Does training in AI affect PhD students' careers? Evidence from France

Pierre Boutros[1] Eliana Diodati[2] Michele Pezzoni[1] Fabiana Visentin[3]

1 Cote d'Azur University, Nice, France

2 University of Turin and Collegio Carlo Alberto, Turin, Italy

3 Maastricht University, Maastricht, The Netherlands

SI 2025 Science of Science Founding,
Boston, 17 July 2025

Motivation

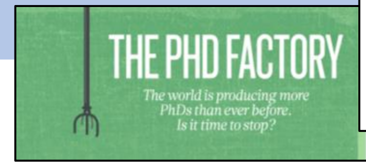
- In the labor market, the demand for workers with AI knowledge is increasing (Acemoglu et al. 2022)
  - What about the supply?
- Educational attainment has been identified as a crucial factor in shaping the labor market supply
  - Countries have invested substantial resources in AI education programs
- PhD graduates are a relevant part of the high-skilled labor force and play a key role in the diffusion of knowledge related to up-to-date technologies like AI (Ahmed et al. 2023; Lane et al. 2024)

What do we know about AI training and graduates' careers?

- Graduates' careers are affected by individual, institutional, and socio-economic factors (Sauermann and Roach 2014; Geuna and Shibayama 2015; Long and McGinnis 1985; Miller et al. 2005; Waldinger 2010)
- What about the content of what they learn?
- PhD students are trained using taxpayers' money, but nearly 50% quit research after graduation (Naddaf 2024)

Are AI-trained PhD students more likely to pursue a research career after graduation?

What do we know about AI training and graduates' career types?

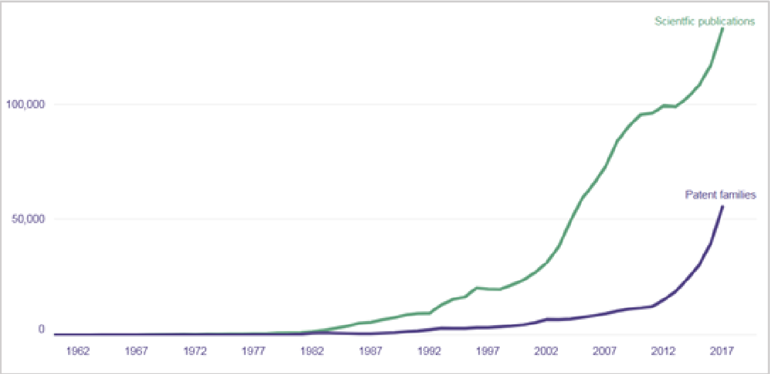


- Decades of policy interventions promoting PhD training have led to an oversupply of graduates (Cyranoski et al. 2011; OECD 2021).
- Graduates tend to prefer academic careers (Sauermann and Roach, 2012); however, not all of them can find a job in academia (OECD 2023; 2021; 2019)
- At the same time, AI knowledge is highly valuable for companies (Acemoglu et al. 2022), leading to a potential brain drain of AI-trained PhD students from academia to private organizations (Jurowetzki et al. 2021; Ahmed et al. 2023)

Are AI-trained PhD students more likely to pursue a research career in the private sector after graduation?

What do we know about AI training and graduates' productivity?

- AI is a new "method of invention" that is reshaping the inventive process in science and technology (Cockburn et al. 2018; Bianchini et al. 2022; European Commission 2023)
- AI applications across all disciplines took off around 2010



Source: WIPO, 2019

- However, we still lack large-scale empirical evidence of the effect of AI learning on researchers' productivity (Lane et al. 2024)

Are AI-trained PhD students more productive?

Discipline heterogeneity

- We conduct our study, distinguishing two types of disciplines:

  - Disciplines in which graduates have a high absorptive capacity of AI knowledge after graduation

    Computer Science

  - Disciplines in which graduates have a low absorptive capacity of AI knowledge after graduation

    Other disciplines

    (Biology, Chemistry, Engineering, Geology, Mathematics, Medicine, and Physics)

Our paper in a nutshell

AI-students vs. non-AI students

1) Are AI students more likely to pursue a research career?

Research career vs. Non-research career

Private sector vs. Public sector

2) Are AI students more likely to pursue a research career in the private sector?

Productivity

3) Are AI students more productive?

Our empirical context

- We conduct our study using data on French PhD graduates during the period 2010-2018

Focus on PhD students:

- In 2018, 44% of PhD graduates stayed in academia (Ministère de l'Enseignement Supérieur et de la Recherche 2024)

Focus on AI:

- Although AI theoretical foundations date back to the 1950s (Haenlein and Kaplan 2019), the number of applications increased sharply during our study period, mainly due to hardware development (WIPO 2019)
- Private AI investment: In 2013-2022, 338 AI startups were founded in France, but big AI giants in the US
- Public AI investment: In 2018, $1.85B was invested in AI R&D

Data

PhD theses

French National Depository, PhD graduates in the STEM fields, 2010-2018

Bibliometric information

OpenAlex

Patents

European Patent Office

After merging the three datasets, we obtained a unique dataset including **35,492 PhD graduates** and **16,298 supervisors**

Our explanatory variable

AI student is a dummy variable that equals 1 if the student writes a PhD thesis with at least one AI keyword. We identified 2,555 (7.20%) AI students

- We merge 2 lists of AI keywords
  - Cockburn et al., 2019 – 39 keywords
  - OECD, 2022 – 189 keywords
- 3 criteria to identify generic AI keywords (keyword frequency, co-occurrence probability, and disciplinary distribution)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Keyword (and variants)	**N. of theses**	**Share**	**Keyword co-occurrence**	**Bio**	**Chem**	**Comp. Sc.**	**Eng**	**Geo**	**Math**	**Med**	**Phys**
~~algorithm~~	4098	33.82%	41.83%	2.19%	0.00%	40.43%	28.86%	2.34%	13.25%	2.49%	9.80%
~~network~~	2442	20.15%	30.59%	17.44%	0.01%	27.81%	23.38%	5.73%	3.19%	8.27%	7.99%
~~robot~~	746	6.16%	67.61%	2.58%	0.01%	38.03%	49.30%	0.00%	1.17%	1.88%	6.57%
~~mapping~~	611	5.04%	36.50%	17.84%	0.03%	21.93%	21.93%	8.02%	6.38%	7.86%	12.11%
~~decision making~~	347	2.86%	45.82%	14.12%	0.04%	36.31%	26.80%	2.02%	2.88%	14.70%	2.88%
machine learning	275	2.27%	85.82%	4.73%	0.02%	61.82%	16.00%	1.09%	8.73%	3.27%	3.27%
sensor network	258	2.13%	74.29%	0.48%	0.00%	56.67%	38.10%	0.48%	1.43%	0.48%	2.38%
image processing	207	1.71%	68.60%	3.38%	0.02%	32.37%	32.37%	0.48%	12.08%	2.90%	16.43%
…	…	…	…	…	…	…	…	…	…	…	…

Our outcomes variables (1)

1) PhD graduates pursuing a research career

- **Research career** is equal to 1 if the PhD student has at least one publication or patent between t+2 and t+6, where t is the defense year

2) Conditional on starting a research career, we use affiliation information to identify if PhD graduates pursue a **career in the private sector**

- **Private career** is equal to 1 if the PhD graduate shows only affiliations to private organizations between t+2 and t+6
- **Private career/collaboration** is equal to 1 if the PhD graduate shows at least an affiliation to private organizations between t+2 and t+6

Our outcomes variables (2)

3) Conditional on starting a research career, we calculate the following bibliometric indicators (between t+2 and t+6):

- Number of publications (**N. publications**)
- Number of AI publications (**N. AI pubs**)
- Inventor of at least one patent (**At least one patent**)
- Number of distinct coauthors **(N. co-auth)**
- At least one non-French affiliation **(Abroad)**
- At least an affiliation with a US or Chinese organization (**U.S. or China**)
- Average citations received per paper **(Citations)**

Descriptive statistics: AI students and research career

- Study sample: 35,492 students

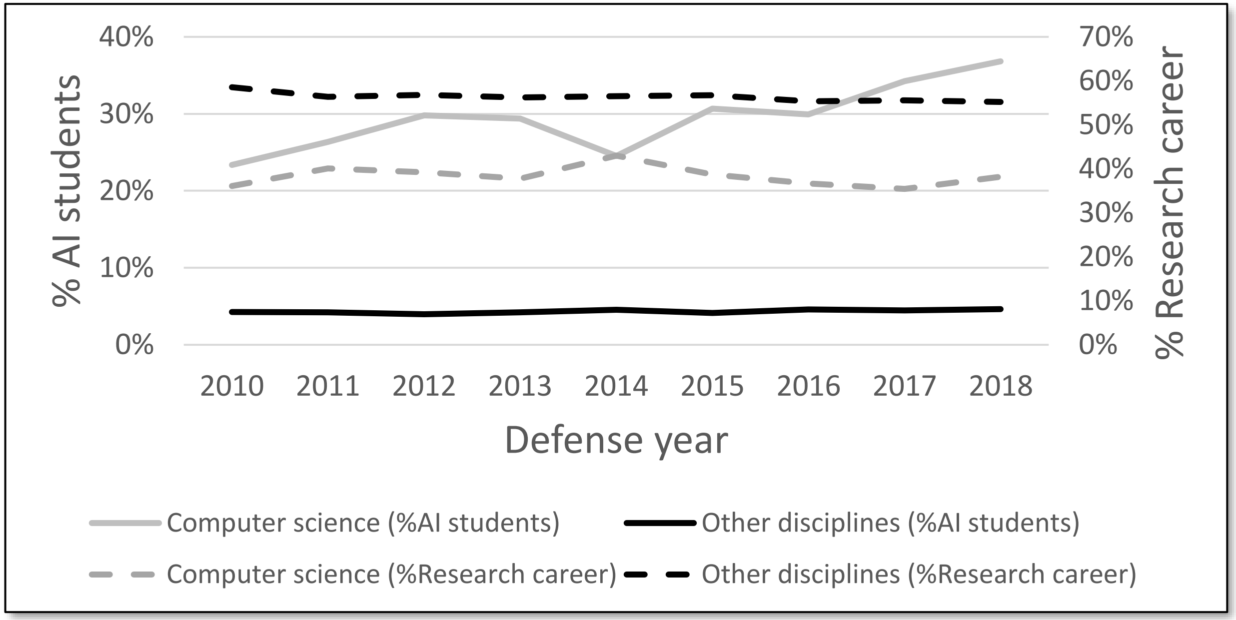
(7.2% AI students, 54.3% research career)

  - Computer science: 3,999 students – 11.3% of the total

  (29.8% AI students, 38.3% research career)

  - Other discipline: 31,493 students – 88.7% of the total

  (4.3% AI students, 56.3% research career)



Descriptive statistics: Computer science

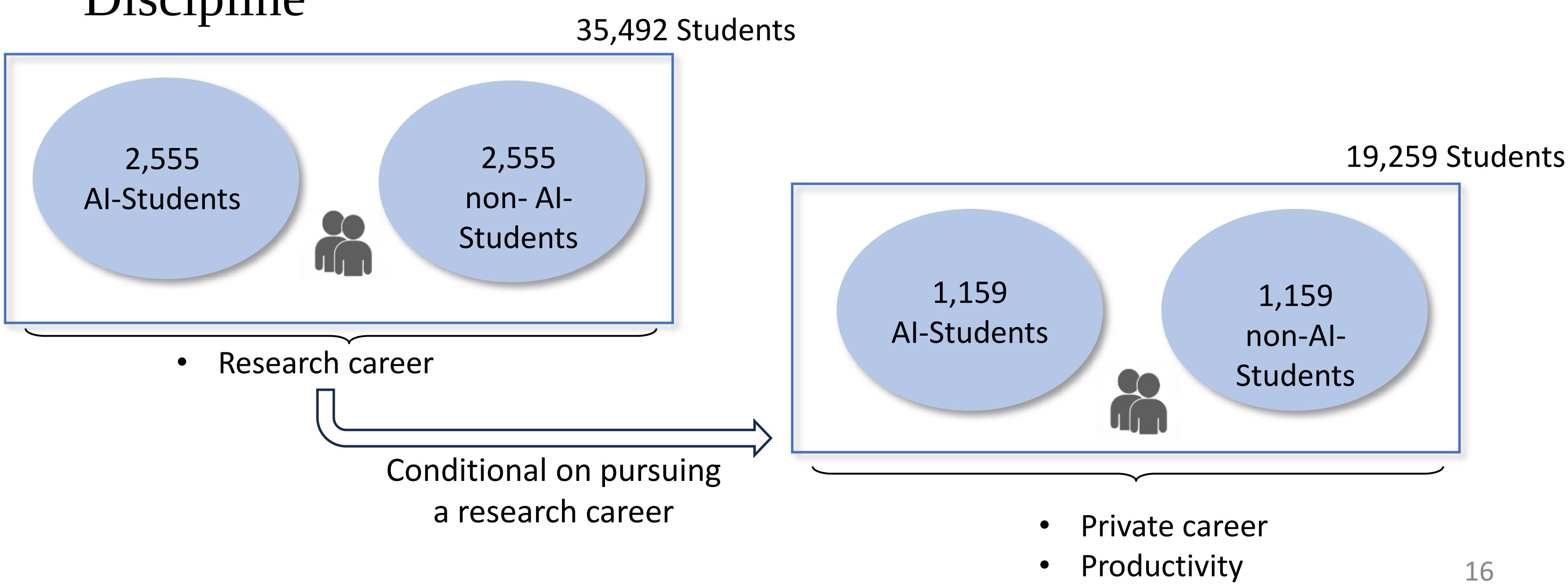
	Non-AI students	AI Students		Non-AI students	AI students
Computer Science (3,999 students)	Mean	Mean	P-value	N. of stud.	N. of stud.
Research career	0.38	0.39	0.64	2,807	1,192
Conditional on pursuing a research career (1,531 students)					
Private career/collaboration	0.20	0.16	0.10	1,068	463
Private career	0.10	0.08	0.12	1,068	463
Research career productivity					
N. of publications	3.96	3.92	0.90	1,068	463
At least one patent	0.11	0.10	0.34	1,068	463
N. AI publications	**0.56**	**1.43**	**0.00**	1,068	463
N. of coauthors	19.62	20.12	0.95	1,068	463
Abroad	0.48	0.47	0.72	1,068	463
U.S. and China	0.12	0.14	0.44	1,068	463
Citations	**10.63**	**16.88**	**0.04**	1,068	463

Descriptive statistics: Other disciplines

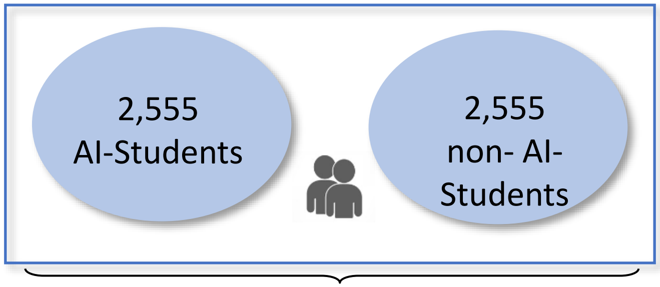
	Non-AI students	**AI Students**		**Non-AI students**	**AI students**
Other disciplines (31,493 students)	Mean	Mean	P-value	N. of stud.	N. of stud.
Research career	**0.57**	**0.51**	**0.00**	30,130	1,363
Conditional on pursuing a research career (17,728 students)					
Private career/collaboration	**0.14**	**0.20**	**0.00**	17,032	696
Private career	**0.05**	**0.09**	**0.00**	17,032	696
Research career productivity					
N. of publications	5.87	5.45	0.11	17,032	696
At least one patent	**0.09**	**0.14**	**0.00**	17,032	696
N. AI publications	**0.15**	**1.21**	**0.00**	17,032	696
N. of coauthors	**53.00**	**28.53**	**0.00**	17,032	696
Abroad	0.48	0.50	0.24	17,032	696
U.S. and China	0.15	0.17	0.26	17,032	696
Citations	16.08	15.22	0.34	17,032	696

Propensity Score Matching

- We match AI students with **similar** non-AI students based on observable characteristics:
  - High productive supervisor, AI supervisor before student enrollment, Supervisor with a private affiliation, PI supervisor, Mentorship experience, University of graduation, Defense year, Discipline



PSM results

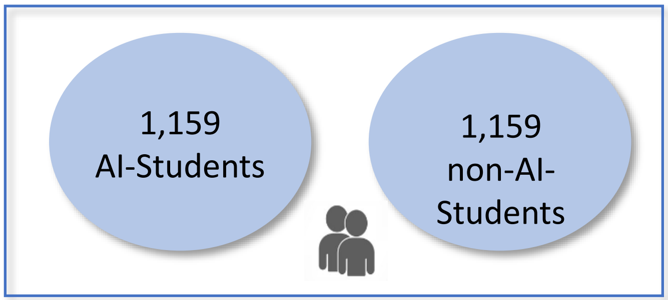


- Research career

	Non-AI Students (2,555)	AI Students (2,555)	P-value
High productive supervisor	0.21	0.21	0.73
AI supervisor before student enrolment	0.63	0.63	0.82
Supervisor with a private affiliation	0.16	0.17	0.31
PI supervisor	0.13	0.15	0.06
Mentorship experience	1.79	1.89	0.20
Bordeaux	0.03	0.03	0.67
Lille	0.04	0.04	0.61
Lyon	0.05	0.05	0.28
Marseille	0.02	0.02	0.29
Montpellier	0.03	0.03	1.00
Nantes	0.02	0.02	0.85
Others	0.45	0.43	0.19
Paris	0.25	0.25	0.54
Rennes	0.04	0.04	0.51
Strasbourg	0.01	0.01	1.00
Toulouse	0.08	0.08	0.64
Defense year 2010	0.09	0.08	0.14
Defense year 2011	0.09	0.09	0.81
Defense year 2012	0.10	0.10	0.64
Defense year 2013	0.11	0.11	0.72
Defense year 2014	0.11	0.11	0.72
Defense year 2015	0.12	0.12	0.97
Defense year 2016	0.12	0.12	0.70
Defense year 2017	0.13	0.14	0.39
Defense year 2018	0.13	0.14	0.62

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

PSM results: Conditional on pursuing a research career



- Private career
- Productivity

	Non-AI Students (1,159)	AI Students (1,159)	P-value
High productive supervisor	0.24	0.22	0.15
AI supervisor before student enrolment	0.64	0.64	0.80
Supervisor with a private affiliation	0.17	0.18	0.51
PI supervisor	0.13	0.16	0.12
Mentorship experience	1.90	1.77	0.23
Bordeaux	0.03	0.03	0.80
Lille	0.04	0.04	0.34
Lyon	0.06	0.06	0.66
Marseille	0.02	0.02	0.44
Montpellier	0.02	0.03	0.69
Nantes	0.02	0.02	1.00
Others	0.42	0.42	0.90
Paris	0.25	0.25	0.96
Rennes	0.04	0.03	0.28
Strasbourg	0.02	0.02	1.00
Toulouse	0.09	0.08	0.88
Defense year 2010	0.08	0.08	0.70
Defense year 2011	0.08	0.09	0.51
Defense year 2012	0.10	0.10	0.95
Defense year 2013	0.11	0.11	0.95
Defense year 2014	0.09	0.10	0.36
Defense year 2015	0.12	0.12	0.80
Defense year 2016	0.13	0.12	0.35
Defense year 2017	0.13	0.14	0.34
Defense year 2018	0.14	0.13	0.40

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

Results: Computer Science

	Non-AI students	**AI Students**		**Non-AI students**	**AI students**
Computer Science (2,384 students)	Mean	Mean	P-value	N. of stud.	N. of stud.
Research career	0.38	0.39	0.56	1,192	1,192
Conditional on pursuing a research career (926 students)					
Private career/collaboration	**0.21**	**0.16**	**0.08**	463	463
Private career	**0.11**	**0.08**	**0.07**	463	463
Research career productivity					
N. of publications	4.35	3.92	0.30	463	463
At least one patent	0.11	0.10	0.39	463	463
N. AI publications	**0.65**	**1.43**	**0.00**	463	463
N. of coauthors	28.18	20.12	0.58	463	463
Abroad	0.48	0.47	0.74	463	463
U.S. and China	0.11	0.14	0.16	463	463
Citations	**9.84**	**16.88**	**0.02**	463	463

Results: Other disciplines

	Non-AI students	AI Students		Non-AI students	AI students
Other disciplines (2,726 students)	Mean	Mean	P-value	N. of stud.	N. of stud.
Research career	0.52	0.51	0.76	1,363	1,363
Conditional on pursuing a research career (1,392 students)					
Private career/collaboration	0.19	0.20	0.64	696	696
Private career	0.08	0.09	0.39	696	696
Research career productivity					
N. of publications	5.22	5.45	0.54	696	696
At least one patent	**0.10**	**0.14**	**0.05**	696	696
N. AI publications	**0.27**	**1.21**	**0.00**	696	696
N. of coauthors	39.63	28.53	0.41	696	696
Abroad	**0.45**	**0.50**	**0.05**	696	696
U.S. and China	**0.13**	**0.17**	**0.08**	696	696
Citations	**12.10**	**15.22**	**0.00**	696	696

Discussion (1)

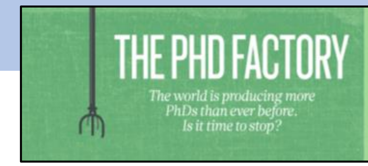

ARTIFICIAL INTELLIGENCE

The growing influence of industry in AI research

Industry is gaining control over the technology's future

- We do not observe any brain drain from public to private
  - Peculiarity of our sample: Limited national demand for AI talents
    - We observe a flow of talents' migration from France to the US and China to work in the private sector *[cor(Private career, US and China)=0.18]*
    - In our timeframe, AI graduates in Computer Science are less attractive for the private sector (they are less likely to conduct applied research than their counterparts in other disciplines)

Discussion (2)



- For the **relationship between AI and productivity**, we observe
  - **Path dependency** in AI research – Graduates keep doing what they have been trained for
  - **No gain in publication quantity** for AI graduates – New topics benefit from an innovative potential, but at the same time, face reluctance in being accepted by the scientific community (this applies also to networking and likelihood to pursue a research career)
  - **High impact of AI graduates post-graduation** – The community is growing and generate potential for a greater number of citations

What's new?

- Contribution to the people-centric framework
  - ''[…] 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, et al. 2024)
- Analysis covering an entire country for an almost 10-year timeframe
  - Privileged access to micro-data from a national comprehensive repository
- Implementation of a PSM strategy to overcome the biases of descriptive evidence

Conclusions

1) **Are AI students more likely to pursue a research career? No**
   - Acquiring AI knowledge does not significantly affect the likelihood of starting a research career

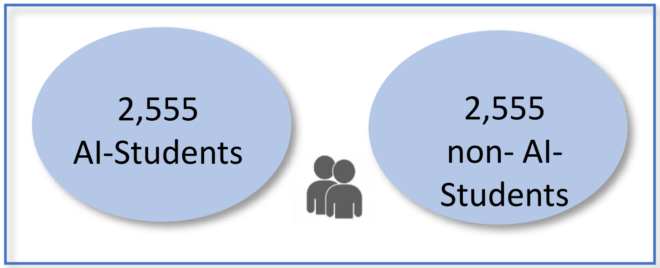
2) **Are AI students more likely to pursue a research career in the private sector? No**
   - On the contrary, in Computer Science, AI-trained students are less likely to pursue a research career in the private sector (no public to private brain drain)

3) **Are AI students more productive? Partially**
   - AI-trained students are more likely to keep working on AI after graduation and receive more citations
   - For disciplines other than Computer Science, AI-trained students are more likely to leave France and to patent

Thank you for your attention!

Backup slides

PSM regression



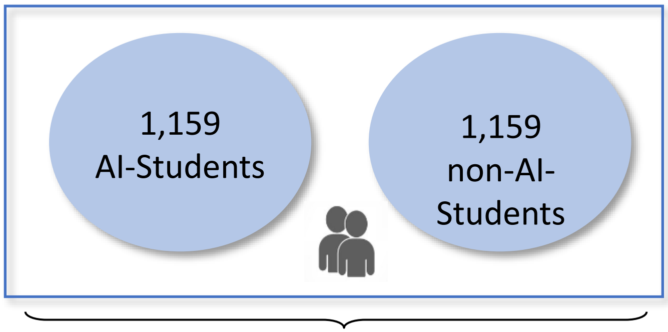
- Research career

Dependent variable: AI student	Estimate	Std. Error	Statistic	P-value
(Intercept)	-3.14***	0.09	-35.75	0.00
High productive supervisor	-0.63***	0.06	-11.23	0.00
AI supervisor before student enrollment	1.88***	0.04	42.13	0.00
Supervisor with a private affiliation	-0.37***	0.06	-6.51	0.00
PI supervisor	-0.28***	0.06	-4.61	0.00
Mentorship experience	0.09***	0.01	9.59	0.00
Bordeaux	-0.44***	0.14	-3.24	0.00
Lille	-0.03	0.12	-0.26	0.80
Lyon	-0.29***	0.10	-2.96	0.00
Marseille	-0.54***	0.15	-3.52	0.00
Montpellier	-0.32**	0.13	-2.40	0.02
Nantes	0.05	0.15	0.32	0.75
Others	0.18***	0.05	3.40	0.00
Rennes	-0.10	0.12	-0.81	0.42
Strasbourg	-0.77***	0.17	-4.47	0.00
Toulouse	0.07	0.09	0.75	0.45
Defense year 2011	-0.04	0.10	-0.34	0.74
Defense year 2012	-0.05	0.10	-0.49	0.63
Defense year 2013	-0.05	0.10	-0.55	0.58
Defense year 2014	-0.13	0.10	-1.32	0.19
Defense year 2015	-0.12	0.10	-1.24	0.22
Defense year 2016	-0.12	0.10	-1.17	0.24
Defense year 2017	-0.08	0.10	-0.79	0.43
Defense year 2018	-0.01	0.10	-0.05	0.96
Pseudo R-squared	0.12			
Number of Observations	35,492			

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.

PSM regression

Conditional on pursuing a research career



- Private career
- Productivity

Dependent variable: AI student	Estimate	Std. Error	Statistic	P-value
(Intercept)	-3.44***	0.13	-26.19	0.00
High productive supervisor	-0.64***	0.08	-7.92	0.00
AI supervisor before student enrollment	1.99***	0.07	30.15	0.00
Supervisor with a private affiliation	-0.34***	0.08	-4.01	0.00
PI supervisor	-0.30***	0.09	-3.36	0.00
Mentorship experience	0.09***	0.01	6.65	0.00
Bordeaux	-0.41**	0.19	-2.14	0.03
Lille	0.11	0.17	0.65	0.52
Lyon	-0.18	0.14	-1.29	0.20
Marseille	-0.50**	0.22	-2.27	0.02
Montpellier	-0.39*	0.20	-1.97	0.05
Nantes	0.12	0.22	0.54	0.59
Others	0.24***	0.08	2.94	0.00
Rennes	-0.11	0.18	-0.63	0.53
Strasbourg	-0.75***	0.24	-3.06	0.00
Toulouse	0.20	0.13	1.57	0.12
Defense year 2011	0.03	0.15	0.21	0.83
Defense year 2012	0.04	0.15	0.29	0.77
Defense year 2013	-0.03	0.15	-0.18	0.86
Defense year 2014	-0.17	0.15	-1.14	0.26
Defense year 2015	-0.13	0.15	-0.86	0.39
Defense year 2016	-0.07	0.15	-0.47	0.64
Defense year 2017	0.06	0.14	0.42	0.67
Defense year 2018	-0.01	0.15	-0.08	0.94
Pseudo R-squared	0.13			
Number of Observations	19,259			

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.

Results: Computer Science + Other disciplines

	Non-AI students	AI Students		Non-AI students	AI students
All disciplines (*5,110* students)	Mean	Mean	P-value	N. of stud.	N. of stud.
Research career	0.45	0.45	0.87	2,555	2,555
Conditional on pursuing a research career (2318 students)					
Private career/collaboration	0.20	0.18	0.46	1,159	1,159
Private career	0.09	0.09	0.61	1,159	1,159
Research career productivity					
N. of publications	4.88	4.84	0.89	1,159	1,159
At least one patent	0.11	0.12	0.30	1,159	1,159
N. AI publications	**0.42**	**1.30**	**0.00**	**1,159**	**1,159**
N. of coauthors	35.06	25.17	0.32	1,159	1,159
Abroad	0.47	0.49	0.20	1,159	1,159
U.S. and China	**0.12**	**0.15**	**0.03**	**1,159**	**1,159**
Citations	**11.20**	**15.88**	**0.00**	**1,159**	**1,159**

Career validation of 100 PhD graduates

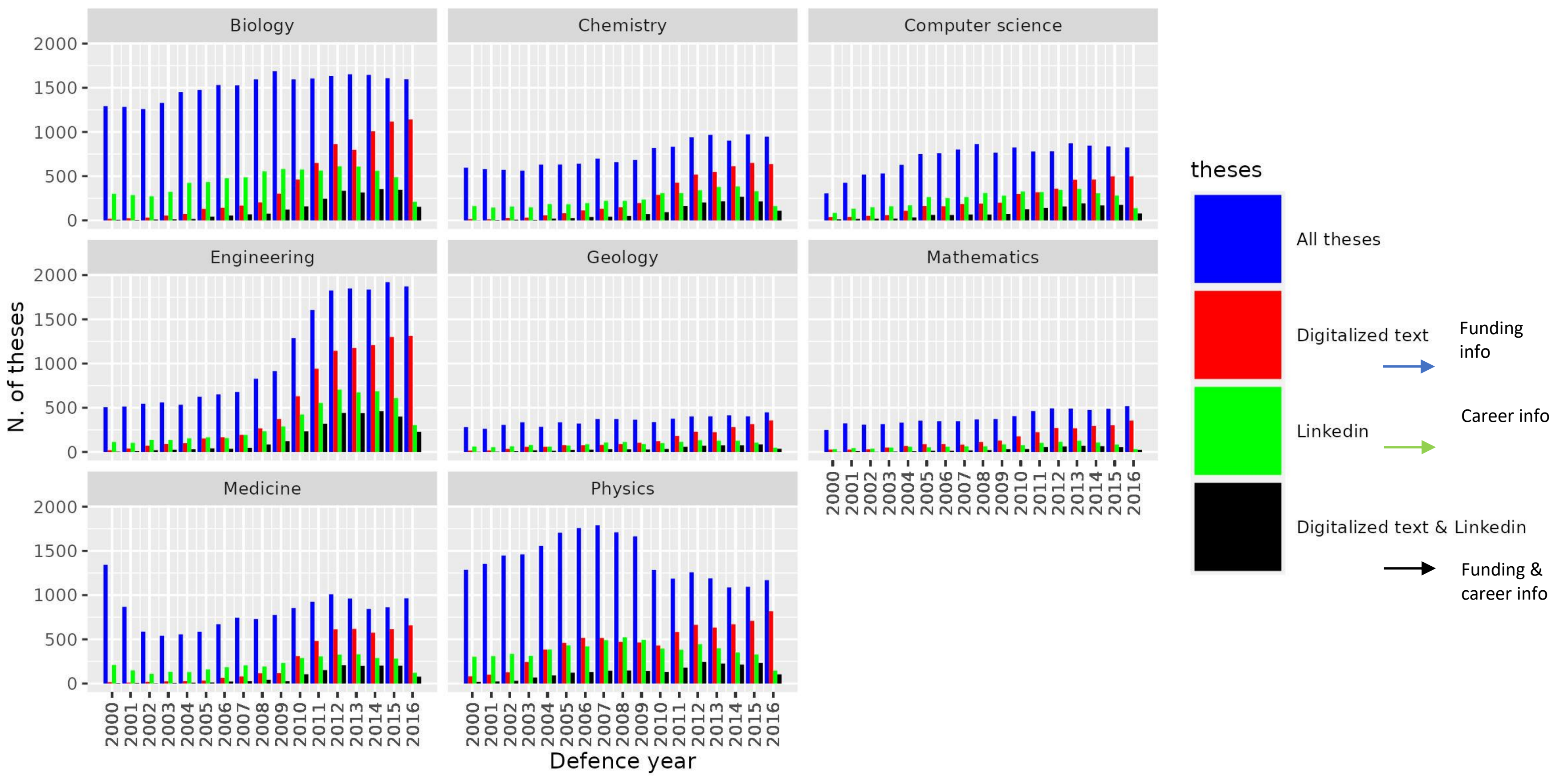
		Classification based on LinkedIn, ResearchGate, and personal websites	
		Research Careers (58 students)	non-Research Careers (42 students)
Classification based on the career proxy obtained from patent and publication data	Research Careers (50 students)	TP = 48	FP = 2
	non-Research Careers (50 students)	FN = 10	TN = 40

Content distance between AI and non-AI theses, by discipline

Discipline	N. of AI theses	N. of non-AI theses	**Average similarity (Standardized)**
Computer science	1976	4598	**0.72**
Mathematics	366	3177	**0.02**
Engineering	1366	12597	**-0.10**
Geology	87	2733	**-0.43**
Physics	297	8688	**-0.46**
Medicine	203	6820	**-0.74**
Chemistry	34	4619	**-0.80**
Biology	210	12192	**-0.88**

NOTE: This table compares the similarity of the content a random sample of 1,000,000 pairs of theses drawn from our study sample. Specifically, it assesses the similarity of the texts of the titles and abstracts of each pair of theses by using a neural network algorithm for text analysis.

Data coverage subsample 2000-2015



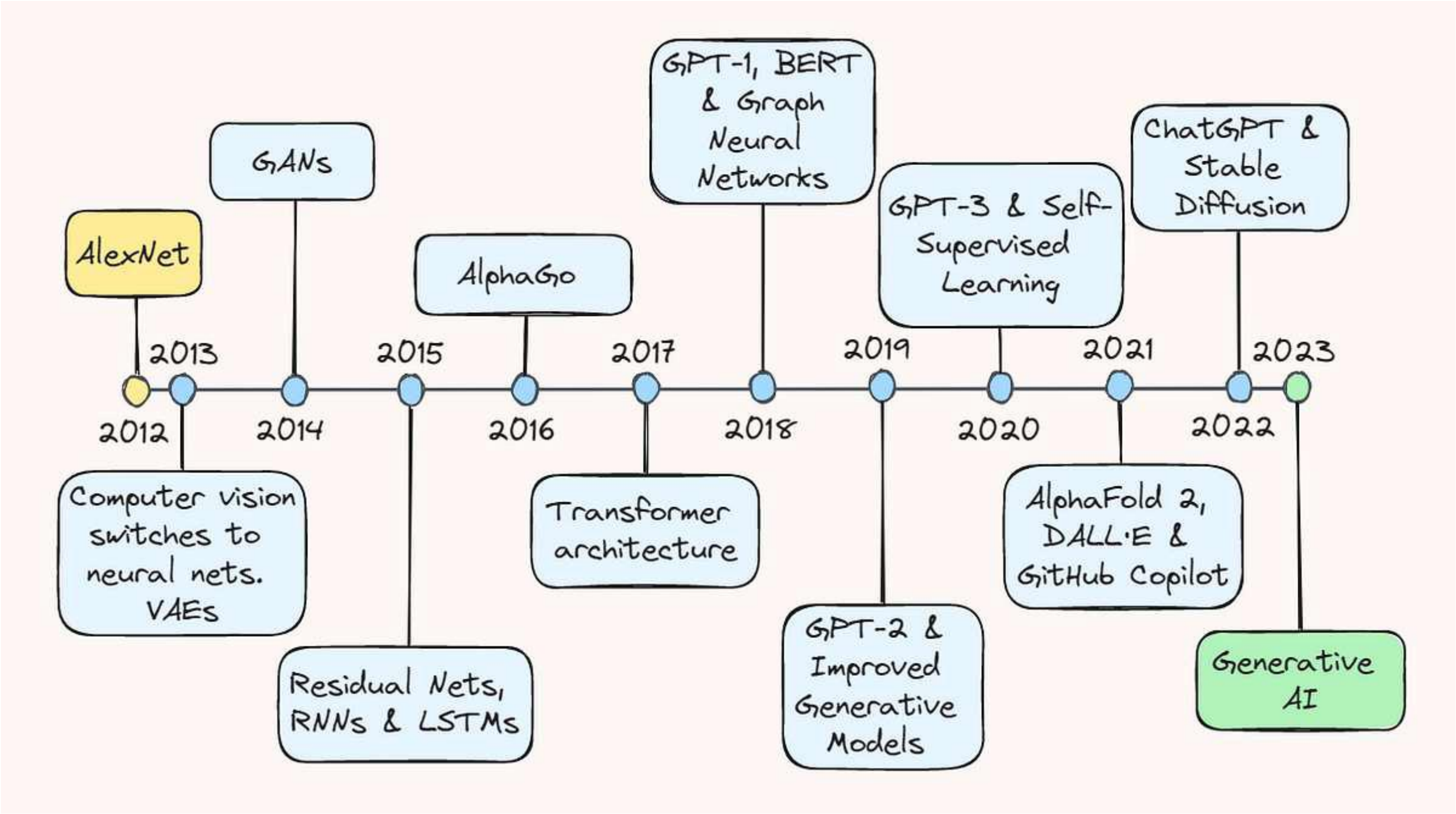
Intensity of AI keywords in the text by research field: Subsample of 1000 theses

	Intensity of AI keywords	N. of theses
Biology	12.79 [AI keywords/10,000 words]	58
Chemistry	8.95 [AI keywords/10,000 words]	7
Computer Science	**33.18 [AI keywords/10,000 words]**	**560**
Engineering	27.90 [AI keywords/10,000 words]	313
Geology	14.21 [AI keywords/10,000 words]	18
Mathematics	16.45 [AI keywords/10,000 words]	140
Medicine	18.10 [AI keywords/10,000 words]	52
Physics	16.74 [AI keywords/10,000 words]	62

Intensity of AI keywords in the text

Thesis classification according to titles and abstracts	Intensity of AI keywords in the full text	N. of theses
AI theses = 1	26.99 [AI keywords/10,000 words]	1,210
AI theses = 0	1.18 [AI keywords/10,000 words]	11,172

Ten Years of AI



[Source: Towards Data Science]