Article

Artificial intelligence tools expand scientists' impact but contract science's focus

https://doi.org/10.1038/s41586-025-09922-y

Received: 2 January 2025

Accepted: 14 November 2025

Published online: 14 January 2026

Check for updates

Qianyue Hao[1], Fengli Xu[1] ✉, Yong Li[1,2] ✉ & James Evans[3,4] ✉

Developments in artificial intelligence (AI) have accelerated scientific discovery[1]. Alongside recent AI-oriented Nobel prizes[2–9], these trends establish the role of AI tools in science[10]. This advancement raises questions about the influence of AI tools on scientists and science as a whole, and highlights a potential conflict between individual and collective benefits[11]. To evaluate these questions, we used a pretrained language model to identify AI-augmented research, with an F1-score of 0.875 in validation against expert-labelled data. Using a dataset of 41.3 million research papers across the natural sciences and covering distinct eras of AI, here we show an accelerated adoption of AI tools among scientists and consistent professional advantages associated with AI usage, but a collective narrowing of scientific focus. Scientists who engage in AI-augmented research publish 3.02 times more papers, receive 4.84 times more citations and become research project leaders 1.37 years earlier than those who do not. By contrast, AI adoption shrinks the collective volume of scientific topics studied by 4.63% and decreases scientists' engagement with one another by 22%. By consequence, adoption of AI in science presents what seems to be a paradox: an expansion of individual scientists' impact but a contraction in collective science's reach, as AI-augmented work moves collectively towards areas richest in data. With reduced follow-on engagement, AI tools seem to automate established fields rather than explore new ones, highlighting a tension between personal advancement and collective scientific progress.

Artificial intelligence (AI) has made considerable strides in recent decades, promising to affect myriad aspects of society, including education[12,13], healthcare[14,15] and industry[16]. Major investments in predictive and generative AI have catalysed society-level debates over the future of AI at home and in the workplace. Perhaps more than any other domain, AI tools have become deeply entwined with the process of knowledge production, yielding findings that attract disproportionate attention in various scientific fields[1]. For example, AlphaFold, which recently earned the 2024 Nobel Prize, learns known protein structures to accurately predict unexplored ones, circumventing the human and experimental cost of conventional structural inference[9,17]. Models improved via deep reinforcement learning have sustained complex plasma configurations in fusion reactors[18] and discovered new, hardware-optimized forms of matrix multiplication that recursively accelerate deep learning itself[19]. Autonomous laboratory systems driven by ChatGPT have helped some chemists and materials scientists upscale the number of adaptive high-throughput experiments[20–22]. Recent developments in large language models have also become increasingly used to assist scientific writing[23–26] and facilitate the distillation of scientific findings, but they raise concerns about weakened confidence in AI-generated content[21,22,27]. Artificial intelligence's increasing capabilities to influence scientific research suggest that it manifests the potential to both increase the productivity of individual scientists and raise the visibility of the science it supports.

Despite the increasing adoption of AI in science, large-scale empirical measurements of AI's scientific impact are limited, and a detailed, dynamic understanding of AI's influence on the entire character of science remains largely unknown. Recent work suggests that AI has brought widespread benefits to individual scientists but may lead to demographic disparity resulting from gaps in AI education[10]. Researchers have also identified evolving citation patterns that signal a changing scientific landscape in AI research[28]. Here we explore the impact of AI in scientific research at different scales, and how the adoption of AI influences both individual scientists' careers and the collective exploration of science as a whole.

We conduct a large-scale quantitative analysis of the impact of AI on scientists and science, covering 41,298,433 research papers spanning from 1980 to 2025 in the OpenAlex dataset[29], with patterns corroborated using the Web of Science[30,31]. Notably, we do not focus on computer science or mathematics—fields which develop AI methodologies directly—but rather on papers that augment research in natural science fields by adopting AI, primarily covering decades that involve the development and deployment of conventional machine learning algorithms, and also extend to a necessarily more preliminary analysis of the latest generative AI techniques. Specifically, we select six representative disciplines that cover the vast majority of natural science contributions: biology, medicine, chemistry, physics, materials science and geology. We then leverage a fine-tuned BERT language model[32,33]

[1]Department of Electronic Engineering, Beijing National Research Center for Information Science and Technology (BNRist), Tsinghua University, Beijing, P. R. China. [2]Zhongguancun Academy, Beijing, P. R. China. [3]Knowledge Lab and Department of Sociology, University of Chicago, Chicago, IL, USA. [4]Santa Fe Institute, Santa Fe, NM, USA. ✉e-mail: fenglixu@tsinghua.edu.cn; liyong07@tsinghua.edu.cn; jevans@uchicago.edu

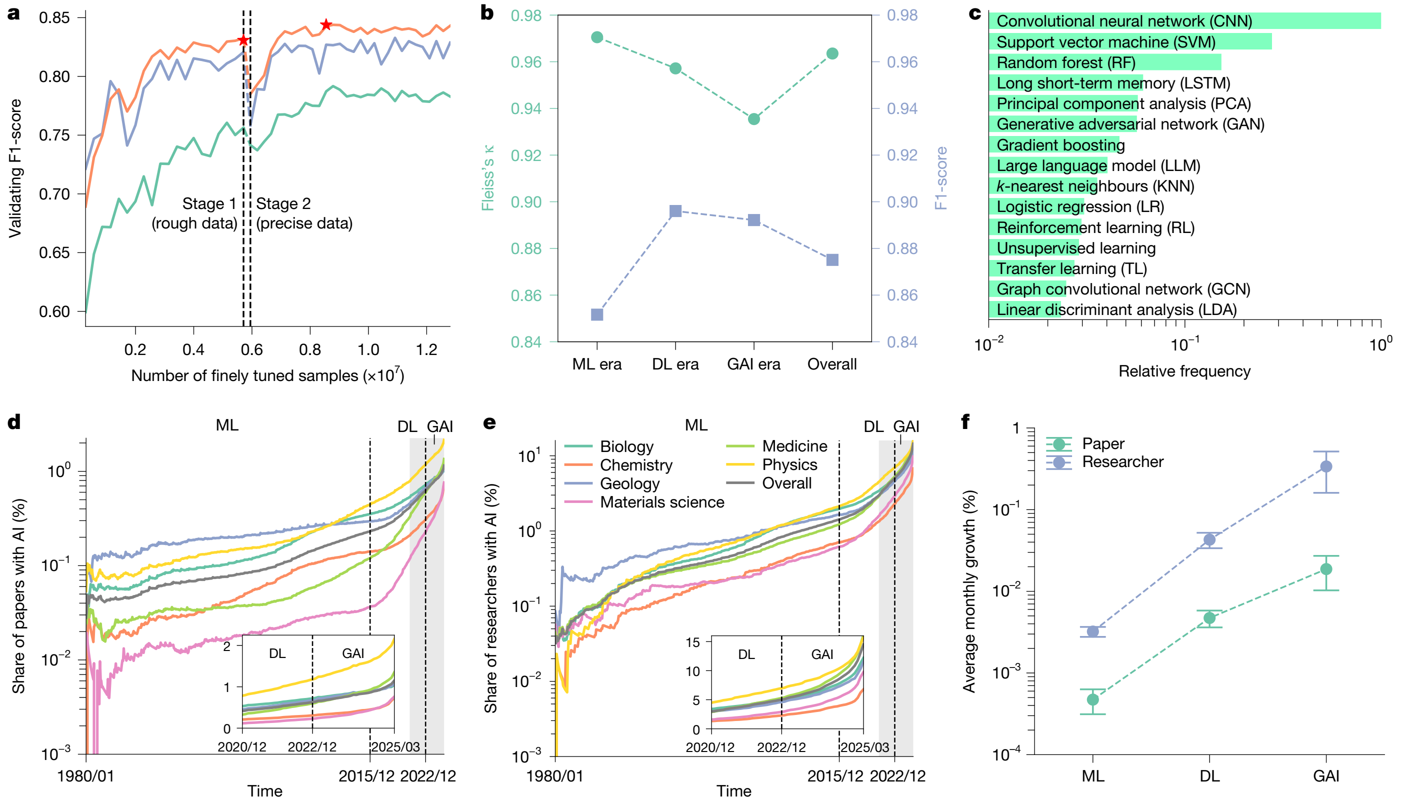


Fig. 1 | Increasing prevalence of AI adoption in science. a, Increasing performance of AI paper identification during the two-stage fine-tuning of BERT pre-trained models, where we use rough training data in stage 1 to evolve precise assessments in stage 2. We independently train two models on titles (green) and abstracts (purple), and then integrate them into an ensemble (orange) that selects the optimal models during both stages (red stars) to identify all relevant papers. **b**, Accuracy evaluation of our identification results by human experts. For samples spanning three eras of AI, experts reached consensus, with $\kappa \geq 0.93$. Our model identification results have strong accuracy in validation against expert-labelled data, with an F1-score ≥0.85. **c**, Relative adoption frequency of the top 15 AI methods across all disciplines for all selected AI development eras. **d**,**e**, The growth of AI-augmented papers (**d**, $n = 41{,}298{,}433$) and AI-adopting researchers (**e**, $n = 5{,}377{,}346$) across machine learning (ML), deep learning (DL) and generative AI (GAI) eras between 1980 and 2025 in selected scientific disciplines. The *y* axes are set to a logarithmic scale. **f**, The average monthly growth rates for AI papers and researchers across the eras of ML, DL and GAI across all selected disciplines ($n = 543$ month observations), where 99% confidence intervals (CIs) are shown as error bars centred at the mean.

to accurately identify such AI-augmented research papers on the basis of their titles and abstracts.

We separate the periods in which AI was predominantly conventional machine learning, deep learning and, most recently, generative designs such as large language models. With abundant data-based evidence across decades of conventional machine learning and deep learning, we validate these AI-based measurements and use them to reveal that the adoption of AI leads to an amplifying effect on the career of individual scientists, accelerating the production and visibility of science produced by those scientists who incorporate AI. Nevertheless, this effect corresponds with a contracted focus within collective science. Measured with 'knowledge extent', the 'diameter' covered by a sampled batch of papers in vector space, AI-driven science spans less topical ground and is associated with a decrease in follow-on scientific engagement, suggesting that AI is currently more likely to focus on existing popular research problems rather than explore new ones. Analyses using currently available data within the latest era of generative AI including large language models reveal consistency with past periods, providing a starting point for further study as generative AI-enhanced science develops over a longer period.

Increasing prevalence of AI in science

Here we focus on research papers using AI methods in various fields of natural science, where we conduct our analysis on the basis of 41,298,433 papers from the OpenAlex dataset[29], covering six representative disciplines: biology, chemistry, geology, materials science, medicine, and physics (Methods). According to the invention of milestone technologies in the trend of AI development, we divide the past decades into three eras, namely, machine learning, deep learning and generative AI (Methods). To identify AI papers in various fields across eras, we fine-tune BERT[32,33], an established language model[34–36], on articles published in explicitly AI-oriented scientific journals and conferences to automatically extract and interpret information from context. Specifically, we use a two-stage fine-tuning process to adapt the pre-trained BERT model to the task of AI paper identification. We first independently train two models based on titles and abstracts of papers, respectively, then ensemble the optimized individual models to identify all selected papers (Fig. 1a, Methods and Extended Data Fig. 1). This approach eliminates the need for manual selection of AI-related trigger words, as demonstrated in previous research[28].

To evaluate the accuracy of our identification, we recruited a team of human experts to validate these results (Methods and Extended Data Fig. 2). The experts formed a strong consensus across their independent annotation of papers sampled at random from the six disciplines mentioned above, achieving an average Fleiss' κ of 0.964 (refs. 37,38). The BERT model attains an average F1-score of 0.875 in an evaluation that uses the expert labels as ground truth. The strong consensus among experts and high quality for identification is consistent across samples from different eras of AI, confirming the reliability of our identification

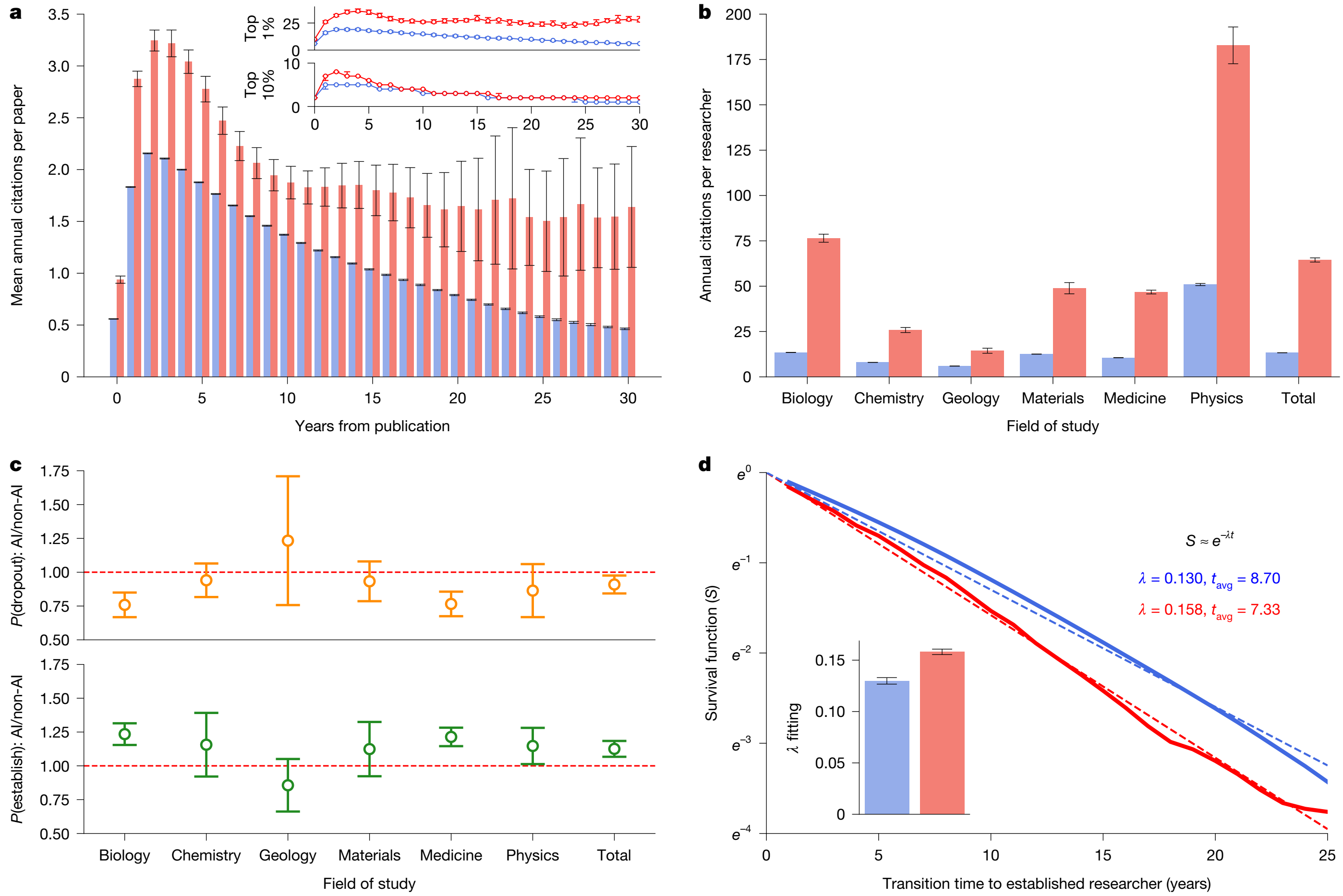


Fig. 2 | AI enlarges paper impact and enhances researcher careers. **a**, Average (insets: top 1% and 10%) annual citations after publication of AI (red) and non-AI (blue) papers (n = 27,405,011), where AI papers attract more citations. **b**, Average annual citations for researchers who use AI and their counterparts who do not (P < 0.001, n = 5,377,346), where researchers who adopt AI receive 4.84 times more citations. **c**, The probability of two role transitions between junior scientists who adopt AI and their counterparts who do not (n = 46 year observations for each field). Junior scientists who adopt AI have a higher probability of becoming established researchers and a lower probability of exiting academia compared with their counterparts who do not adopt AI. **d**, Survival functions for the transition from a junior to an established researcher (P < 0.001, n = 2,282,029). The survival functions can be well-fit with exponential distributions, where junior scientists who adopt AI become established earlier. For all panels, 99% CIs are shown as error bars, with the insets of **a** centred at the 1% and 10% percentiles and other panels centred at the mean. All statistical tests use a two-sided t-test.

accuracy and laying a robust foundation for subsequent analysis (Fig. 1b and Supplementary Tables 1–4). To provide a rationale and explainability for our identification results, we visualize attention strengths in the BERT model with examples, where the model allocates substantial attention to terms such as neural network and large language model, illustrating how the model correctly interprets and accurately identifies AI-related contents from papers published in different eras of AI development (Supplementary Figs. 2 and 3).

In total we identify 310,957 AI-augmented papers, comprising 0.75% of all selected papers. Semantically, the identified AI-related papers tend to combine artificial intelligence and conventional research topics across disciplines (Supplementary Fig. 4). Counting all eras and disciplines collectively, the most commonly adopted AI methods in natural science research include support vector machines and principal component analysis from the machine learning era, and convolutional neural networks and generative adversarial networks from the deep learning era. Large language models, which have emerged in recent years, also rank among the most frequently used methods (Fig. 1c and Supplementary Tables 5–11). Statistically, despite the overall rise in the number of papers published annually across all disciplines[39], the share of AI-augmented papers surged by 10.70 (geology, Z = 348.60, P < 0.001 and degrees of freedom (df) = 1 in a Cochran–Armitage test) to 51.89 (biology, Z = 1,388.70, P < 0.001 and df = 1 in a Cochran–Armitage test) times from 1980 to 2025 (Fig. 1d). Similarly, the proportion of researchers adopting AI has grown even more rapidly: from 135.46 times in geology (Z = 546.81, P < 0.001 and df = 1 in a Cochran–Armitage test) to 362.16 in physics (Z = 2,237.51, P < 0.001 and df = 1 in a Cochran–Armitage test) (Fig. 1e). Meanwhile, growth rates for AI-augmented papers and researchers have accelerated across the three eras (Fig. 1f and Supplementary Figs. 5 and 6). These findings underscore the increasing prevalence and rapid development of AI in science across all disciplines and the importance of understanding AI's impact on scientific research and progress.

AI enhances individual scientists

From statistics across 27,405,011 papers with intact reference records in the OpenAlex dataset, we note that, from the publication date of each paper across subsequent decades, annual citations to AI papers are 98.70% higher than those to non-AI papers on average (Fig. 2a, $t \geq 8.33$, P < 0.001 and df > 10^3 in t-test on any year). In addition to higher annual average citations, the greater scientific impact of AI-augmented papers is also reflected by multiple alternative statistical indicators, including measures of both the highest and lowest annual citation count (Supplementary Fig. 8). Furthermore, AI papers consistently receive more citations, regardless of the era in which they are published (Extended Data Fig. 3, $t \geq 4.06$, $P < 0.001$ and $df > 10^3$ in a t-test on any era). We also examine the distribution of AI-augmented papers across journals of varying Journal Citation Report quantiles[40] (Supplementary Fig. 14). We find that the proportion of AI papers in Q1 journals is 18.60% higher than that of non-AI papers in all journals; in Q2 journals, the AI proportion is 1.59% higher; whereas Q3 and Q4 journals hold a relatively lower proportion of papers with AI ($\chi^2 = 3629.11$, $P < 0.001$ and $df = 3$ in a χ^2-test). These results indicate a heterogeneous distribution of AI-augmented papers across journals, with a higher prevalence in high-impact journals. Paralleled by the attention paid to AI papers, the impact of AI researchers also substantially increases. On average, researchers adopting AI annually publish 3.02 times more papers ($t \geq 47.18$, $P < 0.001$ and $df > 10^3$ in t-test on any discipline) and garner 4.84 times more citations ($t \geq 30.32$, $P < 0.001$ and $df > 10^3$ in t-test on any discipline) than those not adopting AI, with consistency across disciplines and robustness for core researchers with multi-year continuous publication records[41] (Fig. 2b, Extended Data Fig. 4 and Supplementary Fig. 17). Furthermore, when controlling for and comparing scientists with similar early career positions, the enhanced productivity and impact still hold (Supplementary Fig. 16). This suggests that, after accounting for potential selection-biases among researchers with different original achievements that may influence their choice of AI adoption, AI itself contributes to the observed advantages.

To identify the implications of AI adoption on a scientist's career development, we classify the scientists into 'junior' and 'established'; junior scientists are defined as newcomers who have not yet led a research project, whereas established scientists are defined as those who have led one or more research projects (Methods and Extended Data Fig. 5). We extract the career trajectories of 2,282,029 scientists from the dataset, each initially identified as a junior scientist (Methods). The results reveal that AI-augmented research is associated with reduced research team sizes, averaging 1.33 (19.29%) fewer scientists ($t = 20.47$, $P < 0.001$ and $df > 10^3$ in a t-test; Extended Data Fig. 6). Specifically, the average number of junior scientists decreased from 2.89 in non-AI teams to 1.99 (31.14%) in AI teams ($t = 19.02$, $P < 0.001$ and $df > 10^3$ in t-test), whereas the number of established scientists decreased from 4.01 in non-AI teams to 3.58 (10.77%) in AI teams ($t = 20.82$, $P < 0.001$ and $df > 10^3$ in t-test). This indicates that AI adoption primarily contributes to a reduction in the number of junior scientists in teams, whereas the decrease in the number of established scientists is relatively moderate. Given the decline in the number of junior scientists, we further calculate the probability of junior scientists becoming established scientists or leaving academia (Fig. 2c). Across all studied disciplines, the probability that AI-adopting junior scientists become established scientists is 45%, which is 13.64% higher than for their counterparts who do not adopt AI ($t \geq 1.40$, $P < 0.2$ and $df = 90$ in a t-test on four out of six disciplines). This indicates that AI-adopting scientists are associated with increased opportunities to lead research projects and reduced risks of dropping out of academia, thereby experiencing accelerated career transitions from junior to established scientists.

To further quantify this effect, we measure the accelerated career development of junior scientists by using a birth–death model[42] and fitting the model parameter λ with scientists' career trajectories (Fig. 2d and Methods). We find that the anticipated transition time to becoming established is 1.37 years shorter for AI-adopting junior scientists than for their counterparts who do not adopt AI. The expected transition time is 7.33 years for junior scientists who adopt AI ($R^2 = 0.995$) and 8.70 years for those who do not ($R^2 = 0.987$). This demonstrates how AI adoption affords junior scientists with opportunities to lead research projects and become established earlier. Further analysis reveals that this reduction in the transition time for AI-adopting junior scientists to become established is universal across examined disciplines (Extended Data Fig. 7). Moreover, established scientists involved in AI papers are, on average, 10.77% younger than those involved in non-AI papers (Extended Data Fig. 6; $t \geq 2.12$, $P < 0.05$ and $df > 10^3$ in a t-test on most year). Collectively, these findings suggest that AI research receives more attention from academia, and AI-adopting scientists are associated with higher scholarly productivity and impact. In this way, they have a higher probability of becoming established scientists, and at earlier ages, therefore experiencing accelerated career development.

AI contracts science's focus

The accelerating use of AI in science and its impact on individual scientists raises questions about its influence across the entire field of science. To evaluate how AI collectively impacts the frontiers of scientific exploration, we design a measurement to characterize the breadth of scholarly attention represented by a collection of research papers. We use SPECTER 2.0—a specialized text embedding model pre-trained on a large scientific literature corpus and fine-tuned with citation information[36]—to project research articles onto its 768-dimensional embedding space of science (Fig. 3a).

Within this high-dimensional embedding space, we measure knowledge extent as the 'diameter' of vector space covered by a sampled batch of papers, which allows us to compare the coverage of topical ground between AI and non-AI papers in each given domain[43,44] (Fig. 3b and Methods). Compared with conventional research, AI research is associated with a 4.63% contracted median collective knowledge extent across science, which is consistent across all six disciplines (Fig. 3c and Extended Data Fig. 8; $\chi^2 \geq 84.05$, $P < 0.001$ and $df = 1$ in a median test on any discipline). Moreover, when dividing these disciplines into more than two hundred sub-fields, the contraction of knowledge extent can be observed in more than 70% of them (Extended Data Fig. 9). When we compare the median entropy of knowledge distribution between AI and non-AI research in each domain (Fig. 3d), results demonstrate that the knowledge distribution of AI research has a lower entropy ($\chi^2 \geq 79.20$, $P < 0.001$ and $df = 1$ in a median test on any discipline), indicating an increasingly disproportionate focus on specific core problems within established fields.

These results generally highlight an emerging conflict between individual and collective incentives to adopt AI in science, where scientists receive expanded personal reach and impact, but the knowledge extent of entire scientific fields tends to shrink and focus attention on a subset of topical areas. According to analyses on possible factors that may influence the selectivity of AI adoption across different topics, we find that factors such as inherent topicality, original impact and funding priority remain almost unrelated to the disproportionate AI adoption (Supplementary Figs. 22–24). By contrast, data availability seems to be a major impacting factor, where areas with an abundance of data are increasingly and disproportionately amenable to AI research, contributing to the observed concentration within knowledge space (Supplementary Fig. 25).

AI reduces scientific engagement

To analyse mechanisms underlying the conflict between the growing influence of individual papers and researchers and the narrowing of domain knowledge within AI research, we examine the relationship between articles that cite AI and non-AI work. We first examine the knowledge extent of 'paper families', that is, a focal paper and its follow-on citations, which measures the size of the space covered by research derived from each original paper (Fig. 4a and Methods). Results show that the knowledge extent of AI papers' citation families is on average 3.46% more expanded than that of non-AI papers ($t \geq 1.91$, $P \leq 0.1$ and $df > 10^3$ in t-test on 30 out of 32 pairs of data). The contraction of knowledge space in AI research is therefore not attributable to the narrowing of knowledge space that can be derived from each original research work.

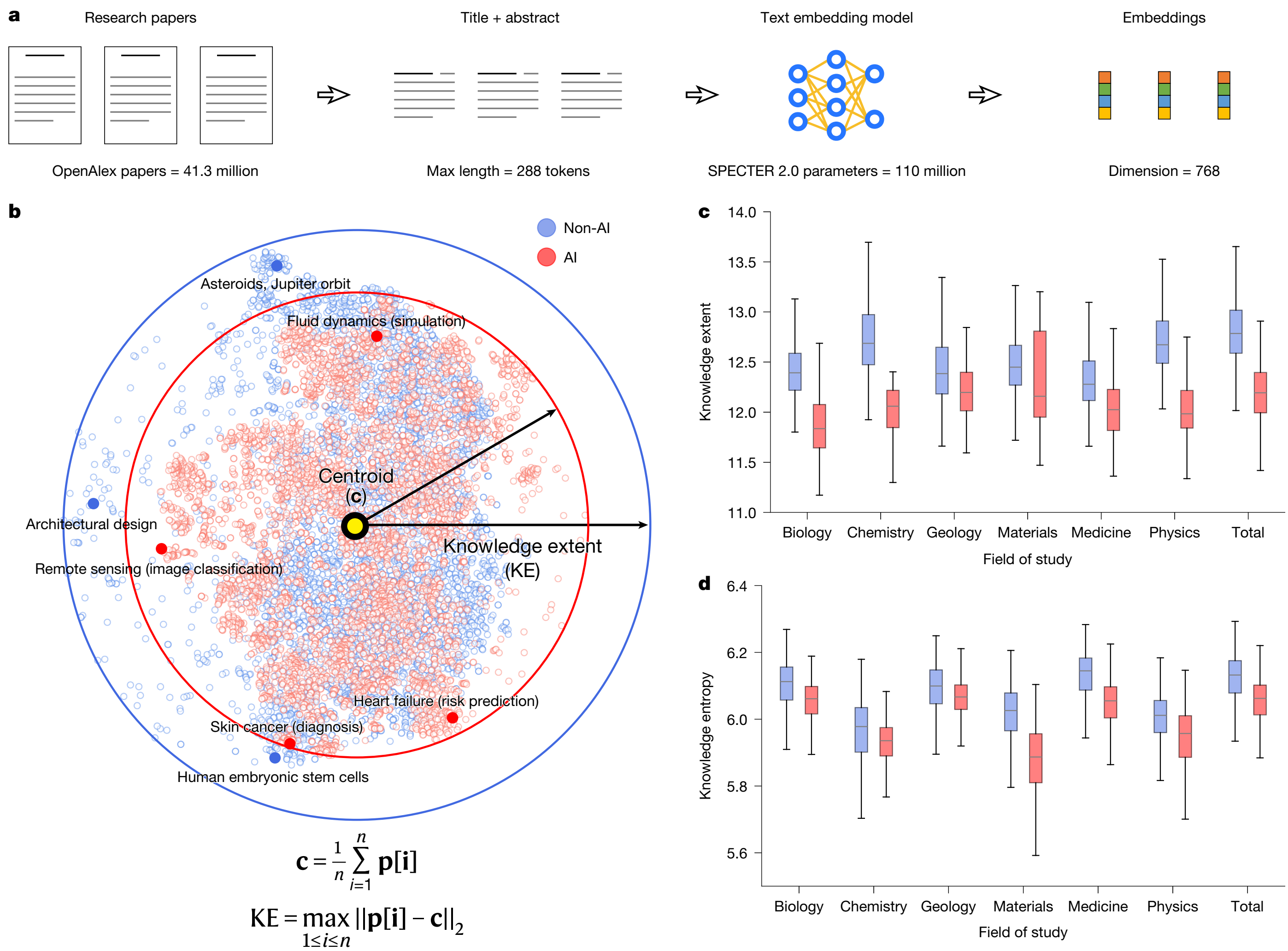


Fig. 3 | AI adoption is associated with a contraction in knowledge extent within and across scientific fields. a, We embed research papers into a 768-dimensional vector space with a pre-trained text embedding model; we then measure the knowledge extent of papers within that space. **b**, For visualization, we use the t-distributed stochastic neighbour embedding (t-SNE) algorithm to flatten the high-dimensional embeddings of a random batch of 10,000 papers (half of which are AI papers) into a two-dimensional plot. As shown by the solid arrows and circular boundaries, the knowledge extent of AI papers (calculated in the unflattened space) is smaller across the entirety of the natural sciences. Furthermore, AI papers are more clustered in knowledge space, indicating a higher concentration on specific problems. **c**, Knowledge extent of AI and non-AI papers in each field ($P < 0.001$, $n = 1{,}000$ samples in each field), where AI research focuses on a more contracted knowledge space. **d**, Knowledge entropy of AI and non-AI papers in each field ($P < 0.001$, $n = 1{,}000$ samples in each field), where AI research has a lower entropy. For panels **c** and **d**, boxplots are centred at the median and bounded at the first and third quartiles (Q1 and Q3), with 1.5 times the interquartile range shown as whiskers from the box. All statistical tests use a median-test.

To further investigate engagement, we examine relationships between papers by measuring the degree of follow-on paper engagement, namely, how frequently citations of the same original paper cite each other (Fig. 4b and Methods). Results demonstrate AI research spawns 22% less follow-on engagement ($t \geq 8.10$, $P < 0.001$ and $\mathrm{df} > 10^3$ in t-test on any discipline), suggesting that AI papers tend to only concentrate on the original paper, rather than forming dense interactions among each other, which is the characteristic of emerging fields[45]. This results in a star-like structure around specific popular research topics, rather than a network of emergent and interconnected research works. Further evidence of this concentration is found in the Matthew effect[46] among AI paper citations across different fields (Fig. 4c and Extended Data Fig. 10). In AI research, a small number of superstar papers dominate the field, with 22.20% of top papers receiving 80% of the citations and the top 54.14% receiving 95% of citations. This unequal distribution leads to a Gini coefficient of 0.754 in citation patterns surrounding AI research, higher than 0.690 for non-AI papers ($t = 27.86$, $P < 0.001$ and $\mathrm{df} = 198$ in t-test), signalling a disparity in recognition.

To further analyse the impact of reduced follow-on engagement, we sample 590,325,130 pairs of papers, where each pair cites the same original work. Among these, 51,723,984 pairs not only cite the same original work but also cite each other (engaged), whereas the remaining pairs do not cite each other (disengaged). We examine distances between these pairs of papers within our 768-dimensional vector space (Fig. 4d) and find that median distance between paper pairs that are disengaged from one another tends to be 18.11% larger than between paper pairs that are engaged with each other. By contrast, the closest disengaged paper pairs are 76.51% closer to one another than the closest engaged paper pairs. Taken together, a pair of disengaged papers commonly focus on less related topics and lie farther apart in the embedding space. Occasionally, however, owing to the lack of reciprocal engagement, it is possible that mutually unaware papers lie very close to each other, which indicates more overlapping research. These findings suggest that AI in science has become more concentrated around popular research topics that become ‘lonely crowds’ with reduced interaction among papers, linking to more

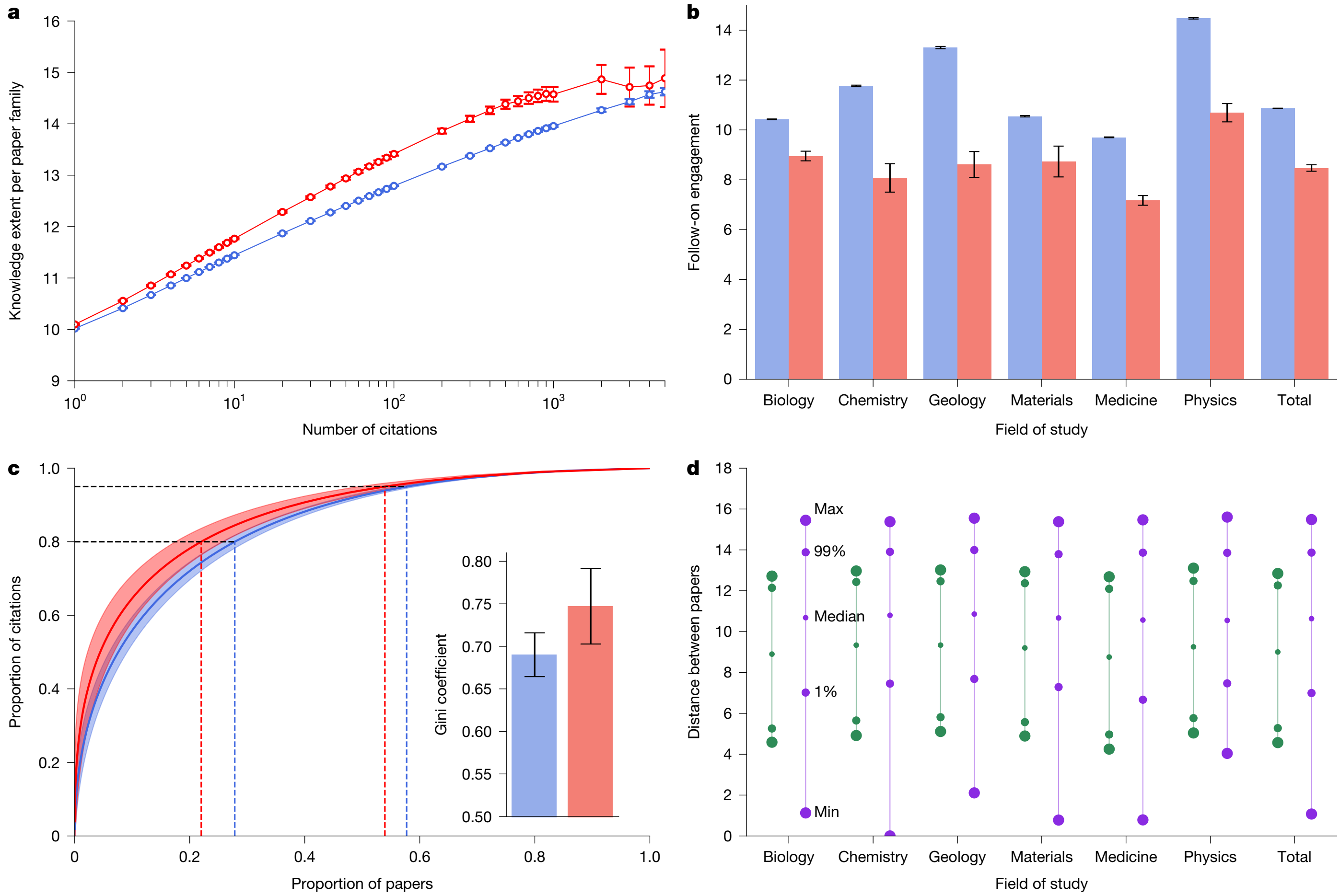


Fig. 4 | Reduced follow-on engagement and more overlapping works in AI research. **a**, Knowledge extent of individual AI (red) and non-AI (blue) paper families, that is, an original paper and its cumulative citations (n = 27,405,011), where the knowledge space of individual AI paper families is broader and grows faster. **b**, Engagement among papers that cite AI versus non-AI papers (P < 0.001, n = 23,342,516), where there are fewer follow-on interactions among papers that cite the same original paper in AI research. **c**, Distribution of citations to AI versus non-AI papers, where AI papers tend to concentrate more on a smaller number of top papers (P < 0.001, n = 100 sampled paper groups). **d**, Distribution of distances between paper pairs that cite the same previous research, with or without citing one another, namely, engaged (green) versus disengaged (purple) (n = 590,325,130 sampled paper pairs). Results show that for papers not engaged with each other, the median distance is larger, but the minimum distance is smaller, indicating a higher probability of overlaps in knowledge space. For all panels, 99% CIs are shown as error bars or error bands centred at the mean. All statistical tests use a two-sided t-test.

overlapping research and a contraction in knowledge extent and diversity across science.

Discussion

Here we perform a large-scale empirical measurement of the effect of adopting AI in science on both individual scientists and scientific communities. We identify three waves of AI adoption in science, which correspond to the dominance of machine learning, deep learning and generative AI, respectively. Each wave is marked with an accelerated AI adoption rate in research papers and authors. In all natural science research fields we studied, we find that individual scientists are increasingly rewarded with expanded academic impact and accelerated career development for incorporating AI assistance in research across each of these waves. On average, AI adoption helps individual scientists publish 3.02 times more papers, receive 4.84 times more citations and become team leaders 1.37 years sooner. This probably results from improved modelling and prediction of field-specific data, resulting in higher performance on recognized benchmarks. The substantial academic benefits of AI use may be a driving force behind its accelerated rate of adoption; however, we also find unintended consequences from the increased prevalence of AI-augmented research. In all fields, AI-augmented research focuses on a narrower scope of scientific topics and reduces the scientific engagement of follow-on research, leading to more overlapping research work that slows the expansion of knowledge. Further, with a greater concentration of collective attention to the same AI papers, the adoption of AI seems to induce authors to converge on the same solutions to known problems rather than create new ones.

These findings raise critical questions for science policy. What are the topics that are most likely to be left behind from AI-augmented research across fields? Those with less available data include critical scientific questions regarding the origins of natural phenomena, where data are necessarily reduced. Accelerating scientific activity under the light cast by highly visible, data-rich phenomena moves science away from many foundational questions and towards operational ones. By driving attention towards the most popular new developments, AI seems to drive problem solution over generation. These issues become particularly concerning in the face of calls to further increase support for AI-augmented science[47,48], coupled with the personal scientific incentives we observe. This could shift collective attention away from new and original questions that lack the data required for AI to demonstrate benefit. It is true that more overlapping attention and a contracted focus may benefit scientific replication and extension, accelerating the emergence of solid and practical solutions to core questions. Insofar as scientific discovery represents a vast and complex landscape, however, concentrating attention on the same developments may increase the likelihood that science becomes fixed on local maxima of scientific explanation and prediction rather than searching in a more broad, decoupled and diverse way.

Although our analysis provides new insight into AI's impact on science, clear limitations remain. Our identification approach—although validated by experts—misses subtle and unmentioned forms of AI use, and our focus on natural sciences excludes important domains in which AI adoption patterns may differ. Moreover, despite consistently suggestive evidence, we cannot fully identify the causal linkage between AI adoption and scientific impact. Nevertheless, our findings demonstrate that currently attributed uses of AI in science primarily augment cognitive tasks through data processing and pattern recognition. Looking forward, these findings illuminate a critical and expansive pathway for AI development in science. To preserve collective exploration in an era of AI use, we will need to reimagine AI systems that expand not only cognitive capacity but also sensory and experimental capacity[49,50], enabling and incentivizing scientists to search, select and gather new types of data from previously inaccessible domains rather than merely optimizing analysis of standing data. The history of major discoveries has been most consistently linked with new views on nature[51]. Expanding the scope of AI's deployment in science will be required for sustained scientific research and to stimulate new fields rather than merely automate existing ones.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-025-09922-y.

1. Wang, H. et al. Scientific discovery in the age of artificial intelligence. *Nature* **620**, 47–60 (2023).
2. Hopfield, J. J. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl Acad. Sci. USA* **79**, 2554–2558 (1982).
3. Hopfield, J. J. Neurons with graded response have collective computational properties like those of two-state neurons. *Proc. Natl Acad. Sci. USA* **81**, 3088–3092 (1984).
4. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
5. Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Commun. ACM*, **60**, 84–90 (2012).
6. Hinton, G. E. & Salakhutdinov, R. R. Reducing the dimensionality of data with neural networks. *Science* **313**, 504–507 (2006).
7. Hinton, G. E. Training products of experts by minimizing contrastive divergence. *Neural Comput.* **14**, 1771–1800 (2002).
8. Kuhlman, B. et al. Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**, 1364–1368 (2003).
9. Jumper, J. et al. Highly accurate protein structure prediction with alphafold. *Nature* **596**, 583–589 (2021).
10. Gao, J. & Wang, D. Quantifying the use and potential benefits of artificial intelligence in scientific research. *Nat. Human Behav.* **8**, 2281–2292 (2024).
11. Evans, J. A. Electronic publication and the narrowing of science and scholarship. *Science* **321**, 395–399 (2008).
12. Adıgüzel, T., Kaya, M. H. & Cansu, F. K. Revolutionizing education with AI: exploring the transformative potential of ChatGPT. *Contemp. Educat. Technol.* **15**, ep429 (2023).
13. Akgun, S. & Greenhow, C. Artificial intelligence in education: addressing ethical challenges in K-12 settings. *AI Ethics* **2**, 431–440 (2022).
14. Meskó, B. & Topol, E. J. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. *npj Digital Med.* **6**, 120 (2023).
15. Loh, H. W. et al. Application of explainable artificial intelligence for healthcare: a systematic review of the last decade (2011–2022). *Comput. Methods Prog. Biomed.* **226**, 107161 (2022).
16. Ahmed, I., Jeon, G. & Piccialli, F. From artificial intelligence to explainable artificial intelligence in industry 4.0: a survey on what, how, and where. *IEEE Trans. Indust. Inform.* **18**, 5031–5042 (2022).
17. Varadi, M. et al. Alphafold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucl. Acids Res.* **50**, D439–D444 (2022).
18. Degrave, J. et al. Magnetic control of tokamak plasmas through deep reinforcement learning. *Nature* **602**, 414–419 (2022).
19. Fawzi, A. et al. Discovering faster matrix multiplication algorithms with reinforcement learning. *Nature* **610**, 47–53 (2022).
20. Boiko, D. A., MacKnight, R., Kline, B. & Gomes, G. Autonomous chemical research with large language models. *Nature* **624**, 570–578 (2023).
21. Stokel-Walker, C. & Van Noorden, R. What ChatGPT and generative AI mean for science. *Nature* **614**, 214–216 (2023).
22. Gilson, A. et al. How does ChatGPT perform on the United States medical licensing examination? The implications of large language models for medical education and knowledge assessment. *JMIR Med. Educat.* **9**, e45312 (2023).
23. Salimi, A. & Saheb, H. Large language models in ophthalmology scientific writing: ethical considerations blurred lines or not at all? *Am. J. Ophthalmol.* **254**, 177–181 (2023).
24. Liang, W. et al. Mapping the increasing use of LLMs in scientific papers. In *Proc. 1st Conference on Language Modeling* (COLM, USA, 2024).
25. Hwang, T. et al. Can ChatGPT assist authors with abstract writing in medical journals? Evaluating the quality of scientific abstracts generated by ChatGPT and original abstracts. *PLoS ONE* **19**, e0297701 (2024).
26. Kobak, D., González-Márquez, R., Horvát, E.-Á. & Lause, J. Delving into LLM-assisted writing in biomedical publications through excess vocabulary. *Sci. Adv.* **11**, eadt3813 (2025).
27. Wojtowicz, Z. & DeDeo, S. *Undermining Mental Proof: How AI Can Make Cooperation Harder by Making Thinking Easier* Vol. 39, 1592–1600 (2025).
28. Frank, M. R., Wang, D., Cebrian, M. & Rahwan, I. The evolution of citation graphs in artificial intelligence research. *Nat. Mach. Intell.* **1**, 79–85 (2019).
29. OpenAlex (OpenAlex, 2025); https://openalex.org/.
30. *Clarivate* (Web of Science, 2025); https://www.webofscience.com.
31. Mongeon, P. & Paul-Hus, A. The journal coverage of web of science and scopus: a comparative analysis. *Scientometrics* **106**, 213–228 (2016).
32. Devlin, J. et al. BERT: pre-training of deep bidirectional transformers for language understanding. In *Proc. 57th Annual Meeting of the Association for Computational Linguistics* 4171–4186 (ACL, Italy, 2019).
33. Wolf, T. et al. Transformers: State-of-the-art natural language processing. In *Proc. 58th Annual Meeting of the Association for Computational Linguistics* 38–45 (ACL, 2020).
34. Beltagy, I., Lo, K. & Cohan, A. SciBERT: a pretrained language model for scientific text. In *Proc. 57th Annual Meeting of the Association for Computational Linguistics* 3613–3618 (ACL, Italy, 2019).
35. Cohan, A., Feldman, S., Beltagy, I., Downey, D. & Weld, D. S. SPECTER: document-level representation learning using citation-informed transformers. In *Proc. 58th Annual Meeting of the Association for Computational Linguistics* 2270–2282 (ACL, 2020).
36. Singh, A., D'Arcy, M., Cohan, A., Downey, D. & Feldman, S. SciRepEval: a multi-format benchmark for scientific document representations. In *Proc. 61st Annual Meeting of the Association for Computational Linguistics* 5548–5566 (ACL, Canada, 2023).
37. Landis, J. R. & Koch, G. G. The measurement of observer agreement for categorical data. *Biometrics* **33**, 159–174 (1977).
38. Fleiss, J. L. Measuring nominal scale agreement among many raters. *Psychol. Bull.* **76**, 378 (1971).
39. Chu, J. S. & Evans, J. A. Slowed canonical progress in large fields of science. *Proc. Natl Acad. Sci. USA* **118**, e2021636118 (2021).
40. *Journal Citation Reports* (Clarivate, 2021); https://jcr.clarivate.com/jcr/home.
41. Ioannidis, J. P., Boyack, K. W. & Klavans, R. Estimates of the continuously publishing core in the scientific workforce. *PloS ONE* **9**, e101698 (2014).
42. Kendall, D. G. Birth-and-death processes, and the theory of carcinogenesis. *Biometrika* **47**, 13–21 (1960).
43. Fortunato, S. et al. Science of science. *Science* **359**, eaao0185 (2018).
44. Milojević, S. Quantifying the cognitive extent of science. *J. Informetrics* **9**, 962–973 (2015).
45. McMahan, P. & Evans, J. Ambiguity and engagement. *Am. J. Sociol.* **124**, 860–912 (2018).
46. Merton, R. K. The matthew effect in science: the reward and communication systems of science are considered. *Science* **159**, 56–63 (1968).
47. Borger, J. G. et al. Artificial intelligence takes center stage: exploring the capabilities and implications of chatgpt and other AI-assisted technologies in scientific research and education. *Immunol. Cell Biol.* **101**, 923–935 (2023).
48. Lawrence, N. D. & Montgomery, J. Accelerating AI for science: open data science for science. *Royal Soc. Open Sci.* **11**, 231130 (2024).
49. King, R. D. et al. The automation of science. *Science* **324**, 85–89 (2009).
50. Burger, B. et al. A mobile robotic chemist. *Nature* **583**, 237–241 (2020).
51. Krauss, A. *Debunking Revolutionary Paradigm Shifts: Evidence of Cumulative Scientific Progress Across Science* Vol. 480, 20240141 (The Royal Society, 2024).

Methods

Dataset and paper selection

In this section we introduce the procedure of selecting the research papers included in our analysis. We conduct our major analyses on OpenAlex[29]—a scientific research database built on the foundation of the Microsoft Academic Graph (MAG)[52,53]. Supported by non-profit organizations, OpenAlex is continuously updated, providing a sustainable global resource for research information. As of March 2025, OpenAlex contains 265.7 million research papers, along with related data about citation, author, institution and so on. Among the massive quantity of papers in the OpenAlex dataset, we select 66,117,158 English research papers published in journals and conferences spanning from 1980 to 2025 and filter out those with incomplete titles or abstracts. We identify the scientific discipline each paper belongs to by making use of the topics noted in OpenAlex, which are extracted using a natural language processing approach that annotates titles and abstracts with Wikipedia article titles as topics sharing textual similarity. In the raw dataset, these topics form a hierarchical structure and each paper is associated with several. Adopting the 19 basic scientific disciplines in MAG[52,53], that is, art, biology, business, chemistry, computer science, economics, engineering, environmental science, geography, geology, history, materials science, mathematics, medicine, philosophy, physics, political science, psychology, and sociology, we trace along the hierarchy and determine to which disciplines each topic belongs. We note that because the original topics of one paper may be retraced to different topics, the scientific discipline of each paper may not be unique. In other words, one paper may span two or more academic disciplines, for example, chemistry and biology, which reflects the common phenomena of borderline or interdisciplinary research[54].

Here we emphasize the adoption of AI methods in conventional natural science disciplines and exclude research developing AI methodologies themselves, separating the influence of AI on science from AI's own invention and refinement. We therefore select biology, medicine, chemistry, physics, materials science and geology as representatives of natural science disciplines, but exclude computer science and mathematics, where most works introducing and developing AI methods are published. We also exclude art, business, economics, history, philosophy, political science, psychology and sociology to focus on how AI is changing the natural sciences and career trajectories in those sciences. Our six natural science disciplines include the majority of OpenAlex articles, resulting in 41,298,433 papers, containing 18,392,040 in biology, 4,209,771 in chemistry and 2,380,666 in geology, 4,755,717 in materials science, 24,315,342 in medicine and 5,138,488 in physics. The selected disciplines cover various dimensions of natural science, representing a broad view of scientific research as a whole.

Divide into three stages of AI development

We divide the history of AI development into three key eras: the traditional machine learning era (1980–2014), the deep learning era (2015–2022) and the generative AI era (2023 to present). We consider 1980 as the start of the traditional machine learning era because several landmark works were published in the 1980s, such as the back-propagating method[55,56]. We regard the deep learning era to have begun in 2015, as indicated by breakthroughs such as ResNet, which enabled the training of ultra-deep neural networks, revolutionizing fields such as computer vision and speech recognition[57]. Finally, we define the generative AI era as beginning in 2023, following the publication of ChatGPT—the first widely used large language model—in December 2022. This marked the advent of large-scale transformer-based models capable of strong generalized performance across a wide range of tasks, sparking new applications in natural language processing and beyond. Each of these transitions was driven by advances in algorithms, computational power and data availability, substantially expanding the capabilities and scope of AI for science.

Design and fine-tune the language model for AI paper identification

Insofar as both a paper's title and abstract contain important information about its content, we independently train two separate models on the basis of paper titles and abstracts, and then integrate the two models into an ensembled one by averaging their outputs. The structure of our natural language processing model for paper identification consists of two parts. The backbone network is a twelve-layer BERT model with twelve attention heads in each layer, and the sequence classification head is a linear layer with a two-dimensional output atop the BERT output. We normalize the two-dimensional output with a softmax function and obtain the probability that the paper involves AI-assistance. We use the BERT model called BERT-base-uncased from Hugging Face[58], which is pre-trained with a large-scale general corpus, and set the maximum length of tokenization to be 16 for titles and 256 for abstracts. We design a two-stage fine-tuning process with training and validation sets, which we extracted from the OpenAlex dataset, to transfer the pre-trained model to our paper identification task. The construction of positive and negative data is different between the two stages. In both stages, we randomly split the positive and negative data into 90% and 10% sets, which correspond to training and validation sets, respectively. We use the training set for model training and use the validation set to select the optimal model. As the numbers of positive and negative cases are unbalanced, we use the bootstrap sample technique on positive cases to balance its number with negative cases at both stages.

In the first stage, we construct relatively coarse positive data, only considering eight typical AI journals and conferences, including *Nature Machine Intelligence*, *Machine Learning*, *Artificial Intelligence*, *Journal of Machine Learning Research* (JMLR), *International Conference on Machine Learning* (ICML), *International Conference on Learning Representations* (ICLR) and the *AAAI Conference on Artificial Intelligence* and *International Joint Conference on Artificial Intelligence* (IJCAI). Among the papers belonging to our chosen six disciplines, we extract all papers published in these venues as positive cases and randomly sample 1% of the remaining papers in our six chosen natural science fields as negative cases, resulting in 26,165 positive and 291,035 negative cases. We fine-tune the pre-trained model for 30 epochs on the training set and select the optimal model according to the F1-score on the validation set.

In the second stage, we construct more precise positive data on the basis of the optimal model obtained in the first stage. We identify papers in the whole OpenAlex dataset and aggregate the results for each venue, obtaining the probability that each venue in OpenAlex is an AI venue by averaging the AI probability for all papers within it. We then select the venues with >80% AI probability and >100 papers as AI venues. We also incorporate venues with 'machine learning' or 'artificial intelligence' in their names. In papers belonging to our six chosen disciplines, we extract all papers published in the selected AI venues as positive cases and randomly sample 1% of those remaining as negative cases, resulting in 31,311 positive and 231,258 negative cases. We then fine-tune the obtained optimal model in the first stage for another 30 epochs with the new training set and select the best model according to F1-score on the new validation set. Finally, we use optimal ensemble models during both stages to identify all papers that use AI to support natural science research from the selected representative natural science disciplines.

Scrutinization of our identification results by disciplinary experts

We arbitrarily sample 220 papers (110 papers × 2 groups) from each of the six disciplines, resulting in twelve paper groups in total. We enlisted twelve experts with abundant AI research experience (Supplementary Table 1) and assigned three different groups of papers to each. Without revealing the classification results obtained from the BERT model, we queried our experts on whether each paper was an

AI paper. In this way, each paper is repeatedly labelled by three distinct experts, and we evaluate the consistency among these experts on the basis of Fleiss's κ (refs. 37,38), which is an unsupervised measurement for assessing the agreement between independent raters. Having confirmed consensus among our experts, we draw the final expert label of each paper from the three experts according to the principle of the minority obeying the majority. We regard the expert labels as ground truth and validate the result of our BERT model against them with the F1-score, which is a supervised measurement of accuracy.

Determine the project leader of papers

Here we define the project leader as the last author of a research paper, in alignment with conventions established by previous studies[59]. To ensure that in most papers, the last author represents the project leader, we examine the fraction of papers that list authors following alphabetical order. First, we directly traverse all selected papers and obtain the prevalence of papers listing authors in alphabetical order, which ranges from 14.87% in materials science to 22.15% in geology. Nevertheless, it is difficult to distinguish whether these papers actually intended to list the authors in alphabetical order or according to their roles, which unintentionally fall in alphabetical order. The latter situation is more likely to occur when there are fewer authors (two or three). To tackle this analytical challenge, we determine the fraction of unintended alphabetical author lists through a Monte Carlo method. We generate ten randomly shuffled copies of the author list for each paper and find that from 13.82% (materials science, $\sigma = 0.02$) to 20.28% (geology, $\sigma = 0.03$) of papers have alphabetically listed authors among the random author lists. This indicates the proportion of 'unintended' alphabetical author lists, and we can derive the actual fraction of papers with intentionally alphabetical author lists by the difference between the above two sets of statistical results. The actual fraction obtained illustrates that only 1.58% of papers across all disciplines intentionally list the authors in alphabetical order (Supplementary Table 12) and therefore, we can, with negligible interference, assume that we can identify last authors as team leaders.

Detect scientists' career role transition

The OpenAlex dataset incorporates a well-designed author name disambiguation mechanism[29], which uses an XGBoost model[60] to predict the likelihood that two authors are the same on the basis of features such as their institutions, co-authors and citations, and then applies a custom, ORCID-anchored clustering process to group their works, assigning a unique ID to each author. Simply using unique IDs, we are able to track a large number of authors at the same time[61], where we depict an individual scientist's career trajectory using a role transition model (Extended Data Fig. 4a) and extract the role transition trajectories for scientists.

First we traverse all selected papers in the six disciplines and extract all the scientists involved in any of these papers. Then, for each individual scientist, we extract all papers in which they have been involved and record the time of their first publication in any role, the time of their first publication as team leader (if ever), and the time of their last publication. We then filter out scientists whose publication records span only a single year. We also filter out those who directly start as established scientists leading research teams without a role transition from junior scientists. Finally, we detect the time that each scientist abandons academic publishing. Considering that one scientist may not publish papers continuously every year, we cannot regard them as having left academia on the basis of their absence in the published record for a single year. We therefore follow the settings used in previous work[62] to use a threshold of three years and regard scientists who have no more publications after 2022 as having exited academia, whereas those who still publish papers after 2022 are considered to have an unclear ultimate status and are excluded from the analysis. Finally, we obtain 2,282,029 scientists in the six disciplines with complete role transition trajectories. We also classify them into AI and non-AI scientists according to whether they have published AI-augmented papers.

Moreover, by analysing author contribution statements collected in previous studies[63,64], we further validate our detection results by examining changes in scientists' self-reported contributions throughout their careers (Extended Data Fig. 4b). Results indicate that junior scientists primarily engage in technical tasks, such as conducting experiments and analysing data, and less in conceptual tasks, such as conceiving ideas and writing papers. Nevertheless, the proportion of conceptual work significantly rises ($P < 0.01$ and df = 1 in a Cochran–Armitage test) during their tenure as junior scientists, reaching saturation at a high level (60% or more) on transition to becoming established scientists. This finding validates our definition of role transition by demonstrating a shift in the nature of scientists' contributions from participating in research projects to leading them.

Estimate the birth–death model for career development of junior scientists

To obtain a more precise quantification of how much AI accelerates the career development of junior scientists, we use a general birth–death model[42]. This type of stochastic process model depicts the dynamic evolution of a population as members join and exit. In our context, it models the role transitions of junior scientists. Specifically, we use two separate birth–death models for junior scientists who eventually become established and for those who leave academia, respectively. Here, 'birth' processes refer to the entry of junior scientists into academic publishing, and 'death' processes symbolize their transition out of the junior stage, either by becoming established scientists or quitting academia. As the entry and exit of each junior scientist are independent from one another, we use Poisson processes to model 'birth' (entry) and 'death' (exit) events, respectively.

The Poisson process is a typical stochastic process model for describing the occurrence of random events that are independent of each other[65]. The mathematical formula of the Poisson process is:

$$P(N(t_0) = k) = \frac{(\lambda_0 t_0)^k}{k!} e^{-\lambda_0 t_0}, t_0 > 0, k = 0, 1, 2, \ldots, \tag{1}$$

where $N(t_0)$ denotes the number of random events that happened before time t_0, and λ_0 is the parameter of the Poisson process, depicting the happening rate of random events. We consider a birth–death model in which birth and death dynamics are both Poisson processes, and rate parameters are μ and ω, respectively. Through mathematical derivation[66], we conclude that the duration time t from birth to death follows an exponential distribution with the parameter $\omega - \mu$, where the exact form of the probability density function is:

$$P(t) = (\omega - \mu) e^{-(\omega - \mu)t}, t > 0. \tag{2}$$

We consider the difference between the two rate parameters $\omega - \mu$ as a whole and fit it with a single parameter λ. The transition time for junior scientists to become established scientists or leave academia then follows the exponential distribution:

$$P(t) = \lambda e^{-\lambda t}, t > 0, \tag{3}$$

and the corresponding survival function is

$$S(t) = 1 - \int_0^t P(u) du = e^{-\lambda t}, t > 0. \tag{4}$$

Hence the average transition time is the conditional expectation of the distribution defined as follows:

$$\bar{t} = E[t|t > 1] = \int_1^\infty t \times P(t)dt = \int_1^\infty t \times \lambda e^{-\lambda t}dt = \frac{1}{\lambda} + 1. \quad (5)$$

We fit the role transition time of scientists with the aforementioned exponential distribution, thereby determining the respective values of λ for AI-adopting junior scientists and their non-AI counterparts. Guided by the underlying mechanism of junior scientists' career development incorporated within the birth–death model, expectations from the model offer a more accurate estimate of the average role transition time.

Measure the knowledge extent of papers

To assess the knowledge extent of a set of research papers within their high-dimensional embeddings

$$\{\mathbf{p[1]}, \mathbf{p[2]}, \ldots, \mathbf{p[n]}\}, \mathbf{p[i]} \in \mathbb{R}^{768}, \quad (6)$$

we first compute the centroid as the mean of their vector locations:

$$\mathbf{c} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{p[i]}. \quad (7)$$

Next, we compute the Euclidean distance from each embedding to the centroid, where the knowledge extent (KE) of the set of papers is defined as the maximum distance or 'diameter' of the vector space covered:

$$\mathrm{KE} = \max_{1 \le i \le n} \|\mathbf{p[i]} - \mathbf{c}\|_2. \quad (8)$$

We note that Euclidean distance is highly correlated with the cosine and related angular distances.

In practice, the number of AI and non-AI papers in each domain differs considerably, introducing bias to the measurement of knowledge extent. To address this issue, we build on past work[44] about cognitive extent, which is a measure of the breadth of a scientific field's cognitive territory, and is quantified by the number of unique phrases—as a proxy for scientific concepts—found within a sampled batch of papers of a given size. For each domain, we randomly sample 1,000 papers from both AI and non-AI categories, compute their respective knowledge extent, and repeat this process 1,000 times. By comparing knowledge extent values across these 1,000 random samples, we ensure that the number of AI and non-AI papers is balanced, making our knowledge extent results comparable.

Measure the knowledge extent of paper families

To measure how much knowledge space can be derived from each original paper, we calculate the knowledge extent of 'paper families', that is, a focal paper and its follow-on citations. Focusing on an original research paper ϕ, which corresponds to a high-dimensional embedding vector $\mathbf{p}_\phi \in \mathbb{R}^{768}$, we extract all n_ϕ research papers that cite this original paper. These papers are sorted chronologically by publication date, from earliest to most recent. The corresponding high-dimensional embeddings of these sorted papers are:

$$\{\mathbf{p}_\phi\mathbf{[1]}, \mathbf{p}_\phi\mathbf{[2]}, \ldots, \mathbf{p}_\phi\mathbf{[n}_\phi\mathbf{]}\}, \mathbf{p}_\phi\mathbf{[i]} \in \mathbb{R}^{768}. \quad (9)$$

Thereby, we calculate knowledge extent covered by the 'paper family' consisting of the original paper ϕ and the first n follow-on papers, citing it ($1 \le n \le n_\phi$) as:

$$\mathrm{KE}_\phi[n] = \max_{1 \le i \le n \le n_\phi} \|\mathbf{p}_\phi\mathbf{[i]} - \mathbf{p}_\phi\|_2. \quad (10)$$

Measure follow-on engagement among papers

To quantify how frequently citations of the same original paper interact with each other, we design a metric called follow-on engagement, building on previous work[45]. For an original paper with n citations, there are at most $\frac{n(n-1)}{2}$ possible citations among these n citing papers if everyone cites all papers published earlier than their own. We then count how many times these n citing papers actually cite one another, denoted as k. Our metric for follow-on engagement (EG) is calculated as the ratio of actual to maximum possible citations:

$$\mathrm{EG} = \frac{k}{\frac{n(n-1)}{2}} = \frac{2k}{n(n-1)} = \frac{2k}{n(n-1)} \times 100\,(\%). \quad (11)$$

This metric helps quantify the degree of interactions and collaboration among papers that cite the same original work. Past work has demonstrated a positive association between the ambiguity of a focal work and follow-on engagement[45].

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The OpenAlex dataset for research papers and researchers is available at https://docs.openalex.org/download-all-data/openalex-snapshot. The Web of Science dataset for research papers and researchers is available at https://clarivate.com/academia-government/scientific-and-academic-research/research-discovery-and-referencing/web-of-science/web-of-science-core-collection. The Journal Citation Report dataset for the journal quantile is retrieved from https://jcr.clarivate.com/jcr/browse-journals. The author contribution dataset is available at https://zenodo.org/records/6569339. The pre-trained parameters for the BERT language model are available at https://huggingface.co/docs/transformers. The pre-trained parameters for the SPECTER 2.0 text embedding model are available at https://huggingface.co/allenai/specter2. Source data are provided with this paper.

Code availability

This study used Python 3.11.0 with software packages to conduct data analysis. Required packages are NumPy (v.1.26.4), pandas (v.2.2.3), SciPy (v.1.15.2), scikit-learn (v.1.6.1) and matplotlib (v.3.10.1). The t-SNE algorithm used is imported from the sklearn package. The codes developed in this study can be found at https://github.com/tsinghua-fib-lab/AI-Impacts-Science.

52. *Microsoft Academic Graph* (Microsoft, 2015); https://www.microsoft.com/en-us/research/project/microsoft-academic-graph.
53. *Open Academic Graph* (Aminer, 2020); https://old.aminer.cn/oag-2-1.
54. Porter, A. & Rafols, I. Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. *Scientometrics* **81**, 719–745 (2009).
55. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagating errors. *Nature* **323**, 533–536 (1986).
56. LeCun, Y. et al. Backpropagation applied to handwritten zip code recognition. *Neural Comput.* **1**, 541–551 (1989).
57. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual Learning for image recognition. In *CVPR'16: Proc. 2016 IEEE conference on computer vision and pattern recognition* 770–778 (2016).
58. Face, H. *Bert for Sequence Classification* (Hugging Face, 2025); https://huggingface.co/docs/transformers/model_doc/bert#transformers.BertForSequenceClassification.
59. Sekara, V. et al. The chaperone effect in scientific publishing. *Proc. Natl Acad. Sci. USA* **115**, 12603–12607 (2018).
60. Chen, T. & Guestrin, C. XGBoost: a scalable tree boosting system. In *KDD'16: Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794 (2016).
61. Hill, R. et al. The pivot penalty in research. *Nature* **642**, 999–1006 (2025).
62. Milojević, S., Radicchi, F. & Walsh, J. P. Changing demographics of scientific careers: the rise of the temporary workforce. *Proc. Natl Acad. Sci. USA* **115**, 12616–12623 (2018).
63. Xu, F., Wu, L. & Evans, J. Flat teams drive scientific innovation. *Proc. Natl Acad. Sci. USA* **119**, e2200927119 (2022).
64. Lin, Y., Frey, C. B. & Wu, L. Remote collaboration fuses fewer breakthrough ideas. *Nature* **623**, 987–991 (2023).
65. Kingman, J. F. C. *Poisson Processes* Vol. 3 (Clarendon, 1992).

66. Meisling, T. Discrete-time queuing theory. *Operat. Res.* **6**, 96–105 (1958).

Acknowledgements This work was supported in part by the National Natural Science Foundation of China (grant no. U23B2030, 23IAA02114 and 62472241), the joint project of Infinigence AI & Tsinghua University, and Tsinghua University-Toyota Research Institute to Y. L. and F.X. J.E. received support from Novo Nordisk Foundation (Simulations of Science for Society), NSF (grant no. 2404109) and the United States Department of Defense (Defense Advanced Research Projects Agency - Modeling and Measuring Scientific Creativity). The funders had no role in study design, data collection, analysis, preparation of or decision to publish the manuscript.

Author contributions F.X., Y.L. and J.E. jointly launched this research and designed the research outline. Q.H. analysed the data and prepared the figures. All authors jointly participated in writing and revising the manuscript.

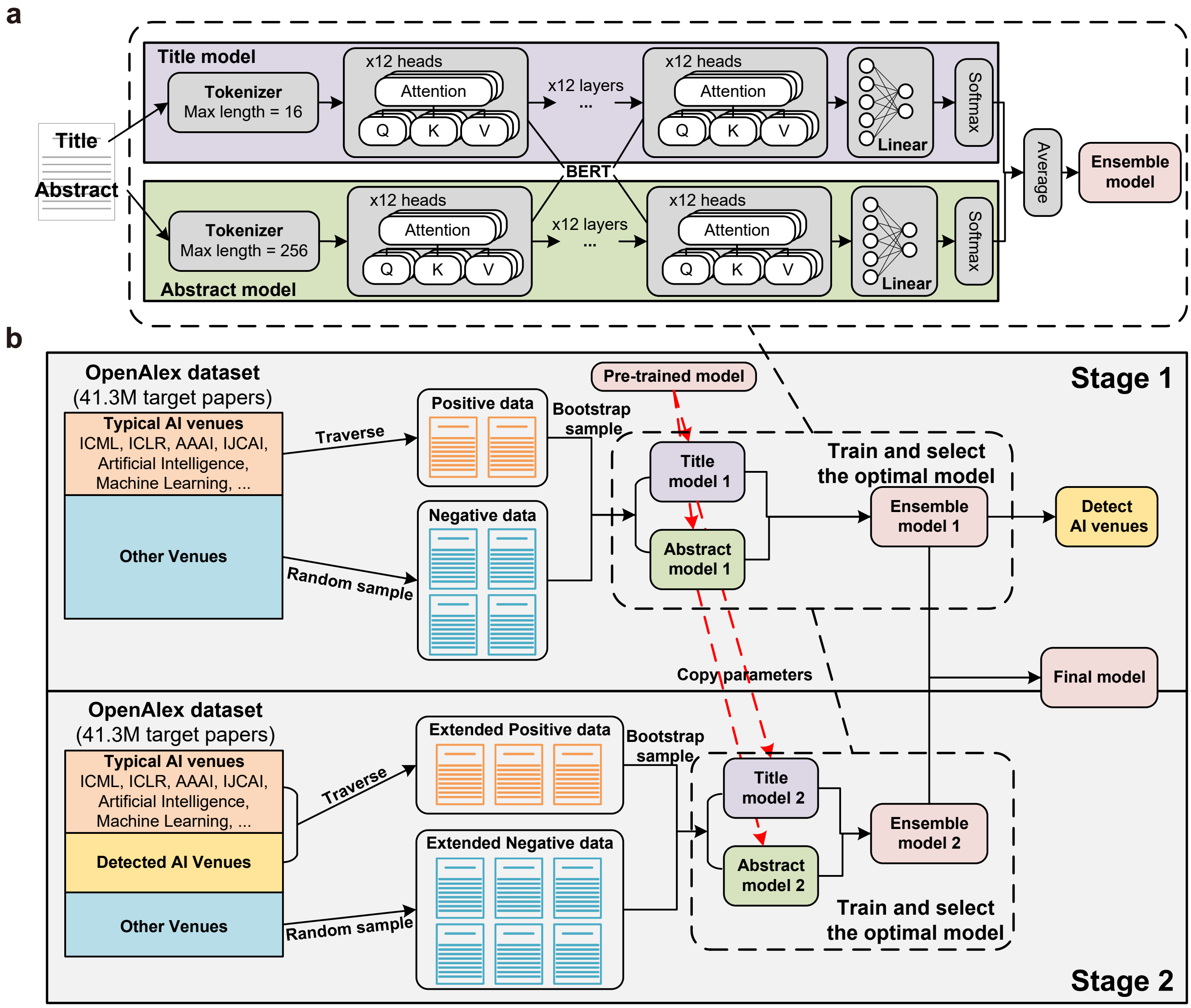
Competing interests J.E. has a commercial affiliation with Google, but Google had no role in the design, analysis, or decision to publish this study. The authors declare no other competing interests.

Additional information

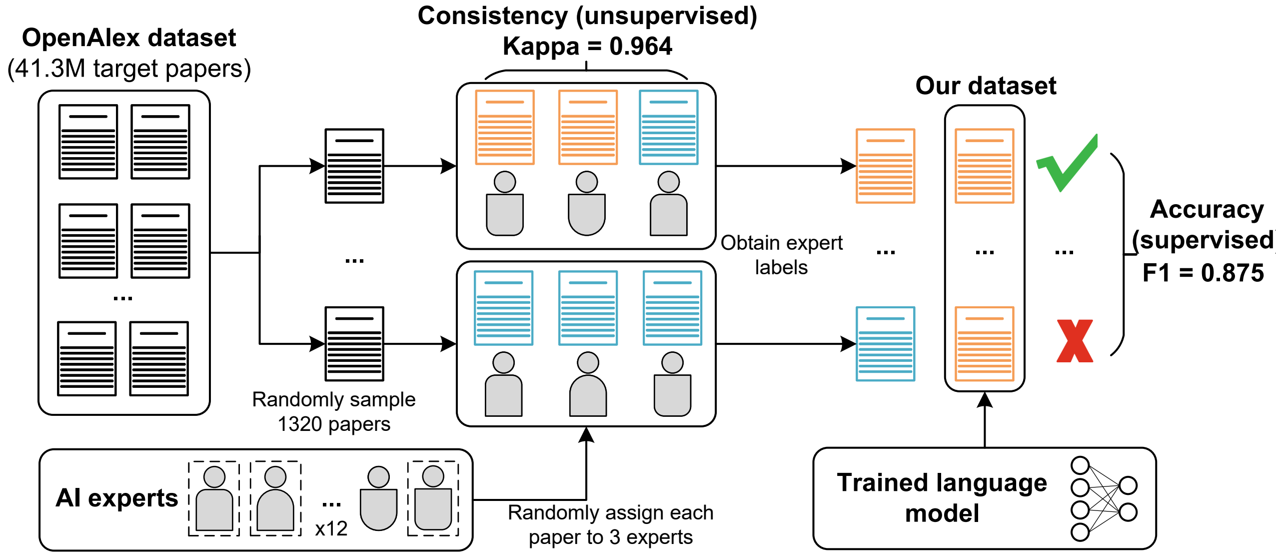
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-025-09922-y.

Correspondence and requests for materials should be addressed to Fengli Xu, Yong Li or James Evans.

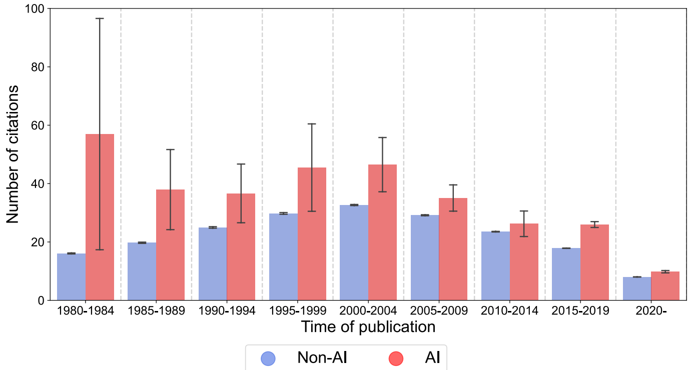
Peer review information *Nature* thanks Giovanni Colavizza, Luis Nunes Amaral, Catherine Shea, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.



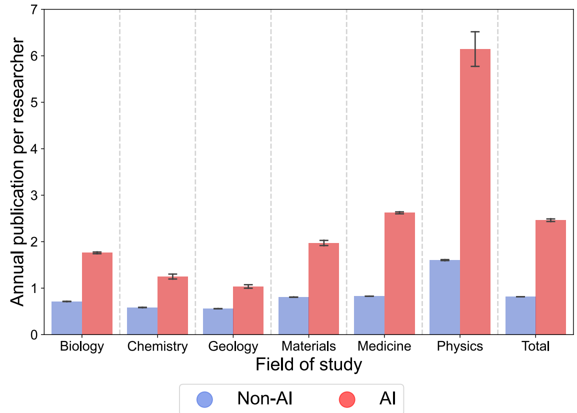
Extended Data Fig. 1 | Illustration for the method of identifying AI usage in research papers with fine-tuned language models. (**a**) Structure of our deployed language model, which consists of the tokenizer, the core BERT model, and the linear layer. (**b**) Procedure of the two-stage model fine-tuning process, where we design specific approaches for constructing positive and negative data at each stage.



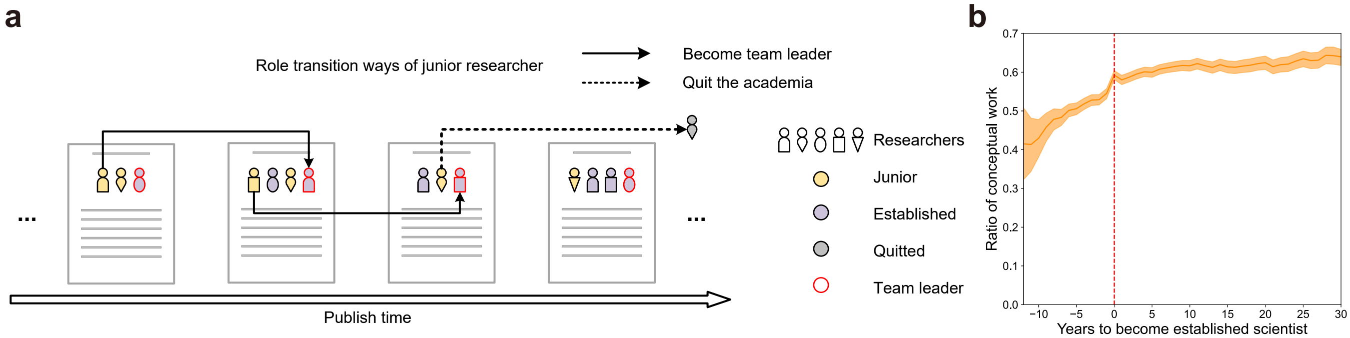
Extended Data Fig. 2 | Procedure of accuracy evaluation via expert evaluation. We randomly sample 1320 papers and delegate three experts to scrutinize the identification results for each paper. We then draw the final expert label of each paper from the three experts according to the principle of the minority obeying the majority and validate the result of the language model with it. Results indicate strong consistency among experts and high accuracy with our identification results.



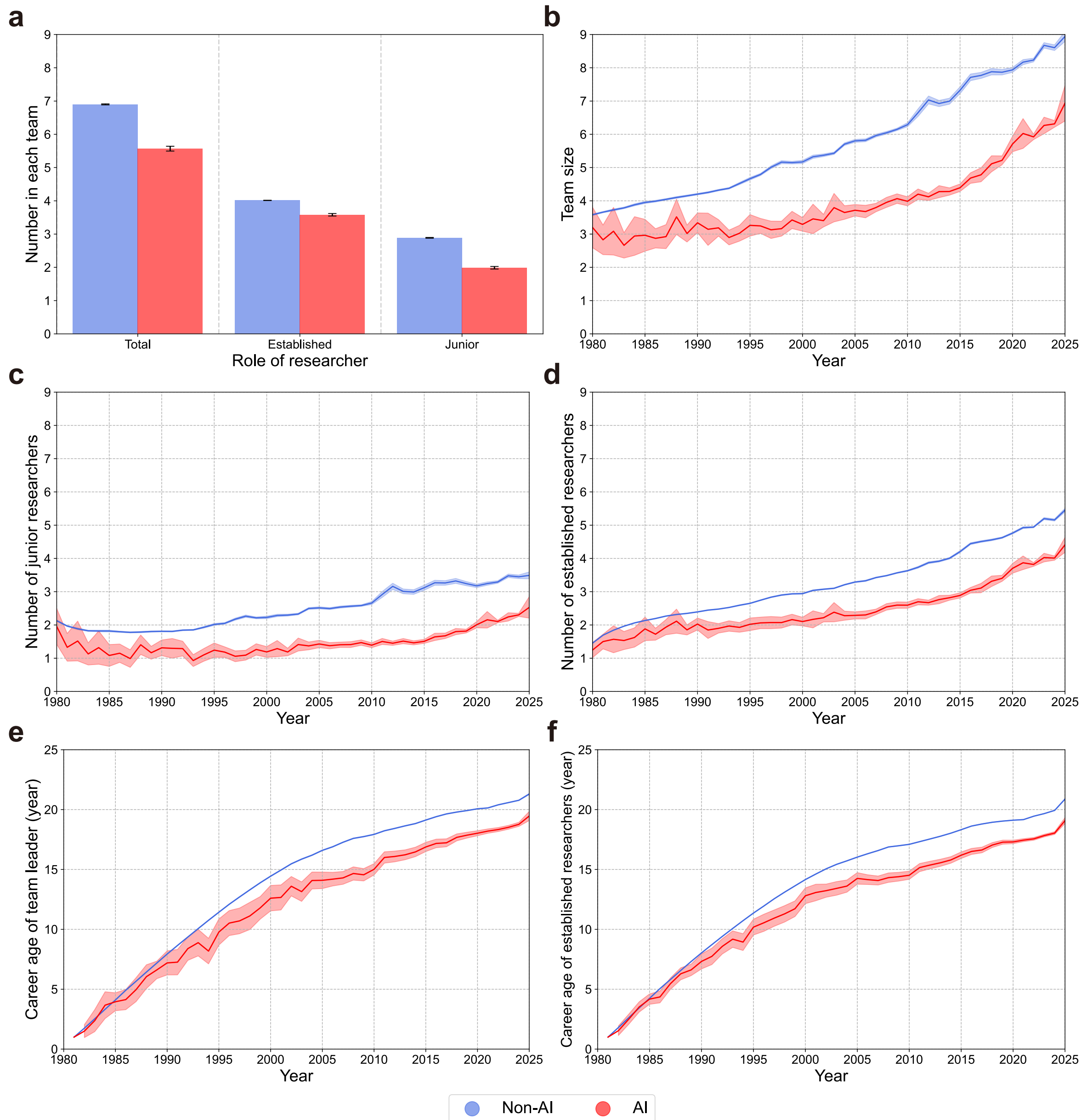
Extended Data Fig. 3 | Comparison of the total citations of AI and non-AI papers published in different eras. Results show that AI papers consistently attract more citations over different eras ($P < 0.001$, $n = 27{,}405{,}011$), indicating a higher academic impact than non-AI papers. 99% CIs are shown as error bars centred at the mean, and the statistical tests use a two-sided t-test.



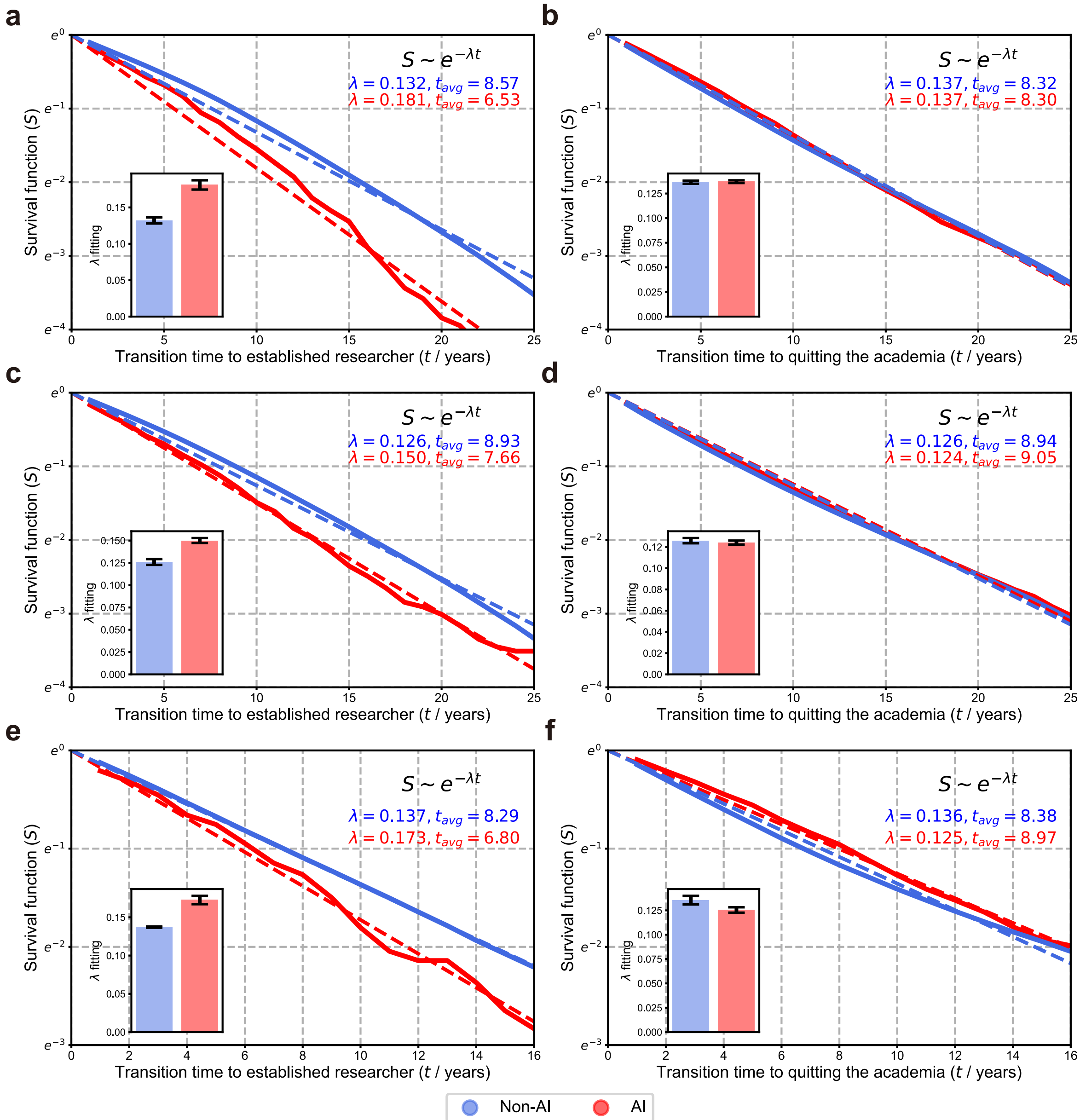
Extended Data Fig. 4 | Annual publications of researchers adopting AI and their counterparts without AI. Results show that in all 6 scientific disciplines, researchers adopting AI are more productive than their counterparts without AI ($P < 0.001$, $n = 5{,}377{,}346$). On average, researchers adopting AI annually publish 3.02 times more papers compared with those not using AI. 99% CIs are shown as error bars centred at the mean, and the statistical tests use a two-sided t-test.



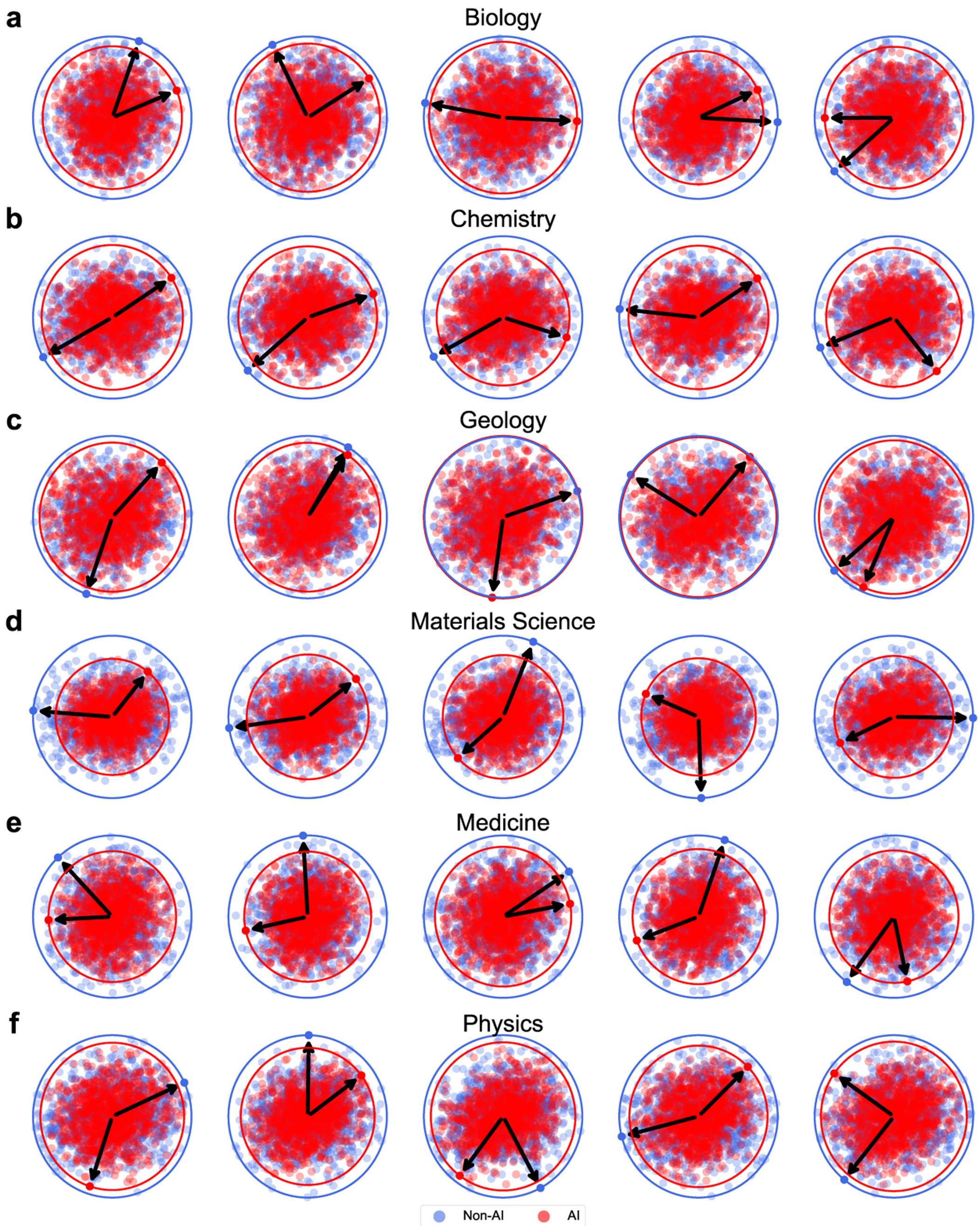
Extended Data Fig. 5 | Scientists' career role transition. (**a**) The career role transition of researchers. We consider the last author of each paper as research project leader and researchers who have been research project leaders as established researchers. Researchers who have yet to lead a research project are junior researchers, and they have two potential role transition pathways in the future: (1) become established researchers (solid arrow), and (2) abandon academia (dashed arrow). (**b**) Change in the ratio of conceptual work across the research career, before and after becoming an established researcher. The ratio increases rapidly before the role transition to established researchers, while it remains stable and high after that transition. 99% CIs are shown as error bands centred at the mean.



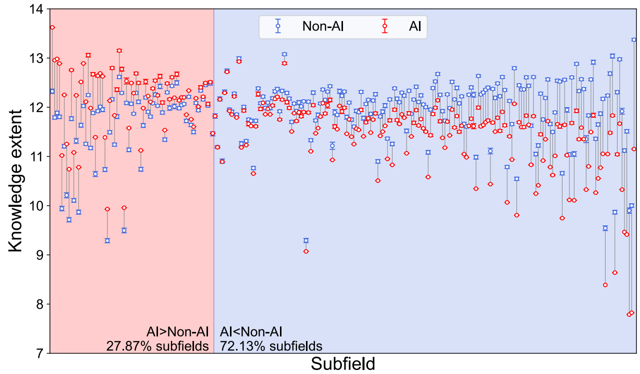
Extended Data Fig. 6 | Team composition of AI and non-AI papers. (**a**) AI research is associated with reduced research team sizes, averaging 1.33 fewer scientists ($P < 0.001$, $n = 33{,}528{,}469$). Specifically, the average number of junior scientists decreased from 2.89 in non-AI teams to 1.99 in AI teams (31.14%), while the number of established scientists decreased from 4.01 to 3.58 (10.77%). (**b**)-(**d**) Change in team size, average number of junior researchers, and average number of established researchers. These findings indicate that within the overall trend of increasing size of scientific research teams, AI adoption primarily contributes to a reduction in the number of junior scientists in teams, while a decrease in the number of established scientists is more moderate. (**e**) The average career age of team leaders in AI and non-AI papers. (**f**) The average career age of all involved established researchers in AI and non-AI papers. Results indicate that AI accelerates the transition from junior to established scientists, enabling AI-adopted researchers to become established at a younger age than those without AI. For all panels, 99% CIs are shown as error bars or error bands centred at the mean. All statistical tests use a two-sided t-test.



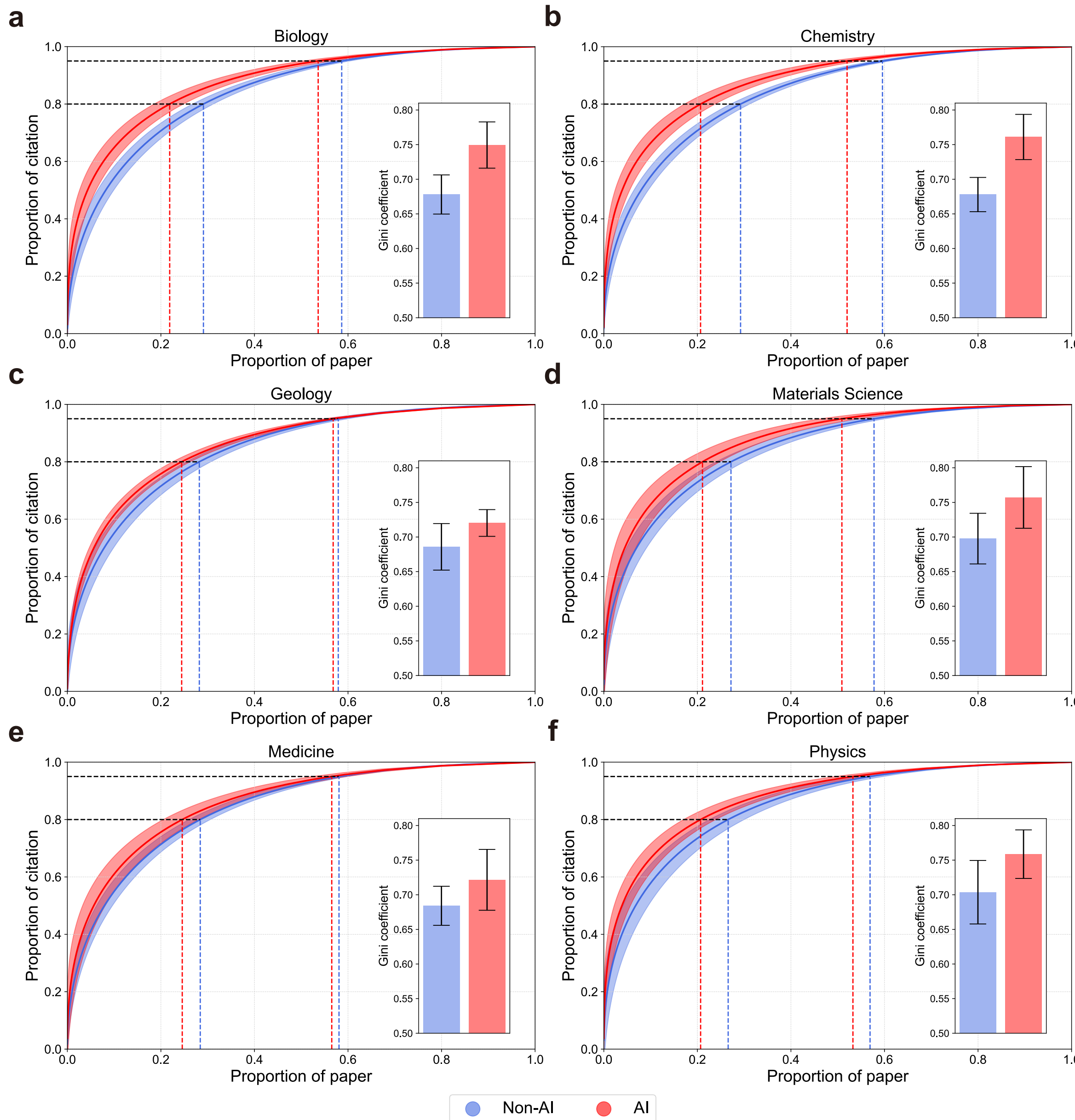
Extended Data Fig. 7 | Model fitting the role transition time of junior scientists. (**a**) (**c**) (**e**) Survival functions for the transition from junior to established researcher in (**a**) biology ($n = 625{,}093$), (**c**) medicine ($n = 1{,}137{,}076$), and (**e**) physics ($n = 120{,}366$). (**b**) (**d**) (**f**) Survival functions for the transition from junior researcher to leave academia in (**b**) biology ($n = 625{,}093$), (**d**) medicine ($n = 1{,}137{,}076$), and (**f**) physics ($n = 120{,}366$). All survival functions can be well-fit with exponential distributions, where the expected time for junior scientists to become established is shorter for those who adopt AI ($P < 0.001$), while the expected time for junior scientists to abandon academia is similar or slightly longer for those who adopt AI. Results indicate that AI not only provides junior scientists opportunities to become established scientists at a younger age, but also reduces the risk of their exiting academia early. For all panels, 99% CIs are shown as error bars centred at the mean. All statistical tests use a two-sided t-test.



Extended Data Fig. 8 | The knowledge extent of AI and non-AI papers. Here we visualize the embeddings of a small random sample of 2,000 papers, half of which are AI papers and half are non-AI papers. To eliminate randomness introduced by the t-SNE algorithm, here we simply pick out the first two dimensions of the high-dimensional embeddings to flatten them into a 2-D plot, and we provide 5 different random batches for each field to ensure robustness. As shown by the solid arrows and circular boundaries, the knowledge extent of AI papers is smaller than that of a comparable sample of non-AI papers, which is consistent across the fields studied in our analysis.



Extended Data Fig. 9 | The knowledge extent of AI and non-AI papers in each subfield. Compared with conventional research, AI research is associated with a shrinkage in the collective knowledge extent of science, where the contraction of knowledge extent can be observed in more than 70% of over two hundred sub-fields (n = 1,000 samples in each subfield). For all subfields, 99% CIs are shown as error bars centred at the mean.



Extended Data Fig. 10 | The Matthew effect in citations to AI and non-AI papers. In AI research, a small number of superstar papers dominate the field, with approximately 20% of top papers receiving 80% of citations and 50% receiving 95%. This unequal distribution leads to a higher Gini coefficient in citation patterns surrounding AI research ($P < 0.001$, $n = 100$ sampled paper groups for each discipline). Such disparity in the recognition of AI papers is consistent across all fields examined. For all panels, 99% CIs are shown as error bars or error bands centred at the mean. All statistical tests use a two-sided t-test.

nature portfolio

Corresponding author(s): Fengli Xu, Yong Li, James Evans

Last updated by author(s): Nov 4, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	The study directly downloaded data snapshot from the websites.
Data analysis	This study used python 3.11.0 with software packages to conduct data analysis. Required packages are numpy==1.26.4, pandas==2.2.3, scipy==1.15.2, sklearn==1.6.1, matplotlib==3.10.1. The used t-SNE algorithm is imported from the sklearn package. The codes developed in this study can be found at https://github.com/tsinghua-fib-lab/AI-Impacts-Science.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The OpenAlex dataset for research papers and researchers is available at https://docs.openalex.org/download-all-data/openalex-snapshot.

The Web of Science (WoS) dataset for research papers and researchers is available at https://clarivate.com/academia-government/scientific-and-academic-research/research-discovery-and-referencing/web-of-science/web-of-science-core-collection.
The Journal Citation Reports (JCR) dataset for the journal quantile is retrieved from https://jcr.clarivate.com/jcr/browse-journals.
The author contribution dataset is available at https://zenodo.org/records/6569339.
The pre-trained parameters for the BERT language model are available at https://huggingface.co/docs/transformers.
The pre-trained parameters for the SPECTER 2.0 text embedding model are available at https://huggingface.co/allenai/specter2.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	This study uses historical research articles to quantitatively analyze the impact of artificial intelligence on scientific research.
Research sample	This study includes 41.3 million global research articles on natural science from the OpenAlex dataset at https://openalex.org.
Sampling strategy	We analyze the full sample of research articles published from 1980 to 2025 archived in OpenAlex.
Data collection	Data collection is directly based on existing datasets. The researchers did not collect the data themselves.
Timing	The OpenAlex dataset was accessed on April 2025, and other auxiliary datasets (WoS and JCR) and models (BERT and SPECTER 2.0) were accessed on November 2024.
Data exclusions	N/A
Non-participation	N/A
Randomization	N/A

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	N/A
Novel plant genotypes	N/A
Authentication	N/A