Does Training in AI Affect PhD Students' Careers? Evidence from France

Pierre Boutros
Université Côte d'Azur, CNRS, GREDEG, France
pierre.boutros@univ-cotedazur.fr

Eliana Diodati
University of Torino, Italy
Collegio Carlo Alberto, Turin, Italy
eliana.diodati@unito.it

Michele Pezzoni
Université Côte d'Azur, CNRS, GREDEG, France
Observatoire des Sciences et Techniques, HCERES, Paris, France
michele.pezzoni@univ-cotedazur.fr

Fabiana Visentin
UNU-MERIT, Maastricht University, the Netherlands
f.visentin@maastrichtuniversity.nl

Abstract: The rise of Artificial Intelligence (AI) technologies urges us to better understand their impact on the labor market. This paper is the first to analyze the supply of individuals with AI training facing the labor market. We estimate the effect of AI training on the careers of the whole population of French PhD students in STEM who graduated between 2010 and 2018. To assess the unbiased effect of AI training, we compare the careers of PhD students trained in AI, with those of a control sample of similar students with no AI training. We find that AI training does not affect the probability of pursuing a research career after graduation. However, for those students who have AI training during the PhD and pursue a research career after graduation, we observe a clear path dependence in continuing publishing on AI topics. Moreover, in Computer Science, AI-trained students are less likely to end up in private organizations after graduation than their non-AI counterparts. AI training is also detrimental to patenting activity in Computer Science while it fosters patenting in other disciplines. Finally, AI training stimulates mobility abroad after graduation but only in disciplines other than Computer Science.

1. Introduction

Artificial Intelligence (AI) technologies are profoundly transforming our social and economic systems, spreading to most disciplines (Agrawal, Gans, and Goldfarb 2019).

The growing importance of AI has made the knowledge related to those technologies a highly valuable asset for individuals competing for job positions (Squicciarini and Nachtigall 2021; Acemoglu et al. 2022; Lane, Owen-Smith, and Weinberg 2024). Extant studies have focused their attention on the demand side of the labor market, documenting companies' increasing interest in individuals with AI knowledge (Squicciarini and Nachtigall 2021; Acemoglu et al. 2022; Acemoglu and Restrepo 2018). However, little is known about the supply of individuals with AI knowledge.

The literature investigating wage differentials identifies individuals' education attainment as crucial in shaping the labor market supply (Goldin and Katz 2007; Furman 2018). Similarly, in the context of AI, education is considered a key factor for addressing the mismatch between workers' existing skills and the new skills required by AI technologies (Agrawal, Gans, and Goldfarb 2019). Therefore, over the last decade, countries' higher education systems have invested substantial resources in developing AI education programs ('AI Index Report 2021' 2021). Part of those resources has been devoted to train PhD students in AI at the frontier of science and technology.[1]

After completing their AI training, PhD students are expected to disseminate AI knowledge throughout society (Arora et al. 2023; Cyranoski et al. 2011; Mangematin and Robin 2003; Zolas et al. 2015). Studying graduates' careers allows us to tackle three fundamental issues that animate the current debate on the AI labor market. The first issue regards the effect of undertaking distinctive training, i.e., AI training, on the opportunity to take full advantage of graduates' qualifications by starting a research career (Cyranoski et al. 2011). The second issue regards the brain drain of AI-trained students from academia to the private sector due to the higher salaries and resources rewarding AI knowledge in the private sector (Jurowetzki et al. 2021). The third issue regards the impact of AI training on researchers' productivity after graduation (Bianchini, Müller, and Pelletier 2022).

[1] For instance, The *Agence Nationale de Recherche* (ANR) in France is co-financing 200 PhD contracts in AI at the national level in 2022: https://anr.fr/en/call-for-proposals-details/call/call-for-programmes-phd-contracts-in-artificial-intelligence/. Moreover, The *Interdisciplinary Institutes for Artificial Intelligence* (3IA) is securing support for research specialized centers in AI: https://instituts-3ia.fr/.

In our study, we ask *Does AI training affect graduates' probability of starting a research career? Does AI training affect graduates' probability of ending up in the private sector*? and *Does AI training affect graduates' scientific productivity, visibility, scientific network, and international mobility?*

We explore discipline heterogeneity by distinguishing between disciplines in which AI knowledge can be easily acquired after graduation, i.e., Computer Science, and those in which learning AI during the PhD training is expected to be a unique occasion for acquiring AI knowledge, i.e., Biology, Chemistry, Engineering, Geology, Mathematics, Medicine, and Physics. Indeed, the theoretical, methodological, and empirical knowledge foundations developed during a PhD in Computer Science are similar to the ones required to easily absorb AI knowledge after graduation (Nelson and Winter 1982; Wesely M. Cohen and Levinthal 2000; Wesley M. Cohen and Levinthal 1990). Therefore, in Computer Science, not undertaking AI-training during the PhD is not an obstacle to learning AI after graduation. On the contrary, for PhD graduates in disciplines other than Computer Science, e.g., a PhD in Biology, the theoretical, methodological, and empirical knowledge foundations developed during the PhD are far from the ones required to easily absorb AI knowledge after graduation, making the PhD period a unique occasion for acquiring AI knowledge.

To conduct our empirical analyses, we draw on a unique dataset covering 46,657 French PhD students in STEM disciplines who graduated between 2010 and 2018. We start by identifying students who acquired AI knowledge during their PhD training as those using AI-related keywords in their theses.

Establishing a causal link between AI training and careers is challenging because students who acquire AI knowledge during the PhD training might differ in many ways from students who do not acquire AI knowledge. The differences between the two groups may also affect their careers after graduation. We overcome this concern by implementing a Propensity Score Matching approach. For each PhD student who undertakes AI training, we find a 'control' student with similar characteristics but not undertaking AI training. We match AI students with similar controls considering the supervisor's characteristics, such as mentorship experience, publication records, previous AI knowledge, connections with the private sector, and fundraising experience. We also consider student's characteristics such as the university, scientific discipline, and defense year.

We find that AI training does not significantly affect the probability of pursuing a research career after graduation, both in Computer Science and other disciplines. However, AI-trained students in Computer Science are less likely to end up in private research organizations after

graduation than their non-AI counterparts. We do not observe significant differences between AI and non-AI-trained students when looking at their productivity after graduation measured by standard bibliometric indexes such as the number of publications, citations received, and coauthors. However, we observe for all disciplines a significant path dependence in continuing publishing on AI topics after graduation for those students who have acquired AI knowledge during the PhD. Moreover, having AI training is detrimental to patenting activity in Computer Science while fostering it in other disciplines. Finally, we observe that AI training stimulates mobility to other countries after graduation in disciplines other than Computer Science.

Our study contributes to the literature by embracing a recently suggested approach to studying the effects of governments' efforts in developing AI. As claimed by Lane and colleagues, "Any attempt to describe the economy-wide impact of public investment in AI would involve identifying the people at the heart of these investments" (Lane, Owen-Smith, and Weinberg 2024, 303). We consider PhD students with AI training as "the people at the heart of these investments," and by investigating their careers we track their "economy-wide impact". Doing so, we shed light on the supply of workers in the AI labor market, assessing how acquiring specific AI knowledge during training shapes careers. We present empirical evidence on three key issues animating the current debate on the AI labor market: the effect of undertaking distinctive training in AI on the probability of starting a research career after graduation, the potential brain drain of AI-trained students from academia to industry, and the effect of AI training on graduates' productivity. Additionally, we take advantage of micro-level individual data from an entire country to conduct an unbiased assessment of the relationship between AI training and post-graduate careers.

2. Theoretical framework

As an emerging general-purpose technology, AI is gradually reshaping the labor market, influencing employability, productivity, and mobility. This section discusses the existent literature focused on the impacts of AI on career trajectories (Section 2.1) and, more generally, on the evolution of the PhD graduates' career patterns (Section 2.2). Moreover, it explores how AI training is expected to influence PhD graduates' career outcomes across different scientific disciplines (Section 2.3).

2.1 The impact of AI on PhD graduates' careers

The integration of AI knowledge into the labor market creates a distinct set of competencies that enhance employability. Recent studies show that AI knowledge constitutes a highly

valuable asset for the workforce. For instance, using online job posting data, Squicciarini and Nachtigall (2021) and Acemoglu et al. (2022) show a relevant and gradual increase in AI-related job demand across all sectors since 2010. These trends encompass not only AI-specific jobs, requiring advanced skills such as Natural Language Processing (NLP) and deep learning skills but also job positions based on soft and generic AI-related competencies, such as communication, problem-solving, and creativity (Squicciarini and Nachtigall 2021).

While the literature has mainly focused on the demand for AI skills, the supply side is still unexplored. Education plays a crucial role in addressing the mismatch between the supply and demand of AI skills (Agrawal, Gans, and Goldfarb 2019), and governments around the world are promoting AI education at all levels ('AI Index Report 2021' 2021). In this context, doctoral programs are particularly important, as PhD graduates represent the highly advanced segment of the workforce and they are expected to disseminate AI knowledge throughout society (Arora et al. 2023; Cyranoski et al. 2011; Zolas et al. 2020; Mangematin and Robin 2003).

Further research is needed to understand whether public spending on AI education is effectively aligning the workforce composition to this increasing demand for AI knowledge. As suggested by Lane and colleagues, identifying individuals benefiting from AI public investments can help understand the economy-wide impact of these investments (Lane, Owen-Smith, and Weinberg 2024). Evaluating the career trajectories of individuals involved in AI education can provide valuable insights into the diffusion of this knowledge between academia and industry (Ahmed, Wahed, and Thompson 2023).

Recent studies in the economics of science are beginning to explore how AI is reshaping scientific discoveries and productivity across all domains of science and innovation (Bianchini, Müller, and Pelletier 2022; European Commission 2023; Cockburn, Henderson, and Stern 2018). Jurowetzki et al. (2021) provide empirical evidence of the *brain drain* from academia to the private sector by scientists with AI knowledge, confirming the industry's increasing interest in these skills. But little is known about the next generation of scientists recently trained at the frontier of knowledge. According to the 'AI Index Report 2021' (2021), in the last decades, the majority of AI PhD graduates in North America preferred to work in industry. However, as AI is spreading across all STEM disciplines, it is important to consider not only PhD graduates in AI-related disciplines, i.e., Computer Science, but also those in other STEM disciplines who have integrated AI.

2.2 PhD graduates' career trajectories

In most advanced countries, PhD programs have traditionally been designed as the highest educational degree, primarily aimed at training future generations of academics. The main emphasis has been on developing advanced scientific expertise and on preparing graduates to be at the frontier of knowledge for teaching and research jobs within universities. Students are generally selected through competitive application processes and are required to complete a high-level research thesis within three to six years, depending on the national context (OECD 2019). During this training period, students acquire a unique set of skills and competencies that constitute essential components of the professional profile for entering the labor market.

In recent decades, career options for junior researchers have increasingly gone beyond academia to include research and non-research jobs in industry, public institutions, and self-employment (Cyranoski et al. 2011; Geuna and Shibayama 2015; OECD 2023; Sarrico 2022). According to a recent study, almost 50% of researchers leave scientific careers within a decade, particularly women (Naddaf 2024). In France, only 44% of PhD graduates in 2018 stayed in academia three years after graduation, and 17% moved to research careers in the private sector (Ministère de l'Enseignement supérieur et de la Recherche 2024).

Several factors have contributed to this shift. One major factor is the increasing supply and the low demand for PhD in academia, which has been part of public and academic debate over the last decade (The Economist 2016; Cyranoski et al. 2011; Larson, Ghaffarzadegan, and Xue 2014; Malloy, Young, and Berdahl 2021; Sarrico 2022). This mismatch is connected to the rise of the knowledge economy (Foray and Lundvall 1998) and the resulting public policies promoting higher education and research aimed at training a highly skilled labor force (OECD 2021b). Universities often have incentives to produce more PhDs than the academic labor market can absorb (Larivière 2011). However, this oversupply creates intense competition for academic positions, exacerbating job insecurity and pushing many graduates to pursue careers outside academia (Andalib, Ghaffarzadegan, and Larson 2018; Bonnal and Giret 2010; Sarrico 2022; Geuna and Shibayama 2015).

Moreover, lower salaries in academia compared to other sectors make alternative careers more attractive (Sauermann and Roach 2014; OECD 2021b), causing many graduates to leave academia, potentially resulting in the loss of the most talented individuals (Waaijer et al. 2017; Geuna and Shibayama 2015; Auriol 2010). This is especially the case in disciplines where science and technology overlap (Stokes 1997; Ahmed, Wahed, and Thompson 2023), and researchers may prefer to renounce their freedom in conducting research for better-paid positions in industry (Aghion, Dewatripont, and Stein 2008).

Outside academia, the growing number of highly qualified junior researchers contrasts with stable R&D expenditures, meaning that not all of them will secure jobs in the research sector (OECD 2021a). Moreover, the transition outside of research is often challenging due to the mismatch between doctoral education and the needs and priorities of the industry (Cyranoski et al. 2011). PhD graduates also face challenges in aligning their highly specialized training with the broader needs of the industry. Companies often report that PhD graduates have overly specialized and theoretical preparation (OECD 2023). These issues are particularly critical because students, on average, start their PhD studies at 29 years old (OECD 2019), with very large differences across countries. Entering the private sector labor market in their thirties can be challenging (Auriol 2010; Auriol, Misu, and Freeman 2013), unless PhD graduates have developed specific skills that are attractive to companies and come from scientific disciplines closely aligned with the industry needs.

2.3 Discipline heterogeneity

The impact of AI knowledge on PhD career trajectories may vary by discipline. To explain these differences, we refer to the human capital theory (Becker 1964).

According to the former, investments in education and skill acquisition, like AI training, can improve individuals' performance leading to higher marginal returns in the labor market. However, the effort to absorb this knowledge and the returns when entering the labor market may differ across disciplines.

For PhD graduates in Computer Science, where AI and machine learning courses are usually included in their curricula, AI knowledge tends to be a standard competence, meaning most graduates have similar skills and may not necessarily distinguish themselves in the job market. AI knowledge can be easily acquired at any moment, even after graduation, because the required preparation and expertise are similar and complementary to AI knowledge (Nelson and Winter 1982; Wesley M. Cohen and Levinthal 1990). Therefore, in Computer Science, not undertaking AI training during the PhD is not an obstacle to learning it after graduation. On the contrary, for graduates in other STEM disciplines, where AI is still less common (Jordan and Mitchell 2015), the acquisition of AI knowledge may represent an important investment, leading to higher marginal returns when entering the labor market in both research and non-research roles. Indeed, for those PhD graduates, the theoretical, methodological, and empirical knowledge foundations developed during the PhD are far from the ones required to easily absorb AI knowledge after graduation, making the PhD period a unique occasion for acquiring AI knowledge.

3. Data

3.1 Data sources

Our study sample merges four data sources to create a unique dataset on French PhD graduates in the STEM disciplines. We extract detailed information on doctoral dissertations from *L'Agence Bibliographique de l'Enseignement Supérieur* (ABES), the French public institute responsible for maintaining the bibliographic records of French universities. The extraction process allows us to gather a large collection of STEM theses from 2010 to 2018, including the student's name, graduation institution, defense year, supervisor name, thesis title and abstract, and scientific discipline.

To collect bibliometric information, we match the names of students and supervisors with the authors listed in the *OpenAlex*[2] database and assign the corresponding journal articles.[3] For each identified journal article, we retrieve its title and abstract, publication year, and author list.

We collect information about research funding from the full list of individual grants awarded by the *Agence Nationale de la Recherche* (ANR)[4], the national funding agency in France.

For information on patents, we match the names of the PhD students and supervisors with the inventors' names listed on patent applications at the European Patent Office (EPO). For each patent, we retrieve the application filing year, the inventor's name and country, and the type of the applicant's institution.[5]

After merging the four datasets mentioned above, we obtained a unique dataset including 46,657 PhD students who graduated from French institutions between 2010 and 2018. Our sample also includes 20,703 distinct supervisors.

3.2 Tracing AI knowledge

To identify theses with AI content, we create a comprehensive list of AI keywords extracted from two recognized lists. The first list, containing 189 AI keywords, is taken from a recent OECD report aimed at developing a methodology for identifying AI-related publications and

---

[2] OpenAlex is a free and open catalog of the global research system. It contains "metadata for 209M works (journal articles, books, etc.); 213M disambiguated authors; 124k venues (places that host works, such as journals and online repositories)" (Priem, Piwowar, and Orr 2022). Compared to other scientific databases like Elsevier's Scopus and Clarivate's Web of Sciences, OpenAlex has about twice the coverage, an open and fast API, and is completely free. Moreover, OpenAlex is updated weekly, ensuring relevant and up-to-date information.

[3] More details about the matching progress can be found in Appendix A.

[4] https://anr.fr/

[5] While bibliography and patent data do not capture individuals who do not publish or patent post-graduation, efforts to match a subset of PhD graduates in our dataset with LinkedIn profiles have demonstrated a low level of success, limiting the utility of such data beyond traditional publication and patent metrics.

patents (Baruffaldi et al. 2020; Calvino et al. 2023). The second list, with 39 AI keywords, is extracted from a study aiming at identifying AI publications (Cockburn, Henderson, and Stern 2018). After refining the list[6], we obtain 272 keywords. In our analysis, students trained in AI are the ones whose thesis contains AI-related keywords. We calculate the corresponding variable, *AI student*, as equal to one if the student's thesis contains AI-related keywords in its title or abstract, zero otherwise. Overall, we identify 3,561 theses with AI content, and an equivalent number of *AI students*, out of 46,657 students in our study sample. 3,561 theses represent 7.63% of the total number of theses.

3.3 Tracing students' careers

To trace careers, we calculate the dummy variable *Research career,* which equals one if the student has at least one journal publication or patent application[7] within five years after the defense year, and zero otherwise.[8] We assume that students active in publishing or patenting are those who undertake a research career and are employed in universities, public or private research institutions, or companies' R&D departments. We consider the five years after the defense year (t) as the years elapsed between $t+2$ and $t+6$. Looking at articles or patents starting from $t+2$ allows us to avoid considering as outcomes relevant to trace the career, publications or patents that are the result of the thesis work but appear after graduation due to publication delays. To assess the impact of AI training on PhD graduates who pursue a research career, we consider several outcomes measured within the same time frame, i.e., between $t+2$ and $t+6$. Specifically, we examine if they pursue a research career solely in private organizations. To do so, we define the variable *Private career* as a dummy variable that equals one if all their affiliations reported in the publication or patent documents published between $t+2$ and $t+6$ refer only to private organizations. We also define the variable *Private career/collaboration* as a dummy variable that equals one if all the students' affiliations reported in the publication or patent documents published between $t+2$ and $t+6$ refer to at least one private organization. Finally, we measure the productivity of the students who undertake a research career using seven measures calculated between $t+2$ and $t+6$: the total number of publications (*N. publications*), the number of AI-related publications (*N. AI publications*), if they hold at least

[6] We include the plural version of keywords that were initially presented in their singular form, and we drop those that are overly generic. More details can be found in Appendix B.
[7] We consider patent applications at the European Patent Office.
[8] In Appendix D, we validate our proxies for career outcomes by analyzing the information reported on LinkedIn for a subset of 100 PhD students.

one patent (*At least one patent)*, the number of coauthors they have (*N. of coauthors)*, if they have at least one international affiliations (*Abroad)*, if they have ties to institutions in the U.S. or China (*U.S. or China*), and the total number of citations received (*Citations*).

3.4 Descriptive statistics

We distinguish the 46,657 PhD students who graduated in the STEM disciplines in France between 2010 and 2018 into two groups: graduates in Computer Science and graduates in other disciplines, i.e., Biology, Chemistry, Engineering, Geology, Mathematics, Medicine, and Physics. Table 1 shows a consistent annual increase in the share of AI theses in Computer Science. The percentage of AI theses reached 37.52% of total theses in 2018. During the same period, the share of AI theses in other STEM disciplines remains substantially lower but is gradually increasing, reflecting the growing use and integration of AI technologies across different disciplines (Jordan and Mitchell 2015). Additionally, a decline in the proportion of students starting a research career after graduation can be observed across all disciplines, which could be attributed to the oversupply of PhDs, leading graduates to seek employment outside of research positions (Cyranoski et al. 2011). When comparing the two groups, graduates in Computer Science are less likely to pursue a research career compared to their peers from other STEM disciplines.

Table 1: Share of AI thesis and share of students pursuing a research career by year

	Computer Science (5,512 theses)		*Other disciplines* (41,145 theses)	
Defense year	AI student	Research career	AI student	Research career
2010	23.38%	39.92%	4.41%	60.01%
2011	26.68%	41.96%	4.27%	58.08%
2012	28.91%	41.64%	4.48%	57.28%
2013	28.86%	39.64%	4.77%	57.08%
2014	26.14%	44.88%	4.57%	57.50%
2015	30.19%	38.05%	4.74%	57.28%
2016	31.02%	39.21%	4.51%	57.32%
2017	33.90%	36.43%	4.73%	55.50%
2018	37.52%	38.34%	5.12%	54.58%
All	30.03%	39.86%	4.63%	57.09%

To further explore these initial observations, Table 2 presents detailed descriptive statistics on career trajectories and productivity metrics for students who graduated with and without acquiring AI knowledge during the PhD. We observe that AI graduates and non-AI graduates do not show a significant difference in the probability of doing a research career in Computer Science, while AI graduates have a lower probability of pursuing a research career in other disciplines. In these other STEM disciplines, AI graduates also show a preference for

collaboration with private research institutions or committing exclusively to private-sector research. Graduates with AI knowledge are likely to continue publishing in AI during their career, both in Computer Science and other disciplines. Graduates with AI knowledge in other disciplines have a higher probability to patent, but at the same time, less collaborations and citations. The differences observed between the two groups align with our hypothesis stating that the impact of acquiring AI knowledge differs across disciplines.

Table 2: Career type and productivity by AI thesis content and discipline

	Non-AI students	AI Students	P-value	N. of Non-AI students	N. of AI students	N. of theses
Computer Science						
Research career	0.39	0.41	0.25	3,857	1,655	5,512
Conditional on pursuing a research career (Research career=1)						
Private career/collaboration	0.19	0.17	0.25	1,518	679	2,197
Private career	0.09	0.08	0.22	1,518	679	2,197
Research career productivity						
N. of publications	3.75	4.06	0.31	1,518	679	2,197
At least one patent	0.11	0.09	0.16	1,518	679	2,197
N. AI publications	0.54	1.40	0.00	1,518	679	2,197
N. of coauthors	11.89	13.39	0.22	1,518	679	2,197
Abroad	0.48	0.48	0.84	1,518	679	2,197
U.S. and China	0.13	0.12	0.68	1,518	679	2,197
Citations	5.41	6.48	0.08	1,518	679	2,197
Other disciplines						
Research career	0.57	0.52	0.00	39,239	1,906	41,145
Conditional on pursuing a research career (Research career=1)						
Private career/collaboration	0.15	0.18	0.01	22,502	988	23,490
Private career	0.04	0.08	0.00	22,502	988	23,490
Research career productivity						
N. of publications	5.75	5.44	0.27	22,502	988	23,490
At least one patent	0.09	0.14	0.00	22,502	988	23,490
N. AI publications	0.15	1.14	0.00	22,502	988	23,490
N. of coauthors	26.73	22.10	0.00	22,502	988	23,490
Abroad	0.48	0.51	0.10	22,502	988	23,490
U.S. and China	0.15	0.17	0.11	22,502	988	23,490
Citations	11.12	8.87	0.00	22,502	988	23,490

NOTE: all career-related variables are calculated in $[t+2, t+6]$, where t is the thesis defense year.

4. Methodology

4.1 Estimation strategy

Although Table 2 informs us about the relationship between acquiring AI knowledge during the PhD and subsequent career outcomes, it does not allow us to claim a causal relationship between AI knowledge acquisition and career outcomes. Indeed, it is possible that students who acquire AI knowledge during their PhD training have different characteristics compared to those who do not, and these differences might influence also their careers. To mitigate this problem in estimating the effect of acquiring AI knowledge during the PhD training, we employ a Propensity Score Matching (PSM) approach. We proceed in two steps.

First, for the entire population of French PhD students in STEM, we estimate a logit model predicting the probability of observing an AI student, based on observable characteristics (Equation 1).

$$\text{AI student}_i = \beta_0 + \beta_1 X_i + \beta_2 L_i + \beta_3 D_i + \varepsilon_i$$

Equation 1: Probability of observing an AI student

The predicted probability allows us to compare the career outcome of each student with AI training, i.e., the treated, with the one of a non-AI student with a similar probability of showing AI training, i.e., the control.[9] We consider as relevant observable characteristics used to calculate the probability of receiving AI training the supervisor characteristics (X_i), location of the university (L_i), defense year (D_i), and discipline (F_i). Then, for this first analysis, we compare the probability of starting a research career (*Research career*) for AI students and the matched non-AI students.

Second, for the sample of PhD students who start a research career (*Research career* = 1), we estimate a logit model predicting the probability of observing an AI student, based on observable characteristics (Equation 2).

$$\text{AI student}_i|_{\text{Research career}_i=1} = \gamma_0 + \gamma_1 X_i + \gamma_2 L_i + \gamma_3 D_i + \tau_i$$

Equation 2: Probability of observing an AI student, conditional on starting a research career

Similarly to the previous matching exercise (Equation 1), the probability predicted using the estimates of Equation 2, allows us to compare the career outcomes of students with AI training with those of a control group of non-AI students with a similar probability of showing AI training based on a set of observable characteristics, i.e., X_i, L_i, D_i, and F_i as in Equation 1. Differently from the previous matching exercise, we consider as career outcomes the variables *Private career, Private career/collaboration, N. of publications*, *N. AI publications*, *At least one patent*, *N. of coauthors*, *Abroad*, *U.S. or China*, and *Citations*.

[9] We use the *Nearest Neighbor* approach *without replacement* to match the treated and control groups. This technique ensures that each control unit is matched exclusively to one treated unit based on the closest propensity score, avoiding potential bias from reusing controls (Smith 1997). To further improve our matching exercise, we perform an exact matching on the disciplines.

4.2 Student and supervisor characteristics

This section describes the set of observable characteristics that we use in Equations 1 and 2 to estimate the probability of observing an AI student which allows us to match AI students with a similar control group of non-AI students.

As supervisor characteristics (X_i), we include *Mentorship experience*. For each supervisor, we calculate *Mentorship experience* by counting the total number of supervisor's students who successfully defended their thesis from 1990 until the focal student's enrollment. To analyze the supervisor's academic output, we consider all their publications from the beginning of their career up to the enrollment of the focal student.[10] We then define a *High productive supervisor* as one who ranks in the top 25% for productivity. We also identify an *AI supervisor before student enrollment* as one having at least one AI-related publication before the enrollment of the student. A *Supervisor with a private affiliation* is defined as a supervisor who has at least one affiliation with a private organization her publications or patents. *PI supervisor* is a dummy variable that equals one if the supervisor is the principal investigator of an ANR[11] grant before the student enrollment in the PhD program.

We also consider the location (L_i) of the university where both the student and the supervisor are based. To do so, we calculate dummy variables grouping universities from major French cities. For example, universities in Paris are combined into one group, represented by the *Paris* dummy variable. We create ten dummy variables for the ten cities that host the majority of the PhD students in France, with all other institutions that are not located in these cities falling under the residual category *Other*. In our analysis, we use the dummy variable *Paris* as our reference group. Regarding the *Defense year* (D_i), we introduce a set of dummy variables that denote the years when students defended their theses. The dummy variable *Defense year 2010* is our reference group. Lastly, we define eight dummy variables representing the major disciplines (F_i) of study in which the supervisor mentors the student. These disciplines include *Mathematics, Engineering, Computer Science, Physics, Geology, Medicine, Biology,* and *Chemistry.*

---

10 Despite having comprehensive data on PhD graduates, information about the students' entry year into their PhD programs was missing. To address this, we estimate the entry year to be three years prior to their defense year. According to the national statistics for STEM disciplines in France (Pommier et al. 2022), the typical duration of PhD training is three years of research plus one additional year dedicated to the thesis defense. Therefore, we define the PhD training period as spanning from t-3 to t, where t represents the defense year.

11 ANR is the main funding agency in France.

4.3 Finding a control sample

This section is in two parts. The first part aims to find a control sample of non-AI students in the entire population of French students. The second part aims to find a control sample of non-AI students among the students who started a research career after graduation. In both parts, we estimate the respective probability of being trained in AI during the PhD, i.e., propensity score, and assess the quality of our matching exercises to ensure that each AI student is matched with a similar non-AI student.

4.3.1 Matching AI students with a control sample of non-AI students in the entire population of French students

Table 3 reports the estimates of Equation 1 used to predict the propensity scores. The results suggest that students supervised by highly productive supervisors, supervisors with private affiliations, and principal investigators on research grants are less likely to write an AI thesis. On the other hand, students are more likely to develop an AI thesis if their supervisor has AI experience before their enrollment and has more supervision experience. We also find that students from universities in *Toulouse* show the highest probability of writing an AI thesis.

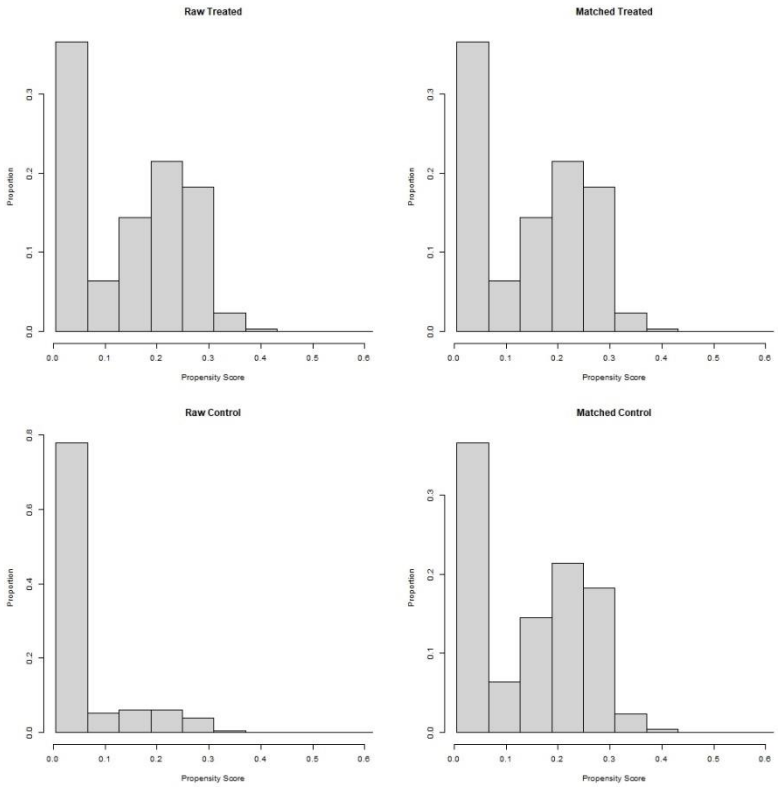
Table 3: PSM equation, probability of writing an AI thesis in the entire population of French students

Dependent variable: AI student	Estimate	Std. Error	Statistic	P-value
(Intercept)	-3.08	0.07	-41.27	0.00
High productive supervisor	-0.59	0.05	-12.35	0.00
AI supervisor before student enrollment	1.87	0.04	49.57	0.00
Supervisor with a private affiliation	-0.33	0.05	-6.71	0.00
PI supervisor	-0.26	0.05	-5.05	0.00
Mentorship experience	0.07	0.01	9.20	0.00
Bordeaux	-0.43	0.12	-3.56	0.00
Lille	0.04	0.10	0.43	0.67
Lyon	-0.22	0.08	-2.63	0.01
Marseille	-0.45	0.13	-3.56	0.00
Montpellier	-0.30	0.11	-2.73	0.01
Nantes	0.05	0.13	0.37	0.71
Others	0.15	0.05	3.19	0.00
Rennes	-0.10	0.10	-0.94	0.35
Strasbourg	-0.93	0.15	-5.99	0.00
Toulouse	0.18	0.07	2.42	0.02
Defense year 2011	-0.06	0.09	-0.67	0.50
Defense year 2012	-0.04	0.09	-0.46	0.64
Defense year 2013	-0.04	0.08	-0.42	0.67
Defense year 2014	-0.10	0.09	-1.20	0.23
Defense year 2015	-0.06	0.08	-0.75	0.46
Defense year 2016	-0.10	0.09	-1.20	0.23
Defense year 2017	-0.04	0.08	-0.52	0.60
Defense year 2018	0.07	0.08	0.87	0.38
Pseudo R-squared	0.12			
Number of Observations	46,657			

NOTE: The dummy variables *Paris* and *Defense year 2010* are the reference groups. The coefficients reported in the table are the coefficients of the Logit model. Discipline dummy variables are not included because we set an "exact matching" on the disciplines.

We use the predictions from the logit model estimated in Table 3 to calculate the probability of observing an AI student for both the AI students, i.e., the treated group, and the non-AI students, i.e. the control group. As shown in Figure 1, the distribution of the propensity scores of the AI students and the non-AI students is different before the matching (*Figure 1, Raw treated* versus *Figure 1, Raw control*), while becomes very similar after the matching (*Figure 1, Matched treated* versus *Figure 1, Matched control*), indicating a high-quality matching exercise.

Figure 1: Propensity score matching distribution before and after the matching exercise in the entire population of French students



To further assess the quality of our matching exercise, Table 4 presents the average characteristics of the two groups of AI and non-AI-matched students after the matching exercise. Specifically, we compare the average characteristics of the 3,561 AI students with the average characteristics of a control group of 3,561 non-AI students. Table 4 reveals no statistically significant differences between the individuals in each group after the matching exercise (the P-values of the tests of the difference between averages are always above the standard significance threshold of 0.1), indicating that we successfully created two comparable groups of students.

Table 4: PSM balancing in the entire population of French students

	Non-AI students (3,561)	AI Students (3,561)	P-value
High productive supervisor	0.21	0.21	0.91
AI supervisor before student enrollment	0.62	0.62	0.83
Supervisor with a private affiliation	0.15	0.16	0.14
PI supervisor	0.14	0.15	0.11
Mentorship experience	1.82	1.91	0.12
Bordeaux	0.02	0.02	1.00
Lille	0.03	0.04	0.17
Lyon	0.06	0.06	0.68
Marseille	0.02	0.02	0.87
Montpellier	0.03	0.03	0.56
Nantes	0.02	0.02	0.27
Others	0.44	0.42	0.14
Paris	0.25	0.25	0.68
Rennes	0.04	0.04	0.71
Strasbourg	0.01	0.01	0.83
Toulouse	0.09	0.09	0.93
Defense year 2010	0.09	0.08	0.37
Defense year 2011	0.09	0.08	0.56
Defense year 2012	0.1	0.1	0.85
Defense year 2013	0.11	0.11	0.82
Defense year 2014	0.1	0.11	0.46
Defense year 2015	0.12	0.12	0.77
Defense year 2016	0.12	0.12	0.64
Defense year 2017	0.13	0.14	0.53
Defense year 2018	0.14	0.14	0.63

NOTE: Discipline dummy variables are not included because we set an “exact matching” on the disciplines.

4.3.2 Matching AI students with a control sample of non-AI students, conditional on pursuing a research career

Table 5 reports the estimates of Equation 2 used to predict the propensity scores. The sample size in this analysis decreased from 46,657 to 25,687 students because we selected only those students who started a research career. Although Equation 2 is estimated on a different sample than the one of Equation 1, the findings in Table 5 are similar to those reported in Table 4. Students are less likely to write an AI thesis if they are supervised by highly productive supervisors, supervisors with affiliations to private institutions, or PIs on research grants. However, the probability increases for students whose supervisors have prior AI experience or more supervision experience. Also, students graduating from universities in *Toulouse* show the highest probability of writing an AI thesis.

Table 5: PSM equation. Probability of writing a thesis in AI, conditional on starting a research career

Dependent variable: AI thesis	Estimate	Std. Error	Statistic	P-value
(Intercept)	-3.32	0.11	-30.85	0.00
High productive supervisor	-0.65	0.07	-9.34	0.00
AI supervisor before student enrollment	1.99	0.05	36.32	0.00
Supervisor with a private affiliation	-0.29	0.07	-4.02	0.00
PI supervisor	-0.30	0.07	-4.12	0.00
Mentorship experience	0.08	0.01	6.60	0.00
Bordeaux	-0.41	0.17	-2.38	0.02
Lille	0.18	0.14	1.27	0.20
Lyon	-0.22	0.12	-1.80	0.07
Marseille	-0.48	0.18	-2.63	0.01
Montpellier	-0.34	0.16	-2.10	0.04
Nantes	0.23	0.17	1.36	0.17
Others	0.19	0.07	2.84	0.00
Rennes	-0.14	0.15	-0.93	0.35
Strasbourg	-0.84	0.21	-4.00	0.00
Toulouse	0.26	0.10	2.45	0.01
Defense year 2011	0.00	0.13	-0.01	0.99
Defense year 2012	0.04	0.12	0.30	0.77
Defense year 2013	-0.04	0.12	-0.32	0.75
Defense year 2014	-0.13	0.12	-1.02	0.31
Defense year 2015	-0.11	0.12	-0.89	0.38
Defense year 2016	-0.11	0.12	-0.91	0.36
Defense year 2017	0.02	0.12	0.20	0.84
Defense year 2018	0.04	0.12	0.36	0.72
Pseudo R-squared	0.13			
Number of Observations	25,687			

NOTE: The dummy variables *Paris* and *Defense year 2010* are the reference groups. The coefficients reported in the table are the coefficients of the Logit model. Discipline dummy variables are not included because we set an "exact matching" on the disciplines.

As in the previous matching exercise, Figure 2 shows similar propensity score distributions for AI students, i.e., the treated group, and non-AI students, i.e., the control group, after the matching *(Figure 2, Matched treated* versus *Figure 2, Matched control)*. Moreover, Table 6 shows no statistically significant differences between the 1,667 AI students and the similar 1,667 non-AI students.

Figure 2: Propensity score matching distribution before and after the matching exercise, conditional on starting a research career after graduation

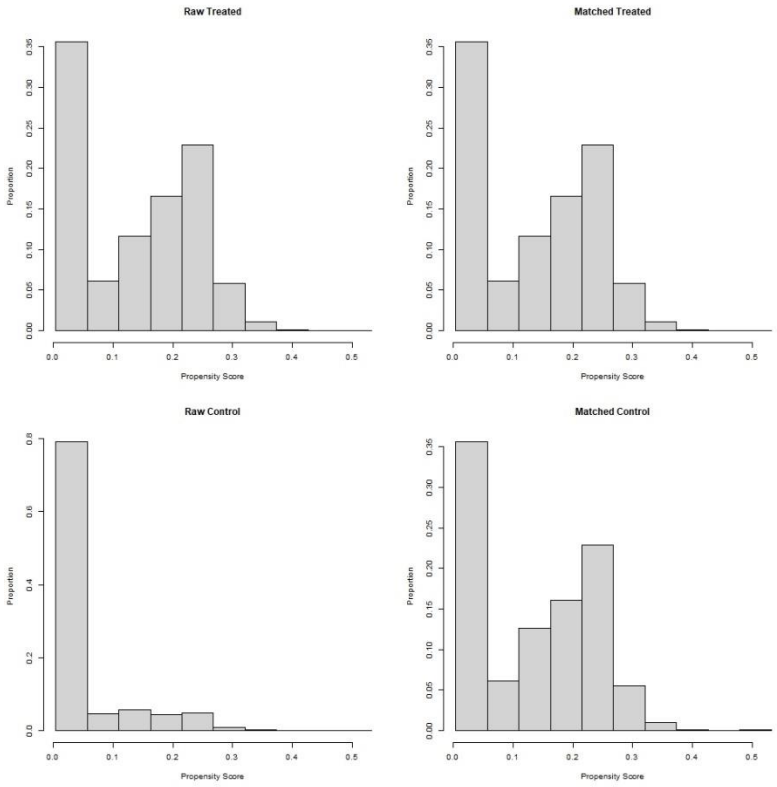


Table 6: PSM balancing, conditional on starting a research career

	Non-AI Students (1,667)	AI Students (1,667)	P-value
High productive supervisor	0.21	0.21	0.77
AI supervisor before student enrollment	0.62	0.63	0.69
Supervisor with a private affiliation	0.15	0.17	0.19
PI supervisor	0.15	0.16	0.36
Mentorship experience	1.74	1.75	0.90
Bordeaux	0.03	0.02	0.39
Lille	0.03	0.04	0.20
Lyon	0.05	0.06	0.65
Marseille	0.02	0.02	0.38
Montpellier	0.02	0.03	0.15
Nantes	0.03	0.03	0.68
Others	0.43	0.41	0.36
Paris	0.25	0.24	0.78
Rennes	0.04	0.04	0.47
Strasbourg	0.01	0.02	0.78
Toulouse	0.08	0.09	0.50
Defense year 2010	0.08	0.08	0.85
Defense year 2011	0.09	0.09	0.58
Defense year 2012	0.12	0.11	0.28
Defense year 2013	0.12	0.11	0.75
Defense year 2014	0.11	0.11	0.87
Defense year 2015	0.11	0.11	0.66
Defense year 2016	0.13	0.12	0.31
Defense year 2017	0.12	0.13	0.32
Defense year 2018	0.12	0.13	0.43

NOTE: Discipline dummy variables are not included because we set an "exact matching" on the disciplines.

5. Results

In this section, we present the results of our empirical analysis investigating the effects of acquiring AI knowledge during the PhD on career outcomes.

Table 7 shows that having AI knowledge does not significantly affect the probability of starting a research career (*Research career*), both in *Computer Science* and in *other disciplines*. However, conditional on starting a research career, AI-trained graduates in Computer Science are less likely to work in private organizations (*Private career*) than their non-AI counterparts. Similarly, they are less likely to collaborate with private organizations (*Private career/collaboration*) than their non-AI counterparts. Surprisingly, the lower attractiveness of private organizations emerging from our analysis is not aligned with the extant literature (Ahmed et al., 2023). Nonetheless, the discrepancy between our results and the literature can be explained by the specificity of the French institutional context in which the amount of resources invested in AI by the French government might have increased the attractiveness of

research careers related to AI in the public sector. When considering their publication activity after graduation, we observe that AI-trained graduates are more likely to continue publishing on AI-related topics (*AI publications*), both in *Computer Science* and *Other disciplines*. This result aligns with the results in the extant literature showing that PhD graduates tend to work in research areas closely related to their PhD topic (Shibayama 2019). Interestingly, for *Computer Science* graduates, acquiring AI knowledge during the PhD does not affect the number of articles published after graduation (*N. of publications*), the size of the professional network (*N. of coauthors*), the probability of having a career abroad (*Abroad* and *U.S. and China*), and the citations received (*Citations*). For graduates in *other disciplines* than Computer Science, having AI training during the PhD increases the probability of having a career abroad (*Abroad* and *U.S. and China*). Our results are in line with those of Franzoni, Scellato, and Stephan (2012), who suggest that highly skilled scientists are more internationally mobile. According to our analysis, graduates with AI knowledge are more likely to show U.S. and Chinese affiliations, two countries leading the worldwide investments in AI technologies (OECD 2017; WIPO 2019). Concerning patenting activity, graduates in *Computer Science* with AI knowledge have a lower probability of patenting than their counterparts without AI knowledge. On the contrary, graduates in *Other disciplines* than Computer Science with AI knowledge show a higher probability of patenting than their non-AI counterparts. The lower probability of patenting for *Computer Science* graduates with AI knowledge might be explained by their preference for an open-source environment to disclose their results, more practical than patents to keep the rapid pace of innovation in emerging technologies such as AI (WIPO 2019).

Table 7: Average career outcomes of AI and non-AI students after the Propensity Score Matching

	Non-AI students	AI students	P-value	N. of Non-AI students	N. of AI students	N. of theses
Computer Science						
Research career	0.40	0.41	0.50	1,655	1,655	3,310
Conditional on pursuing a research career (Research career=1)						
Private career/collaboration	0.21	0.17	0.08	679	679	1,358
Private career	0.10	0.08	0.09	679	679	1,358
Research career productivity						
N. of publications	3.99	4.06	0.86	679	679	1,358
At least one patent	0.13	0.09	0.04	679	679	1,358
N. AI publications	0.72	1.40	0.00	679	679	1,358
N. of coauthors	13.10	13.39	0.85	679	679	1,358
Abroad	0.48	0.48	0.91	679	679	1,358
U.S. and China	0.13	0.12	0.81	679	679	1,358
Citations	5.88	6.48	0.43	679	679	1,358
Other disciplines						
Research career	0.51	0.52	0.60	1,906	1,906	3,812
Conditional on pursuing a research career (Research career=1)						
Private career/collaboration	0.19	0.18	0.64	988	988	1,976
Private career	0.09	0.08	0.33	988	988	1,976
Research career productivity						
N. of publications	4.93	5.44	0.15	988	988	1,976
At least one patent	0.10	0.14	0.01	988	988	1,976
N. AI publications	0.28	1.14	0.00	988	988	1,976
N. of coauthors	19.12	22.10	0.13	988	988	1,976
Abroad	0.42	0.51	0.00	988	988	1,976
U.S. and China	0.12	0.17	0.00	988	988	1,976
Citations	8.42	8.87	0.48	988	988	1,976

NOTE: all career-related variables are calculated in *[t+2, t+6]*, where *t* is the thesis defense year.

6. Conclusion

In this paper, we study the career paths of 46,657 French PhD students who graduated in the STEM disciplines between 2010 and 2018. Specifically, we examine how acquiring AI knowledge during the PhD influences the likelihood of pursuing a research career after graduation. We start by identifying AI students as those who used AI-related keywords in their thesis abstracts or titles. Then, we trace the PhD graduates' career trajectories by analyzing their publications and patent documents. Tracing the graduates' careers allowed us to investigate how AI training during the PhD affects the probability of pursuing a research career after graduation. Moreover, conditional on pursuing a research career, we investigated how AI training affects the probability of working in private research organizations after graduation and how it affects the students' productivity, professional networks, and mobility.

We find that having AI knowledge does not significantly affect the probability of starting a research career. However, conditional on starting a research career, graduates in Computer Science with AI knowledge are less likely to start a career in private research organizations. We

also find that AI-trained graduates are more likely to continue publishing on AI-related topics. AI students in other disciplines than Computer Science are more likely to have careers abroad after graduation than their non-AI counterparts. Finally, we find differences in patenting behavior across disciplines. Students in Computer Science with AI training are less likely to patent than their non-AI counterparts, while students with AI training in other STEM disciplines are more likely to patent than their non-AI counterparts.

Our paper adds to the literature on the economics of science that explores the effects of AI on scientific research. Specifically, we provide new insights into how the acquisition of AI knowledge impacts early-stage scientists' productivity. Furthermore, we contribute to the labor economics literature by analyzing the career trajectories of individuals who graduated with a specific set of knowledge highly demanded in the private sector. Our contribution is also methodological. We show how implementing an identification strategy that identifies the unbiased effect of AI training might change the empirical evidence provided by descriptive statistics. For instance, according to our descriptive statistics, researchers in Computer Science with AI training who pursue a research career are as likely as their non-AI counterparts to work in private research organizations after graduation. This descriptive evidence is challenged when we apply a PSM approach. Indeed, the PSM approach shows that researchers in Computer Science with AI training are less likely than their non-AI counterparts to work in private research organizations after graduation. This latter result contradicts the idea of a brain drain of AI talents from the public to the private sector proposed in other works (Jurowetzki et al. 2021; Ahmed, Wahed, and Thompson 2023).

Our paper is not without limitations. The first limitation concerns our proxies for career paths based on publication and patent information. Although our proxies seem to be in line with the career described in LinkedIn data, future studies might want to consider different proxies for career paths, such as through conducting surveys on a subset of graduates. The second limitation regards the use of keywords to identify AI-related documents. Although this method is widely used in literature, a future extension of this work is to use machine learning techniques to identify AI content. The third limitation concerns the missing information about the careers of PhD graduates who do not publish or patent after graduation. Although we know from official statistics that 92% of PhD graduates in France are employed (Ministère de l'Enseignement supérieur et de la Recherche 2024), we cannot trace the careers of those who leave research and stop publishing and patenting.

References

Acemoglu, Daron, David Autor, Jonathon Hazell, and Pascual Restrepo. 2020. 'AI and Jobs: Evidence from Online Vacancies'. Working Paper. Working Paper Series. National Bureau of Economic Research. https://doi.org/10.3386/w28257.

———. 2022. 'Artificial Intelligence and Jobs: Evidence from Online Vacancies'. *Journal of Labor Economics* 40 (S1): S293–340. https://doi.org/10.1086/718327.

Acemoglu, Daron, and Pascual Restrepo. 2018. 'The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment'. *American Economic Review* 108 (6): 1488–1542. https://doi.org/10.1257/aer.20160696.

Aghion, Philippe, Mathias Dewatripont, and Jeremy C. Stein. 2008. 'Academic Freedom, Private-Sector Focus, and the Process of Innovation'. *The RAND Journal of Economics* 39 (3): 617–35. https://doi.org/10.1111/j.1756-2171.2008.00031.x.

Agrawal, Ajay, Joshua Gans, and Avi Goldfarb. 2019. 'Economic Policy for Artificial Intelligence'. *Innovation Policy and the Economy* 19 (January):139–59. https://doi.org/10.1086/699935.

Ahmed, Nur, Muntasir Wahed, and Neil C. Thompson. 2023. 'The Growing Influence of Industry in AI Research'. *Science* 379 (6635): 884–86. https://doi.org/10.1126/science.ade2420.

Andalib, Maryam A., Navid Ghaffarzadegan, and Richard C. Larson. 2018. 'The Postdoc Queue: A Labor Force in Waiting'. *Systems Research and Behavioral Science* 35 (6): 675–86. https://doi.org/10.1002/sres.2510.

Arora, Ashish, Sharon Belenzon, Larisa C. Cioaca, Lia Sheer, and Hansen Zhang. 2023. 'The Effect of Public Science on Corporate R&D'. Working Paper. Working Paper Series. National Bureau of Economic Research. https://doi.org/10.3386/w31899.

Auriol, Laudeline. 2010. 'Careers of Doctorate Holders: Employment and Mobility Patterns'.

Acemoglu, Daron, David Autor, Jonathon Hazell, and Pascual Restrepo. 2022. 'Artificial Intelligence and Jobs: Evidence from Online Vacancies'. *Journal of Labor Economics* 40 (S1): S293–340. https://doi.org/10.1086/718327.

Acemoglu, Daron, and Pascual Restrepo. 2018. 'The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment'. *American Economic Review* 108 (6): 1488–1542. https://doi.org/10.1257/aer.20160696.

Aghion, Philippe, Mathias Dewatripont, and Jeremy C. Stein. 2008. 'Academic Freedom, Private-Sector Focus, and the Process of Innovation'. *The RAND Journal of Economics* 39 (3): 617–35. https://doi.org/10.1111/j.1756-2171.2008.00031.x.

Agrawal, Ajay, Joshua S. Gans, and Avi Goldfarb. 2019. 'Artificial Intelligence: The Ambiguous Labor Market Impact of Automating Prediction'. *Journal of Economic Perspectives* 33 (2): 31–50. https://doi.org/10.1257/jep.33.2.31.

Ahmed, Nur, Muntasir Wahed, and Neil C. Thompson. 2023. 'The Growing Influence of Industry in AI Research'. *Science* 379 (6635): 884–86. https://doi.org/10.1126/science.ade2420.

'AI Index Report 2021'. 2021. https://aiindex.stanford.edu/report/.

Andalib, Maryam A., Navid Ghaffarzadegan, and Richard C. Larson. 2018. 'The Postdoc Queue: A Labour Force in Waiting'. *Systems Research and Behavioral Science* 35 (6): 675–86. https://doi.org/10.1002/sres.2510.

Arora, Ashish, Sharon Belenzon, Larisa Cioaca, Lia Sheer, and Hansen Zhang. 2023. 'The Effect of Public Science on Corporate R&D'. w31899. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w31899.

Auriol, Laudeline. 2010. 'Careers of Doctorate Holders: Employment and Mobility Patterns'. Paris: OECD. https://doi.org/10.1787/5kmh8phxvvf5-en.

Auriol, Laudeline, Max Misu, and Rebecca Ann Freeman. 2013. 'Careers of Doctorate Holders: Analysis of Labour Market and Mobility Indicators'. Paris: OECD. https://doi.org/10.1787/5k43nxgs289w-en.

Baruffaldi, Stefano, Brigitte van Beuzekom, Hélène Dernis, Dietmar Harhoff, Nandan Rao, David Rosenfeld, and Mariagrazia Squicciarini. 2020. 'Identifying and Measuring Developments in Artificial Intelligence: Making the Impossible Possible'. OECD Science, Technology and Industry Working Papers 2020/05. https://www.oecd-ilibrary.org/science-and-technology/identifying-and-measuring-developments-in-artificial-intelligence_5f65ff7e-en.

Becker, Gary S. 1964. 'Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education, First Edition'. NBER. https://www.nber.org/books-and-chapters/human-capital-theoretical-and-empirical-analysis-special-reference-education-first-edition.

Bianchini, Stefano, Moritz Müller, and Pierre Pelletier. 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.

Bonnal, Liliane, and Jean-François Giret. 2010. 'Determinants of Access to Academic Careers in France'. *Economics of Innovation and New Technology* 19 (5): 437–58. https://doi.org/10.1080/10438590903434851.

Calvino, Flavio, Chiara Criscuolo, Hélène Dernis, and Lea Samek. 2023. 'What Technologies Are at the Core of AI?: An Exploration Based on Patent Data'. OECD Artificial Intelligence Papers 6. Vol. 6. OECD Artificial Intelligence Papers. https://doi.org/10.1787/32406765-en.

Cockburn, Iain, Rebecca Henderson, and Scott Stern. 2018. 'The Impact of Artificial Intelligence on Innovation'. w24449. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w24449.

Cohen, Wesely M., and Daniel A. Levinthal. 2000. 'Absorptive Capacity: A New Perspective on Learning and Innovation *'. In *Strategic Learning in a Knowledge Economy*. Routledge.

Cohen, Wesley M., and Daniel A. Levinthal. 1990. 'Absorptive Capacity: A New Perspective on Learning and Innovation'. *Administrative Science Quarterly* 35 (1): 128–52. https://doi.org/10.2307/2393553.

Cyranoski, David, Natasha Gilbert, Heidi Ledford, Anjali Nayar, and Mohammed Yahia. 2011. 'Education: The PhD Factory'. *Nature* 472 (7343): 276–80.

European Commission. 2023. *Trends in the Use of AI in Science: A Bibliometric Analysis.* LU: Publications Office. https://data.europa.eu/doi/10.2777/418191.

Foray, Dominique, and Bengt-Åke Lundvall. 1998. 'The Knowledge-Based Economy: From the Economics of Knowledge to the Learning Economy'. In *The Economic Impact of Knowledge*. Routledge.

Franzoni, Chiara, Giuseppe Scellato, and Paula Stephan. 2012. 'Foreign-Born Scientists: Mobility Patterns for 16 Countries'. *Nature Biotechnology* 30 (12): 1250–53. https://doi.org/10.1038/nbt.2449.

Furman, Jason. 2018. 'Should We Be Reassured If Automation in the Future Looks Like Automation in the Past?' In *The Economics of Artificial Intelligence: An Agenda*, 317–28. University of Chicago Press. https://www.nber.org/books-and-chapters/economics-artificial-intelligence-agenda/should-we-be-reassured-if-automation-future-looks-automation-past.

Geuna, Aldo, and Sotaro Shibayama. 2015. 'Moving Out of Academic Research: Why Do Scientists Stop Doing Research?' In *Global Mobility of Research Scientists*, edited by

Aldo Geuna, 271–303. San Diego: Academic Press. https://doi.org/10.1016/B978-0-12-801396-0.00010-7.
Goldin, Claudia, and Lawrence Katz. 2007. 'The Race between Education and Technology: The Evolution of U.S. Educational Wage Differentials, 1890 to 2005'. w12984. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w12984.
Jordan, M. I., and T. M. Mitchell. 2015. 'Machine Learning: Trends, Perspectives, and Prospects'. *Science* 349 (6245): 255–60. https://doi.org/10.1126/science.aaa8415.
Jurowetzki, Roman, Daniel Hain, Juan Mateos-Garcia, and Konstantinos Stathoulopoulos. 2021. 'The Privatization of AI Research(-Ers): Causes and Potential Consequences -- From University-Industry Interaction to Public Research Brain-Drain?' arXiv. https://doi.org/10.48550/arXiv.2102.01648.
Lane, Julia, Jason Owen-Smith, and Bruce A. Weinberg. 2024. 'How to Track the Economic Impact of Public Investments in AI'. *Nature* 630 (8016): 302–4. https://doi.org/10.1038/d41586-024-01721-1.
Larivière, Vincent. 2011. 'On the Shoulders of Students? The Contribution of PhD Students to the Advancement of Knowledge'. *Scientometrics* 90 (2): 463–81. https://doi.org/10.1007/s11192-011-0495-6.
Larson, Richard C., Navid Ghaffarzadegan, and Yi Xue. 2014. 'Too Many PhD Graduates or Too Few Academic Job Openings: The Basic Reproductive Number R0 in Academia'. *Systems Research and Behavioral Science* 31 (6): 745–50. https://doi.org/10.1002/sres.2210.
Malloy, Jonathan, Lisa Young, and Loleen Berdahl. 2021. 'Ph.D. Oversupply: The System Is the Problem'. *Inside Higher Ed*, 2021. https://www.insidehighered.com/advice/2021/06/22/how-phd-job-crisis-built-system-and-what-can-be-done-about-it-opinion.
Mangematin, V, and S Robin. 2003. 'The Two Faces of PhD Students: Management of Early Careers of French PhDs in Life Sciences'. *Science and Public Policy* 30 (6): 405–14. https://doi.org/10.3152/147154303781780209.
Ministère de l'Enseignement supérieur et de la Recherche. 2024. 'L'État de l'Enseignement Supérieur, de La Recherche et de l'Innovation En France'.
Naddaf, Miryam. 2024. 'Nearly 50% of Researchers Quit Science within a Decade, Huge Study Reveals'. *Nature*, October. https://doi.org/10.1038/d41586-024-03222-7.
Nelson, Richard R., and Sidney G. Winter. 1982. *An Evolutionary Theory of Economic Change*. Digitally reprinted. Cambridge, Mass.: The Belknap Press of Harvard Univ. Press.
OECD. 2017. *OECD Science, Technology and Industry Scoreboard 2017: The Digital Transformation*. OECD Science, Technology and Industry Scoreboard. OECD. https://doi.org/10.1787/9789264268821-en.
———. 2019. 'What Are the Characteristics and Outcomes of Doctoral Graduates?' Paris: OECD. https://doi.org/10.1787/8389c70e-en.
———. 2021a. 'Challenges and New Demands on the Academic Research Workforce'. Paris: OECD. https://doi.org/10.1787/72f6f879-en.
———. 2021b. 'Reducing the Precarity of Academic Research Careers'. Paris: OECD. https://doi.org/10.1787/0f8bd468-en.
———. 2023. 'Promoting Diverse Career Pathways for Doctoral and Postdoctoral Researchers'. Paris: OECD. https://doi.org/10.1787/dc21227a-en.
Pommier, Sylvie, Mossadek Talby, Marie Auffray-Seguette, Marc Dalaut, Hendrik Eijsberg, Perrine Elshawish, and Hamida Muller. 2022. 'Le Doctorat En FranceRegards Croisés Sur La Formation Doctorale'. Research Report. Réseau National de Collèges Doctoraux. https://hal.science/hal-03494721.

Priem, Jason, Heather Piwowar, and Richard Orr. 2022. 'OpenAlex: A Fully-Open Index of Scholarly Works, Authors, Venues, Institutions, and Concepts'. arXiv. https://doi.org/10.48550/arXiv.2205.01833.

Sarrico, Cláudia S. 2022. 'The Expansion of Doctoral Education and the Changing Nature and Purpose of the Doctorate'. *Higher Education* 84 (6): 1299–1315. https://doi.org/10.1007/s10734-022-00946-1.

Sauermann, Henry, and Michael Roach. 2014. 'Not All Scientists Pay to Be Scientists: PhDs' Preferences for Publishing in Industrial Employment'. *Research Policy* 43 (1): 32–47. https://doi.org/10.1016/j.respol.2013.07.006.

Squicciarini, Mariagrazia, and Heike Nachtigall. 2021. 'Demand for AI Skills in Jobs: Evidence from Online Job Postings'. Paris: OECD. https://doi.org/10.1787/3ed32d94-en.

Stokes, Donald E. 1997. *Pasteur's Quadrant: Basic Science and Technological Innovation*. Washington: Brookings Institution Press.

The Economist. 2016. 'Why Doing a PhD Is Often a Waste of Time'. Medium. 29 December 2016. https://medium.economist.com/why-doing-a-phd-is-often-a-waste-of-time-349206f9addb.

Waaijer, Cathelijn J. F., Rosalie Belder, Hans Sonneveld, Cornelis A. van Bochove, and Inge C. M. van der Weijden. 2017. 'Temporary Contracts: Effect on Job Satisfaction and Personal Lives of Recent PhD Graduates'. *Higher Education* 74 (2): 321–39. https://doi.org/10.1007/s10734-016-0050-8.

WIPO. 2019. *WIPO Technology Trends 2019 – Artificial Intelligence*. Unknown. https://doi.org/10.34667/TIND.29084.

Zolas, Nikolas, Nathan Goldschlag, Ron Jarmin, Paula Stephan, Jason Owen- Smith, Rebecca F. Rosen, Barbara McFadden Allen, Bruce A. Weinberg, and Julia I. Lane. 2015. 'Wrapping It up in a Person: Examining Employment and Earnings Outcomes for Ph.D. Recipients'. *Science* 350 (6266): 1367–71. https://doi.org/10.1126/science.aac5949.

Zolas, Nikolas, Zachary Kroff, Erik Brynjolfsson, Kristina McElheran, David N. Beede, Cathy Buffington, Nathan Goldschlag, Lucia Foster, and Emin Dinlersoz. 2020. 'Advanced Technologies Adoption and Use by U.S. Firms: Evidence from the Annual Business Survey'. Working Paper 28290. Working Paper Series. National Bureau of Economic Research. https://doi.org/10.3386/w28290.

Paris: OECD. https://doi.org/10.1787/5kmh8phxvvf5-en.

Auriol, Laudeline, Max Misu, and Rebecca Ann Freeman. 2013. 'Careers of Doctorate Holders: Analysis of Labor Market and Mobility Indicators'. Paris: OECD. https://doi.org/10.1787/5k43nxgs289w-en.

Baruffaldi, Stefano, Brigitte van Beuzekom, Hélène Dernis, Dietmar Harhoff, Nandan Rao, David Rosenfeld, and Mariagrazia Squicciarini. 2020. 'Identifying and Measuring Developments in Artificial Intelligence: Making the Impossible Possible'. OECD Science, Technology and Industry Working Papers 2020/05. Vol. 2020/05. OECD Science, Technology and Industry Working Papers. https://doi.org/10.1787/5f65ff7e-en.

Becker, Gary S. 1964. 'Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education, First Edition'. NBER. https://www.nber.org/books-and-chapters/human-capital-theoretical-and-empirical-analysis-special-reference-education-first-edition.

Berdahl, Loleen, Jonathan Malloy, and Lisa Young. 2020. 'Faculty Perceptions of Political Science PhD Career Training'. *PS: Political Science & Politics* 53 (4): 751–56. https://doi.org/10.1017/S1049096520000839.

Bianchini, Stefano, Moritz Müller, and Pierre Pelletier. 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.

Bonnal, Liliane, and Jean-François Giret. 2010. 'Determinants of Access to Academic Careers in France'. *Economics of Innovation and New Technology* 19 (5): 437–58. https://doi.org/10.1080/10438590903434851.

Calvino, Flavio, Chiara Criscuolo, Hélène Dernis, and Lea Samek. 2023. 'What Technologies Are at the Core of AI?: An Exploration Based on Patent Data'. OECD Artificial Intelligence Papers 6. Vol. 6. OECD Artificial Intelligence Papers. https://doi.org/10.1787/32406765-en.

Cockburn, Iain, Rebecca Henderson, and Scott Stern. 2018. 'The Impact of Artificial Intelligence on Innovation'. w24449. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w24449.

Cockburn, Iain M., and Rebecca M. Henderson. 1998. 'Absorptive Capacity, Coauthoring Behavior, and the Organization of Research in Drug Discovery'. *The Journal of Industrial Economics* 46 (2): 157–82.

Cohen, Wesley M., and Daniel A. Levinthal. 1990. 'Absorptive Capacity: A New Perspective on Learning and Innovation'. *Administrative Science Quarterly*, 128–52.

Cyranoski, David, Natasha Gilbert, Heidi Ledford, Anjali Nayar, and Mohammed Yahia. 2011a. 'Education: The PhD Factory'. *Nature* 472 (7343): 276–79. https://doi.org/10.1038/472276a.

———. 2011b. 'Education: The PhD Factory'. *Nature* 472 (7343): 276–80.

Dasgupta, Partha, and Paul A. David. 1994. 'Toward a New Economics of Science'. *Research Policy*, Special Issue in Honor of Nathan Rosenberg, 23 (5): 487–521. https://doi.org/10.1016/0048-7333(94)01002-1.

European Commission. 2023. *Trends in the Use of AI in Science: A Bibliometric Analysis*. LU: Publications Office. https://data.europa.eu/doi/10.2777/418191.

Foray, Dominique, and Bengt-Åke Lundvall. 1998. 'The Knowledge-Based Economy: From the Economics of Knowledge to the Learning Economy'. In *The Economic Impact of Knowledge*. Routledge.

Franzoni, Chiara, Giuseppe Scellato, and Paula Stephan. 2012. 'Foreign-Born Scientists: Mobility Patterns for 16 Countries'. *Nature Biotechnology* 30 (12): 1250–53. https://doi.org/10.1038/nbt.2449.

Furman, Jason. 2018. 'Should We Be Reassured If Automation in the Future Looks Like Automation in the Past?' In *The Economics of Artificial Intelligence: An Agenda*, 317–28. University of Chicago Press. https://www.nber.org/books-and-chapters/economics-artificial-intelligence-agenda/should-we-be-reassured-if-automation-future-looks-automation-past.

Geuna, Aldo, and Sotaro Shibayama. 2015. 'Moving Out of Academic Research: Why Do Scientists Stop Doing Research?' In *Global Mobility of Research Scientists*, edited by Aldo Geuna, 271–303. San Diego: Academic Press. https://doi.org/10.1016/B978-0-12-801396-0.00010-7.

Goldin, Claudia, and Lawrence F. Katz. 2007. 'The Race between Education and Technology: The Evolution of U.S. Educational Wage Differentials, 1890 to 2005'. Working Paper. Working Paper Series. National Bureau of Economic Research. https://doi.org/10.3386/w12984.

Green, Andrew. 2024. 'Artificial Intelligence and the Changing Demand for Skills in the Labor Market'. OECD Artificial Intelligence Papers 14. Vol. 14. OECD Artificial Intelligence Papers. https://doi.org/10.1787/88684e36-en.

Jordan, M. I., and T. M. Mitchell. 2015. 'Machine Learning: Trends, Perspectives, and Prospects'. *Science* 349 (6245): 255–60. https://doi.org/10.1126/science.aaa8415.

Jurowetzki, Roman, Daniel Hain, Juan Mateos-Garcia, and Konstantinos Stathoulopoulos. 2021. 'The Privatization of AI Research(-Ers): Causes and Potential Consequences --

From University-Industry Interaction to Public Research Brain-Drain?' arXiv. https://doi.org/10.48550/arXiv.2102.01648.
Lane, Julia. 2023. *The Industry of Ideas: Measuring How Artificial Intelligence Changes Labor Markets*. American Enterprise Institute. https://search.issuelab.org/resources/42165/42165.pdf.
Lane, Julia, Jason Owen-Smith, and Bruce A. Weinberg. 2024. 'How to Track the Economic Impact of Public Investments in AI'. *Nature* 630 (8016): 302–4. https://doi.org/10.1038/d41586-024-01721-1.
Lane, Marguerita, and Anne Saint-Martin. 2021. 'The Impact of Artificial Intelligence on the Labor Market: What Do We Know so Far?' Paris: OECD. https://doi.org/10.1787/7c895724-en.
Larivière, Vincent. 2011. 'On the Shoulders of Students? The Contribution of PhD Students to the Advancement of Knowledge'. *Scientometrics* 90 (2): 463–81. https://doi.org/10.1007/s11192-011-0495-6.
Larson, Richard C., Navid Ghaffarzadegan, and Yi Xue. 2014. 'Too Many PhD Graduates or Too Few Academic Job Openings: The Basic Reproductive Number R0 in Academia'. *Systems Research and Behavioral Science* 31 (6): 745–50. https://doi.org/10.1002/sres.2210.
Malloy, Jonathan, Lisa Young, and Loleen Berdahl. 2021. 'Ph.D. Oversupply: The System Is the Problem'. *Inside Higher Ed*, 2021. https://www.insidehighered.com/advice/2021/06/22/how-phd-job-crisis-built-system-and-what-can-be-done-about-it-opinion.
Mangematin, V, and S Robin. 2003. 'The Two Faces of PhD Students: Management of Early Careers of French PhDs in Life Sciences'. *Science and Public Policy* 30 (6): 405–14. https://doi.org/10.3152/147154303781780209.
Naddaf, Miryam. 2024. 'Nearly 50% of Researchers Quit Science within a Decade, Huge Study Reveals'. *Nature*, October. https://doi.org/10.1038/d41586-024-03222-7.
Nelson, Richard R., and Sidney G. Winter. 1982. *An Evolutionary Theory of Economic Change*. Digitally reprinted. Cambridge, Mass.: The Belknap Press of Harvard Univ. Press.
OECD. 2017. *OECD Science, Technology and Industry Scoreboard 2017: The Digital Transformation*. OECD Science, Technology and Industry Scoreboard. OECD. https://doi.org/10.1787/9789264268821-en.
———. 2019. 'What Are the Characteristics and Outcomes of Doctoral Graduates?' Paris: OECD. https://doi.org/10.1787/8389c70e-en.
———. 2021a. 'Challenges and New Demands on the Academic Research Workforce'. Paris: OECD. https://doi.org/10.1787/72f6f879-en.
———. 2021b. 'Reducing the Precarity of Academic Research Careers'. Paris: OECD. https://doi.org/10.1787/0f8bd468-en.
———. 2023. 'Promoting Diverse Career Pathways for Doctoral and Postdoctoral Researchers'. Paris: OECD. https://doi.org/10.1787/dc21227a-en.
Perkmann, Markus, and Kathryn Walsh. 2007. 'University–Industry Relationships and Open Innovation: Towards a Research Agenda'. *International Journal of Management Reviews* 9 (4): 259–80. https://doi.org/10.1111/j.1468-2370.2007.00225.x.
Pommier, Sylvie, Mossadek Talby, Marie Auffray-Seguette, Marc Dalaut, Hendrik Eijsberg, Perrine Elshawish, and Hamida Muller. 2022. 'Le Doctorat En FranceRegards Croisés Sur La Formation Doctorale'. Research Report. Réseau National de Collèges Doctoraux. https://hal.science/hal-03494721.
Priem, Jason, Heather Piwowar, and Richard Orr. 2022. 'OpenAlex: A Fully-Open Index of Scholarly Works, Authors, Venues, Institutions, and Concepts'. arXiv. https://doi.org/10.48550/arXiv.2205.01833.

Rogers, Everett M. 2003. *Diffusion of Innovations*. Fifth edition. New York London Toronto Sydney: Free Press.

Rogers, Everett M., Arvind Singhal, and Margaret M. Quinlan. 2014. 'Diffusion of Innovations'. In *An Integrated Approach to Communication Theory and Research*, 432–48. Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9780203887011-36/diffusion-innovations-everett-rogers-arvind-singhal-margaret-quinlan.

Sarrico, Cláudia S. 2022. 'The Expansion of Doctoral Education and the Changing Nature and Purpose of the Doctorate'. *Higher Education* 84 (6): 1299–1315. https://doi.org/10.1007/s10734-022-00946-1.

Sauermann, Henry, and Michael Roach. 2014. 'Not All Scientists Pay to Be Scientists: PhDs' Preferences for Publishing in Industrial Employment'. *Research Policy* 43 (1): 32–47. https://doi.org/10.1016/j.respol.2013.07.006.

Shibayama, Sotaro. 2019. 'Sustainable Development of Science and Scientists: Academic Training in Life Science Labs'. *Research Policy* 48 (3): 676–92. https://doi.org/10.1016/j.respol.2018.10.030.

Smith, Herbert L. 1997. 'Matching With Multiple Controls to Estimate Treatment Effects in Observational Studies'. *Sociological Methodology* 27 (1): 325–53. https://doi.org/10.1111/1467-9531.271030.

Squicciarini, Mariagrazia, and Heike Nachtigall. 2021. 'Demand for AI Skills in Jobs: Evidence from Online Job Postings'. Paris: OECD. https://doi.org/10.1787/3ed32d94-en.

Stokes, Donald E. 1997. *Pasteur's Quadrant: Basic Science and Technological Innovation*. Washington: Brookings Institution Press.

'The AI Index Report'. 2021. 2021. https://aiindex.stanford.edu/ai-index-report-2021/.

The Economist. 2016. 'Why Doing a PhD Is Often a Waste of Time'. Medium. 29 December 2016. https://medium.economist.com/why-doing-a-phd-is-often-a-waste-of-time-349206f9addb.

Waaijer, Cathelijn J. F., Rosalie Belder, Hans Sonneveld, Cornelis A. van Bochove, and Inge C. M. van der Weijden. 2017. 'Temporary Contracts: Effect on Job Satisfaction and Personal Lives of Recent PhD Graduates'. *Higher Education* 74 (2): 321–39. https://doi.org/10.1007/s10734-016-0050-8.

WIPO. 2019. *WIPO Technology Trends 2019 – Artificial Intelligence*. Unknown. https://doi.org/10.34667/TIND.29084.

Zolas, N., N. Goldschlag, R. Jarmin, P. Stephan, J. O.- Smith, R. F. Rosen, B. M. Allen, B. A. Weinberg, and J. I. Lane. 2015. 'Wrapping It up in a Person: Examining Employment and Earnings Outcomes for Ph.D. Recipients'. *Science* 350 (6266): 1367–71. https://doi.org/10.1126/science.aac5949.

Appendix

Appendix A: Publications and patents attribution to students and supervisors

In this appendix, we describe the matching procedure used to link OpenAlex authors and patent inventors with PhD students and supervisors.

Matching OpenAlex authors with PhD students and supervisors

We collect bibliometric information about the PhD students and their supervisors using OpenAlex. We start by converting their students' names listed in the doctoral dissertation dataset and the OpenAlex to lowercase, removing punctuation, and handling any special characters. Then, we match their names with the OpenAlex authors' names. Among the 62,917 students in the doctoral dissertation dataset, we successfully matched 40,514 PhD students to unique OpenAlex author profiles, while 18,687 students were not matched with any profile. 3,716 PhD students were matched with multiple profiles. For supervisors, we applied a similar matching strategy.

After matching, we filtered the results according to multiple criteria. For students, we dropped student-author pairs if the author's first publication occurred more than two years before the student's enrollment in the PhD program or more than ten years after her defense year. The rationale behind these filters is to ensure that the author's publications are realistically aligned with the student's academic timeline. For supervisors, we dropped supervisor-author pairs if the author's first publication was older than five years before the supervisor's first PhD supervision, in line with the French requirement for all supervisors to hold an HDR (*Habilitation à Diriger des Recherches).* After filtering the student-author pairs and the supervisor-author pairs according to the two preceding criteria, we obtained 46,657 student matches with a unique OpenAlex profile. We use this sample of students as the study sample for our analysis.

Matching patent inventors with PhD students and supervisors

The European Patent Office (EPO) patent data is extracted from the Patstat database. The EPO dataset contains inventors' personal information, such as person ID, names, addresses, countries, and sectors (i.e., company, government, non-profit organization, university, hospital, and unknown). We cleaned inventors' names removing punctuations, accents, special characters, middle initials, standalone single letters, double spaces, and content within parentheses. Similarly, we cleaned students' names from the doctoral dissertation dataset. To avoid incorrect matches, matches are filtered based on the timeline of patenting activity relative

to the student's thesis defense year. An inventor whose first patent occurred more than 6 years before the student's thesis defense year is excluded. To avoid potential ambiguities, cases where a student is linked to multiple inventors are removed to retain unique matches. Similarly, we cleaned supervisors' names and matched them with inventors' names. In this case, we drop inventor-supervisor pairs in which the inventor has filled their last patent more than 10 years before the year of their first thesis supervision and inventor-supervisor pairs in which the inventor has filled their first patent more than 30 years after their first thesis supervision. As before, unique supervisor-inventor matches are kept. Sectoral information is considered to classify the applicants reported in the students' and supervisors' patents into public (government, nonprofit, university, and hospital) and private applicants (companies, individuals, unknown). Country information is collected to identify U.S. or Chinese patent applicants.

Overall, 2398 students (5.13 % of the student sample) are matched with a unique inventor profile. This students' patent activity shows applicants predominantly concentrated in the private sector.

Appendix B: AI theses identification strategy

In this appendix, we detail the methodology used to identify AI-related theses. The literature does not agree on a single definition of AI, as it includes a wide range of technologies (Agrawal et al., 2019; Vannuccini and Prytkova, 2023). Sometimes, the definition can be quite narrow, focusing only on specific areas like symbolic systems, robotics, and machine learning (Cockburn et al., 2018), while others define AI more broadly, encompassing a wider set of technologies (Baruffaldi et al., 2020; Calvino et al., 2023). In this paper, we adopt a broad definition of AI, identifying documents with AI content through keywords present in titles or abstracts. We develop our list of keywords by merging two widely recognized sources: one from an OECD report with 189 AI keywords (Baruffaldi et al., 2020; Calvino et al., 2023), and another from an economic study with 39 AI keywords (Cockburn et al., 2018). After combining these lists and adding plural forms of the keywords, we refined our list to 281 unique AI keywords. The complete list of keywords is reported in Table B1.

We used these keywords to analyze 46,657 PhD theses in STEM disciplines in France between 2010 and 2018 and identified 3,561theses (7.63%) as AI-related. According to Bianchini et al. (2022), some keywords used in the literature are too generic to accurately identify AI-related content. This is mainly why we see this high proportion of AI theses. Our broad approach resulted in classifying documents with marginal AI relevance, such as those containing "robotics".

To improve our keyword list, we evaluated each keyword by three metrics: the number and share of theses containing the keyword, its co-occurrence probability with other AI keywords, and its distribution across different disciplines. The rationale behind this approach is that if an AI keyword is used to identify a high number of theses, rarely co-occurs with other AI keywords in the same document, and is spread across various disciplines, then the keyword is considered to be too generic to identify AI-related documents, therefore, we drop the keyword from our list. We identified overly generic keywords using Table B2 after aggregating keywords that refer to the same concept. For instance, we group the keywords “robotics”, “robot”, “robots”, and “robotic” under the same concept “robot”. Five concepts—"algorithm," "network," "mapping," "robot," and "decision making"—were deemed too generic due to high frequency (columns 1-2), low co-occurrence with other AI keywords (column 3), and widespread use across various disciplines (columns 4-11). Keywords associated with these concepts were also removed from our list.

After this refinement, we ended up with 272 keywords. Using this revised list, we identified 3,561 AI theses, representing 7.63% of the total theses analyzed. This approach ensures a more precise identification of AI documents by excluding overly generic terms.

Table B1: Full list of keywords

action recognition	cognitive computing	feature extraction	image matching	machine learning	particle swarm optimisation	similarity learning
activity recognition	cognitive modelling	feature learning	image processing	machine translation	pattern analysis	simultaneous localisation
adaboost	collaborative filtering	feature selection	image recognition	machine vision	pattern recognition	social robot
adaptive boosting	collaborative systems	fuzzy c	image retrieval	mapping	pedestrian detection	social robots
adversarial network	collision avoidance	fuzzy logic	image segmentation	mapreduce	physical symbol system	sparse representation
adversarial networks	community detection	fuzzy number	independent component analyses	markovian	q learning	spectral clustering
algorithm	computational intelligence	fuzzy numbers	independent component analysis	memetic algorithm	random field	speech recognition
algorithms	computer vision	fuzzy set	inductive monitoring	memetic algorithms	random fields	speech recognitions
ambient intelligence	convolutional neural network	fuzzy sets	industrial robot	meta learning	random forest	speech to text
ant colony	convolutional neural networks	gaussian mixture model	industrial robots	motion planning	random forests	stacked generalization
ant colony optimisation	crowdsourcing and human computation	gaussian mixture models	instance based learning	multi agent system	rankboost	statistical relational learning
artificial bee colony algorithm	cyber physical system	gaussian process	intelligence augmentation	multi agent systems	recommender system	stochastic gradient
artificial intelligence	cyber physical systems	generative adversarial network	intelligent agent	multi label classification	recommender systems	supervised learning
artificial neural network	data mining	generative adversarial networks	intelligent agents	multi label classifications	recurrent neural network	support vector machine
artificial neural networks	decision making	genetic algorithm	intelligent classifier	multi layer perceptron	recurrent neural networks	support vector machines
association rule	decision tree	genetic algorithms	intelligent classifiers	multi objective evolutionary	regression tree	support vector regression
autoencoder	decision trees	genetic programming	intelligent infrastructure	multi objective optimisation	regression trees	support vector regressions
autoencoders	deep belief network	gesture recognition	intelligent infrastructures	multi objective optimisations	reinforcement learning	swarm intelligence
autonomic computing	deep belief networks	gradient boosting	intelligent software agent	multi sensor fusion	relational learning	swarm optimisation
autonomous vehicle	deep convolutional neural	gradient tree boosting	intelligent software agents	multi sensor fusions	robot	symbol processing
autonomous vehicles	deep learning	graphical model	k means	multi task learning	robot systems	symbolic error analyses
backpropagation	deep neural network	graphical models	kernel learning	naive bayes classifier	robotic	symbolic error analysis
bayesian belief network	deep neural networks	hebbian learning	knowledge representation and reasoning	naive bayes classifiers	robotics	symbolic reasoning
bayesian belief networks	dictionary learning	hidden markov model	latent dirichlet allocation	natural gradient	robots	systems and control theory
bayesian learning	differential evolution	hidden markov models	latent semantic analyses	natural language generation	rough set	temporal difference learning
bayesian network	differential evolution algorithm	hierarchical clustering	latent semantic analysis	natural language processing	rule based learning	text mining
bayesian networks	dimensionality reduction	high dimensional data	latent variable	natural language understanding	rule learning	text to speech
biped robot	dynamic time warping	high dimensional feature	latent variables	natural languages	self organising map	topic model
biped robots	emotion recognition	high dimensional input	layered control systems	nearest neighbour algorithm	semantic web	trajectory planning
blind signal separation	ensemble learning	high dimensional space	learning automata	nearest neighbour algorithms	semi supervised learning	trajectory tracking
bootstrap aggregation	evolutionary algorithm	human action recognition	legged robot	network	sensor data fusion	transfer learning
brain computer interface	evolutionary algorithms	human activity recognition	legged robots	neural network	sensor data fusions	unmanned aerial vehicle
brain computer interfaces	evolutionary computation	human robot interaction	link prediction	neural networks	sensor fusion	unmanned aerial vehicles
chatbot	extreme learning machine	human robot interactions	link predictions	neural turing	sensor fusions	unsupervised learning
chatbots	extreme learning machines	humanoid robot	logic theorist	neuromorphic computing	sensor network	variational inference
classification tree	face recognition	humanoid robotics	logitboost	non negative matrix factorisation	sensor networks	vector machine
classification trees	facial expression recognition	humanoid robots	long short term memory	object detection	sentiment analyses	vector machines
cluster analyses	factorisation machine	image alignment	long short term memory lstm	object recognition	sentiment analysis	virtual assistant
cluster analysis	factorisation machines	image classification	lstm	obstacle avoidance	service robot	virtual assistants
cognitive automation	feature engineering	image grammars	machine intelligence	optimal search	service robots	visual servoing
						xgboost

Table B2: AI concept statistics

Concepts*	(1) N. of theses**	(2) Share	(3) Keyword co-occurrence	(4) Biology	(5) Chemistry	(6) Computer science	(7) Engineering	(8) Geology	(9) Mathematics	(10) Medicine	(11) Physics
algorithm	7169	33.66%	42.01%	2.40%	0.00%	38.71%	29.68%	2.65%	12.88%	2.55%	10.38%
network	4242	19.92%	30.58%	16.31%	0.00%	26.85%	25.55%	5.40%	3.11%	8.25%	8.37%
robot	1378	6.47%	66.33%	2.65%	0.00%	35.44%	49.31%	0.13%	1.13%	3.53%	7.19%
mapping	1096	5.15%	34.76%	19.34%	0.02%	20.99%	20.89%	9.95%	5.57%	7.57%	12.23%
decision making	604	2.84%	45.20%	12.09%	0.02%	34.27%	28.81%	2.48%	3.15%	15.07%	3.15%
machine learning	456	2.14%	85.53%	4.61%	0.01%	58.11%	16.67%	1.54%	9.65%	3.29%	5.04%
sensor network	445	2.09%	72.95%	0.27%	0.00%	50.82%	42.08%	0.82%	1.91%	0.55%	3.55%
image processing	377	1.77%	63.66%	3.18%	0.01%	28.91%	35.81%	1.86%	9.81%	3.45%	16.18%
neural network	360	1.69%	70.77%	12.31%	0.04%	24.92%	28.31%	4.00%	4.62%	14.46%	11.08%
genetic algorithm	250	1.17%	57.26%	3.32%	0.01%	20.75%	54.77%	1.66%	3.32%	1.24%	11.62%
computer vision	233	1.09%	83.69%	0.43%	0.00%	58.37%	28.76%	0.43%	7.30%	1.29%	3.43%
data mining	170	0.80%	74.12%	5.29%	0.03%	62.94%	15.29%	4.12%	3.53%	4.12%	2.94%
multi agent system	148	0.69%	58.14%	0.00%	0.00%	61.24%	31.78%	0.00%	3.88%	2.33%	0.78%
support vector machine	128	0.60%	87.20%	6.40%	0.05%	33.60%	39.20%	2.40%	6.40%	5.60%	5.60%
artificial intelligence	118	0.55%	79.66%	1.69%	0.01%	62.71%	24.58%	1.69%	2.54%	5.08%	1.69%
markovian	118	0.55%	41.53%	1.69%	0.01%	19.49%	12.71%	0.00%	50.00%	3.39%	11.86%
random field	112	0.53%	59.80%	0.00%	0.00%	15.69%	51.96%	4.90%	19.61%	1.96%	5.88%
semantic web	107	0.50%	55.14%	0.00%	0.00%	93.46%	4.67%	0.00%	0.00%	1.87%	0.00%
artificial neural network	93	0.44%	71.59%	5.68%	0.06%	20.45%	38.64%	12.50%	1.14%	5.68%	15.91%
decision tree	92	0.43%	68.60%	6.98%	0.08%	38.37%	18.60%	4.65%	2.33%	6.98%	19.77%
fuzzy logic	91	0.43%	73.63%	2.20%	0.02%	36.26%	52.75%	0.00%	2.20%	4.40%	2.20%
supervised learning	91	0.43%	90.11%	3.30%	0.04%	63.74%	14.29%	4.40%	12.09%	1.10%	1.10%
random forest	89	0.42%	82.93%	9.76%	0.12%	31.71%	25.61%	12.20%	17.07%	1.22%	2.44%
humanoid robot	88	0.41%	91.89%	2.70%	0.04%	39.19%	52.70%	0.00%	1.35%	1.35%	2.70%
natural language processing	88	0.41%	68.18%	1.14%	0.01%	89.77%	5.68%	0.00%	0.00%	3.41%	0.00%
pattern recognition	88	0.41%	54.55%	19.32%	0.22%	38.64%	18.18%	3.41%	3.41%	7.95%	5.68%
bayesian network	87	0.41%	75.00%	4.17%	0.06%	43.06%	43.06%	0.00%	5.56%	2.78%	1.39%
deep learning	83	0.39%	96.39%	1.20%	0.01%	60.24%	24.10%	0.00%	7.23%	6.02%	1.20%
evolutionary algorithm	79	0.37%	72.86%	4.29%	0.06%	45.71%	35.71%	0.00%	1.43%	0.00%	10.00%
image segmentation	79	0.37%	75.95%	5.06%	0.06%	32.91%	31.65%	3.80%	11.39%	5.06%	10.13%
recommender system	79	0.37%	79.69%	0.00%	0.00%	90.63%	3.13%	0.00%	4.69%	1.56%	0.00%
feature extraction	72	0.34%	94.44%	1.39%	0.02%	50.00%	31.94%	1.39%	5.56%	6.94%	2.78%
graphical model	72	0.34%	79.37%	1.59%	0.03%	44.44%	30.16%	0.00%	19.05%	0.00%	4.76%
reinforcement learning	72	0.34%	87.50%	9.72%	0.14%	59.72%	15.28%	0.00%	2.78%	11.11%	1.39%
autonomous vehicle	71	0.33%	78.13%	0.00%	0.00%	48.44%	45.31%	0.00%	3.13%	0.00%	3.13%
feature selection	64	0.30%	93.75%	3.13%	0.05%	50.00%	25.00%	1.56%	14.06%	4.69%	1.56%
convolutional neural network	57	0.27%	98.08%	1.92%	0.04%	59.62%	30.77%	0.00%	3.85%	1.92%	1.92%
gaussian mixture model	57	0.27%	83.02%	1.89%	0.04%	45.28%	32.08%	0.00%	7.55%	7.55%	5.66%
object detection	57	0.27%	82.46%	0.00%	0.00%	54.39%	29.82%	0.00%	8.77%	0.00%	7.02%
unsupervised learning	56	0.26%	96.43%	5.36%	0.10%	55.36%	16.07%	0.00%	16.07%	0.00%	7.14%
speech recognition	54	0.25%	59.26%	5.56%	0.10%	72.22%	11.11%	0.00%	3.70%	3.70%	3.70%
human robot interaction	51	0.24%	97.87%	0.00%	0.00%	63.83%	31.91%	0.00%	0.00%	0.00%	4.26%
k means	51	0.24%	84.31%	3.92%	0.08%	23.53%	29.41%	7.84%	17.65%	11.76%	5.88%
hidden markov models	50	0.23%	82.00%	2.00%	0.04%	38.00%	32.00%	2.00%	26.00%	0.00%	0.00%
image classification	50	0.23%	92.00%	0.00%	0.00%	70.00%	14.00%	6.00%	4.00%	0.00%	6.00%
object recognition	50	0.23%	58.00%	14.00%	0.28%	46.00%	10.00%	0.00%	14.00%	12.00%	4.00%
dimensionality reduction	49	0.23%	77.55%	8.16%	0.17%	38.78%	28.57%	0.00%	16.33%	4.08%	4.08%
visual servoing	46	0.22%	84.78%	0.00%	0.00%	26.09%	67.39%	0.00%	2.17%	0.00%	4.35%
brain computer interface	45	0.21%	57.89%	7.89%	0.21%	42.11%	28.95%	0.00%	0.00%	18.42%	2.63%
trajectory tracking	45	0.21%	80.00%	2.22%	0.05%	26.67%	60.00%	0.00%	0.00%	0.00%	11.11%
unmanned aerial vehicle	45	0.21%	72.73%	0.00%	0.00%	34.09%	47.73%	6.82%	2.27%	0.00%	9.09%
hidden markov model	42	0.20%	76.19%	4.76%	0.11%	47.62%	23.81%	2.38%	14.29%	2.38%	4.76%
machine translation	42	0.20%	52.38%	0.00%	0.00%	97.62%	0.00%	0.00%	2.38%	0.00%	0.00%
gaussian process	41	0.19%	39.02%	0.00%	0.00%	9.76%	34.15%	0.00%	51.22%	2.44%	2.44%
hierarchical clustering	41	0.19%	53.66%	21.95%	0.54%	21.95%	17.07%	7.32%	9.76%	12.20%	7.32%
face recognition	40	0.19%	77.50%	2.50%	0.06%	50.00%	30.00%	0.00%	2.50%	15.00%	0.00%
high dimensional data	40	0.19%	92.50%	2.50%	0.06%	32.50%	10.00%	0.00%	40.00%	10.00%	0.00%
independent component analysis	40	0.19%	70.00%	15.00%	0.38%	20.00%	35.00%	2.50%	7.50%	10.00%	10.00%
motion planning	39	0.18%	92.31%	2.56%	0.07%	38.46%	48.72%	0.00%	7.69%	0.00%	2.56%
community detection	37	0.17%	100.00%	0.00%	0.00%	83.78%	0.00%	0.00%	5.41%	0.00%	8.11%
image retrieval	36	0.17%	86.11%	0.00%	0.00%	83.33%	13.89%	0.00%	0.00%	0.00%	2.78%
latent variable	33	0.15%	72.41%	6.90%	0.24%	34.48%	13.79%	0.00%	34.48%	10.34%	0.00%
sparse representation	33	0.15%	84.85%	0.00%	0.00%	21.21%	63.64%	3.03%	12.12%	0.00%	0.00%
ambient intelligence	32	0.15%	68.75%	0.00%	0.00%	71.88%	25.00%	0.00%	3.13%	0.00%	0.00%
cluster analysis	32	0.15%	50.00%	28.13%	0.88%	12.50%	15.63%	6.25%	9.38%	25.00%	3.13%
cyber physical system	32	0.15%	53.33%	0.00%	0.00%	73.33%	23.33%	0.00%	0.00%	0.00%	3.33%
trajectory planning	31	0.15%	93.55%	0.00%	0.00%	29.03%	45.16%	0.00%	19.35%	0.00%	6.45%
industrial robot	30	0.14%	100.00%	0.00%	0.00%	12.50%	79.17%	0.00%	4.17%	0.00%	4.17%
collision avoidance	29	0.14%	75.86%	0.00%	0.00%	41.38%	51.72%	0.00%	0.00%	0.00%	6.90%
deep neural network	29	0.14%	100.00%	3.85%	0.15%	69.23%	23.08%	0.00%	0.00%	3.85%	0.00%
obstacle avoidance	29	0.14%	86.21%	0.00%	0.00%	37.93%	51.72%	0.00%	0.00%	6.90%	3.45%
dictionary learning	26	0.12%	88.46%	0.00%	0.00%	34.62%	26.92%	3.85%	26.92%	7.69%	0.00%
mapreduce	26	0.12%	69.23%	0.00%	0.00%	84.62%	15.38%	0.00%	0.00%	0.00%	0.00%
recurrent neural network	25	0.12%	88.00%	0.00%	0.00%	56.00%	16.00%	0.00%	12.00%	4.00%	12.00%
ant colony	23	0.11%	91.30%	0.00%	0.00%	47.83%	30.43%	0.00%	17.39%	0.00%	4.35%
transfer learning	22	0.10%	95.45%	0.00%	0.00%	45.45%	40.91%	0.00%	13.64%	0.00%	0.00%
fuzzy set	21	0.10%	63.16%	0.00%	0.00%	63.16%	26.32%	0.00%	10.53%	0.00%	0.00%
gesture recognition	20	0.09%	80.00%	0.00%	0.00%	65.00%	30.00%	0.00%	5.00%	0.00%	0.00%
collaborative filtering	19	0.09%	84.21%	0.00%	0.00%	94.74%	5.26%	0.00%	0.00%	0.00%	0.00%
differential evolution	18	0.08%	66.67%	11.11%	0.62%	16.67%	50.00%	5.56%	5.56%	5.56%	5.56%
action recognition	17	0.08%	88.24%	0.00%	0.00%	76.47%	11.76%	0.00%	11.76%	0.00%	0.00%
activity recognition	17	0.08%	76.47%	0.00%	0.00%	70.59%	17.65%	0.00%	0.00%	0.00%	11.76%
emotion recognition	17	0.08%	70.59%	0.00%	0.00%	41.18%	29.41%	0.00%	0.00%	23.53%	5.88%
text mining	17	0.08%	76.47%	5.88%	0.35%	82.35%	5.88%	0.00%	0.00%	5.88%	0.00%
biped robot	16	0.08%	100.00%	0.00%	0.00%	25.00%	75.00%	0.00%	0.00%	0.00%	0.00%
semi supervised learning	16	0.08%	93.75%	6.25%	0.39%	56.25%	25.00%	0.00%	6.25%	0.00%	6.25%
dynamic time warping	15	0.07%	86.67%	0.00%	0.00%	40.00%	46.67%	6.67%	0.00%	6.67%	0.00%
memetic algorithm	15	0.07%	80.00%	6.67%	0.44%	73.33%	13.33%	0.00%	6.67%	0.00%	0.00%
spectral clustering	15	0.07%	73.33%	0.00%	0.00%	40.00%	26.67%	0.00%	20.00%	0.00%	6.67%
image matching	14	0.07%	78.57%	0.00%	0.00%	42.86%	28.57%	0.00%	21.43%	0.00%	7.14%
stochastic gradient	14	0.07%	100.00%	7.14%	0.51%	42.86%	21.43%	0.00%	28.57%	0.00%	0.00%
kernel learning	13	0.06%	76.92%	7.69%	0.59%	53.85%	23.08%	0.00%	7.69%	0.00%	0.00%
autonomic computing	12	0.06%	58.33%	0.00%	0.00%	91.67%	8.33%	0.00%	0.00%	0.00%	0.00%
regression tree	12	0.06%	50.00%	33.33%	2.78%	0.00%	8.33%	25.00%	33.33%	0.00%	0.00%
support vector regression	12	0.06%	66.67%	0.00%	0.00%	25.00%	66.67%	0.00%	0.00%	0.00%	8.33%

adaboost	11	0.05%	90.91%	0.00%	0.00%	54.55%	36.36%	0.00%	9.09%	0.00%	0.00%
fuzzy c	11	0.05%	100.00%	0.00%	0.00%	27.27%	36.36%	9.09%	0.00%	9.09%	18.18%
generative adversarial network	11	0.05%	100.00%	9.09%	0.83%	27.27%	45.45%	0.00%	9.09%	0.00%	9.09%
image recognition	11	0.05%	90.91%	9.09%	0.83%	54.55%	18.18%	0.00%	9.09%	9.09%	0.00%
sentiment analysis	11	0.05%	72.73%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%
genetic programming	10	0.05%	80.00%	10.00%	1.00%	60.00%	20.00%	0.00%	0.00%	0.00%	10.00%
link prediction	10	0.05%	100.00%	10.00%	1.00%	80.00%	10.00%	0.00%	0.00%	0.00%	0.00%
lstm	10	0.05%	90.00%	10.00%	1.00%	70.00%	20.00%	0.00%	0.00%	0.00%	0.00%
pattern analysis	10	0.05%	30.00%	50.00%	5.00%	0.00%	0.00%	10.00%	0.00%	10.00%	30.00%
pedestrian detection	10	0.05%	90.00%	0.00%	0.00%	50.00%	40.00%	0.00%	0.00%	10.00%	0.00%
service robot	10	0.05%	100.00%	0.00%	0.00%	80.00%	10.00%	0.00%	0.00%	0.00%	10.00%
backpropagation	9	0.04%	88.89%	0.00%	0.00%	44.44%	33.33%	0.00%	0.00%	0.00%	22.22%
deep convolutional neural	9	0.04%	100.00%	0.00%	0.00%	55.56%	33.33%	0.00%	11.11%	0.00%	0.00%
ensemble learning	9	0.04%	100.00%	0.00%	0.00%	66.67%	11.11%	11.11%	0.00%	11.11%	0.00%
human action recognition	9	0.04%	100.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%
humanoid robotics	9	0.04%	100.00%	11.11%	1.23%	11.11%	66.67%	0.00%	0.00%	0.00%	11.11%
sensor fusion	9	0.04%	88.89%	0.00%	0.00%	44.44%	55.56%	0.00%	0.00%	0.00%	0.00%
association rule	8	0.04%	87.50%	0.00%	0.00%	87.50%	0.00%	12.50%	0.00%	0.00%	0.00%
collaborative systems	8	0.04%	62.50%	0.00%	0.00%	75.00%	12.50%	0.00%	0.00%	0.00%	12.50%
facial expression recognition	8	0.04%	87.50%	0.00%	0.00%	62.50%	37.50%	0.00%	0.00%	0.00%	0.00%
long short term memory	8	0.04%	100.00%	0.00%	0.00%	62.50%	37.50%	0.00%	0.00%	0.00%	0.00%
autoencoder	7	0.03%	100.00%	0.00%	0.00%	85.71%	0.00%	0.00%	0.00%	0.00%	14.29%
machine vision	7	0.03%	100.00%	0.00%	0.00%	42.86%	42.86%	0.00%	0.00%	0.00%	14.29%
multi sensor fusion	7	0.03%	42.86%	0.00%	0.00%	71.43%	28.57%	0.00%	0.00%	0.00%	0.00%
high dimensional space	6	0.03%	100.00%	0.00%	0.00%	66.67%	0.00%	0.00%	16.67%	16.67%	0.00%
intelligent agent	6	0.03%	83.33%	0.00%	0.00%	83.33%	16.67%	0.00%	0.00%	0.00%	0.00%
multi label classification	6	0.03%	66.67%	0.00%	0.00%	83.33%	16.67%	0.00%	0.00%	0.00%	0.00%
multi objective optimisation	6	0.03%	83.33%	0.00%	0.00%	33.33%	66.67%	0.00%	0.00%	0.00%	0.00%
natural language generation	6	0.03%	66.67%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%
neuromorphic computing	6	0.03%	66.67%	0.00%	0.00%	16.67%	50.00%	0.00%	0.00%	0.00%	33.33%
sensor data fusion	6	0.03%	100.00%	0.00%	0.00%	50.00%	33.33%	0.00%	16.67%	0.00%	0.00%
swarm intelligence	6	0.03%	100.00%	0.00%	0.00%	66.67%	0.00%	16.67%	0.00%	0.00%	16.67%
text to speech	6	0.03%	83.33%	0.00%	0.00%	66.67%	16.67%	0.00%	0.00%	0.00%	16.67%
high dimensional feature	5	0.02%	100.00%	0.00%	0.00%	60.00%	20.00%	0.00%	20.00%	0.00%	0.00%
image alignment	5	0.02%	80.00%	0.00%	0.00%	20.00%	20.00%	0.00%	20.00%	40.00%	0.00%
knowledge representation and reasoning	5	0.02%	60.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%
meta learning	5	0.02%	60.00%	0.00%	0.00%	80.00%	0.00%	0.00%	0.00%	20.00%	0.00%
naive bayes classifier	5	0.02%	100.00%	0.00%	0.00%	40.00%	60.00%	0.00%	0.00%	0.00%	0.00%
natural languages	5	0.02%	60.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%
q learning	5	0.02%	100.00%	0.00%	0.00%	60.00%	40.00%	0.00%	0.00%	0.00%	0.00%
simultaneous localisation	5	0.02%	100.00%	0.00%	0.00%	20.00%	80.00%	0.00%	0.00%	0.00%	0.00%
topic model	5	0.02%	100.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%
variational inference	5	0.02%	80.00%	20.00%	4.00%	60.00%	0.00%	0.00%	20.00%	0.00%	0.00%
bayesian learning	4	0.02%	75.00%	25.00%	6.25%	25.00%	50.00%	0.00%	0.00%	0.00%	0.00%
classification tree	4	0.02%	50.00%	0.00%	0.00%	25.00%	25.00%	25.00%	0.00%	25.00%	0.00%
fuzzy number	4	0.02%	100.00%	0.00%	0.00%	0.00%	66.67%	33.33%	0.00%	0.00%	0.00%
legged robot	4	0.02%	100.00%	0.00%	0.00%	50.00%	50.00%	0.00%	0.00%	0.00%	0.00%
multi layer perceptron	4	0.02%	100.00%	0.00%	0.00%	50.00%	25.00%	0.00%	0.00%	25.00%	0.00%
particle swarm optimisation	4	0.02%	100.00%	0.00%	0.00%	25.00%	50.00%	0.00%	0.00%	0.00%	25.00%
cognitive automation	3	0.01%	33.33%	0.00%	0.00%	33.33%	33.33%	0.00%	0.00%	33.33%	0.00%
cognitive computing	3	0.01%	100.00%	0.00%	0.00%	0.00%	33.33%	0.00%	0.00%	33.33%	33.33%
extreme learning machine	3	0.01%	66.67%	0.00%	0.00%	33.33%	66.67%	0.00%	0.00%	0.00%	0.00%
feature learning	3	0.01%	100.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%
latent dirichlet allocation	3	0.01%	66.67%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%
multi task learning	3	0.01%	66.67%	0.00%	0.00%	66.67%	0.00%	0.00%	33.33%	0.00%	0.00%
natural gradient	3	0.01%	100.00%	0.00%	0.00%	66.67%	33.33%	0.00%	0.00%	0.00%	0.00%
natural language understanding	3	0.01%	33.33%	0.00%	0.00%	66.67%	33.33%	0.00%	0.00%	0.00%	0.00%
optimal search	3	0.01%	66.67%	0.00%	0.00%	0.00%	33.33%	0.00%	0.00%	0.00%	66.67%
vector machine	3	0.01%	100.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%
virtual assistant	3	0.01%	100.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%
cognitive modelling	2	0.01%	50.00%	0.00%	0.00%	50.00%	0.00%	0.00%	0.00%	50.00%	0.00%
deep belief network	2	0.01%	100.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%
evolutionary computation	2	0.01%	100.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%
hebbian learning	2	0.01%	100.00%	50.00%	25.00%	0.00%	50.00%	0.00%	0.00%	0.00%	0.00%
human activity recognition	2	0.01%	100.00%	0.00%	0.00%	50.00%	50.00%	0.00%	0.00%	0.00%	0.00%
latent semantic analysis	2	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	50.00%	50.00%
relational learning	2	0.01%	100.00%	0.00%	0.00%	50.00%	0.00%	0.00%	0.00%	50.00%	0.00%
similarity learning	2	0.01%	100.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%
statistical relational learning	2	0.01%	100.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%
symbolic reasoning	2	0.01%	50.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%
adversarial network	1	0.00%	100.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%
artificial bee colony algorithm	1	0.00%	100.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%
bayesian belief network	1	0.00%	100.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%
gradient boosting	1	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%
high dimensional input	1	0.00%	100.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%
instance based learning	1	0.00%	100.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%
logitboost	1	0.00%	100.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%
machine intelligence	1	0.00%	100.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%
multi objective evolutionary	1	0.00%	100.00%	100.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
multi objective optimisations	1	0.00%	100.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%
rankboost	1	0.00%	100.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%
rough set	1	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%
social robot	1	0.00%	100.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%
speech to text	1	0.00%	100.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%
systems and control theory	1	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%

NOTE: *Concepts are sets of similar keywords. For instance, the terms “robotics”, “robot”, “robots”, and “robotic” refer to the same concept “robot”. **The keyword may co-occur with other keywords in the same document. Therefore, the same theses might be counted for different keywords introducing double counting for the variable “*N. of theses*”.

Appendix C: Empirical analysis conducted without separating Computer Science from Other disciplines

In this appendix, we provide additional descriptive statistics for our data, along with the descriptive statistics and results presented in the main text but applied to all disciplines.

Table C1 shows that the share of AI students has been increasing over the years, while the share of PhD graduates pursuing research careers has declined.

Table C2: AI students and research careers

		All disciplines (46,657 theses)	
Defense year	AI students	Research career	N. of students
2010	6.78%	57.50%	4202
2011	6.73%	56.31%	4472
2012	7.13%	55.59%	5080
2013	7.52%	55.09%	5295
2014	7.15%	55.99%	5315
2015	7.71%	55.04%	5463
2016	7.43%	55.32%	5759
2017	8.27%	53.18%	5850
2018	9.62%	52.33%	5221
All	7.63%	55.05%	46,657

Table C2 presents descriptive statistics on the subsample of PhD graduates who pursue a research career, in a time frame between *t+2* and *t+6*. The table shows that an increasing number of PhD graduates are engaging with private research institutions after graduation. Productivity levels, measured by the number of publications, patents, and citations, are decreasing. Additionally, more graduates are publishing AI-related papers, likely due to the rise in the number of graduates with AI knowledge, while fewer of them have foreign affiliations.

Table C2: Conditional on starting a research career, type of career and productivity by year

Defense year	Private career /collaboration	Private career	N. publications	At least one patent	N. AI publications	N. of coauthors	Abroad	U.S. or China	N. of citations	N. of students
2010	14.24%	4.88%	5.89	9.77%	0.21	23.14	48.68%	15.77%	13.20	2,416
2011	14.69%	4.69%	5.81	9.61%	0.19	22.83	47.93%	14.18%	12.29	2,518
2012	13.60%	4.32%	6.10	8.64%	0.21	25.75	49.72%	15.83%	13.52	2,824
2013	15.53%	5.04%	5.69	9.43%	0.22	26.11	51.42%	15.53%	13.63	2,917
2014	15.22%	5.04%	5.62	9.27%	0.22	25.54	49.50%	14.62%	13.61	2,976
2015	15.46%	4.89%	5.65	10.64%	0.27	26.40	49.58%	16.36%	11.36	3,007
2016	16.16%	5.43%	5.40	10.67%	0.30	26.17	46.20%	14.47%	8.91	3,186
2017	15.11%	5.34%	5.24	8.42%	0.28	26.26	45.61%	12.95%	5.75	3,111
2018	14.49%	5.23%	4.90	7.43%	0.27	24.77	43.92%	13.14%	3.59	2,732
All	14.99%	5.00%	5.58	9.34%	0.24	25.32	48.05%	14.75%	10.57	25,687

Table C3 presents detailed descriptive statistics on career trajectories and productivity metrics for students who graduated with and without AI knowledge across all disciplines. We observe that graduates with AI knowledge are significantly less likely to pursue a research career, in line with the findings of Ahmed, Wahed, and Thompson (2023) who argue that AI graduates tend to choose industry roles over research careers. Among those who do pursue a research career, the data show that graduates with AI knowledge are more likely to collaborate with private research institutions or engage exclusively in private-sector research without any public affiliations. Regarding productivity measures, the descriptive statistics indicate that graduates with AI knowledge are more likely to patent and publish papers with AI content. However, they tend to publish less, collaborate with fewer coauthors, and receive fewer citations than their non-AI peers.

Table C3: Career type and productivity by AI thesis content and discipline

	Non-AI students	AI Students	P-value	N. of Non-AI students	N. of AI students	N. of theses
All disciplines						
Research career	0.56	0.47	0.00	43,096	3,561	46,657
Conditional on pursuing a research career (Research career=1)						
Private career/collaboration	0.15	0.18	0.00	24,020	1,667	25,687
Private career	0.05	0.08	0.00	24,020	1,667	25,687
Research career productivity						
N. of pubs	5.63	4.88	0.00	24,020	1,667	25,687
At least one patent	0.09	0.12	0.00	24,020	1,667	25,687
AI publications	0.17	1.25	0.00	24,020	1,667	25,687
N. coauthors	25.79	18.55	0.00	24,020	1,667	25,687
Abroad	0.48	0.50	0.17	24,020	1,667	25,687
U.S. and China	0.15	0.15	0.82	24,020	1,667	25,687
Citations	10.76	7.90	0.00	24,020	1,667	25,687

In Table C4, we present the results of the PSM for all disciplines. The results are similar to the ones presented in Section 5 and show that acquiring AI knowledge during the PhD training period does not affect the probability of starting a research career after graduation. However, for those who do, they are less likely to work in private research institutions. Regarding productivity measures, we find that AI-trained graduates are significantly more likely to keep publishing in AI, more likely to have affiliations abroad, and more likely to collaborate with research institutions based in the U.S. and China.

Table C4: PSM results for all disciplines

	Non-AI students	AI students	P-value	N. of Non-AI students	N. of AI students	N. of theses
All disciplines						
Research career	0.46	0.47	0.41	3,561	3,561	7,122
Conditional on pursuing a research career (Research career=1)						
Private career / collaboration	0.19	0.18	0.14	1,667	1,667	3,334
Private career	0.10	0.08	0.07	1,667	1,667	3,334
Research career productivity						
N. of pubs	4.55	4.88	0.20	1,667	1,667	3,334
At least one patent	0.11	0.12	0.45	1,667	1,667	3,334
AI publications	0.46	1.25	0.00	1,667	1,667	3,334
N. coauthors	16.67	18.55	0.15	1,667	1,667	3,334
Abroad	0.44	0.50	0.00	1,667	1,667	3,334
U.S. and China	0.12	0.15	0.02	1,667	1,667	3,334
Citations	7.38	7.90	0.30	1,667	1,667	3,334

Appendix D: Career manual checks using LinkedIn data

We construct careers and scientific outcomes using patent and publication data. PhD graduates who pursue a research career are those who publish at least one paper or file a patent in the 5 years after graduation. In this appendix we conduct manual checks for 100 PhD students using LinkedIn, ResearchGate, and personal websites aiming to assess the reliability of our proxies for students' career.

According to LinkedIn, ResearchGate, and personal websites, 56 PhD graduates pursued research careers and 44 had non-research career, including 11 graduates with no online information, whom we consider as non-researchers due to the lack of online presence. Our career proxy calculated with patent and publication data predicted 52 graduates as having research careers and 48 as having non-research careers. Table D1 the results of our manual checks.

Table D1: Classification of PhD graduates into Research vs. Non-Research Careers

		Classification based on LinkedIn, ResearchGate, and personal websites	
		Research Careers (56 students)	non-Research Careers (44 students)
Classification based on the career proxy obtained from patent and publication data	Research Careers (52 students)	TP = 42	FP = 10
	non-Research Careers (48 students)	FN = 14	TN = 34

NOTE: TP stands for True Positives, TN stands for True Negatives, FN stands for False Negatives, and FP stands for False Positives.

To assess the reliability of our career proxy, we calculate two performance measures: precision and recall. To do so, we used the values reported in Table D1 that include the researchers having a research career correctly identified by our career proxy (True positives, TP), the one not having a research career correctly identified by our career proxy (True negatives, TN), and the errors made: False negatives (FN) and False positives (FP). We calculate precision and recall as follows:

$$Precision = \frac{TP}{TP + FP} = \frac{42}{42 + 10} = 80.77\%$$

$$Recall = \frac{TP}{TP + FN} = \frac{42}{42 + 14} = 75\%$$

We found that 80.77% of the students identified as having a research career by our career proxy were correct, while 75% of the students having actually a research career according to LinkedIn, ResearchGate, and personal websites were correctly identified.

Acknowledgments: We thank participants of the KID summer school, Nice, France; and the PhD Workshop in Economics of Innovation, Complexity, and Knowledge (WICK), Collegio Carlo Alberto, Turin, Italy. We are also grateful to the French government for its financial support of this work through the grant ANR-24-CE26-1154.

Competing interest: The authors disclose no relevant financial interests related to the research in this paper.