuvre des pratiques interrogatives professionnelles. Cependant, elles remettent également en question le fonctionnement de l'outil d'IA, cherchant à comprendre le niveau d'incertitude statistique associé à un résultat, en collaboration avec le data scientist.

Le data scientist gère l'opacité des résultats sous un angle statistique, notamment lors de la compréhension des données dans les étapes 2, 9 et 11 (voir la sous-partie 1.1.3). Bien qu'il se familiarise avec les notions cliniques (voir l'étape 1, sous-partie 1.1.2), il ne développe pas de pratiques interrogatives professionnelles approfondies, préférant laisser cette tâche aux experts médicaux. Il leur laisse ainsi le rôle de distinguer les variables cliniques pertinentes issues de son approche fondée sur les données.

Notre étude de cas illustre comment le dialogue entre ces deux types de pratiques améliore l'interprétation des résultats de l'outil d'IA. Le data scientist commence par des pratiques interrogatives sur l'outil, avant de laisser place à des pratiques interrogatives professionnelles. Dans l'extrait d'interview ci-dessous, DataExp explique comment il a écrit le code pour analyser les résultats de l'outil d'IA issu de l'approche fondée sur l'expérience. Par exemple, il utilise un test de normalité* pour vérifier les conditions des lois statistiques sous-jacentes à l'outil d'IA. Après plusieurs tentatives infructueuses, il attribue l'échec des résultats à des facteurs externes à l'outil, comme l'hétérogénéité des patients.

Nous : Tu as réussi à détecter les variables qui donnaient des faux positifs ?

DataExp : Oui. J'ai déterminé les variables qui génèrent des faux positifs et ça ne change rien de les enlever. Les scores sont encore plus mauvais car il faut prendre en compte l'importance sur la performance de l'algorithme.

Du coup, il faut réussir à enlever les variables qui n'ont pas d'importance dans la prédiction, mais qui sont génératrices de faux positifs. J'ai fait beaucoup de bricolage. J'ai testé quelques choses [montre le code] mais c'était très exploratoire. Tu sors la boîte à outils et tu grattes le fond. Là, [montre le code] j'avoue que... Non, là, c'était une catastrophe. Le score de normalité, tout ça... Vraiment ce n'était pas important. Il y avait plein de trucs que j'ai supprimés ensuite.

Il y a quand-même certains facteurs explicatifs qui disparaissent ou qui apparaissent en fonction des données d'entraînement. Ça veut dire que les données ne sont pas stables. Enfin, une des interprétations qu'on peut avoir c'est que la base de données n'est pas très stable, il y a trop d'hétérogénéité. C'est ce que me répète souvent DatAdvice.

Inversement, les experts métier commencent par des pratiques interrogatives professionnelles qui conduisent à des pratiques interrogatives sur l'outil d'IA. Par exemple, lors du questionnement sur la variable « tumorale-pulmonaire » générée par l'outil d'IA avec l'approche fondée sur les données, l'experte médicale tente d'expliquer cette variable en utilisant ses connaissances cliniques. Elle éprouve des difficultés à interpréter la variable en raison de l'incertitude quant à l'origine possible de la tumeur.

MedExp : La variable « tumorale-pulmonaire », j'ai un peu de mal à comprendre si c'est la tumeur primitive, ou une localisation secondaire. C'est un peu difficile. Est-ce que sur chaque association, il y a des effectifs de patients qui correspondent ? Pour savoir ce qu'on peut garder.

DataExp : Oui, je peux récupérer les effectifs. À chaque fois que la variable ressort, l'effectif est au moins supérieur ou égal à 50.

MedExp : D'accord, ok.

En général, le dialogue entre les experts métier et le data scientist conduit à une progression des pratiques interrogatives utilisées pour comprendre le résultat de l'outil d'IA, passant des connaissances générales à des questions spécifiques, que nous schématisons en quatre étapes (figure 15).

Pratiques interrogatives sur l'outil générales

Utilisées au début par le data scientist pour valider la fiabilité statistique des résultats de l'outil d'IA.



Pratiques interrogatives professionnelles générales

Utilisées par les experts métier pour relier les résultats aux connaissances et mécanismes établis dans la littérature médicale ou tirés de l'expérience clinique.

Pratiques interrogatives sur l'outil spécifiques

Initiées par les experts métier et auxquelles le data scientist y répond avec des analyses plus spécifiques pour comprendre l'importance et la pertinence des résultats.



Pratiques interrogatives professionnelles spécifiques

Mises en œuvre par les experts médicaux en examinant directement les données liées aux patients affectés pour identifier les liens causaux potentiels.

Figure 15 : Schéma de la progression des pratiques interrogatives

Initialement, le data scientist mobilise des pratiques interrogatives sur l'outil « générales » pour vérifier les résultats d'un point de vue statistique. Ces pratiques prennent en compte des indicateurs généraux tels que la valeur de p (*p-value* en anglais) et les conditions de validité du modèle statistique. Malgré l'opacité du résultat, celui-ci semble correct.

Ensuite, les experts métier mobilisent des pratiques interrogatives professionnelles « générales ». Ils explorent les résultats sous un angle clinique afin de les relier aux mécanismes établis par la littérature ou l'expérience médicale. L'interprétation du résultat est rapide s'il est cohérent avec les savoirs des experts métier.

Cependant, lorsque ces pratiques interrogatives se révèlent insuffisantes, elles conduisent à des pratiques interrogatives sur l'outil « spécifiques ». Le data scientist y répond, même si les experts métier les posent en premier. L'objectif est de déterminer l'importance du résultat par rapport au modèle (sa force prédictive) ou aux données (le nombre de patients concernés). Le résultat est écarté s'il ne semble pas important, ou si l'effort pour l'interpréter semble

disproportionné, autrement dit lorsque l'opacité du résultat devient trop importante par rapport à son effet prédictif.

Lorsque le résultat ne peut pas être ignoré, les experts médicaux mobilisent des pratiques interrogatives spécifiques. Elles adoptent diverses stratégies pour comprendre le résultat. Par exemple, dans l'extrait ci-dessous, elles examinent les patients concernés par le résultat, espérant trouver des liens causaux qui auraient pu passer inaperçus.

DataExp : [Montre un tableau avec de nombreuses données] Ça, c'est un résumé statistique. Ça servira pour interpréter les variables. Par exemple quand vous prenez la variable « Tabac » tout en haut, au milieu, quand la modalité est élevée, c'est-à-dire quand un patient fume, il a 30% de chance de développer une toxicité grave. Pour rappel, il y a 20% de toxicité grave dans le jeu de données parmi l'ensemble des patients.

[...] [Montre la figure 16]

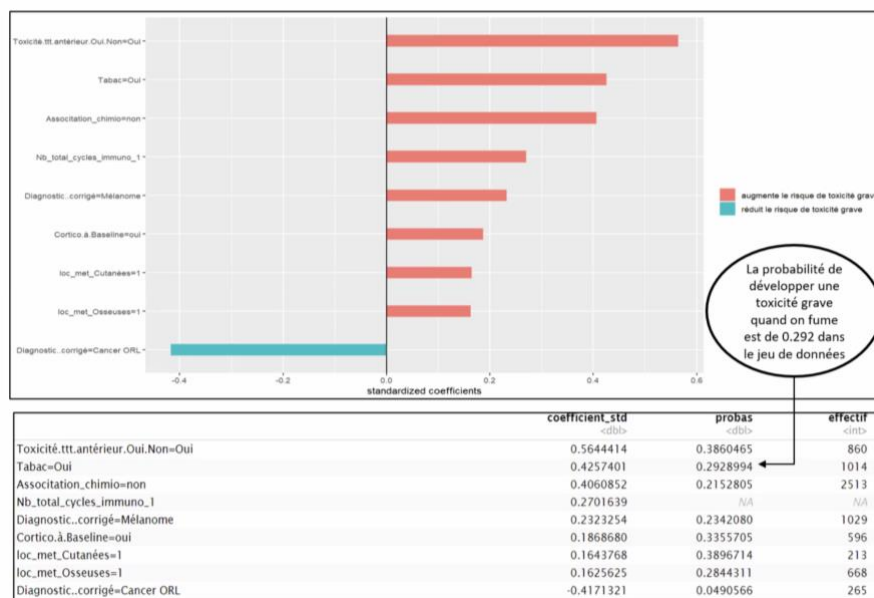


Figure 16 : Diapositive de DataExp sur le résultat de l'outil d'IA issu de l'approche fondée sur les données

La dernière fois, tu m'avais également demandé d'avoir un graphique plus intuitif pour l'interprétation, et je pense que celui-ci l'est un peu plus. En rouge, tu as les cas des variables qui augmentent le risque de toxicité grave, et en bleu, les cas où la variable diminue la probabilité de développer une toxicité.

MedExp : Oui c'est ça. Ce qui est bien, c'est qu'on voit effectivement les marqueurs sur le clinicien. On peut tout de suite se dire : « Attention

au patient qui fume, qui a déjà fait de la toxicité à l'immuno et ce n'est pas un cancer ORL car paf ça va tomber ! » Et donc ça, c'est pas mal.

OrgExp : Ça ne te surprend pas ?

MedExp : Non. La toxicité antérieure, ça, c'est un facteur de risque connu. Ça c'est sûr. Le tabac... *[Réfléchit]* Je pense qu'il y a un lien avec le cancer du poumon. Mais cela dit, c'est un antigène qui stimule aussi le système immunitaire. Donc ce n'est pas forcément étonnant. *[Réfléchit]* Et, on sait que ceux qui ont plus d'effets graves de toxicité, ce sont ceux qui ont été soumis à des agressions extérieures, donc forcément « poumon » et « mélanome » ça ressort dans le modèle statistique. En revanche, le fait d'avoir des « corticoïdes à la baseline », c'est un peu moins intuitif. C'est un peu plus compliqué à justifier. Après, c'est toujours pareil, il faut voir le nombre de patients concernés. Tu en as 596, c'est ça ?

DataExp : Oui, c'est ça. En tout cas, dans le jeu de données, il y a en a 33%... Enfin la probabilité de développer une toxicité quand on a des « corticoïde à la baseline » est de 33%. Ce qui est beaucoup plus élevé que les 20% de base. La colonne des probabilités, ce sont vraiment des statistiques descriptives sur le jeu de données. *[Il fait référence à la figure 16]*

MedExp : Ok, ça ce n'est pas forcément évident à interpréter. Il faut qu'on regarde avec CoExp, éventuellement les doses de corticoïdes pour ces patients.

Cette étude de cas confirme et complète les résultats du chapitre précédent (chapitre 4) concernant la notion de pratiques interrogatives. Les identifier dans le cadre du projet KOVAK témoigne de l'engagement des acteurs envers l'outil d'IA. L'usage de l'outil d'IA favorise une collaboration qui renforce les capacités de l'expert métier.

Nous retrouvons aussi les deux types de pratiques interrogatives identifiées précédemment : les pratiques interrogatives professionnelles et les pratiques interrogatives sur l'outil. Les experts métier ont tendance à mobiliser les premières, tandis que le data scientist privilégie les secondes, sans toutefois établir une division stricte. Nous distinguons également les pratiques interrogatives entre celles qui sont « générales » et celles qui sont « spécifiques ».

Enfin, l'étude de cas nous aide à comprendre le lien entre les pratiques interrogatives et le type d'augmentation suscité par l'usage de l'outil d'IA, comme nous l'explorerons ci-dessous.

1.2.2. Type d'augmentation procuré par l'usage de l'outil d'IA

Le projet KOVAK fait référence à des usages de l'outil d'IA dans la recherche médicale. Derrière l'appellation « outil d'IA », nous évoquons une série de modèles statistiques entraînés. Ces modèles commencent sous forme de code source. En définissant leurs hyperparamètres, le data scientist leur donne une direction pour croître. Ensuite, l'entraînement avec les données de la cohorte les fait évoluer jusqu'à ce qu'ils deviennent des modèles statistiques matures. Ainsi, les modèles (comme les modèles BIC, BIC adaptatif, etc.) se transforment pour devenir des outils d'IA possédant des équations empiriques adaptées à la situation simulée par les données. L'usage de l'outil d'IA dans la recherche médicale vise à trouver et améliorer cette équation empirique pour répondre à la question initiale.

Nous appliquons le cadre de conformation/exploration (Moisdon 2005a) pour éclairer la portée de l'augmentation procurée par ces usages. Dans le projet KOVAK, l'usage des outils d'IA répond à un processus de recherche, distinct d'un acte de gestion classique. Cependant, il implique la collecte, l'analyse et l'interprétation de grandes quantités de données, tout comme la prise de décision en gestion repose souvent sur l'analyse de données pour optimiser les stratégies et opérations. Il aide à identifier de nouvelles avenues de recherche ou à optimiser des expérimentations, ce qui correspond à une forme de planification stratégique.

L'outil d'IA médiatise la situation représentée par les données de la cohorte, offrant des perspectives influençant la prévision et la prise de décision. Nous examinons comment l'outil d'IA normalise et crée des connaissances, optimise la prise de décision et génère de nouvelles possibilités d'action.

Notre étude de cas montre que l'usage de l'outil d'IA entraîne une augmentation des capacités des experts métier dans le domaine de la conformation. L'extrait suivant illustre comment les experts médicaux cherchent à confirmer des connaissances existantes, car « parfois, il faut enfoncer quelques portes ouvertes pour le prouver ». Les résultats de l'outil d'IA renforcent leur compréhension à deux niveaux : chaque variable qui apparaît dans le modèle final confirme les connaissances sur un facteur explicatif, et le modèle final précise l'importance relative de ces facteurs dans l'apparition d'effets indésirables.

Ainsi, l'usage de l'outil d'IA normalise les connaissances, les renforce, les rend plus robustes et plus précises. Il augmente les capacités des experts métier à confirmer des intuitions et des savoirs empiriques, issus de la pratique.

OrgExp : Et avant d'évoquer la stratégie de publication, tu as des surprises dans les variables ? C'est quelque chose qui correspond totalement à ce que tu...

MedExp : Non, c'est totalement ce à quoi on s'attendait. On sait que les patients avec mélanomes développent un peu plus d'effets secondaires, et évidemment ce qui ont déjà eu des toxicités avant, ils vont en développer davantage, c'est assez intuitif. Mais parfois, il faut enfoncer quelques portes ouvertes pour le prouver. Et on sait que la durée d'administration a un lien avec la toxicité, donc c'est assez cohérent. Il y a le sexe masculin qui sort comme facteur protecteur, et il y a certaines méta-analyses qui montrent que la femme était un peu plus à risque de faire des toxicités.

OrgExp : D'accord, en fait, c'est sur la prédiction et le poids relatif des variables, une manière automatique d'avoir une information supplémentaire sur la prédiction.

MedExp : Oui, tout à fait.

Ce type d'usage dans la conformation se voit particulièrement dans la première partie du projet KOVAK, avec l'approche fondée sur l'expérience. Les experts médicaux emploient une logique très clinique, demandant au data scientist de confirmer les variables observées. À ce stade, les capacités des outils d'IA sont peu mises en avant et pourraient être substituées par des modèles statistiques plus basiques.

En revanche, l'augmentation des capacités des experts métier dans le registre de l'exploration reste marginale. La génération d'un modèle statistique montre la probabilité de corrélations entre les données, mais ne donne pas d'explications causales nouvelles. De plus, le modèle statistique se limite à des corrélations, sans pouvoir donner une explication causale inédite.

Cependant, l'usage de l'outil d'IA, en tenant compte des actions des autres parties prenantes, peut conduire à des découvertes plus générales. Par exemple, l'outil d'IA montre que les processus infectieux jouent un rôle plus important que prévu dans le risque de toxicité, ce qui n'est pas vérifié dans la connaissance médicale actuelle. Ainsi, l'outil d'IA peut tester de nouvelles hypothèses sans lien de causalité déterminé, orientant la recherche médicale subséquente.

Dans la troisième partie du projet KOVAK, l'approche mixte, l'usage de l'outil d'IA est davantage exploratoire. Les échanges entre experts et l'utilisation de variables fondées sur les données mais compréhensibles et actionnables par les experts de terrain génèrent un outil d'IA statistiquement valide et pertinent pour les professionnels de santé. Cette reconfiguration génère des résultats qui dépassent les attentes initiales des experts médicaux.

Il semble également possible de concrétiser davantage le potentiel d'exploration par d'autres usages des outils d'IA. Les moments clés des changements d'approche du projet KOVAK auraient pu donner lieu à des courants de recherche plus exploratoires. Par exemple, face aux scores fiables de l'outil d'IA (étape 5), le data scientist aurait pu opter pour une enquête plus approfondie sur l'hétérogénéité des patients, conduisant à de nouveaux apprentissages.

De même, les experts médicaux ont écarté les variables inconnues en faveur de celles déjà établies. L'exemple ci-dessous illustre comment la familiarité l'emporte sur l'exploration des différences entre les prévisions théoriques et les résultats de l'outil d'IA. Cependant, une recherche plus approfondie aurait pu modifier leur compréhension clinique.

DataExp : Chaque antécédent médical est composé de plusieurs phrases. Dans ce modèle, quand dans une phrase il y a « mélanome » et « ganglion » en même temps, ça va augmenter le risque de développer de la toxicité.

MedExp : D'accord. Pour l'item 3, « tabagisme – pa1 », dans la pratique ils sont souvent associés, donc est-ce qu'on ne peut pas regrouper cet item sous forme de « tabac » ? Ton modèle est rassurant, mais ça n'a pas de sens de dire que le « tabagisme » et « pa1 » sont associés. C'est logique, les deux vont toujours ensemble. Ou, dans le modèle, c'est vraiment l'association des deux ? La variable « tabac » seule n'apparaît pas ?

CoExp : J'avais déjà créé une colonne « tabac » avec des 0 ou des 1. Tu peux directement te servir de cette colonne.

DataExp : Si je suis arrivé à cette variable c'est grâce au nouveau jeu de données que j'ai construit, avec l'idée de chercher les mots qui apparaissent souvent avec « tabagisme » et il faut que ces mots-là soient pertinents statistiquement. Mais sur la pertinence clinique, moi, je n'ai pas vraiment l'expertise pour juger si c'est intéressant de considérer telle ou telle association. J'ai besoin de vous pour filtrer à ce niveau-là.

MedExp : Ok, donc ça n'a pas de sens de séparer « tabagisme » et « q1 » pour le modèle, car c'est le jeu de données qui s'est construit comme ça. Mais je pense qu'on peut le remplacer par la variable « tabac », tout simplement, pour que ça soit plus clair au niveau clinique.

En somme, le projet KOVAK montre que l'usage de l'outil d'IA dans la recherche médicale renforce les connaissances existantes (augmentation des capacités dans le registre de la conformation) plutôt que de susciter des découvertes (augmentation des capacités dans le registre de l'exploration). L'usage des outils d'IA précise la force des facteurs influençant l'apparition d'effets indésirables, consolidant et standardisant les connaissances pour produire des scores de risque pour chaque patient. Confirmer ces connaissances est fondamental pour que les oncologues gagnent en confiance dans le diagnostic, prennent de meilleures décisions et préparent des stratégies thérapeutiques adaptées aux patients susceptibles de développer des effets indésirables.

Cependant, le projet KOVAK montre une première avancée dans l'usage de l'outil d'IA dans le registre de l'exploration, grâce à la collaboration entre experts métier et data scientists. Ces découvertes restent limitées, et le projet aurait pu adopter des usages plus exploratoires en combinant différemment les capacités offertes par cette technologie à l'expertise médicale. De nombreuses explications peuvent être avancées : contraintes temporelles, culture de la recherche médicale et du système de soins, etc.

De même, le projet KOVAK montre une première avancée dans l'usage de l'outil d'IA dans le registre de l'exploration, grâce à la collaboration entre experts métier et data scientists. Cependant, ces découvertes, comme la mise en avant de corrélations pas encore discutées dans la littérature médicale, restent limitées. Le projet KOVAK aurait pu adopter des usages des outils d'IA plus exploratoires en combinant différemment les capacités offertes par cette technologie à l'expertise médicale. De nombreuses explications peuvent être avancées : contraintes temporelles, culture de la recherche médicale et du système de soins, etc.

Dans l'esprit de la « doctrine d'usage » des outils de gestion dans le registre de l'exploration (Moison 2005a), nous proposons des moyens de concrétiser davantage cette seconde dimension. Les pratiques interrogatives semblent pertinentes pour envisager les usages des outils d'IA dans le registre de l'augmentation de nouvelles capacités. Nous examinons comment les pratiques interrogatives s'articulent avec l'augmentation des capacités de l'expert métier dans le registre de la conformation et comment elles peuvent accroître les capacités dans le registre de l'exploration.

1.2.3. Articulation entre pratiques interrogatives et conformation/exploration

Deux hypothèses émergent pour expliquer la limitation de l'apprentissage lors de l'usage de l'outil d'IA dans la recherche médicale. D'une part, la cohorte utilisée pourrait restreindre les opportunités d'apprentissage. Elle pourrait ne pas contenir d'éléments « surprenants » en raison de la nature du problème ou de la manière dont elle a été constituée. Comme l'outil d'IA final affiche un score F1 de 0.6 (1 étant le maximum), il semble que d'autres variables devraient être prises en compte pour modéliser le phénomène.

Ce problème se généralise aux cohortes prospectives : les experts médicaux collectent des données liées à des hypothèses et à une manière de voir le problème, réduisant ainsi la possibilité de découvrir de nouveaux facteurs, puisque l'outil d'IA se contente de valider les connaissances existantes.

D'autre part, la manière dont les pratiques interrogatives sont mobilisées, que nous proposons de dénommer le « mode de gestion par normalisation », pourrait également restreindre les possibilités d'apprentissage. Cette configuration des pratiques interrogatives, qui s'apparente à un entonnoir allant du général au particulier, tend à réduire les principaux résultats de l'outil d'IA à des connaissances établies (figure 15).

Les pratiques interrogatives sur l'outil agissent comme un rasoir d'Ockham, écartant les résultats en dehors de certains critères prédéfinis. De même, les pratiques interrogatives professionnelles relient les résultats de l'outil d'IA aux connaissances existantes et les rejettent autrement, sauf s'ils sont jugés trop importants pour être ignorés. Les pratiques interrogatives professionnelles spécifiques, qui sont les plus susceptibles de générer de nouvelles connaissances, ne sont invoquées qu'en dernier recours.

Les experts métier utilisent ces pratiques pour mieux comprendre les résultats et les relier à des mécanismes connus, mais ils les écartent en l'absence d'explications satisfaisantes. L'exception à cette règle est rare et se produit lorsque les pratiques interrogatives sur l'outil fournissent suffisamment de certitude statistique pour que l'ensemble de la communauté médicale les accepte.

Nos deux hypothèses, la limitation due à la cohorte et celle liée à la manière de mobiliser les pratiques interrogatives, peuvent coexister. Le mode de gestion par normalisation des pratiques interrogatives souligne les opportunités d'apprentissage significatives lorsque les résultats de l'outil d'IA sont remarquables mais demeurent sans interprétation classique. Une autre cohorte, recueillant des variables supplémentaires, pourrait révéler des éléments surprenants rendant le modèle final plus pertinent. À l'inverse, d'autres configurations des pratiques interrogatives pourraient favoriser davantage d'apprentissages, pour autant que la cohorte offre de telles possibilités.

Pour favoriser les apprentissages, nous proposons deux autres modes de gestion. Le premier mode, « par approfondissement », se concentre sur les résultats de l'outil d'IA qui ne répondent pas aux critères statistiques. Il est pertinent lorsque les pratiques interrogatives sur l'outil mettent en lumière une limite statistique. Dans ce cas, nous suggérons d'utiliser des pratiques interrogatives spécifiques sur l'outil pour mieux caractériser le résultat, puis d'appliquer des pratiques interrogatives professionnelles pour comprendre cette limite du point de vue pratique.

Pour illustrer ce mode de gestion, reprenons l'étape 5. Le data scientist est confronté à des scores fiables de l'outil d'IA et cherche à les comprendre pour améliorer le modèle. Après avoir utilisé de nombreuses pratiques interrogatives sur l'outil pour caractériser les résultats, il mobilise des pratiques interrogatives professionnelles. Il en conclut que le problème réside dans l'hétérogénéité des patients. Plutôt que de se contenter de cette conclusion, il pourrait envisager de réintroduire des groupes de patients spécifiques dans l'entraînement de l'outil d'IA pour vérifier si cette démarche améliore les modèles, contribuant ainsi à l'acquisition de connaissances sur les profils types.

Le second mode de gestion, « par opposition », s'intéresse aux divergences entre les résultats de l'outil d'IA et les connaissances établies. Lorsque les pratiques interrogatives professionnelles remplacent un résultat de l'outil d'IA par des connaissances déjà établies, le mode de gestion par opposition suggère d'utiliser des pratiques interrogatives sur l'outil pour identifier les différences entre la prévision théorique et le résultat empirique. Ensuite, des pratiques interrogatives professionnelles spécifiques permettraient d'explorer ces différences plus en détail.

Reprenons l'exemple précédent pour illustrer ce mode de gestion. Les experts médicaux excluent la variable « tabagisme – pa1 » au profit de la variable préétablie « tabac ». Les pratiques interrogatives professionnelles les amènent à normaliser le résultat de l'outil d'IA en utilisant des connaissances déjà existantes. Cependant, en mobilisant des pratiques interrogatives sur l'outil, les différences entre ces deux variables pourraient être mises en évidence. Il se pourrait que l'effet du tabac dépende d'autres données à prendre en compte, ou que certains profils de patients soient plus sensibles à cet effet.

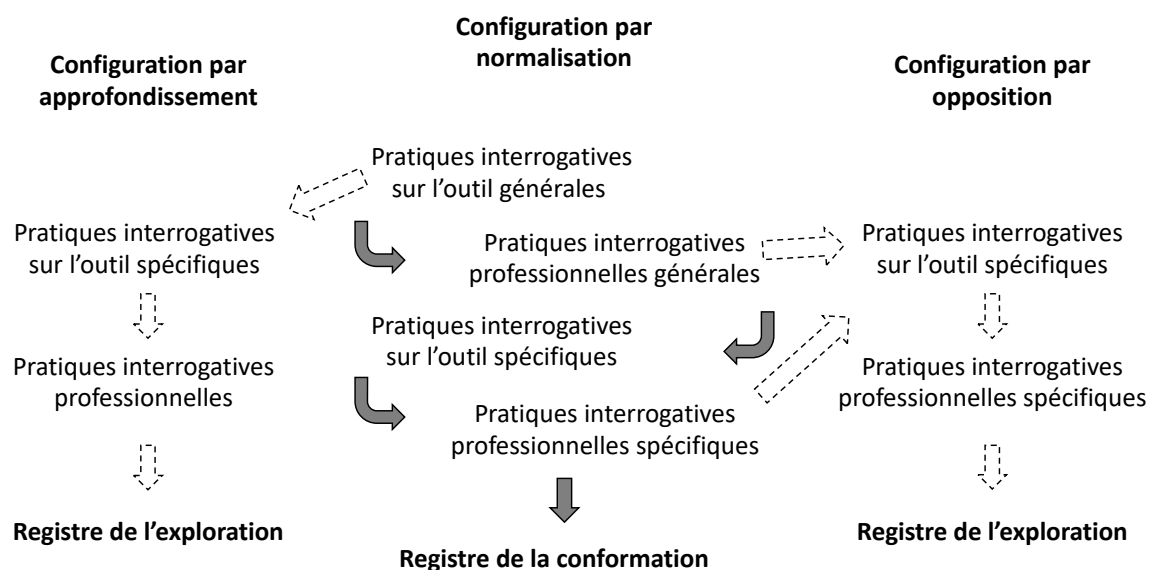


Figure 17 : Schéma des possibles progressions des pratiques interrogatives

En conclusion, le mode de gestion par normalisation tend à limiter l'apprentissage et à renforcer les connaissances existantes. Pour favoriser l'exploration et l'apprentissage, il est essentiel d'adopter des modes de gestion par approfondissement et par opposition, en mobilisant des pratiques interrogatives spécifiques et en remettant en question les connaissances établies lorsque les résultats de l'outil d'IA divergent de ces dernières.

En explorant l'usage d'un outil d'IA dans la construction de connaissances, nous examinons les pratiques interrogatives mises en place par les experts métier et les data scientists pour interpréter et comprendre les résultats abstraits générés. Nous caractérisons plus en détail les pratiques interrogatives professionnelles et celles sur l'outil en distinguant celles qui sont générales de celles qui sont spécifiques.

Ensuite, comme illustré dans les figures 15 et 17, nous associons ces pratiques interrogatives avec le cadre de conformation/exploration. Nous identifions ainsi trois types d'articulations : un mode de gestion par normalisation, relevant du registre de la conformation, ainsi que des modes de gestion par approfondissement et par opposition, qui s'inscrivent dans le registre de l'exploration de nouvelles connaissances.

À l'instar de l'étude de cas précédente, RADO, dans le projet KOVAK nous remarquons que la notion de pratiques interrogatives et le cadre d'analyse conformation/exploration tendent à mettre l'accent sur l'humain. Les pratiques interrogatives sont mobilisées par l'expert et il s'agit de l'augmentation de l'expert.

Cependant, dans ce second cas, l'outil d'IA adopte une posture différente. Il ne « double » pas la lecture du radiologue. Au contraire, son rôle dans la construction de connaissances vient compléter celui des autres acteurs. Par conséquent, les données empiriques nous amènent à adopter un cadre théorique qui envisage l'outil d'IA comme un autre acteur, au même titre que les acteurs humains (Anthony, Bechky, et Fayard 2023).

2. Rôle de l’outil d’IA dans la construction de connaissances

Pour comprendre le rôle de l’outil d’IA dans le projet KOVAK, nous faisons appel au processus d’enquête de Charles S. Peirce en raison de ses travaux épistémologiques qui se focalisent sur la construction sociale des connaissances (Lorino 2014). Ceux-ci apportent en effet des contributions majeures aux études sur les processus, les pratiques, le dialogisme, la performativité discursive et la sociomatérialité (Lorino 2014; 2018).

Plus particulièrement, nous abordons le processus d’enquête à travers sa théorie des inférences et son concept de quasi-esprit (2.1). Ce nouveau cadre théorique nous aide à voir l’outil d’IA dans la recherche médicale comme un acteur de l’enquête détenant un rôle spécifique, celui de produire des inférences inductives (2.2).

2.1. Le processus d’enquête selon Charles S. Peirce

Philosophe et logicien américain du XIXe siècle, Charles Sanders Peirce a contribué de manière significative à divers domaines, dont celui de la philosophie pragmatiste. Une courte biographie est proposée en annexe 9 afin de contextualiser son parcours intellectuel et mieux saisir l’envergure de ses travaux.

Pour éclairer notre étude de cas, nous examinons la théorie des inférences (2.2.1) et le concept de quasi-esprit (2.2.2) qui convergent vers le concept d’enquête pragmatiste – un concept qui sera développé et attribué par la suite à Dewey (Lorino 2018). Une brève introduction sur le fondement de la pensée peircienne est proposée ci-dessous en préambule.

Peirce cherche à dépasser la philosophie cartésienne et la division entre l’esprit (*res cogitans*) et le corps (*res extensa*) qui affirme que les idées sont des entités distinctes, et supérieures, à la matière. Pour Peirce, les phénomènes mentaux et physiques sont des aspects d’un même processus continu de production de sens (la sémiotique). Il considère la pensée comme fonctionnant par l’intermédiaire de signes, et cette perspective mène à sa théorie de la sémiotique triadique.

La sémiotique triadique :

Selon Peirce, la signification est une relation triadique entre trois éléments :

- **Le Representamen (R)** : le signe, un objet perceptible ou tangible, qui représente quelque chose d’autre, comme une carte routière.
- **L’Objet (O)** : ce que le signe représente, comme un concept ou le réseau de routes dans l’exemple précédent.
- **L’Interprétant (I)** : le signe qui émerge selon la situation. Par exemple, la carte routière fait émerger des chemins (pour un conducteur), des flux de marchandises (pour un logisticien) ou

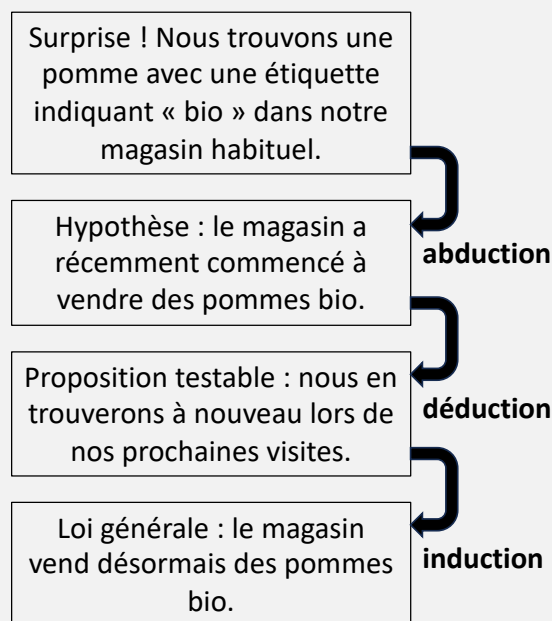
des processus d'enquête et de correction d'erreurs, grâce à des inférences déductives, inductives et abductives (Peirce 1878e) que nous présentons dans un premier temps (2.1.1).

Brièvement, les trois mécanismes de construction de la connaissance sont :

- L'**abduction** : une hypothèse est formulée pour expliquer des observations surprenantes. Elle seule crée de nouvelles connaissances. Par exemple, si nous trouvons une pomme avec une étiquette indiquant « bio » et que nous ignorons que des pommes bio étaient vendues ici, nous pourrions alors abduire que ce magasin a récemment commencé à vendre des pommes bio.

- La **déduction** : un processus logique où une conclusion nécessaire découle de prémisses et qui permet de traduire des hypothèses en propositions testables. Elle développe les connaissances. Par exemple, si notre magasin a commencé à vendre des pommes bio, alors nous en déduisons pouvoir en trouver lors de nos prochaines visites.

- L'**induction** : un principe général est tiré de cas spécifiques. Elle consolide les connaissances. Par exemple, si lors de nos prochaines visites, les pommes sont bio, nous induisons que le magasin vend désormais des pommes bio.



Le processus d'enquête se fait en mobilisant ces inférences. Or, celles-ci ne sont pas exclusivement mobilisées par des acteurs humains. Le processus de production de connaissances se fait à travers ce que Peirce appelle des quasi-esprits, qui incluent les outils d'IA (2.1.2).

Le concept de **quasi-esprit** (ou *quasi-mind* en anglais) se réfère à l'idée que la signification et l'interprétation ne sont pas seulement des propriétés des esprits humains, mais sont

également attribuées à des systèmes ou des processus plus simples et mécanisés. Par exemple, un thermostat est un quasi-esprit parce qu'il interprète des stimuli (température) et effectue une action appropriée (allumer ou éteindre le chauffage).

Nous rejoignons ainsi Skagestad (2004) et P. Steiner (2013) qui mobilisent une approche post-cognitiviste sur l'intelligence (Haugeland 1993; Clark 2008; Clark et Chalmers 1998) pour comprendre les outils d'IA. Cependant, là où leurs travaux restent conceptuels et ancrés dans la tradition philosophique, nous adoptons ce cadre pour éclairer nos observations du projet KOVAK, à la manière des études sur les usages des outils d'IA.

2.1.1. *Épistémologie peircienne : inférences abductives, inductives et déductives*

D'après Peirce, le processus de l'enquête pragmatiste, qu'il appelle dans un premier temps la méthode scientifique, repose sur la théorie des inférences, qui décrit trois mécanismes nécessaires et suffisants (l'un ne pouvant être la somme des deux autres) pour produire de la connaissance : l'abduction, la déduction et l'induction. Cependant, avant d'exposer ces mécanismes, il décrit le point de départ de l'enquête : l'émergence d'un doute qui pousse la pensée à produire une connaissance.

La pensée pour construire une croyance

Pourquoi pensons-nous ? D'après Peirce, nous cherchons à répondre à un doute. La pensée émerge de l'émotion du doute : « *l'irritation produite par le doute nous incite à faire des efforts pour atteindre l'état de croyance. Je nommerai cette série d'efforts l'enquête.* »³⁷ (1877, 8). Cette enquête a pour seul objectif d'établir une croyance, synonyme pour lui d'opinion et de connaissance.

La croyance, en tant que résultat de la pensée, est donc issue d'un processus de signification (de sémiotique triadique), impliquant que la croyance se forme au niveau collectif, qu'elle produit une action, et qu'elle finit en une habitude. La pensée suit une cascade sémiotique, ce qui fait appel à différents interprétants, et donc à un réseau de significations partagées et d'expériences collectives. Ensuite, la croyance est intimement liée à l'action (absence de dualité idées/action). Peirce affirme même que croire, c'est agir : *le sentiment de croyance est*

³⁷ *The irritation of doubt causes a struggle to attain a state of belief. I shall term this struggle inquiry, though it must be admitted that this is sometimes not a very apt designation*

une indication plus ou moins sûre qu'il s'est enraciné en nous une habitude d'esprit qui déterminera nos actions. »³⁸ (1877, 5). Ainsi, avec le temps, la croyance devient une habitude.

Peirce s'intéresse également à la manière dont nous pensons. Il distingue quatre « méthodes » (1878a) pour établir des idées. D'abord, la *méthode de ténacité* consiste à choisir une opinion et à la maintenir coûte que coûte. Ensuite, la *méthode d'autorité* est utilisée par un groupe d'individus pour imposer un dogme à une communauté, parfois par la force. La *méthode a priori* adopte des croyances « agréables à l'esprit » et en tire une série de conclusions, comme proposé par Descartes. Après avoir montré les insuffisances de ces trois méthodes, Peirce propose la *méthode scientifique* comme le seul moyen pour arriver à une croyance vraie, une croyance au réel. Malgré son appellation, plus tard remplacée par « enquête », toute personne peut la mobiliser, peu importe le doute à dissiper.

La méthode scientifique pour construire une croyance vraie

La croyance n'équivaut pas nécessairement à la vérité. Elle est l'interprétation de la réalité et, comme toute interprétation, elle peut être fautive (une croyance en une fiction). Pour distinguer une croyance vraie d'une croyance fautive, Peirce (1878a) propose de la mettre à l'épreuve, de voir dans l'expérience si elle fonctionne.

Elle est l'interprétation de la réalité, et comme toute interprétation, elle peut être fautive et en cela peut être une croyance à la fiction. Pour distinguer une croyance vraie d'une croyance fautive, Peirce propose l'expérience : une croyance vraie est une croyance qui fonctionne.

Il faut donc pouvoir tester ses croyances et seule la méthode scientifique le permet, car elle est fondée sur la vérification. Elle procède par expérimentations et les protocoles pour les réaliser sont extérieurs à un individu. Ils peuvent être critiqués et améliorés par la communauté. Ainsi, « *la croyance prédestinée à réunir finalement tous les chercheurs est ce que nous appelons le vrai, et l'objet de cette croyance est le réel.* »³⁹ (1878a, 300).

La vérité se construit à partir du réel par expérimentations et croyances successives. Si une croyance issue de la méthode scientifique n'est pas toujours vraie, l'enquête pragmatiste garantit de pouvoir vérifier cette croyance, et de la rendre vraie à terme.

³⁸ *The feeling of believing is a more or less sure indication of there being established in our nature some habit which will determine our actions.*

³⁹ *The opinion which is fated to be ultimately agreed to by all who investigate, is what we mean by the truth, and the object represented in this opinion is the real*

Organisation de la méthode scientifique

Pour établir la méthode scientifique, Peirce s'appuie sur les travaux des logiciens, d'Aristote à Kant. L'état des connaissances de son époque se résume à deux types de raisonnements : « 1) *explicatif, analytique ou déductif* ; 2) *ampliatif, synthétique ou (au sens large) inductif*. »⁴⁰ (1878c, 711). Les premiers, les raisonnements déductifs, ont été longuement étudiés, mais ils sont limités en ce qu'ils ne produisent pas de nouvelles connaissances qui ne soient déjà contenues dans les axiomes du raisonnement.

Peirce rend hommage à Kant pour avoir mis en avant les seconds : « *Emmanuel Kant a posé la question suivante : 'Comment les jugements synthétiques a priori [antérieurs à l'expérience] sont-ils possibles ?'. (...) [Par le simple fait d'avoir posé la question,] la philosophie courante de l'époque a été ébranlée et détruite, et une nouvelle époque de son histoire a commencé.* »⁴¹ (1878c, 717).

Cependant, Peirce remarque que les raisonnements synthétiques se composent de deux types d'inférences fondamentalement différentes : l'induction et l'abduction (qu'il appelle 'hypothèse' dans ses premiers textes). Elles se ressemblent assez pour les avoir confondues tout ce temps : « *L'analogie entre l'hypothèse et l'induction est si forte que certains logiciens les ont confondues.* »⁴² (1878d, 475).

L'abduction consiste à imaginer la possibilité d'une règle à partir d'un exemple ou de quelques indices : « *L'hypothèse se forme quand nous constatons une circonstance très curieuse qui s'expliquerait en supposant qu'il s'agit du cas d'application d'une certaine règle, et que nous adoptons cette supposition.* »⁴³ (1878d, 472). Elle est le seul type d'inférence qui crée de nouvelles connaissances, car elle passe d'un type de faits à un autre : « *l'hypothèse déduit des faits d'une certaine nature à des faits d'une autre nature.* »⁴⁴ (1878d, 481). Elle procure ainsi une explication, un lien de cause à effet entre ces deux types d'événements. Cependant,

⁴⁰ 1. *Explicative, analytic, or deductive*; 2. *Amplificative, synthetic, or (loosely speaking) inductive*.

⁴¹ *Immanuel Kant asked the question 'How are synthetical judgments a priori [meaning prior to experience] possible?' ... [By] the mere asking of it, the current philosophy of that time was shattered and destroyed, and a new epoch in its history was begun.*

⁴² *The analogy of hypothesis with induction is so strong that some logicians have confounded them*

⁴³ *Hypothesis is where we find some very curious circumstance, which would be explained by the supposition that it was a case of a certain general rule, and thereupon adopt that supposition.*

⁴⁴ *Hypothesis infers from facts of one kind to facts of another*

cette explication n'est qu'une supposition. Elle est donc peu robuste : « *En règle générale, l'hypothèse est un argument faible.* »⁴⁵ (1878d, 473).

De plus, elle peut être reconnue par l'émotion qu'elle provoque. Elle a une valeur esthétique et narrative, créant un fort sentiment d'harmonie : « *le résultat [de l'hypothèse] est une perturbation unique et harmonieuse que j'appelle une émotion.* »⁴⁶ (1878d, 482). En effet, la recherche est un processus narratif (Lorino, Tricard, et Clot 2011; Lorino et Tricard 2012) dans lequel l'abduction tente de retisser les fils de l'expérience à travers un récit plausible pour rester dans le flux de la pensée et de l'expérience (Lorino 2014).

L'induction fait appel à l'énumération, mais il est moins question de compter, que de mesurer, c'est-à-dire d'établir les liens entre les nombres. « *L'induction consiste à généraliser à partir d'un certain nombre de cas où quelque chose est vrai, et à en déduire que la même chose est vraie pour toute une classe.* »⁴⁷ (1878d, 481). Elle confirme les faits, sans apporter de nouvelles connaissances : « *l'essence d'une induction est de déduire d'un ensemble de faits un autre ensemble de faits similaires.* »⁴⁸ (1878d, 481).

Selon Peirce, les méthodes statistiques sont l'essence même du raisonnement inductif (Peirce 1931, 2.268) et un outil clé pour les enquêtes (Lorino 2014). Par conséquent, les outils d'IA, reposant sur de telles méthodes, semblent être des acteurs de choix dans le processus de production de connaissances comme nous le développons ci-dessous.

Ainsi, Peirce établit trois types d'inférences complémentaires et distinctes : l'abduction qui suggère une hypothèse ; la déduction qui en tire diverses conséquences, et l'induction qui les met à l'épreuve. Qui mobilise ces inférences ? Ce ne sont pas seulement les humains, mais tout quasi-esprit (Peirce 1931, 4.551), dont les outils d'IA font partie.

2.1.2. Les outils d'IA comme des quasi-esprits

D'après Peirce, la sémiotique, c'est-à-dire le processus de signification se fait par des quasi-esprits (Peirce 1931, 4.551). Par conséquent, l'enquête pragmatiste est construite non seulement par des acteurs humains, mais aussi par des acteurs non-humains comme d'autres

⁴⁵ *As a general rule, hypothesis is a weak kind of argument*

⁴⁶ *The result is a single harmonious disturbance which I call an emotion*

⁴⁷ *Induction is where we generalize from a number of cases of which something is true, and infer that the same thing is true of a whole class.*

⁴⁸ *The essence of an induction is that it infers from one set of facts another set of similar facts*

espèces animales (par exemple, les rats de laboratoire), des entités matérielles (comme un thermomètre), des lois physiques, etc.

Pour mieux comprendre le rôle de l'outil d'IA dans notre étude de cas, et de manière plus générale dans les organisations, nous détaillons comment Peirce conçoit les machines dans le processus d'enquête, et plus particulièrement les ancêtres des outils d'IA avec les « Machines Logiques » de Jevons et Marquand (Peirce 1887).

Les Machines Logiques effectuent des opérations de logique formelle grâce à des engrenages. Elles servent à Peirce pour illustrer que la logique dépasse la pensée individuelle (et sa boîte crânienne). Il généralise son argument en donnant l'exemple des instruments de chimie, qu'il considère comme des extensions de la pensée : « *Un appareil destiné à réaliser une expérience physique ou chimique est également une machine à raisonner, avec cette différence qu'il ne dépend pas des lois de l'esprit humain, mais de la raison objective incarnée dans les lois de la nature. Ce n'est donc pas une figure de style que de dire que les alambics et les cucurbites du chimiste sont des instruments de pensée ou des machines logiques.* »⁴⁹ (Peirce 1887, 168). Par conséquent, toute machine est une machine à raisonner, car elle participe à la création de sens à travers la médiation sémiotique.

Cependant, il remarque une distinction entre la pensée intracrânienne et la pensée extracrânienne, issue de deux limites fondamentales : « *Toute machine à raisonner, c'est-à-dire toute machine, a deux impuissances inhérentes.* »⁵⁰ (Peirce 1887, 169).

Selon Peirce, la première limite est un choix de conception : la machine est dépourvue d'initiative et d'originalité pour rester sous le contrôle de nos volontés (en particulier celles du concepteur) : « *D'abord, elle est dépourvue de toute originalité, de toute initiative. Elle ne peut pas trouver ses propres problèmes, elle ne peut pas se nourrir elle-même. Elle ne peut s'orienter entre différentes procédures possibles. [...] Ce n'est pourtant pas un défaut de la machine ; nous ne voulons pas qu'elle fasse ses propres affaires, mais les nôtres.* »⁵¹ (Peirce 1887, 169). Par

⁴⁹ *A piece of apparatus for performing a physical or chemical experiment is also a reasoning machine, with this difference, that it does not depend on the laws of the human mind, but on the objective reason embodied in the laws of nature. Accordingly, it is no figure of speech to say that the alembics and cucurbits of the chemist are instruments of thought, or logical machines.*

⁵⁰ *Every reasoning machine, that is to say, every machine, has two inherent impotencies.*

⁵¹ *In the first place, it is destitute of all originality, of all initiative. It cannot find its own problems; it cannot feed itself. It cannot direct itself between different possible procedures. (...) This, however, is no defect in a machine; we do not want it to do its own business, but ours.*

exemple, ChatGPT et les autres outils d'IA générative ne font que répondre à des « prompts », c'est-à-dire à des instructions données par les utilisateurs.

La seconde limite est structurelle. « *En second lieu, la capacité d'une machine a des limites absolues ; elle a été conçue pour faire une certaine chose, et elle ne peut rien faire d'autre.* »⁵²(Peirce 1887, 169). Par exemple, le fonctionnement automatique des outils d'IA reste limité à ce pour quoi ils ont été programmés. Il ne s'agit pas nécessairement d'un choix de conception, comme la première limite, mais d'une incapacité fondamentale de toute machine. Cette subtilité s'éclaire grâce à la notion de contrôle chez Peirce.

La véritable distinction entre la machine et l'être humain réside dans le « contrôle » et en particulier dans l'autocontrôle (ou la réflexivité). Selon Peirce, les capacités de contrôle varient sur un continuum⁵³.

Toute machine est capable d'un certain niveau de contrôle. Par exemple, le mouvement cybernétique* a doté des machines de boucles de rétroaction* pour moduler leur action et mieux s'adapter à leur environnement. Cependant, les mécanismes de régulation des machines ne sont pas à la hauteur des capacités de l'être humain. Par exemple, les outils d'IA fondés sur l'apprentissage machine sont limités par leurs hyperparamètres malgré leur capacité à s'entraîner sur des données. C'est pourquoi, les outils d'IA actuels sont qualifiés de « faibles », c'est-à-dire qu'ils ne sont capables que de résoudre des tâches spécifiques et ne peuvent s'adapter à de nouvelles tâches une fois la phase d'entraînement terminée.

Bien que certains puissent croire que des IA générales puissent un jour exister (capables donc de dépasser leur conditionnement initial), Larson (2021) expose pourquoi les théories actuelles nous en empêchent, peu importe le perfectionnement des méthodes actuelles : elles sont incapables de mécaniser les inférences abductives.

En revanche, l'esprit humain possède un niveau de contrôle élevé : il fait preuve d'autocontrôle et d'autocorrection, c'est-à-dire d'un contrôle sur le contrôle (ou une autocritique du contrôle). Ce contrôle que l'individu exerce sur sa propre conduite l'amène à réviser et à faire évoluer ses normes et ses principes, lui octroyant des « idéaux esthétiques et moraux »⁵⁴ (P. Steiner 2013, 272).

⁵² *In the second place, the capacity of a machine has absolute limitations; it has been contrived to do a certain thing, and it can do nothing else.*

⁵³ Voir le passage 5.533 pour une description des différents degrés.

⁵⁴ aesthetic and moral ideals

Les êtres humains sont dénués de finalités spécifiques – chacun cherche le sens de sa vie et les réponses évoluent au cours du temps. Or, une finalité est nécessaire pour établir un critère sur lequel approuver, modifier ou critiquer des méthodes de raisonnement. Par exemple, pour approuver la phase d'entraînement d'un outil d'IA, son comportement sera jugé en fonction d'une finalité, de la tâche à accomplir pour laquelle il a été conçu. En revanche, sans finalité spécifique, l'être humain est capable de réviser et rejeter ses finalités antérieures. Nous observons là un parallèle avec la « croyance vraie » qui peut être développée par mises à l'épreuve successives.

Jusqu'ici, Peirce (1887) propose un regard sur les humains et les machines. Les deux sont des quasi-esprits qui participent à l'enquête pragmatiste. Malgré cela, les humains ont une capacité d'auto-contrôle supérieure aux machines, pouvant ainsi dépasser leurs finalités existantes pour s'adapter à de nouvelles situations. Cependant, cet autocontrôle a besoin des machines (et autres quasi-esprits) pour s'exercer : *« l'esprit sans aide est également limité à l'égard [de la capacité d'auto-contrôle] comme à d'autres ; mais l'esprit qui travaille avec un crayon et beaucoup de papier n'a pas de telles limites. Il va de l'avant, et quelles que soient les limites que l'on peut assigner à sa capacité aujourd'hui, elles peuvent être dépassées demain. »*⁵⁵ (Peirce 1887, 171).

Autrement dit, les humains sont interdépendants avec les machines. La capacité d'autocontrôle chez l'être humain ne provient pas de caractéristiques inhérentes à l'individu, à son individualité et à sa subjectivité, mais grâce à une *transformation par la pratique de signes qui exercent un contrôle* (Colapietro 1989). Par conséquent, l'intelligence humaine est artificielle. Elle s'appuie constamment sur des artefacts de tous types, dont les outils d'IA.

La vision de Peirce prend donc en compte un cercle entre quasi-esprits : *« les humains et les mots s'éduquent réciproquement ; chaque augmentation de l'information d'un humain implique et est impliquée par une augmentation correspondante de l'information d'un mot. »*⁵⁶ (Peirce 1868, 157). Un corollaire immédiat, lié à notre question de recherche, consiste à chercher également par quels moyens les outils d'IA peuvent bénéficier de la collaboration avec les experts métier.

⁵⁵ *The unaided mind is also limited in this as in other respects; but the mind working with a pencil and plenty of paper has no such limitation. It presses on and on, and whatever limits can be assigned to its capacity to-day, may be over-stepped to-morrow.*

⁵⁶ *Men and words reciprocally educate each other; each increase of a man's information involves and is involved by, a corresponding increase of a word's information.*

Nous éclairons désormais notre étude de cas à la lumière de ces développements.

2.2. La place de l'outil d'IA dans l'enquête pragmatiste

Pour comprendre le projet KOVAK à l'aide de l'enquête pragmatiste, nous mettons en évidence les principaux quasi-esprits qui y participent et les inférences qu'ils effectuent. Nous considérons alors l'outil d'IA comme un expérimentateur (2.2.1), définissant le rôle du Data Scientist (2.2.2) et celui de l'expert métier (2.2.3) dans le processus de construction de connaissances (2.2.4).

2.2.1. L'outil d'IA comme expérimentateur

L'outil d'IA se résume à un quasi-esprit qui effectue des inférences inductives, et de manière plus générale, son usage principal au sein du projet KOVAK consiste à confirmer des connaissances plutôt qu'à en produire de nouvelles.

Les abductions sont faites par les concepteurs de la cohorte (étape 0). Choisir quel type de données collecter sur un phénomène constitue déjà une hypothèse. Le choix d'une variable suppose que celle-ci est un Representamen judicieux, capable de représenter l'objet étudié dans la perspective de l'étude. Par exemple, dans la cohorte du projet KOVAK, la présence de métastases est supposée contribuer aux effets secondaires de l'immunothérapie.

Sous la forme de modèles, les outils d'IA créent des associations entre les données collectées. À l'aide de méthodes statistiques, ils évaluent la force et la forme des abductions réalisées précédemment. Ils agissent comme des expérimentateurs testant les hypothèses sur un phénomène. L'outil d'IA effectue donc des inférences inductives.

Le test effectué par l'outil sert à valoriser une hypothèse, à confirmer (ou infirmer) une connaissance. Il normalise celle-ci : il attribue des coefficients aux variables, des indices d'incertitude, et une myriade d'autres métriques. Par exemple, la présence de métastases chez un patient se transforme en « une probabilité x avec un indice d'incertitude y d'effets secondaire à l'issue de l'immunothérapie ».

Les limites des inductions de l'outil d'IA restent à définir pour situer le rôle de l'outil dans l'enquête pragmatiste. La première limite est commune à toute machine. Certes, l'outil d'IA excelle dans l'ingéniosité des calculs et se distingue de technologies antérieures par son degré d'auto contrôle. Par exemple, il dépasse la simple méthode statistique car il est capable de mobiliser ces méthodes pour s'adapter au mieux à un problème donné. Pourtant, ceci se fait dans les limites de ses hyperparamètres. L'outil d'IA manque l'initiative de proposer (ou trouver) de nouvelles données.

La seconde limite que nous identifions est propre aux outils d'IA d'apprentissage machine. Ce type de quasi-esprit semble agnostique à un domaine, au contraire par exemple

d'instruments traditionnels propres à réaliser des expériences en chimie. Dans l'interprétation des signes, les outils d'IA interviennent à un niveau plus abstrait car les signes qu'ils traitent sont numérisés. Ces signes peuvent faire référence à tout type de domaine, comme la chimie, la physique, etc. Il demeure néanmoins restreint par la quantité et la qualité des données, c'est-à-dire des interprétations que d'autres quasi-esprits auront fait en amont, comme en témoigne l'extrait ci-dessous.

MedExp : Bon, après... Tant pis. Ce sont les données de la vie. On ne peut pas non plus faire beaucoup mieux.

DataExp : Vu ce que font les modèles boîte noire, comme XBOST, ça me laisse penser que ce sont vraiment les limites supérieures. Si ce type de modèle n'arrive pas à atteindre un bon seuil, avec les modèles linéaires, ça va être compliqué.

MedExp : Non, mais... Écoute, je pense qu'il n'y a pas d'erreur sur le remplissage de la base de données de la cohorte. Ce qui rend compliqué les choses pour toi, DataExp, c'est que ce sont des données de la vraie vie. Ce n'est pas une population uniforme. Tu as plein de types de cancers différents et plein de traitements différents. Je pense que ça joue. Si on avait mis que des mélanomes qui se ressemblaient, et plein de variables, tu aurais pu nous trouver un modèle encore plus performant, mais... C'est le jeu. Rétrospectivement, je pense que tu n'aurais pas pu faire mieux. Il faut donc acter ça. Dans les variables que tu as, elles ne sont pas toutes dépendantes d'un type tumoral. Je trouve ça bien.

Cette seconde limite a également une conséquence en aval du processus de production de connaissances. Comme le définit Lorino (2014, 2018), la médiation sémiotique actualise le passé et le futur, d'autres lieux spatiaux et sociaux, ainsi que les points de vue des autres dans la production de sens ici et maintenant. Cependant, en raison de son niveau d'abstraction, l'outil d'IA produit des résultats qui doivent encore être interprétés par les autres acteurs afin de reconnaître les expériences passées auxquelles les résultats font référence et les futurs possibles qu'ils ouvrent. D'où l'importance de comprendre le rôle des autres acteurs en lien avec l'outil d'IA dans le projet KOVAK.

2.2.2. Vérification de l'expérience par le Data Scientist

L'outil d'IA ne réalise qu'une partie de l'inférence inductive. Pour aller jusqu'au bout de la démarche, il est nécessaire d'abord de vérifier l'expérience menée, de s'assurer qu'elle respecte un protocole défini et que les résultats sont obtenus par des opérations conformes à

la méthode choisie. Par exemple, les outils d'IA peuvent « surapprendre » (*overfitting* en anglais) sur un jeu de données, c'est-à-dire que le modèle obtenu est trop ajusté aux données d'entraînement et que ses résultats ne sont pas fiables pour des observations futures.

Le rôle principal du data scientist est de vérifier l'expérience menée par l'outil d'IA. Dans un premier temps, il définit l'expérience à mener par l'outil d'IA. Lors de la compréhension de la situation (étape 1) et la compréhension des données (étapes 2, 9 et 11), il s'approprie le phénomène à analyser et les données disponibles. Pendant la préparation des données (étapes 3, 6 et 10), il met en place les conditions statistiques nécessaires à la réalisation de l'expérience. Il définit ainsi les protocoles expérimentaux et les moyens pour les réaliser à travers l'outil d'IA, comme le choix entre un modèle linéaire ou non-linéaire.

Ensuite, le data scientist poursuit son rôle de vérification en examinant les résultats produits par l'outil d'IA (phases d'évaluation aux étapes 5, 8 et 12). Par exemple, il utilise les statistiques descriptives effectuées précédemment pour évaluer la conformité des variables obtenues avec le protocole initial. Plus généralement, comme le montre l'échange ci-dessous, les protocoles sont adaptés à l'analyse statistique. Le data scientist contribue ainsi à la phase de confirmation de l'induction en vérifiant les conditions expérimentales.

DataExp : Après avoir effectué les 30 découpages [différents des données d'entraînement], j'ai pu déterminer les variables qui revenaient le plus souvent [en fonction des résultats pour chaque découpage]. J'ai pris les plus présentes avec leur coefficient associé.

MedExp : Le nombre de découpage que tu as fait en entraînement est arbitraire ou c'est quelque chose d'habituel dans les modèles prédictifs ? Pourquoi 30 ? Et pas 10 ou 100 ?

DataExp : C'est une valeur par défaut. J'aurais bien voulu en faire 100, mais vraiment ça prend trop de temps de calcul. À partir de 30, on peut déjà commencer à faire des agrégations. Donc on peut commencer à faire des statistiques, comme calculer la médiane ou la moyenne.

Ce processus de vérification met en lumière les pratiques interrogatives sur l'outil, c'est-à-dire qui mobilisent des connaissances sur son fonctionnement. Les pratiques interrogatives générales sur l'outil concernent la vérification des conditions expérimentales par rapport à de normes et de conventions statistiques, prouvant la fiabilité du modèle final.

Lorsqu'elles sont spécifiques, elles évaluent l'importance d'une variable particulière par rapport à l'ensemble du modèle. Elles déterminent la nécessité de considérer cette variable

par rapport au modèle et au contexte spécifique d'application, s'orientant davantage vers une vérification externe de l'expérience.

Qu'elles soient générales ou spécifiques, les pratiques interrogatives sur l'outil visent à vérifier l'expérimentation réalisée par l'outil d'IA. Mais quel sens donner à ces expériences ? C'est au tour de l'expert métier de poursuivre l'enquête pragmatiste.

2.2.3. Validation de l'expérience par l'expert métier

À la fin du processus inductif, l'expert métier intervient pour valider l'expérience. Cette étape implique de vérifier que les résultats obtenus correspondent aux objectifs fixés et répondent effectivement au doute initial.

Au début de l'enquête, les experts médicaux définissent le problème à résoudre en s'appuyant sur leur expérience et les exigences de la pratique clinique (étape 1 sur la compréhension de la situation). Ils transforment également les expériences en données numériques lors de la constitution de la cohorte (étape 0), établissant ainsi les conditions pratiques pour la validité de l'expérience. Par exemple, les variables générées par DataExp avec l'approche fondée sur les données doivent être revues par les experts médicaux pour s'assurer qu'elles correspondent aux conditions pratiques (étape 11).

Les experts médicaux interviennent de nouveau lors de la phase d'évaluation (étapes 5, 8 et 12). Ils veillent à ce que les variables soient cohérentes avec les connaissances médicales en utilisant des pratiques interrogatives professionnelles. Ils s'assurent que les résultats concordent avec la littérature clinique et qu'ils sont pertinents pour la pratique médicale. Par exemple, les experts médicaux vérifient que les facteurs identifiés puissent aider les praticiens à identifier les patients nécessitant une attention particulière après l'immunothérapie à cause des effets indésirables*.

Ce processus de validation engage également des formes d'abduction. Les experts médicaux entreprennent un travail de contextualisation, situant les variables issues de l'outil d'IA dans un réseau de connaissances et d'expériences partagées. Ces connaissances permettent aux professionnels de santé de mieux agir face à une situation. Comme le montre l'exemple ci-dessous, les experts médicaux ont besoin de pouvoir « *formuler pour que ce soit compréhensible par les cliniciens* ». Une explication narrative de cause à effet doit pouvoir être mise en avant.

MedExp : [En lisant la liste des variables d'un des modèles présentés par DataExp] C'est un peu bizarre « mélanome – lésions ». Je ne sais pas à quoi ça correspond sur le plan clinique. C'est un peu compliqué. En tout cas, avoir des « métastases » dans le cadre d'un « mélanome » ok, on peut garder. On peut aussi garder [énonce plusieurs variables].

Cependant, « tumorale – cellules », pareil, je ne sais pas à quoi ça correspond. Tu as une idée CoExp ?

CoExp : Peut-être les cellules du cancer ? Normalement dans les antécédents médicaux, quand on parle de « cellules », c'est pour définir le type de cancer. Aussi, j'ai un petit doute pour la variable « pulmonaire – osseux ». Ça pourrait être soit des « lésions », soit des « adénocarcinomes pulmonaires » avec une métastase osseuse. Ça aussi, il faut bien préciser si ce sont des lésions avec un primitif ou une métastase. Ça pourrait être les deux.

[...]

MedExp : En fonction de la variable, tu sais à quels patients cela correspond ?

DataExp : Oui, je peux récupérer le numéro du patient en codant un peu.

MedExp : Du coup, on peut regarder le jeu de données.

[...]

MedExp : Avec « Pet », j'ai un peu du mal. C'est si tu as une fixation, mais des ganglions pet. Ça veut dire ça, je pense. Tu as mis ça, CoExp, dans la base ? Le fait que le pet scanner soit positif ?

CoExp : Je peux voir ça dans la colonne des antécédents médicaux, mais je ne me souviens pas si j'avais mis si les patients avaient eu un petscan ou non. Après ça dépend du compte rendu des médecins s'ils indiquent l'information.

MedExp : Ok, il faut voir si ça correspond à quelque chose. Pareil pour « lésion – pulmonaire ». Je pense que cela optimise le fait d'avoir une localisation métastatique pulmonaire. Ton modèle les retrouve, donc il faut juste le formuler pour que ça soit compréhensible par les cliniciens. Avec CoExp on va voir si c'est cliniquement significatif. C'est peut-être juste une question de formulation, tout simplement.

En développant l'analyse davantage, nous identifions deux types d'abductions dans notre matériau, les abductions « surcodées » et les abductions « sous-codées » (Eco 1983).

Lorsque le résultat de l'outil d'IA est facilement interprétable, c'est-à-dire lorsqu'une seule explication semble apparaître en raison d'un contexte clair, alors l'abduction est surcodée.

Selon la typologie des abductions d'Eco, elle correspond à la situation où « *la loi est donnée automatiquement ou semi-automatiquement. [...] Il est très important de supposer que même l'interprétation à travers des codes suppose un effort d'abduction, aussi minimal soit-il. [...] Il semble que généralement nous faisons ce genre de travail interprétatif automatiquement, mais si par hasard on vit dans un milieu international où les gens sont censés parler différentes langues, on se rend compte que le choix n'est pas radicalement automatique.* »⁵⁷ (Eco 1983, 206). Ce type d'abduction apparaît surtout lors de la mobilisation de pratiques interrogatives professionnelles générales. L'interprétation réussie ne demande pas de pratiques interrogatives supplémentaires.

En revanche, les pratiques interrogatives professionnelles spécifiques appellent souvent à des abductions « sous-codées », c'est-à-dire lorsque « *la règle doit être sélectionnée parmi une série de règles équiprobables mises à notre disposition par la connaissance du monde actuel [...] Puisque la règle est sélectionnée comme étant la plus plausible parmi de nombreuses autres, mais qu'il n'est pas certain qu'elle soit la « correcte » ou non, l'explication n'est qu'envisagée, en attendant de nouveaux tests.* »⁵⁸ (Eco 1983, 206).

Lorsque les experts mobilisent les pratiques interrogatives professionnelles spécifiques, ils cherchent des causes probables, soit déjà établies dans la littérature, soit en lien avec leur expérience pratique. La loi à laquelle le résultat s'applique exige plus de recherche, en adoptant des stratégies propres à chaque situation. Par exemple, ils examinent les caractéristiques des patients concernés par le résultat à interpréter. Ils cherchent à établir quelle règle correspond le mieux.

Enfin, la typologie d'Eco (1983) nous indique qu'il existe deux autres types d'abductions possibles : les abductions créatives et les méta-abductions. L'abduction créative formule de nouvelles hypothèses n'ayant pas été envisagées auparavant. Elle est souvent associée à des découvertes majeures qui transforment la compréhension existante dans un domaine. Cette démarche conduit à la méta-abduction, qui évalue si les nouvelles hypothèses ou lois générées correspondent à la réalité telle que nous l'expérimentons. Ces abductions exigent une nouvelle

⁵⁷ *The law is given automatically or semiautomatically. [...] It is very important to assume that even interpreting through codes presupposes an abductional effort, however minimal. [...] It seems that usually we do this kind of interpretive labor automatically, but if by chance one is living in an international milieu in which people are supposed to speak different languages one realizes that the choice is not radically automatic.*

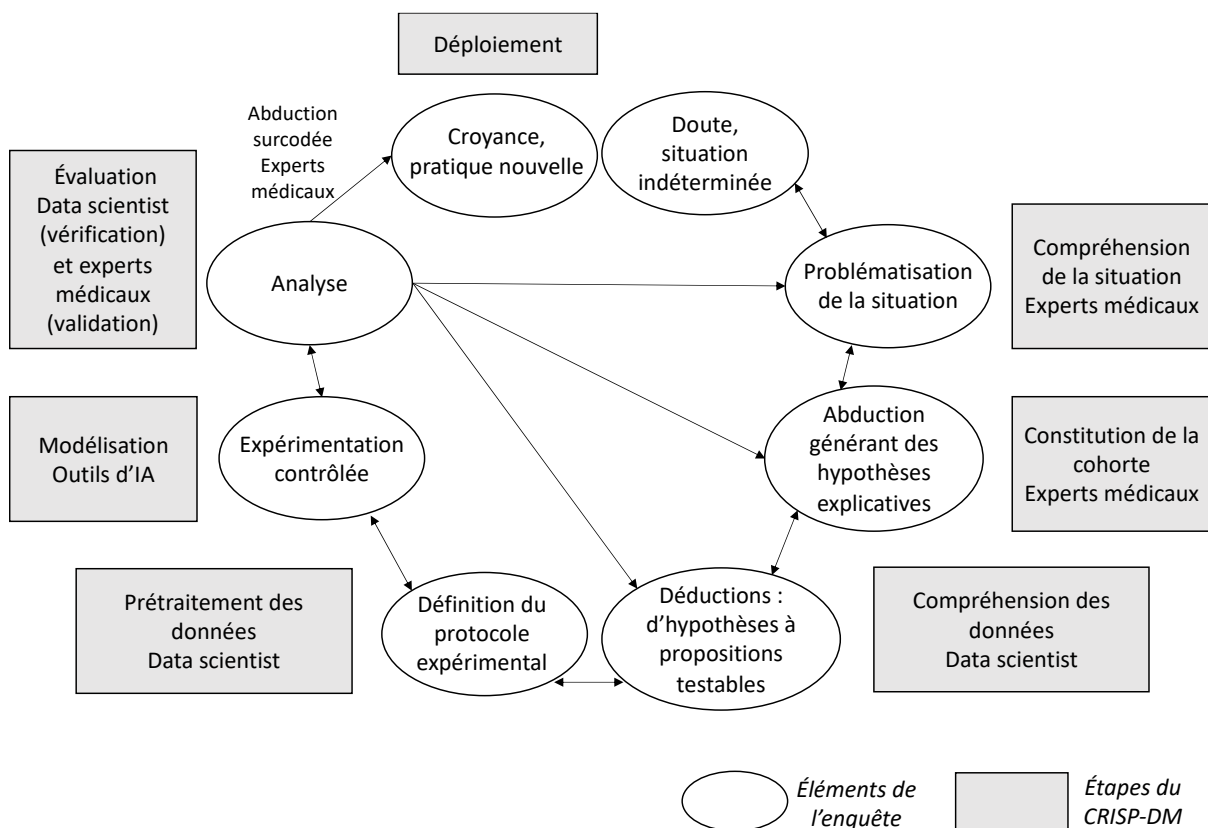
⁵⁸ *The rule must be selected from a series of equiprobable rules put at our disposal by the current world knowledge [...] Since the rule is selected as the more plausible among many, but it is not certain whether it is the "correct" one or not, the explanation is only entertained, waiting for further tests.*

enquête pour éprouver cette nouvelle croyance, contrairement aux abductions surcodées et sous-codées, dont les hypothèses découlent de l'expérience et des connaissances existantes. Cependant, nous n'avons pas observé de telles abductions dans le projet KOVAK.

2.2.4. Relecture du projet KOVAK selon l'enquête pragmatiste

À travers la théorie des inférences et le concept de quasi-esprit de Peirce, le processus d'enquête pragmatiste apporte une nouvelle perspective au projet KOVAK. Il révèle l'usage des outils d'IA dans la recherche médicale comme une collaboration avec des experts. À partir des abductions effectuées précédemment, l'outil d'IA effectue une partie du processus d'induction, complété par le data scientist, tandis que les experts médicaux concluent par des abductions, pour finir de construire un sens aux résultats obtenus et adopter une croyance plus robuste sur la situation.

Nous avons élaboré un schéma inspiré du modèle de l'enquête pragmatiste par Lorino (2018, 112) pour illustrer ce processus, permettant de mieux saisir le rôle et l'interaction des outils d'IA dans le cadre de l'enquête (figure 18).



Grâce à ce schéma, nous voyons que la collaboration avec les outils d'IA dans la recherche médicale aide à la production de connaissances, en validant des hypothèses, ce que nous avons précédemment associé à une augmentation dans le registre de la conformation (Moison 2005a).

En fonction des résultats de l'étape d'évaluation, les acteurs reviennent sur des étapes précédentes pour améliorer le modèle statistique. Dans le projet KOVAK, ce retour sur en amont de l'enquête se fait jusqu'au niveau des déductions (les propositions testables) à partir de la base de données. Autrement dit les hypothèses à l'origine de la conception de la cohorte ne sont pas modifiées.

Pourtant, les éléments de l'enquête (Lorino 2018, 112) ainsi que la typologie des abductions (Eco 1983) suggèrent que l'analyse des expériences est capable d'engendrer de nouvelles hypothèses explicatives, voire de modifier la problématisation initiale de la situation.

Si, dans une certaine mesure, nous assistons à la production de nouvelles connaissances, par exemple lors de l'approche mixte, MedExp et COEXP identifient de nouveaux facteurs, l'étape de l'évaluation pourrait transformer la compréhension du problème et motiver la collecte de nouvelles données, ouvrant la voie à davantage de nouvelles connaissances.

Nous analysons le projet KOVAK comme un processus de construction de connaissances, en utilisant l'approche pragmatiste de Charles S. Peirce. Selon Peirce, la production de connaissances repose sur trois types de raisonnement : l'abduction (formuler des hypothèses), la déduction (tirer des conclusions logiques) et l'induction (généraliser à partir de cas spécifiques), qui peuvent être effectués par des quasi-esprits, c'est-à-dire des entités capables d'interpréter des signes, comme les humains et les outils d'IA.

L'outil d'IA devient ainsi un partenaire d'enquête aux côtés des humains, en particulier des data scientists et des experts médicaux.

Nous considérons l'outil d'IA comme un « nouvel expérimentateur » jouant un rôle crucial dans la mise à l'épreuve des hypothèses. Il teste les hypothèses en analysant les données pour confirmer ou infirmer ce que l'on pense savoir. Par exemple, il vérifie si la présence de certaines données correspond à des attentes fondées sur les hypothèses formulées par des concepteurs de la cohorte.

Le data scientist vérifie les expériences menées par l'outil d'IA. Il configure et contrôle les expériences menées : il s'assure que les données sont bien préparées et que les méthodes utilisées sont appropriées. Il examine les résultats obtenus par l'outil d'IA pour garantir leur pertinence statistique grâce à des pratiques interrogatives sur l'outil.

L'expert médical valide les résultats en fonction de leur pertinence clinique grâce à des pratiques interrogatives professionnelles. En outre, il s'assure que les conclusions de l'outil d'IA sont utiles et pertinentes pour la pratique médicale.

Ensemble, ces acteurs collaborent pour produire de nouvelles connaissances. L'outil d'IA aide à confirmer des hypothèses, le data scientist veille à la rigueur des méthodes, et l'expert

médical apporte son expertise pour interpréter les résultats de manière significative. Cette collaboration montre comment les outils d'IA peuvent être intégrés dans la recherche pour améliorer la compréhension et la découverte de nouvelles informations. Ces aspects sont approfondis dans la seconde partie du chapitre 6.

Chapitre 6 : l'Expert-Machine Collaboration Community pour dépasser les modèles de collaboration humain-IA

En adoptant une démarche compréhensive, nous avons construit des études de cas pour approfondir les modes de gestion de la collaboration entre humains et outils d'IA. Ces modes couvrent les moyens pour orchestrer la collaboration entre les êtres humains et les outils d'IA, ce qui inclut la définition des rôles et responsabilités (Siemon 2022), la coordination des tâches (Ren, Chen, et Qiu 2023), la communication et l'interaction entre acteurs (Stowers et al. 2021; Ren, Chen, et Qiu 2023) ainsi que le contrôle et l'amélioration des processus collaboratifs (Siemon 2022).

Dans le contexte du système de soins, les experts métier restent responsables des décisions prises, et donc du rôle de l'outil d'IA censé augmenter leurs capacités en réduisant l'incertitude. La littérature souligne le besoin de métacognitions (Jussupow et al. 2021) des experts, en raison de l'opacité des outils d'IA, pour une collaboration engagée. Les pratiques interrogatives recensent comment les experts questionnent les résultats des outils d'IA pour les intégrer dans leur réflexion (Lebovitz, Lifshitz-Assaf, et Levina 2022).

Notre analyse a repris cette notion pour établir comment les modes de gestion participent à l'augmentation des capacités des experts. Nous avons identifié diverses pratiques interrogatives et leurs combinaisons, représentant des modes de gestion de la collaboration et des bénéfices pour l'expert.

Nous présentons d'abord la synthèse de ces résultats, incluant l'évolution des cadres théoriques mobilisés (1). Notre dernier cadre, la théorie de l'enquête pragmatiste de Peirce, propose un modèle pour dépasser la vision traditionnelle de la collaboration humain-machine, l'Expert Machine Collaboration Community ou EMC2 (2).

1. Synthèse des résultats des études de cas

La discussion entre les résultats des études de cas RADO et KOVAK nous conduit à privilégier la vision peircienne qui englobe le courant de l'instrumentation située (1.1). Nous présentons ensuite la typologie des pratiques interrogatives identifiées à travers ces deux études de cas (1.2). Enfin, nous examinons les configurations de ces pratiques interrogatives qui mènent à d'autres formes de modes de gestion de la collaboration (1.3).

1.1. Discussion entre les résultats des deux études de cas

RADO et KOVAK éclairent les usages des outils d'IA dans le système de soins. Nos études révèlent l'opacité des résultats produits par les outils d'IA. Les experts métier s'efforcent de

gérer ces résultats pour les intégrer de manière rationnelle dans leur processus de compréhension. Ils utilisent des compétences métacognitives (Jussupow et al. 2021) pour intégrer ces outils dans leur activité, reliant leurs connaissances aux résultats par des « pratiques interrogatives » confirmant leur engagement dans la collaboration (Lebovitz, Lifshitz-Assaf, et Levina 2022).

Les réponses des experts métier diffèrent selon les contextes : certains ignorent les résultats (Pachidi et al. 2021; Valentine et Hinds 2022; Lebovitz, Lifshitz-Assaf, et Levina 2022), tandis que d'autres inventent des récits ad hoc, risquant des décisions erronées (Elish et Watkins 2020; Borch et Hee Min 2022). Les pratiques interrogatives favorisent la collaboration humain-machine.

Cependant, les pratiques interrogatives seules ne suffisent pas à expliquer les bénéfices de cette collaboration. Lebovitz, Lifshitz-Assaf, et Levina (2022, 14) indiquent que « *ces pratiques offrent aux professionnels un moyen de valider les résultats de l'IA, malgré l'opacité rencontrée, et aboutissent à un mode d'augmentation humain-IA engagé.* »⁵⁹ Elles ajoutent que « *comprendre les explications de l'IA nécessite un investissement en temps et en ressources* »⁶⁰ (Lebovitz, Lifshitz-Assaf, et Levina 2022, 15). Toutefois, les bénéfices précis de cette collaboration restent flous.

Avant de présenter la taxonomie des pratiques interrogatives, nous examinons les cadres théoriques pour comprendre comment la collaboration humain-machine peut servir l'expert métier à la lumière de nos études de cas.

RADO utilise le cadre d'analyse de la conformation/exploration. Comparer les outils d'IA à des outils de gestion permet de mieux comprendre l'augmentation qu'ils procurent. Cependant, ce cadre montre des difficultés d'application dans le projet KOVAK (1.1.1). Grâce à cette seconde étude de cas, nous développons l'enquête pragmatiste de Peirce, offrant une lecture plus englobante de nos terrains (1.1.2).

1.1.1. Apports et réorientations du cadre conformation/exploration

Dans l'étude RADO, le cadre d'analyse de la conformation/exploration (Moison 2005a) examine l'usage de l'outil d'IA dans l'analyse des mammographies. Cet outil peut soit confirmer, soit contredire l'avis du radiologue, optimisant la gestion de l'attention. Il offre une assurance et permet de caractériser des cas complexes, réduisant le temps de diagnostic et

⁵⁹ *These practices provided professionals a way of validating AI results, despite experiencing opacity, and resulted in an engaged mode of human-AI augmentation.*

⁶⁰ *Making sense of AI explanations requires an investment of time and resources.*

améliorant la qualité du diagnostic en réduisant les faux négatifs, concordant avec des études quantitatives (Rodríguez-Ruiz et al. 2019; van Winkel et al. 2021; Raya-Povedano et al. 2021).

Face à ces observations, nous avons considéré que l'outil d'IA simplifie la complexité de la situation en offrant une abstraction de la zone identifiée « à risque » grâce à quelques indicateurs, à l'instar des outils de gestion (Berry 1983). Pour caractériser l'augmentation de cet outil, nous avons opté pour le courant sur l'instrumentation située. Le cadre d'analyse conformation/exploration identifie l'augmentation de l'outil d'IA principalement dans le registre de la conformation : l'usage de l'outil d'IA donne aux experts métier la capacité de mieux effectuer leur travail, optimisant ainsi leur activité au niveau de la qualité, du coût et du temps.

Ce cadre montre aussi que l'outil d'IA peut augmenter les capacités des experts en reconfigurant leur activité et en adoptant des approches différentes grâce aux apprentissages générés par son utilisation. Bien que marginal, l'augmentation des capacités dans le registre de l'exploration est possible, comme observé ailleurs (Béjean, Kletz, et Moisdon 2018).

L'analyse de RADO, via les pratiques interrogatives et le cadre de conformation/exploration, propose un cadre pour l'augmentation des capacités d'agir des experts (Perez-Torrents et Minvielle 2023). Ce cadre situe les pratiques interrogatives sur deux registres : conformation et exploration, distinguant deux niveaux d'augmentation : agir plus efficacement et agir différemment.

Ce cadre fournit une compréhension claire de l'augmentation des capacités d'agir des experts face à l'opacité des outils d'IA. Il est simple à comprendre et ancré dans des éléments tangibles, comme les pratiques interrogatives, et encourage à faire évoluer ces pratiques vers une discussion incluant plus d'acteurs pour développer la dimension exploratoire de l'outil d'IA.

Cependant, la seconde étude de cas montre des difficultés d'application de ce cadre. Le projet KOVAK illustre la mobilisation de pratiques interrogatives par les experts métier et les data scientists. L'outil d'IA aide les experts à vérifier les hypothèses en les transformant en facteurs prédictifs, permettant des décisions plus éclairées.

Cette étude souligne que l'outil d'IA, bien que vu comme un simple outil (Anthony, Bechky, et Fayard 2023), apparaît davantage comme un partenaire dans le projet KOVAK, collaborant avec les acteurs humains pour construire des connaissances, et non comme un simple médiateur.

1.1.2. Les études de cas à l'aune des inférences peirciennes

La difficulté à utiliser le cadre d'analyse conformation/exploration pour expliquer notre seconde étude de cas et ses modes de gestion de la collaboration humain-machine nous a

poussés à adopter une autre approche théorique fondée sur le concept de l'enquête pragmatiste de Peirce.

Ce nouveau cadre est particulièrement adapté pour plusieurs raisons. D'abord, les travaux de Peirce et du pragmatisme ont inspiré l'approche des réseaux sociotechniques (Muniesa 2015), ce qui confère à cette approche une cohérence globale pour théoriser les usages des outils d'IA. Par exemple, la vision peircienne considère les acteurs non-humains à travers le concept de Quasi-Esprits (Peirce 1931, 4.551), ce qui se manifeste dans le projet KOVAK.

Ensuite, la typologie des inférences de Peirce (1877; 1878a; 1878b; 1878c; 1878d; 1878e; 1887), au cœur de son concept d'enquête, explique le fonctionnement des outils d'IA à un niveau approprié. Par exemple, cette typologie distingue l'approche symbolique de l'apprentissage machine (Larson 2021) : les premiers fonctionnent selon des inférences déductives, tandis que les seconds fonctionnent selon des inférences inductives.

De plus, comme nous le montrons avec l'analyse du projet KOVAK, en se situant au niveau des inférences logiques, nous pouvons mieux identifier les rôles et responsabilités des différents acteurs, les moyens de coordination et les processus de contrôle et d'amélioration de la collaboration. Nous soulignons ainsi une limite fondamentale de l'outil d'IA : l'absence d'abductions de la machine. Si l'outil d'IA confirme des hypothèses par induction, il est nécessaire de prendre en compte les abductions (et les déductions) des autres acteurs.

Enfin, la vision peircienne intègre le cadre d'analyse de la conformation/exploration (Moison 2005a). En effet, dans le projet KOVAK, nous montrons comment l'induction de l'outil d'IA mène au registre de la conformation puisque cette inférence teste une abduction. Leur puissance de calcul et la sophistication des algorithmes produisent une grande ingéniosité (Peirce 1878e, 478). Grâce à la connaissance testée, les acteurs peuvent interpréter plus facilement les situations et agir en conséquence. Nous retrouvons donc le rôle des outils de gestion comme moyens de réduire la complexité d'une situation à quelques indicateurs simples (Berry 1983).

En relisant la première étude RADO, nous illustrons la puissance explicative du cadre peircien. Dans le registre de la conformation, l'outil d'IA améliore les capacités d'agir du radiologue en répondant aux exigences de production. Par exemple, lorsque l'outil d'IA et le radiologue ne détectent pas d'anomalie, l'usage de l'outil d'IA conforte (configuration AA), et lorsque l'outil d'IA repère une anomalie inaperçue par le radiologue (configuration β), l'outil incite le radiologue à scruter davantage la situation complexe.

À la lumière de la vision peircienne, nous voyons que l'outil d'IA effectue toujours des inférences inductives. Il quantifie les hypothèses sur la présence ou l'absence d'une zone suspecte avec des valeurs chiffrées, se traduisant par un score de risque. En d'autres termes, il réalise une forme particulière d'expérimentation, fondée sur des données numériques et se

manifestant par une série d'indicateurs. Ces inductions mettent à l'épreuve des hypothèses formulées par les radiologues.

Nous avons identifié comme mécanisme principal le mode de gestion par normalisation, qui alterne les pratiques interrogatives sur l'outil et les pratiques professionnelles pour valider des connaissances et renforcer les capacités d'agir des experts métier (registre de la conformation).

Le mode de gestion par normalisation vise à conclure le nœud sémiotique pour confirmer ou infirmer l'induction (et donc l'hypothèse initiale). En d'autres termes, l'usage de l'outil d'IA vise à accélérer un processus d'enquête linéaire, optimisant ainsi la confirmation des connaissances. La croyance du radiologue est liée à l'existence d'une zone à risque (ou à l'absence de risque), ce qui le conduit à recommander des examens complémentaires en fonction du niveau de certitude, pouvant aller jusqu'à la biopsie en cas de certitude très forte.

Dans RADO, nous avons déterminé que l'outil d'IA offre peu d'apprentissages au radiologue pour modifier son activité. Malgré la mobilisation des pratiques interrogatives pour gérer l'opacité des résultats de l'outil d'IA, le radiologue éprouve des difficultés à prendre du recul sur son fonctionnement global. Par exemple, le Dr S avait insisté pour obtenir un retour de notre part afin de mieux évaluer la pertinence de son usage de l'outil par rapport à son contexte d'utilisation.

La vision peircienne corrobore ces affirmations et approfondit l'analyse. La seconde étude de cas KOVAK a mis en lumière la nécessité, au moment du nœud sémiotique, de remettre en question les hypothèses initiales et l'analyse globale de la situation pour favoriser le développement d'apprentissages. En articulant différemment les pratiques interrogatives, telles qu'on les retrouve dans le mode de gestion par opposition ou dans le mode de gestion par approfondissement, le nœud sémiotique peut être orienté vers de nouvelles abductions ou de nouvelles perspectives sur la situation initiale.

Selon la théorie, ces modes de gestion pourraient engendrer de nouvelles connaissances et des apprentissages qui, à terme, pourraient se traduire par des capacités d'action différentes. Cependant, un tel processus d'enquête nécessite un dialogue entre les concepteurs et les experts métier. De plus, du point de vue peircien, l'outil d'IA est également considéré comme un participant à l'enquête. Il devrait donc être capable de développer des pratiques interrogatives pour intégrer les connaissances des autres experts dans ses processus. Cette perspective nous conduit à repenser la notion de pratiques interrogatives. Avant d'élaborer ces déductions, nous présentons d'abord les pratiques interrogatives identifiées sur le terrain.

1.2. Typologie des pratiques interrogatives

Nos résultats identifient deux types de pratiques interrogatives : les pratiques interrogatives professionnelles (1.2.1) et les pratiques interrogatives sur l'outil (1.2.2). Chaque catégorie se subdivise en deux niveaux : générales et spécifiques.

1.2.1. Les pratiques interrogatives professionnelles

Les pratiques interrogatives professionnelles éclairent les résultats de l'outil d'IA en les mettant en relation avec les connaissances métier, comme les connaissances médicales et l'expérience clinique avec les patients. Contrairement aux pratiques sur l'outil, elles ne considèrent pas le processus de production des résultats et sont indépendantes du fonctionnement interne de l'outil d'IA. Les experts métier développent ce type de pratiques au cours de leur formation et les utilisent pour évaluer des situations et justifier leurs jugements auprès de leurs pairs.

Ces pratiques établissent un lien entre le résultat de l'outil d'IA et le contexte empirique où il s'applique. Elles servent à valider le résultat et à évaluer sa pertinence en fonction de l'activité située, indépendamment du processus algorithmique. Les experts utilisent principalement ces pratiques pour valider les résultats et gérer l'opacité de l'outil d'IA en trouvant une causalité au résultat.

Par exemple, dans l'extrait ci-dessous, issu de la première étude de cas (cas 59, jour 6), le Dr S utilise des pratiques interrogatives professionnelles pour établir un diagnostic. Il note les facteurs de risque de la patiente, tels que sa précédente opération et son âge avancé. En examinant la mammographie, il détecte une distorsion architecturale suspecte mais constate qu'il n'y a pas d'évolution en la comparant avec des clichés antérieurs. Il poursuit avec une échographie sans repérer de zone anormale. Même si l'outil d'IA affiche un risque élevé, ses pratiques interrogatives professionnelles lui permettent de relier ce risque à une distorsion bénigne.

Le Dr S, après avoir consulté les comptes rendus, déclare : « La patiente a déjà été opérée du sein gauche. Elle a 80 ans. Ensuite, sur la mammographie, il y a cette zone de distorsion architecturale au sein droit. On va voir si les anciens clichés arrivent. Voilà. Oui, ça ne s'est pas franchement modifié depuis 2021. Je vais voir ce qui se passe sur l'échographie. »

Il part sans regarder les résultats de l'outil d'IA.

Il revient, nous lui montrons que le score de l'outil d'IA est à 10. Dr S : « Oui, c'est noté. Il est aussi embêté par ce qui se passe ici en arrière

du mamelon. Sur l'échographie, il n'y a rien. On va voir ce qui se passe l'année prochaine. »

Il rédige le compte rendu, dans lequel il précise qu'il y a une distorsion architecturale sans évolution depuis 2021, et il conclut en recommandant un suivi annuel.

Souvent, les experts métier mobilisent les pratiques interrogatives professionnelles : ils commencent par des questionnements généraux, reliant le résultat à des connaissances établies comme des mécanismes bien connus, des effets rapportés dans la littérature ou des expériences cliniques régulières. Ils interprètent ces résultats avec des abductions surcodées (Eco 1983), où une seule explication s'impose.

Par exemple, dans l'extrait suivant, le Dr S explique pourquoi l'outil d'IA ne donne pas un score élevé à une zone potentiellement inquiétante (cas 57, jour 5). Le score élevé est lié à des calcifications bénignes, mais il est agréablement surpris de l'absence d'un faux positif sur une autre zone qui aurait pu troubler des médecins moins expérimentés. Celle-ci présente des formes qui pourraient se confondre avec des masses, mais que le Dr S identifie avec certitude comme des kystes en raison de leur forme et de leur localisation dans le sein. Ici, nous observons une abduction surcodée (Eco 1983), c'est-à-dire une interprétation semi-automatique reposant sur la règle : « zones aux contours réguliers => présence de kystes ».

Par exemple, dans l'extrait suivant, le Dr S explique pourquoi l'outil d'IA ne donne pas un score élevé à une zone potentiellement inquiétante (cas 57, jour 5). Le score de risque élevé est seulement lié à des calcifications que le Dr S sait bénignes. Cependant, le radiologue est agréablement surpris de l'absence d'un faux positif sur une autre zone qui aurait pu troubler des médecins moins expérimentés. Celle-ci présente des formes qui pourraient se confondre avec des masses, mais que le Dr S identifie avec certitude comme des kystes en raison de leur forme et de leur localisation dans le sein. Ici, nous observons une abduction surcodée (Eco 1983), c'est-à-dire une interprétation semi-automatique reposant sur la règle : « zones aux contours réguliers => présence de kystes ».

L'outil d'IA affiche un score général de 9, indiquant la présence de calcifications.

Dr S : « C'est intéressant car l'IA n'est pas affolée par les différentes masses des kystes. Certainement parce que les contours sont réguliers. Elle est très spécifique ! »

Les pratiques interrogatives professionnelles spécifiques sont mobilisées lorsque les experts rencontrent des difficultés à interpréter les résultats. Contrairement aux pratiques générales, fondées sur des connaissances établies, ces pratiques spécifiques impliquent l'adaptation de stratégies propres à chaque situation. Par exemple, dans le projet KOVAK, les experts examinent l'ensemble des données des patients liés à un résultat difficile à interpréter pour établir des explications ad hoc ou des abductions sous-codées (Eco 1983).

Le Dr S fournit un exemple de pratiques interrogatives spécifiques avec le cas 72 (jour 7). La patiente présente une situation complexe avec de nombreux tests précédemment réalisés sur des zones inquiétantes mais bénignes. Lors de cet examen, le Dr S évalue une nouvelle zone sur une masse opaque du côté gauche. L'outil d'IA n'indique rien d'alarmant pour cette zone, mais le médecin utilise d'autres méthodes de visualisation, comme la tomographie localisée, pour mieux observer cette zone. Pour les zones à risque détectées par l'outil d'IA, il ne trouve pas d'explication claire, mais s'appuie sur des raisonnements, comme la comparaison entre les côtés droit et gauche pour repérer une distorsion architecturale, et les analyses de l'IRM qui ne montrent rien d'alarmant.

Le Dr S examine les anciens comptes rendus : « Elle a eu un suivi régulier, puis on lui a trouvé de tout. On lui a fait une biopsie à droite qui s'est avérée bénigne, une mastose et ce n'était rien de méchant. Maintenant, il y a un doute sur une masse opaque côté gauche. Il y a trois jours, je l'ai reçue pour faire une IRM qui n'a rien donné. Je la renvoie à la mammographie avec une tomosynthèse localisée voir ce que cela donne. Par ailleurs, elle n'a pas de facteurs de risque. »

Il examine les clichés analogiques réalisés dans d'autres centres.

Il examine les mammographies 2D. « La masse a disparu, c'est peut-être un kyste ? »

Le score de l'outil d'IA affiche 10. Dr S : « C'est intéressant, il détecte des calcifications et des masses, mais seulement du côté droit. Elles sont à 78/100, ce qui est élevé. Cependant, du côté droit, je pense qu'il n'y a rien car j'ai fait une IRM. C'est certainement une glande compacte. En regardant du côté gauche en symétrie, il y a une petite perte d'architecture, mais bon. »

Il regarde la tomosynthèse localisée : « C'est vraisemblablement un kyste, on le voit là, il a diminué de taille. Cependant, côté gauche... » Il zoome sur le cliché. « L'imagerie localisée 3D ça permettra de faire la part des choses entre le cancer et la glande. Là, on est dans la zone où

les glandes sont plus compactes, mais je ne vois rien. Nous suivrons cette zone dans 6 mois avec une mammographie et une IRM. »

Ainsi, ces exemples illustrent comment les experts métier naviguent entre les pratiques interrogatives générales et spécifiques pour gérer l'opacité des résultats de l'outil d'IA et renforcer la fiabilité de leurs diagnostics.

1.2.2. Les pratiques interrogatives sur l'outil

Les pratiques interrogatives sur l'outil se concentrent sur la validation de l'outil d'IA en respectant un protocole lié à son fonctionnement algorithmique. Contrairement aux pratiques interrogatives professionnelles, elles se focalisent sur le modèle statistique sous-jacent et le comportement de l'outil plutôt que sur l'adéquation du résultat avec le terrain. Elles mobilisent des connaissances sur le fonctionnement de l'outil d'IA, notamment des notions statistiques applicables à d'autres outils d'IA, ainsi que des connaissances empiriques liées aux limites de l'outil dans un contexte spécifique.

Le protocole des pratiques interrogatives sur l'outil peut être formel, impliquant des conditions statistiques précises basées sur des lois mathématiques, ou informel, tenant compte des conditions d'utilisation extrêmes selon les contextes d'application, comme la nécessité de minimiser les faux négatifs pour certains diagnostics médicaux.

Nos données montrent que les data scientists utilisent principalement ces pratiques pour vérifier les résultats d'un point de vue statistique. Par exemple, dans la deuxième étude de cas, une discussion entre DataExp et OrgaExp porte sur les métriques et les seuils choisis pour déterminer la validité de l'outil d'IA.

DataExp : Là, on est à 0.6. Dans les précédents modèles, on était aux alentours de 0.5.

OrgaExp : Ah, tu es à 0.6, je croyais que tu étais à 0.8 ?

DataExp : Non, ça c'était avec une autre métrique. Avec la première métrique, je pense qu'on s'est beaucoup améliorés aussi.

OrgaExp : Quelle est la nouvelle métrique ?

DataExp : C'est le F1-score. Avant on utilisait l'AUC.

OrgaExp : Mais sur l'AUC, tu es à combien ? Sur celui-là tu ne sais pas ?

DataExp : Je ne sais pas encore...

OrgaExp : Tu dois être élevé.

DataExp : Là, ça doit être au moins à 0.95.

OrgaExp : Mais en fait, ça, c'est très présentable.

DataExp : Oui, oui, là, oui, c'est bon.

OrgaExp : Et il n'y a pas de... Il y a trop de faux positifs, c'est ça ?

DataExp : Oui, c'est ça.

OrgaExp : Et pourquoi du coup ? Ça veut dire quoi ?

DataExp : Quand je parle de faux positifs, je parle d'un score qui s'appelle la « précision ». Elle était de 0,4 ou 0,38. Maintenant elle est autour de 0,5. C'est plus en termes d'amélioration. Je ne dis pas que c'est très bon. Mais en termes de progression, ça n'a rien à voir.

Nous avons également remarqué que les experts métier adoptent ces pratiques interrogatives sur l'outil, les utilisant comme filtre pour évaluer la validité des résultats des outils d'IA, déterminer l'importance de les inclure dans leur réflexion, et donc de devoir gérer leur opacité. Par exemple, lors du cas 4 (jour 1), le Dr S décide de ne pas attendre le résultat de l'outil d'IA car la patiente dispose d'une prothèse mammaire et il a remarqué que, dans ces situations, l'analyse de l'algorithme est souvent aberrante.

Dr S : « Pour elle, ça n'a pas changé depuis 2019/2020. Voilà, tic-tac ! Qu'est-ce que dit l'IA ? Toujours rien ? Bon, ça prend une éternité ! En même temps, l'IA a du mal à interpréter les cas avec des prothèses mammaires, donc ça ne sert à rien de l'attendre. »

Il rédige ensuite le compte rendu.

Dr S : « Bon elle est à 9. On va voir ce qu'il a indiqué. Alors ce sont des masses et il ne les montre que sur un cliché. Il met un score local de 44 ce qui n'est pas très élevé. »

Les pratiques interrogatives sur l'outil générales démontrent la fiabilité des résultats et du modèle sous-jacent. Elles relient les résultats à des règles et standards statistiques, comme l'acceptation de valeurs p inférieures à 0,1, 0,05 ou 0,001 selon les niveaux de confiance souhaités. Ces pratiques confirment la véracité des résultats grâce à des protocoles explicites.

Ainsi, les pratiques interrogatives générales s'inscrivent dans la continuité de l'induction entamée par l'outil d'IA. Elles viennent valider l'entraînement du modèle. Par conséquent, lorsque les outils d'IA sont déjà entraînés, comme dans notre étude de cas sur les soins courants, ces pratiques interrogatives ne sont plus nécessaires.

Dans l'extrait suivant, issu d'un entretien avec DataExp, nous observons comment la comparaison des performances entre des modèles linéaires et des modèles non-linéaires constitue une forme de pratiques interrogatives sur l'outil générales : elles montrent la pertinence et la possibilité d'amélioration des modèles linéaires en fonction d'une base de données particulière. DataExp s'en sert également pour déterminer l'utilité des variables extraites des antécédents médicaux par l'approche fondée sur les données.

DataExp : En termes de résultats, il y a les différents modèles que j'ai utilisé. Les deux premiers à gauche sont les modèles linéaires, les trois derniers ce sont des modèles « boîtes noire ». Mais ça permet de voir à quel point les modèles linéaires sont bons, un peu à la limite supérieure.

Comme tu peux le voir, quand je considère les antécédents médicaux, il y a vraiment un bond dans les scores. Pour préciser, en termes de prédiction, au-delà de l'amélioration des scores, ce qui diminue beaucoup c'est le nombre de faux positifs. Les faux positifs diminuent vraiment quand je considère les antécédents médicaux. »

Les pratiques interrogatives sur l'outil spécifiques évaluent l'importance d'un résultat ou d'une variable particulière, tandis que les pratiques générales s'intéressent à la cohérence globale du modèle. Même si le modèle de l'outil d'IA est considéré comme fiable, la valeur de chaque variable peut être sujette à discussion. Ces pratiques attribuent des scores pertinents en fonction du contexte d'usage de l'outil d'IA, et non de règles générales applicables à tous les contextes. Par exemple, dans le projet KOVAK, ces scores spécifiques déterminent les efforts à consacrer pour interpréter une variable donnée, nécessitant des pratiques interrogatives professionnelles spécifiques. Dans les soins courants, ces valeurs sont intégrées par les développeurs d'IA sous forme de scores locaux, aidant le radiologue à décider de prendre en compte ou non le résultat.

Dans l'exemple suivant (cas 75, jour 7), nous observons des pratiques interrogatives spécifiques s'appuyant sur les scores locaux (calcifications à 44/100) et des pratiques spécifiques développées par le Dr S en fonction de son expertise : en prédisant le résultat de l'outil d'IA, il estime comprendre le fonctionnement de l'outil, évitant ainsi de mobiliser des pratiques interrogatives professionnelles alors que le score affiché est très élevé.

Dr S à la manipulatrice : « Mme X, est-ce qu'elle a eu un cancer du sein ? »

Manipulatrice : « Oui, à droite, je ramène le dossier. »

Dr S : « Pouvez-vous faire un profil droit svp ? »

La manipulatrice part prendre le cliché supplémentaire, pendant que le Dr S examine la tomographie et compare les mammographies antérieures.

Dr S : « J'ai demandé un profil strict du côté de l'opération pour vérifier qu'il n'y a pas de signes de récidence. Je fais partie de l'école classique. »

Après un moment d'analyse, le Dr S : « Je pense que l'IA va la mettre à 10 car il y a une anomalie architecturale. »

L'outil d'IA affiche un score de 10. Le Dr S : « C'est bien ce que j'attendais, il y a une masse à 96/100, ce qui est très élevé, comme s'il y avait un cancer. C'est dans la zone d'opération. Ensuite, à gauche il détecte des calcifications à 44/100, ce qui n'est pas méchant. »

L'ensemble de ces pratiques interrogatives est synthétisé dans le tableau 5. Chaque pratique représente un mode de gestion de la collaboration humain-machine, permettant à l'expert métier de gérer l'opacité de l'outil d'IA et d'intégrer ses résultats dans leur réflexion. Nos études montrent que certaines combinaisons de ces pratiques interrogatives forment des modes de gestion supplémentaires, optimisant ainsi les bénéfices de la collaboration humain-machine.

Pratiques interrogatives du point de vue de l'expert métier	
Pratiques interrogatives professionnelles	Pratiques interrogatives sur l'outil
Ces pratiques utilisent les connaissances spécifiques d'un domaine pour interpréter les résultats fournis par un outil d'IA, en se concentrant sur leur adéquation avec le contexte d'application et en ignorant la manière dont ils ont été générés.	Ces pratiques se concentrent sur la vérification des résultats de l'outil d'IA par rapport à un protocole, en mobilisant des connaissances sur le fonctionnement de l'outil, notamment des notions statistiques. Le protocole peut être formel ou informel.
Pratiques interrogatives professionnelles générales	Pratiques interrogatives sur l'outil générales
Elles associent les résultats de l'outil d'IA à des mécanismes connus, des résultats bien établis dans la littérature, ou des	Elles démontrent la fiabilité de l'outil d'IA en fonction de normes et de conventions statistiques. Ces pratiques sont

expériences cliniques courantes. Lorsque le contexte est clair, une seule explication s'impose (abduction surcodée), réduisant ainsi l'incertitude liée à l'opacité des résultats grâce à des causalités vérifiées.	particulièrement pertinentes pendant l'entraînement des outils d'IA, car elles participent à la vérification du modèle. Elles réduisent l'incertitude liée à l'opacité en prouvant la fiabilité du modèle.
Pratiques interrogatives professionnelles spécifiques	Pratiques interrogatives sur l'outil spécifiques
Lorsque les experts peinent à interpréter un résultat, ils adaptent des stratégies en fonction des situations spécifiques. Cette analyse est plus approfondie et personnalisée (abductions sous-codées), permettant de mieux comprendre des résultats complexes.	Elles évaluent l'importance d'une variable particulière par rapport au modèle, déterminant la nécessité de la considérer. Elles réduisent l'opacité en établissant la valeur du résultat dans un contexte spécifique.

Tableau 5 : Synthèse des pratiques interrogatives identifiées sur le terrain

1.3. Les configurations des pratiques interrogatives, modes de gestion de la collaboration

La synthèse de nos deux études de cas révèle que l'utilisation de l'outil d'IA contribue à un processus de production de connaissances, que ce soit pour prendre des décisions dans les soins courants ou pour établir des connaissances en recherche médicale. Conformément à la théorie peircienne, ce processus débute par un état de doute et aboutit à une conviction. Dans l'étude RADO, la prise en charge médicale répond à une incertitude concernant l'état de santé de la patiente (existence d'un cancer du sein) et vise à former une conviction pour orienter l'action appropriée (suivi biannuel ou biopsie). De même, le projet KOVAK cherche à résoudre une incertitude (quels facteurs entraînent des effets indésirables ?) et aboutit à l'établissement de croyances permettant de passer à l'action (identification des patients à risque et mise en place d'une stratégie thérapeutique).

Ce processus est communément désigné comme une « enquête » par les pragmatistes. John Dewey remplace les termes de doute et de croyance par ceux de situations indéterminées et déterminées : « *L'enquête est la transformation contrôlée ou dirigée d'une situation indéterminée en une situation si déterminée dans ses distinctions et relations constitutives qu'elle convertit les éléments de la situation originelle en un tout unifié* » (Dewey 1986, 108).

Au-delà de ces nuances, les pragmatistes reconnaissent un processus sémiotique commun qui génère des interprétations et permet une action, telles que des prises de décision.

Par conséquent, nous pouvons concevoir les pratiques interrogatives comme des composantes de cette enquête, offrant des interprétations face à l'opacité des résultats de l'outil d'IA et les intégrant dans une réflexion globale. Elles émergent du doute (ou de la situation indéterminée) engendré par l'opacité des résultats de l'outil d'IA et conduisent à une conviction (situation déterminée). Ainsi, chaque pratique interrogative constitue un univers unique et contextuel.

Plutôt que de nous attarder sur chacune d'elles, nous nous concentrons sur l'habitude qui en découle, c'est-à-dire la propension à agir d'une certaine manière qui généralement aboutit à une conviction suffisamment proche de la vérité pour que les actions entreprises soient cohérentes avec la réalité, reflétant fidèlement l'expérience vécue. Ces habitudes se manifestent sous la forme de modes de gestion de la collaboration humain-machine que nous mettons en lumière.

En combinant nos deux études de cas et les théories sous-jacentes, nous identifions une propension à agir que nous désignons comme le mode de gestion par normalisation (1.3.1). Ce mode de gestion peut englober les quatre types de pratiques interrogatives décrites et les utiliser dans une séquence visant à valider l'hypothèse initiale de l'enquête.

Cependant, nos données empiriques et les cadres théoriques nous permettent de distinguer deux autres types de modes de gestion, à savoir le mode de gestion par approfondissement (1.3.2) et le mode de gestion par opposition (1.3.3). Bien que moins évidents sur le terrain, ces deux modes de gestion émergent grâce à la théorie, illustrant comment ils orientent le processus d'enquête vers de nouvelles hypothèses et une compréhension renouvelée de la situation, entraînant ainsi des apprentissages et une amélioration des capacités des experts métier.

1.3.1. Mode de gestion par normalisation

Le résultat de l'outil d'IA est le fruit d'une inférence inductive. L'outil d'IA effectue une induction grâce aux données empiriques. Dans le cas de KOVAK, cette induction concerne les facteurs de risque, et l'outil d'IA évalue leur robustesse en attribuant des valeurs numériques à ces variables. Les connaissances ne sont pas encore totalement fiables, et l'objectif de l'usage de l'outil d'IA est précisément de les vérifier et de mettre en évidence celles qui sont les plus pertinentes en fonction des patients dans la cohorte. Dans le cas de la mammographie, les connaissances concernent la probabilité que certaines structures sur l'image soient des signes de cancer. L'outil d'IA est déjà entraîné à cette tâche. Son rôle est alors d'évaluer dans quelle mesure les structures apprises lors de son entraînement se manifestent dans de nouvelles mammographies. Dans les deux cas, cependant, l'expérience réalisée par l'outil d'IA,

matérialisée par un résultat, doit être analysée par des intervenants humains pour finaliser le processus d'inférence inductive et parvenir à une croyance.

Comme nous le détaillons ci-dessous, le mode de gestion par normalisation commence par des pratiques interrogatives générales sur l'outil, puis des pratiques interrogatives professionnelles générales. Si une opacité persiste, elle se poursuit avec des pratiques interrogatives spécifiques sur l'outil, et enfin des pratiques interrogatives professionnelles spécifiques.

Tout d'abord, la pratique interrogative générale sur l'outil vérifie le résultat de l'outil d'IA, son adéquation à un protocole expérimental. Cette pratique est observée dans les deux études de cas, bien qu'elle ne soit explicitement mentionnée que dans le projet KOVAK. Pour les outils d'IA déjà entraînés, comme dans le cas de la radiologie, cette vérification a été effectuée en amont par les concepteurs et n'est plus nécessaire. De plus, la démonstration de la fiabilité de l'outil d'IA constitue l'un des principaux arguments de vente.

Ensuite, la pratique interrogative professionnelle générale confirme le résultat de l'outil d'IA en fonction du contexte, des croyances et des actions subséquentes. Elle permet de conclure l'enquête en produisant une abduction surcodée si une règle unique et évidente émerge du résultat. Par exemple, lorsque l'outil d'IA détecte des formations calciques, si le radiologue observe qu'elles ont une morphologie régulière, il conclut qu'elles sont bénignes.

Si cette pratique se révèle insuffisante pour dissiper l'opacité du résultat de l'outil d'IA, la pratique interrogative spécifique sur l'outil évalue l'importance d'une variable par rapport au modèle. Par exemple, le radiologue utilise les statistiques fournies par la documentation technique pour évaluer la signification d'un certain score en termes de probabilité de faux positifs, ou il s'appuie sur son expérience pour juger que le résultat sera insignifiant dans un contexte spécifique.

Si le résultat semble significatif, alors la pratique interrogative professionnelle spécifique s'efforce de formuler d'autres explications, appelées abductions sous-codées, en utilisant des stratégies adaptées à chaque situation. Par exemple, les intervenants peuvent analyser les données pertinentes pour le résultat, si elles sont accessibles, bien que cela soit rare pour les outils d'IA déjà entraînés.

Ce mode de gestion, impliquant de deux à quatre types de pratiques interrogatives, conclut la construction de connaissances dont le résultat de l'outil d'IA est un processus intermédiaire. Il teste l'induction de l'outil d'IA en vérifiant et en validant son résultat.

Il fait progresser l'enquête de manière linéaire (voir figure 19), considérant l'usage de l'outil d'IA comme une augmentation dans le registre de la conformation. La façon de mobiliser les pratiques interrogatives gère l'opacité de l'outil d'IA en recherchant une règle causale permettant d'interpréter le résultat (par abduction surcodée ou sous-codée). Si le résultat ne

se conforme pas à un protocole ou aux connaissances établies, il est exclu. Ce processus est efficace, aboutissant rapidement à la formation de croyances. Il oriente la médiation de l'expérience par l'outil d'IA de manière binaire : le résultat est interprétable, confirmant une croyance et permettant l'action, ou il est rejeté. Cette normalisation du processus de production de connaissances est utile dans un contexte d'incertitude, où une expertise professionnelle forte et des enjeux importants sont en jeu. Cependant, elle limite les possibilités d'apprentissage.

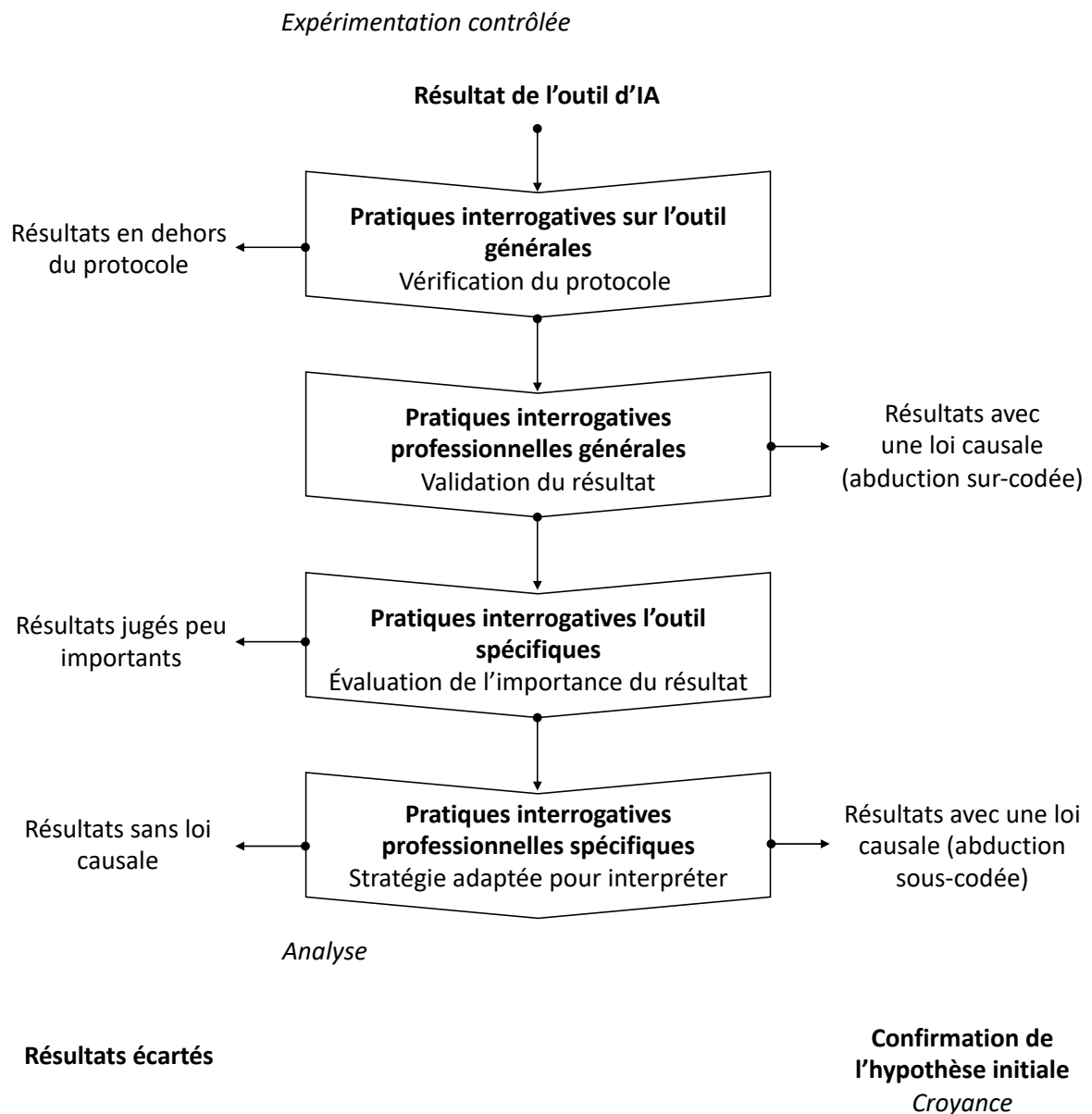


Figure 19 : Mode de gestion par normalisation

Le mode de gestion précédent favorise l'efficacité dans la création de connaissances. Il exclut les résultats qui ne correspondent pas à l'objectif préétabli. Lors de l'analyse de l'expérience contrôlée (le résultat de l'outil d'IA), cette configuration des pratiques

interrogatives oriente le résultat vers des abductions surcodées et sous-codées pour conclure l'enquête.

En revanche, un mode de gestion favorisant l'efficacité (utilisation optimale des données) serait axé sur l'apprentissage. Pour développer ce type de mode de gestion, nous revenons au nœud sémiotique constitué par l'étape d'analyse. Cette étape offre des alternatives à la simple validation (ou exclusion) des résultats. Par exemple, dans le projet KOVAK, nous observons comment l'analyse réintègre les déductions et les expérimentations réalisées. Les médiations produites à ce stade peuvent aller plus loin et conduire à de nouvelles abductions, comme pour l'une des variables finales, issue de l'approche fondée sur les données. De même, les pratiques interrogatives pourraient amener une nouvelle façon de problématiser la situation.

1.3.2. Mode de gestion par approfondissement

Le mode de gestion par approfondissement mobilise des pratiques interrogatives pour interpréter les résultats surprenants de l'outil d'IA. Ces résultats, empiriques, découlent des inductions produites par l'outil d'IA. Comme nous l'avons vu dans le projet KOVAK, lorsque le data scientist développe une approche fondée sur les données, les experts médicaux peinent à comprendre les résultats produits et ne réussissent pas à établir de liens avec leurs connaissances. Ces découvertes mettent en lumière les lacunes théoriques et incitent à développer de nouvelles abductions, à envisager de nouvelles règles pour expliquer le résultat.

Selon la typologie d'Eco (1983), le nœud sémiotique peut mener à deux types supplémentaires d'abduction : l'abduction créative et la méta-abduction. L'abduction créative consiste à formuler de nouvelles hypothèses ou lois qui n'avaient pas été envisagées auparavant, nécessitant une pensée originale et créative. Ce type d'inférence est souvent associé à des découvertes majeures qui transforment la compréhension existante dans un domaine, similaire à un changement de paradigme scientifique selon Kuhn (1962). Cette démarche conduit à la méta-abduction, qui évalue si les nouvelles hypothèses ou lois générées correspondent à la réalité telle que nous l'expérimentons. Ces abductions exigent une nouvelle enquête pour éprouver cette nouvelle croyance, contrairement aux abductions surcodées et sous-codées, dont les hypothèses découlent de l'expérience et des connaissances existantes.

Les pratiques interrogatives professionnelles sont nécessaires pour produire ces nouvelles abductions. D'habitude, elles génèrent des abductions surcodées et sous-codées : elles relient les résultats des outils d'IA aux connaissances professionnelles comme les connaissances médicales et l'expérience avec les patients. Ces pratiques peuvent aller au-delà pour développer de nouvelles hypothèses et envisager de nouveaux mécanismes. En acceptant que les connaissances soient faillibles et sujettes à des lacunes, elles peuvent aider à identifier de nouvelles sources de causalité, ce qu'on appelle l'abduction créative. De même, elles peuvent

évaluer si ces hypothèses s'appliquent concrètement grâce à leur familiarité avec le terrain, ce qui constitue la méta-abduction.

Cependant, pour mobiliser les pratiques interrogatives professionnelles dans ce sens, il est nécessaire d'identifier un résultat de l'outil d'IA propice à ce type d'abduction, afin d'éviter de formuler des hypothèses trop invraisemblables. Comme le montre le cas de KOVAK, avant d'initier ces pratiques interrogatives professionnelles, il est utile de recourir aux pratiques interrogatives sur l'outil. Ces dernières mettent en évidence les résultats trop significatifs pour être ignorés comme de simples artefacts statistiques. Par exemple, pour s'assurer de la pertinence des résultats produits, le data scientist procède d'abord à une « validation croisée » : il divise les données empiriques entre des données d'entraînement et des données de test de manière itérative pour évaluer la capacité du modèle à généraliser sur des données non vues. Le risque qu'une variable soit issue d'un hasard statistique diminue avec le nombre d'apparitions au cours des itérations, et à l'inverse, les variables qui n'apparaissent qu'une ou deux fois sur les 30 itérations réalisées par modèle sont supprimées. Par conséquent, lorsqu'une variable est sélectionnée à la fin de ce type de processus, nous supposons que le mode de gestion par approfondissement gagnerait à être employé.

1.3.3. Mode de gestion par opposition

Le mode de gestion par opposition permet de remettre en question la compréhension de la situation. D'une part, les pratiques interrogatives générales sur l'outil informent sur le nombre d'hypothèses manquantes. En vérifiant l'adéquation du modèle, tel que le taux de prédiction, elles montrent à quel point les données sont insuffisantes pour une compréhension totale. Même si les statistiques indiquent que le modèle sera toujours sous-optimal, l'écart par rapport aux données aide à évaluer le potentiel de nouvelles découvertes et la nécessité de repenser la situation différemment. D'autre part, les résultats qui ne correspondent pas au protocole expérimental représentent autant d'hypothèses non vérifiées. Comprendre pourquoi l'hypothèse initiale n'a pas été confirmée peut aider à réexaminer le problème.

Ainsi, le mode de gestion par opposition prend en compte l'autre issue possible du nœud sémiotique : changer la problématisation de la situation. Cette phase initiale de l'enquête est cruciale car elle définit ce qui est pertinent dans l'analyse du problème, en distinguant l'essentiel de l'accessoire (Dewey 1986), en précisant la portée temporelle et spatiale de l'enquête, ainsi que les acteurs impliqués en tant qu'enquêteurs et ceux exclus (Lorino et Tricard 2012). L'approche pragmatiste implique une co-émergence de problèmes et de solutions afin d'éviter de résoudre des problèmes non pertinents. Dewey (1986) met en garde contre l'erreur de rationalisation consistant à isoler et à considérer abusivement une pure abstraction comme une réalité.

Ainsi, le mode de gestion par opposition mobilise des pratiques interrogatives pour interpréter les résultats contradictoires de l'outil d'IA. Il fournit des pistes de réflexion pour formuler une nouvelle problématisation de la situation.

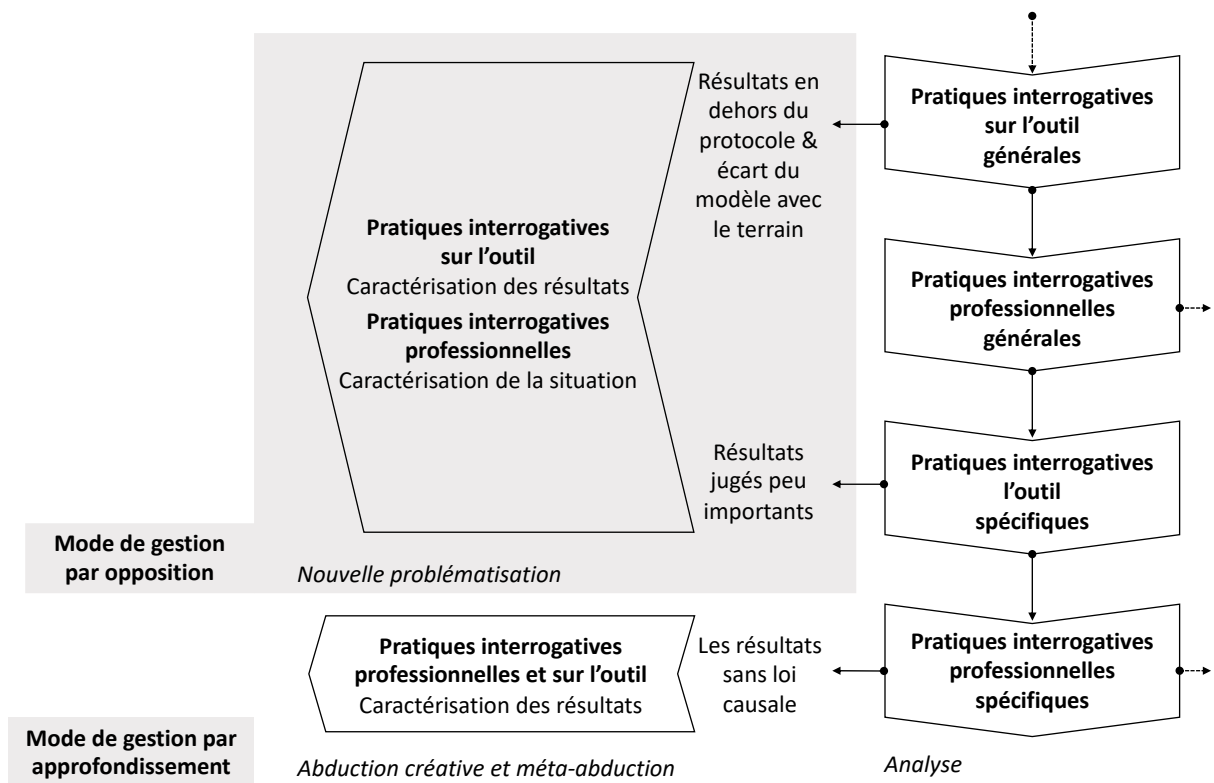


Figure 20 : Modes de gestion par opposition et par approfondissement

Nos deux études de cas montrent que, pour gérer la collaboration humain-machine au service de l'expertise métier, les pratiques interrogatives (Lebovitz, Lifshitz-Assaf, et Levina 2022) sont essentielles. En plus de démontrer un engagement de la part de l'expert métier dans la collaboration avec l'outil d'IA, elles permettent de gérer l'opacité de ses résultats et de construire des connaissances plus robustes.

Nous avons ainsi enrichi la notion de pratiques interrogatives par une typologie qui distingue celles mobilisant des connaissances professionnelles, extérieures à l'outil d'IA et se développant avec l'expertise des acteurs dans leur cœur de métier, et les pratiques interrogatives sur l'outil, nécessitant de se familiariser avec certaines particularités de l'apprentissage machine, dont la nature statistique des résultats.

Si ces pratiques interrogatives représentent un premier type de modes de gestion de la collaboration, nous avons également mis en avant que ces pratiques se combinent et forment des modes de gestion plus complexes. Le mode de gestion par normalisation aide à concrétiser des connaissances, tandis que les modes de gestion par opposition et par

approfondissement enrichissent la compréhension de la situation, apportant de nouvelles connaissances.

Mais l'élément le plus important qui ressort dans la discussion entre les deux études de cas est l'évolution de notre perspective théorique sur la collaboration entre les experts métier et les outils d'IA. Pour intégrer les pratiques interrogatives au sein d'un modèle de collaboration humain-machine et pour identifier les apports des modes de gestion à celui-ci, nous avons d'abord mobilisé le cadre de conformation/exploration (Moisdon 2005a). Ce dernier s'est avéré utile pour analyser le premier cas, mais malgré ses apports, nous avons opté pour la taxonomie des inférences de Peirce (Peirce 1878e), qui est plus englobante, et permet notamment de couvrir les phénomènes observés dans les deux études de cas.

Cette nouvelle perspective considère les outils d'IA comme des partenaires, offrant des déductions théoriques que nous intégrons maintenant au sein du modèle de collaboration qui conclut cette thèse : l'Expert Machine Collaboration Community (EMC2).

2. L'Expert Machine Collaboration Community, une collaboration ai niveau de la communauté d'enquête

La contribution de notre thèse se concrétise autour du modèle de l'Expert Machine Collaboration Community (EMC2). Ce modèle de collaboration humain-machine est issu des observations des deux études de cas et des déductions amenées par les pratiques interrogatives et la vision peircienne.

Pour présenter ce modèle, nous commençons par étayer le changement conceptuel qu'amène la vision peircienne, c'est-à-dire de considérer l'outil d'IA comme un acteur (2.1). Ensuite, nous positionnons l'EMC2 au regard de la littérature (2.2) et proposons des suggestions pour le concrétiser sur le terrain (2.3).

2.1. Considérer l'outil d'IA comme un acteur

Comme l'illustre la seconde étude de cas, l'outil d'IA peut être vu comme un partenaire au même titre que les autres acteurs humains (Anthony, Bechky, et Fayard 2023). Cette perspective conduit à reconsidérer la notion de pratiques interrogatives (2.1.1), ouvrant la voie à de nouveaux types de pratiques centrées sur l'outil d'IA (2.1.2).

2.1.1. Extension de la notion de pratiques interrogatives

Nos observations du projet KOVAK suggèrent de considérer l'outil d'IA comme un acteur collaborant avec un autre expert. Cette approche est reconnue comme une nécessité en sciences de gestion (Anthony, Bechky, et Fayard 2023). De même, dans l'ingénierie de la collaboration, de Vreede et Briggs (2019) ont considéré que « *les agents artificiels deviendront des membres à part entière des équipes* »⁶¹ (De Vreede et Briggs 2019, 103).

La littérature sur la collaboration entre experts de divers domaines montre qu'ils doivent se comprendre mutuellement pour travailler ensemble (Carlile 2004; Maguire, Hardy, et Lawrence 2004; Hardy, Lawrence, et Grant 2005; Levina 2005). Par ailleurs, la littérature sur l'expertise professionnelle suggère que ces connaissances sont souvent tacites (Polanyi 1997; 2009; Von Nordenflycht 2010; Briscoe et Murphy 2012; Gray et Silbey 2014; Huisling 2014).

Appliqués à la collaboration homme-machine, l'aspect tacite des connaissances des experts métier ne semblait pas être un problème majeur lors de la conception de l'outil d'IA jusqu'à présent. Alors que les algorithmes fondés sur l'approche symbolique, comme les systèmes

⁶¹ *artificial agents will become fully functional members of teams*

experts, nécessitent une description explicite des raisonnements des professionnels (Hatchuel et Weil 1992), l'apprentissage machine permet de reproduire l'expertise sans devoir l'expliquer. Par exemple, pour entraîner l'outil d'IA à l'analyse des mammographies, des experts ont labellisé de nombreuses radiographies en entourant les zones suspectes (Lebovitz, Levina, et Lifshitz-Assa 2021), sans avoir besoin d'expliquer pourquoi telle zone était suspecte et quels mécanismes physiologiques étaient en jeu.

Cependant, cet avantage lors de la conception devient une limite dans l'usage de la collaboration humain-machine. Par exemple, dans l'étude de cas RADO, certaines divergences entre l'outil d'IA et les décisions des radiologues sont systématiques : l'outil d'IA se trompe dans la classification des calcifications au point que certains médecins finissent par ignorer ses résultats dans ces contextes précis.

Pour assurer une collaboration plus approfondie, les outils d'IA devraient être en mesure de comprendre, dans une certaine mesure, les choix des experts métier. Ce constat nous conduit à revenir à la notion de pratiques interrogatives pour l'étendre.

Les pratiques interrogatives ont été conçues comme un moyen pour l'expert métier de gérer l'opacité des résultats de l'outil d'IA, en reconnaissant que l'outil d'IA possède un certain niveau d'expertise (Lebovitz, Lifshitz-Assaf, et Levina 2022). Cependant, cette notion restait centrée sur le professionnel dans la collaboration humain-machine, rendant les pratiques interrogatives unidirectionnelles.

Grâce à la nouvelle perspective adoptée, nous proposons d'explorer comment les outils d'IA pourraient être dotés de leurs propres pratiques interrogatives et ainsi intégrer les raisonnements des experts, continuant à apprendre et à s'améliorer de manière plus efficace.

2.1.2. Voies de recherche pour de nouvelles pratiques interrogatives

En nous appuyant sur les données du terrain et la perspective peircienne, nous déduisons deux axes de recherche pour enrichir les pratiques interrogatives liées à l'outil d'IA : les pratiques interrogatives portant sur le modèle et celles portant sur le contexte.

Les pratiques interrogatives sur le modèle visent à évaluer la cohérence entre le modèle sous-jacent de l'outil d'IA et les connaissances des experts métier. Par exemple, dans le domaine de l'aide à la lecture mammographique, les radiologues ont relevé plusieurs cas où les résultats de l'outil d'IA sont systématiquement erronés, comme dans le cas des prothèses mammaires, ou génèrent trop de faux positifs, comme dans le cas des calcifications régulières. Ces récurrences ne sont identifiées que dans la pratique courante, lorsque l'outil d'IA est confronté à de nouvelles données. Les pratiques interrogatives sur le modèle pourraient identifier ces schémas récurrents. Par conséquent, l'outil d'IA pourrait affiner son modèle en développant des corrélations supplémentaires lorsqu'il est confronté à de telles situations, optimisant ainsi la confiance dans les conclusions produites.

Ces améliorations traduisent une forme d'explicabilité des connaissances des experts métier. En examinant les contextes dans lesquels l'outil d'IA développe des optimisations à son modèle, ainsi que la nature de ces améliorations, les experts métier pourraient enrichir leur expertise et formuler de nouvelles hypothèses. Par exemple, dans le projet KOVAK, l'approche fondée sur les données a révélé de nouvelles variables, optimisant ainsi le modèle statistique. Parmi ces variables, certaines ont confirmé des connaissances existantes, tandis que d'autres ont mis en lumière des facteurs dont il serait intéressant de comprendre les mécanismes favorisant les effets secondaires de l'immunothérapie.

En réfléchissant aux pratiques interrogatives sur le modèle dans le cadre de l'aide à l'analyse mammographique, nous sommes amenés à réévaluer le rôle du concepteur et la position de l'IA en tant qu'acteur au sein du processus de prise en charge. Encourager les pratiques interrogatives sur l'outil d'IA concernant le diagnostic radiologique ne remet pas en cause l'expertise des professionnels de santé, mais vise plutôt à renforcer le fonctionnement initial de l'outil d'IA. Ces pratiques interrogatives visent à lever l'opacité dans les situations où l'outil ne fonctionne pas de manière optimale, permettant ainsi aux concepteurs d'améliorer ses performances en intégrant ces nouvelles informations.

Les pratiques interrogatives sur le contexte visent quant à elles à saisir les spécificités du terrain sur lequel l'IA est déployée. Par exemple, entre les établissements du Dr S et du Dr G, des différences importantes sont observées dans les profils cliniques des patientes, les besoins des professionnels de santé et les activités entreprises. De même, les données sur lesquelles l'IA a été entraînée proviennent souvent d'une population de patientes différente, et les experts médicaux impliqués dans sa conception ont des méthodes de travail liées à leur expérience spécifique. Le projet KOVAK rencontre des problèmes similaires avec la cohorte de patients utilisée. Malgré la volonté de généraliser, les IA sont limitées à la modélisation de phénomènes très locaux.

Pour réduire l'écart entre la situation d'entraînement de l'outil d'IA et celle dans laquelle il est appliqué, les pratiques interrogatives sur le contexte cherchent à identifier les différences empiriques. Bien qu'une amélioration immédiate des performances de l'outil d'IA à travers ces informations soit difficile à envisager, elles sont cruciales pour une nouvelle itération de l'outil. Les différences empiriques pourraient aider à reconsidérer la situation et à prendre en compte de nouvelles variables, voire à imaginer d'autres façons d'utiliser l'outil d'IA. Par exemple, dans le projet KOVAK, il aurait été pertinent d'intégrer des variables socioéconomiques et comportementales, comme recommandé dans le cadre de la gestion de la personnalisation des soins à grande échelle (Minvielle 2018). Ainsi, ces pratiques interrogatives pourraient transformer l'outil d'IA et contribuer à élargir le nœud sémiotique.

Les pratiques interrogatives, qu'elles soient axées sur le modèle ou sur le contexte, représentent un engagement des acteurs dans la collaboration humain-machine. Cependant,

considérées séparément, elles ne constituent qu'un premier mode de gestion de cette collaboration. C'est leur combinaison et leur configuration qui font progresser l'enquête, en conduisant soit à la confirmation des connaissances existantes, soit à l'acquisition de nouvelles connaissances. Dans la suite, nous présenterons l'ensemble des modes de gestion de la collaboration d'EMC2.

Pratiques interrogatives du point de vue de l'outil d'IA	
Pratiques interrogatives sur le modèle	Pratiques interrogatives sur le contexte
Analyse des divergences entre le modèle sous-jacent d'un outil d'IA et les connaissances des experts métier. Elles permettent d'identifier les récurrences de résultats erronés ou de faux positifs rencontrés afin de développer des corrélations plus précises.	Approches visant à comprendre les spécificités du terrain. Elles examinent les différences entre contextes, tels que les profils cliniques des patients, en vue d'une future itération des outils. Cela permet de repenser les situations, d'intégrer de nouvelles variables et de trouver d'autres applications des outils d'IA.

Tableau 6 : Synthèse des pratiques interrogatives du point de vue de l'outil d'IA

2.2. Construction de l'EMC2

À la lumière de nos résultats, y compris les déductions sur les pratiques interrogatives du point de vue de l'outil d'IA, nous répondons à notre question de recherche sur les modes de gestion appropriés pour une collaboration entre experts métier et outils d'IA en proposant le modèle de l'Expert Machine Collaboration Community (EMC2) (2.2.1). Grâce à la perspective peircienne adoptée, nous situons également la collaboration humain-IA au sein d'un ensemble d'acteurs plus large, appelé la communauté d'enquête (community of inquiry) (2.2.2).

2.2.1. Modes de gestion composant le modèle de l'EMC2

Grâce à une approche compréhensive, qui s'appuie sur un dialogue entre études de cas et cadres théoriques, nous construisons l'EMC2, un modèle de collaboration humain-machine entre experts métier et outils d'IA. Les modes de gestion de cette collaboration sont décrits ci-dessous et synthétisés dans la figure 21.

L'expert métier, en l'occurrence les médecins dans nos études, a la responsabilité de prendre des décisions critiques en intégrant les résultats fournis par l'outil d'IA. En effet, la responsabilité finale de la décision repose sur lui.

Dans cette collaboration, l'outil d'IA agit comme un assistant ou un support pour les professionnels de la santé. Par exemple, il est utilisé pour analyser des images, comme les mammographies, et fournir des scores de risque fondés sur les données d'entraînement. Il aide à identifier des zones suspectes qui nécessitent une attention particulière de la part du radiologue. Dans le cadre de la recherche médicale, son rôle est encore plus marqué, car il prend en charge une partie de l'analyse des données, ce qui permet de gagner un temps considérable.

L'outil d'IA n'a pas de responsabilité légale ou clinique, car il ne fait que fournir des suggestions. Cependant, il doit être conçu pour que ses modèles sous-jacents répondent à des standards statistiques rigoureux.

La communication entre humain et outil d'IA se fait principalement via des interfaces utilisateur où l'outil fournit des scores et des visualisations, que ce soit sur des zones suspectes dans des mammographies (étude de cas RADO) ou par des graphiques montrant clairement les facteurs de risque et leur importance relative (projet KOVAK). Cependant, comme le montre la seconde étude de cas, l'accès au code est essentiel si l'expert métier veut des réponses plus poussées.

En cas de divergences entre l'avis de l'expert métier et le résultat de l'outil d'IA, une première réponse consiste à mobiliser des pratiques interrogatives, allant jusqu'au mode de gestion par normalisation. Le but est de contrôler le processus de collaboration pour accélérer la confirmation des connaissances. Ce moyen permet d'écarter ou de prendre en compte le résultat de l'outil d'IA en fonction des connaissances établies. L'usage de l'outil d'IA correspond ainsi à l'aboutissement de l'inférence inductive vers la consolidation des connaissances.

Il existe cependant un autre type de réponse visant à améliorer le processus collaboratif. Cette réponse, plus difficile à mettre en place, consiste à orienter le nœud sémiotique à la fin de l'induction produite par l'outil d'IA pour produire de nouvelles connaissances, soit par une nouvelle compréhension de la situation, soit par la génération de nouvelles hypothèses.

Cette amélioration peut se faire de la part des acteurs humains grâce à des modes de gestion par opposition ou par approfondissement. De plus, nos déductions indiquent que l'outil d'IA pourrait également participer à cet effort par des pratiques interrogatives sur le modèle ou sur le contexte.

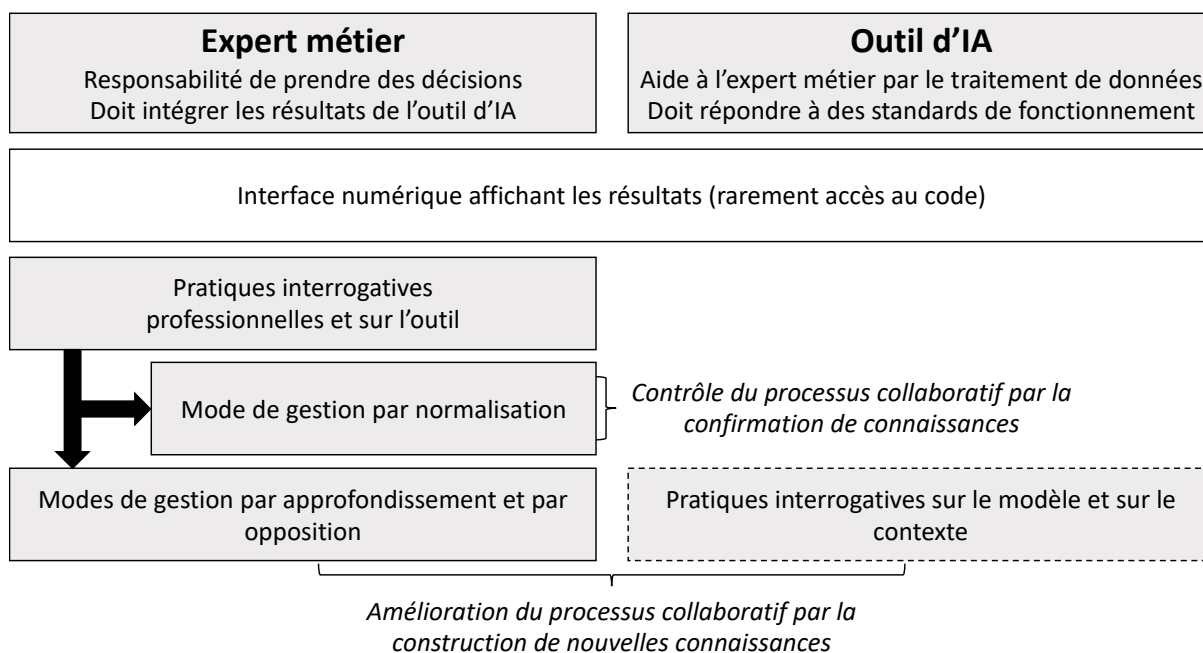


Figure 21 : Modes de gestion du EMC2

En complément, la vision peircienne nous pousse à situer l'EMC2 au niveau d'une enquête collective, ce qui nous amène à considérer l'ensemble des acteurs susceptibles d'y participer.

2.2.2. L'EMC2 dans la communauté d'enquête

L'importance du data scientist dans le projet KOVAK, ainsi que l'absence des concepteurs lors de l'usage de l'outil d'IA dans le projet RADO, nous amènent à mieux situer l'EMC2 dans l'action collective. En d'autres termes, nous souhaitons organiser davantage la collaboration humain-IA en identifiant les acteurs supplémentaires qui devraient y participer, au-delà de l'expert métier et de l'outil d'IA.

Ces considérations résonnent avec le concept d'enquête pragmatiste de Peirce. Dans notre exploration théorique, nous avons inclus les quasi-esprits comme l'ensemble des acteurs participant au processus de construction des connaissances. La question des acteurs impliqués, ou devant l'être, dans l'enquête est approfondie grâce au concept de communauté d'enquête (*community of inquiry*) (Dewey 1966; Evans 2000; Shields 2003; Lorino 2018; Berthelot 2021).

Communauté d'enquête : groupe de personnes engagées collectivement dans un processus de recherche et de réflexion critique, visant à résoudre des problèmes complexes, à prendre des décisions éclairées et à générer de nouvelles connaissances. Cette approche repose sur la coopération, l'interaction, la réflexion critique et le partage d'expériences.

Il est clair, à la lumière de nos résultats, que les concepteurs, y compris les data scientists, font partie intégrante de l'EMC2 et que leur rôle devrait être davantage développé durant l'utilisation de l'outil d'IA. Plus fondamentalement, considérer l'outil d'IA comme un partenaire

dissout même la frontière entre conception et usage, comme l'illustre le projet KOVAK. Ainsi, nous devons examiner comment la culture et la structure d'un organisme façonnent la conception des outils d'IA et comment cela se reflète dans la manière dont les utilisateurs collaborent avec l'outil d'IA (Anthony, Bechky, et Fayard 2023). Ces questions nécessitent de démêler les problèmes liés au pouvoir et à la valeur symbolique entourant l'outil. Il faut également considérer quels récits autour de l'outil d'IA sont développés et/ou contestés, et par qui. Une attention particulière doit être accordée à ceux qui développent le récit autour de l'usage de l'outil d'IA, notamment en ce qui concerne la collaboration.

Si le prisme théorique des réseaux sociotechniques offre une piste intéressante en parlant d'actants, au lieu des catégories de social et de matériel, la perspective peircienne permet d'aller plus loin dans l'analyse, comme le démontre Lorino (2014) en reprenant l'exemple de Latour (1993, 25–32) sur la ceinture de sécurité dans une voiture. Ce dernier lit la situation sociomatérielle comme une séquence de dyades A-B, B-C (conducteur/ceinture, ceinture/moteur, conducteur/moteur), où le couplage de la ceinture et du système de démarrage est un effet mécanique.

Cependant, en considérant la ceinture comme un quasi-esprit exprimant les interprétants passés, nous introduisons la figure du concepteur. Nous pouvons lire la situation comme une triade (geste du conducteur/représentation du concepteur/réactions du moteur), amenant ainsi à des questions telles que : est-ce une stratégie efficace pour améliorer la sécurité ? (Lorino 2014).

Ce raisonnement permet d'aller encore plus loin. Grâce à cette perspective, la collaboration entre l'expert métier et l'outil d'IA apparaît comme un processus « allopoïétique », c'est-à-dire tourné vers l'extérieur. Ainsi, il ne faut pas seulement prendre en compte le concepteur de l'outil, mais également la personne qui est censée en bénéficier. Par exemple, dans le cas RADO, nous pourrions analyser comment l'usage de l'outil d'IA, mobilisant une technologie généralement conçue par des hommes bien portants, influence la prise en charge d'une population féminine et éventuellement en difficulté. Ces considérations invitent ainsi à ouvrir l'analyse aux recherches féministes sur la technologie (Leavy et Harris 2018).

Par conséquent, qui – et comment – inclure dans la communauté d'enquête autour d'une collaboration humain-IA reste une voie de recherche prometteuse, rejoignant la question plus générale de savoir comment faire communauté dans nos organisations (Berthelot 2021).

2.3. Positionnement théorique de l'EMC2

Nous situons maintenant l'EMC2 au sein de la littérature sur les modèles de collaboration humain-machine (2.2.1) et discutons de la manière dont la vision peircienne enrichit les prismes théoriques pour comprendre les usages des outils d'IA (2.2.2).

2.3.1. L'EMC2 comme modèle de collaboration humain-machine

Au sein de la typologie des modèles de collaboration établie dans la revue de littérature, nous pouvons situer l'EMC2 dans le prolongement du tournant post-cognitivist (Skagestad, 1993; Skagestad 1996; Skagestad 2004; P. Steiner 2013) comme indiqué dans la figure 22. Il présente ainsi un type de collaboration humain-IA dans lequel la technologie est une extension de l'intellect humain.

En outre, l'EMC2 répond à l'appel de Babushkina et Votsis (2022) pour une épistémologie hybride, fusionnant les processus cognitifs humains et les procédures algorithmiques des outils d'IA. Cela crée un nouveau paradigme pour la génération, la manipulation et l'interprétation des connaissances. L'EMC2 reconnaît les outils d'IA comme produisant un type unique de produit épistémologique qui complète, au lieu de remplacer, la cognition humaine.

L'EMC2 converge également avec le concept d'augmentation de l'intelligence, adoptant la perspective de Licklider (1960) sur une symbiose humain-ordinateur pour résoudre des problèmes et prendre des décisions complexes. Il cherche à accroître l'intelligence de l'expert métier et à exploiter la technologie pour étendre les capacités cognitives et créatives humaines (Engelbart 1962).

Ainsi, la thèse contribue à cette littérature en proposant un ensemble de modes de gestion plus complet, issu d'observations empiriques. L'EMC2 ne se limite pas à l'augmentation de l'intelligence humaine (Zhou et al. 2021). En explicitant des modes de gestion garantissant une collaboration humain-machine efficace, l'EMC2 reconnaît également la nécessité d'améliorer les outils d'IA.

Enfin, la contribution majeure à cette littérature est une conséquence de notre approche empirique, couplée à la vision peircienne. Grâce à ce dialogue entre le matériau et la littérature, nous soulignons l'importance de la place de la collaboration humain-IA dans

l'organisation, une question qui mériterait d'être approfondie dans de futures recherches.

**Collaboration humain-machine
focalisée sur la tâche à accomplir**

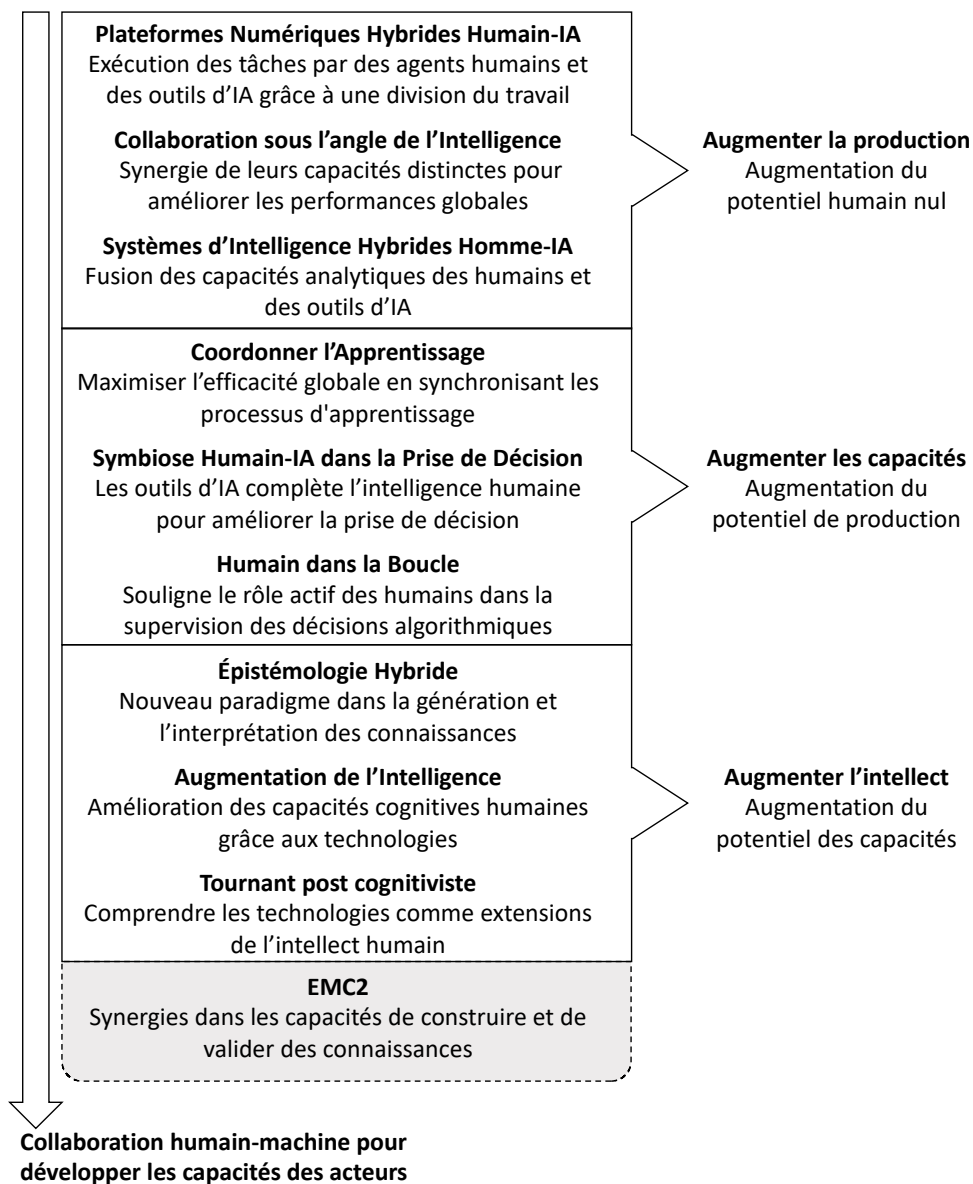


Figure 22 : Positionnement de l'EMC2 dans les modèles de collaboration humain-machine

2.3.2. L'enquête pragmatiste de Peirce et la compréhension des usages des outils d'IA

Face à l'émergence des nouvelles technologies numériques, de nombreux chercheurs appellent à développer des théories plus adaptées. Par exemple, Faraj et Leonardi (2022) soulignent que les concepts traditionnels de « technologie » peinent à saisir la nature dynamique et en constante évolution de la numérisation. En réponse, diverses approches ont été proposées, comme l'approche relationnelle (Bailey et al. 2022; Pakarinen et Huisling 2023) ou l'approche reliant les phases de conception et d'usage (Bailey et Barley 2020; Anthony, Bechky, et Fayard 2023).

Le modèle EMC2, en s'appuyant sur une perspective théorique issue des travaux de Peirce (1878e; 1887), nous semble offrir une avenue de recherche prometteuse pour enrichir les autres approches théoriques dans la compréhension des usages des nouvelles technologies.

À ce titre, nous considérons le concept d'enquête pragmatiste de Peirce particulièrement adapté à l'étude des usages des outils d'IA. D'abord, il dépasse la critique dualiste de la sociomatérialité et approfondit l'analyse des actants et de leurs relations par rapport aux réseaux sociotechniques.

Le terme même de « sociomatérialité », bien qu'il représente une fusion de la « socialité » et de la « matérialité », renvoie inévitablement à un dualisme entre ces deux termes (Lorino 2018). Pour aller plus loin, il est nécessaire de se libérer de cette catégorisation duale. Les réseaux sociotechniques dépassent cette distinction grâce au terme d'actants et à une ontologie relationnelle où les entités à analyser sont les relations entre actants.

Le terme peircien de quasi-esprit (Peirce 1931, 4.551), qui désigne les entités capables de produire des inférences et participant à l'enquête pragmatiste, constitue une alternative. Comme nous l'avons vu précédemment (2.2.2), comparé aux réseaux sociotechniques, la vision peircienne offre une perspective processuelle « allopoïétique » (Lorino 2014), prenant en compte la temporalité d'une situation avec des interprétants passés et des actions tournées vers un futur envisagé.

De plus, le concept d'enquête pragmatiste de Peirce offre une série de concepts particulièrement utiles pour l'analyse des usages des outils d'IA. Si nous avons déjà illustré la puissance explicative du quasi-esprit et de la théorie des inférences pour intégrer l'outil d'IA dans un processus de construction de connaissances, Peirce propose encore d'autres concepts pertinents. Par exemple, Anthony, Bechky, et Fayard (2023) remarquent que la perspective sociomatérielle (et la perspective des réseaux sociotechniques) tend à dire que « *les humains et les artefacts sont mutuellement constitués* »⁶² (Suchman 2007, 268). Cependant, Suchman soutient qu'il faut explorer les asymétries et distinguer « la particularité des acteurs humains »⁶³ (Suchman 2007, 270).

Les notions de contrôle et d'auto-contrôle de Peirce nous semblent offrir dans ce contexte une réponse. Elles permettent d'expliquer pourquoi les outils d'IA sont plus puissants et problématiques que de simples règles statistiques : en raison de leur capacité accrue d'auto-contrôle pour trouver des solutions sans intervention humaine. En d'autres termes, ces

⁶² *humans and artifacts are mutually constituted*

⁶³ *the particularity of human actors*

notions offrent une granularité à la notion d'intentionnalité (Pickering 1995; Leonardi 2012), traditionnellement mobilisée pour dire que les « *machines n'ont pas d'intentionnalité inhérente, indépendamment de la manière dont elles sont employées pour offrir des possibilités aux êtres humains* »⁶⁴ (J. R. Taylor et al. 2000, 137).

Ainsi, la vision peircienne permet de reconsidérer la technologie, la voyant comme une extension des capacités humaines et non comme une entité distincte. Elle éclaire différemment les outils d'IA, montrant à quel point ils sont le fruit d'un collectif d'acteurs, d'une communauté d'enquête qui dépasse les dichotomies classiques de conception/usage, humain/machine ou automatisation/augmentation. En particulier, l'inférence abductive n'est pas le résultat d'un seul individu (ou d'un artefact), mais d'un effort collectif. Peirce dépasse une vision individualisée de la logique. Lorsqu'on se concentre sur le collectif qui développe et utilise les outils d'IA, l'histoire de l'IA décrite dans la première partie du chapitre prend un nouveau sens. Les exemples précédents (l'algorithme MYCIN, le projet AliveCor, etc.) résultent d'un ensemble d'inférences partiellement automatisées par l'outil d'IA et nécessitant la contribution des acteurs humains pour les compléter.

2.4. Deux leviers pour mettre en place de l'EMC2

En considérant que « *toute la fonction de la pensée est de produire des habitudes d'action* »⁶⁵ (Peirce 1992, 131), nous proposons deux leviers d'action pour rendre l'EMC2 plus concret sur le terrain, répondant ainsi à une recherche en lien avec les acteurs et leurs préoccupations. Dans un premier temps, nous évoquons la promotion des modes de gestion auprès des experts métier (2.4.1). Ensuite, nous explorons la possibilité d'intégrer des pratiques interrogatives dans l'outil d'IA (2.4.2).

2.4.1. Promouvoir les modes de gestion de l'EMC2 auprès des experts métier

Nous proposons de promouvoir les pratiques interrogatives et les modes de gestion de la collaboration EMC2 auprès des experts métier par leur propre participation, un encadrement managérial et l'implication des concepteurs d'outils d'IA.

Nos recherches montrent que les experts métier possèdent déjà des pratiques interrogatives professionnelles. Toutefois, sensibiliser davantage à ces pratiques suggère de

⁶⁴ *machine artifacts have no inherent intentionality, independent of their being harnessed to or offering possibilities to humans*

⁶⁵ *The whole function of thought is to produce habits of action*

créer des dispositifs supplémentaires, comme des plateformes ou des forums où les experts peuvent partager leurs expériences, succès et difficultés avec les outils d'IA.

Une difficulté majeure, soulignée par le Dr S, réside dans la crainte du jugement interprofessionnel. Certains métiers, comme la radiologie, sont fondés sur des dynamiques d'autorité, rendant difficile le partage des difficultés et toute forme de vulnérabilité de l'expertise. Ces plateformes devraient donc créer des espaces sûrs, par exemple en adoptant une forme d'anonymat ou en garantissant une modération des contenus.

Les pratiques interrogatives concernant l'outil sont encore émergentes. Aller plus en avant suppose, nous, il semble-t-il, nécessaire de proposer des formations spécifiques. Des ateliers et sessions de formation pourraient mettre en avant les avantages de la collaboration humain-machine selon le modèle EMC2, incluant des études de cas, des simulations et des démonstrations pratiques pour illustrer comment ces pratiques peuvent améliorer les capacités des experts métier.

Les travaux de cette thèse mettent en lumière les modes de gestion découlant des articulations des pratiques interrogatives. Si le mode de gestion de normalisation apparaît assez spontanément, nous soulignons l'importance de mettre en place les modes de gestion d'opposition et d'approfondissement.

De plus, il apparaît crucial d'intégrer l'apprentissage des pratiques interrogatives dès le début de la formation des futurs experts métier pour instaurer une culture d'interrogation et de critique constructive dès les premières étapes de leur carrière.

Ces efforts doivent être soutenus par les managers. Comme l'ont identifié Béjean, Kletz, et Moisdon (2018, 22) concernant la mise en place du dossier patient informatisé à l'hôpital, le manque d'accompagnement est un frein majeur à l'augmentation de la valeur de cet outil dans le registre de l'exploration : *« si [les potentialités d'exploration] ne se sont que très peu réalisées, c'est sans doute du fait d'un effort essentiellement axé sur les processus d'adoption et de déploiement, certes longs et difficiles, mais remettant à plus tard des réflexions sur l'appropriation de l'outil d'un niveau supérieur. »* Le besoin d'accompagnement pour réaliser le potentiel d'apprentissage des outils de gestion (de Vaujany 2005; Ologeanu-Taddei et al. 2015) s'applique également à l'usage des outils d'IA.

Enfin, nous soulignons le rôle des concepteurs pour promouvoir les modes de gestion de la collaboration auprès des experts. Les data scientists, par exemple, possèdent des connaissances plus élaborées sur les pratiques interrogatives des outils. L'outil d'IA pourrait servir de plateforme pour échanger les bonnes pratiques, nécessitant une nouvelle approche dans la conception des outils d'IA.

2.4.2. Concevoir autrement les outils d'IA

Nos observations sur les pratiques interrogatives convergent avec la littérature (Lebovitz, Lifshitz-Assaf, et Levina 2022) en suggérant qu'elles offrent une troisième approche pour gérer l'opacité des outils d'IA, complémentaire mais sans se substituer aux deux premières.

Reposant sur un principe de précaution, un axe de recherche se focalise sur la restriction de l'usage des outils d'IA pour des décisions critiques lorsque la transparence est inatteignable (Gillespie 2014; Burrell 2016; Teodorescu et al. 2021).

La deuxième approche, l'*Explainable Artificial Intelligence** (XAI), vise à améliorer la conception des outils d'IA pour les rendre moins opaques. Cette discipline se concentre sur la phase de conception pour établir des directives visant à réduire l'opacité des outils d'IA, tout en développant des méthodes d'évaluation pour détecter les risques de biais (Meske et al. 2022).

La notion de pratiques interrogatives met en lumière une troisième voie où les professionnels gèrent l'opacité des outils d'IA en situation. Elle se concentre ainsi davantage sur les aspects liés à l'utilisation et aux répercussions organisationnelles, qu'il est également nécessaire de prendre en considération dans l'XAI (Hafermalz et Huysman 2021; Kostick-Quenet et Gerke 2022; Meske et al. 2022).

Plus précisément, l'EMC2 propose d'intégrer les connaissances des experts métier lors de l'utilisation des outils d'IA afin de générer de nouvelles possibilités de conception dans les versions ultérieures. En plus d'améliorer la performance des outils, cette approche vise à créer des espaces de partage de connaissances entre les outils d'IA et les experts métier.

Certains outils d'IA fonctionnent sur un mode d'apprentissage par renforcement, qui se rapproche des pratiques interrogatives car ils peuvent mettre à jour les modèles en tenant compte de nouvelles données. Toutefois, cette technique présente des limites, notamment son inadéquation avec d'autres formes d'apprentissage machine et le manque de contrôle sur l'évolution de l'outil, pouvant conduire à des modèles aberrants.

Pour pallier ces limites, nous proposons de renforcer la supervision humaine dans la mise à jour des outils d'IA, maximisant ainsi l'apport de connaissances des experts métier. Par exemple, dans l'analyse mammographique, il serait utile de caractériser les zones suspectes en tenant compte non seulement de la forme, de la régularité et de la position des anomalies, mais aussi des nuances plus fines spécifiques à certains cas. Cela nécessite de recueillir les retours des experts métier et de s'engager dans des inférences abductives pour améliorer le modèle statistique de l'outil d'IA dans les directions souhaitées.

Nous suggérons également de mettre en place des pratiques interrogatives sur le contexte. Dans l'analyse mammographique, par exemple, il serait bénéfique d'intégrer des données

contextuelles en dehors de l’image, comme l’historique médical de la patiente et les raisonnements derrière les conclusions des examens. Cette intégration permettrait de découvrir des corrélations plus pertinentes et de développer des théories plus précises sur les symptômes du cancer du sein.

Pour que ces pratiques soient efficaces, il est crucial de créer des synergies entre experts métier et concepteurs d’outils d’IA. Les data scientists, qui possèdent des connaissances approfondies sur les pratiques interrogatives des outils, doivent travailler en étroite collaboration avec les experts métier pour adapter les outils d’IA aux besoins spécifiques des utilisateurs finaux. Cette collaboration pourrait se matérialiser par des ateliers, des formations et des plateformes d’échange, favorisant une co-construction des solutions.

En résumé, la conception des outils d’IA doit aller au-delà des aspects purement techniques pour inclure des solutions organisationnelles qui favorisent l’interrogation continue et la supervision humaine. En intégrant les pratiques interrogatives sur le modèle et le contexte, et en créant des synergies entre experts métier et concepteurs, nous pouvons développer des outils d’IA plus transparents, performants et adaptés aux besoins des utilisateurs. Cette approche enrichit le modèle EMC2 en le situant dans un cadre d’IA explicable et contextuelle, assurant ainsi une meilleure intégration et utilisation des technologies d’IA sur le terrain.

Promouvoir les modes de gestion de l’EMC2 auprès des experts métier	Intégrer les connaissances des experts métier mobilisées pendant l’utilisation
1. Amélioration des pratiques interrogatives professionnelles et sur l’outil	1. Recueillir les retours des experts métier plus systématiquement
2. Mis en avant des modes de gestion par approfondissement et par opposition.	2. Approfondir des collaborations entre experts métier et concepteurs
→ Besoin d’impliquer les managers et les concepteurs	→ Besoin d’intégrer conception et usage de l’outil d’IA

Figure 23 : Recommandations managériales

Conclusion

Si les progrès techniques permettent de sortir les outils d'IA fondés sur l'apprentissage machine des laboratoires de recherche pour les intégrer dans les organisations, il reste encore à déterminer comment le faire efficacement. En effet, le perfectionnement de ces outils d'IA leur permet de rivaliser avec des experts métier sur des tâches spécifiques, promettant des changements structurels dans les modes de production, une promesse suscitant autant de craintes que d'espoirs. Notre thèse contribue à établir des modes de gestion de la collaboration humain-IA, et en ouvrant à une perspective collective de cette collaboration applicable à l'ensemble de l'organisation.

Les outils d'IA d'apprentissage machine découlent d'un paradigme algorithmique conçu pour dépasser l'expertise tacite des professionnels et modéliser des phénomènes à travers des données et des méthodes statistiques. En effet, cette technologie se distingue par sa capacité à apprendre à partir de données. Contrairement aux logiciels traditionnels qui suivent des instructions précises codées par des humains, les outils d'IA peuvent ajuster leurs comportements et améliorer leurs performances à mesure qu'ils traitent plus de données. Les outils sont alors capables d'automatiser des tâches cognitives comme la reconnaissance d'images, la compréhension du langage naturel, et la prise de décision, parfois même à un niveau équivalent à celui d'un professionnel. Cependant, la façon de produire les résultats des outils d'IA reste difficile à cerner, ce qui peut poser un problème à l'heure de prendre des décisions importantes.

Les outils d'IA ont plus de difficultés à transformer radicalement certains secteurs qu'il n'y paraît, notamment le système de soins (Perez-Torrents 2022). Ce secteur présente un contexte de collaboration humain-IA qui peut être jugé comme un cas extrême : en raison de la haute expertise requise, de l'incertitude et des risques associés aux décisions, ainsi que des pressions de production, le résultat de cette collaboration est lourd de conséquences ; en parallèle, les résultats des outils d'IA sont souvent opaques et difficiles à intégrer pour les professionnels de santé, qui restent garants des jugements critiques.

Bien que la littérature propose de nombreux modèles de collaboration humain-machine pour les outils d'IA, ceux-ci, essentiellement conceptuels, s'avèrent peu ancrés dans le vécu des acteurs. Les recherches en Théorie des Organisations et en Gestion des Systèmes d'Information montrent que ces outils d'IA ont des caractéristiques particulières à prendre en compte, comme l'opacité des résultats, qui peuvent aider à faire évoluer les théories sur l'usage de la technologie. De plus, les usages observés des outils d'IA suscitent une variété de réponses des acteurs, démontrant la nécessité d'ancrer tout modèle de collaboration humain-machine sur des données empiriques.

Notre thèse s'inscrit dans cette perspective, en partant de la notion de pratiques interrogatives. Face à nos deux études de cas d'usages d'outils d'IA dans le système de soins (RADO et KOVAK), nous cherchons à développer la notion de pratiques interrogatives en des modes de gestion plus complets. Cette question se veut à la fois descriptive, ancrée dans l'observation des acteurs, et normative en proposant une collaboration entre experts métier et outils d'IA efficace et responsable.

Lors de notre démarche de recherche, nous avons établi des premiers résultats grâce au cadre d'analyse de la conformation/exploration en analysant notre première étude de cas RADO (Moisdon 2005a). Cependant, la seconde étude de cas KOVAK nous a conduit à faire émerger le concept d'enquête pragmatiste de Peirce. La synthèse de cette discussion entre terrain et littérature, nous a enfin amené à établir un modèle de modes de gestion de la collaboration homme-machine nommé EMC2.

En raison de notre méthode compréhensive, ces résultats sont limités. Comme toute étude de cas, les connaissances produites restent à consolider, notamment par des approches quantitatives. Nous rencontrons également des limites sur l'échantillon des données, qui, malgré les stratégies de triangulation, reste limité à des acteurs ayant souhaité répondre aux questions ou être observés. Par exemple, des radiologues moins expérimentés n'ont pas souhaité être observés, ce qui aurait pu révéler d'autres comportements vis-à-vis des outils d'IA.

De plus, notre recherche théorique s'est orientée pour répondre au mieux à notre matériau. Si cette démarche nous a permis de faire dialoguer et de comparer des cadres théoriques pertinents, nous ne prétendons pas avoir réalisé une étude exhaustive sur les possibilités, et nous reconnaissons avoir ressenti une influence plus marquée par les théories post-cognitivistes. Outre ce biais de sélection des cadres théoriques, nous reconnaissons que l'identification d'un problème à partir du terrain rend plus difficile la progression d'une théorie en particulier. Nous avons surtout mobilisé et appliqué des théories, mais nous aurions souhaité faire progresser ces concepts davantage.

Malgré ces limites, nous pensons avoir contribué à la communauté académique et à la société en apportant des perspectives sur un objet technique complexe et encore difficile à cerner. Nous résumons ces contributions en deux grandes parties.

Dans un premier temps, nous espérons que notre thèse contribue aux apports du pragmatisme pour les sciences de gestion. Comme le remarque Lorino (2018), les idées de Peirce ont historiquement exercé une influence limitée sur les études en sciences de gestion. On retrouve plus récemment une influence en marketing (Ogilvie et Mizerski 2011) et en sociologie des marchés financiers (Muniesa 2007), ainsi qu'en théorie des organisations (Czarniawska 1999a; 1999b) et en méthodologie (Alvesson et Kärreman 2007; Van Maanen,

Sørensen, et Mitchell 2007; Hansen 2008; Locke, Golden-Biddle, et Feldman 2008; Alvesson et Sköldbberg 2018). A l'intersection des domaines de la Gestion des Systèmes d'Information et des Théories des Organisations, notre modèle EMC2 vise à dépasser les dichotomies habituelles, entre augmentation et automatisation, entre conception et usage, et plus largement entre humain et machine.

Dans un second temps, nous espérons que notre thèse aide les acteurs sur le terrain à mieux appréhender les outils d'IA. Pour illustrer ces apports, nous montrons comment nous avons bénéficié de nos propres résultats en utilisant ChatGPT.

Nous revenons d'abord sur le premier usage détaillé dans l'introduction. Nous mobilisons des pratiques interrogatives professionnelles générales pour vérifier les propositions de ChatGPT. Ensuite, à l'aide de pratiques interrogatives sur l'outil spécifiques, nous excluons celles que nous avons identifiées comme des artefacts de l'outil dans notre contexte de travail (par exemple, il tend à ajouter « essentiellement » partout et propose l'anglicisme « basé sur »). Ainsi, nous établissons un mode de gestion par normalisation. L'outil d'IA permet d'effectuer une induction par rapport aux règles d'écriture, et nous confirmons ou excluons les résultats qu'il produit pour achever cette micro-enquête. Par conséquent, nous pouvons corriger beaucoup plus rapidement et de manière sûre, réduisant le coût cognitif de la relecture. Cependant, cet usage de ChatGPT nous permet seulement de confirmer des connaissances, non d'en construire de nouvelles.

Plus intéressant est de comprendre comment l'usage de ChatGPT peut nous aider à compléter un processus abductif, et donc à nous apporter des apprentissages. D'abord, remarquons que l'interaction avec l'outil se fait en langage vernaculaire. Qui plus est, cet outil d'IA possède une certaine réflexivité, dans le sens où, dans certains contextes, nous pouvons lui demander les informations dont il a besoin pour exécuter au mieux une tâche. Nous pouvons ainsi nous engager dans une forme de dialogue pour réfléchir hors de nos schémas de pensée classiques. Autrement dit, l'outil d'IA s'offre comme partenaire pour nous aider à verbaliser nos idées, nous permettant de creuser une question (mode de gestion par approfondissement) ou d'ouvrir le champ à de nouvelles perspectives (mode de gestion par opposition).

Par exemple, des retours sur de précédentes versions du manuscrit indiquaient que la double complexité du terrain, entre le domaine de l'Intelligence Artificielle et celui de la santé, rendait celui-ci difficile à comprendre. Par curiosité, nous avons demandé des solutions à ChatGPT pour rendre le texte plus accessible sans perdre certaines nuances importantes du terrain. Nous avons alors mobilisé des pratiques interrogatives professionnelles pour caractériser notre terrain, puis des pratiques interrogatives sur l'outil pour trier les réponses offertes et continuer l'échange avec cet outil d'IA. Nous avons ainsi abouti à la mise en place d'un glossaire auquel ChatGPT a également participé en identifiant les termes à ajouter et les

descriptions à inclure. Nous avons donc mobilisé des pratiques interrogatives dans un mode de gestion par opposition, nous permettant de remettre en question notre compréhension de la situation (tension entre simplifier trop le texte ou le laisser inaccessible). Cet apprentissage nous a permis d'aborder autrement le travail de rédaction.

Cependant, à l'heure de conclure la rédaction, nous devons formuler une limite à l'usage de ChatGPT. Bien qu'il puisse être utilisé comme correcteur ou partenaire d'échanges intellectuels, son bénéfice principal réside dans son accessibilité. En termes de qualité, nous avons trouvé les corrections et les remarques de nos amis bien plus précieuses, aussi bien par leur sensibilité stylistique que par leur regard différent. De même, les échanges avec nos pairs sont incommensurables par rapport à ChatGPT. La technologie reste une extension de l'intelligence humaine et peut s'avérer très avantageuse, notamment dans un contexte de production, mais elle reste un moyen. Nous n'avons trouvé de vraie richesse nulle part ailleurs que dans le dialogue avec autrui.

Glossaire

Terme	Description
Algorithme LASSO (Least Absolute Shrinkage and Selection Operator)	Une méthode pour rendre un modèle de prédiction plus simple et plus précis en éliminant certaines informations moins importantes.
Apprentissage machine (Machine Learning)	Méthode d'analyse de données qui automatise la création de modèles statistiques.
Apprentissage non supervisé	Apprentissage machine où les modèles apprennent à partir de données non étiquetées.
Apprentissage par renforcement	Apprentissage où un agent apprend à réaliser des tâches par essais et erreurs.
Apprentissage profond (Deep Learning)	Apprentissage machine fondé sur des réseaux très denses de « neurones » artificiels, permettant de saisir des structures complexes dans les données.
Apprentissage supervisé	Apprentissage machine où les modèles sont entraînés sur des données étiquetées.
Données massives (Big Data)	Grands ensembles de données complexes, difficiles à traiter avec des méthodes traditionnelles.
Cohorte prospective	Une étude de recherche où un groupe de personnes (cohorte) est suivi sur une période pour observer les résultats ou les effets de certaines variables ou traitements. Contrairement aux études rétrospectives, elle collecte des données au fil du temps.
Computationalisme	Théorie selon laquelle la pensée humaine est une forme de calcul.
Connexionnisme	Approche en IA qui modélise les processus mentaux comme des réseaux de connexions simples.
Courbe ROC (Receiver Operating Characteristic)	Un graphique qui aide à visualiser et à évaluer à quel point un modèle de prédiction est précis.

CRISP-DM (Cross-Industry Standard Process for Data Mining)	Un modèle de processus standard pour la conduite de projets de développement des outils d'IA (voir annexe 7).
Cybernétique	Science de la communication et du contrôle, notamment dans les machines et les organismes vivants.
Data scientist	Un professionnel spécialisé dans l'analyse, l'interprétation, et l'exploitation de données complexes. Il utilise différentes techniques pour générer des informations et aider à la prise de décisions.
Effets indésirables	Effets secondaires ou réactions négatives qui peuvent se produire lors de l'utilisation d'un médicament ou d'un traitement.
Explainable Artificial Intelligence ou l'Intelligence Artificielle Explicable ou XAI	Domaine de recherche pour développer des outils d'IA transparents et dont les décisions peuvent être comprises par les humains.
Heuristiques	Techniques adaptées des connaissances empiriques et permettant de résoudre des problèmes complexes rapidement, avec des solutions pratiques et non optimales.
Hyper-paramètres	Ce sont comme des réglages que l'on ajuste avant de commencer à entraîner un outil d'IA pour aider à obtenir de meilleurs résultats.
Immunothérapie	Un type de traitement médical qui aide le système immunitaire du corps à combattre des maladies, souvent utilisé pour traiter le cancer.
Médecine fondée sur les preuves (evidence-based medicine ou EBM)	Approche médicale théoriquement fondée sur des preuves scientifiques rigoureuses.
Médecine personnalisée ou de précision	Approche médicale visant à personnaliser le traitement en fonction des caractéristiques individuelles.
(Non) linéaire (modèle, algorithme d'apprentissage, régression...)	Un type de modèle statistique qui suppose une relation (non) linéaire entre les variables indépendantes et la variable dépendante.
Neurone artificiel	Unité de base dans un réseau de neurones, inspiré des neurones du cerveau humain.
Normalité (test de)	Vérification de la distribution normale des données, c'est-à-dire d'une distribution « en cloche », une condition souvent requise pour l'utilisation de théorèmes statistiques.

Médecine centrée sur le patient (patient-centered care)	Approche de la médecine où les préférences, besoins et valeurs du patient guident toutes les décisions. La centricité-patient est définie comme « <i>le fait de placer le patient au premier plan dans une relation ouverte et durable avec celui-ci, afin de réaliser de manière respectueuse et compatissante la meilleure expérience et le meilleur résultat pour cette personne et sa famille</i> » (Yeoman et al. 2017, 80)
Réduction de la dimensionnalité	Une méthode pour simplifier les données complexes en ne gardant que les parties les plus importantes, rendant l'analyse plus facile et plus claire.
Réseaux de neurones artificiels	Modèles informatiques inspirés du réseau neuronal du cerveau humain.
Rétroaction (Feedback)	Processus où l'information sur les résultats d'une action est retournée à l'origine de l'action.
Score AUC (Area Under the Curve)	Une façon de mesurer à quel point un modèle de prédiction est bon pour distinguer entre deux options, comme dire si un courriel est spam ou non.
Score F1	Un score qui combine deux aspects importants d'un modèle de prédiction : sa précision et sa capacité à ne pas manquer de résultats importants (le rappel).
Systèmes experts	Programmes informatiques simulant le jugement et le comportement d'experts humains.
Traitement automatique du langage naturel (TALN)	Branche de l'intelligence artificielle qui traite de l'interaction entre les ordinateurs et le langage humain, en particulier comment programmer les ordinateurs pour traiter et analyser de grandes quantités de données en langage naturel.
Variables indépendantes/variables dépendantes	Une variable indépendante est manipulée ou contrôlée pour observer son effet sur une variable dépendante, la variable d'intérêt mesurée dans l'étude. Dans le projet KOVAK la variable dépendante est l'apparition d'effets indésirables.

Liste des figures

Figure 1 : Plan de thèse	19
Figure 2 : Ngram Viewer sur les vagues liées à l'IA	23
Figure 3 : Résultat d'une spectrographie de masse de l'ester méthylique de l'acide abiétique	31
Figure 4 : Graduation des modèles de collaboration humain-IA.....	89
Figure 5 : Représentation des concepts de la sociomatérialité (Leonardi, 2013)	111
Figure 6 : Schéma situant les approches théoriques selon leur façon de concevoir la réalité	113
Figure 7 : Schéma situant les approches théoriques selon leur conception de l'agency ..	115
Figure 8 : Description du poste d'analyse des mammographies du Dr G.....	137
Figure 9 : Exemple de template (diagnostic 96, jour 8).....	142
Figure 10 : Schéma temporel des actions prises entre les acteurs dans la prise en charge	157
Figure 11 : Schéma spatial des actions prises entre les acteurs dans la prise en charge ..	157
Figure 12 : Schéma de la narration du projet KOVAK selon le modèle CRISP-DM	184
Figure 13 : Interaction durée d'administration et toxicité	191
Figure 14 : Comparaison des modèles adaptatifs BIC (en haut) et LASSO (en bas) : influence des variables sur les effets indésirables (en orange les favorisant)	193
Figure 15 : Schéma de la progression des pratiques interrogatives	200
Figure 16 : Diapositive de DataExp sur le résultat de l'outil d'IA issu de l'approche fondée sur les données	201
Figure 17 : Schéma des possibles progressions des pratiques interrogatives.....	208
Figure 18 : Schéma du projet KOVAK et de ses acteurs selon les éléments de l'enquête. 226	
Figure 19 : Mode de gestion par normalisation	244
Figure 20 : Modes de gestion par opposition et par approfondissement.....	247
Figure 21 : Modes de gestion du EMC2	254
Figure 22 : Positionnement de l'EMC2 dans les modèles de collaboration humain-machine	257

Figure 23 : Recommandations managériales	262
Figure 24 : Extrait d'un transcrit d'une consultation MYCIN pour un patient à Stanford (Pt538) (Buchanan et Shortliffe, 1984, 709).....	274
Figure 25 : Schéma des approches structurelles sur l'usage de la technologie dans l'organisation (Leonardi 2013).....	278

Liste des tableaux

Tableau 1 : Comparaison de l'ontologie et l'épistémologie des approches théoriques	113
Tableau 2 : Comparaison des notions de matérialité, de social, de sociomatérialité, d'agency et de pratiques	114
Tableau 3 : Comparaison des performances des radiologues avec et sans les outils d'IA (Conant et al. 2019)	160
Tableau 4 : Configurations des diagnostics du Dr S	163
Tableau 5 : Synthèse des pratiques interrogatives identifiées sur le terrain.....	241
Tableau 6 : Synthèse des pratiques interrogatives du point de vue de l'outil d'IA.....	252
Tableau 7 : Entretiens avec les radiologues	282
Tableau 8 : Entretiens avec les concepteurs d'outils d'IA	283
Tableau 9 : Codage de RADO	284
Tableau 10 : Entretiens liés au projet KOVAK	285
Tableau 11 : Codage du matériau du projet KOVAK	286

Annexes

Annexe 1 : Présentation de l'algorithme MYCIN

L'algorithme MYCIN est conçu dans les années 1970, à l'Université de Stanford, par Buchanan et Shortliffe dans le but d'aider au diagnostic des infections liées au sang et de prodiguer des conseils thérapeutiques sur les antibiotiques associés. La brève narration de ce projet illustre les problèmes et caractéristiques saillantes des systèmes experts*.

Avec DENDRAL, Buchanan éprouve l'intérêt de représenter algorithmiquement des savoirs spécifiques. Il s'associe avec Shortliffe, alors étudiant en médecine qui découvre l'IA dans un cours en informatique. Les deux commencent leur collaboration avec un projet pour envoyer des messages d'alertes automatiques aux médecins sur des interactions médicamenteuses dangereuses.

Après six mois de discussion, Buchanan et Shortliffe réduisent la portée du projet à de la surveillance et des messages d'alerte pour les prescriptions d'antibiotiques. Malgré une ambition réduite, le projet demande de connecter trois bases de données distinctes de l'Université de Stanford (la pharmacie, le laboratoire d'analyses et le système de bactériologie) ainsi que l'intégration de nombreuses connaissances médicales pour espérer corriger les décisions thérapeutiques inappropriées et proposer des combinaisons pertinentes d'antimicrobiens.

Pour contourner la première difficulté, le projet passe d'un système de surveillance à une interface dans laquelle le médecin interagit : il rentre les données du patient et reçoit un avis. Ici, contrairement à DENDRAL et à la chimie analytique, la question traitée est beaucoup plus incertaine et complexe. Une fois le résultat de la spectrographie de masse obtenue, les informations sur la molécule recherchée ne changent pas.

Cependant, les informations sur les patients sont limitées et elles arrivent au goutte-à-goutte en fonction des consultations et des diagnostics réalisés. Pour rendre la tâche plus difficile, si la chimie analytique comprend la théorie derrière les signaux produits par la spectrographie de masse, les connaissances cliniques sur les médicaments et leurs interactions sont très empiriques et ne cessent d'évoluer faute d'un modèle théorique fort.

Cette instabilité conduit les chercheurs à séparer les « propriétés médicales », c'est-à-dire un ensemble de « faits », et les opérations associées à celles-ci, c'est-à-dire l'ensemble de règles applicables sur ces connaissances (Buchanan and Shortliffe 1984, 671) : « *la flexibilité d'un algorithme semble accrue en séparant les connaissances médicales des procédures qui manipulent et raisonnent avec ces connaissances. Nous pensons qu'en rendant la*

représentation plus flexible, il serait plus facile de construire des algorithmes plus puissants dans des domaines où les algorithmes croissent par accréation. »⁶⁶.

Le cœur du projet réside alors dans la captation des connaissances métier et leur traduction pour que l'algorithme les comprenne. Les chercheurs se mettent à interviewer des médecins experts et à transformer leurs connaissances en règles, objets et des systèmes ontologiques (façons de représenter le savoir). À cause de sa difficulté, cette étape décisive pour la construction d'un système expert sera connue comme le « goulot d'étranglement de l'acquisition des connaissances ».

Le projet prend une nouvelle envergure, sollicitant de nombreuses personnes pour coder plus de 500 règles, avec des nuances sur la certitude des connaissances. MYCIN, le premier système expert, voit le jour.

Conçu pour être utilisé par des médecins et non des informaticiens, l'interface est soignée. Par exemple, les règles et les connaissances peuvent être modifiées par des utilisateurs sans expertise informatique. De même, lors de l'établissement du diagnostic, l'algorithme recherche dans sa base de données les informations sur le patient et demande des compléments d'information au médecin si elles font défaut. Enfin, l'algorithme est capable de fournir, dans un anglais compréhensible, les étapes du raisonnement avec les connaissances et les règles qui amènent à proposer un traitement.

La figure 24 illustre cette capacité. Pour établir des recommandations thérapeutiques, MYCIN pose une série de questions. Dans l'exemple de la figure, à la question 35, il demande si le patient a été infecté de la méningite à l'hôpital. Le médecin s'interroge sur la raison de cette requête (« WHY »). Le programme renvoie alors les différentes prémisses établies qui mènent à la règle qu'il applique (« RULE545 » rédigée par Yu). Le médecin décide ensuite de demander le raisonnement qui conduit l'algorithme à déduire qu'il s'agit d'une origine bactérienne (« HOW »). MYCIN liste alors les six règles qui lui permettent d'aboutir à sa conclusion.

⁶⁶ *The overriding principle guiding us was the belief that the flexibility of a program was increased by separating medical knowledge from procedures that manipulate and reason with that knowledge. We believed that by making the representation more flexible, it would be easier to build more powerful programs in domains where programs grow by accretion.*

33) Does Pt538 live in a crowded environment?
 ** NO

34) Do you suspect recent alcoholic history in Pt538?
 ** NO

35) Is meningitis a hospital-acquired infection?
 ** WHY

[i.e. WHY is it important to determine whether the infection was acquired while the patient was hospitalized?]

[2.0] This will aid in determining the organisms (other than those seen on cultures or smears) which might be causing the infection.
 It has already been established that

[2.1] the infection which requires therapy is meningitis, and
 [2.2] organisms were not seen on the stain of this pending CSF culture, and
 [2.3] the type of the infection is bacterial, and
 [2.4] neurosurgery is not one of the relevant items from the history of Pt538, and
 [2.5] the age of Pt538 is greater than 10 days

Therefore, if
 [2.6] the infection was acquired while the patient was hospitalized then
 there is evidence that the organisms (other than those seen on cultures or smears) which might be causing the infection is e.coli (.75) staphylococcus-coag-pos (.3) pseudomonas-aeruginosa (.3) klebsiella-pneumoniae (.5)

Author: Yu
 [RULE545]

[back to question 35. . .]

** HOW 2.3

[i.e. HOW was it established that the type of the infection is bacterial?]

The following rules concluded about:
 the likelihood that the type of the infection is bacterial

	Rule	Yes	No
[2.7]	RULE148	(.19)	
[2.8]	RULE500	(.51)	
[2.9]	RULE501	(.95)	
[3.0]	RULE502	(.97)	
[3.1]	RULE526	(.98)	
[3.2]	RULE504	(.97)	

Figure 24 : Extrait d'un transcrit d'une consultation MYCIN pour un patient à Stanford (Pt538) (Buchanan et Shortliffe, 1984, 709)

Les médecins font une évaluation plutôt positive de MYCIN. Ses recommandations thérapeutiques ne font pas l'unanimité, mais elles sont acceptées par une majorité alors que la discorde entre spécialistes est fréquente dans ce domaine. Pourtant, l'algorithme ne sera jamais utilisé dans la pratique clinique pour trois raisons, toujours d'actualité avec les outils d'IA.

Tout d'abord, les explications fournies ne suffisent pas à résoudre les questions éthiques et juridiques liées à l'utilisation d'ordinateurs en médecine, notamment la répartition de la responsabilité en cas de mauvais diagnostic. Ensuite, en termes d'infrastructure informatique, l'installation de l'algorithme s'avère compliquée. Enfin, saisir les informations des patients demande beaucoup de temps et décourage toute utilisation quotidienne. MYCIN n'a pas l'effet escompté sur le domaine médical.

En revanche, son apport à l'IA s'avère décisif. Il démontre la puissance de l'approche envisagée Feigenbaum pour répondre à des questions complexes.

Annexe 2 : Évolution des prismes théoriques pour comprendre les usages de la technologie

Les prismes théoriques pour comprendre les usages de la technologie exposés dans la thèse (Chapitre 2, 1.1) sont le fruit d'une évolution intellectuelle. Nous retraçons ici quelques « stratifications conceptuelles », à l'image des couches géologiques, qui se sont succédé d'une ère à une autre, pour aboutir à ces approches.

Notre travail de « fouille archéologique » débute en suivant les ramifications bibliographiques des articles mobilisés par ces approches jusqu'à leur source, jusqu'aux travaux fondateurs. Notre recherche s'appuie également sur des études qui extraient des « calottes » de ces stratifications théoriques et qui présentent alors certains moments de cette évolution (Verbeek 2006; Zammuto et al. 2007; Leonardi 2013). Enfin, la « pierre de Rosette » de Leonardi (2012) se révèle essentielle pour parfaire notre travail archéologique, décrivant et distinguant le vocabulaire employé dans ces approches.

La restitution de ces recherches fait l'objet des deux sous-parties suivantes. Elles correspondent à deux grands pans intellectuels, l'un propre aux sciences de gestion et l'autre aux *Science and Technology Studies*. Bien que les deux disciplines bénéficient d'enrichissements mutuels, nous les présentons successivement.

De la théorie de la contingence à des approches plus constructivistes en sciences de gestion

Dans les années 1950, l'École de la contingence se développe en sciences de gestion pour expliquer que la structure organisationnelle résulte de la taille, de l'histoire, de la technologie, et de l'environnement (Saussois 2019). Les études se font au niveau des organisations pour observer ce que le changement technologique provoque. La technologie joue un rôle plutôt déterministe dans le processus d'organisation.

En particulier, Woodward (1958) propose une théorie fondamentale à cette vision. Elle enquête sur une centaine d'entreprises industrielles basées dans le Sussex de 1953 à 1957. Son analyse montre, entre autres, que la technologie limite les possibilités de diviser le travail et d'exercer le contrôle, et donc qu'elle influence la structure organisationnelle.

Dans le même courant, Perrow (1967) approfondit l'aspect déterministe de la technologie sur les fonctions organisationnelles à travers une étude sur les hôpitaux. La structure organisationnelle devrait s'adapter aux technologies pour gagner en efficacité.

Dans les années 1970, les études sur la relation entre technologie et organisation perdent de l'attrait, les facteurs de la taille et surtout de l'environnement gagnent en intérêt (Zammuto et al. 2007), comme le montre l'apparition de la théorie institutionnelle (Meyer et Rowan 1977), ou celle de la théorie de l'écologie des populations (*population ecology*) (Hannan et Freeman 1987). En outre, l'influence de la technologie sur l'organisation s'efface devant celle de la hiérarchie : la prémisse dominante est que celle-ci s'impose comme le principal mécanisme de coordination et de contrôle en structurant les relations spécifiques d'autorité et le mouvement de l'information (Zammuto et al. 2007).

Dans les années 1980, Barley (1986) propose une nouvelle conception de la relation entre technologie et organisation en rupture de la vision techno-déterministe. Il puise en sociologie la théorie de la structuration qui propose de dépasser la séparation entre une analyse macro des structures et une analyse micro des agents. L'approche unifiée sert à comprendre les systèmes sociaux comme le fruit des interactions entre ces deux niveaux (entre les agents et les structures) (Giddens 1984).

Barley propose ainsi un modèle dans lequel l'introduction d'une nouvelle technologie est le déclencheur d'un changement structurel⁶⁷, qui se consolide à partir de l'agrégation des actions individuelles. En particulier, dans son étude sur l'adoption d'une technologie de scanners dans deux services de radiologie. Barley montre que l'usage de la technologie modifie la façon dont les personnes agissent et notamment leurs rapports de pouvoir. L'accumulation de ces nouveaux agissements provoque des changements dans la structure organisationnelle (voir la section A de la figure 25).

Il remarque cependant que ces changements sont différents et d'intensité variable d'un service de radiologie à l'autre. Barley conclut que la technologie ne détermine pas tant la structure organisationnelle, qu'elle n'aide des acteurs à provoquer un changement.

Avec la même inspiration conceptuelle et des résultats similaires, Poole et DeSanctis (1994) proposent la théorie de la structuration adaptative⁶⁸ : l'usage de la technologie médiatise l'action des acteurs (en leur proposant de nouvelles possibilités) et en conséquence la structure organisationnelle change (voir la section B de la figure 25).

⁶⁷ *Technology-Triggered Structural Change Model*

⁶⁸ *Adaptive Structuration Theory*

De même, Orlikowski (1992) s'inspire de la théorie de la structuration pour proposer le modèle de dualité technologique⁶⁹. Comparé aux deux théories précédentes, elle propose un changement conceptuel majeur.

L'unité d'analyse est différente : ce n'est plus l'action des acteurs qui change (ou est médiatisée) par l'introduction d'une nouvelle technologie, c'est l'*usage de la technologie* lui-même qui est l'action à considérer. L'usage de la technologie n'est plus un « intermédiaire », mais ce qui se passe au niveau micro et qui s'agrège pour donner la structuration organisationnelle au niveau macro (voir la section C de la figure 25).

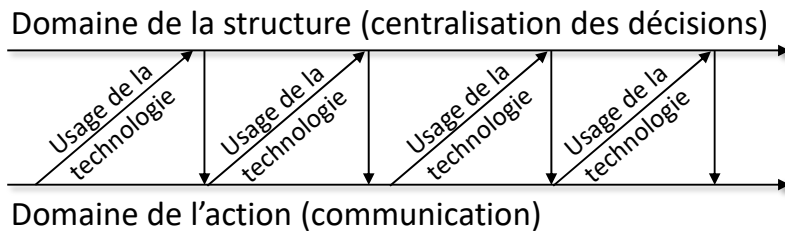
Qui plus est, dans son modèle, l'usage spécifique d'une technologie est d'habitude choisi par les acteurs en position de pouvoir pour renforcer la structure organisationnelle qui les avantage, et ainsi préserver leurs privilèges. Il arrive aussi que d'autres acteurs se rebellent et utilisent autrement la technologie, ce qui provoque un changement organisationnel.

Orlikowski (2000) développe ensuite le modèle « *practice lens* », qui prolonge le précédent (la dualité technologique). L'usage de la technologie au niveau micro reste. En miroir, au niveau macro, Orlikowski propose de regarder la « technologie-en-pratique », au lieu des structures organisationnelles traditionnelles. Ainsi, son modèle fait évoluer le dialogue entre les actions des acteurs et la structure organisationnelle, où l'usage de la technologie se fait l'intermédiaire, à un dialogue entre l'usage de la technologie et la technologie-en-pratique. L'intermédiaire est alors l'interprétation que font les acteurs (voir la section D de la figure 25).

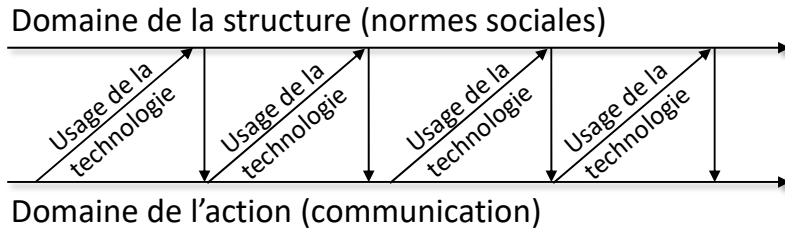
Paradoxalement, alors que la technologie semble au cœur de ce modèle, elle est mise en périphérie. Le point central repose sur les pratiques sociales d'interprétation. Si le modèle est critiqué pour son aspect trop « social », il suscite aussi de nombreux chercheurs à analyser les processus d'interprétation de l'usage d'une technologie et de leur cristallisation dans des formes institutionnalisées (Davidson 2002; Schultze et Orlikowski 2004; Boudreau et Robey 2005; Vaast et Walsham 2005; Baptista 2009).

⁶⁹ *Duality of technology model*

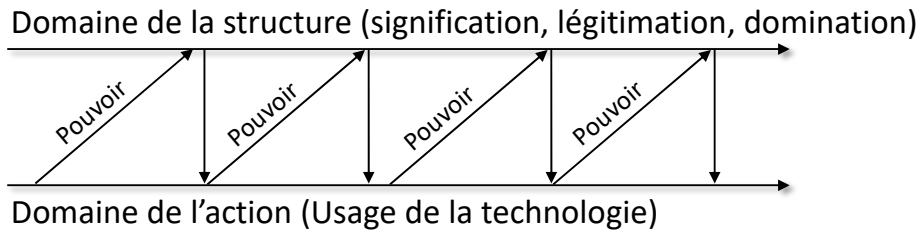
A. Technology-Triggered Structural Change Model (Barley, 1986)



B. Adaptive Structuration Theory (Poole et DeSanctis, 1990)



C. Duality of technology model (Orlikowski, 1992)



D. Practice Lens (Orlikowski, 2000)

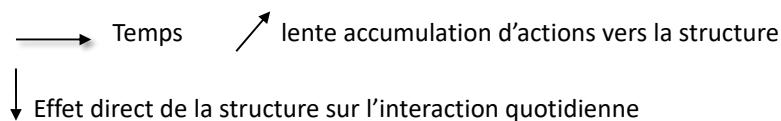
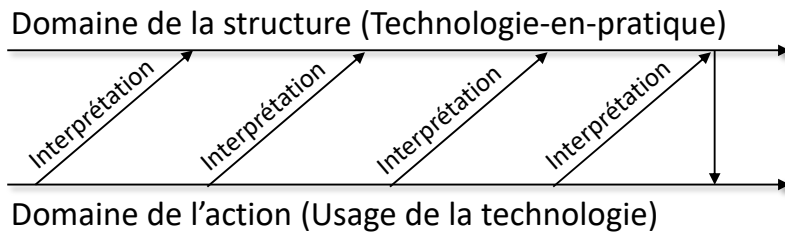


Figure 25 : Schéma des approches structurelles sur l'usage de la technologie dans l'organisation (Leonardi 2013)

L'émergence des études des sciences et des techniques

Les études des sciences et des techniques⁷⁰ sont un mouvement qui émerge dans les années 1960 et qui influence et propose plusieurs concepts d'hybridation technologique. Ce mouvement est d'autant plus intéressant qu'il rompt avec de nombreuses traditions intellectuelles, notamment la sociologie du savoir scientifique qui établit une dichotomie entre nature et culture. Par exemple, l'approche dominante en sociologie des sciences dans les années 1960 et 1970 se concentre sur les scientifiques et leurs institutions (Lemaine et al. 1969), plutôt que sur la manière dont la recherche s'effectue.

L'origine des études des sciences et des techniques peut être située dans les années 1960 avec l'ouvrage de Kuhn (1962), intitulé *La Structure des Révolutions Scientifiques*. Écrit par un physicien de formation converti à l'histoire des sciences, ce livre décrit comment les sciences ont progressé par bonds successifs.

Selon cette perspective, une science commence par avoir plusieurs écoles de pensée en compétition pour expliquer un phénomène. Cependant, l'une d'entre elles finit par l'emporter, notamment en raison de facteurs sociaux et politiques. Cette école adopte alors un paradigme unique, c'est-à-dire une manière particulière de voir le monde qui indique ce qui doit être analysé et comment, et enseigné aux jeunes chercheurs à travers des manuels sans montrer des alternatives. La recherche se concentre alors sur la confirmation de cette théorie en résolvant les énigmes qu'elle propose. Or, avec le temps, l'accumulation des erreurs de mesure, c'est-à-dire des écarts entre la réalité et la théorie, finit par provoquer un moment de rupture, facilité par d'autres facteurs, notamment sociaux et politiques. Un nouveau paradigme est alors proposé et adopté pour résoudre les incohérences. Ainsi, la science avance par « révolutions »⁷¹ et par bonds successifs plutôt que de manière incrémentale et par l'accumulation de savoirs. Les sciences dites « dures » sont donc considérées comme des processus sociaux, enracinées dans leur contexte historique.

L'ouvrage de Kuhn ouvre un nouvel champ de recherche que de nombreuses études anglo-saxonnes approfondissent dans les années 1960 et 1970. Celles-ci sont introduites en France grâce à deux ouvrages dirigés par Callon et Latour (1985; 1982).

Cependant, malgré le changement de perspective qu'elles apportent, Callon et Latour remarquent que ces recherches anglo-saxonnes maintiennent une certaine continuité : les

⁷⁰ *Science and Technology Studies* (STS) en anglais

⁷¹ Voir, par exemple, le passage de la physique newtonienne à celle de la relativité.

historiens des sciences et les sociologues des sciences utilisent sur les mêmes données (articles, archives, transcriptions d'entretiens) et appliquent une vision sociologique à de nombreux domaines scientifiques, sans étudier les spécificités techniques de chaque discipline. En somme, la dichotomie entre nature et culture est préservée, avec des sociologues qui interprètent les sciences naturelles à travers le prisme de la culture

Par conséquent, Callon et Latour s'intéressent aux sciences contemporaines, à la manière dont elles se pratiquent au quotidien, dans leurs controverses, plutôt qu'à une rétrospective historique des faits. Ils adoptent donc une approche qui dépasse le clivage introduit par les modernes entre la nature et la culture.

Latour et Woolgar s'inscrivent dans cette nouvelle perspective avec *La Vie de Laboratoire* (1979). Comme ils le soulignent eux-mêmes : « *Nous nous sommes aperçus très vite que, pour dépasser les limites de la littérature qui nous précédait [évoqué plus haut], il nous fallait non seulement ne pas croire aux scientifiques – ce qui était assez facile –, mais en plus ne pas croire aux sociologues – ce qui était plus ardu.* » (1979, 19-20). Ils s'immergent dans la vie quotidienne d'un laboratoire comme des ethnographes étudiant une tribu exotique. Leur analyse s'appuie sur le principe de symétrie de Bloor (1976), qui soutient que les études sociologiques ou historiques des sciences doivent considérer à la fois le contexte social et le contenu scientifique, quelle que soit la discipline. Dans cette optique, toutes les explications du développement scientifique doivent être symétriques, ce qui signifie que si des explications d'ordre économique sont employées pour expliquer le succès d'une théorie, elles doivent expliquer par la même occasion pourquoi une autre théorie ne marche pas. D'autres recherches parallèles (Lynch 1982) adoptent une approche similaire et contribuent à développer une vision qui étudie le contenu scientifique, en dépassant la séparation entre « dimension cognitive » et « facteurs sociaux ». Les études des sciences et des techniques cherchent à inclure les institutions des sciences étudiées, ainsi que leurs épistémologies, leurs logiques de fonctionnement et leurs méthodes.

À partir du milieu des années 1980, l'objet des études des sciences et des techniques, qui portait initialement sur la construction des faits scientifiques, évolue pour se concentrer sur les artefacts issus des technologies. Ce changement est reflété dans deux ouvrages clés : *The Social Shaping of Technology* (MacKenzie et Wajcman 1999) et *The Social Construction of Technological Systems* (Bijker, Hughes, et Pinch 2012). Ils retracent l'historique d'objets techniques comme la bakélite et l'ampoule, plutôt que des domaines scientifiques.

La vision principale qui émerge de ces travaux est celle d'une « construction sociale de la technologie ». La théorie associée est particulièrement développée par Pinch Hughes et Pinch (2012). Selon cette théorie, le développement des objets technologiques est « ouvert », ce qui signifie que pour une même technologie, différentes formes d'artefacts peuvent être développés et que des facteurs sociaux orientent le produit final. L'interprétation des acteurs

joue un rôle essentiel. À l’instar des domaines scientifiques, les objets techniques résultent de négociations sur l’interprétation des observations. Il est donc nécessaire d’inclure la « flexibilité interprétative » (*interpretative flexibility*) dans la production des objets techniques. Cette interprétation est réalisée par des « groupes sociaux pertinents » (*relevant social group*). Chacun de ces groupes incarne un type d’interprétation : « tous les membres d’un groupe social partagent le même ensemble de significations attachées à un artefact spécifique » (Bijker, Hughes, et Pinch 2012). En d’autres termes, chaque groupe concerné dans le développement d’un artefact propose une définition particulière de la manière dont la technologie fonctionne. Le développement de l’artefact s’arrête lorsque chaque groupe est satisfait (Bijker 1997).

Annexe 3 : Entretiens de RADO

Entretiens avec les radiologues

Identifiant	Type d’établissement	Mode d’entretien	Date de l’entretien	Durée
RAD1	Clinique privée française	Face-à-face	22/06/2022	67 minutes
RAD2	Clinique privée française	Face-à-face	24/07/2022	49 minutes
RAD3	Clinique privée française	Face-à-face	28/10/2022	77 minutes
RAD4	Centre de recherche	Face-à-face	14/11/2022	44 minutes
RAD5	Centre de recherche	Face-à-face	25/11/2022	51 minutes
RAD6	Clinique privée française	Visioconférence	28/03/2023	48 minutes
RAD7	Hôpital publique français	Visioconférence	13/10/2022	46 minutes
RAD8	Hôpital publique français	Visioconférence	06/04/2023	61 minutes
RAD9	Hôpital publique français	Téléphone	29/09/2022	54 minutes
RAD10	Hôpital publique suisse	Téléphone	18/05/2023	62 minutes
RAD11	Grand centre hospitalier américain	Face-à-face	01/08/2023	49 minutes

RAD12	Grand centre hospitalier américain	Visioconférence	02/08/2023	68 minutes
RAD13	Grand centre hospitalier américain	Visioconférence	10/08/2023	53 minutes
RAD14	Clinique américaine privée	Téléphone	28/07/2023	42 minutes

Tableau 7 : Entretiens avec les radiologues

Entretiens avec les concepteurs d'outils d'IA

Identifiant	Rôle	Entreprise	Outil(s) d'IA	Durée de l'entretien
JFR1	Commercial senior	Période 2, Europe, PME	Diagnostic imagerie (1)	11 minutes
JFR2	Technique senior	Période 2, États-Unis, PMI	Diagnostic et traitement du cancer (>5)	13 minutes
JFR3	Groupe dirigeant	Période 2, France, PME	Diagnostic imagerie (3)	7 minutes
JFR4	Commercial junior	Période 2, France, PME	Diagnostic imagerie (3)	11 minutes
JFR5	Technique junior	Période 2, France, PME	Diagnostic imagerie spécialisé (1)	15 minutes
JFR6	Technique senior	Période 2, France, PME	Diagnostic imagerie spécialisé (1)	16 minutes
JFR7	Technique junior	Période 1, France, PME	Reconstitution d'imagerie (4)	9 minutes
JFR8	Commercial junior	Période 1, France, PME	Reconstitution d'image (4)	13 minutes
JFR9	Commercial senior	Période 1, États-Unis, PMI	Diagnostic imagerie (>5)	9 minutes

JFR10	Commercial senior	Période 1, États-Unis, PMI	Diagnostic imagerie (>5)	8 minutes
JFR11	Groupe dirigeant	Période 3, France, PME	Diagnostic imagerie (1)	12 minutes
JFR12	Commercial senior	Période 2, États-Unis, PMI	Infrastructure et aide à la décision (>5)	13 minutes
JFR13	Commercial junior	Période 3, France, PME	Diagnostic imagerie (2)	18 minutes
JFR14	Commercial senior	Période 3, France, PME	Diagnostic imagerie (2)	15 minutes
JFR15	Groupe dirigeant	Période 2, Europe, PME	Infrastructure (1)	19 minutes
JFR16	Groupe dirigeant	Période 3, France, PME	Diagnostic imagerie (1)	39 minutes
JFR17	Groupe dirigeant	Période 3, France, PME	Infrastructure et diagnostic imagerie (1)	7 minutes

Tableau 8 : Entretiens avec les concepteurs d'outils d'IA

Annexe 4 : Codage du matériau de RADO

Codage		
Concept	Définition	Codes illustratifs
Pratiques interrogatives	Des pratiques mises en œuvre par les experts humains pour établir un lien entre leurs propres connaissances et celles de l'outil d'IA (Lebovitz, Lifshitz-Assaf, and Levina 2022, 8).	<ul style="list-style-type: none"> Mobilisation de pratiques médicales Prise de clichés complémentaires Chercher à déterminer l'évolutivité Mobilisation des statistiques liées au fonctionnement de l'outil d'IA Anticipation des résultats de l'outil d'IA Évocation des dysfonctionnements précédents

Outil de conformation	Il ne modifie pas substantiellement les routines existantes (Béjean, Kletz, et Moisdon 2018).	Confirmation du résultat négatif Divergence des conclusions
Outil d'exploration	L'outil devient générateur d'apprentissage collectif en permettant l'exploration de nouvelles possibilités (Béjean, Kletz, and Moisdon 2018).	<i>Non Applicable</i>
Relation radiologue concepteur	Lien établi entre les radiologues et les concepteurs	Motifs de contact et informations échangées Moyens de contact Description de la relation
Relation radiologue patiente	Lien établi entre les radiologues et les patientes	Description des patientes Attitudes face aux patientes

Tableau 9 : Codage de RADO

Annexe 5 : Entretiens liés au projet KOVAK

Identifiant	Rôle	Mode d'entretien	Date de l'entretien	Durée de l'entretien
IMP1	Médecin et data engineer	Visioconférence	06/07/2023	56 minutes
IMP2	Médecin et data scientist	Visioconférence	10/07/2023	67 minutes
IMP3	Médecin et data engineer	Visioconférence	14/07/2023	47 minutes

IMP4	Médecin et data scientist	Présentiel	19/07/2023	71 minutes
IMP5	Médecin et chef de projet	Présentiel	01/08/2023	36 minutes
IMP6	Ingénieur de recherche implémentation	Visioconférence	01/08/2023	56 minutes
IMP7	Data scientist senior	Visioconférence	01/08/2023	39 minutes
IMP8	Médecin et chef de projet	Visioconférence	02/08/2023	63 minutes
IMP9	Ingénieur de recherche implémentation	Présentiel	16/08/2023	59 minutes
IMP10	Médecin et chef de projet	Présentiel	24/08/2023	48 minutes

Tableau 10 : Entretiens liés au projet KOVAK

Annexe 6 : Codage du matériau du projet KOVAK

Codage		
Concept	Définition	Codes illustratifs
Abduction	Suggère une hypothèse (Harris et Hoover 1980, 331).	Conception de la cohorte Explications de causalité Génération d'hypothèses Validation des résultats de l'outil d'IA
Déduction	Tire diverses conséquences (Harris et Hoover 1980, 331).	Prétraitement des données

Induction	Met à l'épreuve les idées (Harris et Hoover 1980, 331).	Estimation de la force d'une variable Vérification des résultats de l'outil d'IA
Pratiques interrogatives	Des pratiques mises en œuvre par les experts humains pour établir un lien entre leurs propres connaissances et celles de l'outil d'IA (Lebovitz, Lifshitz-Assaf, et Levina 2022, 8).	Mobilisation de pratiques médicales Lecture de la base de données Évaluation des performances de l'outil d'IA Mobilisation d'indicateurs statistiques Mobilisation de pratiques algorithmiques
Outil de conformation	Il ne modifie pas substantiellement les routines existantes (Béjean, Kletz, et Moisdon 2018)	Confirmation de facteurs explicatifs connus Estimation de la force des facteurs
Outil d'exploration	L'outil devient générateur d'apprentissage collectif en permettant l'exploration de nouvelles possibilités (Béjean, Kletz, et Moisdon 2018).	Surprises sur les résultats Nouvelle lecture sur la situation

Tableau 11 : Codage du matériau du projet KOVAK

Annexe 7 : Modèle CRISP-DM

Nous expliquons brièvement les étapes du modèle CRISP-DM (Fayyad 1996; Fayyad, Piatetsky-Shapiro, et Smyth 1996b; 1996a; Azevedo et Santos 2008; Shafique et Qaiser 2014; Cazacu et Titan 2021) car elles rendent compte du processus classique de développement des outils d'IA. En pratique, et comme montre notre narration du projet KOVAK (Chapitre 5, 1.1) montre, ces étapes sont « itératives et interactives » (Fayyad, Piatetsky-Shapiro, and Smyth 1996b), voir superposées et entremêlées (Chapitre 3, 2.2.3).

1. Compréhension de la situation : Cette étape implique une compréhension claire du problème à résoudre. Dans KOVAK, cela concerne la reconnaissance des facteurs associés aux effets indésirables* de l'immunothérapie*.

2. **Compréhension des données** : Cela implique d'analyser la base de données et les variables*. Dans KOVAK, cela se traduit par la sélection des variables pertinentes pour la question de recherche et l'observation de leur structure.

3. **Préparation des données** : Cette étape consiste à nettoyer et à organiser les données pour l'analyse. KOVAK montre comment les données ont été préparées pour la modélisation.

4. **Modélisation** : Il s'agit d'appliquer des techniques statistiques ou d'apprentissage machine pour découvrir des motifs (*patterns*) dans les données. Dans KOVAK, différents modèles sont testés pour prédire les effets indésirables.

5. **Évaluation** : Cette étape consiste à évaluer la performance des modèles. Dans le cas de KOVAK, l'évaluation se concentre sur l'efficacité des modèles à prédire correctement les cas d'effets indésirables ainsi que de l'interprétation des variables et de leur adéquation avec les connaissances médicales.

6. **Déploiement** : La dernière étape concerne l'utilisation pratique des résultats de la modélisation. Pour KOVAK, cela implique la rédaction d'un article scientifique et la possible incorporation des résultats dans la pratique médicale.

Annexe 8 : La sémiotique triadique et l'interaction avec le monde

La sémiotique, l'étude des signes et des symboles, explore la création et la communication du sens à travers divers systèmes de signes, qu'ils soient verbaux, visuels ou gestuels. Cette discipline examine la négociation et la compréhension des significations dans divers contextes culturels et sociaux, s'attachant à trois aspects clés : le signifiant (forme du signe), le signifié (ce que le signe représente) et l'interprète (celui qui décode le signe). Influencée par Ferdinand de Saussure et Charles Sanders Peirce, elle trouve son application dans des domaines tels que la linguistique, la philosophie, l'anthropologie, et les études culturelles et médiatiques.

Peirce a révolutionné la sémiotique avec son concept de « sémiotique triadique ». D'autres pragmatistes, comme Mead, James et Dewey, ont également contribué à la théorie sémiotique, chacun avec sa perspective unique⁷² (Lorino 2018).

La sémiotique triadique de Peirce comprend trois éléments interactifs : le representamen (signe), l'objet (référence du signe) et l'interprétant (interprétation du signe). Selon ses propos : « *le REPRESENTAMEN est le sujet d'une relation triadique avec un second appelé son*

⁷² Voir « les symboles signifiants » chez Mead, la « médiation symbolique » chez James, et les « signes et symboles » chez Dewey.

OBJET, POUR un troisième appelé son INTERPRÉTANT, cette relation triadique étant telle que le REPRESENTAMEN détermine son interprétant à entretenir la même relation triadique avec le même objet pour quelque interprétant » (Peirce 1931, 1.541). Ainsi, ces éléments interagissent pour créer du sens. Le representamen représente quelque chose pour quelqu'un (ou quelque chose), « d'un certain point de vue » (Peirce 1931, 2.228).

Pour illustrer son concept, nous reprenons l'exemple du baigneur et de la vague, utilisé par Gilles Deleuze pour expliquer les types de connaissances chez Baruch Spinoza⁷³. Selon la sémiotique triadique, la vague elle-même est le representamen, sa force physique est l'objet, et la réaction du baigneur en tant qu'interprétant façonne la compréhension de cette interaction. Ce processus est une « sémiose ».

La sémiose représente le processus par lequel un signe produit du sens. Sans sémiose, nous restons dans une relation de « secondité », telle qu'un baigneur frappé par une vague sans interprétation. En revanche, la sémiose transforme l'expérience brute en sujet de réflexion et de discussion, avec la pensée agissant comme médiatrice. En effet, « *toute pensée est faite de signes* » (Peirce 1998, 213). La pensée est une construction continue de la réalité, de l'expérience, à travers des interprétations logiques.

Avec la médiation sémiotique, la situation devient compréhensible, ce qui entraîne l'action. La médiation sémiotique est performative. Les signes transforment la situation, influençant les possibilités d'action. La médiation sémiotique propose une « *construction active du futur par un processus d'action situé [...] [L'approche pragmatiste] suggère des processus narratifs qui visent à transformer une situation en une autre à travers un parcours concrètement situé* » (Lorino 2018). Ainsi, le baigneur n'est plus frappé par la vague, il nage avec elle. Lorsque le baigneur interprète ce qui se passe, il y a « tiercéité ». Cette médiation s'inscrit dans l'organisation du monde. La plage devient un espace de loisirs plutôt qu'un lieu de danger.

Les rôles de representamen, objet, et interprétant sont dynamiques et contextuels. Ce qui est un interprétant dans une situation peut devenir un representamen dans une autre. Un baigneur agissant comme interprétant dans une situation peut devenir un objet à interpréter pour un sauveteur dans une autre situation. Par conséquent, l'individu, le « il » du baigneur n'est pas un être subjectif, mais un espace dans lequel les signes sont interprétés, un espace ouvert et situé.

Au contraire, la médiation sémiotique porte l'attention sur le signe, qui acquiert une double dimension : il participe à une situation spécifique et singulière à un moment précis, tout en

⁷³ Gilles Deleuze, « cours sur Spinoza », Vincennes, 17/03/81

renvoyant à des catégories de significations construites socialement, à des abstractions de la situation, à des expériences passées et à des anticipations d'avenir, à des événements et des actions se déroulant ailleurs (Lorino 2018). Le signe « *transmet une idée sur une chose* » (Peirce 1998, 4).

Ainsi, la sémiotique triadique dépasse la perspective subjective, embrassant un collectif élargi de signes partagés et d'expériences. Elle reconnaît que les signes existent au sein d'un réseau de significations partagées et d'expériences collectives. Elle se distingue d'autres approches pragmatistes par son antisubjectivisme. Par exemple, Peirce critique la déviation psychologique de James comme « non fondée et peu fiable » (Peirce 1998, 412). En englobant un collectif élargi, composé de la communauté d'êtres vivants et des éléments matériels qui participent à la production de signes, la sémiotique triadique est un précurseur de la « sociomatérialité ».

La compréhension de la sémiotique triadique est essentielle pour aborder l'épistémologie de Peirce, notamment sa théorie des inférences. Cette théorie s'appuie sur la manière dont les signes façonnent nos processus de raisonnement, y compris les démarches déductives, inductives et abductives. Cependant, la sémiotique triadique se poursuit avec la distinction de chaque fonction (le représentamen, l'objet et l'interprétant) en trois catégories par fonction. L'ensemble forme dix classes de sémiose⁷⁴. L'analyse consiste alors à identifier les signes employés et comment ils s'agencent les uns avec les autres dans le processus de la pensée. Un projet qui dépasse notre ambition dans la compréhension de l'épistémologie peircienne.

Annexe 9 : Éléments biographiques de Charles S. Peirce

Peirce lui-même, de manière quelque peu provocante, déclare : « *s'il y a un lecteur qui ne peut pas comprendre mes écrits, il me faut lui dire qu'aucun effort mental ne l'aidera : toutes ses difficultés viennent du fait qu'il n'a pas d'expérience du monde de problèmes dont je parle* » (Peirce 1931, 3.419). Bien que ces propos puissent être exagérés, ils découlent du développement de la pensée de Peirce, qui se considère avant tout comme un logicien. Ainsi, son œuvre théorique peut être interprétée comme une succession de propositions logiques dont nous ne devrions pas avoir de difficultés à suivre. Or, en accord avec le pragmatisme, dont il est l'un des fondateurs, ce processus logique est intimement lié à l'expérience. Dans cette optique, nous présentons quelques éléments biographiques visant à illustrer sa position

⁷⁴ La combinaison de ces catégories ne mène pas à 27 (3x3x3) classes distinctes de sémiose, mais à dix, en raison de relations spécifiques qui ne produisent pas toutes une combinaison unique.

intellectuelle. Cette présentation est loin d'être exhaustive, certains aspects de sa vie ayant été délibérément omis en raison de leur distance par rapport au sujet traité.

Charles Sanders Peirce est né le 10 septembre 1839 au Massachusetts (États-Unis), de Sarah Hunt Mills et Benjamin Peirce, professeur de mathématiques (et plus tard de physique et d'astronomie) à l'Université d'Harvard. Son père entreprend de lui enseigner lui-même sa discipline, qu'il considère comme la plus importante : les mathématiques étaient pour lui « la science des conclusions nécessaires », comme il l'écrit en préambule de son livre *L'Algèbre logique associative*. Par conséquent, Peirce hérite de cette vision et se consacre à développer le domaine de la logique. Dès l'âge de 16 ans, il commence à lire *Critique de la Raison pure* de Kant, qu'il mémorise intégralement trois ans plus tard. Il étudie également d'autres grands auteurs pour parfaire sa formation en logique. En effet, bien que la théorie de l'inférence soit souvent attribuée à Peirce, ses fondements remontent à la Grèce antique (Deledalle 1987). En 1859, à l'âge de 20 ans, il obtient une licence en mathématiques, puis une maîtrise en 1862 à l'Université de Harvard.

Parallèlement à cet apprentissage, il s'intéresse à la chimie dès l'âge de huit ans. À l'âge de douze ans, Peirce rédige un ouvrage intitulé *Histoire de la chimie*. En plus de souligner son appétence pour la chimie, cet écrit de jeunesse montre déjà comment la discipline historique fait partie intégrante de son ontologie processuelle (Viola 2020, 193) : il développe une conception de l'histoire utilisant les mêmes principes inférentiels et les mêmes présuppositions que d'autres sciences empiriques. Peirce continue à s'intéresser à la chimie et met au point un laboratoire à l'âge de douze ans pour mener des expériences avancées. Il n'est donc pas surprenant qu'après avoir obtenu sa licence en mathématiques, il décroche la première licence en chimie *summa cum laude* de l'Université de Harvard. Sa connaissance approfondie de cette discipline est devenue le socle de sa réflexion, la chimie étant littéralement et métaphoriquement la matière première sur laquelle sa pensée s'est cristallisée. Par exemple, comme nous le verrons plus loin, lorsqu'il réfléchissait à la manière dont la pensée pouvait s'étendre au-delà des limites du cerveau humain pour se répartir parmi les objets, il faisait référence à cette discipline : « Un appareil destiné à réaliser une expérience physique ou *chimique* est aussi une machine à raisonner, avec cette différence qu'il ne dépend pas des lois de l'esprit humain, mais de la raison objective incarnée dans les lois de la nature. Ce n'est donc pas une figure de style que de dire que les *alambics* et les *cucurbites du chimiste* sont des instruments de pensée ou des machines logiques. »⁷⁵ (Peirce, 1887).

⁷⁵ *A piece of apparatus for performing a physical or chemical experiment is also a reasoning machine, with this difference, that it does not depend on the laws of the human mind, but on the objective reason embodied in*

En outre, la chimie lui permet de comprendre l'importance de l'expérimentation dans le raisonnement logique. Il comprend également la nécessité de développer ses sens pour mieux appréhender ces expériences. En tant que jeune adulte, il décide d'affiner son sens du goût et développe son palais avec l'aide d'un sommelier français. Cette formation le dote d'une expertise considérable en vins de Bordeaux, bien qu'elle lui vaille la réputation de grand buveur, ce qui est une tare dans la société puritaine de son époque.

Pendant son séjour à Harvard, il développe une amitié avec James. Avec lui et une douzaine d'anciens étudiants de cette institution, ils organisent de janvier à décembre 1872 ce que Peirce appelle le « Club de Métaphysique » (*The Metaphysical Club*). À cette époque, les États-Unis sortaient de la guerre civile et le pays était en pleine construction, tant sur le plan intellectuel que philosophique. Les membres du groupe souhaitaient contribuer à cette évolution en apportant leur contribution sur le plan philosophique. Leurs discussions, bien que restant informelles, critiquaient la pensée européenne dominante (notamment les philosophes modernes et Descartes) en s'appuyant sur les avancées scientifiques, telles que les théories de l'évolution de Darwin présentées dans *On the Origin of Species* en 1859, et sur les avancées philosophiques, dont l'empirisme britannique et les philosophes allemands contemporains (Hegel et Kant). Ces discussions se sont interrompues lorsque Peirce a dû partir pour une mission liée à son travail au sein du Service Géodésique des États-Unis (*United States Coast and Geodetic Survey*). Cependant, les critiques de la pensée européenne dominante, la mobilisation des découvertes scientifiques, ainsi que le projet social et politique de renforcer une démocratie américaine, affaiblie moralement par la guerre civile, ont constitué les ingrédients de la philosophie américaine à venir : le pragmatisme.

Du point de vue intellectuel, Peirce entretient un lien constant avec la religion. Cette relation laisse des traces dans ses écrits et s'intègre dans son œuvre. Cependant, le détail de cette relation dépasse le cadre de cette thèse et ne contribue pas à son sujet. Pour mieux comprendre sa vision du monde, notons simplement que la pensée de Peirce est imprégnée d'évolutionnisme, une notion visible dans ses écrits à partir de 1878. Son évolutionnisme est inspiré du naturaliste français Lamarck, qui conçoit l'amour cosmique du créateur comme la principale force du changement, plutôt que la sélection naturelle de Darwin ou la théorie mécaniste de Spencer. Cette approche se reflète notamment dans la croyance téléologique de Peirce, qui considère qu'il existe un objectif à atteindre, une destination finale pour la logique.

the laws of nature. Accordingly, it is no figure of speech to say that the alembics and cucurbits of the chemist are instruments of thought, or logical machines.

La carrière professionnelle de Peirce a été moins prestigieuse que ce que pourrait laisser penser la puissance de sa pensée. Il a travaillé pendant plus de trente ans au Service Géodésique, au cours desquels il a réalisé plusieurs études scientifiques. Parallèlement, il a enseigné la logique pendant cinq ans à l'Université Johns Hopkins, notamment à Allan Marquand, qui a contribué au développement des précurseurs de l'ordinateur, comme mentionné plus loin. Son style de vie, y compris sa consommation d'alcool, ainsi que sa pédagogie n'ont pas été bien accueillis par l'établissement, ce qui a entraîné la fin de son contrat. Vers la fin de sa vie, il a accompli divers travaux pour subvenir à ses besoins, avec l'aide financière de ses amis, malgré sa situation précaire.

Il est décédé dans la pauvreté le 9 avril 1914, à l'âge de 74 ans. Sa veuve a vendu ses nombreux écrits à l'Université d'Harvard. En 1964, l'institution a transformé ces écrits en microfilms. En 1967, Robin en a fait un catalogue annoté, qui indiquait environ plus de 100 000 pages pour 1 650 manuscrits inédits. Plus d'un siècle après sa mort, les écrits de Peirce continuent d'être découverts, et de nombreuses entreprises éditoriales s'efforcent de comprendre la portée de l'œuvre de cet esprit iconoclaste. L'actualité de ses idées demeure toujours aussi forte, et le monde académique revient sans cesse aux textes originaux pour en saisir la profondeur.

Bibliographie

Aaltonen, Aleksi, Cristina Alaimo, et Jannis Kallinikos. 2021. « The Making of Data Commodities: Data Analytics as an Embedded Process ». *Journal of Management Information Systems* 38 (2): 401-29.

Abbott, Andrew Delano. 1991. « The future of professions: Occupation and expertise in the age of organization ». *Research in the Sociology of Organizations* 8 (1): 17-42.

———. 2004. « Methods of discovery: Heuristics for the social sciences ».

Abell, Peter. 2004. « Narrative explanation: an alternative to variable-centered explanation? » *Annu. Rev. Sociol.* 30:287-310.

Acemoglu, Daron, et Pascual Restrepo. 2018. « The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment ». *American Economic Review* 108 (6): 1488-1542.

Acs, B., M. Rantalainen, et J. Hartman. 2020. « Artificial Intelligence as the next Step towards Precision Pathology ». *Journal of Internal Medicine* 288 (1): 62-81.

Aggeri, Franck, et Julie Labatut. 2014. « Chapitre 1 / Les métamorphoses de l'instrumentation gestionnaire: Une généalogie des approches par les instruments ». In *L'instrumentation de l'action publique*, 63-94. Presses de Sciences Po.

Ajunwa, Ifeoma, et Daniel Greene. 2019. « Chapter 3 Platforms at Work: Automated Hiring Platforms and Other New Intermediaries in the Organization of Work ». In *Research in the Sociology of Work*, édité par Steve P. Vallas et Anne Kovalainen, 33:61-91. Emerald Publishing Limited.

Akrich, Madeleine. 1992. « The de-scription of technical objects ». In *Shaping technology/building society*, édité par Wiebe E. Bijker et John Law, MIT Press, 205-24. *Studies in Sociotechnical Change*. MIT press.

Akrich, Madeleine, et Bruno Latour. 1992. « A summary of a convenient vocabulary for the semiotics of human and nonhuman assemblies ». In *Shaping technology/building society studies in sociotechnical change*, 259-64. The MIT Press.

Alcaras, Gabriel, et L Larribeau. 2022. « Codes. L'informatique comme elle s'écrit ». *RESET*, no 11 (mars).

Allen, Ryan, et Prithwiraj (Raj) Choudhury. 2022. « Algorithm-Augmented Work and Domain Experience: The Countervailing Forces of Ability and Aversion ». *Organization Science* 33 (1): 149-69.

Almalki, Yassir Edrees, Abdul Qayyum, Muhammad Irfan, Noman Haider, Adam Glowacz, Fahad Mohammed Alshehri, Sharifa K. Alduraibi, et al. 2021. « A Novel Method for COVID-19 Diagnosis Using Artificial Intelligence in Chest X-Ray Images ». *Healthcare* 9 (5): 522.

Almaslukh, Bandar. 2021. « A Lightweight Deep Learning-Based Pneumonia Detection Approach for Energy-Efficient Medical Systems ». Édité par Arun K. Sangaiah. *Wireless Communications and Mobile Computing* 2021 (avril):1-14.

Alvesson, Mats, et Dan Kärreman. 2007. « Constructing mystery: Empirical matters in theory development ». *Academy of Management Review* 32 (4): 1265-81.

Alvesson, Mats, et Kaj Sköldbberg. 2018. *Reflexive Methodology: New Vistas for Qualitative Research*. Third edition. London Los Angeles New Delhi Singapore Washington DC Melbourne: SAGE.

Ananny, Mike. 2016. « Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness ». *Science, Technology, & Human Values* 41 (1): 93-117.

Ananny, Mike, et Kate Crawford. 2018. « Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability ». *new media & society* 20 (3): 973-89.

Angwin, Julia, Jeff Larson, Surya Mattu, et Lauren Kirchner. 2022. « Machine Bias* ». In *Ethics of Data and Analytics*, par Kirsten Martin, 1re éd., 254-64. Boca Raton: Auerbach Publications.

Ansell, David A., et Edwin K. McDonald. 2015. « Bias, Black Lives, and Academic Medicine ». *New England Journal of Medicine* 372 (12): 1087-89.

Anthony, Callen. 2018. « To Question or Accept? How Status Differences Influence Responses to New Epistemic Technologies in Knowledge Work ». *Academy of Management Review* 43 (4): 661-79.

———. 2021. « When Knowledge Work and Analytical Technologies Collide: The Practices and Consequences of Black Boxing Algorithmic Technologies ». *Administrative Science Quarterly* 66 (4): 1173-1212.

Anthony, Callen, Beth A. Bechky, et Anne-Laure Fayard. 2023. « “Collaborating” with AI: Taking a System View to Explore the Future of Work ». *Organization Science*, Janvier.

Araujo, Theo. 2018. « Living up to the Chatbot Hype: The Influence of Anthropomorphic Design Cues and Communicative Agency Framing on Conversational Agent and Company Perceptions ». *Computers in Human Behavior* 85 (août):183-89.

Aritzeta, Aitor, Barbara Senior, et Stephen Swailes. 2005. « Team Role Preference and Cognitive Styles: A Convergent Validity Study ». *Small Group Research* 36 (4): 404-36.

Arnoldi, Jakob. 2016. « Computer Algorithms, Market Manipulation and the Institutionalization of High Frequency Trading ». *Theory, Culture & Society* 33 (1): 29-52.

Ashok, Mona, Rohit Madan, Anton Joha, et Uthayasankar Sivarajah. 2022. « Ethical Framework for Artificial Intelligence and Digital Technologies ». *International Journal of Information Management* 62 (février).

Auerswald, Philip E. 2017. *The code economy: a forty-thousand-year history*. New York, NY: Oxford University Press.

Autor, David H, Lawrence F Katz, et Melissa S Kearney. 2006. « The Polarization of the U.S. Labor Market ». *American Economic Review* 96 (2): 189-94.

Ayache, Magali, et Hervé Dumez. 2011a. « Le codage dans la recherche qualitative une nouvelle perspective? » *Le libellio d’Aegis* 7 (2-Eté): 33-46.

———. 2011b. « Réflexions en forme de réponses: à propos du codage dans la recherche qualitative ». *Le Libellio d’Aegis* 7 (3-Automne): 29-34.

Aytes, Ayhan. 2012. « Return of the Crowds: Mechanical Turk and Neoliberal States of Exception ». In *Digital Labor*. Routledge.

Babic, Boris, Sara Gerke, Theodoros Evgeniou, et I. Glenn Cohen. 2021. « Beware Explanations from AI in Health Care ». *Science* 373 (6552): 284-86.

Babushkina, Dina, et Athanasios Votsis. 2022. « Epistemo-Ethical Constraints on AI-Human Decision Making for Diagnostic Purposes ». *Ethics and Information Technology* 24 (2): 22.

Badré, Adrien, Li Zhang, Wellington Muchero, Justin C. Reynolds, et Chongle Pan. 2021. « Deep Neural Network Improves the Estimation of Polygenic Risk Scores for Breast Cancer ». *Journal of Human Genetics* 66 (4): 359-69.

Baer, William C. 1986. « Expertise and Professional Standards ». *Work and Occupations* 13 (4): 532-52.

Bailey, Diane E., et Stephen R. Barley. 2011. « Teaching-Learning Ecologies: Mapping the Environment to Structure Through Action ». *Organization Science* 22 (1): 262-85.

Bailey, Diane E., Samer Faraj, Pamela J. Hinds, Paul M. Leonardi, et Georg von Krogh. 2022. « We Are All Theorists of Technology Now: A Relational Perspective on Emerging Technology and Organizing ». *Organization Science* 33 (1): 1-18.

Bailey, Diane E., Paul M. Leonardi, et Stephen R. Barley. 2012. « The Lure of the Virtual ». *Organization Science* 23 (5): 1485-1504.

Balasubramanian, Natarajan, Yang Ye, et Mingtao Xu. 2022. « Substituting Human Decision-Making with Machine Learning: Implications for Organizational Learning ». *Academy of Management Review* 47 (3): 448-65.

Bansal, Gagan, Besmira Nushi, Ece Kamar, Walter S. Lasecki, Daniel S. Weld, et Eric Horvitz. 2019. « Beyond accuracy: The role of mental models in human-AI team performance ». In *Proceedings of the AAAI conference on human computation and crowdsourcing*, 7:2-11.

Bansal, Veena, et Shubham Shukla. 2021. « Exploring Big Data Analytics Adoption using Affordance Theory ». In *Proceedings of the 23rd International Conference on Enterprise Information Systems*, 131-38.

Baptista, João. 2009. « Institutionalisation as a Process of Interplay between Technology and Its Organisational Context of Use ». *Journal of Information Technology* 24 (4): 305-19.

Barad, Karen. 2003. « Posthumanist Performativity: Toward an Understanding of How Matter Comes to Matter ». *Signs: Journal of Women in Culture and Society* 28 (3): 801-31.

———. 2007. *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning*. Duke University Press.

Barile, Sergio, Paolo Piciocchi, Clara Bassano, Jim Spohrer, et Maria Cristina Pietronudo. 2019. « Re-defining the Role of Artificial Intelligence (AI) in Wiser Service Systems ». In *Advances in Artificial Intelligence, Software and Systems Engineering*, édité par Tareq Z. Ahram, 787:159-70. *Advances in Intelligent Systems and Computing*. Cham: Springer International Publishing.

Barley, Stephen R. 1986. « Technology as an Occasion for Structuring: Evidence from Observations of CT Scanners and the Social Order of Radiology Departments ». *Administrative Science Quarterly* 31 (1): 78-108.

Barley, Stephen R., et Pamela S. Tolbert. 1991. « Introduction: At the intersection of organizations and occupations ».

Barley, William C., Jeffrey W. Treem, et Timothy Kuhn. 2018. « Valuing Multiple Trajectories of Knowledge: A Critical Review and Agenda for Knowledge Management Research ». *Academy of Management Annals* 12 (1): 278-317.

Barocas, Solon, et Andrew D. Selbst. 2016. « Big Data's Disparate Impact ». *California Law Review* 104:671.

Barrett, Michael, Eivor Oborn, et Wanda Orlikowski. 2016. « Creating Value in Online Communities: The Sociomaterial Configuring of Strategy, Platform, and Stakeholder Engagement ». *Information Systems Research* 27 (4): 704-23.

Barrett, Michael, Eivor Oborn, Wanda J. Orlikowski, et JoAnne Yates. 2012. « Reconfiguring Boundary Relations: Robotic Innovations in Pharmacy Work ». *Organization Science* 23 (5): 1448-66.

Beane, Matt, et Wanda J. Orlikowski. 2015. « What Difference Does a Robot Make? The Material Enactment of Distributed Coordination ». *Organization Science* 26 (6): 1553-73.

Beane, Matthew. 2019. « Shadow Learning: Building Robotic Surgical Skill When Approved Means Fail ». *Administrative Science Quarterly* 64 (1): 87-123.

Bechky, Beth A. 2003. « Sharing Meaning Across Occupational Communities: The Transformation of Understanding on a Production Floor ». *Organization Science* 14 (3): 312-30.

———. 2020. « Evaluative Spillovers from Technological Change: The Effects of “DNA Envy” on Occupational Practices in Forensic Science ». *Administrative Science Quarterly* 65 (3): 606-43.

Bechky, Beth A., et Daisy E. Chung. 2018. « Latitude or Latent Control? How Occupational Embeddedness and Control Shape Emergent Coordination ». *Administrative Science Quarterly* 63 (3): 607-36.

Bechmann, Anja, et Geoffrey C Bowker. 2019. « Unsupervised by Any Other Name: Hidden Layers of Knowledge Production in Artificial Intelligence on Social Media ». *Big Data & Society* 6 (1).

Becker, A. 2019. « Artificial intelligence in medicine: What is it doing for us today? » *Health Policy and Technology* 8 (2): 198-205.

Becker, Howard Saul, et Charles C. Ragin. 1992. *What is a case?: exploring the foundations of social inquiry*. Cambridge University Press Cambridge.

Béjean, Mathias, Frédéric Kletz, et Jean-Claude Moisdon. 2018. « Création de valeur organisationnelle et technologies de l'information à l'hôpital : le cas du dossier patient informatisé ». *Gestion et management public* Volume 6 / n° 4 (2): 9-24.

Béjean, Mathias, Frédéric Kletz, Jean-Claude Moisdon, et Claude Sicotte. 2015. « Informatisation incrémentale ou de rupture ? Le cas du dossier patient hospitalier »: *Journal de gestion et d'économie médicales* Vol. 33 (7): 445-67.

Belbin, R Meredith. 2010. *Management Teams*. 0 éd. Routledge.

Benjamin, Ruha. 2019. « Assessing Risk, Automating Racism ». *Science* 366 (6464): 421-22.

———. 2020. *Race after technology: abolitionist tools for the New Jim Code*. Cambridge, UK ; Medford, MA: Polity.

Benlian, Alexander, Martin Wiener, W. Alec Cram, Hanna Krasnova, Alexander Maedche, Mareike Möhlmann, Jan Recker, et Ulrich Remus. 2022. « Algorithmic Management: Bright and Dark Sides, Practical Implications, and Research Opportunities ». *Business & Information Systems Engineering* 64 (6): 825-39.

Berry, Michel. 1983. « Une technologie invisible-L'impact des instruments de gestion sur l'évolution des systèmes humains ».

Berthelot, Vincent. 2021. « (En)quête de communauté : penser et construire une communauté vivante avec John Dewey ». Phdthesis, Université Paris sciences et lettres.

Beyer, David. 2016. *Future of Machine Intelligence*. Place of publication not identified: O'Reilly Media, Inc.

Bez, Sea Matilda, Irène Georgescu, et Mohammad Saleh Farazi. 2022. « Democratizing Healthcare: Opportunities for Value Creation through Patient Feedback Platforms ».

Bijker, Wiebe E. 1997. *Of Bicycles, Bakelites, and Bulbs: Toward a Theory of Sociotechnical Change*. MIT Press.

Bijker, Wiebe E., Thomas P. Hughes, et Trevor Pinch, éd. 2012. *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*. The MIT Press.

Bloor, David. 1976. *Knowledge and Social Imagery*. University of Chicago Press.

Blumer, Herbert. 1986. *Symbolic Interactionism: Perspective and Method*. University of California Press.

Bogert, Eric, Aaron Schecter, et Richard T. Watson. 2021. « Humans Rely More on Algorithms than Social Influence as a Task Becomes More Difficult ». *Scientific Reports* 11 (1).

Bollier, David. 2017. « Artificial intelligence comes of age: The promise and challenge of integrating AI into cars, healthcare, and journalism ». Washington, DC: The Aspen Institute.

Bolter, Jay David. 1991. *Writing space: The computer, hypertext, and the history of writing*. Writing space: The computer, hypertext, and the history of writing. Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Boon, Mieke. 2020. « How Scientists Are Brought Back into Science—The Error of Empiricism ». In *A Critical Reflection on Automated Science*, édité par Marta Bertolaso et Fabio Sterpetti, 1:43-65. *Human Perspectives in Health Sciences and Technology*. Cham: Springer International Publishing.

Borch, Christian, et Bo Hee Min. 2022. « Toward a Sociology of Machine Learning Explainability: Human–Machine Interaction in Deep Neural Network-Based Automated Trading ». *Big Data & Society* 9 (2).

Bostrom, Nick. 2014. *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.

Boudreau, Marie-Claude, et Daniel Robey. 2005. « Enacting Integrated Information Technology: A Human Agency Perspective ». *Organization Science* 16 (1): 3-18.

Boyacı, Tamer, Caner Canyakmaz, et Francis De Véricourt. 2023. « Human and Machine: The Impact of Machine Input on Decision Making Under Cognitive Limitations ». *Management Science*, mars.

Boyd, Danah, et Kate Crawford. 2012. « Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon ». *Information, Communication & Society* 15 (5): 662-79.

Brachten, Florian, Felix Brünker, Nicholas R. J. Frick, Björn Ross, et Stefan Stieglitz. 2020. « On the Ability of Virtual Agents to Decrease Cognitive Load: An Experimental Study ». *Information Systems and E-Business Management* 18 (2): 187-207.

Braga, Adriana, et Robert Logan. 2017. « The Emperor of Strong AI Has No Clothes: Limits to Artificial Intelligence ». *Information* 8 (4): 156.

Breemen, Auke JJ van, Jozsef I. Farkas, et Janos J. Sarbo. 2011. « Knowledge representation as a tool for intelligence augmentation ». In *Computational Modeling and Simulation of Intellect: Current State and Future Perspectives*, 321-41. IGI Global. <https://www.igi-global.com/chapter/computational-modeling-simulation-intellect/53311>.

Briscoe, Forrest, et Chad Murphy. 2012. « Sleight of Hand? Practice Opacity, Third-Party Responses, and the Interorganizational Diffusion of Controversial Practices ». *Administrative Science Quarterly* 57 (4): 553-84.

Brynjolfsson, Erik. 2022. « The Turing Trap: The Promise & Peril of Human-Like Artificial Intelligence ». *Daedalus* 151 (2): 272-87.

Brynjolfsson, Erik, et Andrew McAfee. 2014. *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton & Company.

Buchanan, Bruce G., Edward A. Feigenbaum, et Joshua Lederberg. 1971. A heuristic programming study of theory formation in science. 145. Citeseer.

Bucher, Taina. 2012. « Want to Be on the Top? Algorithmic Power and the Threat of Invisibility on Facebook ». *New Media & Society* 14 (7): 1164-80.

Bunderson, J. S., et K. M. Sutcliffe. 2002. « Comparing Alternative Conceptualizations of Functional Diversity in Management Teams: Process and Performance Effects. » *Academy of Management Journal* 45 (5): 875-93.

Buolamwini, Joy, et Timnit Gebru. 2018. « Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification ». In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, 77-91. PMLR.

Burrell, Jenna. 2016. « How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms ». *Big Data & Society* 3 (1).

Burt, Andrew. 2019. « The AI Transparency Paradox ». *Harvard Business Review*, 13 décembre 2019.

Burton-Jones, Andrew. 2014. « What Have We Learned from the Smart Machine? » *Information and Organization* 24 (2): 71-105.

Burton-Jones, Andrew, et Olga Volkoff. 2017. « How Can We Develop Contextualized Theories of Effective Use? A Demonstration in the Context of Community-Care Electronic Health Records ». *Information Systems Research* 28 (3): 468-89.

Callon, Michel. 1985. *Les scientifiques et leurs alliés*. Traduit par Dominique Ebnöther. Paris: Pandora.

———. 1986. « Éléments pour une sociologie de la traduction: La domestication des coquilles Saint-Jacques et des marins-pêcheurs dans la baie de Saint-Brieuc », *L'année sociologique*, 36:41.

Campello, Victor M., Polyxeni Gkontra, Cristian Izquierdo, Carlos Martin-Isla, Alireza Sojoudi, Peter M. Full, Klaus Maier-Hein, et al. 2021. « Multi-Centre, Multi-Vendor and Multi-Disease Cardiac Segmentation: The M&Ms Challenge ». *IEEE Transactions on Medical Imaging* 40 (12): 3543-54.

Carlile, Paul R. 2002. « A Pragmatic View of Knowledge and Boundaries: Boundary Objects in New Product Development ». *Organization Science* 13 (4): 442-55.

———. 2004. « Transferring, Translating, and Transforming: An Integrative Framework for Managing Knowledge Across Boundaries ». *Organization Science* 15 (5): 555-68.

Carlin, C., A. Taylor, I. van Loon, G. McDowell, S. Burns, P. McGinness, et D.J. Lowe. 2021. « Role for artificial intelligence in respiratory diseases-chronic obstructive pulmonary disease ». *Journal of Hospital Management and Health Policy* 5 (September).

Carroll, Micah, Rohin Shah, Mark K. Ho, Tom Griffiths, Sanjit Seshia, Pieter Abbeel, et Anca Dragan. 2019. « On the utility of learning about humans for human-ai coordination ». *Advances in neural information processing systems* 32.

Cazacu, Mihaela, et Emilia Titan. 2021. « Adapting CRISP-DM for social sciences ». *BRAIN. Broad Research in Artificial Intelligence and Neuroscience* 11 (2Sup1): 99-106.

CCNE, et CNPEN. 2022. « Diagnostic Médical et Intelligence Artificielle : Enjeux Éthiques. Avis commun du CCNE et du CNPEN, Avis 141 du CCNE, Avis 4 du CNPEN ». Avis 141 du CCNE, Avis 4 du CNPEN.

Cellard, Loup. 2022. « Algorithms as Figures: Towards a Post-Digital Ethnography of Algorithmic Contexts ». *New Media & Society* 24 (4).

Challen, Robert, Joshua Denny, Martin Pitt, Luke Gompels, Tom Edwards, et Krasimira Tsaneva-Atanasova. 2019. « Artificial Intelligence, Bias and Clinical Safety ». *BMJ Quality & Safety* 28 (3): 231-37.

Chandler, Alfred D. 1993. *The Visible Hand: The Managerial Revolution in American Business*. Harvard University Press.

Chen, W., W. He, J. Shen, X. Tian, et X. Wang. 2023. « Systematic analysis of artificial intelligence in the era of industry 4.0 ». *Journal of Management Analytics* 10 (1): 89-108.

Chhotray, Vasudha, et Gerry Stoker. 2009. *Governance theory and practice: a cross-disciplinary approach*. Basingstoke [England] ; New York: Palgrave Macmillan.

Choo, Chun Wei. 1991. « Towards an information model of organizations ». *The Canadian Journal of Information Science* 16 (3): 32-62.

Choudhury, Prithwiraj, Ryan T. Allen, et Michael G. Endres. 2021. « Machine Learning for Pattern Discovery in Management Research ». *Strategic Management Journal* 42 (1): 30-57.

Christin, Angèle. 2020. « The Ethnographer and the Algorithm: Beyond the Black Box ». *Theory and Society* 49 (5-6): 897-918.

Clark, Andy. 2008. *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. 1re éd. Oxford University Press New York.

- Clark, Andy, et David Chalmers. 1998. « The Extended Mind ». *Analysis* 58 (1): 7-19.
- Cohen, Lisa E., et Sara Mahabadi. 2022. « In the Midst of Hiring: Pathways of Anticipated and Accidental Job Evolution During Hiring ». *Organization Science* 33 (5): 1938-63.
- Coiera, Enrico. 2015. *Guide to Health Informatics*. Third edition. Boca Raton: CRC Press, Taylor & Francis Group.
- Colapietro, Vincent Michael. 1989. *Pierce's Approach to the Self: A Semiotic Perspective on Human Subjectivity*. SUNY Series in Philosophy. Albany, NY: State university of New York press.
- Conant, Emily F., Alicia Y. Toledano, Senthil Periaswamy, Sergei V. Fotin, Jonathan Go, Justin E. Boatsman, et Jeffrey W. Hoffmeister. 2019. « Improving Accuracy and Efficiency with Concurrent Use of Artificial Intelligence for Digital Breast Tomosynthesis ». *Radiology: Artificial Intelligence* 1 (4): e180096. <https://doi.org/10.1148/ryai.2019180096>.
- Cram, W. Alec, et Martin Wiener. 2020. « Technology-mediated Control: Case Examples and Research Directions for the Future of Organizational Control ». *Communications of the Association for Information Systems*, 70-91.
- Creative Destruction Lab, réal. 2016. Geoff Hinton: On Radiology. <https://www.youtube.com/watch?v=2HMpRXstSvQ>.
- Cresswell, Kathrin M, Allison Worth, et Aziz Sheikh. 2010. « Actor-Network Theory and Its Role in Understanding the Implementation of Information Technology Developments in Healthcare ». *BMC Medical Informatics and Decision Making* 10 (1): 67.
- Cui, Miao, et David Y. Zhang. 2021. « Artificial Intelligence and Computational Pathology ». *Laboratory Investigation* 101 (4): 412-22.
- Cydulka, Rita K., Charles L. Emerman, et Nicholas J. Jouriles. 1996. « Evaluation of Resident Performance and Intensive Bedside Teaching during Direct Observation ». *Academic Emergency Medicine* 3 (4): 345-51.
- Czarniawska, Barbara. 1999a. « Management She Wrote: Organization Studies and Detective Stories ». *Studies in Cultures, Organizations and Societies* 5 (1): 13-41.
- . 1999b. *Writing Management: Organization Theory as a Literary Genre*. Oxford University Press.
- . 2021. « How to shadow organizing ». In *Organizational Ethnography*, Edward Elgar, Cheltenham, 45-58.
- D'Adderio, Luciana. 2011. « Artifacts at the Centre of Routines: Performing the Material Turn in Routines Theory ». *Journal of Institutional Economics* 7 (2): 197-230.
- Dagognet, François. 1973. *Écriture et iconographie. Problèmes et controverses*. Paris: Vrin.
- Dai, Tinglong, et Sridhar Tayur. 2022. « Designing AI-augmented Healthcare Delivery Systems for Physician Buy-in and Patient Acceptance ». *Production and Operations Management* 31 (12).
- Dane, Erik, Kevin W. Rockmann, et Michael G. Pratt. 2012. « When Should I Trust My Gut? Linking Domain Expertise to Intuitive Decision-Making Effectiveness ». *Organizational Behavior and Human Decision Processes* 119 (2): 187-94.

Das, Asha, Madhu S. Nair, et S. David Peter. 2020. « Computer-Aided Histopathological Image Analysis Techniques for Automated Nuclear Atypia Scoring of Breast Cancer: A Review ». *Journal of Digital Imaging* 33 (5).

Daugherty, Paul R., et H. James Wilson. 2018. *Human+ machine: Reimagining work in the age of AI*. Harvard Business Press.

Davenport, Thomas H., et Julia Kirby. 2016. « Just how smart are smart machines? »

David, Albert. 1998. « Outils de gestion et dynamique du changement ».

Davidson, Elizabeth J. 2002. « Technology Frames and Framing: A Socio-Cognitive Investigation of Requirements Determination ». *MIS Quarterly* 26 (4): 329-58.

Davis, Gerald F. 2015. « What might replace the modern corporation: Uberization and the web page enterprise ». *Seattle UL Rev.* 39:501.

———. 2016. « Can an Economy Survive Without Corporations? Technology and Robust Organizational Alternatives ». *Academy of Management Perspectives* 30 (2): 129-40.

Davis, Stanley M. 1987. *Future perfect*. Reading, Mass: Addison-Wesley.

D’Cruz, Premilla, et Ernesto Noronha. 2016. « Positives Outweighing Negatives: The Experiences of Indian Crowdsourced Workers ». *Work Organisation, Labour and Globalisation* 10 (1).

De Vreede, Gert-Jan, et Robert O. Briggs. 2019. « A Program of Collaboration Engineering Research and Practice: Contributions, Insights, and Future Directions ». *Journal of Management Information Systems* 36 (1): 74-119.

Deledalle, Gérard. 1987. « Quelle Philosophie Pour La Sémiotique Peircienne? Peirce et La Sémiotique Grecque » 63 (3-4): 241-52.

Deleuze, Gilles, et M. Foucault. 1986. « Deleuze présente pouvoir et force comme équivalentes:«le pouvoir est un rapport de forces» ». Minit.

Dellermann, Dominik, Philipp Ebel, Matthias Söllner, et Jan Marco Leimeister. 2019. « Hybrid Intelligence ». *Business & Information Systems Engineering* 61 (5): 637-43.

Denzin, Norman K. 2009. *The research act: a theoretical introduction to sociological methods*. Routledge. New York: AldineTransaction.

DeSanctis, Gerardine, et Marshall Scott Poole. 1994. « Capturing the complexity in advanced technology use: Adaptive structuration theory ». *Organization science* 5 (2): 121-47.

Desrosières, Alain, et Sandrine Kott. 2005. « Quantifier ». *Genèses* 58 (1): 2.

Dewey, John. 1966. « Democracy and education ». Jo Ann Boydston (ed.). *The Middle Works of John Dewey 9:1899-1924*.

———. 1986. *Logic: The Theory of Inquiry*. J. A. Boydston (ed.), *The Later Works, 1925–1953, Vol. 12 (Original work published 1938.)*. Carbondale Edwardsville: Southern Illinois University Press.

Diakopoulos, Nicholas. 2015. « Algorithmic Accountability: Journalistic Investigation of Computational Power Structures ». *Digital Journalism* 3 (3): 398-415.

———. 2020. « Computational news discovery: Towards design considerations for editorial orientation algorithms in journalism ». *Digital Journalism* 8 (7): 945-67.

- Dick, Stephanie. 2019. « Artificial Intelligence ». *Harvard Data Science Review*, juin.
- Dietvorst, Berkeley J., et Soaham Bharti. 2020. « People Reject Algorithms in Uncertain Decision Domains Because They Have Diminishing Sensitivity to Forecasting Error ». *Psychological Science* 31 (10): 1302-14.
- Dietvorst, Berkeley J., Joseph P. Simmons, et Cade Massey. 2015. « Algorithm Aversion: People Erroneously Avoid Algorithms after Seeing Them Err. » *Journal of Experimental Psychology: General* 144 (1): 114-26.
- Dodge, Martin, et Rob Kitchin. 2007. « The Automatic Management of Drivers and Driving Spaces ». *Geoforum* 38 (2): 264-75.
- Donovan, Joan, Robyn Caplan, Jeanna Matthews, et Lauren Hanson. 2018. « Algorithmic Accountability: A Primer ». Report. Data & Society Research Institute.
- Dourish, Paul. 2016. « Algorithms and Their Others: Algorithmic Culture in Context ». *Big Data & Society* 3 (2): 1-11.
- Du mode d'existence des outils de gestion. 1997. Paris: Seli Arslan.
- Dumez, Hervé. 2011. « Qu'est-ce que la recherche qualitative ? » *Le Libellio d'AEGIS* 7 (4-Hiver): 47-58.
- . 2012. « Qu'est-ce que l'abduction, et en quoi peut-elle avoir un rapport avec la recherche qualitative ». *Le libellio d'Aegis* 8 (3): 3-9.
- . 2013. « Qu'est-ce que la recherche qualitative ? Problèmes épistémologiques, méthodologiques et de théorisation ». *Annales des Mines - Gérer et comprendre* 112 (2): 29.
- . 2021. *Méthodologie de la recherche qualitative: les 10 questions clés de la démarche compréhensive*. 3e éd. Les spécialités du sup. Paris: Vuibert.
- . 2023. « Réflexions sur l'étude de cas (explorer, écrire, théoriser) ». *Recherches Qualitatives*, no 28, 70-80.
- Dumez, Hervé, et Étienne Minvielle. 2017. « L'e-santé rend-elle la démocratie sanitaire pleinement performative ? » *Systèmes d'information & management* 22 (1): 9.
- Dumez, Hervé, et Camille Toussaint. 2022. « Le grand oubli des sciences de gestion et de la théorie des organisations ». *Le Libellio d'AEGIS* 18 (1): 7-19.
- Dumond, Jean-Paul. 2022. *Refonder les organisations de santé: vers l'hôpital post-industriel*. Paris: Seli Arslan.
- Dwivedi, Yogesh K., Laurie Hughes, Elvira Ismagilova, Gert Aarts, Crispin Coombs, Tom Crick, Yanqing Duan, et al. 2021. « Artificial Intelligence (AI): Multidisciplinary Perspectives on Emerging Challenges, Opportunities, and Agenda for Research, Practice and Policy ». *International Journal of Information Management* 57 (avril).
- Eckstein, Harry. 2000. « Case Study and Theory in Political Science ». In *Case Study Method: Key Issues, Key Texts*, édité par Roger Gomm, Martyn Hammersley, et Peter Foster. SAGE.
- Eco, Umberto. 1983. « Horns, hooves, insteps: Some hypotheses on three types of abduction ». In *The sign of three: Dupin, Holmes, Peirce*, édité par Umberto Eco et Thomas A. Sebeok, 198-220. *Advances in semiotics*. Bloomington: Indiana University Press.

Ekins, Sean. 2016. « The Next Era: Deep Learning in Pharmaceutical Research ». *Pharmaceutical Research* 33 (11): 2594-2603.

Elbanna, Amany R. 2016. « Doing Sociomateriality Research in Information Systems ». *ACM SIGMIS Database: The DATABASE for Advances in Information Systems* 47 (4): 84-92.

Elder, Alexandra, Christina Ring, Kerry Heitmiller, Zena Gabriel, et Nazanin Saedi. 2021. « The Role of Artificial Intelligence in Cosmetic Dermatology—Current, Upcoming, and Future Trends ». *Journal of Cosmetic Dermatology* 20 (1): 48-52.

Ellahham, Samer, Nour Ellahham, et Mecit Can Emre Simsekler. 2020. « Application of Artificial Intelligence in the Health Care Safety Context: Opportunities and Challenges ». *American Journal of Medical Quality* 35 (4): 341-48.

Elsbach, Kimberly D. 2003. « Relating Physical Environment to Self-Categorizations: Identity Threat and Affirmation in a Non-Territorial Office Space ». *Administrative Science Quarterly* 48 (4): 622-54.

Elson, J S, Douglas C Derrick, et Gina S Ligon. 2020. « Trusting a Humanoid Robot: Exploring Personality and Trusting Effects in a Human-Robot Partnership ». In . Hawaii, US.

Engel, F., F. Kletz, et Jean-Claude Moisdon. 1997. « L'instrument multiprise ou 'les métamorphoses des ICR' ». In *Du mode d'existence des outils de gestion, Les instruments de gestion à l'épreuve de l'organisation*, 114-33.

Engel, François, Frédéric Kletz, Jean-Claude Moisdon, et Dominique Tonneau. 2000. « La démarche gestionnaire à l'hôpital, le PMSI, tome 2: La régulation du système hospitalier ». Paris, Seli Arslan.

Engelbart, Douglas C. 1962. « Augmenting human intellect: A conceptual framework ». SRI Project No. 3578 (AUGMENT,3906,) AFOSR-3223. Washington DC: Director of Information Sciences, Air Force Office of Scientific Research.

Espeland, Wendy Nelson, et Mitchell L. Stevens. 2008. « A Sociology of Quantification ». *European Journal of Sociology* 49 (3): 401-36.

European Commission, Directorate-General for Communications Networks, Content and Technology. 2019. *Ethics Guidelines for Trustworthy AI*. Publications Office of the European Union.

European Parliament, Directorate General for Parliamentary Research Services. 2022. *Artificial Intelligence in Healthcare: Applications, Risks, and Ethical and Societal Impacts*. LU: Publications Office.

Evans, Karen G. 2000. « Reclaiming John Dewey: Democracy, Inquiry, Pragmatism, and Public Management ». *Administration & Society* 32 (3): 308-28.

Farah, Line, Juliette M. Murriss, Isabelle Borget, Agathe Guilloux, Nicolas M. Martelli, et Sandrine I.M. Katsahian. 2023. « Assessment of Performance, Interpretability, and Explainability in Artificial Intelligence–Based Health Technologies: What Healthcare Stakeholders Need to Know ». *Mayo Clinic Proceedings: Digital Health* 1 (2): 120-38. <https://doi.org/10.1016/j.mcpdig.2023.02.004>.

Faraj, Samer, et Paul M Leonardi. 2022. « Strategic Organization in the Digital Age: Rethinking the Concept of Technology ». *Strategic Organization* 20 (4): 771-85.

Faraj, Samer, Stella Pachidi, et Karla Sayegh. 2018. « Working and Organizing in the Age of the Learning Algorithm ». *Information and Organization* 28 (1): 62-70. <https://doi.org/10.1016/j.infoandorg.2018.02.005>.

Fayard, Anne-Laure, et Anca Metiu. 2014. « The Role of Writing in Distributed Collaboration ». *Organization Science* 25 (5): 1391-1413.

Fayyad, Usama M., Gregory Piatetsky-Shapiro, et Padhraic Smyth. 1996a. « From data mining to knowledge discovery in databases ». *AI magazine* 17 (3): 37-37.

———. 1996b. « The KDD process for extracting useful knowledge from volumes of data ». *Communications of the ACM* 39 (11): 27-34.

Fazal, Mohammad Ihsan, Muhammed Ebrahim Patel, Jamie Tye, et Yuri Gupta. 2018. « The Past, Present and Future Role of Artificial Intelligence in Imaging ». *European Journal of Radiology* 105 (août):246-50.

Fetzer, James H. 1990. *Artificial Intelligence : Its Scope and Limits*. Dordrecht ; Boston : Kluwer Academic Publishers.

Fihn, Stephan, Suchi Saria, Eneida Mendonça, Seth Hain, Michael Matheny, Nigam Shah, Hongfang Liu, et Andrew Auerbach. 2019. « Deploying AI in clinical settings ». In *Artificial intelligence in health care: The hope, the hype, the promise, the peril*, édité par Michael Matheny, Sonoo Thadaneys Israni, et Danielle Whicher. Vol. 145.

Forsblad, Peter. 1984. « Observation for what? » In *Leaders and Managers: International Perspectives on Managerial Behaviour and Leadership*, édité par D. Hunt, Dian-Marie Hosking, Chester A. Schriesheim, et R Stewart, 200-203. New York: Pergamon: Elsevier.

Foucault, Michel. 2003. *Surveiller et punir: naissance de la prison*. Bibliothèque des Histoires. Paris: Gallimard.

Frank, Alejandro Germán, Lucas Santos Dalenogare, et Néstor Fabián Ayala. 2019. « Industry 4.0 Technologies: Implementation Patterns in Manufacturing Companies ». *International Journal of Production Economics* 210 (avril):15-26.

Freidson, Eliot. 1970. *Professional Dominance: The Social Structure of Medical Care*. Transaction Publishers.

Fügener, Andreas, Jörn Grahl, Alok Gupta, et Wolfgang Ketter. 2022. « Cognitive Challenges in Human–Artificial Intelligence Collaboration: Investigating the Path Toward Productive Delegation ». *Information Systems Research* 33 (2): 678-96.

Fügener, Andreas, Jörn Grahl, Alok Gupta, et Wolfgang Wahlster. 2021. « Will Humans-in-the-Loop Become Borgs? Merits and Pitfalls of Working with AI ». *MIS Quarterly* 45 (3): 1527-56.

Gal, Uri, Tina Blegind Jensen, et Mari-Klara Stein. 2020. « Breaking the Vicious Cycle of Algorithmic Management: A Virtue Ethics Approach to People Analytics ». *Information and Organization* 30 (2).

Galetsis, P., K. Katsaliaki, et S. Kumar. 2019. « Values, Challenges and Future Directions of Big Data Analytics in Healthcare: A Systematic Review ». *Social Science & Medicine* 241 (novembre).

Gao, Ruijiang, Maytal Saar-Tsechansky, Maria De-Arteaga, Ligong Han, Min Kyung Lee, et Matthew Lease. 2021. « Human-AI Collaboration with Bandit Feedback ».

Gawande, Atul. 2011. « The Checklist Manifesto: How to Get Things Right ». *Journal of Nursing Regulation* 1 (4): 64.

Geertman, Stan, Qingming Zhan, Andrew Allan, et Christopher Pettit, éd. 2019. *Computational Urban Planning and Management for Smart Cities. Lecture Notes in Geoinformation and Cartography*. Cham: Springer International Publishing.

Geiger, R. Stuart. 2014. « Bots, bespoke, code and the materiality of software platforms ». *Information, Communication & Society* 17 (3): 342-56.

George, Alexander L., et Andrew Bennett. 2005. *Case studies and theory development in the social sciences*. mit Press.

Gerke, Sara. 2021. « Health AI for good rather than evil? The need for a new regulatory framework for AI-based medical devices ». *Yale J. Health Pol'y L. & Ethics* 20:432.

Gerke, Sara, Boris Babic, Theodoros Evgeniou, et I. Glenn Cohen. 2020. « The Need for a System View to Regulate Artificial Intelligence/Machine Learning-Based Software as Medical Device ». *Npj Digital Medicine* 3 (1): 53.

Gerring, John. 2004. « What Is a Case Study and What Is It Good For? ». *American Political Science Review* 98 (2): 341-54.

Gibson, James J. 1982. *Reasons for Realism: Selected Essays*. Édité par Edward Reed et Rebecca Jones. *Resources for Ecological Psychology*. Hillsdale, N.J: Lawrence Erlbaum.

———. 2014. *The Ecological Approach to Visual Perception: Classic Edition*. New York: Psychology Press.

Gichoya, Judy Wawira, Imon Banerjee, Ananth Reddy Bhimireddy, John L Burns, Leo Anthony Celi, Li-Ching Chen, Ramon Correa, et al. 2022. « AI Recognition of Patient Race in Medical Imaging: A Modelling Study ». *The Lancet Digital Health* 4 (6).

Gill, T. Grandon. 1995. « Early Expert Systems: Where Are They Now? ». *MIS Quarterly* 19 (1): 51.

Gillespie, Tarleton. 2012. « Can an Algorithm Be Wrong? ». *Limn*, 2012.

———. 2014. « The Relevance of Algorithms ». In *Media Technologies*, édité par Tarleton Gillespie, Pablo J. Boczkowski, et Kirsten A. Foot, 167-94. The MIT Press.

Girin, Jacques. 1989. « L'opportunisme méthodique dans les recherches sur la gestion des organisations ». Communication à la journée d'étude la recherche action en action et en question, AFCET, collège de systémique, Ecole centrale de Paris.

———. 1990. « L'analyse empirique des situations de gestion: éléments de théorie et de méthode ». *Epistémologies et sciences de gestion* 1990:141-82.

Gitelman, Lisa, éd. 2013. « Raw Data » Is an Oxymoron. The MIT Press.

Glaser, Barney G., Anselm L. Strauss, et Elizabeth Strutzel. 1968. « The Discovery of Grounded Theory; Strategies for Qualitative Research ». *Nursing Research* 17 (4): 364.

Gluyas, Heather. 2015. « Patient-centred care: improving healthcare outcomes ». *Nursing Standard (2014+)* 30 (4): 50.

Gomez-Uribe, Carlos A., et Neil Hunt. 2016. « The Netflix Recommender System: Algorithms, Business Value, and Innovation ». *ACM Transactions on Management Information Systems* 6 (4): 1-19.

Good, Irving John. 1966. « Speculations Concerning the First Ultraintelligent Machine ». In *Advances in Computers*, 6:31-88. Elsevier.

Goodman, Tyler, Michael E. Miller, Christina F. Rusnock, et Jason Bindewald. 2016. « Timing within human-agent interaction and its effects on team performance and human behavior ». In *2016 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, 35-41. San Diego, CA, USA: IEEE.

Goyal, Manu, Thomas Knackstedt, Shaofeng Yan, et Saeed Hassanpour. 2020. « Artificial Intelligence-Based Image Classification Methods for Diagnosis of Skin Cancer: Challenges and Opportunities ». *Computers in Biology and Medicine* 127 (décembre)..

Granovetter, Mark. 2005. « The Impact of Social Structure on Economic Outcomes ». *Journal of Economic Perspectives* 19 (1): 33-50.

Grassegger, Written Hannes, et Mikael Krogerus. 2017. « The Data That Turned the World Upside Down ». *Vice Motherboard*, 28 janvier 2017.

Gray, Garry C., et Susan S. Silbey. 2014. « Governing Inside the Organization: Interpreting Regulation and Compliance ». *American Journal of Sociology* 120 (1): 96-145.

Grosman, Jérémy, et Tyler Reigeluth. 2019. « Perspectives on Algorithmic Normativities: Engineers, Objects, Activities ». *Big Data & Society* 6 (2).

Gulshan, Varun, Lily Peng, Marc Coram, Martin C. Stumpe, Derek Wu, Arunachalam Narayanaswamy, Subhashini Venugopalan, et al. 2016. « Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs ». *JAMA* 316 (22): 2402.

Guo, Eileen, et Karen Hao. 2022. « This Is the Stanford Vaccine Algorithm That Left out Frontline Doctors* ». In *Ethics of Data and Analytics*, par Kirsten Martin, 1re éd., 6-9. Boca Raton: Auerbach Publications.

Guo, Yiting, Ximing Yin, De Liu, et Sean Xin Xu. 2020. « “She is not just a computer”: Gender Role of AI Chatbots in Debt Collection ».

Gusterson, Hugh. 2004. *People of the Bomb: Portraits of America’s Nuclear Complex*. U of Minnesota Press.

Haas, Aurore. 2015. « Crowding at the frontier: boundary spanners, gatekeepers and knowledge brokers ». Édité par Dr Martine Sigal Prof Claude Paraponaris. *Journal of Knowledge Management* 19 (5): 1029-47.

Hacking, Ian. 1983. *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge University Press.

Haenlein, Michael, et Andreas Kaplan. 2019. « A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence ». *California Management Review* 61 (4): 5-14.

Hafermalz, Ella, et Marleen Huysman. 2021. « Please Explain: Key Questions for Explainable AI research from an Organizational perspective ». *Morals & Machines* 1 (2): 10-23.

Haibe-Kains, Benjamin, George Alexandru Adam, Ahmed Hosny, Farnoosh Khodakarami, Massive Analysis Quality Control (MAQC) Society Board of Directors, Thakkar Shraddha, Rebecca Kusko, et al. 2020. « Transparency and Reproducibility in Artificial Intelligence ». *Nature* 586 (7829): E14-16.

Hamburg, Margaret A., et Francis S. Collins. 2010. « The Path to Personalized Medicine ». *New England Journal of Medicine* 363 (4): 301-4.

Hammersley, Martyn. 1990. « What's Wrong with Ethnography? The Myth of Theoretical Description ». *Sociology* 24 (4): 597-615.

Hampson, Grace, Adrian Towse, Steven D Pearson, William B Dreitlein, et Chris Henshall. 2018. « Gene Therapy: Evidence, Value and Affordability in the US Health Care System ». *Journal of Comparative Effectiveness Research* 7 (1): 15-28.

Hang, Haiming, et Zhifeng Chen. 2022. « How to Realize the Full Potentials of Artificial Intelligence (AI) in Digital Economy? A Literature Review ». *Journal of Digital Economy* 1 (3): 180-91.

Hannan, Michael T., et John Freeman. 1987. « The Ecology of Organizational Founding: American Labor Unions, 1836-1985 ». *American Journal of Sociology* 92 (4): 910-43. <https://doi.org/10.1086/228587>.

Hansen, Hans. 2008. « Abduction ». In *The sage handbook of new approaches in management and organization*, 454-63. SAGE Publications Inc.

Hao, Karen. 2020. « The Coming War on the Hidden Algorithms That Trap People in Poverty ». *MIT Technology Review*, 4 décembre 2020. <https://www.technologyreview.com/2020/12/04/1013068/algorithms-create-a-poverty-trap-lawyers-fight-back/>.

Haraway, Donna. 2006. « A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late Twentieth Century ». In *The Transgender Studies Reader*. Routledge.

Hardy, Cynthia, Thomas B. Lawrence, et David Grant. 2005. « Discourse and collaboration: The role of conversations and collective identity ». *Academy of management review* 30 (1): 58-77.

Harvey, H. Benjamin, et Vrushab Gowda. 2020. « How the FDA Regulates AI ». *Academic Radiology* 27 (1): 58-61.

Harwell, Drew. 2022. « A Face-Scanning Algorithm Increasingly Decides Whether You Deserve the Job * ». In *Ethics of Data and Analytics*. Auerbach Publications.

Hatchuel, Armand. 2017. « Les métamorphoses de la confiance dans l'échange marchand Petite histoire des compteurs d'eau ». In *Research*, n° 6. Classiques Garnier.

Hatchuel, Armand, et Jean-Claude Moisdon. 1993. « Modèles et apprentissage organisationnel ». *Cahiers d'Economie et de Sociologie Rurales (CESR)* 28: 17-32.

———. 1997. « Modèles d'aide à la décision ou modèles d'organisation? » In *Du mode d'existence des outils de gestion*, 46-67.

Hatchuel, Armand, et Hughues Molet. 1986. « Rational modelling in understanding and aiding human decision-making: About two case studies ». *European Journal of Operational Research, OR and Microcomputers Miscellaneous OR Applications*, 24 (1): 178-86.

- Hatchuel, Armand, et Benoit Weil. 1992. « L'expert et le système ». *Economica*, Paris 192.
- Haugeland, John. 1993. « Mind embodied and embedded ». In *Mind and Cognition: 1993 International Symposium*, édité par Yu-Houng H. Houg et J. Ho, 233-67. *Academica Sinica*.
- Haute Autorité de Santé. 2022. « Annexe 6. Informations descriptives spécifiques à fournir pour les fonctionnalités du dispositif médical s'appuyant sur des procédés d'apprentissage automatique (technologies relevant du champ de l'intelligence artificielle) ».
- Hays, James, et Alexei A. Efros. 2007. « Scene Completion Using Millions of Photographs ». *ACM Transactions on Graphics* 26 (3): 4.
- Heaton, J. B., N. G. Polson, et J. H. Witte. 2017. « Deep Learning for Finance: Deep Portfolios ». *Applied Stochastic Models in Business and Industry* 33 (1): 3-12. <https://doi.org/10.1002/asmb.2209>.
- Hedström, Peter, et Peter Bearman. 2009. *The Oxford Handbook of Analytical Sociology*. OUP Oxford.
- Hedström, Peter, et Richard Swedberg. 1998. *Social Mechanisms: An Analytical Approach to Social Theory*. Cambridge University Press.
- Hern, Alex. 2019. « Apple Made Siri Deflect Questions on Feminism, Leaked Papers Reveal ». *The Guardian*, 6 septembre 2019, sect. Technology. <https://www.theguardian.com/technology/2019/sep/06/apple-rewrote-siri-to-deflect-questions-about-feminism>.
- Herreid, Clyde Freeman. 1998. « What Makes a Good Case? Some Basic Rules of Good Storytelling Help Teachers Generate Student Excitement in the Classroom ». *Journal of College Science Teaching* 27 (3).
- Hodges, Andrew. 2014. *Alan Turing: The Enigma: The Book That Inspired the Film The Imitation Game*. Princeton, New Jersey: Princeton University Press.
- Hoffman, Kelly M., Sophie Trawalter, Jordan R. Axt, et M. Norman Oliver. 2016. « Racial Bias in Pain Assessment and Treatment Recommendations, and False Beliefs about Biological Differences between Blacks and Whites ». *Proceedings of the National Academy of Sciences* 113 (16).
- Hollebeek, Linda D., Jodie Conduit, Jill Sweeney, Geoffrey Soutar, Ingo O. Karpen, Wade Jarvis, et Tom Chen. 2016. « Epilogue to the Special Issue and Reflections on the Future of Engagement Research ». *Journal of Marketing Management* 32 (5-6): 586-94.
- Hosny, Ahmed, Chintan Parmar, John Quackenbush, Lawrence H. Schwartz, et Hugo J. W. L. Aerts. 2018. « Artificial Intelligence in Radiology ». *Nature Reviews Cancer* 18 (8): 500-510.
- Huang, Ming-Hui, Roland Rust, et Vojislav Maksimovic. 2019. « The Feeling Economy: Managing in the Next Generation of Artificial Intelligence (AI) ». *California Management Review* 61 (4): 43-65.
- Huang, Ming-Hui, et Roland T. Rust. 2018. « Artificial Intelligence in Service ». *Journal of Service Research* 21 (2): 155-72.
- Huising, Ruthanne. 2014. « The Erosion of Expert Control Through Censure Episodes ». *Organization Science* 25 (6): 1633-61.

Huising, Ruthanne, et Susan S. Silbey. 2018. « From Nudge to Culture and Back Again: Coalface Governance in the Regulated Organization ». *Annual Review of Law and Social Science* 14 (1): 91-114.

Hung, Shin-Yuan. 2003. « Expert versus Novice Use of the Executive Support Systems: An Empirical Study ». *Information & Management* 40 (3): 177-89.

Hunkenschroer, Anna Lena, et Christoph Luetge. 2022. « Ethics of AI-Enabled Recruiting and Selection: A Review and Research Agenda ». *Journal of Business Ethics* 178 (4): 977-1007.

Hutchby, Ian. 2001. « Technologies, Texts and Affordances ». *Sociology* 35 (2): 441-56.

———. 2013. *Conversation and Technology: From the Telephone to the Internet*. John Wiley & Sons.

Huws, Ursula, Neil Spencer, et Simon Joyce. 2016. « Crowd Work in Europe: Preliminary Results from a Survey in the UK, Sweden, Germany, Austria and the Netherlands », décembre.

Ihde, Don. 1990. « Technology and the lifeworld: From garden to earth ». <https://philpapers.org/rec/IHDTAT-3>.

Institut national du cancer. 2023. « Panorama des cancers en France – Édition 2023 ». *Etat des lieux et des connaissances / Epidémiologie*.

Introna, Lucas D. 2016. « Algorithms, Governance, and Governmentality: On Governing Academic Writing ». *Science, Technology, & Human Values* 41 (1): 17-49.

Introna, Lucas D., et Niall Hayes. 2011. « On Sociomaterial Imbrications: What Plagiarism Detection Systems Reveal and Why It Matters ». *Information and Organization* 21 (2): 107-22.

Israni, Ellora Thadaney. 2017. « Opinion | When an Algorithm Helps Send You to Prison ». *The New York Times*, 26 octobre 2017, sect. Opinion. <https://www.nytimes.com/2017/10/26/opinion/algorithm-compas-sentencing-bias.html>.

Jain, Hemant, Balaji Padmanabhan, Paul A. Pavlou, et T. S. Raghu. 2021. « Editorial for the Special Section on Humans, Algorithms, and Augmented Intelligence: The Future of Work, Organizations, and Society ». *Information Systems Research* 32 (3): 675-87.

Jaiswal, Akanksha, C. Joe Arun, et Arup Varma. 2022. « Rebooting Employees: Upskilling for Artificial Intelligence in Multinational Corporations ». *The International Journal of Human Resource Management* 33 (6): 1179-1208.

Jarrahi, Mohammad Hossein. 2018. « Artificial Intelligence and the Future of Work: Human-AI Symbiosis in Organizational Decision Making ». *Business Horizons* 61 (4): 577-86.

Jarrahi, Mohammad Hossein, Gemma Newlands, Min Kyung Lee, Christine T. Wolf, Eliscia Kinder, et Will Sutherland. 2021. « Algorithmic Management in a Work Context ». *Big Data & Society* 8 (2).

Jaton, Florian. 2020. *The constitution of algorithms: ground-truthing, programming, formulating*. Inside technology. Cambridge, Massachusetts: The MIT Press.

Jessup, Sarah A, Wright-Patterson Afb, et Anthony M Gibson. 2020. « Investigating the Effect of Trust Manipulations on Affect over Time in Human-Human versus Human-Robot Interactions ». In . Hawaii, US.

Jiang, Fei, Yong Jiang, Hui Zhi, Yi Dong, Hao Li, Sufeng Ma, Yilong Wang, Qiang Dong, Haipeng Shen, et Yongjun Wang. 2017. « Artificial Intelligence in Healthcare: Past, Present and Future ». *Stroke and Vascular Neurology* 2 (4): 230-43.

Jobert, Bruno. 1994. « Le tournant néo-libéral en Europe: l'Harmattan ». Paris.

Joerges, B., et Terry Shinn. 2001. *Instrumentation Between Science, State and Industry*. Springer Science & Business Media.

Johnson, Carolyn Y. 2022. « Racial Bias in a Medical Algorithm Favors White Patients over Sicker Black Patients* ». In *Ethics of Data and Analytics*, par Kirsten Martin, 1re éd., 10-12. Boca Raton: Auerbach Publications.

Johnson, Matthew, et Alonso H. Vera. 2019. « No AI Is an Island: The Case for Teaming Intelligence ». *AI Magazine* 40 (1): 16-28.

Jonker-Hoffrén, Paul. 2021. « Bridging the Gap between Critical Algorithm Studies and the Social Sciences: The Case of Platform Work ». *Informaatiotutkimus* 40 (3).

Jordan, M. I., et T. M. Mitchell. 2015. « Machine Learning: Trends, Perspectives, and Prospects ». *Science* 349 (6245): 255-60. <https://doi.org/10.1126/science.aaa8415>.

Jussupow, Ekaterina, Kai Spohrer, et Armin Heinzl. 2022. « Radiologists' Usage of Diagnostic AI Systems: The Role of Diagnostic Self-Efficacy for Sensemaking from Confirmation and Disconfirmation ». *Business & Information Systems Engineering* 64 (3): 293-309.

Jussupow, Ekaterina, Kai Spohrer, Armin Heinzl, et Joshua Gawlitza. 2021. « Augmenting Medical Diagnosis Decisions? An Investigation into Physicians' Decision-Making Process with Artificial Intelligence ». *Information Systems Research* 32 (3): 713-35.

Kaplan, Alan, Hui Cao, J. Mark FitzGerald, Nick Iannotti, Eric Yang, Janwillem W.H. Kocks, Konstantinos Kostikas, et al. 2021. « Artificial Intelligence/Machine Learning in Respiratory Medicine and Potential Role in Asthma and COPD Diagnosis ». *The Journal of Allergy and Clinical Immunology: In Practice* 9 (6): 2255-61.

Kaplan, Andreas, et Michael Haenlein. 2019. « Siri, Siri, in My Hand: Who's the Fairest in the Land? On the Interpretations, Illustrations, and Implications of Artificial Intelligence ». *Business Horizons* 62 (1): 15-25.

Kaplan, B. 1997. « Review: Addressing Organizational Issues into the Evaluation of Medical Systems ». *Journal of the American Medical Informatics Association : JAMIA* 4 2:94-101.

———. 2001. « Evaluating informatics applications - some alternative approaches: theory, social interactionism, and call for methodological pluralism ». *International journal of medical informatics* 64 1:39-56.

Kaplan, B., et N. Shaw. 2004. « Future directions in evaluation research: people, organizational, and social issues. » *Methods of information in medicine* 43 3:215-31.

Karabarbounis, Loukas, et Brent Neiman. 2014. « The Global Decline of the Labor Share* ». *The Quarterly Journal of Economics* 129 (1): 61-103.

Kellogg, Katherine C., Melissa A. Valentine, et Angéle Christin. 2020. « Algorithms at Work: The New Contested Terrain of Control ». *Academy of Management Annals* 14 (1): 366-410. <https://doi.org/10.5465/annals.2018.0174>.

Kellogg, Katherine C. 2009. « Operating Room: Relational Spaces and Microinstitutional Change in Surgery ». *American Journal of Sociology* 115 (3): 657-711.

Kelly, Christopher J., Alan Karthikesalingam, Mustafa Suleyman, Greg Corrado, et Dominic King. 2019. « Key Challenges for Delivering Clinical Impact with Artificial Intelligence ». *BMC Medicine* 17 (1): 195.

Khadpe, Pranav, Ranjay Krishna, Li Fei-Fei, Jeffrey T. Hancock, et Michael S. Bernstein. 2020. « Conceptual Metaphors Impact Perceptions of Human-AI Collaboration ». *Proceedings of the ACM on Human-Computer Interaction* 4 (CSCW2): 1-26.

Khanijahani, Ahmad, Shabnam Iezadi, Sage Dudley, Megan Goettler, Peter Kroetsch, et Jama Wise. 2022. « Organizational, Professional, and Patient Characteristics Associated with Artificial Intelligence Adoption in Healthcare: A Systematic Review ». *Health Policy and Technology* 11 (1).

Kim, Sooyong, et Byungwoong Kwon. 2022. « A Analysis on the Role of AI Recommendation System in Social Media through Actor-Network Theory ». *The Korean Society of Culture and Convergence* 44 (9): 117-29.

Kitchin, Rob. 2014. « Big Data, New Epistemologies and Paradigm Shifts ». *Big Data & Society* 1 (1).

———. 2017. « Thinking Critically about and Researching Algorithms ». *Information, Communication & Society* 20 (1): 14-29.

Kline, Stephen Jay. 2020. *Conceptual Foundations for Multidisciplinary Thinking*. Stanford University Press.

Kluttz, Daniel N., et Deirdre K. Mulligan. 2019. « Automated Decision Support Technologies and the Legal Profession ». *Berkeley Technology Law Journal* 34 (3): 853-90.

Knight, Will. 2017. « The Dark Secret at the Heart of AI ». *MIT Technology Review*, 11 avril 2017.

Koenig, GERARD. 2009. « L'étude de cas à visée infirmationniste ». *Le Libellio d'Aegis* 5 (4): 7-13.

Kooiman, Jan. 2008. « Exploring the Concept of Governability ». *Journal of Comparative Policy Analysis: Research and Practice* 10 (2): 171-90.

Koppel, Ross, et Suzanne Gordon, éd. 2017. *First, Do Less Harm: Confronting the Inconvenient Problems of Patient Safety*. Cornell University Press.

Kose, Ilker, Mehmet Gokturk, et Kemal Kilic. 2015. « An Interactive Machine-Learning-Based Electronic Fraud and Abuse Detection System in Healthcare Insurance ». *Applied Soft Computing* 36 (novembre):283-99.

Kostick-Quenet, Kristin M., et Sara Gerke. 2022. « AI in the Hands of Imperfect Users ». *Npj Digital Medicine* 5 (1): 197.

Krämer, Nicole C., Astrid Von Der Pütten, et Sabrina Eimler. 2012. « Human-Agent and Human-Robot Interaction Theory: Similarities to and Differences from Human-Human Interaction ». In *Human-Computer Interaction: The Agency Perspective*, édité par Marielba Zacarias et José Valente De Oliveira, 396:215-40. *Studies in Computational Intelligence*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Krittanawong, Chayakrit, HongJu Zhang, Zhen Wang, Mehmet Aydar, et Takeshi Kitai. 2017. « Artificial Intelligence in Precision Cardiovascular Medicine ». *Journal of the American College of Cardiology* 69 (21): 2657-64.

Krogh, Georg von. 2018. « Artificial Intelligence in Organizations: New Opportunities for Phenomenon-Based Theorizing ». *Application/pdf. Academy of Management Discoveries* 4 (4): 404-9.

Kuhn, Thomas S. 1962. *The structure of scientific revolutions*. 1st edition. Chicago, Illinois: University of Chicago Press.

Kwon, Seok-Woo, Emanuela Rondi, Daniel Z. Levin, Alfredo De Massis, et Daniel J. Brass. 2020. « Network Brokerage: An Integrative Review and Future Research Agenda ». *Journal of Management* 46 (6): 1092-1120.

La Science telle qu'elle se fait, une anthologie de la sociologie des sciences de langue anglaise. 1982. Paris: Pandore.

Lampel, Joseph, et Henry Mintzberg. 1996. « Customizing Customization ». *Sloan Management Review* 38 (janvier).

Lange, Ann-Christina, Marc Lenglet, et Robert Seyfert. 2019. « On Studying Algorithms Ethnographically: Making Sense of Objects of Ignorance ». *Organization* 26 (4): 598-617.

Lapointe et Rivard. 2005. « A Multilevel Model of Resistance to Information Technology Implementation ». *MIS Quarterly* 29 (3): 461.

Larson, Erik J. 2021. *The Myth of Artificial Intelligence: Why Computers Can't Think the Way We Do*. Cambridge, Massachusetts: The Belknap Press of Harvard University Press.

Lascombes, Pierre, et Patrick Le Galès. 2004. « Gouverner par les instruments ».

Lash, Scott. 2007. « Power after Hegemony: Cultural Studies in Mutation? » *Theory, Culture & Society* 24 (3): 55-78.

Latour, Bruno. 1992. « Where are the missing masses? The sociology of a few mundane artifacts ». In *Shaping technology/building society: Studies in sociotechnical change*, 1:225-58. Cambridge.

———. 1993. *We Have Never Been Modern*. Cambridge, Mass: Harvard University Press.

———. 1996. « On actor-network theory: A few clarifications ». *Soziale Welt* 47 (4): 369-81.

Latour, Bruno, et Steve Woolgar. 1979. *Laboratory life: the social construction of scientific facts*. Sage library of social research ; v. 80. Beverly Hills: Sage Publications.

Lave, Jean, et Etienne Wenger. 1991. *Situated Learning: Legitimate Peripheral Participation*. 1re éd. Cambridge University Press.

Le Ludec, Clément, Maxime Cornet, et Antonio A Casilli. 2023. « The Problem with Annotation. Human Labour and Outsourcing between France and Madagascar ». *Big Data & Society* 10 (2): 20539517231188723.

Leavy, Patricia, et Anne Harris. 2018. *Contemporary Feminist Research from Theory to Practice*. Guilford Publications.

Lebovitz, Sarah, Natalia Levina, et Hila Lifshitz-Assa. 2021. « Is AI Ground Truth Really True? The Dangers of Training and Evaluating AI Tools Based on Experts' Know-What ». *MIS Quarterly* 45 (3): 1501-26.

Lebovitz, Sarah, Hila Lifshitz-Assaf, et Natalia Levina. 2022. « To Engage or Not to Engage with AI for Critical Judgments: How Professionals Deal with Opacity When Using AI for Medical Diagnosis ». *Organization Science* 33 (1): 126-48.

Lee, Jae-woo, Hyun-sun Lim, Dong-wook Kim, Soon-ae Shin, Jinkwon Kim, Bora Yoo, et Kyung-hee Cho. 2018. « The development and implementation of stroke risk prediction model in National Health Insurance Service's personal health record ». *Computer methods and programs in biomedicine* 153:253-57.

Lee, Min Kyung, Daniel Kusbit, Evan Metsky, et Laura Dabbish. 2015. « Working with Machines: The Impact of Algorithmic and Data-Driven Management on Human Workers ». In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 1603-12. Seoul Republic of Korea: ACM.

Legg, Shane, et Marcus Hutter. 2007. « Universal Intelligence: A Definition of Machine Intelligence ». *Minds and Machines* 17 (4): 391-444.

Lenglet, Marc. 2011. « Conflicting codes and codings: How algorithmic trading is reshaping financial regulation ». *Theory, Culture & Society* 28 (6): 44-66.

Leonardi. 2011. « When Flexible Routines Meet Flexible Technologies: Affordance, Constraint, and the Imbrication of Human and Material Agencies ». *MIS Quarterly* 35 (1): 147.

Leonardi, Paul M. 2012. « Materiality, Sociomateriality, and Socio-Technical Systems: What Do These Terms Mean? How Are They Different? Do We Need Them? » In *Materiality and Organizing: Social Interaction in a Technological World*, édité par Bonnie A. Nardi, Jannis Kallinikos, et Paul M. Leonardi. OUP Oxford.

———. 2013. « Theoretical Foundations for the Study of Sociomateriality ». *Information and Organization* 23 (2): 59-76.

Leotsakos, A., H. Zheng, R. Croteau, J. M. Loeb, H. Sherman, C. Hoffman, L. Morganstein, et al. 2014. « Standardization in Patient Safety: The WHO High 5s Project ». *International Journal for Quality in Health Care* 26 (2): 109-16.

Levina, Natalia. 2005. « Collaborating on Multiparty Information Systems Development Projects: A Collective Reflection-in-Action View ». *Information Systems Research* 16 (2): 109-30.

Levina et Vaast. 2005. « The Emergence of Boundary Spanning Competence in Practice: Implications for Implementation and Use of Information Systems ». *MIS Quarterly* 29 (2): 335.

Li, David, Kulamakan Kulasegaram, et Brian D. Hodges. 2019. « Why We Needn't Fear the Machines: Opportunities for Medicine in a Machine Learning World ». *Academic Medicine* 94 (5): 623-25.

Li, Jiaoyang, Lingxiao Zhou, Yi Zhan, Haifeng Xu, Cheng Zhang, Fei Shan, et Lei Liu. 2022. « How Does the Artificial Intelligence-Based Image-Assisted Technique Help Physicians in Diagnosis of Pulmonary Adenocarcinoma? A Randomized Controlled Experiment of Multicenter Physicians in China ». *Journal of the American Medical Informatics Association* 29 (12): 2041-49.

Licklider, Joseph CR. 1960. « Man-computer symbiosis ». IRE transactions on human factors in electronics, no 1, 4-11.

Lijphart, Arend. 1971. « Comparative Politics and the Comparative Method ». American Political Science Review 65 (3): 682-93.

Lindebaum, Dirk, Mikko Vesa, et Frank Den Hond. 2020. « Insights From “The Machine Stops to Better Understand Rational Assumptions in Algorithmic Decision Making and Its Implications for Organizations ». Academy of Management Review 45 (1): 247-63.

Little, Greg, Lydia B. Chilton, Max Goldman, et Robert C. Miller. 2010. « TurKit: Human Computation Algorithms on Mechanical Turk ». In Proceedings of the 23rd Annual ACM Symposium on User Interface Software and Technology, 57-66. New York New York USA: ACM.

Liu, Han, Vivian Lai, et Chenhao Tan. 2021. « Understanding the Effect of Out-of-Distribution Examples and Interactive Explanations on Human-AI Decision Making ». Proceedings of the ACM on Human-Computer Interaction 5 (CSCW2): 1-45.

Liu, Xiaoxuan, Livia Faes, Aditya U Kale, Siegfried K Wagner, Dun Jack Fu, Alice Bruynseels, Thushika Mahendiran, et al. 2019. « A Comparison of Deep Learning Performance against Health-Care Professionals in Detecting Diseases from Medical Imaging: A Systematic Review and Meta-Analysis ». The Lancet Digital Health 1 (6): e271-97.

Liu, Yun-En, Travis Mandel, Emma Brunskill, et Zoran Popovic. 2014. « Trading Off Scientific Knowledge and User Learning with Multi-Armed Bandits. » In EDM, 161-68.

Locke, Karen, Karen Golden-Biddle, et Martha S. Feldman. 2008. « Perspective—Making Doubt Generative: Rethinking the Role of Doubt in the Research Process ». Organization Science 19 (6): 907-18.

Loftus, Tyler J., Patrick J. Tighe, Amanda C. Filiberto, Philip A. Efron, Scott C. Brakenridge, Alicia M. Mohr, Parisa Rashidi, Gilbert R. Upchurch, et Azra Bihorac. 2020. « Artificial Intelligence and Surgical Decision-Making ». JAMA Surgery 155 (2): 148.

Lorino, Philippe. 2007. The Instrumental Genesis of Collective Activity: The Case of an ERP Implementation in a Large Electricity Producer. ESSEC.

———. 2014. « Peirce ». In The Oxford Handbook of Process Philosophy and Organization Studies., In J. Helin, T. Hernes, D. Hjorth, and R. Holt (eds.), 143-65. Oxford, UK: Oxford University Press.

———. 2018. Pragmatism and organization studies. Oxford University Press.

Lorino, Philippe, et J. C. Peyrolle. 2005. « Démarche pragmatiste et mise en processus dans les situations de gestion ». Entre connaissance et organisation: l'activité collective, 220-29.

Lorino, Philippe, et Régine Teulier. 2005. Entre connaissance et organisation : l'activité collective. Recherches. La Découverte.

Lorino, Philippe, et Benoît Tricard. 2012. « The Bakhtinian theory of chronotope (time-space frame) applied to the organizing process ». Perspectives on process organization studies 2:201-34.

Lorino, Philippe, Benoît Tricard, et Yves Clot. 2011. « Research methods for non-representational approaches to organizational complexity: The dialogical mediated inquiry ». Organization studies 32 (6): 769-801.

Loukissas, Yanni A. 2019. All data are local: thinking critically in a data-driven society. Cambridge, Massachusetts: The MIT Press.

Lowe, Andrew, Anthony C. Norris, A. Jane Farris, et Duncan R. Babbage. 2018. « Quantifying Thematic Saturation in Qualitative Data Analysis ». *Field Methods* 30 (3): 191-207.

Luo, Xueming, Marco Shaojun Qin, Zheng Fang, et Zhe Qu. 2021. « Artificial Intelligence Coaches for Sales Agents: Caveats and Solutions ». *Journal of Marketing* 85 (2): 14-32.

Mackenzie, Adrian. 2007. « Protocols and the irreducible traces of embodiment: The Viterbi algorithm and the mosaic of machine time ». R. Hassan & RE Purser (Eds.) 24 (7): 89-106.

MacKenzie, Donald A. 2014. « A sociology of algorithms: High-frequency trading and the shaping of markets ». Preprint. School of Social and Political Science, University of Edinburgh.

———. 2019. « How Algorithms Interact: Goffman's 'Interaction Order' in Automated Trading ». *Theory, Culture & Society* 36 (2): 39-59.

MacKenzie, Donald A., et Judy Wajcman, éd. 1999. *The social shaping of technology*. 2nd ed. Buckingham [Eng.] ; Philadelphia: Open University Press.

Maedche, Alexander, Christine Legner, Alexander Benlian, Benedikt Berger, Henner Gimpel, Thomas Hess, Oliver Hinz, Stefan Morana, et Matthias Söllner. 2019. « AI-Based Digital Assistants: Opportunities, Threats, and Research Perspectives ». *Business & Information Systems Engineering* 61 (4): 535-44.

Maguire, Steve, Cynthia Hardy, et Thomas B. Lawrence. 2004. « Institutional entrepreneurship in emerging fields: HIV/AIDS treatment advocacy in Canada ». *Academy of management journal* 47 (5): 657-79.

Mahdi, S.S., G. Battineni, M. Khawaja, R. Allana, M.K. Siddiqui, et D. Agha. 2023. « How does artificial intelligence impact digital healthcare initiatives? A review of AI applications in dental healthcare ». *International Journal of Information Management Data Insights* 3 (1).

Malik, Ashish, Pawan Budhwar, Hrishi Mohan, et Srikanth N. R. 2023. « Employee Experience –the Missing Link for Engaging Employees: Insights from an MNE's AI-Based HR Ecosystem ». *Human Resource Management* 62 (1): 97-115.

Malik, Ashish, Pawan Budhwar, Charmi Patel, et N. R. Srikanth. 2022. « May the Bots Be with You! Delivering HR Cost-Effectiveness and Individualised Employee Experiences in an MNE ». *The International Journal of Human Resource Management* 33 (6): 1148-78.

Malik, Ashish, Tuyet-Mai Nguyen, et Pawan Budhwar. 2024. « Towards a Conceptual Model of AI-Mediated Knowledge Sharing Exchange of HRM Practices: Antecedents and Consequences ». *IEEE Transactions on Engineering Management*, 1-30.

Mannion, Russell, et Mark Exworthy. 2017. « (Re) Making the Procrustean Bed? Standardization and Customization as Competing Logics in Healthcare ». *International Journal of Health Policy and Management* 6 (6): 301-4.

March, James G. 1991. « Exploration and Exploitation in Organizational Learning ». *Organization Science* 2 (1): 71-87.

Marchand, Donald A., et Joe Peppard. 2013. « Why IT Fumbles Analytics ». *Harvard Business Review*, janvier 2013.

Markoff, John. 2016. *Machines of Loving Grace: The Quest for Common Ground between Humans and Robots*. First Ecco paperback edition. New York: Ecco.

Martin, Roger L. 2009. *The design of business: why design thinking is the next competitive advantage*. Boston, Mass: Harvard Business Press.

Martinelli, Arianna, Andrea Mina, et Massimo Moggi. 2021. « The Enabling Technologies of Industry 4.0: Examining the Seeds of the Fourth Industrial Revolution ». *Industrial and Corporate Change* 30 (1): 161-88.

Mayo, Ray Cody, Daniel Kent, Lauren Chang Sen, Megha Kapoor, Jessica W. T. Leung, et Alyssa T. Watanabe. 2019. « Reduction of False-Positive Markings on Mammograms: A Retrospective Comparison Study Using an Artificial Intelligence-Based CAD ». *Journal of Digital Imaging* 32 (4): 618-24.

McCarthy, John, Marvin L. Minsky, Nathaniel Rochester, et Claude E. Shannon. 2006. « A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955 ». *AI Magazine* 27 (4): 12-12.

McCorduck, Pamela. 2004. *Machines who think: a personal inquiry into the history and prospects of artificial intelligence. 25th anniversary update*. Natick, Mass: A.K. Peters.

McCulloch, Warren S., et Walter Pitts. 1943. « A Logical Calculus of the Ideas Immanent in Nervous Activity ». *The Bulletin of Mathematical Biophysics* 5 (4): 115-33.

McDonald, Seonaidh. 2005. « Studying actions in context: a qualitative shadowing method for organizational research ». *Qualitative research* 5 (4): 455-73.

McKinney, Scott Mayer, Marcin Sieniek, Varun Godbole, Jonathan Godwin, Natasha Antropova, Hutan Ashrafian, Trevor Back, et al. 2020. « International Evaluation of an AI System for Breast Cancer Screening ». *Nature* 577 (7788): 89-94

Meske, Christian, Enrico Bunde, Johannes Schneider, et Martin Gersch. 2022. « Explainable Artificial Intelligence: Objectives, Stakeholders, and Future Research Opportunities ». *Information Systems Management* 39 (1): 53-63.

Meyer, John W., et Brian Rowan. 1977. « Institutionalized Organizations: Formal Structure as Myth and Ceremony ». *American Journal of Sociology* 83 (2): 340-63.

Midler, Christophe. 1993. *L'auto qui n'existait pas: management des projets et transformation de l'entreprise*. Paris: Dunod.

Miller, D. Douglas, et Eric W. Brown. 2018. « Artificial Intelligence in Medical Practice: The Question to the Answer? ». *The American Journal of Medicine* 131 (2): 129-33.

Minsky, Marvin, et Seymour A. Papert. 1969. *Perceptrons: An Introduction to Computational Geometry*. 2. print. with corr. Cambridge/Mass.: The MIT Press.

Mintzberg, Henry. 1970. « Structured observation as a method to study managerial work ». *Journal of management studies* 7 (1): 87-104.

———. 1993. *Structure et dynamique des organisations*. Traduit par Pierre Romelaer. Paris, France: Les éditions d'organisation.

Minvielle, Étienne. 2018. *Le patient et le système. En quête d'une organisation sur mesure. Approches innovantes du parcours de santé*. Paris: Seli Arslan.

Minvielle, Étienne, Mathias Waelli, Claude Sicotte, et John R. Kimberly. 2014. « Managing Customization in Health Care: A Framework Derived from the Services Sector Literature ». *Health Policy* 117 (2): 216-27.

Mirbabaie, Milad, Stefan Stieglitz, Felix Brünker, Lennart Hofeditz, Björn Ross, et Nicholas R. J. Frick. 2021. « Understanding Collaboration with Virtual Assistants – The Role of Social Identity and the Extended Self ». *Business & Information Systems Engineering* 63 (1): 21-37.

Mittal, Nitin, P. Lowes, R. Ronanki, J. Wen, et S. Sharma. 2017. « Machine intelligence: Technology mimics human cognition to create value ». *Tech Trends*, 14.

Möhlmann, Mareike, et Ola Henfridsson. 2019. « What People Hate About Being Managed by Algorithms, According to a Study of Uber Drivers ». *Harvard Business Review*, 30 août 2019. <https://hbr.org/2019/08/what-people-hate-about-being-managed-by-algorithms-according-to-a-study-of-uber-drivers>.

Möhlmann, Mareike, Lior Zalmanson, Ola Henfridsson, et Robert Wayne Gregory. 2021. « Algorithmic Management of Work on Online Labor Platforms: When Matching Meets Control ». *MIS Quarterly* 45 (4): 1999-2022.

Möhlmann, Mareike, Bentley University, Carolina Alves De Lima Salge, University of Georgia, Marco Marabelli, et Bentley University. 2023. « Algorithm Sensemaking: How Platform Workers Make Sense of Algorithmic Management ». *Journal of the Association for Information Systems* 24 (1): 35-64.

Moisdon, Jean-Claude. 2005a. « 12. Comment apprend-on par les outils de gestion ? Retour sur une doctrine d'usage ». In *Entre connaissance et organisation : l'activité collective*, 239-50. La Découverte.

———. 2005b. « Sur la largeur des mailles du filet. Savoirs incomplets et gouvernement des organisations ». In *Gouvernement, Organisation et Gestion. L'héritage de Michel Foucault*. Montréal: Presses de l'Université Laval.

———. 2007. « De la difficulté de prévoir la trajectoire d'un outil de gestion. » *Revue Sciences de Gestion*, no 64.

———. 2012. « Le paradoxe de la boîte noire. Réformes hospitalières et organisation ». *Droit et société* n°80 (1): 91.

———. 2017. « PMSI, T2A et gestion interne ». *Journal de gestion et d'économie médicales* Vol. 35 (2): 51-66.

Morley, J., C.C.V. Machado, C. Burr, J. Cows, I. Joshi, M. Taddeo, et L. Floridi. 2020. « The Ethics of AI in Health Care: A Mapping Review ». *Social Science and Medicine* 260.

Mueller, Benjamin, Uta Renken, et Gijs Van Den Heuvel. 2016. « Get Your Act Together: An Alternative Approach to Understanding the Impact of Technology on Individual and Organizational Behavior ». *ACM SIGMIS Database: The DATABASE for Advances in Information Systems* 47 (4): 67-83.

Muniesa, Fabian. 2007. « Market Technologies and the Pragmatics of Prices ». *Economy and Society* 36 (3): 377-95.

———. 2014. *The Provoked Economy: Economic Reality and the Performative Turn*. 1re éd. Routledge.

———. 2015. « Actor-Network Theory ». In *International Encyclopedia of the Social & Behavioral Sciences*, 80-84. Elsevier.

Muñoz, Katrina A., Kristin Kostick, Laura Torgerson, Peter Zuk, Lavina Kalwani, Clarissa Sanchez, Jennifer Blumenthal-Barby, Eric A. Storch, et Gabriel Lázaro-Muñoz. 2021. « Pressing Ethical Issues in Considering Pediatric Deep Brain Stimulation for Obsessive-Compulsive Disorder ». *Brain Stimulation* 14 (6): 1566-72.

Murray, Alex, Jen Rhymer, et David G. Sirmon. 2021. « Humans and Technology: Forms of Conjoined Agency in Organizations ». *Academy of Management Review* 46 (3): 552-71.

Musiani, Francesca. 2013. « Governance by Algorithms ». *Internet Policy Review* 2 (3).

Mutch, Alistair. 2013. « Sociomateriality—Taking the wrong turning? ». *Information and organization* 23 (1): 28-40.

Muzio, Daniel, David M. Brock, et Roy Suddaby. 2013. « Professions and Institutional Change: Towards an Institutional Sociology of the Professions ». *Journal of Management Studies* 50 (5): 699-721.

Newell, Allen, et Herbert A. Simon. 1972. *Human Problem Solving*. Englewood Cliffs, N.J.: Prentice-Hall.

Nguyen, Tuyet-Mai, et Ashish Malik. 2022. « Impact of Knowledge Sharing on Employees' Service Quality: The Moderating Role of Artificial Intelligence ». *International Marketing Review* 39 (3): 482-508.

Nikolaidis, Stefanos, et Julie Shah. 2013. « Human-robot cross-training: Computational formulation, modeling and evaluation of a human team training strategy ». In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 33-40. Tokyo, Japan: IEEE.

Norman, Don. 2014. *Things That Make Us Smart: Defending Human Attributes in the Age of the Machine*. New York: Diversion Books.

Norman, Donald A. 1991. « Cognitive artifacts ». *Designing interaction: Psychology at the human-computer interface* 1 (1): 17-38.

Nunn, Jeremy. 2018. « How AI is transforming HR departments ». *Forbes*, 2018. <https://www.forbes.com/sites/forbestechcouncil/2018/05/09/how-ai-is-transforming-hr-departments/>

Obermeyer, Ziad, et Ezekiel J. Emanuel. 2016. « Predicting the Future — Big Data, Machine Learning, and Clinical Medicine ». *New England Journal of Medicine* 375 (13): 1216-19.

Obermeyer, Ziad, Brian Powers, Christine Vogeli, et Sendhil Mullainathan. 2019. « Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations ». *Science* 366 (6464): 447-53.

Ogilvie, Madeleine, et Katherine Mizerski. 2011. « Using Semiotics in Consumer Research to Understand Everyday Phenomena ». *International Journal of Market Research* 53 (5): 651-68.

Okhuysen, Gerardo A., et Beth A. Bechky. 2009. « 10 Coordination in Organizations: An Integrative Perspective ». *Academy of Management Annals* 3 (1): 463-502.

Ologeanu-Taddei, Roxana, Karine Gauche, David Morquin, et Rodolphe Bourret. 2015. « La capacité d'appropriation, une capacité organisationnelle immatérielle négligée dans l'adoption des systèmes d'information et de gestion ». *Innovations* 47 (2): 79-100.

Olssen, Mark. 2004. « Neoliberalism, Globalisation, Democracy: Challenges for Education ». *Globalisation, Societies and Education* 2 (2): 231-75.

Olthof, Allard W., Peter M.A. Van Ooijen, et Mohammad H. Rezazade Mehrizi. 2020. « Promises of Artificial Intelligence in Neuroradiology: A Systematic Technographic Review ». *Neuroradiology* 62 (10): 1265-78.

O'Neil, Cathy. 2016. *Weapons of math destruction: how big data increases inequality and threatens democracy*. First edition. New York: Crown.

Orlikowski, Wanda J. 2000. « Using technology and constituting structures: A practice lens for studying technology in organizations ». *Organization science* 11 (4): 404-28.

———. 2007. « Sociomaterial practices: Exploring technology at work ». *Organization studies* 28 (9): 1435-48.

———. 2010. « The sociomateriality of organisational life: considering technology in management research ». *Cambridge Journal of Economics* 34 (1): 125-41.

Orlikowski, Wanda J., et Daniel Robey. 1991. « Information Technology and the Structuring of Organizations ». *Information Systems Research* 2 (2): 143-69.

Orlikowski, Wanda J., et Susan V. Scott. 2008. « 10 Sociomateriality: Challenging the Separation of Technology, Work and Organization ». *Academy of Management Annals* 2 (1): 433-74.

Orr, Julian E. 1996. *Talking about machines: an ethnography of a modern job*. Collection on technology and work. Ithaca, N.Y: ILR Press.

Oshima Lee, Emily, et Ezekiel J. Emanuel. 2013. « Shared Decision Making to Improve Care and Reduce Costs ». *New England Journal of Medicine* 368 (1): 6-8.

Pachidi, Stella, Hans Berends, Samer Faraj, et Marleen Huysman. 2021. « Make Way for the Algorithms: Symbolic Actions and Change in a Regime of Knowing ». *Organization Science* 32 (1): 18-41.

Pakarinen, Pauli, et Ruthanne Huising. 2023. « Relational Expertise: What Machines Can't Know ». *Journal of Management Studies*, mars.

Pakdemirli, Emre. 2019. « Artificial Intelligence in Radiology: Friend or Foe? Where Are We Now and Where Are We Heading? ». *Acta Radiologica Open* 8 (2).

Pantanowitz, Liron, Uno Wu, Lindsey Seigh, Edmund LoPresti, Fang-Cheng Yeh, Payal Salgia, Pamela Michelow, et al. 2021. « Artificial Intelligence–Based Screening for Mycobacteria in Whole-Slide Images of Tissue Samples ». *American Journal of Clinical Pathology* 156 (1): 117-28.

Paranjape, Ketan, Michiel Schinkel, Rishi Nannan Panday, Josip Car, et Prabath Nanayakkara. 2019. « Introducing Artificial Intelligence Training in Medical Education ». *JMIR Medical Education* 5 (2).

Parker, Charlie, Sam Scott, et Alistair Geddes. 2019. « Snowball sampling ». *SAGE research methods foundations*.

Paschen, Jeannette, Matthew Wilson, et João J. Ferreira. 2020. « Collaborative Intelligence: How Human and Artificial Intelligence Create Value along the B2B Sales Funnel ». *Business Horizons* 63 (3): 403-14.

Pasquale, Frank. 2015. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press.

Pearl, Judea, et Dana Mackenzie. 2018. *The book of why: the new science of cause and effect*. New York: Basic Books.

Peirce, Charles S. 1868. « Some Consequences of Four Incapacities ». *Journal of Speculative Philosophy* 2:140-57.

Peirce, Charles Sanders. 1931. *The Collected Papers of Charles Sanders Peirce*. C. Hartshorne and P. Weiss (vols. 1–6) and A. Burks (vols. 7–8). Cambridge, MA: Harvard University Press.

Peng, Jialin, et Ye Wang. 2021. « Medical Image Segmentation With Limited Supervision: A Review of Deep Network Models ». *IEEE Access* 9:36827-51.

Perez-Torrents, Joël. 2022. « Pourquoi la révolution de l'IA médicale n'aura-t-elle peut-être jamais lieu ? » *polytechnique-insights.com*, 15 novembre 2022. <https://www.polytechnique-insights.com/tribunes/digital/pourquoi-la-revolution-de-lia-medicale-naura-t-elle-peut-etre-jamais-lieu/>.

Perez-Torrents, Joël, et Étienne Minvielle. 2023. « Comment les usages des outils d'Intelligence Artificielle augmentent-ils la capacité d'agir des radiologues ? » *Journal de gestion et d'économie de la santé* 3 (3): 185-203.

Perrow, Charles. 1967. « A framework for the comparative analysis of organizations ». *American sociological review*, 194-208.

Pfaff, H, E Driller, N Ernstmann, U Karbach, C Kowalski, F Scheibler, et O Ommen. 2010. « Standardization and Individualization in Care for the Elderly: Proactive Behavior Through Individualized Standardization ». *Open Longevity Science*, no 4, 51-57.

Phillips, Christopher J. 2020. « Precision Medicine and Its Imprecise History ». *Harvard Data Science Review*, janvier.

Pickering, Andrew. 1995. *The Mangle of Practice: Time, Agency, and Science*. University of Chicago Press.

Pine, B. Joseph. 1993. « Mass Customizing Products and Services ». *Planning Review* 21 (4): 6-55.

Polanyi, Michael. 1997. *Personal knowledge: towards a post-critical philosophy*. London: Routledge.

———. 2009. *The Tacit Dimension*. Édité par Amartya Sen. Chicago, IL: University of Chicago Press.

Poplin, Ryan, Avinash V. Varadarajan, Katy Blumer, Yun Liu, Michael V. McConnell, Greg S. Corrado, Lily Peng, et Dale R. Webster. 2018. « Prediction of Cardiovascular Risk Factors from Retinal Fundus Photographs via Deep Learning ». *Nature Biomedical Engineering* 2 (3): 158-64.

Preuer, Kristina, Günter Klambauer, Friedrich Rippmann, Sepp Hochreiter, et Thomas Unterthiner. 2019. « Interpretable Deep Learning in Drug Discovery ». In *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, édité par Wojciech Samek, Grégoire

Montavon, Andrea Vedaldi, Lars Kai Hansen, et Klaus-Robert Müller, 11700:331-45. Lecture Notes in Computer Science. Cham: Springer International Publishing.

Price, W. 2019. « Medical AI and contextual bias ». *Harv. JL & Tech.* 33:65.

Price, W. Nicholson, Sara Gerke, et I. Glenn Cohen. 2019. « Potential Liability for Physicians Using Artificial Intelligence ». *JAMA* 322 (18): 1765.

Prikshat, Verma, Mohammad Islam, Parth Patel, Ashish Malik, Pawan Budhwar, et Suraksha Gupta. 2023. « AI-Augmented HRM: Literature Review and a Proposed Multilevel Framework for Future Research ». *Technological Forecasting and Social Change* 193 (août):122645.

Puranam, Phanish. 2021. « Human–AI Collaborative Decision-Making as an Organization Design Problem ». *Journal of Organization Design* 10 (2): 75-80.

Qiu, Lingyun, et Izak Benbasat. 2009. « Evaluating Anthropomorphic Product Recommendation Agents: A Social Relationship Perspective to Designing Information Systems ». *Journal of Management Information Systems* 25 (4): 145-82.

Rabardel, Pierre. 1995. *Les hommes et les technologies; approche cognitive des instruments contemporains*. Armand colin.

———. 2005. « 13. Instrument, activité et développement du pouvoir d’agir ». In *Entre connaissance et organisation: l’activité collective*, 251-65. Recherches. La Découverte.

Raghavan, Manish, Solon Barocas, Jon Kleinberg, et Karen Levy. 2020. « Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices ». In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 469-81. Barcelona Spain: ACM.

Rai, Arun, Panos Constantinides, et Saonee Sarker. 2019. « Next generation digital platforms: toward human-AI hybrids ». *MIS Quarterly* 43 (1): iii-ix.

Raisch, Sebastian, et Sebastian Krakowski. 2021. « Artificial Intelligence and Management: The Automation-Augmentation Paradox ». *Academy of Management Review* 46 (1): 192-210.

Rancour, Pastrice. 1996. « Introducing Medical Students to Psycho-Oncology ». *Psycho-Oncology* 5 (1): 55-57.

Raya-Povedano, José Luis, Sara Romero-Martín, Esperanza Elías-Cabot, Albert Gubern-Mérida, Alejandro Rodríguez-Ruiz, et Marina Álvarez-Benito. 2021. « AI-Based Strategies to Reduce Workload in Breast Cancer Screening with Mammography and Tomosynthesis: A Retrospective Evaluation ». *Radiology* 300 (1): 57-65.

Razorthink Inc. 2019. « 4 Major Challenges Facing Fraud Detection; Ways to Resolve Them Using Machine Learning ». *Razorthink AI (blog)*. 25 avril 2019. <https://medium.com/razorthink-ai/4-major-challenges-facing-fraud-detection-ways-to-resolve-them-using-machine-learning-cf6ed1b176dd>.

Ren, Minglun, Nengying Chen, et Hui Qiu. 2023. « Human-Machine Collaborative Decision-Making: An Evolutionary Roadmap Based on Cognitive Intelligence ». *International Journal of Social Robotics* 15 (7): 1101-14.

Roberts, Sarah T. 2019. *Behind the Screen*. Yale University Press.

Rodríguez-Ruiz, Alejandro, Elizabeth Krupinski, Jan-Jurre Mordang, Kathy Schilling, Sylvia H. Heywang-Köbrunner, Ioannis Sechopoulos, et Ritse M. Mann. 2019. « Detection of Breast

Cancer with Mammography: Effect of an Artificial Intelligence Support System ». *Radiology* 290 (2): 305-14.

Romanow, Darryl, Arun Rai, et Mark Keil. 2018. « CPOE-Enabled Coordination: Appropriation for Deep Structure Use and Impacts on Patient Outcomes ». *MIS Quarterly* 42 (1): 189-A11.

Rosenblat, Alex, et Luke Stark. 2016. « Uber's Drivers: Information Asymmetries and Control in Dynamic Work ». *SSRN Electronic Journal*.

Rouse, William B., et James C. Spohrer. 2018. « Automating versus Augmenting Intelligence ». *Journal of Enterprise Transformation* 8 (1-2): 1-21.

Rudin, Cynthia. 2019. « Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead ». *Nature Machine Intelligence* 1 (5): 206-15.

Rundle, Chandler W., Parker Hollingsworth, et Robert P. Dellavalle. 2021. « Artificial Intelligence in Dermatology ». *Clinics in Dermatology* 39 (4): 657-66.

Russell, Stuart J., et Peter Norvig. 2021. *Artificial intelligence: a modern approach*. Fourth edition. Pearson series in artificial intelligence. Hoboken: Pearson.

Rye, Colleen Beecken, et John R. Kimberly. 2007. « The Adoption of Innovations by Provider Organizations in Health Care ». *Medical Care Research and Review* 64 (3): 235-78.

Sachon, M., et I. Boquet. 2017. « KUKA: Planning for the future of automation, IESE Business School Case ». Barcelona, Spain: Universidad de Navarra.

Sadigh, Dorsa, Nick Landolfi, Shankar S. Sastry, Sanjit A. Seshia, et Anca D. Dragan. 2018. « Planning for Cars That Coordinate with People: Leveraging Effects on Human Actions for Planning and Active Information Gathering over Human Internal State ». *Autonomous Robots* 42 (7): 1405-26.

Salehi, Hadi, et Rigoberto Burgueño. 2018. « Emerging Artificial Intelligence Methods in Structural Engineering ». *Engineering Structures* 171 (septembre):170-89.

Samulowitz, Anke, Ida Gremyr, Erik Eriksson, et Gunnel Hensing. 2018. « "Brave Men" and "Emotional Women": A Theory-Guided Literature Review on Gender Bias in Health Care and Gendered Norms towards Patients with Chronic Pain ». *Pain Research and Management* 2018:1-14.

Saussois, Jean-Michel. 2019. *Théories des organisations*. La Découverte.

Schneider, Petra, W. Patrick Walters, Alleyn T. Plowright, Norman Sieroka, Jennifer Listgarten, Robert A. Goodnow, Jasmin Fisher, et al. 2020. « Rethinking Drug Design in the Artificial Intelligence Era ». *Nature Reviews Drug Discovery* 19 (5): 353-64.

Schuetz, Sebastian, et Viswanath Venkatesh. 2020. « The Rise of Human Machines: How Cognitive Computing Systems Challenge Assumptions of User-System Interaction ». *SSRN Scholarly Paper*. Rochester, NY.

Schultze, Ulrike, et Wanda J. Orlikowski. 2004. « A Practice Perspective on Technology-Mediated Network Relations: The Use of Internet-Based Self-Serve Technologies ». *Information Systems Research* 15 (1): 87-106.

Scott, Susan V., et Wanda J. Orlikowski. 2013. « Sociomateriality — Taking the Wrong Turning? A Response to Mutch ». *Information and Organization* 23 (2): 77-80. h

———. 2014. « Entanglements in Practice: Performing Anonymity Through Social Media ». *MIS Quarterly* 38 (3): 873-93.

Seaver, Nick. 2017. « Algorithms as Culture: Some Tactics for the Ethnography of Algorithmic Systems ». *Big Data & Society* 4 (2).

Seeber, Isabella, Eva Bittner, Robert O. Briggs, Triparna De Vreede, Gert-Jan De Vreede, Aaron Elkins, Ronald Maier, et al. 2020. « Machines as Teammates: A Research Agenda on AI in Team Collaboration ». *Information & Management* 57 (2): 103174.

Seidel, Stefan, Nicholas Berente, Aron Lindberg, Kalle Lyytinen, et Jeffrey V. Nickerson. 2018. « Autonomous Tools and Design: A Triple-Loop Approach to Human-Machine Learning ». *Communications of the ACM* 62 (1): 50-57.

Seyyed-Kalantari, Laleh, Guanxiong Liu, Matthew McDermott, Irene Y. Chen, et Marzyeh Ghassemi. 2020. « CheXclusion: Fairness Gaps in Deep Chest X-Ray Classifiers ». In *Biocomputing 2021*, 232-43. Kohala Coast, Hawaii, USA: WORLD SCIENTIFIC.

Shestakofsky, Benjamin. 2017. « Working Algorithms: Software Automation and the Future of Work ». *Work and Occupations* 44 (4): 376-423.

Sheth, Deepa, et Maryellen L. Giger. 2020. « Artificial Intelligence in the Interpretation of Breast Cancer on MRI ». *Journal of Magnetic Resonance Imaging* 51 (5): 1310-24.

Shields, Patricia M. 2003. « The Community of Inquiry: Classical Pragmatism and Public Administration ». *Administration & Society* 35 (5): 510-38.

Shipman, Marten Dorrington. 2014. *The limitations of social research*. Routledge.

Siemon, Dominik. 2022. « Elaborating Team Roles for Artificial Intelligence-Based Teammates in Human-AI Collaboration ». *Group Decision and Negotiation* 31 (5): 871-912.

Simon, Herbert A. 1947. *Administrative behavior*. United States: Macmillan.

———. 1997. *Models of Bounded Rationality: Empirically Grounded Economic Reason*. MIT Press.

Sit, Cherry, Rohit Srinivasan, Ashik Amlani, Keerthini Muthuswamy, Aishah Azam, Leo Monzon, et Daniel Stephen Poon. 2020. « Attitudes and Perceptions of UK Medical Students towards Artificial Intelligence and Radiology: A Multicentre Survey ». *Insights into Imaging* 11 (1): 14.

Skagestad, Peter. 1996. « The mind's machines: The Turing machine, the Memex, and the personal computer ». *Semiotica* 111 (3-4): 217-43.

———. 2004. « 10 Peirce's Semeiotic Model of the Mind ». In *The Cambridge Companion to Peirce*, édité par C. J. Misak, 241. Cambridge University Press.

Snow, Charles C., et James B. Thomas. 1994. « Field research methods in strategic management: contributions to theory building and testing* ». *Journal of Management Studies* 31 (4): 457-80.

Snyder, Neil, et William F. Glueck. 2019. « How managers plan—the analysis of managers' activities ». In *Managerial Work*, 227-33. Routledge.

Sowa, John. 1984. *Conceptual Structures: Information Processing in Mind and Machine The Systems Programming Series*.

Spanaki, K., U. Sivarajah, M. Fakhimi, S. Despoudi, et Z. Irani. 2022. « Disruptive technologies in agricultural operations: a systematic review of AI-driven AgriTech research ». *Annals of Operations Research* 308 (1-2): 491-524.

Spohrer, Jim. 2016. « Innovation for jobs with cognitive assistants: a service science perspective ». *Disrupting*, 132.

Spohrer, Jim, et Guruduth Banavar. 2015. « Cognition as a Service: An Industry Perspective ». *AI Magazine* 36 (4): 71-86.

Spohrer, Jim, Alessio Giuiusa, Haluk Demirkan, et David Ing. 2013. « Service Science: Reframing Progress with Universities ». *Systems Research and Behavioral Science* 30 (5): 561-69.

Stanley, N., J. Manthorpe, G. Bradley, A. Alaszewski, J. Cheetham, et M. Kazi. 1998. « Researching community care assessments: a pluralistic approach ». *The Working of Social Work*, edited by J. Cheetham and M. Kazi,—. London: Jessica Kingsley.

Star, Susan Leigh, et James R. Griesemer. 1989. « Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39 ». *Social Studies of Science* 19 (3): 387-420.

Steiner, Christopher. 2012. *Automate this: how algorithms came to rule our world*. New York: Portfolio/Penguin.

Steiner, Pierre. 2013. « C.S. Peirce and Artificial Intelligence: Historical Heritage and (New) Theoretical Stakes ». In *Philosophy and Theory of Artificial Intelligence*, édité par Vincent C. Müller, 5:265-76. *Studies in Applied Philosophy, Epistemology and Rational Ethics*. Berlin, Heidelberg.

Stoeckle, John D. 1988. « Reflections on Modern Doctoring: [Introduction] ». *The Milbank Quarterly* 66:76-91.

Stohl, Cynthia, Michael Stohl, et Paul M. Leonardi. 2016. « Digital Age | Managing Opacity: Information Visibility and the Paradox of Transparency in the Digital Age ». *International Journal of Communication* 10 (0): 15.

Stowers, Kimberly, Lisa L. Brady, Christopher MacLellan, Ryan Wohleber, et Eduardo Salas. 2021. « Improving Teamwork Competencies in Human-Machine Teams: Perspectives From Team Science ». *Frontiers in Psychology* 12 (mai):590290.

Sturm, Timo, Jin P. Gerlach, Luisa Pumplun, Neda Mesbah, Felix Peters, Christoph Tauchert, Ning Nan, et Peter Buxmann. 2021. « Coordinating Human and Machine Learning for Effective Organization Learning ». *MIS Quarterly* 45 (3): 1581-1602.

Subbaswamy, Adarsh, et Suchi Saria. 2019. « From Development to Deployment: Dataset Shift, Causality, and Shift-Stable Models in Health AI ». *Biostatistics*, novembre, kxz041.

Suchman, Lucy A. 1987. *Plans and situated actions: The problem of human-machine communication*. Cambridge university press.

———. 2007. *Human-machine reconfigurations: plans and situated actions*. 2nd ed. Cambridge ; New York: Cambridge University Press.

Sukums, F., D. Mzurikwao, D. Sabas, R. Chaula, J. Mbuke, T. Kabika, J. Kaswija, et al. 2023. « The use of artificial intelligence-based innovations in the health sector in Tanzania: A scoping review ». *Health Policy and Technology* 12 (1). <https://doi.org/10.1016/j.hlpt.2023.100728>.

Sutton, Robert I., et Barry M. Staw. 1995. « What Theory is Not ». *Administrative Science Quarterly* 40 (3): 371-84.

Sutton, Steve G., Matthew Holt, et Vicky Arnold. 2016. « “The Reports of My Death Are Greatly Exaggerated”—Artificial Intelligence Research in Accounting ». *International Journal of Accounting Information Systems* 22 (septembre):60-73.

Talley, Edmund M, David Newman, David Mimno, Bruce W Herr, Hanna M Wallach, Gully A P C Burns, A G Miriam Leenders, et Andrew McCallum. 2011. « Database of NIH Grants Using Machine-Learned Categories and Graphical Clustering ». *Nature Methods* 8 (6): 443-44.

Tang, R., L. De Donato, N. Bešinović, F. Flammini, R.M.P. Goverde, Z. Lin, R. Liu, T. Tang, V. Vittorini, et Z. Wang. 2022. « A literature review of Artificial Intelligence applications in railway systems ». *Transportation Research Part C: Emerging Technologies* 140.

Taskin, Laurent, et Gabriel Van Bunnem. 2015. « Knowledge Management through the Development of Knowledge Repositories: Towards Work Degradation ». *New Technology, Work and Employment* 30 (2): 158-72.

Taylor, James R., Carole Groleau, Lorna Heaton, et Elizabeth Van Every-Taylor. 2000. *The Computerization of Work: A Communication Perspective*. 1st edition. Thousand Oaks, Calif.: SAGE Publications, Inc.

Taylor, Matthew, Greg Marsh, Diane Nicol, et Paul Broadbent. 2017. « Good work: the Taylor review of modern working practices ». Government report or paper TD/TNC 129.269. London, England: Department for Business, Energy and Industrial Strategy.

Teodorescu, Mike HM, Lily Morse, Yazeed Awwad, et Gerald C. Kane. 2021. « Failures of Fairness in Automation Require a Deeper Understanding of Human-ML Augmentation. » *MIS quarterly* 45 (3).

Thrift, Nigel, et Shaun French. 2002. « The Automatic Production of Space ». *Transactions of the Institute of British Geographers* 27 (3): 309-35.

Tiercelin, Claudine. 1984. « Peirce on Machines, Self Control and Intentionality ». In *The Mind and the Machine: Philosophical Aspects of Artificial Intelligence*, 99-113. Chichester: Ellis Horwood Ltd. and New York: John Wiley.

Timmermans, Stefan, et Rene Almeling. 2009. « Objectification, Standardization, and Commodification in Health Care: A Conceptual Readjustment ». *Social Science & Medicine* 69 (1): 21-27.

Tong, Siliang, Nan Jia, Xueming Luo, et Zheng Fang. 2021. « The Janus Face of Artificial Intelligence Feedback: Deployment versus Disclosure Effects on Employee Performance ». *Strategic Management Journal* 42 (9): 1600-1631.

Topol, Eric J. 2019a. *Deep medicine: how artificial intelligence can make healthcare human again*. First edition. New York: Basic Books.

———. 2019b. « High-Performance Medicine: The Convergence of Human and Artificial Intelligence ». *Nature Medicine* 25 (1): 44-56.

Triberti, Stefano, Ilaria Durosini, et Gabriella Pravettoni. 2020. « A “Third Wheel” Effect in Health Decision Making Involving Artificial Entities: A Psychological Perspective ». *Frontiers in Public Health* 8 (avril):117.

Tschang, Feichin Ted, et Esteve Almirall. 2021. « Artificial Intelligence as Augmenting Automation: Implications for Employment ». *Academy of Management Perspectives* 35 (4): 642-59.

Tsoukas, Haridimos. 2009. « A Dialogical Approach to the Creation of New Knowledge in Organizations ». *Organization Science* 20 (6): 941-57.

Turing, Alan Mathison. 1936. « On computable numbers, with an application to the Entscheidungsproblem ». *J. of Math* 58 (345-363): 5.

———. 1950. « Computing Machinery and Intelligence ». *MIND, New Series*, 59 (236): 433-60.

Vaast, Emmanuelle, et Geoff Walsham. 2005. « Representations and actions: the transformation of work practices with IT use ». *Information and Organization* 15 (1): 65-89.

Valentine, Melissa, et Rebecca Hinds. 2022. « How Algorithms Change Occupational Expertise by Prompting Explicit Articulation and Testing of Experts’ Theories ». *SSRN Electronic Journal*.

Vamathevan, Jessica, Dominic Clark, Paul Czodrowski, Ian Dunham, Edgardo Ferran, George Lee, Bin Li, et al. 2019. « Applications of Machine Learning in Drug Discovery and Development ». *Nature Reviews Drug Discovery* 18 (6): 463-77.

Van Baalen, Sophie, Mieke Boon, et Petra Verhoef. 2021. « From Clinical Decision Support to Clinical Reasoning Support Systems ». *Journal of Evaluation in Clinical Practice* 27 (3): 520-28.

Van De Water, Henny, Kees Ahaus, et Robert Rozier. 2008. « Team Roles, Team Balance and Performance ». *Journal of Management Development* 27 (5): 499-512.

Van Maanen, John, Jesper B. Sørensen, et Terence R. Mitchell. 2007. « The Interplay between Theory and Method ». *Academy of Management Review* 32 (4): 1145-54.

Vaujany, François-Xavier de. 2005. *De la conception à l’usage: Vers un management de l’appropriation des outils de gestion*. Éditions EMS.

———. 2006. « Pour une théorie de l’appropriation des outils de gestion : vers un dépassement de l’opposition conception-usage: » *Management & Avenir* n° 9 (3): 109-26.

———. 2022. *Apocalypse managériale: promenade à Manhattan de 1941 à 1946 puis au-delà*. Paris: les Belles lettres.

Verbeek, Peter-Paul. 2006. « Materializing Morality: Design Ethics and Technological Mediation ». *Science, Technology, & Human Values* 31 (3): 361-80.

Vickery, Jamie, Paul Atkinson, Leesa Lin, Olivier Rubin, Ross Upshur, Eng-Kiong Yeoh, Chris Boyer, et Nicole A Errett. 2022. « Challenges to Evidence-Informed Decision-Making in the Context of Pandemics: Qualitative Study of COVID-19 Policy Advisor Perspectives ». *BMJ Global Health* 7 (4).

Villani, Cédric, Marc Schoenauer, Yann Bonnet, Charly Berthet, Anne-Charlotte Cornut, François Levin, et Bertrand Rondepierre. 2018. « Donner un sens à l’intelligence artificielle : pour une stratégie nationale et européenne ».

- Viola, Tullio. 2020. Peirce on the Uses of History. De Gruyter.
- Von Nordenflycht, Andrew. 2010. « What Is a Professional Service Firm? Toward a Theory and Taxonomy of Knowledge-Intensive Firms ». *Academy of Management Review* 35 (1): 155-74.
- Vygotski, Lev Semionovitch. 1930. « La méthode instrumentale en psychologie ». *Vygotsky aujourd'hui*, 39-47.
- Waardenburg, Lauren, Marleen Huysman, et Anastasia V. Sergeeva. 2022. « In the Land of the Blind, the One-Eyed Man Is King: Knowledge Brokerage in the Age of Learning Algorithms ». *Organization Science* 33 (1): 59-82.
- Wagner, Erica L., Jodie Moll, et Sue Newell. 2011. « Accounting Logics, Reconfiguration of ERP Systems and the Emergence of New Accounting Practices: A Sociomaterial Perspective ». *Management Accounting Research* 22 (3): 181-97.
- Wagner-Pacifici, Robin, John W Mohr, et Ronald L Breiger. 2015. « Ontologies, Methodologies, and New Uses of Big Data in the Social and Cultural Sciences ». *Big Data & Society* 2 (2).
- Walker, Charles R., Robert H. Guest, et Arthur N. Turner. 2017. *The foreman on the assembly line*. Routledge.
- Wang, Weiguang, Guodong (Gordon) Gao, et Ritu Agarwal. 2023. « Friend or Foe? Teaming Between Artificial Intelligence and Workers with Variation in Experience ». *Management Science*, octobre.
- Wang, Yi, Hai-Shu Ma, Jing-Hui Yang, et Ke-Sheng Wang. 2017. « Industry 4.0: A Way from Mass Customization to Mass Personalization Production ». *Advances in Manufacturing* 5 (4): 311-20.
- Waring, Justin, Davina Allen, Jeffrey Braithwaite, et Jane Sandall. 2016. « Healthcare Quality and Safety: A Review of Policy, Practice and Research ». *Sociology of Health & Illness* 38 (2): 198-215.
- Watson, David. 2019. « The Rhetoric and Reality of Anthropomorphism in Artificial Intelligence ». *Minds and Machines* 29 (3): 417-40.
- Webel, Henry E., Talia B. Kimber, Silke Radetzki, Martin Neuenschwander, Marc Nazaré, et Andrea Volkamer. 2020. « Revealing Cytotoxic Substructures in Molecules Using Deep Learning ». *Journal of Computer-Aided Molecular Design* 34 (7): 731-46.
- Weick, Karl E. 1974. « The Nature of Managerial Work. »
- Weick, Karl E., et Karlene H. Roberts. 1993. « Collective Mind in Organizations: Heedful Interrelating on Flight Decks ». *Administrative Science Quarterly* 38 (3): 357-81.
- Weick, Karl E., Kathleen M. Sutcliffe, et David Obstfeld. 2005. « Organizing and the Process of Sensemaking ». *Organization Science* 16 (4): 409-21.
- Weking, Jörg, Maria Stöcker, Marek Kowalkiewicz, Markus Böhm, et Helmut Krcmar. 2020. « Leveraging Industry 4.0 – A Business Model Pattern Framework ». *International Journal of Production Economics* 225 (juillet).
- Westie, Frank R. 1957. « Toward closer relations between theory and research: A procedure and an example ». *American Sociological Review* 22 (2): 149-54.

- Whitehouse, Nick. 2020. « INSIGHT: Legal AI 2.0 Is Now Appealing to the Masses ». 2020.
- Whyte, William Foote. 1984. *Learning from the field: A guide from experience*. 123. Sage.
- Wiener, Martin, W. Alec Cram, et Alexander Benlian. 2023. « Algorithmic Control and Gig Workers: A Legitimacy Perspective of Uber Drivers ». *European Journal of Information Systems* 32 (3): 485-507.
- Wiener, Norbert. 1948. *Cybernetics or Control and Communication in the Animal and the Machine*. The MIT Press.
- Wiggers, Kyle. 2020. « Google's Breast Cancer-Predicting AI Research Is Useless without Transparency, Critics Say ». *VentureBeat (blog)*. 14 octobre 2020. <https://venturebeat.com/ai/googles-breast-cancer-predicting-ai-research-is-useless-without-transparency-critics-say/>.
- Winkel, Suzanne L. van, Alejandro Rodríguez-Ruiz, Linda Appelman, Albert Gubern-Mérida, Nico Karssemeijer, Jonas Teuwen, Alexander J. T. Wanders, Ioannis Sechopoulos, et Ritse M. Mann. 2021. « Impact of Artificial Intelligence Support on Accuracy and Reading Time in Breast Tomosynthesis Image Interpretation: A Multi-Reader Multi-Case Study ». *European Radiology* 31 (11): 8682-91.
- Winner, Langdon. 2007. « Do Artifacts Have Politics? » In *Computer Ethics*. Routledge.
- Winograd, Shmuel, et Jack D. Cowan. 1963. *Reliable computation in the presence of noise*. MIT Press Cambridge, Mass.
- Wirtz, Jochen, Paul G. Patterson, Werner H. Kunz, Thorsten Gruber, Vinh Nhat Lu, Stefanie Paluch, et Antje Martins. 2018. « Brave New World: Service Robots in the Frontline ». *Journal of Service Management* 29 (5): 907-31.
- Wood, Alex J. 2018. « Powerful Times: Flexible Discipline and Schedule Gifts at Work ». *Work, Employment and Society* 32 (6): 1061-77.
- Wood, Alex J, Mark Graham, Vili Lehdonvirta, et Isis Hjorth. 2019. « Good Gig, Bad Gig: Autonomy and Algorithmic Control in the Global Gig Economy ». *Work, Employment and Society* 33 (1): 56-75.
- Woodward, Joan. 1958. *Management and technology*. 3. HM Stationery Office.
- Wu, Eric, Kevin Wu, Roxana Daneshjou, David Ouyang, Daniel E. Ho, et James Zou. 2021. « How Medical AI Devices Are Evaluated: Limitations and Recommendations from an Analysis of FDA Approvals ». *Nature Medicine* 27 (4): 582-84.
- Wu, Nan, Jason Phang, Jungkyu Park, Yiqiu Shen, Zhe Huang, Masha Zorin, Stanislaw Jastrzebski, et al. 2020. « Deep Neural Networks Improve Radiologists' Performance in Breast Cancer Screening ». *IEEE Transactions on Medical Imaging* 39 (4): 1184-94.
- Wulczyn, Ellery, Kunal Nagpal, Matthew Symonds, Melissa Moran, Markus Plass, Robert Reihs, Farah Nader, et al. 2021. « Predicting Prostate Cancer Specific-Mortality with Artificial Intelligence-Based Gleason Grading ». *Communications Medicine* 1 (1): 10.
- Yin, Robert K. 2011. *Applications of case study research*. Los Angeles London New Delhi Singapore Washington DC Melbourne: SAGE.
- . 2018. *Case Study Research and Applications: Design and Methods*. Sixth edition. Los Angeles London New Delhi Singapore Washington DC Melbourne: SAGE.

Youyou, Wu, Michal Kosinski, et David Stillwell. 2015. « Computer-Based Personality Judgments Are More Accurate than Those Made by Humans ». *Proceedings of the National Academy of Sciences* 112 (4): 1036-40.

Yu, Kun-Hsing, Andrew L. Beam, et Isaac S. Kohane. 2018. « Artificial Intelligence in Healthcare ». *Nature Biomedical Engineering* 2 (10): 719-31.

Zammuto, Raymond F., Terri L. Griffith, Ann Majchrzak, Deborah J. Dougherty, et Samer Faraj. 2007. « Information Technology and the Changing Fabric of Organization ». *Organization Science* 18 (5): 749-62.

Zanzotto, Fabio Massimo. 2019. « Viewpoint: Human-in-the-loop Artificial Intelligence ». *Journal of Artificial Intelligence Research* 64 (février):243-52. <https://doi.org/10.1613/jair.1.11345>.

Zhang, Zhewei, Youngjin Yoo, Kalle Lyytinen, et Aron Lindberg. 2021. « The Unknowability of Autonomous Tools and the Liminal Experience of Their Use ». *Information Systems Research* 32 (4): 1192-1213.

Zheng, Nan-ning, Zi-yi Liu, Peng-ju Ren, Yong-qiang Ma, Shi-tao Chen, Si-yu Yu, Jian-ru Xue, Ba-dong Chen, et Fei-yue Wang. 2017. « Hybrid-Augmented Intelligence: Collaboration and Cognition ». *Frontiers of Information Technology & Electronic Engineering* 18 (2): 153-79.

Zhou, Lina, Souren Paul, Haluk Demirkan, Lingyao (Ivy) Yuan, Jim Spohrer, Michelle Zhou, et Julie Basu. 2021. « Intelligence Augmentation: Towards Building Human- machine Symbiotic Relationship ». *AIS Transactions on Human-Computer Interaction* 13 (2): 243-64.

Ziewitz, Malte. 2016. « Governing Algorithms: Myth, Mess, and Methods ». *Science, Technology, & Human Values* 41 (1): 3-16.

Zuboff, Shoshana. 1988. *In the age of the smart machine: The future of work and power*. Basic Books, Inc.

———. 2019. *The age of surveillance capitalism: the fight for a human future at the new frontier of power*. First Trade Paperback Edition. New York: PublicAffairs.

Titre : Gérer la collaboration entre l'expert métier et l'Intelligence Artificielle. Deux études de cas dans le système de soins.

Mots clés : collaboration ; humain-IA ; Intelligence Artificielle ; expert métier ; système de soins

Résumé : L'Intelligence Artificielle (IA) mobilise des méthodes d'apprentissage machine pour automatiser la création de modèles statistiques complexes. Ces outils d'IA peuvent accomplir des tâches avec des performances similaires à celles des experts, mais la manière dont les résultats sont produits reste souvent opaque, posant des défis pour leur intégration dans les organisations. Cette thèse explore la collaboration entre les experts métier et les outils d'IA dans le contexte du système de soins.

Dans le système de soins, cette collaboration est cruciale en raison de la responsabilité professionnelle et morale des décisions médicales et de leur incertitude inhérente. Si les outils d'IA promettent de répondre aux tensions de ce système en offrant une personnalisation accrue des soins à moindre coût, ils suscitent également des craintes et leur adoption reste à concrétiser.

Notre démarche empirique produit deux études de cas illustrant ces dynamiques. RADO porte sur les usages d'un outil d'IA par des radiologues pour l'analyse mammographique, visant à améliorer leur activité. La seconde étude de cas, KOVAK, examine comment une équipe de recherche médicale utilise des outils d'IA pour analyser des données de cohortes de patients.

Nous utilisons un premier cadre d'analyse pour observer comment les experts intègrent leurs connaissances avec les résultats de l'outil d'IA, démontrant ainsi un engagement dans la collaboration. Un deuxième cadre caractérise une double nature de l'usage de l'outil d'IA, entre optimisation d'une activité et apprentissages, et montre l'augmentation des capacités des radiologues grâce à cette collaboration. Un troisième cadre, fondé sur les travaux de Peirce sur l'enquête pragmatiste, considère l'outil d'IA comme un partenaire dans la construction des connaissances.

Nous proposons un modèle de collaboration, l'EMC2 (Expert Machine Collaborative Community). Il intègre différents modes de gestion de la collaboration expert-IA facilitant ainsi une meilleure intégration de cette collaboration au sein des organisations.

Cette thèse contribue à la littérature sur les modèles de collaboration humain-IA par des modes de gestion issus d'une démarche empirique. Elle enrichit aussi la littérature sur les usages des outils d'IA en spécifiant des pratiques interrogatives et en appliquant le concept d'enquête pragmatiste.

Title : Managing the collaboration between domain experts and Artificial Intelligence: Two case studies in the healthcare system.

Keywords : collaboration ; human-AI ; Artificial Intelligence ; domain expert ; healthcare system

Abstract: Artificial Intelligence (AI) leverages machine learning methods to automate the creation of complex statistical models. These AI tools can perform tasks with expert-level proficiency, yet the precise way results are generated often remains opaque, posing challenges for their integration into organizations. This thesis explores the collaboration between domain experts and AI tools within the healthcare system.

In the healthcare system, human-AI collaboration is particularly critical due to the professional and moral responsibilities inherent in medical decisions and their inherent uncertainty. While AI tools promise to address the tensions in this system by offering increased personalization of care at lower costs, they also raise concerns, and their real-world adoption remains to be fully realized.

Our empirical approach includes two case studies, illustrating these dynamics. RADO focuses on the use of an AI tool by radiologists for mammographic analysis, aiming to enhance their activities. KOVAK,

examines how a medical research team uses AI tools to analyze patient cohort data.

A first analytical framework observes how experts leverage their knowledge to incorporate AI results, thereby demonstrating an engaged collaboration. A second one identifies the dual nature of the AI tool use, between a way to optimize an activity or generate learning. A third one, based on Peirce's work on pragmatic inquiry, considers the AI tool as a partner in the knowledge construction process. We propose our collaboration model, EMC2 (Expert Machine Collaborative Community). It integrates various modes of managing expert-AI collaboration, thus facilitating better integration of this collaboration within organizations.

This thesis contributes to the literature on human-AI collaboration models by defining management modes derived from an empirical approach. It also contributes to the literature on AI tool usage by specifying interrogative practices and by applying the concept of pragmatic inquiry.