HAL open science

A meso-scale cartography of the AI ecosystem

Floriana Gargiulo, Sylvain Fontaine, Michel Dubois, Paola Tubaro

To cite this version:

Floriana Gargiulo, Sylvain Fontaine, Michel Dubois, Paola Tubaro. A meso-scale cartography of the AI ecosystem. Quantitative Science Studies, 2023, 4 (3), pp.574-593. 10.1162/qss_a_00267 . hal-04367071

HAL Id: hal-04367071

https://hal.science/hal-04367071v1

Submitted on 2 Jan 2024

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

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

Distributed under a Creative Commons Attribution 4.0 International License

an open access journal

Check for updates

Citation: Gargiulo, F., Fontaine, S., Dubois, M., & Tubaro, P. (2023). A meso-scale cartography of the AI ecosystem. *Quantitative Science Studies*, *4*(3), 574–593. https://doi.org/10.1162/qss_a_00267

DOI: https://doi.org/10.1162/qss_a_00267

Peer Review: https://www.webofscience.com/api/gateway/wos/peer-review/10.1162/qss_a_00267

Supporting Information: https://doi.org/10.1162/qss_a_00267

Received: 23 December 2022
Accepted: 24 June 2023

Corresponding Author:
Floriana Gargiulo
floriana.gargiulo@cnrs.fr

Handling Editor:
Vincent Larivière

Copyright: © 2023 Floriana Gargiulo, Sylvain Fontaine, Michel Dubois, and Paola Tubaro. Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.

The MIT Press

RESEARCH ARTICLE

A meso-scale cartography of the AI ecosystem

Floriana Gargiulo[1], Sylvain Fontaine[1], Michel Dubois[1], and Paola Tubaro[2]

[1]GEMASS, CNRS, 59 rue Pouchet 75017, Paris
[2]CREST, CNRS, Institut Polytechnique de Paris, 5 avenue Le Chatelier, Palaiseau

Keywords: artificial intelligence, science cartography, science of science

ABSTRACT

Recently, the set of knowledge referred to as "artificial intelligence" (AI) has become a mainstay of scientific research. AI techniques have not only greatly developed within their native areas of development but have also spread in terms of their application to multiple areas of science and technology. We conduct a large-scale analysis of AI in science. The first question we address is the composition of what is commonly labeled AI, and how the various subfields within this domain are linked together. We reconstruct the internal structure of the AI ecosystem through the co-occurrence of AI terms in publications, and we distinguish between 15 different specialties of AI. Furthermore, we investigate the spreading of AI outside its native disciplines. We bring to light the dynamics of the diffusion of AI in the scientific ecosystem and we describe the disciplinary landscape of AI applications. Finally we analyze the role of collaborations for the interdisciplinary spreading of AI. Although the study of science frequently emphasizes the openness of scientific communities, we show that collaborations between those scholars who primarily develop AI and those who apply it are quite rare. Only a small group of researchers can gradually establish bridges between these communities.

1. INTRODUCTION

Artificial intelligence (AI) is increasingly recognized as a vector of technological and scientific innovation (Bianchini, Müller, & Pelletier, 2022; Cockburn, Henderson, & Stern, 2018) with a potentially strong impact on economic growth (Aghion, Jones, & Jones, 2018). A *Nature* editorial (Nature, 2019) describes it as one of the scientific events that shaped the last decade: "Few fields are untouched by the machine-learning revolution, from materials science to drug exploration; quantum physics to medicine."

The latest developments in AI, mostly as a result of the rise of Deep Learning (DL), provide a unique potential to extract information from the unprecedented sources of data currently largely available in many scientific and technological domains. AI has been described as enabling a general paradigm shift toward a data-immersive science (King, Rowland et al., 2009; Kitchin, 2014), based on smart machines able to grasp the hidden patterns and relationships from large masses of data. The origins of AI are usually traced back to a renowned workshop held in 1956 in Dartmouth, where a group of scientists first used this term to define their research activities and identify a distinct research area. Although AI is now commonly associated with artificial neural network methods, marking the very advent of connectionism and the automatic learning paradigm (Cardon, Cointet, & Mazières, 2018), AI was situated at this time in a disciplinary landscape that shaped both current neuroscience and cognitive science, where the intelligence was rather conceived as fixed knowledge representations that are

Downloaded from http://direct.mit.edu/qss/article-pdf/4/3/574/2200626/qss_a_00267.pdf by guest on 02 January 2024

symbolically manipulated in the brain (Haugeland, 1985). In this context, according to the definition given later in 2004 by John McCarthy, promoter of the Dartmouth workshop, AI "is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable." (McCarthy, 2004)

This top-down definition, like similar ones (Annoni, Benczur et al., 2018; OECD, 2019; WIPO, 2019), emphasizes the overall goals of AI but leaves open the actual meaning of "intelligence," the scope of the AI domain, and the relationships between AI and the existing structure of scientific knowledge.

Although the first question, related to the definition of intelligence, has raised multiple debates and controversies in the context of AI epistemology (McCarthy, 1981), and in closely related research areas such as cognitive sciences and neuroscience (Hassabis, Kumaran et al., 2017), our study sheds light on the second aspect and aims to model the structure of the AI research area through a bottom-up approach based on a cartography of AI-related scientific publications.

Science mapping, or science cartography, is a common method in science and technology studies, allowing us to represent scientific knowledge, the actors involved in scientific discovery, and the connections among concepts and actors (Chen, 2017). Science cartography uses visual tools to make explicit the roles of the actors according to different metrics and indicators, and allows us to read the complexity of bibliometric data sets. These techniques have made further progress through the more recent approach of science of science (Fortunato, Bergstrom et al., 2018), which uses statistical methods in complex systems analysis to represent the scientific ecosystem dynamically, aiming to identify the mechanisms underlying its growth.

In most science cartography studies, bibliometric corpora are built on the basis of the classifications and the field categories provided by scientific search engines (Web of Science, Scopus, etc.). However, this approach misses interdisciplinary papers: in this case, papers applying AI methods that are not published in AI journals, but rather in disciplinary journals. We define our approach as "bottom-up" because we start from building the corpus from the lowest level, searching for AI-related keywords and rebuilding the field mesoscale structure (the subfields of AI) and the disciplinary composition of the AI ecosystem (journals and disciplines) from an unstructured corpus of AI papers developing or applying AI techniques.

Along these lines, we address two research questions:

1. What is the structure of AI as a research area (i.e., the different specialties within AI and their development over time)?
2. How has AI knowledge been dynamically embedded in traditional scientific fields?

Previous works provide partial answers to these questions. For example Frank, Wang et al. (2019) build a classification based on citation networks among the list of AI subfields, defined through the categories adopted in Microsoft Academic Graph (MAG). However, this categorization does not allow a deep understanding of which AI terms are explicitly present in each subdomain. Baruffaldi, van Beuzekom et al. (2020) also propose a taxonomy of AI methods from various data sources, such as scientific publications referenced in Scopus, open source software available on GitHub, and patents. Although it shows variations in the intensity of research activity related to some flagship AI classes since the late 1990s, this work does not address the historical relations between them. Bianchini, Müller, and Pelletier (2020) provide a mapping of the spreading of DL practices in science and describe in great detail the

Downloaded from http://direct.mit.edu/qss/article-pdf/4/3/574/2200626/qss_a_00267.pdf by guest on 02 January 2024

geographical and disciplinary spreading of DL, but do not address the connections of DL with other AI practices as well. Other studies do so but only from the viewpoint of specific disciplines (Baum, Yu et al., 2021).

We start by establishing the semantic diversity of AI, gathering a large list of terms that are related to it without distinction of context, generality, timing, or other criteria. Assembling a suitable set of keywords to drive a bibliometric search is in itself a complex task, simplified in our case by the possibility of relying on multiple glossaries of AI accessible online—first of all, the Wikipedia AI glossary, which also contains synonyms for several terms. We therefore create a list of AI keywords by mining a large number of AI glossaries available on the web, which represent how a variety of actors, all dealing with AI, draw its boundaries. These terms represent the semantic building blocks of AI.

Whether we describe it as a body of knowledge, practices, or tools, AI is a dynamic phenomenon that has experienced several phases in its evolution over time. As scientific innovation in general (Uzzi, Mukherjee et al., 2013) can be viewed as a cumulative process where novelty arises from the recombination of existing building blocks, a dynamical view of AI can be seen as the formation of its specialties through the recombination of its foundational terms. Interdisciplinary exchanges also play a central role in scientific innovation, proposing new possible building blocks and thereby opening the "adjacent possible" of scientific discoveries (Kauffman, 2000; Monechi, Ruiz-Serrano et al., 2017). Likewise, extensions of AI arise from the recombination of pre-existing applications and by interactions with other research areas: Consider the example of DL, resulting from AI research on artificial neural networks recombined with connectionist approaches in cognitive science. For this reason, understanding the embedding of AI in the scientific ecosystem provides fundamental information to grasp the dynamic processes behind its current rise.

In the last decade, a spectacular increase of the application of AI techniques in several and diverse scientific domains has been observed, above all in relation with the development of DL (Bianchini et al., 2020, 2022). The common understanding of this phenomenon is that AI is spreading from its "native" disciplines (mainly computer science, mathematics, and statistics), where its key tools were designed, to a series of applications in various fields of knowledge. This distinction between native disciplines and application disciplines can be found in several studies (Bianchini et al., 2022; Cockburn et al., 2018).

In line with this literature, we assume that foundational AI algorithms and theorems are mostly published in journals in statistics, mathematics, and computer science. Therefore, it is on the basis of journals' classification that we retain this definition of statistics, mathematics, and computer science as "native disciplines." This does not rule out the possibility that authors from other disciplines, such as cognitive science, may produce and publish AI papers in "native" journals. Lack of systematic information prevents us from making any assumption about authors' backgrounds.

In what follows, we analyze a large corpus of scientific papers published between 1970 and 2017, extracted from Microsoft Academic Graph through the AI keywords used by the authors, and analyzed with different relational structures among the scientometric data (keyword co-occurrence network, authors' collaboration network). We opt for uncontrolled keywords-oriented extraction to explore the production of AI-related papers in international academic research, with maximal disciplinary diversity. Although a variety of methods exist for assessing scientometrically the robustness of a thematic list of keywords, such as the one exposed in Huang, Schuehle et al. (2015), their applications involve the risk of shrinking the list to comprise only keywords associated with the "native disciplines" of AI, characterized by specific

Downloaded from http://direct.mit.edu/qss/article-pdf/4/3/574/2200626/qss_a_00267.pdf by guest on 02 January 2024

vocabularies and cocitation patterns. Indeed, the main aim of these methods is to remove noise that may arise from publications that do not develop or apply AI tools but take them as objects of study, as happens, for example, in some critical humanities and social science analyses.

Therefore in our case, these methods could lose useful information about the potential diffusion of AI tools in some disciplines. We thus keep the raw list of keywords we extracted from the web to reduce this risk, and for robustness, we compare it to those, already validated, of Baruffaldi et al. (2020); WIPO (2019).

To characterize the keywords used in corpus selection, we define their hierarchical structure to distinguish core AI terms (mainly specific algorithms and techniques) from more peripheral ones. We first focus on the definition of the meso-scale structure of AI, namely on the identification of the specialties of AI, their interactions, and their temporal patterns (Section 3.1). Second, we analyze how AI is globally spreading in multiple research areas or disciplines. Very similar to a scientometric work dedicated to this subject (Arencibia-Jorge, Vega-Almeida et al., 2022), we observe a first phase of concentration of AI in the "native" disciplines of computer science, mathematics, and statistics at the end of the 1980s, after the so-called "AI Winter," with the emergence of expert systems and the decline of symbolic AI. These disciplines remain responsible for the production of AI literature until today. However, a spreading phase started in the last decade, corresponding to the development of DL, as AI knowledge started to be largely applied to several other disciplines (Section 3.2). We also reveal the disciplinary patterns associated with the different specialties of AI. We observe, for example, that only a few specialties (such as dimensionality reduction techniques and DL) reached a high degree of diversity in the application ecosystem. Finally, we highlight the collaboration mechanisms responsible for knowledge transfer from the originating domains to applications. We notice indeed that very few collaborations exist between researchers in disciplines that create AI and researchers in disciplines that only (or mainly) apply AI. Transfer of AI knowledge is largely ascribable to a core of multidisciplinary researchers mutually interacting both with AI developers and with researchers in applied disciplines (Section 3.3).

2. DATA AND METHODS

2.1. AI Terms

There are several definitions of AI, and each of them implies a different perimeter of the terms or lexical units associated with it. With the effort of defining the scope of AI, several and diverse actors involved in its production, such as the OECD (Baruffaldi et al., 2020), made available online glossaries containing lists of associated keywords, with the objective of identifying the variety of terms that it covers. In particular, Wikipedia has a large list of its pages connected to AI, including synonyms.

We started by extracting the content of the Wikipedia index page[1], after which we performed a Google query searching for "AI glossary," "AI keywords," "AI terms," and "AI concepts." We obtained a set of more than 20 specific glossaries, for example those developed by Google[2], by independent specialists[3], and by publishers[4].

We built our original list of terms from all the keywords from these web resources, removing duplicates and lemmatizing words. We manually cleaned the list of keywords, removing very

[1] https://en.wikipedia.org/wiki/Glossary_of_artificial_intelligence.
[2] https://developers.google.com/machine-learning/glossary.
[3] https://machinelearning.wtf/.
[4] https://link.springer.com/content/pdf/bbm:978-3-319-94878-2/1.pdf.

Downloaded from http://direct.mit.edu/qss/article-pdf/4/3/574/2200626/qss_a_00267.pdf by guest on 02 January 2024

general words not strictly related to AI (such as "software," "algorithm," and "self-management"). The final list includes 594 terms, mostly bigrams or trigrams, with different levels of generality. There are general terms, such as "machine learning" and specific algorithmic procedures, such as "word2vec." The full list of terms is reported in the appendix.

2.2. The Bibliometric Data Set

The bibliometric data set on which this article is based starts from a recent data dump of the Microsoft Academic Graph (MAG), disambiguated and made available by M. Färber on the Zenodo platform[5] (Färber, 2019). From this data set, we first selected all the papers including any of the previously identified 594 AI terms in their abstracts or title (2,737,813 papers with associated metadata). From this set, we kept only the papers published after 1970. This choice could appear too strict, missing almost 2 decades of early AI research, but it avoids the heterogeneities that would result from differences in editorial policies and scientific infrastructure in that period compared to today, notably in terms of peer reviewing. Additionally, we retain only studies published in or before 2017 because of a possible bias in the MAG database for later entries, which can be guessed from an unmotivated decrease in the total number of papers. We further filtered this data set to the papers published in journals or conferences indexed in the Web of Science (WoS), getting a final set of metadata for 1,090,138 papers.

We associate each of these papers two supplementary attributes with respect to the MAG metadata: the disciplinary fields, according to the first label (which is indeed the more specific) in the WoS classification of journals and conferences, and the list of AI keywords contained in their abstracts. To build the authors' collaboration network we used the disambiguated authors' identifiers provided by Färber and Ao (2022) in the latest version of the MAG database.

To summarize, the bibliometric corpus that we have constituted starting from the MAG data set is a collection of documents having the following attributes (Figure 1): list of AI keywords contained in the abstract and/or title, publication year, list of authors, journal (or conference), and disciplinary field derived from the journal's classification and categorical structure in the WoS.

2.3. The Network Structures

With these data, we reconstruct two different network structures: the keyword co-occurrence network (KCON) and the author collaboration network (ACN).

These networks are directly built from the documents of the corpus, as described in Figure 1. For the keyword co-occurrence, each document that contains more than a keyword represents a hyperedge of the graph, namely, a fully connected subgraph among all the present keywords. These hyperedges are merged together and each link gets a weight corresponding to the number of hyperedges where it is present. The procedure is identical for the authors' collaborations. Both these graphs are undirected weighted structures.

The keyword co-occurrence graph, KCON, has 535 nodes and 24,358 edges (some less frequently used AI terms were indeed disconnected from the larger component and have therefore been omitted in the rest of the analysis). A this network is extremely dense (density = 0.17) and the weights are very heterogeneous, we first apply a disparity filter on the original

[5] https://zenodo.org/record/4617285#.YlUoaS2ubs0.

Downloaded from http://direct.mit.edu/qss/article-pdf/4/3/574/2200626/qss_a_00267.pdf by guest on 02 January 2024

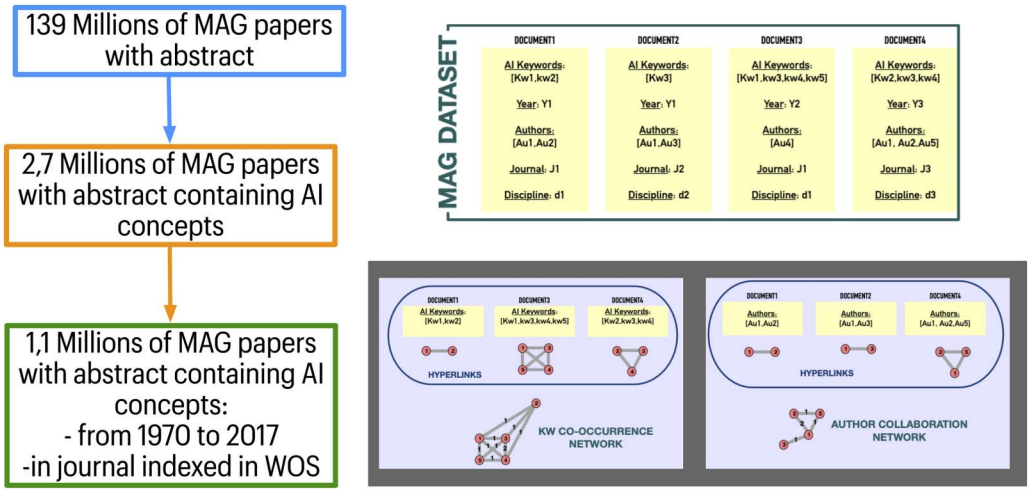


Figure 1. The data set. Left: filtering process of the MAG corpus. Top right: Structure of the AI corpus. Bottom right: Building process of the keywords co-occurrence network and of the author collaboration network.

graph (Serrano, Boguñá, & Vespignani, 2009) to get the relevant connections and simplify the partitioning of the structure. The filtered graph DKCON has 3,276 edges. The author collaboration network (ACN) has 103,175 nodes and 453,137 edges.

2.4. The Disciplinary Distance Matrix

To reconstruct a distance matrix between all the WoS disciplines, we start again from the whole MAG data set, filtered on the WoS journals. To reduce computation time, which would be very significant if we analyzed a single large snapshot from this extremely large data set, we followed a procedure of producing several independent samples. Specifically, we extracted 10,000 random samples of 100,000 papers. For each paper in each random sample, we extracted the list of all the referenced papers and, from the latter, the set of unique WoS disciplines relative to the references. From these lists of disciplines, following the same procedure adopted for the KCON and the ACN networks, we build the co-citation structure of disciplines in the sample.

Because the weighting structure of this graph (the number of co-occurrences w_{ij}) is strongly related to the relative frequency of each discipline, we implement a similarity measure based on the pointwise mutual information between the nodes (disciplines):

$$\mathrm{pmi}_{ij} = \max\left(2\log_2\left(\frac{w_{ij}}{\sum_k(w_{ik})\sum_k(w_{jk})}\right), 0\right) \tag{1}$$

The pointwise mutual information ranges indeed between –1 and 1; in our case, the negative values, representing a very uncorrelated situation, are put to zero to obtain an indicator ranging from 0 to 1. From this similarity measure we obtain a distance matrix whose values are given by: $D_{ij} = 1 - pmi_{ij}$.

We repeat this computation for all the 10,000 samples, and the final distance matrix is derived from the average values on all the samples.

Downloaded from http://direct.mit.edu/qss/article-pdf/4/3/574/2200626/qss_a_00267.pdf by guest on 02 January 2024

2.5. The Spreading Indicators

We measure how a corpus is concentrated around the so-called native disciplines of AI (as above, computer science, mathematics, and statistics) with a measure inspired from solid body mechanics, the moment of inertia:

$$m_I = \sum_i \frac{n_i}{N_{tot}} \min\left(D_{i,CS}, D_{i,Matb}, D_{i,Stat}\right)^2 \tag{2}$$

where i covers all the disciplines present in the corpus. If the moment of inertia is small, the corpus is very concentrated around the native disciplines; if not, it is largely diffused in the disciplinary ecosystem.

To measure how AI is represented in a discipline we compare the number n_i^{AI} of AI papers produced in this discipline with an expected value given by the share of publications in the given discipline ($s_i = N_{tot}^i / N_{tot}$, extracted from the whole MAG corpus) multiplied by the total number of AI publications. We define therefore the AI score of a discipline:

$$z_i = \frac{n_i^{AI} - N_{tot}^{AI} s_i}{n_i^{AI} + N_{tot}^{AI} s_i} \tag{3}$$

This measure ranges between –1 and 1. Positive high values of this indicator indicate that AI is more represented in the discipline than in a case in which diffusion followed a random process, and vice versa. The same measure also applies at the level of journals.

Finally, we compute for each author in the corpus an AI score, A_I, given by the fraction of papers published by the author in the native disciplines of computer science, mathematics, and statistics.

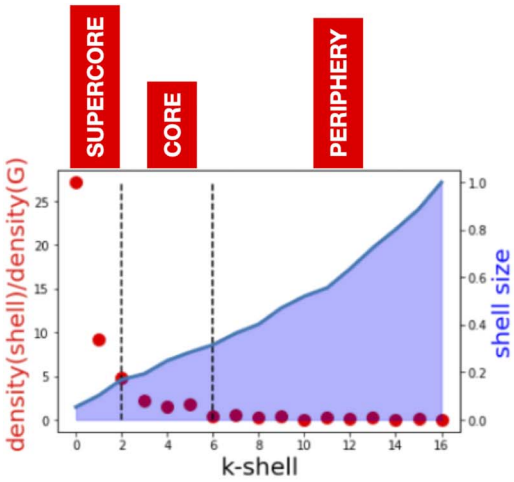
3. RESULTS

3.1. The Specialties of AI

We first use the co-occurrence of keywords in abstracts (DKCON as described above) to describe the specialties, or thematic diversity, of AI.

As we pointed out in the data presentation, the keywords used in the query have different levels of "generality." We first use the filtered keyword co-occurence graph (DKCON) to identify the hierarchy of dependencies between keywords. We build the k-shell structure of the graph and we calculate the internal density of each shell, compared to the density of the whole DKCON graph. This analysis allows us to distinguish three dimensions: the super core, the core, and the periphery. Figure 2 shows that the first two shells are very dense: They include a group of 25 keywords largely used and tightly connected among them. We call these first two shells the "super-core." This class contains general AI categories ("artificial intelligence," "machine learning," "DL," "neural networks") and very popular classes of methods ("random forest," "support vector machine"). The internal density decreases suddenly starting from the third shell and goes to zero in the most external shells, starting from the seventh one. We define shells 3–6 as the core and the last ones as the periphery. The core also contains general methods (such as "cluster analysis," "particle swarm optimization," and "stochastic gradient descent") but less connected among them and hierarchically depending on supercore terms (namely connected to the corpus only through supercore terms). The periphery mostly contains specific algorithms and specific methods not connected among them but just to the more central cores.

Downloaded from http://direct.mit.edu/qss/article-pdf/4/3/574/2200626/qss_a_00267.pdf by guest on 02 January 2024



SUPER-CORE	CORE
neural networks, machine learning, **classifier** , training data, **deep learning**, artificial intelligence, support vector machine, **genetic algorithm**, **data mining**, feature extraction, artificial neural network, **computer vision**, pattern recognition, artificial neural networks, support vector machines, feature selection, principal component analysis, image classification, supervised learning, random forest, face recognition, **dimensionality reduction**, overfitting , predictive model	convolutional neural network, **natural language processing**, information retrieval, **reinforcement learning**, convolutional neural networks, recurrent neural network, knowledge representation, **speech recognition**, **expert system**, **robotics** , convolutional neural network (cnn), search algorithm, particle swarm optimization, object detection, machine translation, multi-agent system, evolutionary algorithm, decision support system, hidden layer, recurrent neural networks, maximum likelihood estimation, k-means clustering, gaussian mixture model, autoencoder , unsupervised learning, naive bayes, stochastic gradient descent, conditional random fields, latent dirichlet allocation, evolutionary computation, hidden markov models, markov chain monte carlo, bag-of-words , computational intelligence, minimax , inference engine, singular value decomposition, feature learning, word2vec , linear discriminant analysis, machine vision, cluster analysis, intelligent agent, distributed representation

Figure 2. AI term hierarchy.

We apply to the DKCON graph the well-known Louvain community detection algorithm (Blondel, Guillaume et al., 2008) and identify the presence of 15 meso-scale structures that correspond to a partitioning of the network at the level of specialties (Figure 3): expert systems, natural language processing, dimensionality reduction, data mining, classifiers, neural networks, robotics, genetic algorithms, speech recognition, logic programming, face recognition,

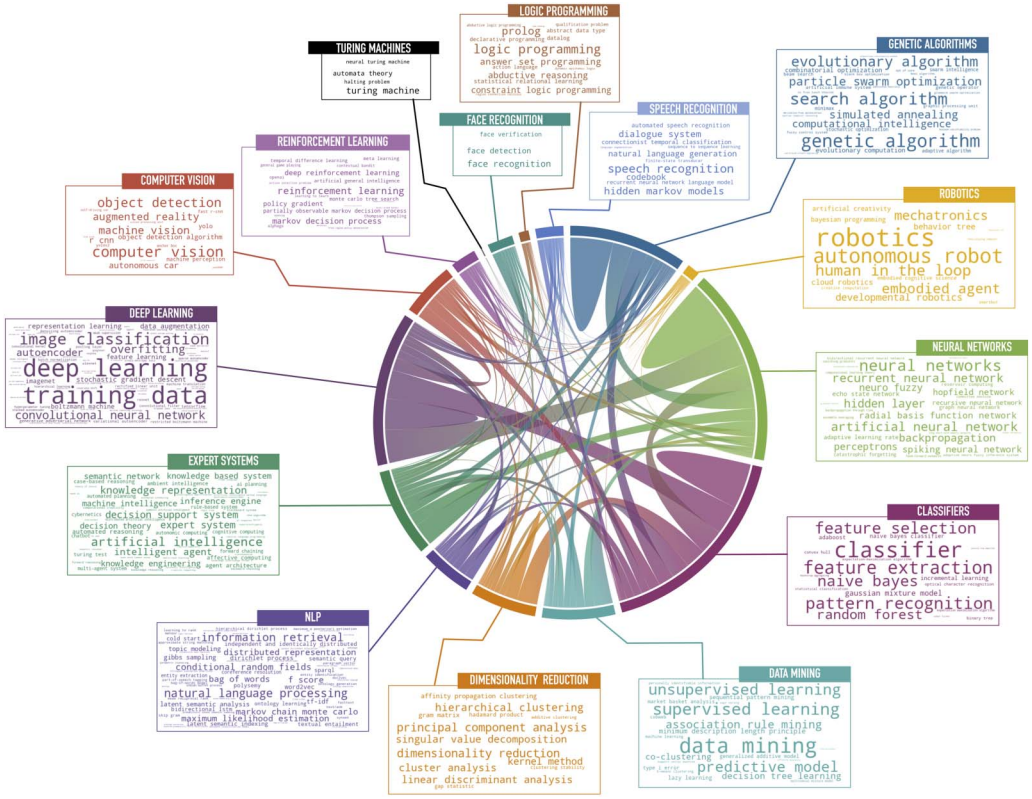


Figure 3. AI specialties and their relationships.

Downloaded from http://direct.mit.edu/qss/article-pdf/4/3/574/2200626/qss_a_00267.pdf by guest on 02 January 2024

Turing machines, reinforcement learning, computer vision, and DL. These structures are labeled according to their internal concept characterized by the term with the highest core position.

Some of these specialties show a significant degree of openness, demonstrating a flow of knowledge from one domain to the other. For example, the DL specialty has several semantic relationships with both neural networks and classifiers. Classifiers and dimensionality reduction are strictly related, the latter being a particular form of classification. Computer vision has connections with the classifier class but also with DL, as indeed computer vision was one of the first application domains of DL techniques.

The different AI specialties are also characterized by different temporal patterns that better define the temporality of the knowledge flows between them. In Figure 4 we show the annual time series of the number of publications in each area. AI general terms, as well as Turing machine and logic programming (symbolic AI), widely diffused in the early days, disappeared after the AI winter—around 1995—but the "expert systems" specialty (together with agent-based systems) started to emerge. Just afterward, we observe the rapid growth of specialties such as neural networks, data mining, optimization, and face recognition. Finally, the last two decades have seen the rapid decline of specialties such as optimization and dimensionality reduction, parallel to the extremely fast development of DL. Analyzing the relationships among specialties, although optimization research does not seem to enter new combinations with keywords

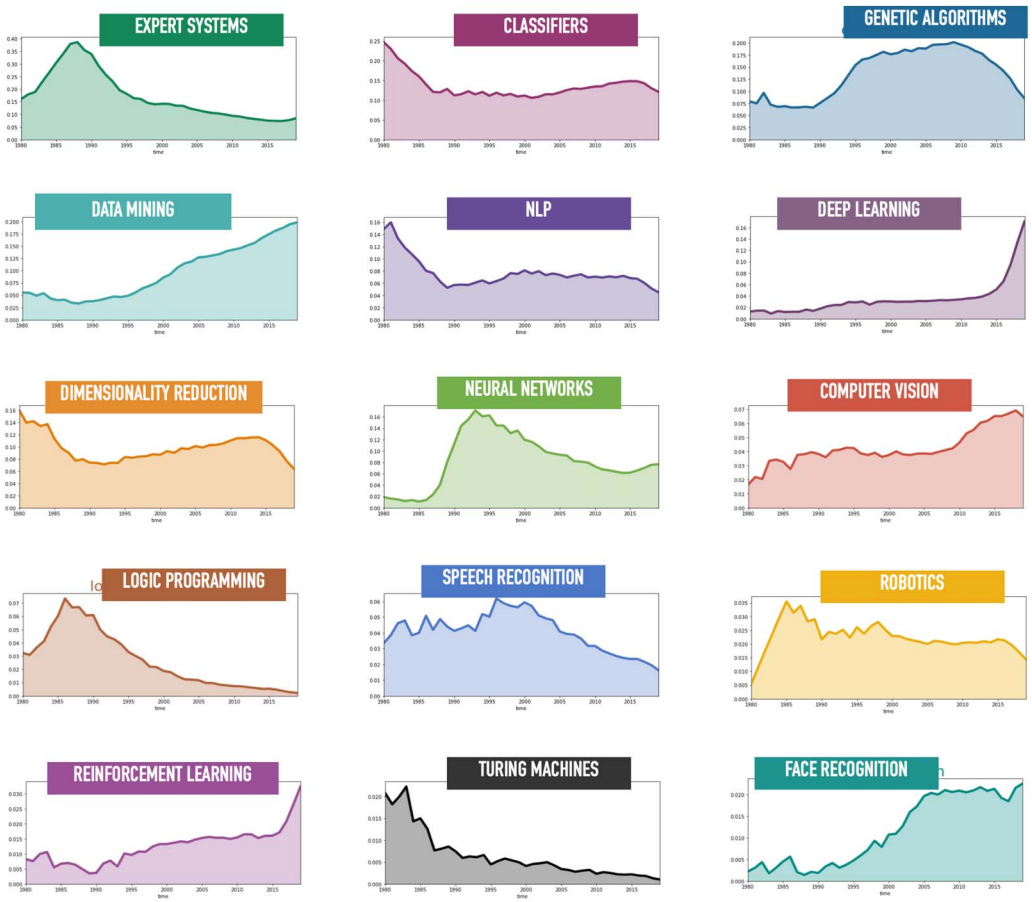


Figure 4. The timeline of AI specialties.

Downloaded from http://direct.mit.edu/qss/article-pdf/4/3/574/2200626/qss_a_00267.pdf by guest on 02 January 2024

in emerging areas (indicating a gradual fading of research interest in this domain), dimensionality reduction is being gradually recombined with DL concepts.

This global picture of the historical patterns of AI specialties, and in particular the striking emergence of DL, also describes the "competition" dynamics and the shift from two "dominant designs" driving AI research from the symbolic to the connectionist approach, well documented in Cardon et al. (2018).

3.2. AI: From Development to Applications

The study of our database highlights an important, and perhaps somewhat surprising, fact: Almost half of the publications included (48%) are associated with disciplines outside the native fields of computer science, mathematics, or statistics. This section strictly focuses on this subset of the corpus: the applied side of AI.

For each year starting from 1970, we calculate the moment of inertia of the corpus. As outlined, this indicator measures the dispersion of the corpus around the native disciplines. With this measure, the historical dynamics of AI appears as an oscillation between periods marked by forms of disciplinary dispersion followed by periods of disciplinary concentration.

Figure 5A shows that before 1988, AI was present in numerous disciplines beyond computer science, mathematics, and statistics (high moment of inertia). The so-called native disciplines were not the exclusive founders of AI, whose origins appear to be much more

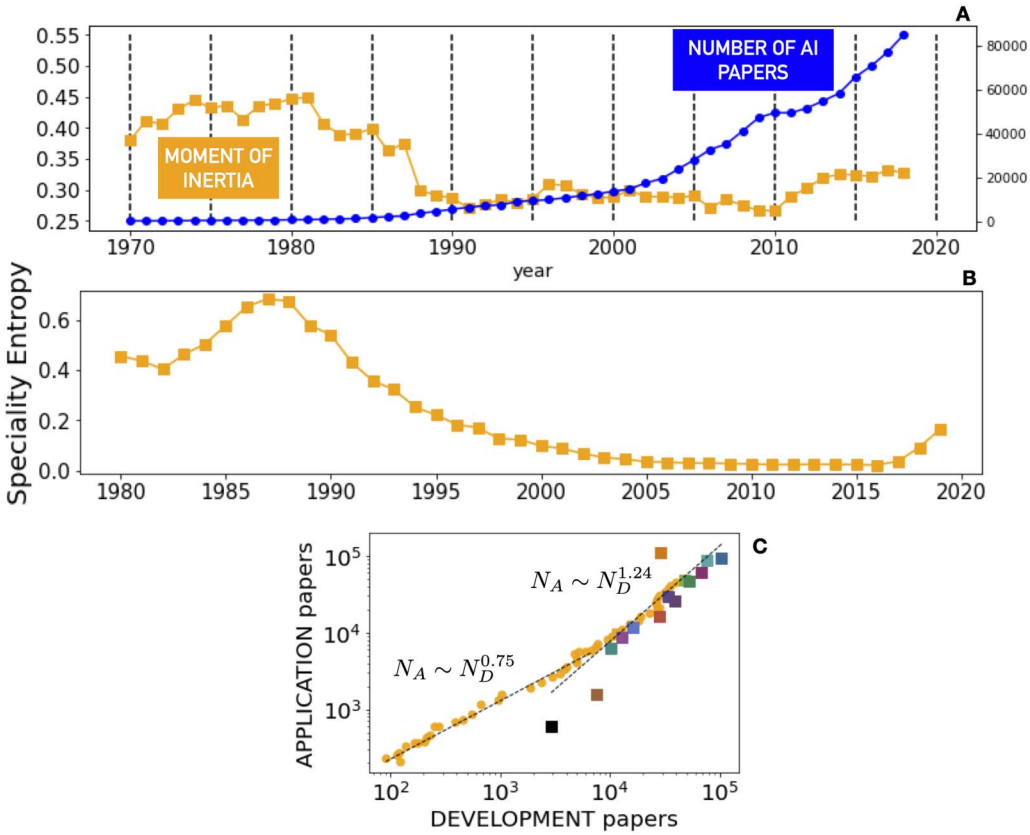


Figure 5. AI development versus applications. A: Yellow line: moment of inertia. Blue line: total number of AI publications. B: relative entropy of the dispersion of AI into specialties. C: number of application papers versus number of papers in native disciplines. Each yellow point represents 1 year. Each square represents temporal aggregation at the level of AI specialties.

Downloaded from http://direct.mit.edu/qss/article-pdf/4/3/574/2200626/qss_a_00267.pdf by guest on 02 January 2024

interdisciplinary, with inputs from, among others, engineering, philosophy, and psychology. In 1988, a phase of concentration around the three native disciplines began, reaching a maximum in 2010 (low m_I). After 2010, the moment of inertia starts to increase again, indicating the gradual spreading of AI knowledge to other disciplinary domains, more distant from the native disciplines. Of note, the recent diffusion process started with a delay of around 10 years after the take-off of scientific production in AI (around the year 2000).

Therefore, we observe cycles: a first phase of disciplinary diversity in the AI ecosystem, then concentration at the time of the AI winter and the emergence of the expert systems tradition, followed by a recent diffusion process (linked to the renewed interest in AI connected to DL applications).

In Figure 5C we analyze the relationship between the number of papers in the native disciplines and the number of application papers. We do this analysis both at the level of years (yellow points) and at the level of specialties (colored squares). The scaling shows the presence of two different regimes: When the number of native papers (in computer science, mathematics, and statistics) is low (<10,000) we observe a sublinear regime, in that applications grow more slowly than the development of new concepts and tools. When the number of papers in native disciplines is high (and this trend is also confirmed by the aggregated values in terms of specialties) we are in a superlinear regime: The number of applications grows faster than the development production (i.e., each paper in a native discipline gives rise to more than one application paper).

These cycles of disciplinary attention to AI are strongly connected to the technological cycles (Anderson & Tushman, 1990) of the epistemic production of AI. In Figure 5B we represent the relative entropy (KL divergence) of the dispersion of AI into its specialties with respect to the time-aggregate relative size of the specialties (which, given the exponential growth in terms of papers, almost represents the actual scenario).

In the first phase, we observe a high concentration in a few specialties, in particular, observing Figure 4, expert systems and dimensionality reduction. The concentration decreases quickly starting from the 1990s, reaching the baseline of the actual scenario. This peak in what, following a tradition of research in innovation studies, may be called the *technological trajectory of AI* (Dosi, 1982), represents a moment of paradigm transition, well explained in Cardon et al. (2018), where symbolic AI gets overtaken by the new paradigm of neural networks (and afterward, DL). The impact of this paradigm change spills over to the applied disciplines, which progressively lose interest in symbolic AI technologies. The new paradigm of neural networks, with all its connected specialties (such as face recognition, data mining, and computer vision), becomes the dominant design (Anderson & Tushman, 1990) from the early 2000s, when a new phase of gradual incremental progress of the emerging paradigm starts.

The 10-year delay that we observe between the development phase and the applications could be explained as a case of lock-in dynamics (Arthur, 1989; Lascialfari, Magrini, & Cabanac, 2022): It could be that in those 10 years, the costs of switching to the new AI technologies (based on neural networks and deep learning) for scientists in, say, medicine or psychology was too high—in terms of unavailability of adequate training, of low computational power etc.—so that they preferred to stick to more traditional methods and techniques.

After describing the aggregate scenario, we explore the disciplinary composition of the AI applied ecosystem. As could be expected, Figure 6 shows that technological disciplines (such as engineering, robotics, and imaging) are the sectors in which AI is more largely overrepresented.

Downloaded from http://direct.mit.edu/qss/article-pdf/4/3/574/2200626/qss_a_00267.pdf by guest on 02 January 2024

Figure 6. AI application disciplinary landscape. Upper left: AI score versus fraction of AI papers for all disciplines. The size of the points is proportional to the total number of AI publications. Color indicates the WoS disciplinary category. Upper right: boxplot of the AI score of the journals in the six WoS disciplinary categories. Lower boxes: list of the top 10 journals publishing AI papers in each category.

Some technical medical disciplines, such as neuroimaging and medical informatics, are also rapidly adopting AI methodologies. Our disciplinary AI score shows that the physical sciences are not always well positioned. For example, AI techniques are less prevalent in physics than in some social sciences fields such as (following WoS classification) management, geography, and linguistics. Only the arts and humanities are consistently underrepresented.

This pattern can be also described at the granularity scale of journals where we observe a dominance of AI in technology and multidisciplinary outlets. It is important here to keep in mind that having excluded conferences for which categorization is less fine, the physical sciences, life sciences, and social sciences are mostly at the same level. However in all categories, the journals that publish most AI-related papers are among those that specifically focus on computational methods. In the multidisciplinary journal landscape, especially noticeable is the presence of journals related to complex systems, which like AI can be seen as a technological platform, with multiple contact points with AI techniques (Li Vigni, 2021).

Concerning the diffusion process of AI we can see that the disciplinary ranking has been quite stable since the late 1990s (Figure 7). To measure distance among rankings, we use the "ranked Jaccard similarity" introduced in (Gargiulo, Caen et al., 2016). In the lower plots of Figure 7 we can observe some prototypical trajectories of disciplines that experienced an important change in the ranking from the 1990s until today. Some disciplines (above all in social science) experienced a strong decrease being strongly connected to declining AI

Downloaded from http://direct.mit.edu/qss/article-pdf/4/3/574/2200626/qss_a_00267.pdf by guest on 02 January 2024

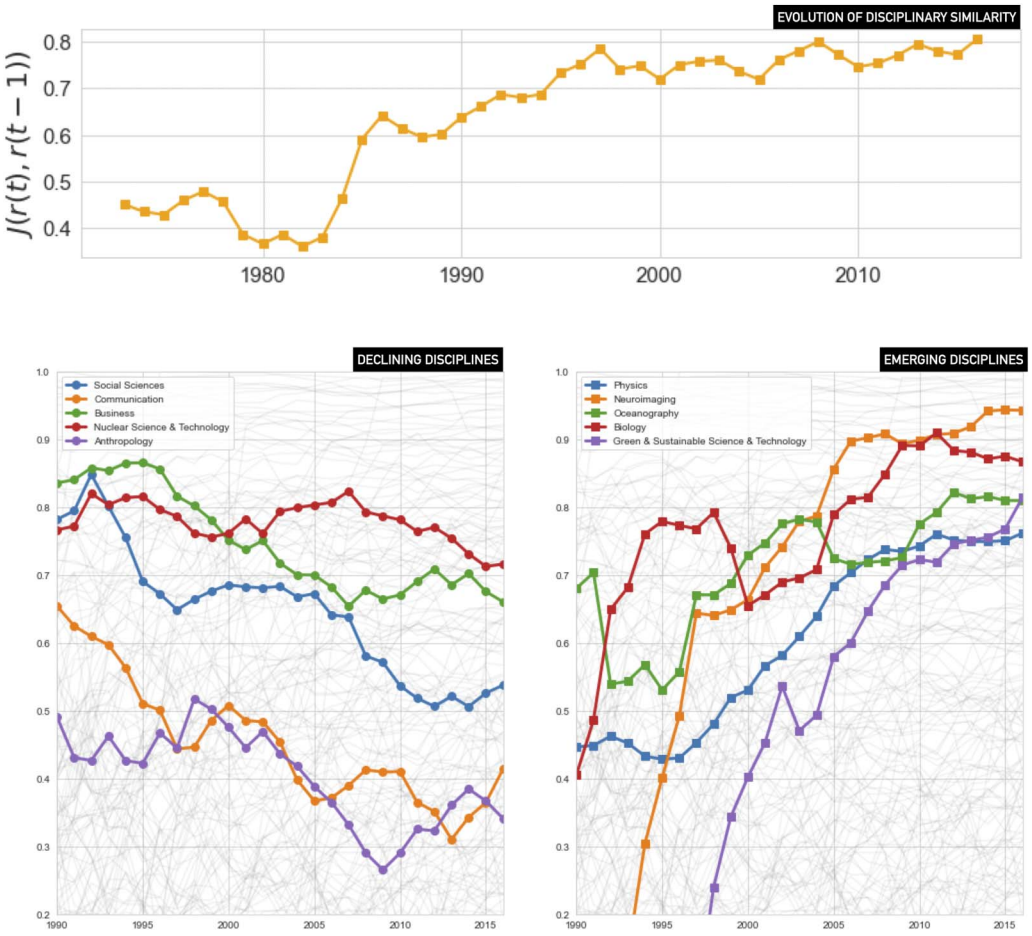


Figure 7. Disciplinary ranking of AI. Upper plot: ranked Jaccard similarity between disciplinary ranks in two subsequent years. Lower plots: Disciplines declining in the ranking and climbing the ranking.

specialties such as symbolic AI and expert systems. Disciplines such as physics and biology show a periodic growth (with a constant trend) in AI adoption, but others, such as neuroimaging and green and sustainability technologies, display a sudden climbing of the ranking after their creation.

Examining specialties in more detail (Figure 8), we observe that "dimensionality reduction" is generally the most widespread in applications, in quantitative terms (fraction of applications with respect to native papers) and in terms of disciplinary distance from native disciplines. Instead, the performance of "optimization" (the more represented specialty in terms of total number of publications, as shown by the size of the point) is high in terms of fraction of applications, but very low in terms of the moment of inertia; namely, it is largely applied in disciplines that are close to the native ones.

The different knowledge domains have very diverse profiles in terms of the adoption of AI specialties. In the arts and humanities, applications are mostly related to expert systems. The social sciences have strong interest in four AI specialties, namely machine learning, dimensionality reduction, expert systems, and natural language processing (NLP). Physical sciences, as well as life sciences and multidisciplinary frameworks, adopt dimensionality reduction, classifiers, and machine learning. Technology disciplines have a much more uniform

Downloaded from http://direct.mit.edu/qss/article-pdf/4/3/574/2200626/qss_a_00267.pdf by guest on 02 January 2024

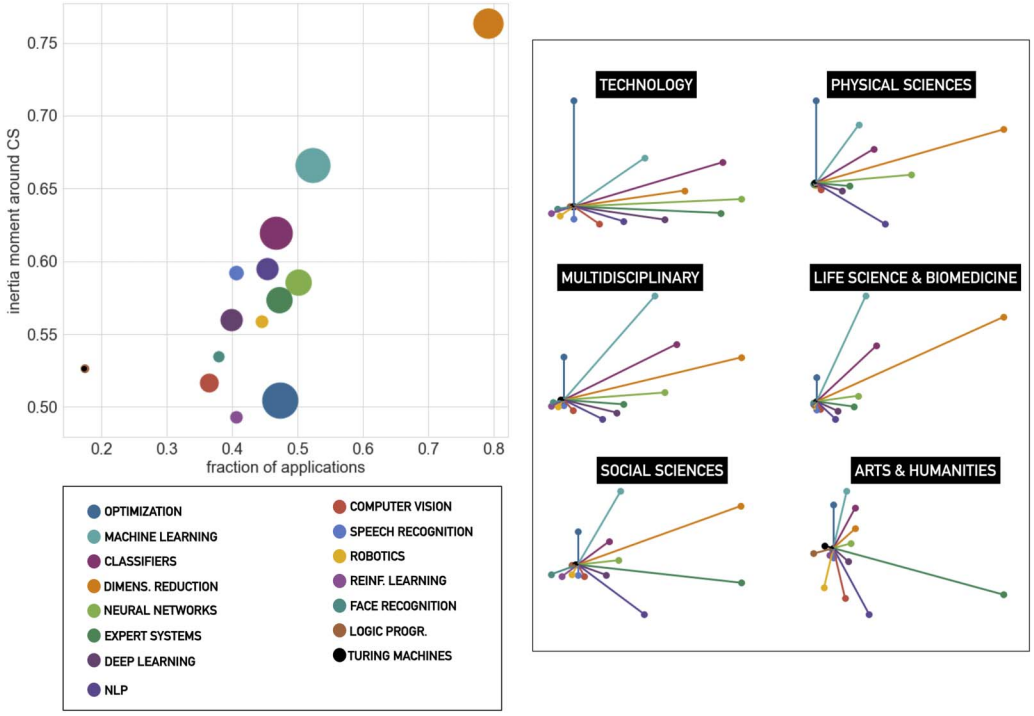


Figure 8. The disciplinary landscape of AI specialties. Left: Moment of inertia around native disciplines versus fraction of applications, for all AI specialties. Right: share of all the different AI specialties in the main knowledge domains.

distribution of AI specialties. Optimization is relevant only for technology and, to a lesser extent, for the physical sciences.

3.3. Authors' Collaborations in the AI Landscape

The last part of this study analyzes the collaboration patterns driving the diffusion of AI. The basic question we address is whether the writing process of papers applying AI involves the direct collaboration of AI developers and experts in the application domains.

We assign to each author an AI score, A_I, as described. The distribution of authors' AI scores shows two distinct peaks on the extremes, one around 0 (authors never publishing in AI journals) and the second around 1 (authors only publishing in AI journals). We divide the authors into groups according to the quartiles of their AI score: The first group (Q0) only contains authors publishing in out-disciplinary journals ($A_I = 0$), the last one authors only publishing in AI journals (Q3, $A_I = 1$), with the two intermediate groups (Q1 and Q2) containing authors having respectively AI scores lower and higher than 0.7 (Figure 9A).

From the histograms in Figure 9D we can also observe that authors in the first group (Q0) mostly publish in biomedical disciplines, but also astrophysics. Authors in Q1 are mostly specialized in interdisciplinary mathematical and technological applications (such as mathematical biology, operation research, energy, and telecommunications). Finally, Q2 includes mostly researchers in general engineering.

Looking at Figure 9C, authors in Q0 have the highest moment of inertia around native disciplines but at the same time have a low global level of interdisciplinarity. They interact with few disciplines, quite close to each other. Q1 includes most authors involved in interdisciplinary

Downloaded from http://direct.mit.edu/qss/article-pdf/4/3/574/2200626/qss_a_00267.pdf by guest on 02 January 2024

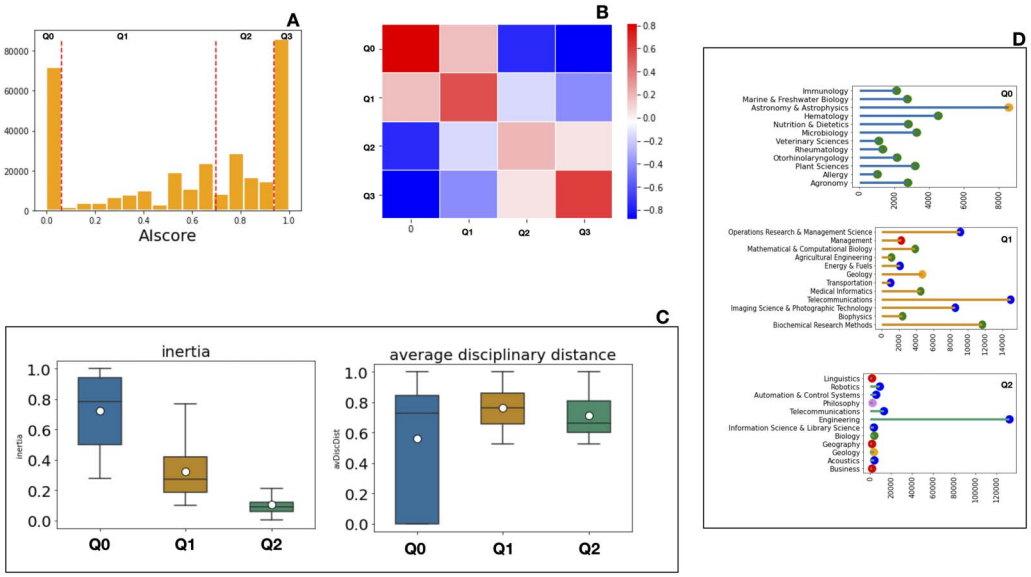


Figure 9. AI collaborations. A: Distribution of the authors' AI scores, and identification of the quartiles. B: Collaborations between groups of authors. The map refers to the comparison with the multinomial expectations for the links. C: Box plots of the moment of inertia and of the interdisciplinary distance for each group of authors. D: Top 10 disciplines (sorted according to the disciplinary AI score in the group) in which the authors of each group specialize. The color of the points represents the knowledge domain associated to the discipline.

collaborations. Q2 authors publish in disciplines close to the native ones (low moment of inertia) but at a quite large distance from them.

Then we analyze the author collaboration network (ACN), the weighted graph structure where all authors present in the database are linked to all the other authors they collaborated with. Weight is given by the number of joint publications. Aggregating at the level of the four AI score groups of the authors, we calculate the fraction of collaborations between each class and all the others. To avoid the bias that might result from the different sizes of the groups, we compare these values to the expected multinomial distribution of these links (Figure 9B). As we can observe from the figure, AI authors (those in the third and fourth quartiles of the AI score distribution, Q2 and Q3) mostly collaborate among themselves. Authors in Q2, as observed before, are mostly involved in engineering disciplines and are therefore more connected to AI developers. In contrast, disciplinary experts (Q0 and Q1) have few collaborations with AI developers (Q3) and collaborate among them.

This highlights a clear separation between the scholars who do research *on* AI and those who do research *with* AI. Application of AI to other disciplines is not primarily driven by direct collaborations between authors with different backgrounds. However, there are some direct collaborations among the second and the third groups. To some extent, authors in these groups constitute a bridge between the theoretical development of AI and the disciplinary applications.

To characterize the authors who are responsible for the connections among the different groups (i.e., the "brokers" that promote interdisciplinarity), we assign to each author a connectivity score, estimating how their collaborations are distributed across the four groups. This indicator is based on the relative entropy RH_i (Kullback–Leibler divergence) and measures the extent to which the collaborators of an author are localized within a single group (high relative entropy value) or are spread, exactly as the size distribution of the groups, across the four groups (relative entropy value next to zero). The distribution of this indicator is

Downloaded from http://direct.mit.edu/qss/article-pdf/4/3/574/2200626/qss_a_00267.pdf by guest on 02 January 2024

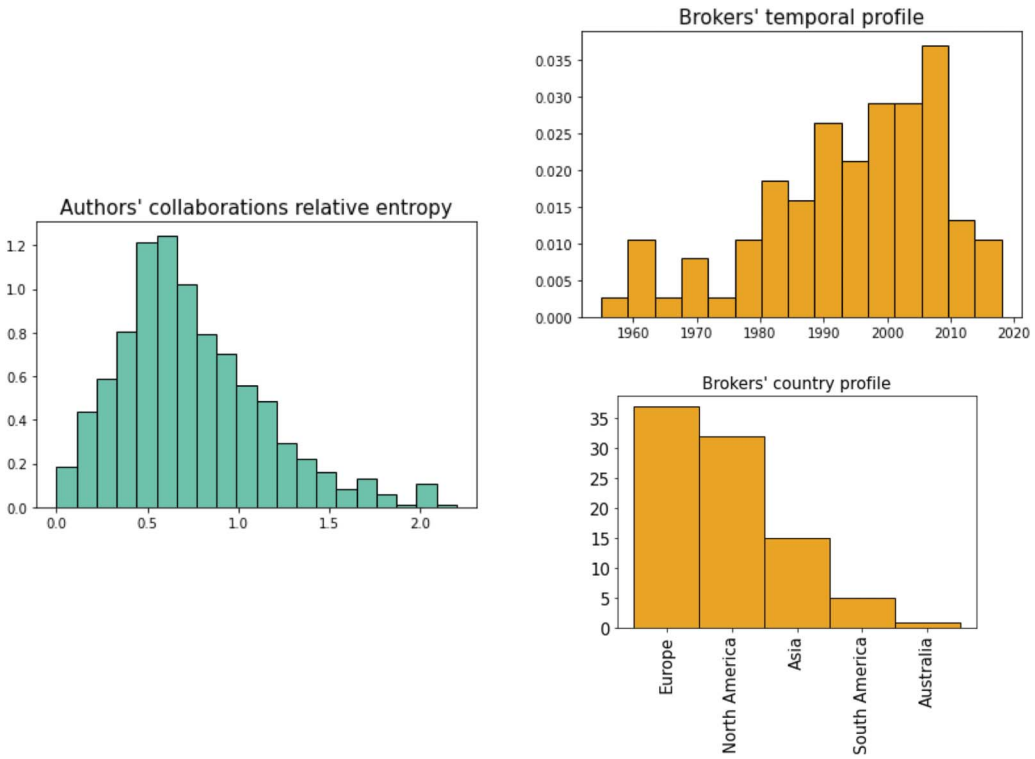


Figure 10. Characterizing brokers' profiles. Left: Distribution of the relative entropy indicator (RH_i) for all authors. Upper right: Distribution of the first publication year for the brokers. Lower right: Affiliation countries for the brokers.

reported in the left plot of Figure 10. The fact that this measure is highly concentrated on high values (authors' collaborations are mostly internal to a group) is in agreement with the observations relative to Figure 9.

We define as top "brokers" the 100 authors with the lowest relative entropy value ($RH < 0.019$). To give a first characterization of these authors we extract their profiles using the OpenAlex APIs (https://openalex.org).

This information provides a very first characterization of these "brokers": They mostly entered academia since the late 1990s (as we can observe from the distribution of the first publication years of the brokers in the upper right plot of Figure 10). Most of them are from Europe and the United States. Some 15% of the brokers are from private companies (mostly Google and Amazon).

4. DISCUSSION

We have built a bottom-up cartography of AI, based on an analysis of the scientific production on AI, from 1970 until 2017.

We started from an expert-based definition of AI through the identification of all the terms that different actors involved in AI studies use to structure the vocabulary of this research field (collected as mentioned, through an all-encompassing analysis of the web pages containing AI glossaries).

We show the effects of what historians of science call the "AI winters," periods in which interest in AI declined. We also show that the AI scientific platform came out of these crises

Downloaded from http://direct.mit.edu/qss/article-pdf/4/3/574/2200626/qss_a_00267.pdf by guest on 02 January 2024

thanks to the innovation dynamics arising from the recombination of specialties: for example, expert systems in the 1990s and DL around 2010 (DL was initially defined in 2006 and largely spread in 2012, with the ImageNet challenges).

The emergence of innovative specialties is partially due to the recombination of pre-existing building blocks from other domains but also from the dynamics of interdisciplinary participation.

We therefore studied the historical spread of AI into traditional disciplines. After the foundation of AI in a very interdisciplinary context, we identified a first phase of disciplinary "concentration" at the time of the AI winter during which the development of AI knowledge, previously distributed among several fields, condensed on the so-called native disciplines: computer science, mathematics, and statistics. Only more recently, starting from 2010, do we observe a new interdisciplinary phase of AI, with its gradual spreading to a larger number of fields where it is applied.

Transfer of AI knowledge from development to application is mediated by scientists in multiple fields of study, notably applied mathematical fields (for example mathematical biology), geology, biophysics, and some applied engineering fields. A relevant role has been played in the last decade by multidisciplinary journals where applied AI papers from several fields were published.

This study has offered a bottom-up definition of AI, building a dynamic cartography of this domain from its published traces. Its deeper intent is, more generally, to open the way toward a more ambitious effort to understand the role of AI in the transformation of the scientific ecosystem. For example, the structure of AI specialties would require an in-depth qualitative study based on interviews of the actors involved in each of them, in order to investigate their overall perception of AI and their positioning in this quantitative landscape. A study of this type would be necessary to globally assess the extent to which AI can be really defined as a scientific platform (Li Vigni, 2021) with a well-defined research program and objectives.

In this paper, we have mostly focused on the presence of AI terms in applied disciplines. We have adopted the designation of "native" AI disciplines from the current literature (Bianchini et al., 2022; Cockburn et al., 2018) but our findings challenge it by showing how, by its historical origins, AI was rather an interdisciplinary research area. This interdisciplinary contribution was mostly evident in the historical practices commonly known as symbolic systems. Later, different scientific fields have become, in turn, the central originating domains and applicators of AI knowledge; for example, operations research, which was for a long time one of the core actors of AI applications related to expert systems. A deeper historical analysis of the disciplines that developed AI, and the specialties of AI they focused on, would be worth studying.

One way to investigate this question would be based on disciplinary case studies. A discipline can indeed be transformed by the introduction of a new set of knowledge, expanding its possibilities. Likewise, serendipitous interactions with external fields could spark new ideas. For example, neuroscience could be considered in principle as an originating domain of AI, notably concerning the development of neural network architectures, but the centrality of neuroscience journals in AI scientific production will need to be ascertained in detail.

This study can also provide some interesting considerations and open research directions concerning science policies. The emergence of AI specialties, competition between the connectionist and the symbolic approaches, and the effects of the AI winters largely depend on research funding programs (Cardon et al., 2018). In the last decade, several funding agencies

Downloaded from http://direct.mit.edu/qss/article-pdf/4/3/574/2200626/qss_a_00267.pdf by guest on 02 January 2024

have promoted research in AI as a central challenge. This is probably one of the drivers of the recent interdisciplinary diffusion of studies applying AI techniques, as has also been the case concerning research on the COVID-19 pandemic (Abbonato, Bianchini et al., 2023; Chen & See, 2020). However, the data used for this study are insufficient for an estimation of the role of such funding programs in promoting AI-based research. This kind of information could be addressed, for example, by exploiting the Dimensions data set (https://www.dimensions.ai).

Another possible driver of the growth of AI applications can be more deeply connected to technological advancements due to the development of popular programming languages (such as Python or R), allowing an easy implementation of AI techniques even for users with basic informatic knowledge. This hypothesis could be tested by analyzing online forums, such as Stack Overflow.

Finally, another hypothesis could be the more and more frequent creation of interdisciplinary master's level courses allowing students, in different disciplines, to be exposed to AI methodologies from the beginning of their studies. The question of the transformation (or adaptation) of university programs with AI, and computation in general, can be explored with a qualitative approach based on investigations of university tracks in diverse disciplinary institutions.

On the other hand, as the number of papers using AI increases, the impact of this research in the scientific ecosystem is still unclear. It has been shown in Abbonato et al. (2023) that interdisciplinary papers using AI in COVID-19 research are not "performing" well in terms of the classical scientometrics indicators, and in particular concerning the number of citations, but they have good capacity for breaking disciplinary barriers, being cited by several other disciplines. This aspect should be central when research schemes promote this kind of interdisciplinary collaboration, which could be risky if evaluated on the basis of the usual conventional metrics.

The "risk," in terms of impact, connected to high interdisciplinarity could also explain the collaboration patterns observed in this paper, where disciplinary "homophily" seems to dominate collaborations.

Therefore, to promote interdisciplinarity, funding programs alone may not be a sufficient incentive, without deep reflection on the need of defining new criteria to evaluate researchers' careers that cross disciplinary borders.

This paper gives therefore important hints on how to navigate the AI scientific ecosystem in order to select potentially interesting case studies for subsequent analyses.

AUTHOR CONTRIBUTIONS

Floriana Gargiulo: Conceptualization, Formal analysis, Methodology, Visualization, Writing—original draft. Sylvain Fontaine: Data curation, Formal analysis, Writing—review & editing. Michel Dubois: Conceptualization, Writing—review & editing. Paola Tubaro: Conceptualization, Writing—review & editing.

COMPETING INTERESTS

The authors have no competing interests.

FUNDING INFORMATION

This work is partially founded by the ANR project ScientIA. The PhD fellowship of S.F. is founded by the CNRS-MiTi program EpiAI.

Downloaded from http://direct.mit.edu/qss/article-pdf/4/3/574/2200626/qss_a_00267.pdf by guest on 02 January 2024

DATA AVAILABILITY

The data set used in this manuscript is extracted from the latest Microsoft Academic Graph dump, available at https://zenodo.org/record/4617285#.YlUoaS2ubs0.

REFERENCES

Abbonato, D., Bianchini, S., Gargiulo, F., & Venturini, T. (2023). Questioning the impact of AI and interdisciplinarity in science: Lessons from Covid-19. *arXiv:2304.08923*. https://doi.org/10.48550/arXiv.2304.08923

Aghion, P., Jones, B. F., & Jones, C. I. (2018). Artificial intelligence and economic growth. In A. Agrawal, J. Gans, & A. Goldfarb (Eds.), *The economics of artificial intelligence: An agenda* (pp. 237–282). Chicago, IL: University of Chicago Press.

Anderson, P., & Tushman, M. L. (1990). Technological discontinuities and dominant designs: A cyclical model of technological change. *Administrative Science Quarterly*, *35*(4), 604–633. https://doi.org/10.2307/2393511

Annoni, A., Benczur, P., Bertoldi, P., Delipetrev, B., De Prato, G., ... Junklewitz, H. (2018). *Artificial intelligence: A European perspective*. https://doi.org/10.2760/11251

Arencibia-Jorge, R., Vega-Almeida, R. L., Jiménez Andrade, J. L., & Carrillo-Calvet, H. (2022). Evolution and multidisciplinarity of artificial intelligence research. *Scientometrics*, *127*, 5139–5158. https://doi.org/10.1007/s11192-022-04477-5

Arthur, W. B. (1989). Competing technologies, increasing returns, and lock-in by historical events. *Economic Journal*, *99*(394), 116–131. https://doi.org/10.2307/2234208

Baruffaldi, S., van Beuzekom, B., Dernis, H., Harhoff, D., Rao, N., ... Squicciarini, M. (2020). *Identifying and measuring developments in artificial intelligence: Making the impossible possible*. OECD Science, Technology and Industry Working Papers.

Baum, Z. J., Yu, X., Ayala, P. Y., Zhao, Y., Watkins, S. P., & Zhou, Q. (2021). Artificial intelligence in chemistry: Current trends and future directions. *Journal of Chemical Information and Modeling*, *61*(7), 3197–3212. https://doi.org/10.1021/acs.jcim.1c00619, PubMed: 34264069

Bianchini, S., Müller, M., & Pelletier, P. (2020). Deep learning in science. *arXiv:2009.01575*. https://doi.org/10.48550/arXiv.2009.01575

Bianchini, S., Müller, M., & Pelletier, P. (2022). Artificial intelligence in science: An emerging general method of invention. *Research Policy*, *51*(10), 104604. https://doi.org/10.1016/j.respol.2022.104604

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

Cardon, D., Cointet, J.-P., & Mazières, A. (2018). La revanche des neurones: L'invention des machines inductives et la controverse de l'intelligence artificielle. *Réseaux*, *5*(211), 173–220. https://doi.org/10.3917/res.211.0173

Chen, C. (2017). Science mapping: A systematic review of the literature. *Journal of Data and Information Science*, *2*(2), 1–40. https://doi.org/10.1515/jdis-2017-0006

Chen, J., & See, K. C. (2020). Artificial intelligence for Covid-19: Rapid review. *Journal of Medical Internet Research*, *22*(10), e21476. https://doi.org/10.2196/21476, PubMed: 32946413

Cockburn, I. M., Henderson, R., & Stern, S. (2018). The impact of artificial intelligence on innovation: An exploratory analysis. In A. Agrawal, J. Gans, & A. Goldfarb (Eds.), *The economics of artificial intelligence: An agenda* (pp. 115–146). Chicago, IL: University of Chicago Press. https://doi.org/10.7208/chicago/9780226613475.001.0001

Dosi, G. (1982). Technological paradigms and technological trajectories: A suggested interpretation of the determinants and directions of technical change. *Research Policy*, *11*(3), 147–162. https://doi.org/10.1016/0048-7333(82)90016-6

Färber, M. (2019). The Microsoft Academic knowledge graph: A linked data source with 8 billion triples of scholarly data. In *International Semantic Web Conference* (pp. 113–129). Cham: Springer. https://doi.org/10.1007/978-3-030-30796-7_8

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. https://doi.org/10.1162/qss_a_00183

Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., ... Barabási, A.-L. (2018). Science of science. *Science*, *359*(6379), eaao0185. https://doi.org/10.1126/science.aao0185, PubMed: 29496846

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. https://doi.org/10.1038/s42256-019-0024-5

Gargiulo, F., Caen, A., Lambiotte, R., & Carletti, T. (2016). The classical origin of modern mathematics. *EPJ Data Science*, *5*(1), 26. https://doi.org/10.1140/epjds/s13688-016-0088-y

Hassabis, D., Kumaran, D., Summerfield, C., & Botvinick, M. (2017). Neuroscience-inspired artificial intelligence. *Neuron*, *95*(2), 245–258. https://doi.org/10.1016/j.neuron.2017.06.011, PubMed: 28728020

Haugeland, J. (1985). *Artificial intelligence: The very idea*. Cambridge, MA: MIT Press. https://doi.org/10.2307/2219930

Huang, Y., 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. https://doi.org/10.1007/s11192-015-1638-y

Kauffman, S. A. (2000). *Investigations*. Oxford: Oxford University Press.

King, R. D., Rowland, J., Oliver, S. G., Young, M., Aubrey, W., ... Clare, A. (2009). The automation of science. *Science*, *324*(5923), 85–89. https://doi.org/10.1126/science.1165620, PubMed: 19342587

Kitchin, R. (2014). Big data, new epistemologies and paradigm shifts. *Big Data & Society*, *1*(1), 1–12. https://doi.org/10.1177/2053951714528481

Lascialfari, M., Magrini, M.-B., & Cabanac, G. (2022). Unpacking research lock-in through a diachronic analysis of topic cluster trajectories in scholarly publications. *Scientometrics*, *127*(11), 6165–6189. https://doi.org/10.1007/s11192-022-04514-3

Li Vigni, F. (2021). Complexity sciences: A scientific platform. *Science & Technology Studies*, *34*(4), 30–55. https://doi.org/10.23987/sts.97027

McCarthy, J. (1981). Epistemological problems of artificial intelligence. In B. L. Webber & N. J. Nilsson (Eds.), *Readings in artificial*

Downloaded from http://direct.mit.edu/qss/article-pdf/4/3/574/2200626/qss_a_00267.pdf by guest on 02 January 2024

intelligence (pp. 459–465). Elsevier. https://doi.org/10.1016/B978-0-934613-03-3.50035-0

McCarthy, J. (2004). What is artificial intelligence? https://www-formal.stanford.edu/jmc/whatisai

Monechi, B., Ruiz-Serrano, A., Tria, F., & Loreto, V. (2017). Waves of novelties in the expansion into the adjacent possible. *PLOS ONE*, *12*(6), e0179303. https://doi.org/10.1371/journal.pone.0179303, PubMed: 28594909

Nature. (2019). The scientific events that shaped the decade. *Nature*, *576*(7787), 337–338. https://doi.org/10.1038/d41586-019-03857-x, PubMed: 31853092

OECD. (2019). Artificial intelligence in society. https://www.oecd.org/publications/artificial-intelligence-in-society-eedfee77-en.htm

Serrano, M. Á., Boguñá, M., & Vespignani, A. (2009). Extracting the multiscale backbone of complex weighted networks. *Proceedings of the National Academy of Sciences*, *106*(16), 6483–6488. https://doi.org/10.1073/pnas.0808904106, PubMed: 19357301

Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013). Atypical combinations and scientific impact. *Science*, *342*(6157), 468–472. https://doi.org/10.1126/science.1240474, PubMed: 24159044

WIPO. (2019). *WIPO technology trends 2019: Artificial intelligence*. Geneva: World Intellectual Property Organization.

Downloaded from http://direct.mit.edu/qss/article-pdf/4/3/574/2200626/qss_a_00267.pdf by guest on 02 January 2024