

RESEARCH

Open Access



Epistemic integration and social segregation of AI in neuroscience

Sylvain Fontaine^{1*}, Floriana Gargiulo¹, Michel Dubois¹ and Paola Tubaro²

*Correspondence:
sylvain.fontaine@cnrs.fr

¹ GEMASS, CNRS-Sorbonne Université, 59-61 rue Pouchet, 75017 Paris, France

² CREST, CNRS-IPP, 5 avenue Henry Le Chatelier, 91120 Palaiseau, France

Abstract

In recent years, Artificial Intelligence (AI) shows a spectacular ability of insertion inside a variety of disciplines which use it for scientific advancements and which sometimes improve it for their conceptual and methodological needs. According to the transverse science framework originally conceived by Shinn and Joerges, AI can be seen as an instrument which is progressively acquiring a universal character through its diffusion across science. In this paper we address empirically one aspect of this diffusion, namely the penetration of AI into a specific field of research. Taking neuroscience as a case study, we conduct a scientometric analysis of the development of AI in this field. We especially study the temporal egocentric citation network around the articles included in this literature, their represented journals and their authors linked together by a temporal collaboration network. We find that AI is driving the constitution of a particular disciplinary ecosystem in neuroscience which is distinct from other subfields, and which is gathering atypical scientific profiles who are coming from neuroscience or outside it. Moreover we observe that this AI community in neuroscience is socially confined in a specific subspace of the neuroscience collaboration network, which also publishes in a small set of dedicated journals that are mostly active in AI research. According to these results, the diffusion of AI in a discipline such as neuroscience didn't really challenge its disciplinary orientations but rather induced the constitution of a dedicated socio-cognitive environment inside this field.

Keywords: Science of science, Artificial Intelligence, Neuroscience, Research-technology

Introduction

In recent years, Artificial Intelligence (AI) has been continuously spreading across science. The associated worldwide scientific production and the amount of dedicated funding programs for technological developments supported both by academia and industry, are spectacularly growing (Gao and Wang 2023; Baruffaldi et al. 2020; Liu et al. 2021), and the outputs of such research are touching a variety of disciplines that are more and more citing it (Frank et al. 2019; Arencibia-Jorge et al. 2022). Though commonly associated to disciplines such as mathematics, statistics and computer science, AI originated as an interdisciplinary research area, and is currently spreading to a growing number of fields (Gargiulo et al. 2023).

AI knowledge and tools are amenable to applications in such welcoming disciplines which use them for scientific advancements, and which sometimes adapt and improve them in turn for their use. Bianchini et al. (2022) studied especially such mutual improvement dynamics at a macroscopic scale of science by focusing on the diffusion of neural networks' concepts and tools in numerous disciplines. Qualified as a *general method of invention* by these authors because of its non-disciplinary status, AI is thus expanding the *adjacent possible* of the purposes of the concerned disciplines, and is therefore the source of innovation in the latter (Kauffman 2000; Monechi et al. 2017).

In this paper we address the process of penetration of AI inside a given field of research. We focus on neuroscience, which are part of such fields largely receiving AI, according to Gargiulo et al. (2023). We choose especially this field as it has a particular history which is closely intertwined with the AI one. Indeed neuroscience and AI are continuously providing feed-backs to each other toward the comprehension of the human brain and the artificial reproduction of some cognitive processes (Hassabis et al. 2017; Savage 2019). This lead us to also assume that neuroscience is not only a field that receives AI, but also contribute to reshape AI knowledge, tools and practices, thus expanding the *adjacent possible* of the latter.

Along these lines, we focus on two dynamical mechanisms underlying the appropriation of AI by neuroscience and then the possible knowledge transfer between the two fields:

- The cognitive embedding of AI into the multidisciplinary context of neuroscience:
 1. what is the disciplinary landscape around neuroscience research which is using AI?
 2. how does such research fit in the disciplinary objectives of neuroscience?
- The diffusion throughout the neuroscience community:
 3. who are the actors leading AI research in this field?
 4. does AI-related knowledge widespread throughout the whole scientific community?

In what follows we answer these questions by analyzing a large corpus of neuroscience articles published between 1970 and 2019, extracted with a journal-oriented query from the Microsoft Academic Knowledge Graph (MAG) (Färber 2019). We distinguish AI-related publications from others with a dedicated keyword filter applied to their titles and abstracts, as in Gargiulo et al. (2023). After discussing in Sect. 2 the intertwined history of AI and neuroscience, and also the main concepts mobilized in this paper, we detail in Sect. 3 the building of some relational structures and some metrics given by the corpus that will be analyzed afterwards, namely its temporal egocentric citation network (composed of bibliographical references and citations of each paper), its peer-reviewed journals, and its temporal collaboration network, featured with the disciplinary profiles of the involved scientists.

In Sects. 4.1 and 4.2, we identify within the citation network the disciplines that are shaping the AI research in neuroscience, and the extent to which they differ from

those that are characterizing other neuroscience's subfields unrelated to AI. We especially show that some disciplines related to STEM have become the main influence of AI research since the 1990s, by replacing progressively the conventionally used ones in neuroscience and commonly associated with biomedical and clinical research. The disciplines impacted by AI research over time are also shifting from neuroscience-related clinical and biomedical fields to STEM ones. These first results thus show that AI research is building itself as a subfield (or specialty) inside neuroscience, which draws upon a disciplinary basis that is different from the other neuroscience subfields. We also confirm this partial cognitive integration in Sect. 4.3, which reveals the spreading of AI in almost all the journals of the database, but with different patterns of promotion of such research. In particular, we show that the journals that are publishing a lot of AI-related works are invested by researchers who contributed to AI research in general (inside or outside neuroscience). By regarding the temporal co-authorship network of the field, we show in the following Sect. 4.4 that the AI community includes two main academic profiles: (1) researchers trained in neuroscience – mainly within the biomedical and clinical parts –, who publish few AI-related works, and (2) researchers trained in STEM-related disciplines and having moved to neuroscience, who contribute much more to AI research in the welcoming field than those described by profile (1). Putting all the AI practitioners together, we then show that they are more and more isolated in the collaboration network since 1970.

Finally, by considering AI as a *research-technology* spreading across science, as originally conceived by Shinn and Joerges (2002) and formalized later by Hentschel (2015), we discuss in Sect. 5 the partial diffusion observed empirically in our results. With this framework we also discuss the similarity and differences between the local development dynamics of AI within this particular field of research, and the global one within the whole science system that is depicted in Gargiulo et al. (2023).

Literature review

AI and neuroscience: toward the mutual expansion of their *adjacent possible*

Artificial Intelligence (AI) commonly refers to both a research program and, more generally, a set of complex computer-based programs which aim to mimic human mind processes with high reckoning power. Although its foundations are mainly associated with STEM disciplines, mainly mathematics, statistics and computer science (Gargiulo et al. 2023), AI is constantly evolving alongside neuroscience by maintaining a virtual circle of mutual improvement (Hassabis et al. 2017).

Indeed neuroscience brings in a first place empirical confirmations of 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 are at the roots of AI-related algorithms, notably the case of the bio-inspired AIs such as artificial neural networks and their numerous versions, which became *biologically plausible* with neuroscience. In particular, the 1980s and early 1990s marked also the launching of the first body-scanner machines applying positron emission tomography (PET) and functional magnetic resonance imaging (fMRI), both of which led to important discoveries on 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 brain sciences, even if some debates are persisting within cognitive science about the representations of knowledge and the logical operations to process them in interaction with the real world (McCarthy 1981; Andler 1990; Perconti and Plebe 2020)

In the recent *big data* era, a panel of AI tools enables the efficient processing of large datasets composed of various kinds of biomedical data (electroencephalogram, MRI, biomarkers tracking, movement recordings, psychological survey, etc.), acquired from important clinical trials and cohorts for studying brain damages (Gopinath 2023). They are especially part of the improvement of the diagnosis of various neuro-degenerative diseases and of the attribution of potential dedicated treatments, if they exist.

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

Conceiving AI as a *research-technology* in science

In a socio-historical perspective of science, Shinn and Joerges (2002) propose the notion of *research-technology*, called also *transverse science*, to describe the dynamics of science since the end of the 19th century, now largely based on instrumentation for experimental or empirical investigations, especially in physical and life sciences. Within such a research regime, the production of knowledge is conditioned by an instrument that is designed in a specific research environment before being disseminated outside the latter (Shinn and Joerges 2002; Marcovich and Shinn 2020). Such an instrument requires the contribution of a dedicated socio-cognitive workforce composed of a variety of actors (scientists, technicians, promoters, administrators,...), to develop a dedicated technological culture of it, ie. a set of associated knowledge and practices shared by everyone in the community, without necessarily claiming a common professional identity related to it. These criteria shape a social group that could be also 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 the established disciplinary boundaries in order to provide their expertise for the resolution of diverse scientific problems.

The research-technology associated with an instrument is thus a dynamical entity, with changing social and institutional organizations at different moments of its development. Based on Shinn and Joerges (2002), Hentschel (2015) proposes four stages to describe the life of a given research-technology in a historical perspective, ie. throughout a long time period: *prehistory*, *exploration*, *optimization* and *diffusion*.

The first two stages are often associated together in the designing and testing phases of a given instrument. According to these authors, the underlying process needs

temporarily a closure of the community working on it, thus fostering the creation of a dedicated scientific field, often interdisciplinary, in which the instrument is also a research object.

Shinn and Joerges (2002) also highlight the criterion of *genericity* of an instrument: at later stages of its development, it could be adapted for disciplinary research contexts other than those in which it has been originally designed, or for a variety of applications outside scientific research. The two last development steps of Hentschel are thus representing such dynamics toward this final step. In particular, the diffusion phase requires a relaxation of the disciplinary boundaries within which the instrument was designed during the prehistoric and exploratory phase, thus implying an openness of the actors in the community, who promote it in different fields of research within academia or industry. This is the case of technologies related to laser beams or X-ray emission, now used widely from experimental physics to medical research and standard practices, even in daily objects such as smartphones (Hentschel 2015).

According to the analysis of the AI research ecosystem provided by Gargiulo et al. (2023), AI seems to embrace some characteristics of an instrument embedded in a global research-technology program within science. Through the analysis of the development of AI in neuroscience, throughout a longer time frame (1970–2019), we discuss in Sect. 5 to what extent this framework applies in the case of the development and the diffusion of AI in a single field of research.

Data and methods

Extraction and preprocessing of the data

MAG is a scientometric database that is well used in *science of science*, and especially for drawing the main trends of research within AI research (Gargiulo et al. 2023), measuring their impact in different scientific disciplines (Gao and Wang 2023), identifying the main actors involved in such research (Frank et al. 2019) and how they are organized across the world (Tang et al. 2022), among others. Much bigger than the traditional databases such as the *Web of Science* (WOS) and *Scopus* (Visser et al. 2021), MAG covers not only the articles published in peer-reviewed journals, but also the publications within conference proceedings, books, preprints, etc., for a total of about 260 million scientific publications in all disciplines (in 2022).

MAG is organized as a big network of files containing different metadata to describe a single publication, such as title, abstract, publication date and publisher (Färber 2019). These metadata are embedded as strings or numerical identifiers in a set of dump files linked together by link files, all of them being available on the data deposit platform Zenodo Färber (2020).

To build an exhaustive database of the neuroscience literature from MAG, we draw upon a list of journals that we extracted from WOS, with the help of the database *SCImago Journal Rank* (SJR). We choose to use SJR in addition of WOS in order to extend the scope of neuroscience too narrow in the classification of the last one. Indeed the former is covering more neuroscience journals than the latter: in 2021, WOS referenced only 281 journals labeled as *Neuroscience* while SJR referenced 608 journals labeled as such. Moreover journals could be labeled differently by the two databases. Thus, with the help of SJR, we can retrieve in WOS some journals that are not labeled as

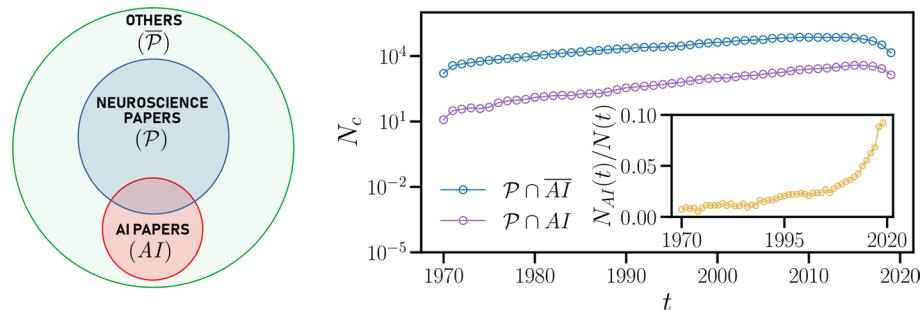


Fig. 1 Left: classification of the papers in the extracted corpus. The intersection between the blue and red zones corresponds to the set $\mathcal{P} \cap \text{AI}$ in the main text, while the blue zone excluded from this intersection represents the set $\mathcal{P} \cap \overline{\text{AI}}$. The green set $\overline{\mathcal{P}}$ includes all papers added after the building of the ego-centered citation network that are not published in neuroscience journal. Right: cumulative number N_c of publications in the two main subsets considered in this paper. The inset shows the instantaneous part of AI-related publications in the whole neuroscience corpus

Neuroscience in this database but that are in fact publishing papers treating neuroscience topics. For instance, the well-recognized journal *Neurocomputing*, while labeled with both *Neuroscience* and *Computer Science* in SJR, is labeled only as *Computer Science* by WOS. Using SJR thus contributes to increasing the diversity of neuroscience-related subjects treated by the articles in the final dataset under study.

After having extracted the set of journals labeled as *Neuroscience* by WOS, we add the set of journals also labeled as such by SJR, and remove duplicates occurring in both of them. We thus get 421 journals.

Here we focus only on peer-reviewed journals because of their easy availability in WOS database, although we are aware that AI research could be largely published in other publication media such as preprint archives and conference proceedings (Wainer et al. 2013).

Then we extract the publications from MAG with an ISSN identifier matching procedure based on this list of journals. These publications are brought together in a set denoted by \mathcal{P} in the following. In order to distinguish AI-related works in our neuroscience dataset, we apply a selection criterion such that they must include at least one AI-related keyword in their title or abstract (see the SI of Gargiulo et al. (2023) for the complete list of keywords). This subset is denoted $\mathcal{P} \cap \text{AI}$. All neuroscientific studies outside this set are called $\mathcal{P} \cap \overline{\text{AI}}$.

In the following we keep only the papers that were published in the period running from 1970 until 2019, and that have at least 10 references and at least 10 generated citations. Finally the sets $\mathcal{P} \cap \overline{\text{AI}}$ and $\mathcal{P} \cap \text{AI}$ include respectively 829,317 and 26,374 papers, spread over the period mentioned as illustrated in Fig. 1. Therefore, among all this research in neuroscience, only 3% contain AI-related keywords. The inset of this figure exhibits especially a slow growth of the share of AI-related publications in neuroscience at the very end of the 1980s, stabilizing around 1995 and followed by a very rapid growth from 2007 until 2019. This plateau around 1995 suggests a second, prolonged *AI winter* in neuroscience, characterized here by stable interest in AI research but not waning, unlike in other disciplines or research fields (Cardon et al. 2018; Schuchmann 2019). The following period of important growth of this share, started in 2007, suggests also the

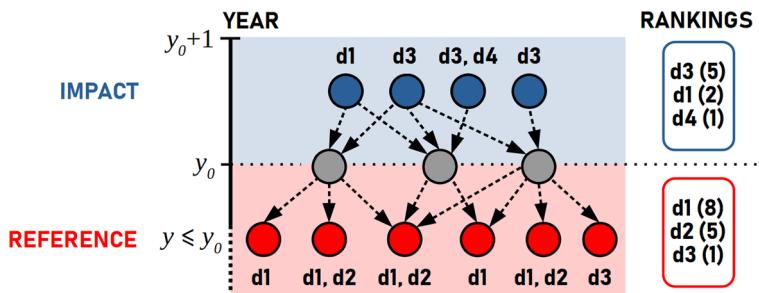


Fig. 2 Citation ecosystem centered here around three papers (grey) published at year y_0 and that are included either in $\mathcal{P} \cap AI$ or $\mathcal{P} \cap \bar{AI}$. One dashed arrow represents the citation of a target paper by a source one. Hence red points are the papers that are *cited* by the papers of our corpus (*reference*) while the blue ones are *citing* them (*impact*). The rankings are shown in decreasing order with the associated number of citations of each JSC d_i

main influence of the rise of deep learning techniques, a well-known trend shared by almost all the sciences (Cardon et al. 2018; Gargiulo et al. 2023). The share of AI publications in neuroscience reaches only 10% of the number of publications at its highest stage situated at the end of the studied period, which means that the use of the AI keywords of our list in neuroscience remains rather limited, even today.

Capturing the disciplinary landscape

Building of a citation ecosystem

After gathering the dataset under study, we build an exhaustive citation ecosystem around neuroscience publications, as illustrated in Fig. 2. We use the general reference link file of MAG – which is actually a giant citation network – for retrieving the references and citations of our identified papers. We thus add such papers citing or cited in those in our dataset, as long as they are published in a journal that belongs to the WOS journals database – we conserve here 84% of the original citation network centered on the neuroscience dataset \mathcal{P} . The added papers could therefore be published either in neuroscience publications, therefore in the set \mathcal{P} , or in others such that they join the set denoted by $\bar{\mathcal{P}}$, ie. the green border zone in Fig. 1. With the journal classification of WOS, we then enrich these sets of references and citations by assigning them a set of Journal Subject Categories (JSC) associated to their journals if they are referenced in this database, which is a synonym of a large well-recognized discipline in science.

Then, for each corpus $\mathcal{P} \cap AI$ and $\mathcal{P} \cap \bar{AI}$, we count the annual number of citations obtained by each JSCs represented in the global set of bibliographic references used by their papers. Doing this year by year over the period 1970–2019, we therefore build a temporal ranking of fields of study that are most cited by each corpus. For example, in Fig. 2, the papers cited by one corpus at year y_0 (grey points) are mainly published in journals labeled by the JSC d_1 with 8 citations, followed by the JSCs d_2 with 5 citations and d_3 with only one (red points and 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 following ones, depending on the citations received by each JSC at these years.

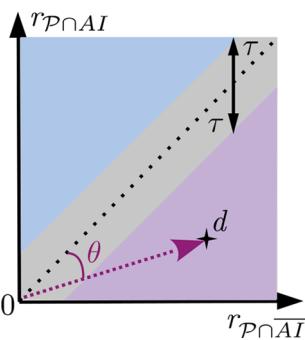


Fig. 3 Coordinate system to characterize the distribution of disciplines in the references and citations of the corpora $\mathcal{P} \cap \bar{\mathcal{A}I}$ and $\mathcal{P} \cap \mathcal{A}I$

We do the same for the JSCs that are most citing our papers of interest up to one year after their official publication year. We choose this temporal period to avoid the high time dispersion of the generated citations in the studied temporal period, the citation impact by the papers published in 1970 being eventually higher than those published in 2019.

In this way we assess from which JSCs are issued at first the generated citations before eventually spreading into a broader disciplinary landscape. Finally, by looking at both reference and impact side, we represent two disciplinary landscapes associated respectively with the corpora $\mathcal{P} \cap \bar{\mathcal{A}I}$ and $\mathcal{P} \cap \mathcal{A}I$.

Representation of the disciplinary landscape

Following the procedure described in Sect. 3.2.1, and according to all the couples of references' and citations' numbers associated with all the represented JSCs in $\mathcal{P} \cap \bar{\mathcal{A}I}$ and $\mathcal{P} \cap \mathcal{A}I$, we associate with each JSC d two pairs of ranks indicating its weight either in references or in citations of the two corpora at year t , denoted respectively as $r_R^d(t) = (r_{R,\mathcal{P} \cap \bar{\mathcal{A}I}}^d(t), r_{R,\mathcal{P} \cap \mathcal{A}I}^d(t))$ and $r_I^d(t) = (r_{I,\mathcal{P} \cap \bar{\mathcal{A}I}}^d(t), r_{I,\mathcal{P} \cap \mathcal{A}I}^d(t))$. In case of a missing JSC in one of the two corpora at a given year t , we fill the missing rank by a maximum value set at 100.

With one of such pair of ranks (either for references or citations), we can locate one discipline d in a 2D space of rankings, as shown in Fig. 3. In this coordinate system, the lower the value of one axis, the better the associated rank and the higher the number of collected citations along this 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 $\mathcal{P} \cap \mathcal{A}I$ and $\mathcal{P} \cap \bar{\mathcal{A}I}$ (in grey) as the region close to the diagonal where the two rankings of the fields of study inside it are almost the same and not significantly varying over time. This zone is comprised between the two lines of respective equations $r_{\mathcal{P} \cap \mathcal{A}I} = r_{\mathcal{P} \cap \bar{\mathcal{A}I}} + \tau$ (above the diagonal) and $r_{\mathcal{P} \cap \mathcal{A}I} = r_{\mathcal{P} \cap \bar{\mathcal{A}I}} - \tau$ (below the diagonal), with τ a parameter set as 10 in this paper.

The blue zone over the common interest area corresponds to the space where the disciplines have a better rank in $\mathcal{P} \cap \bar{\mathcal{A}I}$ than in $\mathcal{P} \cap \mathcal{A}I$. In the case of references, such disciplines in this zone are more cited by the corpus $\mathcal{P} \cap \bar{\mathcal{A}I}$ than by $\mathcal{P} \cap \mathcal{A}I$, and in the case

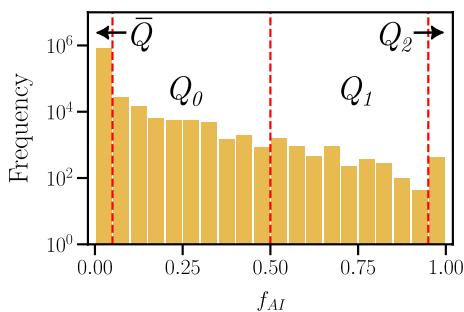


Fig. 4 Distribution of the AI score f_{AI} of the authors

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, ie. 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. 4.2, we apply this per decade periods instead of single years, ie. we build a decennial 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.

Capture use of AI by neuroscientists

Classification of authors according to their AI involvement

The dataset \mathcal{P} alone would not be sufficient for a study of social dynamics of scientists inside neuroscience. Indeed the latter would have had published in disciplines or fields of research other than neuroscience or in journals that are not labeled as such in WOS. We thus add another level of granularity by extracting all the missing publications that have been published by the authors who belong to \mathcal{P} . Notice that some of these publications can have been published in the neuroscience journals extracted in our list of 421 journals (see Sect. 3.1), since we focus only on the more impacting authors who have published papers with at least 10 generated citations (and having at least 10 references). From this extended set of papers we deduce the duration of the scientific life of each author, and we select first those who began after 1940. In addition, because of the poor disambiguation of some authors in the provided MAG dump-files, where some names are associated with centuries-long career and far too many publications, we select only the authors having a reasonable duration of academic life, here up to 50 years.

With these added data, we also capture the real activity (before 2019) of each author in AI research, not only in neuroscience but also in other fields of research by attributing them an AI score. We define it as the share of AI-related papers published before 2019, $f_{AI}(a) = n_a^{P \cap AI} / n_a^{tot}$. To prevent the unexpected effect of accumulation of scientists having an AI score equal to 1 due to a single publication in their very short career, we filter a second time the authors dataset by considering only those who have published at least 3 papers. We thus consider in the following 886,074 scientists, in which 188,325 (only 16% of the scientists) have published at least one AI-related paper in neuroscience. According to the f_{AI} distribution shown in Fig. 4, we divide this set of authors in four parts, namely \bar{Q} ($f_{AI} = 0$; $N_a = 697,749$), Q_0 ($f_{AI} \in (0, 0.5)$; $N_a = 182,925$), Q_1 ($f_{AI} \in [0.5, 1]$; $N_a = 4977$) and Q_2 ($f_{AI} = 1$; $N_a = 423$).

Collaboration network

We then build a temporal co-signature network TCN from the set of papers \mathcal{P} . Two authors i and j publishing together at year t are linked by an edge weighted by the number of common publications at this year $w_{ij}(t)$. All co-signatures in all papers published in year t shape a weighted snapshot called $\text{TCN}(t)$. In order to unveil the structure of the collaboration at macro-scale in the whole period 1970–2019, we build the weighted time-aggregated co-signature network ACN, 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 TCN, $E = \{\tilde{w}_{ij} \forall (i,j) \in V^2 | \tilde{w}_{ij} = \sum_t w_{ij}(t)\}$. ACN is finally composed of 871,282 nodes and 7,420,423 weighted edges, and is divided into 17,599 independent connected components, with a giant one including 93% of the nodes. The smallest components exhibit also a non-negligible representation of authors belonging to the quartiles Q_1 and Q_2 , respectively 25,5% and 26,5% of the population in these quartiles. Since the authors belonging to them are supposed to drive the AI research in neuroscience, we choose to keep all the communities for our analysis, even smaller than the giant component.

Here the collaboration network is built only among the most impacting scientists of our dataset. Indeed we draw on the one hand the collaboration network directly from the set of neuroscience papers that are the most impacting in terms of citations, and on the other hand we focus on the authors who published at least 3 articles up to 2019. Therefore we disregard the short-terms collaborations among very early-career scientists, short-career scientists or extra-academic ones, who are involved in one or two publications with too small impact but who can potentially drive forward innovation.

Intensity of collaborations between f_{AI} quartiles

In this section we detail the general computation of z-scores which will be used in Sect. 4.4.2 to indicate the propensity of collaboration in the time-aggregated collaboration network ACN.

We first consider a static collaboration network $G = \{S, E\}$ with S 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 Sect. 3.3.1, the distribution is divided in four parts $\{\bar{Q}, Q_0, Q_1, Q_2\}$, from the least to the most AI-expert. Therefore one author is belonging to one given group Q given its 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 one 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 $\{\tilde{G}_k\} = \{(S, \tilde{E}_k)\}$, with $k = \{1, \dots, N\}$, which are all based on the same set of scientists but have different sets of edges obtained by perturbing the network with a uniform shuffling which conserves the degree distribution of the scientists. By applying the same clustering of scientists in Q groups for each network as for the real network G , we draw a set of matrices $\{\tau_k\}$ corresponding to the share of edges between the Q groups within the randomized networks $\{\tilde{G}_k\}$. From this set we extract on the one hand the average matrix τ_{sim} where one 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 one 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 one element is defined as follows:

$$z(Q_i, Q_j) = \frac{\tau_{obs}(Q_i, Q_j) - \tau_{sim}(Q_i, Q_j)}{\sigma(Q_i, Q_j)}. \quad (1)$$

This standardization is used here to test the over-representation 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 through randomization. Applying the whole procedure before on the network ACN built in Sect. 3.3.2, with $N = 100$, gives the Fig. 13B. Figure 13A is also drawn by applying the same method with same N on each temporal snapshot of the temporal collaboration network TCN (see also Sect. 3.3.2), and by aggregating the groups Q_0, Q_1 and Q_2 into a single one called Q_i (see Fig. 13A) that includes all the scientists with at least one AI-related publication. Since TCN is undirected, the 4×4 matrices $\tau_{obs}, \tau_{sim}, \sigma$ and z are therefore reduced to scalars defined with only the two groups \bar{Q} and Q_i .

Results

Citation homogenisation of the AI research specialty 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, that are here represented as the AI-related publications and the non-AI ones within our neuroscience dataset.

After having divided the neuroscientific papers in two sub-corpora, one including the AI-related papers ($\mathcal{P} \cap AI$) and another including non-AI ones ($\mathcal{P} \cap \bar{AI}$), we build for each of them two disciplinary rankings which could vary over time, one for the disciplines appearing in their respective references ($r^R(t)$), called *reference ranking* in the following, and another for those appearing in the respective citations they have received ($r^I(t)$), called *citation ranking* in the following. The reference ranking associated with one given corpus summarizes its main influential fields of study, thus the disciplinary structure on which it draws over time, while the citation ranking indicates the fields that are impacted at first by the corpus (for more detail, see Sect. 3.2.1 above).

We then compare macroscopically at each year the reference rankings of the two corpora ($r_{\mathcal{P} \cap AI}^R(t)$ and $r_{\mathcal{P} \cap \bar{AI}}^R(t)$) and also their citation rankings ($r_{\mathcal{P} \cap AI}^I(t)$ and $r_{\mathcal{P} \cap \bar{AI}}^I(t)$) by using a common similarity metric J provided by Gargiulo et al. (2016, see their supplementary materials). This last measure, comprised between 0 and 1, evaluates how similar are two given rankings r_A and r_B : if $J(r_A, r_B) = 1$, the two rankings are exactly the same, ie. 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 , whatever their respective ranks.

Figure 5 shows the evolution of two such similarity measures, a first one for comparing the rankings of disciplines cited by the corpora $\mathcal{P} \cap AI$ and $\mathcal{P} \cap \bar{AI}$ (blue curve) and a second one for comparing the rankings of the disciplines that are citing them over time (red curve).

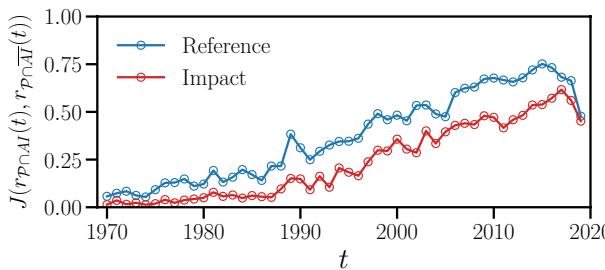


Fig. 5 Instantaneous similarity between the references' or citations' (also called impact) rankings of the two corpora $\mathcal{P} \cap AI$ and $\mathcal{P} \cap \bar{AI}$ at year t

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

Neuroscience and AI: two different disciplinary landscapes

We now study the difference between the disciplinary environments of the AI-related corpus ($\mathcal{P} \cap AI$) and the non-AI one ($\mathcal{P} \cap \bar{AI}$) in our neuroscience dataset. In particular, by considering the individual disciplines that are commonly cited by (or are citing) both of the two corpora, we compare to what extent these disciplines are actually cited by (are citing) each of them. The goal of such an analysis is to identify the disciplines from which each corpus prefers to find information, ie. their references, and those into which they spread their knowledge, ie. their citations.

By applying the framework described in Sect. 3.2, we associate with each represented discipline d in the references or citations of the corpora $\mathcal{P} \cap \bar{AI}$ and $\mathcal{P} \cap AI$, their time-aggregated ranks over decades, $(\tilde{r}_{\mathcal{P} \cap \bar{AI}}^d(T), \tilde{r}_{\mathcal{P} \cap AI}^d(T))$, with $T = [t_0, t_0 + 10]$ a given time period where $t_0 \in \{1970, 1980, 1990, 2000, 2010\}$. These ranks are built with the total numbers of references to the discipline d made by the corpora (references), or with the total numbers of citations given by d to the corpora (citations, or impact) during the time period T .

We then locate the cited or citing disciplines in the 2D space of rankings associated with each corpus respectively, as shown in Fig. 6. In one of the maps drawn on this figure, one colored disc represents a specific discipline d that is located with its respective ranks in each corpus during the considered time period, in references (Fig. 6A) or citations (Fig. 6B). Lower the value of one axis, better the associated rank, and then higher the number of collected citations along this axis. From these maps we define the common interest area between $\mathcal{P} \cap AI$ and $\mathcal{P} \cap \bar{AI}$ as the region close to the diagonal where the two rankings of the fields of study are almost the same and not significantly varying over time. These fields are colored in grey in these maps. The fields that are more

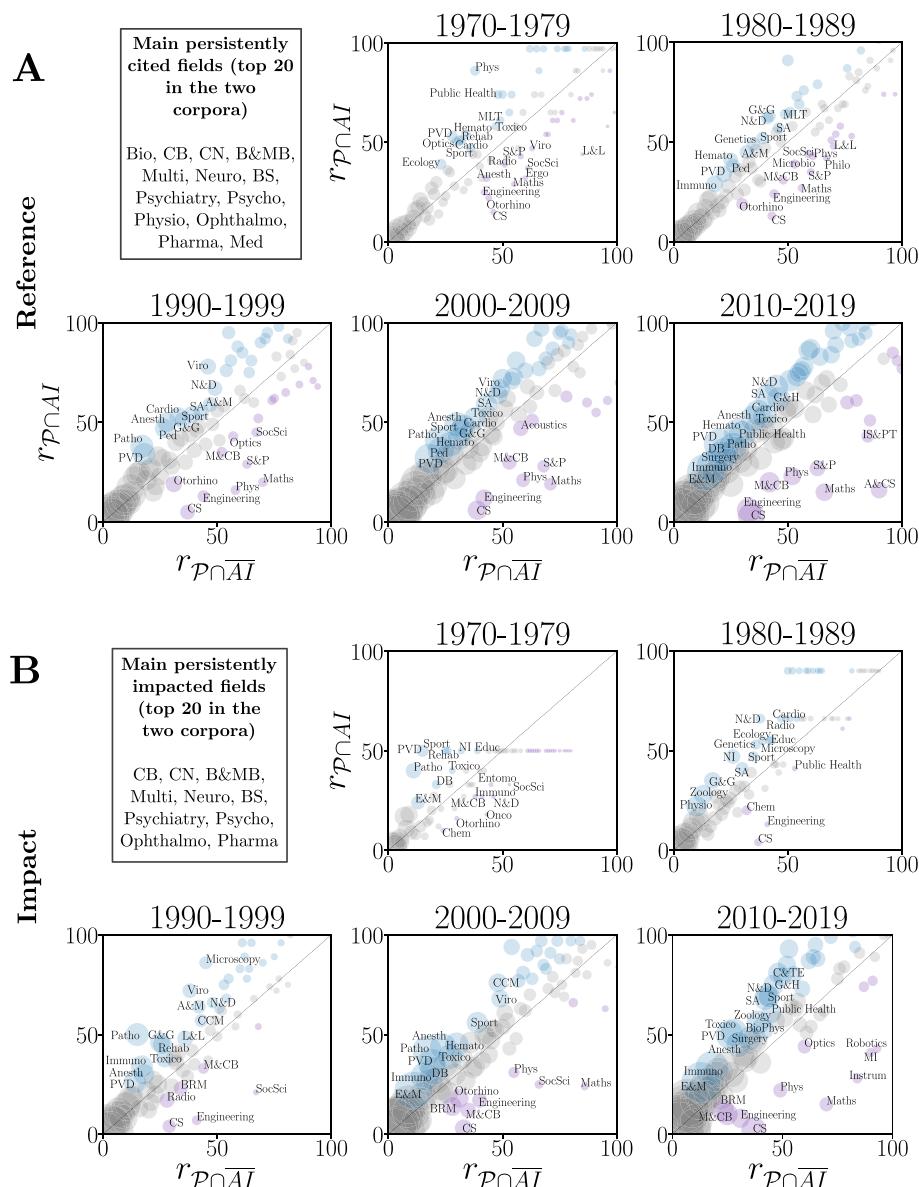


Fig. 6 Time-aggregated ranking maps of the fields of study involved in publications *cited* by both of the studied corpora $\mathcal{P} \cap \mathcal{A}I$ and $\mathcal{P} \cap \bar{\mathcal{A}}I$ (**A**), and in publications that are *citing* them (**B**). The dashed lines show the diagonal where the rankings are exactly the same in the two corpora. Only are shown the most significant disciplines, with the condition that they commonly appear in the two corpora. The sizes of the discs, based on their empirical number of citations in the two different corpora, are normalized to compare their respective citation weights within each corpus. Grey points correspond to disciplines confined in the area between the two lines of respective equations $r_{\mathcal{P} \cap \mathcal{A}I} = r_{\mathcal{P} \cap \bar{\mathcal{A}}I} + \tau$ (above the diagonal) and $r_{\mathcal{P} \cap \mathcal{A}I} = r_{\mathcal{P} \cap \bar{\mathcal{A}}I} - \tau$ (below the diagonal), with $\tau = 10$, and with no important variations of positions from one period to one another. The most persistent ones over the 5 represented decades, 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 $\mathcal{P} \cap \bar{\mathcal{A}}I$ corpus while purple ones are those for the corpus $\mathcal{P} \cap \mathcal{A}I$. Abbreviations of the apparent disciplines are given in Appendix C

dispersed around the diagonal are represented with two different colors, the blue ones having a better rank in $\mathcal{P} \cap \bar{\mathcal{A}}I$ than in $\mathcal{P} \cap \mathcal{A}I$ and the purple ones having a better rank in $\mathcal{P} \cap \mathcal{A}I$ than in $\mathcal{P} \cap \bar{\mathcal{A}}I$.

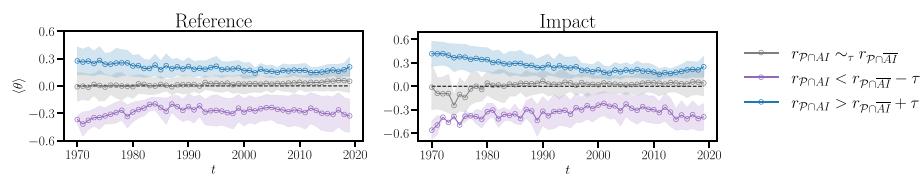


Fig. 7 Temporal average angles produced by the disciplines in each area of reference's (left) and impact's (right) diagrams of Fig. 6, with respect to the diagonals represented here by a dashed line at $(\theta) = 0$. The colored area covering the curves are representing their respective standard deviation from the mean angle. The angles are expressed in radians

To characterize each of these three areas around the diagonal in the references (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 $\mathcal{P} \cap \bar{AI}$ 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 we compute inside each zone and at each year the average angle $\langle \theta \rangle$ and the standard deviation from the latter, both represented in Fig. 7.

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 references' and impact's sides. According to the common ones that are more persistent over decades inside the references and received citations (see upper left boxes in Fig. 6A and B), 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 to more technological aspects of neuroscience, such as *Computer Science*, *Engineering*, *Radiology*, *Neuroimaging* and *Audiology & Speech-Language pathology*. A description of the evolution of the disciplinary landscape inside this zone is detailed in Appendix B.

The special disciplinary ecosystem of the non-AI corpus $\mathcal{P} \cap \bar{AI}$, in blue in the rankings maps of Fig. 6, 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. 7, the mean angles and the standard deviations of the disciplines in the associated zone are asserting a global concentration of them toward the common interest area, which includes on 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's angle plot of Fig. 7 a deviation from the diagonal of the disciplines in this zone of references since 2010, that 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 1980 (see Fig. 7), these references preferred by $\mathcal{P} \cap AI$ become more 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. 6A and B¹). Moreover, these special disciplines that cite most the corpus $\mathcal{P} \cap AI$ in Fig. 6B are as dispersed as the most cited ones by this corpus in Fig. 6A, 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 in purple in the impact side of Fig. 7). 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. 8 shows the rise of *Neuroimaging* and *Radiology* both on references' and impact's sides, as already observed in the common interest area, and also the progressive domination of *Computer Science*. This figure shows also the decrease of 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 with 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, that 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 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.

¹ Fig. 6B 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.

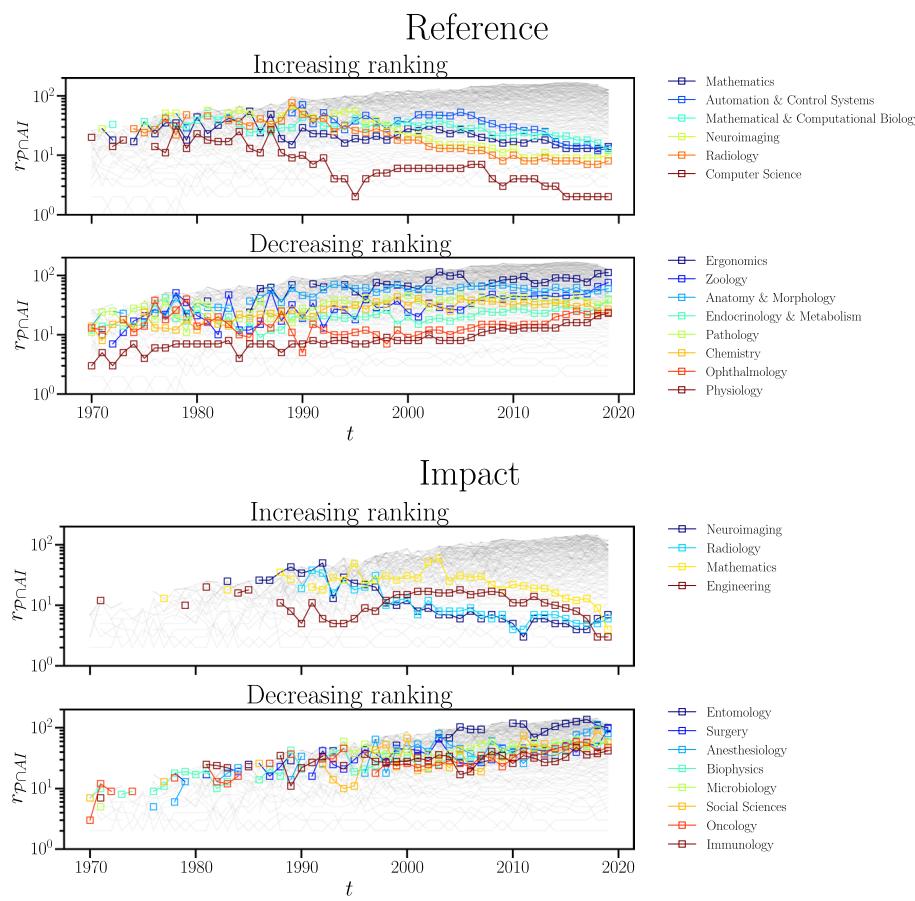


Fig. 8 Time evolution of the ranking $r_{P \cap AI}$ of the disciplines that are mostly represented by the corpus $\mathcal{P} \cap AI$ in the references (top) and citations (bottom), i.e. situated in the lower zone of the maps shown in Fig. 6. Only the curves with most significant evolution are highlighted with colors

An AI literature confined in a small set of journals

Another factor of differentiation of the AI research from the core of neuroscience lies on 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.

For each journal, based on its respective entire publication history, we compute the temporal share $a_{AI}(t)$ of AI-related publications (called also *temporal AI activity*), and its global share a_{AI} of such publications since its launch year (called also just *AI activity*). These scores are represented in Fig. 9. We then identify the journals with a_{AI} higher than 10%, whose names and launch years are given in the table in the bottom of this figure.

According to the evolution of the AI activities of neuroscience journals, 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 which is active in neuroscience since the late 1980s (Cardon et al. 2018). They are representing 32,7% of the scientific production of AI in neuroscience during the whole period 1970–2019,

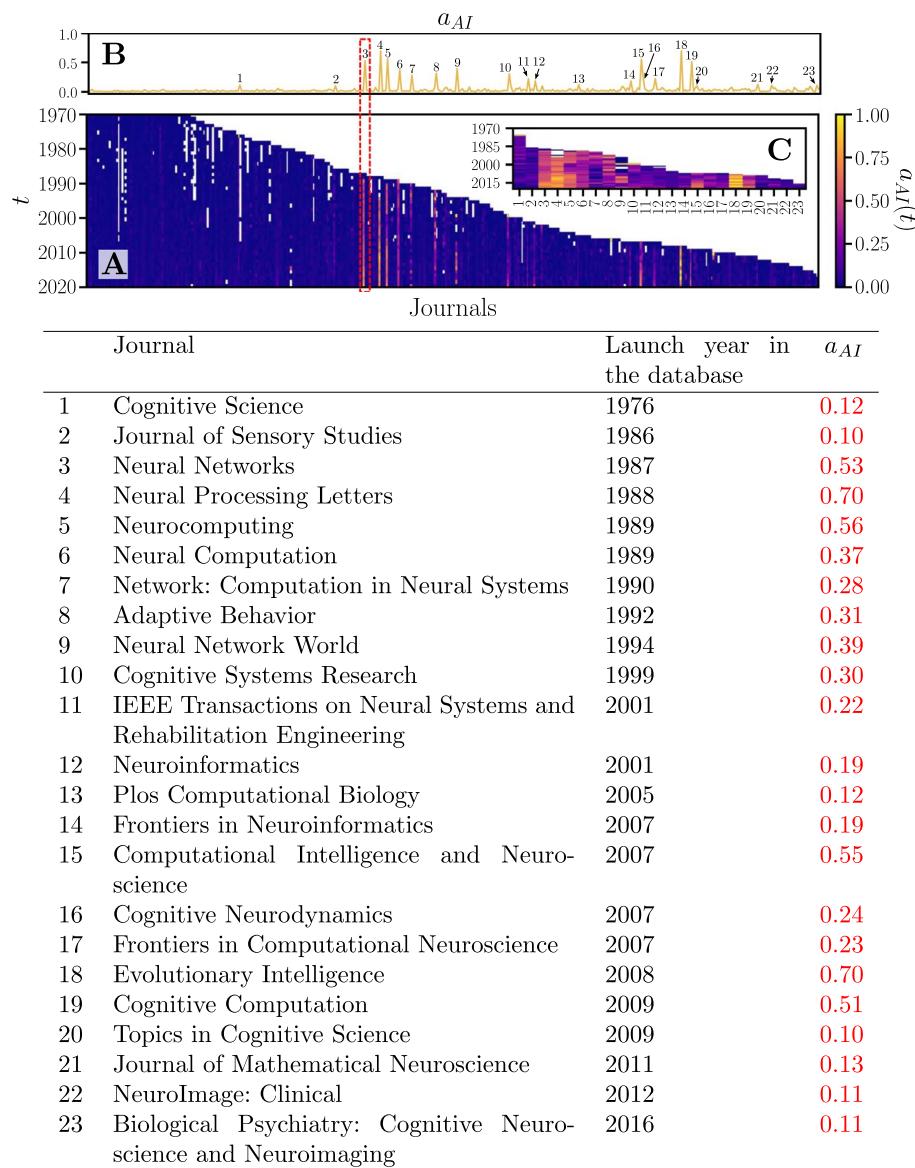


Fig. 9 Top diagrams: **A:** temporal AI activity of the journals. One square at the position (j, t) of the diagram is the share of AI publications in the journal j at time t . One column of the plot is the AI activity of one journal. **B:** distribution of the global AI activity of the journals over all their publications up to 2019. The horizontal axis is corresponding to the journals axis of the bottom 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 bottom table

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. 9) 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. 9) 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

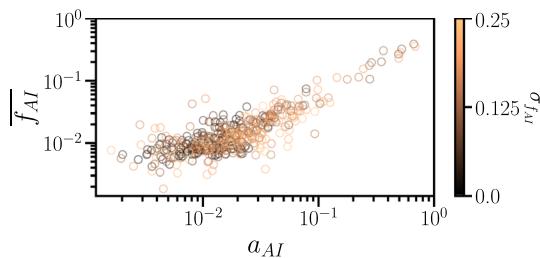


Fig. 10 Correlation plot between the global AI activity a_{AI} of the journals appearing in \mathcal{P} and the average AI scores \bar{f}_{AI} of the authors who are published in it between 1970 and 2019. Each point represents one journal in the horizontal a_{AI} space, colored by the standard deviation of the f_{AI} scores of their associated authors. The maximum average value of \bar{f}_{AI} is situated around 0.4

no. 7 and 9 (respectively *Network: Computation in Neural Systems* and *Neural Network World*) are showing well-sustained activity on AI research. This period falls commonly 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 are 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 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. 10 shows a linear correlation between the mean AI-activity in the journals and the average of the AI scores f_{AI} of the authors that have published at least one paper in these journals. The tail at highest a_{AI} (higher than 0.1) is corresponding to the top 15 of the most AI-active journals given in Fig. 9. This particular result thus unveils attraction of the authors publishing the most AI-related works of our dataset in the journals with a high AI-activity in time. This suggests that AI research in neuroscience is done by a specialized scientific community inside this field with its own journals for communicating results.

Who are the scientists making AI in neuroscience?

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

In this part we focus on the profiles of the scientists doing AI in neuroscience (in Fig. 4, included in one of the quartiles Q_0 , Q_1 or Q_2) and how they are inserted into the global authorship landscape of neuroscience. In particular, we compare these AI practitioners with the other scientists in neuroscience who never published AI-related papers

(included in the quartile \bar{Q}) under two aspects, namely (1) their *disciplinary background* and (2) their *disciplinary trajectory* in academia.

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 three first 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. 11. In this way we assess for each quartile the main native specialties in which authors have first published.

Figure 11 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. 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. 6). These two profiles best represent the disciplinary spectrum of neuroscience itself. We notice in addition 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 of 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. 11 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 rising of a group of scientists that are specialized into computation in neuroscience. Notice that the behavior of the *Neurosciences* curve within Q_1 in Fig. 11 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. 12. Since the backgrounds are built from the publications at a given year, the authors who began their career in 2019 would have a disciplinary profile corresponding to it. For avoiding 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.

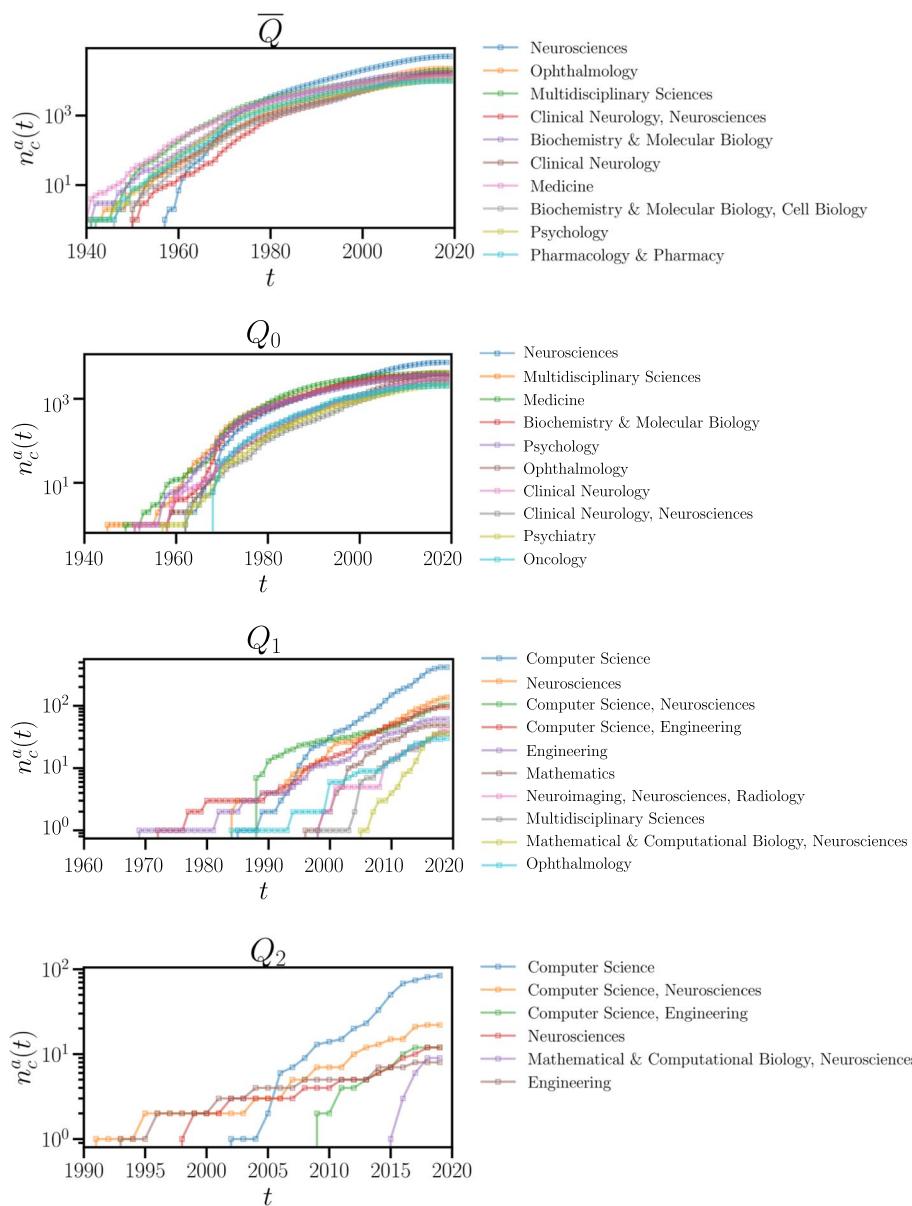


Fig. 11 Cumulative number of new authors per disciplinary background for each quartile. For one plot, the point of the curve of the specialty d at year t is the number of authors who have published their first articles inside d during their very first year of academic career, since the year of first 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

As a confirmation of the disciplinary background shown in Figs. 11, 12 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

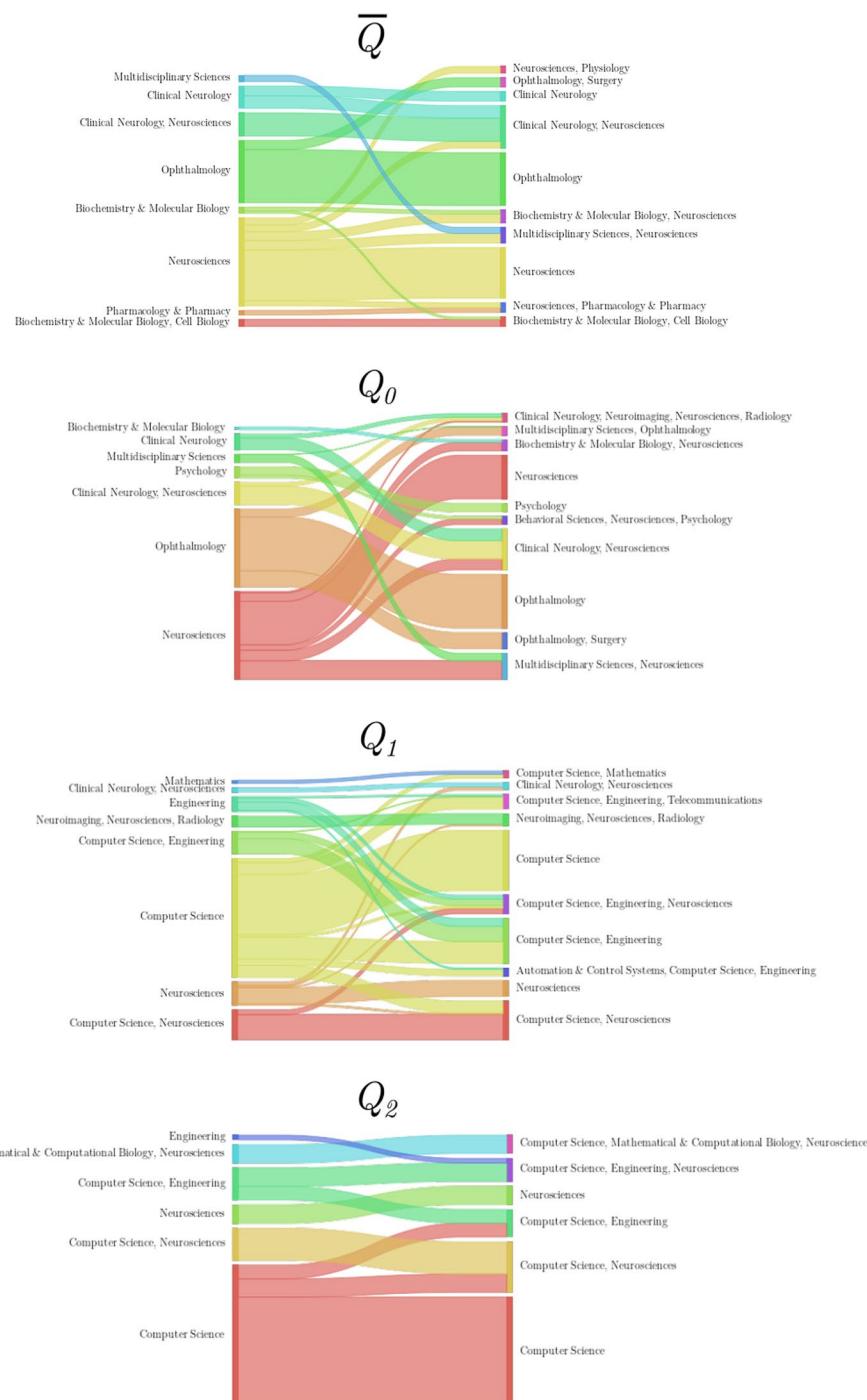


Fig. 12 Significant disciplinary trajectories of the authors belonging in each quartile. On each plot lie on the left the most significant disciplinary backgrounds, and on the right lie the career-related disciplines up to 2019

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

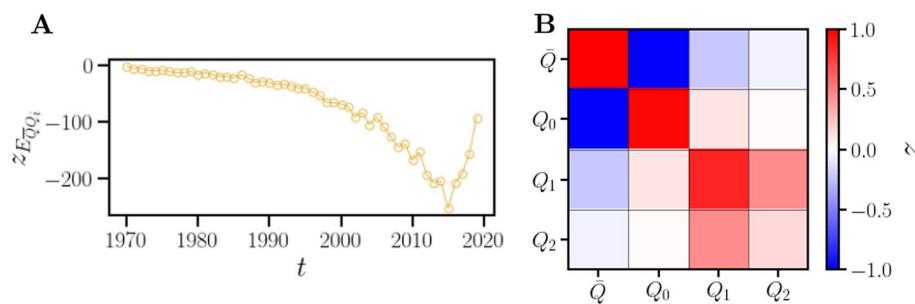


Fig. 13 **A** temporal z-score applied to the edges in TCN between the scientists in \bar{Q} and those in the other Q_i , here aggregated together under the notation Q_i . **B** Z-score matrix of the edges between each quartile in the time-aggregated collaboration network ACN. The values are normalized with the absolute maximum one in the matrix

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. 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*, *Neurosciences*, *Radiology*. This means that these scientists are trained into a disciplinary context centered around neuroscience. 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. 12, for instance *Computer Science*, *Engineering* leading to *Computer Science*, *Engineering*, *Neurosciences*. 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 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 represents 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).

Structure of the collaboration network between quartiles

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

We first consider the temporal collaboration network TCN built in Sect. 3.3.2, and we evaluate the temporal standardized share of edges between scientists belonging respectively to \bar{Q} and to all other quartiles Q_i (see Sect. 3.3.3 for the calculation of such score). This score is showed in Fig. 13A, 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 one computed from several random distributions of all the edges between the authors in TCN. This fact thus shows that the AI practitioners within neuroscience and the other neuroscientists maintain few collaborations as TCN 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 indicates that the neuroscientists making AI, ie. in Q_i and not in \bar{Q} , are shaping an almost independent community inside neuroscience by widening progressively a gap with other neuroscientists belonging to the set \bar{Q} . Furthermore, concerning especially the authors in Q_0 , who have similar disciplinary backgrounds and trajectories as the authors in \bar{Q} (see Figs. 11 and 12), 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 divide is confirmed with Fig. 13B, which represents the panorama of the links shared between the different quartiles in ACN (see also Sect. 3.3.3). This diagram represents a stabilized situation observed since the end of the 1990s (see Fig. 16 in Appendix D), 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. 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 impacted directly by neuroscience in Fig. 5, as well as with the temporal evolution of the disciplinary landscapes in Fig. 6, both indicating meaning that the quartiles Q_i export their knowledge in the whole neuroscience as much as the quartile \bar{Q} .

We also observe in Fig. 13B a polarization within the subset 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. 11, 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 appears to be the most interdisciplinary by assuring the bridge between these differentiated profiles inside AI community. These other *outsiders* in neuroscience are especially driving the diffusion of AI in neuroscience from computation to medical and clinical applications, given their disciplinary trajectories shown in Fig. 12.

Discussion

In this article we have explored the penetration of AI in neuroscience through the co-development of the two fields and especially the construction of a technological specialty centered around AI inside neuroscience. To do this we have conducted a scientometric analysis of an exhaustive bibliometric database that aims to represent at best neuroscience research between 1970 and 2019. This analysis relies on many indicators.

First, we have conducted a comparative analysis of the egocentric citation networks associated with respectively AI research and non-AI research within neuroscience. We have shown that the sets of disciplines influencing respectively the AI-related works and the non-AI ones tend to be similar over time, as well as those impacted by these works. AI thus becomes epistemically embedded in neuroscience over time, by reproducing citation patterns that are characterizing the field – ie. toward the disciplines that structure the field –, and by increasingly impacting the disciplines that show the most interest in this field – including neuroscience itself.

However, we have shown at micro-scale a progressive cognitive differentiation of the AI research from the non-AI one within neuroscience, based on both the employed bibliographic references in each of the two fields and their respective ensemble of impacted disciplines. We especially have observed specialization of AI research toward computer science, mathematics and engineering, while the core of neuroscience research draws upon biomedical and clinical research fields related to it. This differentiation is also produced by the neuroscience journals landscape, in which a small set of 23 journals, the most active in AI research in neuroscience, are representing those STEM fields in references and are gathering the scientists who are the most active in AI in neuroscience as well. Furthermore, by observing the fields impacted by the two corpora under study, we show that AI research targets rather the same fields of research as those cited by it. The epistemic integration of AI into neuroscience that was observed at first glance thus remains partial because of a specialty demarcation process of the former from the latter.

In a second step we explained this specialization by conducting a relational approach based on the scientific profiles of the authors of the studied database. By distinguishing the AI practitioners in neuroscience from the other neuroscience specialists, we have shown that the first tend to not maintain links with the second in the temporal co-authorship network including the main collaborations in the field since 1970. With the previous results we concluded that AI researchers are outsiders in neuroscience. We especially 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 are shaping the neuroscience field since the 1940s and who have a low AI activity, and a smallest second wave which emerged around the 1980s and that is including 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, its members keeping to publish within their original disciplines that are mainly STEM ones.

This social polarization inside the AI practitioners in neuroscience suggests that AI becomes over the years a set of technologies that need to be shaped not only by neuroscientists themselves, but also with the help of scientists coming from outside neuroscience, or from within the discipline but with an interdisciplinary background, and who present specific expertise about AI itself. Since we considered AI as a global research-technology in science (Shinn and Joerges 2002; Marcovich and Shinn 2020; Hentschel

2015), these results thus are a consequence of the diffusion of AI outside its originating STEM disciplines and throughout the science system (Gargiulo et al. 2023). The second wave of AI researchers described before represents the mobility of such experts toward 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.

These results thus illustrate quite well the generic property of AI when applied in neuroscience, which is producing a social and cognitive differentiation inside the latter (Shinn and Joerges 2002). However we could question the relevance of this criterion in this case, where only 3% of publications in our dataset involves AI according to our keyword filter introduced in Sect. 3.1. This leads especially to ask how spread is AI across the topics covered by neuroscience over time, ie. if it is present in all sub-fields of neuroscience or concentrated around a few ones. With such a topic space, and by reusing data about authors, the co-signature network and the citation network, we could evaluate more precisely how much universal AI is inside neuroscience through the distributions of authors and citations in this topic space, and then deduce the propensity of AI to fit with the knowledge and methods associated with some topics rather than others.

Nonetheless, the dynamical process of integration of AI in neuroscience exhibits some differences regarding the global history of AI in science depicted in Gargiulo et al. (2023). Although the development of AI in science is statistically characterized by a disciplinary closure around STEM disciplines between 1980 and 2010, AI continued its interactions with neuroscience for its own epistemic purposes in this period, as shown by a large number of journals created in this period and that are very active in AI research (see Sect. 4.3), and by its global citational impact in the discipline (see Sects. 4.2 and 4.1). However, a social closure was simultaneously occurring and accentuated inside the field, where not everyone is finally using AI at all, even in recent days. According to Shinn and Joerges (2002), this is typical of a differentiation process observed in the conception phase of an instrument.

The penetration of AI inside a single discipline thus could also be described as another underlying dynamical process of development of the associated knowledge and instruments following the four steps of Hentschel (2015), included inside the diffusion phase of the original instrument. For instance, the AI developed inside neuroscience, potentially different from the originally what is produced in STEM disciplines, would be also adapted to furnish other capabilities in neuroscience first, but also in other disciplines or fields of research afterwards if they judge it useful for their own goals. Under such an hypothesis, we could explore more precisely on the one hand the expansion of adjacent possible of neuroscience caused directly by AI, and on the other hand the expansion of the adjacent possible of other fields that have been influenced later by an adapted AI that was designed in neuroscience (Kauffman 2000; Monechi et al. 2017; Bianchini et al. 2022).

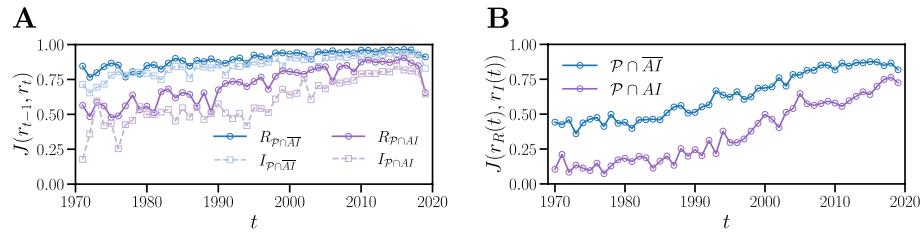


Fig. 14 **A:** temporal self-similarity of disciplinary rankings associated with either the references of one specific corpus (R) or the citations it has received (I), either for the AI-related corpus ($\mathcal{P} \cap AI$) or for the non-AI one ($\mathcal{P} \cap \overline{AI}$). More precisely, one point at time t is the similarity between the ranking at time t and the former one at $t - 1$. **B:** instantaneous similarity between the references used by one corpus and its produced citational impact

In addition to the limitations mentioned throughout this paper, especially for the building of our database described in Sect. 3, we should also consider diversity of types of publications that could differ from one discipline to another, and the manner they impact different 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 subsists in other media for communicating research results, such as conference proceedings and pre-prints sharing platforms such as *arXiv*, commonly used by mathematicians, physicists and computer scientists (Wainer et al. 2013). Future research may clarify whether our results hold if we consider the impact of AI produced within neuroscience through these other publication outlets.

Finally, this paper intends to be a road map 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. A comparative work would then be required.

Appendix A

Self-similarities of references and citations, and disciplinary concentration in the AI and non-AI corpora

This section focuses on the cognitive development of the research field associated with the AI corpus ($\mathcal{P} \cap AI$) and the non-AI one ($\mathcal{P} \cap \overline{AI}$), based on the same similarity measure introduced in Sect. 4.1. We compute for each corpus two such measures that could vary over time:

1. the temporal self-similarity of one given ranking r , namely the similarity between this ranking at time t and the same at the previous time $t - 1$, denoted as $J(r_{t-1}, r_t)$; we apply it for references' and citations' rankings in each corpus,
2. the temporal disciplinary concentration of one corpus, namely the similarity between its references ranking r_R and its citations one r_I at a given time t , denoted as $J(r_R(t), r_I(t))$,

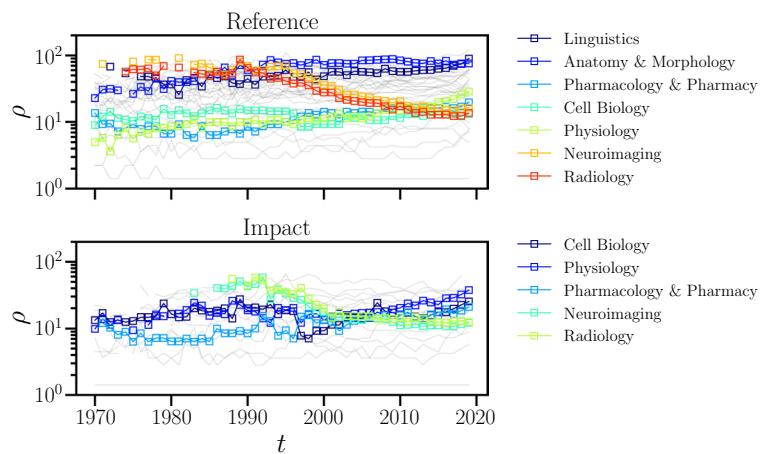


Fig. 15 Time evolution of the distance ρ of the disciplines included in the common interest area, as defined and shown in Fig. 6. Only the most significant curves are highlighted with colors

The evolution of these two indices are represented in Fig. 14A and B respectively.

According to Fig. 14A, the references on which the AI corpus and the non-AI one draw upon respectively (solid lines) are both consolidating with time toward their highest respective 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 MAG. In addition, while the self-similarity of citations in the corpus $\mathcal{P} \cap \overline{AI}$ (blue dashed line) follow the same trend as its corresponding self-similarity of references, the self-similarity of citations in the corpus $\mathcal{P} \cap \overline{AI}$ (purple dashed line) is rather chaotic and varying around 50% between 1970 and the late 1990s, then it increases toward around 80% until 2019. This means that the short-term impacted fields in the two corpora are also consolidating toward rather the same ones in recent days. More precisely, Fig. 14B shows a growing concentration of references' and impact's rankings inside each corpus over the time period under study, ie. the impacted disciplines and those appearing in references become more and more similar, hence a research inspired by itself toward itself. This shows also the progressive disciplinary homogenisation of the two corpora $\mathcal{P} \cap AI$ and $\mathcal{P} \cap \overline{AI}$ independently, and therefore a consolidation of the entire field of neuroscience.

Appendix B

Disciplinary composition of the common interest area of the AI-related and non-AI corpora
 This section focuses on the disciplines inside the common interest area of the AI and non-AI corpora defined in Sect. 4.2. According to Fig. 6 in the main text, each discipline d appearing in the references and/or in the citations of these corpora is located in a 2D

space by 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. From these coordinates we compute the distance of the disciplines from the origin of the map (point with ranks (0,0)), denoted as ρ in the following. Figure 15 shows the evolution of this distance for some disciplines, which are here exhibiting the most significant variations (increasing or decreasing) over the years since the 1970s.

This figure shows especially a spectacular rise since around 1995 of the references' 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.

Appendix C

Web of Science categories' abbreviations table

| WOS Categories | Abbreviations |
|--------------------------------------|---------------|
| Acoustics | Acoustics |
| Anatomy & Morphology | A &M |
| Anesthesiology | Anesth |
| Anthropology | Anthrop |
| Automation & Control Systems | A &CS |
| Biophysics | BioPhys |
| Biochemical Research Method | BRM |
| Biotechnology & Applied Microbiology | Biotech |
| Cardiac & Cardiovascular System | Cardio |
| Chemistry | Chem |
| 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 |

| WOS Categories | Abbreviations |
|---|----------------------|
| Hematology | Hemato |
| Imaging Science & Photographic Technology | IS &PT |
| Immunology | Immuno |
| Instruments & Instrumentation | Instrum |
| Language & Linguistics | L &L |
| Mathematical & Computational Biology | M &CB |
| Mathematics | Maths |
| Mechanics | Mech |
| Medical Informatics | MI |
| Medical Laboratory Technology | MLT |
| Microbiology | Microbio |
| Microscopy | Microscopy |
| Neuroimaging | NI |
| Nutrition & Dietetics | N &D |
| Operations Research & Management Science | OR &MS |
| Optics | Optics |
| Otorhinolaryngology | Otorhino |
| Pathology | Patho |
| Pediatrics | Ped |
| Peripheral Vascular Diseases | PVD |
| Philosophy | Philo |
| Physics | Phys |
| Plant Sciences | Plant |
| 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 |

Appendix D

Temporal evolution of the inter-quartile collaboration network

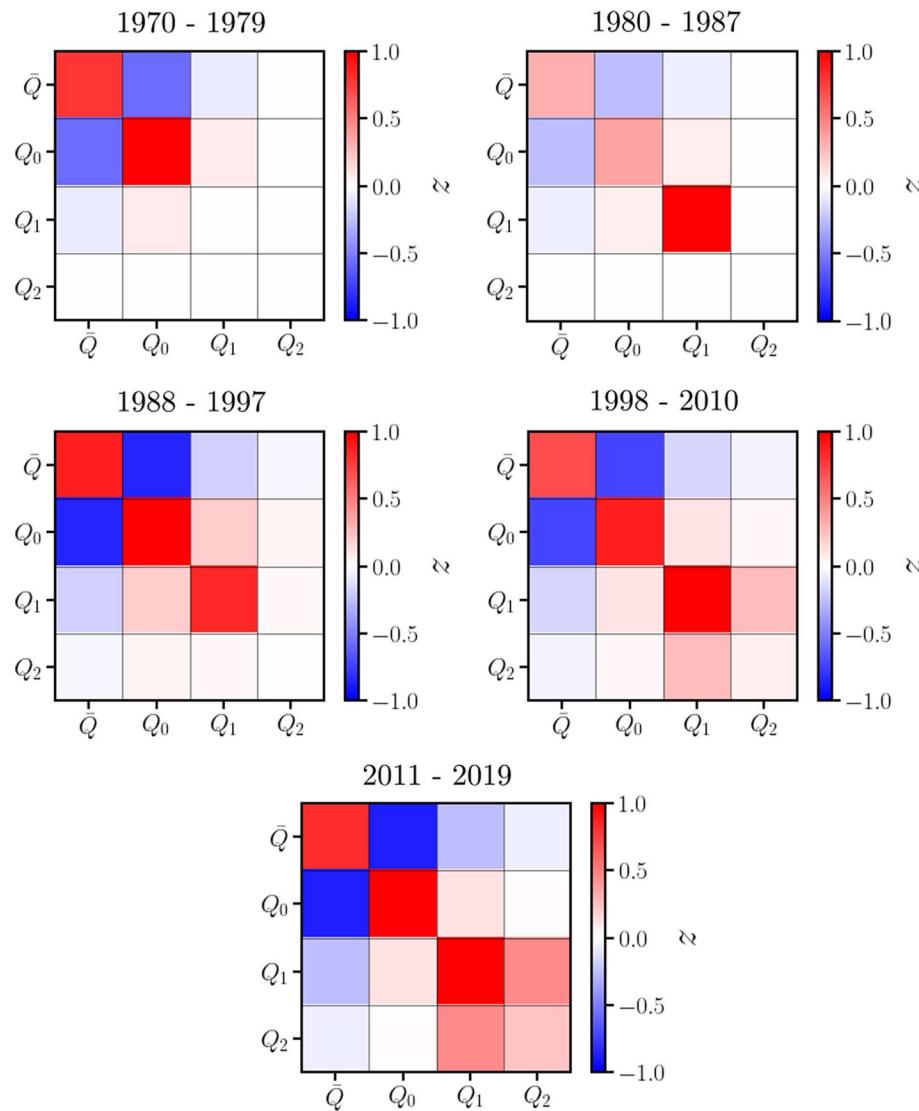


Fig. 16 Evolution of the z-score matrix of the share of edges between f_{AI} quartiles. The first line, corresponding to the period 1970–1987, refers to the situation where the quartile Q_2 has not yet emerged in our dataset. The separation of \bar{Q} and the other quartiles Q_i 's occurs qualitatively in the period 1988–1997, when Q_2 was emerging and was beginning to publish in neuroscience. The ending period spanning from 1998–2010, which concentrates 91% of the 13,786,616 unweighted single edges in the temporal collaboration network TCN, is very similar to the configuration shown in the main text, especially in Fig. 13 which was computed on the basis of all the period 1970–2019

Acknowledgements

We thank Abdelghani Maddi for very useful discussions.

Author Contributions

S.F. collected and analyzed the data, discussed the results, wrote the manuscript. F.G. conceived the research, collected the data, discussed the results, wrote the manuscript. M.D. discussed the results, wrote the manuscript. P.T. discussed the results, wrote the manuscript. All authors have read and approved the manuscript.

Funding

S.F. is funded through a CNRS-MITI PhD grant within the project "Epistemic Impact of Artificial Intelligence in Science" (EpiAI). This research has been partially supported by the ANR grant ScientIA (ANR-21-CE38-0020).

Availability of Data and Materials

The datasets used for the this study are available at: <https://doi.org/10.5281/zenodo.10777508>. The codes for analyzing them are available at: https://github.com/sySyMC/AI_in_Neuroscience_EpistemicIntegration_SocialSegregation.

Declarations**Ethics approval and consent to participate:**

Not applicable.

Consent for publication:

Not applicable.

Competing interests

The authors declare that they have no conflict of interest.

Received: 2 October 2023 Accepted: 6 April 2024

Published online: 17 April 2024

References

- Andler D (1990) Connexionisme et cognition: à la recherche des bonnes questions. *Rev Synth* 111(1–2):95–127
- Arencibia-Jorge R, Vega-Almeida RL, Jiménez Andrade JL, Carrillo-Calvet H (2022) Evolution and multidisciplinarity of artificial intelligence research. *Scientometrics* 127(2):1–20
- 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)
- Bianchini S, Müller M, Pelletier P (2022) Artificial intelligence in science: an emerging general method of invention. *Res Policy* 51(10)
- 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
- Cooper RP, Shallice T (2010) Cognitive neuroscience: the troubled marriage of cognitive science and neuroscience. *Top Cogn Sci* 2(3):398–406
- Frank MR, Wang D, Cebrian M, Rahwan I (2019) The evolution of citation graphs in artificial intelligence research. *Nat Mach Intell* 1(2):79–85
- Färber M (2019) The microsoft academic knowledge graph: a linked data source with 8 billion triples of scholarly data. In: Ghidini C et al. (eds) The semantic web—ISWC 2019 (vol 11779, pp 113–129). Springer International Publishing
- Färber M (2020) Microsoft academic graph records. <https://zenodo.org/record/3936556>
- Gao J, Wang D (2023) Quantifying the benefit of artificial intelligence for scientific research. Preprint at: <https://arxiv.org/abs/2304.10578>
- Gargiulo F, Caen A, Lambiotte R, Carletti T (2016) The classical origin of modern mathematics. *EPJ Data Sci* 5(1):26
- Gargiulo F, Fontaine S, Dubois M, Tubaro P (2023) A meso-scale cartography of the AI ecosystem. *Quant Sci Stud* 4(3):574–593
- Gopinath N (2023) Artificial intelligence and neuroscience: An update on fascinating relationships. *Process Biochem* 125:113–120
- Haas PM (1992) Introduction: epistemic communities and international policy coordination. *Int Organ* 46(1):1–35
- Hassabis D, Kumaran D, Summerfield C, Botvinick M (2017) Neuroscience-inspired artificial intelligence. *Neuron* 95(2):245–258
- Hentschel K (2015) A periodization of research technologies and of the emergency of genericity. *Stud Hist Philos Mod Phys* 52:223–233
- Kauffman SA (2000) Investigations. Oxford University Press, Oxford
- Lake BM, Ullman TD, Tenenbaum JB, Gershman SJ (2017) Building machines that learn and think like people. *Behav Brain Sci* 40(e253):1–72
- 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
- Marcovich A, Shinn T (2020) Science research regimes as architectures of knowledge in context: a 'longue durée' comparative historical sociology of structures and dynamics in science. *Soc Sci Inf* 59(2):310–328
- McCarthy J (1981) Epistemological problems of artificial intelligence. *Read Artif Intell*, pp 459–465. Elsevier
- Monechi B, Ruiz-Serrano Á, Tria F, Loreto V (2017) Waves of novelties in the expansion into the adjacent possible. *Plos One*, 12(6)
- Perconti P, Plebe A (2020) Deep learning and cognitive science. *Cognition* 203:12
- Roth C (2008) Réseaux épistémiques : formaliser la cognition distribuée. *Sociologie du Travail* 50(3):353–371
- Savage N (2019) How AI and neuroscience drive each other forwards. *Nature* 571(7766):15–17
- Schuchmann S (2019) History of the second AI winter. Medium. <https://towardsdatascience.com/history-of-the-second-ai-winter-406f18789d45>

- Sedooka A, Steffen G, Paulsen T, Darbellay F (2015) Paradoxe identitaire et interdisciplinarité : un regard sur les identités disciplinaires des chercheurs. *Nat Sci Sociétés* 23(4):367–377
- Shinn T, Joerges B (2002) The transverse science and technology culture: dynamics and roles of research-technology. *Soc Sci Inf* 41(2):207–251
- Tang X, Li X, Ma F (2022) Internationalizing AI: evolution and impact of distance factors. *Scientometrics* 127(1):181–205
- Visser M, Van Eck NJ, Waltman L (2021) Large-scale comparison of biblio-graphic data sources: scopus, web of science, dimensions, crossref, and microsoft academic. *Quant Sci Stud* 2(1):20–41
- Wainer J, Eckmann M, Goldenstein S, Rocha A (2013) How productivity and impact differ across computer science subareas. *Commun ACM* 56(8):67–73

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.