

# Industry, Experts, or Industry Experts? Academic Sourcing in News Coverage of AI

**Authors:** J. Scott Brennen, Anne Schulz, Philip N. Howard, and Rasmus Kleis Nielsen

## Key Findings

Part hype, part hysteria, part thoughtful debate, the public conversation around artificial intelligence (AI) continues to evolve. Commentators, journalists, politicians, and publics now debate myriad topics concerning AI, from the benefits and risks of automation, to the role of AI in geopolitics and warfare, to the hidden biases of algorithmic discrimination.

News media do much to shape this evolving public discussion of AI. Not only do they provide a space for public conversation, but news reporting and commentary also provide and mediate evidence, arguments, values, and meanings. Similarly, the people who populate news content as sources and story subjects give shape to reporting and commentary by providing examples and anecdotes, giving testimony, providing context, and evaluating claims.

Within news articles about science and technology, academic researchers frequently serve as story subjects and sources, providing useful first-hand experience and specialised technical understanding. For coverage of topics like artificial intelligence, where reporting is heavily influenced by industry hype (Chuan et al. 2019) and future expectations (Natale and Ballatore 2017), academic researchers can provide a unique and independent perspective, often distinct from that of industry sources, politicians, or

policy makers. Specifically, researchers can provide detailed and realistic assessments of the capabilities, limitations, risks, and potential of AI.

Building on a previous finding that academics make up around 17% of sources in news articles on AI (Brennen et al. 2018), this factsheet looks more closely at who are the academics serving as sources for news stories in the UK and the US. Specifically, it examines and models the news mentions of the 150 most cited academic experts in artificial intelligence in both the UK and the US. Analysis of these data reveals three main findings:

- News mentions are concentrated on a very small number of high-profile scholars: in both countries the ten most mentioned scholars account for more than 70% of news mentions in the sample.
- The researchers who appear most regularly in the news are not necessarily the most widely cited by their academic peers and many of them have strong ties to industry. The 16% of scholars across countries who are affiliated with industry account for 64.8% of all news mentions, but only 18.3% of academic citations. Scholars with an industry affiliation are more likely to have many times more news mentions in both the UK and the US. We find no statistically significant relation between academic citations and news mentions in this sample.

- Many of the scholars who feature most frequently in the news are men, and the overwhelming majority of all AI researchers who appear in the news are men. Only 6% of the sample are women, who account for 6.7% of news mentions and 4.2% of Google Scholar citations.

Given the breadth and diversity of AI research across academia, public conversation would benefit from journalists including and highlighting a wider set of the most prominent academics to provide first-hand expertise on the actual capabilities, limitations, and risks of AI research. At the same time, as more and more academic AI researchers are becoming affiliated with companies, it remains unclear what influence industry affiliation has on the way AI researchers publicly share and discuss their work.

## General Overview

Over the past several decades, scientists and scientific organisations across fields have increasingly embraced the value of publicly communicating their research (Rödger et al. 2012). This trend appears to be occurring within AI research as well. In an interview, Michael Wooldridge, who is the Head of the Department of Computer Science at the University of Oxford and who appears in this data set as one of the most cited AI scholars in the UK, observed a similar shift over a long career in computer science, noting, 'now I think part of my role is to be a commentator on what's going on in AI'.

Research on the public engagement of science has identified a number of reasons why scientists say they participate in the public communication of their work (Besley et al. 2018). Some researchers describe a feeling of obligation or duty to communicate their research to the public, especially when their research is funded with public money (Sharman and Howarth 2017). David Hogg, Professor of Artificial Intelligence at the University of Leeds, who is also included in these data, explained in an interview that he feels

*a desire to give an accurate picture of what's actually happening, what the limitations are, and the way in which it works. I want people to get a better understanding.*

Kate Devlin, Senior Lecturer in Social and Cultural Artificial Intelligence at King's College London, said she feels motivated to counteract pervasive misperceptions about AI. She noted in an interview

that when she speaks to the public about her work,

*I'm trying to show that it's much more nuanced than would be suggested by the media coverage. And my own personal goal is to try and myth-bust things, so I like to go in and say, right, you might think this, but that's a really mistaken view and here's the actual truth behind it. So it's kind of the truth behind the headline sort of thing.*

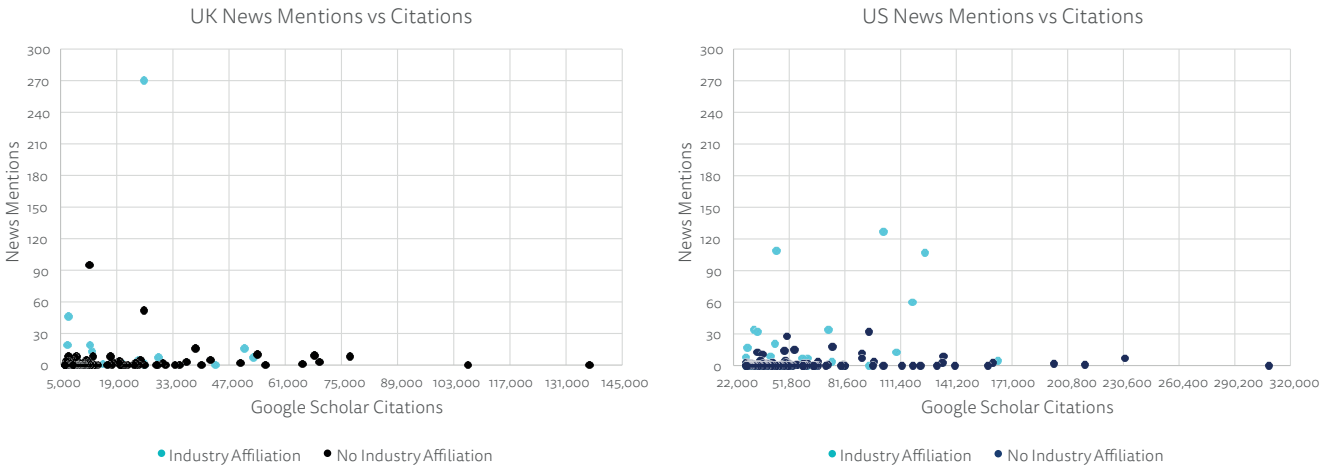
Decades of research on news sources has shown that reporters regularly turn to official and elite sources. This holds for reporting on science, technology, and health too, where scientists have been shown to have significant power both to generate stories and to shape how articles are written and framed (Nisbet and Lewenstein 2002). The influence of institutional and elite sources has grown as more reporting on science, technology, and health is done by generalists with little expertise in those areas (Dunwoody 2008). Less is known, however, about academic sourcing in reporting on computer science or AI.

## Methods

To better understand which academic scholars appear in news reporting on AI, we identified the 150 most cited academic-affiliated AI experts in both the UK and the US. To do so, we searched Google Scholar profiles for self-reported specialists in AI, machine learning, deep learning, or neural networks who have academic email addresses. We then collected the number of mentions for each of these scholars in 34 UK and 80 US major news outlets in a news database, from 5 September 1980 to 14 September 2019. We did so by searching variants of each scholar's name, along with a set of keywords related to AI. We also used publicly available information, including institutional webpages, CVs, news articles, and Google Scholar pages to determine each scholar's professional field affiliation, length of career, gender, and industry affiliation. Finally, we modelled news mentions based on the five factors we collected using a negative binomial regression. Notably, this approach means this is not a study of all academics working on AI, but rather a study of the most prominent scholars in the field according to Google Scholar. We also completed a series of interviews with UK-based technology journalists and AI researchers to better understand and contextualise these findings. Please see the methodological appendix for a fuller discussion of our methods.

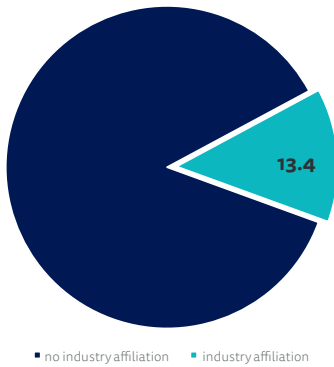
## Results

**Figure 1: News Mentions and Google Scholar Citations**

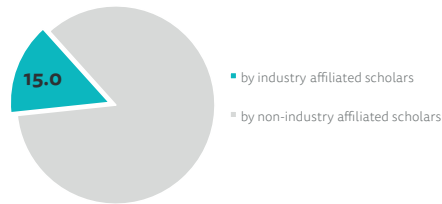
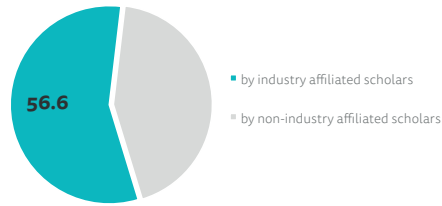


**Figure 2: Industry-Affiliated AI Scholars in UK News**

Share of industry affiliation among the 150 most cited UK-based AI scholars on Google Scholar



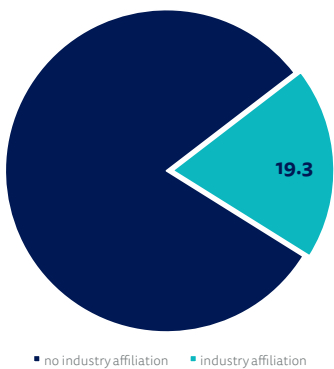
Share of news mentions



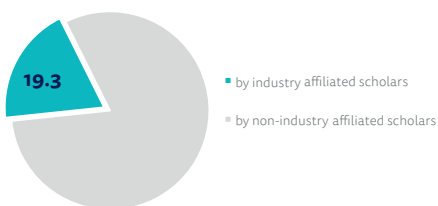
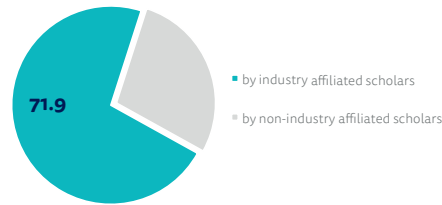
Share of Google Scholar citations

**Figure 3: Industry-Affiliated AI Scholars in US News**

Share of industry affiliation among the 150 most cited US-based AI scholars on Google Scholar



Share of news mentions



Share of Google Scholar citations

**Table 1: Selected Descriptive Statistics**

	🇬🇧 UK			🇺🇸 US		
	N	% of Total News Mentions	% of Total Google Scholar Citations	N	% of Total News Mentions	% of Total Google Scholar Citations
Most Google Scholar Citations	10	7.5	27.4	10	3.6	20.2
Most News Mentions	10	76.9	9.1	10	70.0	7.9
All Industry-Affiliated Scholars	20	56.6	15.0	29	71.9	19.3
Female Scholars	10	1.4	5.4	8	11.4	3.9
Scholars with No News Mentions	92	0	54.7	84	0	52.1

### CITATIONS AND MENTIONS

The data suggest that the scholars with the most Google Scholar citations are *not*, by and large, the scholars with the most news mentions. Notably, many of the most highly cited scholars had few media mentions (see Figure 1). The 10 most cited scholars in the UK and US, despite accounting for 27.4% and 20.2% of all academic citations in the sample, account for only 7.5% and 3.6% of news mentions (see Table 1). Importantly, our statistical model did not find a relationship between the number of Google Scholar citations and news mentions (see methodological appendix).

### INDUSTRY AFFILIATION

Across these data, industry affiliation is strongly correlated to news mentions. Our model identified industry affiliation as the strongest predictor of news mentions (see methodological appendix). While the model contains a large degree of uncertainty, it suggests that on average industry-affiliated AI scholars receive 34.6 (UK) times and 20.6 (US) times as many mentions as AI scholars without industry affiliation.

Looking more broadly at the data, industry-affiliated researchers (20 in the UK and 29 in the US) account for 56.6% of news mentions in the UK and 71.9% in the US. As comparison, those same industry researchers only account for 15% and 19.3% of Google Scholar citations (see Figures 2 and 3).

While these data do not offer a clear explanation for the strong association observed here between industry affiliation and news mentions, two broader trends may provide some insight. First, many companies – especially the large tech companies – have large public relations and communication efforts that help

drive coverage of industry-related research, initiatives, and events. In an interview, one UK technology journalist estimated, ‘I probably get at least one email that mentions something about artificial intelligence every day.’ More, there is indication that industry PR and media relations efforts *work* (Göpfert 2007). While some journalists in our interviews strongly denied writing stories from press releases, others reported doing so regularly. As journalists face cutbacks, time pressures, and declining resources, there is reason to believe some turn both to easy-to-write stories from industry press releases and to industry sources (Schäfer 2017).

Second, large tech companies have been hiring prominent AI researchers. After some pushback that companies were decimating academic departments, some companies have now been offering academics dual positions. One UK journalist observed,

*DeepMind have also bought up everybody, every academic in the field ends up working at DeepMind, which is a bit of a problem... the sheer power of Google is a worry, however good their intentions may be.*

For some academics, this involves taking a leave of absence from their academic positions. For others, it means splitting their time between academia and industry. Ultimately, this trend could mean that, instead of industry affiliation driving news mentions and citations, news mentions and citations could be driving industry affiliation.

### GENDER

The data also suggest a notable gender imbalance in both the field of AI research and news coverage. Across both countries only 18 of the 300 most cited

researchers we analysed, or 6%, are female. These 18 women account for 4.2% of Google Scholar citations and 6.7% of news mentions in the sample. Notably, the US and UK data diverge slightly: 8 women in the US make up 11.4% of all US news mentions and 10 women in the UK make up 1.4% of news mentions in our sample. These findings belie recent efforts to increase representation in AI research by ensuring that female researchers are included on conference panels and in expert groups consulting with governments. Similarly, many of the major think tanks and research groups working on the social impact of AI are headed by women, and recently there have been a series of major conferences that feature the contributions of women to AI development.

There is, however, indication that some outlets have begun to take steps to increase representation of female sources. In an interview, one UK tech journalist confirmed that their outlet had begun tracking the genders of sources in published articles to help journalists be more mindful of ‘unconscious biases’ in sources because ‘you always end up going back to certain people’. At the same time, the model found that for US scholars, being female is associated with about four times more news mentions.

## CONCENTRATION

News mentions are highly concentrated in a small number of scholars (see Figure 1). The top 10 scholars in the UK and US account for 76.9% and 70% of all mentions, respectively. Notably, those same 10 scholars in each country only account for 9.1% and 7.9% of Google Scholar citations (see Table 1).

Across both countries only 12 scholars have more than 30 news mentions. Together, they account for 64.1% of all news mentions, but only 5.9% of citations. Notably, 10 of the 12 are men, and 11 have worked as faculty members in university departments (although two are not computer scientists, but philosophers). Finally, 9 of the 12 have strong industry ties.

In contrast to these heavily sourced scholars, 61.7% of the sampled UK and 56% of US scholars have no news mentions at all. These groups nonetheless account for 54.7% and 52.1% of all Google Scholar citations in the sample (see Table 1). All of the scholars in the sample are leaders in their fields and are all likely pursuing important research worthy of coverage or could serve as an important source of independent, accurate, and timely insight about the state of research.

These data suggest that journalists return again and again to a very small number of scholars. Increasing the

diversity of sources even from within a single company could help bring new and deeper understanding of ongoing AI research.

Both existing research and interviews completed with UK technology journalists for this project offer some insight into this concentration of news sources and subjects. First, as journalists face significant time pressures in their day-to-day reporting, there is a strong incentive to return to sources who have proven to provide quick, useful, and interesting quotes. Cultivating new sources is both time intensive and risky – they may or may not provide usable quotes. One UK-based technology journalist observed:

*There’s the typical journalistic thing, particularly in broadcasting, of once you find somebody who has got a plausible manner and speaks well, you kind of stick to them for maybe too long and maybe you should be looking for alternative voices.*

At the same time, several journalists who work at tabloid or digital native outlets reported in interviews that they are required to meet regular traffic targets. When attempting to reach traffic targets on tight deadlines, there is pressure to return to reliable topics. While one technology journalist observed that ‘there’s no telling what’s going to work or not; you throw a lot of stuff at a wall’, part of the expertise of being a technology journalist is knowing the sorts of topics and subjects that will draw traffic.

At the same time, journalists often look both to other outlets and to press releases for article topics and sources. Doing so helps ensure that journalists do not miss important and popular stories, while providing an easy and reliable set of sources for their own articles.

Taking this a step further, some journalists admitted to using quotes published in other news articles or press releases. One UK tabloid technology reporter described this practice this way:

*So, you’re not stealing, but you’re basically bridging the line with stealing, because you’re not just like copying and pasting, you’re writing it in your own way, but it’s using information that someone else has gathered.*

This practice means that, once a researcher appears in one article, they are likely to appear in a number of other articles, as journalists reconfigure existing articles for their own publications.

## CONCLUSIONS AND RECOMMENDATIONS

This analysis should not be seen as a criticism of those

scientists who appear in news reporting. Likewise, although existing research suggests that some scholars may be hesitant to speak with media (Dudo 2015), many of the patterns described here appear to be more a function of journalistic practices. As journalists face notable time and resource pressures, there are strong incentives to pursue sources who provide ready-made stories and frames or who have proven to be quick, reliable, and interesting. That is to say, journalists return to sources who have both technical expertise and expertise as news sources. Likewise, facing these pressures, many journalists turn to easy-to-write stories either about industry initiatives or drawn directly from industry press releases. While university press offices are increasingly professionalising (Schäfer 2017), many still cannot duplicate the efforts of industry press offices.

Although these data provide only a partial view of news sourcing of AI researchers, they suggest some cause for concern. While there are growing expectations that tech journalists help the public understand the impact and potential of new technologies and hold the technology industry to account, our research suggests that tech reporting continues to be deeply reliant on industry. In our previous work, we found that 33% of unique sources in coverage of AI in the UK were from industry, and just 17% from academics. In this factsheet, we have shown that a large number of the academics who appear in news coverage of AI also have strong ties to industry.

This is not always a problem, but in light of the earlier failure of uncritical and industry-led technology reporting to question the power and influence of social media platforms as they developed, it is worth asking: will coverage of AI be more independent, informative, and investigative from the outset, or is there a risk that technology journalism will repeat the same mistakes? As AI is promoted and embraced across sectors, it is important that technology journalists critically interrogate not only industry claims about the potential of AI, but also the ethics, risks, and biases of AI-based systems.

Given the huge variety of AI research now occurring both in and outside of industry, journalists would be well served to work to develop new and diverse sources for their reporting, including from a wider range of independent academics. Increasing the diversity of sources and story subjects could help provide broader, richer, and potentially more critical insight into the many pressing public problems and opportunities surrounding artificial intelligence.

## References

- Besley, J. C., Dudo, A., Yuan, S. 2018. 'Scientists' Views about Communication Objectives', *Public Understanding of Science* 27(6), 708–30.
- Brennen, J. S., Howard, P. N., Nielsen, R. K. 2018. *An Industry-Led Debate: How UK Media Cover Artificial Intelligence*. Oxford: Reuters Institute for the Study of Journalism.
- Chuan, C. H., Tsai, W-H. S., Cho, S. Y. 2019. 'Framing Artificial Intelligence in American Newspapers.' *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* (AIES '19). ACM, New York, NY, 339–344.
- Dudo, A. 2015. 'Scientists, the Media, and the Public Communication of Science', *Sociology Compass* 9(9), 761–75.
- Dunwoody, S. 2008. 'Science Journalism', in M. Bucchi and B. Trench (eds), *Handbook of Public Communication of Science and Technology*. London: Routledge, 15–26.
- Göpfert, W. 2007. 'The Strength of PR and the Weakness of Science Journalism', in M. W. Bauer and M. Bucchi (eds), *Journalism, Science and Society*. New York: Routledge, 215–26.
- Natale, S., Ballatore, A. 2017. 'Imagining the Thinking Machine: Technological Myths and the Rise of Artificial Intelligence' *Convergence: The International Journal of Research into New Media Technologies*, 1–16. <https://doi.org/10.1177/1354856517715164>
- Nisbet, M. C., Lewenstein, B. V. 2002. 'Biotechnology and the American Media: The Policy Process and the Elite Press, 1970 to 1999', *Science Communication* 23(4), 359–91.
- Rödger, S., Franzen, M., Weingart, P. (eds). 2012. *The Sciences' Media Connection: Public Communication and its Repercussions*. London: Springer.
- Schäfer, M. S. 2017. 'How Changing Media Structures are Affecting Science News Coverage', in K. H. Jamieson, D. M. Kahan, and D. A. Scheufele (eds), *The Oxford Handbook of the Science of Science Communication*. Oxford: Oxford University Press, 51–9.
- Sharman, A., Howarth, C. 2017. 'Climate Stories: Why do Climate Scientists and Sceptical Voices Participate in the Climate Debate?', *Public Understanding of Science*, 26(7), 826–42.

## Acknowledgements

The authors would like to thank Richard Fletcher for his insight, guidance, and patience, as well as Christina Koster, Simge Andi, Sílvia Majó-Vázquez, Seth Lewis, Lea Hellmueller, Sascha Hölig, and the rest of the research, communications, and administration teams at the Reuters Institute for the Study of Journalism.

### ABOUT THE AUTHORS

**J. Scott Brennen** is a Research Fellow at the Reuters Institute for the Study of Journalism and the Oxford Internet Institute at the University of Oxford.

**Anne Schulz** is a Research Fellow at the Reuters Institute for the Study of Journalism at the University of Oxford.

**Philip N. Howard** is the Director of the Oxford Internet Institute and a Professor of Sociology, Information and International Affairs.

**Rasmus Kleis Nielsen** is the Director of the Reuters Institute for the Study of Journalism and Professor of Political Communication at the University of Oxford.



Published by the Reuters Institute for the Study of Journalism as part of the Oxford Martin Programme on Misinformation, Science and Media, a three-year research collaboration between the Reuters Institute, the Oxford Internet Institute, and the Oxford Martin School.

## Methodological Appendix

### DATA COLLECTION

The data described in this factsheet were collected as follows. On 14 September 2019, we searched Google Scholar profiles for self-identified academically affiliated experts in AI. We queried Google Scholar profiles for scholars whose profiles include the labels artificial intelligence, AI, machine learning, deep learning, or neural networks, as well as for academic email addresses, identified by adding ‘ac.uk’ or ‘edu’ to the search terms. For each country we collected the names and citation count of the 150 most cited scholars. We removed a small number of names from these lists that did not appear to be real people. We also removed Alan Turing from the UK sample. Although he had hundreds of news mentions, he died 65 years ago. While other scholars in the sample are also deceased, Alan Turing represents an extreme case. Complete and reliable data from one UK scholar could not be guaranteed, and so that scholar was removed from the sample, dropping the number of UK scholars in the sample to 149. Because our starting point was to identify the most prominent academics using Google Scholar and then look for mentions of them in news articles, we were unable to identify the news mentions of those outside of the 150 most cited in the UK and the US. Therefore, our findings apply to the media coverage surrounding a relatively small group of highly cited academics and not necessarily to the media coverage of AI as a whole.

We chose Google Scholar citations as a rough measure of field prominence. It should be noted that, while there is a long history of using the number of citations of academics in this way, it is in some ways a limited measure of a scholar’s standing in the community. In addition to indexing a range of publications of varying qualities, Google Scholar counts all citations across fields, rather than only within an area of expertise. This means that authors of textbooks and popular books benefit from boosts to their citations. Also, some scholars, especially those working in experimental physics, work in teams of hundreds or thousands of scholars and produce articles that attract huge numbers of citations. In some collaborations, the norm is to include all members of the collaboration on each publication. Also, some Google Scholar profiles studied here appear to include publications from scholars with similar names, further increasing the number of citations.

We rely on how experts choose to present themselves through self-applied labels in Google Scholar. This

means that we likely missed scholars who either do not have a Google Scholar page associated with their institutional email, or who have not been labelled as researching some aspect of AI. Also, we chose not to attempt to assess the accuracy or validity of these labels. We also relied on scholars having an academic email account for membership in these data. Academic affiliation is not always clear. Some industry researchers have academic affiliations – or at least academic email addresses. For example, Demis Hassabis of DeepMind has an academic email address listed in Google Scholar. We did not investigate whether this email address remains active – or the full nature of his relationship with the university.

Next, for each set of scholars, we searched Factiva’s collections of Major News and Business Sources for the US or UK for all news mentions of each name, along with the same AI-related keywords above over all dates until 14 September 2019. Before searching, we removed all business-only news sources and all that only distribute press releases (e.g. Dow Jones Institutional News) from the source collections. We searched variants of each scholar’s name, such as with and without middle names and/or initials. We also searched the source lists for just the collection of AI-related terms to get a better idea of the total number of news articles related to AI in the entire source collections over the entire time frame; we found 22,664 US and 38,793 UK articles. While completing searches, we selected the option to remove all similar articles, in order to limit duplicates.

For each scholar in the list we identified four additional pieces of information. Based on name, picture, and other published information, we labelled each scholar as presenting as male or female. This approach is undoubtedly problematic. It required us to make a determination based on outward presentation, rather than actually understanding how each scholar identifies.

We also attempted to determine if each scholar is currently affiliated with a for-profit business. To do so, we looked at personal, institutional, and industry websites, CVs, resumes, and Google Scholar pages. For some scholars, this was a clearer determination than for others. Many scholars are associated with small start-up projects. We did not count these associations as industry affiliation. Instead, we coded industry affiliation only for those with clear paid employment for a for-profit company.

While our sample includes many of the most well-known computer scientists working on AI, it also



includes some physicists, biologists, and geneticists who say they use AI approaches in their work. Rather than make judgements about who should or should not be included, we included a variable for professional field identification. These were mostly assessed through CVs or through institutional affiliation. We collapsed these into a simple dichotomous variable: 0 for being in computer science or related field, 1 for being in another field, such as high energy physics, biology, or genetics.

Finally, we used personal and institutional websites, CVs, and Google Scholar pages to determine the number of years since each scholar’s first publication.

### INTERVIEWS

This factsheet draws on interviews with 17 UK-based technology journalists and 5 UK-based AI researchers. Interviews with technology journalists also inform an academic article currently under journal review. Interview protocol went through ethics approval at the University of Oxford. Semi-structured interviews lasted between 30 and 120 minutes and addressed a range of topics. For journalists, interview questions addressed work histories and day-to-day reporting practices. For researchers, questions addressed research and public communication experiences.



Many of the journalists were granted anonymity as per their request. Researchers named in the report gave express permission. They were also given a chance to approve quotations. Journalists were identified through article by-lines and snowball sampling. Researchers were identified from news mentions and institutional affiliations.

### DATA ANALYSIS

In addition to undertaking a series of descriptive analyses, we modelled the number of news mentions based on the five factors collected: Google Scholar citations, years researching, gender, industry affiliation, and professional discipline.

The dependent variable in this study (number of news mentions) is a count variable. Usually, for count responses Poisson-type regressions are applied. But, because we found signs of overdispersion (i.e. the variance of our outcome variable is larger than its mean), we use negative binomial regressions that account for such data. We ran the analyses in R using the MASS package in W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S* (4th edn, New York: Springer, 2002). Table 2 summarises the results of these analyses for the UK and the US.

**Table 2: Negative Binomial Regression Predicting the Number of News Mentions for UK- and US-Based AI Scholars.**

	 UK			 US		
Factor	Exp(b)	S.E.	P	Exp(b)	S.E.	P
Google Scholar Citations	1.00	0.00	0.06	1.00	0.00	0.25
Non-Comp Sci (Ref=yes)	4.81	0.58	0.01	1.40	0.42	0.42
Industry Affiliation (Ref=yes)	34.57	0.58	0.00	20.64	0.43	0.00
Gender (Ref=female)	0.21	1.03	0.13	4.09	0.70	0.04
Years Research Active	1.07	0.02	0.00	1.08	0.01	0.00
Pseudo-R <sup>2</sup> (Cragg-Uhler)	0.21			0.32		
N	149			150		

Note. Exp(b) = Exponentiated coefficients. Values < 1 indicate a negative effect, values > 1 indicate a positive effect; S.E. = standard error.