The Privatization of AI Research(-ers): Causes and Potential Consequences

– From university-industry interaction to public research brain-drain? –

Roman Jurowetzki[ϕ], Daniel S. Hain[ϕ], Juan Mateos-Garcia[†], and Konstantinos Stathoulopoulos[†]

[ϕ] *Aalborg University Business School, DK*
[†] *Nesta, UK*

January 18, 2021

Abstract: In this paper, we analyze the causes and discuss potential consequences of a perceived privatization of AI research, particularly the transition of AI researchers from academia to industry. We explore the scale of the phenomenon by quantifying transition flows between industry and academia, and providing a descriptive account and exploratory analysis of characteristics of industry transition. Here we find that industry researchers and those transitioning into industry produce more impactful research as measured by citations. Using a survival regression approach we identify mechanisms that trigger these university-industry transitions focusing on researcher characteristics, performance and research field as documented in bibliographic data. We find that, researchers working within the field of deep learning as well as those with higher average impact tend to transition into industry. These findings highlight the importance of strengthening academic research in public organizations within AI to balance a potential dominance of private companies and to maintain public supervision of the development and application of this technology.

Keywords: AI, university-industry interaction, researcher careers, private research, bibliometrics

1 Introduction

"The regulatory environment for technology is often led by the people who control the technology"

— *Zoubin Ghahramani*

In December 2020, renowned AI researcher Timnit Gebru was dismissed from her position as Ethics Co-Lead in Google Brain, Google's AI research unit. The reason was a disagreement with senior management about a conference paper where she and her co-authors outline the limitations and risks of large language models that have come to dominate AI research and become an important component of Google's technical infrastructure (Hao, 2020b). More specifically, the paper highlighted growing concerns about the fairness of models trained on biased and noisy internet data, their substantial environmental impacts and their limited ability to *understand* language as compared to *generate* plausible-reading text. Gebru's dismissal created an uproar in the AI research community - As of 21st December 2020, a letter in her support has garnered 7,000 signatures, including 2,600 Google employees.[1]

This controversy illustrates the increasing role that industrial labs are playing in AI research, where they are not only advancing new AI techniques but also studying the ethical risks and socio-economic impacts of AI systems. It also underscores the possibility that these labs may discourage their employees from pursuing research agendas that are not aligned with their commercial interests, potentially resulting in the development of AI technologies that are unfair, unsafe or unsuitable beyond the use-cases of the companies that build them. Ultimately, it bolsters the case for boosting AI research capabilities in academia and government in order to ensure that public interests can continue playing an active role in shaping and monitoring the trajectory of powerful AI systems. However, strong industry demand for AI researchers with advanced technical skills may create a brain drain from academia into industry that shrinks the pool of talent available for public interest AI research.

[1] https://googlewalkout.medium.com/standing-with-dr-timnit-gebru-isupporttimnit-believeblackwomen-6dadc300d382

In this paper, we use bibliographic data to measure this flow of researchers from academia to industry, study the factors driving it and consider its potential consequences. In doing this we provide, to the best of our knowledge, the first comprehensive quantitative analysis of AI researcher flows between academia and industry, contributing to the evidence base for science policies aimed at ensuring that AI evolves following a trajectory that is consistent with the public good.[2]

2 Background

2.1 Recent evolution of AI research

Last decade has witnessed an unprecedented acceleration in the development, diffusion and application of methods and technologies from the field of machine learning (ML) and Artificial Intelligence (AI). This has been driven by breakthroughs in the development of deep learning (DL) algorithms (LeCun et al., 2015) trained on a growing amount of public and private data (Einav and Levin, 2014). Since 2012 in particular, AI has flourished in academia and industry alike (Arthur, 2017), and is considered as a likely candidate to become a general purpose technology (GPT Trajtenberg, 2018; Goldfarb et al., 2019; Klinger et al., 2018).

2.2 Private sector participation in AI research

Private companies native to the digital economy such as Google and Facebook are playing an increasingly important role in basic AI research activities that used to be the domain of academia. For example, at the 2019 "Neural Information Processing Systems" (NeurIPS) conference, the main annual conference in AI and DL, Google research accounted for 167 of the accepted full papers (fractionalized by the number of authors), more than twice the amount of the second most represented institution, Stanford University (82 full papers). In addition to the growing amount of research

[2] Gofman and Jin (2019) also studies the AI brain drain but with a specific focus on the transitions of professors in AI universities and the subsequent (negative) impacts that their transition into industry has on academic entrepreneurship.

output generated, industry is also playing a dominant role in the creation of research tools, platforms, and frameworks. While the first DL frameworks - Theano and Caffe - emerged out universities, today's most popular frameworks for deep learning - Tensorflow (GoogleBrain) and PyTorch (Facebook AI) - have been developed by corporate players.

This change in the centre of gravity of AI research from academia to industry is also reflected in the career trajectory of researchers. Many star-scientists in the field of DL have over time moved to full- or part-time industry affiliations, for instance Geoffrey Hinton (Google), Yann LeCun (Facebook AI), Ian Goodfellow (Apple, prev. Google Brain), Zoubin Ghahramani (Uber AI) or Ruslan Salakhutdinov (Apple). The scale of movement of AI researchers from academia to industry has let to concerns about an "AI brain drain" (Sample, 2017; Gofman and Jin, 2019).

There are multiple (complementary) potential explanations for the increasing participation of private sector companies in basic research activities that may be expected to generate spillovers that benefit the competition.

1. Modern AI methods need to be trained on large datasets and computational infrastructures that have already been collected by these companies and may be difficult to transfer to researchers in academia for technical, data protection and privacy reasons.

2. There may be a disconnect between the type of AI research undertaken in academia and the needs of industry (Arora et al., 2020) as a consequence of innovation systems failures (Gustafsson and Autio, 2011) leading private companies to take basic research activities "in their own hands".

3. AI systems are increasingly becoming tightly integrated into the cloud infrastructure of private sector companies in order to help them address their own needs as well as those of third-party clients - the development of such systems may be easier to undertake in-house. In doing this, companies also seek to establish their AI systems as a de facto standard that increases the competitiveness of

complementary platforms and cloud computing services.

4. The opportunity to continue doing basic research and publishing results openly helps private sector companies attract top talent that is intrinsically attracted to environments where it is possible to conduct creative, "blue-skies" research and gain academic esteem.

5. Large technology companies with a substantial degree of market power are able to internalize many of the externalities generated by basic research by, for example, recruiting researchers, developers and engineers who have built up their AI skills using open source tools and research results generated in industry, and acquiring start-ups that sell AI-driven products and services.

2.3 Potential risks of AI privatization

In general terms, there are several reasons to worry about an encroachment on public research agendas by the private sector. Increasing participation of private sector organisations in basic research could lead to a potential homogenisation of public and private research spheres as academic researchers respond to financial incentives to commercialise their work in a way that limits its spillovers (David, 2003; David and Hall, 2006). Further, there is no guarantee that market-led opportunities correspond to social needs (Archibugi and Filippetti, 2018) or that they take into account technology's externalities and broader (perhaps longer-term) socioeconomic impacts.

If anything, industry-driven and dominated technological development tends to favor solutions which can be monetized in the short term, utilize incumbents' accumulated capabilities, resources, infrastructure, and other types of competitive advantages, thus making them less inclusive and posing higher barriers for new entrants (Hain and Jurowetzki, 2017). Ultimately, all this restricts the scope to steer technological development in a way that is aligned with societal goals. As biologist Paul Berg wrote in relation to the Asilomar conference that led to a moratorium on genetic modification of humans: "the best way to respond to concerns created by emerging knowledge or early-stage technologies is for scientists from publicly funded institutions to find common cause with the wider public about the best way to regulate - as early as possible. Once scientists from corporations begin to dominate the research enterprise, it will simply be too late" (Berg, 2008).

All these concerns are heightened in the case of AI because of its potentially pervasive impact. As a strong candidate for one of the near-future's GPTs, AI technologies are expected to cause major disruptions across multiple domains, from communication, production, transport to education and health, and more broadly socioeconomic dynamics – for example around public attitudes to privacy, autonomy or the right to an explanation for a decision. As a still emerging technology, AI's dominant technological trajectories are about to be established but there are increasing concerns about certain aspects of the industry-sponsored DL trajectory that has driven recent advances in the field.

Training deep neural networks requires enormous amounts of data and computing power (Marcus, 2018; Russell, 2019), often exclusively available to large industry players and costly in terms of energy use and carbon emissions (Strubell et al., 2019). While platforms and frameworks provided by industry (such as Tensorflow or PyTorch) dramatically decrease entry barriers and advance collective progress, the direction of search and effort along this trajectory reinforces the data and computation hungry DL paradigm. Strong demand for data has led researchers to exploit large online corpora that are increasingly being shown to incorporate a variety of gender and racial biases that are subsequently transmitted into the trained models and their outputs (Paullada et al., 2020). In the field of natural-language-processing (NLP), pretrained language models in need of enormous resources such as "Bidirectional Encoder Representations from Transformers" (BERT, GoogleAI Devlin et al., 2018) have become the *de facto* standard for research and industry alike, shifting attention and resources away from other "leaner" techniques - this concern was at the heart of the censored Timnit Gebru paper mentioned in the introduction.[3]

[3] It should be noted that a comparably big and resource intensive model (GPT-2 Radford et al., 2019) has been open-sourced by the nonprofit research lab OpenAI, which aims at counterbalancing corporate AI with a public-spirited approach o technology development. Interestingly, as the costs of basic AI research have increased, OpenAI been criticized for becoming more secretive and

Further downstream, a growing number of economists have expressed concerns that left in the hands of the private sector, AI's trajectory may evolve towards what is described as "the wrong kind of AI" (Acemoglu and Restrepo, 2019) which displaces workers without material impacts on productivity. Activists and critical scholars point, on their side, at the evidence of racial and gender biases in AI applications (Zou and Schiebinger, 2018). The opacity of DL systems and their propensity to experience important declines in performance when exposed to situations outside their training set (D'Amour et al., 2020) have raised questions about their suitability for high-stakes domains such as health (Marcus, 2018).

Ultimately, a shrinking space for high-impact public research about AI technologies is likely to lead to a loss in attention and knowledge, hampering the capacity of public authorities to regulate and utilize them, and limiting the extent to which they can be deployed in areas where there are less commercial incentives.[4] Already today, algorithmic bias, has been identified as one such problem, where technologies are clearly in conflict with social values and regulations but lack of technological insight are hindering regulation (Sweeney, 2013; Hajian et al., 2016; Zou and Schiebinger, 2018; Clark and Hadfield, 2019).

2.4 Studying university-industry transitions of AI researchers

Here, we analyze the causes and discuss potential consequences of this ongoing privatization of AI research, focusing particularly on the transition of AI researchers from academia to industry. We start by assessing the scale of the phenomenon by measuring transition flows between industry and academia, and providing a descriptive account and exploratory analysis of characteristics of industry transition, research topics, and temporal dynamics.

Having done this, we estimate the importance of various mechanisms that trigger

aggressive in fundraising in order to keep pace with their corporate competitors (see Hao, 2020a).

[4] In recent years public agencies have launched numerous funding calls and initiatives to support AI. Yet, it remains questionable to which extent research in such a dynamic and competitive domain can be supported within existing bureaucratic frameworks to an extent that would allow it to compete with private AI labs.

these university-industry transitions including researcher characteristics, performance and field of activity as documented in bibliographic data. Researchers with a preference for a corporate lifestyle, financial incentives, and less "taste for science" (Roach and Sauermann, 2010) may self-select into particular fields of research aligned with the interests of the industry through an (*researcher-push* mechanism. Further, the increasing demand for data and infrastructure in particular fields of AI research (e.g., DL) result in a *technology-push* providing incentives for AI researchers to seek for industry affiliation in order to get access to necessary resources beyond the capacities most universities offer Ahmed and Wahed (2020). Lastly, industry might indeed attempt to play a more active role in shaping the trajectories of AI research by either by recruiting star AI researchers *per se*, or researchers associated with central key-technologies - we refer to this as an *industry-pull* mechanism. To assess the relative importance of these factors we deploy a survival model where we estimate the probability that academic AI researchers will transition into industry. In doing this, we test the effect of a range of researcher characteristics related to their preferences for academia, their topical focus, and academic success.

Finally, we attempt to quantify the effect of university industry transition on researchers productivity (proxied through citations). To do this, we match industry transitions with similar peers that remained in academia. Here, we leverage insights on the mechanism triggering academia-industry transitions identified in the previous step. In a difference-in-difference analysis we investigate the impact of transitions.

3 Data and Methods

3.1 Data

We collect data from Microsoft Academic Graph (MAG), a scientometric database with more than 232 million academic documents (Wang et al., 2019). We leverage MAG's Fields of Study (FoS), a six-level topical hierarchy (Shen et al., 2018), to query its API with a list of hand-picked fields of study that cover key techniques in modern

AI research; *machine learning*, *deep learning* and *reinforcement learning.*

We bound the timeframe of our analysis between 2000 and 2020 and retrieve the academic publications containing at least one of the queried FoS. In total, we collect 786,118 AI research papers alongside their metadata such as citation count, publication year and venue, title and abstract, fields of study, author names and affiliations. We find that 1,165,913 scholars have developed or used AI methods in their research which has been published in 10,653 journals and presented at 3,150 conferences.

To investigate this paper's main objective, we aim at constructing the affiliation history of all researchers to be found as (co-) authors of the identified deep-learning paper. To do so, we leverage affiliation information to be found on the papers to identify 10,381 unique affiliations and construct the affiliation history for all authors. Then, we infer the type of an affiliation (industry or non-industry) using an expansive list of terms related to academic institutions and governmental agencies and find that 80.73% are non-industry affiliations. We use this variable to identify the authors' academia-industry transitions.[5]

3.2 Econometric strategy

To investigate the phenomenon of university-industry transition in AI research, we structure our analysis in three steps. First, we perform a basic exploratory data analysis to determine the magnitude and characteristics of academia-industry transitions as well as the effects it has on performance as measured by publication numbers and citations.

Second, we aim at identifying the drivers of university-industry transition. We here assume the transition of academics to industry to not happen at random, but rather being subject to self-selection by the researchers (technology-push) as well as external selection by potential employers (industry-pull). Using the affiliation history of all deep learning researchers which either remain in academia or at one point transit to

[5] This is complemented, in our exploratory data analysis, with an alternative strategy where we match researcher affiliations with the Global Research Identifier (GRID) database using the method described in Klinger et al. (2018), providing information about the character of an organisation (in particular, whether it is a private company or an educational institution).

industry, we perform a survival analysis (Cox proportional hazard model) where we consider researcher characteristics, their research interactions and overall pre-transition academic performance as potential candidates for transition drivers.

Third, we perform a regression analysis of the consequences of university-industry transition in terms of research performance. To address the assumed endogenous selection of researchers that transit to industry (ca. 10%), we apply the following strategy to mimic a (quasi-) experiential setting. For every researcher that undergoes an university-industry transition, we perform a propensity-score matching (PSM) procedure to find their most similar counterpart among peers which remained in academia throughout their observed career.[6] We then for every academia-remained an "artificial transition" point, which we define to happen after the same number of periods observed as the actual transition of their academia-industry matched peer. By doing so, we aim at constructing an empirical setting that allows us to tackle the question: "What would have happened to the researcher if she would have remained in academia?". Using this matched sample, we then perform a difference-in-difference regression analysis, where we contrast the effect of university-industry transitions of researchers which undergo this transition with peers that remained in academia.

3.3 Variables

In the following, we describe the construction of and rational behind the variables used in the set of survival models (transition drivers) as well as the difference-in-difference models (impact of university-industry transition). A compact summary thereof can be found in Table 1. To tackle remaining endogeneity concerns, all independent and control variables are lagged by one year.

[6] We match researchers on their main field of study, mean number of annual publications and received citations, and gender. We also enforce that matched researchers need to have the exactly same number of periods observed in our sample.

Table 1: Variable Description

Variable	Model	Description
		Dependent Variables
transition	Surv.	Dummy indicating the year of academia-industry transition.
citation_{rank}	DiD	Percentage rank of researcher's received citations in the corresponding year.
		Independent Variables
DeepLearning	Surv.	Dummy variable for researcher's publication of min. 1 deed learning paper in corresponding year.
cent^{dgr}	Surv.	Researcher's degree centrality in overall co-publication network.
$\text{cent}^{dgr-ind}$	Surv.	Researchers degree centrality in industry co-publication network.
switcher	DiD	Dummy variable indicating researcher to at one point undergo a university-industry transition.
transited	DiD	Dummy variable indicating the researcher has undergone a university-industry transition.
transited^{t}	DiD	Number of years since researcher's university-industry transition.
		Control Variables
seniority	Surv., DiD	Years since first observed publication.
gender	Surv., DiD	Dummy variable for researcher's gender (fermale = 0, male = 1)
paper^{n}	Surv., DiD	Number of researcher's publications in corresponding year.
cit_{ln}^{cum}	Surv.	Cumulative number of researchers citations (natural logarithm).
StudyField	Surv., DiD	Categorical control for most popular field of study in the researcher's publications.
Year	Surv., DiD	Categorical control for the corresponding year.

Dependent Variables

The **dependent variable** in the survival analysis (transition drivers) is a dichotomous indicator which takes the value of *zero* in the years a researcher has been affiliated with academia in the previous year and continuous to do so in this year, and takes the value of *one* in the year the researcher's first changes to a corporate affiliation. To do so, we use the affiliation information found in the researcher's published papers in the corresponding year. In order to avoid short term affiliations (eg. project based co-affiliation, internship, visiting researcher programs), we compute the affiliation of researchers on annual basis, and assign it to the institution found on most papers published by the researcher in the corresponding year.

When analysing the effect of university-industry transitions on researcher's career in a difference-in-difference regression, we use the percentage-rank of the researcher's received citations in the corresponding year (cit^{rank}) as dependent variable to approximate research performance and impact. Here, *zero* corresponds the researcher with the lowest and *one* to researcher with the highest citation rank in the corresponding year.

Independent Variables

We construct additional **independent variables** in the following way:

DeepLearning: A dummy variable indicating that the researcher published at least one paper in the corresponding year which includes the MAG field-of-research tag for either "Deep Learning" or one of the most related tags.[7] Since deep learning represents a field of research where access to large amounts of data and computing power represents one of the main competitive advantages, we expect deep learning researchers to be more likely to undergo a university-industry transition in their career path (researcher-push and industry-pull).

cent^{dgr}**:** The authors degree-centrality in the co-publication network of papers published in the corresponding year. Edges are weighted by the number of researchers per paper, so that an increasing number of authors on a paper leads to a decreasing edge-weight. This variables approximates the researcher's current embeddedness within the research community. We would expect researchers that are more embedded in the community to be better networked and influential and therefore attractive for industry recruiters (industry-pull).

cent$^{dgr-ind}$**:** The authors degree-centrality in the co-publication network of papers published in the corresponding year, where only edges to researchers with a current industry-affiliation are included. This variable approximates the researcher's proximity to industry actors. We expect researchers that are already collaborating actively with industry to be more likely to transition into industry.

paper^{n}**:** The number of papers (co-) authored by the researcher in the corresponding year, approximating the quantity of the researcher's current

cit$_{ln}^{cum}$**:** Accumulated number of citations to the researcher's current and historical publications. Assuming cumulative citations to have an decreasing marginal effect, we transform this variable's value by its' natural logarithm.

[7] In this case, we include the field tags that most often co-occur together with "Deep Learning" in our corpus. These are Recurrent neural network, Time delay neural network, Types of artificial neural networks, Deep neural networks, Autoencoder, Deep belief network

Control Variables

We approximate **Seniority** by the number of years since we observe a researcher's first publication in the data.[8] We also control for the researcher's **Gender**, which we infer from their name using the GenderAPI. This dummy variable takes the value of one for male researchers, and zero for female. We also include categorical controls for the researchers main MAG **field-of-research** (Shen et al., 2018), where we assign the MAG field which is most often found within the categories of her publications in the corresponding year. Finally, to cover time-dependent exogeneous effects, we also for the current year.

3.4 Descriptive statistics

Table 2 provides descriptive statistics and Table 3 the corresponding correlation matrix on our full dataset.

Table 2: Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
switch	83,002	0.047	0.212	0	0	0	1
affiliation	83,002	0.809	0.393	0	1	1	1
seniority	83,002	8.136	4.169	3	5	11	21
gender	83,002	0.872	0.334	0	1	1	1
DeepLearning	83,002	0.113	0.316	0	0	0	1
paper_n	83,002	1.431	2.460	0	0	2	76
citation_rank	83,002	0.319	0.381	0	0	0.7	1
cit_n_cum	83,002	2.417	1.673	0	1.050	3.527	9.630
cent_dgr	83,002	1.190	2.332	0	0	1.5	65
cent_dgr_ind	83,002	1.094	2.028	0	0.000	1.357	56.343

[8]Note that due to the our sample only including publications from earliest 2000, this variable is left-censored by our starting point.

Table 3: Correlation Matrix

	(1)	(2)	(3)	(4)	(5)	6)	(7)	(8)	(9)
(1) switch									
(2) affiliation	-0.01∗								
(3) seniority	-0.08∗	-0.02∗							
(4) gender	-0.01	0.01	0.03∗						
(5) dl_researcher	-0.01	-0.06∗	0.17∗	0.02∗					
(6) paper_n	-0.06∗	0.02∗	0.21∗	0.03∗	0.29∗				
(7) citation_rank	-0.05∗	-0.01	0.10∗	0.02∗	0.27∗	0.25∗			
(8) cit_n_cum	-0.02∗	-0.03∗	0.18∗	0.04∗	0.23∗	0.31∗	0.12∗		
(9) cent_dgr	-0.04∗	-0.01∗	0.23∗	0.03∗	0.45∗	0.66∗	0.38∗	0.45∗	
(10) cent_dgr_ind	-0.04∗	0.00	0.22∗	0.03∗	0.34∗	0.61∗	0.22∗	0.45∗	0.78∗

$^{*}p<0.001$

4 Exploratory Data Analysis

4.1 Thematic and organisational trends

We begin our exploration of the data by considering the overall evolution of activity (Figure 1), level of company participation in research (Figure 2) and thematic focus of different organisation types (Figure 3).

Figure 1 shows the growth in the levels of AI research in recent years specially driven by a fast increase in the levels of research involving deep learning techniques, which have gone from accounting for a negligible amount of AI research in 2012 to ca 30% of all the papers published in 2019.

Figure 2 focuses on the level of company participation in AI research. It shows that papers involving authors with a company affiliation have started capturing a larger share of research since the 2010s. This is consistent with the idea that private sector organisations are playing a stronger role in AI research although, at least in overall volume of activity they are very far from dominant.

In Figure 3 we look at the share of all papers involving an educational institution or a company in a year that contain a field of study (focusing on the 20 most frequently occurring fields of study in the data). We note in particular that deep learning was over-represented in private sector research by comparison to academia but educational

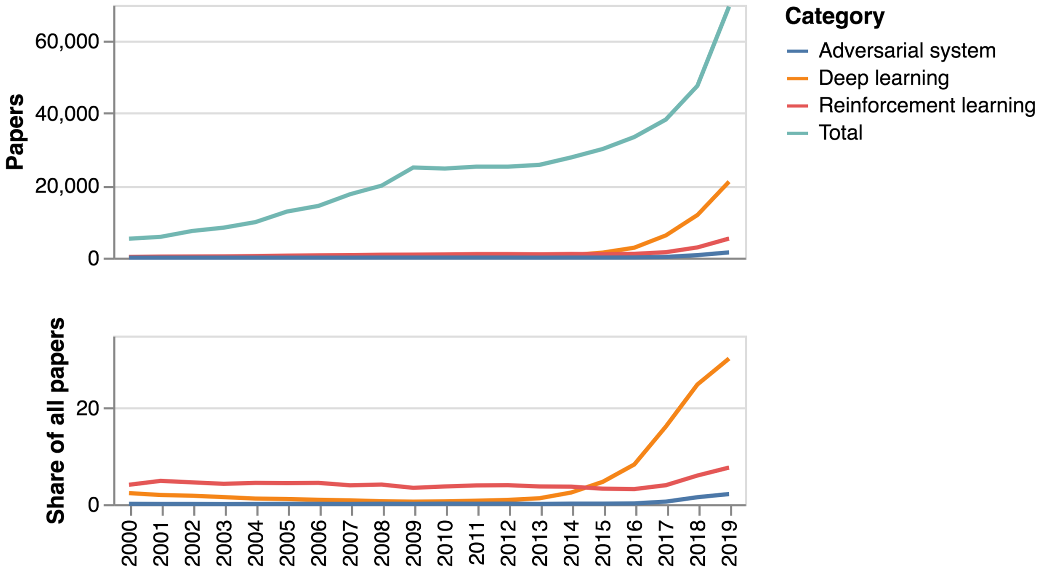

Figure 1: Evolution of activity in absolute terms for the corpus and selected fields of study (top panel) and as a share of all papers for selected fields of study (bottom panel)

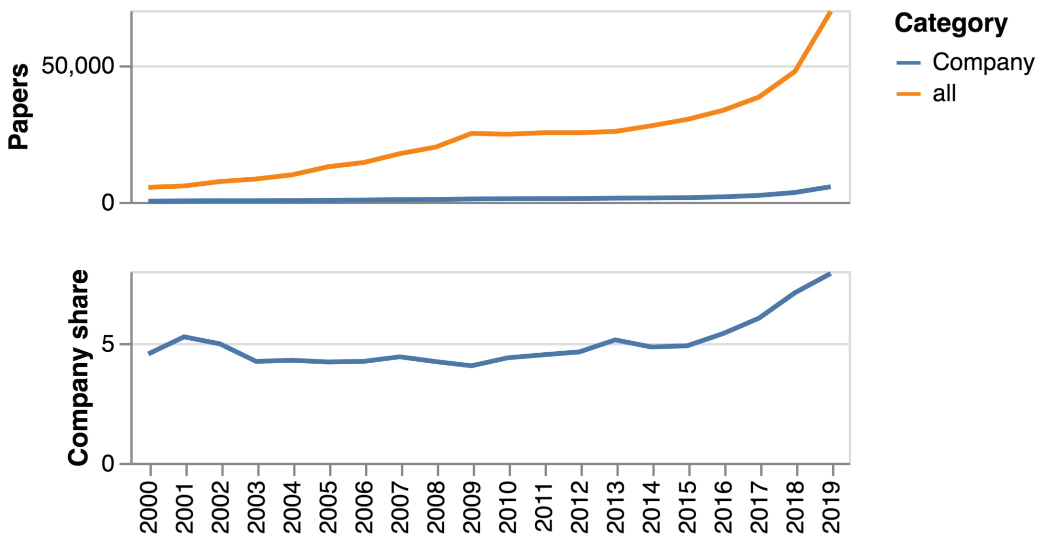

Figure 2: Evolution of organisational participation in absolute terms (top panel) and share of papers with company participation (bottom panel)

institutions seem to have caught up in recent years. Companies are also more active in reinforcement learning and, more broadly, computer science topics - this could also be linked to the finding elsewhere in the literature that private sector companies specialise in more scalable and computationally demanding techniques ((Klinger et al., 2020; Ahmed and Wahed, 2020)

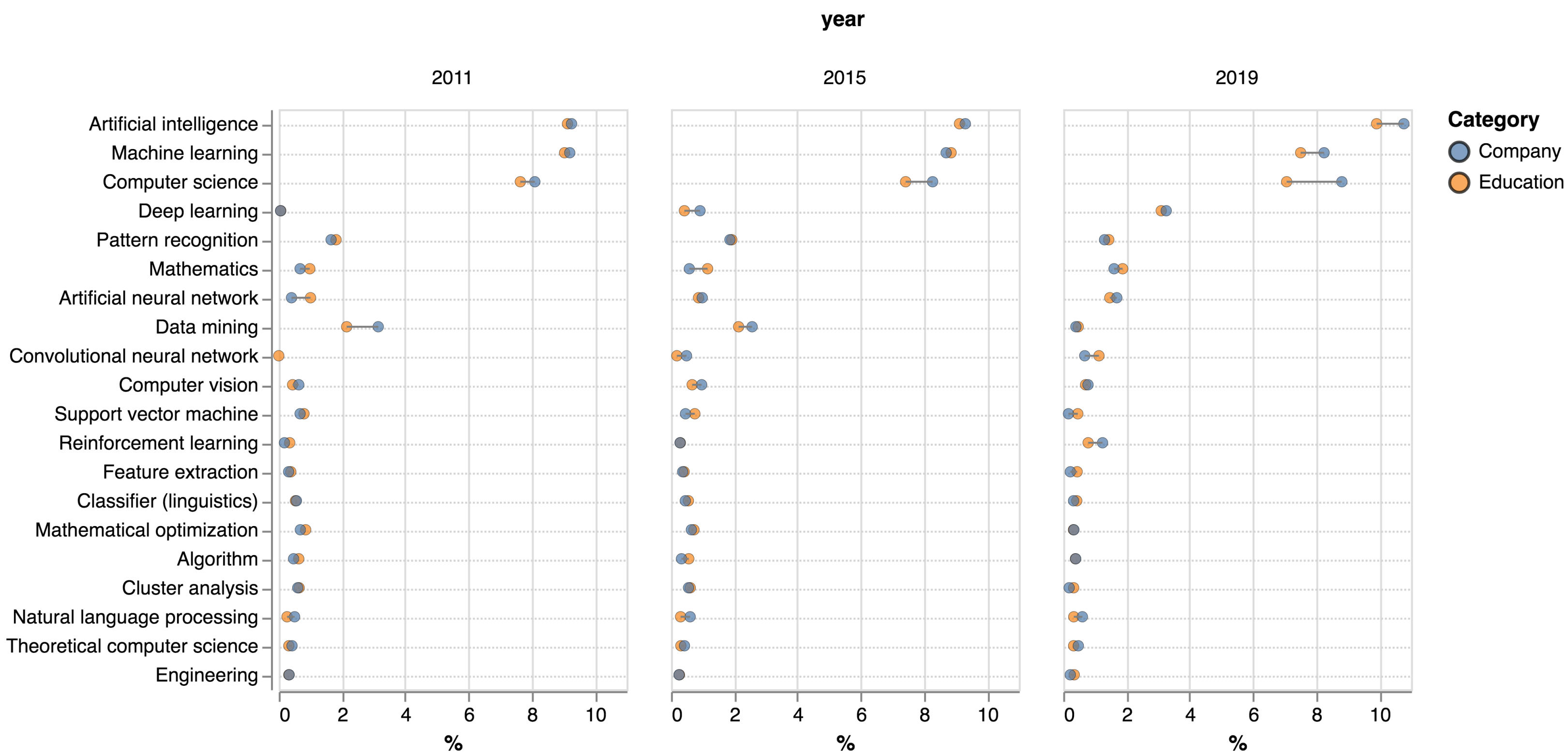

Figure 3: Share of all papers involving companies (orange point) and educational institutions (blue point) that have been assigned a field of study each year.

4.2 Transition trends

We move on to analyse the dynamics of researcher transitions, going from a macro picture that considers all transition types in the data (Figure 4) to focus on researcher flows between university and industry (Figure 5) distinguishing between academic institutions in different positions of Nature's global university rankings (Figure 6) and finally considering the main educational sources and industrial destinations of AI research talent (Figure 7).

Figure 4 shows the evolution in levels of transition by transition type in total (top panel) and as share of all transitions (bottom panel). It shows rapid growth in the number of all types of researcher transitions within the AI ecosystem while underscoring that researcher mobility between academic institutions is the dominant type of transition, reflecting the prevalence of educational institutions in terms of publications.

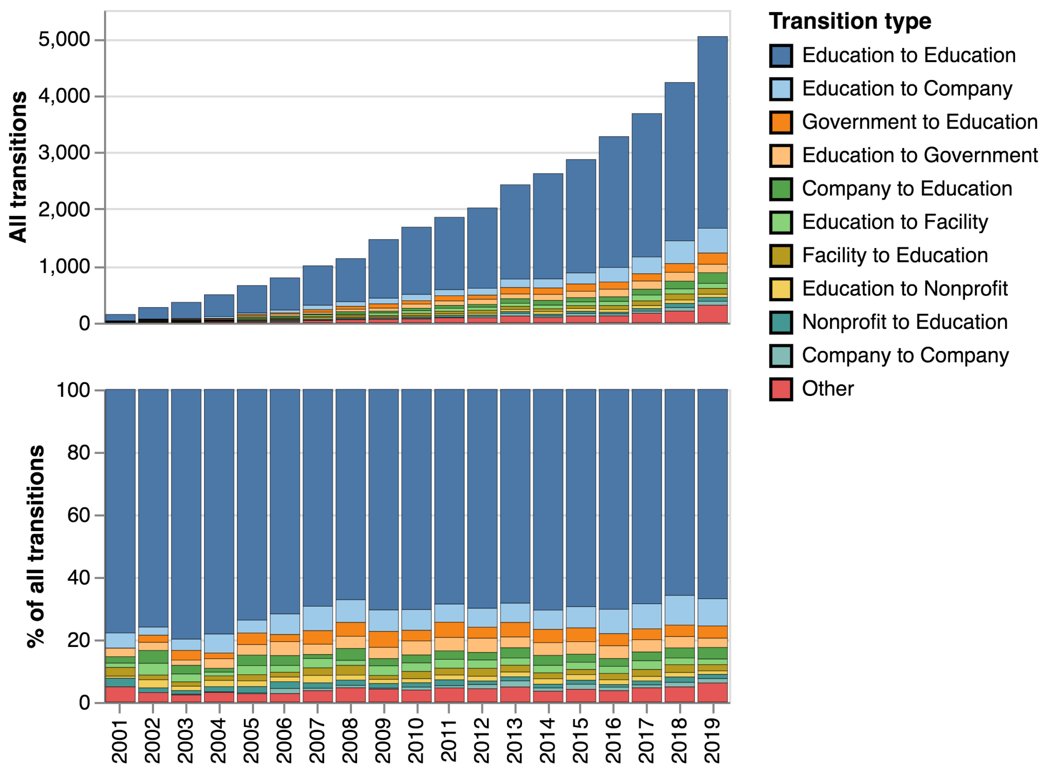

Figure 4: Number of researcher transition by types (top panel) and transition types as a share of the total (bottom panel).

In Figure 5 we concentrate on researcher transitions between educational institutions and industry taking into account that flows can go in either direction. Our analysis shows that in net terms, researcher flow favours industry (consistent with the hypothesis of a 'brain drain') from academia to industry but also that there is a non-trivial number of industry researchers transitioning into academia. One potential explanation for this is that having moved into industry, academic researchers may not like the environment and decide to return to the public sector.

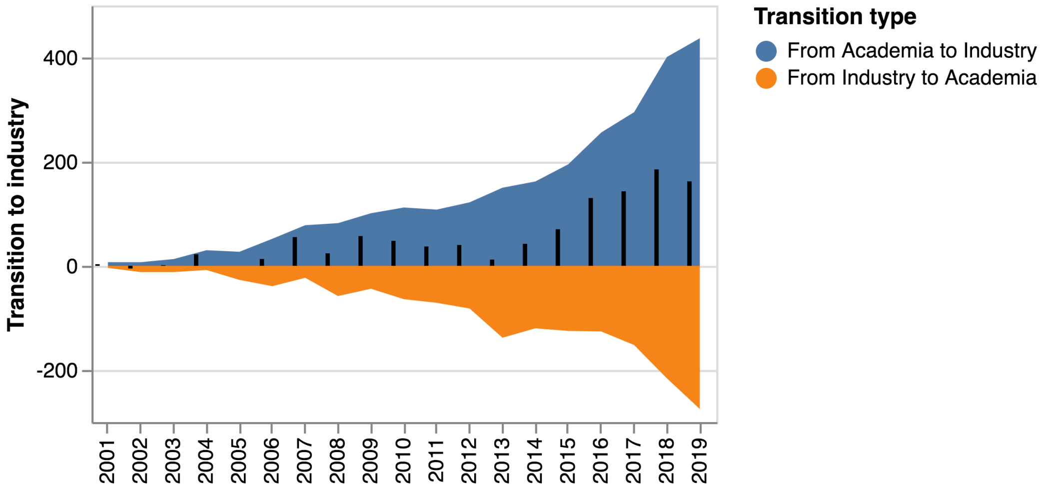

Figure 5: Researcher transitions between education and industry (blue area) and industry and education (orange area). Net flow in black bars.

When looking at labour flows between academia and industry it is important to take into account the prestige of the organisations involved, which could be seen as a rough proxy for the 'quality' of the researchers involved. To do this, we have fuzzy-matched institution names from Microsoft Academic Graph with the 2020 Nature Index, ranking institutions based on the quality of their research in the Natural Sciences.[9]. In Figure 6 we present the share of transitions from institutions in different positions of the ranking into industry (the Nature Index only includes 500 institutions so those below are labelled as 'unranked'. The chart shows a clear and strong correlation between a university's prestige and its propensity to experience a flow of researchers into industry. In particular, 25% of the AI researcher transitions from institutions in the top 5 of the Nature Index were into industry - this suggests that industry tends to attract AI

[9]https://www.natureindex.com/annual-tables/2020/institution/academic/all

researchers from elite institutions, perhaps with ‘super-star’ qualities.

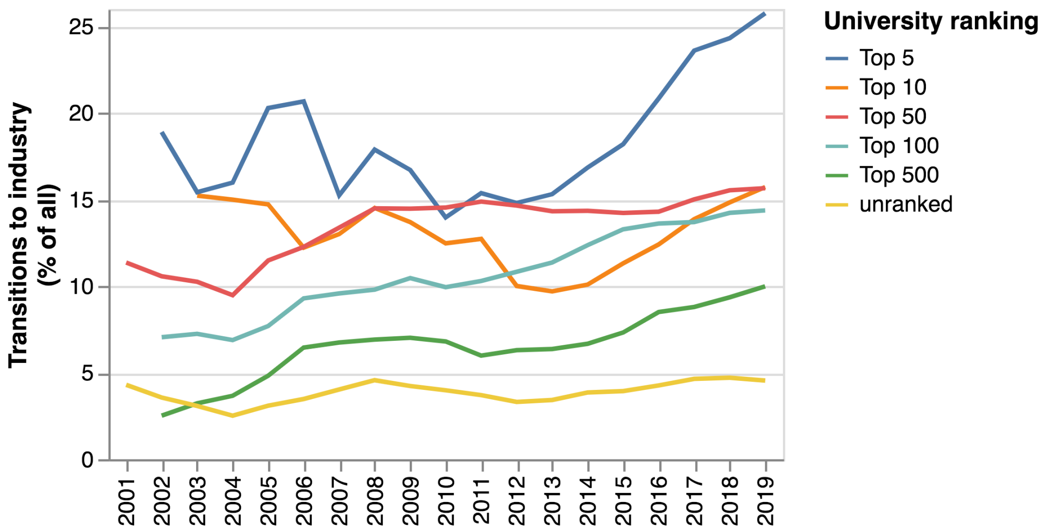

Figure 6: Share of all transitions from education to industry by year and position of university in Nature University ranking.

Figure 7 drills down further to consider what are the top educational sources of talent moving into industry and what are the top industrial destinations. We see that the top academic sources of talent in the vertical axis are prestigious institutions such as Carnegie Mellon, Stanford, Princeton, MIT etc. The top destinations for AI talent (in the horizontal axis) are tech companies, and particularly Google. We note the rapid growth in the share of *all AI researcher transitions* from source institutions into Google between an early period (before 2015) and a late period (after 2015) - in many cases Google accounts for more than 10% of all researcher transitions into industry from top institutions. We also see that Facebook has rapidly growth in importance as a destination for AI researcher talent since 2015.

4.3 Characterisation of academic researchers transitioning into industry

To observe the phenomenon of university-industry transition in AI research, we first aim at identifying transiting researchers as well as the point in time where this transition happens in their career paths. We leverage bibliographic metadata of publications in the MAG database to identify the author affiliation as to be found in publications.

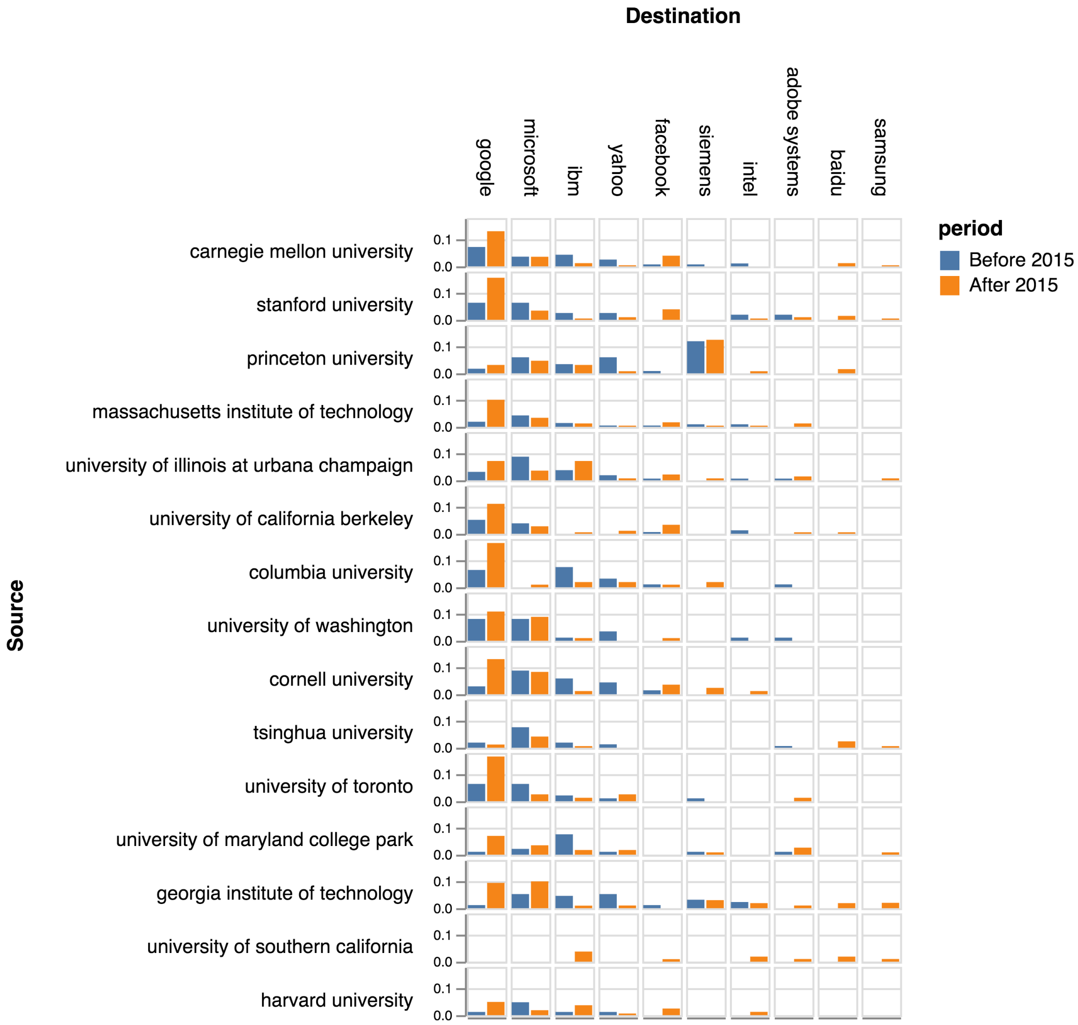

Figure 7: Share of all transitions from a source academic institution (vertical axis) accounted by a destination company (horizontal axis) in pre-2015 and post-2015 period.

Since the main unit of analysis in the following exercises is the researcher-year level, in case of multiple affiliations of an author during one year we assign the researchers to the type of affiliation to be found on most publications. In case of a draw, we prioritize affiliations in the order they are mentioned on the publication.

In doing so, we are able to identify three distinct research-career profiles over time: (i.) academia-only, (ii.) industry-only, and (iii.) university-industry transitions. We define the latter as researchers which started their career in academia, but at one point become mainly associated with industry for at least two consecutive years. We choose two years to avoid a overestimation of transitions due to sabbaticals, internships, and other short time industry affiliations. We do not further differentiate between additional career paths, for instance "academia returnees" or "serial switchers". To derive meaningful information regarding the researchers' career paths, we also exclude researchers that could not be observed in the MAG data for at least five years. Furthermore, due to the timeliness of the phenomenon under research, we exclude researchers which have the last time been observed before 2015.

This leads us to the following picture of the composition of AI research.

Table 4: Number and characteristics of author types

author_type	n	share	paper_mean	cit_mean	gender
academia	54113	0.89	0.96	1.33	0.82
industry	1837	0.03	0.79	2.78	0.85
switcher	4751	0.08	1.18	4.23	0.86

Table 4 reports counts and mean values for characteristics and publication performance for the different researcher groups. Overall our restricted sample contains 60.701 individual AI researchers with approximately 90 percent in academia, 3 percent in industry and 8 percent transitioning from academia to industry in the course of their career. As for the academic productivity, we observe that AI researchers in industry are least productive with regard to numbers of papers produced per year, which is not surprising given that output in industry is measured differently than in academia. However, their impact in terms of received citations per year is on average double that of academic researchers. Transitioning AI scholars show the highest values for quantity and impact. This could indicate "cherry picking" by the private sector. As for the *gender* of the scholars, we can see that the field is mainly populated by men, with the highest share of women in employed in academia.

5 Econometric analysis

In a next step, we attempt to investigate the drivers and mechanisms of university-industry transition within AI research.

5.1 Drivers of switching - Survival analysis

In the following we perform a survival analysis (Cox-proportional hazard model) on the probability to transit from university to industry.

Table 5 reports the results of a Cox proportional-hazard model, modelling the probability of an researcher to undergo a university-industry transition in the corresponding period. Panel (1) only includes the control variables, panel (2) additionally the deep learning dummy, panel (3) adds the network-related independent variables, panel (4) the research-performance related independent variables, and finally model (5) includes all variables jointly.

In model (1) we observe no significant effect for seniority, while the coefficient for gender is negative and significant on the 1% level, indicating female researchers to be more likely to transit to a career in industry. This effect remains persistent for all following models.

The DeepLearning variable included in model (2) has a relatively high positive coefficient, significant on the 1% level. This is in line with our initial expectations that the characteristics of this particular research field make a transition to industry more attractive. This is in line with the above observation of the strong engagement of companies within deep learning.

In model (3) we see that centrality within the co-authorship network in general increases the probability of transition, while–somewhat surprising–the centrality only

Table 5: Cox Proportional Hazard Regression: Probability of university-industry transition

	Dependent variable:				
	Industry Transition				
	(1)	(2)	(3)	(4)	(5)
seniority	-13.358^{***}	-13.355^{***}	-13.346^{***}	-13.418^{***}	-13.423^{***}
	(5.113)	(5.099)	(5.068)	(5.100)	(5.113)
gender	0.230^{***}	0.227^{***}	0.227^{***}	0.184^{***}	0.183^{***}
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
DeepLearning		0.568^{***}			0.177^{***}
		(0.017)			(0.017)
cent^{dgr}			0.034^{***}		-0.002
			(0.002)		(0.003)
$\text{cent}^{dgr-ind}$			0.037^{***}		-0.043^{***}
			(0.002)		(0.003)
paper^{n}				-0.066^{***}	-0.054^{***}
				(0.003)	(0.003)
cit^{rank}				0.540^{***}	0.574^{***}
				(0.020)	(0.020)
cit^{cum}_{ln}				0.376^{***}	0.386^{***}
				(0.004)	(0.004)
Study Field Control	Yes	Yes	Yes	Yes	Yes
Year Control	Yes	Yes	Yes	Yes	Yes
N	479,093	479,093	479,093	479,093	479,093
Pseudo R^2	0.174	0.174	0.175	0.188	0.188
Wald Test	$4{,}500^{***}$	$6{,}196^{***}$	$5{,}177^{***}$	$12{,}454^{***}$	$12{,}927^{***}$
LR Test	$91{,}412^{***}$	$91{,}871^{***}$	$92{,}129^{***}$	$99{,}825^{***}$	$99{,}967^{***}$
Score (Logrank) Test	$36{,}988^{***}$	$37{,}392^{***}$	$37{,}634^{***}$	$45{,}079^{***}$	$45{,}253^{***}$
Note:	*p<0.05; **p<0.01; ***p<0.001, standard errors in parentheses				

to industry-affiliated researchers decrease this probability.

Model (4) reveals the impact of research performance related variables, indicating the average citation rank as well as cumulative citation numbers to increase the probability of transition, while the number of papers published decreases it. This may be an indication for industry to favour quality over quantity in terms of research output of transitioning scholars and thus provide further support for the 'cherry-picking' hypothesis.

Finally, when including all variables jointly in model (5), most observed effects remain roughly unchanged. The only exception here is the centrality in the co-authorship network, which remains significant yet changes the coefficient's direction from positive to negative.

5.2 Consequences of switching - Difference-in-Difference analysis

Finally, we attempt to investigate the consequences of university industry transition in terms of research performance. Therefore, Table 6 reports the results of a regression analysis, where we investigate the effect of university-industry transitions on scientific productivity, which we approximate by a researcher's annual citation rank.

We perform this analysis in a difference-in-difference setting, where we compare the development of scientific performance of researchers which undergo a university-industry transition (treated) with their counterparts remaining in academia. The dependent variable here is the researcher's annual citation rank ($citation^{rank}$) as a three year moving average.

Due to self-selection into an industry career are expected to be systematically different from their peers remaining in academia. We address this issue by performing a difference-in-difference analysis containing the following steps. First, we perform a nearest neighbor matching, where we match every researcher in the sample which at one point transits to industry with a peer which is only observed with academic affiliations. We match these pairs on their field of study, gender, mean number of papers published and citations received per year. We additionally require the matched pair to b observable for the exactly same number of periods.

Having done so, we attempt to empirically transform this observational study into a quasi-experimental econometric setting. In a difference-in-difference analysis, one usually matches an observations subject to an intervention (treatment) with a similar one which did not experience this intervention. However, our sample is not stratified and subject to left and right censoring, and furthermore the university-industry transition happens at different points in time and at different stages of their career for each researcher, we cannot define one intervention point across the sample. Rather, we create a 'pseudo-treatment' time for every researcher remaining in academia which is equal to the observation period in which their matched university-industry *switcher* transits (variable *transited*). We furthermore create a variable indicating the years since this transit takes place ($transited_t$). Otherwise, the models include a similar selection of independent and control variables as former ones.

Table 6 reports the results of this set of regressions. The first panel (1) includes only control variables plus the dichotomous variable indicating researcher that at one point undergo the university transition (*switcher*) and the periods after the transition has taken place (*transited*). In the next panel (2) we include further controls for the researchers overall ($cent^{dgr}$) and industry ($cent^{dgr-ind}$) centrality within the co-citation network. In the following panel (3), we turn our attention to the effect of university-industry transitions by including the number of periods since the researcher has transited to industry ($transited_t$) as well as interaction between *switcher* and the variables indicating the post transition period. This enables us to identify differences in $citation^{rank}$ between researchers after their transition has taken place, as compared to peers remaining in academia with otherwise similar characteristics, and thereby isolate the effect of university-industry transitions on research performance. The final panel (4) includes all variables jointly.

The main findings of this analysis can be seen in the interaction terms. Here, we can observe – and this is overall in line with the findings of the survival analysis – that transitioning scholars are higher positioned in terms of their citation rank. An

Table 6: Difference-in-Difference Regression: Effect of university-industry transition

	Dependent variable:			
	citation_{rank}			
	(1)	(2)	(3)	(4)
switcher	-0.016^{***}	-0.013^{***}	-0.016^{***}	-0.013^{***}
	(0.002)	(0.002)	(0.002)	(0.002)
transited	0.029^{***}	0.029^{***}	-0.0004	0.013^{***}
	(0.003)	(0.003)	(0.004)	(0.003)
seniority	-0.001^{**}	-0.006^{***}	-0.002^{***}	-0.006^{***}
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
gender	0.015^{***}	0.009^{***}	0.015^{***}	0.009^{***}
	(0.003)	(0.003)	(0.003)	(0.003)
cent^{dgr}		0.026^{***}		0.025^{***}
		(0.001)		(0.001)
$\text{cent}^{dgr-ind}$		0.047^{***}		0.047^{***}
		(0.001)		(0.001)
transited_t			0.008^{***}	0.004^{***}
			(0.001)	(0.001)
switcher$*$transited	0.050^{***}	0.053^{***}	0.080^{***}	0.079^{***}
	(0.004)	(0.003)	(0.005)	(0.005)
switcher*transited_t			-0.007^{***}	-0.006^{***}
			(0.001)	(0.001)
Study Field Control	Yes	Yes	Yes	Yes
Year Control	Yes	Yes	Yes	Yes
N	83,364	83,364	83,364	83,364
R^2	0.223	0.387	0.224	0.387
$\bar{R}^2$	0.222	0.386	0.224	0.387
Note:	*p<0.05; **p<0.01; ***p<0.001, standard errors in parentheses			

interesting effect can be observed when looking at the dynamic effect on the citation rank of the transition. Here scholars' citation rank appears to decline over time, indicating that on average those switching into industry end up having less academic impact over time. In future versions of this paper we will have a more thorough look at this effect as well as at further outcomes of transitions including thematic development, diversity and variation on co-authorship constellations.

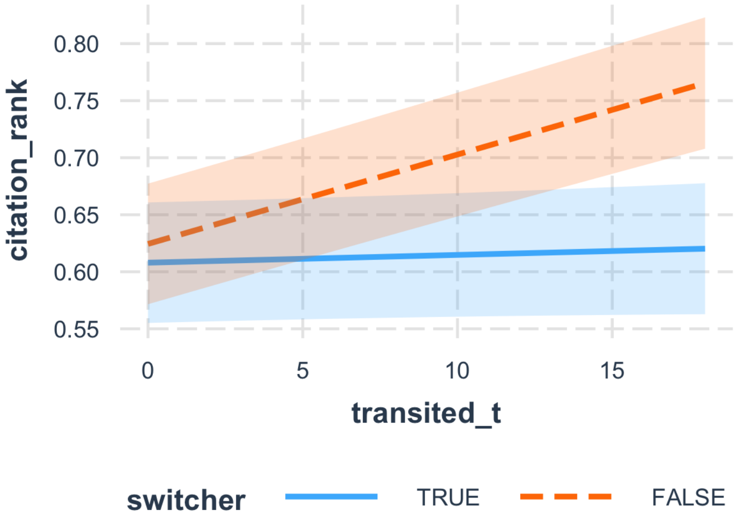

Figure 8: Interaction plot

6 Discussion and Conclusion

Studying career paths of AI researchers, we shed light on the interplay between academic and corporate research in this field, investigating potential public brain-drain. The primary aim of this research is to inform science policy discussions around the development and application of AI technologies. Our findings show that AI research within private organizations has more impact as measured by citations, confirming our argumentation.

The survival analysis shows that researchers working within the field of deep learning have a much higher likelihood to transition to industry confirming the initially presented evidence. Scholars producing lower numbers of papers with higher impact

– which we interpret as prioritising quality rather than quantity – are also more likely to transition. Interestingly, stronger embedding in the research community seems to marginally reduce the transition probability when paper impact is accounted for.

On average researches working in industry receive twice the amount of citations as compared to scholars in academia, while publishing less. In an industry with a rather short distance from research to deployment in products and services, it is not surprising that industry players are better at selecting and funding promising research, and promoting its relevance. However, looking at the results from the diff-in-diff analysis we also see some indication that over time researchers switching into industry stagnate in terms of academic impact. This may suggest that they fare similarly to start-ups that get acquired and absorbed by large companies with their technological agendas that may be more interested in implementation and exploitation of existing technology rather than exploration of entirely new trajectories. Looking into the recent developments in NLP one could argue that this is not the case – the majority of breakthrough developments (i.e. large scale models) came out of industrial labs in the recent years. On the other hand – and that brings us to the story that we mentioned in the introduction – one could argue that these models are in line with interest of large companies while leaner approaches in state-of-the-art language processing have not been pursued.

Overall the results suggest that the anecdotal evidence of industry hiring leading AI researchers away from academic institution can be supported by bibliographic data.

Future versions of this paper will look into further effects of researcher transitions into industry, examining for instance potential thematic change, diversity of themes as well as co-authors.

While strong contributions to research from private companies are commendable, it is key to understand where complementary public investments in R&D can contribute to favourable long-term outcomes. More specifically it is important to make sure that public research organization remain an attractive workplace for talented AI researchers. This requires investments in equipment, research funding as well as well coordinated frameworks that allow these scholars to contribute to the development of this technology and promote their contributions. AI is a strong contender for being a general purpose technology and therefore a lot of the development and application is conditioned by coordination between different stakeholders across disciplines. In practise that means that it may be unlikely for public research institutes to compete with industry in the development of framework and platforms (Tensorflow, PyTorch co. are here to stay). However, public interest related to e.g. fairness, security and accessibility has to be made explicit within legal guidelines and enforced. Here it is key to support interdisciplinary research looking into AI application to understand where commercial AI may need better alignment with societal goals. Industrial AI labs today often have divisions looking into AI ethics, yet the case of Timnit Gebru is an interesting example of a clear clash between an honest academic interest in the matter and the commercial perspective of executives.

References

Acemoglu, D. and Restrepo, P. (2019). The wrong kind of ai? artificial intelligence and the future of labor demand. Technical report, National Bureau of Economic Research.

Ahmed, N. and Wahed, M. (2020). The de-democratization of ai: Deep learning and the compute divide in artificial intelligence research.

Archibugi, D. and Filippetti, A. (2018). The retreat of public research and its adverse consequences on innovation. *Technological Forecasting and Social Change*, 127:97–111.

Arora, A., Belenzon, S., Patacconi, A., and Suh, J. (2020). The changing structure of american innovation: Some cautionary remarks for economic growth. *Innovation Policy and the Economy*, 20(1):39–93.

Arthur, B. (2017). Where is technology taking the economy? *McKinsey Quarterly*, (October):1–12.

Berg, P. (2008). Asilomar 1975: Dna modification secured. *Nature*, 455(7211):290–291.

Clark, J. and Hadfield, G. K. (2019). Regulatory markets for ai safety. *arXiv preprint arXiv:2001.00078*.

D'Amour, A., Heller, K., Moldovan, D., Adlam, B., Alipanahi, B., Beutel, A., Chen, C., Deaton, J., Eisenstein, J., Hoffman, M. D., et al. (2020). Underspecification presents challenges for credibility in modern machine learning. *arXiv preprint arXiv:2011.03395*.

David, P. A. (2003). 8 innovation and europe's academic institutions–second thoughts about embracing the bayh–dole regime. *This page intentionally left blank*, page 251.

David, P. A. and Hall, B. H. (2006). Property and the pursuit of knowledge: Ipr issues affecting scientific research. *Research Policy*, 35(6).

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Einav, L. and Levin, J. (2014). Economics in the age of big data. *Science*, 346(6210):1243089.

Gofman, M. and Jin, Z. (2019). Artificial intelligence, human capital, and innovation. *Human Capital, and Innovation (August 20, 2019)*.

Goldfarb, A., Gans, J., and Agrawal, A. (2019). *The Economics of Artificial Intelligence: An Agenda*. University of Chicago Press.

Gustafsson, R. and Autio, E. (2011). A failure trichotomy in knowledge exploration and exploitation. *Research Policy*, 40(6):819–831.

Hain, D. S. and Jurowetzki, R. (2017). Incremental by design? on the role of incumbents in technology niches. In *Foundations of Economic Change*, pages 299–332. Springer.

Hajian, S., Bonchi, F., and Castillo, C. (2016). Algorithmic bias: From discrimination discovery to fairness-aware data mining. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 2125–2126, New York, NY, USA. ACM.

Hao, K. (2020a). The messy, secretive reality behind openai's bid to save the world. *MIT Technology Review*, March/April.

Hao, K. (2020b). "I started crying": Inside Timnit Gebru's last days at Google.

Klinger, J., Mateos-Garcia, J., and Stathoulopoulos, K. (2018). Deep learning, deep change? mapping the development of the artificial intelligence general purpose technology. *arXiv preprint arXiv:1808.06355*.

Klinger, J., Mateos-Garcia, J. C., and Stathoulopoulos, K. (2020). A narrowing of ai research? *SSRN Electronic Journal*.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436.

Marcus, G. (2018). Deep learning: A critical appraisal. *arXiv preprint arXiv:1801.00631*.

Paullada, A., Raji, I. D., Bender, E. M., Denton, E., and Hanna, A. (2020). Data and its (dis)contents: A survey of dataset development and use in machine learning research.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., and Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI Blog*, 1(8):9.

Roach, M. and Sauermann, H. (2010). A taste for science? phd scientists' academic orientation and self-selection into research careers in industry. *Research policy*, 39(3):422–434.

Russell, S. (2019). *Human compatible: Artificial intelligence and the problem of control*. Penguin.

Sample, I. (2017). 'We can't compete': why universities are losing their best AI scientists.

Shen, Z., Ma, H., and Wang, K. (2018). A web-scale system for scientific knowledge exploration. *arXiv preprint arXiv:1805.12216*.

Strubell, E., Ganesh, A., and McCallum, A. (2019). Energy and policy considerations for deep learning in nlp. *arXiv preprint arXiv:1906.02243*.

Sweeney, L. (2013). Discrimination in online ad delivery. *Queue*, 11(3):10:10–10:29.

Trajtenberg, M. (2018). Ai as the next gpt: a political-economy perspective. Technical report, National Bureau of Economic Research.

Wang, K., Shen, Z., Huang, C.-Y., Wu, C.-H., Eide, D., Dong, Y., Qian, J., Kanakia, A., Chen, A., and Rogahn, R. (2019). A review of microsoft academic services for science of science studies. *Frontiers in Big Data*, 2:45.

Zou, J. and Schiebinger, L. (2018). Ai can be sexist and racist—it's time to make it fair.