



HAL
open science

**The diffusion of Artificial Intelligence into Neuroscience.
A multiscale approach of the genericity of a
research-technology**

Sylvain Fontaine

► **To cite this version:**

Sylvain Fontaine. The diffusion of Artificial Intelligence into Neuroscience. A multiscale approach of the genericity of a research-technology. History, Philosophy and Sociology of Sciences. Sorbonne Université, 2024. English. NNT : 2024SORUL126 . tel-04928992

HAL Id: tel-04928992

<https://theses.hal.science/tel-04928992v1>

Submitted on 4 Feb 2025

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.

Sorbonne Université
École Doctorale 433 : Concepts et Langages
Groupe d'Étude des Méthodes de l'Analyse Sociologique de la Sorbonne

THÈSE
pour obtenir le grade de
DOCTEUR DE SORBONNE UNIVERSITÉ
Discipline : Sciences sociales et philosophie de la connaissance

Présentée et soutenue publiquement par :

Sylvain Fontaine

le 6 décembre 2024

**The diffusion of Artificial Intelligence into
Neuroscience**
A multiscale approach of the genericity of a research-technology

Sous la direction de :

M. Michel DUBOIS (<i>directeur</i>)	Directeur de recherche CNRS GEMASS, Sorbonne Université
Mme Floriana GARGIULO (<i>co-encadrante</i>)	Chargée de recherche CNRS GEMASS, Sorbonne Université
Mme Paola TUBARO (<i>co-encadrante</i>)	Directrice de recherche CNRS CREST, ENSAE, IPP

Membres du jury :

M. Stefano BIANCHINI (<i>examinateur</i>)	Maître de conférence HDR BETA, Université de Strasbourg
Mme Laura HERNÁNDEZ (<i>examinatrice</i>)	Maîtresse de conférence HDR LPTM, CY Cergy Paris Université
M. Gianluca MANZO (<i>président/examinateur</i>)	Professeur des Universités GEMASS, Sorbonne Université
Mme Béatrice MILARD (<i>rapporteuse</i>)	Professeure des Universités LISST, Université Toulouse Jean Jaurès
M. Camille ROTH (<i>rapporteur</i>)	Directeur de recherche CNRS CAMS, EHESS
Mme Cassidy R. SUGIMOTO (<i>examinatrice</i>)	Professeure Georgia Institute of Technology, USA

« *Dieu est mort mais deux et deux font quatre.* »
 (“*God is dead but two plus two still makes four.*”)

Bernard Charbonneau, *Le Totalitarisme industriel* (2019)

Remerciements

Elon has apparently one genuine technical paper indexed by Google Scholar [and] published in the Journal of Medical Internet Research in 2019. The authors? "Elon Musk and Neuralink". I'm sure the scientists who hide behind this collective name are super happy about that. I just hope they won't die bitter and forgotten.

Answer of Yann LeCun
to Elon Musk on the social
network X, July, 1st 2024

Venant clôturer un parcours universitaire (quelque peu périlleux) entre physique fondamentale et études des sciences et techniques, cette thèse doit beaucoup à de nombreuses personnes qui m'ont accompagné, soutenu de près ou de loin et mis en confiance lors de mon passage à la sociologie, qu'elles appartiennent ou non au monde professionnel de la recherche. Au sortir de la rédaction de ce manuscrit, il m'est alors important de leur adresser quelques mots.

Tout d'abord, au cours de ma licence et mon master en physique fondamentale, et surtout à travers trois stages estivaux entre 2018 et 2020, j'ai expérimenté la pratique de la recherche en systèmes complexes sociaux appliquée à des sujets divers et variés et mobilisant des outils formels et computationnels relevant de la physique théorique. Je tiens alors à remercier ici, en premier lieu, les encadrantes et encadrants de ces stages marqués par une dimension interdisciplinaire sous-jacente qui m'a tout de suite plu, chronologiquement Cécile Appert-Rolland, Timoteo Carletti et Laura Hernández. Je remercie une nouvelle fois le second pour m'avoir conseillé durant l'année 2019 de contacter une « certaine » Floriana Gargiulo, spécialiste en analyse de réseaux sociaux dans un laboratoire parisien de sociologie.

En effet, ce tournant « sociologique » dans mon parcours universitaire a été initié par la rencontre de cette dernière en tout début d'année 2020, alors en master à l'Université Paris-Saclay. Floriana, un immense merci pour ton soutien sans faille, et pour m'avoir

guidé sur le terrain des sciences sociales computationnelles, de la complexité et de la scientométrie. Cette aventure sociologique n'aurait été complète sans l'encadrement régulier de Michel Dubois et Paola Tubaro. Vous m'avez tout deux convaincu de suivre la dernière année du master de sociologie contemporaine de Sorbonne Université, et surtout introduit dans un environnement disciplinaire dont j'avais tout à apprendre, sans pour autant « renier » ce que j'avais acquis de ma formation universitaire initiale. Chacun à votre manière, vous m'avez fait découvrir ces domaines passionnants que sont les études des sciences et de l'intelligence artificielle, et avez contribué à développer chez moi une appétence toujours plus forte pour les sciences sociales. Pour tout cela, je vous remercie chaleureusement.

Je remercie l'ensemble de mon jury de soutenance, voulu éminemment multidisciplinaire, pour leurs commentaires à propos de la recherche défendue dans ce manuscrit et leurs suggestions d'amélioration de ce dernier en vue de sa poursuite : Béatrice Milard et Camille Roth en tant que rapporteure et rapporteur ; Stefano Bianchini, Laura Hernández (encore merci d'avoir accepté d'avoir participé à ce jury), Gianluca Manzo et Cassidy R. Sugimoto en tant qu'examinatrices et examinateurs.

Je remercie les institutions ayant appuyé ces recherches, surtout d'un point de vue financier. D'abord le Centre National de la Recherche Scientifique (CNRS), dont la Mission pour les Initiatives Transverses et Interdisciplinaires (sous la bannière du projet *EpiAI*) m'a accordé une bourse doctorale après l'année de master susmentionnée, ensuite le projet *ScientIA* (dont le logo fait même envier les fans de Star Wars les plus assidus) financé par l'Agence Nationale de la Recherche (ANR), sans qui je n'aurais pu assister ni présenter mes travaux lors de conférences internationales reconnues, et qui a permis l'acquisition de l'ordinateur portable ayant servi à l'intégralité des analyses présentée dans ce manuscrit. Je remercie l'école doctorale Concepts et Langages de Sorbonne Université, pour son suivi lors de mon parcours doctoral et également pour son soutien financier à quelques missions ponctuelles, qu'elles soient des conférences ou des formations.

Un grand merci aux chercheurs et chercheuses de l'équipe retenue pour le terrain présenté dans cette thèse, et auprès de laquelle j'ai collecté des informations inestimables sur la recherche en intelligence artificielle appliquée à des enjeux scientifiques de taille liés à la détection précoce de maladies neurodégénératives.

*
* *

Tout en menant les recherches ayant conduit à ce présent manuscrit, mon parcours doctoral a été pour moi l'occasion de combler de sérieuses lacunes en sociologie des sciences, intelligence artificielle, sciences des données et bibliométrie, par l'apprentissage (accéléré)

de certains savoirs et méthodes de recherche propres à ces domaines. De par mon rattachement au Groupe d'Étude des Méthodes de l'Analyse Sociologique de la Sorbonne (GEMASS), j'ai ainsi bénéficié d'un environnement de recherche stimulant m'ayant permis de me construire une culture interdisciplinaire au carrefour de ces grands domaines, et ce grâce à de multiples discussions que j'ai pu avoir avec l'ensemble de ses membres qui m'ont suggéré de nombreuses lectures et transmis leur passion pour la sociologie, sans oublier les séminaires et événements ponctuels, plus ou moins formels, que j'ai pu coorganiser avec certains d'entre eux.

Le GEMASS étant implanté sur deux sites géographiquement séparés par quelques stations de métro, j'ai surtout pris mes marques dans l'un de ces derniers, l'UAR Pouchet (pour Unité d'Appui à la Recherche) – dotée par ailleurs d'un toit-terrasse prisé pour de nombreuses pauses café. Parmi les collègues de cette partie de l'unité, je tiens à remercier Jean-François pour nos deux années à animer ensemble le séminaire interne du GEMASS, et pour avoir accepté de faire partie de mon comité de suivi de thèse accompagné de Marion Maisonobe, que je remercie également au passage. Un remerciement tout particulier va également à Georgie, notre gestionnaire de laboratoire avec qui j'ai réussi à régler je ne sais combien de soucis administratifs, même dans les moments difficiles de la « simplification » et de la numérisation des services du CNRS, notamment avec l'arrivée de l'environnement *ESR Mission* courant juillet 2023 (et ses composantes désormais célèbres *Étamine* et *Notilus*). Merci à Dominique, qui m'a permis de donner cours à l'Université Mohammed VI Polytechnique (UM6P) de Rabat. Sans oublier Abdelghani, Alexandra, Catherine, Christelle et Louis-André. Merci enfin aux membres réguliers du *thésarium* pour animer le fond du couloir du GEMASS : Nicolas, Raphaël, Sandy, Victor (à la fin de sa thèse), ainsi que les chercheurs en contrat cours passés par le laboratoire dans le cadre d'un stage ou d'un post-doc, notamment Andrea, Aurélien et Emmanuel.

Je remercie nos voisins de palier du Centre Internet et Société (CIS), qui m'ont permis d'assister à de riches séminaires en sociologie du numérique, en particulier Antoine pour toutes nos péripéties dans et hors les murs de la recherche (un grand merci, chef), Fabrizio pour nos riches échanges sur les diffusions technologiques en sciences, et Tommaso pour ses nombreux coups de pouce en cartographie des connaissances. Un petit clin d'œil va à Nicolas Bresch de l'Institut de Recherche sur l'Architecture Antique (IRAA), aujourd'hui à la retraite mais qui nous a beaucoup fait voyager le temps d'une pause café sur les anciens navires de la Grèce antique. Sans oublier tout le personnel logistique de l'UAR sans qui cette structure ne pourrait tourner. Je tiens notamment à remercier Hasna, Mazyar et Rachid, qui ont réglé pas mal de soucis de connectique au thésarium, Jean-René et Thomas pour l'organisation matérielle de la soutenance, ainsi que le personnel d'Elior en charge de la cantine de l'UAR pour nous concocter de bons petits plats malgré les problèmes récurrents en cuisine, et qui garantissent un temps de pause nécessaire pour

se détacher de la recherche le temps de quelques minutes (et parce qu'il est important de bien manger pour que les cerveaux érudits continuent de chauffer !).

Le site Pouchet ayant dû fermer ses portes pour causes de travaux en pleine période olympique au cours de l'été 2024, je remercie l'Institut des Systèmes Complexes en Île-de-France pour son accueil dans un cadre studieux afin de parachever la rédaction de ce manuscrit. Je remercie tout particulièrement les réguliers collègues de bureau sur place, Franck et Jean-Yves (le deuxième également au CIS), qui ont égayé ce quotidien estival et sont parvenus (peut-être malgré eux) à me sortir épisodiquement, et pour mon plus grand bien, de cette « bulle » qu'impose l'exercice de rédaction de thèse.

Au-delà de l'UAR Pouchet, comme évoqué plus haut, le GEMASS est également implanté à la Maison de la Recherche de Sorbonne Université. Ce lieu de recherche et de cours accueille notamment le personnel enseignant du parcours universitaire de sociologie au sein de la Sorbonne, dont la plupart m'ont enseigné des pans passionnant de la discipline il y a maintenant trois à quatre ans. Merci beaucoup à Béate (qui nous a malheureusement quittés en 2023), Renaud, Cyril, Élise, Gérald, Marie, Marta, Pierre, Pierre-Marie, Sébastien et Solenne. La Maison de la Recherche accueille également presque l'intégralité des doctorantes et doctorants de l'unité (les *Serpentards* !), enseignant également au sein de la licence et du master de sociologie de la Sorbonne, et avec qui nous avons partagé de nombreux verres et petits-déjeuners. Merci à Alix, Camille, Élixa, Ganfen, Hana, les trois Hugo, Jeanne, Julie, Laurie-Anne, Lucas, Manon, Mathis, Margot, Melchior, Romain, Salomé, Simha, Sofiane et Zara. Pas nécessairement sociologues, je remercie également le Centre d'expérimentation en méthodes numériques pour les Recherches en Sciences Humaines et Sociales (CERES), unité d'appui de la Faculté des Lettres dirigée par Gaël Lejeune, pour ses invitations à quelques séminaires et les suggestions fructueuses de ses membres faites à mon travail lorsqu'il s'est agi d'entrer dans le domaine du traitement automatique du langage, dont ils sont experts. Merci notamment à Caroline et Julien pour leurs invitations à ces événements qui constituent, à mon sens, des initiatives bénéfiques au développement multidisciplinaire des humanités numériques.

Au moment d'écrire ces lignes, et bien que cela occasionne une répétition pour les concernés, je pense surtout à des compagnons de route depuis le master de sociologie devenus de grands amis : Antoine, Nicolas et Raphaël. Vous avez su me transmettre d'abord et avant tout votre amour de la sociologie ainsi que de précieux conseils pour avancer sereinement dans ma conversion de la physique à la sociologie, et pour cela je vous remercie chaleureusement. C'est à vos côtés que je place une petite dédicace à nos deux lieux de fréquentes retrouvailles, le *Mayflower* et l'*Octopussy*.

Durant mon doctorat, j'ai eu également l'opportunité de faire un court et intense séjour de recherche au Sony Computer Science Lab de Rome, partenaire du projet *ScientIA*. Je

remercie tout particulièrement Vittorio Loreto et Gabriele Di Bona pour leur accueil dans ce laboratoire et m'avoir partagé leurs points de vue pertinents sur la modélisation des processus d'innovation et de créativité, dans un contexte scientifique ou non. Thank you again, Gabriele, for having found some time to review so quickly some parts of this thesis.

*
* *

Sans être dévoué entièrement au travail, beaucoup de personnes m'ont permis de garder les pieds sur terre hors les murs de la recherche.

Je remercie tout d'abord le « club lecture » (étendu depuis un an avec sa variante ciné !) composé de Julia, Léa, Marie, Mélanie, Omar et Ulysse, pour avoir proposé des ouvrages passionnants venus s'ajouter à une bonne pile d'articles et de livres sur l'intelligence artificielle, la sociologie des sciences et j'en passe.

Je remercie très chaleureusement l'Association Pour l'Accompagnement des Jeunes et de la Famille (APAJF), ses adhérents et ses bénévoles qui la font vivre, et dans laquelle j'ai surtout participé à l'aide aux devoirs des plus petits (cinq ans déjà !). Dominique, Farook, Ibtissem, Jean-Pierre, Ramata, Reffka, Sonia et tant d'autres, merci pour tout ce que vous faites au quotidien pour la ville de Massy, la cohésion et le vivre-ensemble. Une pensée pour les enfants qui, même bruyants et peu concentrés après l'école, apportent joie, gaieté et insouciance après une rude journée de travail.

Merci à Jérémy pour nos péripéties en tout genre depuis le Magistère de physique d'Orsay, que ce soit en région parisienne, au pays basque ou ailleurs. Un grand merci également à Marion, Philippine et Inès, surtout la dernière pour tous ses encouragements jusqu'au bout !

Romain, merci à toi depuis le temps qu'on se connaît. Chacun en prise avec nos activités de recherche respectives, nos retrouvailles deviennent de plus en plus précieuses.

Je dédie enfin cette thèse à ma famille qu'il m'est coutumier de remercier à chaque fin d'année universitaire conclue par un stage ou un mémoire, pour son soutien en toutes circonstances, même pour une reconversion « hasardeuse » : mes parents Noëlle et Arnaud, mes frères Louis et Paul, Pauline qui m'accompagne depuis maintenant sept ans, sans oublier ma belle-famille, Françoise, Marc et Géraud. Vraiment merci pour tout ce que vous m'apportez au quotidien malgré les (trop) longues distances qui nous séparent.

Table of contents

Introduction	1
Conceptualization and research questions	5
Methodological positioning	8
Outline of the thesis	11
1 Artificial Intelligence: a sociotechnical, multifaceted object	13
1.1 A brief history of AI	15
1.2 The academic work of AI captured by sociology	20
1.3 How to study the diffusion of AI in science? An overview	23
1.3.1 At the macro- and meso-scale: AI in the scientific system	24
1.3.2 At the micro-scale: the social and epistemic diffusion of AI	25
1.4 An analytical framework: Conceiving AI as a <i>research-technology</i> in science	26
2 A journey in neuroscience: collection and preprocessing of quantitative and qualitative data	31
2.1 A multidisciplinary domain interacting with AI	32
2.2 Extraction of a neuroscientific corpus and construction of data structures .	34
2.2.1 Building of a first exhaustive database of neuroscience articles from the MAG	34
2.2.2 Detecting mentions of AI in selected publications	39

2.2.3	Selected data structures	40
2.2.4	Selection of a subset of the extracted neuroscience database	44
2.2.5	Final dataset and discussion	50
2.3	Qualitative data collection from a laboratory fieldwork	54
2.3.1	The Aramis team: a window on AI research for neuroscience	54
2.3.2	Choosing a semi-structured interview format	55
3	Domain adaptation and social segregation of AI research in neuroscience	59
3.1	Presentation of the mobilized framework	60
3.1.1	Method for comparing the disciplinary influence and impact of the AI-related and non-AI research over time	60
3.1.2	Computation of the intensity of collaborations in the aggregated co-authorship network	62
3.1.3	The AI-activity of the journals	64
3.2	The situation of AI within the neuroscience disciplinary landscape	64
3.2.1	Citation homogenization of the AI-related publications with general neuroscience	64
3.2.2	Beyond the citation homogenization: AI and non-AI publications lie on different disciplinary bases	67
3.3	Who are the scientists making AI in neuroscience?	72
3.3.1	Disciplinary profiles in each quartile of the f_{AI} distribution	73
3.3.2	Segregation between AI practitioners and non-AI ones	77
3.4	An AI literature confined in a small set of specialized journals	80
3.5	Discussion	82
3.6	Summary	86
4	Exploring the epistemic integration of AI in the neuroscience knowledge landscape	89
4.1	Methodology	90
4.1.1	Representing the knowledge space of neuroscience	90

4.1.2	Measuring the epistemic articulation of AI-related and general neuroscience concepts	92
4.1.3	Measuring the diffusion of AI-related work in the knowledge map	95
4.2	Where is AI situated in the knowledge landscape of neuroscience?	96
4.3	The limited integration of AI into the various concept networks of neuroscience	100
4.4	The limited citation spreading of AI-related publications in neuroscience	102
4.4.1	AI-related articles diffuse less than non-AI ones across the domain	102
4.4.2	AI remains confined to local knowledge subspaces of neuroscience	105
4.5	Discussion	106
4.6	Summary	109
5	Complementing quantitative findings: The interdisciplinary work of AI in the Aramis team	111
5.1	An “immersion” in the core of Aramis research	112
5.1.1	An international research in the context of the <i>big data</i> era	113
5.1.2	Collaborations outside academia	117
5.2	A dichotomy of scientific profiles	119
5.2.1	Computer scientists looking for concrete applications	119
5.2.2	The medical profession in need of effective computing methods	120
5.2.3	Computer scientists at the service of clinicians	123
5.2.4	Testing interdisciplinarity	127
5.3	The diffusion of AI in neuroscience research: between promise and reality	129
5.3.1	A still recent bubble	130
5.3.2	An unclear but promising concept	132
5.3.3	The need to be up-to-date to maintain visibility	133
5.3.4	A practical tool to help you stand out in clinical research	134
5.3.5	A future medical assistance tool?	135
5.4	Discussion and conclusion	138

Conclusion	141
Summary of the results	141
Limitations, perspectives of improvement and future work	145
Discussion of the chosen analytical framework to study the diffusion of AI within neuroscience	145
Refining the neuroscience corpus	146
Studying the capability of AI to create knowledge	148
Applying our framework to other fields of research	149
Appendices	151
A Chapter 2 appendices	151
A.1 AI-related keywords table	151
A.2 List of journals retained to build the neuroscience corpus	156
A.3 Web of Science categories' abbreviations table	159
B Chapter 3 appendices	161
B.1 Citation patterns of the AI research in neuroscience	161
B.2 Temporal evolution of the inter-quartile collaboration network	164
C Chapter 4 appendices	166
C.1 Technical elements for the construction of the neuroscience knowl- edge map	166
C.2 Temporal Jaccard similarity between the topical universe of the knowledge clusters and that presented by the AI-related papers within them	176
C.3 Dynamic citation radius of gyration of the papers published in 1976 and 1989	176
D Chapter 5 appendices	178
D.1 Interview grid	178
Bibliography	183
Detailed abstract	201

Introduction

Since the introduction of this terminology in the late 1950s, Artificial Intelligence (AI) has undergone a significant evolution, giving rise to a multitude of technological applications in recent times, whether originating from laboratories or industrial warehouses, deployed or prototypical: Embedded systems for automated transportation;¹ humanoid robots for nursing,² playing football,³ or just for impressive stunts performing;⁴ manufacturing machines; conversational agents;⁵ face recognition on smartphones; recommendation systems on online social networks; and so forth. While this list is far from being complete, these recent advances in AI demonstrate its growing importance in our everyday lives, by first automatizing and accelerating a variety of complex tasks that were originally performed by humans themselves, then by creating novel tools favoring the emergence of new uses,

¹Some of them have already been integrated in public rail transportation vehicles, and more recently, into prototypes of automated cars, such as those proposed by the Google and Tesla companies.

²In order to address the shortage of nursing staff to care for its growing elderly population, Japan has designed several prototypes of nursing robots able to transport patients and to assist with their bedding, through its national laboratories RIKEN. The most famous one is *Robear*, which was first conceived in 2015.

³In order to boost global research in robotics, an annual soccer competition between humanoid robot teams, called *RoboCup*, has been created to test the advancements in this domain. The research underpinning this competition notably aims to integrate into such robots diverse cognitive operations that are required to move and play soccer, such as appropriating the space of the soccer pitch, keeping the ball and interacting with other robots in order to pass the ball or not. The objective is to surpass the capabilities of a human team (one day). The well-know NAO, which was first conceived in 2007 by the French company Aldebaran, figures among the models presented at this competition figures.

⁴Commissioned by the Defense Advanced Research Projects Agency (DARPA) of the United States of America, the company Boston Dynamics has designed a humanoid robot called *Atlas* that is able to run and to jump over numerous obstacles. A 2021 demonstration of Atlas in motion is available at the following link: <https://www.youtube.com/watch?v=tF4DML7FIWk>.

⁵The first conversational agent *Chat-GPT* (for *Chat Generative Pre-trained Transformer*), launched by OpenAI in November 2022, has encountered such a large success that rival *Big Tech* companies have rapidly released their own ones, including *Gemini* by Google DeepMind, first released as *Bard* in April 2023. These major events have contributed to launch a “race” for the most competitive, powerful Large Language Model (LLM) worldwide. In line with this trend, the French start-up Mistral AI has also provided its own eponymous LLM in September 2023, one of the most powerful one in France, available in numerous versions.

production, and consumption modes, as evidenced by recent prowesses in the digital and web worlds.

Despite a sporadic development punctuated by winters and springs, from the previously successful expert systems to the current deep neural networks (Cardon et al., 2018), artificial intelligence, especially its digital forms, has constantly aroused the curiosity of social sciences. From the physical construction of high-performance computing infrastructures to the symbolic representations of knowledge produced by intelligent algorithms manipulating data, which then return information or actions at the request of users through their personal computers, AI is a complex chain of many interdependent subsystems (Crawford and Joler, 2018), whose multiple issues can be the subject of separate studies. In particular, as it has become a central component of the socioeconomic development of multiple nations⁶ in recent days, almost concomitantly with the rise of deep learning techniques (Klinger et al., 2018; LeCun et al., 2015) and its applications in diverse fields since the early 2010s, notably in the processing of large-scale image and audio databases for pattern or speech recognition and classification, the contemporary forms of AI have raised tricky questions regarding its societal impact. The latter concern ethical and transparency challenges (Burrell, 2016; Jobin et al., 2019; Kerr et al., 2020), the transformation of labor and employment (Boyd and Holton, 2018; Frank, Autor, et al., 2019; Graetz and Michaels, 2018), especially in the context of the recent platformization of economy and work (Le Ludec et al., 2020; Miceli et al., 2024), and the perpetuation and reinforcement of existing inequalities of various kinds, as well as the emergence of new ones (Beer, 2017; Joyce et al., 2021; Polcumpally, 2021), among many others. Meanwhile, these last dimensions are giving rise to special sociotechnical imaginaries in society (Sartori and Bocca, 2023), for instance the augmentation of humans in various daily tasks, which are often viewed positively, but also the replacement of workers and the inability to distinguish a humanoid robot from a human, which is especially related to the fear of a future “technological singularity”⁷ achieved by AI. Finally, these challenges tend to deconstruct the notion of AI as a purely technological phenomenon in favor of a sociotechnical perspective, at the interface of multiple “technical and social practices, institutions and infrastructures, pol-

⁶As AI is now a worldwide challenge, its wide impact across many societal sectors (including education, research, digital services, manufacturing, etc.) has led many national organizations to produce reports in order to evaluate the current state of research on the subject and the production of associated technological tools, and to delineate the contours of future research and applicative enterprise, as well as the future of work. This is notably the case in Australia (Hajkowicz et al., 2022), France (Aghion and Bouverot, 2024; Villani et al., 2018), China (CISTP, 2018), which is establishing itself as a world leader in AI research and applications (Baruffaldi et al., 2020), and the United States of America (The White House, 2019). Some international organizations and companies have also provided reports at the global scale, such as the Organisation for Economic Co-operation and Development (OECD), where Baruffaldi et al. (2020) have conducted a bibliometric study of the state of the art of recent AI research.

⁷A scenario in which AI would be more intelligent and powerful than humans, and potentially able to annihilate them. This specific imaginary is largely shaped and maintained by multiple cultural productions (books, movies, video games, and so forth), although it is not yet sufficiently developed in its current form to reach this technical stage.

itics and culture” (Crawford, 2021, p.8), and which attracts significant interest from the broad field of *science and technology studies* (STS).

Scientific activity is no exception to this increasing AI hype, as testified by an abundant literature that demonstrates its wide diffusion at various scales into the science system. Supported by research programmes and strong public-private partnerships in recent years, AI has indeed become an essential part of basic research, in almost all scientific disciplines (Gao and Wang, 2023) and in the research and development sector within specialized companies (Ahmed et al., 2023; Frank, Wang, et al., 2019), with proven technical applications in several areas (Xu et al., 2021).

Furthermore, by being part of the recent rise of a data-driven research mode, especially its facets related to machine learning and neural network algorithms that allow its efficient processing, AI has itself raised a debate about its ability to foster innovation, even to shape a new *paradigm* of knowledge production, as Kitchin (2014) has stipulated. Some authors have suggested that AI is effectively competing with well-established methods and technical tools within various scientific disciplines that choose to adopt it for some of their research branches (Bianchini et al., 2022; Cockburn et al., 2018). Without claiming that AI will completely replace the traditional analytical frameworks used in these disciplines in the near future, crucial questions have nevertheless arisen about how precisely AI might transform these traditional, disciplinary-specific ways of doing research. For instance, following the categorization of scientific creative processes proposed by Boden (1998), does AI act as an intermediary, facilitating *combinations*⁸ of pre-existing knowledge from distinct disciplines, thus contributing to the reshaping of traditional disciplinary boundaries at the cognitive level? Or, by introducing novel analytical frameworks into a given discipline, is it able to open up new fields of knowledge to be explored, either within its original conceptual basis (*exploration*) or in new ones corresponding to novel research paths (*transformation*)?⁹

These broad challenges are the subject of an ongoing investigation of the diffusion and epistemic impact of AI in science, of which this thesis is a part. Combining two projects, named *ScientIA*¹⁰ and *EpiAI*,¹¹ this study is conducted in the GEMASS lab, where the present author of this manuscript has pursued his doctoral journey, with the support of

⁸Jones, Mukherjee, Stringer and Uzzi (2015; 2013) have examined this aspect by quantifying the impact of *conventional* or *atypical*, innovative combinations of prior knowledge, especially with a large-scale citation network analysis considering individual scientific publications, while ignoring their respective disciplinary affiliations.

⁹Kauffman (2000) has analogously conceptualized such phenomenon as the expansion of the *adjacent possibles* of the discipline.

¹⁰Project funded by the French Agence Nationale de la Recherche (ANR) and dedicated to the study of applications and impact of AI in science. More details are available at the following address: <https://anr.fr/Projet-ANR-21-CE38-0020>.

¹¹Official name of the doctoral fellowship funded by the Mission pour les Initiatives Transverses et Interdisciplinaires (MITI) of CNRS.

other partners. In line with a major sociological research on the subject (Z. Liu, 2021; Woolgar, 1985), it intends first to offer recent, updated perspectives on AI within the scientific sphere. But what interests us more is twofold. On the one hand, these projects aim to examine the extent to which AI is disseminated in science, especially by identifying the research domains in which it is most prevalent today and the dynamics of its insertion within them. On the other hand, referring to the aforementioned questions, they seek to investigate whether AI has contributed (or not) to transforming the methodological and knowledge frameworks of such domains as they develop, as well as their respective organisations of scientific work and institutions, notably through the launch of dedicated scientific teams and laboratories within these domains, academic journals and conferences for communicating results, funding, university teaching programmes, perhaps learned societies, etc.

The first of the last two axes of this enterprise, namely the large-scale diffusion of AI in science, has already been addressed in a recent scientometric study that drew a dynamic cartography of the AI-related literature published since the 1970s, resulting in a publication that has been co-authored by the author of this thesis (Gargiulo, Fontaine, et al., 2023). In this paper, AI is viewed as a set of concepts, mainly algorithms and computational methods, that we can recover in scientific publications accessible through standard bibliometric databases. It especially advances two main results that have informed the main research lines of this thesis. First, AI has originated within an interdisciplinary research environment before experiencing a disciplinary narrowing around three main domains from the early 1980s, namely mathematics, statistics, and computer science. This narrowing has come to an end in the 2010s in favour of a new diffusion to a growing number of external domains. This long, albeit transient, transformation of the disciplinary environment surrounding AI research, which was necessary for its recent theoretical foundations in connectionist approaches, underlines that AI has become a set of specialized knowledge and tools designed by groups of scientists populating the aforementioned domains, alongside engineering ones, and then applied by specialists outside these domains for their own research purposes.

Although not conceived as such in our paper, this finding supports an *instrumental* conception of AI, which is typical of a *research-technology* regime, as defined by Shinn and Joerges (2002). Alternatively called *transverse science*, this analytical reading grid enables the comprehension of AI as a research instrument, here defined as a set of knowledge and technological devices capable of enhancing research and created in a moving socio-epistemic environment situated at the interface of diverse disciplinary (sub-)communities, which draw upon multiple knowledge universes and support institutions. As it is able to navigate throughout these communities, the AI research environment may prefer some of them at different stages of its development for many reasons, including the alignment of AI

with external disciplinary research objectives or the fulfillment of specific technical needs for diverse research. More importantly, according to these authors, the results produced within such a socio-epistemic environment are devoted to be disseminated beyond this environment throughout science, and to be adapted to the research objectives of different disciplines, at which stage the instrument achieves a *generic* character (Hentschel, 2015). This conceptual framework, especially the notion of *genericity*, is particularly suited to describing AI as spreading ideas and technologies in science. For example, the aforementioned stage of diffusion of AI since 2010 suggests that the latter tends to become generic after an extensive development within computer science, mathematics, and statistics.

However, a further analysis conducted in this paper has revealed that the distribution of AI-related publications across bounded disciplines remains rather uneven. It thus appears that some of the latter are more inclined to use AI for their own goals than others, such as astrophysics, engineering, medical imaging, and geography, as opposed to those in the arts, humanities, and social sciences. This finding thus challenges the supposed genericity character mentioned above, and leads to question the reasons that would foster the insertion of AI within a given scientific discipline, such as its cognitive proximity with the latter, a prior vernacular IT-based¹² culture that allows to easily master AI, fruitful collaborations within specific research arenas due to the permeability of existing disciplinary boundaries, or something else.

Conceptualization and research questions

Following the aforementioned questioning that has emerged in the course of the investigation outlined above, this thesis seeks to gain insight into the partial integration of AI in science, which concerns only some disciplines, as described in our previous paper (Gargiulo, Fontaine, et al., 2023). In line with an extensive literature on the diffusion of innovation in science and beyond (see, for example, Acemoglu et al., 2016; Cheng et al., 2023; Katz et al., 1963; Uzzi et al., 2013; Zhai et al., 2017), we need to propose an appropriate analytical framework for measuring AI diffusion in terms of various disciplinary subtleties, in order to explain the aforementioned pattern.

As previously stated, this thesis conceptualizes AI as a *research-technology* enterprise (Hentschel, 2015; Marcovich and Shinn, 2012; Shinn and Joerges, 2002), with a particular focus on the final phase of its development, namely *genericity*, here understood as a marker of both the integration of related knowledge, technologies, and practices into

¹²Information Technology, which encompasses all the knowledge related to computer science and all the possible technological infrastructures used to perform various digital work at different scales, from small ones with individual computers to heavy ones with supercomputers.

several different disciplines,¹³ and their subsequent diffusion throughout these disciplines. More precisely, we propose to define genericity with three criteria, namely 1) the capacity of the instrument to adapt and be used in various disciplinary research orientations (termed *disciplinary adaptation*), 2) the disciplinary mobility of the practitioners of this instrument in different research environments, involving their ability to initiate collaborations with non-specialists in order to make them adopt AI in their regular research practices (*social adoption*), and 3) its capacity to incorporate the conceptual ensemble of the discipline within which it is used, thereby becoming a significant part of the creation of new knowledge in it (*epistemic integration*) (Cheng et al., 2023). We thus assume that AI is *generic* in a given discipline if it fulfills these criteria.

Initially motivated by applying this model to examine the spreading dynamics of AI within a range of disciplines that were presumed to be distant in terms of knowledge and research methods, and not particularly aligned with mathematics, statistics, computer science and engineering *a priori*, this thesis instead focuses on a broad research domain actually gathering multiple disciplines (or some parts of them) around common topics, neuroscience. Although recognized as a unified entity in science, with dedicated laboratories and research facilities, an international Society for Neuroscience¹⁴ (a French one¹⁵ also exists), and benefiting from a dedicated terminology in standard academic journals' classifications available on major bibliometric databases, neuroscience is actually divided into several sub-domains¹⁶ involving many disciplines. For example, computational neuroscience, which aims to model the neural activity and information transmission throughout the central nervous system, brings together computer scientists, mathematicians, physicists and biologists, while neuropsychology, which explores the correlations between human behavior and diverse physiological brain and body signals, gathers several fields of psychology and clinical specialties, notably neurology and psychiatry. Therefore, for the

¹³As Sugimoto and Weingart (2015) have highlighted, the notion of *discipline* is subject to a wide range of definitions, which have led to numerous alternative names in a substantial literature. These multiple definitions indeed depend on the dimensions along which their creators observe the scientific activity and how they describe its structure. Commonly, a discipline may designate a unified group of scientists working on separating from others by sharing a cognitive basis, practices, and traditions, by developing a common communication language and media, such as journals and conferences, and by institutionalizing and making recognized and legitimized within the science system under a common identity. However, despite the accuracy of the attempt of definition, it encompasses many aspects that are rather difficult to combine and measure together in practice, even in sociological or scientometric research. Although the name “discipline” has already been mentioned earlier in the present introduction, also alternatively named as “research domains” or “research areas” to designate distant communities of knowledge, here we restrict our definition to a group of scientists who claim a common identity and are gathered around coherent, bounded sets of knowledge and academic journals. Although not universal, this last definition selects only some of the aforementioned dimensions we deem useful to address the research questions outlined in this thesis, as well as meeting scientometric requirements, as detailed later in this manuscript.

¹⁴Further details on its governance and actions can be found on the following website: <https://www.sfn.org/>.

¹⁵See also: <https://www.neurosciences.asso.fr/>.

¹⁶The Wikipedia webpage gives an exhaustive list of existing research branches: <https://en.wikipedia.org/wiki/Neuroscience> (consulted in September 13, 2024, last modified in August 16, 2024).

sake of clarity, we designate neuroscience as a multidisciplinary research *domain*, or alternatively as a *field* of research,¹⁷ which covers all studies of human brain cognition and the central nervous system, of their physiological mechanisms in the human body, their psychological and behavioral translations, and the potential damage of this important body system.

Moreover, neuroscience and AI share some parts of their respective history since the emergence of the latter during the 1940s and 1950s, with the support of cybernetics and distributed cognition on neural networks at that time (Hassabis et al., 2017; Macpherson et al., 2021). Without delving too far into the past, and despite the apparent proximity of neuroscience to AI's foundations as outlined in the precited literature, we aim to trace the development of AI within contemporary neuroscience since the early 1970s, by considering the latter as an exogeneous field of application of AI. We assume this (big) approximation, given that AI-related publications within the neuroscience literature account for only 3% of it between 1970 and 2020, as we will show further in this manuscript, especially in Chapter 2. Despite this relatively modest proportion of such publications, neuroscience remains one of the research fields with the highest concentration of AI-related publications in science, according to Gargiulo et al. (2023). This hence renders neuroscience a suitable case study for investigating the diffusion of this instrument within a given host research domain.

As the main research orientations and paradigms of both AI and neuroscience may have changed from these times up to recent days, it seems plausible to hypothesize that the former has received varying degrees of attention from the latter in the course of their respective development – and vice versa –, which may have accelerated or slowed down the genericity dynamics of AI into neuroscience. An underlying goal of this study is therefore to identify the periods in which successful diffusion of AI has occurred (if they exist), suggesting momentary compatibilities of the main research orientations of neuroscience and AI, high momentary mobilities of AI *research-technologists*¹⁸ towards neuroscience and the launch of new collaborations that would foster the appropriation of AI by neuroscientists, and potential reconfigurations of various neuroscientific conceptual ensembles and the creation of new knowledge spaces to be explored due to the epistemic integration of AI within this field at specific times. This thesis thus intends to assess the extent to which AI is generic in all the neuroscience, or whether it is beneficial only to some of its parts that would correspond to distinct disciplinary areas or smaller specialties. In this case, the first genericity criterion of disciplinary adaptation can thus be relaxed in

¹⁷This designation is in line with the classification standards of the scientometric databases used throughout this thesis. As it will be discussed in the following lines, neuroscience is either a *field of study* for the Microsoft Academic Knowledge Graph (MAG), or simply a *field* for the recent OpenAlex topic categorization (OurResearch, 2024).

¹⁸Designated as such by Marcovich, Joerges and Shinn (2002, 2012).

favor of a more holistic one, termed *domain adaptation* in what follows, in order to study the impact of AI in neuroscience as a whole.

With this case study, this thesis will answer the two following major research questions, which will be developed further into sub-questions:

- Is AI being adopted by the whole neuroscientific community, or is it solely aimed to specialists in related knowledge and technologies?
- Is AI really becoming generic as it has developed in this domain since the 1970s?

Methodological positioning

How might the three genericity criteria presented above be measured in a field as large as neuroscience over a period of 50 years? As these criteria correspond to multiple observation scales, different methodologies are therefore required to approach them.

A first, mainly derived from the field of quantitative science studies (Fortunato et al., 2018), consists of an extensive scientometric analysis of a vast corpus of scientific publications (articles, conference proceedings, patents, preprints, and other materials) representing neuroscience since the 1970s, in order to construct a preliminary overview of the diffusion of AI within the field. Accessible through numerous bibliometric databases, such as the Web of Science (WOS), Scopus, Semantic Scholar, Dimensions, and the recently launched OpenAlex, these publications are wealthy collections of various metadata from which we can build aggregated structures modeling some knowledge encoded in our corpus. In particular, the sets of authors who have co-signed publications reveal together a large dynamic collaboration network, which is commonly used as a proxy for studying the scientific community associated with the field under study; the sets of bibliographical references and citations both allow the linking of publications in order to study knowledge transfer from one to another and the disciplinary structure of the field; the textual elements of the papers (titles and abstracts) can be probed by natural language processing algorithms in order to provide keywords and expressions that represent concepts, whose co-occurrences within the publications originate a semantic network, which provides insights into the knowledge structure of the field and its evolution over time. Thus, in order to situate AI within these structures, it is first necessary to delineate a specific subset of AI-related neuroscientific publications, which is characterized by its own disciplinary environment (recall that neuroscience is multidisciplinary), its own scientists involved, and its conceptual basis within the field. Then, with the two obtained corpora of AI-related and non-AI publications, we propose to explore the insertion dynamics of AI in

neuroscience through the interactions between the respective networked structures built from these corpora (citation, collaboration, and semantic networks). For instance, one of our objectives is to determine whether the aforementioned features extracted from AI-publications tend to be incorporated into, or even merged with, those of neuroscience, or whether they remain distinct from the domain or behind the field despite its insertion (e.g., within a dedicated research subarea). This can be achieved with the help of various numerical indicators that allow to describe and model some knowledge diffusion dynamics on these data structures.

To tackle these technical challenges, a comprehensive corpus of neuroscientific articles published between 1970 and 2020 has been extracted from a bibliometric database that is no longer available, the Microsoft Academic Knowledge Graph (MAG). The construction of this corpus, as well as the numerous data structures used throughout the following lines, including the citation and co-authorship networks, are developed in Chapter 2. In order to validate or test the robustness of some results introduced throughout this thesis, we have also used the more recent database OpenAlex, which reinvests the original MAG database, curates it, updates it, and adds new metadata of interest since 2022 (Priem et al., 2022).

With such a corpus, the macro-scale approach exposed above enables us to evaluate the degree of genericity of AI within neuroscience, according to the three criteria defined in the previous conceptualization part. First, using a large-scale, temporal citation network, we will study the adaptation of AI to the disciplinary ecosystem of neuroscience, asking in particular whether AI has attracted some related disciplines situated outside the disciplinary context of the host field since the 1970s, and whether it has contributed to the potential reorganization of all these disciplines, established or newly introduced, around new research subjects that respond to some epistemic objectives of neuroscience. Second, using a co-authorship network, we will observe the social adoption of AI through the dissemination of related techniques throughout the neuroscientific community, and we will specifically identify spreader scientists driving such a dynamics – should it emerge – in order to examine their respective profiles, with a particular focus on their disciplinary backgrounds and trajectories, as well as the occurrence of a “methodological shift” towards AI in their papers published over the course of their careers. Finally, using a semantic network, we will study the manner in which AI-related knowledge interlinks with the neuroscience conceptual network over time, which provides some hints to evaluate the epistemic integration criterion. The latter study will be completed through the application of natural language processing techniques applied to corpus’ titles and abstracts, using, in particular a recent BERT-inspired large language model (LLM) trained with scientific databases (Cohan et al., 2020; Devlin et al., 2019), in order to model the knowledge structure of neuroscience through a dynamic cartography (resembling a geographic

map). Such a knowledge representation would be beneficial to confirm the analyses of the first and third exposed genericity criteria, by exploring the epistemic subspaces of neuroscience where AI is mostly implanted.

However, such a quantitative, data-driven approach that only focuses on the knowledge embedded in a big scientific corpus neglects several important dimensions of the construction of AI at the micro-scale. In particular, with regard to the criterion of social adoption, mentions of AI concepts in articles do not necessarily reflect its complete penetration in daily research practices of all the neuroscientists who have co-signed AI-related works. Indeed, accurately assessing this criterion requires identifying the precise use of digital tools by researchers in teams or laboratories, how these researchers perceive these new research instruments, what the tasks concretely operationalized by these scientists are and how they are distributed within the scientific team under study, and finally how these elements could together contribute to forge a dedicated professional identity, or a community of experts around a common *technological culture* (Forsythe, 1993a). Although recent advances in scientometrics, social network analysis, natural language processing, and sentiment analysis are now able to capture these dimensions in some metadata of scientific publications, albeit with limited scope, we can approach them more precisely through laboratory ethnographies and interviews with scientists themselves in various fields of research that are welcoming AI (Laudel and Gläser, 2007).

To this end, we have conducted interviews in spring 2021 with scientists affiliated with a multidisciplinary team, Aramis, which is a member of the Institut du Cerveau (ICM, or Paris Brain Institute in English) based at the Pitié-Salpêtrière Hospital in Paris. Combining academic and clinical research, the team specializes in the field of computational neuroscience, notably in the precise prediction of the onset of neurodegenerative diseases (mainly Alzheimer's, Parkinson's, Huntington's) with the help of in-depth analyses of large, heterogeneous medical data acquired by worldwide clinical trials. Consequently, these research purposes motivate the design of technical applications, often in the form of software enabling the earlier detection of patients as soon as first warning signals occur, and to better monitor their evolution and care them. Gathering specialists from computer scientists to neurologists combining both medical and research activities, such a team allows us to enter the realm of AI research practiced within the neuroscience universe. Although this investigation remains somewhat limited to evaluate the large-scale disciplinary adaptation and epistemic integration of AI in neuroscience, as it reflects only the state of AI research in a single laboratory at a given time, it offers valuable insights into the social adoption of AI in terms of daily research practices and interdisciplinary work (Marcovich and Shinn, 2011; Sedooka et al., 2015; Stokols et al., 2008), notably through personal narratives that depict interactions between AI experts and non-specialists.

Falling within the domain of STS (Hackett, Parker, et al., 2017) and drawing upon effective methods from *computational social science* (Conte et al., 2012; Lazer et al., 2009) and *big data* (Bastin and Tubaro, 2018; Do et al., 2022), this thesis thus adopts a top-down, mixed-methodological, multi-data approach to study the genericity of AI in neuroscience.

Outline of the thesis

The manuscript is divided into five chapters, completed at its very end with appendices that mainly expose additional confirmation results and methodological details.

In Chapter 1, we introduce a concise history of the concept of AI, an overview of the literature on the diffusion of AI, and a detailed section on the *research-technology* regime that has been mentioned earlier in this introduction and adopted to represent AI as a research instrument in science. The following Chapter 2 echoes the preceding section on methodological positioning and exposes the successive steps of collection and preprocessing of scientometric data and qualitative interviews required to address the main research questions that have been established above.

The following three chapters examine the data and offer responses to these questions. Chapter 3 focuses on the first two genericity criteria described above, namely disciplinary integration and social adoption, and is partly related to a work published (and revised here) by the author of this thesis (Fontaine et al., 2024a). Drawing upon the evolution of the citation and collaborative organization of neuroscience, it introduces, on the one hand, the degree of embedding of AI into the multidisciplinary context of neuroscience and, on the other hand, its diffusion throughout the neuroscience community. This chapter will also observe the evolution of the internal disciplinary configuration that shapes neuroscience over time, and whether the gradual implantation of AI modifies it. More precisely, these goals translate into the two following sets of questions:

1. What is the disciplinary landscape around neuroscience research which is using AI? How does such research fit into the disciplinary objectives of neuroscience?
2. Who are the actors leading AI research in this field? Is AI-related knowledge widespread throughout the whole scientific community?

Chapter 4 presents a knowledge cartography of neuroscience, generated with advanced lexical embedding techniques, in order to identify the knowledge areas where AI is mainly

used and to assess whether the results of AI research within these areas are disseminated to the rest of neuroscience. We especially ask the following research questions:

3. Which neuroscience areas are attracting AI knowledge and devices the most? Does AI fit into the conceptual universe of these areas?
4. How are the AIs produced in these subfields promoted into the whole discipline?

Through these questions, we will test the hypothesis that AI-related knowledge and technologies are better suited to address some neuroscience subjects than others over time, as well as their ability to be transferred to different subfields of neuroscience.

Chapter 5 introduces some perspectives derived from the aforementioned short-term investigation conducted in spring 2021 within the Aramis team. Through interviews' excerpts reflecting some pieces of research activities led by the scientists within this team, we aim to answer the following research questions:

5. What kinds of scientists use AI? Do some of them have a prior scientific background that facilitates its handling?
6. Has AI transformed the practices of involved researchers?
7. How do they perceive AI, especially in terms of the future of scientific and medical work?

We finally conclude this thesis in a final dedicated part that summarizes the main findings according to the main questions that were posed above. We also suggest some perspectives and future work to complement this research.

Artificial Intelligence: a sociotechnical, multifaceted object

Même si cela peut paraître extravagant [...], il n'est pas si sûr que résoudre la moitié des Principia Mathematica [A. N. Whitehead et B. Russell] nécessite une plus grande intelligence qu'aller chercher son courrier dans sa boîte aux lettres.

Borrowed from (Vayre, 2021)

Artificial Intelligence (AI) proves to be multiple in the scientific field (Z. Liu, 2021). Originally conceived as a whole field of research with a dedicated research programme, as proposed by its pioneers at its very beginning (McCarthy et al., 2006 [1955]), AI has subsequently evolved into a set of digital technologies that are mainly used by multi-skilled, proficient individuals who share a common set of technical practices and a common metalanguage to solve various typical problems in the scope of computer science and mathematics (often related to optimization, control theory or decision making) and who are able to work in a variety of disciplines, and even various professional environments (Forsythe, 1993a; Kirtchik, 2019). Being not only a set of conceptual and technical expressions anchored in today's daily research practices in academic laboratories, AI is also, at the macro-scale, a broad *technoscience* whose development effectively takes place not only in the academic sphere, but also within data-producing companies, mainly the Big Five tech companies¹ (Ahmed et al., 2023; Crawford and Joler, 2018) and with the support of public policies to deploy AI-related technologies in various societal sectors, such as education, healthcare, agriculture, transportation, and national defense, as listed in the French parliamentary report by Villani et al. (2018). Even since its inception (Cardon et al., 2018; Vayre, 2021), AI has been shaped by a multitude of actors, academic and non-academic, who lead and work in corresponding research and technological deployment, and who

¹Alphabet (including Google), Amazon, Apple, Meta (including Facebook) and Microsoft.

contribute to develop interaction arenas² between diverse professional worlds, whether in the design or promotion of AI in society. Driven by strong technological, economic promises³ formulated by industries and public policies, such technoscientific arenas influence particularly the ways of conducting and organizing research within dedicated teams or laboratories, the allocation of funding, the implementation of technological infrastructures, and also the establishment of dedicated university courses to meet the staffing needs of academic research and industrial engineering.

In light of these assertions, it can be seen that AI has undergone an expansion in scope since its creation, accompanied by a notable semantic shift, thus offering various interpretations of it. Consequently, before delving into the methods and results constituting the core of the manuscript, this chapter provides a state of the art of social studies of AI in order to situate our discussion in this field. After a brief summary on the history of AI-related ideas and its main advancements since the 1950s, as presented in Sect. 1.1, we expose some studies on the social construction of AI in Sect. 1.2 since its first major applications in the 1970s and 80s – although they did not encounter long-term success. Then, in Sect. 1.3, we review the different frameworks to apprehend the diffusion of AI in science at various scales, including the whole science field and individual disciplines. Finally, we detail in Sect. 1.4 the notion of *research-technology* that has been presented in the Introduction. This concept enables the analysis of the social, institutional, and epistemic organization of a type of research that is committed to the conception of an instrument for various disciplinary purposes outside the socio-epistemic context in which it has been developed, and also outside academia – such as laser beams, X-rays, or computers (Hentschel, 2015; Marcovich and Shinn, 2017). This conceptualization of AI inside a *research-technology* enterprise allows various interpretations of the phenomena we can observe in both the bibliometric data representing the neuroscience literature and the interviews conducted with neuroscientists in contact with AI.

²As an example, Rosental (2002) has studied the arena of *demonstrations* of software and machines at dedicated conferences, in order to stimulate the interest of industrial investors.

³In the context of the recent rise of technoscientific fields, Cointet, Joly and Raimbault (2021, 2016) have studied the construction of a specialty, namely synthetic biology, which gathers heterogeneous profiles, from biologists to engineers, in order to design prototypical devices of DNA assembling. Given the numerous potential applications, not only in genomic and sequencing but also beyond, an approach encompassing AI as such is valuable for a wide range of future works on the subject, which would focus on the production arena outside standard academic research, and whose potential applications are not limited to natural fields of applications of AI, mainly related to engineering and computer science. In this sense, the notion of *technoscience* is similar to the *research-technology* regime advanced by Shinn and Joerges (2002).

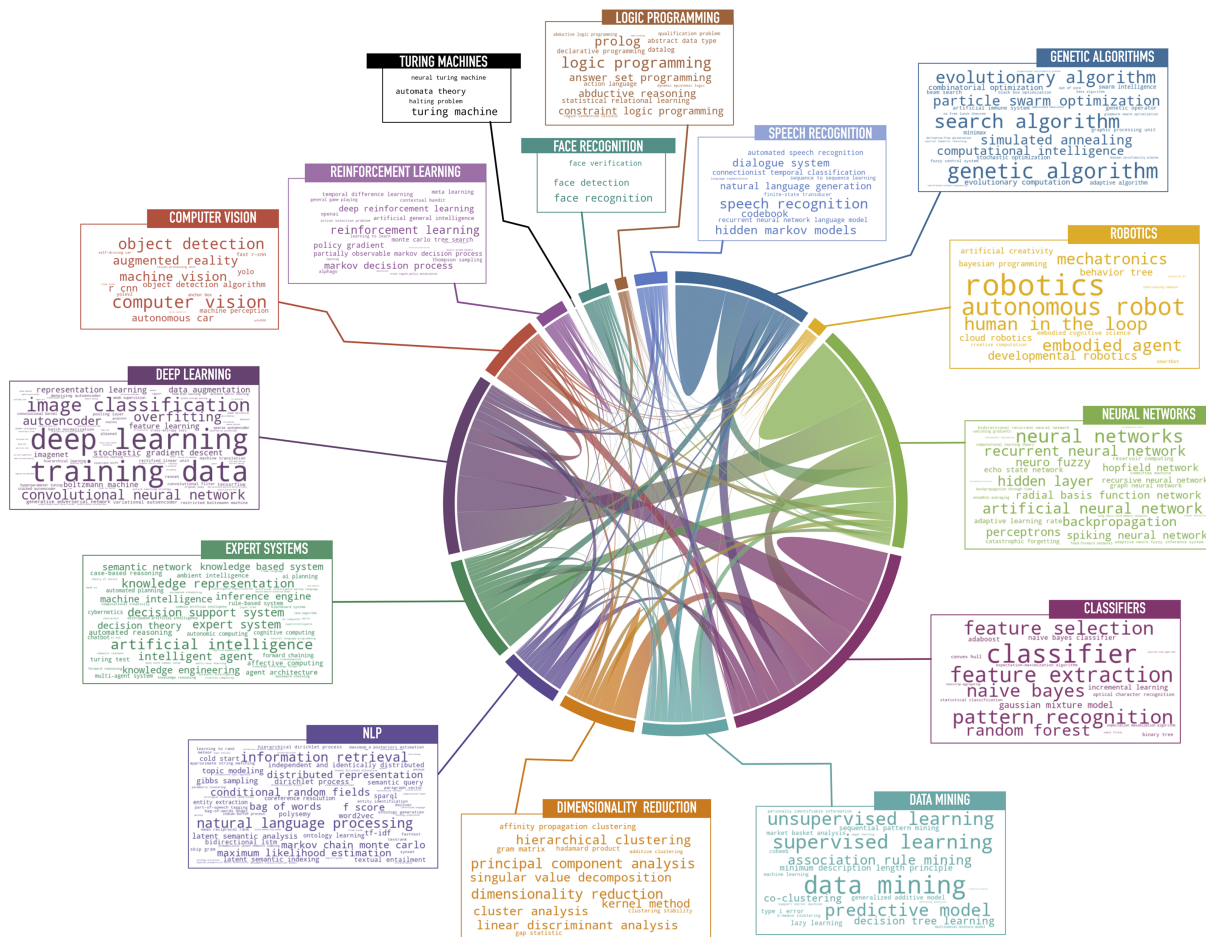


Figure 1.1: State of the art of the scientific literature on AI published between 1970 and 2017, aggregated⁴ into 15 specialties. A link between two classes is the number of terms they share, based on a co-occurrence network of AI-related terms in this literature. Source: Gargiulo et al. (2023).

1.1 A brief history of AI

Known today as a combination of robotics, software and algorithms, as shown in Fig. 1.1, AI is first and foremost the result of a rapid, multidisciplinary scientific advancement on a human scale. A plethora of books and reviews on the subject report that its development trajectory can be divided into three main movements that dominated the scientific arena at different times, namely *cybernetics*, *symbolic* and *connectionist*, in the wake of numerous controversies and tensions that have opposed them – especially the last two (Andler, 1990; Buchanan, 2005; Cardon et al., 2018; Ganascia, 1993; Olazaran, 1996; Smith, 2019; Vayre, 2021).

⁴Alternatively, by analogy with the studies of knowledge flows between scientific specialties (Di Bona et al., 2023; Sun and Latora, 2020), Liu et al. (2020) have proposed a “citation involved Hierarchical Dirichlet Process” model to construct the evolution path of some AI-related fields of study, which are subject to several reconfigurations of their scope over time as new concepts are added or removed, thus implying a potential semantic turn of these fields. For instance, the field of *machine learning*, which has first emerged between 1996 and 1999 according to their model, has broadened its scope by successively attracting the field of *bayesian networks* and some parts of *pattern recognition [for] image analysis* during

The construction of AI first owes much to *cybernetics*, a multidisciplinary research movement at the crossroads of biology and information science, whose aim was to understand the mechanisms of information transmission between the components of a system, such as living organisms or robots. Norbert Wiener was one of the prominent figures of this innovative movement, who established the principle of *feedback loop*,⁵ namely the capability of a system to adapt its future behavior and actions according to the information circulating between its sub-systems, thus inducing interactions between micro and macroscales of the system, or coming from the outside. Cybernetics encountered great interest in the 1940s and early 1950s, thanks to the Macy Conferences that laid its formation, whose members represented the “first [wave of] cybernetics” (Léon, 1996). After an attempt to legitimize itself in the international scientific community as a new metalanguage capable of unifying all the sciences (Bowker, 1993; Gerovitch, 2001), the movement gradually faded in the face of the simultaneous advent of neuroscience and, in particular, cognitive science. Nevertheless, cybernetics preceded the emergence of a “general systems science”, which could also be compared to the field of complexity science (Simon, 1962), and also helped to establish the field of robotics. Many achievements marked this stage, notably the derivation by Warren S. McCulloch and Walter H. Pitts (1943) of the first mathematical models of signal transmission on different neural circuit architectures, and the conception by Frank Rosenblatt (1958) and his team at the Cornell Aeronautical Laboratory of the *perceptron*, which is a machine based on the former neural network models applied to the formulation of responses after excitation of a retina that sends electronic signals to the neural circuit, with a learning process based on a back-propagation of information between the two intermediate layers constituting the circuit (see Fig. 1.2). Cybernetics is thus a precursor of the *connectionist* movement of AI, the field of information processing in the brain and machines under a paradigm centered on learning.

In the same vein as cybernetics, which was promoted by the aforementioned Macy conferences, the notion of “artificial intelligence” *per se* has entered the scientific field with another conference, the 1956 Dartmouth conference (McCarthy et al., 2006 [1955]). Officially known as the *Dartmouth Summer Research Project on Artificial Intelligence*, this multidisciplinary assembly laid the foundations for the first programmatic milestones of AI research, focusing on the dual challenge of devising ways to simulate “our intelligence” (first goal) in order to better understand it (second goal). The following decade saw a general enthusiasm despite some dissonant voices (Schwartz, 1989), with the emergence

this period, then others from *natural language systems and machine translation* between 1999 and 2002. This dynamic framework could enrich the static one introduced by Gargiulo et al. (2023), in particular by observing the dynamic entanglement relationships between their 15 AI specialties over time. The framework of Lobbé et al. (2021) would also be interesting in this respect.

⁵This notion is often derived from a well-known story. Having contributed to the American arms race during the Cold War and to the design of automatic guidance systems for airborne missiles, Wiener wondered how the trajectory of such a missile, the angle and the direction of the gun that would propel it, should be adjusted according to the trajectories of enemy aircraft detected by radar.

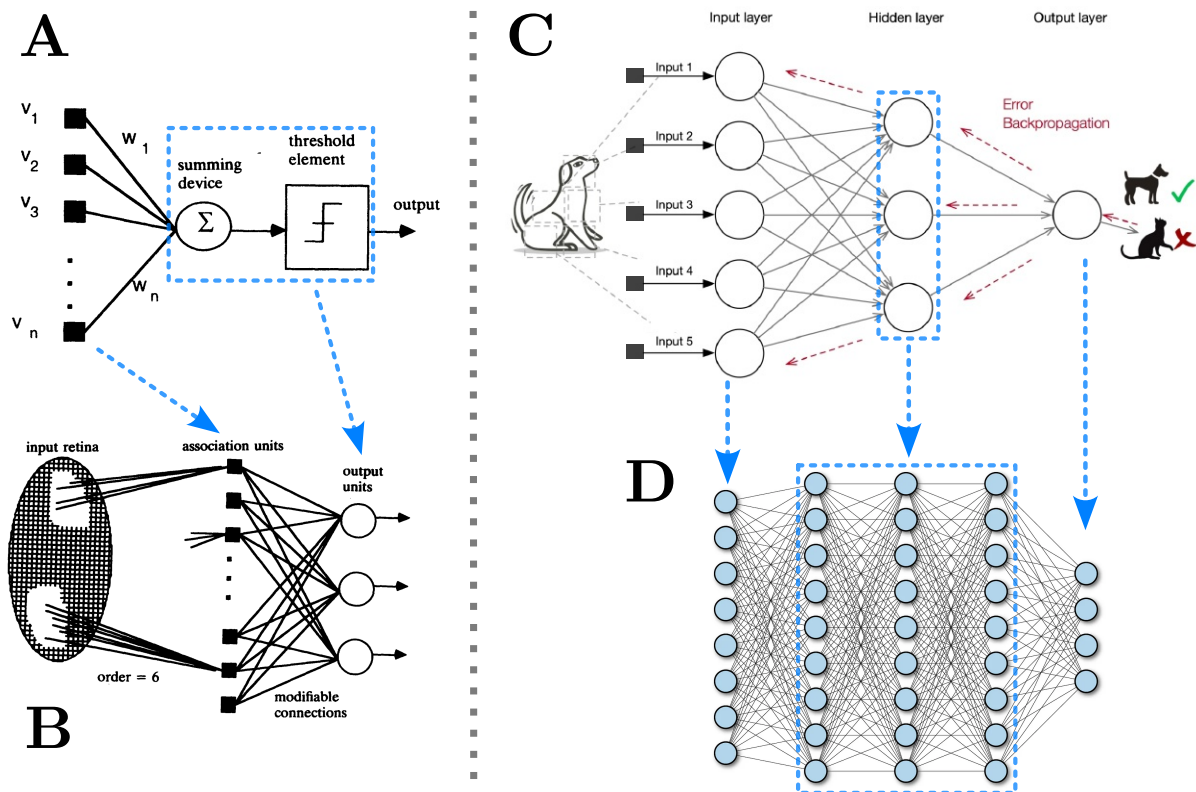


Figure 1.2: **A**: Representation of the functioning of a single neuron, the elementary unit of a neural network (blue box). The neuron performs a calculation given a set of inputs. **B**: Schematic of Rosenblatt’s perceptron (1958), composed of neurons aligned in a single layer and receiving information from the retina. **C**: Representation of a more complex architecture, with 9 neurons distributed in 3 layers (*input, hidden, output*) interacting forward – toward the output – and backward – back to the first layers. This last feature is mainly at the foundations of the back-propagation of errors that could make the model when learning to correctly predict the classification of the items of a given input dataset (Rumelhart et al., 1986). Rosenblatt’s original perceptron is simply based on the hidden and output layers, with the same mechanisms of forward and backward propagation of information between them. Current advanced *deep learning* methods, albeit often more complex, are grounded on a multilayer perceptron architecture, with a multiplication of hidden layers (**D**) – here they are numbered to 3, but it could be more – each having its own representation of the input data at given moments of the learning process (LeCun et al., 2015). Sources: (**A,B**) Olazaran (1996), (**C**) Cardon et al. (2018) and (**D**) Terry-Jack (2019).

of speculative discourses about the possible future of AI and the possible power of an intelligent programme, as exemplified by the two pioneers Herbert A. Simon and Allen Newell, who firmly believed that by the 1960s and 1970s AI would be able to prove mathematical theorems on its own and even create new ones, thus replacing humans in many tasks (Ganascia, 1993; Vayre, 2021), and also Pamela McCorduck and Edward A. Feigenbaum, for whom AI would become an integral part of our lives, living alongside humans (Schwartz, 1989).

Despite its relative youth, an AI-related “promise-based economy” has gradually taken hold, with vast funding programmes from private companies and governments (Vayre,

2021). For example, the emerging field of Natural Language Processing (NLP) has attracted the interest of the United States during the Cold War, which have funded a major research programme to translate secret, encrypted Soviet documents – a kind of legacy of Turing’s work on *Enigma* machines for the Germans during the Second World War. Although the geopolitical and industrial context of wartime has often proved favourable to the development of numerous computational AI techniques or precursors of AI, American attempts at NLP and many potential military applications have ended up in numerous dead ends, never leaving the laboratories, largely due to the lack of sufficient technologies to perform heavy computations and to store large databases in the working memory of available computers (Cardon et al., 2018). This failure led the US government to stop funding all research into machine translation in the mid-1960s. The 1973 *Lighthill Report* also pointed out the unfulfilled promises of the AI research programme that had started earlier in the United Kingdom. Because of these failures, the period between the mid-1960s and mid-1970s is commonly referred to as the *first AI winter*. Nonetheless, this was only a temporary setback, as other research was also being launched at the same time, albeit much less tremendous, but it did give rise to *expert systems*.

In contrast to the competing connectionist approach of AI, the architecture of expert systems is based on a set of logically interconnected rules (e.g., if A is equivalent to B and B is equivalent to C, then A is equivalent to C), so that the reasoning consists of navigating through a decision tree that starts and ends with predefined features. These knowledge and reasoning structures are then used to interpret the data in a given collection, thereby mimicking an ideal intelligent reasoning process of an expert solving a complex problem. First developed in the 1950s and 1960s during the rise of knowledge engineering, these systems have thus proved to be very useful in assisting decision-makers and have rapidly attracted the attention of industrial investors and research actors. The Logic Theorist and the General Problem Solver (GPS), developed respectively in 1955 and 1959 respectively by Herbert A. Simon, Cliff Shaw and Allen Newell for proving mathematical theorems, the Dendral project (Lindsay et al., 1993) for automatic detection of molecular structures in materials based on mass spectrometry and Nuclear Magnetic Resonance (NMR), and the MYCIN software, which was designed to administer an antibiotic based on the symptoms of visiting patients (Shortliffe and Buchanan, 1975), are four of the best-known expert systems that have brought the so-called *symbolic AI* or *GOFAI* (*Good Old-Fashion AI*) to the fore (Haugeland, 1985; Smith, 2019). These systems, and symbolic AI in general, were successful until the 1990s, promoted in particular by a deligitimization operation by Marvin Minsky and Samuel Papert through their book entitled *Perceptrons*, published in 1969 (Cardon et al., 2018; Olazaran, 1996; Vayre, 2021). However, most of these GOFAI have encountered various obstacles in their pursuit of achieving intelligence reproduction. These include the encoding of a single mode of reasoning (often hypothetico-deductive, sometimes inductive), the limited and non-adaptive knowledge bases implanted by their

designers, which reflect reductive representations of the world – Cardon et al. (2018) called such knowledge universes as “toy worlds” – and the inability of the researchers of the time to implement tacit, common sense knowledge in a computer, considered proof of a “real” intelligence at the time (Bloomfield, 1988).

Despite the promising outcomes of research on expert systems, which resulted in patents, and despite their major limitations in terms of the types of problems they can solve, AI received a “second cybernetic breath” in the 1980s due to the contribution of cognitive sciences, which brought together biologists, physicists and electronic engineers. The gradient backpropagation algorithm of Rumelhart et al. (1986) is the most famous example of the advent of this *connectionist* era, as they have reinvested the pioneering work of Rosenblatt (1958) on the perceptron by generalizing the learning process to any number of neuronal layers (see Fig. 1.2). The following decades have also seen the emergence of increasingly sophisticated neural network architectures for data classification, which learn to perform their assigned task with a *training* phase requiring either annotated data (*supervised learning*, the classes within the training dataset are known and the algorithm is trained to predict them, and to generalize the prediction process to other datasets it has not yet “seen”), or unannotated ones (*unsupervised learning*, the classes within the training dataset are unknown). Convolutional neural networks⁶ (CNN), recurrent neural networks⁷ (RNN), generative adversarial network⁸ (GAN) and the newer transformers⁹ appears among the most prominent classes of artificial neural networks, which are now commonly grouped under the term *deep learning* (LeCun et al., 2015; see also the handbook of Zhang et al. (2021), pp.251-348, pp.349-439, pp.949-963, pp.440-503 for the respective detailed introduction of these algorithms’ classes).

Although not yet hegemonic in the 1990s and 2000s, as it faced to a *second AI winter* at

⁶A convolutional neural network (CNN) is a complex architecture especially designed for two- or three-dimensional image classification and segmentation. It comprises a succession of layers that learn specific features within the training images to accurately identify a target object, such as details confined to a small area of an image and involving only a few pixels.

⁷A recurrent neural network (RNN) is a standard multilayer neural network with a memory component. In such a network, a neuron in a given layer has two outputs, one that is passed on to the neurons in the next layer, and others that are stored and reused in combination with other inputs in subsequent time steps, thus directly influencing the outputs given by one neuron to the next. RNNs are therefore pretty useful for analyzing time series and highly correlated datasets, the most commonly used type being the Long Short-Term Memory (LSTM).

⁸Generative Adversarial Networks (GAN) are used to generate content with two competing neural networks: a first one is dedicated to generating new data *per se* and a second one is trained to evaluate whether the new content “exists” in the real world, according to its knowledge stored in its training database. More precisely, the first network would adapt its new generations according to the successive evaluations of the second one, in order to generate a content as close to reality as possible. The most famous model is StyleGAN, on which the website *This person does not exist* is based. Each time the page is refreshed, the website returns a different facial image of people who (supposedly) do not exist, but which looks remarkably like a photo of someone who does. The aforementioned website: <https://www.thispersondoesnotexist.com>.

⁹One of them is used and explained later in Chapter 4.

this time (Cardon et al., 2018), connectionist AI will gain visibility from the 2010s with the advent of technologies capable of processing¹⁰ voluminous databases, especially massively parallel computing – while expert systems were only sequential programmes. The recent rise of a *big data* era which is characterized by the accumulation of so much data that it can no longer be analysed by hand or with standard statistical techniques, has especially contributed to a profound transformation of AI-related research, which is now guided by a “data-driven scientific discovery” paradigm to analyze and reveal the structures of large databases (Bianchini et al., 2022; Kitchin, 2014; see also Frégnac, 2017 for a similar situation in the brain sciences). In particular, such a mode of knowledge production is organized along standard pipelines, with successive crucial steps that are necessary to process these data, from their construction (Jaton and Vinck, 2016) to the design of specific AI algorithms to analyze them, often for classification purposes (Crawford and Trevor, 2019). All of these tasks are carried out by different professionals who are confined in delineated sets of institutions, technical norms, values, and worldviews that are ultimately inscribed in the final data processing pipeline.

Finally, the goals of AI research seem to have moved away from the initial research programme and ambitions first set out by the protagonists of the 1956 Dartmouth conference, from the desire to create a *strong AI*, autonomous and interacting with us in society, to the development of a regulated *weak AI* that assists (Kerr et al., 2020) or *augments* us in our tasks (De Vignemont, 2020; Jordan, 2019) and that offers a revision of the general organization of labor in all sectors (Ganascia, 1993). Jordan (2019) ironically summarizes this situation as follows (p.9): “In an interesting reversal, it is Wiener’s intellectual agenda that has come to dominate in the current era, under the banner of McCarthy’s terminology.”

1.2 The academic work of AI captured by sociology

As outlined in the very first lines above, AI is an object that is grasped by social sciences at the intersection of many specialties, including science and technology studies (STS), but also innovation studies, research policy, sociology of work and organization, and political science. In their recent review and call for a new sociology of AI, Joyce et al. (2021) have identified two main themes addressed by the social sciences towards AI, namely 1) the underlying process of conception of AIs prior to their release and introduction in society, particularly through the data management and coding practices that underpin

¹⁰We note in particular the prowess of AlexNet (Krizhevsky et al., 2012), coded into the computer’s graphics processing units (GPU), which brought CNN to the forefront of computer vision with its impressive performance in the 2012 ImageNet Large Scale Visual Recognition Challenge (ILSVRC), which is a competition for image classification based on the ImageNet database. Indeed, AlexNet beat the previous winner in this competition by reducing its prediction error by approximately 10 points.

the social construction of algorithms, and 2) the influence of current uses of AI on the evolution of all components of the social world, e.g., norms, values, modes of consumption, institutions, policies, among others (what the authors have called the “social shaping of AI in practice”). In another review, Liu (2021) has preferred a categorization of social studies of AI into three perspectives related to the “type” it can endow, namely 1) “scientific AI”, which we could associate with the first thematic set mentioned above, 2) “technical AI”, focusing on its applications in many other contexts – not only scientific – and 3) “cultural AI”, which we could associate with the second theme of the aforementioned review.

Without going into the details of this vast literature, here we select some works that emphasize the social construction of AI, from the laboratories to its experimental dissemination in some fields that are classic case studies in social studies of AI – speech recognition, knowledge engineering, medical decision-making, radiology, etc. Some of these works constitute relevant premises for a general sociology of AI, which initially led to a sociology of expert machines. The works presented below complement others already mentioned in the previous section of this chapter, which have focused more on the history of AI ideas through the controversies about cognition and the underlying processes that originate the human mind in the brain, which have induced a polarization of the AI community throughout its history, which has now almost disappeared.

As soon as the first “intelligent” systems were developed, Woolgar (1985), in a kind of manifesto, has first suggested some relevant ways to study AI in scientific research from a perspective related to the sociology of science. After first proposing three manners of investigating the social fabric of AI according to the standard purposes of constructivist sociology of science, namely 1) building a sociology of AI researchers and their motives, 2) studying the outputs of AI research, and 3) studying human-machine interactions, whose human actions of requesting and physically tinkering with the machine are at the center of such analysis, Woolgar has swept these aside in order to introduce a fourth way of doing a sociology of AI, which considers the latter as a being placed on the same stage as humans and endowed with a social sense, or *agency* (Alač et al., 2011), without saying so in his text. Summarized by Woolgar as the study of “the community composed of ‘*expert machines and machine experts*’ ”, this rupture (which seems like a refusal) of the human-machine opposition in terms of social trait, is inherent to the era of expertise surrounding AI in his time, when the discourses and imaginaries, often consensual, together indicated a “rhetoric of progress” towards the creation of an expert machine that would be equal to humans.

Nevertheless, the first perspective of Woolgar’s programme, i.e., a sociology of AI practitioners, has been the subject of a more prominent literature, even at the time of expert systems. In particular, the construction of the knowledge given to reasoning machines,

the process of judging what is a good knowledge, the tensions between different actors, and the issues this may imply, since the machine may merely reproduce most of the biases that the designers would bring to it (Bechmann and Bowker, 2019), have been at the heart of many investigations within the community of knowledge engineers, involving either a review of their internal discourses (Bloomfield, 1988; Schwartz, 1989) or ethnographic fieldwork in their laboratories (Forsythe, 1993a; Forsythe, 1993b). The latter in particular have shown that the practice of AI within dedicated teams leads to the building of a dedicated *technological culture*. They have also inspired more recent studies that are not centered on the production of AI itself, but on semantically gravitating objects such as *algorithms* and *big data* that have escaped the lab and are now diffusing into society (boyd danah and Crawford, 2012; Burrell and Fourcade, 2021; Kirtchik, 2019).

So far, the constructivist programme of AI in laboratories has understood this notion solely as an engineering enterprise, a set of computational techniques to enhance the production of new knowledge. However, this focus tends to see AI as disconnected from potential applications in disciplinary contexts other than itself, here expert systems or knowledge engineering. Other work, on the other hand, proposes leaving the laboratory to observe AI in action within other areas. As already outlined in the previous section, one of these areas, which has produced a lot of work at the time of expert systems and continues to do so today, is medical support.

During the experimental deployment of the INTERNIST-I software, which specializes in inferring diagnoses from patients' symptoms, Weaver (1986) has described the implications of introducing an assistive expert system into the medical routine, from enhancing the practice of physicians to the risk of the latter's gradual disappearance in favour of an intelligent computer, which could jeopardize the human factor at the heart of the medical practice and thus reduce the quality of patient care. In the same vein, radiologists have recently been subjected to the deployment of experimental AI software for image segmentation and the detection of anomalous bodies on radiographs, and have therefore been widely scrutinized as they express a cautious attitude towards the software when used within the framework of standard practices – AI may indeed return misinterpretations of an identified pathology, thus distorting a diagnosis – and a great concern about the replacement of their work by a machine. In particular, the entry of such intelligent computers into medical imaging strains the tacit knowledge at the heart of radiologists' medical training (Anichini and Geffroy, 2021; Gaglio and Loute, 2023; Matuchansky, 2019; Mignot and Schultz, 2022; Pesapane et al., 2018), and raises transparency issues when the software involved produces dubious conclusions. The latter are often expressed in a “black box” rhetoric to denote the lack of knowledge of the algorithm and the inability to follow its reasoning process, especially in the case of artificial neural networks working with parallel computing and both deductive and inductive reasoning (Topol, 2019; Winter

and Carusi, 2022).

These limitations, especially transparency, are also highlighted in a recent survey conducted by Fecher et al. (2023) with researchers using the LLM Chat-GPT in various research areas. They have advanced three further concerns about the use of such LLMs in research (p.6), namely their “lack of creativity”, the rapid “outdatedness” of the algorithm, as a single version is trained with a specific knowledge base at a given time, and the “unspecificity”, as they return too generic, sometimes disorganized responses. More generally, the perception of AI by academic researchers or other actors welcoming AI is mainly oriented towards its ethical use in diverse contexts and “opening the black box”, which has given rise to the field of *explainable AI*, also abbreviated *XAI* (Burrell, 2016; Cupe, 2018; Gilpin et al., 2019; Savage, 2022).

1.3 How to study the diffusion of AI in science? An overview

The research programme for a sociology of AI, as presented above, has thus far sought to examine some microscopic aspects of the spreading of AI in science, mainly in the lens of the underlying scientific work behind AI and the reconfiguration of scientific practices in the light of the arrival of new technologies in specific disciplines. However, by focusing exclusively on special machines and algorithms applied, often experimentally, in a given scientific and technological context, the aforementioned works do not trace the long-term impact of these tools on the newly produced knowledge in these contexts, nor the extent of the real adoption in their methodological and conceptual frameworks.

In the framework of the science of science (Fortunato et al., 2018), an abundant literature has been produced for parts of the aforementioned purposes, mainly providing large-scale numerical analyses of the scientific textual production we can find in various bibliometric databases accessible on the Web, including peer-reviewed articles, conference proceedings, patents, grant proposals, university courses, among others. Drawing on some methods from scientometrics, research policy, and social network analysis, this literature depicts the historical and current situation of the diffusion of AI in science, in particular its extent in the research fields that choose to use it for some of their tasks, the nature of the impact of AI in the latter – e.g., does it transform their methodologies and knowledge over time – and the factors that explain such an adoption of AI by the actors of a given research area.

Here we detail some analytical approaches from this literature that have kept our at-

tention to how to conceive the mechanisms of diffusion of AI in science, and how it might interact with other research domains. Moreover, these approaches have the advantage of being applicable not only to AI but also to any knowledge, technology, or innovation that spreads in science. To this end, we distinguish two scales of analysis, which we illustrate with some case studies.

1.3.1 At the macro- and meso-scale: AI in the scientific system

A first kind of analysis relies on the macroscopic study of the overall science system, asking how AI fits into the latter and how the former might change the state of the latter (and vice versa). As an exemplary case, Gao et al. (2023) have conducted an extensive lexicometric analysis on a huge dataset from 19 disciplines, including not only standard papers but also patents and course syllabi from various American universities. Through a measure of the cooccurrence of some AI concepts with other terms that characterize these disciplines over time, defined in the paper as impact, this paper illustrates the ability of AI to be embedded in the conceptual universes of several disciplines.

Alternatively, other authors have focused specifically on the AI literature, interrogating the historical reshaping of AI research, whose contours, influences, and impact on science may change over time. From a bibliometric corpus of AI literature published between 1970 and 2017, we and Gargiulo et al. (2023) have established a comprehensive categorization of AI tools and concepts embedded in the textual elements of this literature (see Fig. 1.1), and have mapped the disciplinary ecosystem of AI research and the preference of some fields for only a few types of AI methods. We have also shown a retreat of research fields that have been at the heart of the creation of AI-related knowledge since its very beginning, such as medicine, psychology, philosophy, and logic, which have been replaced by fields related to computer science, mathematics, and statistics. Frank et al. (2019) have come to the same conclusions with a similar AI corpus, exhibiting a reconfiguration of the citation network within the adjacent field of computer science, in which AI becomes increasingly central in all its subfields through a growing intensity of citation feedback from them to AI. They have also shown that such a reconfiguration is part of the engineering and industrial turn of AI (Ahmed et al., 2023), where Big Tech companies have recently emerged at the forefront of AI and transformed the pattern of citation from diverse fields of research toward STEM-oriented¹¹ works and for broad industrial applications.

Thus, in all these papers, AI-related knowledge seems to directly reshape science, not only at the level of its inscribed knowledge in articles, but also at the level of the organization of research, with the rise of the private sector. These papers also bring supplementary

¹¹Science, Technology, Engineering and Mathematics.

insights to explain the shift in the programmatic orientations of AI (Cardon et al., 2018), already described in Sect. 1.1, with the stronger influence of STEM-related disciplines and short-term technological applications on an industrial scale than of basic research emanating from computer science, cognitive science, and neuroscience, mainly conducted in academia (Klinger et al., 2022; “The new NeuroAI” 2024).

1.3.2 At the micro-scale: the social and epistemic diffusion of AI

However, the aforementioned approaches suffer from ignoring important factors that could help or hinder the diffusion of AI-related ideas across science. One of the most important (and perhaps naive) factors is the cognitive proximity of AI-related knowledge to other knowledge associated with certain fields of research (F. Liu et al., 2024). For example, as Gargiulo et al. (2023) have shown, since the 1980s, contemporary AI knowledge and methodologies have tended to fall more within computer science, mathematics, and statistics, while failing to attract the social sciences and humanities, which seem to be progressively missing the contemporary connectionist turn,¹² despite being at the heart of the very first AI research programmes in the late 1950s (Frank, Wang, et al., 2019).

In a recent paper, Bianchini et al. (2023) have provided a partial explanation for the above observation, by proposing a unified framework based on several factors that characterize scientists in order to predict their ability to adopt AI. They have highlighted four complementary effects, namely the promotion of AI within the scientists’ institutions, especially through the easy accessibility of High Performance Computing (HPC) facilities, their collaboration with AI practitioners (if they are not one of them), their academic status (age, position and reputation) and their “taste for exploration” (March, 1991; Singh et al., 2024; Zeng et al., 2019), which is linked to their respective “disciplinary diversity” (sic.) encoded in their publications. This last factor is particularly interesting, as it is seen as the ability to bridge the traditional disciplinary or conceptual boundaries that are imposed by the institutions or the collaborations, as opposed to the resistance to adoption, also called *intellectual inertia*, which is often expressed by scientists who

¹²This conclusion about the social sciences and humanities must be put into perspective. Indeed, it reflects a statistical bias, because the number of publications in the disciplines associated with these domains, according to these bibliometric databases, is much lower than that of the in natural and formal sciences combined, and because the former have different publication standards from those prevailing in the latter, which favor a multiplication of scientific publications within these domains. Faced with this conclusion, we can retort that sociology has been interested in AI-related tools since the late 1980s, especially in the domain of simulation of artificial societies, also called agent-based modelling (Bainbridge et al., 1994; Carley, 1996). More recently, powerful LLMs for translation and speech recognition have incomparably accelerated the transcription of interviews or the textual analysis of large corpus of interviews or documents (Cointet and Parasie, 2018; Do et al., 2022). Finally, AI has also contributed to the augmentation of the humanities, giving rise in particular to the field of *digital humanities*.

occupy the highest positions in the hierarchy of science and who continue to support the traditional paradigms of knowledge creation within their field¹³ (Barber, 1961; Foster et al., 2015; Shinn, 1980).

However, although these factors, especially the last one, seem convincing in explaining the diffusion of AI, this narrative, which mainly emphasizes institutional, infrastructural and collaborative dimensions, still omits the crucial detail mentioned above that interests us, namely the fact that scientists are often agents constrained within disciplinary or epistemic boundaries,¹⁴ rather than versatile ones capable of adapting to different research contexts in which AI would be prominent. In this view, we suggest that the adoption of AI by some scientists may be subject to the initial epistemic condition of the welcoming domain to which they belong, which would be intrinsically favorable to the application of AI – such as computer science and engineering. This is precisely one of the aims of this thesis, which we will explore more specifically in neuroscience and its variety of subfields (especially in Chapter 4), which are characterized by different knowledge frameworks that may be more or less suitable to apply AI.

1.4 An analytical framework: Conceiving AI as a *research-technology* in science

In a socio-historical perspective of science, Shinn and Joerges (2002) have proposed the notion of *research-technology*, also called *transverse science*, to describe the dynamics of science since the end of the 19th century, now largely based on instrumentation for experimental or empirical investigation, especially in the physical and life sciences. Within such a research regime, the production of knowledge is conditioned by an instrument that is developed in a specific research organization before being disseminated outside it (Marcovich and Shinn, 2012; Shinn and Joerges, 2002), for instance, X-rays, now widely used for medical radiography, and NMR, now used for Magnetic Resonance Imaging¹⁵ (MRI) and spectroscopy for material studies. According to these authors, a research-technology is defined by three criteria, namely 1) an *interstitial* socio-institutional arena, 2) *metrol-ogy*, here referred to as a common language and vocabulary to grasp the instrument and its actions, as well as the means by which its performances must be measured and con-

¹³This view, which stipulates a competition between the attempt to adopt a novel technology and the resistance to adoption, comes from a long lineage of innovation diffusion studies within econometrics. For more details, the reader is invited to refer to (Arthur, 1989; Brancheau and Wetherbe, 1990; Robertson, 1967; Rogers, 1983).

¹⁴This aspect could, nonetheless, be encoded in the variables of collaboration established by Bianchini et al. (2023) in their paper.

¹⁵This process allows observations of the activation of different areas of the brain and types of neurons during a cognitive task, such as speaking, writing, remembering or logical reasoning.

trolled (such as the notions of *benchmark*, *time complexity*, *representation*, *training data*, in the case of machine learning, among others), and 3) *genericity*.

Such an instrument requires the contribution of a dedicated socio-cognitive workforce composed of a variety of actors (scientists, technicians, promoters, administrators, etc.) in order to develop a dedicated technological culture of it, delineated by a set of associated knowledge and practices shared by everyone in the community. These shape a social group that could also be defined as an *epistemic community* (Haas, 1992; Roth, 2008), although the research-technology framework relaxes the socio-cognitive boundaries that are specific to such communities. According to Shinn and Joerges (2002), the actors mentioned above are indeed able to move between different research environments and across established disciplinary boundaries in order to provide their expertise to the resolution of diverse scientific problems. As a “referent” affiliation of its practitioners, the instrument gives rise to an *interstitial* arena at the frontier of traditional disciplinary boundaries, and thus to a common professional identity and related expertise, even when its practitioners are working to implant such an instrument in a given research context outside of this arena. Through this mobility of its specialists, the research arena created by the instrument can be seen as a moving entity, able to navigate towards various disciplinary, institutional, epistemic or organizational (i.e., how the scientific work is organized) contexts that would suit its development at different times.

The research-technology regime thus extends various notions that emphasize the central role of a component (a given research object, a method, a technical instrument as above, etc.) that brings together many different disciplines for a common goal, such as the *trading zone* (Galison, 1997; Grauwin et al., 2012), a *scientific platform* (Li Vigni, 2021b), or an *interdiscipline* (Frickel, 2004). In particular, the first two concepts presuppose that the actors involved in an interdisciplinary enterprise do not renounce their original disciplinary affiliation, the former being a kind of transitory path in their research career before returning to the latter, which does not favor the construction of an autonomous specialty destined to be institutionalized in the scientific field – according to the criteria given in (Chubin, 1976; Wray, 2005). In contrast, as mentioned above, the *research-technologists* (sic.) refer rather to the instrument itself, as a kind of scientific specialty that promotes both the creation of knowledge and a technological culture. Moreover, both the *interdiscipline* and the *scientific platform* suggest that such an interdisciplinary arena crystallizes disciplinary tensions, through the more or less assertive domination of a particular field over the research orientations of the enterprise. Even if the research-technology framework does not mention the possible emergence of internal tensions between the different communities gathered around the instrument, such an interstitial arena can be preferentially compared to an interdiscipline in this sense, so that its research scope may migrate from one preferred context to another over time, as mentioned above – and which, fur-

thermore, may be the consequence of a growing monopoly of one of the aforementioned communities, or a group of them, over all the others.

We also note that the central arena of the research-technology is not confined to a purely academic context, as its development could be constrained by other challenges coming from industry or policy makers showing particular interest in further use outside science. In this sense, it could also be related to the notion of technoscience given by Hackett et al. (2004), although we will not explore the openness of an instrument like AI to non-scientific arenas in this thesis.

Based on Shinn and Joerges (2002), Hentschel (2015) has suggested four stages to describe the life of a given research-technology in a historical perspective, i.e., throughout a long time period: *prehistory*, *exploration*, *optimization* and *diffusion*. The first two stages are often associated together in the design and testing phases of a given instrument. According to these three authors, the underlying process needs temporarily a closure of the community working on it, thus fostering the creation of a dedicated interdisciplinary field wherein the instrument, initially dedicated to further applications, also becomes a research object. In this context, a specific paradigm emerges to apprehend the functioning of the instrument, corresponding to the aforementioned *metrology* criterion.

At later stages in its development, the tool could be adapted to disciplinary research contexts outside the environment in which it has been originally designed, or to a variety of applications outside scientific research – Shinn and Joerges have evoked the duality *dis-embedding/re-embedding* to describe this process. The two last development steps of Hentschel thus represent such dynamics toward this final step, called the *genericity* criterion. This notion designates the diffusion of not only the instrument, but also of all its knowledge, vocabulary, imaginaries and practices around it, thus “building and promoting a sort of social, technical, and cognitive universality”¹⁶ (Shinn and Joerges, 2002, p.209). In particular, the diffusion phase requires a relaxation of its boundaries within which the instrument was designed during the prehistoric and exploratory phase, thus implying an openness of the actors of the community, who promote¹⁷ it in different fields of research throughout academia or industry.

According to the analysis of the AI research ecosystem provided by Gargiulo et al.

¹⁶Within this rhetoric of universality, here understood as the large diffusion of recent AI-related knowledge and technologies in various applicative domains, especially *deep learning*, Cockburn et al. (2018) have proposed the notion of *general-purpose method of invention* to define AI in this sense. We can also refer to (Bowker, 1993), which exposes the desire for universality expressed by the first cybernetic paradigm, notably through the establishment of a discursive arena where a new meta-language has emerged that was intended to unify all domains of science – as mathematics was and still is in the natural and formal sciences.

¹⁷In the context of synchrotron development, Söderström et al. (2022) have explored such a promoting stage in the lens of the creation of multidisciplinary journals dedicated to disseminating related instruments, such as beamlines, to a wide audience.

(2023), AI seems to embrace the previously mentioned characteristics of an instrument embedded in a global research-technology programme within science, especially when it has entered into its contemporary connectionist phase since the late 1980s and was confined, at least in their basic branch, into computer science, mathematics and statistics (see previous Sect. 1.3), before spreading throughout science and providing methodological support to various fields of research, thus suggesting a certain genericity. This thesis also considers AI as such, focusing on the last aforementioned stage of its development, which seems to occur recently. It especially intends to discuss the insertion of AI into a single field of research, here neuroscience, and to what extent and by what mechanisms it penetrates the social practices and epistemic frameworks of this domain. Along these lines, this thesis will also discuss the limitations of such a framework to describe AI.

A journey in neuroscience: collection and preprocessing of quantitative and qualitative data

“The Data—It Is Me!” (“Les données—c’est Moi!”)

Ronald E. Day, in Cronin and Sugimoto (2014)

As detailed in Introduction, we have selected neuroscience as a case study to apply our genericity model for describing the integration of AI-related knowledge and technical tools into a field of research. Although we are not experts in this field and have very little knowledge of it, this thesis also sets out to describe our discovery of this field of research and the gradual construction of a specific culture around it in the course of our investigation. Following a traditional approach in the sociology of science (and in science studies in general), this manuscript also attempts to convey to the reader, as faithfully as possible, many aspects of neuroscience through several types of data, which constitute broad observation windows on this moving field, in which AI may have influence. As part of this endeavour, this particular chapter focuses on the production of these research materials, in particular how they have been collected and processed before being analyzed.

After a brief introduction to neuroscience and their interaction with AI, in Sect. 2.1, the second main Sect. 2.2 describes the step-by-step construction of a consistent corpus of neuroscience publications, which will be subjected to a scientometric analysis in the following two chapters. Sect. 2.2.1 presents the bibliometric database from which we extract the corpus, namely the Microsoft Academic Knowledge Graph (MAG), the method of extraction of the elements of the corpus and, finally, the collected metadata of interest. In particular, we use the titles and abstracts of the articles to immediately identify those that mention AI-related expressions, with a keyword matching process detailed in Sect. 2.2.2. Then, from this corpus, we manually build two relational structures and additional databases that encode various types of knowledge about the field under study, such as the underlying co-authorship network, which approaches the social organization of the

scientific community working on the advancement of the field, and the citation network between the extracted articles and beyond, which allows to study the patterns of knowledge transmission across this community. Their respective constructions are depicted in Sect. 2.2.3, followed by a presentation in Sect. 2.2.4 of various filtering and cleaning steps that could affect these networks, in particular the distribution of the citations and collaborations inherited from AI-related works. The final dataset is summarized and discussed in Sect. 2.2.5.

In a final Sect. 2.3, in order to qualitatively analyze AI work in an academic laboratory, which intends to provide supplementary insights into the genericity criteria established in the Introduction (especially domain adaptation and social adoption) and whose main results are exposed in Chapter 5, we introduce the fieldwork we have chosen, namely the Aramis team (in Sect. 2.3.1), and the methodology we have adopted to interview some of its members and to access the information needed for our analysis (in Sect. 2.3.2). We have specifically relied on a questionnaire that explored, among other aspects, the professional careers of these scientists, their daily use of AI, and their definition and perception of this research instrument at work (see App. D.1 for more details).

2.1 A multidisciplinary domain interacting with AI

Neuroscience encompasses all the scientific work carried out to understand the human nervous system and its various types of dysfunctions and pathologies, which today, unfortunately, are often incurable these days. The nervous system comprises two regions, the *central*, which includes the brain and spinal cord, and the *peripheral*, composed of the nerves and ganglia inside the last two organs, allowing information and stimuli to circulate between them and the other organs of the human body. The first region remains the most studied within the field, mainly because it is where most of the triggers for numerous disabling pathologies are observed, including neurodegenerative diseases such as Parkinson's, Alzheimer's and Huntington's,¹ multiple sclerosis, depression, attention deficit or obsessive-compulsive disorders and other traumas.² Because of its intrinsic proximity to the world of healthcare, neuroscience brings together basic and clinical research whose results are then deployed (with precise protocols) in the regular medical practice with the help of an intermediate phase, the translational research (Stokols et al., 2008; see also the case of AI and neuroimaging in Allen et al., 2019).

In recent days, neuroscience claims to be multidisciplinary. Previously a domain solely

¹Huntington's disease differs from Parkinson's and Alzheimer's in that its initiating factors are mainly of genetic kind.

²The following website provides a classification of all existing referenced neurological disorders: <https://www.merckmanuals.com/home/brain,-spinal-cord,-and-nerve-disorders> (visited on July, 14 2024).

concerned with biology and anatomical medicine, neuroscience has broadened its field of investigation thanks to scientific and technological advances in other disciplines, notably chemistry, physics, mathematics, and computer science, which have led to the discovery of fine mechanisms at various scales within the central nervous system – molecular, cellular, neuronal, neural networks, brain areas, the organ itself, to the entire body. Magnetic Resonance Imaging (MRI), Positron Emission Tomography³ (PET), electroencephalography⁴ (EEG), or recent optogenetics,⁵ are notable examples. And let us not forget the increasing prevalence of AI, which aims to simulate and even reproduce the brain’s mental processes.

Although its foundations are mainly associated with STEM disciplines, mainly mathematics, statistics, and computer science (Gargiulo, Fontaine, et al., 2023), AI is constantly evolving alongside neuroscience, both maintaining a continuous cycle of mutual improvement (Hassabis et al., 2017). Indeed, neuroscience brings in a first place empirical confirmation to some theoretical models that reproduce parts of mental processes, and that were first imagined, analytically derived, and computationally simulated by (neuro)psychologists within the field of cognitive science (Cooper and Shallice, 2010; Lake et al., 2017). Most of these models serve as the foundation for some AI-related algorithms, especially in the case of bio-inspired AIs such as artificial neural networks and their numerous versions, which have become *biologically plausible* thanks to neuroscience. In particular, the deployment of the first body scanners based on PET and functional MRI (fMRI) in the 1980s and early 1990s led to important discoveries about the functional biological mechanisms in the human brain that are induced by complex cognitive tasks (Cooper and Shallice, 2010). This contributed to the rise of the connectionist paradigm that is now dominant in the brain sciences, although some debates still persist within cognitive science about the representations of knowledge and the logical operations mobilized to process them in interaction with the real world (Andler, 1990; McCarthy, 1981; Perconti and Plebe, 2020).

More than an assistive technology, AI is thus a powerful tool for understanding and reproducing our intelligence, from its physical and biological workings in the central nervous system to the production of specific motions and behaviours in a given situation, through complex information transmission mechanisms activated by stimuli inside or outside the body. In the recent era of *big data*, a panel of AI tools now allows efficient processing of

³Similar to MRI, PET is a technique that draws upon another nuclear process to observe the metabolic activity of an organ, particularly its glucose production. It is widely used in neuroimaging to observe metabolic failures in the brain, which are partly responsible for the onset of neurodegenerative diseases.

⁴Like the electrocardiogram, the electroencephalogram measures the electrical activity of the brain. It is an alternative way of observing, through a signal, the different areas of activation in the brain during a task or in response to an electrical stimulus in order to test movement reflexes.

⁵Optogenetics is a technique of light stimulation of neurons to study their activity and functioning in the brain.

large datasets composed of various kinds of biomedical data (electroencephalogram, MRI, biomarker tracking, movement recordings, psychological surveys, etc.) acquired from major clinical trials and cohorts for studies of brain damage (Gopinath, 2023; Macpherson et al., 2021). These tools are particularly part of the improvement of the diagnosis of various neurodegenerative diseases and of the attribution of potential dedicated treatments, if they exist.

This virtual feedback loop, which is well documented in the neuroscience literature, is thus inducing a reinforcement dynamics of both AI and the neuroscience that receives it. From this assertion, AI and neuroscience are expanding each other's *adjacent possible* (Kauffman, 2000; Monechi et al., 2017), i.e., one domain expands the field of possibilities that have yet to be explored in the other by blending with the pre-existing knowledge and practices that characterize the latter. According to this model, one domain thus reshapes the knowledge space of the other.

2.2 Extraction of a neuroscientific corpus and construction of data structures

2.2.1 Building of a first exhaustive database of neuroscience articles from the MAG

2.2.1.1 The Microsoft Academic Knowledge Graph

The MAG is a bibliometric database that was provided freely by Microsoft and available through an API until the end of 2021, and that covers various kinds of scientific publications, namely journal articles, conference proceedings, patents, datasets and books. In 2020, this database has referenced approximately 240 million scientific publications of these types, in all disciplines, grabbed throughout the entire Semantic Web, thus making it much larger and more comprehensive than the traditional databases used for bibliometric purposes and research evaluation, notably the *Web of Science* (WOS) and *Scopus* (Visser et al., 2021).

Thanks to the openness of these data, some scientists and engineers have downloaded and cleaned the entire MAG database at different periods, in order to share it freely as complete data dumps in dedicated repositories. The most popular snapshots are distributed by Färber (2019) on the Zenodo⁶ platform with four versions between 2018 and

⁶The deposit is located at the following address: <https://zenodo.org/records/4617285>.

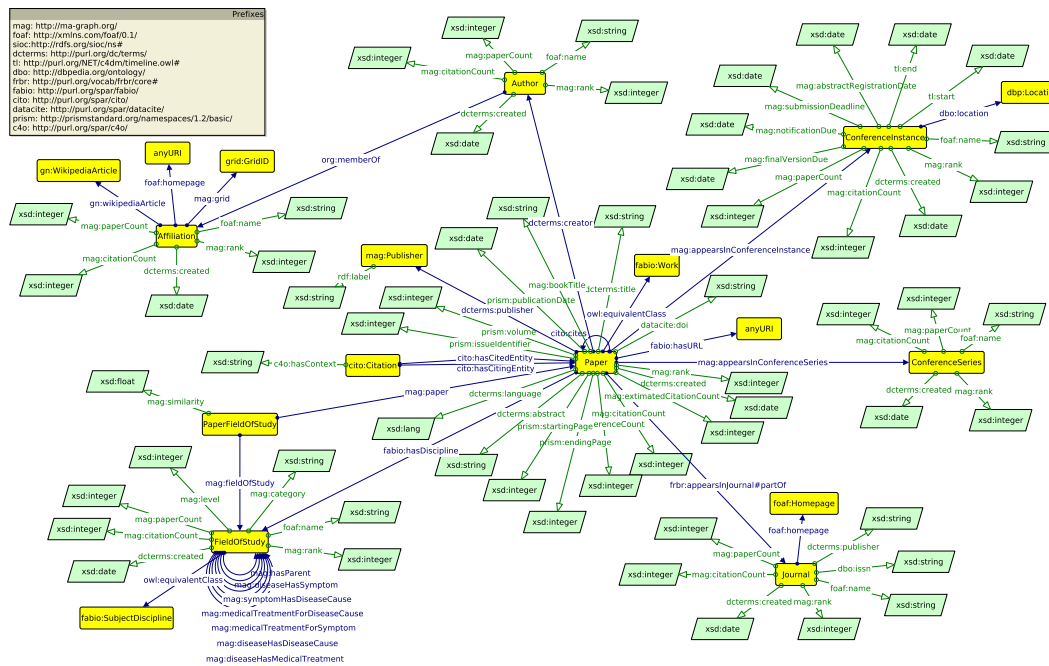


Figure 2.1: Organization of the metadata associated with one single paper referenced in the MAG. Source: Färber (2019).

2021. In what follows, we use the version called 2020-05-29 on this website, completed with some supplementary files provided in the next version 2020-06-19, which were missing in the former. This database is the underlying basis of the recent OpenAlex platform,⁷ an API⁸ launched in the very beginning of 2022 and maintained by the non-profit association OurResearch (Priem et al., 2022). As this platform has undergone multiple and simultaneous modifications since its launch – such as the addition of publications, large disambiguation operations on authors, progressive replacement of the *Fields of Study* taxonomy of articles by a new one around *Concepts* or *Topics*, etc. – we will use this API occasionally, depending on the needs identified during the thesis, in particular to verify the consistency of some metadata provided by the MAG database.

As illustrated in Fig. 2.1, the MAG is organized as a network of objects, represented with yellow rectangles, that are associated with a large set of various scientific contributions. Here we detail only those that are mobilized throughout this thesis.

The basic unit of this knowledge graph is the *Paper*, representing a scientific textual output described by 24 variables (identified with outgoing green and blue arrows from the considered rectangle), including *title*, *abstract*, *citationCount* and *publicationDate*. One element described by such a unit is published either in a *Journal* or in a conference

⁷More details are available at the following address: <https://help.openalex.org/>.

⁸Application Programming Interface. Often provided by the database maintainer (an organization or a single person), this service allows any user, through a dedicated website or direct programming on their personal computer, to query some features of that database stored on another computer or on a shared remote server. These features are then sent by the latter to the user users on their own machines.

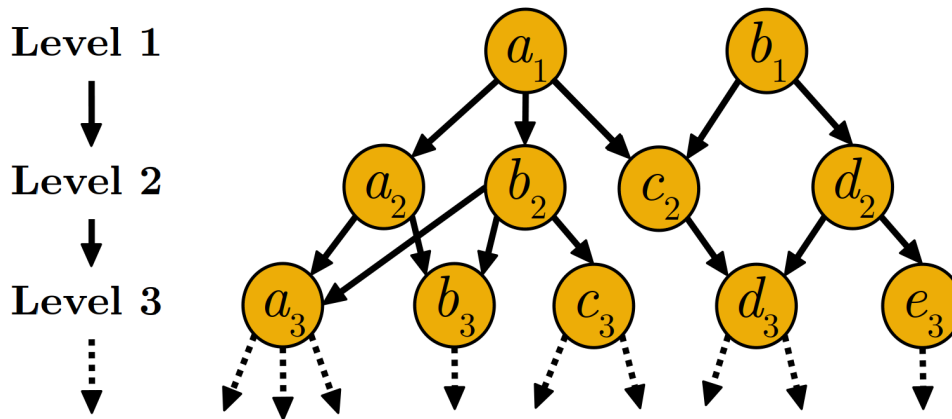


Figure 2.2: Schematic representation of the MAG’s **Fields of Study** network, here truncated at the first three levels and with two origins, namely the concepts a_1 and b_1 , which then include the concepts at level 2, including the following at level 3, and so on.

proceeding, the latter being referenced by two objects, namely the **ConferenceSeries** (the editor or collection that publishes the contribution as a paper) and the **ConferenceInstance** (summarizing mainly the name, date, and place of the physical conference where the contribution was presented).

All the papers referenced in the MAG are labeled with one or several **FieldsOfStudy** – analogously called *concepts* throughout this thesis. These fields of study, derived with Natural Language Processing (NLP) and Machine Learning (ML) techniques, are organized in the MAG as a hierarchical directed acyclic graph (DAG), as shown in Fig. 2.2. More specifically, as indicated in this figure, this hierarchy relies on the *level* parameter. This last discrete variable, comprised between 0 and 5, indicates the degree of specificity of a given field of study within science, from the most general, such as discipline or sub-discipline levels – values 0 and 1 respectively – to the most precise, such as research subjects or keywords extracted from abstracts – values 4 and 5 respectively. Such fields of study are thus proxies to circumscribe vast or small knowledge domains.

For the special needs of interpreting the neuroscientific knowledge map shown in Chapter 4, whose construction is more broadly detailed in App. C.1, we complete this thematic classification with another one launched in early 2024 by OpenAlex, according to which each paper in the database is labeled with a *primary topic* and two others (sorted by a matching score), which are embedded into **Topic** objects. These topics are leaves of a four-level hierarchical classification that is actually a tree:⁹ Level 0 is composed of *domains*, each of which is divided into *fields* at level 1, which are subdivided into *subfields* at level 2, which contain a given set of *topics* at level 3. More robust than the MAG’s fields of study, this topical taxonomy will be used to confirm some of the results initially obtained with the former. In particular, on 30 June 2024, we have performed an extraction from

⁹See also: <https://docs.openalex.org/api-entities/topics>.

OpenAlex of such topics associated with the papers in our corpus, and we conserve only the *primary* one for each of them – with their own respective subfield, field, and domain.

Finally, each paper in the MAG has a list of authors whose characteristics are embedded in the **Author** object, such as their respective **Affiliation(s)**, which constitute another distinct object of the database.

Each of these entities is described by its own set of metadata in separate files written in RDF¹⁰ format to avoid the accumulation of complex metadata in a single main file. These files are represented in Fig. 2.1 by a yellow rectangle from which the names of the attached variables emanate with outgoing green arrows. The Zenodo data dump used also provides some correspondence files to link the identifiers associated with two distinct objects when the combination of the latter into a single file could occupy an important memory space – represented with directed outgoing blue arrows between the main objects in the aforementioned figure. For example, to reach the fields of study labelling a given paper, we have to navigate through a link file called `PaperFieldsOfStudy.nt` in order to obtain the identifiers of these fields and retrieve their characteristics encoded in the variables stored in the file of the object `FieldsOfStudy`. This organization is suitable for analyzing different objects separately and for reconstructing datasets by mixing only some of them, according to the user’s appreciation.

2.2.1.2 Extraction of neuroscience articles

In order to extract a coherent set of neuroscience publications from the MAG, we choose to start from a list of peer-reviewed journals that are representative of the field under study. As essential vectors for communicating new discoveries, journals also constitute an arena for establishing and consolidating the existence of a community within the entire scientific landscape (Chubin, 1976), which is organized around shared knowledge, methods, practices and values (what Cole and Zuckerman (1975) called the *cognitive consensus*). Among a variety of approaches,¹¹ such as keyword matching within titles and

¹⁰ *Resource Description Format*, which is useful for representing knowledge networks from the web.

¹¹ In addition to journal-oriented extraction, several alternative methods exist for building a bibliometric corpus representing a delimited research domain. A first type would be keyword-oriented. With the help of a comprehensive list of keywords that summarize the main concepts around a specific field, we can access an exhaustive corpus of papers that match some keywords or expressions in their respective title, abstract or main text (Ying Huang et al., 2015; N. Liu et al., 2021). Such an uncontrolled procedure has been used by us and Gargiulo et al. (2023) to map AI research since the 1980’s. However, this technique alone is not sufficient to get the structure of the field we want to study. Indeed some keywords would be too generic and would therefore appear in many other research fields with different scientific communities and epistemic frameworks. This approach thus requires supplementary filters to select the papers of interest, as in the method introduced by Rimbault et al. (2016), who combined both the keyword queries and the delimitation of the topics they studied with the concept taxonomy of the MAG. These approaches are thus suitable for bounding very precise subfields with few publications, but they could not be applied to describe large fields with blurred boundaries or with a multidisciplinary character – such as

abstracts, and the selection of the field of study of interest within the MAG (with some limitations), this extraction technique remains rather common among bibliometricians and sociologists to delimit well-established scientific fields. More precisely, we use the disciplinary taxonomies¹² provided by the traditional WOS database and the *SCImago Journal Rank* (SJR), which both categorize only the journals and conference venues they reference. Each source is labeled with one or two Journal Subject Categories (JSC), which are broad, well-recognized disciplines in science – the combination of two of them for a given journal indicates its interdisciplinary positioning. In the following, we therefore assume that the papers published in these referenced journals directly inherit the JSC(s) of the latter.

More practically, we use the SJR in addition to the WOS for many reasons. First, they are both freely available on the Web (a crucial technical criterion!). Second, the former provides another measure of impact, also called SJR, which enhances the visibility of some journals that may have a lower impact factor within the WOS. Third, because of the different disciplinary classifications provided by the two databases, the SJR covers more neuroscience journals than the WOS. Indeed, in 2021, the WOS referenced only 281 journals labeled as “neuroscience” while the SJR referenced 608 journals labeled as such. We consider here the neuroscience journals referenced by the WOS and other neuroscience journals referenced as such by the SJR but labeled differently by the WOS. For example, the well-recognized journal *Neurocomputing*, while affiliated with both “neuroscience” and “computer science” in the SJR, is labeled only as “computer science” by the WOS. Using the SJR thus contributes to increasing the diversity of neuroscience-related subjects treated by the articles in the final dataset under study. For the sake of consistency, we lie only on the WOS disciplinary classification afterwards.

Then, we extract the desired publications from the MAG with an ISSN identifier match-

neuroscience. A second type of approach would be based on established disciplinary or research subjects classifications, such as those provided by the main bibliometric databases. Although in this thesis we use the journal classification of the WOS and SJR, we recall that the MAG has its own classification of fields of study, which only to individual papers (as mentioned in the previous subsection), thus allowing a direct extraction of the articles labeled with the field “Neuroscience”, situated at level 1 in the MAG field hierarchy. However, within the MAG, an article can be labeled with an undefined number of fields of study (from 0 to 10), so that its label “Neuroscience” could be surrounded by others unrelated to it, thus challenging its affiliation to this precise field. To be sure, we have to manually check the content of its title, abstract or journal, whose scope might be far away from neuroscience. This field-oriented approach thus risks compromising the coherence of the domain we wish to circumscribe. Moreover, such a query of the MAG would disproportionately increase the number of publications, possibly including irrelevant ones, and thus the size (and noise) of the citation and collaboration networks described in the main text.

¹²We remind that these bibliometric classifications are often criticized, notably the WOS for its limited regional and linguistic coverage, the low visibility it confers to the social sciences and humanities (Mongeon and Paul-Hus, 2016; Visser et al., 2021), and the occurrence of some discrepancies between the actual scope of journals and the fields or disciplines attributed to them, based on some of the knowledge structures built on it, such as its journal-to-journal citation network (Leydesdorff, 2006; Leydesdorff and Rafols, 2009), hence the production of diverse alternative classifications, mostly at the level of articles instead of journals (Glänzel and Schubert, 2003; Milojević, 2020).

ing procedure based on this list of journals. In the following, we only keep the papers published in the period running from 1970 until 2020, which results in a dataset of 2,052,292 publications disseminated in 479 journals, recalled in App. A.2. We denote this set by $\mathcal{P}_{original}$.

2.2.2 Detecting mentions of AI in selected publications

In a recent paper, we and Gargiulo et al. (2023) have used a list of 594 keywords to identify AI-related papers within the MAG, in order to draw a cartography of the nature of contemporary AI since the 1980s and to examine the mechanisms of its diffusion at various scales across a broad disciplinary landscape. This list, provided in their supplementary material and appended in App. A.1, is based on the compilation of multiple glossaries available on the Web,^{13,14,15,16} which were compared with other similar lists provided in expert reports edited by various organizations (Baruffaldi et al., 2020, pp.66-67; Hajkowicz et al., 2022, p.57; WIPO, 2019, pp.146-149). It also represents a particularly rich semantic landscape on AI, which considers most of its forms, knowledge, and technological applications throughout its contemporary history since the early 1950s, when the AI research programme was initiated.

We use this list of AI-related keywords to distinguish AI-related works from others within $\mathcal{P}_{original}$. In particular, we apply a selection criterion such that the neuroscience works dealing with AI must include at least one AI-related keyword in their respective title or abstract (or both at the same time). We count 52,056 AI-related papers in $\mathcal{P}_{original}$, which represent 2.5% of the dataset.

We notice, however, that this method faces many limitations. First, it misses the context of the use of AI expressed in the neuroscience articles, which could concern the application of a software, the design of a piece of code or a robot, the discussion of a method or its critics, a methodological review, among others. Nevertheless, we believe that the mere mention of AI in the title and/or abstract of these articles is a good proxy of the presence of AI within the domain under study, regardless of the context of its mention. Indeed, as long as an AI-related term figures explicitly in an article, its authors are thus demonstrating their interest in AI.

Second, some keywords in this list could be too generic or not semantically related to AI. By using the recent topic classification of scientific publications provided by OpenAlex,

¹³https://en.wikipedia.org/wiki/Glossary_of_artificial_intelligence

¹⁴<https://developers.google.com/machine-learning/glossary>

¹⁵<https://machinelearning.wtf/>

¹⁶<https://link.springer.com/content/pdf/bbm:978-3-319-94878-2/1.pdf>

which has been produced with a deep learning classifier lying on a multilingual BERT (OurResearch, 2024), we find that 82% of the AI-related papers of our dataset are not categorized with a topic inherited to the *subfield* of AI, although they actually mention keywords that are intimately linked to famous accomplishments of AI research such as, in primary positions, the expert system *Mycin*, *neural networks* and the first convolutional neural network *Lenet*. However, this proportion decreases to 50% when using MAG's classification of field of study at levels 1, 2 and 3 only, with AI-related fields defined as concepts closely related to *Artificial Intelligence* or *Machine Learning* in MAG's concept network, both located at level 1. This large gap, we believe, could be due both to the large number of AI-related concepts in MAG (4902) compared with the number of AI-related topics in OpenAlex (77), and to the number of possible concepts allowed by MAG to label articles, which is unlimited at first glance, whereas OpenAlex labels publications with a maximum of 3 topics. These reasons would therefore allow a greater number of articles to be labeled as AI with the MAG's classification than with OpenAlex's, which let us suggest that the former offers a broader delimitation of the AI domain than the latter.

And third, such a method may exclude some publications that actually deal with an AI-related research topic or use an AI-related method without explicitly mentioning a single keyword from the aforementioned list in their title or abstract. Still using OpenAlex's classification, we find that 0.6% of non-AI-related articles in our dataset are categorized with a topic inherited from the AI *subfield*, whereas this proportion rises to 5% when we use the MAG one. This partly confirms the reasons mentioned above for the discrepancies between MAG and OpenAlex.

Although we have chosen the keyword matching method, while not unique, the list still needs to be validated effectively, notably through the intervention of human expertise in order to check the actual semantic proximity of each keyword to AI, which often requires meticulous manual annotation (which would be very time-consuming for a very large corpus such as ours).

2.2.3 Selected data structures

2.2.3.1 Capturing the disciplinary landscape through an egocentric citation ecosystem

One of the main goals of this thesis is to reconstruct the disciplinary environment of the whole neuroscience, and another – included in the previous one – that surrounds the particular contributions involving AI within this field of research. To do this, we need,

on the one hand, the bibliographic references of the papers, as they constitute a proxy¹⁷ for the influence mobilized by the field under study – they also testify to the cognitive structure of the latter (Cole and Zuckerman, 1975; Griffith, H. Small, et al., 1974; H. Small and Griffith, 1974; Sugimoto and Weingart, 2015) – and, on the other hand, the citations accumulated by them, since they indicate the direct use of their related outputs by other research, within or outside the neurosciences.

After gathering the neuroscience dataset $\mathcal{P}_{original}$ described in the previous section, we build an exhaustive citation ecosystem around the neuroscience publications within it, as illustrated in Fig. 2.3 for a single paper (represented by a gray circle). We use the general reference link file of the MAG, a giant temporal, directed, unweighted citation network in order to retrieve the direct references and citations related to our identified papers. The latter are represented by the red and blue circles, respectively, in Fig. 2.3. The latter are represented in Fig. 2.1 by the object `cito:Citation` and the three blue directed relations `cito:hasCitedEntity`, `cito:hasCitingEntity` and `cito:cites`. Since some of these references and citations are outside the set $\mathcal{P}_{original}$, we extract the metadata of these missing papers, which join a set called $\bar{\mathcal{P}}$ in the citation scheme in Fig. 2.3.

We then enrich this other extracted ensemble of papers by assigning to them a set of JSCs from the WOS disciplinary journal classification, which are associated with their journals if they are referenced in this database. The papers without an associated JSC are thus automatically discarded, on the one hand for the sake of consistency by using at most the WOS disciplinary taxonomy, and on the other hand to discard papers considered as noise with little or no information about their publisher or their disciplinary affiliation.

Our final citation network therefore contains 120,018,653 directed edges, with 44,555,248 of them being circumscribed within $\mathcal{P}_{original}$ – representing 37% of all these edges.

¹⁷We recall here that the signification of a citation given by one scholarly paper to another remains a major case study throughout the histories of the sociology of science (Bornmann and Daniel, 2008) and scientometrics (Leydesdorff, 1998). Indeed, citations are not only a reference to a discovery that the authors find interesting for their own research purposes (H. G. Small, 1978), but also a tool to increase the symbolic credit of the scientists associated with the cited works, whose main translations are their recognition in the field in which they work, the capacity to obtain research grants and other sources of funding, among other various rewards (Bourdieu, 1975; Merton, 1968). More recently, in the context of research evaluation, the citations received by scientists have become central in measuring their individual performance and ranking laboratories, universities, and other research organizations in order to target and prioritize financing to the most “promising” ones (Cronin and Sugimoto, 2014; Gingras, 2016). In this thesis, we take the citation as a direct intellectual influence for an article or a direct impact on another. Considering a citation as such thus enables us to assess how AI-related knowledge is transmitted throughout the neuroscience literature.

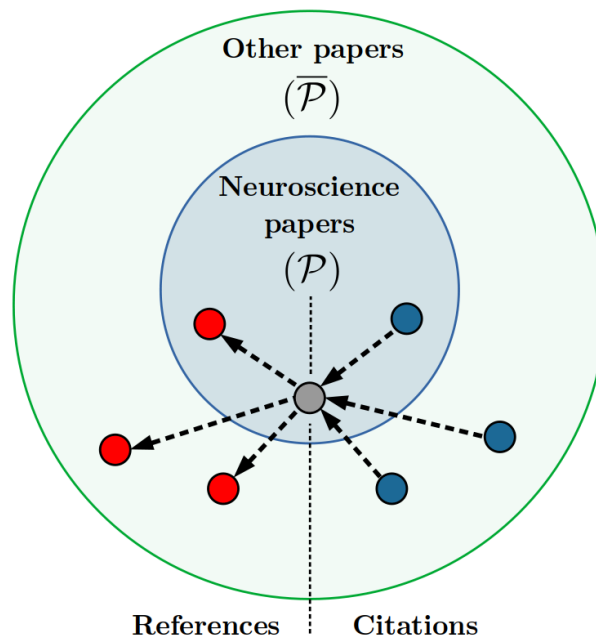


Figure 2.3: Schematic representation of the citation network, centered on the set of neuroscientific papers, here called \mathcal{P} in the blue area. The subset $\bar{\mathcal{P}}$ in the green area is the subset of additional extracted papers that cite or are cited by \mathcal{P} but do not belong to the latter. The small circles represent papers, where the gray one is a *focal* one belonging to \mathcal{P} , which cites the red ones (*references*) and is cited by the blue ones (*citations*). These references and citations belong to either \mathcal{P} or $\bar{\mathcal{P}}$.

2.2.3.2 Building of the collaboration network and a dataset of neuroscientists

In Chapter 3, we will explore the social composition of neuroscience through the profiles of its scientists and the underlying collaborative structure captured by co-signatures within the papers (Newman, 2001). These elements will allow us to identify the potential specificities of the actors involved in some neuroscience subfields where AI is involved, as well as how these actors drive (or not) the diffusion of the technology within this social space. In this section, we detail the construction of these data structures that will be used later in this thesis, which is illustrated in Fig. 2.4.

First, we build a temporal co-authorship network \mathcal{T} from the papers in $\mathcal{P}_{original}$. Two authors i and j publishing together in year t are connected by an edge weighted by the number of joint publications in that year, denoted $w_{ij}(t)$. All co-signatures in all papers published in year t form a weighted snapshot called $\mathcal{T}(t)$. To study the structure of the collaborations at the macroscale over the entire period 1970-2019, we also build the weighted time-aggregated co-signature network $\mathcal{A}_{original}$, which includes the set V of all authors appearing in the dataset, and a set of weighted edges obtained by summing all the weights of the edges appearing in each snapshot of \mathcal{T} , $E = \{\tilde{w}_{ij}, \forall (i, j) \in V^2 | \tilde{w}_{ij} = \sum_t w_{ij}(t)\}$. $\mathcal{A}_{original}$ thus includes 2,611,456 authors, of whom 68,759 (2.7%) have published alone and without any collaboration during the period studied. We discard these outsiders in the

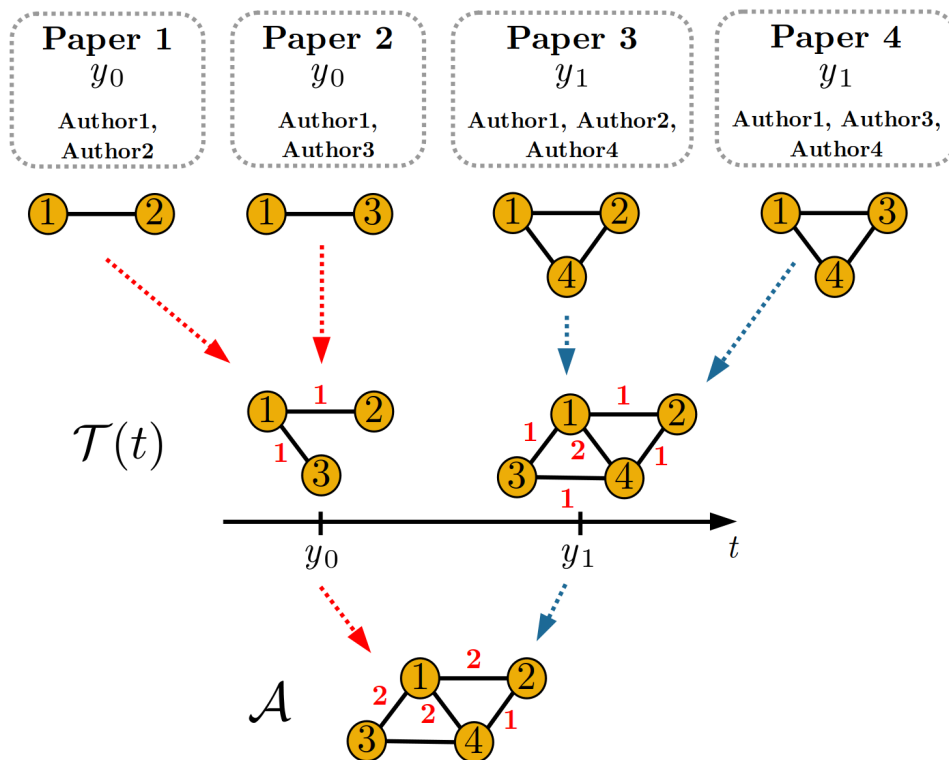


Figure 2.4: Schematic representation of the construction of the temporal collaboration network \mathcal{T} and its time-aggregated version \mathcal{A} , based on 4 articles published in years y_0 or y_1 . The weights of the edges in \mathcal{T} and \mathcal{A} are indicated by red numbers alongside these edges.

following, as their isolation renders them irrelevant to a potential social diffusion process of AI-related knowledge and technologies across the neuroscience community.

Second, we carry out a further extraction from the MAG to gather all the publications of the aforementioned authors, in order to establish several indicators allowing to assess their respective profiles. We especially retain, for a given author:

1. The year of his/her very first publication(s), as a proxy for the beginning of his/her academic career, denoted as y_0 ,
2. The JSCs associated with the journals in which the articles in y_0 were published, which represent the disciplinary fields in which the author was trained. We call these JSCs the *disciplinary background*,
3. The year of his/her last publication(s) before 2020, as a proxy for the end of his/her academic career – or his/her continuation if it is 2020 – denoted as y_f ,
4. The JSCs associated with the journals in which the articles have been published throughout his/her whole career (until 2020), which represent his/her global *disciplinary profile*,
5. The total number of publications published between y_0 and y_f , representing his/her productivity,

6. His/her share of AI publications between y_0 and y_f , which represents his/her level of affinity with AI.

In particular, indicators 2 and 4 will be used in Chapter 3 to build the disciplinary trajectories of the authors in our dataset, which will be compared with their respective levels of affinity with AI in neuroscience (indicator 6).

2.2.4 Selection of a subset of the extracted neuroscience database

A main issue encountered in our study is the huge size of our datasets – especially the citation network – which limits the memory usage of our computers when performing some heavy numerical operations on them. In the following chapters of this thesis, we will base all our analysis on a particular core of publications in the neuroscience corpus $\mathcal{P}_{original}$ that meet the criterion below, denoted $C1$:

Criterion selection C1

The selected articles must have at least 10 bibliographic references and at least 10 generated citations since their publication year (until 2020).

These papers join a subset called \mathcal{P}_{C1} in the following lines. When selecting this subset, we thus discard many papers and all their associated features that are part of the citation network and the collaboration network presented in the previous sections, such as the authors who are not present in the papers belonging to \mathcal{P}_{C1} , some papers that are only cited by the discarded ones, and the papers that only cite the latter. These discarded features could therefore impact the shape of these networks.

In this section, we introduce a comparison of the time-aggregated citation network and the time-aggregated collaboration network inherited from $\mathcal{P}_{original}$ and \mathcal{P}_{C1} , respectively, for two purposes. First, we evaluate how similar the topologies of these networks are before ($\mathcal{P}_{original}$) and after (\mathcal{P}_{C1}) the application of the selection criterion $C1$. Second, since the number of AI-related papers within $\mathcal{P}_{original}$ is quite small, as already mentioned in Sect. 2.2.2, we evaluate how much the main regions of these networks containing AI-related publications are affected by this selection.

Our analysis is based on the comparison of simple global indicators describing the different networks (original and filtered), although we are conscious that more advanced methods exist to control the multiscale similarity of the topologies of the compared networks

under study (Tantardini et al., 2019).¹⁸ As the types of our citation and collaboration networks are different – directed and unweighted for the former, undirected and weighted for the latter – we detail in the next two sections the specific indicators we use to compare the topologies of these two networks during filtering. Based on the new dataset \mathcal{P}_{C1} , we also introduce a special filter to select the profiles of the authors involved in the papers within this set, thus affecting the structure of the collaboration network a second time.

2.2.4.1 Impact on the citation network

Here, we report whether some global properties of the time-aggregated citation network are preserved before and after the selection of papers with at least 10 bibliographic references and 10 citations. Tab. 2.1 summarizes some descriptive metrics associated with the networks built, respectively, from $\mathcal{P}_{original}$ (before selection) and \mathcal{P}_{C1} (after selection). In addition to the share of AI-related papers and citation links related to them, we include the *transitivity* of the networks, here the average of the *clustering coefficients* of all articles within the undirected versions of the citation networks. More precisely, the clustering coefficient of a paper is defined as the proportion of *triangles* (a set of three fully connected nodes) actually observed among all possible *triads* in which the node is involved, thus including neighbors that are not connected to each other (Menczer et al., 2020, pp.55-58). With this metric indicating the structure of the local neighborhood of a paper, transitivity therefore indicates the probability for one paper to be connected to another one knowing that both have a common neighbor.

We also plot in Fig. 2.5 the total degree distribution of all papers within each of the two citation networks. We define the total degree k as the sum of the number of references of an individual paper, indicated as the *out-degree* k_{out} , and the number of citations received, indicated as the *in-degree* k_{in} – as indicated by the directions of the citation arrows in Fig. 2.3. Notice that the criterion $C1$ established earlier, that is to say $k_{in} \geq 10$ and $k_{out} \geq 10$ for a given paper within $\mathcal{P}_{original}$ is not equivalent to a k -core decomposition, i.e., the incremental removal of all nodes with degree k in order to conserve only the subgraph with nodes exhibiting a minimal degree $k + 1$. Indeed, as removing papers not satisfying $C1$, we also remove some edges related to the papers satisfying effectively $C1$ within the original citation network, so that the final total degree of the latter within the filtered network may be less than 20. Hence the left bound of the distribution of Fig. 2.5 situated at 0.

¹⁸The methods introduced by Tantardini et al. (2019), if they are freely available and easy to implement with Python, exhibit some difficulties in scaling up very large networks like ours, in particular for *NetLSD* (Tsitsulin et al., 2018) and the computation of the *Network Portrait Divergence* (Bagrow and Bollt, 2019), among those tested. In addition, some of these methods are not generalizable to all types of networks – directed, weighted or a combination of the two. Thus, we limit our topological analysis to common metrics that are easily computable on the huge networks we possess.

	N_0	$N_{0,AI}$	N_1	E_0	$E_{0,AI}$
$\mathcal{P}_{original}$	2,052,292	52,056	11,236,818	97,286,873	3,805,467
\mathcal{P}_{C1}	855,691	26,374	10,157,654	77,899,796	2,852,268
Loss	-58%	-49%	-9.6%	-20%	-25%

	$N_{0,AI}/N_0$	$E_{0,AI}/E_0$	Transitivity
$\mathcal{P}_{original}$	0,025	0,039	0,075
\mathcal{P}_{C1}	0,031	0,037	0,089

Table 2.1: Comparison of the egocentric citation networks around the original database $\mathcal{P}_{original}$ and the filtered one \mathcal{P}_{C1} . The *Loss* row is the fraction of items removed from the set X , ie. $Loss = (X_{filtered} - X_{original})/X_{original}$. Papers without citation between 1970 and 2020 are discarded. Transitivity refers to the global clustering coefficient of undirected versions of the citation networks. **Notations:** N_0 : Number of neuroscience papers in either $\mathcal{P}_{original}$ or in \mathcal{P}_{C1} , which are the nodes of the citation networks constructed from these sets; $N_{0,AI}$: Number of papers mentioning AI in their title and/or abstract; N_1 : Number of additional papers citing or being cited by the papers in either $\mathcal{P}_{original}$ or \mathcal{P}_{C1} , but not belonging to them; E_0 : Number of directed citation edges in the networks; $E_{0,AI}$: Number of directed citation edges related only to the AI-related papers in the networks (references and citations together).

According to the top table in Tab. 2.1, 42% of all neuroscience publications extracted in Sect. 2.2.1.2 concentrates 80% of the edges. For AI-related publications within neuroscience, 51% of them concentrate 75% of the edges. Thus, after a significant loss of papers caused by the application of $C1$, a still significant part of the citation edges remains conserved. The preservation of a still important number of links is also testified by the number N_1 of supplementary papers collected to enrich the egocentric citation network, whose only 9.6% are lost after the selection. In addition, the share of AI publications among all neuroscience publications referenced in the bottom table in Tab. 2.1, as well as the share of citation links associated with AI publications and the transitivity, remain rather similar before and after the selection. The total degree distribution of the global citation network shown in Fig. 2.5, is also preserved, which is also a signal of the concentration of links around the selected core. Only some papers before the elbow around $k \sim 10^2$ and in the long tail of the distribution at very high k (higher than 10^3) are discarded.

More precisely, we discard the papers exhibiting either a total degree k lower than 10 or higher than 10. Among the latter, they could have either more than 10 references and less than 10 citations (first category), or less than 10 bibliographic references and more than 10 citations (second category). We assume that the papers in the first category do not contribute to the diffusion of important knowledge in our corpus because of their limited citation impact. On the contrary, the papers in the second category, although they have a very small bibliography, which complicates the identification of their disciplinary or knowledge affiliation, could produce a large citation impact¹⁹ in neuroscience and beyond

¹⁹Although it does not figure in the extracted corpus $\mathcal{P}_{original}$, the pioneering work extending the

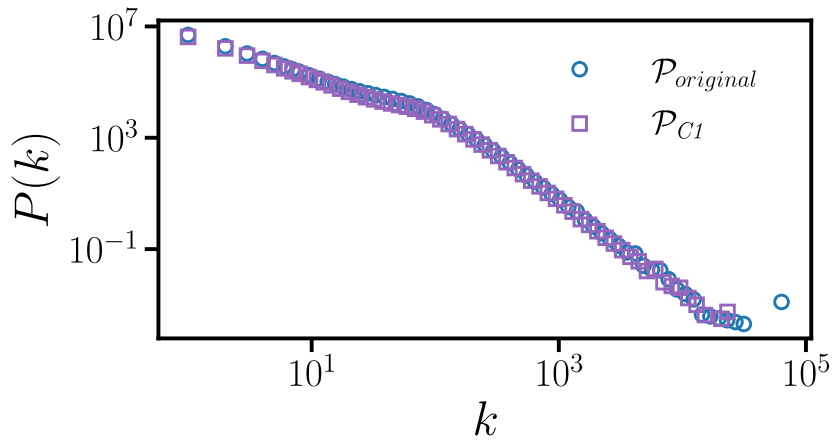


Figure 2.5: Log-binned total degree distributions of the citation network before and after applying of the selection criterion $C1$.

that we miss with our selection process. Animated by computational issues due to the size of the mobilized datasets, and with the aim of sharing and making reproducible the research presented in this work, we reject some crucial works that could be at the heart of the dissemination of important knowledge in our corpus, especially those within our small AI-related subcorpus.

In short, even if we have previously characterized the networks without the use of more sophisticated metrics that more accurately summarizes their respective true topology and their possible similarities, such as those exposed by Tantardini et al. (Tantardini et al., 2019), the basic metrics we used above indicate a conservation of some fundamental properties of the citation network after application of the $C1$ filter, which is a rather good sign for further analyses. For all these reasons, we decide to focus the latter only on the core of neuroscience publications within the subset \mathcal{P}_{C1} .

2.2.4.2 Impact on the collaboration network

Here we analyze whether some global properties of the time-aggregated collaboration network $\mathcal{A}_{original}$ are preserved before and after the selection criterion discussed above. In what follows, we especially define *AI practitioners* as neuroscientists with at least one AI-related publication during their academic career.

Applying the selection criterion $C1$ based on the citation characteristics of the papers, we obtain a sub-collaboration network called \mathcal{A}_{C1} , whose main global properties are summarized in Tab. 2.2. However, a quick look at the author properties within \mathcal{A}_{C1} –

backpropagation model for learning automata by Rumelhart et al. (1986), published as a letter for *Nature*, is a particularly edifying example with its only 4 references for its almost 38,000 citations in 2024, according to the platform *Google Scholar*.

as assigned in Sect. 2.2.3.2 – reveals a poor disambiguation²⁰ of the authors referenced within the MAG database: some distinct researchers with exactly the same first names and last names may have been “merged” by the database into a single fictitious author who owns an extended set of publications corresponding to the sum of all their scientific productions. In such a case, it becomes arduous to distinguish between the different physical researchers who are involved separately in some of these publications and not in others. This therefore implies that some authors, who are aggregations of two or more different persons, present implausible features of a standard academic career, such as a career length Δy of about a century and an excessively high number of publications. Conversely, many other authors have very few publications (one or two at most), which causes some issues for the subsequent reconstruction of authors’ disciplinary trajectories. Indeed, on the one hand, a small number of publications is certainly not sufficient to build an individual scientific trajectory, and on the other hand, it is not representative of a solid advancement of an academic career.

We therefore overcome these problems by proposing a second selection criterion $C2$, applied only to the authors of the collaboration network \mathcal{A}_{C1} . Although not based on a second disambiguation process of the authors, this new criterion relies on several numerical parameters, namely a maximum length of their academic career, a minimum year at the beginning of their career, and a minimum number of publications to have. More precisely:

Criterion selection C2

The scientists selected must have had their first publication(s) at least in 1940, with a career no exceeding 50 years and a minimum number of publications of 3.

We have chosen the above parameter thresholds for several reasons. First, the authors who started their scientific careers in 1940 may have had a significant impact on the field at least by the 1970s, the beginning of the period under study, especially during the rise of expert systems. Second, we set the maximal duration of an academic career at 50 years because it represents a reasonable period of time from the Ph.D. graduation of a researcher to his/her retirement or accession to an emeritus status, although we

²⁰Authors disambiguation designates various computational operations to distinguish two or more authors who could have exactly the same name and first name in a given database. Since these operations are not done by the traditional bibliometric database (WOS, Scopus, APS, etc.), bibliometricians often handle this process after the extraction of their database, which is based on various factors that could distinguish homonymic authors, such as their affiliations during his/her career, their recurring collaborations, or their individual citation network (Schulz et al., 2014; Sinatra et al., 2016, pp.4-6 of their supplementary material). The originality of the MAG lies in the fact that the disambiguation process is already executed, based on an unsupervised classifier compiling various features related to the authors (Färber and Ao, 2022), but with some imperfections that need to be corrected afterwards. Nonetheless, we use the dataset available in the version 2020-06-19 of the MAG Zenodo repository at the address <https://zenodo.org/records/4617285>.

	N_a	$N_{a,AI}$	E	E_{AI}
$\mathcal{A}_{original}$	2,611,456	–	37,354,666	–
\mathcal{A}_{C1}	1,278,468	202,101	13,786,616	3,136,119
Loss($\mathcal{A}_{original} \rightarrow \mathcal{A}_{C1}$)	-51%	–	-63%	–
\mathcal{A}_{C2}	871,282	186,406	10,785,916	2,839,313
Loss($\mathcal{A}_{C1} \rightarrow \mathcal{A}_{C2}$)	-31%	-7.8%	-22%	-9.5%

	$N_{a,AI}/N_a$	E_{AI}/E	Transitivity
$\mathcal{A}_{original}$	–	–	0,001
\mathcal{A}_{C1}	0,16	0,23	0,003
\mathcal{A}_{C2}	0,21	0,26	0,003

Table 2.2: Comparison of the collaboration network at different steps of filtration, following the successive applications of the selection criteria $C1$ and $C2$. The authors publishing alone and with no collaboration between 1970 and 2020 are discarded. The *Loss* rows are the fractions of items removed from the set X before and after one step of selection, i.e. $Loss = (X_{filtered} - X_{original})/X_{original}$. Due to the too high number of publications to download for authors within the set of papers $\mathcal{P}_{original}$, which are necessary to know their respective number of AI-related publications, the quantities related to AI are not mentioned for the original collaboration network $\mathcal{A}_{original}$. **Notations:** N_a : total number of authors in each collaboration network \mathcal{A}_i ; $N_{a,AI}$: number of authors in each network, who have at least one AI-related publication throughout their career; E : total number of co-signature edges, i.e. $E = \sum_{i,j \in authors} w_{ij}$; E_{AI} : total number of co-signature edges involving only AI practitioners.

lose the contribution of such emeriti who, due to their age and career progression, may be *opinion leaders* in science (Zuckerman and Merton, 1972) and continue to produce fruitful, influential work even after their 50-year academic careers. And third, we set the minimum number of publications at 3 because it is representative of the scholarly production of a fresh Ph.D.-graduate researcher, even an early career one.

These authors meeting this criterion $C2$ belong to a subset of the collaboration network \mathcal{A}_{C1} , called \mathcal{A}_{C2} in the following lines. Tab. 2.2 also indicates the potential loss of authors and links after the application of this last criterion $C2$.

This table shows that, although the application of the two successive selection processes removes a lot of authors and collaborations, a significant proportion of those who have published at least one AI-related article still remain in the co-authorship network, with also a remarkably stable number E_{AI} of collaborations involving at least one AI practitioner. In particular, the application of the $C2$ filter increases the presence of these AI contributors as well as their collaborations, as shown in the lower table. We notice that the number of collaborations E_{AI} between authors who have published at least one AI-related article reaches about a quarter of all edges of the two filtered networks \mathcal{A}_{C1} and \mathcal{A}_{C2} . The following chapters will explore how these scientists who have been touched by AI at some point are organized, as well as the dynamics of such social diffusion over time, in particular by considering different types of AI practitioners according to their

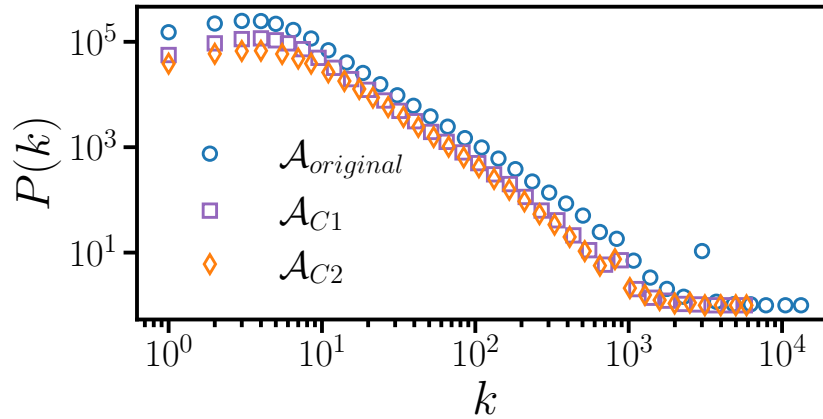


Figure 2.6: Log-binned weighted degree distributions of the authors in the datasets $\mathcal{A}_{original}$ and its two different filtered versions \mathcal{A}_{C1} and \mathcal{A}_{C2} .

level of expertise in the subject. Furthermore, the transitivity of the network, although very small, remains unchanged after the entire filtration process.

We also plot in Fig. 2.6 the degree distribution of collaboration after the two filtering steps $C1$ and $C2$. The $C1$ criterion helps to homogenize the collaboration network by removing the (often poorly disambiguated) authors located in the long tail of the degree distribution. We notice that the second selection $C2$ does not change the shape of the distribution, thus demonstrating global conservation of the collaboration structure within our dataset \mathcal{P}_{C1} .

Finally, notice that the collaboration network is built only from the most impactful scientists in our dataset. Indeed, on the one hand, we have drawn the collaboration network directly from the set of neuroscience papers that are the most influential in terms of citations, i.e. \mathcal{P}_{C1} , and on the other hand, we have focused on the authors who have published at least 3 articles by 2019 and who join the collaboration network \mathcal{A}_{C2} . Therefore, we did not consider short-term collaborations with early-career, small-career, or non-academic researchers who are involved in one or two publications with too low an impact but who can potentially drive forward innovation.

2.2.5 Final dataset and discussion

In this section, we have followed the pipeline illustrated in Fig. 2.7 to build a database representing the neuroscience literature between 1970 and 2020, including 855,691 articles with several metadata of interest for the studies introduced in the next chapters, and completed with an egocentric citation network and a collaboration network. This set of papers, finally called \mathcal{P} in the rest of the manuscript, is distributed over 421 journals listed in App. A.2. Some parts of the dataset – all except titles, abstracts, fields of study

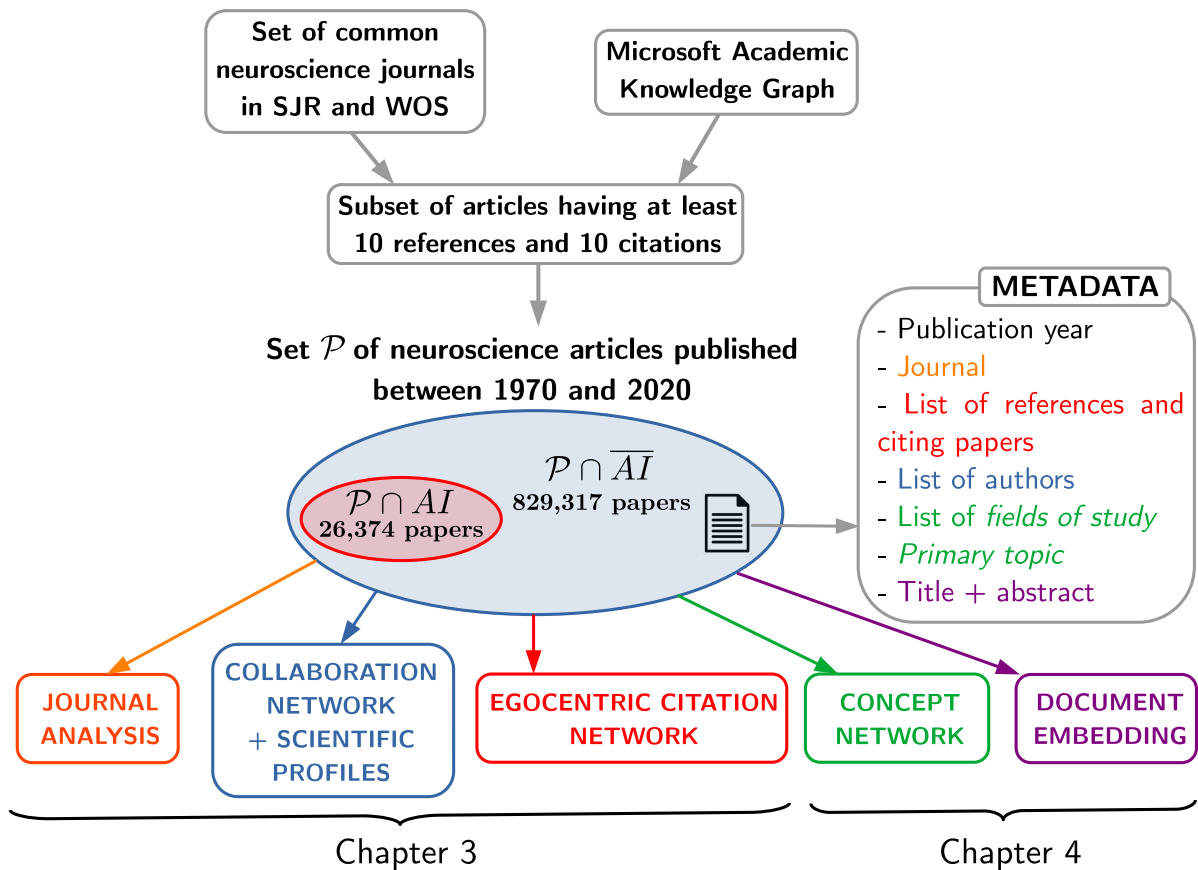


Figure 2.7: Building steps of the dataset of neuroscience articles used throughout the thesis, especially the two following chapters.

and topics – have been shared on Zenodo in the repository referenced by (Fontaine et al., 2024b) and have been used in an article co-written by the author of this thesis, which is the subject of Chapter 3 (Fontaine et al., 2024a).²¹

More precisely, the two subsets of AI-related papers ($\mathcal{P} \cap \overline{AI}$) and non-AI ones ($\mathcal{P} \cap AI$) contain 829,317 and 26,374 papers, respectively, spread over the above period according to Fig. 2.8. Of all these high-impact research papers in neuroscience, only 3% contain AI-related keywords. The inset of this figure exhibits in particular a slow growth of the share of AI-related publications in neuroscience at the very end of the 1980s, which stabilizes around 1995 and is followed by a very rapid growth from 2007 to 2019. This plateau around 1995 suggests a second, prolonged *AI winter* period in neuroscience, characterized here by a stable interest in AI research, but not a decline, unlike in other disciplines or research fields (Cardon et al., 2018; Schuchmann, 2019). The subsequent period of important growth in this share, starting in 2007, also suggests the main influence of the rise of deep learning techniques, a well-known trend shared by almost all sciences (Cardon

²¹We have used the Python programming language to perform the analysis outlined in Chapter 3. The code and results are presented in the form of Jupyter notebooks, which are freely available at: https://github.com/syMC/AI_in_Neuroscience_EpistemicIntegration_SocialSegregation. As the results of Chapter 4 are currently provisional, codes and associated notebooks are still under construction and revision at the time of writing this thesis.

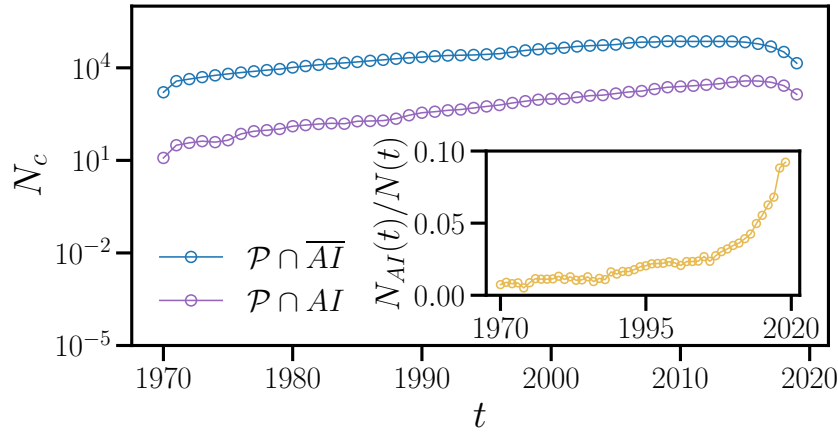


Figure 2.8: Cumulative number N_c of papers in the two main subsets considered in this paper. The inset shows the instantaneous fraction of AI-related publications in the whole neuroscience corpus.

et al., 2018; Gargiulo, Fontaine, et al., 2023). The share of AI publications in neuroscience reaches only 10% of the number of publications at its highest level situated at the end of the period studied, which means that the use of the AI-related keywords (from our list) in neuroscience remains rather limited even today.

Thanks to the two selection criteria described above, we will also consider a set of 886,074 scientists, of which 188,325 (only 16% of the scientists) have published at least one AI-related paper in neuroscience. In Sect. 2.2.3.2 we have built a special indicator to capture the activity of each author in AI research in general, that is, not only in neuroscience but also in other fields of research. We have especially defined this indicator as the proportion of AI-related papers published before 2019, $f_{AI}(a) = n_a^{\mathcal{P} \cap AI} / n_a^{tot}$. According to the f_{AI} distribution shown in Fig. 2.9, we divide this set of authors in four parts, namely \overline{Q} ($f_{AI} = 0$; $N_a = 697,749$), Q_0 ($f_{AI} \in (0, 0.5)$; $N_a = 182,925$), Q_1 ($f_{AI} \in [0.5, 1)$; $N_a = 4,977$) and Q_2 ($f_{AI} = 1$; $N_a = 423$). Thus, by constructing these quartiles, neuroscientists in \overline{Q} are assumed never to have collaborated with those in Q_2 between 1970 and 2020, but each member of these two groups may have collaborated with scientists belonging to the other two quartiles Q_0 and Q_1 . Moreover, only scientists in the latter two quartiles may have collaborated with scientists belonging to either \overline{Q} or Q_2 . We adopt this classification of authors in the next chapter.

Nevertheless, we warn the reader that this dataset is not as exhaustive as we would have liked and does not perfectly reflect the state of the field during the 50-year period retained, due to many biases expressed throughout this section. These biases could affect the way AI spreading within neuroscience is studied further. While some of these have already been addressed above, some implications of others are detailed below.

Firstly, recall that we focus only on neuroscience articles published in peer-reviewed

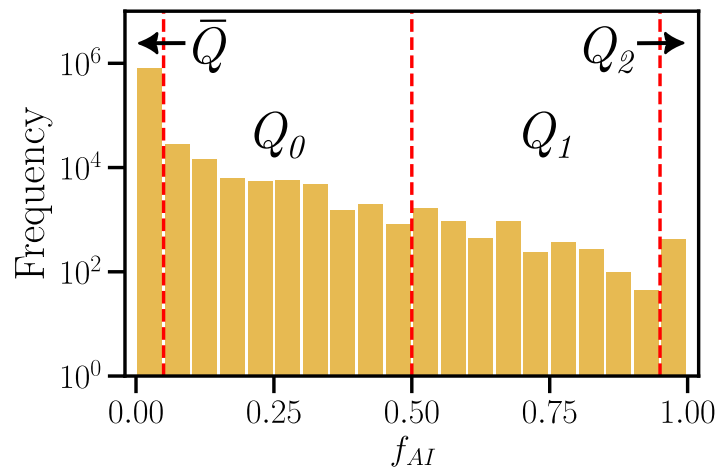


Figure 2.9: Distribution of the AI score f_{AI} of the authors.

journals. Indeed, we have discarded important types of publications that could be at the heart of the dissemination of AI in neuroscience, such as conference proceedings, which account for a significant part of publications in the field of computer science (Meyer et al., 2009; Wainer et al., 2013), which is adjacent to AI, the preprints available in dedicated repositories, e.g. arXiv, and also patents. Nonetheless, a simple query to the OpenAlex API²² unveils that neuroscience results are largely communicated in this format, so we can reasonably assume that we are capturing, in a bibliometric sense, the most important part of AI-related research output published according to neuroscience publication standards. The method advanced in this section is therefore a consistent, though not unique, way of measuring the degree of appropriation of AI by the field of neuroscience.

Another important point is that, although we are focusing on the most widely used format for communicating neuroscientific results, the analysis introduced in the next chapters will rely on a particular subset of journals classified in the WOS, which is rather selective about the sources it chooses to reference. Although the latter now shows an expansion of its international and disciplinary coverage (Birkle et al., 2020), we lose some contributions that are not published in journals referenced by this database, because of their languages, their recent launch, or their too low impact factor. Here, we thus have chosen to draw our analysis upon a neuroscience literature that is supposed to drive the most circulation of ideas across academia, i.e. published in journals with a significant impact factor and mostly written in English, according to the WOS.

In summary, the \mathcal{P} dataset and its ecosystem represent a particular scientific production

²²We call the OpenAlex API on May 24, 2024 to get the distribution of output types (according to the classification given by CrossRef) related to the field *Neuroscience* between 1970 and 2020. Among the 4,943,457 fund publications, 84% of them are *journal articles*, about 7% are *book chapters*, 3% are *proceedings articles* and 1% are *posted-contents*, including preprints – which are removed from OpenAlex once they have been published in a peer-reviewed journal or conference proceedings. The remaining 5% are of various other types that are not significant.

within neuroscience, namely that which is published as articles in journals referenced by traditional bibliographic databases and has achieved a certain citation success.

2.3 Qualitative data collection from a laboratory field-work

2.3.1 The Aramis team: a window on AI research for neuroscience

According to its 2019 activity report,²³ Aramis has been founded in 2012²⁴ and is based at the Institut du Cerveau (ICM) since 2014 as a “project team”²⁵ of the Institut National de Recherche en Informatique et en Automatique (INRIA), under joint supervision of the Centre National de la Recherche Scientifique (CNRS), the Institut National de la Santé Et de la Recherche Médicale (INSERM) and Sorbonne University. This team is also part of the Centre d’Acquisition et de Traitement d’Images pour la maladie d’Alzheimer (CATI).²⁶ Its members, mainly computer scientists, statisticians, and clinicians who are also engaged in regular medical practice, are in charge of designing numerical and statistical models to predict the evolution of brain diseases, particularly neurodegenerative pathologies. Using AI algorithms based on machine learning (ML) and deep learning (DL), researchers need large clinical databases to train, verify, and validate their models. These models are then implemented in specialized, often *open-source*²⁷ software designed to help not only researchers but also medical practitioners to more easily identify pathologies from acquired images. As part of translational research (Allen et al., 2019; Stokols et al., 2008) conducted in ICM, allowing the deployment of academic outputs into regular medical practice directly in the hospital, Aramis works closely with other clinical research teams at the institute, sharing knowledge and predominant IT-based skills.

²³This report is available at the following address for the year 2019: <https://raweb.inria.fr/rappportsactivite/RA2019/aramis/uid3.html>.

²⁴The first Aramis activity report, written in 2013 and available on INRIA’s Explora-Web portal, confirms its foundation during 2012.

²⁵Translated from the french expression “équipe projet”, used especially at INRIA.

²⁶National and international network for sharing neuroimaging techniques, created in 2011 as part of the French Alzheimer’s Plan. See at: <https://cati-neuroimaging.com>.

²⁷A software is *open-source* when its original code is freely and publicly shared on a collaborative platform. In this way, any user or developer can freely modify, customize, and even improve the software to enhance its performance, and then redistribute it on the same platform afterwards.

2.3.2 Choosing a semi-structured interview format

A qualitative survey, based on semi-structured interviews using the questionnaire in App. D.1 (initially in French), has been conducted between March and April 2021 with eight scientists in ICM. Seven are members of the Aramis team and exhibit together diverse scientific profiles, from computer scientists and mathematicians conceiving statistical or AI-based models and programming them, to clinicians using them for clinical research. An eighth person, with a status of statistician and not member of Aramis, has also been interviewed for his/her close relationships with the various ICM's teams. All of them are referred to hereafter by an interviewee code. Table 2.3 below indicates their gender, academic status, and institutional affiliations, alongside the date and duration of their corresponding interview.

Due to the lock-down restrictions imposed by the French health emergency policy, effective from March 17, 2020 and still active during the fieldwork in the next year, ethnographic observations in a hospital were not feasible, and the interviews have all been conducted on a virtual face-to-face mode using Zoom software. A significant number of technical hitches have disturbed interviews, mainly due to poor internet connections. From simple image cuts to time lags between the two speakers, these events have led to some confusion or repetition of what one speaker had just said to the other. These virtual sessions thus differ from a conversation that two individuals could have physically in another space – e.g., in the street or at work – despite the effort made by the interviewer (here the author of the present thesis) to ensure that these interviews are “real-like” conversations in which the interviewees can feel at ease without necessarily having the impression of being subjected to a simple questionnaire.

This interview format adopted by the interviewer (in fact, the author of this thesis) allows to capture wealthy personal information revealed by the interviewees on their respective perception of AI in their neuroscientific research work, but has some consequences in terms of interpretations given in Chapter 5. In fact, the interviewer can follow his questionnaire according to precise themes while allowing himself to be carried along by interviewee's discourse, which he may find interesting for his own fieldwork, with the risk, however, of getting caught up by his/her personal opinions and perhaps sharing them. This may obviously influence the way the interviewer asks the questions and, consequently, the interviewee's answer, which may hinder any (idealistic) semblance of objectivity. Digressions are then observed as well as answers to questions – within the questionnaire or not – that the interviewer did not explicitly ask, thus raising new questions and hypotheses that can be explored in future interviews. Two exchanges thus are completely different. Moreover, unlike a multiple-choice grid or a structured interview,²⁸

²⁸The structured interview format tends to be adopted for large national statistical survey based on

Interviewee	Gender	Status	Affiliations	Date of the interview	Duration
I1	Male	Post-doctoral researcher in neuroimaging & genetic data analysis	ICM, QIMB	March 8, 2021	1h28
I2	Male	R&D engineer & data scientist	ICM, INRIA	March 19, 2021	1h
I3	Male	Neurologist & researcher	AP-HP, ICM, IM2A	March 23, 2021	42min
I4	Male	Post-doctoral researcher in machine learning applied to neuroscience	ICM, INRIA	March 26, 2021	57min
I5	Female	Neuropsychologist & PhD candidate	AP-HP, ICM	March 30, 2021	58min
I6	Female	Statistician	AP-HP (Biostatistic team of La Pitié-Salpêtrière Hospital), IM2A	April 6, 2021	1h57
I7	Male	Neurologist & PhD candidate	AP-HP, ICM	April 7, 2021	1h23
I8	Female	Full-time researcher in computational neuroimaging	CNRS, ICM	April 13, 2021	1h06

Table 2.3: Interviewees’ summary. All interviews have been conducted in French and remotely, due to the COVID-19 pandemic, with the Zoom video-conference software. Abbreviations: **AP-HP**: Assistance Publique-Hôpitaux de Paris; **CNRS**: Centre National de la Recherche Scientifique; **ICM**: Institut du Cerveau; **IM2A**: Institut de la Mémoire et de la Maladie d’Alzheimer; **INRIA**: Institut National de Recherche en Informatique et en Automatique; **QIMB**: Queensland Institute for Molecular Bioscience (Brisbane, Australia).

the order of the questions is rarely respected, making it difficult to observe any potential regularities. Respondents' discourse and the way they implicitly answered a question in the original interview grid are also crucial for a detailed understanding of their opinions. Full or partial transcriptions, duly annotated²⁹ in French, have therefore been essential to the success of this work.

Finally, in addition to being a discovery of a scientific world whose (precise) research topics, the nature of the work effectively done in laboratories, in what institutional context(s) – at the crossroads of academic and clinical research, often conducted within international consortia and in dialog with regular medical practice and health policy – were unknown to us, these interviews have also enabled us to appropriate, even build for ourselves a scientific culture (Laudel and Gläser, 2007) of neuroscientific research, especially in the domain of computational neuroscience, as is done at Aramis.

a representative sample. Individuals are then asked one by one, usually by phone or email, following a questionnaire whose answers are often yes or no (or neutral).

²⁹We have manually transcribed these interviews following the method suggested by Rioufreyt (2016). Although not added to this thesis, we can provide the transcriptions on request.

Domain adaptation and social segregation of AI research in neuroscience

In this chapter, using the database built in Chapter 2 and the additional metrics exposed in the following Sect. 3.1, we analyze two AI genericity criteria, namely domain adaptation and social adoption, through two dynamic mechanisms underlying the appropriation of AI by neuroscience and then the possible knowledge transfer between the instrument and the field.

The first one concerns the level of embedding of AI into the multidisciplinary context of neuroscience, which we study through a special representation of the disciplinary landscape of the host field, based on its bibliographic references and citations. Such an enterprise aims to situate the AI-focused research subarea within this disciplinary landscape. This is the object of Sect. 3.2, in which we first build two citation ecosystems surrounding, respectively, the whole neuroscience and the particular AI research within it, in order to evaluate at macro- and micro-scales the common traits of these two ecosystems afterwards.

The second mechanism, object of Sect. 3.3, concerns the social diffusion of AI throughout the neuroscience community, which we study not only with the collaboration network of the field but also with the disciplinary trajectories of the neuroscientists involved in AI research. These trajectories will enable, in particular, identifying the scientific profiles who are more inclined to use AI than others. With their location within the whole neuroscience collaboration network, we will also assess the ability of these authors to spread AI across the field.

Finally, in Sect. 3.5, we discuss these two mechanisms, in particular whether the configurations of the citations and the collaborations around AI research within neuroscience foster or hinder its diffusion throughout the field.

Some elements of this chapter are borrowed from the article by Fontaine et al. (2024a), partly written by the author of this thesis.

3.1 Presentation of the mobilized framework

3.1.1 Method for comparing the disciplinary influence and impact of the AI-related and non-AI research over time

In this section, we propose a method to describe the articulation of AI research with neuroscience, based on the aggregation of bibliographic references and citations according to their respective WOS-related JSCs. With the following exposed framework, we can identify the most influential disciplines of neuroscience – as it is a multidisciplinary domain – and also those that are at the source of AI-related research within the field. This framework also allows us to identify the disciplines that cite this research the most.

3.1.1.1 Building temporal disciplinary rankings representing preferred bibliographic references and preferred citations

We start with the egocentric citation networks of the two neuroscience corpora $\mathcal{P} \cap AI$ and $\mathcal{P} \cap \overline{AI}$ built in Sect. 2.2.5 of the previous chapter, i.e., the corpus containing AI-related papers and that without AI-related ones, respectively. As drawn in Sect. 2.2.3.1, these citation ecosystems consider both the bibliographic references and the citations of the focal papers within one of the two corpora. We consider such a citation ecosystem as a proxy for the disciplinary environment surrounding the two corpora we want to compare. In particular, we build these ecosystems with the WOS JSCs associated with the papers constituting the global set of bibliographic references of the corpora (or their citations). Here, we consider only the first level¹ of the disciplinary taxonomy of journals given by WOS in 2021, which are broad disciplines in science (recalled in App. A.3).

First, for each corpus, we count the annual number of citations received by each of these JSCs within its own set of bibliographic references. Doing this year-by-year over the period 1970-2019, we therefore build a temporal ranking of the disciplines that are most cited

¹This taxonomy actually comprises two levels, which apply only to some broad disciplines. For example, within the JSC *Computer Science* (at level 1) we can distinguish some main areas of the field, such as *Artificial Intelligence*, *Theory and Methods*, *Information Systems* and *Cybernetics* (at level 2). This is also the case for many other JSCs at level 1. Aggregating all these subareas into a single discipline introduces a bias at the microscale, as they could be related to different research practices and manners of adopting (or not) AI-related knowledge and technologies. Here, for the sake of clarity, we choose to rely on the first level of JSCs of the presented results.

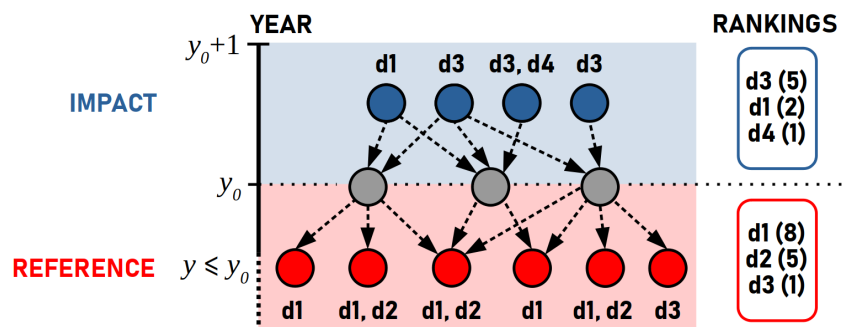


Figure 3.1: Citation ecosystem centered here around three papers (gray) published in year y_0 and included in either $\mathcal{P} \cap AI$ or $\mathcal{P} \cap \overline{AI}$. A dashed arrow represents the citation of a target paper by a source one. Thus, the red circles are the papers that are *cited* by the papers in our corpus (*reference*), while the blue ones are the papers that *cite* them (*impact*). The rankings are shown in descending order with the corresponding number of citations of each JSC d_i .

by each corpus. For example, in Fig. 3.1, the papers cited by a corpus in year y_0 (gray points) are mainly published in journals labeled by JSC d_1 with 8 citations, followed by JSC d_2 with 5 citations, and d_3 with only one citation (red points and the corresponding red-contoured ranking). Hence, the disciplinary reference ranking $r(y_0) = \{d_1, d_2, d_3\}$ with each JSC sorted according to their respective ranks mentioned above. However, this ranking could be different before this year y_0 and also in the next ones, depending on the citations received by each JSC in those years.

We specify that we use such JSC rankings for the two corpora instead of the absolute number of citations to allow comparison of their respective citation ecosystems, because AI research occupies a small place in the field of neuroscience (3%), with a much smaller number of references and citations than those associated with neuroscience research away from AI.

We do the same for the JSCs that most cite our papers of interest up to one year after their official publication year. We have chosen this time period to avoid the high temporal dispersion of the generated citations over the whole time period under study, where the citation impact of the papers published in 1970 may eventually be greater than those published in 2019. In this way, we assess from which JSCs the generated citations first originate, before eventually spreading to a broader disciplinary landscape.

Finally, by looking at both the reference and the impact sides, we represent two disciplinary landscapes associated with the corpora $\mathcal{P} \cap \overline{AI}$ and $\mathcal{P} \cap AI$, respectively. More precisely, the reference ranking associated with a given corpus summarizes its main influential fields of study, and thus the disciplinary structure on which it draws over time, while the citation ranking indicates the fields that are at first impacted by the corpus.

3.1.1.2 Building of a comparative disciplinary cartography between AI-related and non-AI research

Following the procedure described above, and according to all the pairs of references' and citations' numbers associated with all the JSCs represented in $\mathcal{P} \cap \overline{AI}$ and $\mathcal{P} \cap AI$, we associate with each JSC d two pairs of ranks indicating its weight either in references or in citations of the two corpora in year t , denoted respectively as $r_R^d(t) = (r_{R, \mathcal{P} \cap \overline{AI}}^d(t), r_{R, \mathcal{P} \cap AI}^d(t))$ and $r_I^d(t) = (r_{I, \mathcal{P} \cap \overline{AI}}^d(t), r_{I, \mathcal{P} \cap AI}^d(t))$. In case of a missing JSC in one of the two corpora in a given year t , we fill the missing rank with a maximum value set to 100.

With such a pair of ranks (either for references or citations), we can locate a discipline d in a 2D space of rankings, as shown in Fig. 3.2. In this coordinate system, the lower the value of an axis, the better the associated rank, and the higher the number of citations collected along that axis. The angle θ indicates the deviation of d from the diagonal in this space, where the ranks are exactly the same in each corpus.

From this map we define the *common interest area* between $\mathcal{P} \cap AI$ and $\mathcal{P} \cap \overline{AI}$ (in gray) as the region close to the diagonal where the two rankings of the fields of study within it are almost the same and do not vary significantly over time. This zone is comprised between the two lines of respective equations $r_{\mathcal{P} \cap AI} = r_{\mathcal{P} \cap \overline{AI}} + \tau$ (above the diagonal) and $r_{\mathcal{P} \cap AI} = r_{\mathcal{P} \cap \overline{AI}} - \tau$ (below the diagonal), where τ is a parameter set to 10.

The blue zone above the common interest area corresponds to the space where disciplines have a better rank in $\mathcal{P} \cap \overline{AI}$ than in $\mathcal{P} \cap AI$. In the case of references, such disciplines in this zone are cited more by the corpus $\mathcal{P} \cap \overline{AI}$ than by $\mathcal{P} \cap AI$, and in the case of citations received by each of the two corpora, such disciplines in this zone cite more the papers in $\mathcal{P} \cap \overline{AI}$ than those in $\mathcal{P} \cap AI$. The purple zone under the common interest area is therefore the opposite case, i.e., in the case of references (for example), the disciplines in this zone are more cited by the corpus $\mathcal{P} \cap AI$ than by $\mathcal{P} \cap \overline{AI}$.

In Sect. 3.2.2 we apply this geometric construction per decade periods instead of single years, i.e., we build a decade ranking space from the total number of citations given or received by the represented JSCs in $\mathcal{P} \cap \overline{AI}$ and $\mathcal{P} \cap AI$ within the period under study.

3.1.2 Computation of the intensity of collaborations in the aggregated co-authorship network

In this section, we detail the general computation of an indicator that will be used in Section 3.3.2 to indicate the propensity to collaboration between neuroscientists in our

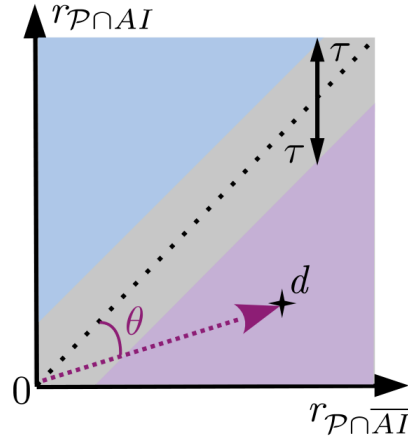


Figure 3.2: Coordinate system used to characterize the distribution of disciplines in the references and citations of the corpora $\mathcal{P} \cap \overline{AI}$ and $\mathcal{P} \cap AI$.

database.

We first consider the static collaboration network $\mathcal{A} = (S, E)$, whose construction is summarized in Sect. 2.2.5 of Chapter 2, where S is the set of scientists and E the set of edges between them. Each scientist in S is identified by their respective AI score f_{AI} . According to Section 2.2.5, the distribution is divided into four parts $\{\overline{Q}, Q_0, Q_1, Q_2\}$, from the least AI expert to the most expert. Therefore, an author belongs to a given group Q , given his or her own AI score. By aggregating the scientists into such Q groups, we compute the 4×4 matrix τ_{obs} of effective edges between these groups, where an element is written as $\tau_{obs}(Q_i, Q_j) = |\{(u, v) | u \in Q_i, v \in Q_j\}|$.

Then we create a set of N alternative collaboration networks $\{\tilde{\mathcal{A}}_k\} = \{(S, \tilde{E}_k)\}$, with $k = \{1, \dots, N\}$, all of which are based on the same set of scientists but have different sets of edges obtained by perturbing the network with a uniform shuffling that conserves the degree distribution of scientists. By applying the same clustering of scientists into Q groups for each network as for the real network \mathcal{A} , we draw a set of matrices $\{\tau_k\}$ corresponding to the fraction of edges between the Q groups within the randomized networks $\{\tilde{\mathcal{A}}_k\}$. From this set, we extract, on the one hand, the average matrix τ_{sim} , where an element is defined as $\tau_{sim}(Q_i, Q_j) = \frac{1}{N} \sum_k \tau_k(Q_i, Q_j)$, and on the other hand the standard deviation matrix σ , where an element is defined as $\sigma(Q_i, Q_j) = \frac{1}{N-1} \sum_k (\tau_k(Q_i, Q_j) - \tau_{sim}(Q_i, Q_j))^2$.

We finally compare the empirical matrix τ_{obs} of the real network with the simulated one τ_{sim} through a z -score matrix where an element is defined as follows:

$$z(Q_i, Q_j) = \frac{\tau_{obs}(Q_i, Q_j) - \tau_{sim}(Q_i, Q_j)}{\sigma(Q_i, Q_j)}. \quad (3.1)$$

Here, this standardization is used to test the over- or under-representation of a given number of edges between two groups with respect to the corresponding average simulated

value, which represents an ideal situation produced by randomization. Applying the whole procedure above to the network \mathcal{A} , with $N = 100$, gives Fig. 3.11B. Fig. 3.11A is also drawn by applying the same method with the same N to each temporal snapshot of the temporal collaboration network \mathcal{T} (see also Section 2.2.3.2 for its building), and by aggregating the groups Q_0 , Q_1 and Q_2 into a single one called Q_i (see Fig. 3.11A), which includes all scientists with at least one AI-related publication. Since \mathcal{T} is undirected, the 4×4 matrices τ_{obs} , τ_{sim} , σ and z are therefore reduced to scalars defined only by the two groups \bar{Q} and Q_i .

3.1.3 The AI-activity of the journals

For each of the 421 journals listed in our database drawn in Chapter 2 and recalled in App. A.2, we retrieve its entire publication history from the MAG to compute its temporal share $a_{AI}(t)$ of AI-related publications, also called *temporal AI activity* in the following lines, and its global share a_{AI} of such publications since its founding year, more simply called *AI activity*.

3.2 The situation of AI within the neuroscience disciplinary landscape

3.2.1 Citation homogenization of the AI-related publications with general neuroscience

First of all, we explore the integration of AI technologies and knowledge within neuroscience through the dynamical interaction between the respective disciplinary environment of the two fields – AI and neuroscience. We especially compare year by year the basis of reference and citations that are shaping these fields, here represented as the AI-related publications and the non-AI ones within our neuroscience dataset.

For the two subcorpora $\mathcal{P} \cap AI$ and $\mathcal{P} \cap \bar{AI}$, and following the method exposed in Sect. 3.1.1.1, we build two disciplinary rankings that may vary over time, one for the disciplines appearing in their respective references ($r^R(t)$), hereafter called *reference ranking*, and another for those appearing in the respective citations they have received ($r^I(t)$), called *citation ranking* in the following. We then use a common similarity metric J provided by Gargiulo et al. (2016, see their supplementary materials), comprised between 0 and 1, to evaluate how similar two given rankings r_A and r_B are: if $J(r_A, r_B) = 1$, the two

rankings are exactly the same, i.e., containing the same elements with the same respective ranks; conversely, if $J(r_A, r_B) = 0$, the elements included in r_A are not in r_B , regardless of their respective ranks. In particular, we compute three time-varying measures:

1. the temporal self-similarity of a given ranking r , namely the similarity between this ranking at time t and the same one at the previous time $t - 1$, denoted as $J(r_{t-1}, r_t)$; by applying it to the reference and citation rankings in each corpus, we evaluate the dispersion or the reinforcement of the set of disciplines constituting these rankings independently,
2. the temporal disciplinary concentration of a corpus, namely the similarity between its reference ranking r_R and its citation ranking r_I at a given time t , denoted by $J(r_R(t), r_I(t))$,
3. the temporal similarity of the rankings of the references (or citations) of the two corpora, denoted as $J(r_{\mathcal{P} \cap AI}^R(t), r_{\mathcal{P} \cap \overline{AI}}^R)$ (and $J(r_{\mathcal{P} \cap AI}^I(t), r_{\mathcal{P} \cap \overline{AI}}^I)$ for the citations); this measure compares how much similar the respective citation ecosystems of the two corpora are over time.

The evolution of the first two indices is represented in Figs. 3.3A and 3.3B respectively. According to the former, the references on which the AI corpus and the non-AI one respectively rely (solid lines) both consolidate over time towards their respective highest values in 2017, but not at the same speed. Indeed, the non-AI corpus lies on almost the same set of disciplines from year to year, while the AI-related one grows from a low similarity in 1970 (around 50%) to a stable set with a high similarity (around 90%) between 2015 and 2017. The observed decrease after 2017 is mainly due to the lack of data grabbed by the MAG. In addition, while the self-similarity of citations in the corpus $\mathcal{P} \cap \overline{AI}$ (blue dashed line) follows the same trend as the corresponding self-similarity of references, the self-similarity of citations in the corpus $\mathcal{P} \cap AI$ (purple dashed line) is rather chaotic, fluctuating around 50% between 1970 and the late 1990s, and then increasing to around 80% by 2019. This means that the respective sets of short-term impacted fields in the two corpora are consolidating around the same disciplines in recent days. More precisely, Fig. 3.3B shows a growing concentration of reference and impact rankings within each corpus over the time period under study, i.e., the impacted disciplines and those appearing in the references are becoming more and more similar, hence a research inspired by itself towards itself. This also shows the progressive disciplinary homogenization of the two corpora $\mathcal{P} \cap AI$ and $\mathcal{P} \cap \overline{AI}$ independently, and thus a consolidation of the entire field of neuroscience.

Whereas Fig. 3.3 shows some measures characterizing the independent evolution of the citation ecosystems of AI-related and non-AI research in neuroscience, respectively,

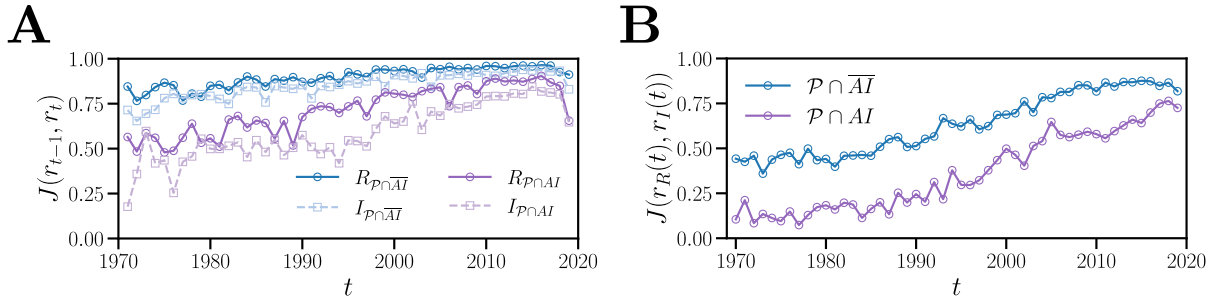


Figure 3.3: **A**: Temporal self-similarity of disciplinary rankings associated with either the references of one specific corpus (R) or the citations it has received (I), either in the AI-related corpus ($\mathcal{P} \cap AI$) or in the non-AI one ($\mathcal{P} \cap \overline{AI}$). More precisely, a point at time t is the similarity between the ranking at time t and the preceding one at $t - 1$. **B**: Instantaneous similarity between the references used by a corpus and the citation impact of that corpus.

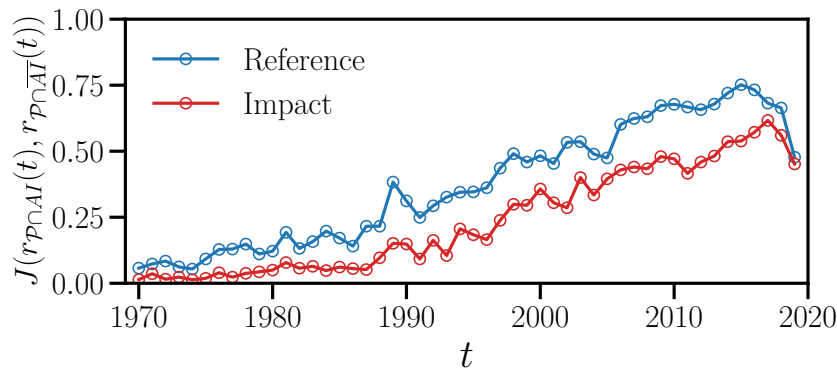


Figure 3.4: Instantaneous similarity between the reference or citation rankings of the two corpora $\mathcal{P} \cap AI$ and $\mathcal{P} \cap \overline{AI}$ at year t .

Fig. 3.4 offers a view of the evolution of temporal similarity measures of the third kind, which confront the rankings associated with the two corpora. A first one compares the rankings of the disciplines cited by the corpora $\mathcal{P} \cap AI$ and $\mathcal{P} \cap \overline{AI}$ (blue curve), and a second one compares the rankings of the disciplines that are citing them over time (red curve).

This figure shows that the disciplines cited by $\mathcal{P} \cap AI$ remain quite different from those cited by $\mathcal{P} \cap \overline{AI}$. However, the almost linear growth of the similarity of references over time is a signal of bibliographic homogenization. In this figure, we also observe a similar behavior for the second similarity measure related to the disciplines citing the two corpora, which at first glance could be divided into two stages. First, the similarity was stabilized at a very low value between 1970 and 1987, i.e., the two corpora were cited by two very different sets of disciplines, then it steadily increases since 1988, i.e., the sets of disciplines showing interest in the AI-related and the non-AI literature, respectively, have become more intertwined over time.

In summary, as the two corpora grow around reinforced sets of disciplines (see Fig. 3.3), the last two rankings shown in Fig. 3.4 together exhibit a progressive entanglement of the

two disciplinary ecosystems that characterize the two corpora $\mathcal{P} \cap AI$ and $\mathcal{P} \cap \overline{AI}$.

3.2.2 Beyond the citation homogenization: AI and non-AI publications lie on different disciplinary bases

We complete the previous ranking analysis by focusing on the difference between the microscopic disciplinary composition of the AI-related corpus ($\mathcal{P} \cap AI$) and the non-AI one ($\mathcal{P} \cap \overline{AI}$) in our neuroscience dataset. In particular, by considering the individual disciplines that are simultaneously cited by (or cite) both of the two corpora, we compare the extent to which these disciplines are actually cited by (or cite) each of them. The goal of such an analysis is to identify the disciplines from which each corpus prefers to find information, i.e., its references, and those to which it spreads its knowledge, i.e., its citations.

By applying the framework described in Section 3.1.1.1, we associate with each represented discipline d within the references or citations of the corpora $\mathcal{P} \cap \overline{AI}$ and $\mathcal{P} \cap AI$, their time-aggregated ranks over decades, $(\tilde{r}_{\mathcal{P} \cap \overline{AI}}^d(T), \tilde{r}_{\mathcal{P} \cap AI}^d(T))$, with $T = [t_0, t_0 + 10)$ a given time period where $t_0 \in \{1970, 1980, 1990, 2000, 2010\}$. These ranks are built using the total number of references to the discipline d made by the corpora, or with the total number of citations given by d to the corpora (the impact) during the time period T .

We then locate the cited or citing disciplines in the 2D ranking space shown in Fig. 3.5. In one of the maps drawn in this figure, a colored disk represents a specific discipline d that appears in each corpus with its respective ranks during the considered time period, either in references (Fig. 3.5A) or in citations (Fig. 3.5B). The lower the value along one axis, the better the associated rank, and then the higher the number of citations collected along that axis. From these maps, we define the common interest area between $\mathcal{P} \cap AI$ and $\mathcal{P} \cap \overline{AI}$ as the region near the diagonal where these fields are colored in gray. The disciplines that are more dispersed around the diagonal are represented in two different colors, the blue having a better rank in $\mathcal{P} \cap \overline{AI}$ than in $\mathcal{P} \cap AI$ and the purple having a better rank in $\mathcal{P} \cap AI$ than in $\mathcal{P} \cap \overline{AI}$.

The position of a point on a ranking map in Fig. 3.5 is set with respect to both ranks in the subsets $\mathcal{P} \cap AI$ and $\mathcal{P} \cap \overline{AI}$. We calculate the size $s_d(T)$ of the point associated with d as the square root of the weighted average of the number of citations it receives in each set:

$$s_d(T) = \sqrt{\frac{\tilde{c}_{\mathcal{P} \cap \overline{AI}}^d(T)w_{\mathcal{P} \cap \overline{AI}}^d(T) + \tilde{c}_{\mathcal{P} \cap AI}^d(T)w_{\mathcal{P} \cap AI}^d(T)}{w_{\mathcal{P} \cap \overline{AI}}^d(T) + w_{\mathcal{P} \cap AI}^d(T)}}, \quad (3.2)$$

where $\tilde{c}_k^d(T)$ is the total number of citations appearing for the JSC d in the given time

period T , and $w_k^d(T)$ is the proportion of citations obtained by d in the subset k for the considered time period T . This weighted average combines the level of occurrence of each JSC in $\mathcal{P} \cap AI$ and $\mathcal{P} \cap \overline{AI}$, as it could have collected the same number of citations in each corpus while being ranked differently within them. The square root especially flattens out the high dispersion of the values of s for the sake of visibility and comparison between the two corpora.

To characterize each of these three areas around the diagonal in the references (or citations) of each corpus, we use the annual couples of ranks $r_R^d(t)$ ($r_I^d(t)$) of each represented discipline d in $\mathcal{P} \cap \overline{AI}$ and $\mathcal{P} \cap AI$, and we compute the temporal angle $\theta(t)$ that indicates its temporal deviation from the diagonal, here corresponding to the null angle $\theta = 0$. Then, within each zone for each year, we compute the average angle $\langle \theta \rangle$ and the standard deviation from the latter, both represented in Fig. 3.6.

The disciplinary composition of the common interest area, concentrated around the diagonals of all the maps, includes rather the same fields at each decade, both on reference and impact sides. According to the common ones that are more persistent over decades inside the references and received citations (see upper left boxes in Figs. 3.5A and 3.5B), the core of the observed citation dynamics lies in neuroscience and is composed by disciplines that are mainly associated with medicine and biomedical research, such that *Biochemistry & Molecular Biology*, *Behavioral Sciences*, *Clinical Neurology*, *Physiology*, *Cell Biology*, *Psychology*, *Psychiatry* and *Ophthalmology*. This zone is thus coinciding with the definition of neuroscience given by neuroscientists themselves, namely it “include[s] all fields that are involved with the study of the brain, the behaviors that it generates, and the mechanisms by which it does so, including cognitive neuroscience, systems neuroscience and psychology” (Hassabis et al., 2017, p.245).

This zone is also accompanied by disciplines whose ranks are more variable and that are associated with more technological aspects of neuroscience, such as *Computer Science*, *Engineering*, *Radiology*, *Neuroimaging* and *Audiology & Speech-Language pathology*. Fig. 3.7 represents the annual disciplinary evolution of this zone with the distance ρ of the disciplines d inside this zone from the origin of the 2D ranking maps (i.e., the point with ranks (0,0)), according to its location with the coordinates $(r_{\mathcal{P} \cap AI}^d, r_{\mathcal{P} \cap \overline{AI}}^d)$ associated with its respective ranks in the AI and non-AI corpora: $\rho = \sqrt{(r_{\mathcal{P} \cap AI}^d)^2 + (r_{\mathcal{P} \cap \overline{AI}}^d)^2}$. In particular, Fig. 3.7 shows a spectacular rise since around 1995 of the reference rankings associated with *Radiology* and *Neuroimaging* in both AI and non-AI corpora, as well as a growing impact of neuroscience articles published in this period on these disciplines. This trend thus testifies to the diffusion of these technologies in scientific and medical practices associated with neuroscience.

The special disciplinary ecosystem of the non-AI corpus $\mathcal{P} \cap \overline{AI}$, in blue on the ranking

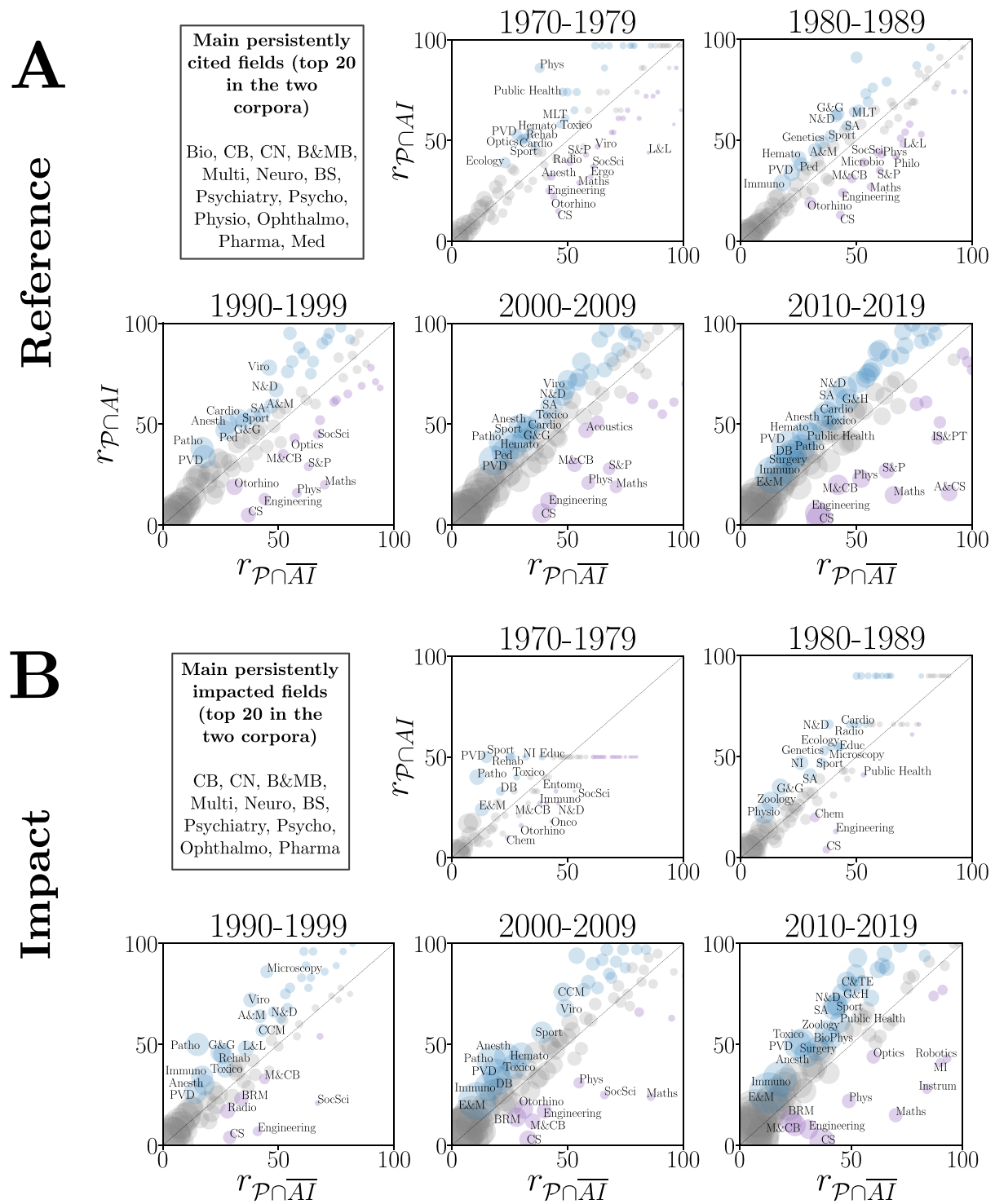


Figure 3.5: Time-aggregated ranking maps of the fields of study involved in publications that are *cited by* both of the studied corpora $\mathcal{P} \cap AI$ and $\mathcal{P} \cap \bar{AI}$ (**A**), and in publications that *cite* them (**B**). The dashed lines show the diagonal where the rankings are exactly the same in the two corpora. Only the most significant disciplines are shown, with the condition that they must appear simultaneously in both corpora. As mentioned in the main text, the sizes of the disks are normalized based on their empirical number of citations in the two different corpora in order to compare their respective citation weights within each corpus. The gray points correspond to disciplines that are situated in the area between the two lines of respective equations $r_{\mathcal{P} \cap AI} = r_{\mathcal{P} \cap \bar{AI}} + \tau$ (above the diagonal) and $r_{\mathcal{P} \cap AI} = r_{\mathcal{P} \cap \bar{AI}} - \tau$ (below the diagonal), with $\tau = 10$, and that shows no important variations of their respective positions from one period to one another. The most persistent disciplines over the 5 decades represented, with a rank lower than 20, are mentioned in the upper left box of each figure **A** and **B**. Finally, blue points are the most preferred disciplines of the $\mathcal{P} \cap \bar{AI}$ corpus, while purple ones are those of the $\mathcal{P} \cap AI$ corpus. Abbreviations of the apparent disciplines are provided in App. A.3.

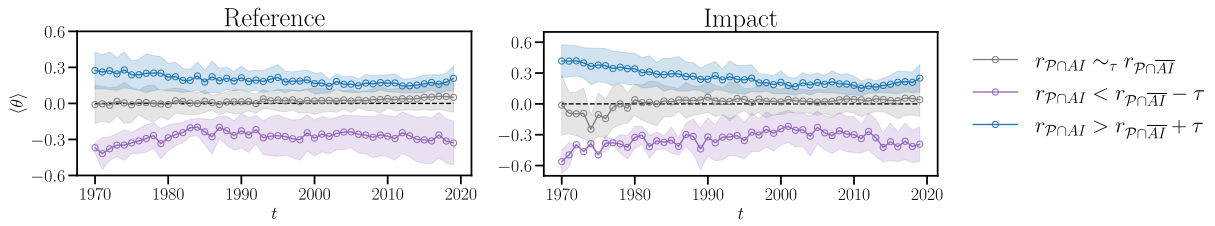


Figure 3.6: Temporal average angles produced by the disciplines in each area of the reference (left) and impact (right) diagrams of Fig. 3.5, with respect to the diagonals represented here by a dashed line at $\langle \theta \rangle = 0$. The colored area covering the curves represents their respective standard deviation from the mean angle. Angles are expressed in radians.

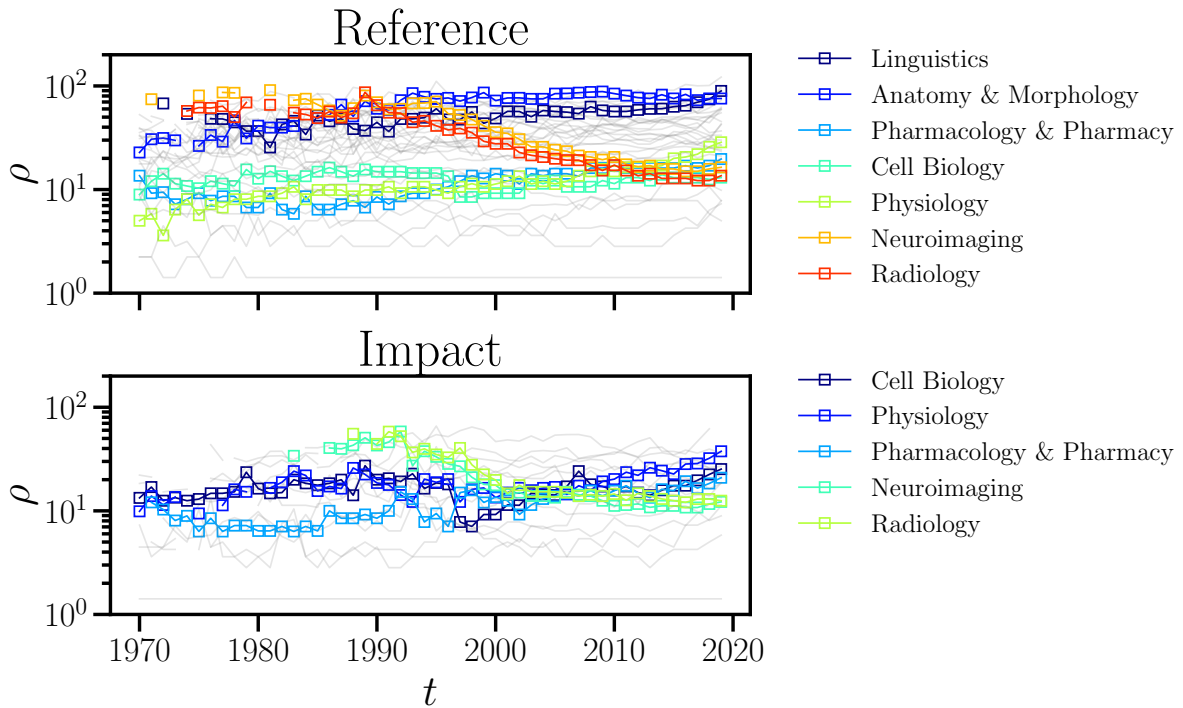


Figure 3.7: Time evolution of the distance ρ of the disciplines included in the common interest area, as defined and shown in Fig. 3.5. Only the curves with the most significant evolutions are highlighted with colors.

maps of Fig. 3.5, is also centered around biomedical fields of study that tend to be close to the disciplines that characterize the whole neuroscience as mentioned above. In particular, as shown in Fig. 3.6, the mean angles and standard deviations of the disciplines in the associated zone both assert a global concentration of them toward the common interest area, which includes, on the one hand, the most influential fields of research (reference), and on the other hand, the core of those that show the same interest for the two studied corpora (impact). We also notice in the reference angle plot of Fig. 3.6 a deviation from the diagonal of the disciplines in this zone of references since 2010, which is also shared by the disciplines in the common interest area of references. This suggests a recent shift of references shared by the two corpora toward fields of study preferred by the non-AI corpus.

Conversely, the special disciplinary ecosystem of $\mathcal{P} \cap AI$ evolves differently by representing the mathematical, computational, and technological part of neuroscience since 1970. The regular references to *Computer Science*, *Physics*, *Statistics & Probability*, *Mathematical & Computational Biology* and *Engineering* show a large influence of technological-oriented research in this particular AI research in neuroscience. Progressively concentrated toward the common interest area between 1970 and the late 1980s (see Fig. 3.6), these references preferred by $\mathcal{P} \cap AI$ become further away from the common interest area and more dispersed after this period, thus indicating a cognitive differentiation of references on which the AI-related corpus $\mathcal{P} \cap AI$ is drawing upon from the non-AI corpus $\mathcal{P} \cap \overline{AI}$. In addition, while neuroscience and associated medical fields – as *Clinical Neurology* and *Neuroimaging* – remain the primary stakeholders in the AI research conducted within it, the latter appears to be of varied interest, since the 1980s, to a subset of disciplines which do not place as much emphasis on works in the non-AI corpus and which are common to those cited preferentially by the AI-corpus over time, such as *Computer Science*, *Engineering* and *Mathematics for Computational Biology*, the last one exhibiting especially a spectacular increase of its own rank between 1990 and 2019 (see Figs. 3.5A and B²). Moreover, these special disciplines that cite most the corpus $\mathcal{P} \cap AI$ in Fig. 3.5B are as dispersed as the most cited ones by this corpus in Fig. 3.5A, and remain over time less close to the common interest area than the disciplines preferring the other corpus $\mathcal{P} \cap \overline{AI}$ (see the corresponding annual evolution of the average angle $\langle \theta \rangle$, colored purple on the impact side of Fig. 3.6). All these results thus show that AI research in neuroscience is situated in a particular disciplinary environment that is not shared by the core of the neuroscience field.

By regarding the disciplines with the most significant rank evolution within the AI ecosystem $\mathcal{P} \cap AI$, Fig. 3.8 shows the rise of *Neuroimaging* and *Radiology* both on reference and impact sides, as already observed in the common interest area, and also the progressive domination of *Computer Science*. This figure also shows the decrease in the influence of fields of research linked to the disciplinary orientations of neuroscience and especially of the corpus $\mathcal{P} \cap \overline{AI}$. This is also a sign of a progressive differentiation of the research supported through AI from the rest of neuroscience. The case of *Physiology*, which contributed to the foundations of neuroscience (Cooper and Shallice, 2010), is particularly eloquent in its distancing in time from the highest positions in the reference ranking of the corpus $\mathcal{P} \cap AI$ since the 2000s. This observed retreat of master neuroscience disciplines, which had a strong influence in the building of AI in this field since the 1970s, suggests that a social transformation occurred in this special research during the studied temporal period. Perconti and Plebe (2020) mention such a transformation, in which

²Fig. 3.5B shows also a particular proximity of the AI-related corpus with the field of chemistry, here represented with the JSCs *Chemistry* and *Biochemical Research Methods*, the latter giving more citations than the former since 1990.

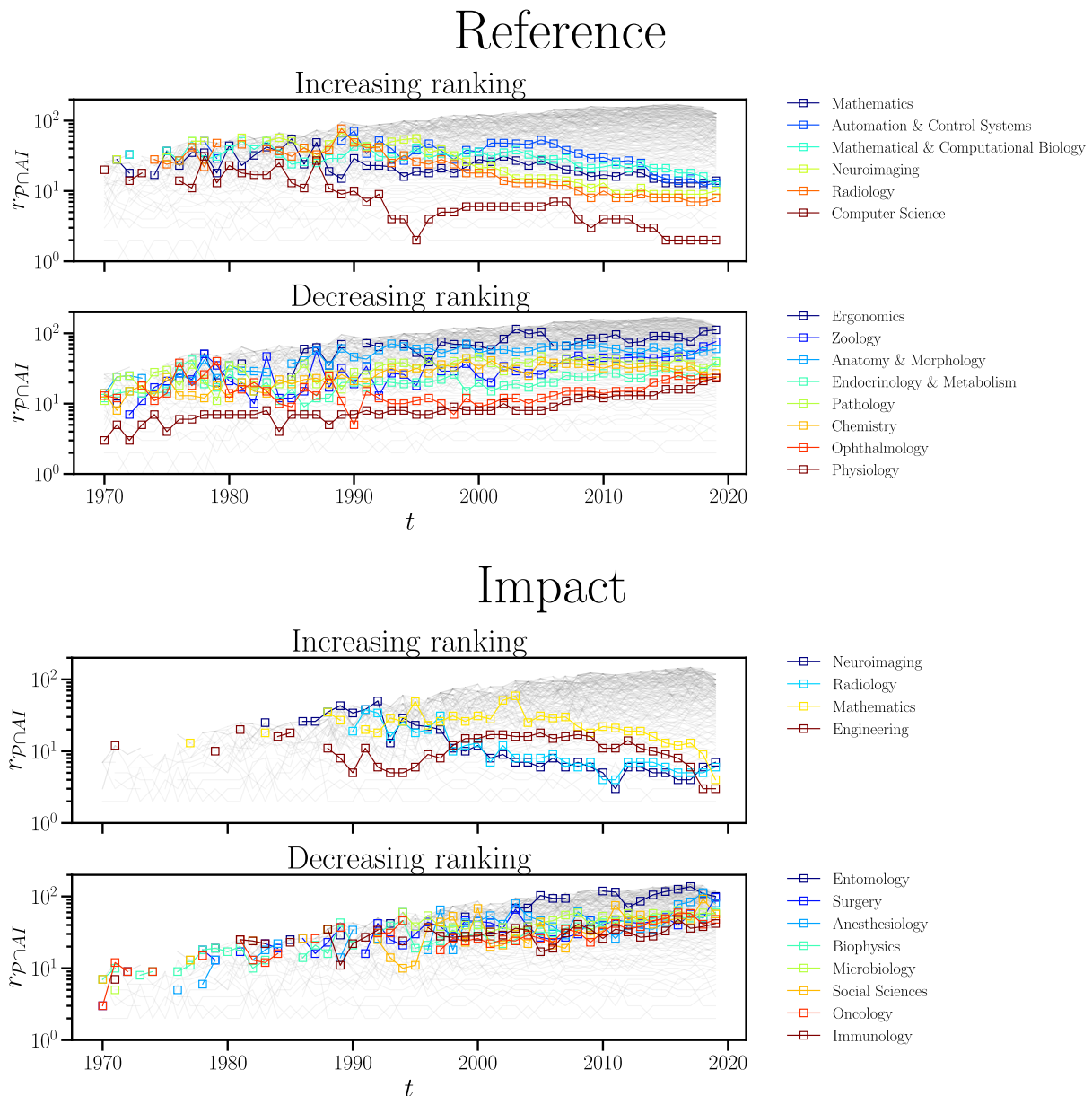


Figure 3.8: Time evolution of the ranking $r_{\mathcal{P} \cap AI}$ of the disciplines that are mostly represented by the corpus $\mathcal{P} \cap AI$ in the references (top) and citations (bottom), ie. situated in the lower zone of the maps shown in Fig. 3.5. Only the curves with most significant evolution are highlighted with colors.

AI in neuroscience was a matter for biomedical specialists before becoming an object of study and technological developments for engineers. This will be shown empirically in the following sections.

3.3 Who are the scientists making AI in neuroscience?

In this part, we focus on the profiles of the scientists doing AI in neuroscience and how they are inserted into the global authorship landscape of neuroscience. In particular, we

compare these AI practitioners with the other neuroscience researchers who have never published AI-related papers under three aspects, namely (1) their *disciplinary background*, (2) their *disciplinary trajectory* in academia, and (3) their surrounding collaborations within neuroscience. Here, we lie on the quartile classification of authors according to their respective share of AI-related publications (see Fig. 2.9 in Chapter 2), such as the authors gathered in the quartile \bar{Q} have a null AI-score $f_{AI} = 0$, those in Q_0 having $f_{AI} \in (0, 0.5)$, those in Q_1 having $f_{AI} \in [0.5, 1)$ and those in Q_2 having $f_{AI} = 1$.

3.3.1 Disciplinary profiles in each quartile of the f_{AI} distribution

Inspired by (Sugimoto and Cronin, 2012) (but with much less available data on the status of the authors), we define the disciplinary background of one author as the set of unique disciplines corresponding to the journals in which he/she has published in his/her first year of academic life, namely the year of his/her very first publication(s). For example, a fictitious author publishing his/her first three papers in the same year y_0 in two different journals labeled with disciplines $\{d_1\}$, $\{d_1, d_2\}$ and $\{d_3\}$ respectively, would have a disciplinary background composed of disciplines $\{d_1, d_2, d_3\}$. In this way, another fictitious author who has published only one paper in his/her first year of academic life (actually the most frequent situation, occurring 64% of times in the dataset) would have a disciplinary background composed only of the discipline(s) labeling the corresponding journal.

Then we compute for each quartile the temporal cumulative number of new scientists trained in each represented disciplinary background in this quartile, as shown in Fig. 3.9. In this way, we assess for each quartile the main native specialties in which authors have first published.

Fig. 3.9 shows that the profiles included in Q_0 and \bar{Q} overall the period 1940-2019 are very similar. They are mainly confined in biomedical research around neuroscience, as well as in *Multidisciplinary Sciences* which is represented by 91 international journals with a broad topical diversity, such as *Scientific Report*, the *Journal of the Royal Society Interface* and *Plos One*. We recognize also the main fields of research that are shaping the common interest area of the disciplinary ecosystems of the two separated AI and non-AI corpora (see the most persistent disciplines in references and citations in Fig. 3.5). These two profiles best represent the disciplinary spectrum of neuroscience itself. We also notice the spectacular increase of the *Neurosciences* curve in the two plots until the 1970s (after having emerged in 1957 for \bar{Q} and in 1962 for Q_0), followed by a quasi-linear progression until 2015. This boom in neuroscience profiles in these quartiles suggests that modern neuroscience is progressively being institutionalized as a well-structured discipline in science.

The profiles of Q_1 and Q_2 shown in Fig. 3.9 are at the opposite of the previous ones, coming at most from fields of research related to *Computer Science*. The specialty *Computer Science*, *Neuroscience*, which emerged in 1988 for Q_1 and in 1991 for Q_2 , indicates the rise of a group of scientists that are specialized in computation in neuroscience. Notice that the behavior of the *Neurosciences* curve within Q_1 in Fig. 3.9 appears later than computational and engineering profiles, suggesting also that AI-related knowledge and technological tools penetrate progressively the global neuroscience field.

From the subset of authors belonging to the most frequent disciplinary background in each quartile, we consider their respective disciplinary profile as the disciplines corresponding to the journals in which they publish throughout their scientific life – that would not be ended for the youngest still publishing in 2019. We therefore draw from these two parameters – disciplinary background and career-related disciplines – the typical disciplinary trajectories in each quartile, which are shown in Fig. 3.10. Since the backgrounds are built from the publications in a given year, the authors who began their career in 2019 would have a disciplinary profile corresponding to it. To avoid an over-representation of some confined disciplinary trajectories due to these newcomers, we therefore consider only the authors who began at the latest in 2018. For the sake of clarity, we select only the most significant trajectories.

As a confirmation of the disciplinary background shown in Fig. 3.9, Fig. 3.10 shows that the studied authors in \bar{Q} and Q_0 are involved in a similar disciplinary landscape centered around *Ophthalmology*, *Clinical Neurology* and general *Neurosciences*, while Q_1 and Q_2 are more confined into *Computer Science* and *Engineering*. In particular, by considering neuroscience as the subset of disciplines composed of the WOS disciplines *Neurosciences*, *Clinical Neurology* and *Neuroimaging*, we observe that \bar{Q} and Q_0 are more involved in that field of research, with respectively 77% of the authors in the first and 78% of those in the second with a disciplinary profile that includes one or more of the fields of research associated with neuroscience. On the contrary, 45% of the authors in Q_1 and 42% of the authors in Q_2 have a disciplinary profile that includes such JSCs. These two last quartiles are therefore mainly detached of the neuroscience goals given the disciplinary backgrounds and profiles of their main respective authors. These results thus show that neuroscience is a field that brings together heterogeneous profiles who seem to serve different epistemic objectives inside and outside neuroscience (Sedooka et al., 2015).

Nonetheless we notice two special things about these last quartiles, which can be interpreted in the light of the classification of interdisciplinary profiles of Sedooka et al. (2015) First the authors who began into the specialties including one or more neuroscience-related disciplines tends to continue in the same field of research, which is sometimes interdisciplinary, such as the combinations *Computer Science*, *Neurosciences* and *Neuroimaging*,

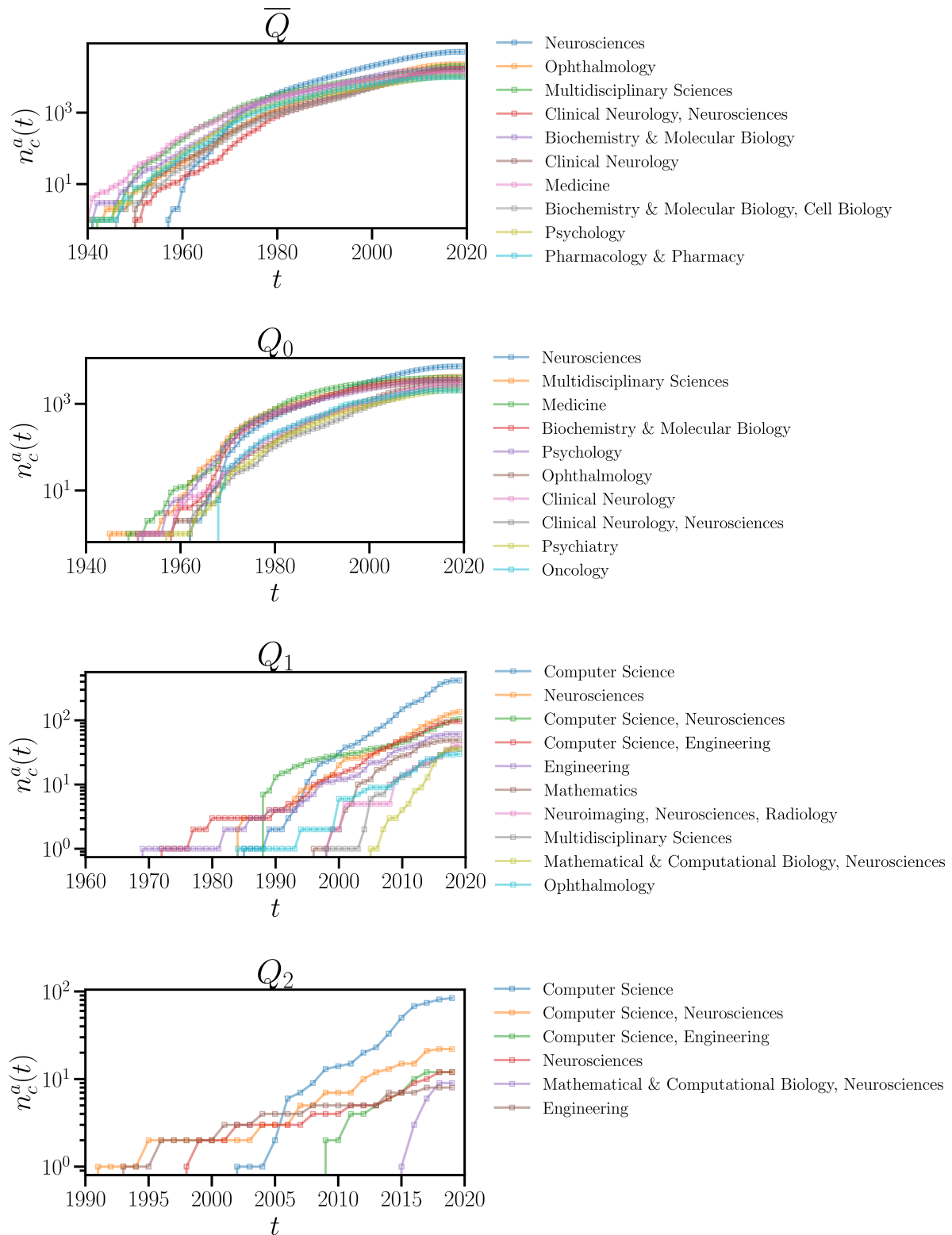


Figure 3.9: Cumulative number of new authors per disciplinary background for each quartile. For one plot, the point of the curve of the specialty d at year t is the number of authors who have published their first articles inside d during their very first year of academic career, since the year of first appearance of d . Only the top 10 native specialties in 2019 are shown in \bar{Q} , Q_0 and Q_1 , and the top 6 for Q_2 because of the insignificance of the following ones.



Figure 3.10: Significant disciplinary trajectories of the authors belonging in each quartile. On each plot lie on the left the most significant disciplinary backgrounds, and on the right lie the career-related disciplines up to 2019.

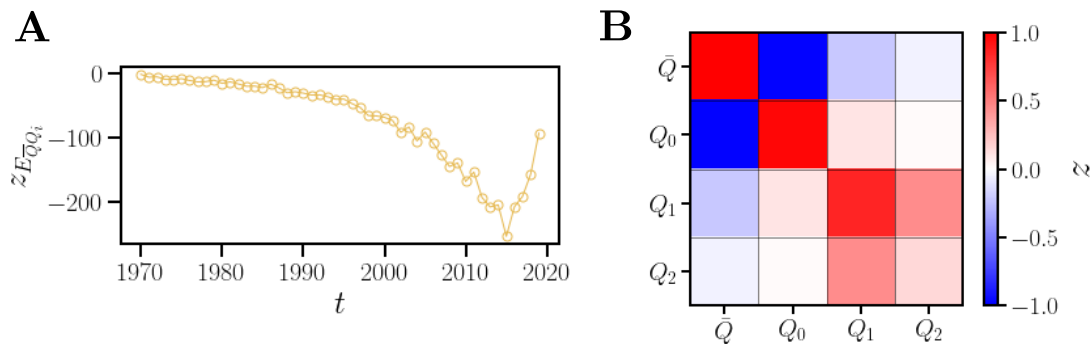


Figure 3.11: **A:** Temporal z -score applied to the edges in \mathcal{T} between the scientists in \bar{Q} and those in the other Q_i , here aggregated together under the notation Q_i . **B:** Z -score matrix of the edges between each quartile in the time-aggregated collaboration network \mathcal{A} . The values are normalized with the absolute maximum one in the matrix.

Neurosciences, Radiology. This means that these scientists are trained in a disciplinary context centered around neuroscience, with a rather technical component – whom Sedooka et al. call “interdisciplinary natives”. Second, some authors who began into *Computer Science, Engineering* or mathematics-related disciplines – who seem to be more distant from neuroscience in their disciplinary backgrounds – continue into neuroscience, as shown by the combination of their originating field(s) of research with neuroscience ones in their disciplinary profiles in Fig. 3.10, for instance *Computer Science, Engineering* leading to *Computer Science, Engineering, Neurosciences* – the “interdisciplinary migrant”. The late emergence of interdisciplinary profiles in computer science and neuroscience, who are more involved in neuroscience in general, such as those in Q_2 , also testify that the AI community under study is taking root in the global neuroscience landscape, the more recent profiles in neuroscience becoming *insiders* in this new technological specialty. All these scientists thus represent a special labor force for neuroscience whose main expertise lies in AI, and more generally in mathematical, computational, and technological tools (Perconti and Plebe, 2020).

3.3.2 Segregation between AI practitioners and non-AI ones

We investigate in the following how AI practitioners are distributed within the neuroscience community and how the collaborations between them are shaped, especially between the different kinds of scientific profile described before.

We first consider the temporal collaboration network \mathcal{T} built in Section 2.2.3.2, and we evaluate the temporal standardized share of edges between scientists belonging, respectively, to \bar{Q} and to all other quartiles Q_i (see Section 3.1.2 for the calculation of such score). This score is showed in Fig. 3.11A, from which we deduce two facts.

First, all its values are lower than 0 over time, meaning that the number of edges

between the two ensembles \bar{Q} and $\{Q_0, Q_1, Q_2\}$, is lower than the average computed from several random distributions of all edges between the authors in \mathcal{T} . This fact thus shows that AI practitioners within neuroscience and other neuroscientists maintain few collaborations as \mathcal{T} grows over time. Second, this situation worsens over time, with a steady decline until the 1990s followed by a steeper one toward 2015. Although this tendency has reversed since 2015, these results indicate that the neuroscientists making AI, ie, in Q_i and not in \bar{Q} , are shaping an almost independent community within neuroscience by progressively widening the gap with other neuroscientists belonging to the set \bar{Q} . Furthermore, concerning especially the authors in Q_0 , who have disciplinary backgrounds and trajectories similar to the authors in \bar{Q} (see Figs. 3.9 and 3.10), the results advanced above also suggest that the scientists in Q_0 became progressively *outsiders* from the subset \bar{Q} in the history of neuroscience by moving closer to AI – even if the scientists in Q_0 have few AI-related publications. This social separation around AI research does not produce, however, a strict knowledge divide in neuroscience, as demonstrated before with the temporal similarity index of disciplines directly impacted by neuroscience in Fig. 3.4, as well as with the temporal evolution of the disciplinary landscapes in Fig. 3.5, both indicating that the quartiles Q_i export their knowledge to the whole neuroscience as much as the quartile \bar{Q} .

This social divide is confirmed with Fig. 3.11B, which represents the panorama of the links shared between the different quartiles in the time-aggregated collaboration network \mathcal{A} (see also Section 3.1.2). This diagram represents a stabilized situation observed since the end of the 1990s (see Fig. B.3 in App. B.2), where the scientists in Q_i cosign more with one another than with \bar{Q} while the scientists in the second ensemble prefer to collaborate with one another as well. More precisely, it exhibits a polarization within the ensemble $\{Q_0, Q_1, Q_2\}$ of AI practitioners in neuroscience. Indeed, the authors belonging to Q_1 and Q_2 are more strongly connected together than with those in Q_0 . This can be explained by the disciplinary proximity of the authors in Q_1 and Q_2 observed in Fig. 3.9, especially the prominence between the late 1980s and the early 2000s of full computer scientists and hybrid profile publishing in journals labeled as *Computer Science*, *Neurosciences*. Furthermore, the links between Q_1 and Q_0 are much more important than those between Q_2 and Q_0 . The scientists in Q_1 thus appear to be the most interdisciplinary in ensuring the bridge between these differentiated profiles within the AI community. These other *outsiders* in neuroscience are especially good candidates to drive the diffusion of AI in neuroscience from computation to medical and clinical applications, given their disciplinary trajectories shown in Fig. 3.10.

As they are gradually sidelining from the neuroscience community, which is mostly composed of scientists who are not familiar with AI, we test in the following whether the AI practitioners described above remain in the core of the collaboration network – which

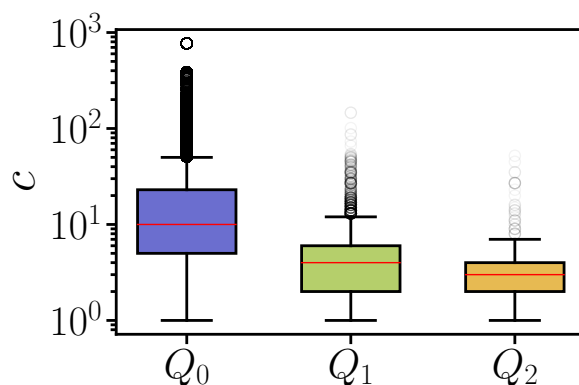


Figure 3.12: Distribution by group Q_i of the core number c of the AI practitioners in the time-aggregated collaboration network \mathcal{A} . The red lines inside each box represent the medians, which amounts to 10 for Q_0 , 4 for Q_1 , and 3 for Q_2 .

could foster the social diffusion of AI – or move to the periphery of the collaboration network – which, on the contrary, could hinder this process. To this end, we perform a k -core decomposition of \mathcal{A} to derive the core number c of each AI practitioner within neuroscience, i.e., their membership in a subgraph of degree at most c (the c -shell) but not in the next core of minimal degree $c + 1$ (Batagelj and Zaversnik, 2003). The lower the metric c , the farther away from the core the individual author is. The distribution of this measure by the Q_i groups is shown in Fig. 3.12. We note here that the maximum core number at the end of the k -core decomposition of \mathcal{A} is 770, which is reached by 302 individuals in quartile Q_0 (0.02% of its own population).

Knowing this last information, Fig. 3.12 shows that the scientists in Q_0 are closer to the core of the collaboration network – as they show a disciplinary proximity with the profiles of the non-AI specialists within \bar{Q} – than the scientists in the other two quartiles Q_1 and Q_2 , both of which exhibit a concentration of their core number distribution around their respective medians, far from the median of Q_0 . Furthermore, we notice that 93% of the population of Q_0 lies within the giant component of \mathcal{A} , while 74% of the scientists in Q_1 and 72% of those in Q_2 do, the rest of the last two quartiles remaining isolated from the core of the collaboration network (perhaps fragmented in smaller groups), which is concentrated within the giant component. The great majority of the collaborations between these AI practitioners thus take place mainly in the periphery of the network. Since these two quartiles show strong connections together, as already shown in Fig. 3.11B, we can conclude that the most AI experts in neuroscience are situated in a (rather) distant periphery of the whole collaboration network.

In summary, in addition to the strong segregation phenomenon already shown in Fig. 3.11 and the important disciplinary disparities between scientists with the highest AI expertise and those with the lowest or no such expertise (see the results of the previous Sect. 3.3.1), the above findings show that AI practitioners are distancing from

the neuroscience collaborative core over time, as observed in Fig. 3.12. This phenomenon thus appears to hinder the diffusion of AI-related ideas throughout the community under study.

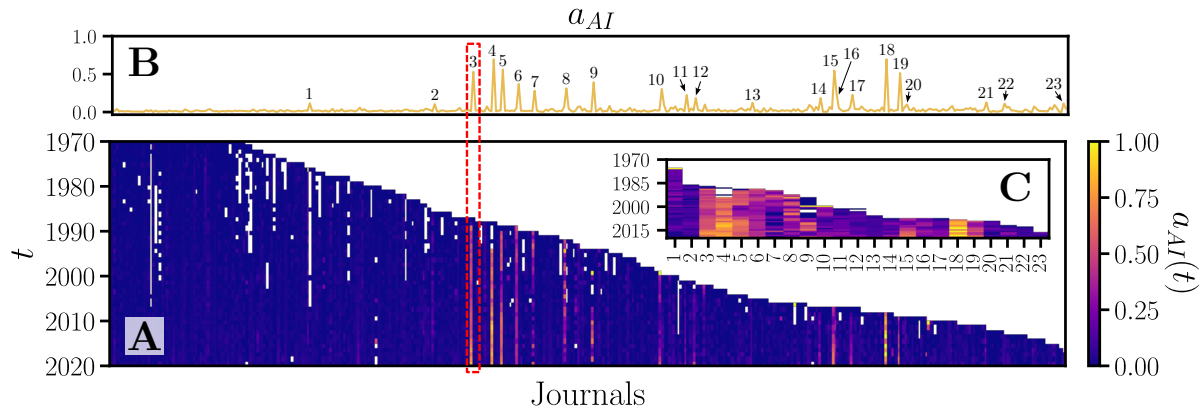
3.4 An AI literature confined in a small set of specialized journals

Another factor of differentiation of AI research from the core of neuroscience lies in the set of journals in which the former are mainly published. Here, we analyze how AI is promoted within the 421 journals included in the whole neuroscience dataset and who are the authors heading to the neuroscience journals publishing most AI research.

Fig. 3.13 offers an answer to the first aspect, by summarizing the evolution of the AI activities of all the neuroscience journals listed in our dataset. It shows especially that the AI-related research in this field is concentrated around a small subset of journals providing development in computational techniques mainly linked to neural networks and cognition. These journals are quite representative of the connectionist wave of AI that has been active in neuroscience since the late 1980s (Cardon et al., 2018). They represent 32.7% of the scientific production of AI in neuroscience during the whole period 1970-2019, according to our dataset, the rest being distributed among the 398 other journals with a much smaller AI-activity (less than 10%).

The spectrum of a_{AI} (top yellow curve in Fig. 3.13) shows that the launches of the most AI-active journals in the neuroscience field are concentrated around three periods. The first one spans the period 1987-1994 with especially 7 journals (no. 3 to 9 in the table of Fig. 3.13) whose a_{AI} is higher than 28% and whose scopes are oriented toward computational neuroscience and the use of neural network formalism for complex calculations, especially the simulation of cognition on neural systems. These journals, except no. 7 and 9 (respectively *Network: Computation in Neural Systems* and *Neural Network World*) are showing well-sustained activity on AI research. This period commonly falls into the second “AI winter”, when AI-research funding and the production of scientific results and associated technological solutions were at their lowest for a second time (Cardon et al., 2018; Schuchmann, 2019). Paradoxically, neuroscience is especially active in such research in this period, as shown by the strong, long-lasting AI-activity of these journals created in this period.

The second period includes the journals 10, 11 and 12, which were launched between 1999 and 2001. The journal *IEEE Transactions on Neural Systems and Rehabilitation*



	Journal	Launch year in the database	a_{AI}
1	Cognitive Science	1976	0.12
2	Journal of Sensory Studies	1986	0.10
3	Neural Networks	1987	0.53
4	Neural Processing Letters	1988	0.70
5	Neurocomputing	1989	0.56
6	Neural Computation	1989	0.37
7	Network: Computation in Neural Systems	1990	0.28
8	Adaptive Behavior	1992	0.31
9	Neural Network World	1994	0.39
10	Cognitive Systems Research	1999	0.30
11	IEEE Transactions on Neural Systems and Rehabilitation Engineering	2001	0.22
12	Neuroinformatics	2001	0.19
13	Plos Computational Biology	2005	0.12
14	Frontiers in Neuroinformatics	2007	0.19
15	Computational Intelligence and Neuroscience	2007	0.55
16	Cognitive Neurodynamics	2007	0.24
17	Frontiers in Computational Neuroscience	2007	0.23
18	Evolutionary Intelligence	2008	0.70
19	Cognitive Computation	2009	0.51
20	Topics in Cognitive Science	2009	0.10
21	Journal of Mathematical Neuroscience	2011	0.13
22	NeuroImage: Clinical	2012	0.11
23	Biological Psychiatry: Cognitive Neuroscience and Neuroimaging	2016	0.11

Figure 3.13: Top diagrams: **A**: Temporal AI activity of the journals. A square at position (j, t) in the diagram is the share of AI publications in the journal j at time t . One column of the plot is the AI activity of a single journal. **B**: Distribution of the global AI activity of the journals over all their publications until 2019. The horizontal axis corresponds to the journal axis of the lower plot A, as shown by the red dashed zone. **C**: Zoomed activities of the 23 most active journals in AI research, with a global AI activity higher than 10%. They are indicated in the lower table.

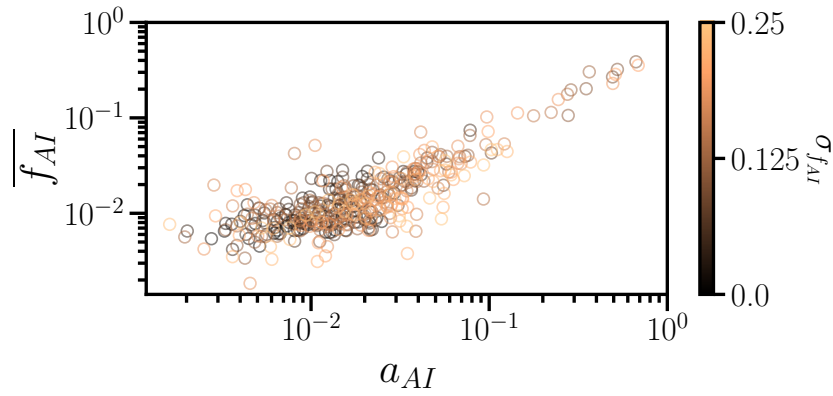


Figure 3.14: Correlation plot between the global AI activity a_{AI} of the journals appearing in \mathcal{P} and the average AI scores $\overline{f_{AI}}$ of the authors who are published in them between 1970 and 2019. Each point represents one journal in the horizontal a_{AI} space, colored by the standard deviation of the f_{AI} scores of their associated authors. The maximum average value of $\overline{f_{AI}}$ is situated around 0.4.

Engineering is especially oriented toward the development of computational methods and technological tools to capture the neural activity of the brain. This period is followed by another until 2007 which did not lead to the launch of AI-active journals.

Finally, the third period following the latter, between 2007 and 2009, saw the launch of numerous journals (no. 14 to 20) in a particularly short time, that are fostering research at the crossroads of neuroscience and cognitive science. It includes especially the most AI active journal of our dataset, *Evolutionary Intelligence*, whose scope is oriented toward evolutionary computation, a subset of the field of optimization. Its temporal activity is also strongly sustained over the whole time period covered in our datasets.

Aggregating over time, Fig. 3.14 shows a linear correlation between the mean AI-activity in the journals and the average of the AI scores f_{AI} of the authors who have published at least one paper in these journals. The tail at highest a_{AI} (higher than 0,1) corresponds to the top 15 of the most AI-active journals referenced in Fig. 3.13. This particular result thus unveils the attraction of the authors publishing the most AI-related works of our dataset in the journals with high AI-activity over time. With all the results introduced in the previous sections, it shows as well that AI research in neuroscience is done by a specialized scientific community inside this field with its own journals for communicating results.

3.5 Discussion

As previously assumed, the success of the diffusion of an instrument (in the form of technical knowledge and devices) in science, which is the last stage of its development

within a research-technology enterprise (Shinn and Joerges, 2002), depends greatly on its ability to achieve *genericity* in a variety of fields of research. Considering AI as such an instrument, this chapter has been devoted to examining two particular dimensions of the penetration of AI into neuroscience – here supposedly exogenous to the socio-epistemic environment in which it has been created – namely *domain adaptation* and *social adoption*.

First, let us examine what we have found about the first aforementioned genericity criterion at the heart of this chapter. The Jaccard similarities exposed in Sect. 3.2.1 show that the citation environment of AI embeds in this neuroscience at macroscale over time, by reproducing the main citation patterns that are characterizing this domain – i.e., toward the disciplines structuring it – and by increasingly impacting the disciplines that show the most interest in it – including neuroscience itself. Therefore, the global disciplinary environment of AI research seems to be inscribed at first glance into the neuroscience arena. However, a second look at the level of the individual disciplines constituting this arena reveals a demarcation process of such an AI-related subdomain within neuroscience, with a differentiated disciplinary ecosystem composed not only of the latter and other related disciplines, which remain the first influence and recipients of the outputs produced by the studied AI research, but also around STEM-related disciplines such as computer science, mathematics, statistics, physics, engineering, and instrumentation, which are in the periphery of the rest of the neuroscience citation ecosystem. This disciplinary demarcation is also operationalized by the constitution of a small set of journals, albeit labeled “neuroscience” by the WOS and SJR bibliometric classifications, which promotes the results of such research to an expert audience.

These first findings thus demonstrate that these AI-related publications within neuroscience seem to serve different disciplinary interests: a pronounced technological orientation and the production of formal knowledge from various fields within the natural sciences, while still serving the purposes of the host discipline, with citations firstly oriented towards the top disciplines constituting the neuroscience ecosystem, from chemistry and molecular biology to health-oriented disciplines, such as ophthalmology, pharmacology, physiology, psychiatry, and psychology. AI as produced in neuroscience can be seen in this case as a *trading zone*, an intermediate place at the crossroads of these multiple disciplines divided into these two main epistemic orientations (Galison, 1997; Grauwin et al., 2012), thus suggesting the existence of transfers of knowledge and technical tools between them. Although not addressed in this thesis, this last point could be investigated in further research through the construction of chronological knowledge flows between publications, as done in various ways in (Di Bona et al., 2023; Franceschet and Colavizza, 2020; Schäfermeier et al., 2022).

These two main disciplinary orientations are crystallized in the profiles of the scientists

involved in the specific workforce producing the instrument. Indeed, we have distinguished two classes in this group of scientists, namely a first wave of outsiders, including authors who are trained in the main disciplines that have shaped neuroscience since the 1940s and who have a low number of AI-related publications, and a smaller second wave, which emerged around the 1980s and includes authors who are trained in other disciplines that are not represented in the former group, such as computer science and engineering, and who exhibit the highest activities in AI research in general (inside and outside neuroscience). Furthermore, this second group is not the most involved in the field of neuroscience, as its members continue to publish within their original disciplines, which are mainly STEM.

This social polarization within AI practitioners in neuroscience suggests that, over the years, AI is becoming a set of research technologies that needs to be shaped not only by neuroscientists themselves but also with the help of scientists who come either from outside neuroscience, or within it but with an interdisciplinary background and presenting a specific expertise in AI itself. These observations reflect the specialization and engineering turn of AI in science since the 1980s and 1990s, which have already been widely documented in the literature (Ahmed et al., 2023; Frank, Wang, et al., 2019; Kirtchik, 2019; Klinger et al., 2022; “The new NeuroAI” 2024). Besides, the second wave of AI researchers described before represents the mobility of such experts to other fields of research, in order to propose and integrate the associated knowledge and technologies to achieve some of the disciplinary objectives of these receiving fields or, less ambitiously, to solve some technical problems that could not be solved with more conventional tools. According to Joerges, Marcovich, and Shinn (Marcovich and Shinn, 2012; Shinn and Joerges, 2002), these characteristics thus confer to these scientists a *research-technologist* status, engaged in the diffusion of AI outside their originating STEM disciplines and throughout the science system. However, most of them remain oriented by their individual disciplinary framework, with their own epistemic objectives that would not necessarily be compatible with those of neuroscience. This suggests that, for these scientists gathered under the combined banner of engineering, computer science, and AI, neuroscience seems to be only one area of application of AI among others, thus making it an episodic, *transitory* step in their entire careers – this will be detailed later with the fieldwork introduced in Chapter 5. This interpretation, however, faces a main limitation directly given by the WOS journal classification used to establish the disciplinary profiles of these AI practitioners (see Fig. 3.10 in Sect. 3.3.1). Further work mobilizing the textual content of the articles in our database and another topical classification, such as the OpenAlex’s, would be much appropriate for this purpose.

In addition, this AI workforce within neuroscience, especially the most expert in Q_1 and Q_2 , is situated on the boundary of the collaboration network and is increasingly separated

from the rest of the neuroscience community between 1970 and 2015, with a few links still maintained. According to the citation ecosystem of the specific AI research in neuroscience depicted above, this social distancing process reflects an epistemic closure around AI, around which the involved scientists show sustained collaborations, while continuing to publish mainly in their respective disciplines of origin. Although not explored here (and, more broadly, not in this thesis), this epistemic closure could be a sign of the creation of an organized, *coherent* social group specializing in AI inside a neuroscience, gathered not only around the same journals, knowledge basis and repeated collaborations (Chubin, 1976; Griffith and Mullins, 1972; Wray, 2005), but also around common bibliographical references (Griffith, H. Small, et al., 1974; H. Small and Griffith, 1974), individual figures consolidating an emerging scientific specialty (Cole and Zuckerman, 1975), and reinforced citations between collaborating authors (Espinosa-Rada et al., 2024; Wallace et al., 2012). In this respect, and inspired by the latter two references, a future research direction would be to combine the collaboration structure and the citation network between authors in order to interpret our preliminary results as a development of such an AI specialty within neuroscience. However, such an endeavor, especially the building of the inter-scientist citation network, would require all the authors of our original, unfiltered database, thus directly challenging the filtering criteria established in Chapter 2. Nevertheless, this specific data structure would constitute a valuable tool to study the social diffusion of AI in neuroscience without focusing solely on the effect of diffusion on the co-authorship network, as done in this present chapter.

Futhermore, it would also be worth testing this diffusion across the neuroscience community by comparing the results of our data-driven analysis since 1970 with a model describing the dynamics of adoption of AI within this community, in order to possibly predict when the complete adoption of the tools by the whole community would happen, or at least an evolution trend for the next few years (if its hype is not disrupted by the arrival of another, even more powerful technology). In this line of thought, Carley (1989, 1991) has proposed to model such a social transfer of knowledge by considering a network of agents, each possessing a knowledge base that would be adapted as they share knowledge with their neighbors. In the same vein, Roth and Cointet (2010) have suggested a kind of extension of this mechanism, which would be effective as long as the knowledge bases of the collaborating agents are similar.³ This previous contribution may be compared to a complex contagion model (Cencetti et al., 2023; Centola, 2015). This type of model generalizes epidemiological models on complex networks, where the transmission of information or innovation is effective with only creation of links between scientists, by considering the transmission mechanism as resulting from the combination of many other

³A combination of the frameworks suggested by Carley (1989, 1991) and Roth and Cointet (2010) would lead to a model comparable to various opinion models with bounded confidences, such as the heterogeneous Hegselmann-Krause model recently examined by Hernández, Perrier, and Schawe (2020, 2024).

factors that could foster the adoption of the innovation under study by scientists and reinforce research collaborations around this innovation – what Centola (2015) has called a *consolidation*⁴ mechanism. Actually, such a model could provide evidence to support the aforementioned epistemic closure hypothesis. However, regardless of the model chosen, we believe that AI would be socially adopted within neuroscience through a knowledge transfer dynamic of the tools starting from the highest experts in AI, i.e., the last two quartiles of outsiders Q_1 and Q_2 described in Sect. 3.3, to the lowest ones in Q_0 , who are the most inserted in the community and would therefore play a pivotal role in promoting (or hindering) the diffusion across the whole domain, mainly represented by members in the last quartile \bar{Q} who have never used or mentioned AI at all.

We also notice that, although the development of AI in science is statistically characterized by a disciplinary closure around STEM-related disciplines between 1980 and 2010, according to Gargiulo et al. (2023), AI still has continued its interactions with neuroscience for its own epistemic purposes during this period, as evidenced by a large number of journals created during this period that are very active in AI research (see Section 3.4), and by its global citation impact in the discipline (see Sections 3.2.2 and 3.2.1).

3.6 Summary

In this chapter, we have shown the dynamics of the disciplinary and social structures related to AI research within the field of neuroscience since 1970. In particular, while this specific research area is integrated into the disciplinary environment of neuroscience and is situated at the nexus of multiple disciplines (not only related to the latter) that also benefit from the knowledge and technologies it produces, it is nonetheless progressively differentiating itself from the host domain, as evidenced by both: 1) the constitution of a second epistemic orientation towards a STEM-related disciplinary environment, which shapes contemporary AI research in the whole science since the late 1980s (Gargiulo, Fontaine, et al., 2023), and 2) a social segregation of AI practitioners within the neuroscience community, especially those with a computer science or technological background, gathered at the periphery of neuroscience around a few journals that publish most of the

⁴This mechanism, strongly depending on the topology of the social network under study, comes directly in opposition to the well-known Granovetter’s hypothesis (1973) according to which the diffusion is facilitated by the presence of few bridges (called *weak ties* by this author) between social groups where individuals share similar traits and are densely and strongly linked, i.e., strong homophily and consolidation within each social group. Instead, Centola (2015) has proposed that the diffusion is successful for several combinations of the two last parameters’ intensities. Notably, one of them, in contrast to Granovetter’s model, exhibits strong homophily (i.e., formation of densely linked social groups) and moderate consolidation, thus allowing the creation of numerous interactions between the different formed social groups. Nevertheless, this consolidation is not always observed empirically, as demonstrated by the work of Manzo et al. (2018) on the diffusion of some pottery’s practices among different rural communities.

AI-related work in the field.

This pattern of development of AI within neuroscience thus confirms its *adaptation* to the field under study (as defined in the Introduction), but does not testify to a widespread *social adoption* of the associated instruments by neuroscientists, which does not support the idea of a *generic* AI in the domain under study. Moreover, the current configuration of this AI-related research subarea seems to hinder the circulation of its results throughout the whole field. But is this really the case when focusing on the encoded knowledge within the articles under study?

The next Chapter 4 proposes to go beyond the discipline-aggregated representations of the citation and social organization of neuroscience depicted in this chapter by representing the field of neuroscience as an evolving body of knowledge embedded in the textual elements of its publications, i.e., their titles and abstracts. With such a representation, we will be able to identify, in particular, which research subfields, addressing various topics covered by neuroscience, AI has entered over time and whether it has incorporated the core of their respective conceptual frameworks, which is a translation of our third *genericity* criterion.

Exploring the epistemic integration of AI in the neuroscience knowledge landscape

In the previous chapter, we have demonstrated that neuroscience remains the primary stakeholder of the AIs produced within its own disciplinary environment. However, this analysis did not identify which epistemic components of neuroscience actually mobilize AI. Indeed, the disciplinary aggregated citation network used in the previous chapter significantly obscures the knowledge organization of neuroscience. In addition to its inherently multidisciplinary nature, this field of research may also be structured around subfields, research topics, and encoded knowledge within the published articles – which may involve interactions between multiple disciplines. This leads us, therefore, to identify the knowledge areas in which AI is most developed, which means that the latter might be more suited to address some neuroscience topics than others, and to evaluate the scope of the outputs of the AI research across the entire field, in other words, whether the associated AI knowledge is then disseminated in all the neuroscientific research subfields with their own respective conceptual frameworks. Such an ability of AI to transfer into different neuroscience subfields and thus to be a multipurpose tool in any epistemic framework would also be a signal of a genericity fully achieved within the knowledge base of the host field (Hentschel, 2015; Shinn and Joerges, 2002).

In this chapter, we tackle the aforementioned questions by drawing a temporal cartography of the neuroscientific knowledge space embedded in our corpus built in Chapter 2. Such a dynamic map is valuable for locating AI research contributions (corpus $\mathcal{P} \cap AI$) in the different epistemic regions of the neuroscientific landscape (explored in Sect. 4.2). In Sect. 4.1, we detail the numerical construction of this map and introduce other data structures used throughout this chapter to assess the third genericity criterion of AI, *epistemic integration*. By mobilizing semantic networks representing the conceptual universes of the inferred subfields of the map, we will study the conceptual articulation of AI-related knowledge with others associated with general neuroscience purposes (in Sect 4.3). And

with an internal citation network between the articles used to draw the map, we will analyze the diffusion of AI through the knowledge space of neuroscience (in Sect. 4.4). We finally discuss all the results in Sect. 4.5.

4.1 Methodology

4.1.1 Representing the knowledge space of neuroscience

Inspired by recent efforts to map scientific knowledge as faithfully as possible (González-Márquez et al., 2024; F. Liu et al., 2024; Singh et al., 2024), we need a large language model (LLM) capable of transforming the textual metadata of the articles in our corpus, here their titles and abstracts, into vectors that are easier to manipulate numerically, representing their knowledge and the contextual features in which it is stated (through specific sentence constructions or recurrent word associations), in order to evaluate the degree of lexical similarity between two given papers. The well-known *Bidirectional Encoder Representations from Transformers*, also known simply as BERT (Devlin et al., 2019), successfully achieves such a task. This model, now available in different versions for different usages, relies on various intricate *attention*-based mechanisms that shape a *transformer* (Vaswani et al., 2017), an architecture that is mainly used to predict word associations between sets of tokens lying in different layers of a large neural network. They require a huge pre-training step that consists of generating consistent sentences given the previous ones in a text (*next sentence prediction*), and of inferring a word of the right type (with also the right gender and number) given the surrounding others in a sentence, a paragraph or a longer piece of text (*masked language modeling*) – the latter process partly generalizes the also famous lexical embedding models *word2vec* (Mikolov et al., 2013) and *Glove* (Pennington et al., 2014). With such architecture and training, the BERT transformers are able to learn from scratch the morphosyntactic properties of the language(s) employed in the entire training corpus and to build their own vocabulary basis (Martin et al., 2020). The learning process of such an algorithm is, therefore, commonly said *self-supervised*. They are also able to encode these features into (often) high-dimensional context-aware vectors that represent a piece of text within a given corpus.

Here, we especially use the SPECTER model (Cohan et al., 2020), a BERT-based sentence transformer inherited from the SciBERT model (Beltagy et al., 2019), which has been trained on a scientific corpus mixing computer science, biology, and medical science. In particular, SPECTER shows a better disciplinary comprehension of the scientific documents than SciBERT, since the former has been trained on a corpus where two given elements can be linked by a citation relation. Instead of generating an embedding of a

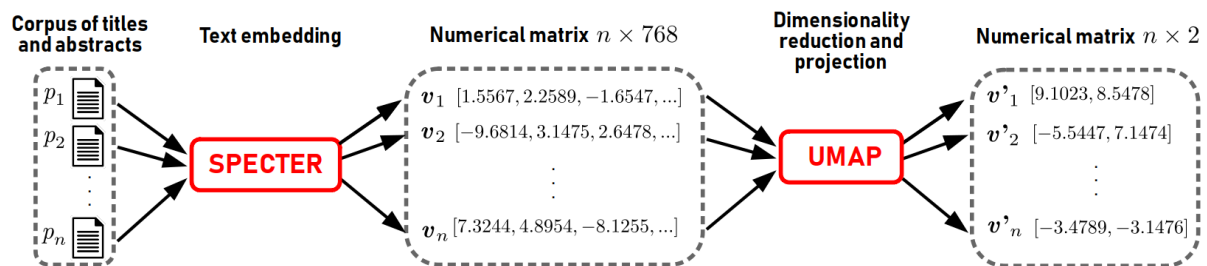


Figure 4.1: Schematic representation of the pipeline used to generate the neuroscience knowledge map below. First, SPECTER converts the textual elements of the articles within our corpus (called p_i) into 768-dimensional vectors each (v_i), then UMAP transforms the latter into 2-dimensional vectors (v'_i) given the data structure provided by SPECTER. The final list of 2D points are used to draw the map.

document based only on its own semantic context (*intra*-document context), SPECTER also considers the semantic context of other documents related to the former by citations (*inter*-document context).

In order to produce a synthetic cartography of neuroscience knowledge, we first apply SPECTER to the textual elements of our corpus \mathcal{P} , here the titles and abstracts of its papers, in order to obtain their respective contextual embedding, which result as vectors of 768 floating elements each. We then reduce these vectors to a two-dimensional Euclidean space with a *Uniform Manifold Approximation and Projection* (UMAP),¹ which aims to preserve the proximity (or similarity) between the initial vectors expressed in the higher-dimensional space, here of dimension 768 (McInnes, Healy, and Melville, 2020). This lower-dimensional space is thus an approximation of the original semantic space that characterizes all of our neuroscience work. This procedure is illustrated in Fig. 4.1. As UMAP projects our high-dimensional data into an Euclidean space, this step finally allows us to plot the two-dimensional points in such a space, as shown in Fig. 4.3 – and compute standard measures on the map produced without adapting them to a particular metric that would describe a more complex differential manifold. A paper is represented by a single point. An aggregate of such points very close to each other in a specific region of this map thus forms a vocabulary subspace where the knowledge inscribed in the associated articles is similar. If such a subspace is isolated in the maps, the papers belonging to it thus compose a single research topic.

Since the initialization of the UMAP projection is random, and because of the stochastic property of the algorithm,² we remind the reader that the reduction obtained is not unique. Nevertheless, since UMAP preserves the local proximity of the projected points, the reductions obtained after several realizations of the algorithm would have the same shape in a two-dimensional plane but oriented or organized differently, thus not affecting

¹We use the eponymous Python package `umap`, whose operation is detailed on the following webpage: https://umap-learn.readthedocs.io/en/latest/how_umap_works.html.

²We refer the reader to: <https://umap-learn.readthedocs.io/en/latest/reproducibility.html>.

a further clustering process on these reduced data. Therefore, we propose here an investigation of one realization of this projection. App. C.1.1 details further analyses of the robustness of the UMAP embedding according to various features, such as the location of the nearest neighbors of each paper and their topical proximity – based on the recent `Topic` classification of scientific papers given by OpenAlex.

Then, we apply a *Hierarchical Density-Based Spatial Clustering of Applications with Noise*, or HDBSCAN (McInnes, Healy, and Astels, 2017), to this set of “reduced” points. This method first returns a condensed dendrogram whose sticks are clusters of points, which could be merged or divided by varying a density threshold λ . The partitioning of the dataset is detailed in App. C.1.2. We retain nine clusters, whose names are established with the distributions of the `MAG Fields of study`, and `OpenAlex Topics` and `Keywords` related to the papers within them, as detailed in App. C.1.3.

Finally, we represent the reduced dataset as a density plot in Fig. 4.3, with clusters delineated by solid colored lines. We draw on the same dataset and partition to examine a series of temporal snapshots of this map, shown in Fig. 4.4.

4.1.2 Measuring the epistemic articulation of AI-related and general neuroscience concepts

4.1.2.1 Conceptual similarity between AI-related publications and others in a knowledge cluster

Here, we build a measure of the integration of AI-related knowledge into the clusters of the built-up neuroscientific knowledge maps, based on the *fields of study* (analogously called *concepts*) labeling the articles within these clusters. This indicator is the same as that used in Chapter 3 to study the homogenization of citations from the AI subarea into the neuroscience disciplinary landscape. In this case, it aims to indicate the concordance of AI-specific research programmes and topics (embodied here by the fields of study labeling the papers) with those held by the considered neuroscience cluster over the course of its evolution.

For a given cluster C of the map, we first compute the annual frequency of occurrence of each field of study labeling the papers within that cluster, resulting in an evolving set $F_t = \{n_f\}_t$ of frequencies n per field of study f in a given year t . For example, given four papers p_1, p_2, p_3 and p_4 in such a group of clusters, labeled respectively with the sets of fields of study $\{f_1\}$, $\{f_1, f_2\}$, $\{f_3\}$ and $\{f_1, f_3\}$, we get the frequencies $\{n_{f_1} : 3, n_{f_2} : 1, n_{f_3} : 2\}$. We then sort the fields with such frequencies to get a *concept ranking*

highlighting the most important fields that characterize the papers within C , from the most frequent concepts at the top to the least frequent ones at the bottom in year t . With the previous example, we thus get the ranking $\{1 : f_1, 2 : f_3, 3 : f_2\}$. We then derive the temporal concept ranking of C by computing such rankings for all years between 1970 and 2019, denoted as $r_C(t)$.

We also compute the temporal concept ranking of the subset of AI-related publications within C – which are also members of the subset $\mathcal{P} \cap AI$, called $r_C^{AI}(t)$.

Finally, we apply the method provided by Gargiulo et al. (2016, see their supplementary material) to obtain a temporal Jaccard similarity $J_c(r_C(t), r_C^{AI}(t)) := J_c(C, t)$, which compares these two rankings $r_C(t)$ and $r_C^{AI}(t)$ with a value comprised between 0 and 1. If $J_c(C, t) = 1$, the rankings are exactly the same in year t , i.e. they include the same elements with the same ranks in year t ; conversely, if $J_c(C, t) = 0$, the rankings are completely different, i.e., they do not contain the same elements at all.

4.1.2.2 Assessing the centrality of AI-related concepts in each cluster’s concept network

Unlike the previous metric, the following one designed in this section is intended to assess the degree of entanglement of AI-related fields of study within the concept network of a given cluster, in particular by determining their location within such a network – in the core or in the periphery. To this end, we consider only the fields of study situated at levels 2 or 3 of the MAG field network labeling the articles, so that we cover 93% of the papers having at least two concepts at these levels. With such a selection, we conserve 96.5% of the AI-related papers in the subset $\mathcal{P} \cap AI$.

The concept network describing the knowledge universe of a given cluster is framed under the cumulative process as follows and shown in Fig. 4.2. Starting from the co-occurrence network of level-2 and/or level-3 fields of study present in the cluster’s papers published in 1970, denoted as $G_0 = (V_0, E_0)$ – where V_0 the set of fields of study and E_0 the set of co-occurrence edges between the latter – and the next one embodied in the papers published in 1971, denoted as G_1 , we construct the cumulative concept network for the year 1971 as the network $\tilde{G}_1 = G_0 \cup G_1 = (\tilde{V}_1, \tilde{E}_1)$, where $\tilde{V}_1 = V_0 \cup V_1$ and $\tilde{E}_1 = E_0 \cup E_1$. We generalize this recursive relation to any year t as $\tilde{G}_t = G_{t-1} \cup G_t = (V_{t-1} \cup V_t, E_{t-1} \cup E_t)$, so that the final graph \tilde{G}_{2020} is the entire concept network of the cluster spanning the period 1970-2020. This temporal dynamic of concept cooccurrence thus gives rise to an evolving semantic network revealing the topical evolution of the cluster under study, where the meaning of any concept depends on its neighboring ones, with which its links may be modified (added, strengthened, or removed) over time (Cheng et al., 2023; Rule et al.,

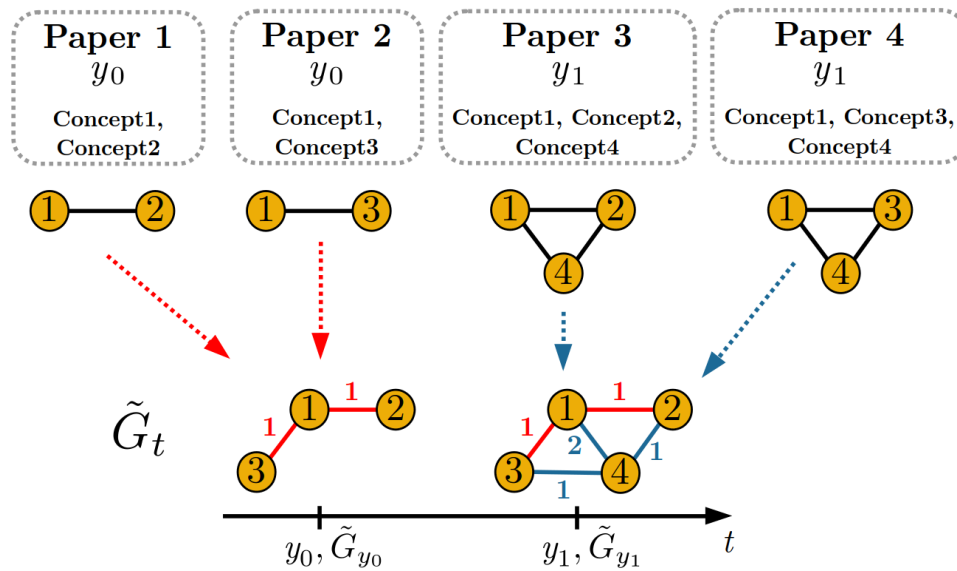


Figure 4.2: Schematic representation of the building of the temporal concept network, denoted as \tilde{G}_t as in the main text, based on four articles published in years y_0 or y_1 . The weights of the edges are indicated by the numbers next to them.

2015).

This cumulative representation of associated fields of study within the clusters allows us to track the location of AI-related ones as new concepts and links are added to the network describing the cluster over time, i.e., as the cluster's knowledge universe expands. For example, for an AI-related concept situated in the core of the network in a given year – which is a signal of close conceptual ties with its neighbors, and therefore of a semantic proximity to them – do we observe a perpetuation of its position through the reinforcement of existing links in subsequent years, or its retreat toward the periphery? We could also ask the opposite in the case of a peripheral AI concept advancing toward the core as the network grows, which would be a signal not only of its increasing use but also of its ability to be grafted onto concepts of different kinds.

We define this location of concepts within the studied network by their core number obtained by a k -core decomposition of the network, i.e., their membership to a subgraph with a degree that is at most c (the c -shell) but not in the next core with minimal degree $c + 1$ (Batagelj and Zaversnik, 2003). We normalize this metric by the maximum core number of the network, i.e., $\tilde{c} = c/c_{max}$, such that the closer to 0 the indicator \tilde{c} , the farther away from the core the concept is – and *vice versa* when moving closer to 1. We will refer to this indicator as *coreness* in the following, and we will focus especially on the values exhibited by the AI-related fields of study within each cluster of the neuroscience knowledge map.

4.1.3 Measuring the diffusion of AI-related work in the knowledge map

4.1.3.1 Building of the citation network

We propose to analyze the diffusion of AI-related knowledge on the neuroscience cartography drawn above through the underlying citation network between the mapped articles. We first define $G = (V, E)$ the citation network between all these papers, such that $V = \{p_i\}_{i=1, \dots, N}$ is the set of papers, which are identified by their respective cluster C_k ($k = 0, \dots, 8$), and $E = \{(p_i, p_j)\}_{i, j \in [1, \dots, N]^2}$ the set of unique directed links between the papers in V , e.g., a link (p_i, p_j) denoting that p_i cites p_j .

4.1.3.2 Individual citation radius of gyration of the papers

Here, we propose a measure of the ability of a paper to influence a wide range of the neuroscience knowledge space, based on its individual radius of gyration (RoG) generated by other articles citing it – observed in 2020. This quantity is expressed as the square root of the moment of inertia associated with the publications citing a focal paper i :

$$r_g^i = \sqrt{\frac{1}{N_i} \sum_{j \in p_i} d_{ij}^2}, \quad (4.1)$$

where p_i is the set of papers citing i (with cardinality N_i) and d_{ij} is the spatial distance between papers i and j in the knowledge map, i.e., the paper i is considered as the citation *center of mass*. This RoG is also useful for assessing the diffusion or concentration of citations in a local vocabulary space or beyond, and with what intensity.

However, an article located at one of the edges of the map – top, bottom, left, or right extremities – may have a larger RoG than another article situated rather at the center of that map. Indeed, as a result of the construction of the map, the articles near the center are situated in a middle mass of knowledge that is almost equidistant from all others, which makes their citation diffusion of knowledge across the entire map less difficult than that of articles living in highly specific knowledge and vocabulary clusters, such as C_0 , C_1 , C_2 , C_3 , and C_4 in Fig. 4.3. For example, a paper in the cluster C_4 that cites another in C_1 must have spanned a larger lexical gap than another paper also citing C_1 but located in C_7 .

To compare the RoG between the clusters, we normalize the observed individual RoG of a given article by another hypothetical one, its *maximum RoG*, which is a function of the maximal distance between this article and the one farthest away from it on the map,

indicated by d_i^{max} below. By considering the same number of papers citing the article i , but all hypothetically located at a distance d_i^{max} , the application of Eq. 4.1 gives:

$$r_{g,max}^i = \sqrt{\frac{1}{N_i} \sum_{j \in p_i} (d_i^{max})^2} = \sqrt{\frac{1}{N_i} \times N_i (d_i^{max})^2} = d_i^{max}. \quad (4.2)$$

This maximum RoG is thus analogous to the maximum knowledge area that an article could cover if cited by these very distant papers on the map.

Then, we derive this alternative RoG used throughout this chapter, which is actually the intensity of the observed citation spread compared to its own possible maximal one in the neuroscience map:

$$\tilde{r}_g^i = \frac{r_g^i}{r_{g,max}^i}. \quad (4.3)$$

Since its values are comprised between 0 and 1, the closer the new RoG is to 1, the wider the diffusion, and conversely, the closer it is to 0, the narrower the diffusion.

In what follows, we will refer to this metric to compare the respective spreading ability of AI-related ($\mathcal{P} \cap AI$) and non-AI ($\mathcal{P} \cap \overline{AI}$) papers across the knowledge landscape of neuroscience. Here, we only consider publications that have received at least 3 citations until 2020.

4.2 Where is AI situated in the knowledge landscape of neuroscience?

Fig. 4.3 shows the cartography of all the knowledge encoded in the textual metadata of articles in our neuroscience database spanning the period 1970-2020, built with the method exposed in Sect. 4.1.1. More precisely, this map represents the spatial distribution of these articles according to their own vocabulary, represented here as points in a broad lexical space. For example, the vocabulary employed in the cluster *Eyes & Vision* is quite distinct from that used in the cluster *Mathematics for Connectionist AI*, both of which are also different from the vocabulary employed in *Parkinson's disease* area. The names of the different subspaces delimited by colored lines – the clusters – are derived from the most frequent concepts associated with the papers in them, which are provided by three topical classifications, namely the MAG fields of study, and OpenAlex's topics and keywords (see App. C.1.3 for their definitions). They are also listed in Tab. 4.1. We used grayscale density level lines to highlight the heterogeneous density of papers in this knowledge space. Only AI-related publications are plotted as red dots on this map.

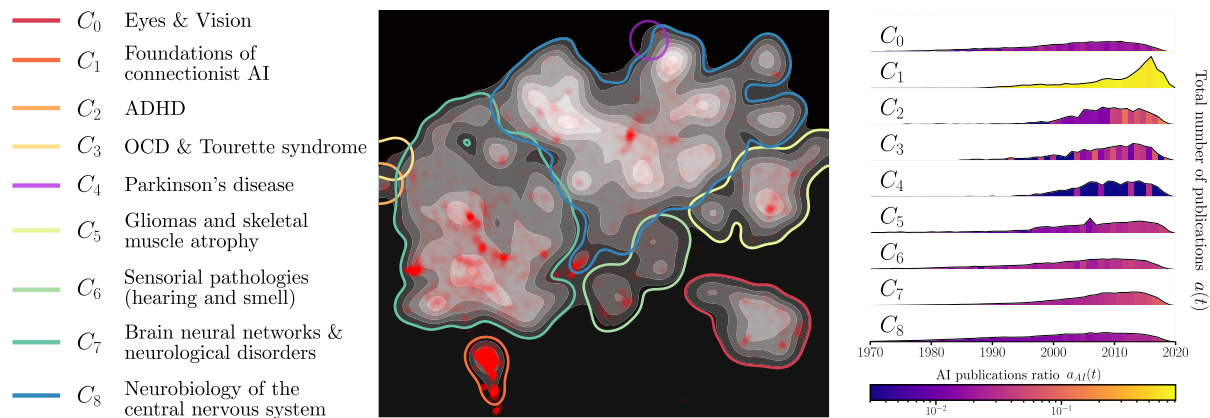


Figure 4.3: Neuroscience knowledge map. In the center is the time-aggregated knowledge map of neuroscience between 1970 and 2020. The grayscale represents the density of articles in this space: the lighter the zone, the denser the local concentration of papers, and vice versa when moving towards through darker shades of gray. The clusters delimited by colored lines, whose names are given in the legend on the left of this figure, have been drawn with the largest level line obtained with the HDBSCAN algorithm. The red dots are the AI-related papers distributed in this knowledge space. On the right are plotted the temporal evolution of the size of the clusters, normalized by their own size in the time-aggregated map on the left. The vertical axes of all plots are standardized and do not exceed 0.15. The colored area under these curves represents the temporal evolution of the share of AI-related publications within these clusters.

The spatial distribution of AI papers in this map shows at first glance that AI is able to integrate itself into different knowledge subspaces of neuroscience and also to link to a variety of knowledge across the discipline, thus again demonstrating the adaptation of AI to neuroscience epistemic objectives and a certain genericity that could also be denoted as epistemic.

However, the temporal evolution of the number of AI publications in each cluster, shown in the right-hand plots of Fig. 4.3, along with the distributions of AI publications among the clusters summarized in Tab. 4.1, both testify to the uneven distribution of AI papers among the clusters. For example, cluster C_1 condenses 22.5% of all AI publications into a much smaller knowledge space than that covered by cluster C_8 , which nonetheless exhibits an equivalent share of such publications (24.1%). The first cluster thus shows a very high density of AI articles with very similar vocabularies, while such publications belonging to the second cluster are more spread out in its broad lexical space - with some small redder subspaces of high local density, however. Moreover, according to Tab. 4.1, the leftmost clusters on the map, grouping C_1 , C_2 , C_3 and C_7 , aggregate 59% of all the AI-related papers, thus demonstrating that the core of such publications is situated in a particular knowledge subspace of neuroscience, focused on the study of neurodegenerative diseases and the formal foundations of contemporary AI-related models. The cluster C_1 , although separated from the continuum of papers ranging from neurological disorders studies (represented by C_7) to neurobiological studies of the central nervous system (represented by C_8), remains quite close to the former, with some bridges between them through other AI

publications. This suggests that the subfield gathering the studies of the mathematical and computational foundations of connectionist AI, which could be related to computational neuroscience, preferentially maintains some links with the subfield of neurological damage studies represented by cluster C_7 .

Cluster	Name	Share of AI papers (%)
C_0	Eyes and vision studies	3,7
C_1	Mathematical and computational foundations of connectionist AI	22,5
C_2	Attention-Deficit/Hyperactivity Disorders	0,33
C_3	Obsessive-Compulsive Disorders and Tourette syndrome studies	0,14
C_4	Parkinson's disease	0,02
C_5	Gliomas and skeletal muscle atrophy studies	5,76
C_6	Studies of sensorial pathologies	3,02
C_7	Brain neural networks and neurological disorders studies (epilepsy and schizophrenia)	36,1
C_8	Studies of neurobiological mechanisms in the central nervous system	24,1

Table 4.1: Distribution of AI-related articles across the clusters of the knowledge map. For example, C_0 owns 3.7% of all the AI publications in our dataset.

We complete the map of Fig. 4.3 by plotting its temporal decomposition in Fig. 4.4, in order to approximate the evolution of the neuroscience subfields and the location of AI in the different configurations shaped by them between 1970 and 2020. With these maps, we identify three phases in the development of the main field of neuroscience.

The first, spanning the period 1970-1984, marks the consolidation of the rightmost clusters of the time-aggregated map, namely the research around the central nervous system (C_8), the skeletal muscle diseases (C_5), sensorial pathologies (C_6) and eyes and vision (C_0) – with a growing number of publications inside them, as already shown in Fig. 4.3. The second phase, running from 1985 to 2009, shows the lexical expansion of these clusters, the substantial growth of the subfield related to the studies of brain neural networks and neurological disorders (C_7), and the birth of the knowledge universe around connectionist AI (C_1), especially during the subperiod 1985-1989. We also observe the unification of the knowledge subspace of all clusters except C_0 and C_1 during this second period 1985-2009.

The third and last stage, running from 2010 to 2020, shows a division of the aforementioned continuum of clusters into two parts whose boundaries no longer meet, a first one including the clusters C_2 , C_3 , C_7 (neurological disorders studies) and some subsets of the cluster related to the studies of the central nervous system (C_8), and a second one including the remaining clusters C_4 , C_5 , C_6 , and C_8 . This period also signs the decline of

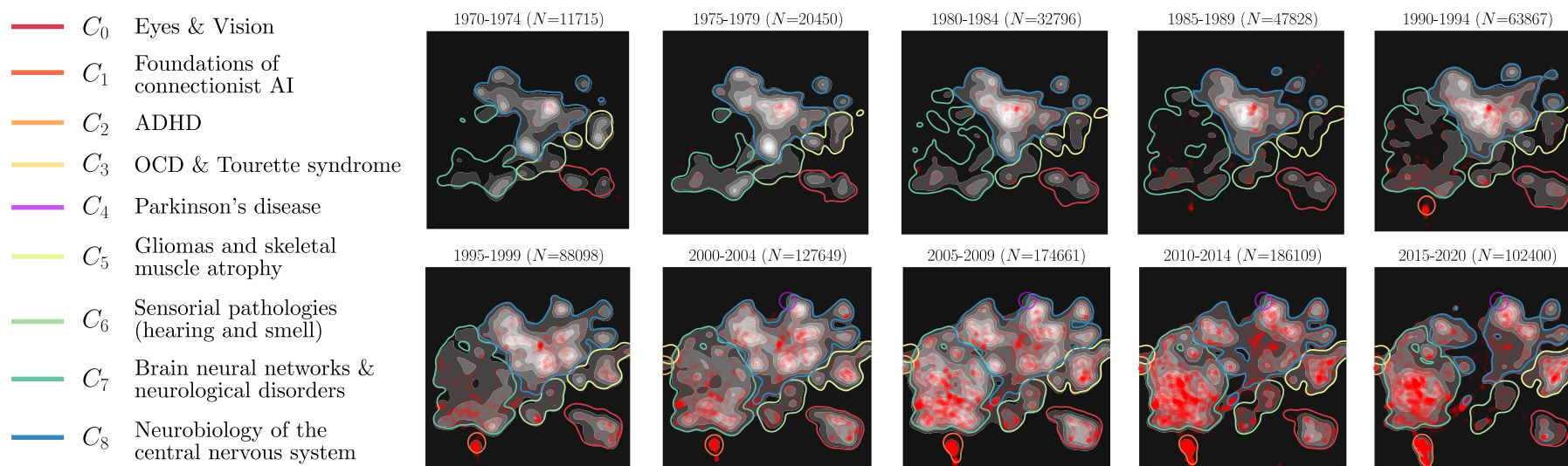


Figure 4.4: Evolution of the neuroscience knowledge space and its clusters, plotted every 5 years between 1970 and 2020. N is the number of publications in the corresponding period. The red dots are those related to AI. The clusters delineated by colored lines, whose names are given in the legend on the left of this figure, are the same as those delimited in the time-aggregated map in Fig. 4.3.

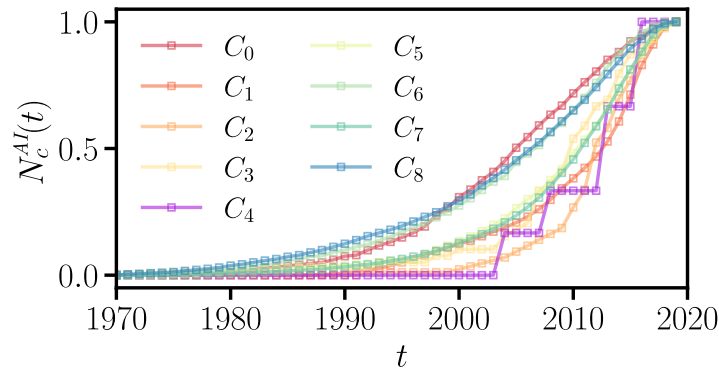


Figure 4.5: Cumulative number of AI publications per cluster, normalized by their respective maximum value reached in the year 2019 in order to compare their respective growth trend. The stepped shape of the C_4 curve is due to the very limited number of AI-related publications within it.

the subfields of eyes and vision studies and sensory pathologies.

These temporal snapshots of the knowledge space in Fig. 4.4 also suggest an acceleration in the growth of AI research since this period 1985-1989. This trend is confirmed by the cumulative number of AI publications per cluster shown in Fig. 4.5. While C_0 , C_6 and C_8 exhibit a sustained growth of AI publications within them, all others show a faster growth of such scientific production in them during the 2000s, the most spectacular one having happened in C_2 (ADHD studies), which has particularly been fast after the 2010s. The pre-existing knowledge and the dominant vocabulary in these last clusters seem to favor the development of AI within their own research context, for example, by attracting and pairing external AI-related concepts with their own conceptual basis, or by allowing the emergence of other AI-related ones – which is more difficult to establish. The next section is dedicated to examine such microscopic aspects related to the organization of knowledge in all clusters of the map.

4.3 The limited integration of AI into the various concept networks of neuroscience

Through the MAG fields of study assigned to each article in our neuroscience dataset \mathcal{P} , which are also identified by their respective knowledge cluster memberships, we delve below into the first dimension evoked in the very last lines of the previous section, namely the entanglement of AI concepts within the conceptual universe of these clusters. Here, we report this articulation with two findings.

The first focuses on the macroscopic comparison between the set of concepts promoted

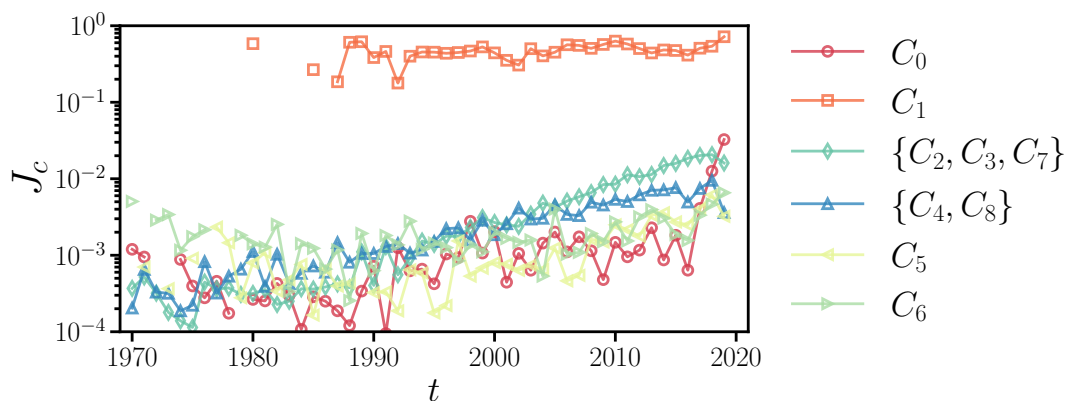


Figure 4.6: Jaccard ranking similarity between the main fields of study associated with the papers that are members of a given cluster C_i (or in a set of clusters, indicated between brackets) and those associated with the AI-related publications within the same cluster. The small peripheral clusters C_2 , C_3 and C_4 , whose weak presence of AI papers over the years causes noisy, sometimes non-continuous similarity trends, are merged into large closer clusters, here C_7 for the first two and C_8 for the third.

by the articles in a given cluster and another set composed only of concepts promoted by the AI-related articles in that cluster. To this end, we quantitatively evaluate the similarity of these sets by computing the Jaccard ranking similarity J_c between these two sets (for the computation of this indicator, see Sect. 4.1.2.1), whose results are shown in Fig. 4.6. We test the robustness of the plotted trends by applying the same methodology to the network composed of OpenAlex topics, whose results are shown in App. C.2 – they also confirm the first results, and thus (again) the proximity between the two conceptual classifications of MAG and OpenAlex. Since the MAG fields of study at levels 2 and 3 are analogous to research topics, we can thus compare the set of concepts in a cluster as a kind of research programme supported by the subfield represented by the cluster. According to Fig. 4.6, the conceptual similarities between AI-related research and all others represented in each cluster clearly distinguish cluster C_1 – with an average similarity of 47% since 1987 – from all the others, whose values have been rising steadily since the 1990s, while remaining very low – less than the 3% reached in C_0 . We thus can conclude that AI, through its associated publications, is embedded in conceptual universes that significantly differ from the main ones represented by the clusters at first glance, except for C_1 , whose case is quite obvious according to the considerations outlined above.

The second analysis uses the cumulative concept network of the clusters in order to track the location of AI-related concepts within it over time – in the core or in the periphery –, which is summarized by the mean temporal corenesses of AI-related concepts plotted in Fig. 4.7 for each cluster (see Sect. 4.1.2.2 for the computation of this metric). More interestingly, it shows a general retreat of AI from the core to the periphery of the concept networks in all clusters over the years, even for the cluster C_1 . In particular, while AI becomes less central in the respective conceptual framework of the largest knowledge

clusters C_0 , C_5 , C_6 , C_7 and C_8 over the period 1970-2020 and seems to migrate to a more distant periphery, it suffers a considerable decline in clusters C_1 , C_2 and C_3 , where it was rather central at the time of their respective emergence in the field of neuroscience. This is especially the case for the last two aforementioned clusters, where AI's coreness decreases sharply over a short period of time since the early 2000s, 20 and 22 years, respectively. Although AI participated in the creation of these clusters and thus contributed to lay the first milestones of the conceptual universes of these clusters, the progressive addition of new fields of study and connections between them seems to have transformed these conceptual universes and pulled AI away during the last 20 years. Nevertheless, only AI's coreness is increasing in C_4 between 2015 and 2016, showing a progressive interest of the subfield studying Parkinson's disease for AI-related knowledge.

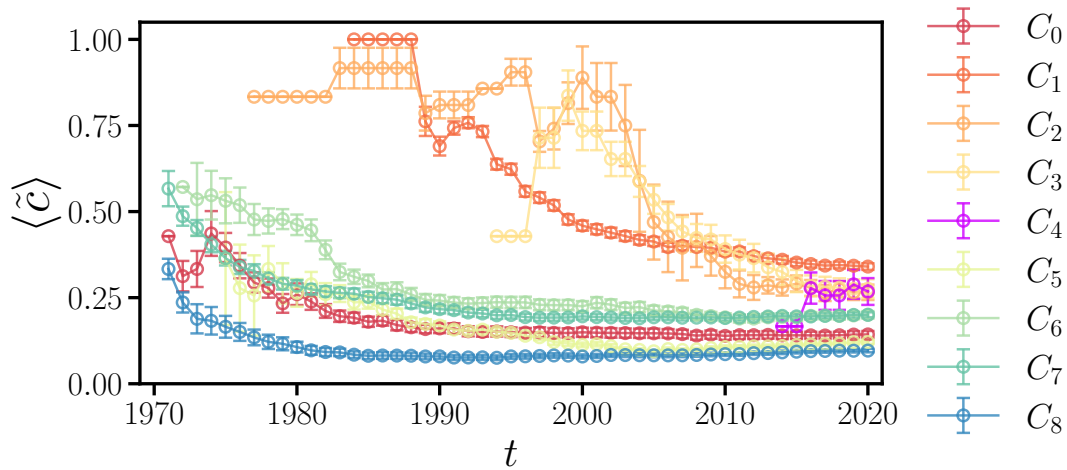
The behavior of the coreness of AI within the cluster C_1 is particularly intriguing. Indeed, while the majority of its publications are based on AI and are limited to a very small vocabulary subspace, as previously shown by the cartography in Figs. 4.3 and 4.4, the very numerous AI-related concepts attached to its publications compose only 15% of the population of unique concepts within this cluster in 2020. This suggests that this cluster also faces a conceptual reconfiguration, so that the AI concepts that contributed to shape this cluster open up to other external ones that are not related to AI but close to knowledge domains centered around the studies of neurological pathologies (especially within C_7). The case of C_1 , as well as those of C_2 and C_3 depicted above, thus illustrate the adaptation of AI-centered clusters to larger conceptual universes.

4.4 The limited citation spreading of AI-related publications in neuroscience

4.4.1 AI-related articles diffuse less than non-AI ones across the domain

In this part, we investigate the ability of AI-related articles ($\mathcal{P} \cap AI$) to spread their knowledge across neuroscience by comparing their citation impact in different knowledge subspaces of this field with that of non-AI articles ($\mathcal{P} \cap \overline{AI}$).

Starting from the citation network between the papers in the dataset, and by selecting only those with at least 3 citations, we compute for each of them the spatial radius of gyration (RoG) produced by the positions of their respective citing papers in the knowledge space, following the method introduced in Sect. 4.1.3.2. We average these measures per year, distinguishing the AI-related papers from the non-AI ones, as shown



Cluster	$\langle \tilde{c} \rangle$ (year of apparition)	$\langle \tilde{c} \rangle$ (2020)	Number of AI-related concepts	Share of AI-related concepts
C_0	0.43 (1970)	0.14	209	0.02
C_1	1 (1983)	0.34	1062	0.16
C_2	0.83 (1977)	0.27	55	0.03
C_3	0.43 (1993)	0.28	37	0.03
C_4	0.17 (2013)	0.27	4	0.01
C_5	0.37 (1974)	0.12	376	0.02
C_6	0.57 (1971)	0.2	283	0.02
C_7	0.57 (1970)	0.2	1245	0.04
C_8	0.33 (1970)	0.1	590	0.02

Figure 4.7: Temporal average coreness $\langle \tilde{c} \rangle$ of the AI-related concepts in the cumulative temporal concept network of each cluster – normalized by the highest coreness returned by the k -core decomposition of the network at a given year t . The higher the coreness $\langle \tilde{c} \rangle$, the closer the AI-related concepts are to the core of the network, and vice versa. We only consider AI-related concepts included in the giant component of each cluster’s concept network. The error bars are the standard errors produced by the distribution of corenesses of all AI-related concepts within each cluster at each year, i.e., the ratio of the standard deviation to the square root of the number of entities present in the distribution. In the table below the plots are indicated, for each cluster, the value of the coreness of AI in the cluster’s inception year, the final coreness of AI in 2020, the number of unique AI-related concepts in 2020, and the share of such concepts among all the others in 2020.

in Fig. 4.8.

The average RoG of the citations received by the non-AI papers, which first decreases slightly until the 1980s, increases almost linearly since this period, while the evolution of the measure applied to the AI-related papers fluctuates with high amplitudes³, especially before the early 1990s. Overall, the latter remains rather below the former after the 1990s, thus meaning that AI-related articles diffuse less than the non-AI ones within neuroscience over the time period studied.

³This behavior could be explained by the very heterogeneous distribution of these papers during the studied temporal period – 87% of the papers are published after 1990 – and their high dispersion in the conceptual space – as shown by a standard error that is significantly higher before the 1990s than after.

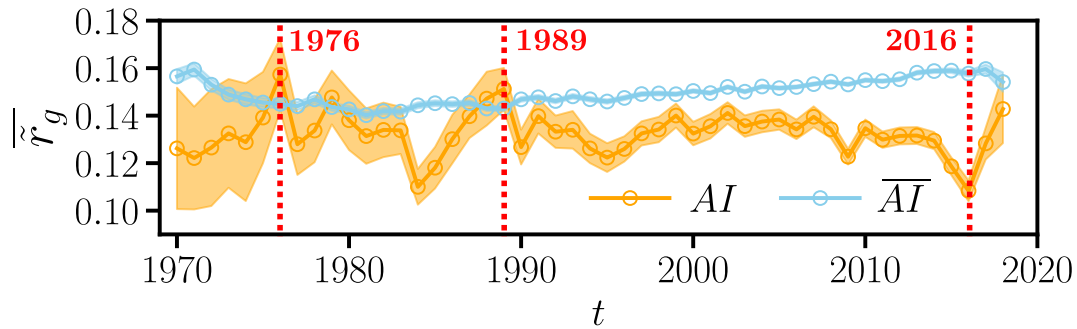


Figure 4.8: Mean temporal citation radius of gyration of the neuroscience papers, AI-related (orange) and non-AI (blue). The colored areas around the curves are their respective standard errors, expressed here as the ratio of the standard deviation of a sample in a given year to the size of that sample. We only consider the papers with at least 3 citations.

Following the average citation RoG of the AI-related articles, the lexical coverage of the publications citing the AI-related work between the 1970s and the 1990s alternates between phases of low (pits) and high (peaks) citation diffusion. In particular, the curve shows two peaks higher than the curve of the non-AI papers in 1976 and 1989. These peaks correspond to two well-known stages in the history of AI, as shown by the evolution of the use of AI-related concepts in Fig. 4.9, namely the symbolic and the connectionist ones. In particular, we notice the rise of *Backpropagation* and stochastic gradient descent techniques on artificial neural networks in the late 1980s, especially popularized by Rumelhart et al. (1986), and that is directly inherited from the reinvestment of the concept of *Perceptron* – single or multilayer (Rosenblatt, 1958). This will lead to the rise of neural network techniques and later to deep learning ones (Cardon et al., 2018). This connectionist phase fosters the development of neuroscience through the formalization of cognitive processes under (artificial) neural networks, which are useful to study the development of neurological disorders in recent days (see Section 4.2). Some further results, detailed in App. C.3, also show that the related findings during the two studied periods have achieved a kind of success, as testified by the rapid citation spread of AI-related ideas across the neuroscience knowledge space.

However, we notice in Fig. 4.8 a strong decrease in the RoG between 2014 and 2016, meaning that the AI-related knowledge and tools embedded in these papers are beneficial only to their close lexical neighborhood. In fact, among the 3,739 papers published between 2015 and 2016 during the decrease in the mean RoG, 37% emanates from studies on neurological disorders and brain neural networks (C_7) and 35% from the subfield developing connectionist AI knowledge and tools (C_1), while the rest is distributed among the other clusters and noise from the knowledge map. This distribution supports one of the phenomena already evoked in the previous chapter, namely the specialization of AI research around the subfield designing AI itself in neuroscience (cluster C_1), accompanied by other applicative subfields mainly confined to cluster C_7 .

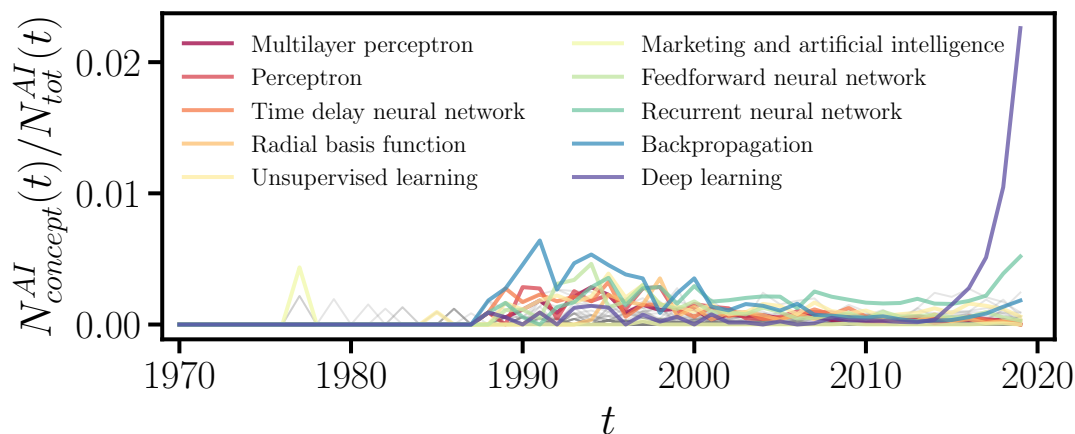


Figure 4.9: Share of AI-related fields of study provided by MAG among the AI-related articles in neuroscience over the years. The top 10 are highlighted in colors, others are in gray. The concept *Marketing and artificial intelligence* is inherited from another one called *Intelligent Decision Support System*, namely a subfield of symbolic AI that was *in vogue* between the 1970s and the 1990s. The algorithms populating this class of AI are based on complex knowledge representations in order to provide some assistance to decision makers. One of the most important algorithms in the medical and clinical domains during the aforementioned period, including neuroscience, was the so-called *Mycin*, which is also widely mentioned in the AI-related papers of our database during the late 1970s.

The late decrease in the mean RoG associated with AI-related concepts is particularly intriguing because it falls within a period of well-documented, high diffusion of AI in science, associated with the rise of deep learning techniques (Cardon et al., 2018; Gargiulo, Fontaine, et al., 2023), which reasonably suggests that any paper using or mentioning such concepts would find a wide audience throughout the whole knowledge landscape of neuroscience. We will see in the next subsection that this decrease and, more generally, the low RoG of AI-related articles are parts of a larger phenomenon of concentration of AI-related neuroscientific knowledge within the clusters from which they originate, which does not guarantee the epistemic genericity of AI within the domain under study.

4.4.2 AI remains confined to local knowledge subspaces of neuroscience

As suggested in the previous subsection, the spreading of the AI instrument across the entire neuroscience knowledge space remains quite limited over the years. In this section, we explore at microscale the clustered distribution of the citation RoGs of all articles in our dataset, as well as the time-aggregated citation network centered on the AI-related works and aggregated into clusters. Both are represented in Fig. 4.10.

According to Fig. 4.10A, which complements the previous Fig. 4.8, the publications dealing with AI within all the clusters of the neuroscience knowledge map diffuse as much

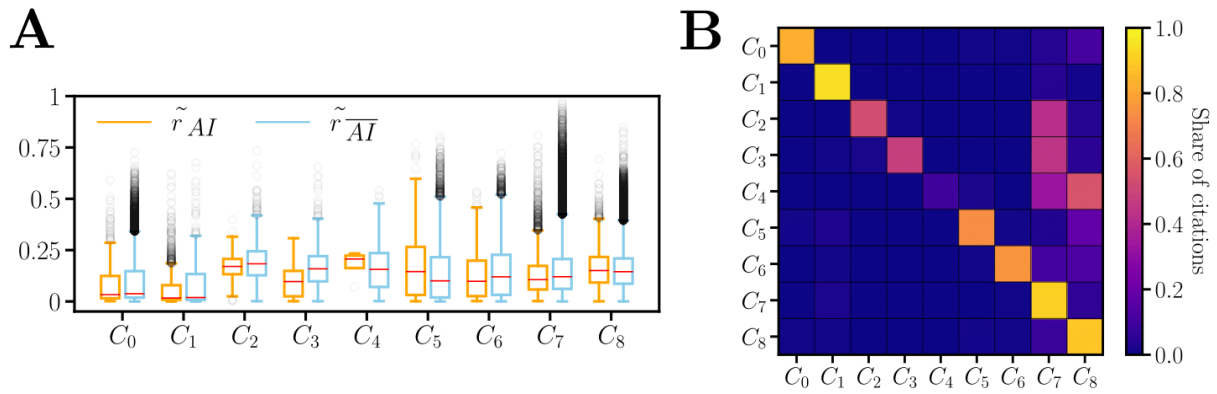


Figure 4.10: **A**: Time-aggregated distribution of the citation RoGs of the neuroscience papers within each cluster, AI-related (orange) and non-AI (blue). The red lines in the boxes indicate the medians of the distributions. Here, we only consider the papers with at least 3 citations. **B**: AI-centered citation matrix between the clusters of the neuroscience knowledge map shown in Fig. 4.3 (between 1970 and 2019). A cell (C_i, C_j) is the number of papers in C_i that cite the AI-related articles in C_j . The results are normalized by row, so the sum of the score for a row is 1.

or less than the non-AI ones, except for the small peripheral clusters C_4 (*Parkinson's disease*) and C_5 (*Skeletal Muscle Disease*), which both show the highest citation RoGs for the AI knowledge produced in them. This low diffusion is especially observed in the clusters where AI has been mainly developed in recent years (see Section 4.2), namely *Foundations for connectionist AI* (C_1), *ADHD studies* (C_2), *OCD and Tourette syndrome studies* (C_3) and *Studies of neurological disorders* (C_7).

The AI-centered citation matrix in Fig. 4.10B also shows a strong phenomenon of self-citation within each cluster, except for the small peripheral clusters C_2 , C_3 , and C_4 , which are also influenced by the largest clusters C_7 and C_8 that constitute the core of the knowledge map. The AI-related knowledge produced in a given cluster thus preferentially impacts that cluster – and/or its neighbors – and not the others.

4.5 Discussion

How does AI-related knowledge develop and spread in neuroscience? Through the analysis of a dynamic knowledge map of neuroscience generated with advanced document embedding techniques applied to the textual elements (titles and abstracts) within our corpus built in Chapter 2, which are also connected by citations, we have shown that AI is found in every lexical regions of the map, but it meets a greater success within subfields that have begun to grow significantly since the 1990s, and that are related to the mathematical and computational foundations of contemporary connectionist AI, and to the studies of the human brain neural network and various neurological disorders that lead to physio-

logical and/or cognitive impairments. In particular, the presence of the majority (59%) of the total number of AI-related publications in these regions tends to confirm the implantation of artificial neural network architecture as an underlying paradigm in the most recent research domains of neuroscience. These first findings provide some empirical evidence to support the considerations of neuroscientists themselves about the potentialities of applying AI to their own research goals, and especially about the reinforcing feedback between the host research field and the AI instrument towards their mutual development (Gopinath, 2023; Hassabis et al., 2017; Macpherson et al., 2021; “The new NeuroAI” 2024).

In an analysis of the set of fields of study labeling the papers within each knowledge cluster of the map, which constitute the delimited conceptual framework of the latter, we have found that AI is not integrated in the various conceptual arcs of the domain, except for the small, dense, constantly isolated knowledge subspace related to the formal conception of connectionist AI and computational neuroscience, which includes 22.5% of the AI-related production in our corpus – and which also exhibits a strongly confined citation network, as shown in Fig. 4.10B, similar to an echo chamber. Finally, this cartography supports the argument of the dual epistemic orientation of AI research in the discipline that was mentioned previously in Chapter 3: The aforementioned island of knowledge would represent the epistemic environment in which AI serves the purposes of the so-called STEM disciplines, while the rest of the AI-related publications spread across the entire map represent its epistemic orientation towards neuroscience itself, thus demonstrating a certain epistemic genericity of AI throughout the entire domain (Marcovich and Shinn, 2012; Shinn and Joerges, 2002).

However, although AI is one of the cornerstones of the conceptual edification of all the neuroscience subfields, it progressively moves away from the core of all the concept networks representing these subfields as they expand over time. This result thus challenges the seemingly widespread distribution of AI exhibited by the knowledge mapping. Indeed, the lexical proximity of AI-related articles to other works within each cluster, confronted with the organization of the concepts attached to them, does not allow us to firmly establish the genericity of AI produced in neuroscience. This finding leads instead to support the fact that AI is a general method that could be applied everywhere in neuroscience (Cockburn et al., 2018), without being a source of new concepts that could enrich the pre-existing conceptual frameworks of all its subfields.

Besides, this progressive disembedding of AI from the core of all conceptual frameworks of neuroscience suggests that a social component involved in this AI-related research is progressively pushed to the periphery of the collaboration network, as shown in the previous chapter, thus testifying the rise of a high specialization of AI towards a more formal

and technology-oriented epistemic orientation, as also depicted in (Klinger et al., 2022) for the global science system. Although we have not addressed here the location of the neuroscientists on this map, a further work would consist of projecting the co-authorship network studied in the previous chapter onto this map – by locating the authors with an aggregated position derived from their respective publications – such that we could identify the knowledge subspaces in which the AI practitioners are located, and clearly validate or not the aforementioned hypothesis of a specialization of AI practitioners towards a dedicated conceptual framework distinct from the rest of neuroscience. Such a mapping would also allow us to delineate specific socio-epistemic environments in neuroscience, i.e., sub-networks of collaborations organized around a common knowledge base, so that we could trace the potential social bridges between them that would promote the diffusion of particular knowledge from one epistemic community to another, in particular AI, and also identify the actors involved in such bridges. As in (Roth and Cointet, 2010), we could also build a socio-conceptual network linking neuroscientists to the fields of study related to their articles, in order to delineate the conceptual universe that is preferred by the AI practitioners, given their level of expertise on the notion and their membership cluster.

Furthermore, with the citation network between the articles, we have constructed a measure of the citation scope of an individual paper, the citation radius of gyration (RoG), which enables us to dynamically identify some time periods when its encoded knowledge has or has not spread beyond its local lexical neighborhood. In particular, we have unveiled alternating phases of increase and decrease in the average citation scope depicted over time that coincide with well-known events that structure the evolution of AI and its impact on science, such as the development of expert systems in the late 1970s and the rise of artificial neural networks in the late 1980s. Although AI is present in all of the neuroscience research subfields revealed by our data, we have shown that these subfields locally disseminate their own AI publications, with very little circulation of knowledge between the subfields. Thus, AI is only beneficial for the neuroscience subfields in which it has been produced. This assumption also indicates that the applications of AI remain localized in specific topics and do not seem to be transferable to other knowledge subspaces of neuroscience, thus suggesting the multiplication of distinct AI research dedicated to different purposes. Similarly, we could compare this situation with the embedding algorithm SPECTER (Cohan et al., 2020) that we have used to draw the knowledge map of neuroscience: although it is based on a BERT-based sentence transformer, a very common architecture to perform natural language processing tasks or document classification, it has been specifically designed to enrich the classifications of scientific texts only, through the training data used on the one hand, and the inter-document similarity provided by citations on the other hand. Therefore, such a new AI inherited from BERT should not be applied to corpora other than scientometric ones.

Although the clustering provided by HDBSCAN (see Fig. 4.3) remains relevant for studying the confinement of AI-oriented citations within large lexical areas, as shown in Fig. 4.10B, the RoG distributions in Fig. 4.10A, on the contrary, suggest that very different lexical subareas, with their own and probably disjoint citation networks (composed of papers with very low RoG), also coexist within these clusters, like small specialties (Wray, 2005). It would then be worth refining the clustering in order to better observe the bridges between smaller lexical areas represented by few articles, such as the small clusters C_2 , C_3 , and C_4 in Fig. 4.3, whose AI-related publications find interest in nearby lexical areas within the larger clusters C_7 and C_8 . Moreover, we have focused on direct citations of AI-related publications, but we could imagine a more complex dynamics of impact involving citation chains towards the different clusters in time, from which we could compute the number of citations needed to reach a particular cluster from another, and evaluate potential bifurcations through fields outside neuroscience before reintegrating the latter – following the exemplary frameworks depicted in (Brahim et al., 2021; Yong Huang et al., 2018).

4.6 Summary

Using a dynamic knowledge cartography of neuroscience drawn with advanced document embedding techniques, we have traced the shaping of the field from 1970 to 2020 through the emergence and the development of lexically consistent subfields, and we have located the AI-related publications within the latter. The vast distribution of scientific articles involving AI on the drawn map indicates an apparent epistemic genericity of the related knowledge and technologies across neuroscience, while being less and less aligned with the conceptual universes of neuroscience subfields as they develop, as we have shown with conceptual structures based on the co-occurrence network of fields of study provided by MAG and that label the papers of our dataset. AI thus demonstrates its ability to be applied in all the subfields delimited in our study, which include various knowledge areas for the largest one, but it does not intend to create new knowledge and conceptual universes within them, except for a small lexical region that is dedicated to it.

These contrasting findings complete the perspectives evoked in the previous Chapter 3, according to which AI is well inserted in the multidisciplinary context of neuroscience, while its practitioners, regardless of their level of expertise, experience together a segregation within the whole field. Here, AI seems to be generic *in application*, as demonstrated by its spread in various lexical areas covering the knowledge space of neuroscience, which is a signal of its adaptation in almost all epistemic frameworks of the domain, but not generic *in conceptualization*, as it withdraws from their conceptual arcs that could give rise

to new theories and research paradigms (Lemaine, 1980; Kuhn, 2018 [1978]). Paradoxically, AI seems to lose this second genericity over time in neuroscience and consequently does not seem to replace all the methodological frameworks present in the field.

Besides, the last study depicted in this chapter, which focused on the citation network centered around AI-related publications, has shown that AI-related publications are mainly beneficial for the knowledge domains close to the one in which they have been produced. AI thus remains highly specialized when used or produced within a specific knowledge region and is not transferable from one region to another. This is especially the case for machine learning or deep learning models and algorithms, which are highly dependent on the data used to train them before performing predictions for very precise purposes. A complex AI architecture shaped and trained to predict the course of Alzheimer's disease in a patient of a certain age would also not be able to predict the progression of Parkinson's or Huntington's disease.

Complementing quantitative findings: The interdisciplinary work of AI in the Aramis team

So far, Chapters 3 and 4 have focused on the scientometric aspects of the genericity of AI in neuroscience on a macroscopic scale. In particular, they have yielded notable elements linked to the three criteria we have set to describe it for this thesis: a confirmed insertion of AI in the majority of the knowledge subfields of neuroscience, while neither meeting a great success within the scientific community (through notably a segregation of AI practitioners, in few numbers), nor articulating with the conceptual arcs of these subfields. However, these major trends, which have been drawn dynamically since 1970, miss important elements that often emerge at micro-scale during the scientific work, and that could drive the diffusion of the instrument across the field and subsequently increase its genericity.

For example, the disciplinary trajectories of most AI experts unveiled in Chapter 3 suggest that such *research-technologists*, primarily trained outside neuroscience, must engage in an interdisciplinary arena situated at the frontier of STEM disciplines and neuroscience to disseminate AI within the latter. In such an enterprise, more nuanced interactional elements than simple collaborations between research-technologists and neuroscience “natives” may occur during scientific work, including conversation and negotiation in the laboratory itself regarding the use of the novel instrument in place of more traditional, established ones. This would, in turn, condition the social adoption of the new tools, as well as the production of future scientific facts (Latour and Woolgar, 1996, p.155).

Complementing the previous chapters, this chapter explores this kind of trading zone, here portrayed by a particular research team we have chosen for our sociological fieldwork, Aramis (part of the Institut du Cerveau, indicated also by ICM). In order to describe the internal organization of work within this team and the social construction of AI in an interdisciplinary context, we have conducted eight interviews, following the methods

exposed in Chapter 2 and the questionnaire provided in App. D.1. The interviewees are designated with their respective numbers as given in Tab. 2.3 of the aforementioned chapter.

This chapter is structured as follows. The first section presents an overview of the research conducted within Aramis, as reported by all interviewees. The second section examines the organization of tasks and interactions of scientific staff, with a particular focus on the distinction between computer scientists and engineers, who are specialists in data analysis, programming, and other various IT tasks, and medical professionals involved in parallel research activities. This analysis aims to elucidate the practices of interdisciplinary collaboration in the context of AI in neuroscience. The third and final section turns to the disparate perceptions of AI held by the two latter groups and the factors that might explain them.

5.1 An “immersion” in the core of Aramis research

In addition to the few elements gleaned from the Internet and briefly described in the Chapter 2, the interviews conducted are a valuable and comprehensive resource that enriches our knowledge of the research carried out within the team under study. Aramis indeed differs from the other ICM teams in terms of the research topics it proposes and contributes to the development of free and open-source software that can be used by other researchers and medical staff, whether or not they are involved in clinical research. Consequently, the team is somewhat charged to create or enhance various digital techniques for processing images of the human brain, predicting the evolution of neurodegenerative diseases with the objective of providing personalized patient monitoring, and modeling information transmission mechanisms in the brain.

In this regard, the team assumes a dual role, with missions that are inextricably linked and subject to constant feedback. On the one hand, it contributes to a vast field of knowledge related to neuroimaging by employing complex mathematical formalisms, including graph theory to build a model of a biological network of neurons and differential manifolds to mathematically describe the shapes present in three-dimensional brain images – or even more if we consider other quantities added to the Cartesian reference frame. In particular, differential manifolds can be used to segment areas of interest in an image (in other words, to outline shapes) or to recalibrate multiple images from the same patient sample. This may entail cleaning them up and aligning them (for example, putting them back in the correct position or orienting them in a specific direction) in order to facilitate comparison. On the other hand, Aramis furnishes clinicians with state-of-the-art digital tools that implement these mathematical and computational theories, thereby

facilitating the execution of laborious analyzes of voluminous medical databases within the context of research based on extensive international clinical cohorts. In the distant future, the software developed will also be used to assist practitioners in making diagnoses and determining the appropriate treatment to administer. Aramis' research is therefore firmly situated within a translational research framework, whereby findings from basic research in mathematics, computer science, and even physics are translated into applications in clinical research and, potentially, into everyday medical practice (Allen et al., 2019; Stokols et al., 2008).

The codes and software designed by Aramis may contain certain forms of AI that are more or less similar to simple machine learning methods based on large databases. For example, a programme may autonomously learn to calibrate images or identify specific regions of the brain according to certain criteria. The first subsection below provides an inventory of the main data sources available and used by Aramis members to train such algorithms. The second subsection presents an overview of the collaborative efforts between Aramis and prominent industrial groups or start-ups engaged in the advancement of research and development (R&D) initiatives, especially the production of treatments or therapeutic innovation that can effectively address the challenges posed by neurodegenerative diseases.

5.1.1 An international research in the context of the *big data* era

As observed in most scientific disciplines at this time, neuroscience is experiencing a significant increase in the volume of medical data that must be processed. The majority of this data are typically obtained from large-scale clinical trials designed to study the potential onset and progression of neurodegenerative diseases in various individuals of varying ages, physical, psychological and socio-demographic conditions. These subjects are monitored over several years with precise medical examinations conducted at regular time intervals, since the effects of most neurodegenerative diseases are rarely instantaneous and are assessed over (very) long time periods.

The patient cohorts are supervised by international research consortia or local groups, depending on whether the subjects originate from the same region or hospital environment. Particularly involved in the study of Alzheimer's disease, Aramis researchers mainly use data from such cohorts to statistically study the emergence of early symptoms of this disease and their progression in patients. The clinical data processed by the team are mainly sourced from the internationally renowned *Alzheimer's Disease Neuroimaging Initiative* (ADNI) cohort, in addition to several French ones, the largest of which is

MEMENTO,¹ which was launched as part of the French government's Alzheimer's Plan between 2008 and 2012 and comprises approximately 2,300 subjects. Several clinicians practicing in La Pitié-Salpêtrière Hospital have also studied a subsample of MEMENTO subjects as part of the *INveStIGATION of AlzHeimer's PredicTors in subjective memory complainers-Pre Alzheimer's Disease* study (INSIGHT-preAD),² which was initiated by the Institut National de la Santé et de la Recherche Médicale (INSERM) and Assistance Publique-Hôpitaux de Paris (AP-HP) in 2014. The subjects in this subsample underwent further examinations to those initially submitted by MEMENTO's medical investigation protocol.

5.1.1.1 The nature of used data

The data stored in the large clinical research databases, which often have the name of the study cohort, encompass various forms. In particular, they include imaging data acquired using techniques already mentioned at the beginning of Chapter 2, in particular scans acquired with Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET) techniques, which are the most well-known and whose working is often very costly in terms of time and money. Moreover, only a limited number of medical centers around the world are equipped with the appropriate machines and scanners to perform such medical examinations. The corresponding data thus are initially acquired at some of these centers that are part of the cohort study in question – these data are therefore said *multicentric* – and then pooled internationally with other centers that do not have the necessary equipment. As the machines may differ from one center to another – for example, in terms of brand or component obsolescence –, data are acquired according to thorough protocols for setting up the machines, under the same examination conditions as possible for each subject selected, in order to guarantee a semblance of consistency and minimize experimental discrepancies when pooling the data in other centers. Biomarkers that describe certain biological properties of the brain, such as its volume, that of the hippocampus and frontal lobe, and other indicators of its metabolic activity, are then extracted from the raw data, which are often colored images of the brain including various attached metadata. The latter images enable the observation of not only brain regions activated by a specific task performed by the subject at the clinician's request but also lesions in the organ caused by the onset of a neurodegenerative disease, such as amyloid residues around neuronal cells in patients suffering from Alzheimer's disease.

¹Further information can be found at the following address: <https://www.fondation-alzheimer.org/la-recherche/cohorte-memento/>.

²Further information on this subject can be found at the following address: <https://presse.inserm.fr/lancement-dune-etude-visant-a-mieux-comprendre-les-facteurs-de-declenchement-de-la-maladie-dalzheimer/15483/>. A report was published in 2018: <https://aphp.fr/contenu/letude-insight-montre-lexistence-de-mecanismes-de-compensation-chez-les-sujets-porteurs-de>.

Other data that are not derived from neuroimaging are also acquired. Several questionnaires³ are mobilized to evaluate cognitive abilities in subjects with or without neurodegenerative diseases, with a particular focus on memory, language, and concentration disorders. These questionnaires are administered to patients during neuropsychological tests in the presence of a duly qualified clinician. Lastly, medical databases contain socio-demographic data, including age, sex, socio-professional category for subjects who are still working, and in some cases, race, and genetic data, the latter allowing for the investigation of the influence of the genome on the onset of certain degenerations.

All of the aforementioned data can be compiled and explored together, as exemplified by the subfield of *brain mapping*, which provides cerebral activity maps under some parameters – of physiological and cognitive kinds – and the recent *disease course mapping*, which aims to monitor the progression of triggered mechanisms responsible for the initiation and installation of a neurodegenerative disease. Figure 5.1 shows an example of the latter, which is a study produced by Aramis on the progression of Alzheimer’s disease in some patients of the ADNI cohort.

5.1.1.2 How are medical data shared?

Aramis’ research programme, which aims to use high-performance processing software to format large-scale health data, is helping to make these data accessible to researchers around the globe.

In France specifically, the use of health data in research and clinical practice is subject to a strict legal framework. Such data are collected, anonymized, and shared on dedicated hosting platforms with the consent of the patients monitored by a healthcare organization, in accordance with French law and bioethics regulations. On these platforms, specialist computer scientists and clinicians, who are often involved in the design of the cohort under study or in a related clinical trial, clean and homogenize the deposited data, with the aim of minimizing acquisition-related bias and rendering the data comparable for future statistical use in clinical research or in specific medical situations. The main platforms mentioned by the interviewees are the Health Data Hub⁴ (*Plateforme de Données de Santé*), the UK Biobank, the central neuroimaging platform CATI⁵ (in french, Centre

³From the interviews of some Aramis members and some of their scientific publications, we identified three commonly used tests: the Mini-Mental State (MMSE) or Folstein test (Derouesné et al., 1999), the 16-item Free Recall-Indicated Recall (RL-RI) test (Dartinet and Martinaud, 2005), the Rapid Frontal Efficiency Battery (BREF) (Van der Linden et al., 2004) and the Alzheimer’s Disease Assessment Scale-Cognitive Subscale (ADAS-Cog) (Rosen et al., 1984), which is dedicated to detect the symptoms of Alzheimer’s disease in patients.

⁴See at the following address: <https://www.health-data-hub.fr>.

⁵See at the following address: <https://cati-neuroimaging.com/>.

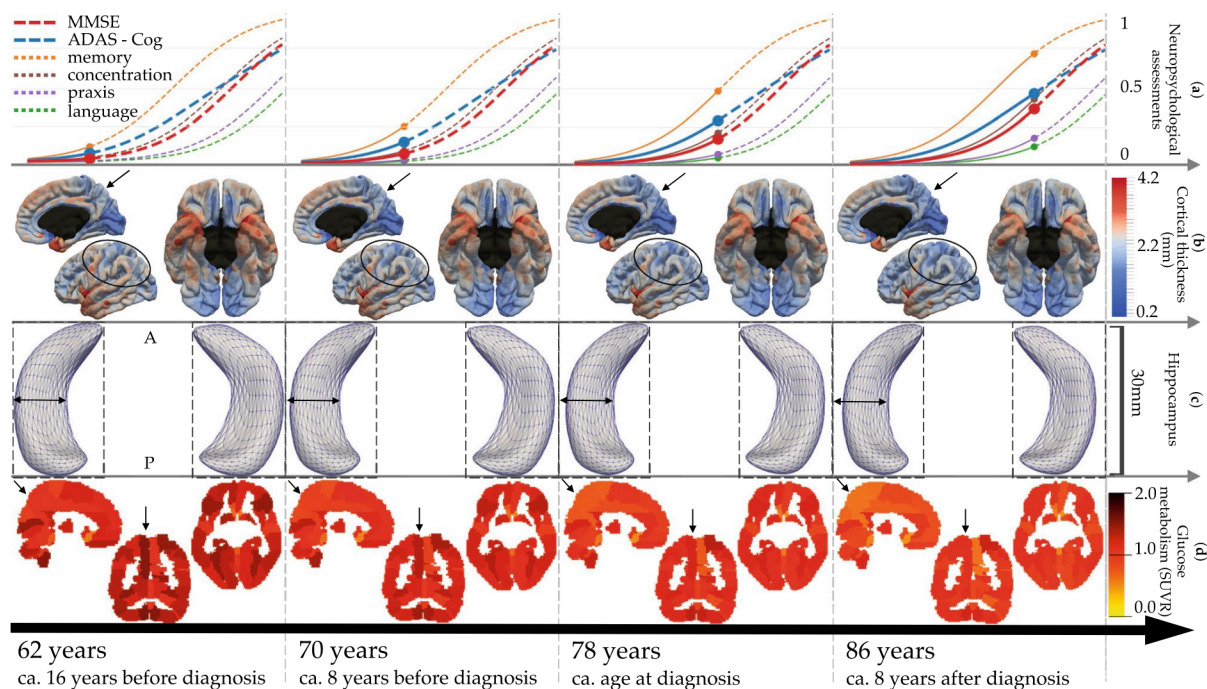


Figure 5.1: Illustration of the functionalities of the *AD Course Map* software, which combines various data to forecast the onset of Alzheimer’s disease. At four disease stages, representative of the average ages of the affected patients in the cohorts, are indicated the results of two neuropsychological tests designed to assess their cognitive decline, MMSE and ADAS-Cog, the latter being decomposed into four sub-scores – memory, concentration, praxis, language – (first row), the progression of multiple atrophies in delineated regions of the brain (second row), the deformation of the hippocampus (third row), and the intensity of hypometabolism (an abnormal, inferior metabolic activity compared to healthy patients at these ages) in different regions of the brain. Source: Koval et al. (2021).

d’Acquisition et de Traitement d’Images), CleanWEB,⁶ the Research Electronic Data Capture (REDCap),⁷ and the AP-HP Health Data Warehouse. In the context of clinical research, the data available on these platforms must be first controlled by moderators who certify their quality and validity before authorizing (or rejecting) their uses and investigations in various research, such as the Commission Nationale de l’Informatique et des Libertés (CNIL) and the Comité Éthique et Scientifique pour les Recherches, les Études et les Évaluations dans le domaine de la Santé (CESREES).

Other health data sharing solutions currently exist, particularly within the economic data market. However, these solutions are weakly regulated and face serious privacy concerns (Chen et al., 2019; Loukides and Lorica, 2019), which have been (in part) addressed by the General Data Protection Regulation (GDPR, or RGPD in French) that is effective at the European level since 2018. Nevertheless, much remains to be done in the construction of a universal legal framework for the transaction of these very particular data.

⁶French server created by the *Télémedecine Technologie* company, which is used to store and share data entry files or electronic metadata in the context of multicentric clinical trials.

⁷This internal AP-HP server is comparable to CleanWeb.

5.1.2 Collaborations outside academia

A logical continuation of the academic research led by ICM is R&D, here mainly oriented towards the development of innovative therapeutic solutions. This kind of research is primarily conducted by research departments of major industrial groups, which aims to improve the targeting of patients suffering from cognitive disorders at varying stages of progression and thereby to administer them an appropriate treatment. Interviewee I3 (neurologist and researcher, AP-HP, ICM, IM2A) is familiar with these collaborations, as evidenced in the dialogue below:

I3: [We, Aramis, entertain] partnerships with local industrial groups and start-ups that are incubated at the Institut du Cerveau where I work. There are two main partners with whom I work, a company called Genius Healthcare⁸ and another one ... Carthera.⁹ [The latter] is a company that develops innovative treatment tools for neuro-oncology, which we are trying to deploy in the field of neurodegenerative diseases, based on the use of ultrasound. ... Those are the on-site partners, I would say, and I also have ongoing partnerships with what could be called ‘Big Pharma’, namely industrial pharmaceutical partners that develop molecules against Alzheimer’s disease or related disorders, and from which [some of their members] come to test these molecules at our center.

SF: Are you working with them on the development of artificial intelligence algorithms?

I3: Not really. ... I mean, I know they are developing them, but in-house. I know that the director of our team has initiated a collaboration with Sanofi specifically for that, namely using artificial intelligence to identify patients for this or that molecule, for this or that drug, in [the context of a research on] Parkinson’s disease.

SF: Are any other pharmaceutical companies involved in your projects?

I3: No. In any case, [for] artificial intelligence ... it is Sanofi and that’s it. ... In terms of, I would say, medical care, there are more industrial groups working with us. There’s also Roche, Biogen...

SF: Do they promote funding for your research? If you don’t mind my asking...

I3: Yes, Roche is actually supporting us in establishing a cohort of patients from whom we are collecting a lot of data that will help us validate the artificial intelligence tools [developed in Aramis].

From this conversation, we learn that the main industrial partners of Aramis, except Sanofi, are not collaborating with the latter in the development of AI-related algorithms and software, as they are able to do so independently (in addition of easily patenting them) with their own knowledge and technological resources. As interviewees I4 (post-doc, ICM, INRIA) and I7 (neurologist and PhD candidate, AP-HP, IM2A) have highlighted in other discussions, pharmaceutical companies are especially seeking knowledge about

⁸Genius Healthcare is a company which creates therapeutic video games and health-specialized serious games. Further information are available at the following address: www.genius-healthcare.com.

⁹See at the following address: carthera.eu.

cohort data that have been collected mainly within university hospitals and public research laboratories. Although they can produce their own data, they still call on academic research to find subjects and set up test samples for therapeutic trials:

No company has any interest in promoting this kind of study because it is really, extremely expensive, running into millions [of euros], and they don't get any return on it just by looking at people. So [these studies are often conducted] within public research. However, that is not always true; sometimes there are private-public partnerships. On the other hand, pharma companies [and a lot] of industrial actors perform such trials, and when they do, they have patients and they submit them some medical tests, and in the end they have patient data. But if we [the public research sector] are not mentioned in the data use clauses, we don't have access to the data and we have to go back to the customers, get them sign papers again, in order to get their agreement. (I4, post-doc, ICM, INRIA)

Furthermore, according to I3, Aramis also works with start-ups at the ICM, which serve as privileged intermediaries between public and private research and are incubated in the iPEPS structure. However, most of the researchers interviewed have told us that they did not at all. Indeed, despite ICM's desire to bring together scientific research and entrepreneurship, these researchers admit that they are not interested in start-ups because they do not understand neither their approach towards innovation nor the way they reuse – and therefore transform – the knowledge produced by the ICM's scientific teams into concrete applications. The interviewee I8 (full-time researcher, CNRS, ICM) firmly affirms this trend: “For me, it is not clear what start-ups want, so I don't want to spend too much time on it.” While reinforcing this trend, the interviewee I1 (post-doc, ICM, QIMB) raises a completely different aspect, mainly of discursive kind, which adds to this lack of understanding. By highlighting the “*business-like*” character of the iPeps manager, he has indeed demonstrated a discrepancy between ICM researchers and entrepreneurs that is crystallized here by a specific vocabulary used by the latter but that is not always readily comprehensible to the former – the reverse may also be true. This aspect reflects an interactional barrier between two professional worlds cohabiting within the same institution.

These aspects thus challenge the translational nature of research at ICM, specifically in the context of applications of AI research related by our interviews, in particular through the inability of actors of both sides, i.e., academic research, and start-ups and industry, to move from one environment to another. Although I3 seems to be in the best position to do this, the rest of the researchers interviewed in the team do not participate in these collaborations on their own (including I1 and I8). Thus, at the time of this present investigation, such intermediary arenas seem to be only at a nascent stage, established by senior actors in the institution (clinicians and computer scientists), who have a broad vision of the scientific problem led by both the academic and clinical research and other

industrial partners, and who are able to bridge the gap between these two worlds with their different approaches to innovation, and who can attract different actors from these two environments towards such an arena.

5.2 A dichotomy of scientific profiles

As highlighted in Chapter 1, one of the main assets of the *research-technologists*, that is to say, the researchers engaged in the conception and the diffusion of AI in various scientific fields, is their ability to move between different research fronts. This presupposes an underlying ability to adapt to the needs of these research, or a prior interdisciplinary training that articulates different knowledge, research practices and epistemic cultures (Knorr-Cetina, 1999) – for example, in both neuroscience and AI.

As briefly mentioned in Sect. 2.3 of Chapter 2, and according to the composition of the team on its website and the interviewee’s narratives, Aramis’ members can be divided into two distinct groups: *computer scientists* and *clinicians*. In order to enrich the similar categorization provided by the data analysis of the large-scale neuroscience collaboration network presented in the previous Chapter 3, this section therefore proposes to recall and describe more precisely the typical profile representing each group, as well as the main social interactions between the two groups in the context of specific research projects mentioned by the interviewees. Since the two categories implicitly refer to different disciplinary representation,¹⁰ not only in terms of research culture and scientific knowledge, but also in terms of epistemic affinity (or proximity) to the instrument AI (its digital tools, programming, etc.), this part also aims to elucidate the nature of the interdisciplinary work produced by the aggregated actions of each of these researchers, given their respective intellectual trajectories and their insights into a common scientific challenge.

5.2.1 Computer scientists looking for concrete applications

The first group of scientists within the team includes individuals who propose modeling, simulations, algorithms, and software for the effective analysis and processing of medical data. Combined under the generic banner *computer scientists*, they are in fact mainly graduates of general mathematics and statistics courses at leading engineering schools. For instance, the interviewees I1 and I8 have both completed a PhD track on the digital processing of neuroimaging data and have followed subsidiary training in biomedical research, in epidemiology and biomedical engineering, respectively. Interviewee I4, who is

¹⁰For the author of this present thesis, but it may not be the same for the reader.

currently a post-doctoral student, only turned to medical applications during his doctoral studies, thanks to data science, which he discovered “*by chance*” during an internship and another dedicated training programme. As the team also needs non-researchers for other purposes, some computer scientists arrive in Aramis with a career that is often far removed from research and biomedical sciences, like the interviewee I2 (R&D engineer, ICM, INRIA), who has been employed as a R&D engineer within Aramis after a brief stint within the Institut National de l’Information Géographique et Forestière (IGN).¹¹ The interviewees I2 and I4 are also singular cases in that they have met biomedical research and neuroscience at a relatively advanced stage in their studies and/or professional careers, as they have not been trained in these domains. Regardless of an underlying, inherent vocation, they have decided to join these domains after a substantial reflection on their respective subjects of interest, while simultaneously wanting to work exclusively with computational tools and data. This versatile, “multi-applicative” nature of these profiles clearly shows that biomedical research, as well as neuroscience, were not initially a priority for them:

Following [my stint at IGN], I decided to reorient myself a little bit, well to change sector a little bit because ... I was getting a bit farther away from the technical side and I wanted to stay close to it. So I joined the Aramis team [to this end]. (I2, R&D engineer, ICM, INRIA)

[W]hen I entered [engineering] school I said to myself, “I really don’t want to do maths and physics and continue [on this path], I want to have a closer grip on the real world.” So I started to do a degree in economics in a university at the same time [and] I thought about doing a double degree with a business school. Nothing predestined me [to do this] and nothing made me want to do it. But in retrospect, I think I wanted to do all that because I wanted to apply what I knew and not let it remain a bit esoteric. Then I arrived a little bit by chance in a US computer science research lab for an internship ..., and I just realized: “in fact, I can use everything I have learned since [the beginning of my studies], via data science and machine learning.” And particularly at this time, I wasn’t into computer science, so it was not my thing and I didn’t know that it could be a way of applying ... equations, models, algorithms or whatever. And I think that’s what I liked about it. ... That’s the story as I tell it right now, at the time I’m not sure I was that conscious; I liked it for a while, I thought it was cool and it was nice. (I4, post-doc, ICM, INRIA)

5.2.2 The medical profession in need of effective computing methods

The second group of scientists at Aramis comprises hospital or independent practitioners with additional research activities. In a significantly lower number than the computer scientists – with only four out of a total team of forty members in 2021 –, these neurologists

¹¹French geographic information institute.

or neuropsychologists by training and profession, who will be named as *clinicians* in the following, are interested in the numerical methods developed by Aramis for in-depth studies on a few clinical aspects of some neurodegenerative diseases, including Alzheimer's. However, they do not participate in work involving conceptual frameworks in mathematics and computer science, as they lack the requisite expertise. Consequently, clinicians often remain detached from the modeling of observed phenomena, the actual writing and optimization of algorithms, and software development. This is particularly true for interviewee I3:

Honestly, as far as I'm concerned, I wouldn't call myself an artificial intelligence researcher, [I mean that] I know the principles of this research, I understand the ins and outs in broad terms, but I'm personally incapable of developing a decision algorithm and optimizing it. On the other hand, I can talk to artificial intelligence researchers without too much difficulty... Well, even if I don't have the knowledge to understand the math behind it, I do understand what the algorithms do, and the data they work with. (I3, neurologist and researcher, AP-HP, IM2A)

However, this apparent lack of technical AI-related knowledge expressed by clinicians does not prevent them from bringing their expertise in medical knowledge (neurology, neuropsychology, physiopathology, etc.) to assist computer scientists and mathematicians with their more formal approach, in particular in the data used to train the algorithms, which computer scientists do not fully assimilate, as the interviewee I2 aptly highlights:

There's really a big time to understand the data, [of] understanding the subtleties around these data. It's a phase where you really have to read a lot... well, [of] clinical things, ask clinicians for questions to understand the subtleties of these data. (I2, R&D engineer, ICM, INRIA)

Interviewee I5, a neuropsychologist working on her thesis on anosognosia¹² and memory complaints¹³ in the early stages of Alzheimer's disease, affirms the need for computer scientists and clinicians to share knowledge in the context of clinical research:

SF: Do you participate directly with these teams in the creation of models, codes if necessary, or in the improvement of the software?

I5: For code creation, never. ... Maybe I will use the code, but I don't write it at all, because I'm completely incapable! But before... if they [computer scientists] want to implement something, for example, then maybe they call me and tell me:

¹²Anosognosia is a neuropsychological disorder, a cognitive pre-degeneracy characterized by an absence of awareness of one's own illness.

¹³Memory complaints are one of the earliest symptoms of Alzheimer's disease, which, however, not always manifests in patients who develop the disease at a later time. These complaints consist of memory loss that the patient himself or herself recalls during neuropsychological interviews or tests in the presence of a healthcare professional.

“Is it better to use this score or the other one? Which is the most sensitive, the one that varies the most, that isn’t very reliable, etc.?” And that is when I give them my idea, which would help them to create the algorithm.

This clinician also uses the *Leaspy*¹⁴ software developed by Aramis as part of her research, which demonstrates the reciprocity between clinicians and computer scientists in terms of exchanged expertise. Aramis’ clinicians thus contribute indirectly to the development of medical AI designed by computer scientists for subsequent uses in research or clinical routine. However, as pointed out by the interviewee I7, who is a neurologist working on the genetic origins of rare neurodegenerative diseases such as amyotrophic lateral sclerosis¹⁵ and fronto-temporal dementia¹⁶ in the framework of his doctoral thesis, “they [clinicians] are not so accustomed to using algorithms”, and for a good reason: most clinicians do not have advanced quantitative training – including those interviewed. Nonetheless, some of them do have necessary coding skills to analyze their own clinical data, often relying on pre-designed libraries and sets of robust routines coded with various programming languages. In other words, the only computer-related work that clinicians perform seems to be restricted to the mere use of digital tools that are already developed by computer scientists, without necessarily delving into the details of these computational programmes, or even modifying them to their convenience. For example, to calculate genetic risk scores associated with the emergence of some neurodegenerative diseases, interviewee I7 extensively programs with ready-to-use “toolbox” coded with R or Matlab language. However, he concedes in the following passage that his programmer’s skills are rapidly becoming limited:

I don’t have the information and expertise to carry out this research on my own, that is also why I’m part of the Aramis team, and I work with other PhD students and researchers. ... [A]ctually, working ... on machine learning algorithms isn’t always easy for me, as I haven’t had any advanced training in that. (I7, neurologist and PhD candidate, AP-HP, ICM)

Our interviewees have also shared another solution to compensate for this lack of technical skills. In the context of clinical studies involving large cohorts of subjects with or without neurodegenerative diseases, the clinical staff often rely on statisticians with expertise in medical data, who are working directly on-site at the hospital and able to

¹⁴Acronym of *LEARNING Spatiotemporal Patterns in Python*, which is a Python package for longitudinal health data analysis developed by Aramis, and which is mainly based on advanced statistical and machine learning techniques. The library can be downloaded from the shared directory of computer programs proposed by Aramis on the GitLab platform: <https://gitlab.com/icm-institute/aramislab/leaspy>.

¹⁵Also known as Charcot’s and Lou Gehrig’s disease in French and English, respectively, this neurodegenerative disease slowly affects muscle tissue until complete paralysis is reached.

¹⁶Affecting regions of the brain around the frontal lobe, this disease results in a range of symptoms including language, behavior and personality disorders. Memory disorders are also observed, though to a lesser extent than in Alzheimer’s disease.

produce in-depth quantitative studies that clinicians would be unable to conduct themselves alone. The aim of these statisticians is often to statistically determine risk factors responsible for the onset of these diseases. The interviewee I6, one of such professionals, briefly explains this approach in the following passage:

My regular job is to help researchers, physicians, nurses, or whatever, who have a problem [for their research] and therefore need a statistician to answer their question. So, either they have already collected the data and are well advanced in their study, or they come to see me at the very beginning to write the research protocols ... or, when they want to do retrospective studies and go back to the archives, they come to see me to learn a bit how to build the database, to think about how we are going to do the stats¹⁷ and whether or not it is worth doing the stats analysis. (I6, statistician, AP-HP, IM2A)

Although they are not affiliated with any particular team and are spread throughout the hospital site on short-term contracts, and in very few numbers in rather isolated departments, these statisticians are of crucial importance for rigorous research:

[I]t's always a pain to find a statistician; in fact researchers have trouble finding a statistician... and most of the time they don't find one so they do their own stats and they do anything, and that's normal because they haven't necessarily been trained for it. Once again, it is a job, so if you are just doing comparisons of averages, [or just calculating] p -tests or Wilcoxon [tests], that is a basic thing to do and OK, why not, but when it comes to models that are a little more complicated, [it] is better to come visit a stat[istician]. And when they find one, they don't let him/her go! (I6, statistician, AP-HP, IM2A)

In conclusion, like the statisticians at La Pitié-Salpêtrière Hospital, the computer scientists at Aramis serve (among many other functions) as a “support computing team” whose goals are to assist clinicians in their regular research (and maybe medical) practices. This need for simple, effective computational methods is a clear manifestation of the demand expressed by clinicians for the expertise that they do not all possess.

5.2.3 Computer scientists at the service of clinicians

The clinician-researchers in the Aramis team are the key contacts between the computer scientists and the medical staff who are qualified to deliver diagnoses. Due to their professional experience in the hospital environment, they are aware of the potential of AI technologies in their routine medical practice, which motivates them to closely guide the corresponding research in order to obtain tools that align with their needs. Interviewee

¹⁷Diminutive of “statistics”.

I3 does not hesitate to express the technological expectations of the medical community when responding to a question asked by the interviewer about the software he would likely use in his medical practice:

SF: How does the programme you're using work? Is it just a computer programme with an interface, you enter data in it and something comes out?

I3: That's it. That's it, exactly. It produces a prediction of diagnosis and evolution. [...] The software would also fetch the data from a system that we use for our computerized patient records. That's what we would like to do: we would like the process to be more fluid, to consume as little effort as possible for the medical community, to be a truly integrated approach to current clinical practice and not add extra steps to integrate the data, enter it into a system, and so on. Because they are pretty time-consuming, the ergonomic nature of these tools is really important to consider, I think.

The few interviews conducted thus reveal an implicit subordination relationship between clinicians and computer scientists, which is accentuated by the simple fact that the Aramis team is based directly on a hospital site. In this way, the latter work first for a medical cause, rather than the other way around – whereby the medical cause is used to enrich AI-related knowledge. The comments of interviewee I8 below corroborate these apparent unilateral relations for knowledge creation, namely the translation of clinicians' expectations into formal languages that can lead to a software application in radiology:

SF: Do these radiologists, for example, help you create codes or algorithms?

I8: When I developed the method, yeah I got in touch with a neuroradiologist. ... [W]hen I proposed my research programme to CNRS, I basically had something in mind, I discussed it with her and she said, "Oh well, no, but in fact I would prefer that". So, it is true that, by talking things over, you realize what everyone wants and you adapt, for sure. But then when it comes to algorithms, [...] we have to translate what [the radiologists] can tell us, because they won't necessarily have the... they won't know what it can correspond in concrete terms [for mathematicians and statisticians]. But yes, interaction with them is very important for me because, in any case, the tools we develop are intended for them, so if they don't meet their expectations, there is no point.

IT and digital neuroimaging specialists then become the clinicians' little hands and must adapt to the demands of the medical profession. To use the expression of interviewee I2, a "culture shock"¹⁸ can occur for a computer scientist immersed in the world of clinical research if he or she has not been trained in the domain beforehand. New doctoral students or computer engineers who are new to the field must familiarize themselves with the main principles of biology, physiopathology, and the issues surrounding the clinical data they use, in order not to fall into the trap of misunderstanding and risk producing

¹⁸Translation of the French expression "choc culturel".

scientific aberrations. In this respect, I2 mentions a clinical research's way of thinking that is significantly different from what he experienced during his engineering career, in regards to his own representation of the scientific investigation process:

Even when I worked with the neuropsychologist [interviewee I5], she often started from... she really had her hypotheses, she had her ideas and the result had to be [the latter]. [It's] a bit the inverse, in a way, of the scientific approach. (I2, R&D engineer, ICM, INRIA)

Conversely, not all clinicians adapt to IT and do not make coding their primary activity. As mentioned above by the interviewees I2, I5 and I7, the clinicians remain quite distant from computer programming, despite their affiliation to a team that places it at the core of its research activities. A long discussion with the interviewee I4 concerning a tool he developed during his Ph.D. studies to forecast the progression trajectories of a neurodegenerative disease based on various medical data exemplifies the distinction in the roles played by computer scientists and clinicians in typical Aramis research. His research is divided into two main stages. The first consists of developing models capable of learning such predictions from a voluminous set of medical data (see Sect. 5.1.1.1). This stage essentially concerns computer scientists, and clinicians help them understand the training data. For example, I4 first modeled the phenomenon under study by constructing a customized data structure, then determined a cost function, and finally coded an algorithm to optimize this cost function in order to best represent the data and subsequently generate an accurate statistical prediction. It is precisely at this stage that computer scientists are best positioned to offer proposals. The second stage of I4's research consists of validating the learning algorithm that has been devised, usually with the help of both R&D engineers – including I2 – who are supervised by researchers, and clinicians at the end of the process, whose job is to promote the software to external organizations outside the domain of medical research. In this approach, which could be described as a *demo* (a demonstration) in the sense of Rosental (2002), the researchers aim to patent the software and implement it in current clinical practice:

[W]e have a R&D approach ... [Recently, our process] looks a bit more like a research ... [and] development within a company. We try to promote [the software] to the various institutions, academic partners, and laboratories with which we work. I'm doing much less research in fact. I'm more in charge of a team of four-five engineers with whom we work collectively to deploy this kind of tool. So, always with a scientific objective in mind, we try to validate what it is made for. One of our projects, for example, is to deploy a tool for understanding the evolution of patients suspected of having Alzheimer's disease, and in fact these tools are not at all for patients, they're for neurologists. We are [also] working with an anthropologist who is trying to understand what changes [these tools] may bring about in the way neurologists deal with their patients... . (I4, post-doc, ICM, INRIA)

At the end of this process, the interviewee I3 endows a mediating role in order to promote AI to the major French public health and personal data protection authorities:

In fact, I have two roles: I have a role in recovering data of interest to develop and validate these algorithms [and] to ensure that they are robust, then a role of implementing them in everyday practice. More recently, for example, we had a discussion with INSERM to promote a research protocol we would like to set up at the hospital. It's a project to implement these tools in everyday practice, a tool we would call *Predictema*. From a regulatory point of view, this tool is called a medical device,¹⁹ because it is a software that processes data and provides information that we didn't have before. So now we need to validate it [the medical device] in clinical research and get the necessary agreement from various agencies, that is to say the CNIL,²⁰ the ANSM,²¹ which also deals with medical devices, personal protection committees... So ... I have a direct impact on submitting [the device] to the regulatory authorities and then ensuring that the protocol runs smoothly in the clinic, because we are in contact with patients and we have to explain them the protocol properly, explain what it's all about, what we are going to do, obtain their [informed] consent... So all this is a biomedical research approach in which I think I bring expertise compared to researchers who are very... IT specialists, but who are not in contact with patients, and there is a real need for both aspects. Because you can have the best tool in the world, but if it's not applicable in real life at all, then we won't do anything from it. (I3, neurologist and researcher, AP-HP, IM2A)

This privileged status as a spokesperson for these organizations at the national level is indicative of his detailed understanding of the medical field and the issues surrounding the use of algorithms in medical practice, which computer scientists seem to lack. With his status, he therefore reinforces his position as a team leader, thereby accentuating this subordination of clinicians to computer scientists. The latter is even more apparent when it comes to the publication of scientific results. During a conversation on his transition from pure computer science to clinical research, interviewee I4 talks about the different ways computer scientists and clinicians publish and communicate their findings, which are crucial elements during his gradual adaptation stage into biomedical research:

We have just published something on Huntington's disease, so it's really for clinical neurologists, and we spent a lot of time [writing it]. In fact, we realized that our writing and our way of explaining what we do were not suited at all. And it took a lot of work to get it published, but we finally did it! ... [For another] article in review ... it was a huge relief for us, because it means that we have made progress at least on that part, which is giving our knowledge to another community. But we were helped! In this case, the doctors really took from us the pen and we dictated to them, well [they] translated it. In fact, thanks to this pooling ... we were able to do this more easily. ... There's another thing that's quite funny, *open science*, which comes from *open data*, I mean the codes that are freely available [and] the aim is to do something for everyone. Sometimes I went to conferences on biology or

¹⁹Translation of the French, juridical expression “dispositif médical”.

²⁰Commission Nationale d'Informatique et Liberté, in French.

²¹Agence Nationale de Sécurité du Médicament, in French.

neuroscience stuff, where you can find whole workshops on open science, you know, and people are still talking about: “Are we going to make papers freely available?” And I was like: “What do you mean? Yes, of course!” In fact, I’m trying to understand, well I understood that a year ago, but I didn’t know that there were still communities which didn’t make their papers open-access when they submitted them, whereas it is evident for us. ... I think [this way] is very “hard science”.²² (I4, post-doc, ICM, INRIA)

As scientific results are primarily destined for clinicians, the entire process of writing a scientific article is carried out by neurologists and neuropsychologists in order to ensure that the resulting publication is accessible to their professional community. A priori, it is impossible for computer scientists to bring doctors outside their strict disciplinary constraints, by overturning the codes for writing and transmitting knowledge in clinical neurology. Furthermore, the aforementioned excerpt illustrates that he omits to address the regulatory and personal data concerns that impede the unrestricted accessibility of research materials and findings, as previously highlighted by interviewee I3. All of this reflects a disciplinary gap in the way science is done and knowledge is disseminated.

5.2.4 Testing interdisciplinarity

Promoted and facilitated by institutional arrangements promoting dialogue between computer scientists and customers, in particular the partnership between INRIA, INSERM, CNRS, and AP-HP, Aramis is a place that encourages dialogue within a heterogeneous scientific population engaged in producing original interdisciplinary knowledge that falls within common research goals. As defined by Stokols et al. (2008), interdisciplinarity can be understood as the interaction of several disciplines resulting in the creation of knowledge that incorporates different conceptual and methodological frameworks. Such an enterprise contributes to softening traditional disciplinary boundaries for a common scientific purpose, which allows the mobility of researchers at different career stages and trained in various environments.

As witnessed by Aramis, in particular through the interviews collected and the results mentioned above, we recover the main characteristics of this interdisciplinary model in the organization of scientific activities within the team. Indeed, each member brings his or her own knowledge and skills within a well-defined disciplinary paradigm. A neurologist does not improvise himself/herself as a computer scientist, as shown by their lack of programming and software development skills. Similarly, a computer scientist is not a

²²Translated from the french expression “sciences dures”, which refers generically to basic, natural sciences around mathematics, physics, engineering and computer science, at the difference of biomedical sciences – even humanities and social sciences.

recognized clinician who is also a member of a medical order or council,²³ although the latter is more involved in this interdisciplinary approach than clinicians without research activities. However, this research model is challenged by the fact that no one in the team seems to really master simultaneously the two disciplinary expertise enabling them to think differently about the same scientific problem. In particular, this leads to misunderstandings between the two typological groups, as I2 (computer scientist) suggests when relating a working session with I3 (clinician):

At a presentation ... I think he [interviewee I3] had misunderstood what we had presented. We had tried to present the biases in the predictions of our model. ... [M]aybe it was the way we presented it, it was a bit confusing at some times. But I have the impression that the fact that he didn't understand [what we have presented] made him... yeah, it made him a bit nervous, ... he had misunderstood things and that made him a bit nervous. ... We were ready to use the method but we [finally] weren't, well our results tried to validate our method, and he had the impression that we wanted to go a bit faster, which wasn't the case. ... After that, ... sometimes I feel that, if you're not careful, it can quickly get bad to you, because if you don't take the time to ask questions about the data you have, you can quickly find yourself saying something stupid, in which case the neurologist will correct you. Because there are always subtleties in the data we have, it's not pure data. There's always some interpretative work to be done before you can use it, and if you use it out of the blue, like some people do in the spirit of pure machine learning, ... you're going to be talking nonsense. (I2, R&D engineer, ICM, INRIA)

This disciplinary dialogue between computer scientists and clinicians therefore appears at first sight to be difficult when researchers are directly questioned about the results produced by computer science. Furthermore, a gap still remains between the computational reality of the accuracy of the results that models and algorithms can produce, and clinicians' desire to obtain certain, error-free results. The story of the interviewee I4's below illustrates this last point, focusing in particular on the perception of manipulated data by the two groups:

[W]e [the computer scientists and IT engineers] have a very different conception of data than doctors do, and I really enjoy titillate²⁴ doctors' views on what data is. So here's an example, that's a bit silly but really important... Imagine that you have a patient in front of you, you give him a test and you want to predict what he's going to be in the next five years [whether or not he/she is subject to Alzheimer's disease]. Obviously ... you want a null error. So there's a [onset] score of ... 10 right now and he's actually going to be 20 in five years, so you want to predict that he will [get a score of] 20. But when you look at people's scores, if you give the same person the test on two different days, they [might] have a five-point difference, because they were angry in the morning, they drank a coffee, ... they were ill, whatever, but there's a different variability. In fact, the ambition to find a perfect prediction means predicting things that are independent of the disease [the

²³In France, this is called the "Ordre des Médecins".

²⁴Translated from the French expression said by the interviewee I4, "challenger la vision des médecins".

interviewee waves his hand to illustrate a noisy curve with high variability]. But that's not what we want! It's funny because it is typically the kind of thing we try to [change in the mind of] doctors who see the data as something pure and perfect. ... So when you say to them, "I made a two-point error on the prediction, ... and in fact I'll never do better and I don't want to do better", well, the first thing they will answer to you is: "Yeah, but we want a perfect prediction". (I4, post-doc, ICM, INRIA)

In this passage, in which he seems to be taunting the doctors, I4 supports this real disciplinary separation between the computer scientists and clinicians, which also implies various stereotypes of each towards the other in terms of their conception of AI, computer science, how the problem is posed and, in this particular case, what a model is. The clinicians' quest for purity in the data and models they use is undoubtedly an inherent part of the medical discipline to which they belong. Indeed, if these clinician-researchers, who are first and foremost medical professionals, are supposed to use the (AI-related) software developed by Aramis to provide reliable diagnoses for their patients, then these computer programmes must operate with as few errors as possible. Therefore, a compromise must be found between what computer scientists can achieve within the limits of IT and the ambitious expectations of clinicians.

To conclude this section, this apparent interdisciplinarity in terms of the organization of work and day-to-day interactions within the team underlies a unilateral relationship between the expertise brought by the disciplinary frameworks specific to each of the two groups: computer science expertise is advanced for clinical and biomedical research, but the reverse is not necessarily true, as there is no return or feedback to the former, except for the nature of the data manipulated.

5.3 The diffusion of AI in neuroscience research: between promise and reality

AI has become an essential tool in neuroscience research. However, it is subject to a number of biases specific to the disciplines that use it, resulting in divergent representations across the disciplinary spectrum of scientific research. The accounts gathered during the interview campaign illustrate, albeit on a small scale, this dissonance in neuroscience, which tends to accentuate the disciplinary divide between computer scientists and clinicians, and which in fact hinders interdisciplinarity on the subject. The lack of training in the main AI algorithms during their university studies may help to explain the various points of view later expressed by the respondents.

5.3.1 A still recent bubble

Although AI is currently central in various research, both in terms of its development and improvement and in terms of its everyday use, the Aramis researchers interviewed, whether computer scientists or clinicians, stated that they had not been exposed to AI in their training prior to joining the team, apart from the main machine learning tools that are most successful today. The late democratization of AI in French higher education and research may explain this lack of training among both clinicians and computer scientists. For example, a large number of intelligent algorithms have already been known for several decades, developed, mastered, and even passed on to younger generations of students through training in data science or statistics in a *Big Data* approach, often using supervised or unsupervised automatic learning. However, the term “AI” has only belatedly established itself as a unifying research front in the French scientific landscape and higher education, as evidenced by the government’s announcement of a vast funding programme for AI research throughout the country, launched only in 2018 (Villani et al., 2018). This was well after the interviewee I2 had left his engineering school in 2015, which explains why he had little or no contact with AI during his studies:

Before [entering in engineering school] honestly I didn’t have ... any particular training [in AI]. It wasn’t the generation where... well, we didn’t talk too much about it, we did the maths, ... operational research, stuff like that, but [no AI]. Then at the [school], not very much either, I would say ... it wasn’t yet so fashionable. ... I know that they added a lot of lectures at the school [I frequented] ... on big data, artificial intelligence, machine learning, and so on. ... It’s recent, but it wasn’t really the *boom* (sic.) yet. There’s quite a lot of progress that has been made recently, ... actually in the world of work. (I2, R&D engineer, ICM, INRIA)

AI research therefore developed very rapidly from 2015 onwards, not only in response to the sharp increase in the volume of data to be processed, regardless of the scientific discipline under consideration, but also because of the new performance records set by neural networks in processing these data. For example, the annual ImageNet Large Scale Visual Recognition Challenge (ILSVRC) organized by the ImageNet project in 2012 demonstrated the unrivaled robustness and efficiency of a neural network for image recognition and segmentation. The development of the U-Net convolutional neural network architecture (Ronneberger et al., 2015) also helped to establish the future dominant position of deep learning in biomedical image processing research just before 2015. Interviewee I8, who was also not exposed to AI during her training, nevertheless mentions a craze for this research in the inner circle of image processing specialists:

I think things are improving a little bit, because at the beginning, everyone focused on topics that required the least amount of time for the maximum amount of results.

So that phase has now ended, and all the results that could be obtained very easily have been obtained, but now we have to think a little bit! ... Now things are becoming more interesting, I just see it in the major conference in our field, called MICCAI.²⁵ ... I used to go there since 2013, so 2013, 2014, 2015 were fine. 2015 saw the first [deep learning] methods, but it was still very new, it was all fun! Well, 2016 really started to explode, and 2017, well, I was... In fact, half of the presented papers treated segmentation problems whose [authors said]: “We took this network, we got this result, wow, great!” So in 2017 ... I got bored. 2018 was a bit the same, and then 2019 started to tackle more interesting subjects, such as whether we could model the uncertainty that exists in the network, whether we could predict when the network is going to do well or not, and so on. ... I haven’t really followed the 2020 [session], ... but it has evolved a bit now, so I think the field is starting to produce more interesting things. But a few years ago, 2016 to 2019, pfff... (I8, full-time researcher, CNRS, ICM)

In the end, I8’s intervention portrays a fad in scientific research, which she does not necessarily consider to be a guarantee for the creation of solid knowledge:

Another criticism [that I have of deep learning] is the “cooking”²⁶ aspect. ... I mean, the typical story is that a PhD student has spent six months cooking with his little network, removing layers, adding layers, changing hyperparameters, one day it works, he will publish it and what have we learned? Well, nothing, because all you have to do is add a layer and the whole enterprise falls apart... (I8, full-time researcher, CNRS, ICM)

So in her view, neural networks are toys that no one really knows how to use, as mentioned in this brief comment:

On the one hand, people who aren’t familiar with medical images get into [deep learning] because it’s fun, but they don’t understand the data, and on the other [hand] I think there are also people coming from medical image processing but who don’t know much about deep learning, and it’s a disaster too. (I8, full-time researcher, CNRS, ICM)

All of the aforementioned I8’s comments show how easy it is to integrate deep learning into neuroscience research, mainly because the techniques are easy to learn with implemented, concise, ready-to-use packages on standard programming languages.

²⁵Medical Imaging Computing and Computer Assisted Intervention, which designate both a scientific society and one of the best known conferences for disseminating advances in the domain of medical imaging. For further details: <https://miccai.org/>

²⁶Translated from the French expression “côté/aspect cuisine”, here used by I8 to draw an analogy with the work she observes, where in this context, researchers just have a common basic recipe (the neural network) and modify it by mixing some ingredients (the hyperparameters, the number of layers, of neurons, etc.) whose quantities change, and thus producing a completely new dish (the result).

5.3.2 An unclear but promising concept

As developed in Chapter 1, AI covers a wide range of concepts and definitions that are very well documented in the scientific literature, but which are still the subject of debate in the scientific community. As E. W. Shortliffe rightly pointed out in (Patel et al., 2009), doctors often use machine learning techniques without really knowing that they fall under the umbrella term of “artificial intelligence”. Thus, the various computational methods of automatic or statistical learning do not at first appear to be cognitively related to this concept, which then becomes open to multiple interpretations. The Aramis members interviewed, especially the computer scientists, emphasized this lack of a clear and precise delimitation in the definition of AI, which could be partly attributed to the lack of prior training in the main algorithms and associated methods during their higher education studies, as previously described above. The testimonies of the interviewees I2 and I4 illustrate this last point well, the former accentuating the vagueness of the concept of AI by describing it as a “big word”, the latter in the passage below:

SF: Had you ever heard of artificial intelligence or data science before you went on this Master’s course or internship you told me about?

I4: Um, I don’t know, I would distinguish between AI as it appears in the papers, yeah I’ve heard of it, on the other hand... Well, actually I don’t even know what that means. ... I was interested in the mathematical concepts that were behind it to find out whether I was interested in it or not, whether the way of thinking this science was a bit linked to what I liked to do or not, and I think yes. But only from a technical point of view!

For clinicians, on the other hand, the situation seems quite different. Indeed, interviewees I3, I5 and I7 seem to talk about this concept confidently and concisely without too much difficulty, and do not hesitate to equate AI with machine learning. We might then naively ask why there is such a gap in the attitudes of computer scientists and clinicians to the definition of AI. Interviewee I6, who regularly interacts with clinicians who are not affiliated with Aramis and who are probably less expert in AI than the three above, brings an answer:

What I’ve seen since I arrived in 2017, researchers and doctors often visit me and say: “I want to do AI, machine learning, etc.”, because it’s hype and it’s really sexy (sic.) in the articles. But when I calm them down and see what they really want, machine learning is definitively not what they want to do, actually. They tend to think that now it’s the new thing in statistics, it’s the sexy thing, you really have to do it all the time, whereas in practice ... classical statistics will answer [their] questions. (I6, statistician, AP-HP, IM2A)

I6’s comment, which echoes I8’s earlier comments about a generalized, global craze for AI, shows that clinicians who are unfamiliar with statistical and computational methods

often hold misperceptions about machine learning and, more broadly, about AI itself. Despite this discrepancy between this perception and the reality of computational work (as previously shown in the preceding section), which is often performed by specialists other than clinicians themselves, AI is brandished as a kind of label or brand, especially by the latter. This is exemplified by the aforementioned expression from I6, “AI is sexy ... in the articles”. This results in various dimensions linked to the academic visibility of researchers’ works, in which AI has become a pivotal vector. The following sections provide further details on this point.

5.3.3 The need to be up-to-date to maintain visibility

During an exchange on the growing tendency of students to absolutely want to engage in research that monopolizes deep learning methods, regardless of the concrete application, interviewee I8 shows the extent to which these methods have pervaded minds even in higher education:

Now those who come out of schools have had training [in AI]. It’s perfect, but between the fact that there is very little theoretical support for deep learning and the fact that people like me who got into it because... well, because! If you add all that together, it’s clear that it’s complicated to make meaningful progress. But then, not doing it is also... In a very trivial way, all the people who come out of schools and universities now [and] who want to do a Ph.D. want to do deep learning. ... If I propose a subject with classic registration optimization, I don’t think I will hire anyone. (I8, full-time researcher, CNRS, ICM)

The unfinished sentence “but then, not doing it is also...” is indicative of a widespread phenomenon, namely that researchers are unwillingly thrown into deep learning without any prior training, at the risk of being sidelined from ongoing research. This sidelining can be both financial, as a research topic that does not include deep learning is no longer considered promising enough, and social. A researcher who is reluctant to use deep learning may suffer a loss of recognition and credit from the international scientific community because of his conservative tendency to use old knowledge and methods, which can also be seen as a sign of resistance to novelty and to new progress. To avoid these pitfalls, computer scientists who have not been trained in AI have no choice but to learn some techniques themselves. Interviewee I8, for example, has learned deep learning on the fly “like many” of her colleagues, although she regrets being caught up in this trend:

SF: You’re in a team where researchers are quite into machine learning and deep learning, but you have never practiced any, so how do you manage to fit into that team, that topic?

I8: Actually, even though I had never really practiced it before, ... it's a topic you just can't escape, so you have to get into it. Well, I did, and that was also the idea when I joined the team, where I was going to learn new things, which I did. ... The work I did in my PhD thesis was actually one of the first to do image synthesis. So when all the articles on image synthesis based on deep learning came out, well, I got a lot of them to review, so we are learning little by little!

SF: But then, perhaps your work have also been overshadowed by machine learning in a way, hasn't it?

I8: Yeah... A pretty good example is the work I started on anomaly detection during my thesis, which resulted in a conference paper, and I had always the aim of converting it into a journal article. Except that it took me years to do that, because I always had millions of things to do in parallel, and when deep learning has come along, it has become complicated to make people accept work that doesn't contain deep learning.

This exchange with I8 also reveals the high capacity of AI to obscure (almost) entire areas of research with competing methods in neuroscience – at least in neuroimaging – which in practice, as I8 confides later in the interview, prove to be just as effective. Deep learning thus becomes an inevitable feature within the field of neuroimaging, to the detriment of techniques that had turned marginal at the time of its spread, such as those developed by I8 before she was recruited to Aramis.

5.3.4 A practical tool to help you stand out in clinical research

Perceived as a constraint by some computer scientists who, like interviewee I8, are used to mobilizing their “traditional” methods, AI is being treated quite differently by researchers outside this group. Some clinicians, aware of the potential of AI-integrated algorithms and software, do not hesitate to praise its practicality in their own research. This is particularly true for interviewee I5, who makes extensive use of the *Leaspy* software developed by Aramis. In particular, this software has enabled her to validate a clinical intuition born of her experience with patients, which actually contrasts with the results she could find in the scientific literature on the subject, which was rather poor at the beginning of her doctorate:

When you have a lot of subjects who have had several [medical] visits, who are healthy subjects with mild or severe disorders, hence in different phases of the disease, you make longitudinal models. We can see that the disease is so variable, so one subject may have an early-onset disease and another a later-onset disease, ... one subject may manifest the disease through an initial symptom, and another subject through another symptom. ... So this model, with the software called

Leaspy, allows to homogenize perhaps all these subjects and their differences and variability, and to create a progression of the disease based on the longitudinal visits of each subject. It will place each subject on a timeline (sic.) that has been created on the basis of the subjects we are studying. ... In this way, we were able to trace the progression of the disease. For example, I was able to trace the progression of memory complaints and lack of awareness of disorders' symptoms. In fact, I was able to find this although it's not so obvious if I have used classic statistical models... [Leaspy] removes an error and a bias which both still exist when we study sick subjects who present a disease that is so heterogeneous, so different [from one subject to another], whereas we can really discard all these individual differences to really describe the disease in this way. (I5, neuropsychologist and PhD candidate, AP-HP, ICM)

The analysis tools developed by the computer scientists at Aramis are even a fruitful alternative to the “classical” (and sometimes complex) statistics that some statisticians and clinicians are accustomed to using for clinical studies in large observational cohorts:

SF: But could you have implemented your idea without these tools or algorithms? Would it have been more difficult?

I5: Yes, but no! In fact, as I really wanted to go against this idea advanced in the literature, which is very much followed – it's a very active literature now and I'm saying the complete opposite –, I really needed clear and striking results. ... If I used methods [like] ANOVA,²⁷ and other stuff like that, I can't see anything with averaged [values]. ... [A]nother thing I used was a latent class analysis to really find latent classes which were otherwise mixed together, [but] then by [calculating] average we find nothing. That's why we never find anything in pre-clinical [patients], because we mix up too many things. As a result [with *Leaspy*], I've been able to get clear results that really relate a story and address different aspects of the question. It's not necessary, but if you want to do a qualitative work in my field of pre-clinical research, it's worth having methods that are [a little more advanced than what I already had].

Here, interviewee I5 already set herself apart from her research community by working on symptoms that have received very little attention in the scientific literature on Alzheimer's disease, but really did so by using “new and innovative” AI-inspired computational techniques that will ensure her future recognition by the scientific and medical communities if she continues her research after her Ph.D. graduation.

5.3.5 A future medical assistance tool?

The use of AI in medicine and its influence on the evolution of practices within the biomedical research community are still being debated by many renowned AI experts (Patel et al., 2009). The main conclusion that emerges from these rich discussions is that the algorithms

²⁷Analysis of Variance.

provided by computer scientists are designed to aid doctors in their decision-making or medical acts, but not to replace them, since the results or diagnoses produced must be checked and criticized by an expert medical authority before any practical steps are taken (Pesapane et al., 2018). AI must therefore not challenge the clinical intuition of health practitioners (Matuchansky, 2019), as interviewee I7 argues in the following excerpt:

In my opinion, the diagnosis cannot be based on these algorithms. ... In fact, the diagnosis is a mixture of probabilities, of *a priori*²⁸ (sic.), obviousness but also clinical intuition. There's something that goes beyond all that. To make a diagnosis, all [these algorithms and software are] very useful, but it must be interpreted by clinicians and transmitted to the patient, because ... the communication aspect [is] absolutely important. Sometimes, a diagnosis can be revised, refined, changed over time, ... it's only clinical observation that allows us to end at the most appropriate diagnosis. ... So I think these instruments ... will improve clinical practice for clinicians, but they can't replace it. On the other hand, clinicians need to be properly trained in how to use, read and interpret these algorithms, it's crucial. ... Now, I would like to insist that I believe that, at the current stage, algorithms for combining multimodal data, etc., are more likely to be applied to research, to the identification of *endpoints* (sic.) in therapeutic trials, really in research *outcomes* (sic.), therefore they are mainly destined to the research community. In my opinion, there is still some validation work to be done before algorithms and scores can really be translated from research to clinical practice. (I7, neurologist and PhD candidate, AP-HP, ICM)

Although AI is already a powerful tool in clinical research (see, for example, Anichini and Geffroy, 2021; Ansart et al., 2021; Pesapane et al., 2018; Valliani et al., 2019; Vieira et al., 2017), it encounters negative receptions from some clinicians who are hostile to its use in their day-to-day practice, thus hindering the translational process from basic research to experimental deployment in hospitals. This results from two reasons. On the one hand, they do not necessarily see the point of using software that simply produces predictions for them, as interviewee I5 pointed out in another exchange:

There are some very good doctors, even really well-known ones, but they are not used to the idea of using a computer to do certain things. ... For example, if we wanted to do a study that would enable us to predict a patient's condition in the next three or five years, these doctors would tell us, "I don't see the point of taking my computer, entering the data, his or her current scores, and having an estimate of the scores in five years because I already know, otherwise there may be variations, changes, events that we hadn't foreseen and that have changed my own intuition, but in fact they will also change the computer's estimate." So they were a bit reticent. (I5, neuropsychologist and PhD candidate, AP-HP, ICM)

On the other hand, clinicians may not always trust the results given by the machine, because they do not all know how the machine arrives at the result it produces. Deep

²⁸Said as such during the interviews, which can be translated as *biases* in English.

learning is a typical example, where even computer scientists who are used to manipulate are also unsure on the way to interpret the results returned by such a neural network:

It's very difficult to know the performance of deep learning approaches because the validations are often not sufficiently advanced and it's actually a bit painful. For example, going back to what I was doing in my thesis, which was to correct the attenuation of PET images taken from a PET-MRI scanner, all the research groups that had scanners at that time started to develop [numerical] methods, because [those] supplied by the manufacturer [were] useless. So we got together a dozen centers and conducted a very thorough evaluation of the different methods that each center had developed. ... There was a deep learning method ... that came out at the time, dating from 2017 I think. ... And in fact, the result was that the deep learning method worked as well as the others, it didn't work ten times better, and it had *outliers* (sic.) that were very difficult to predict, so we didn't really understand why it didn't work. Whereas, with the other methods, we could see why they didn't work, and we were able to give an explanation like: "Well, this patient's anatomy is very different, or the density of his or her bones is really very different, so it's normal that the method doesn't work." With the deep learning method, we didn't really know how to explain it. (I8, full-time researcher, CNRS, ICM)

Deep learning and machine learning algorithms are often viewed by both computer scientists and clinicians as "black boxes", subject to the inherent randomness of the probabilistic methods implemented in them - two successive executions of an algorithm do not yield substantially the same result. In response to clinicians' demands for transparency in algorithms and an understanding of their functioning, the Aramis team is committed to improving the *interpretability* of the tools it develops (Burrell, 2016; Gilpin et al., 2019). The objective is to determine, through a series of validation stages, the precise manner in which the algorithm returns a given result. The outcome of this research could, in particular, build clinicians' confidence in the algorithm, which they would use later in clinical research or in their medical practice. As indicated above by the interviewee I7, this deployment must be accompanied by appropriate training in the use and explanation of the approach of the algorithms provided, which would also help to allay clinicians' reservations.

In short, the testimonies of the three clinicians interviewed during this fieldwork converge to the same conclusion: even though AI is already well established in the research process – formal, biomedical, clinical, etc. –, there is currently no clear indication that it will lead to a global consensus among practitioners engaged in intensive clinical use.

5.4 Discussion and conclusion

The Aramis team serves as an interdisciplinary breeding ground for the dissemination of AI in clinical neuroscience research and its future integration into the medical community specializing in neurology. Based on eight interviews and the scientometric analysis provided in the two previous chapters, we have established a typology of scientific profiles, divided into two groups, in order to describe the organization of work within the team and the interactions between them. The team is primarily composed for the most part of researchers in computer science and mathematics, who are not only committed to improving software but also involved in the development of existing software. They demonstrate their creativity by imagining and designing new high-performance methods capable of processing large masses of medical data, with the objective of assisting clinicians in their clinical research and also in their daily medical practice. The latter, who constitute the second group of our typology and who are not necessarily trained in computational methods, thus set the tone for interdisciplinary collaboration with these computer scientists, guiding research in line with the requirements of the medical profession. In this regard, they serve as *ambassadors* of the medical professions within a computer science team (Li Vigni, 2021a), mediating and promoting the research output to the supra-academic and supra-medical instances.

However, this collaboration between the two aforementioned scientific profiles is proving difficult because of difficulties in understanding and even communicating between them, whose research paradigms and visions of AI and the computational work differ. Nevertheless, each requires the insights of the others to make progress in this interdisciplinary endeavor. For example, computer scientists still need the advice and knowledge of clinicians to develop software tailored to their needs. Thus, in light of the conclusions of our scientometric model of genericity drawn in the previous Chapters 3 and 4 with large-scale data analysis over the whole neuroscience, the results introduced in the present chapter corroborate that AI is a research instrument brought by specialists who develop and improve it for the welcoming research environment (*genericity in conceptualization*). These AI experts also adapt to this host research context (that is not their own for some of them) by appropriating its disciplinary culture (*interdisciplinary migrants*, as proposed by Sedooka et al. (2015)). Conversely, clinicians simply appropriate the instrument just for applicative ends (*genericity in application*).

We remain cautious about this last conclusion as they reflect only one research context of this team, which topics are not directly focused on the conceptualization of AI's computational architectures but rather on clinical applications. Due to the limited number of interviews used and the narrow scope of our investigation, we inevitably miss the precise knowledge produced by Aramis, which would require both a micro-scale analy-

sis of the content of the team's publication, along with a much longer on-site fieldwork (a few years) to provide a comprehensive retrospective of the true impact of AI in the team's daily research practice. Such a fieldwork would entail not only the realization of supplementary individual interviews but also observations of the research activity within Aramis and other ICM teams, in order to precisely delineate the different stages in the development of algorithms across basic and translational research, and to bring other insights into the relationships between clinicians and computer scientists within the institute. Drawing upon the reflections of Forsythe (1993a; 1993b), Rosental (2003) and Voskuhl (2004) on the sociology of the development of AI in computer science, who have adopted such an ethnomethodological approach, these observations would also enable an analysis of conversations between members in contact with AI, both formal and informal, which are at the core of the construction of the scientific knowledge in neuroscience and the enhancement of the AI technological instruments.

In the framework of such an extended fieldwork within Aramis, it would also be worth to examine the governance structure of this team, with a specific focus on the status of its members (principal investigator, research scientist, engineer, etc.) given their affinity with AI. The aim of such an inquiry is to ascertain whether this organization has an incidence on the main orientations of the produced research and, if so, whether it contributes to shaping an affirmed professional identity around the research subject and/or the methods used, such as AI, or to reinforce well-established disciplinary boundaries (Louvel, 2015; Sedooka et al., 2015). Indeed, we could suppose that a team (co-)directed by an AI specialist would favor the importation of related knowledge, methods, and practices within the neuroscientific field, thereby engendering a complete epistemic culture. A detailed investigation of the historical construction of the team would then be required.

Thus far, the interviews analyzed in the present chapter support a common narrative, especially at work, which suggests that AI is juxtaposing in a pre-existing well-furnished set of technical tools used in neuroimaging and neuropsychology, and enabling some tasks that would not be possible to carry out without traditional statistical tools (although not necessarily outperforming them). This is notably the case for disease course mapping that has been widely developed throughout these interviews, which aims to forecast the onset of neurodegenerative diseases. Moreover, the growing use of such AI tools is accompanied by a hype around it, as well summarized by one of our interviewees who has stated: "AI is sexy". The corresponding excerpt especially reflects that AI has become a kind of label, engaging multiple intertwined realms beyond the mere academic work, which together build a specific rhetoric of promises of applications in the near future (Raimbault and Joly, 2021). In such realms, AI appears to become an essential keyword, a discursive element employed to attract various audiences for different purposes. For instance, AI targets not only scientists (both computer scientists and clinicians) to maintain or gain visibility

in academia, but also health professionals and industrial companies for concrete clinical applications, and last but not least, funding organizations and journals to guarantee the sustainability of such research they deem promising. All of these actors, situated within diverse arenas of interest for AI, play a pivotal role in strengthening the *academic credibility cycle* of researchers, as first conceptualized by Latour and Woolgar (1996). The successful additional work of persuasion of researchers to these actors thus tends to reinforce the credit of the former within the scientific community in a novel context of a technoscientific model of knowledge production (Li Vigni et al., 2023), which also concerns the AI instrument. Consequently, AI has become an essential to survive in the competitive academic or industrial worlds, due to the mounting pressure to use it in the context of promises of applications. The latter thus reveals itself to be a driver of genericity of AI in science, although it has not yet fully achieved this character, according to the depicted reality of work in laboratories and the scientometric elements previously drawn at the scale of the field.

Summary of the results

As AI develops since the late 1950s, voices have raised (Ganascia, 1993; Schwartz, 1989; Vayre, 2021), and are still resonating today (Cockburn et al., 2018; Köstler and Ossewaarde, 2022; Sartori and Bocca, 2023), about its possible capacity to become pervasive in all sectors of society. Science is no exception to this trend, especially because AI can propose new ways of doing research by reframing current methodological frameworks across all disciplines in order to become data-driven (Kitchin, 2014) and by assisting creation of new knowledge that would not have been possible to produce without it. While being progressively disseminated in science, almost all research domains are nowadays showing interest in AI-related knowledge and technologies for their own epistemic goals.

However, do we observe this alleged generalization of the use of AI in all scientific domains? In this thesis, we have adopted a framework proposed by Shinn and Joerges (2002), from which we conceive AI as an instrument embedded in a *research-technology* regime since the 1950s and destined to be disseminated at all levels of scientific production. Such a final step is called *genericity*, that we have defined, in Introduction, with three criteria, namely *domain adaptation*, *social adoption* and *epistemic integration*, which refer, respectively, to the capacity of AI to be applied in various topics covered by the domain, the growth of the number of AI users within the scientific community representing the domain, and the articulation of AI with the conceptual frameworks structuring the domain. So the aforementioned question could be translated as: according to these criteria, is AI becoming generic in science?

This thesis has attempted to bring an answer to this overarching research question by focusing on the development and diffusion of AI within a single multidisciplinary field of

research, that is, neuroscience. To this end, we have drawn upon a dual methodology in order to capture at various scales the three aforementioned criteria describing our genericity model applied to AI. First, we have traced the long-term impact of AI in neuroscience at the macro-scale over 50 years with a scientometric, quantitative approach (Chapters 3 and 4). Second, we have examined the social construction of AI at the micro-scale within a laboratory through semi-structured interviews with researchers, conducted as part of a fieldwork run in 2021 (Chapter 5).

With the first quantitative approach, we have represented neuroscience with its scientific literature published since the 1970s, available and extracted from a large bibliometric database, the Microsoft Academic Knowledge Graph (MAG). In Chapter 2 of this manuscript, we have especially introduced and discussed an extensive cleaning step of this extracted neuroscience corpus, from which we have retained only articles published in peer-reviewed journals and delineated an AI-related subcorpus with a keyword filter applied on titles and abstracts within this corpus. Then, we have detailed the construction of two temporal networked structures from some of the metadata attached to the articles of the corpus, namely an ego-centered citation network including references and citations related to all of the latter, and the co-signature network from their authors. The first network, capturing the evolution of the disciplinary environment of neuroscience over time, allows to assess the disciplinary embedding of AI-related publications into the neuroscientific ecosystem, while the second one enables us to evaluate the degree of adoption of AI among the neuroscientists and to locate AI practitioners within the whole social network of the community – in its core or its periphery.

These two networks, accompanied with numerical indicators describing them, have been mobilized for the purposes exposed in Chapter 3, centered around the two first aforementioned criteria of genericity of AI spreading across neuroscience. We have shown that AI fits in with the epistemic objectives of neuroscience by sharing its multidisciplinary citation environment. AI thus demonstrates a successful disciplinary adaptation at first glance, but it encounters a limited social adoption, AI practitioners even being gradually segregated in the collaboration structure of the neuroscience community, and publishing in a small set of dedicated scientific journals. In addition, most of these AI experts, especially those entering neuroscience since the late 1990s, have a common disciplinary profile (according to their published papers), that remains anchored into STEM-related fields, including computer science. According to Shinn and Joerges (2002), they are the *AI research-technologists* attracted by neuroscience.

In Chapter 4, we have then introduced other data structures based on different metadata, in order to study more precisely the ensemble of neuroscience knowledge and the capability of AI to articulate within its conceptual base. To this end, we have first

mapped the textual elements (titles and abstracts) of the articles of our corpus on a two-dimensional space representing the state of knowledge of the field under study, based on the combination of document lexical embedding and dimension reduction techniques, respectively, SPECTER (Cohan et al., 2020) and UMAP (McInnes, Healy, and Melville, 2020). We especially have partitioned this knowledge space with a density-based hierarchical clustering, HDBSCAN (McInnes, Healy, and Astels, 2017). Then we have built a conceptual network linking *fields of study* attached to the papers, which represent broad topics in which they can be inserted – according to the topical classification of the MAG database. Such a network allows to assess whether AI-specific concepts are linked, or achieve to link, to major neuroscientific concepts over time, and whether they are central within this conceptual network. We have studied such a network per knowledge area of the neuroscience map given by the aforementioned clustering algorithm.

Still in Chapter 4, we have shown that, although AI is present in almost all the specialties of neuroscience, testifying again a large adaptation of AI to various purposes of the domain, it faces a limited incorporation into their respective conceptual network – even withdrawing from them over time –, except for a neuroscientific subfield dedicated to the mathematical and computational foundations of contemporary connectionist AIs. This suggests that AI is generic *in application*, as it is able to be operationalized in various research contexts exhibiting their own knowledge base and vocabulary, but not *in conceptualization*, as it moves away from the most prominent concepts mobilized in these research contexts over time. This finding, which corroborates the argument of the gradual distancing of AI research topics and *big data* from fundamental research, supported by many authors (Ahmed et al., 2023; Frank, Wang, et al., 2019; Frégnac, 2017; Klinger et al., 2022; “The new NeuroAI” 2024), also demonstrates that the epistemic integration of AI does not seem fully achieved. It also illustrates the conclusions of Cheng et al. (2023) who studied the spread of diverse novelties in different domains at different times, such that the failure of an idea to link with a preexisting set of prominent concepts in a given domain hinders its wider diffusion (of social and epistemic kinds) across the latter. This is precisely the case of AI within neuroscience. Moreover, the discoveries made with AI, as well as the discussions it may arouse, are only relevant in the knowledge areas of neuroscience in which they were produced, thus testifying a limited transferability of the AI-related tools and knowledge across the domain.

According to all of these first results, the AI instrument does not appear to be totally generic in neuroscience, which is rather surprising since the latter constitutes one of the main bases of the recent connectionist “paradigm” of AI. Overall, it does not yet seem to supplant all the methodological frameworks within the domain under study. While AI seems to be spreading within science at large scale (Gao and Wang, 2023; Gargiulo, Fontaine, et al., 2023), which would be a signal of successful diffusion, this quantitative

analysis has shown that it does not become a must when observing the spread of its use in all the components of a given field of research at a lower scale, according to the criteria set in our genericity model.

Finally, in the last Chapter 5, we have complemented these aforementioned quantitative results by reducing our scale of analysis to the micro level, using a more qualitative approach. Specifically, the second research material mentioned above was mobilized, namely interviews with researchers from a research team called Aramis, based at the Paris Brain Institute (ICM) and specializing in the study of neurodegenerative diseases at the intersection of computational neuroscience, neuroimaging, and neuropsychology. This fieldwork has mainly focused on the interdisciplinary aspect of AI work in the context of neuroscience research, particularly through a categorization of scientists into two groups directly inspired by the two profiles widely depicted in the previous scientometric analysis: computer scientists and clinicians. With this fieldwork, this chapter has tried to provide a more nuanced understanding of the genericity criteria, especially domain adaptation and social adoption.

First of all, it should be emphasized that the research conducted in Aramis is at the interface between basic and clinical contexts, with a particular focus on further clinical applications in regular medical practice. Computer scientists thus are a workforce responsible for tasks that clinicians are unable to perform independently, in particular the processing of large amounts of clinical data, where AI appears to be a valuable aid. As the output of the research is primarily intended for an audience that is more composed of clinicians rather than computer scientists, engineers, or technicians, the latter must therefore comply with the requirements not only of medical disciplines but also of health practitioners. In this context, our interviews have suggested that the practice of interdisciplinary collaboration within Aramis is facing challenges, with interactional difficulties between members of the two groups due to their respective different paradigms and approaches to analyzing medical data. Furthermore, they revealed that the term “AI” is more a catch-all designation rather than a well-defined one, which contributes to accentuate the aforementioned point about clinicians’ limited understanding of the work done by computer scientists. During the course of the investigation, AI has also emerged as a kind of brand image, serving as a central mechanism for maintaining academic recognition for both computer scientists and clinicians (Li Vigni et al., 2023).

In summary, this brief investigation has shown that AI is capable of adapting to the main challenges addressed by the team under study, and even reaching early-career clinicians who are engaged in a doctoral programme and have chosen to use it for their respective research. However, our fieldwork still needs to be extended to laboratories with different subjects and approaches (not necessarily data-driven), and where the sci-

entific work differs significantly from this ideal, biased case of Aramis in which AI plays a central role.

Limitations, perspectives of improvement and future work

Discussion of the chosen analytical framework to study the diffusion of AI within neuroscience

As discussed in Chapter 1, AI is defined in multiple ways in the scientific literature, especially in social studies of science. In particular, we have noticed that AI, although not often mentioned clearly, could be associated with either a research instrument, composed of various technical tools semantically related to this notion, or a scientific domain as a whole. Therefore, the research-technology framework and the specific concept of genericity (Shinn and Joerges, 2002) have been chosen to overcome this distinction in order to conceptualize the diffusion of knowledge, research instruments, and practices related to AI, which have been designed and developed in a dedicated research domain. In addition, drawing upon the findings presented in our article characterizing the evolution of AI research in science (Gargiulo, Fontaine, et al., 2023), we have assumed that AI is mainly produced within a research environment dominated by STEM-related domains, especially computer science, mathematics and statistics, before being disseminated throughout others exogenous from the former. This conceptualization thus supposes that the domain under study receiving AI, here neuroscience, is situated outside its development since its inception. Although the literature on the shared history of AI and neuroscience demonstrates the contrary (Cardon et al., 2018; Hassabis et al., 2017; Macpherson et al., 2021), we have maintained this hypothesis with regard to the limited production of AI-related knowledge (see Chapter 2, Fig. 2.8), which suggests that AI has been and continues to be peripheral to the epistemic preoccupations of the neuroscience domain, at least since 1970.

Nevertheless, some concepts that have contributed to many recent achievements in the field of AI have been directly borrowed from pure neuroscientific ones, such as the well-known neural networks. Consequently, we could question the direction of knowledge transfer between AI and neuroscience when the former emerged: Do we attend the diffusion of neuroscience concepts to STEM ones, which have appropriated them afterwards under the calling of AI, or do we attend something else? Therefore, coming back to our main research questions exposed in this thesis, does AI effectively adapt to neuroscience,

or is neuroscience adapting to AI? This finally challenges the validity of the established *domain adaptation* and *epistemic integration* criteria established for this thesis, which both regard AI as integrating neuroscience but not the other way around. A more refined definition of the genericity criteria would thus be valuable in better characterizing the diffusion phenomenon under study.

In the context of the hypothesis (adopted here) that AI represents a technological innovation intended for dissemination in neuroscience, which is here supposed to be not implicated in its development, a future work should invest alternative analytical perspectives to address this diffusion, such as a complex contagion model (Cencetti et al., 2023; Centola, 2015; Manzo et al., 2018), as previously emphasized in the discussion of Chapter 3. Indeed, based on the preliminary identification of some adoption factors of interest to combine, such as the correlation of close collaborations and citation patterns (Espinosa-Rada et al., 2024; Wallace et al., 2012) or the potential methodological, epistemic, cultural proximity of the neuroscientists who adopt AI (see, for example, the work of Roth and Cointet (2010)), such a model would provide precise *explanations* – in the sense of analytical sociology (Manzo, 2010) – for some microscopic mechanisms of the diffusion phenomenon that has been interpreted according to the theoretical framework of research-technology throughout this thesis. This approach would have supported the social diffusion criterion of our genericity model. Moreover, such a contagion mechanism would facilitate the isolation of the profiles of innovators, spreaders, and early or late adopters at different stages of the social diffusion of AI (for some classic examples, see Brancheau and Wetherbe, 1990; Coleman et al., 1957; Robertson, 1967; Rogers, 1983). In this line of inquiry, we could use our data to calibrate the aforementioned model in order to ascertain at which threshold we consider genericity to be achieved.

Refining the neuroscience corpus

As already exposed and discussed in the previous chapters, the results exhibited in this thesis face diverse limitations, that are mainly due to the bibliometric data used throughout this thesis to evaluate the aforementioned genericity criteria of AI inside neuroscience.

An important one, and the most difficult to apprehend in scientometric analysis in our opinion, lies on the plurality of disciplinary or topical classifications of scientific publications available on the web, whose constructions and usage are quite different. In particular, in Chapter 3 we have mobilized the Web of Science (WOS) taxonomy built from a set of referenced journals, which we used to set the disciplinary membership of the papers that cite or are cited by our neuroscientific corpus, without much concern about the finer research subjects attached to the articles. Conversely, in Chapter 4 we have partly relied on

an old, not continued *fields of study* classification provided by MAG, organized in a network that is difficult to understand. Furthermore, the publications in MAG were labeled with a variable number of such fields situated in different hierarchical levels of granularity within the network (a discipline, a specialty, a very small research area, etc.), sometimes with inconsistent associations of very distant fields producing very different knowledge. This thesis has thus articulated different coexisting bibliometric classifications of papers, which deserve to be subsequently unified.

In this work, we have begun to use the MAG bibliometric database before its successor service, OpenAlex, has been launched, which provides improved metadata describing scientific publications. Consequently, the aforementioned classification of fields of study has now been completely replaced by *topics* in OpenAlex, that are organized in a more consistent way on a hierarchical tree from small research topics to large *domains*, through two intermediary classes,¹ namely *subfields* and *fields*. In this database, the publications are only labeled with three topics at most according to a textual-based classifier, one of them being set as a *primary* one, thus limiting inconsistencies on the identification of the research area related to these publications. Although we have exploited this topical taxonomy in App. C.1.1 in order to verify some results provided with the ancient MAG's one and introduced in Chapter 4, we should reconsider some analysis exposed throughout this thesis with this novel classification, especially in Chapters 3 and 4, wherein it can supplant the used WOS one. On the one hand, such a classification would indeed be worth to build more precise intellectual trajectories of the authors, out of the major disciplinary boundaries set by WOS, as done in (F. Liu et al., 2024; Singh et al., 2024; Zeng et al., 2019). On the other hand, it would help to refine the disciplinary ecosystem drawn in this chapter and thus revise the disciplinary adaptation criterion mentioned above.

Moreover, moving back to the methodology used to build the neuroscientific corpus exposed in Chapter 2, we could replace the journal classification of WOS, which reflects a specific disciplinary structure of the research domain we want to study, by the OpenAlex's one in order to define a completely new dataset. As this taxonomy includes a whole field *Neuroscience* divided in several subfields and topics, it would also be useful to refine the cartography of Chapter 4 and its clustering. We could also pursue this rationale by identifying the AI-related papers in the corpus as those exhibiting at least one AI-related topic included in the subfield *Artificial Intelligence* of this classification – combined with topics included in other fields or subfields such as *Computer Vision and Pattern Recognition* and *Natural Language Processing*, that might not be included in the AI subfield. Overall, such a refinement work would be valuable for the sake of consistency, using a single, stable topical classification, especially with that provided by OpenAlex.

¹Notice that these three classes are inherited from the database *Scopus*.

In order to complete our corpus, we should also consider the diversity of publication types, which could differ from one discipline to another, and the manner these different publications impact specific disciplinary communities as well. Indeed, we have shown that most of AI research published in peer-reviewed journals in neuroscience is impacting mainly neuroscience itself, but we could test whether this pattern still remains in other media for communicating research results, such as conference proceedings, preprints published on online platforms such as *arXiv*, commonly used by mathematicians, physicists and computer scientists (Wainer et al., 2013), without forgetting the sharing of various codes and open-source software in dedicated online repositories, such as *GitHub*, that are crucial to the transmission and the development of an AI-centered technological, coding culture within the domain under study. Future research may clarify whether our results hold if we consider the impact of AI produced within neuroscience through these other publication outlets.

Finally, since our analysis lies on the state of AI before mid-2020, and thus before the rise of generative AI since 2022, through notably the launch of Large Language Models (LLM) such as *Chat-GPT* and *Gemini*, this thesis deserves to be extended to a more up-to-date dataset, as the recent prowesses in AI can reconfigure the relationships between AI and the scientists, and therefore the citation, collaboration, conceptual networks used to study the genericity of AI within neuroscience. In addition, as deep neural networks are reaching their peak since the early 2010s, we can suppose that the small number of AI-related works observed in neuroscience over time is due to a (very) long lag time before adoption, which could be happening nowadays, but is still invisible in recent bibliometric data. We thus need to wait a bit before tackling again this broad question of diffusion of AI in science.

Studying the capability of AI to create knowledge

Our genericity criterion of *epistemic integration*, understood as a conceptual articulation of AI knowledge with neuroscientific ones throughout this thesis, can be extended to its capability to generate novelties inside the knowledge foundations of neuroscience. As suggested by Cheng et al. (2023), a novel scientific discovery that is able to link with a preexisting set of knowledge of a research domain is able to spread more in the latter. This mechanism would therefore allow for the reinforcement of a virtuous circle consisting of attracting subsequent novelties within this domain, which is analogous to the phenomenon known as the expansion of *adjacent possibles* (Kauffman, 2000; Monechi et al., 2017). Although AI is not central within the conceptual field of neuroscience, it remains interesting to know whether AI is part of the creation of some knowledge used for neuroscientific purposes.

Many methods can capture such a generation of novelties from a set of concepts, notably knowledge flows through citations between successively published papers in the domain under study (Di Bona et al., 2023; Sun and Latora, 2020) and *phylomemetic trees*² (Lobbé et al., 2021), a framework enabling a simplified visualization of vocabulary inheritance relationships between sets of papers across variable temporal periods. Besides, these evolutionary representations of knowledge flows, which offer a comprehensive view of the whole scientific literature of a given domain, through the merging of some knowledge areas into a single one, the birth of others, and the semantic change of pre-existing ones, would lead us to study the potential competition between AI and other methods and technologies (Arthur, 1989) inside this domain at many moments of its development – for instance, the rise of deep learning techniques in brain imaging, that is nowadays able to outperform standard classifiers like the standard Support Vector Machine (SVM), as shown by Abrol et al. (2021). A comparison of these methods would then be required to choose the most appropriate one.

For other future projects, such methods would be valuable in analyzing the degree of AI usage within individual research teams or laboratories (as Aramis) through their respective set of academic publications since their launch.

Applying our framework to other fields of research

In conclusion, returning to the overarching research question animating this manuscript, we did not address whether AI is generic in other research domains, which are characterized with other knowledge bases and methodological frameworks related to diverse scientific practices and cultures that are not the same as neuroscience's ones, but could favor (or not) the insertion of AI inside them. This thesis thus intends to be a roadmap for further studies of the diffusion of AI in a broad range of disciplines or fields of research that are receptive to it, but probably with different patterns, such as astrophysics, which also deals with heavy, complex images and signals that need to be efficiently processed. We should also consider some domains that are, a priori, far away from AI (according to Gargiulo et al. (2023)), such as social science and humanities. Sociology seems to be a good additional fieldwork for this purpose. A comparative work would then be worthwhile, on the basis of the genericity criteria established for this work.

²Such knowledge structure can be drawn with the *GarganText* software, designed at Paris' Complex Systems Institute (France) by the authors of the aforementioned article.

A Chapter 2 appendices

A.1 AI-related keywords table

evolutionary algorithm	simulated annealing
hierarchical clustering	principal component analysis
unsupervised learning	neural networks
backpropagation	classifier
search algorithm	genetic algorithm
training data	association rule mining
data mining	speech recognition
hidden markov models	adaptive algorithm
reinforcement learning	feature extraction
gaussian mixture model	expectation maximization algorithm
linear discriminant analysis	cluster analysis
swarm intelligence	general game playing
artificial intelligence	knowledge representation
adaptive learning rate	information retrieval
random forest	robotics
decision support system	inference engine
genetic operator	singular value decomposition
kernel method	expert system
boltzmann machine	deep learning
representation learning	incremental learning
catastrophic forgetting	machine vision
natural language processing	predictive model
adaboost	face recognition
feature selection	latent semantic analysis
image classification	pattern recognition
particle swarm optimization	face detection
backward chaining	forward chaining
affinity propagation clustering	perceptrons
computational intelligence	cobweb
dimensionality reduction	combinatorial optimization
prolog	latent semantic indexing
intelligent agent	autonomous robot

Continued on next page

computer vision	markov decision process
echo state network	agent architecture
tf-idf	answer set programming
ambient intelligence	gram matrix
binary tree	machine intelligence
generalized additive model	naive bayes
temporal difference learning	radial basis function network
supervised learning	decision theory
artificial immune system	statistical classification
semantic network	overfitting
hidden layer	partially observable markov decision process
markov chain monte carlo	dialogue system
hopfield network	beam search
gibbs sampling	dirichlet process
hierarchical dirichlet process	stochastic gradient descent
topic modeling	co-clustering
multinomial mixture model	indian buffet process
declarative programming	knowledge engineering
logic programming	conditional random fields
naive bayes classifier	turing test
evolutionary computation	constraint logic programming
distributed representation	codebook
optical character recognition	graphic processing unit
augmented reality	policy gradient
data augmentation	convex hull
autonomic computing	object detection
adaptive neuro fuzzy inference system	maximum a posteriori estimation
decision tree learning	maximum likelihood estimation
stochastic optimization	distance metric learning
minimax	imagenet
distributed artificial intelligence	automated reasoning
thompson sampling	spiking neural network
independent and identically distributed	abstract data type
turing machine	automata theory
object detection algorithm	affective computing
sequential pattern mining	backpropagation through time
bidirectional recurrent neural network	feature learning
alphago	triplet loss function
triplet loss	connectionist temporal classification
word2vec	chatbot
autoencoder	denoising autoencoder
clustering stability	adversarial autoencoder
expectation-maximization algorithm	deep reinforcement learning
monte carlo tree search	halting problem
recurrent neural network language model	superintelligence
maxout	artificial creativity
sparse autoencoder	lenet
cognitive computing	rete algorithm
vq-vae	variational autoencoder
weak supervision	resnet
alexnet	personally identifiable information
capsnet	anchor box
blackboard system	rectified linear unit
bees algorithm	fuzzy control system
no free lunch theorem	similarity learning
artificial general intelligence	machine perception
abductive reasoning	ontology generation

Continued on next page

global average pooling	polysemy
hyperparameter tuning	conditional gan
convolutional filter	natural language generation
textual entailment	neural style transfer
frame language	graph neural network
1x1 convolution	reservoir computing
pooling layer	datalog
leaky relu	exponential linear unit
ontology learning	batch normalization
sobel filter	googlenet
face verification	stochastic convex hull
stacked autoencoder	elman networks
mycin	yolo
coreference resolution	siamese neural network
tensorflow	transhumanism
automated speech recognition	knowledge reasoning
market basket analysis	autonomous car
mechatronics	bidirectional lstm
seq2seq	negative log likelihood
cybernetics	entity extraction
dying relu	fasttext
active learning strategy	sparql
covariate shift	convolutional kernel
textrank	artificial intelligence markup language
abductive logic programming	automated planning
recursive neural network	semantic query
statistical relational learning	ai planning
approximate string matching	hierarchical learning
dynamic epistemic logic	embodied agent
embodied cognitive science	ontology acquisition
neural turing machine	contextual bandit
belief-desire-intention model	lazy learning
deep structured learning	meta learning
cloud robotics	hierarchical softmax
mode collapse	admissible heuristic
openai	minimum description length principle
mean reciprocal rank	junction tree algorithm
automated scheduling	neuromorphic engineering
computational learning theory	bilingual evaluation understudy
deep convolutional generative adversarial network	learning rate decay
doc2vec	fast r-cnn
committee machine	gap statistic
forward reasoning	entity identification
weak ai	developmental robotics
contractive autoencoder	semantic hashing
paragraph vector	zero padding
vggnet	computational creativity
action language	probabilistic latent semantic indexing
bag-of-n-grams	collaborative topic regression
symbolic artificial intelligence	glowworm swarm optimization
creative computation	adam optimizer
bootstrap aggregating	federated learning
chatterbot	tree-lstm
vanishing gradients	yolov2
entity chunking	affinity analysis
peephole connection	multi-modal learning
intelligence amplification	ontology extraction

Continued on next page

rmsprop	capsule neural network
behavior tree	eager learning
logic tree	ensemble averaging
open mind common sense	meteor
hadamard product	machine listening
trust region policy optimization	constructed language
opencog	qualification problem
nouvelle ai	region connection calculus
sequence to sequence learning	artificial emotional intelligence
logit function	synset
propositional satisfiability problem	natural language programming
multilayer lstm	action selection problem
knowledge interchange format	creative computing
attributional calculus	antonym
constituency tree-lstm	internal covariate shift
boolean satisfiability problem	minibatch gradient descent
adadelta	stochastic gradient variational bayes
stochastic hopfield network	name binding
information processing language	mini-batching
garbage in, garbage out	theory of choice
bayesian probabilistic matrix factorization	document classification model
additive clustering	vision processing unit
convolutional stride	language segmentation
tree lstm	cognitive augmentation
action model learning	decentralized artificial intelligence
emotion ai	max-margin loss
artificial conversational entity	ai accelerator
naive semantics	semantic reasoner
computational humor	bayesian programming
yolo9000	wasserstein loss
issue tree	conditional markov model
stanford research institute problem solver	computational cybernetics
driveless car	true quantified boolean formula
multidimensional recurrent neural network	learning rate annealing
nchw	conlang
computer audition	smartbot
magnet loss	behavior informatics
meronym	pca color augmentation
tensor network theory	constrained conditional model
darkforest	convolutional padding
synthetic intelligence	deepmind technologies
he initialization	error-correcting tournaments
child-sum tree-lstm	ebert test
friendly artificial intelligence	syntaxnet
glove embeddings	fuzzy string searching
ldade	deductive classifier
adversarial variational bayes	dynamic k-max pooling
categorical mixture model	inceptionism
differential topic modeling	reinforce policy gradient algorithm
evolving classification function	word2phrase
alternating conditional expectation algorithm	dartmouth workshop
accams	support vector machine
artificial neural network	machine learning
case-based reasoning	neuro fuzzy
recurrent neural network	restricted boltzmann machine
latent dirichlet allocation	machine translation
derivative-free optimization	knowledge based system

Continued on next page

bag of words	feed-forward networks
k-means clustering	zero-shot learning
rule-based system	human in the loop
type i error	multi-armed bandit
learning to rank	finite-state transducer
f score	r cnn
multi-task learning	skip gram
cross-entropy loss	generative adversarial network
cold start	self-driving car
vector-quantized variational autoencoders	multi-agent system
learning to learn	bag-of-words model
part-of-speech tagging	out of core
mixed membership model	max pooling
sequential model-based optimization	long short-term memory networks
black box optimization	named entity recognition
chinese restaurant process	one-dimensional convolution
spatial temporal reasoning	parametric clustering
search-based software engineering	ai complete
continuous bag of words	error-driven learning
pseudo labeling	extractive sentence summarization
time-delayed neural network	moore-penrose pseudoinverse
top-5 error rate	k-fold cross-validation
hessian-free optimization	passive-aggressive algorithm
kl one	dirichlet-multinomial distribution
non-max suppression	multi-swarm optimization
fast-and-frugal trees	chess-playing computer
bias-variance tradeoff	convolutional neural network
support vector machines	support-vector machines
smooth support vector machine	artificial neural networks
online machine learning	convolutional neural networks
case based reasoning	neuro-fuzzy
recurrent neural networks	restricted boltzmann machines
hierarchical latent dirichlet allocation	neural machine translation
derivative free optimization	knowledge-based system
bag-of-words	feed forward networks
k means clustering	zero shot learning
rule based system	human-in-the-loop
type ii error	multi armed bandit
learning-to-rank	finite state transducer
weighted finite-state transducer	weighted finite state transducer
f-score	r-cnn
multi task learning	skip-gram
cross entropy loss	generative adversarial networks
cold-start	self driving car
vector quantized variational autoencoders	multi agent system
learning-to-learn	bag of words model
part of speech tagging	out-of-core
mixed-membership model	k-max pooling
sequential model based optimization	long short term memory networks
black-box optimization	named-entity recognition
named entity recognition in query	nested chinese restaurant process
one dimensional convolution	spatial-temporal reasoning
nonparametric clustering	search based software engineering
ai-complete	continuous-bag-of-words
error driven learning	pseudo-labeling
abstractive sentence summarization	time delayed neural network
moore penrose pseudoinverse	top 5 error rate

Continued on next page

top-1 error rate	k fold cross validation
hessian free optimization	passive aggressive algorithm
kl-one	dirichlet multinomial distribution
non max suppression	multi swarm optimization
fast and frugal trees	chess playing computer
bias variance tradeoff	convolutional neural network (cnn)

A.2 List of journals retained to build the neuroscience corpus

Acs Chemical Neuroscience	Acta Neurobiologiae Experimentalis
Acta Neurologica Scandinavica	Acta Neuropathologica
Acta Neuropathologica Communications	Acta Neuropsychiatrica
Adaptive Behavior	Advances In Child Development And Behavior
Advances In Cognitive Psychology	Advances In The Study Of Behavior
Ageing Research Reviews	Alcohol
Alzheimer's & Dementia	Alzheimer's Research & Therapy
American Journal Of Alzheimer's Disease And Other Dementias	American Journal Of Medical Genetics Part B-Neuropsychiatric Genetics
Anatomy & Cell Biology	Annals Of Clinical And Translational Neurology
Annals Of Neurology	Annals Of Neurosciences
Annals Of The New York Academy Of Sciences	Annual Review Of Neuroscience
Aphasiology	Archives Of Neuroscience
Arquivos De Neuro-Psiquiatria	Asn Neuro
Audiology And Neuro-Otology	Autism Research
Autonomic Neuroscience-Basic & Clinical	Basic And Clinical Neuroscience
Behavioral And Brain Functions	Behavioral And Brain Sciences
Behavioral Neuroscience	Behavioral Sleep Medicine
Behaviour	Behavioural Brain Research
Behavioural Neurology	Behavioural Processes
Biological Psychiatry	Biological Psychiatry-Cognitive Neuroscience And Neuroimaging
Biological Psychology	Biomedical Reports
Biopsychosocial Medicine	Bipolar Disorders
Bmc Neuroscience	Brain & Development
Brain And Behavior	Brain And Cognition
Brain And Language	Brain Behavior And Evolution
Brain Behavior And Immunity	Brain Connectivity
Brain Imaging And Behavior	Brain Impairment
Brain Injury	Brain Pathology
Brain Research	Brain Research Bulletin
Brain Sciences	Brain Stimulation
Brain Structure & Function	Brain Topography
Brazilian Journal Of Medical And Biological Research	British Journal Of Developmental Psychology
British Journal Of Ophthalmology	Bulletin Of Mathematical Biology
Canadian Journal Of Neurological Sciences	Cell Adhesion & Migration
Cell Death & Disease	Cell Death Discovery
Cellular And Molecular Life Sciences	Cellular And Molecular Neurobiology
Cerebellum	Cerebral Cortex
Cerebrovascular Diseases	Cerebrovascular Diseases Extra
Chemical Senses	Chemosensory Perception
Clinical Autonomic Research	Clinical Eeg And Neuroscience
Clinical Neuropathology	Clinical Neurophysiology
Clinical Neurophysiology Practice	Clinical Psychopharmacology And Neuroscience
Cns & Neurological Disorders-Drug Targets	Cognition
Cognitive Affective & Behavioral Neuroscience	Cognitive And Behavioral Neurology
Cognitive Computation	Cognitive Neurodynamics
Cognitive Neuropsychiatry	Cognitive Neuropsychology
Cognitive Neuroscience	Cognitive Processing
Cognitive Science	Cognitive Systems Research
Computational Intelligence And Neuroscience	Constructivist Foundations
Cortex	Cts-Clinical And Translational Science
Current Alzheimer Research	Current Biology
Current Developmental Disorders Reports	Current Eye Research
Current Neurology And Neuroscience Reports	Current Neuropharmacology
Current Neurovascular Research	Current Opinion In Behavioral Sciences
Current Opinion In Neurobiology	Current Opinion In Neurology
Dementia And Geriatric Cognitive Disorders	Dementia And Geriatric Cognitive Disorders Extra
Developmental Cognitive Neuroscience	Developmental Medicine And Child Neurology

Continued on next page

<p>Developmental Neurobiology Developmental Neuroscience Developmental Science Disease Models & Mechanisms Early Intervention In Psychiatry Embo Journal Epilepsy & Behavior Epileptic Disorders</p> <p>European Journal Of Neurology European Neurology Evolutionary Intelligence Experimental Brain Research Experimental Neurology Eye Focus On Autism And Other Developmental Disabilities Frontiers In Aging Neuroscience Frontiers In Cellular Neuroscience Frontiers In Human Neuroscience Frontiers In Molecular Neuroscience Frontiers In Neuroanatomy Frontiers In Neuroinformatics Frontiers In Neuroscience Frontiers In Systems Neuroscience Future Neurology Glia</p> <p>Headache Hearing Research Hormones And Behavior Human Factors</p> <p>I-Perception Ieee Transactions On Neural Systems And Rehabilitation Engineering International Journal Of Behavioral Development International Journal Of Developmental Neuroscience International Journal Of Neuroscience International Journal Of Stroke International Review Of Neurobiology Iranian Journal Of Child Neurology Jaro-Journal Of The Association For Research In Otolaryngology Journal Of Applied Biomedicine Journal Of Cataract And Refractive Surgery Journal Of Chemical Neuroanatomy Journal Of Clinical Neurophysiology Journal Of Clinical Sleep Medicine Journal Of Communication Disorders</p> <p>Journal Of Computational Neuroscience Journal Of Ect Journal Of Experimental Psychology-Human Perception And Performance Journal Of Fluency Disorders Journal Of Individual Differences Journal Of Intellectual Disability Research Journal Of Korean Neurosurgical Society Journal Of Molecular Neuroscience Journal Of Motor Learning And Development Journal Of Neuro-Oncology Journal Of Neurodevelopmental Disorders Journal Of Neurogenetics Journal Of Neuroimmunology Journal Of Neurolinguistics Journal Of Neuromuscular Diseases</p> <p>Journal Of Neurophysiology Journal Of Neuroscience Journal Of Neuroscience Nursing Journal Of Neuroscience Research Journal Of Neurovirology Journal Of Parkinsons Disease</p>	<p>Developmental Neurorehabilitation Developmental Psychobiology Discover Oncology Documenta Ophthalmologica Elife Epilepsia Epilepsy Research European Archives Of Psychiatry And Clinical Neuroscience European Journal Of Neuroscience European Neuropsychopharmacology Evolutionary Psychology Experimental Eye Research Expert Review Of Neurotherapeutics Fluids And Barriers Of The Cns Fortschritte Der Neurologie Psychiatrie Frontiers In Behavioral Neuroscience Frontiers In Computational Neuroscience Frontiers In Integrative Neuroscience Frontiers In Neural Circuits Frontiers In Neuroendocrinology Frontiers In Neurology Frontiers In Synaptic Neuroscience Functional Neurology Genes Brain And Behavior Graefes Archive For Clinical And Experimental Ophthalmology Health Psychology And Behavioral Medicine Hippocampus Human Brain Mapping Human Psychopharmacology-Clinical And Experimental Ideggyogyaszati Szemle-Clinical Neuroscience Integrative Psychological And Behavioral Science</p> <p>International Journal Of Bipolar Disorders International Journal Of Endocrinology International Journal Of Psychophysiology International Neurology Journal Investigative Ophthalmology & Visual Science Iranian Journal Of Psychiatry And Behavioral Sciences Journal Of Alzheimer's Disease</p> <p>Journal Of Brachial Plexus And Peripheral Nerve Injury Journal Of Cerebral Blood Flow And Metabolism Journal Of Clinical And Experimental Neuropsychology Journal Of Clinical Neuroscience Journal Of Cognitive Neuroscience Journal Of Comparative Physiology A-Neuroethology Sensory Neural And Behavioral Physiology Journal Of Contextual Behavioral Science Journal Of Electromyography And Kinesiology Journal Of Eye Movement Research</p> <p>Journal Of Huntingtons Disease Journal Of Integrative Neuroscience Journal Of Intelligence Journal Of Mathematical Neuroscience Journal Of Motor Behavior Journal Of Neural Transmission Journal Of Neurochemistry Journal Of Neuroendocrinology Journal Of Neuroimmune Pharmacology Journal Of Neuroinflammation Journal Of Neurology Journal Of Neuropathology And Experimental Neurology Journal Of Neuropsychology Journal Of Neuroscience Methods Journal Of Neuroscience Psychology And Economics Journal Of Neurosciences In Rural Practice Journal Of Pain Journal Of Pediatric Neurosciences</p>
---	--

Continued on next page

Journal Of Psychiatric Research	Journal Of Psychiatry & Neuroscience
Journal Of Psychophysiology	Journal Of Research On Adolescence
Journal Of Sensory Studies	Journal Of Sexual Aggression
Journal Of Sleep Research	Journal Of The Experimental Analysis Of Behavior
Journal Of The History Of The Neurosciences	Journal Of The International Neuropsychological Society
Journal Of The Neurological Sciences	Journal Of The Peripheral Nervous System
Journal Of Vision	Jove-Journal Of Visualized Experiments
Language Cognition And Neuroscience	Learning & Behavior
Learning & Memory	Learning Disability Quarterly
Luts-Lower Urinary Tract Symptoms	Medical Gas Research
Mental Lexicon	Metabolic Brain Disease
Mind Brain And Education	Mind Culture And Activity
Molecular And Cellular Neuroscience	Molecular Autism
Molecular Brain	Molecular Neurobiology
Molecular Neurodegeneration	Molecular Pain
Molecular Psychiatry	Movement Disorders
Movement Disorders Clinical Practice	Multiple Sclerosis And Related Disorders
Multiple Sclerosis Journal	Multisensory Research
Muscle & Nerve	Nature And Science Of Sleep
Nature Human Behaviour	Nature Neuroscience
Nature Reviews Neurology	Nature Reviews Neuroscience
Nervenarzt	Network-Computation In Neural Systems
Neural Computation	Neural Development
Neural Network World	Neural Networks
Neural Plasticity	Neural Processing Letters
Neural Regeneration Research	Neuro-Oncology Practice
Neurobiology Of Aging	Neurobiology Of Disease
Neurobiology Of Learning And Memory	Neurobiology Of Stress
Neurochemical Research	Neurochemistry International
Neurocomputing	Neurodegenerative Diseases
Neuroendocrinology	Neuroendocrinology Letters
Neuroethics	Neurogastroenterology And Motility
Neurogenetics	Neuroimage
Neuroimage-Clinical	Neuroimmunomodulation
Neuroinformatics	Neurological Research
Neurology And Clinical Neuroscience	Neurology And Therapy
Neurology Asia	Neurology India
Neurology Research International	Neurology-Neuroimmunology & Neuroinflammation
Neuromodulation	Neuromolecular Medicine
Neuromuscular Disorders	Neuron
Neuropathology And Applied Neurobiology	Neuropeptides
Neuropharmacology	Neurophotonics
Neurophysiologie Clinique-Clinical Neurophysiology	Neurophysiology
Neuropsychiatric Disease And Treatment	Neuropsychobiology
Neuropsychologia	Neurorehabilitation And Neural Repair
Neuroreport	Neuroscience
Neuroscience And Biobehavioral Reviews	Neuroscience Letters
Neuroscience Of Consciousness	Neuroscience Research
Neuroscientist	Neurotoxicity Research
Neurotoxicology	Neurotoxicology And Teratology
Nutritional Neuroscience	Open Biology
Open Life Sciences	Ophthalmic And Physiological Optics
Ophthalmic Research	Ophthalmologica
Otology & Neurotology	Pain
Pain Research & Management	Parkinsonism & Related Disorders
Parkinsons Disease	Pediatric Neurology
Peerj	Peptides
Perception	Perceptual And Motor Skills
Pharmacology Biochemistry And Behavior	Pharmacology Research & Perspectives
Phenomenology And The Cognitive Sciences	Physiology & Behavior
Plos Biology	Plos Computational Biology
Pm&R	Pragmatics & Cognition
Prion	Progress In Brain Research
Progress In Neuro-Psychopharmacology & Biological Psychiatry	Progress In Neurobiology
Progress In Neurology And Psychiatry	Progress In Retinal And Eye Research
Psychiatric Genetics	Psychiatry And Clinical Neurosciences
Psychiatry Investigation	Psychiatry Research-Neuroimaging
Psychoneuroendocrinology	Psychophysiology
Purinergic Signalling	Respiratory Physiology & Neurobiology
Restorative Neurology And Neuroscience	Review Journal Of Autism And Developmental Disorders

Continued on next page

Reviews In The Neurosciences	Revista Argentina De Ciencias Del Comportamiento
Revue Neurologique	Schizophrenia Research
Schizophrenia Research-Cognition	Seizure-European Journal Of Epilepsy
Seminars In Neurology	Sexual Medicine
Sleep And Biological Rhythms	Sleep Health
Sleep Medicine Reviews	Sleep Science
Social Cognitive And Affective Neuroscience	Social Neuroscience
South African Journal Of Communication Disorders	Spinal Cord
Spinal Cord Series And Cases	Stress-The International Journal On The Biology Of Stress
Stroke	Synapse
Therapeutic Advances In Neurological Disorders	Thyroid Research
Topics In Cognitive Science	Transactions Of The Indian Institute Of Metals
Translational Behavioral Medicine	Translational Neurodegeneration
Translational Neuroscience	Translational Psychiatry
Translational Stroke Research	Trends In Cognitive Sciences
Trends In Neuroscience And Education	Trends In Neurosciences
Vision Research	Visual Cognition
Visual Neuroscience	Wiley Interdisciplinary Reviews-Cognitive Science
World Journal Of Biological Psychiatry	Zeitschrift Fur Neuropsychologie

A.3 Web of Science categories' abbreviations table

WOS Categories	Abbreviations
Acoustics	Acoustics
Anatomy & Morphology	A&M
Anesthesiology	Anesth
Anthropology	Anthropo
Automation & Control Systems	A&CS
Behavioral Science	BS
Biophysics	BioPhys
Biochemical Research Method	BRM
Biochemistry & Molecular Biology	B&MB
Biotechnology & Applied Microbiology	Biotech
Cardiac & Cardiovascular System	Cardio
Cell Biology	CB
Chemistry	Chem
Clinical Neurology	CN
Computer Science	CS
Critical Care Medicine	CCM
Dentistry	Dentistry
Developmental Biology	DB
Ecology	Ecology
Education	Educ
Endocrinology & Metabolism	E&M
Engineering	Engineering
Entomology	Entomo
Ergonomics	Ergo
Gastroenterology & Hepatology	G&H
Genetics & Heredity	Genetics
Geriatrics & Gerontology	G&G
Hematology	Hemato
Imaging Science & Photographic Technology	IS&PT
Immunology	Immuno
Instruments & Instrumentation	Instrum
Language & Linguistics	L&L
Mathematical & Computational Biology	M&CB
Mathematics	Maths

Continued on next page

WOS Categories	Abbreviations
Mechanics	Mech
Medical Informatics	MI
Medical Laboratory Technology	MLT
Medicine	Med
Microbiology	Microbio
Microscopy	Microscopy
Multidisciplinary Science	Multi
Neuroimaging	NI
Neuroscience	Neuro
Nutrition & Dietetics	N&D
Operations Research & Management Science	OR&MS
Optics	Optics
Ophthalmology	Ophthalmology
Otorhinolaryngology	Otorhino
Pathology	Patho
Pediatrics	Ped
Peripheral Vascular Diseases	PVD
Pharmacology & Pharmacy	Pharma
Philosophy	Philo
Physics	Phys
Physiology	Physio
Plant Sciences	Plant
Psychology	Psy
Public, Environmental & Occupational Health	Public Health
Radiology	Radio
Rehabilitation	Rehab
Social Sciences	SocSci
Sport Sciences	Sport
Statistics & Probability	S&P
Substance Abuse	SA
Telecommunications	Telecom
Toxicology	Toxico
Virology	Viro
Zoology	Zoology

B Chapter 3 appendices

B.1 Citation patterns of the AI research in neuroscience

In Chapter 3, we have analyzed the disciplinary embedding of an ensemble of AI-related articles within neuroscience (corpus \mathcal{P}), specifically through the citation history of this last discipline between 1970 and 2020. The results discussed in this chapter have led us to consider this AI-related subcorpus, denoted here as $\mathcal{P} \cap AI$, as a *transverse* arena subject to an intermediate stage of disciplinary confinement in the course of its development (Hentschel, 2015; Marcovich and Shinn, 2012; Shinn and Joerges, 2002). In particular, this implies the emergence of an epistemic orientation that slightly differs from the disciplinary core of neuroscience, thus suggesting the development of a research subfield within the latter, comparable to a scientific specialty, in the sense of (Chubin, 1976; Wray, 2005). In this appendix, we test this last hypothesis in a different way than that in the core of the manuscript by observing whether or not the citations surrounding this AI subcorpus (bibliographic references and citations of its articles) are concentrated around it over time.

To achieve this, we opt for a macroscopic representation of the citation network surrounding $\mathcal{P} \cap AI$ by aggregating the papers that cite or are cited by this corpus into four groups, as illustrated in Fig. B.1:

- $\mathcal{P} \cap AI$: the corpus under study,
- $\mathcal{P} \cap \overline{AI}$: the neuroscience corpus without AI-related papers,
- $\overline{\mathcal{P}} \cap AI$: AI-related papers outside the neuroscience corpus,
- $\overline{\mathcal{P}} \cap \overline{AI}$: other papers outside the neuroscience corpus and without AI-related keywords in their textual metadata.

Inspired by the method of Mukherjee et al. (2015), we compute the propensity of the corpus $\mathcal{P} \cap AI$ to cite and be cited by each of the four groups in a given year by comparing the real number of citations to/from them and a simulated one resulting from random shufflings of the global citation network in that year – which leads to compute temporal citation z -scores for each group, for both references and citations. Such a measure allows to evaluate the most preferred group(s) cited by $\mathcal{P} \cap AI$ and those that cite the latter preferentially in a given year, with respect to the random shufflings of the citation network that tend to reproduce an ideal situation of uniform mixing of groups among the citations, regardless of the number of papers within each of these groups.

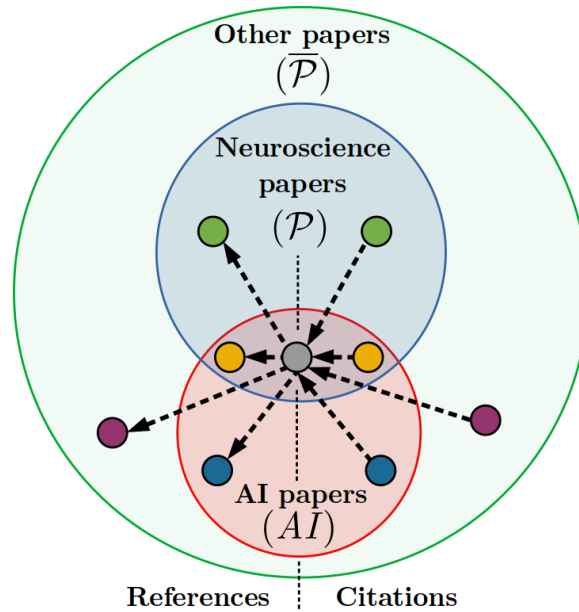


Figure B.1: Egocentric citation network around the AI subcorpus within neuroscience, at the intersection of the neuroscience corpus \mathcal{P} (blue zone) and the subset of AI-related articles (red zone). A focal article of interest is represented by a gray disk. The other colored disks represent the different groups of papers mentioned in the main text, citing the subcorpus (right disks with arrows pointing to the gray one) or being cited by it (left disks with arrows originating from the gray one).

A paper p_i in $\mathcal{P} \cap AI$ is described by a publication year $t_i \in [1970, 2019]$ and a vector of reference (citation) scores $x_i = \{n_i^k\}_{reference}$ ($y_i = \{n_i^k\}_{impact}$), where $n_i^k = c_i^k/c_i$ is the share of citations received (given) by the papers belonging to the group k in the entire set of references (generated citations) of p_i . Given $\mathcal{N}(t)$ the set of papers in the set $\mathcal{P} \cap AI$ published in year t , the average *observed* number of references (citations) to the group k at that time is given by:

$$C_{obs}^k(t) = \frac{1}{|\mathcal{N}(t)|} \sum_{i \in \mathcal{N}(t)} n_i^k. \quad (\text{B.1})$$

To obtain the corresponding score $C_{simu}^k(t)$ over multiple simulated citation networks built from the subset of papers $\mathcal{N}(t)$, we exploit the temporal distribution of references of the original citation network, i.e., the number of citations received by the references published before t over the individual papers published at t within $\mathcal{N}(t)$. To this end, we randomly shuffle the citations according to this temporal distribution, so that we conserve the temporal structure of the whole aggregated network and also the distribution of the number of citations per year. Here we consider $N = 100$ constrained shuffles of the edges of the original citation network built from $\mathcal{N}(t)$, avoiding self-citations and duplicates ones. The new score, averaged over N simulations, is thus:

$$C_{simu}^k(t) = \frac{1}{N} \sum_{i=1, \dots, N} n_i^k(t). \quad (\text{B.2})$$

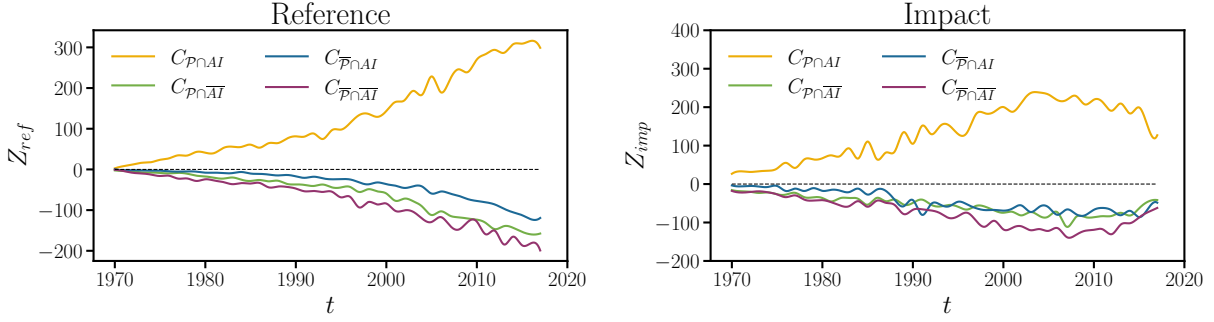


Figure B.2: Temporal z -scores for references (left) and citations (right) associated with the subset $\mathcal{P} \cap AI$.

By also computing the standard deviation $\sigma(t)$ from these simulations, we then compute the z -score z of the citations collected by the group k in year t as:

$$z^k(t) = \frac{C_{obs}^k(t) - C_{simu}^k(t)}{\sigma(t)}. \quad (\text{B.3})$$

This metrics thus assesses the under- ($z^k(t) < 0$) or over-representation ($z^k(t) > 0$) of the citations related to the group k in a year t compared to a situation obtained by chance. This method is independent of the different scientific citation practices over the whole time period between 1970 and 2019, considered here as the number of references.

Repeating afterwards all this procedure for the citations of the four groups to/from the corpus $\mathcal{P} \cap AI$, we obtain all the corresponding trends plotted in Fig. B.2.

The latter shows that both references and citations follow the same pattern. The over-presence of references from the corpus $\overline{\mathcal{P}} \cap AI$, as well as its huge expansion in this period, confirms that the AI-related articles published in neuroscience journals cite more papers also dealing with AI within these journals. Symmetrically, the three other groups are much less favored by the corpus, this situation worsening from year to year. The specific subset of neuroscience publications involving AI is thus autonomous in terms of the references used, sensibly confined within this AI subfield and with its own disciplinary composition (see Chapter 3). We also observe the same behavior for the citations, except that the internal trend of the group $\overline{\mathcal{P}} \cap AI$ reverses after 2000, while the trends of the other groups reverse after 2010, which is a signal of a new diffusion stage outside the AI-related subcorpus, where all the disciplines represented in the citation sub-network of the corpus, and not only neuroscience, together show an increasing interest in the applications of AI-related tools in their own research programs. Beyond the step of *optimization* (Hentschel, 2015), these last periods thus show a small but progressive deconfinement of the AI produced in neuroscience.

B.2 Temporal evolution of the inter-quartile collaboration network

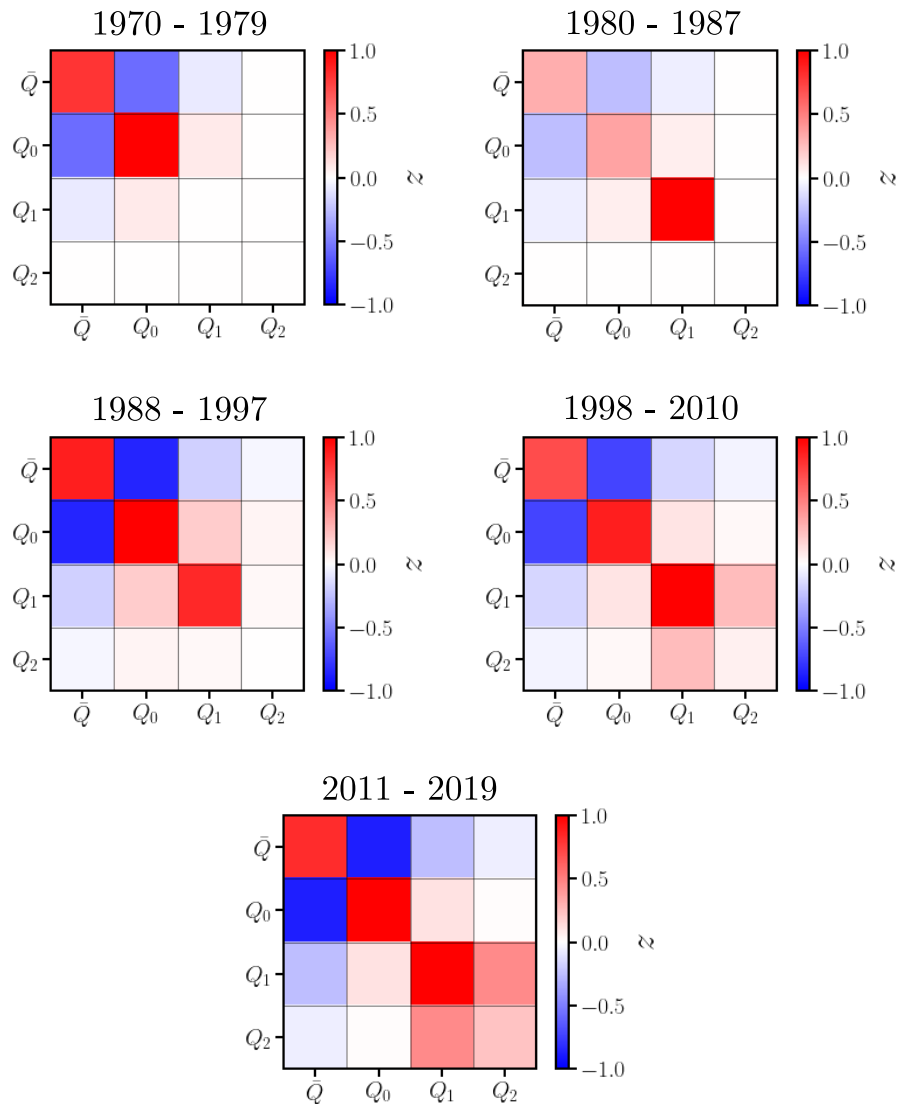


Figure B.3: Evolution of the z-score matrix of the fraction of edges between f_{AI} quartiles.

In Chapter 3 is introduced the propensity to collaborate between four different classes of neuroscientists, according to their respective AI expertise f_{AI} from the lowest to the highest, which is summarized as a z-score matrix in Fig. 3.11B of this chapter. In particular, we have mentioned in the main text of Sect. 3.3.2 that this matrix is drawn on the time-aggregated collaboration network \mathcal{A} of neuroscience between 1970 and 2019. As it might miss the particular structures of the temporal collaboration network \mathcal{T} , we decompose this matrix into 5 periods, as shown in Fig. B.3. In particular, the first line of this figure, corresponding to the period 1970-1987, refers to the situation where the quartile Q_2 has not yet emerged in our dataset. The separation of \bar{Q} from the other quartiles Q_i occurs qualitatively in the period 1988-1997, when Q_2 was emerging and beginning to publish in neuroscience. The final period spanning from 1998 to 2019 (represented by

the last two heatmaps), which concentrates 93% of the 10,785,916 weighted edges in the temporal collaboration network \mathcal{T} , is very similar to the configuration shown in Fig. 3.11 in the main text.

C Chapter 4 appendices

C.1 Technical elements for the construction of the neuroscience knowledge map

C.1.1 Robustness of the UMAP projection

In Chapter 4, with the SPECTER model (Cohan et al., 2020), we have converted the titles and abstracts of the articles in our neuroscience corpus into 768-dimensional vectors summarizing their respective lexical features and scientific knowledge. Then, we have projected these high-dimensional vectors into a latent two-dimensional Euclidean space using the UMAP method (McInnes, Healy, and Melville, 2020). More specifically, we have used the default configuration of the UMAP function of the eponymous Python package `umap`,¹ which operates the projection into a space of the same nature as that described above. As mentioned in Sect. 4.1.1 of this chapter, the projection made by UMAP is not unique. In this appendix, we test the robustness of this projection with two kinds of *nearest neighbors research*, both of which allow to verify whether the points preserve their respective neighbors from the high-dimensional space to the low-dimensional one under study in this manuscript.

The first one merely consists in running such a nearest-neighbors search for each paper of the dataset and comparing its two respective ensembles of neighbors in the two embedding spaces, hereafter simply called SPECTER and UMAP. More precisely, we perform this comparison by computing for each paper its share of common neighbors exhibited in these two spaces. With the module `NearestNeighbors` from the Python package `scikit-learn`, with a neighborhood population to query set to $k = 100$ (which is not too large to avoid covering entire regions of the map) and a calculation mode set to pairwise Euclidean distances, we find an average conservation of 10.8%. This very huge loss after projection, however, does not imply that the papers are not surrounded by other neighbors in the same area of vocabulary and encoded knowledge (or nearby). That is why, in a second time, we verify whether the articles remain in a similar thematic environment after the UMAP projection.

Inspired by the approach given by Singh et al. (2024) and González-Márquez et al. (2024) in their respective technical materials used to scientometrically reconstruct broad research landscapes, we implement a k -nearest neighbors (k NN) procedure to decipher the thematic structure of the SPECTER-embedded corpus and the UMAP-embedded one.

¹This package is available on Pypi and, when using `pip`, must be installed with the command `pip install umap-learn`.

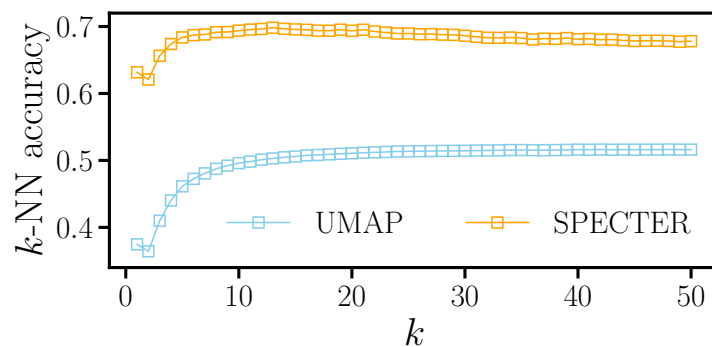


Figure C.1: Accuracy of subfield prediction given by a k -NN algorithm at several number of neighbors k for each embedding space – 768 dimensions for SPECTER and two dimensions for UMAP. The accuracy at a given k is computed as the share of points in the test set whose labeling subfield prediction is correct, i.e., the predicted subfield matches the actually observed one. For SPECTER, 99% of the dataset was used to train the algorithm at each k , and the remaining 1% served to compute the accuracy. For UMAP, we choose 80% of the dataset to train the algorithm, and 20% to test it.

For a given set of points, such a k NN classifier predicts their unique class given the classes labeling their respective neighbors, i.e., it assigns the majority class appearing in the neighborhood of the point under study. To achieve this, we choose to label the papers with their respective *subfield* associated with their *primary topic* extracted from OpenAlex (see their definitions in Sect. 2.2.1.1 of Chapter 2), specifically for two reasons: 1) the *topics* are too numerous in the dataset (2881), and 2) the disciplinary or thematic classifications used throughout the main text of the manuscript label the articles with neither a unique discipline (from WOS JSCs) nor a unique field of study (from MAG), thus compromising the execution of a k NN search. With such a subfield classification, the application of k NN to the SPECTER and UMAP datasets allows to build two confusion matrices summarizing the proportion of true predictions per subfield, which will be compared afterwards.

In what follows, we restrict our analysis to a subset of papers covering 68 subfields, each labeling at least 1000 publications – which accounts for 96% of the original dataset. We thus keep their associated vectors within the two embedded datasets. Before plotting their respective confusion matrices, we choose for each of them the number of neighbors k^* maximizing the accuracy of the predictions returned by the search algorithm. As shown in Fig. C.1, the accuracy for UMAP is limited to about 51% from approximately $k^* = 20$, while SPECTER’s one reaches its maximum of about 70% at $k^* = 12$. We choose these values to run optimized k NN searches within each dataset – with the help of the module `KNeighborsClassifier` of the Python package `scikit-learn`.

The results of the predictions per subfield are shown in Fig. C.2. They indicate that the predictions given by the k NN run over the papers embedded with SPECTER are concentrated around the diagonal of the corresponding confusion matrix, thus testifying that the papers lying in a given subfield are statistically surrounded by neighbors also

labeled with that subfield. The papers labeled with a common subfield are thus statistically included in a common area of knowledge represented by a common vocabulary, which confirms the rather good embedding performance of SPECTER. We recover this trend in the projected UMAP space, but with an accentuation of incorrect predictions in some zones of the confusion matrix, corresponding to those already observed in the SPECTER one, but with a lower intensity. This lack of precision concerns almost all subfields, which show an important part of the predictions towards the subfields *Molecular Biology* (light blue rectangular zone on the two heatmaps), *Cellular and Molecular Neuroscience* and *Cognitive Neuroscience* (grouped in a light green zone on the heatmaps). The neighborhoods of articles belonging to the subfields with high prediction rates to the three aforementioned ones are therefore predominantly populated by at least one of these three subfields, especially in the UMAP space. This suggests, first, that these major subfields cover a large vocabulary basis and therefore a broad part of the neuroscience knowledge landscape, and, second, that they are lexically close to articles labeled with subfields showing them as main predictions in the confusion matrices, such as *Biochemistry* (row 4), *Physiology* (row 12), *Developmental and Educational Psychology* (row 62) and *Experimental and Cognitive Psychology* (row 63). Moreover, because they are the most populated subfields – 15% labeled as *Cognitive Neuroscience*, 14% as *Cellular and Molecular Neuroscience*, and 10% as *Molecular Biology*, with the remaining 71% almost equally distributed among the 65 others – they are more likely to be predicted than any of the other subfields.

In summary, with a k NN algorithm applied to the prediction of the subfields labeling the articles in our neuroscience dataset, we confirm the robustness of the SPECTER embedding of the textual metadata of these articles and also of the UMAP dimension reduction algorithm used to transform the 768-dimensional vectors representing them into two-dimensional ones. Albeit a loss of prediction accuracy before and after projection, we have demonstrated a global conservation of the thematic structure of our corpus, thus enabling us to use the reduced vectors to study the cartography of neuroscience shown in Chapter 4, and also the location of AI-related knowledge in this reduced space.

C.1.2 HDBSCAN dendrogram

In this section, we detail the method retained to partition the knowledge map shown in Fig. 4.3 of Chapter 4 into 9 clusters.

We have drawn upon the main method provided by the Python package `hdbscan` (McInnes, Healy, and Astels, 2017), which returns a dendrogram indicating the different clusters we could get by varying a density parameter, here denoted as λ . Fig. C.3

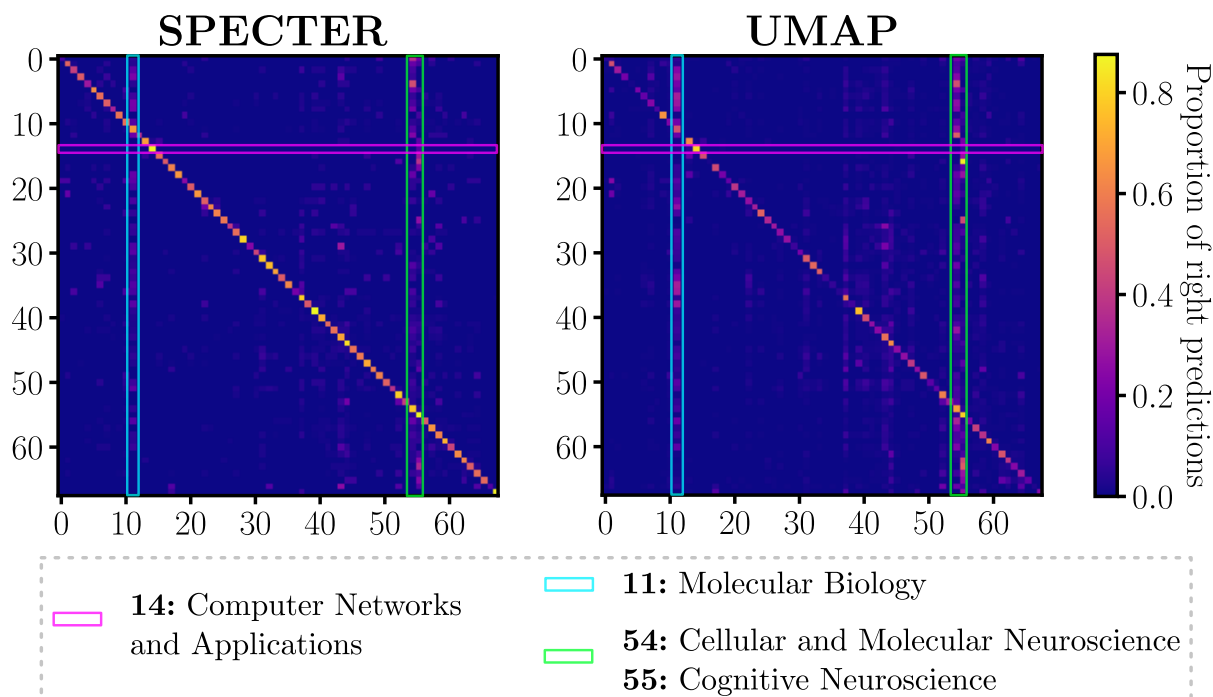


Figure C.2: Confusion matrix of k NN predictions per subfield for the SPECTER-embedded dataset (left) and the UMAP-embedded one (right). The indices of the matrices represent the 68 subfields considered to run the k NN models. The rows are the actual observed subfields and the columns are the predicted ones, sorted as in the rows. Given a subset of papers labeled with an observed subfield i (indexed in rows) within the test set, the cell m_{ij} in each matrix is the share of predictions falling within the subfield j (indexed in columns). For example, among the test papers labeled with the observed subfield *Computer Networks and Communications* (pink-colored row 14 in the heatmaps) in the SPECTER space, 87.5% of them have a correctly predicted subfield, i.e., the observed one, but the remaining 12.5% have a predicted subfield other than their common observed one, here *Cognitive Neuroscience*. In the UMAP space, this is 81% of them that are correctly predicted into the observed aforementioned subfield, against 19% exhibiting other subfields – 14 in total. Therefore, the sum of the prediction shares equals 1 for each observed subfield indexed in rows. For SPECTER, 99% of the dataset was used to train the k NN algorithm with $k^* = 12$, and the remaining 1% served to compute the predictions (with an accuracy of about 70%). For UMAP, we choose 80% of the dataset to train it with $k^* = 20$, and 20% to compute the predictions (with an accuracy of about 51%). The light blue rectangle corresponds to the columns associated with the subfields predicted as *Molecular Biology*, and the light green one to the subfields predicted as *Cellular and Molecular Neuroscience* and *Cognitive Neuroscience*.

illustrates a simple use case of this method, here applied to a set of numerical values picked from a heterogeneous distribution. According to such a dendrogram, choosing a high λ is equivalent to selecting small, local, and dense ensembles of points, such as those located at the very ends of the branches of the clusters C_1 and C_3 , which correspond to the purple and other peaks, respectively, in the lower left plot of the distribution for $\lambda = \lambda_2$. On the contrary, choosing a small λ is equivalent to selecting the largest clusters obtained by merging all the smallest clusters at the bottom of this hierarchy, such as the green cluster in the middle left plot at $\lambda = \lambda_1$, which is the sum of the clusters C_1 and C_2 , and also the yellow one at $\lambda = \lambda_0$ in the top left plot, which is the sum of all the three

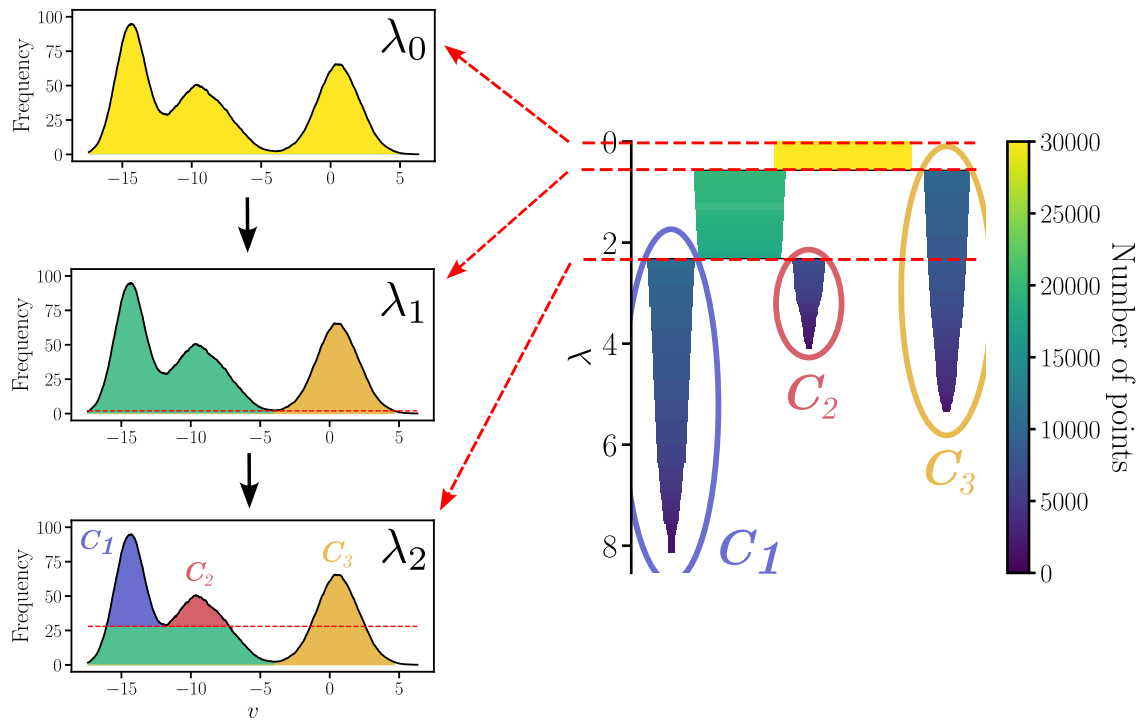


Figure C.3: Schematic illustration of the application of HDBSCAN to a set of 30,000 continuous numbers v on a horizontal axis generated from a non-trivial distribution, represented on the left plots. This distribution is composed of density peaks corresponding to high amounts of points in a small interval of the horizontal axis v , e.g. the leftmost peak (cluster C_1) corresponds to a high frequency of values v picked around -15 . On the contrary, the pits are regions with less selected values v . On the right is plotted the clustering dendrogram returned by the Python module `hdbscan`, where the density threshold parameter is denoted by λ . The clustering has been executed with a minimum neighborhood size of 1000 values for each number and a minimal cluster size of 1000 numbers. A branch is a cluster whose color is the number of values it contains at a given λ – represented also by the varying width when moving down the tree – and whose length is the persistence of the cluster as λ increases. For example, the cluster C_2 is less persistent than the other two clusters C_1 and C_3 , i.e., the former is merged with the sea of noisy points at $\lambda \sim 4$, while the latter continue to exist after that value, but with smaller respective sizes. As λ still increases, C_3 is also considered as noise by the algorithm before C_1 . At a value of $\lambda = \lambda_0 = 0$ of the density parameter, we consider all the points of the distribution, which together shape one unique cluster, colored yellow in the upper left plot. At a value of $\lambda = \lambda_1$, the set of points is split into two clusters, namely the cluster C_3 and another one colored in green in the middle plot. The latter is finally divided into two subparts at $\lambda = \lambda_2$, namely the clusters C_1 (in purple in the bottom plot) and C_2 (in wine red).

clusters. In such a case, the algorithm covers a large amount of points and thus reduces the amount of noise.

We apply this clustering method to our embedded 2D representation of neuroscience papers, i.e., after applying UMAP. Fig. C.4A shows the dendrogram returned by its execution. We manually choose the most separated and persistent clusters on this dendrogram, regardless of their respective size, shape, or intrinsic density, by progressively increasing λ . When setting λ around 5, we divide the dataset into 3 clusters, namely C_0 , C_1 , and a very large one whose partition is shown below in Fig. C.4B. In this way, we obtain

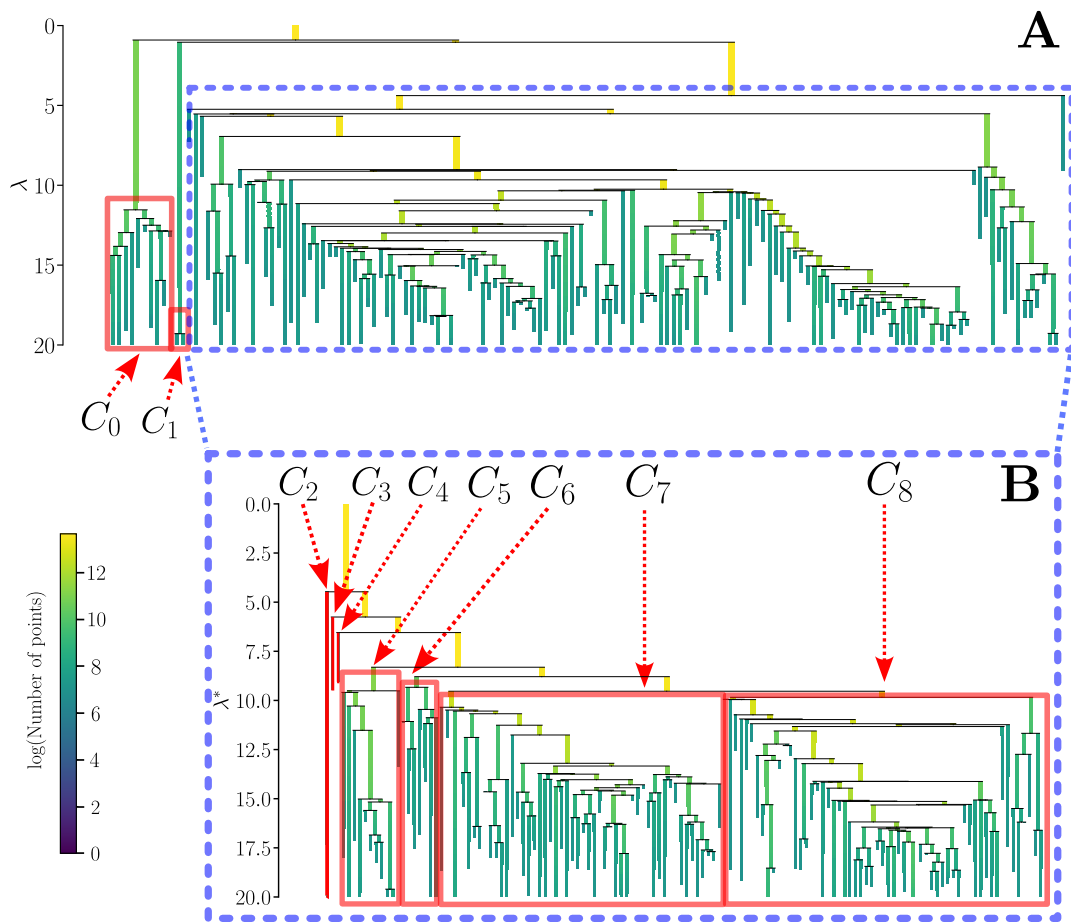


Figure C.4: A: Clustering dendrogram returned by HDBSCAN on our reduced dataset. The program has been executed with a minimum neighborhood size of 100 points for each paper and a minimal size of a cluster of 1000 papers. The density threshold is denoted as λ . A vertical bar is a cluster, whose color represents the number of points within the corresponding cluster, scaled with the logarithmic color bar on the left side, and whose length represents the persistence of the corresponding cluster according to λ , i.e., the height of the associated amount of points in the density map. B: Clustering dendrogram of the large cluster from dendrogram A above, replotted after a second run of HDBSCAN on the subset of data included in this cluster. The density threshold is denoted here as λ^* because it is a translated value of the λ parameter from plot A – the lowest value $\lambda^* = 0$ actually corresponds to $\lambda \sim 1$. The clusters C_2 , C_3 and C_4 are simple bars, colored red for clarity. Population of each cluster: $N_{C_0} = 56,644$; $N_{C_1} = 8,755$; $N_{C_2} = 2,222$; $N_{C_3} = 1,407$; $N_{C_4} = 1,179$; $N_{C_5} = 66,465$; $N_{C_6} = 28,382$; $N_{C_7} = 266,593$; $N_{C_8} = 370,814$. We also count 55,230 papers that are unclassified by HDBSCAN, i.e., that are considered as noise by the clustering algorithm.

the other clusters appearing in Fig. 4.3. The populations of all the clusters are indicated in the caption of Fig. C.4. With such a partition, we allow a total noise rate of about 6.5%. We considered these time-aggregated clusters for the analysis of all the neuroscience knowledge maps shown throughout the main text, in particular the 5-year temporal maps in Fig. 4.4.

Furthermore, choosing a high number of neighbors to perform HDBSCAN leads to very aggregated clusters with inhomogeneous density within them, while choosing a low number of neighbors favors the formation of smaller and more homogeneously distributed

clusters of points – hence the very high number of branches Fig. C.4. We have opted for this second strategy to plot this dendrogram for two reasons: 1) to avoid the identification of large and thematically inconsistent clusters, which could gather too distant knowledge, and 2) to allow a more precise partitioning of the dataset in the future, especially the large clusters C_7 and C_8 .

C.1.3 Definition of the obtained clusters

Since the clusters delimited before are of different sizes, they could either cover a small and consistent lexical region corresponding to a given research specialty in neuroscience (such as C_1 , C_2 , C_3 and C_4), or many different ones and therefore many research areas with strong vocabulary distance at the same time (such as C_7 and C_8). Furthermore, depending on the topical or conceptual taxonomy provided by bibliometric databases and used to label the articles in our dataset, we could misapprehend the contents of these clusters, notably the largest ones. Therefore, in what follows, we compare three such taxonomies in order to approach as best as possible the contents of these clusters, here MAG’s *fields of study*, and OpenAlex’s *topics* and *keywords*, which label the articles. In particular, we indicate in Tab. C.1, for each classification, the five most important elements shared by the articles in each cluster.

For the MAG classification, we complete the list of the most frequent fields of study situated at level 2 of the field network by indicating the most frequent ones at level 3, since the fields located at the former level are too general and do not provide a clear meaning of the cluster on their own. Moreover, we provide this supplementary list because 92.3% of the papers in our neuroscience dataset are tagged with at least one concept situated at this level. We discard the fields situated at the first level of the field networks because they are merely disciplines or broad research domain, and also the fourth and fifth ones because they are more sparsely tagged in our dataset.

Despite the excessive generality of the MAG taxonomy of fields of study – even at the third level – we observe that the three classifications match quite well for the smallest clusters: C_0 ² focuses on the eye and vision studies, C_1 on the design of artificial neural network models and classifiers for various applications, C_2 on the Attention-Deficit/Hyperactivity Disorders (ADHD) – the first term *Methylphenidate* within the list of level-3 fields of study for this cluster is also a molecule found in psychostimulant treatments for such disorders –, C_3 on the neurological studies of the Obsessive-Compulsive Disorder (OCD) and Tourette syndrome, C_4 on the studies of the genetic factors responsible for the onset of Parkinson’s disease, and C_6 on the studies of sensory pathologies related to hearing and smell. The

²A closer look at the following level-2 fields of study within the C_0 ranking reveals the terms *Visual field*, *Lens (optics)* and *Astigmatism*, thus confirming the name of the cluster mentioned in the main text.

	Fields of Study (level 2)	Fields of Study (level 3)	Topics	Keywords
C_0	Diabetes mellitus Population In vivo Cell culture Erg	Retinal Retinal Eye Disease Visual acuity Cornea	Molecular Mechanisms of Retinal Degeneration and Regeneration Global Prevalence and Treatment of Glaucoma Age-Related Macular Degeneration Research Biomechanical Properties of the Cornea and Related Diseases Cataract Surgery Techniques and Complications	Intraocular Pressure Photoreceptor Degeneration Intraocular Lens Implantation Sutureless Intraocular Lens Fixation Ocular Hypertension
C_1	Artificial neural network Nonlinear system Support vector machine Cluster Analysis Convergence (routing)	Exponential stability Recurrent Neural Network Lyapunov function Deep learning Backpropagation	Neural Network Fundamentals and Applications Network Synchronization in Complex Systems Face Recognition and Dimensionality Reduction Techniques Theory and Applications of Extreme Learning Machines Blind Source Separation and Independent Component Analysis	Backpropagation Learning Time Delays Support Vector Machines Feedforward Neural Networks Recurrent Neural Networks
C_2	Attention deficit hyperactivity disorder Cognition Electroencephalography Impulsivity Population	Methylphenidate Comorbidity Neuropsychology El Niño Working memory	Attention-Deficit/Hyperactivity Disorder Analysis of Brain Functional Connectivity Networks Autism Spectrum Disorders Obsessive-Compulsive Disorder and Related Conditions Epidemiology and Management of Bipolar Disorder	Attention-Deficit/Hyperactivity Disorder ADHD Adolescent Brain Development Brain Imaging Attentional Networks
C_3	Obsessive compulsive Tics Cognition Neuroimaging Obsessive-compulsive disorder (OCD)	Tourette syndrome Anxiety Comorbidity Schizophrenia Neuropsychology	Obsessive-Compulsive Disorder and Related Conditions Deep Brain Stimulation for Neurological Disorders Analysis of Brain Functional Connectivity Networks Effects of Brain Stimulation on Motor Cortex Diffusion Magnetic Resonance Imaging	Obsessive-Compulsive Disorder Tourette Syndrome Deep Brain Stimulation Hoarding Behavior Treatment
C_4	Gene Population Diabetes Mellitus	Disease Mutation Genotype	Pathophysiology of Parkinson's Disease Lysosomal Storage Disorders in Human Health and Disease Epigenetic Modifications and Their Functional Implications	Parkinson's Disease Parkinsonism Neurodegeneration

Continued on next page

	Fields of Study (level 2)	Fields of Study (level 3)	Topics	Keywords
	Meta-analysis Cognition	Allele Polymorphism (computer science)	Nurr1/CoREST Pathway in Neuroprotection and Inflammation Deep Brain Stimulation for Neurological Disorders	Dopaminergic Neurons Genetic Risk Factors
C_5	Gene Signal transduction Diabetes mellitus Population Cell	Cancer Apoptosis Transcription factor Receptor Mutation	Gliomas Molecular Mechanisms of Muscle Regeneration and Atrophy Regulation and Function of Microtubules in Cell Division Mitochondrial Dynamics and Reactive Oxygen Species Regulation Myasthenia Gravis and Thymic Tumors Research	Skeletal Muscle Atrophy Glioblastoma mRNA modification Muscle Regeneration Brain Tumor Epidemiology
C_6	Stimulus (physiology) Cochlea Olfaction Inner ear Sensory system	Electrophysiology Hair cell Auditory cortex Stimulation Receptor	Cochlear Neuropathy and Hearing Loss Mechanisms Olfactory Dysfunction in Health and Disease Neuroscience and Genetics of <i>Drosophila Melanogaster</i> Avian Vocal Communication and Evolutionary Implications Impact of Hearing Loss on Cognitive Function	Olfactory Receptors Olfactory System Auditory Processing Inner Ear Development Avian Vocal Communication
C_7	Cognition Stimulus (physiology) Perception Electroencephalography Population	Schizophrenia Epilepsy Visual perception Prefrontal cortex Disease	Neural Mechanisms of Visual Perception and Processing Neural Mechanisms of Cognitive Control and Decision Making Epilepsy and Seizures Analysis of Brain Functional Connectivity Networks Neuronal Oscillations in Cortical Networks	Visual Perception Neuroimaging Data Analysis Perceptual Learning Working Memory Sensory Processing
C_8	Hippocampus Diabetes mellitus Glutamate receptor Spinal cord Neuron	Central nervous system Receptor Hippocampal formation Dopamine Stimulation	Molecular Mechanisms of Synaptic Plasticity and Neurological Disorders Neurobiological Mechanisms of Drug Addiction and Depression Role of Neuropeptides in Physiology and Disease Mechanisms of Alzheimer's Disease Mechanisms and Management of Neuropathic Pain	Neuroinflammation Glutamate Receptors Dopamine Neurodegeneration Cell Signaling

Table C.1: Table summarizing the five most important fields of study (retrieved from MAG classification, at levels 2 and 3), topics and keywords (both retrieved from OpenAlex) per cluster. To fill the second column, we retain only OpenAlex’s *primary topics* of the papers within each cluster. In the third column, we consider the ensemble of all OpenAlex’s keywords associated with the articles within each cluster.

Cluster	Name	N_p
C_0	Eyes and vision studies	56,644
C_1	Mathematical and computational foundations of connectionist AI	8,755
C_2	Attention-Deficit/Hyperactivity Disorders	2,222
C_3	Obsessive-Compulsive Disorders and Tourette syndrome studies	1,407
C_4	Parkinson's disease	1,179
C_5	Gliomas and skeletal muscle atrophy studies	66,465
C_6	Studies of sensorial pathologies	28,382
C_7	Brain neural networks and neurological disorders studies (epilepsy and schizophrenia)	266,593
C_8	Studies of neurobiological mechanisms in the central nervous system	370,814
Noise	–	55,230

Table C.2: Final description of the clusters used throughout the Chapter 4. We also recall the number of papers N_p within the clusters obtained with HDBSCAN.

other three clusters are more thematically diversified: C_5 gathers studies on glioma, an aggressive brain tumor responsible for severe neurological degeneration, and on skeletal muscle atrophy and regeneration, C_7 focuses on the studies of various functions of brain neural networks – vision, hearing, memory, cognition, etc. – and also on schizophrenia and epilepsy disorders, and C_8 ³ on various neurobiological mechanisms that could occur in the central nervous system and that could explain some pathologies – drug addiction, depression, neuropathic pain, among others – and other neurological disorders,⁴ such as Alzheimer's disease and multiple sclerosis.

By compiling such elements associated with their papers, we attribute a unique name to the clusters obtained above, as summarized in Tab. C.2.

³The following terms within the ranking of level-2 fields of study describing cluster C_8 are *Cerebral cortex* and *Synapse*, thus confirming the chosen name of the cluster centered on various studies of the central nervous system.

⁴The following website offers a classification of all the neurological disorders referenced: <https://www.merckmanuals.com/home/brain,-spinal-cord,-and-nerve-disorders> (visited on July, 14 2024).

C.2 Temporal Jaccard similarity between the topical universe of the knowledge clusters and that presented by the AI-related papers within them

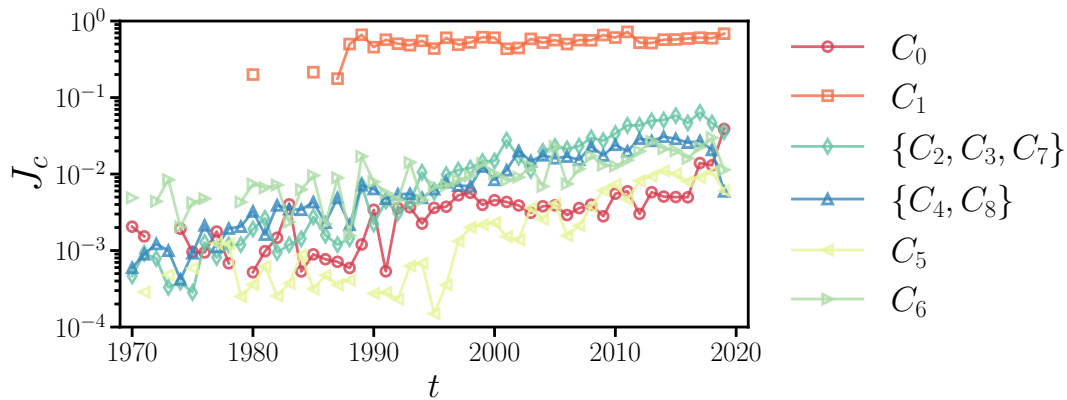


Figure C.5: Jaccard ranking similarity between the main topics (according to the OpenAlex taxonomy) presented by the papers that are members of a given cluster C_i (or in a set of clusters, indicated between brackets) and those associated with the AI-related publications within the same cluster. The small peripheral clusters C_2 , C_3 , and C_4 , whose weak presence of AI papers over the years causes noisy, sometimes non-continuous similarity trends, are joined within their large closer clusters, here C_7 for the first two and C_8 for the third.

C.3 Dynamic citation radius of gyration of the papers published in 1976 and 1989

The temporal average citation radius of gyration (RoG) produced by the location of the citations of a single paper within the neuroscience knowledge map, as computed with the method in Chapter 4 and shown in Fig. 4.8 in the main text (under its normalized definition), is a measure that covers all the citations accumulated by that paper over the years. Thus, a paper published at the very beginning of the time period under study (1970-2019) could have more citations and therefore a better lexical coverage than a paper published more recently. In order to prospect the dynamic citation diffusion of the AI-related papers published in the two specific years when the average RoG exhibits a maximum, namely 1976 and 1989, we explore the evolution of the citation gyration radii of these papers, expressed here as the logarithmic return $\log(1 + R_t)$, where $R_t = (r_g(t) - r_g(t - 1))/r_g(t - 1)$ is the percentage of gain ($R_t > 0$) or loss ($R_t < 0$) in RoG between two consecutive years. This measure, shown in Fig. C.6 below, indicates for a single paper, respectively, the expansion of the lexical coverage exhibited by the papers citing it, or the concentration of these citing papers around the local lexical subspace surrounding the paper under study.

Fig. C.6 shows that most of the papers published in the two years studied (1976 and

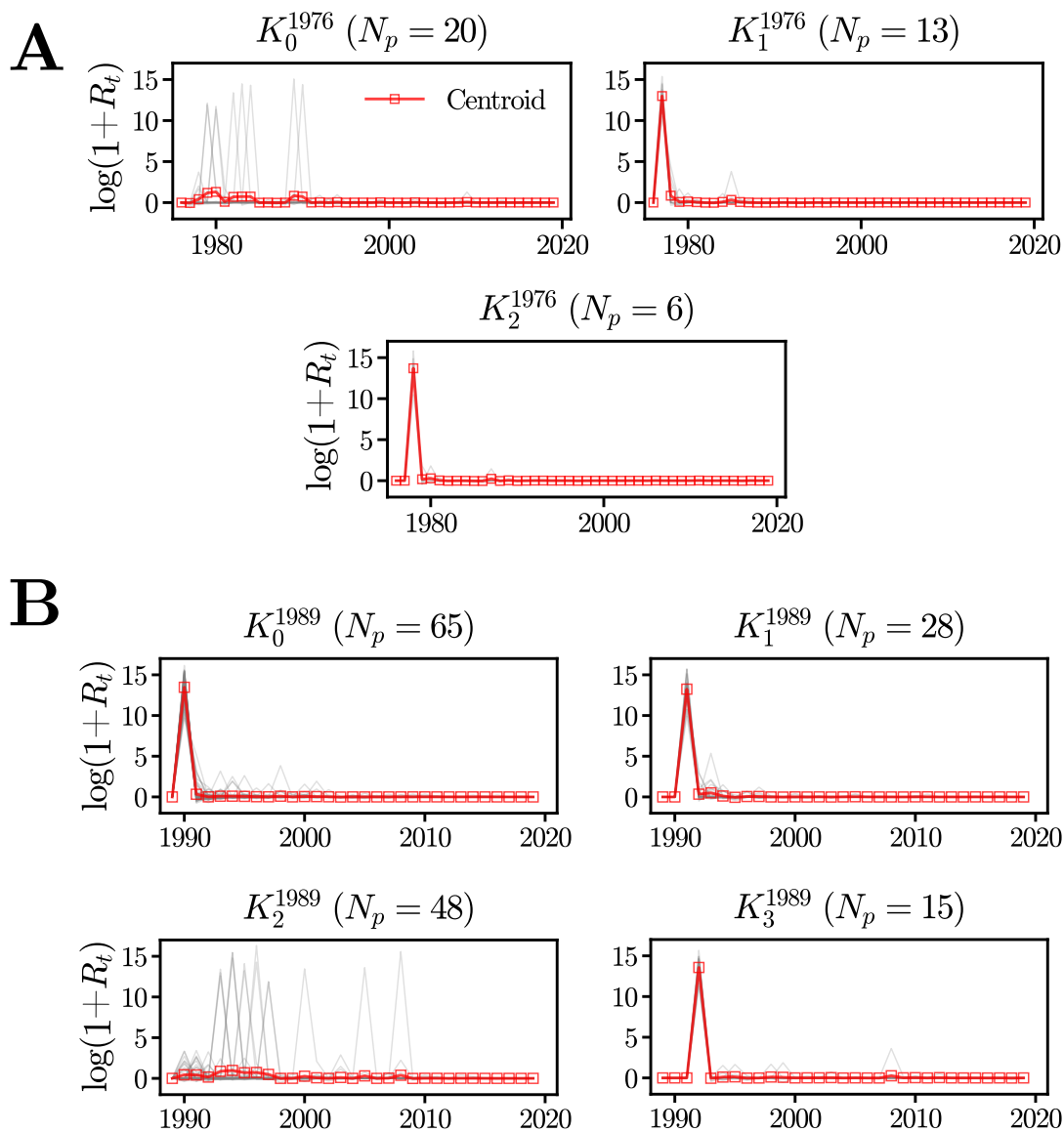


Figure C.6: Temporal log return R_t of the citation RoG of the papers published in 1976 (**A**) and in 1989 (**B**). The time series are aggregated into clusters K_i in each year with a k -means algorithm, based on an optimal number of clusters identified with a standard elbow method. N_p is the number of articles within the studied cluster K_i and therefore the number of different log-returned RoGs plotted for this cluster. The red curves represent the centroid of the cluster time series.

1989) cover their final lexical subspace very quickly after the publication, as shown by the high peaks situated one, two, or at most three years after publication. This trend is represented by the papers included in the clusters K_1^{1976} , K_2^{1976} , K_0^{1989} , K_1^{1989} and K_3^{1989} . The papers included in the clusters K_0^{1976} and K_2^{1989} follow the same trend, but with some peculiarities, such as a variation that is not as important as those in the other clusters in the two or three years following the publication, and also a later impact after publication – four years or more.

D Chapter 5 appendices

D.1 Interview grid

D.1.1 Original questionnaire in french

1. Trajectoires professionnelles

1.1. Formation et parcours professionnel avant entrée dans l'équipe ou l'institut

- Quelle type de formation avez-vous suivi (universitaire, ingénierie...) ?
- À quelle(s) spécialité(s)/disciplines(s) avez-vous été formé (santé, informatique, mathématique...) ?
- Dans quel(s) établissement(s) avez-vous été formé ? (lieu, ville, pays)
- Si vous êtes détenteur d'une thèse, sur quoi portait-elle ? Avec qui l'avez-vous réalisée ? Dans quel(s) laboratoire(s) de tutelle ?
- Avez-vous réalisé des *post-docs* ? Si oui, dans quel(s) laboratoire(s) ? Sinon, quelle profession intermédiaire avez-vous exercée ?

1.2. Sensibilisation à l'IA

- Durant votre cursus, connaissiez-vous déjà l'IA ?
- Avez-vous été initié à l'utilisation de certains outils d'IA, via des cours théoriques ou des stages pratiques ?
- Ou bien vous-êtes vous formé à ces outils en autodidacte ?

1.3. Recherche actuelle

- Quel est votre sujet de recherche actuel ?
- Votre recherche est-elle confidentielle ?
- Quel(s) article(s) de référence conseilleriez-vous pour illustrer vos recherches ? Quel(s) sont ceux que vous banniriez ? Pourquoi ?

2. Utilisation de l'IA

2.1. Focus sur l'IA dans les recherches

- Quel(s) type(s) d'IA manipulez-vous au quotidien ? Plutôt numérique (*machine learning*) ou symbolique (*systèmes experts*) ?

- Travaillez-vous sur la conception ou l'amélioration d'une IA (recherche fondamentale) ou plutôt sur son application médicale ?
- Quelles sont les outils médicaux potentiellement *candidats* à l'implémentation d'un programme d'IA ? (imagerie médicale, aide à la prise de décision, élaboration de diagnostics...)
- Collaborez-vous avec les équipes de recherche clinique ?
- Y a-t-il une validation des programmes conçus auprès de patients sélectionnés, au sein ou hors du site hospitalier ?

2.2. Données d'entraînement utilisées

- Si vous utilisez des données anonymisées (santé publique, génétique...) pour entraîner les algorithmes, quelles sont-elles ?
- D'où proviennent-elles ?
- Comment vous les procurez-vous ? (autorisation auprès d'une instance supérieure, vérification du *bien-fondé* de la recherche en cours...)

3. Perception de l'IA

3.1. Lors de la professionnalisation

- Lors de votre formation ou de votre parcours professionnel, l'IA vous semblait-elle balbutiante ? Ou bien vous semblait-elle déjà répandue dans certaines disciplines scientifiques ?
- Cela dépendait-il des laboratoires ? Par exemple, avez-vous fréquenté des laboratoires dans lesquels l'IA était absente malgré son essor dans la recherche scientifique ces vingt dernières années ?

3.2. Actuellement au travail

- L'IA a-t-elle changé votre quotidien de recherche en général ?
- Plus précisément, a-t-elle changé vos pratiques, créé ou métamorphosé certains savoir-faire ?
- A-t-elle accéléré certains processus automatisés ? A-t-elle remplacé certaines tâches contraignantes que vous deviez faire manuellement ?
- Si tel est le cas, par quoi remplacez-vous ce temps « gagné » ?
- Les outils d'IA que vous utilisez ou inventez créent-ils des connaissances ? Auriez-vous fait les mêmes découvertes sans l'intervention de l'IA ?

3.3. Considérations générales sur l'IA en santé

- Selon vous, l'IA au service de la santé publique pourra-t-elle totalement remplacer les médecins ?

- Selon vous, quels sont les principaux enjeux derrière l'IA (santé publique, développement scientifique...)
- Que pense la communauté médicale autour de vous à propos de l'IA ? Se montre-t-elle méfiante envers l'IA ? Cette dernière serait-elle néfaste à la profession et à l'éthique médicale ? (l'IA source de discrédit de la profession ?)
- Que pensez-vous de la course à l'homme augmenté aujourd'hui en cours ? (normes de représentations corporelles ?)
- Pensez-vous que l'IA dépassera l'homme ?

4. Cadre institutionnel

- Travailler dans le cadre dans un institut largement financé par quelques particuliers vous procure-t-il un confort de recherche (financier et matériel) ? Si oui, cela vous permet-il d'avancer sereinement ?
- Travaillez-vous avec des chercheurs du monde industriel et de la R&D ? Si oui, de quels laboratoire, bureau d'études, ou entreprise proviennent-ils ?
- Travaillez-vous avec des start-ups innovant et brevetant des technologies d'IA ? Si oui, lesquelles ? Font-elles parties de l'institut ?

D.1.2 Translated in english

1. Professional career

1.1. Training and career path before joining the team or the institute

- What type of training did you follow (university, engineering...)?
- In which speciality(ies)/discipline(s) have you been trained (health, medical track, neuroscience, computer science, mathematics...)?
- In which organisms(s) have you been trained (name, town, country)?
- If you are PhD graduate, what was your study object? With whom did you complete it? In which laboratory(ies)?
- Have you completed any *post-docs*? If so, in which laboratory(ies) ? If not, what intermediate profession have you held?

1.2 Meeting with AI

- During your studies, did you already know about AI?
- Have you been initiated to the use of certain AI tools, via theoretical courses or practical training?

- Or did you learn about these tools on your own?

1.3. Current search

- What is your current research topic?
- Is your research confidential?
- What reference article(s) would you recommend to illustrate your research? Which ones would you ban? Why?

2. Use of AI

2.1. Focus on AI in research

- What type(s) of AI do you handle on a daily basis? More data-fed algorithms (*machine learning*) or symbolic ones (*expert systems*)?
- Are you working on the design or improvement of an AI (fundamental research), or rather on its medical application?
- What are the medical tools that are potential candidates for the implementation of an AI programme (medical imaging, decision support, diagnosis, etc.)?
- Do you collaborate with clinical research teams?
- How are programmes designed with selected patients validated, within or outside the hospital site?

2.2. Used training data

- If you use any anonymized data (public health, genetics...) to train your algorithms, what are they?
- How are they produced?
- How do you obtain them (authorization from a higher authority, verification of the validity of the research in progress, etc.)?

3. AI perception by users

3.1. During professionalization

- During your training or your professional career, did AI seems still fledgling to you? Or did it already seem widespread in some scientific disciplines?
- Did it depend on the laboratories? For example, have you been in laboratories where AI was absent despite its growth in scientific research over the last twenty years?

3.2. Currently at work

- Has AI changed your daily research routine in general?
- More specifically, has it changed your practices, created or transformed certain skills?
- Has it sped up some automated processes? Has it replaced certain burdensome tasks that you used to have to do manually?
- If so, what are you replacing this newly available time with?
- Do AI tools you use or invent create knowledge? Would you have made the same discoveries without AI?

3.3. General considerations on AI in healthcare

- In your opinion, will AI in public health be able to completely replace practitioners?
- In your opinion, what are the main issues behind AI (public health, scientific development, etc.)?
- What does the medical community around you think about AI? Are they wary of AI? Would AI be a threat to the medical profession and ethics?
- What do you think of the race to “augmented human” that is currently underway?
- Do you think AI will overtake the human kind?

4. Institutional context

- Does working within an institute largely funded by a few private individuals provide you a kind of research comfort (of financial and material kinds)? If so, does it allow you to move forward serenely?
- Do you work with researchers from industry and R&D? If so, which laboratories, engineering design offices or companies do they come from?
- Do you work with start-ups innovating and patenting AI technologies? If so, which ones? Are they part of the institute?

Bibliography

- Abrol, A., Fu, Z., Salman, M., Silva, R., Du, Y., Plis, S., & Calhoun, V. (2021). Deep learning encodes robust discriminative neuroimaging representations to outperform standard machine learning. *Nature Communications*, *12*(1), 353.
- Acemoglu, D., Akcigit, U., & Kerr, W. R. (2016). Innovation network. *Proceedings of the National Academy of Sciences*, *113*(41), 11483–11488.
- Aghion, P., & Bouverot, A. (2024). *IA : notre ambition pour la France* (Premier Ministre, Ed.; tech. rep.).
- Ahmed, N., Wahed, M., & Thompson, N. C. (2023). The growing influence of industry in AI research. *Science*, *379*(6635), 884–886.
- Alač, M., Movellan, J., & Tanaka, F. (2011). When a robot is social: Spatial arrangements and multimodal semiotic engagement in the practice of social robotics. *Social Studies of Science*, *41*(6), 893–926.
- Allen, B., Seltzer, S. E., Langlotz, C. P., Dreyer, K. P., Summers, R. M., Petrick, N., Marinac-Dabic, D., Cruz, M., Alkasab, T. K., Hanisch, R. J., Nilsen, W. J., Burleson, J., Lyman, K., & Kandarpa, K. (2019). A Road Map for Translational Research on Artificial Intelligence in Medical Imaging: From the 2018 National Institutes of Health/RSNA/ACR/The Academy Workshop. *Journal of the American College of Radiology*, *16*(9), 1179–1189.
- Andler, D. (1990). Connexionnisme et cognition: À la recherche des bonnes questions. *Revue de Synthèse*, *111*(1-2), 95–127.
- Anichini, G., & Geffroy, B. (2021). L'intelligence artificielle à l'épreuve des savoirs tacites. Analyse des pratiques d'utilisation d'un outil d'aide à la détection en radiologie. *Sciences Sociales et Santé*, *39*(2), 43–69.

- Ansart, M., Epelbaum, S., Bassignana, G., Bône, A., Bottani, S., Cattai, T., Couronné, R., Faouzi, J., Koval, I., Louis, M., Thibeau-Sutre, E., Wen, J., Wild, A., Burgos, N., Dormont, D., Colliot, O., & Durrleman, S. (2021). Predicting the progression of mild cognitive impairment using machine learning: A systematic, quantitative and critical review. *Medical Image Analysis*, *67*, 101848.
- Arthur, W. B. (1989). Competing Technologies, Increasing Returns, and Lock-In by Historical Events. *The Economic Journal*, *99*(394).
- Bagrow, J. P., & Boltt, E. M. (2019). An information-theoretic, all-scales approach to comparing networks. *Applied Network Science*, *4*(1), 1–15.
- Bainbridge, W. S., Brent, E. E., Carley, K. M., Heise, D. R., Macy, M. W., Markovsky, B., & Skvoretz, J. (1994). Artificial Social Intelligence. *Annual Review of Sociology*, *20*, 407–436.
- Barber, B. (1961). Resistance by Scientists to Scientific Discovery. *Science, New Series*, *134*(3479), 596–602.
- Baruffaldi, S., van Beuzekom, B., Dernis, H., Harhoff, D., Rao, N., Rosenfeld, D., & Squicciarini, M. (2020). *Identifying and measuring developments in artificial intelligence: Making the impossible possible* (OECD Science, Technology and Industry Working Papers No. 2020). Editions OCDE, Paris.
- Bastin, G., & Tubaro, P. (2018). Le moment big data des sciences sociales. *Revue française de sociologie*, *Vol. 59*(3), 375–394.
- Batagelj, V., & Zaversnik, M. (2003). An $O(m)$ Algorithm for Cores Decomposition of Networks. arXiv. <http://arxiv.org/abs/cs/0310049>
- Bechmann, A., & Bowker, G. C. (2019). Unsupervised by any other name: Hidden layers of knowledge production in artificial intelligence on social media. *Big Data & Society*, *6*(1), 1–11.
- Beer, D. (2017). The social power of algorithms. *Information, Communication & Society*, *20*(1), 1–13.
- Beltagy, I., Lo, K., & Cohan, A. (2019). SciBERT: A Pretrained Language Model for Scientific Text. arXiv. <http://arxiv.org/abs/1903.10676>
- Bianchini, S., Müller, M., & Pelletier, P. (2022). Artificial intelligence in science: An emerging general method of invention. *Research Policy*, *51*(10).
- Bianchini, S., Müller, M., & Pelletier, P. (2023). Drivers and Barriers of AI Adoption and Use in Scientific Research. arXiv. <http://arxiv.org/abs/2312.09843>

- Birkle, C., Pendlebury, D. A., Schnell, J., & Adams, J. (2020). Web of Science as a data source for research on scientific and scholarly activity. *Quantitative Science Studies*, 1(1), 363–376.
- Bloomfield, B. P. (1988). Expert systems and human knowledge: A view from the sociology of science. *AI & Society*, 2(1), 17–29.
- Boden, M. A. (1998). Creativity and artificial intelligence. *Artificial Intelligence*, 103, 347–356.
- Bornmann, L., & Daniel, H.-D. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, 64(1), 45–80.
- Bourdieu, P. (1975). The specificity of the scientific field and the social conditions of the progress of reason. *Social Science Information*, 14(6), 19–47.
- Bowker, G. C. (1993). How to be universal: Some Cybernetics Strategies, 1943-1970. *Social Studies of Science*, 23(1), 107–127.
- Boyd, R., & Holton, R. J. (2018). Technology, innovation, employment and power: Does robotics and artificial intelligence really mean social transformation? *Journal of Sociology*, 54(3), 331–345.
- boyd danah, d., & Crawford, K. (2012). Critical questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5), 662–679.
- Brahim, A. S., Tabourier, L., & Le Grand, B. (2021). A data-driven analysis to question epidemic models for citation cascades on the blogosphere. arXiv. <http://arxiv.org/abs/1306.0424>
- Brancheau, J. C., & Wetherbe, J. C. (1990). The Adoption of Spreadsheet Software: Testing Innovation Diffusion Theory in the Context of End-User Computing. *Information Systems Research*, 1(2), 115–143.
- Buchanan, B. G. (2005). A (very) brief history of Artificial Intelligence. *AI Magazine*, 26(4), 53–60.
- Burrell, J. (2016). How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1), 1–12.
- Burrell, J., & Fourcade, M. (2021). The Society of Algorithms. *Annual Review of Sociology*, 47(1), 213–237.
- Cardon, D., Cointet, J.-P., & Mazières, A. (2018). Neurons spike back. The invention of inductive machines and the artificial intelligence controversy (L. Carey-Libbrecht, Trans.). *Réseaux*, 211(5), 173–220.

- Carley, K. M. (1989). The Value of Cognitive Foundations for Dynamic Social Theory. *Journal of Mathematical Sociology*, 14(2/3), 171.
- Carley, K. M. (1991). A Theory of Group Stability. *American Sociological Review*, 56(3), 331–354.
- Carley, K. M. (1996). Artificial Intelligence within Sociology. *Sociological Methods and Research*, 25(3), 29.
- Cencetti, G., Contreras, D. A., Mancastroppa, M., & Barrat, A. (2023). Distinguishing Simple and Complex Contagion Processes on Networks. *Physical Review Letters*, 130(24), 247401.
- Centola, D. (2015). The Social Origins of Networks and Diffusion. *American Journal of Sociology*, 120(5), 1295–1338.
- Chen, R., Ramsundar, B., & Robbins, R. (2019). Fair value and decentralized governance of data. Retrieved January 20, 2021, from https://github.com/computablelabs/computable/blob/master/whitepaper/computable_whitepaper.pdf
- Cheng, M., Smith, D. S., Ren, X., Cao, H., Smith, S., & McFarland, D. A. (2023). How New Ideas Diffuse in Science. *American Sociological Review*, 88(3), 522–561.
- Chubin, D. E. (1976). The Conceptualization of Scientific Specialties. *The Sociological Quarterly*, 17(4), 448–476.
- CISTP (Ed.). (2018). *China AI Development Report* (tech. rep.). China Institute for Science and Technology Policy at Tsinghua University.
- Cockburn, I., Henderson, R., & Stern, S. (2018). *The Impact of Artificial Intelligence on Innovation* (tech. rep. No. 24449). National Bureau of Economic Research. Cambridge, MA, USA.
- Cohan, A., Feldman, S., Beltagy, I., Downey, D., & Weld, D. S. (2020). SPECTER: Document-level Representation Learning using Citation-informed Transformers. arXiv. <http://arxiv.org/abs/2004.07180>
- Cointet, J.-P., & Parasie, S. (2018). Ce que le big data fait à l'analyse sociologique des textes. *Revue française de sociologie*, 59(3), 533–557.
- Cole, J., & Zuckerman, H. (1975). The Emergence of a Scientific Specialty: the Self-Exemplifying case of the Sociology of Science. In *The Idea of Social Structure. Papers in honor of Robert K. Merton* (Lewis A. Coser Editions, pp. 139–174). New York, Harcourt Brace Jovanovich.
- Coleman, J., Katz, E., & Menzel, H. (1957). The Diffusion of an Innovation among Physicians. *Sociometry*, 20(4), 253–270.

- Conte, R., Gilbert, N., Bonelli, G., Cioffi-Revilla, C., Deffuant, G., Kertesz, J., Loreto, V., Moat, S., Nadal, J. -P., Sanchez, A., Nowak, A., Flache, A., San Miguel, M., & Helbing, D. (2012). Manifesto of computational social science. *The European Physical Journal Special Topics*, 214(1), 325–346.
- Cooper, R. P., & Shallice, T. (2010). Cognitive Neuroscience: The Troubled Marriage of Cognitive Science and Neuroscience. *Topics in Cognitive Science*, 2(3), 398–406.
- Crawford, K. (2021). *Atlas of AI : Power, Politics and the Planetary Costs of Artificial Intelligence*. Yale University Press.
- Crawford, K., & Joler, V. (2018). Anatomy of an AI System: The Amazon Echo as an Anatomical Map of Human Labor, Data and Planetary Resources. AI Now Institute; Share Lab. Retrieved September 8, 2024, from <http://www.anatomyof.ai>
- Crawford, K., & Trevor, P. (2019). Excavating AI: The Politics of Images in Machine Learning Training Sets. Retrieved July 19, 2022, from <https://excavating.ai>
- Cronin, B., & Sugimoto, C. R. (2014). *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact*. The MIT Press.
- Cupe, J. (2018). L'interprétabilité de l'IA - Le nouveau défi des data scientists. Retrieved April 11, 2021, from <https://www.actuia.com/contribution/jean-cupe/linterpretabilite-de-lia-le-nouveau-defi-des-data-scientists/>
- Dartinet, V., & Martinaud, O. (2005). La BREF, une batterie rapide d'évaluation frontale. *NPG Neurologie - Psychiatrie - Gériatrie*, 5(29), 43–46.
- De Vignemont, F. (2020). Un homme augmenté mais à quel prix ? *Cahiers Droit, Sciences & Technologies*, (11), 23–30.
- Derouesné, C., Poitreneau, J., Hugonot, L., Kalafat, M., Dubois, B., & Laurent, B. (1999). Le Mini-Mental State Examination (MMSE) : un outil pratique pour l'évaluation de l'état cognitif des patients par le clinicien. *La Presse Médicale*, 28(21), 1141–1148.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv. <http://arxiv.org/abs/1810.04805>
- Di Bona, G., Bracci, A., Perra, N., Latora, V., & Baronchelli, A. (2023). The concept of decentralization through time and disciplines: a quantitative exploration. *EPJ Data Science*, 12(1), 42.

- Do, S., Ollion, É., & Shen, R. (2022). The Augmented Social Scientist: Using Sequential Transfer Learning to Annotate Millions of Texts with Human-Level Accuracy. *Sociological Methods & Research*, 53(3), 1167–1200.
- Espinosa-Rada, A., Lerner, J., & Fritz, C. (2024). Socio-cognitive Networks between Researchers. arXiv. Retrieved August 1, 2024, from <http://arxiv.org/abs/2407.21067>
- Färber, M. (2019). The Microsoft Academic Knowledge Graph: A Linked Data Source with 8 Billion Triples of Scholarly Data. In C. Ghidini, O. Hartig, M. Maleshkova, V. Svátek, I. Cruz, A. Hogan, J. Song, M. Lefrançois, & F. Gandon (Eds.), *The Semantic Web – ISWC 2019* (pp. 113–129). Cham, Springer International Publishing.
- Färber, M., & Ao, L. (2022). The Microsoft Academic Knowledge Graph enhanced: Author name disambiguation, publication classification, and embeddings. *Quantitative Science Studies*, 3(1), 51–98.
- Fecher, B., Hebing, M., Laufer, M., Pohle, J., & Sofsky, F. (2023). Friend or foe? Exploring the implications of large language models on the science system. *AI & Society*.
- Fontaine, S., Gargiulo, F., Dubois, M., & Tubaro, P. (2024a). Epistemic integration and social segregation of AI in neuroscience. *Applied Network Science*, 9(1), 8.
- Fontaine, S., Gargiulo, F., Dubois, M., & Tubaro, P. (2024b). Epistemic integration and social segregation of AI in neuroscience - Dataset extracted from the Microsoft Academic Knowledge Graph. Zenodo. Retrieved August 28, 2024, from <https://zenodo.org/records/10777508>
- Forsythe, D. E. (1993a). The Construction of Work in Artificial Intelligence. *Science, Technology, & Human Values*, 18(4), 460–479.
- Forsythe, D. E. (1993b). Engineering Knowledge: The Construction of Knowledge in Artificial Intelligence. *Social Studies of Science*, 23(3), 445–477.
- Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., Milojević, S., Petersen, A. M., Radicchi, F., Sinatra, R., Uzzi, B., Vespignani, A., Waltman, L., Wang, D., & Barabási, A.-L. (2018). Science of science. *Science*, 359(6379), 9.
- Foster, J. G., Rzhetsky, A., & Evans, J. A. (2015). Traditions and Innovation in Scientists' Research Strategies. *American Sociological Review*, 80(5), 1–34.
- Franceschet, M., & Colavizza, G. (2020). Quantifying the higher-order influence of scientific publications. *Scientometrics*, 125(2), 951–963.
- Frank, M. R., Autor, D., Bessen, J. E., Brynjolfsson, E., Cebrian, M., Deming, D. J., Feldman, M., Groh, M., Lobo, J., Moro, E., Wang, D., Youn, H., & Rahwan, I. (2019).

- Toward understanding the impact of artificial intelligence on labor. *Proceedings of the National Academy of Sciences*, 116(14), 6531–6539.
- Frank, M. R., Wang, D., Cebrian, M., & Rahwan, I. (2019). The evolution of citation graphs in artificial intelligence research. *Nature Machine Intelligence*, 1(2), 79–85.
- Frégnac, Y. (2017). Big data and the industrialization of neuroscience: A safe roadmap for understanding the brain? *Science*, 358(6362), 470–477.
- Frickel, S. (2004). Building an Interdiscipline: Collective Action Framing and the Rise of Genetic Toxicology. *Social Problems*, 51(2), 269–287.
- Gaglio, G., & Loute, A. (2023). L'émergence d'enjeux éthiques lors d'expérimentations de logiciels d'intelligence artificielle. Le cas de la radiologie. *Réseaux*, 240(4), 145–178.
- Galison, P. (1997). *Image and Logic: A Material Culture of Microphysics*. Chicago, IL, University of Chicago Press.
- Ganascia, J.-G. (1993). *L'intelligence artificielle*. Flammarion.
- Gao, J., & Wang, D. (2023). Quantifying the Benefit of Artificial Intelligence for Scientific Research. arXiv. <https://arxiv.org/pdf/2304.10578.pdf>
- Gargiulo, F., Caen, A., Lambiotte, R., & Carletti, T. (2016). The classical origin of modern mathematics. *EPJ Data Science*, 5(1), 26.
- Gargiulo, F., Fontaine, S., Dubois, M., & Tubaro, P. (2023). A meso-scale cartography of the AI ecosystem. *Quantitative Science Studies*, 1–20.
- Gerovitch, S. (2001). 'Mathematical Machines' of the Cold War: Soviet Computing, American Cybernetics and Ideological Disputes in the Early 1950s. *Social Studies of Science*, 31(2), 253–287.
- Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., & Kagal, L. (2019). Explaining Explanations: An Overview of Interpretability of Machine Learning. arXiv. Retrieved May 31, 2022, from <http://arxiv.org/abs/1806.00069>
- Gingras, Y. (2016). *Bibliometrics and Research Evaluation. Uses and Abuses*. The MIT Press.
- Glänzel, W., & Schubert, A. (2003). A new classification scheme of science fields and subfields designed for scientometric evaluation purposes. *Scientometrics*, 56(3), 357–367.
- González-Márquez, R., Schmidt, L., Schmidt, B. M., Berens, P., & Kobak, D. (2024). The landscape of biomedical research. *Patterns*, 5(6), 100968.

- Gopinath, N. (2023). Artificial intelligence and neuroscience: An update on fascinating relationships. *Process Biochemistry*, *125*, 113–120.
- Graetz, G., & Michaels, G. (2018). Robots at Work. *The Review of Economics and Statistics*, *100*(5), 753–768.
- Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology*, *78*(6), 1360–1380.
- Grauwin, S., Beslon, G., Fleury, E., Franceschelli, S., Robardet, C., Rouquier, J.-B., & Jensen, P. (2012). Complex Systems Science: Dreams of Universality, Reality of Interdisciplinarity. *Journal of the American Society for Information Science and Technology*, *63*(7), 1327–1338.
- Griffith, B. C., & Mullins, N. C. (1972). Coherent Social Groups in Scientific Change: "Invisible colleges" may be consistent throughout science. *Science*, *177*(4053), 959–964.
- Griffith, B. C., Small, H., Stonehill, J. A., & Dey, S. (1974). The Structure of Scientific Literatures II: Toward a Macro- and Microstructure for Science. *Science Studies*, *4*(4), 339–365.
- Haas, P. M. (1992). Introduction: Epistemic Communities and International Policy Coordination. *International Organization*, *46*(1), 1–35.
- Hackett, E. J., Conz, D., Parker, J., Bashford, J., & DeLay, S. (2004). Tokamaks and turbulence: research ensembles, policy and technoscientific work. *Research Policy*, *33*(5), 747–767.
- Hackett, E. J., Parker, J. N., Vermeulen, N., & Penders, B. (2017). The Social and Epistemic Organization of Scientific Work. In *The Handbook of Science and Technology Studies* (Fourth edition, pp. 733–764). MIT Press.
- Hajkowicz, S., Naughtin, C., Sanderson, C., Schleiger, E., Karimi, S., Bratanova, A., & Bednarz, T. (2022). *Artificial intelligence for science. Adoption trends and future development pathways* (tech. rep.). CSIRO Data61. Brisbane, Australia.
- Hassabis, D., Kumaran, D., Summerfield, C., & Botvinick, M. (2017). Neuroscience-Inspired Artificial Intelligence. *Neuron*, *95*(2), 245–258.
- Haugeland, J. (1985). *Artificial intelligence: The very idea*. Cambridge, MIT Press.
- Hentschel, K. (2015). A periodization of research technologies and of the emergency of genericity. *Studies in History and Philosophy of Modern Physics*, *52*, 223–233.

- Huang, Y. [Ying], Schuehle, J., Porter, A. L., & Youtie, J. (2015). A systematic method to create search strategies for emerging technologies based on the Web of Science: illustrated for 'Big Data'. *Scientometrics*, *105*(3), 2005–2022.
- Huang, Y. [Yong], Bu, Y., Ding, Y., & Lu, W. (2018). Number versus structure: towards citing cascades. *Scientometrics*, *117*(3), 2177–2193.
- Jaton, F., & Vinck, D. (2016). Processus frictionnels de mises en bases de données. *Revue d'anthropologie des connaissances*, *11*(4), 489.
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389–399.
- Jordan, M. I. (2019). Artificial Intelligence—The Revolution Hasn't Happened Yet. *Harvard Data Science Review*, *1*(1).
- Joyce, K., Smith-Doerr, L., Alegria, S., Bell, S., Cruz, T., Hoffman, S. G., Noble, S. U., & Shestakofsky, B. (2021). Toward a Sociology of Artificial Intelligence: A Call for Research on Inequalities and Structural Change. *Socius: Sociological Research for a Dynamic World*, *7*, 1–11.
- Katz, E., Levin, M. L., & Hamilton, H. (1963). Traditions of Research on the Diffusion of Innovation. *American Sociological Review*, *28*(2), 237–252.
- Kauffman, S. A. (2000). *Investigations*. Oxford University Press.
- Kerr, A., Barry, M., & Kelleher, J. D. (2020). Expectations of artificial intelligence and the performativity of ethics: Implications for communication governance. *Big Data & Society*, *7*(1), 1–12.
- Kirtchik, O. (2019). Qu'apprend-on des machines apprenantes ? *Zilsel*, *5*(1), 446–468.
- Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, *1*(1), 1–12.
- Klinger, J., Mateos-Garcia, J., & Stathoulopoulos, K. (2018). Deep learning, deep change? Mapping the development of the Artificial Intelligence General Purpose Technology. arXiv. <http://arxiv.org/abs/1808.06355>
- Klinger, J., Mateos-Garcia, J., & Stathoulopoulos, K. (2022). A narrowing of AI research? arXiv. Retrieved April 25, 2024, from <http://arxiv.org/abs/2009.10385>
- Knorr-Cetina, K. (1999). *Epistemic cultures: How the sciences make knowledge*. Cambridge, MA, USA, Harvard University Press.
- Köstler, L., & Ossewaarde, R. (2022). The making of AI society: AI futures frames in German political and media discourses. *AI & Society*, *37*(1), 249–263.

- Koval, I., Bône, A., Louis, M., Lartigue, T., Bottani, S., Marcoux, A., Samper-González, J., Burgos, N., Charlier, B., Bertrand, A., Epelbaum, S., Colliot, O., Allassonnière, S., & Durrleman, S. (2021). AD Course Map charts Alzheimer's disease progression. *Scientific Reports*, *11*(1), 8020.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, *25*.
- Kuhn, T. S. (2018). *La structure des révolutions scientifiques* (L. Meyer, Trans.; Flammarion).
- Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building Machines That Learn and Think Like People. *Behavioral and Brain Sciences*, *40*(e253), 1–72.
- Latour, B., & Woolgar, S. (1996). *La vie de laboratoire. La production des faits scientifiques* (M. Biezunski, Trans.). Editions La Découverte.
- Laudel, G., & Gläser, J. (2007). Interviewing Scientists. *Science, Technology & Innovation Studies*, *3*(2), 91–111.
- Lazer, D., Pentland, A., Adamic, L., Barabási, A.-L., Brewen, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., & Van Alstyne, M. (2009). Computational Social Science. *Science*, *323*(5915), 721–723.
- Le Ludec, C., Wahal, E., Casilli, A., & Tubaro, P. (2020). Quel statut pour les petits doigts de l'intelligence artificielle? Présent et perspectives du micro-travail en France. *Les Mondes du Travail*, (24-25).
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.
- Lemaine, G. (1980). Science normale et science hypernormale: Les stratégies de différenciation et les stratégies conservatrices dans la science. *Revue Française de Sociologie*, *21*(4), 499.
- Léon, J. (1996). Histoire des sciences cognitives : les filiations multiples - À propos de : Aline Péliissier et Alain Tête, prés, et Trad. Sciences cognitives ; Textes fondateurs (1943- 1950). Wiener, Rosenblueth, Bigelow, McCulloch, Pitts, von Neumann, Hebb, Weaver, Shannon, Turing. *Histoire Epistémologie Langage*, *18*(2), 206–210.
- Leydesdorff, L. (1998). Theories of citation? *Scientometrics*, *43*(1), 5–25.

- Leydesdorff, L. (2006). Can scientific journals be classified in terms of aggregated journal-journal citation relations using the Journal Citation Reports? *Journal of the American Society for Information Science and Technology*, 57(5), 601–613.
- Leydesdorff, L., & Rafols, I. (2009). A global map of science based on the ISI subject categories. *Journal of the American Society for Information Science and Technology*, 60(2), 348–362.
- Li Vigni, F. (2021a). Cinq types de travail scientifique « interdisciplinaire ». *Natures Sciences Sociétés*, 29(2), 130–140.
- Li Vigni, F. (2021b). Complexity Sciences: A Scientific Platform. *Science & Technology Studies*.
- Li Vigni, F., Louvel, S., & Raimbault, B. (2023). How do the vectors of scientific credibility reconfigure at the interface between the social worlds? *Revue d'anthropologie des connaissances*, 17(2).
- Lindsay, R. K., Buchanan, B. G., Feigenbaum, E. A., & Lederberg, J. (1993). DENDRAL: A case study of the first expert system for scientific hypothesis formation. *Artificial Intelligence*, 61(2), 209–261.
- Liu, F., Zhang, S., & Xia, H. (2024). Science as exploration in a knowledge landscape: tracing hotspots or seeking opportunity? *EPJ Data Science*, 13(1), 1–20.
- Liu, H., Chen, Z., Tang, J., Zhou, Y., & Liu, S. (2020). Mapping the technology evolution path: a novel model for dynamic topic detection and tracking. *Scientometrics*, 125(3), 2043–2090.
- Liu, N., Shapira, P., & Yue, X. (2021). Tracking developments in artificial intelligence research: constructing and applying a new search strategy. *Scientometrics*, 126(4), 3153–3192.
- Liu, Z. (2021). Sociological perspectives on artificial intelligence: A typological reading. *Sociology Compass*, 15(3), e12851.
- Lobbé, Q., Delanoë, A., & Chavalarias, D. (2021). Exploring, browsing and interacting with multi-level and multi-scale dynamics of knowledge. *Information Visualization*, 21(1), 17–37.
- Loukides, M., & Lorica, B. (2019). How new tools in data and AI are being used in health care and medicine. Retrieved November 9, 2020, from <https://www.oreilly.com/radar/how-new-tools-in-data-and-ai-are-being-used-in-health-care-and-medicine/>

- Louvel, S. (2015). Ce que l'interdisciplinarité fait aux disciplines. Une enquête sur la nanomédecine en France et en Californie. *Revue française de sociologie*, *56*(1), 75.
- Macpherson, T., Churchland, A., Sejnowski, T., DiCarlo, J., Kamitani, Y., Takahashi, H., & Hikida, T. (2021). Natural and Artificial Intelligence: A brief introduction to the interplay between AI and neuroscience research. *Neural Networks*, *144*, 603–613.
- Manzo, G. (2010). Analytical Sociology and Its Critics. *Archives Européenne de Sociologie*, *51*(1), 129–170.
- Manzo, G., Gabbriellini, S., Roux, V., & M'Mbogori, F. N. (2018). Complex Contagions and the Diffusion of Innovations: Evidence from a Small-N Study. *Journal of Archaeological Method and Theory*, *25*(4), 1109–1154.
- March, J., G. (1991). Exploration and Exploitation in Organizational Learning. *Organization Science*, *2*(1), 71–87.
- Marcovich, A., & Shinn, T. (2011). Where is disciplinarity going? Meeting on the borderland. *Social Science Information*, *50*(3-4), 582–606.
- Marcovich, A., & Shinn, T. (2012). Regimes of science production and diffusion: towards a transverse organization of knowledge. *Scientiae Studia*, *10*(spe), 33–64.
- Marcovich, A., & Shinn, T. (2017). How scientific research instruments change: A century of Nobel Prize physics instrumentation. *Social Science Information*, *56*(3), 348–374.
- Martin, L., Muller, B., Suárez, P. J. O., Dupont, Y., Romary, L., de la Clergerie, É. V., Seddah, D., & Sagot, B. (2020). CamemBERT: a Tasty French Language Model, In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*.
- Matuchansky, C. (2019). Intelligence clinique et intelligence artificielle. Une question de nuance. *Médecine/sciences*, *35*(10), 797–803.
- McCarthy, J. (1981). Epistemological Problems of Artificial Intelligence. In *Readings in Artificial Intelligence* (pp. 459–465). Elsevier.
- McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. *AI Magazine*, *27*(4), 3.
- McCulloch, W., & Pitts, W. H. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, *5*, 115–133.
- McInnes, L., Healy, J., & Astels, S. (2017). HDBSCAN: Hierarchical density based clustering. *The Journal of Open Source Software*, *2*(11), 205.

- McInnes, L., Healy, J., & Melville, J. (2020). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv. <http://arxiv.org/abs/1802.03426>
- Menczer, F., Fortunato, S., & Davis, C. A. (2020). *A First Course in Network Science*. Cambridge University Press.
- Merton, R. K. (1968). The Matthew Effect in Science. *Science*, 159.
- Meyer, B., Choppy, C., Staunstrup, J., & van Leeuwen, J. (2009). Research evaluation for computer science. *Communications of the ACM*, 52(4), 31–34.
- Miceli, M., Tubaro, P., Casilli, A. A., Bonniec, T. L., Wagner, C. S., & Sachenbacher, L. (2024). *Who Trains the Data for European Artificial Intelligence?* (Tech. rep.). European Microworkers Communication and Outreach Initiative (EnCOre, 2023-2024). European Parliament; The Left.
- Mignot, L., & Schultz, É. (2022). Les innovations d'intelligence artificielle en radiologie à l'épreuve des régulations du système de santé. *Réseaux*, 232-233(2), 65–97.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. arXiv. <http://arxiv.org/abs/1301.3781>
- Milojević, S. (2020). Practical method to reclassify Web of Science articles into unique subject categories and broad disciplines. *Quantitative Science Studies*, 1(1), 183–206.
- Monechi, B., Ruiz-Serrano, Á., Tria, F., & Loreto, V. (2017). Waves of novelties in the expansion into the adjacent possible. *PLoS ONE*, 12(6).
- Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics*, 106(1), 213–228.
- Mukherjee, S., Uzzi, B., Jones, B., & Stringer, M. (2015). A New Method for Identifying Recombinations of Existing Knowledge Associated with High-Impact Innovation. *Journal of Product Innovation Management*, 33(2), 224–236.
- The new NeuroAI. (2024). *Nature Machine Intelligence*, 6(3), 245–245.
- Newman, M. (2001). The structure of scientific collaboration networks. *PNAS*, 98(2), 404–409.
- Olazaran, M. (1996). A Sociological Study of the Official History of the Perceptrons Controversy. *Social Studies of Science*, 26(3), 611–659.
- OurResearch. (2024). *OpenAlex: End-to-End Process for Topic Classification* (tech. rep.).

- Patel, V. L., Shortliffe, E. H., Stefanelli, M., Szolovits, P., Bellazzi, R., & Abu-Hanna, A. (2009). The Coming of Age of Artificial Intelligence in Medicine. *Artificial Intelligence in Medicine*, 46(1), 18.
- Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global Vectors for Word Representation, In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, Association for Computational Linguistics.
- Perconti, P., & Plebe, A. (2020). Deep learning and cognitive science. *Cognition*, 203, 104365.
- Perrier, R., Schawe, H., & Hernández, L. (2024). Phase coexistence in the fully heterogeneous Hegselmann–Krause opinion dynamics model. *Scientific Reports*, 14(1), 241.
- Pesapane, F., Codari, M., & Sardanelli, F. (2018). Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine. *European Radiology Experimental*, 2(35), 10.
- Polcumpally, A. T. (2021). Artificial intelligence and global power structure: understanding through Luhmann’s systems theory. *AI & Society*.
- Priem, J., Piwowar, H., & Orr, R. (2022). OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts, In *From Global Indicators to Local Applications*, Granada, Spain.
- Raimbault, B., Cointet, J.-P., & Joly, P.-B. (2016). Mapping the Emergence of Synthetic Biology. *PLOS ONE*, 11(9), e0161522.
- Raimbault, B., & Joly, P.-B. (2021). The Emergence of Technoscientific Fields and the New Political Sociology of Science. In K. Kastenhofer & S. Molyneux-Hodgson (Eds.), *Community and Identity in Contemporary Technosciences* (pp. 85–106). Springer.
- Rioufreyt, T. (2016). La transcription d’entretiens en sciences sociales. Enjeux, conseils et manières de faire. <https://shs.hal.science/halshs-01339474/>
- Robertson, T. S. (1967). The Process of Innovation and the Diffusion of Innovation. *Journal of Marketing*, 31(1), 14–19.
- Rogers, E. M. (1983). *Diffusion of innovations (3rd ed.)* New York, New York: Free Press of Glencoe.
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In N. Navab, J. Hornegger, W. Wells, & A. Frangi

- (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (pp. 234–241). Springer, Cham.
- Rosen, W., Mohs, R., & Davis, K. (1984). A new rating scale for Alzheimer’s disease. *The American journal of psychiatry*, *141*(11), 1356–64.
- Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, *65*(6), 386–408.
- Rosental, C. (2002). De la démo-cratie en Amérique. Formes actuelles de la démonstration en intelligence artificielle. *Actes de la recherche en sciences sociales*, *141-142*(1), 110–120.
- Rosental, C. (2003). Certifying Knowledge: The Sociology of a Logical Theorem in Artificial Intelligence. *American Sociological Review*, *68*(4), 623–644.
- Roth, C. (2008). Réseaux épistémiques : formaliser la cognition distribuée. *Sociologie du Travail*, *50*(3), 353–371.
- Roth, C., & Cointet, J.-P. (2010). Social and semantic coevolution in knowledge networks. *Social Networks*, *32*(1), 16–29.
- Rule, A., Cointet, J.-P., & Bearman, P. S. (2015). Lexical shifts, substantive changes, and continuity in State of the Union discourse, 1790–2014. *Proceedings of the National Academy of Sciences*, *112*(35), 10837–10844.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536.
- Sartori, L., & Bocca, G. (2023). Minding the gap(s): public perceptions of AI and socio-technical imaginaries. *AI & Society*, *38*(2), 443–458.
- Savage, N. (2022). Breaking into the black box of artificial intelligence. *Nature*.
- Schäfermeier, B., Hirth, J., & Hanika, T. (2022). Research topic flows in co-authorship networks. *Scientometrics*.
- Schawe, H., & Hernández, L. (2020). When open mindedness hinders consensus. *Scientific Reports*, *10*(1).
- Schuchmann, S. (2019). History of the Second AI Winter. Retrieved April 13, 2023, from <https://towardsdatascience.com/history-of-the-second-ai-winter-406f18789d45>
- Schulz, C., Mazlounian, A., Petersen, A. M., Penner, O., & Helbing, D. (2014). Exploiting citation networks for large-scale author name disambiguation. *EPJ Data Science*, *3*(1), 11.

- Schwartz, R. D. (1989). Artificial Intelligence as a Sociological Phenomenon. *The Canadian Journal of Sociology / Cahiers canadiens de sociologie*, 14(2), 179–202.
- Sedooka, A., Steffen, G., Paulsen, T., & Darbellay, F. (2015). Paradoxe identitaire et interdisciplinarité : un regard sur les identités disciplinaires des chercheurs. *Natures Sciences Sociétés*, 23(4), 367–377.
- Shinn, T. (1980). Orthodoxy and innovation in science: The atomist controversy in French chemistry. *Minerva*, 18(4), 539–555.
- Shinn, T., & Joerges, B. (2002). The Transverse Science and Technology Culture: Dynamics and Roles of Research-technology. *Social Science Information*, 41(2), 207–251.
- Shortliffe, E. H., & Buchanan, B. G. (1975). A model of inexact reasoning in medicine. *Mathematical Biosciences*, 23(3), 351–379.
- Simon, H. A. (1962). The Architecture of Complexity. *Proceedings of the American Philosophical Society*, 106(6), 467–482.
- Sinatra, R., Wang, D., Deville, P., Song, C., & Barabási, A.-L. (2016). Quantifying the evolution of individual scientific impact. *Science*, 354(6312).
- Singh, C. K., Tupikina, L., Lécuyer, F., Starnini, M., & Santolini, M. (2024). Charting mobility patterns in the scientific knowledge landscape. *EPJ Data Science*, 13(1), 1–20.
- Small, H., & Griffith, B. C. (1974). The Structure of Scientific Literatures I: Identifying and Graphing Specialties. *Science Studies*, 4(1), 17–40.
- Small, H. G. (1978). Cited Documents as Concept Symbols. *Social Studies of Science*, 8(3), 327–340.
- Smith, B. C. (2019). *The promise of Artificial Intelligence: Reckoning and Judgement*. MIT Press.
- Söderström, K. R., Åström, F., & Hallonsten, O. (2022). Generic instruments in a synchrotron radiation facility. *Quantitative Science Studies*, 3(2), 420–442.
- Stokols, D., Hall, K. L., Taylor, B. K., & Moser, R. P. (2008). The Science of Team Science : Overview of the Field and Introduction to the Supplement. *American Journal of Preventive Medicine*, 35(2), S77–S89.
- Sugimoto, C. R., & Cronin, B. (2012). Biobibliometric profiling: An examination of multifaceted approaches to scholarship. *Journal of the American Society for Information Science and Technology*, 63(3), 450–468.

- Sugimoto, C. R., & Weingart, S. (2015). The Kaleidoscope of Disciplinarity. *Journal of Documentation*, 71(4), 775–794.
- Sun, Y., & Latora, V. (2020). The evolution of knowledge within and across fields in modern physics. *Scientific Reports*, 10(1), 12097.
- Tantardini, M., Ieva, F., Tajoli, L., & Piccardi, C. (2019). Comparing methods for comparing networks. *Scientific Reports*, 9(1), 17557.
- Terry-Jack, M. (2019). Deep Learning: Convolutional Neural Networks (CNNs). Retrieved August 6, 2024, from <https://medium.com/@b.terryjack/deep-learning-convolutional-neural-networks-cnns-7f7f16341df1>
- The White House. (2019). Artificial Intelligence for the American People. Retrieved November 29, 2021, from <https://trumpwhitehouse.archives.gov/ai/>
- Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44–56.
- Tsitsulin, A., Mottin, D., Karras, P., Bronstein, A., & Müller, E. (2018). NetLSD: Hearing the Shape of a Graph, In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*.
- Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013). Atypical Combinations and Scientific Impact. *Science*, 342(6157), 468–472.
- Valliani, A. A.-A., Ranti, D., & Oermann, E. K. (2019). Deep Learning and Neurology: A Systematic Review. *Neurology and Therapy*, 8(2), 351–365.
- Van der Linden, M., Coyette, F., Poitrenaud, J., Kalafat, M., Calicis, F., Wyns, C., Adam, S., & Membres du GREMEM. (2004). L'épreuve de rappel libre/rappel indicé à 16 items (RL/RI-16). In *L'évaluation des troubles de la mémoire : Présentation de quatre tests de mémoire épisodique (avec leur étalonnage)* (Marseille : Solal).
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need. arXiv. <http://arxiv.org/abs/1706.03762>
- Vayre, J.-S. (2021). Intelligence artificielle : entre science et marché. Quelques éléments sociohistoriques pour mieux comprendre une étrange expérimentation scientifique (1956-1990). *Annales des Mines - Gérer et comprendre*, 145(3), 55–69.
- Vieira, S., Pinaya, W. H. L., & Mechelli, A. (2017). Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications. *Neuroscience & Biobehavioral Reviews*, 74, 58–75.

- Villani, C., Shoenauer, M., Bonnet, Y., Berthet, C., Cornut, A.-C., Levin, F., & Rondepierre, B. (2018). *Donner un sens à l'intelligence artificielle : pour une stratégie nationale et européenne* (Rapport public). Rapport de la mission confiée par le Premier Ministre. Paris.
- Visser, M., van Eck, N. J., & Waltman, L. (2021). Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. *Quantitative Science Studies*, 2(1), 20–41.
- Voskuhl, A. (2004). Humans, Machines, and Conversations: An Ethnographic Study of the Making of Automatic Speech Recognition Technologies. *Social Studies of Science*, 34(3), 393–421.
- Wainer, J., Eckmann, M., Goldenstein, S., & Rocha, A. (2013). How productivity and impact differ across computer science subareas. *Communications of the ACM*, 56(8), 67–73.
- Wallace, M. L., Larivière, V., & Gingras, Y. (2012). A Small World of Citations? The Influence of Collaboration Networks on Citation Practices. *PLOS ONE*, 7(3), e33339.
- Weaver, R. R. (1986). Some implications of the emergence and diffusion of medical expert systems. *Qualitative Sociology*, 9(3), 237–255.
- Winter, P. D., & Carusi, A. (2022). (De)troubling transparency: artificial intelligence (AI) for clinical applications. *Medical Humanities*, 1–10.
- WIPO. (2019). *WIPO Technology Trends 2019: Artificial Intelligence* (tech. rep.). World Intellectual Property Organization. Geneva.
- Woolgar, S. (1985). Why not a sociology of machine? The case of sociology and artificial intelligence. *Sociology*, 19(4).
- Wray, K. B. (2005). Rethinking Scientific Specialization. *Social Studies of Science*, 35(1), 151–164.
- Xu, Y., Liu, X., Cao, X., Huang, C., Liu, E., Qian, S., Liu, X., Wu, Y., Dong, F., Qiu, C.-W., Qiu, J., Hua, K., Su, W., Wu, J., Xu, H., Han, Y., Fu, C., Yin, Z., Liu, M., . . . Zhang, J. (2021). Artificial intelligence: A powerful paradigm for scientific research. *The Innovation*, 2(4), 100179.
- Zeng, A., Shen, Z., Zhou, J., Fan, Y., Di, Z., Wang, Y., Stanley, H. E., & Havlin, S. (2019). Increasing trend of scientists to switch between topics. *Nature Communications*, 10(1), 3439.

- Zhai, Y., Ding, Y., & Wang, F. (2017). Measuring the diffusion of an innovation: A citation analysis. *Journal of the Association for Information Science and Technology*, 69(3), 368–379.
- Zhang, A., Lipton, Z. C., Li, M., & Smola, A. J. (2021). Dive into Deep Learning. arXiv. <https://arxiv.org/abs/2106.11342>
- Zuckerman, H., & Merton, R. K. (1972). Age, Aging and Age Structure in Science. In M. White Riley, M. Johnson, & A. Foner (Eds.), *A Theory of Age Stratification*. New York, Russell Sage Foundation.

English abstract

Over the past decade, Artificial Intelligence (AI) has been increasingly used by a wide range of scientific disciplines to advance their respective research agendas. This phenomenon is the subject of ongoing discussions and criticism, particularly with regard to its potential to supplant other tools traditionally used in these disciplines, and to create new knowledge capable of revolutionizing already well-established theoretical paradigms. In this context, this thesis examines the epistemic impact of AI in scientific research, and specifically in a multidisciplinary field, neuroscience. By identifying and explaining the mechanisms of integration and diffusion of AI within the latter, this thesis aims to determine the extent to which AI is becoming pervasive (or not) in the production of knowledge specific to this research domain. Using a multiscale approach, this thesis is divided into two parts.

Firstly, a numerical analysis of bibliometric data representing the neuroscience literature between 1970 and 2020 demonstrates that AI is applicable to a vast array of topics covered by neuroscience in these years, but fails to merge with the fundamental knowledge that structures the core of their respective conceptual frameworks, and therefore to become a central part of knowledge creation or entire paradigms upheaval. Thus, we argue that AI is *generic* in neuroscience in terms of applications, but not conceptualization. The study of this corpus also reveals that only a small number of scientists, mostly mathematicians, computer scientists, or engineers segregated at the periphery of the entire neuroscience collaboration network, employ AI in their publications. In particular, they tend to publish mostly together and little alongside neuroscientists engaged primarily in medical research, who nevertheless represent the core of the professionals listed in our database.

Secondly, by mobilizing data from semi-structured interviews conducted in 2021 with members of a clinical research team specializing in neuroimaging and computational neuroscience, we demonstrate the interdisciplinary nature of AI development, which involves various collaborators from computer science and medical research. We distinguish two elements. First, AI specialists, as the initial broadcasters of this research tool, are required to align their practices with the expectations of healthcare professionals, and thus transform the research practices they may have acquired in the past before joining the team. Second, although its performance remains limited or comparable to tools that are already used for the scientific tasks carried out in the team, AI is rapidly adopted by all the members of the team, whether they are computer scientists or clinicians, as it contributes to enhance the visibility of the output of their research to other scientists and institutions within academic and industrial realms.

The combination of these two qualitative and quantitative approaches finally demonstrates that AI is far from replacing all the research tools used by neuroscientists, due to its current limited genericity.

Keywords: *Artificial Intelligence, Neuroscience, Research-technology, Genericity, Science and Technology Studies, Computational Social Sciences, Complex Systems*

French abstract

Depuis la dernière décennie, l'intelligence artificielle (IA) se révèle être un outil de choix dans la recherche scientifique, comme en témoigne son utilisation toujours plus accrue dans presque toutes les disciplines. Cette dernière suscite de nombreuses réflexions et critiques, notamment sur sa capacité à supplanter d'autres outils traditionnellement mobilisés dans ces disciplines et à créer de nouvelles connaissances capables de révolutionner des paradigmes théoriques déjà bien établis. Dans ce contexte, cette thèse s'intéresse à l'impact épistémique de l'IA dans la recherche scientifique, et plus particulièrement dans un domaine multi-disciplinaire, les neurosciences. À travers l'identification et l'explication des mécanismes d'intégration et de diffusion de l'IA au sein de ces dernières, cette thèse vise notamment à établir dans quelle mesure elle devient omniprésente (ou non) dans la production des savoirs propres à ce domaine de recherche. S'appuyant sur une méthode multi-échelle, cette thèse se divise en deux temps.

Premièrement, au moyen d'une analyse numérique de données bibliométriques représentant la littérature neuroscientifique entre 1970 et 2020, nous montrons que l'IA s'applique bel et bien à un vaste ensemble de thématiques portées par les neurosciences entre 1970 et 2020 (*généricité en application*), mais peine à s'articuler avec les cadres théoriques

majeurs de ces dernières (*généricité en conceptualisation*), ne lui permettant donc pas de bousculer des paradigmes entiers. L'étude de ce corpus révèle également que quelques scientifiques bien particuliers recourent à l'IA dans leurs écrits : formés et évoluant au sein des mathématiques, de l'informatique ou de divers pans de l'ingénierie, ces derniers publient majoritairement ensemble et peu aux côtés de neuroscientifiques plutôt impliqués dans la recherche médicale et constituant pourtant le cœur des professionnels recensés dans notre base de données.

Deuxièmement, en mobilisant des données d'entretiens semi-directifs réalisés en 2021 auprès de membres d'une équipe de recherche clinique spécialisée dans la neuro-imagerie et les neurosciences computationnelles, nous montrons une réalité du travail interdisciplinaire de construction de l'IA par divers acteurs des recherches informatique et médicale, et ce à travers deux éléments. D'abord, les spécialistes de l'IA, premiers diffuseurs de cet instrument de recherche, doivent se conformer aux attendus des professionnels de santé et ainsi transformer leurs pratiques de recherche qu'ils ont pu acquérir dans le passé avant de rentrer dans l'équipe. Ensuite, l'IA se retrouve progressivement adoptée par l'ensemble des informaticiens et cliniciens affiliés à cette dernière, car elle participe à l'accroissement de la visibilité des recherches produites auprès d'autres scientifiques et institutions représentant les mondes académique et industriel, et ce malgré des performances encore limitées ou similaires à des outils déjà utilisés pour les tâches scientifiques à accomplir dans l'équipe.

Ces deux approches qualitatives et quantitatives montrent finalement que l'IA est loin de remplacer l'ensemble des outils de recherche utilisés par les neurosciences, en raison de sa *généricité* encore limitée.

Mots-clefs : *Intelligence Artificielle, Neurosciences, Recherche technologique, Généricité, Étude des sciences et technologies, Sciences sociales computationnelles, Systèmes complexes*

La diffusion de l'intelligence artificielle dans les neurosciences. Une approche multi-échelle de la généricité d'une recherche technologique

Depuis la dernière décennie, l'intelligence artificielle (IA) se révèle être un outil de choix dans la recherche scientifique, comme en témoigne son utilisation toujours plus accrue dans presque toutes les disciplines. Considérant cette dernière comme une recherche technologique particulière dans les sciences, cette thèse s'intéresse à l'impact épistémique de l'IA dans un domaine multi-disciplinaire, les neurosciences. S'appuyant sur une méthode multi-échelle, recourant à une analyse de données bibliométriques représentant la littérature neuroscientifique entre 1970 et 2020, ainsi qu'à des données d'entretiens semi-directifs réalisés en 2021 auprès de membres d'une équipe de recherche clinique spécialisée dans la neuro-imagerie et les neurosciences computationnelles, cette thèse vise à établir dans quelle mesure l'IA devient omniprésente (ou non), voire générique, dans la production des savoirs propres aux neurosciences, et ce à travers l'identification et l'explication de quelques mécanismes d'intégration et de diffusion de l'IA au sein de ces dernières. Ces deux approches qualitatives et quantitatives montrent finalement que l'IA est loin de remplacer l'ensemble des outils de recherche utilisés par les neurosciences, en raison de sa généricité encore limitée.

Mots-clefs : *Intelligence Artificielle, Neurosciences, Recherche technologique, Généricité, Étude des sciences et technologies, Sciences sociales computationnelles, Systèmes complexes*

The diffusion of Artificial Intelligence into Neuroscience. A multiscale approach of the genericity of a research-technology

Over the past decade, Artificial Intelligence (AI) has been increasingly used by a wide range of scientific disciplines to advance their respective research agendas. Conceptualizing AI as a particular research-technology in science, this thesis examines its epistemic impact in a multi-disciplinary field, neuroscience. Using a multiscale approach that mobilizes bibliometric data representing the neuroscience literature between 1970 and 2020 and data from semi-structured interviews conducted in 2021 with members of a clinical research team specializing in neuroimaging and computational neuroscience, this thesis aims to determine the extent to which AI is becoming pervasive (or not), even generic, in the production of knowledge specific to neuroscience. To this end, this thesis identifies and explains some mechanisms of integration and diffusion of AI within this research domain. The combination of these two qualitative and quantitative approaches finally demonstrates that AI is far from replacing all the research tools used by neuroscientists, due to its current limited genericity.

Keywords: *Artificial Intelligence, Neuroscience, Research-technology, Genericity, Science and Technology Studies, Computational Social Sciences, Complex Systems*