AI & SOCIETY
https://doi.org/10.1007/s00146-021-01284-z

ORIGINAL ARTICLE

Ethical considerations and statistical analysis of industry involvement in machine learning research

Thilo Hagendorff[1] · Kristof Meding[2]

Received: 21 June 2021 / Accepted: 7 September 2021

Abstract
Industry involvement in the machine learning (ML) community seems to be increasing. However, the quantitative scale and ethical implications of this influence are rather unknown. For this purpose, we have not only carried out an informed ethical analysis of the field, but have inspected all papers of the main ML conferences NeurIPS, CVPR, and ICML of the last 5 years—almost 11,000 papers in total. Our statistical approach focuses on conflicts of interest, innovation, and gender equality. We have obtained four main findings. (1) Academic–corporate collaborations are growing in numbers. At the same time, we found that conflicts of interest are rarely disclosed. (2) Industry papers amply mention terms that relate to particular trending machine learning topics earlier than academia does. (3) Industry papers are not lagging behind academic papers with regard to how often they mention keywords that are proxies for social impact considerations. (4) Finally, we demonstrate that industry papers fall short of their academic counterparts with respect to the ratio of gender diversity. We believe that this work is a starting point for an informed debate within and outside of the ML community.

Keywords Machine learning research · Industry influence · Conflict of interest · Gender equality · Public–private partnership

1 Introduction

The number of papers submitted and accepted at the major machine learning (ML) conferences is growing rapidly. Besides submissions from academia, big tech companies like Amazon, Apple, Google, and Microsoft submit a large number of papers. But the influence of these companies on science is unclear. Do they drive trends? What are potential upsides and downsides of industry involvement in ML research? What are the possible ramifications of conflicts of interest? To investigate these topics, namely the industry involvement in ML research and its associated ramifications that range from questions about conflicts of interests, to scientific progress, research agendas, and gender balance, we conducted a statistical data analysis of the field.

Our analysis serves to answer four overarching research questions. First of all, we will develop a quantitative analysis of the proportion of industry, academic, and academic—corporate collaboration papers within the three major ML conferences (from 2015 to 2019), namely the Conference and Workshop on Neural Information Processing Systems (NeurIPS), the International Conference on Machine Learning (ICML), and the Conference on Computer Vision and Pattern Recognition (CVPR). Secondly, we aim to find out whether conflicts of interest are disclosed in those cases in which they are pertinent. Answering these questions will be of importance to assess potential changes in conference policies on transparency statements and to inform discourses on "AI governance" (Daly et al. 2019). Thirdly, we are interested in the role industry papers play with regard to scientific progress and ethical concerns, as well as whether they are, in this respect, any different from academic research. Finally, we investigate gender balances, particularly with regard to the proportions of women working on industry papers. We also discuss our findings in light of recent ethical research and implications for the ML community (Mittelstadt 2019).

✉ Thilo Hagendorff
thilo.hagendorff@uni-tuebingen.de

Kristof Meding
kristof.meding@uni-tuebingen.de

[1] University of Tübingen, Cluster of Excellence "Machine Learning - New Perspectives for Science", Tübingen, Germany

[2] University of Tübingen, Neural Information Processing Group, Tübingen, Germany

Published online: 29 September 2021

In the following paragraphs, we give a theoretical introduction to the mentioned issues and discuss their ethical implications.

2 The ethics of industry funding and conflicts of interest

A concern connected to industry funding is that research agendas are skewed. More applied topics and short-term benefits are favoured over basic science and its potential long-term outcomes (Savage 2017). Moreover, industry funding may affect the very questions researchers choose to tackle. This causes research strands to strongly orient towards corporate interests (Washburn 2008), or, more severely, to the plain distortion or suppression of certain research results to produce favourable outcomes for the respective sponsor. This is called "industry bias" (Lundh et al. 2017; Probst et al. 2016; Krimsky 2013). This bias can occur due to payments for services, the commodification of intellectual property rights, research funding, job offers, startups or companies owned by scientists, consultation opportunities, and the like. Especially criticised is the fact that machine learning conferences are hardly ever free of industry sponsors (Abdalla and Abdalla 2020). These sponsors may in some cases be able to control a conference's agenda to a certain extent.

To delve further into the subject, we will examine conflicts of interest which are a common side effect of industry involvement in academic research in general (Etzkowitz and Leydesdorff 2000; D'Este and Patel 2007; Boardman 2009; Bruneel et al. 2010). A substantial amount of literature is dedicated to reflecting on conflicts of interest that can occur in clinical practice, education, or research (Rodwin 1993; Fickweiler et al. 2017; Thompson 1993). As a consequence of conflicts of interest in research, medical journals require researchers to name funding sources. The public disclosure of funding sources, affiliations, memberships, etc. are supposed to inform those who receive scientific information or advice to fill information gaps. This allows them to assess the information or advice to its full extent.

But what exactly are conflicts of interest? While it is hard to find a universal definition, a common denominator is that conflicts of interest arise when personal interests interfere with requirements of institutional roles or professional responsibilities (Komesaroff et al. 2019). Here, interests can be seen as goals that are aligned with certain financial or non-financial values that have a particular, maybe detrimental effect on decision-making. The coexistence of conflicting interests results in an incompatibility of two or more lines of actions. In modern research settings, dynamic and complex constellations of conflicting interests frequently occur (Komesaroff et al. 2019). For instance, conflicts of interest do not only pose a problem in cases where researchers intentionally follow particular interests that undermine others. Many effects arising from conflicts of interest take effect on subconscious levels (Cain and Detsky 2008; Moore and Loewenstein 2004; Dana and Loewenstein 2003), where actions are rationalized by post hoc explanations (Haidt 2001). Many studies, especially in the field of medical research, show that even when physicians report that they are not biased by financial incentives, they actually are (Orlowski and Wateska 1992; Avorn et al. 1982). This means that despite researchers' belief in their own integrity and the idea that financial opportunities, honorariums, grants, awards, or gifts have no influence in their line of action, opinion, or advice, the influence is, in fact, measurable.

Psychological research has shown that individuals often succumb to various biases that steer their behaviour (Chavalarias and Ioannidis 2010; Ioannidis 2005; Kahneman 2012; Tversky and Kahneman 1974). These are so-called "self-serving biases", meaning that fairness criteria, assumptions about the susceptibility towards conflicts of interest, or other ways of evaluating issues are skewed towards one's own favour (McKinney 1990). One famous self-serving bias is exemplified by the fact that physicians assume that small gifts do not significantly influence their behaviour, while actually, the opposite is true (Brennan et al. 2006). Even small favours elicit the reciprocity principle, meaning that there is a clear influence or bias on an individual's behaviour. These biases are not necessarily associated with lacking moral integrity or even corruptibility. On the contrary, they can be assigned to an "ecological rationality", meaning that an individual's behaviour is adapted to environmental structures and certain cognitive strategies (Arkes et al. 2016; Gigerenzer and Selten 2001). Nevertheless, conflicts of interest can have or actually do have dysfunctional effects on the scientific process. Hence, the scientific community does well in finding a way to deal with them properly. This is mostly done by obliging researchers to disclose conflicts of interest. While this is an accepted method in many scientific fields, it can actually have negative effects. These so-called "perverse effects" are described by Cain et al. (2005) and Crawford and Sobel (1982). On the one hand, Cain and colleagues demonstrate that disclosing conflicts of interest does not lead people to relativize claims by biased experts sufficiently since disclosure can in some cases increase rather than decrease trust. On the other hand, and more importantly, experts who reveal conflicting interests may thus feel free to exaggerate their advice and claims since they have lowered their guilty conscience about spreading misleading or biased information. While transparency statements have side effects, they should certainly not be omitted entirely.

Research on conflicts of interest shows the many facets they possess. Hence, as stated above, it is difficult to come up with a concise definition. However, to stipulate what we mean when using the term "conflicts of interest" throughout the paper, we want to define it as an interference between personal or financial interests and the requirements of professional responsibilities that emerge due to holding a position in both academia and industry.

3 Setting trends

Despite the manifold pitfalls that are caused by the intermingling of academia and industry, studies show that particularly corporate-sponsored research can be very valuable for science itself as well as for society as a whole (Wright et al. 2014). Hence, one has to discuss another concomitant of industry involvement in research, namely industry's potential innovative strength. Industry involvement in the sciences can not only provide more jobs, lead to tangible applications of scientific insights, provide life-enhancing products, increase a society's wealth, but also lead to much-cited papers, and spur trends. Researchers (Wright et al. 2014) have shown that corporate-sponsored inventions resulted in licenses and patents more frequently than federally sponsored ones—although this alone does not mean that industry is more innovative per se. Current research also shows that machine learning research in the private sector tends to be less diverse topic-wise than research in academia (Klinger et al. 2020). Furthermore, corporations are often seeking university partners to widen their portfolio of products, business models, and profit opportunities. This can nudge academic partners to act progressively, towards novel, unprecedented experiments, research ideas, and speculative approaches (Evans 2010). Indirectly, industry funds lead to scientific progress. Research on innovation processes has shown that organizations are typically not innovating internally, but in networks, in social relationships between members of different organizations, in technology transfer offices, science parks, and many other university–industry collaborations (Perkmann and Walsh 2007). These collaborations can emerge via research papers, conferences, meetings, informal information exchange, consulting, contract research, hired graduates, a joint work on patents or licences, etc., and play a vital role in driving innovation processes (Cohen et al. 2002). All in all, scientists' sensitivity towards opportunities of industry funding may cause "deformed" research agenda settings. This does not necessarily mean, though, that trends, innovations, scientific progress, and their positive effects on society are diminished. With our data analysis, we aspire to find out how this constellation is reflected in the field of ML research.

Academic engagement, i.e. the involvement of researchers in university–industry knowledge transfer processes of all kinds, is a common by-product of academic success. Scientists who are well established, more senior, have more social capital, more publications, and more government grants, are at the same time more likely to have industrial collaborators (Perkmann et al. 2013). This is due to the "Matthew effect", meaning that researchers who are already successful in their field of research are more likely to reinforce this success with industry engagements whose returns continuously lead to more academic success. Researchers involved in commercialization activities publish more papers in comparison to their non-patenting colleagues (Fabrizio and Di Minin 2008; Breschi et al. 2007), whereas the economic value of patents can also be used to predict a firm's success in general (Xu et al. 2021). Scientific success in ML research seems to go hand in hand with industry collaborations. However, industry-driven research or research that is intended to be commodified is, in most cases, more secretive and less accessible for the public.

Taking all these considerations into account, a further objective of our data analysis is to scrutinize the innovative strength of industry research. For that purpose, we will not only conduct a citation analysis, but also look at three successful machine learning methods and measure whether industry or academic papers mentioned—and therefore most probably pushed—these methods before they became a commonly used standard tool for machine learning practitioners. Lastly, we analysed proxies of social impact awareness, measured by "social impact terms" such as privacy, fairness, accountability, and security, in industry research and also compare it to academic research.

4 Statistics on gender imbalances

The final issue we are going to investigate is that of gender aspects and their entanglement with industry research. Noticeably, male academics are significantly more likely to have industry partners than female scientists (Perkmann et al. 2013). This finding corresponds to the fact that ML research has a diversity imbalance, indicating that male researchers strikingly outnumber females. Statistics show that only a small share of authors at major conferences are women. The same holds true for the proportion of ML professors, the affliction of tech companies with heavy gender imbalances, women's tendency to leave the technology sector, as well as the fact that they are paid less than men (Myers et al. 2019; Simonite 2018). Further diversity dimensions such as ethnicity, intersexuality, and many other minorities or marginalized groups are often not statistically documented. Tech companies have even thwarted access to diversity figures to attempt to silence employees to highlight the under-representation of women and minorities (Pepitone 2013). All in all, the "gender problem" of the ML sector does not only manifest itself on the level of lacking workforce diversity, but in the functionality of software applications too (Leavy 2018). Despite these rather general observations and statistics, we want to find out whether gender imbalances have a particularly pronounced manifestation in the context of industry ML research. Inspired by previous research on gender imbalances (Andersen et al. 2019), we scrutinize the ratio of female (last) authors in academia and industry papers. This is of importance to prove or disprove common intuitions about the disadvantage against women, which is actually stronger in companies compared to university contexts as our own data will show.

5 Methods

5.1 Analysing 10,772 ML papers

At this point, we will briefly describe the methods we have used to conduct our statistical analysis. More detailed information about the process can be found in the supplementary material. All in all, our analysis focuses on three major ML conferences: ICML, CVPR, and NeurIPS. We downloaded all articles available spanning the years 2015 to 2019 from the respective conference proceedings. Altogether, the data set contains 10,807 papers. The papers were downloaded using the python-tool Beautiful Soup (v. 4.8.2). We extracted the text with pdftotext (v. 0.62.0) and analysed the text with a self-written script. Using this method, we were able to search 10,772 papers (99.7%). Some of the papers are, for example, not searchable because their text is embedded as an image. We are not only analysing the papers themselves, we are also interested in the metadata, namely affiliations and authors. For the analysis of the affiliations, we extracted them from the texts where possible and categorized them into academic and industry affiliations. We ascertained the affiliations by automatically looking at the headers of the papers. This was no problem for NeurIPS or CVPR papers. For these papers, we simply extracted all content before the word "abstract". In most cases, there were no issues. Very rarely, a figure appeared before the abstract or authors changed the standard template. The same procedure worked for ICML 2015 and 2016. However, from 2017 onwards, the affiliations were shown in the lower left corner. No keywords were placed before, only a blank line. This was difficult to parse with our script. We thus decided to keep the first 5000 characters as header for these papers, but split it before the terms "international conference of machine learning", which always ended the listing of authors. We think that this yields only a small amount of false positives if we search for affiliations, since it is most likely that academic and industry institutional terms will appear in the affiliations only.

To get an impression of which institutions publish on NeurIPS, CVPR, and ICML, we followed preexisting analyses:

- https://www.microsoft.com/en-us/research/project/academic/articles/neurips-conference-analytics/
- https://www.reddit.com/r/MachineLearning/comments/bn82ze/n_icml_2019_accepted_paper_stats/
- https://medium.com/@dcharrezt/neurips-2019-stats-c91346d31c8f
- https://www.microsoft.com/en-us/research/project/academic/articles/eccv-conference-analytics/

To prevent us from cherry-picking, we only used terms that appeared in the analyses above.

We performed a non-exclusive classification. Papers may have academic and industry affiliations. It is important to note that we included blanks before and after the text for the UC, UT, MILA, MIT, NEC, and Intel terms to avoid contamination with other words like "admit".

Moreover, we define a paper as academic if it contains one of the following terms (see supplementary material for more information on why we use these terms only):

AMII/California Institute of Technology/College/Ecole/EPFL/ETH Z/Georgia Institute/IIT Bombay/INRIA/Kaist/Massachusetts Institute of Technology/MILA/MIT/MPI/ParisTech/Planck/RIKEN/Technicon/Toyota Technological/TU Darmstadt/UC/Universi/UT Austin/Vector

For the definition of a paper as industry, we use the following terms:

Adobe/AITRICS/Alibaba/Amazon/Ant Financial/Apple/Bell Labs/Bosch/Criteo/Data61/DeepMind/Expedia/Facebook/Google/Huawei/IBM/Intel/Kwai/Megvii/Microsoft/NEC/Netflix/NTT/Nvidia. /OpenAI/Petuum/Qualcomm/Salesforce/Sensetime/Siemens/Tencent/Toyota Centrl/Toyota Research/Trace/Sensetime/Uber/Xerox/Yahoo/Yandex.

Unless otherwise stated, we define a paper as academic if it does not contain an industry term in the affiliation section and a paper as belonging to industry if it does not contain an academic term. A paper is defined as mixed if it contains an academic and an industry affiliation. In total, 90.2% of all papers contain at least one of the terms from academia or industry listed above. These numbers are entirely dependent on the fact that the authors actually declare all their affiliations in the paper. We show that our automatic approach gives sufficient results in Sect. 5.2.

Furthermore, we extracted the authors' names and the titles of the papers directly from the websites, not from pdf documents. For this purpose, we once again used Beautiful Soup. We extracted 41,939 authors. However, many authors have multiple accepted papers, and thus, the number of authors is reduced to 18,060 unique authors by pooling all authors with the same name. Of course, this leads to the effect that different authors with the same name are pooled. We believe that this effect is negligible. For authors with middle names, we kept only the first letter. People vary the ways in which they indicate their middle name, e.g. T., T, or

Tom. All information in text and graphics about the number of authors refers to these unique authors. The genders of the names were determined using the name-to-gender service GenderAPI. GenderAPI offers the highest accuracy of the name-to-gender tools (Santamaría and Mihaljević 2018) and was able to determine the gender of 17,412 authors (96.4%). GenderAPI also provides an estimate of the accuracy. The mean accuracy in our case was 87.1%. Unfortunately, we noticed that most times, GenderAPI fails in the recognition of names from Asian language families. This is a clear bias in the underlying dataset of GenderAPI. Furthermore, we want to acknowledge that some people reject the idea that a name corresponds to a gender. However, we applied the analysis of genders here to gain insight into the inequality of authors' genders on average, not only in single cases. Finally, we downloaded the citations received for each individual. We wrote an automated script to access the Microsoft Academic Knowledge API (Sinha et al. 2015). This was successful for 10,616 papers (98.2%, date of citation download: 03.29.2020). The most common reason for a paper not being found in the database is the use of special characters like λ, etc. in the title.

Further, we extracted the acknowledgements for our conflict of interest analysis. In this particular analysis, we focused on academic papers. In our data sets, we have 6802 papers from academia. Of these papers, 5373 papers (79.0%) contain an acknowledgement section which we were able to parse. We also included both spellings of acknowledgement: "acknowledgement" and "acknowledgment".

Our approach has three (possible) limitations. Firstly, our results should be understood as general and robust trends but not as exact numbers, since it is not possible to extract data from the papers in all cases. A further limitation of our method that is particularly relevant to our analysis of conflicts of interest is that we cannot detect cases where researchers have academic and industry affiliations at the same time but state only one of them in the respective research paper. Moreover, we would like to point out that the data set is smaller for the industry analysis (6802 vs. 731 papers). Small data sets tend to produce extreme results—in both positive and negative directions. Nevertheless, we believe that this is not a problem in our case as our results, as we will see in the following section, are very robust.

5.2 Error bars and statistical modelling

Here, we briefly describe the way we calculated error bars and explain our statistical model approach.

In most of the analysis, we extract ratios which follow a binomial distribution. Following this, we used the Wilson-score approximation of a binomial distribution for the confidence intervals of individual data points (Figs. 1b, 2b, 3b, c, 4a, b). For the calculation of confidence intervals of median citation counts a 1000-fold bootstrap approach was used. This approach is less influenced by the underlying data distribution compared to a parametric approach. These error bars represent the uncertainty of individual data points and should not be confused with the error of groups (academia, industry, etc.)

These confidence intervals are not suitable for the comparison of different data points in our figures, in this case, academic, industry, and mixed papers. For the comparison of

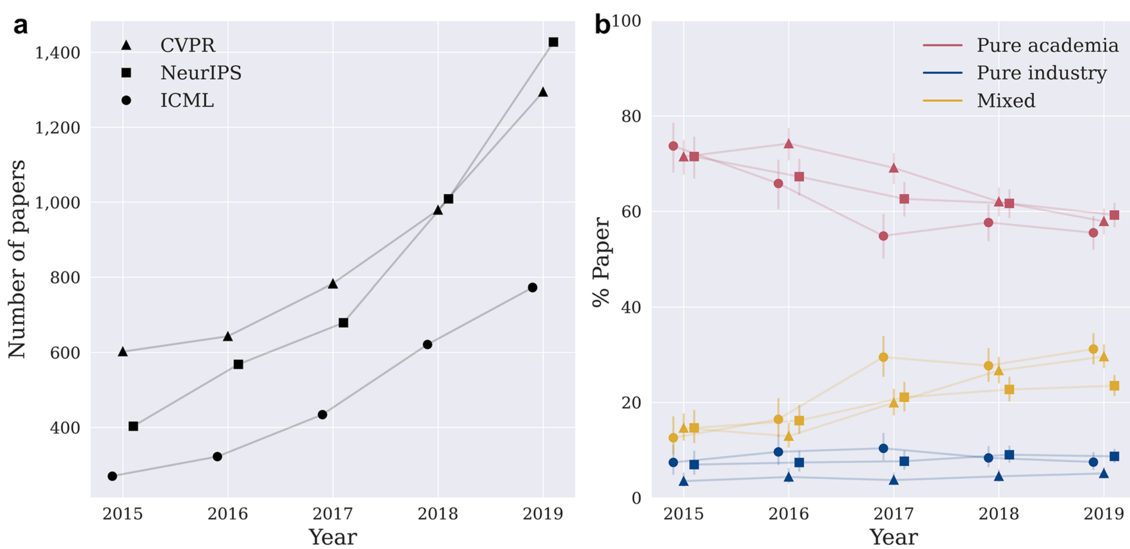

Fig. 1 Progression of the number of papers at major ML conferences (**a**) and (**b**) institutional affiliations. Please note the numbers in (**b**) do not add up to 100% because we were only able to extract this information for 90% of the papers, see methods and supplementary information. Error bars in this example and all following figures illustrate a 95% confidence interval

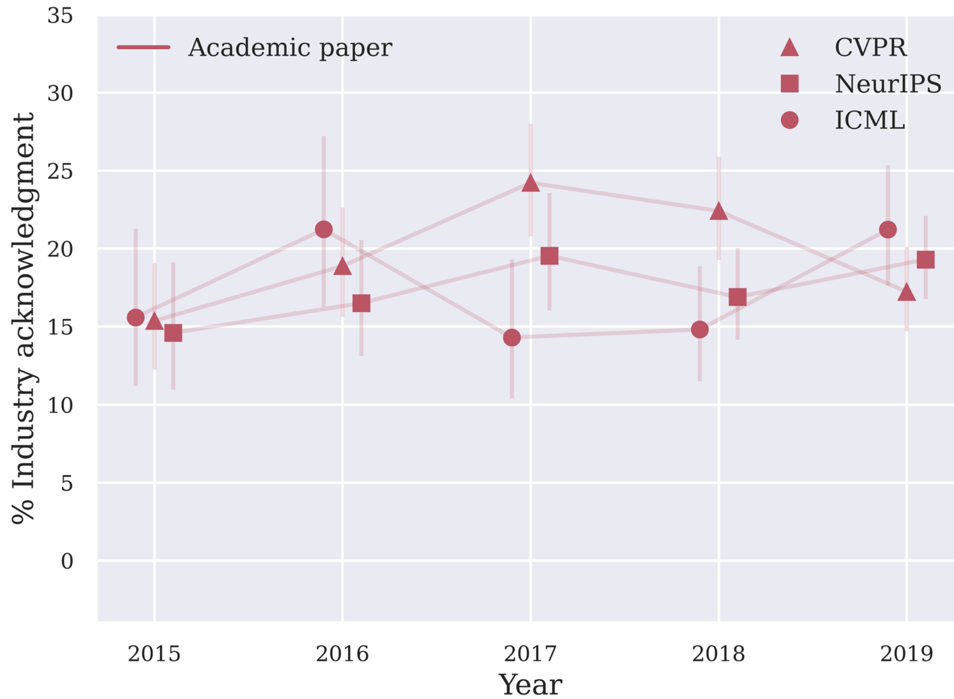

Fig. 2 Percentage of paper from academia which contain an industry acknowledgement

academia and industry as well as trends in time, a (generalized) linear model approach was used (Faraway 2014). We used the metafor-package (v. 2.4-0) in R for all figures except for Fig. 3a. The metafor-package allows us to include the uncertainty of data points. In Fig. 3a we used the negative binomial fitting procedure *glm.nb* from the MASS-package (v. 7.3-51.6), since we are able to obtain an individual citation count to every individual paper.

6 Results

6.1 Subtle conflicts of interest in academia

Figure 1a plots the number of papers accepted at ICML, NeurIPS, and CVPR between 2015 -and 2019. The number of accepted papers is steadily increasing. Figure 1b shows whether the paper includes authors with affiliations from academia, industry, or both. While the ratio of industry papers is stable, an increasing ratio of papers has affiliations from both academia and industry. We obtain with our linear analysis a slope of – 3.7%/year (95% CI: [−4.7, −2.7], $p < 10^{-5}$) for academia, 0.3%/year (95% CI: [−0.03,0.61], $p = 0.07428$) for industry papers and 3.8%/year (95% CI: [2.8,4.9], $p < 10^{-5}$) for mixed papers.

Furthermore, we extracted the acknowledgements of all papers from academia and searched them for terms of industry affiliations (Google, Facebook etc.). This gives us an insight into whether academic papers acknowledge industry funding, grants etc. In fact, we calculated the conditional probability p(industry acknowledgement | academia).

With recourse to the insights from Fig. 1b, there is no doubt that purely academic papers make up the largest part of submissions to all major ML conferences, not industry papers.

6.2 Percentage of paper from academia which contains an industry acknowledgement

However, Fig. 2 shows the ratio of purely academic papers with industry acknowledgements. Roughly 20% of purely academic papers contain an industry acknowledgement. No significant trend in time is found (0.7%/year, 95% CI: [−0.3,1.8] ,$p = 0.18293$). Finally, we also searched for the terms "conflict of interest"—the plural "conflicts of interest", which did not lead to a single finding—and "disclosure" to identify whether such influences are named. Only 3 of more than 10,000 papers contain an explicit conflict of interest statement at all. This inquiry shows that on the one hand, conflicts of interest are present in many academic research papers, while on the other hand, those conflicts are not clearly stated. This further indicates that it is sensible for ML conferences to demand researchers to add transparency statements to their submissions. Nevertheless, our quantitative analysis cannot result in a detailed in-depth analysis of concrete conflicting interests. It must disregard the subtle influences of past funding resources that lie outside of the period of investigation.

All in all, ties between the two social systems (Luhmann 1995)—university and industry—do seem to become tighter.

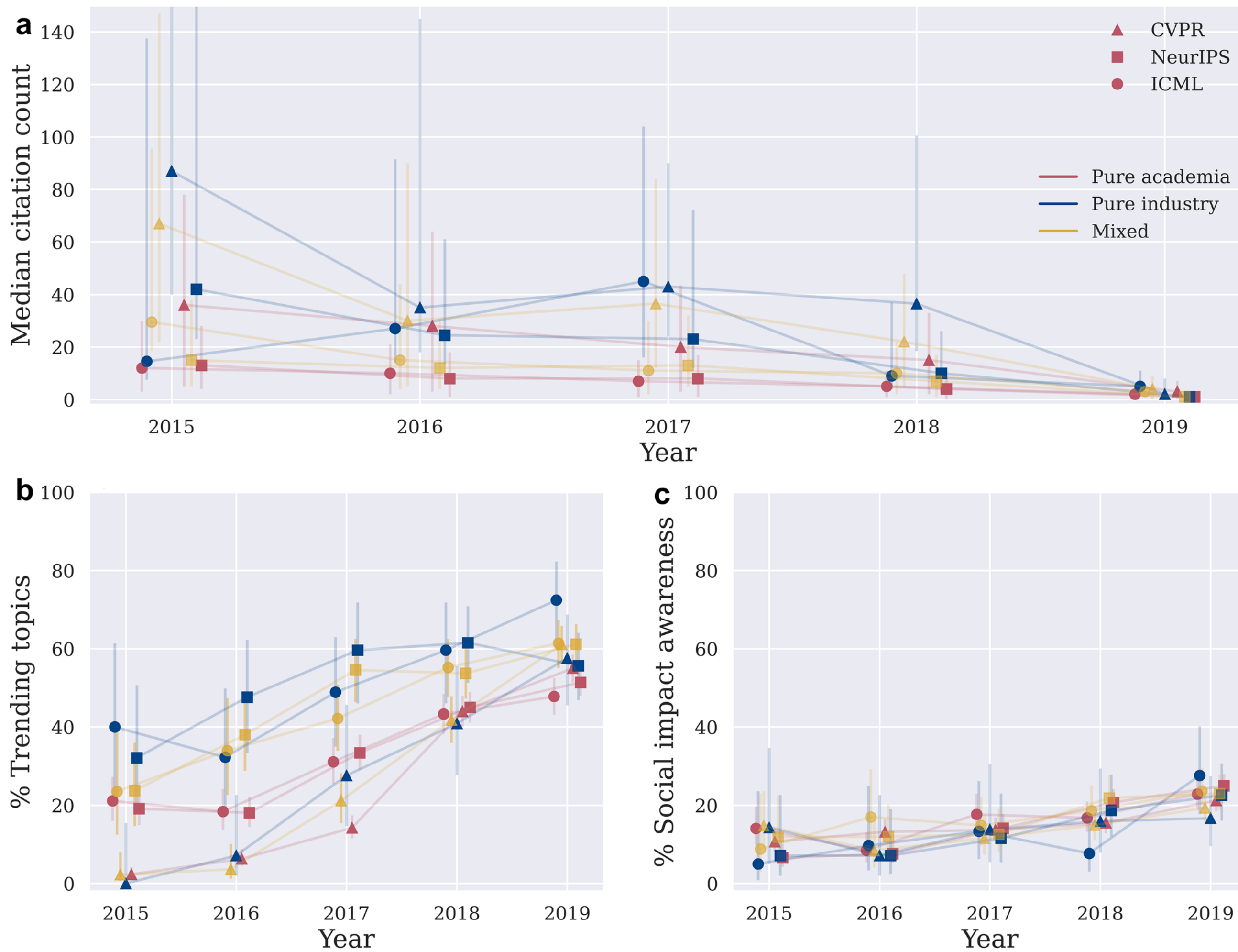

Fig. 3 (**a**) Median number of citations received. (**b**) Ratio of papers from academia and industry with trending topics: 'adversarial' and 'reinforcement' and (**c**) papers with social impact terms

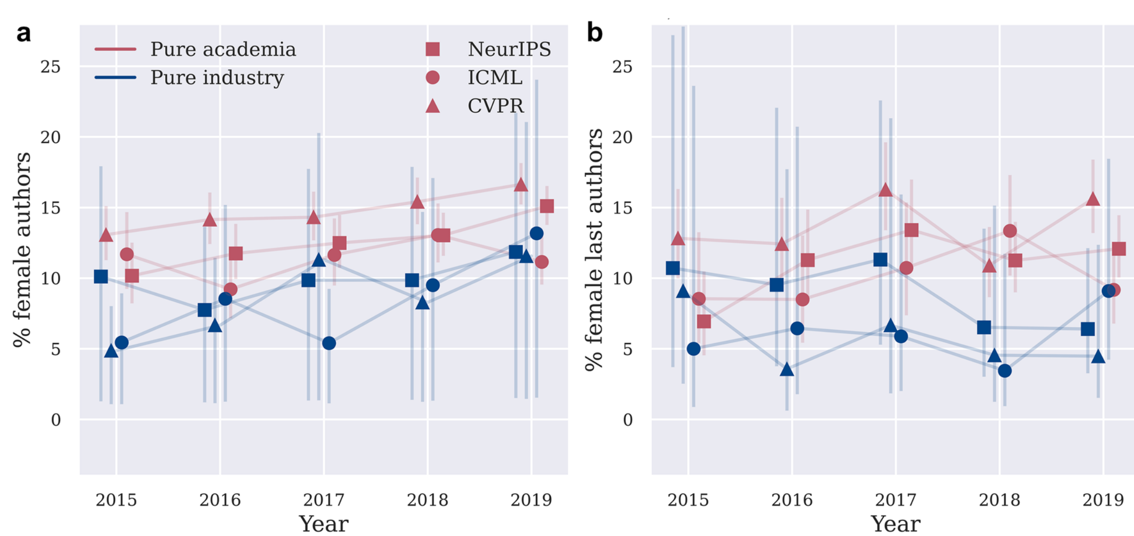

Fig. 4 (**a**) Overall ratio of female authors in academia and industry and (**b**) ratio of last authors

Academic settings are becoming increasingly intertwined with corporate tech environments. Moreover, academic papers with no industry affiliations are slightly on the decrease. This urgently calls for an appropriate approach to dealing with conflicts of interest. However, purely academic papers still make up the largest part of the submissions to all major conferences.

7 Publishing behaviour and impact

Next, we want to find out whether it is academia or industry that mentions machine learning methods in their research papers that, in hindsight, turned out to be very successful. Hence, we use a proxy to find out whether academia or industry is likely to propel important parts of the machine learning field. Thus, we compared industry and academic papers with regard to the average number of citations they possess. The results are shown in Fig. 3a. Our generalized linear model analysis shows that there is a significant difference between academia and industry $p < 2 \cdot 10^{-16}$). Due to the negative binomial model, the estimate coefficient (1.03) has no direct interpretation.

While citation analyses are not particularly credible for papers that were published quite recently, since citations are slow to accumulate, citation analyses gain significance over time. Thus, our analysis clearly shows that industry papers from 2015 were cited far more frequently than academic papers. This trend prevails throughout the following years, albeit on a smaller scale.

To gain further insights into whether it is academia or industry that mentions machine learning methods that, in hindsight, turned out to eminently important for the field, we searched for two terms, the first one being “adversarial”. This, on the one hand, corresponds to the very popular Generative Adversarial Networks invented in 2014 by Goodfellow et al. (2014) and, on the other, to the adversarial attack on neural networks (Szegedy et al. 2013). We also included the term “reinforcement” for reinforcement learning. These are topics of increasing interest to the ML field (Biggio and Roli 2018; Lipton and Steinhardt 2018). The results are shown in Fig. 3b. The linear model analysis shows a difference of 11.2% (95% CI: [5.1,17.2], $p = 0.00030$) between academia and industry. The figure indicates that, with regard to the three methods, academia is lagging roughly 2 years behind industry (ICML and NeurIPS) in terms of how often they mention generative adversarial networks, adversarial attacks, or reinforcement learning. Similar trends can be found for much more frequently used terms like “convolution” and “deep” in supplementary figure A1.

In addition, we are interested in whether social aspects are becoming more interesting to the ML community. However, with quantitative analysis, it is somewhat difficult to measure social impact awareness in academia and industry. To at least approximate social impact awareness and make it measurable as far as possible, we included terms from the social impact category of NeurIPS 2020 (safety, fairness, accountability, transparency, privacy, anonymity, security) and added “ethic” as well as “explainab”. We call these terms social impact terms. Figure 3c shows the results. Overall, we can see, that the ML community has paid more attention to these terms in the course of the past years (3.3%/year, 95% CI: [2.6,4.0], $p < 10^{-5}$). But while one may assume that academic papers put a stronger focus on social impact issues in comparison to industry research, this intuition does not hold true, at least when using the number of mentions of the listed social impact terms as a proxy for social impact issues. Only a small non-significant difference of 2.6% (95% CI: [−0.3,5.6], $p = 0.07670$) between academia and industry is found. The amount of social impact terms is more or less equally shared between academia and industry. A clear limitation of our analysis is that one could also include ethics-of-AI related conferences like AIES or FaccT to get a more complete picture of the ethical awareness of the machine learning community as a whole (Birhane et al. 2021). Our analysis only holds true for the landscape of the three selected conferences. Furthermore, one could object that the mere word count of a single term does not allow the conclusion on the ethical “superiority” or “inferiority” of organizations. However, we think that the overall trends of our analysis give a rough hint of the insignificant differences between academia and industry in terms of the intensity with which ethical issues are discussed. Discussing them, though, does not necessarily mean that, for instance, corporate actions are also actually aligned with the discussed issues.

Especially with regard to the results from Fig. 3a, we can show that industry papers have higher citation rates compared to academic papers every year, giving evidence for the high scientific relevance of industry papers. There is no question that industry papers receive greater attention from the scientific community than academic papers. A confound of this analysis is that one may assume that academic researchers, who are strikingly successful, are likely to be hired by ML companies, which then causes industry papers to have more citations on average than academic papers. Thus, it is difficult to state whether industry research has a more scientific impact because of the industry context itself or because of companies’ strategic hiring policies and the corresponding migration of successful university researchers to companies.

8 Gender equality: industry is behind academia

Finally, we analysed the contribution of male and female authors to ML conferences. We only focus on the difference between purely academic and purely industry papers since we are not able to assign the individual affiliations in mixed papers. Figure 4a shows the ratio of female to all authors across the conferences, indicating a slight increase in the ratio of female authors across the three major ML conferences.

The ratio of female authors increased by 0.9%/year (95% CI: [0.6,1.2], $p < 10^{-5}$). It is somewhat noticeable, though, that female authors are less represented in industry papers compared to academic papers. Our linear model analysis shows a difference of 3.9% (95% CI: [2.7,5.0], $p < 10^{-5}$) between academia and industry. Unfortunately, with the means of our quantitative analysis, we are not able to explain why the differences between academia and industry with regards to gender ratios occur. It may be due to direct discrimination against women in the industry. Another explanation that is perhaps more plausible is that companies are selecting researchers who already possess a certain amount of scientific success and reputation, which causes indirect discrimination of women due to the overrepresentation of successful male researchers.

Albeit not a universal practice, being the last author is a privileged position in the author list that typically corresponds to the principal investigator or the most senior author. Apart from that, papers may have multiple advising authors. Notwithstanding these exceptions, we assume that in most cases last authors are sole principal investigators. Taking up previous research (Andersen et al. 2019; Mohammad 2020) and going into further detail, Fig. 4b shows the ratio of female last authors compared to all last authors. No significant trend in time is found here (0.5%/year 95% CI: [−0.02,1], $p = 0.06145$), although our model analysis reveals that there is a difference of 5.6% (95% CI: [3.8,7.5], $p < 10^{-5}$) between industry and academic papers.

Taking up the results from our analysis, we see that female authors are less represented in industry papers than in academic papers. The results are in line with other studies, claiming that the proportion of women in ML research and in the number of workforces at major tech companies is typically hovering between 10 and 20 percent (Yuan and Sarazen 2020). A recent study by Mohammad (2020) that was dedicated to natural language processing research also looked at disparities in authorship and found that 29.2% of first authors and 25.5% of last authors are female. These numbers are a bit higher than ours (one has to consider that natural language processing research also contains disciplines like linguistics, psychology, and social science, and not just computer science and machine learning, though). The authors themselves state that the reported percentages for many other computer science sub-fields are significantly lower. Notwithstanding that, according to Mohammad (2020), the percentages have not changed during the last two decades, and papers with female first authors are cited less than male first authors, giving a clear sign of the enduring shortcomings in gender equality. In our dataset we can confirm this finding. Papers with a male first author get on average 12.5 more citations ($p = 4.9 \cdot 10^{-6}$) than papers with a female first author (24.6 vs. 37.1).

Despite our analysis that looks at male–female ratios, we are fully aware of the fact that gender equality is only one dimension in a broader spectrum of diversity (Hopkins 1997). It is obvious that other types of diversity, like ethnicity (asian, black, hispanic, white, other), nationality, age, etc. could also be analysed with respect to their differences in industry and university contexts. But since it is not possible to reliably yield this information from our data set, we refrained from analysing other types of diversity besides gender.

9 Conclusion

The scientific success of ML research lured an increasing amount of industry partners to coalescence with academia. The growing number of papers stemming from academic-corporate collaboration is an indication of this. Medical journals already require researchers to name conflicts of interest. The ML community is slowly following this demand and obliges researchers to add transparency statements to their work, at least at some conferences and journals. Further efforts to introduce transparency declarations are to be welcomed, while at the same time, a responsible interpretation of these declarations is required to ensure that disclosure brings about the intended effects (Loewenstein et al. 2012). This seems reasonable, especially against the backdrop of an increasing number of academic-corporate collaborations and academic papers with industry acknowledgements.

Up to now, though, only a handful of papers that were published in the proceedings of the analysed conferences voluntarily add conflicts of interest sections. On a related note, it is difficult to describe concrete ramifications on lines of action, opinions, or advice. In medical research, tangible and relatively direct influences from the pharmaceutical industry can be picked up. In ML research, industry influences are fuzzier and hard to monitor. Hence, the concrete consequences of existing conflicts of interest can only be discovered by more in-depth, qualitative empirical social research. One can assume that in ML research ramifications mostly affect research agendas so that scientists consciously or unconsciously steer their research in a direction that is most valuable for corporate interests or commercialization processes of all kinds. This bias can also potentially suppress certain research results to avoid unfavourable outcomes that are nonpractical to those interests or processes. After all, universities and companies follow different "symbolically generalized communication media" (i.e. money or truth, see (Luhmann 1995)), which can make it difficult for researchers with corporate cooperation to act in accordance with only one of those goals. In this context, it is important to keep in mind that even small gifts or favours elicit the reciprocity principle, meaning that individual behaviour is under the influence of an industry bias.

Despite the issue of conflicting interests, our data analysis provides evidence for the fact that industry-driven research has a measurable impact and is setting research trends. This insight stands in contrast to the rather industry-critical discourse on conflicts of interest and proves the positive impact industry-driven research has on scientific progress in ML. In line with this insight, we show that industry papers receive significantly more citations than research from academia, which is a clear sign that corporate ML research is of high importance for the scientific community. Besides the great attention that is directed towards industry papers, we demonstrate that these papers are not solely oriented towards technical issues and collected clues that they do not omit to discuss social aspects of technology. The amount of social impact terms that we used as a proxy to measure the significance of social aspects is more or less equally distributed between academic and industry papers.

Tangible problems, however, occur in view of diversity shortcomings. We show that the ratio of female authors compared to male authors of conference papers indicates a slight improvement in gender equality over time. But overall, the proportion of women in ML research is quite small. This holds especially true with respect to industry research. Here, amendments are necessary, mainly comprising the creation of more inclusive workplaces, changes in hiring practices, but also an end of pay and opportunity inequalities (Crawford et al. 2019). In contrast to issues like innovative strength or citations, industry has a lot of catching up to do here.

In summary, we provide quantitative evidence for the increasing influence tech companies have on ML research. Our analysis reveals three main insights that can inform and differentiate future policies and principles of research ethics. Firstly, the analysis shows that besides the growing number of academic-corporate collaborations, conflicts of interest are not disclosed sufficiently. Secondly, it proves that industry-led papers are not only a strong driving force for promising scientific methods, but possess significantly more citations than academic papers. Thirdly, we provide further evidence for the need to improve gender balance in ML research, especially in industry contexts.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00146-021-01284-z.

Acknowledgements We would like to thank Felix Wichmann, Ulrike von Luxburg, Sarah Fabi, and Cornelius Schröder for helpful comments on the manuscript and discussions. Additionally, we thank our two anonymous reviewers for constructive feedback.

Funding Thilo Hagendorff was supported by the Cluster of Excellence "Machine Learning – New Perspectives for Science" funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – Reference Number EXC 2064/1 – Project ID 390727645. Kristof Meding was supported by the German Research Foundation (DFG): SFB 1233, Robust Vision: Inference Principles and Neural Mechanisms, TP 3, Project ID 276693517. Open Access funding enabled and organized by Projekt DEAL.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Abdalla M, Abdalla M (2020) The Grey Hoodie Project: Big Tobacco, Big Tech, and the Threat on Academic Integrity. arXiv 1–9

Andersen JP, Schneider JW, Jagsi R, Nielsen MW (2019) Gender variations in citation distributions in medicine are very small and due to self-citation and journal prestige. Elife 8:1–17

Arkes HR, Gigerenzer G, Hertwig R (2016) How bad is incoherence? Decision 3(1):20–39

Avorn J, Chen M, Hartley R (1982) Scientific versus commercial sources of influence on the prescribing behavior of physicians. Am J Med 73(1):4–8

Biggio B, Roli F (2018) Wild patterns: ten years after the rise of adversarial machine learning. Pattern Recognit 84:317–331

Birhane A, Kalluri P, Card D, Agnew W, Dotan R, Bao M (2021) The values encoded in machine learning research. arXiv 1–28

Boardman PC (2009) Government centrality to university-industry interactions: university research centers and the industry involvement of academic researchers. Res Policy 38(10):1505–1516

Brennan TA, Rothman DJ, Blank L, Blumenthal D, Chimonas SC, Cohen JJ, Goldman J et al (2006) Health industry practices that create conflicts of interest. A policy proposal for academic medical centers. JAMA 295(4):429–433

Breschi S, Lissoni F, Montobbio F (2007) The Scientific productivity of academic inventors: new evidence from italian data. Econ Innov New Technol 16(2):101–118

Bruneel J, D'Este P, Salter A (2010) Investigating the factors that diminish the barriers to university-industry collaboration. Res Policy 39(7):858–868

Cain DM, Detsky AS (2008) Everyone's a little bit biased (even physicians). JAMA 299(24):2893–2895

Cain DM, Loewenstein G, Moore DA (2005) The dirt on coming clean: perverse effects of disclosing conflicts of interest. J Legan Stud 34(1):1–25

Chavalarias D, Ioannidis JPA (2010) Science mapping analysis characterizes 235 biases in biomedical research. J Clin Epidemiol 63(11):1205–1215
Cohen WM, Nelson RR, Walsh JP (2002) Links and impacts: the influence of public research on industrial R&D. Manag Sci 48(1):1–23
Crawford K, Dobbe R, Dryer T, Fried G, Green B, Kaziunas E, Kak A et al (2019) AI now 2019 report. New York. https://www.ainowinstitute.org/AI_Now_2019_Report.pdf. Accessed 22 Sep 2021
Crawford VP, Sobel J (1982) Strategic information transmission. Econometrica 50(6):1431–1451
D'Este P, Patel P (2007) University–industry linkages in the UK: what are the factors underlying the variety of interactions with industry? Res Policy 36(9):1295–1313
Daly A, Hagendorff T, Hui L, Mann M, Marda V, Wagner B, Wang W, Witteborn S (2019) Artificial intelligence, governance and ethics: global perspectives: the Chinese University of Hong Kong Faculty of Law Research Paper No. 2019–15. SSRN Electron J 1–41
Dana J, Loewenstein G (2003) A social science perspective on gifts to physicians from industry. JAMA 290(2):252–255
Etzkowitz H, Leydesdorff L (2000) The dynamics of innovation: from national systems and 'mode 2' to a triple helix of university–industry–government relations. Res Policy 29(2):109–123
Evans JA (2010) Industry induces academic science to know less about more. Am J Sociol 116(2):389–452
Fabrizio KR, Minin AD (2008) Commercializing the laboratory: faculty patenting and the open science environment. Res Policy 37(5):914–931
Faraway JJ (2014) Linear models with R. CRC Press, Boca Raton
Fickweiler F, Fickweiler W, Urbach E (2017) Interactions between physicians and the pharmaceutical industry generally and sales representatives specifically and their association with physicians' attitudes and prescribing habits: a systematic review. BMJ Open 7(9):1–12
Gigerenzer G, Selten R (eds) (2001) Bounded rationality: the adaptive toolbox. The MIT Press, Cambridge
Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. In: Ghahramani Z, Welling M, Cortes C, Lawrence ND, Weinberger KQ (eds) Advances in neural information processing systems 27, pp 2672–2680. ACM Association for Computing Machinery, New York
Haidt J (2001) The emotional dog and its rational tail: a social intuitionist approach to moral judgment. Psychol Rev 108(4):814–834
Hopkins WE (1997) Ethical dimensions of diversity. Sage, Thousand Oaks
Ioannidis JPA (2005) Why most published research findings are false. PLoS Med 2(8):696–701
Kahneman D (2012) Thinking, fast and slow. Penguin, London
Klinger J, Mateos-Garcia J, Stathoulopoulos K (2020) A narrowing of Ai research? arXiv 1–58
Komesaroff PA, Kerridge I, Lipworth W (2019) Conflicts of interest: new thinking, new processes. Intern Med J 49(5):574–577
Krimsky S (2013) Do financial conflicts of interest bias research? Sci Technol Hum Values 38(4):566–587
Leavy S (2018) Gender bias in artificial intelligence. In: Abraham E, Nitto ED, Mirandola R (eds) Proceedings of the 1st international workshop on gender equality in software engineering. ACM Press, New York, pp 14–16
Lipton ZC, and J Steinhardt J (2018) Troubling Trends in Machine Learning Scholarship. arXiv Preprint arXiv:1807.03341
Loewenstein G, Sah S, Cain DM (2012) The unintended consequences of conflict of interest disclosure. JAMA 307(7):669–670
Luhmann N (1995) Social systems. Stanford University Press, Redwood City
Lundh A, Lexchin J, Mintzes B, Schroll JB, Bero L (2017) Industry sponsorship and research outcome. Cochrane Database Syst Rev 2:1–143
McKinney WP (1990) Attitudes of internal medicine faculty and residents toward professional interaction with pharmaceutical sales representatives. JAMA 264(13):1693–1697
Mittelstadt B (2019) Principles alone cannot guarantee ethical Ai. Nat Mach Intell 1(11):501–507
Mohammad SM (2020) Gender gap in natural language processing research: disparities in authorship and citations. arXiv 1–12
Moore DA, Loewenstein G (2004) Self-interest, automaticity, and the psychology of conflict of interest. Soc Justice Res 17(2):189–202
Orlowski JP, Wateska L (1992) The effects of pharmaceutical firm enticements on physician prescribing patterns. there's no such thing as a free lunch. Chest 102(1):270–273
Pepitone J (2013) Black, female, and a silicon valley 'Trade Secret'. CNN. https://www.money.cnn.com/2013/03/17/technology/diversity-silicon-valley/index.html. Accessed 22 Sep 2021
Perkmann M, Walsh K (2007) University–industry relationships and open innovation: towards a research agenda. Int J Manag Rev 9(4):259–280
Perkmann M, Tartari V, McKelvey M, Autio E, Broström A, D'Este P, Fini R et al (2013) Academic engagement and commercialisation: a review of the literature on university-industry relations. Res Policy 42(2):423–442
Probst P, Knebel P, Grummich K, Tenckhoff S, Ulrich A, Büchler MW, Diener MK (2016) Industry bias in randomized controlled trials in general and abdominal surgery: an empirical study. Ann Surg 264(1):87–92
Rodwin MA (1993) Medicine, money and morals: physicians' conflicts of interest. Oxford University Press, New York
Santamaría L, Mihaljević H (2018) Comparison and benchmark of name-to-gender inference services. PeerJ Comput Sci 4:e156
Savage N (2017) Industry links boost research output. Nature 552(7683):S11–S13
Simonite T (2018) AI is the future—but where are the women? Wired. https://www.wired.com/story/artificial-intelligence-researchers-gender-imbalance/. Accessed 22 Sep 2021
Sinha A, Shen Z, Song Y, Ma H, Eide D, Hsu B-J, Wang K (2015) An overview of Microsoft academic service (Mas) and applications. In: Proceedings of the 24th international conference on World Wide Web, pp 243–46
Szegedy C, Zaremba W, Sutskever I, Bruna J, Erhan D, Goodfellow I, Fergus R (2013) Intriguing properties of neural networks. arXiv Preprint arXiv:1312.6199
Thompson DF (1993) Understanding financial conflicts of interest. N Engl J Med 329(8):573–576
Tversky A, Kahneman D (1974) Judgment under uncertainty: heuristics and biases. Science 185(4157):1124–1131
Washburn J (2008) University Inc: the corporate corruption of higher education. Basic Books, New York
Myers WS, Whittaker SM, Crawford K (2019) Discriminating systems: gender, race, and power in Ai. AI Now
Wright BD, Drivas K, Lei Z, Merrill SA (2014) Technology transfer: industry-funded academic inventions boost innovation. Nature 507(7492):297–299
Xu S, Mariani MS, Lü L, Napolitano L, Pugliese E, Zaccaria A (2021) Citations or dollars? Early signals of a firm's research success. http://arxiv.org/abs/2108.00200
Yuan Y, Sarazen M (2020) Exploring gender imbalance in Ai: numbers, trends, and discussions. Medium. https://www.medium.com/syncedreview/exploring-gender-imbalance\\-in-ai-numbers-trends-and-discussions-33096879bd54. Accessed 22 Sep 2021

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