

Artificial Intelligence Tools Expand Scientists' Impact but Contract Science's Focus

(Just accepted by *Nature*, to be online soon)

Qianyue Hao¹, Fengli Xu^{1*}, Yong Li^{1,2*}, and James Evans^{3,4*}

¹Department of Electronic Engineering, Tsinghua University, Beijing & 100084, P. R. China.

²Zhongguancun Academy, Beijing & 100094, P. R. China.

³Knowledge Lab and Department of Sociology, University of Chicago, Chicago & IL 60637, USA.

⁴Santa Fe Institute, Santa Fe & NM 87501, USA.

*Corresponding authors. Emails: fenglixu@tsinghua.edu.cn, liyong07@tsinghua.edu.cn, jevans@uchicago.edu

ABSTRACT

Development in Artificial Intelligence (AI) has accelerated scientific discovery¹. Alongside recent AI-oriented Nobel prizes^{2–9}, these trends establish the role of AI tools in science¹⁰. This advancement raises questions about the potential influences of AI tools on scientists and science as a whole, and highlights a potential conflict between individual and collective benefits¹¹. To evaluate, we used a pretrained language model to identify AI-augmented research, with an F1-score of 0.875 in validation against expert-labeled data. Using a dataset of 41.3 million research papers across natural science 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 scientist's engagement with one another by 22.00%. Thereby, AI adoption in science presents a seeming paradox—an expansion of individual scientists' impact but a contraction in collective science's reach—as AI-augmented work moves collectively toward areas richest in data. With reduced follow-on engagement, AI tools appear to automate established fields rather than explore new ones, highlighting a tension between personal advancement and collective scientific progress.

Introduction

Artificial intelligence (AI) has made significant strides in recent decades, promising to impact myriad aspects of society, including education^{12,13}, healthcare^{14,15}, and industry¹⁶. Major investments in predictive and generative AI have catalyzed 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 fields¹. For example, AlphaFold learns known protein structures to accurately predict the unexplored ones, circumventing the capital and human cost of conventional structural inference and recently granted a 2024 Nobel Prize^{9,17}. Models improved via deep reinforcement learning have become tuned to contain complex fusion reactions¹⁸ and discovered new, hardware-optimized forms to matrix multiplication that recursively accelerate deep learning itself¹⁹. Autonomous laboratory systems driven by ChatGPT have helped some chemists and material scientists upscale the number of adaptive high-throughput experiments^{20–22}. Moreover, recent developments in large language models are making them increasingly incorporated in assisting scientific writing^{23–26}, facilitating the distillation of scientific findings, but also raising concerns about weakened confidence in AI-generated content^{21,22,27}. AI's increasing capabilities to influence scientific research suggest that it manifests potential to both increase the productivity of individual scientists and raise the visibility of 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 impact on the entire character of science remains largely unknown. Recent work suggests AI has brought widespread benefits to individual scientists but may lead to demographic disparity resulting from gaps in AI education¹⁰. Researchers have also identified evolving citation patterns that signal a changing scientific landscape in AI research²⁸. Here we seek to explore the impact of AI in scientific research at different scales by posing the question: How does the adoption of AI influence 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 dataset²⁹, with patterns corroborated using the Web of Science^{30,31}.

Notably, we do not focus on computer science or mathematics, fields that develop AI methodologies directly, but rather on papers that augment research in natural science fields by adopting AI, primarily covering decades involving development and deployment of conventional machine learning algorithms and also extending 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 the BERT language model^{32,33} to accurately identify such AI-augmented research papers based on their titles and abstracts.

We separate the periods in which AI was predominantly conventional machine learning, deep learning, and most recently generative designs including 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, bringing acceleration in 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. Meanwhile, analyses using currently available data within the latest era of generative AI including large language models reveal a preliminary consistency with prior periods, providing a starting point for further study as generative AI develops over a longer period.

Results

Increasing prevalence of AI in science

In this investigation, we focus on research papers using AI methods in various fields of natural science, where we conduct our analysis based on 41,298,433 papers from the OpenAlex dataset²⁹, covering six representative disciplines: biology, chemistry, geology, materials science, medicine, and physics (Methods M1). According to the invention of milestone technologies in the trend of AI development, we divide the past decades into three eras, namely machine learning (ML), deep learning (DL), and generative AI (GAI) (Methods M2). 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 for automatically extracting and interpreting information from context. Specifically, we employ 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 M3, and Extended Data Fig. 1). This approach eliminates the need for manual selection of AI-related trigger words, as demonstrated in previous research²⁸.

To evaluate the accuracy of our identification, we recruited a team of human experts to validate these results (Methods M4 and Extended Data Fig. 2). The experts demonstrate strong consensus across their independent annotation of papers sampled at random from the six disciplines mentioned above, achieving an average Fleiss’ Kappa (κ) of 0.964^{37,38}. The BERT model attains an F1-score of 0.875 in an evaluation that uses the expert labels as ground truth. Meanwhile, 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 accuracy and laying a robust foundation for our subsequent analysis (Fig. 1b and Supplementary Table S1–S4). 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 era of AI development (Supplementary Fig. S2–S3).

In total, we identify 310,957 AI-augmented papers, comprising 0.75% of all selected papers. Semantically, the identified AI-related papers turn out to be around topics combining artificial intelligence and conventional research topics across disciplines (Supplementary Fig. S4). 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 ML era, and convolutional neural networks and generative adversarial networks from the DL era. The large language model, which has emerged in recent years, also ranks among the most frequently utilized methods (Fig. 1c and Supplementary Table S5–S11). Statistically, despite the overall rise in the number of papers published annually across all disciplines³⁹, the share of AI-augmented papers surged by 10.70 (geology, $Z = 348.60$, $p < 0.001$ and $df = 1$ in Cochran-Armitage test) to 51.89 (biology, $Z = 1388.70$, $p < 0.001$ and $df = 1$ in 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 Cochran-Armitage test) to 362.16 in physics ($Z = 2237.51$, $p < 0.001$ and $df = 1$ in Cochran-Armitage test) (Fig. 1e). Meanwhile, growth rates for AI-augmented papers and researchers have accelerated across the three eras (Fig. 1f and Supplementary Fig. S5–S6). These findings underscore the increasing prevalence and fast development of AI in science across all disciplines and the importance of understanding AI’s impact on scientific research and progress.

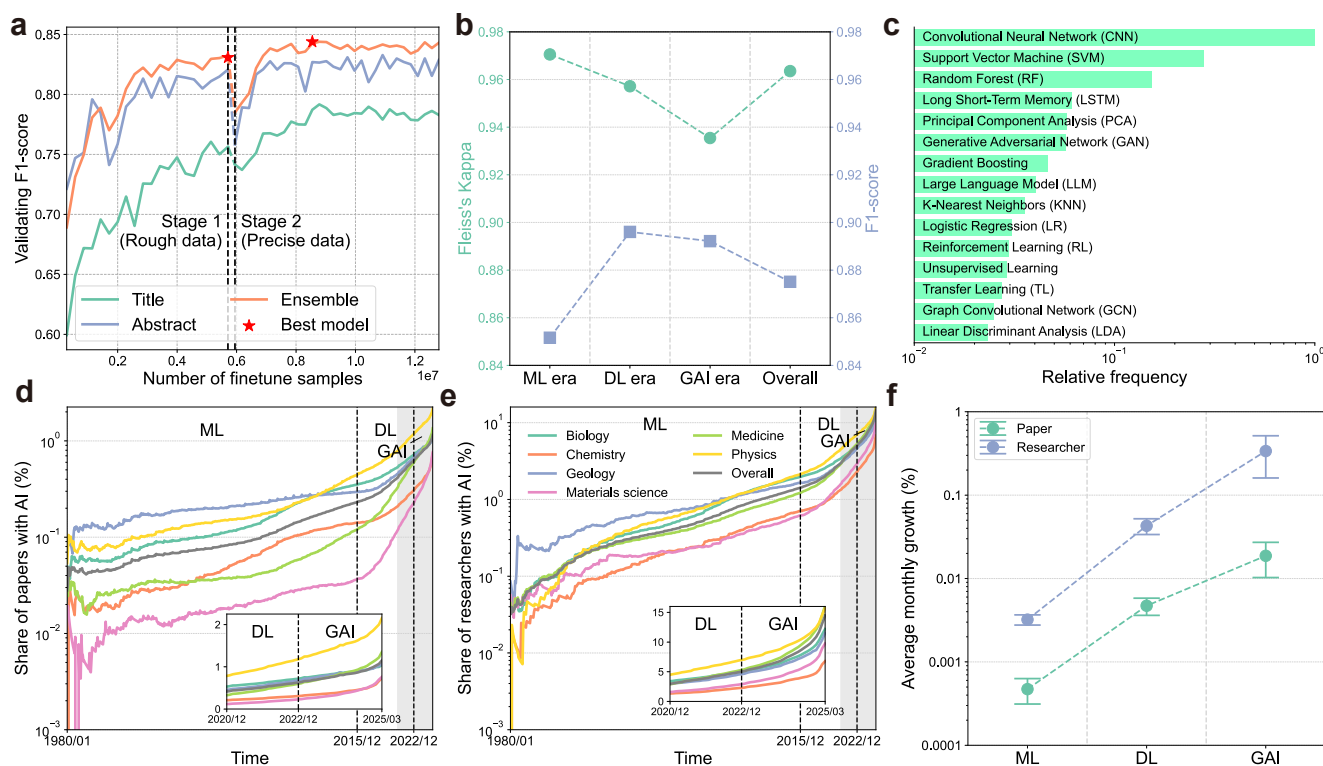


Figure 1. Increasing prevalence of AI adoption in science. (a) Increasing performance of AI paper identification during the two-stage fine-tuning of the BERT pre-trained models, where we use rough training data in Stage 1 to evolve precise assessments in Stage 2. We independently train two models based on titles and abstracts, respectively, and then integrate them into an ensemble that selects the optimal models during both stages (red stars) to identify all selected papers. (b) Accuracy evaluation of our identification results by human experts. For samples spanning three eras of AI, experts reached consensus with Fleiss' Kappa (κ) ≥ 0.93 . Our model identification results have strong accuracy in validation against expert-labeled 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-adopted researchers (e, $n = 5,377,346$) across the eras of ML, DL, and GAI between 1980 and 2025 in selected scientific disciplines. The y-axes are set to log-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% CIs are shown as error bars centred at the mean.

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 until decades later, annual citations to AI papers are 98.70% higher than 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 higher scientific impact of AI-augmented papers is also reflected by multiple alternative statistical indicators about top and bottom annual citation count (Supplementary Fig. S8). Also, AI papers published at different eras consistently receive more citations (Extended Data Fig. 3, $t \geq 4.06$, $p < 0.001$ and $df > 10^3$ in t-test on any era). Furthermore, we examine the distribution of AI-augmented papers across journals of varying Journal Citation Report (JCR) quantiles⁴⁰ (Supplementary Fig. S14). We find that the proportion of AI papers in Q1 journals is 18.60% higher than non-AI ones in all journals, and in Q2 journals, the AI proportion is 1.59% higher, while Q3 and Q4 journals hold a relatively lower proportion of papers with AI ($\chi^2 = 3629.11$, $p < 0.001$ and $df = 3$ in 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) compared with those not adopting AI, with consistency across disciplines and robustness for core researchers with multi-year continuous publication records⁴¹ (Fig. 2b, Extended Data Fig. 4 and Supplementary Fig. S17). Furthermore, when controlling for and comparing scientists with similar early-career positions, the enhanced productivity and impact still hold (Supplementary Fig. S16). This suggests that, after accounting for potential selection-biases among researchers with

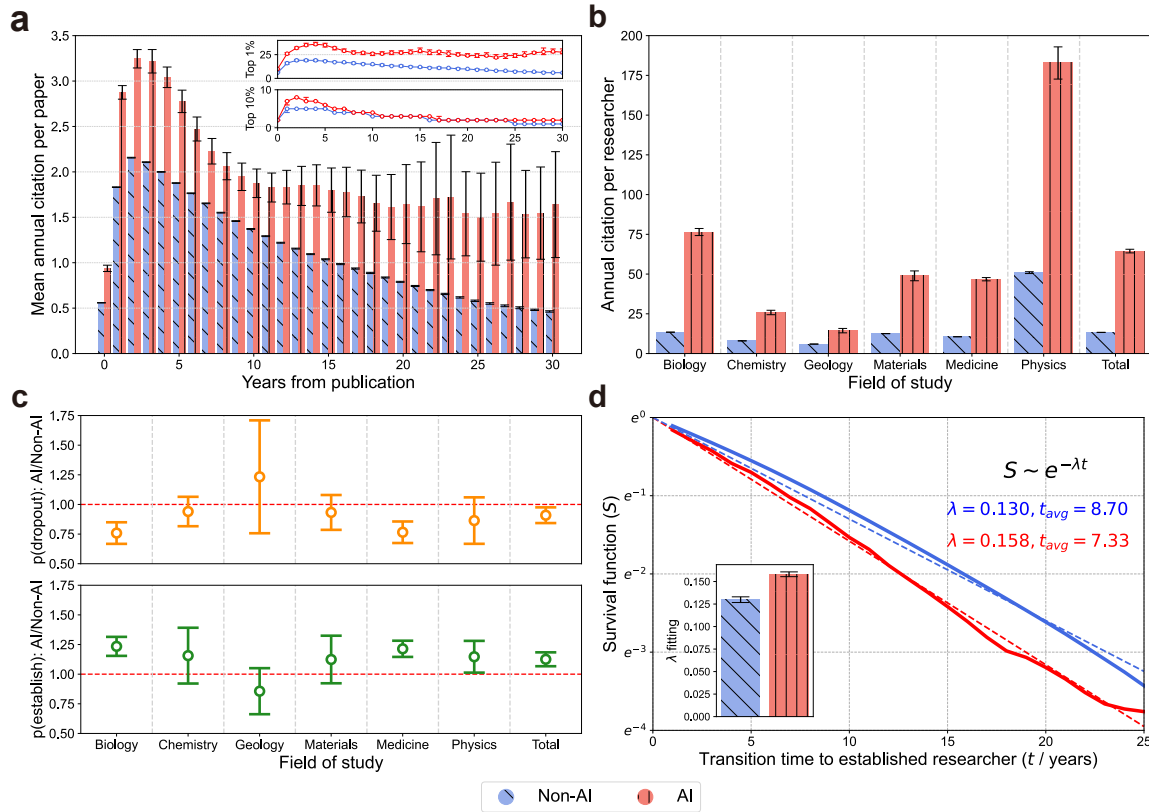


Figure 2. AI enlarges paper impact and enhances researcher careers. (a) Average (insets: top 1% and 10%) annual citations after publication of AI and non-AI papers ($n = 27,405,011$), where AI papers attract more citations. (b) Average annual citations for researchers adopting AI and their counterparts without AI ($p < 0.001, n = 5,377,346$), where researchers adopting AI garner 4.84 times more citations than their counterparts without AI. (c) The probability of two role transitions between junior scientists adopting AI and their counterparts without AI ($n = 46$ year observations for each field). Junior scientists adopting AI have a higher probability of becoming established researchers and a lower probability of exiting academia compared with their counterparts without AI. (d) Survival functions for the transition from junior to established researcher ($p < 0.001, n = 2,282,029$). The survival function can be well-fit with exponential distributions, where junior scientists adopting AI become established earlier than their counterparts without AI. For all panels, 99% CIs are shown as error bars, with the insets of panel (a) centred at the 1% and 10% percentiles and other panels centred at the mean. All statistical tests use a two-sided t-test.

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 scientist's career development, we classify the scientists into "junior" and "established", where junior scientists are defined as newcomers who have not yet led a research project, whereas established scientists refer to those who have led one or more research projects (Methods M5 and Extended Data Fig. 5). We extract 2,282,029 career trajectories of scientists from the dataset, each initially identified as a junior scientist (Methods M6). 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 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), while the number of established scientists decreased from 4.01 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, while 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 for AI-adopted junior scientists to transition to established scientists is 45.00%, 13.64% higher than for their counterparts who do not adopt AI ($t \geq 1.40, p < 0.2$ and $df = 90$ in t-test on 4 out of 6 disciplines). This indicates that AI-adopted scientists are associated with increased opportunities to lead research projects and reduced risks of dropping out from 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 employing a birth-death model⁴² and fitting the model parameter λ with scientists' career trajectories (Fig. 2d and Methods M7). We find that the anticipated transition time to becoming established scientists is 1.37 years shorter for AI-adopted junior scientists compared to their counterparts. The expectation of transition time is 7.33 years for junior scientists adopting AI ($R^2 = 0.995$) and 8.70 years for those without ($R^2 = 0.987$). This demonstrates how AI adoption affords junior scientists opportunities to lead research projects and become established earlier. Further analysis reveals that this reduction in the transition time for AI-adopted junior scientists to become established ones is universal across examined disciplines (Extended Data Fig. 7). Moreover, the established scientists involved in AI papers are, on average, 10.77% younger than those in non-AI papers (Extended Data Fig. 6, $t \geq 2.12$, $p < 0.05$, and $df > 10^3$ in 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 become established scientists with higher probability and at earlier ages, experiencing accelerated career development.

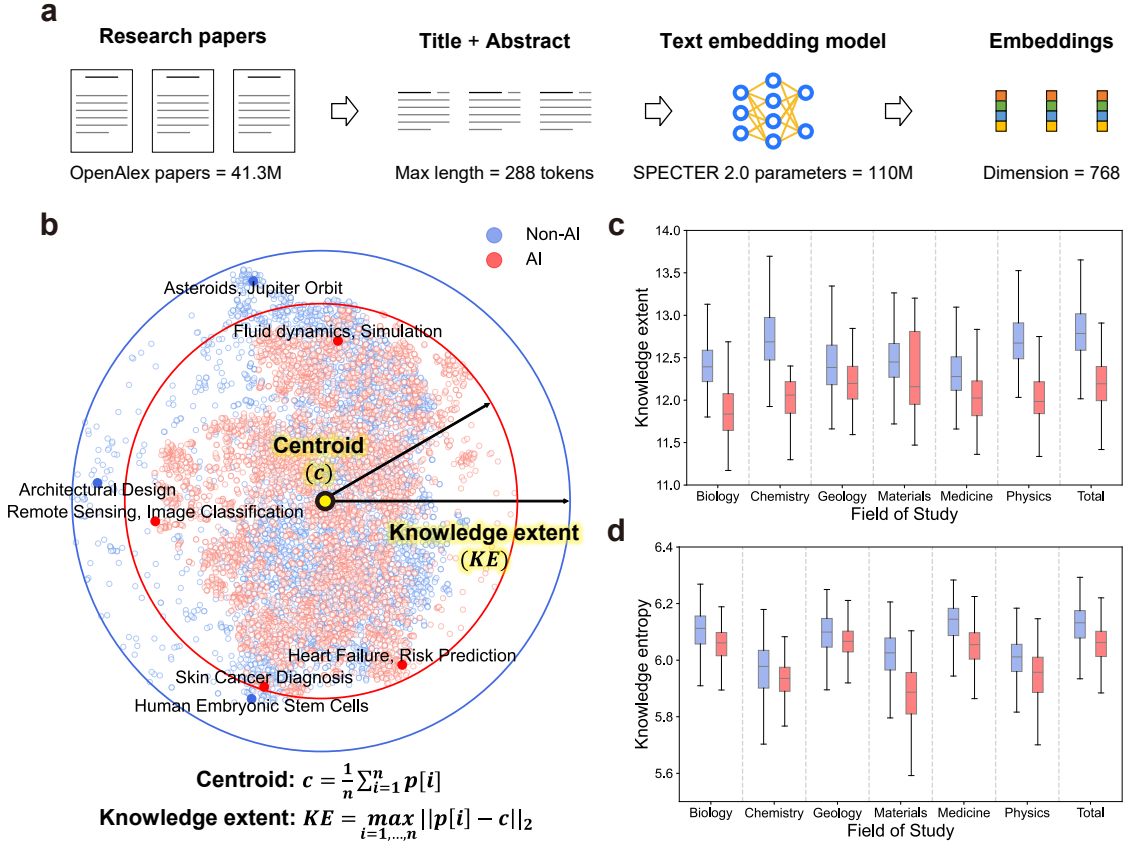


Figure 3. AI adoption is associated with a contraction in knowledge extent within and across scientific fields. (a) We embed research papers into a 768-dimensional vector space with a pre-trained text embedding model, then measure the knowledge extent of papers within that space. (b) For visualization, we use the t-SNE algorithm to flatten the high-dimensional embeddings of a random batch of 10,000 papers, half of which are AI papers and half are non-AI papers, into a 2-D 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 natural science, and AI papers are more clustered in knowledge space, indicating more 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 quartile (Q1 and Q3), with 1.5 times of the inter-quartile range (IQR) shown as whiskers from the box. All statistical tests use a median-test.

AI contracts science's focus

The accelerating use of AI in science and its impact on individual scientists raises questions about its influences across the entire scientific field. 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 employ SPECTER 2.0, a specialized text embedding model pre-trained on a large scientific literature corpus and fine-tuned with citation information³⁶, to project research articles onto this 768-dimensional embedding space of science (Fig. 3a).

Within the high-dimensional embedding space, we design the measurement of knowledge extent (KE), which is 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 M8). Compared to 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 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 sub-fields (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 median-test on any discipline), indicating an increasingly disproportionate focus on specific problems rather than across entire fields.

Generally, these results 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 like inherent topicality, original impact, and funding priority, remain almost unrelated to the disproportionate AI adoption (Supplementary Fig. S22-S24). In contrast, data availability appears 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. S25).

AI reduces scientific engagement

In order to analyze 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”, i.e., 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 M9). Results show that the knowledge extent of AI papers’ citation families is on average 3.46% more expanded compared to 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). Therefore, the contraction of knowledge space in AI research is not attributable to the narrowing of knowledge space that can be derived from each original research work.

To further investigate, 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 M10). Results demonstrate AI research spawns 22.00% less follow-on engagement ($t \geq 8.10$, $p < 0.001$ and $df > 10^3$ 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 fields⁴⁵. 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 effect⁴⁶ 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 $df = 198$ in t-test), signaling a disparity in recognition.

To further analyze 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), while the remaining pairs do not cite each other (disengaged). We examine distances between these pairs of papers within the 768-dimensional vector space (Fig. 4d) and find that median distance between paper pairs disengaged from one another tends to be 18.11% larger than between paper pairs engaged with each other. In contrast, the closest disengaged paper pairs are 76.51% closer to one another than the closest paper pairs engaged with one another. Taken together, a pair of disengaged papers commonly focus on less related topics and lie farther apart in the embedding space. Occasionally, however, due 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 overlapping research and a contraction in knowledge extent and diversity across science.

Discussion

In this study, 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 adoptions in science, corresponding with the dominance of machine learning, deep learning, and large language models. Each wave is marked with an accelerated AI adoption rate

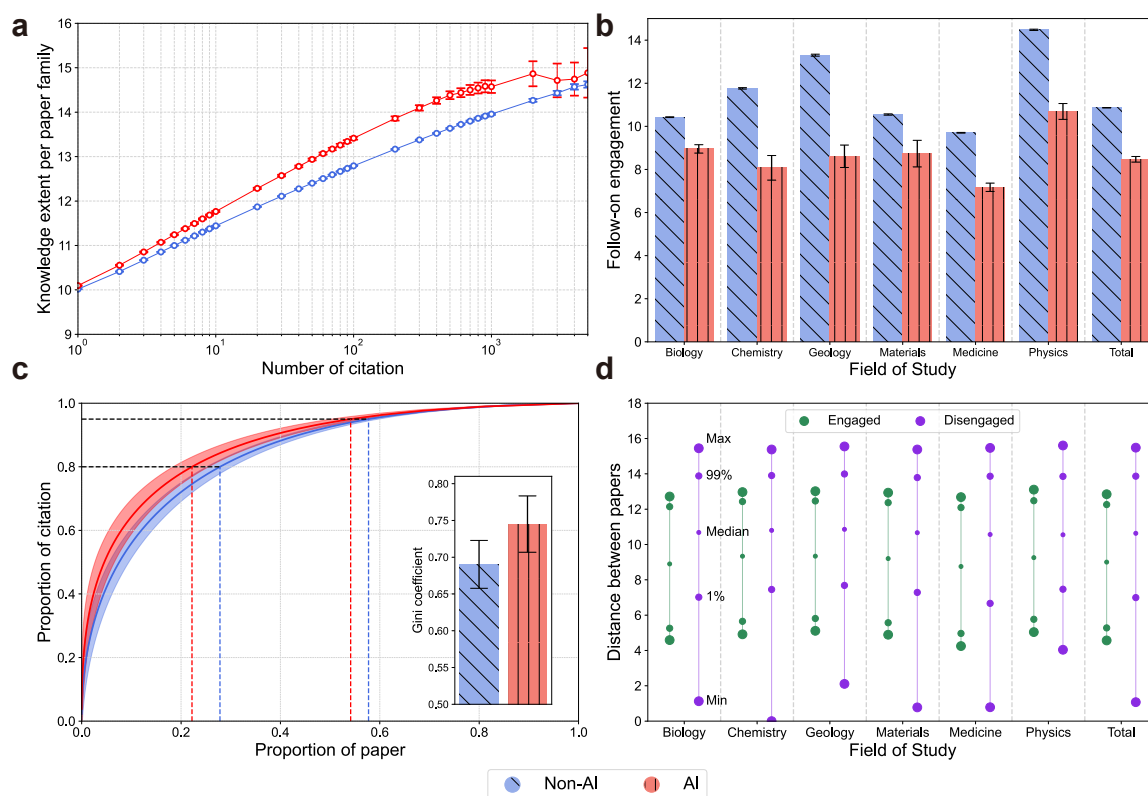


Figure 4. Reduced follow-on engagement and more overlapping works in AI research. (a) Knowledge extent of individual AI and non-AI paper families, i.e., an original paper and its cumulative citations ($n = 27,405,011$), where the knowledge space of individual AI paper families is broader and grows faster. (b) Engagement among papers that cite AI vs. 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 vs. 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 prior research, with or without citing one another—engaged versus disengaged ($n = 590,325,130$ 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 overlapping 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.

in research papers and authors. We find that individual scientists are increasingly rewarded with expanded academic impact and accelerated career development for incorporating AI assistance in research across these waves and in all natural science research fields we studied. On average, the use of AI helps individual scientists publish 3.02 times more papers, receive 4.84 times more citations, and become team leaders 1.37 years earlier. This substantial academic benefit may be a driving force behind the accelerated rate of AI 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 works that slows the expansion of knowledge. Further, with a greater concentration of collective attention to the same AI papers, the adoption of AI appears to induce authors to engage in collective hill-climbing¹, catalyzing solutions to known problems rather than creating new ones.

These findings raise critical questions for science policy. What are the topics most likely 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 lamp post” of highly visible, data-rich phenomena moves science away from many foundational questions and towards operational ones. By driving attention toward the most popular new developments, AI appears 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

¹The metaphor of “collective hill-climbing” describes a situation where researchers act like a group of climbers all scaling the same popular mountain from the same route. Because both the “path” of a known approach and the “peak” of an anticipated solution are constrained, this collective rush leads to “crowding” and may discourage the search for other, potentially higher “mountains” representing new questions and answers.

scientific incentives we demonstrate. 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 specific 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.

While our analysis provides new insight into AI's impact on science, clear limitations remain. Our identification approach, though validated by experts, misses subtle and unmentioned forms of AI use, and our focus on natural sciences excludes important domains where 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 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 nature⁵¹. 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.

Methods

M1. Dataset and Paper selection

In this section, we introduce the procedure of selecting the research papers included in our analysis. In this paper, we conduct our major analyses based on OpenAlex²⁹. OpenAlex is a scientific research database built upon 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 265.7M research papers, along with related data about citation, author, institution, etc. 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 utilizing the topics contained 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 the Microsoft Academic Graph (MAG)^{52,53}, i.e., 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 to 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, e.g., chemistry and biology, which reflects the common phenomena of borderline or interdisciplinary research⁵⁴.

In this paper, 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. Therefore, we select biology, medicine, chemistry, physics, materials science, and geology as representatives of natural science disciplines, while we 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, in order to focus on how AI is changing the natural sciences and career trajectories in the sciences. Our 6 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, 5,138,488 in physics. The selected disciplines cover various dimensions of natural science, representing an broad view of scientific research as a whole.

M2. Divide three stages of AI development

We divide the history of AI development into three key eras: the traditional machine learning (ML) era (1980–2014), the deep learning (DL) era (2015–2022), and the generative (GAI) era (2023–present). We consider 1980 as the start of the traditional machine learning era because several landmark researches were published in the 1980s, such as the back-propagating method^{55,56}. We regard that the deep learning era began in 2015, as marked by breakthroughs including ResNet, which enabled the training of ultra-deep neural networks, revolutionizing fields including computer vision and speech recognition⁵⁷. Finally, we divide the GAI era to begin in 2023 with the publication of ChatGPT, a representative large language model, in December 2022, which saw 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.

M3. Design and fine-tune the language model for AI paper identification

Insofar as both a paper's title and the abstract contain important information about its content, we independently train two separate models based on paper titles and abstracts, and then integrate the two models into an ensemble one by averaging their outputs. The structure of our NLP model for paper identification consists of two parts. The backbone network is a 12-layer BERT model with 12 attention heads in each layer, and the sequence classification head is a linear layer with a 2-dimensional output atop the BERT output. We normalize the 2-dimensional output with a softmax function and obtain the probability that the paper involves AI-assistance. We utilize the BERT model named "bert-base-uncased" from Hugging Face⁵⁸, 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% to correspondingly obtain training and validation sets. We use the training set for model training and employ the validation set to select the optimal model. Because 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, i.e., *Nature Machine Intelligence*, *Machine Learning*, *Artificial Intelligence*, *Journal of Machine Learning Research*, *International Conference on Machine Learning (ICML)*, *International Conference on Learning Representations (ICLR)*, *AAAI Conference on Artificial Intelligence*, and *International Joint Conference on Artificial Intelligence (IJCAI)*. Among the papers belonging to our chosen 6 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 data. 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 based on the obtained optimal model in the first stage. We identify papers in the whole OpenAlex dataset and aggregate the results for each venue, obtaining the probability for each venue across OpenAlex to be 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 6 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. Then, we 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 utilize 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.

M4. Scrutinize our identification results by disciplinary experts

We arbitrarily sample 220 papers (110 papers $\times$ 2 groups) from each of the 6 disciplines, resulting in 12 paper groups in total. We enlisted 12 experts with abundant AI research experience (Supplementary Table S1) and assigned 3 different groups of papers to each. Without revealing the identification results obtained with the BERT model, we queried our experts whether each paper was an AI paper. In this way, each paper is repeatedly labeled by three distinct experts, and we evaluate the consistency among these experts based on Fleiss' Kappa coefficient (κ)^{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 it with the F1-Score, which is a supervised accuracy measurement of accuracy.

M5. 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 studies⁵⁹. To ensure that in most papers, the last authors represent the project leader, we examine the fraction of papers that list authors following alphabetical order. First, we directly traverse all selected papers and obtain that 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 are really intended to list the authors in alphabetical order or the list of authors according to their roles, which just happen to unintentionally fall in alphabetical order. The latter situation is more likely to occur when there are fewer authors, i.e., two or three. To tackle this analytical challenge, we determine the fraction of "unintended" alphabetical author lists through a Monte Carlo method. We generate 10 randomly shuffled copies of the author

list for each paper and obtained that from 13.82% (materials science, standard deviation $\sigma = 0.02$) to 20.28% (geology, standard deviation $\sigma = 0.03$) 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 S12), and therefore, we can, with negligible interference, assume that we can identify last authors as team leaders.

M6. Detect scientists’ career role transition

The OpenAlex dataset incorporates a well-designed Author Name Disambiguation (AND) mechanism²⁹, which utilizes an XGBoost model⁶⁰ to predict the likelihood that two authors are the same based on features like their institutions, co-authors, and citations, and then applies a custom, ORCID-anchored clustering process to group their works, assigning a unique ID for each author. Simply utilizing unique IDs, we are able to track a large number of authors at the same time⁶¹, 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 6 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. Subsequently, we 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 based on their absence in the published record for a single year. Therefore, we follow the settings in previous research⁶² to use a threshold of 3 years and regard scientists who have no more publications after 2022 as having exited academia, while those who still publish papers after 2022 are considered to have an unclear ultimate status and are excluded from analysis. Finally, we obtain 2,282,029 scientists in the 6 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 analyzing 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 a large proportion of technical tasks, such as conducting experiments and analyzing 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 Cochran-Armitage test) during their tenure as junior scientists, reaching saturation at a high level (60% or more) upon 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.

M7. 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 employ a general birth-death model⁴². 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 those who leave academia, respectively. Here, “birth” processes refer to the entry of junior scientists into academia, and “death” processes symbolize their transition out of the junior stage, either by becoming established scientists or quitting academia. Because 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 other⁶⁵. The mathematical formula of the Poisson process is:

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

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 where birth and death dynamics are both Poisson processes, and rate parameters are μ and ω , respectively. Through mathematical derivation⁶⁶, 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:

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

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

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

and the corresponding survival function is

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

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

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

We fit the role transition time of the scientists with the aforementioned exponential distribution, thereby determining the respective values of λ for AI-adopted 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.

M8. Measure the knowledge extent of papers

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

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

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 of the set of papers is defined as the maximum distance or “diameter” of the vector space covered:

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

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 differ substantially, introducing bias to the measurement of knowledge extent. To address this issue, we build on prior work⁴⁴ about cognitive extent². 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.

M9. Measuring the knowledge extent of paper families

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

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

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

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

²Cognitive extent is a measure of the breadth of a scientific field's cognitive territory. It is quantified by the number of unique phrases, as a proxy for scientific concepts, found within a sampled batch of papers with given size.

M10. 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 prior work⁴⁵. 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 is calculated as the ratio of actual to maximum possible citations:

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

This metric helps quantify the degree of interactions and collaboration among papers that cite the same original work. Prior work has demonstrated a positive association between the ambiguity of a focal work and follow-on engagement⁴⁵.

Extend data figures

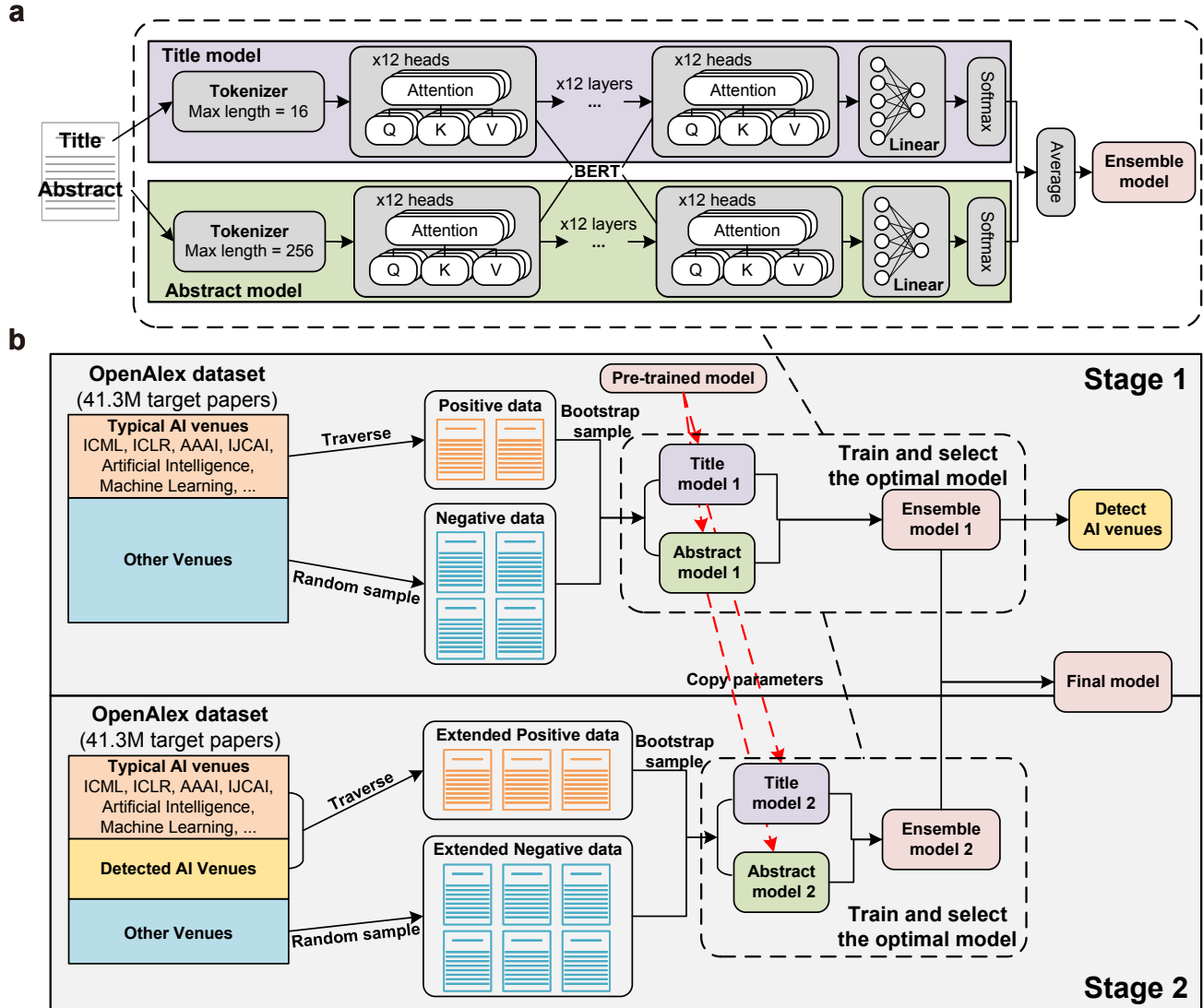


Figure 1. Illustration for the method of identifying AI usage in research papers with fine-tuned language models. (a) Structure of our deployed language model, which consists of the tokenizer, the core BERT model, and the linear layer. **(b)** Procedure of the two-stage model fine-tuning process, where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

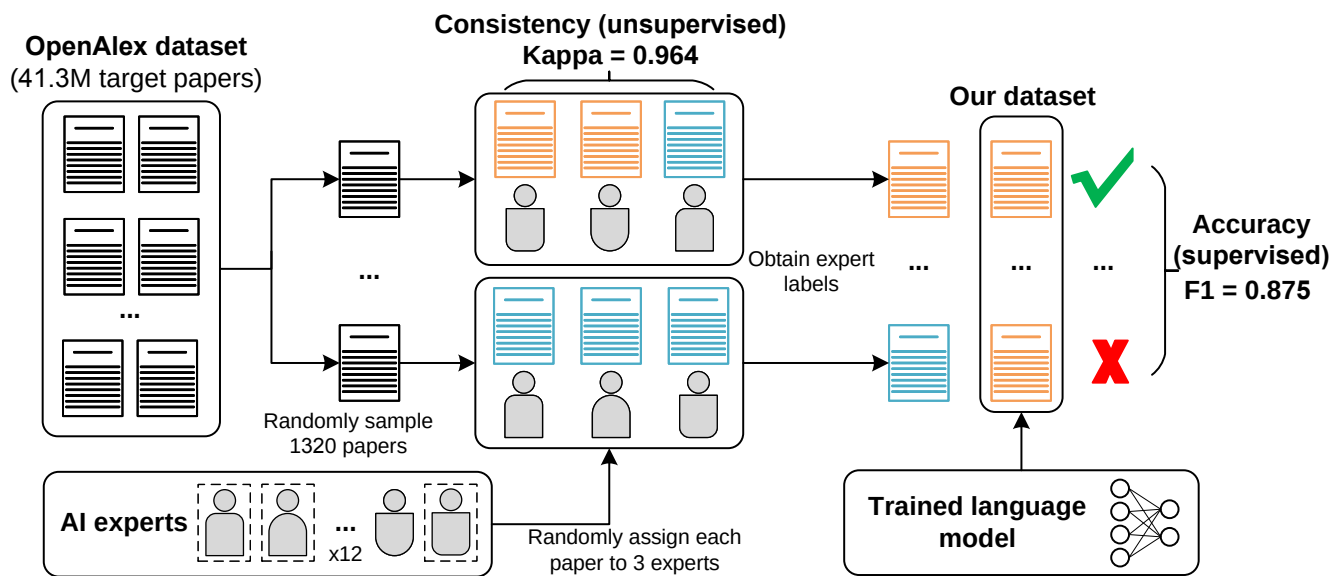


Figure 2. Procedure of accuracy evaluation via expert evaluation. We randomly sample 1320 papers and delegate three experts to scrutinize the identification results for each paper. We then draw the final expert label of each paper from the three experts according to the principle of the minority obeying 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.

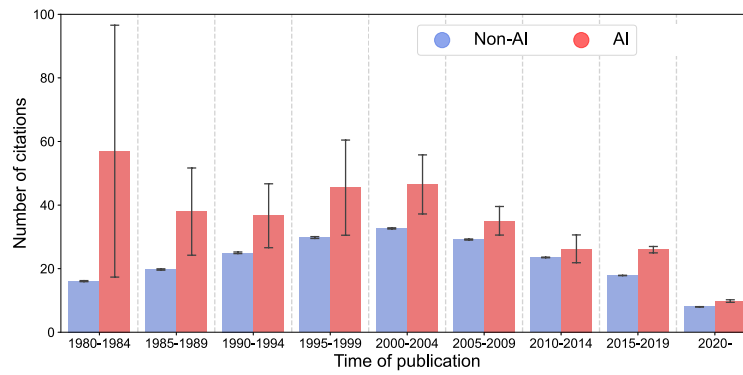


Figure 3. Comparison of the total citations of AI and non-AI papers published in different eras. The results show that AI papers consistently attract more citations over different eras ($p < 0.001, n = 27,405,011$), indicating a higher academic impact than non-AI papers. 99% CIs are shown as error bars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

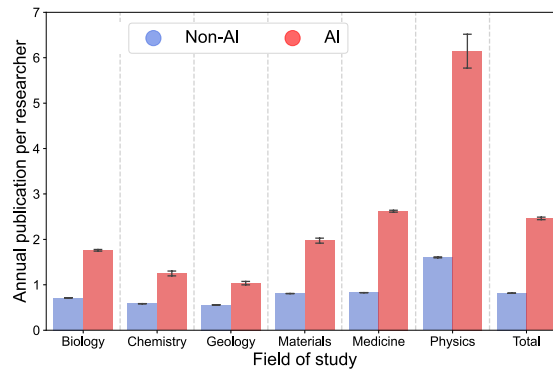


Figure 4. Annual publications of researchers adopting AI and their counterparts without AI. Results show that in all 6 scientific disciplines, researchers adopting AI are more productive than their counterparts without AI ($p < 0.001, n = 5,377,346$). On average, researchers adopting AI annually publish 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.

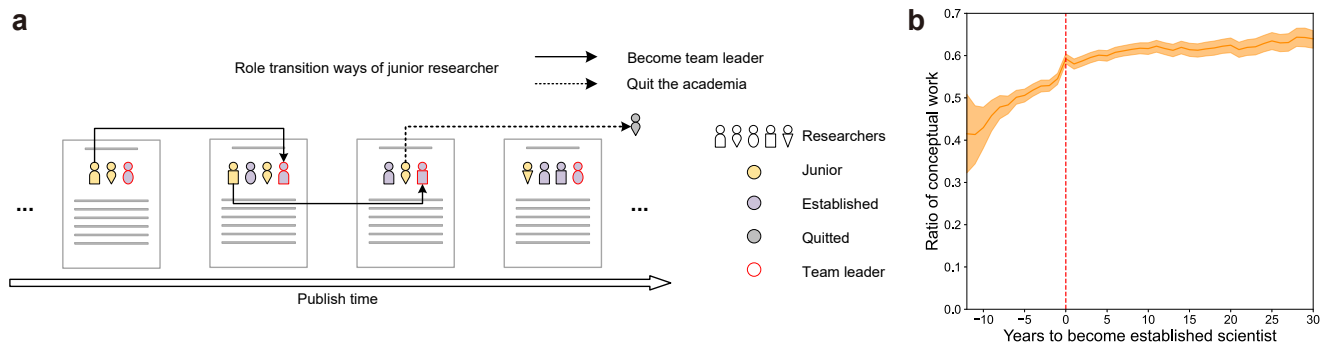


Figure 5. Scientists' career role transition. (a) The career role transition of researchers. We consider the last author of each paper as research project leader and researchers who have been research project leaders as established researchers. Researchers who have yet to lead a research project are junior researchers, and they have two potential role transition pathways in the future: (1) become established researchers (solid arrow), and (2) abandoning 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.

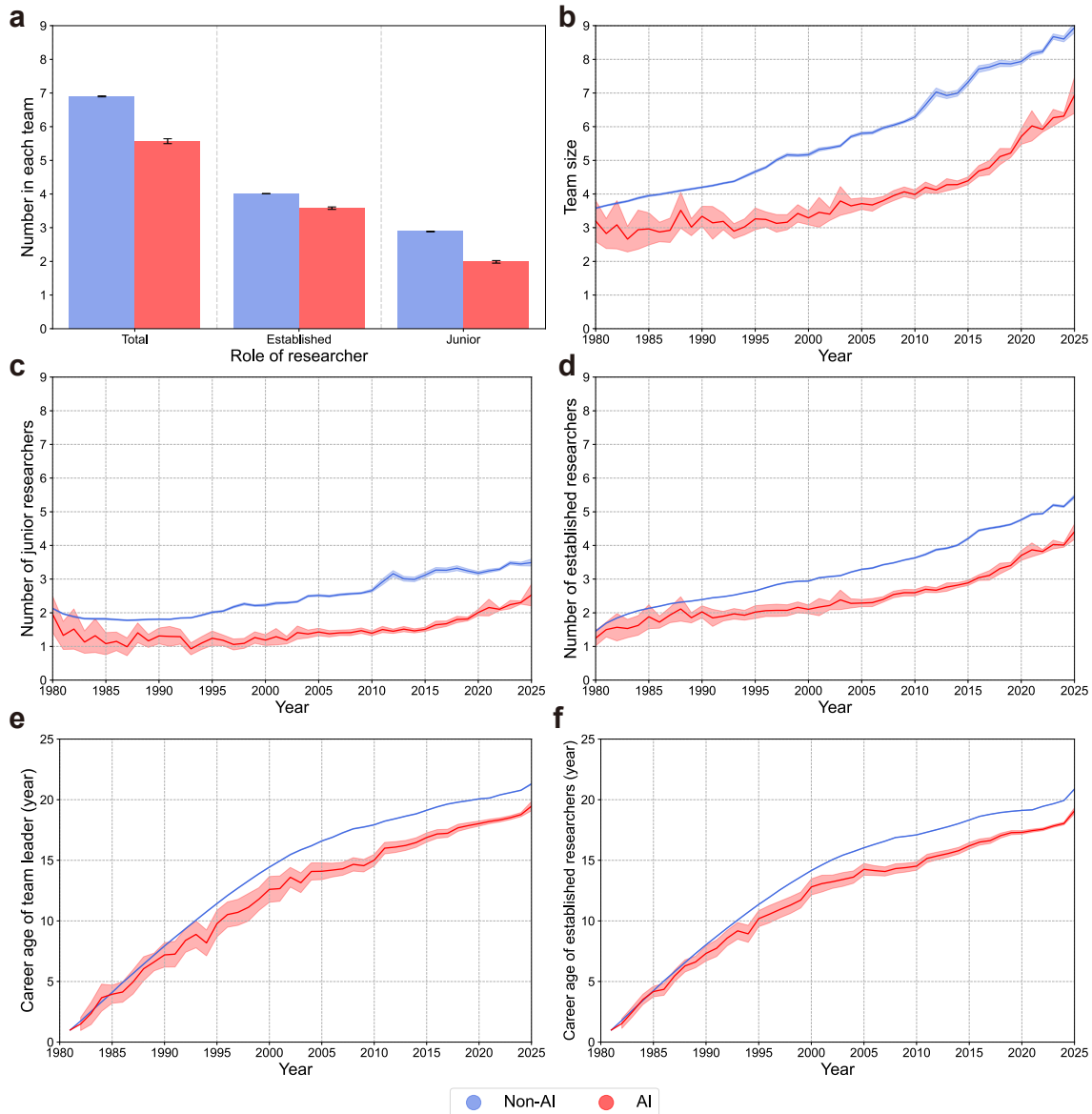


Figure 6. Team composition of AI and non-AI papers. (a) AI research is associated with reduced research team sizes, averaging 1.33 fewer scientists ($p < 0.001, n = 33,528,469$). Specifically, the average number of junior scientists decreased from 2.89 in non-AI teams to 1.99 in AI teams (31.14%), while 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.

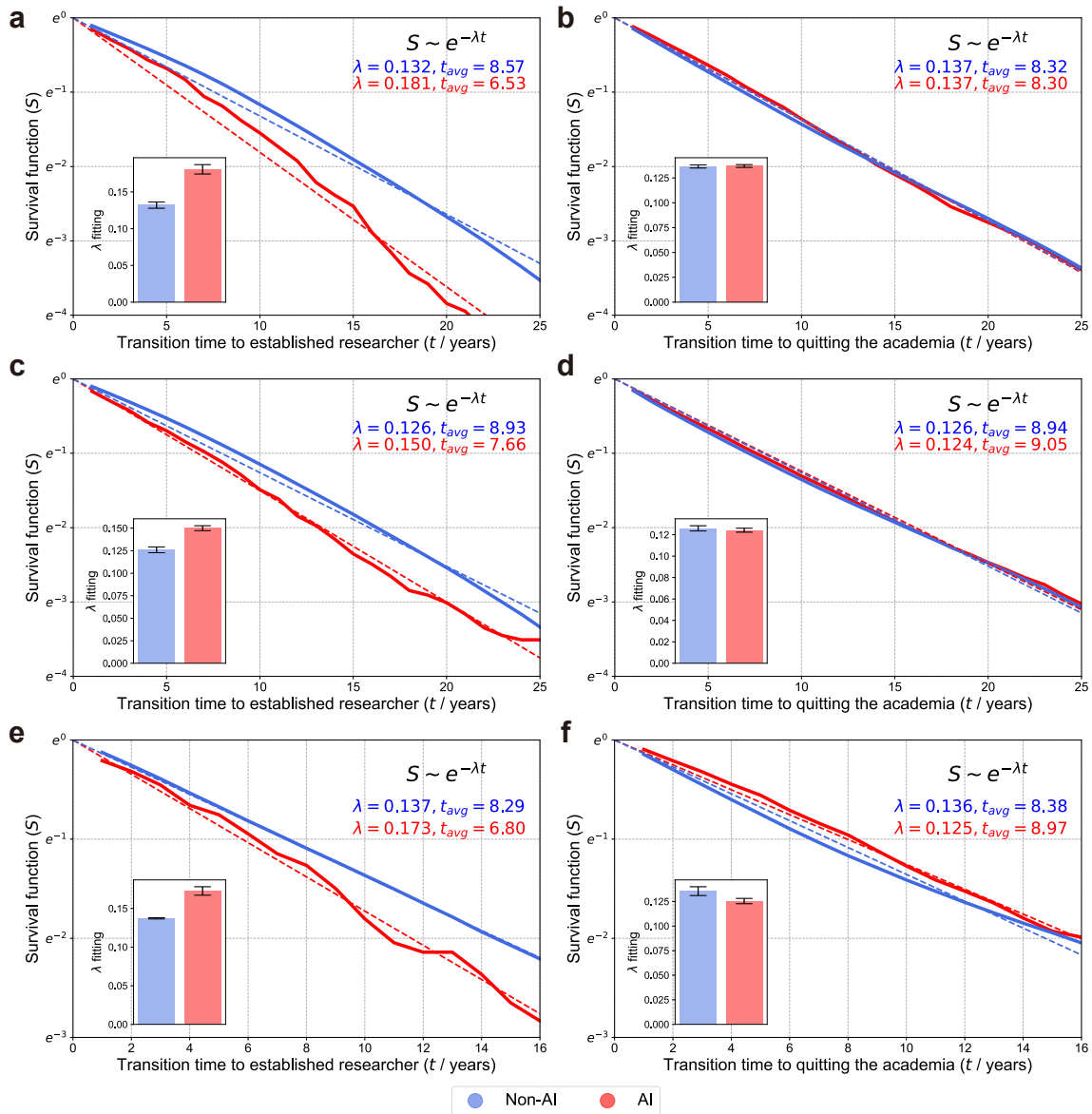


Figure 7. Model fitting the role transition time of junior scientists. (a) (c) (e) Survival functions for the transition from junior to established researcher in biology ($n = 625,093$), medicine ($n = 1,137,076$), and physics ($n = 120,366$). (b) (d) (f) Survival functions for the transition from junior researcher to leave academia in biology ($n = 625,093$), medicine ($n = 1,137,076$), and physics ($n = 120,366$). All survival function 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.

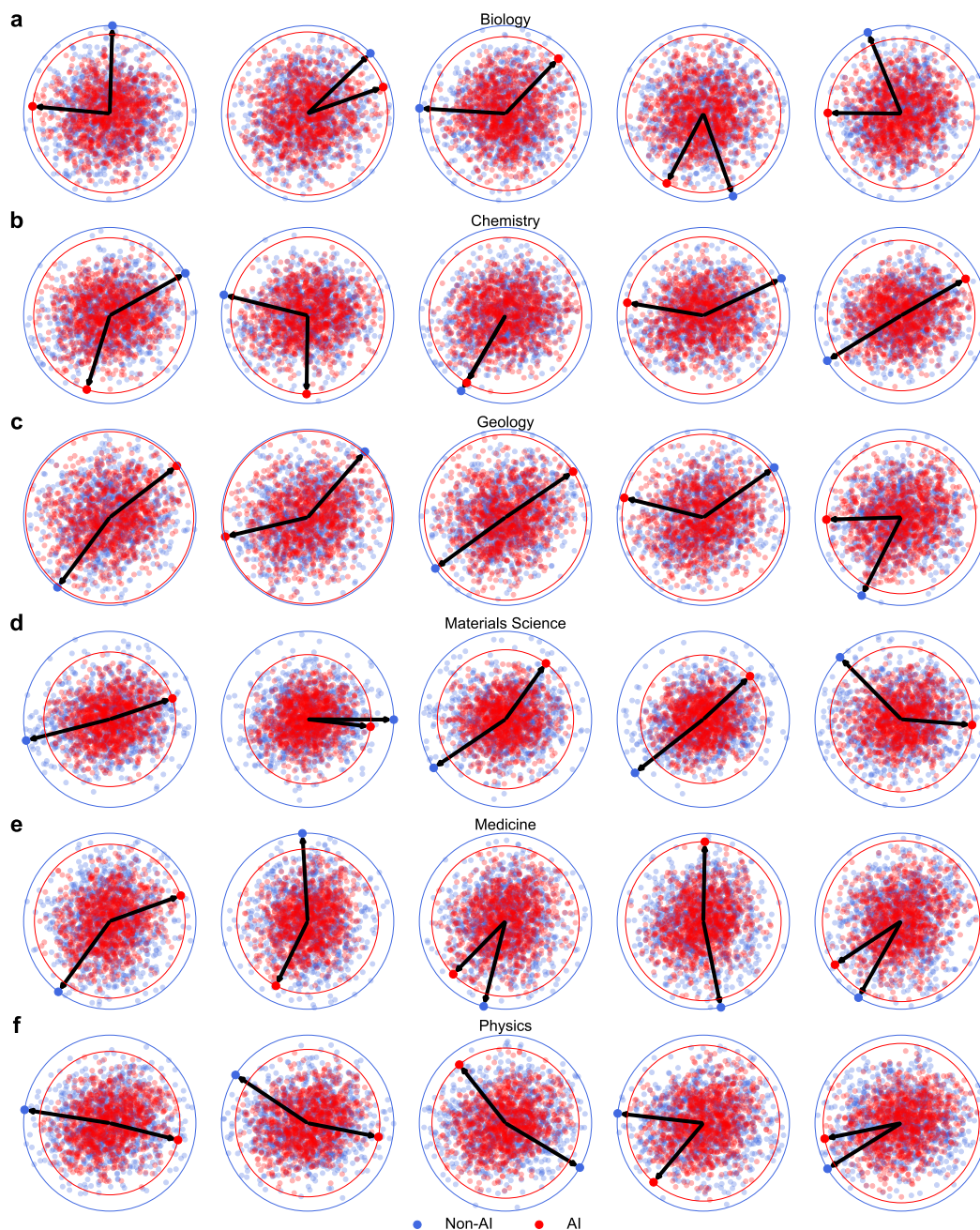


Figure 8. The knowledge extent of AI and non-AI papers. Here we visualize the embeddings of a small random sample of 2,000 papers, half of which are AI papers and half are non-AI papers. To eliminate randomness introduced by the T-SNE algorithm, here we simply pick out the first two dimensions of the 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.

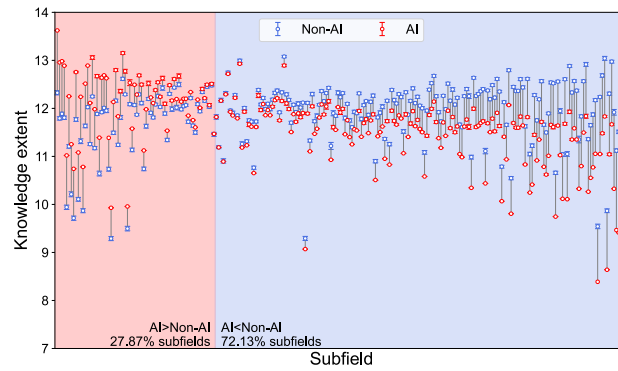


Figure 9. The knowledge extent of AI and non-AI papers in each subfield. Compared with conventional research, AI research is associated with a shrinkage in the collective knowledge extent of science, where the contraction of knowledge extent can be observed in more than 70% of over two hundred sub-fields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

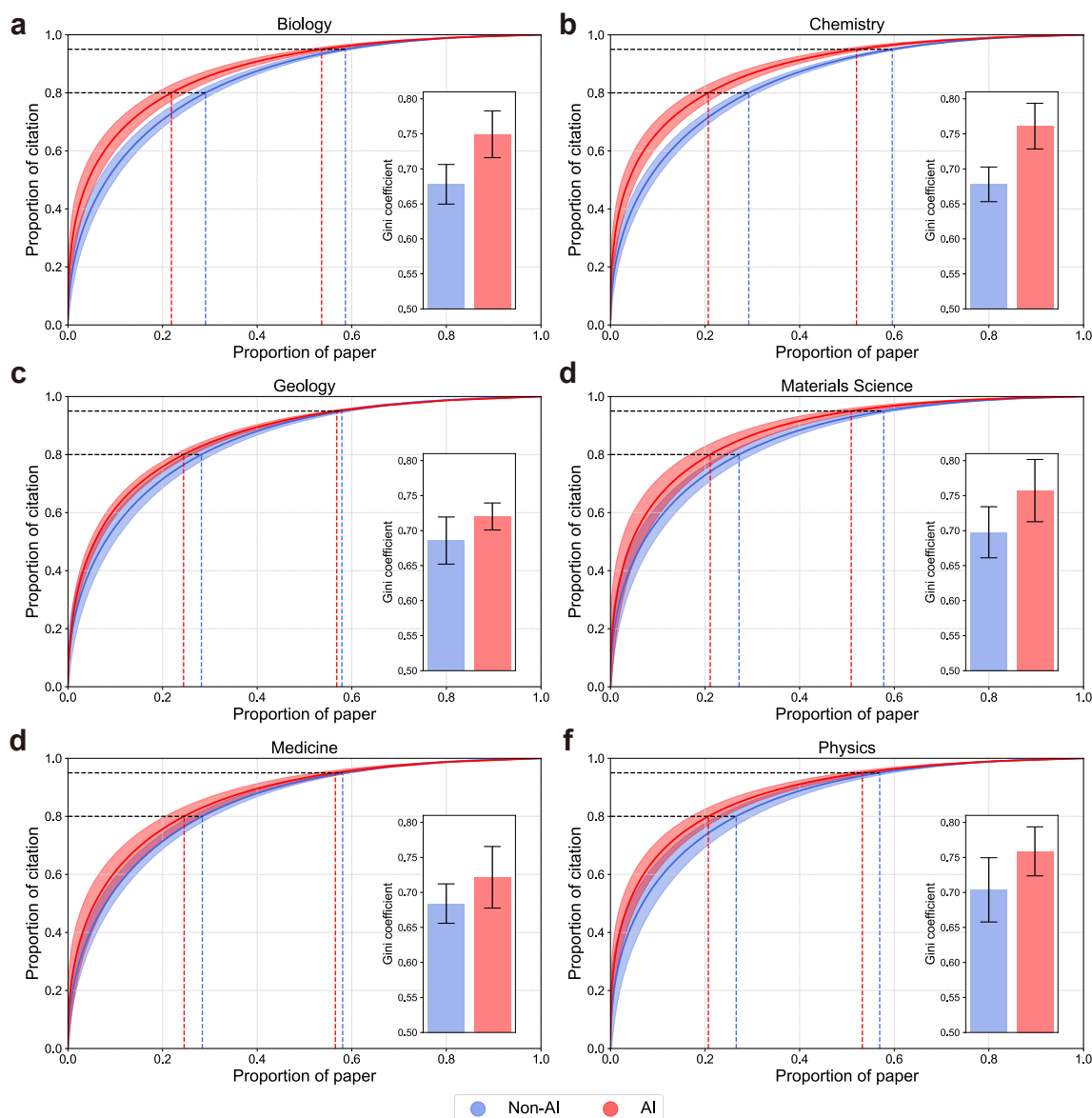


Figure 10. The Matthew effect in citations to AI and non-AI papers. In AI research, a small number of superstar papers dominate the field, with approximately 20% of top papers receiving 80% of citations and 50% receiving 95%. This unequal distribution leads to a higher GINI coefficient in citation patterns 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.

Supplementary Notes

1 Identifying artificial intelligence in scientific research

To thoroughly investigate how the adoption of AI impacts scientific research, the very rudimentary step is to identify the AI papers from massive research papers published over the past decades. In this section, we discuss the details of AI paper identification, covering the model and method design (Section 1.1), the accuracy evaluation (Section 1.2), the statistical (Section 1.4) and semantic (Section 1.3) features of identified AI papers along the trend of scientific research over decades.

1.1 Identification model and method

As described in Methods M2 of the main text, we design a two-stage fine-tuning process to leverage the pre-trained BERT model to identify papers that use AI to support natural science research. In the first stage, we construct coarse positive data with typical AI journals and conferences to fine-tune the pre-trained model. In the second stage, we identify papers in the whole dataset with the obtained optimal model in the first stage and aggregate the results for each venue. We then select the venues with $> 80\%$ AI probability and > 100 papers as AI venues for positive data, and we also incorporate venues with “machine learning” or “artificial intelligence” in their names. Finally, we utilize 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. To illustrate the final identification results, we show the proportion and number of identified AI papers in each venue in Supplementary Fig. S1.

1.2 Identification accuracy evaluation

To evaluate the accuracy of our identification, we recruited a team of human experts with abundant AI research experience (Supplementary Table S1) to validate the results. We sample 2 groups of papers at random from each of the 6 disciplines, resulting in 12 paper groups in total, where samples in each group span all three eras of AI (Supplementary Table S2). We assign three different experts to independently label each sampled paper and measure the consistency among experts based on Fleiss' Kappa coefficient³⁷. Results exhibit an overall Fleiss' Kappa of 0.964, and each of the three eras of AI has a Fleiss' Kappa greater than 0.93 (Supplementary Table S3), indicating strong consensus across the independent annotation of distinct experts. Taking the expert labels as ground truth and validating the identification results of our BERT model against it, our model reaches an overall F1-score of 0.875, and the F1-scores of the three eras of AI are all greater than 0.85 (Supplementary Table S4). This indicates the consistent high quality of our AI paper identification, which lays a robust foundation for our subsequent analysis.

To provide rationale and explainability for our identification results, we offer multiple identification examples from different eras of AI and visualize the average attention strengths within our title and abstract BERT models. Our example for the ML era is a medicine paper with AI published in *PNAS* 2006⁶⁷, where our model allocates substantial attention to terms such as “independent component analysis” (Supplementary Fig. S2a). In the DL era, our first example is a chemistry paper with AI published in *Nature* 2018⁶⁸, and our second is a biology paper with AI published in *Nature* 2021⁹, where our model allocates substantial attention to terms such as “neural network” (Supplementary Fig. S2bc). Our example of the generative AI era is a biology paper using AI and published in *Science* 2023⁶⁹, where the model allocates substantial attention to terms such as “large language model” (Supplementary Fig. S2d). These illustrate how our identification model correctly interprets and accurately identifies various AI-related contents from papers published in different eras of AI.

Furthermore, to gain a deeper understanding of our identification results, we present an example in which the model makes a mistake (Supplementary Fig. S3). The sample paper, published in *Géoscience*⁷⁰, is incorrectly classified by the model as AI-related, but human experts confirm that the article does not involve AI methodologies. As we illustrate, the model assigns substantial attention to terms like “representing” and “deep”, which are commonly associated with AI research. In the context of this article, however, “representing” refers to general expression or depiction, which is unrelated to representation learning in AI, and “deep” is used to describe the extent of uncertainty rather than indicating a deep neural network or similar AI concepts. This example illustrates an edge case where the identification model may produce misclassifications. Nevertheless, given the high F1-score observed in our evaluation of identification accuracy, such cases appear to be rare, and the effect on our subsequent analyses minimal.

1.3 Profile of the identified AI papers

To better illustrate how AI has been applied in natural science research across different disciplines and AI development eras, we present a semantic overview of the major research topics and the primary AI methods employed. For research topics, AI-related studies tend to cluster around areas that integrate artificial intelligence with conventional disciplinary subjects (Supplementary Fig. S4). For example, in the field of medicine, the topics “Radiology, Nuclear Medicine and Imaging” and “Computer Vision and Pattern Recognition” rank among the most prominent. This reflects the widespread application of computer vision techniques to enhance research in medical imaging.

For adopted AI methods, we extracted phrases from the abstracts of identified AI papers and calculated the frequency of phrases related to specific AI techniques. We identified and listed the top 10 most frequently used AI methods for each discipline and AI era (Supplementary Tables S5–S11). It is worth noting that in earlier years, due to the limited number of AI papers, the number of commonly used AI methods may be fewer than ten. The results show that during the ML era, the most frequently applied AI methods in natural science research included Artificial Neural Networks (ANN), Principal Component Analysis (PCA), and Support Vector Machines (SVM). In the DL era, Convolutional Neural Networks (CNN), a landmark of the time, dominated the landscape, while traditional methods such as SVM continued to see widespread adoption. Since the beginning of the generative AI era in 2023, Large Language Model (LLM) has increasingly appeared among the most commonly used AI methods, with their frequency rankings steadily rising across disciplines. The evolution of AI methods in natural science research reflects the development and transformation of AI technologies over the past decades.

1.4 Increasing prevalence of AI adoption in science

In the main text, the proportion of papers and researchers adopting AI exhibit exponential growth over the past decades. To further illustrate and analyze this upward trend, we separately plot the proportion of papers and researchers adopting AI in each discipline with a log-scale y-axis and estimate the growth rate with exponential fitting in each era (Supplementary Fig. S5–S6). As results show, the proportion of papers and researchers adopting AI fit to straight regions in the log-scale plot, indicating an exponential growth trend. Meanwhile, the growth rate of AI papers and AI researchers in all disciplines increased progressively from the ML to the DL and generative AI eras, as estimated by the exponential fitting, which underscores the increasing prevalence and fast development of AI in science.

Beyond the general upward trend of AI adoption in science, we further investigate how the recent emergence of generative AI influences this trajectory. We calculate the monthly increase in the proportion of AI papers and researchers across disciplines in recent years (Supplementary Fig. S7). Results show that following the release of ChatGPT in December 2022, which we regard as the beginning of the generative AI era, growth rates in the proportion of papers and researchers initially remain consistent with prior trends. A period of time later, however, these growth rates begin to exhibit marked acceleration across disciplines, providing evidence for the impact of generative AI advancement. Meanwhile, it also aligns with intuitive expectation: the mass adoption of generative AI in natural science research requires time, and there is an inherent delay for generative AI-based research to pass through peer review and achieve publication.

2 Extended analyses on how AI impacts individual science

In the main text, we illustrate that AI research attracts more attention from academia, and AI-adopting scientists are more likely to achieve higher scholarly productivity and impact. In this section, we discuss in more detail how AI impacts individual science, covering the effect of AI on individual papers (Section 2.1) and individual scientists' careers (Section 2.2).

2.1 The effect of AI on individual papers

As demonstrated in the main text, AI-related papers, on average, receive higher annual citation counts than non-AI papers. Given that citation distributions are empirically right-skewed^{71,72}, we adopt multiple statistical indicators beyond the average annual citation counts to ensure more robust conclusions (Supplementary Fig. S8–S9). For the right tail of the distribution, namely papers with high citation counts, we observe that from year of publication and across subsequent decades, citations for the top 1st and 5th-percentile of AI papers exceed those of non-AI papers by 152.39% and 48.27%, respectively. For the left tail of the distribution, namely papers with low citation counts, we find that, over the same time period, the proportion of AI papers receiving fewer than three and fewer than five citations each year is 2.49% and 2.46% lower, respectively, than non-AI papers. Taken together, these findings suggest that AI papers not only tend to receive higher citations but are also less likely to become low-impact publications, indicating the enhanced visibility and academic attention garnered by AI research.

In addition to the “velocity” of citation, shown above as the number of citations received per year after publication, we also compare the “acceleration” of citations, namely the year-over-year change in annual citation counts, between AI and non-AI papers (Supplementary Fig. S10). Results reveal that during the initial post-publication years when annual citations are generally increasing, AI papers exhibit a faster acceleration in citation growth. In the subsequent period, after reaching peak citation “velocity”, AI papers also exhibit a more rapid deceleration in annual citations. Over the longer term, the changing patterns of annual citations to AI and non-AI papers converge, with no significant differences observed. Nevertheless, throughout the entire citation lifecycle, the citation “velocity” of AI papers consistently remains higher than that of non-AI papers.

Empirically, review articles, editorial pieces, and other types of special publications tend to exhibit markedly different citation patterns compared to original research papers. To verify the robustness of our finding that AI papers garner higher academic impact, we distinguish these special publications from those reporting original research and replicate our analyses exclusively on the latter (Supplementary Fig. S11–S12). Specifically, we filter publications labeled as “article” in the OpenAlex dataset²⁹, thereby excluding journals dedicated to reviews as well as publication types such as letters, editorials, and erratum.

To further eliminate a small number of review articles that may still be included in non-review journals, we additionally filter all papers with “review” or “survey” in their titles. As a result, these special pieces account for 9.26% of the 27,405,011 publications with intact reference records from our original analysis. After excluding them, we obtained a refined dataset of 24,867,012 original research papers. On this subset, AI papers still receive higher visibility and academic attention than their non-AI counterparts across the previously introduced statistical indicators.

In our analysis of annual citations for AI and non-AI papers, we find that both types of papers continue to be cited even 20 or more years after publication. Given that fewer papers, especially AI-related ones, were published several decades ago and that most scholarly works cite “historical” papers to situate their research within a broader context, it is important to understand whether referenced papers are being cited as foundational citations or simply “throwaway” citations that the scholarly crowd uses more colloquially. To address this, we follow an established methodology for distinguishing between core and superficial citations⁷³, and we compute the times AI and non-AI papers are cited as core citations in each year following publication (Supplementary Fig. S13). Results show that the proportion of AI papers to be cited as core citations tends to decrease over time, which aligns with the intuitive expectation that older papers are more likely to be cited for historical context rather than as active intellectual foundations. Nevertheless, foundational influence can still be observed in decades-old papers. For example, two recent studies published in 2020 and 2021, which used AI techniques to predict diabetes^{74,75}, both cite a 1988 paper on the ADAP learning algorithm⁷⁶ as a foundational reference. Both newer papers build upon the prior algorithm to design more advanced ensemble models. Moreover, in terms of absolute counts of being core citations, AI papers still surpass non-AI papers by 98.70%, reflecting the heightened academic attention attracted by AI research.

AI represents a technology that started small and then became the most prominent method in multiple fields. As such, AI papers exhibit a distinctive citation pattern compared to non-AI papers. To investigate whether this citation pattern is unique to AI or also characteristic of other emerging technologies that started small and then became widely appreciated and used, we conduct a comparative analysis using Nanotechnology as case study (Supplementary Fig. S14). We identified Nanotechnology-related research by detecting the presence of the word “nano” in the titles of all publications. As shown, Nanotechnology was sparsely studied in its early stages but began to gain widespread attention across disciplines such as biology, chemistry, and physics beginning in the 1990s. We observe that Nanotechnology exhibits a citation pattern partially similar to AI: Nanotechnology papers are still cited as core citations many years following publication, although the proportion of being such core citations declines over time. In contrast to AI, however, both total number of citations and the frequency of being core citations for Nanotechnology papers eventually decline to levels indistinguishable from non-Nanotechnology papers, suggesting that nanotechnology does not exhibit the same sustained higher academic influence as AI.

To provide a more comprehensive assessment of the impact of AI papers, we analyze additional indicators beyond citation-related statistics. First, we examine the distribution of AI papers across journals of varying Journal Citation Report (JCR) quantiles⁴⁰ (Supplementary Fig. S15). We find that the proportion of AI papers in Q1 journals is 18.60% higher than non-AI ones in all journals, and in Q2 journals, the AI proportion is a scant 1.59% higher, while Q3 and Q4 journals hold a relatively lower proportion of papers with AI. These results indicate a heterogeneous distribution of AI-augmented papers across journals, with a higher prevalence in high-impact journals. Second, we examine the influence of AI and non-AI papers from the perspective of disruption^{77,78} (Supplementary Fig. S16). Despite higher citation counts for AI papers, we find that their disruption scores are lower. This suggests that while AI papers tend to influence a larger number of subsequent studies, the higher impact is primarily developmental rather than disruptive, which means that they contribute to advancing (and completing) existing fields rather than initiating new ones. This observation is consistent with our later finding that AI contracts science’s focus.

2.2 The effect of AI on individual scientists’ careers

As shown in the main text, scientists who adopt AI tend to achieve higher scholarly productivity and impact. However, an alternative explanation for this observation is whether these advantages are driven by AI adoption itself, or whether scientists who are already on a trajectory toward greater productivity and impact are simply more likely to adopt AI. To investigate this, we conducted a match-and-comparison analysis of scientists with similar early-career productivity and impact trajectories, but who differ in their subsequent AI adoption behavior (Supplementary Fig. S17). Specifically, we filter 11,019 scientists who began adopting AI in the third year of their careers and match them with 1,926 scientists who exhibited comparable annual citation counts during their first three years but never adopted AI. The comparison reveals that starting from the third year, when the former group began adopting AI, a divergence in annual citation counts emerges between the two groups. By the tenth year of their careers, the scientists who adopted AI in their third year have 24.45% higher annual citations than their non-AI-adopting counterparts with similar early trajectories. Similarly, their annual productivity in the tenth year is 9.61% higher than the matched group of non-AI-adopters with comparable early-career productivity. The same pattern holds when analyzing a second cohort: 9,837 scientists who adopted AI in the fifth year of their careers were compared with peers who had similar citation and productivity levels in their first five years but never adopted AI. Taken together, these findings suggest that for scientists with

comparable early-career positions, adopting AI itself contributes to their subsequent advantages in productivity and impact.

In our main analysis, we include 2.3 million scientists with complete career trajectories and show that those adopting AI are more productive and receive more citations. Prior research suggests that only a relatively small subset of scientists continue publishing over long periods, however, forming what has been recognized as the global core of active researchers⁴¹. To examine whether our findings hold for this persistent group, we identified 1,495,265 researchers with uninterrupted publication records over at least five consecutive years and 525,716 researchers over at least ten consecutive years, accounting for 52.30% and 18.39% of all researchers in our dataset, respectively (Supplementary Fig. S18). We then compared the productivity and citation impact of AI-adopting versus non-AI-adopting researchers within these subsets. Among those with at least five consecutive years of publications, researchers adopting AI published 2.40 times more papers and received 3.88 times more citations per year than their non-AI counterparts, with consistent patterns observed across disciplines. Similarly, for researchers with at least ten consecutive years of publications, who are more productive and receive more citations than researchers with five consecutive years of publications, AI adopters published 2.41 times more papers and received 4.31 times more citations annually. These results further confirm the positive impact of AI adoption on the career progression of both the continuously publishing core scientists and broader population of normal scientists.

In our career transition model of researchers, we set a threshold of 3 years and regard scientists who have no more publications after 2022 as having exited academia, where the threshold setting is consistent with previous research⁶². To ensure the robustness of our findings, we switch to different thresholds in detecting the dropout of researchers and replicate our results (Supplementary Fig. S19). As we illustrate, when using different dropout thresholds of 2 or 4 years, the probability for AI-adopted junior scientists to transition to established scientists is consistently higher than for their counterparts who do not adopt AI, and the anticipated transition time to becoming established scientists is shorter for AI-adopted junior scientists compared to their counterparts. These robust results underscore the role of AI in bringing about increased opportunities for junior scientists to lead research teams and reducing the risks of their leaving academia. To assess the appropriateness of our threshold for detecting researcher dropout, we analyzed the distribution of gap year durations in researcher careers (Supplementary Fig. S20). Here, a period of gap years refers to a temporary interruption in a researcher's publication activity, followed by a subsequent resumption of publishing. Results show that 44.67% of all gap periods lasted only one year, and 76.94% lasted no more than three years. Therefore, when choosing a three-year threshold for identifying dropout events, it can appropriately capture the majority of cases where researchers temporarily paused and then resumed publication activity. Additionally, as we illustrate, both shorter and longer thresholds yield similar overall results. Nevertheless, a threshold that is too short would misclassify many researchers with temporary gaps as dropouts, while a threshold too long would result in a large number of researchers being classified as having uncertain status and excluded from analysis, thereby reducing the sample size. Taking these considerations into account, we adopt a three-year threshold as a balanced and robust choice.

To better understand whether the adoption of AI affects all groups of scientists uniformly, we provide demographic information on the people who are leaving the field (Supplementary Fig. S21). First, we utilized the "institution" field in the OpenAlex dataset to determine researchers' affiliations and categorized them into demographic groups based on institutional attributes. On the one hand, we categorize institutions by type, such as companies, educational institutions, government agencies, etc. Researchers affiliated with educational institutions exhibit the lowest dropout rates, although differences in dropout rates across institution types are relatively modest. Notably, across all institutional types, researchers who adopt AI show similarly reduced dropout probabilities. On the other hand, we categorize institutions by geographic region, including Africa, Europe, Asia-Pacific, etc. We find that researchers based in Africa and South America have higher dropout probabilities compared with those in Europe, Asia-Pacific, and North America. Furthermore, while AI adoption is associated with reduced dropout rates for researchers in regions such as Asia-Pacific and North America, the benefit is far less pronounced for those in Africa and South America. Second, we follow established methods in prior work and infer gender and ethnicity information based on author names⁷⁹. The results show that White and API (Asian or Pacific Islander) researchers have lower dropout probabilities compared to their Hispanic and Black counterparts. Meanwhile, AI adoption is associated with reduced dropout rates for White and API researchers, while the benefit is much less pronounced for Black researchers. We also observe heterogeneity in the effect between male and female researchers. Male researchers tend to have lower original dropout probabilities and receive greater benefit from AI adoption. In contrast, female researchers exhibit higher original dropout probabilities and gain relatively less from adopting AI. These findings provide preliminary evidence regarding the heterogeneous impact of AI on scientific careers across different demographic groups, suggesting that the benefits of AI are unequally distributed, but the causes and consequences of this heterogeneity merit further investigation.

3 Extended analyses on how AI impacts the overall science

In the main text, we illustrate the conflict between individual and collective incentives to adopt AI in science, where individual scientists receive expanded personal reach and impact, but the knowledge extent of entire scientific fields shrinks to a narrower focus. In this section, we discuss in more detail how AI impacts the entire knowledge extent. First, we illustrate the benefit of

topic diversity (Section 3.1). Second, we discuss multiple factors that may impact the selectivity of AI adoption in different fields and thereby lead to the observed outcome of narrowed focus of AI-augmented research (Section 3.2).

3.1 The benefit of topic diversity

Given our observation of the contracted knowledge extent of AI-augmented research, a natural question arises: Is a broader knowledge space, namely a greater diversity of research topics, indeed beneficial to the overall advancement of scientific knowledge? To examine this, we categorize our selected papers into 252 distinct sub-fields and compute, for each sub-field, the average citation count and disruption score^{77,78} across different eras of AI (Supplementary Fig. S22). Results show that, across different sub-fields and eras of AI, neither citation (as a proxy for impact) nor disruption (as a proxy for novelty) is obviously correlated with the sub-field's knowledge extent. Specifically, the absolute values of Pearson's r are below 0.1 in almost all cases. This indicates that higher topic diversity within a subfield does not dissipate the sub-field's intellectual energy and reduce its scholarly impact or innovative capacity. Therefore, maintaining a broader diversity of research topics does not hinder a sub-field's performance; rather, it likely offers more opportunities for advancing the scientific frontier and provides researchers with a wider array of investigation choices, ultimately benefiting the state of collective knowledge.

3.2 Selectivity of AI adoption across different topics

To gain a deeper understanding of our main findings that "AI in science has become more concentrated around some popular research topics" and that "AI-augmented research focuses on a narrower scope", we examine whether external factors might influence the selectivity of AI adoption across different topics, potentially leading to their disproportionate representation.

Topicality. Certain topics may inherently lend themselves more to AI applications, potentially making them more likely to be represented in AI-assist research. To investigate this possibility, we analyzed the topical correlations between AI and non-AI research (Supplementary Fig. S23). Based on citation relationships between AI and non-AI papers, we found that AI research across various fields is more frequently cited by non-AI studies than by AI studies themselves. This suggests that topics in AI research influence non-AI research rather than forming isolated, self-referential clusters of AI literature. There is no obvious difference in topic selection between AI and non-AI research, indicating that inherent topicality does not account for the disproportionate prevalence of AI-augmented research in different fields.

Original impact. Another possibility concerns whether the topics being squeezed out by AI are likely to be marginalized anyway. That is, does AI tend to concentrate on more fruitful areas rather than topics with lower prior and potential impact? To evaluate this, we categorize our selected papers into 1,883 topics according to the OpenAlex taxonomy. We calculate the average citation count and disruption score^{77,78} of non-AI papers within each topic, obtaining proxies for the topic's original influence and novelty. We then examine, across topics, the correlation between the degree of AI penetration and the original influence and novelty across different AI development eras, where the absolute values of Pearson's r are below 0.1 in all cases (Supplementary Fig. S24). The results show that across topics and AI eras, neither original citation nor original disruption is obviously correlated with the proportion of AI adoption in that topic. This suggests that AI adoption does not selectively favor topics based on their original impact. Instead, AI homogeneously penetrates into topics with both high and low original influence.

Funding priority. Another potential factor that might influence the selectivity of AI adoption across different topics is funding priority, which means that funding agencies may tend to directly support AI research more in some specific research topics. To evaluate this, we merge information from the "grants" field in the OpenAlex dataset and acknowledgment texts in the Web of Science (WOS) dataset^{30,31}, obtaining 31,115,808 coded funding data acknowledgments from 32,437 funding agencies. By categorizing our selected papers into topics as mentioned above, we calculate the deviation of the probability that AI papers are funded in each topic from the average level and obtain the indicator of funding priority to the topics. We then examined, across topics, correlation between the degree of AI penetration and funding priority across different AI development eras, where the absolute values of Pearson's r are below 0.1 in all cases (Supplementary Fig. S25). Results show that across topics and AI eras, there are heterogeneous funding priorities, but funding priority is not obviously correlated with proportion of AI adoption, suggesting that policy-makers' choices do not obviously affect AI adoption on selective topics.

Data abundance. As discussed in the main text, the knowledge extent of entire scientific fields tends to shrink to focus attention on areas most amenable to AI research, such as those with an abundance of data. To directly evaluate the impact of data abundance on the selectivity of AI adoption across topics, we extract and quantify the appearance frequency of data-related terms, such as "data" and "dataset", in titles and abstracts of papers within each topic. We then normalize these frequencies to serve as an indicator of data abundance for each topic. We then examine, across topics, correlation between the degree of AI penetration and data abundance across different AI development eras (Supplementary Fig. S26). Results show that, as expected, data abundance is diverse across topics, while across AI eras, data abundance is positively and significantly ($p < 0.001$ for all eras) correlated with the proportion of AI adoption, suggesting that AI is more likely to be adopted in topics with abundant data resources. Notably, the correlation of AI-use with frequency of mentioned data resources rises with the size and intensity of AI-models from 0.24 in the machine learning era to 0.36 in the deep learning era to 0.43 in the large model era.

In general, data abundance is a major external factor influencing the selectivity of AI adoption across different topics, contributing to the observed disproportionate representation within knowledge space and the contraction of scientific focus, whereas the other external variables discussed above appear to be largely unrelated. This finding elucidates factors underlying the heterogeneous adoption of AI across topics and provides valuable insights into how AI can be better leveraged to promote sustainable and comprehensive scientific development.

Given the selectivity of AI across topics, we further examined the consistency of our findings regarding the contracted knowledge extent of AI-augmented research in different topics. We conducted stratified analyses to control for the effects of the external factors discussed above, which may impact the finding as confounding variables (Supplementary Fig. S27). Results show that when we control each external factor by partitioning them into 10 tiers based on magnitude, contraction in knowledge extent for AI-augmented research remains evident within each tier. Regardless of whether a topic is more or less likely to be selected for AI adoption, our findings hold consistently: existing AI-augmented research within the topic tends to cover a contracted knowledge space compared to non-AI research.

4 Robustness analyses

4.1 Robustness of results in the generative AI era

Because generative AI has not been around for very long and there remains a lead time before publications using them appear, the available data regarding generative AI is necessarily limited for analysis. To verify the robustness of our general findings over decades of AI development to the specific era of generative AI, we replicated our findings with only the subset of papers published during the generative AI era. Of the 41,298,433 publications in our OpenAlex and Web of Science analyses that cover six disciplines, 4,797,614 are published in the generative AI era. As we show in Supplementary Fig. S28-S30, all results in the generative AI era are consistent with those from prior decades of AI development.

This separate analysis of AI from the generative AI era could exhibit limitations due to the lack of data caused by the relatively short history of generative AI, including LLMs. One major point is that our method of detecting scientists' career role transitions requires scientists to maintain a certain length of publication history, which is not feasible separately in the short generative AI era. Therefore, we are not able to replicate the analysis regarding the career development of individual scientists. Although reaching quantitatively consistent conclusions, with the limited data currently available, results in the generative AI era may appear less significant than corresponding results in our general findings and serve as a starting point for future research as these new models develop and are deployed in new ways. As large, generative models evolve over longer periods of time and produce richer data, additional evaluations should be pursued, further revealing how generative AI impacts scientific development consistent with traditional machine learning techniques or representing new directions.

4.2 Robustness of results on the Web of Science dataset

To verify the robustness of our findings on a more restricted dataset of high-quality papers, we replicated our analysis using the subset of articles from the OpenAlex that also appear in the Web of Science (WOS) dataset^{30,31}. To extract the publications in WOS, we subsetting the OpenAlex publications that could be linked to the WOS dataset based on a shared DOI or a PubMed Identifier (PMID). Of the 41,298,433 publications in our OpenAlex analyses that cover six disciplines, 23,576,370 could be linked to a WOS publication. As we show in Supplementary Fig. S31-S33, all of the results on the WOS dataset are consistent with those from the OpenAlex dataset in the main text.

4.3 Robustness of distance calculations in the high-dimensional paper embedding space

Because multiple important analyses in the main text are presented based on the distance calculated in the paper embedding space, it is crucial to ensure reliability of our distance calculations for high-dimensional vectors. To evaluate this, we utilize a principal components analysis (PCA) approach. We first identified the number of principal components required to explain 90% of the total variance in the original 768-dimensional embeddings, which we determined to be 135. We then projected the original embeddings onto these principal components and recalculated distances in this reduced-dimensional space to replicate our analysis. As we show in Supplementary Fig. S34-S35, all results in the reduced-dimensional space are consistent with those from the original embedding space reported in the main text.

In addition to replicating our analysis in the reduced-dimensional space, we further tested the sensitivity of our distance calculations. Using the same PCA method, we reduce the original embeddings to 86, 135, and 197 dimensions, which correspond to the principal components that explain 80%, 90%, and 95% of the total variance, respectively. We then randomly sample 1,000 NonAI-NonAI, 1,000 AI-AI, and 1,000 NonAI-AI paper pairs and calculate the distance for each pair within each of the reduced-dimensional spaces. As we show in Supplementary Fig. S36, distances in the various reduced-dimensional spaces are highly correlated with those in the original space, with Pearson's r greater than 0.95 ($p < 0.001$ for all groups).

Besides the sensitivity to embedding dimension, we also check the robustness regarding different distance measurements. We substitute all Euclidean distance into Cosine distance and replicate our analyses. As we show in Supplementary Fig. S37,

all results with Cosine distance are consistent with the original ones with Euclidean distance in the main text. These results demonstrate the robustness of our distance calculations based on paper embeddings, ensuring the reliability of our distance-based analyses and suggesting the potential for future research to leverage these distance metrics for further analyses.

Supplementary Figures

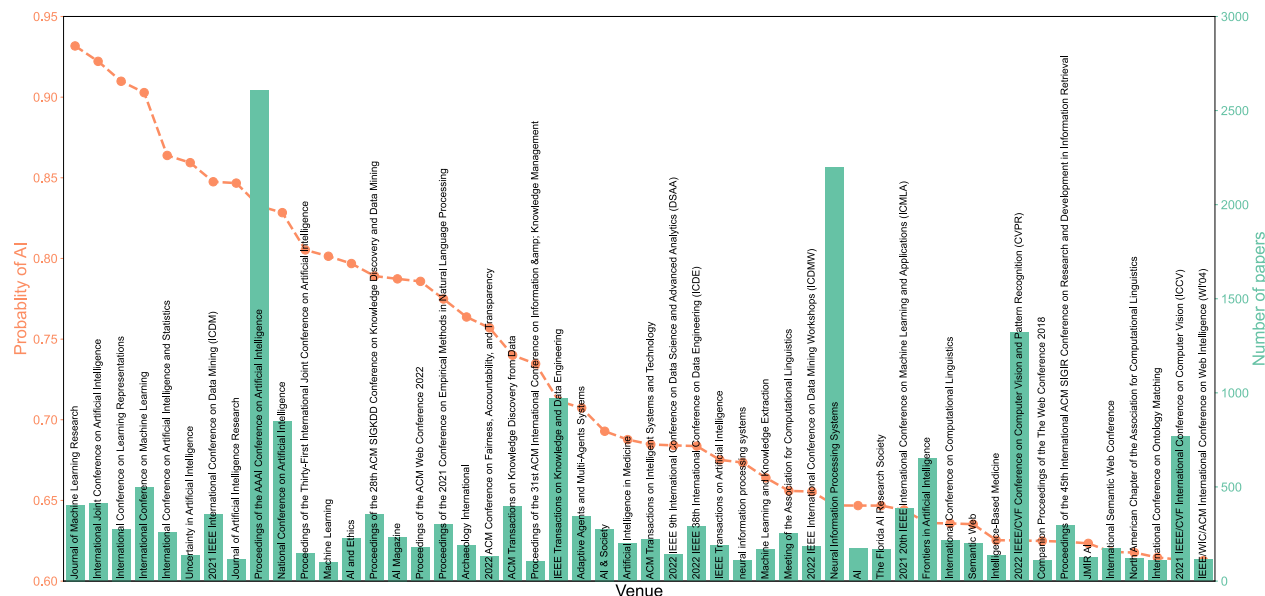


Figure 1. Probability of AI (orange) and number of papers (green) for selected venues. Final identification results combine the best models in both stages of fine-tuning. Venues are ordered according to the probability of AI papers within them.

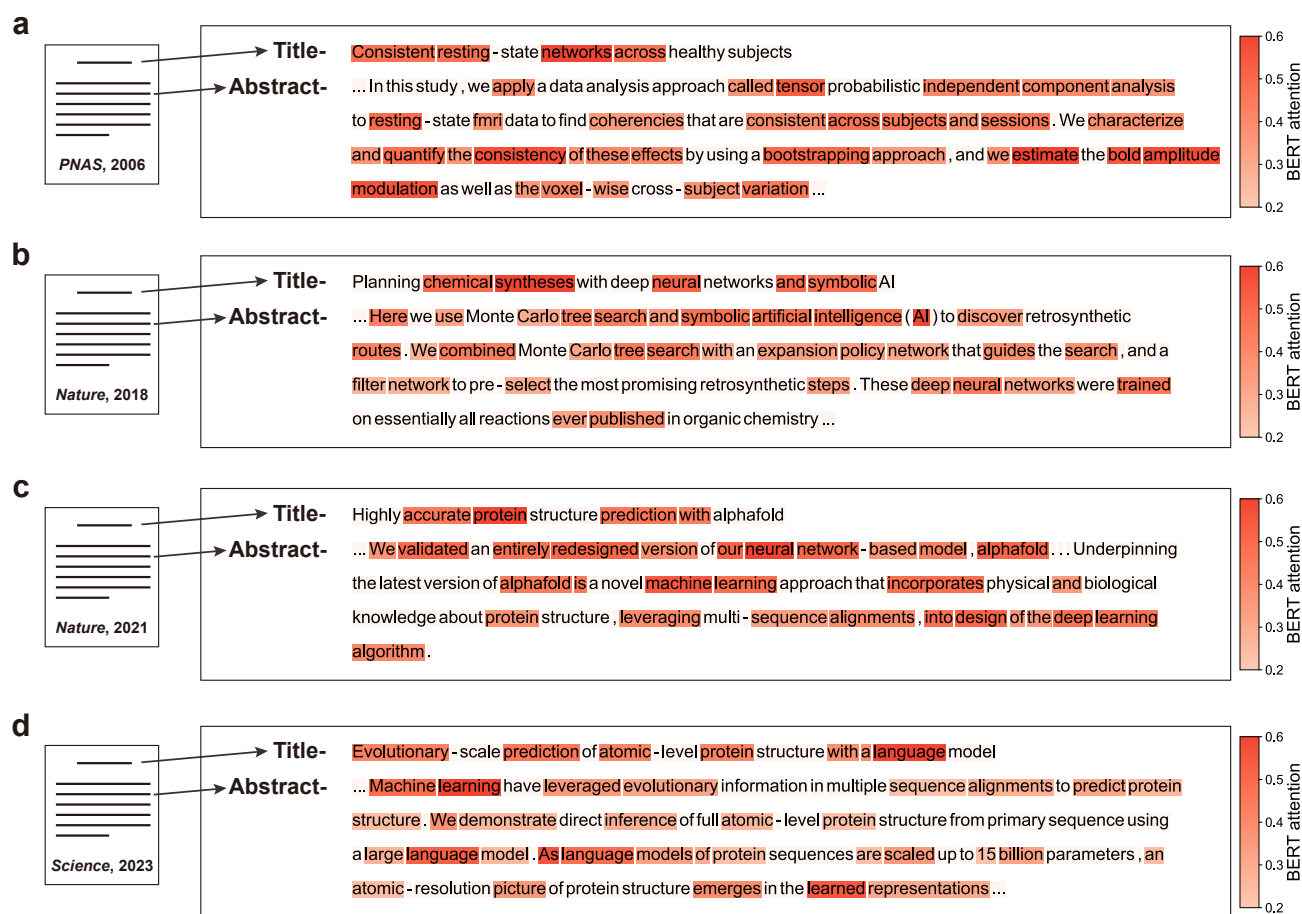


Figure 2. Correct examples of AI paper identification across different eras of AI development. (a) Example from the ML era, a medicine paper with AI published in *PNAS* 2006. (b) Example from the DL era, a chemistry paper with AI published in *Nature* 2018. (c) Example from the DL era, a biology paper with AI published in *Nature* 2021. (d) Example from the GAI era, a biology paper with AI published in *Science* 2023.

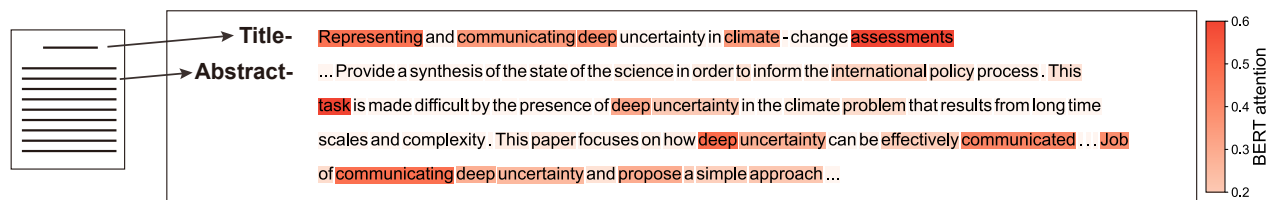


Figure 3. An incorrect example of AI paper identification. The example is a geology paper without AI published in 2006, which is mistaken to be AI-related by the identification model.

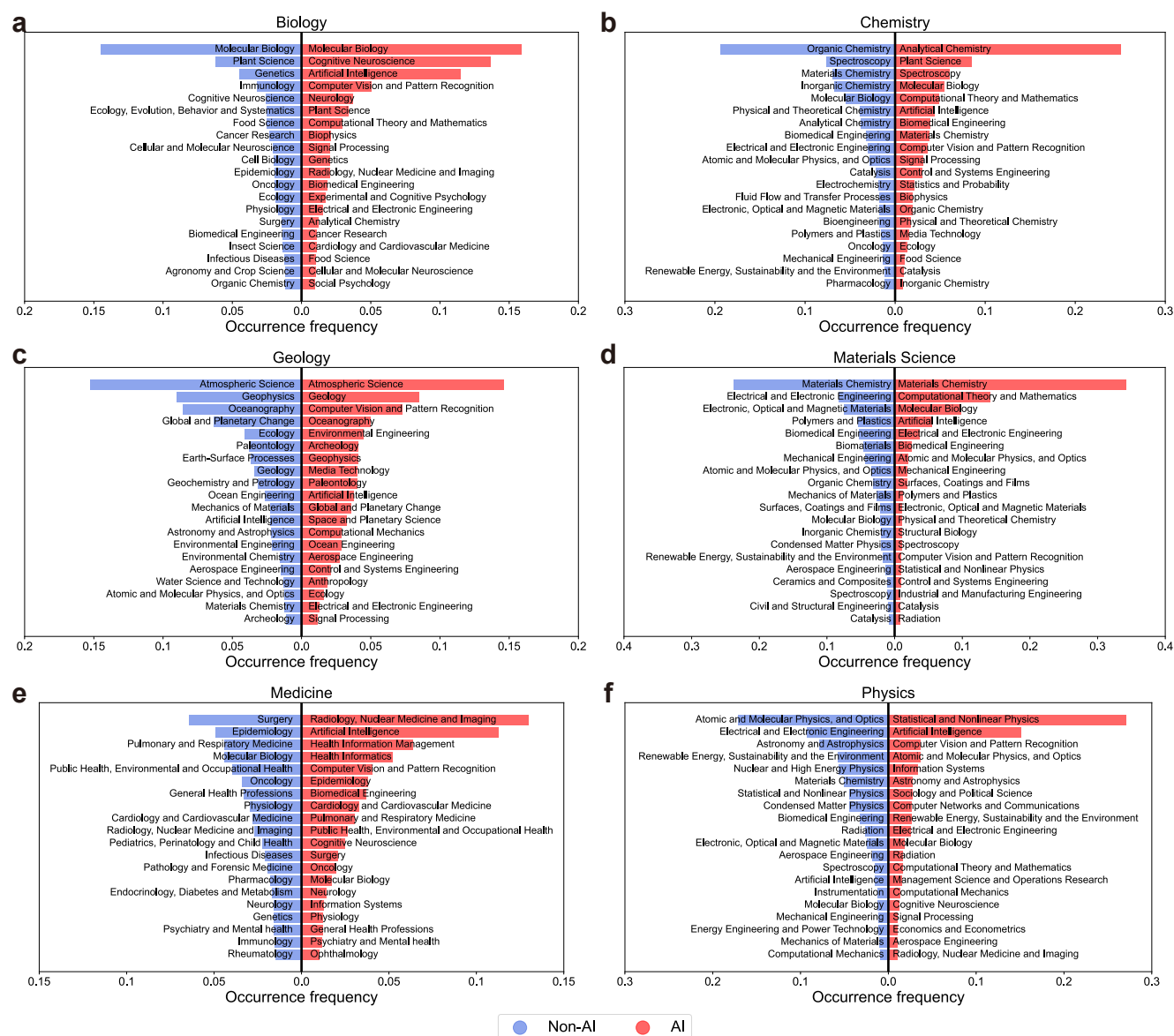


Figure 4. Topics with top occurrence frequency in AI and non-AI papers. Results show that AI’s primary contribution to conventional research fields is around computer science and machine learning algorithms. Across disciplines, identified AI papers turn out to be selected combinations of conventional research topics and AI-related techniques, including “Artificial Intelligence”, “Computer Vision and Pattern Recognition”, and “Computational Theory and Mathematics”.

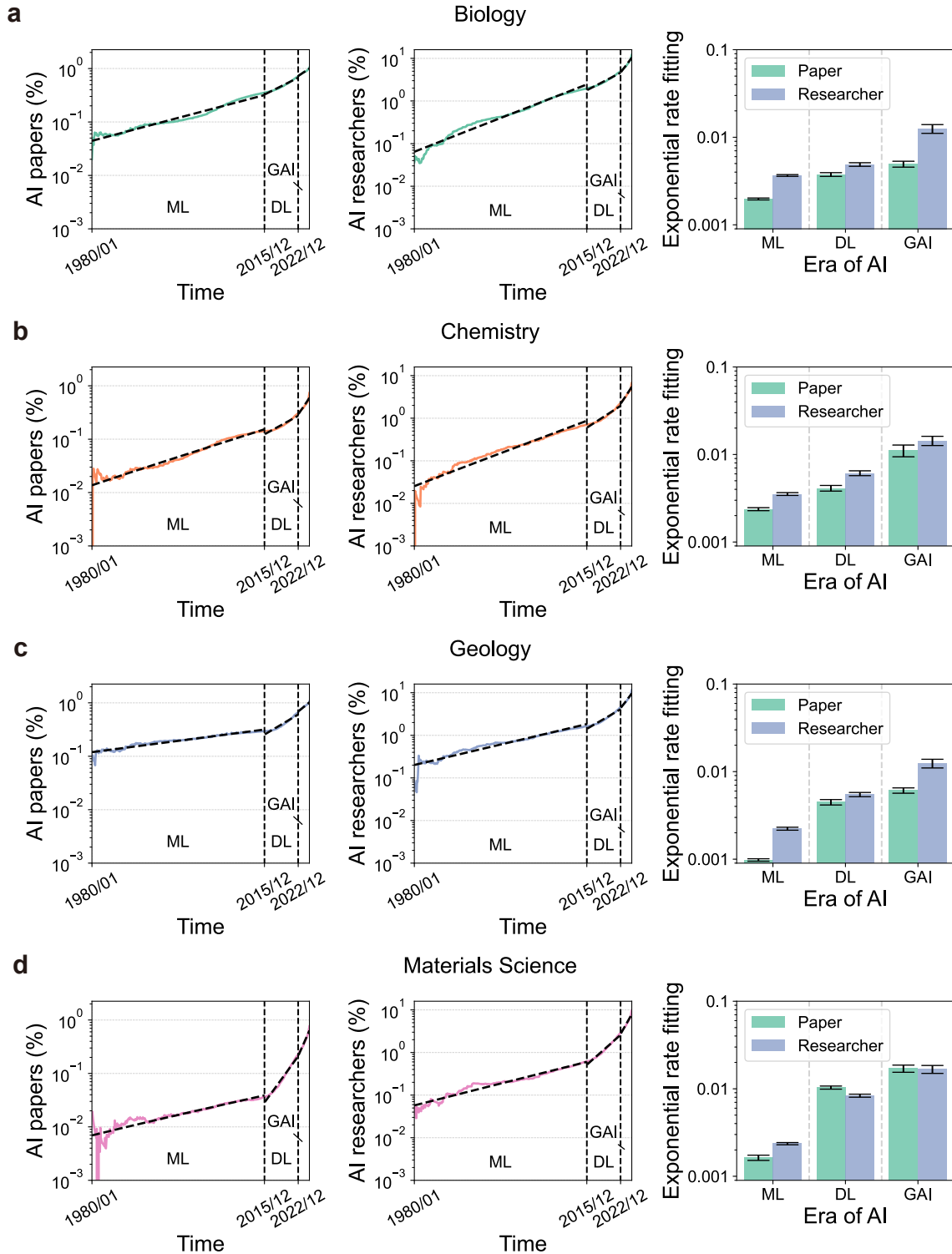


Figure 5. Proportion of papers and researchers adopting AI in each discipline with log-scale y-axis and estimated growth rate with exponential fitting for each era. Proportion of papers and researchers adopting AI fit to straight regions in the log-scale plot, indicating high precision of the exponential fitting. Meanwhile, estimated by the exponential fitting, the growth rate of AI papers and AI researchers in all disciplines increased progressively from the ML era to the DL and GAI eras. For the estimated growth rates ($n = 543$ month observations), 99% CIs are shown as error bars centred at the mean.

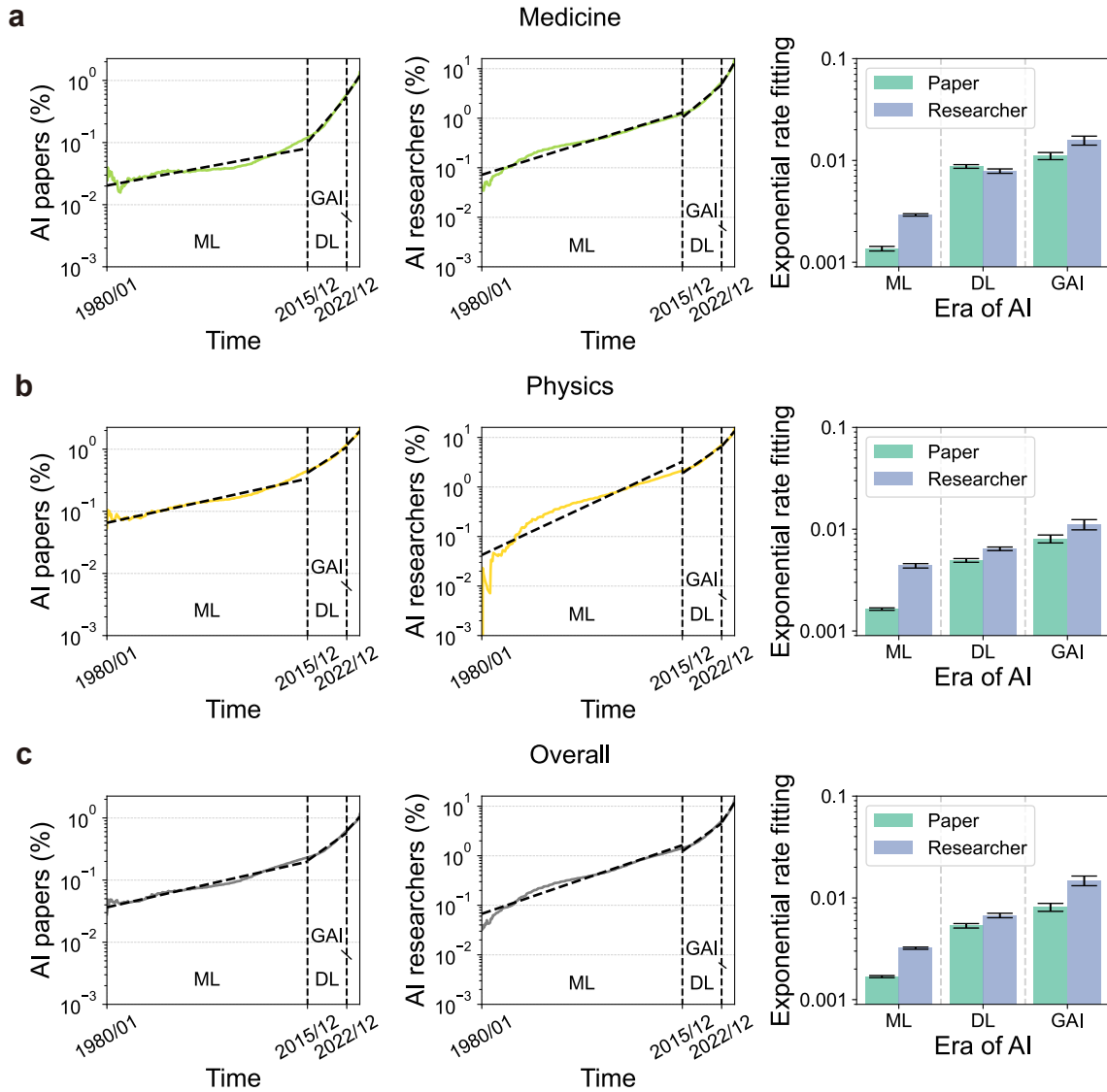


Figure 6. (Continued from Supplementary Fig. S5) The proportion of papers and researchers adopting AI in each discipline with log-scale y-axis and estimate the growth rate with exponential fitting in each era. The proportion of papers and researchers adopting AI fit to straight regions in the log-scale plot, indicating high precision of the exponential fitting. Meanwhile, estimated by the exponential fitting, the growth rate of AI papers and AI researchers in all disciplines increased progressively from the ML era to the DL and GAI eras. For the estimated growth rates ($n = 543$ month observations), 99% CIs are shown as error bars centred at the mean.

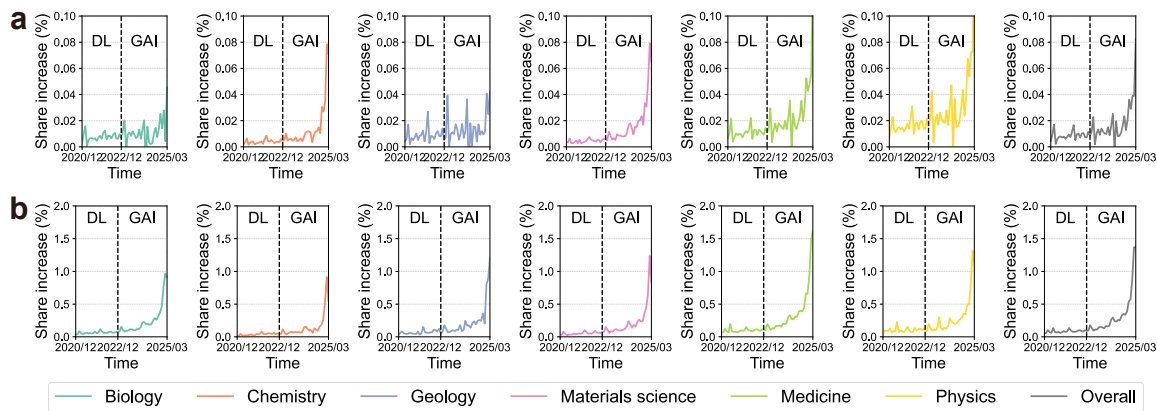


Figure 7. The monthly increase in the proportion of AI (a) papers and (b) researchers across disciplines in recent years. Following the release of ChatGPT in December 2022, growth rates in the proportion of papers and researchers initially remain consistent with prior trends. A period of time later, these growth rates begin to exhibit a marked acceleration across disciplines, providing evidence for the impact of generative AI advancement and use.

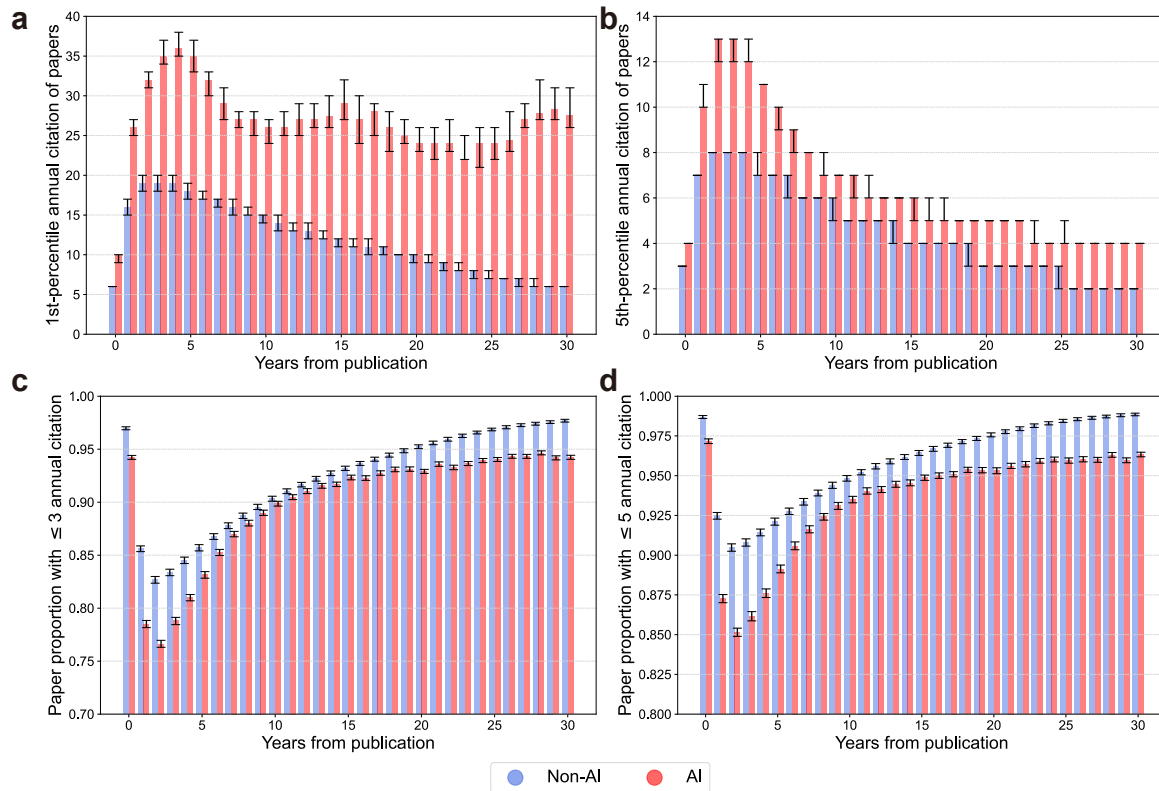


Figure 8. Alternative statistical indicators for the annual citation comparison between AI and non-AI papers. (a) Top 1st-percentile annual citation for AI and non-AI papers from year of publication ($n = 27,405,011$). (b) Top 5th-percentile annual citation for AI and non-AI papers from year of publication ($n = 27,405,011$). (c) Proportion of AI and non-AI papers receiving fewer than three citations each year from year of publication ($n = 27,405,011$). (d) Proportion of AI and non-AI papers receiving fewer than five citations each year from year of publication ($n = 27,405,011$). For all panels, 99% CIs are shown as error bars centred at the corresponding percentiles or proportions.

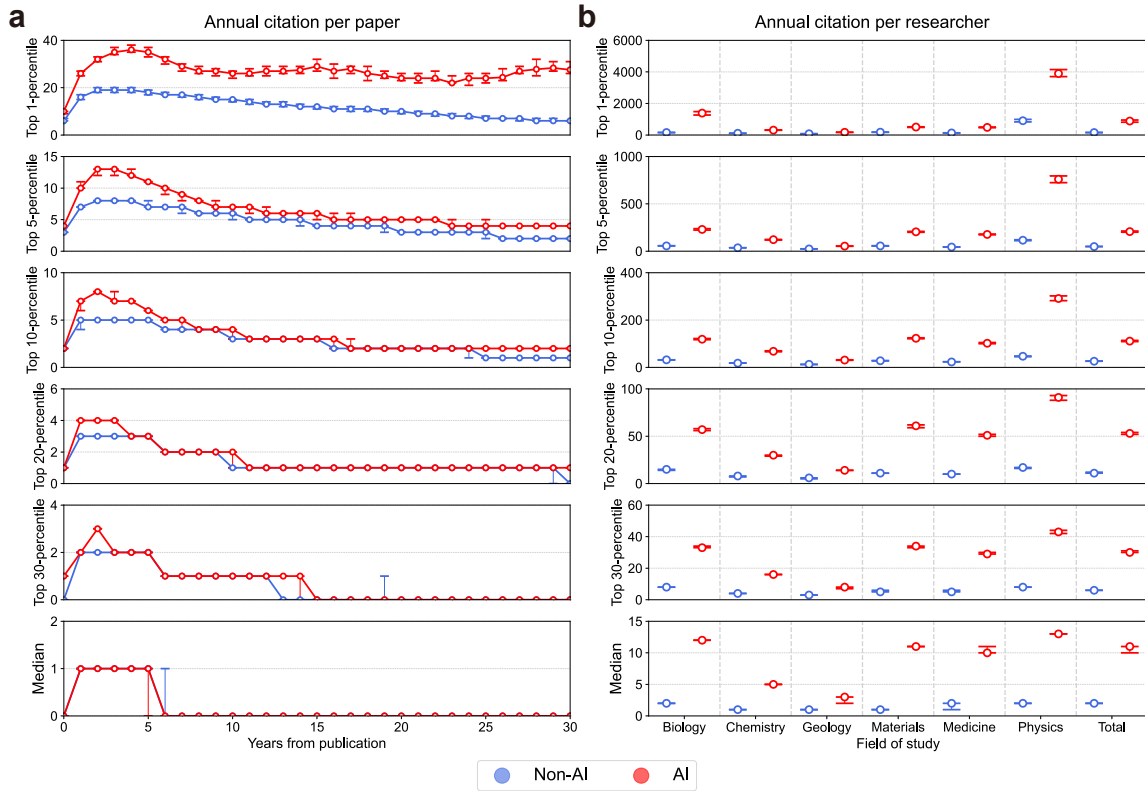


Figure 9. Different percentile statistics for the annual citation comparison between AI and non-AI papers. (a) Annual citations after publication of AI and non-AI papers ($n = 27,405,011$). **(b)** Annual citations for researchers adopting AI and their counterparts without AI ($n = 5,377,346$). Consistently indicated by different percentile statistics, AI papers and researchers attract more citations. For all panels, 99% CIs are shown as error bars centred at the corresponding percentiles.

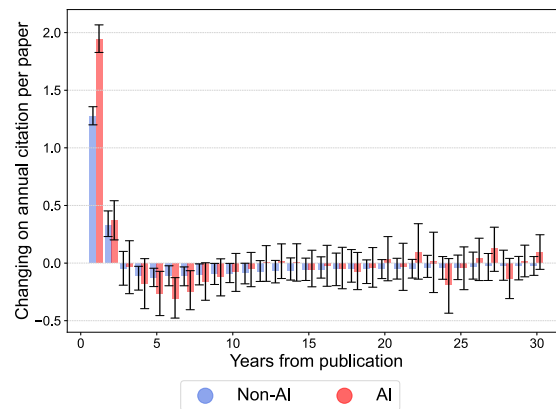


Figure 10. Comparison of “acceleration” of citation, namely year-over-year change in annual citation counts, between AI and non-AI papers. AI papers experience a faster acceleration in citation growth during the initial post-publication years, and exhibit a more rapid deceleration in annual citations after reaching peak annual citations ($n = 27,405,011$). Throughout the entire citation lifecycle, the annual citation of AI papers consistently remains higher than that of non-AI papers. 99% CIs are shown as error bars centred at the mean.

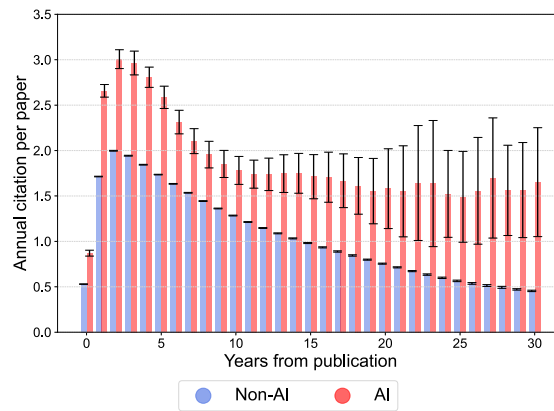


Figure 11. Annual citations after the publication of AI and non-AI original research papers, excluding review articles, editorial pieces and other special publications. Results show that AI papers attract more citations, indicating higher academic impact than papers without AI ($n = 24,867,012$). Results are consistent with the overall statistics in the main text. 99% CIs are shown as error bars centred at the mean.

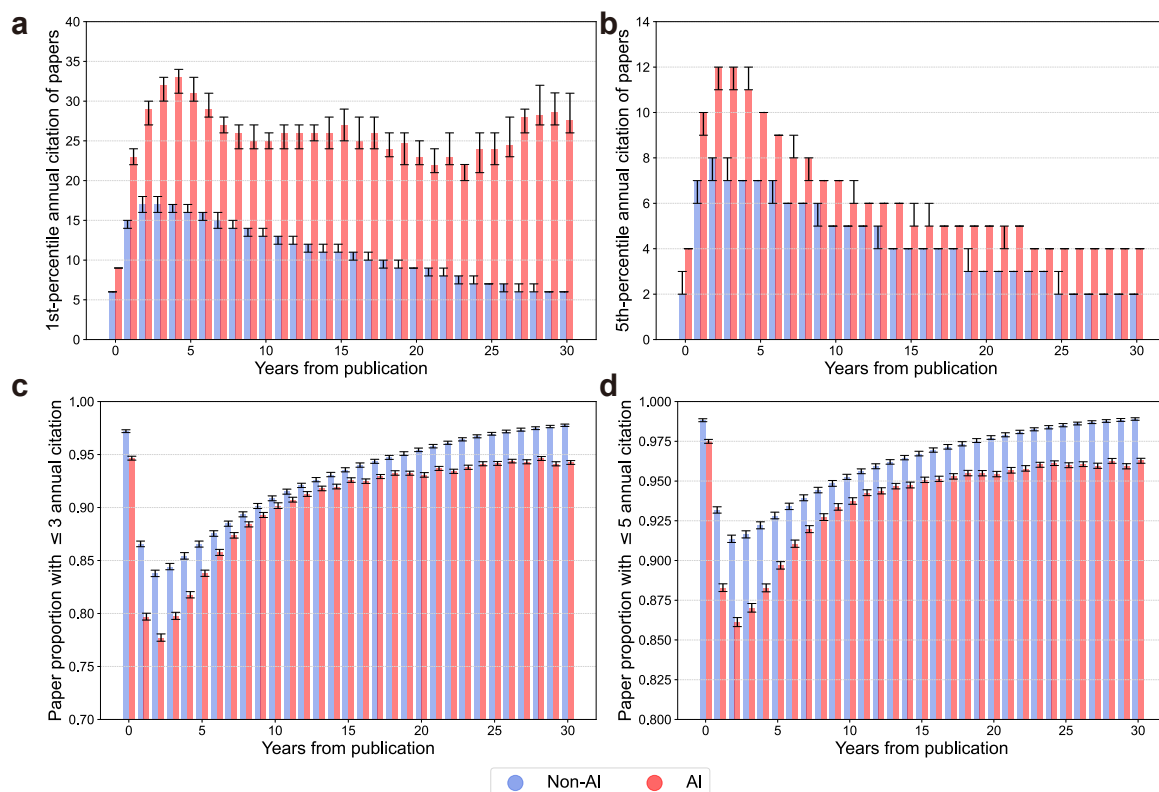


Figure 12. Alternative statistical indicators for the annual citation comparison between AI and non-AI original research papers, excluding review articles, editorial pieces and other special publications. (a) Top 1st-percentile annual citation for AI and non-AI papers from year of publication ($n = 24,867,012$). **(b)** Top 5th-percentile annual citation for AI and non-AI papers from year of publication ($n = 24,867,012$). **(c)** Proportion of AI and non-AI papers receiving fewer than three citations each year from year of publication ($n = 24,867,012$). **(d)** Proportion of AI and non-AI papers receiving fewer than five citations each year from year of publication ($n = 24,867,012$). Results are consistent with the overall statistics in the main text. For all panels, 99% CIs are shown as error bars centred at the mean.

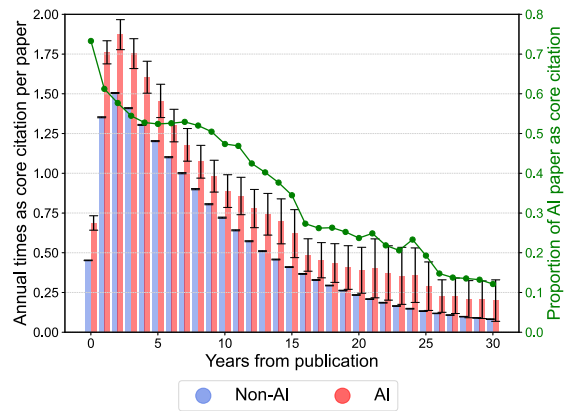


Figure 13. Distinguishing between core and superficial citations. Results show that the proportion of AI papers cited as core citations tends to decrease over time (green line), while foundational influence can still be observed in decades-old papers ($n = 27,405,011$). In terms of the absolute count of instances being among the core citations in future papers, AI papers still outperform non-AI papers (red and blue bars), reflecting their enhanced academic impact. 99% CIs are shown as error bars centred at the mean.

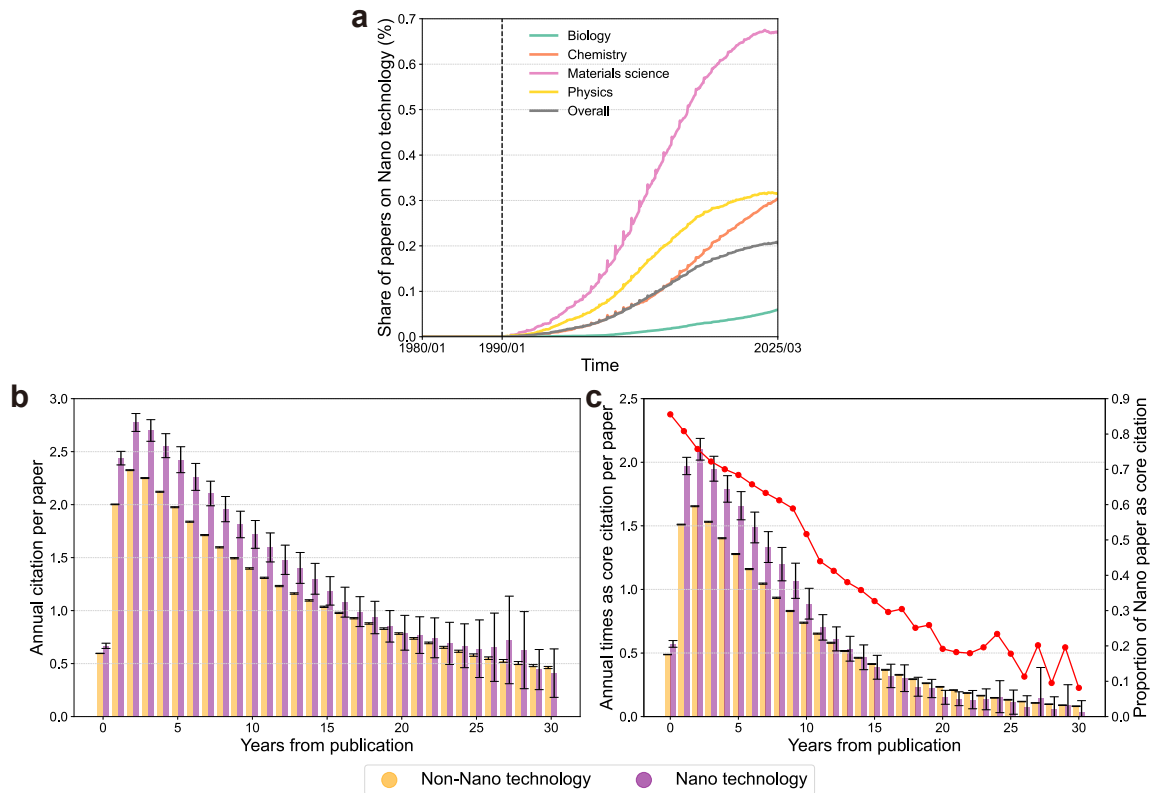


Figure 14. Comparison between citation patterns of Nanotechnology and non-Nanotechnology papers. (a) Growth of Nanotechnology over time in different disciplines. **(b)** Annual citations after publication of Nanotechnology and non-Nanotechnology papers ($n = 17,976,303$). **(c)** Annual times of being core citation following publication of Nanotechnology and non-Nanotechnology papers ($n = 17,976,303$). Nanotechnology papers are still cited as core citations many years after publication, with the proportion of being among the core citations declining over time. In contrast to AI, both total number of citations and the frequency of being core citations for Nanotechnology papers eventually decline to levels indistinguishable from non-Nanotechnology papers. For panels (b) and (c), 99% CIs are shown as error bars centred at the mean.

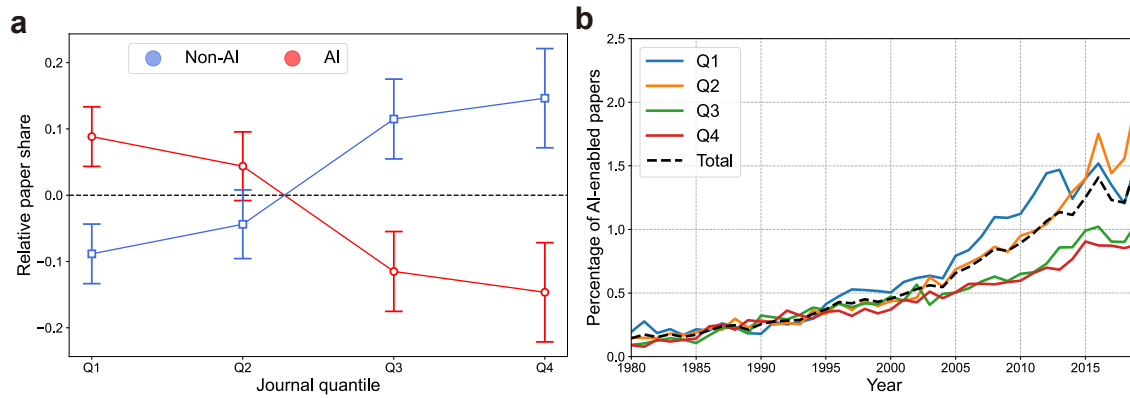


Figure 15. Distribution of AI papers across journals of varying Journal Citation Report (JCR) quantiles. (a)

Comparison of the relative share of AI and non-AI papers published in different journals ($n = 11,098$), where 99% CIs are shown as error bars centred at the mean. **(b)** The change in the percentage of AI papers in journals with different JCR quantiles. The percentages of AI papers rise in all journals, and the percentages of AI papers in Q1 (blue) and Q2 (orange) journals are higher than the total percentage in all journals (black). These results indicate that AI-augmented papers are more likely to be published in high-impact journals (Q1 and Q2) than papers without AI. These statistics are obtained based on papers published before 2021, comparable with the 2021 JCR quantile data we used.

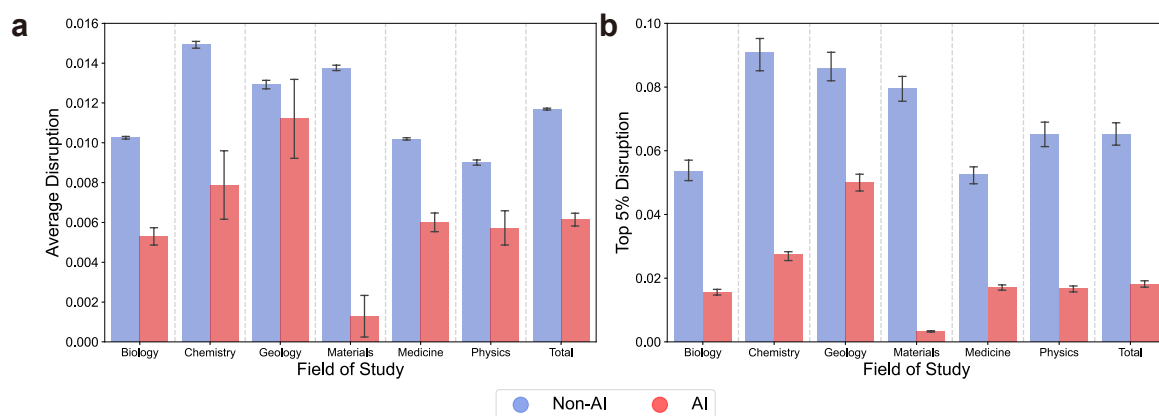


Figure 16. Comparison of disruption between AI and non-AI papers. (a) Average disruption of AI and non-AI papers in each field ($p < 0.001, n = 23,199,583$). **(b)** Top 5th-percentile disruption of AI and non-AI papers in each field ($p < 0.001, n = 23,199,583$). 99% CIs are shown as error bars centred at the mean or the 5% percentile. All statistical tests use a two-sided t-test.

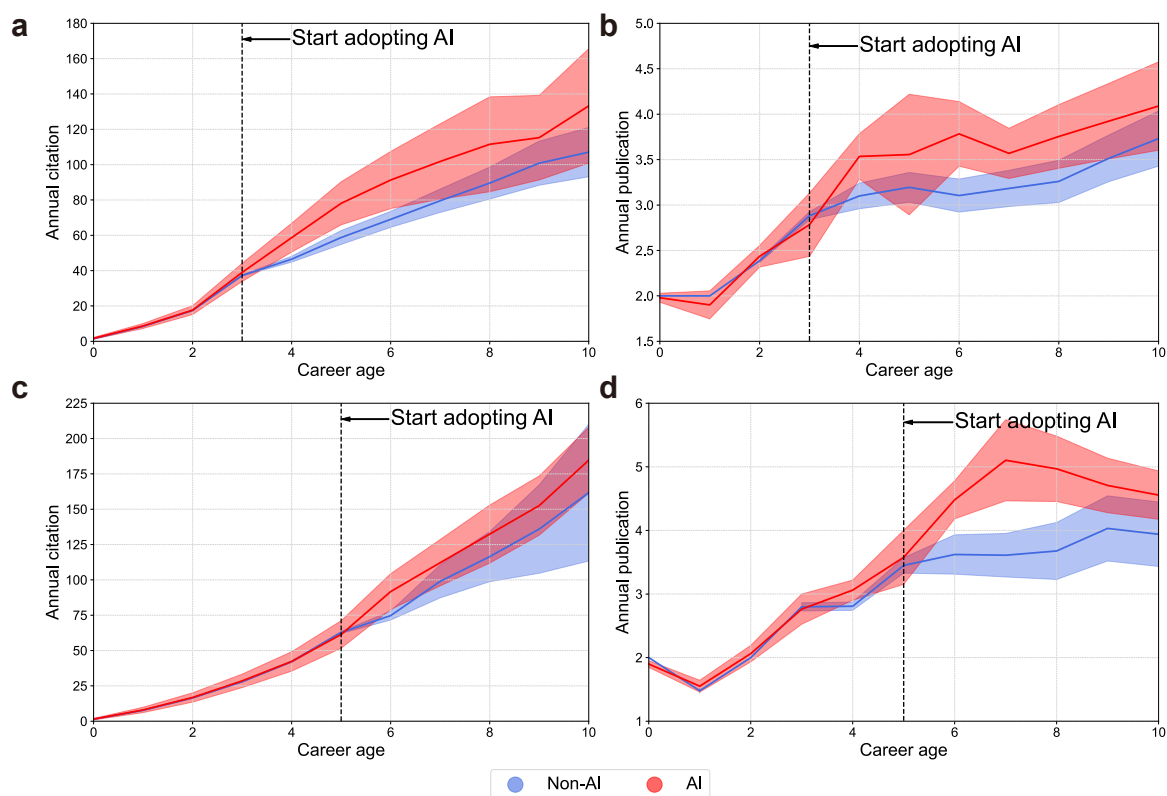


Figure 17. Match-and-comparison analysis on the role of AI in driving the advantages of scientific in productivity and scholarly impact. (a) Comparison between scientists who began adopting AI in the third year of their careers and matched counterparts who exhibited comparable annual citation counts during their first three years but never adopted AI. (b) Comparison between scientists who began adopting AI in the third year of their careers and matched counterparts who exhibited comparable annual productivity during their first three years but never adopted AI. (c) Comparison between scientists who began adopting AI in the fifth year of their careers and matched counterparts who exhibited comparable annual citation counts during their first three years but never adopted AI. (d) Comparison between scientists who began adopting AI in the fifth year of their careers and matched counterparts who exhibited comparable annual productivity during their first three years but never adopted AI. The results suggest that for scientists with comparable early-career positions, adopting AI itself contributes to their subsequent advantages in productivity and scholarly impact. For all panels, 99% CIs are shown as error bands centred at the mean.

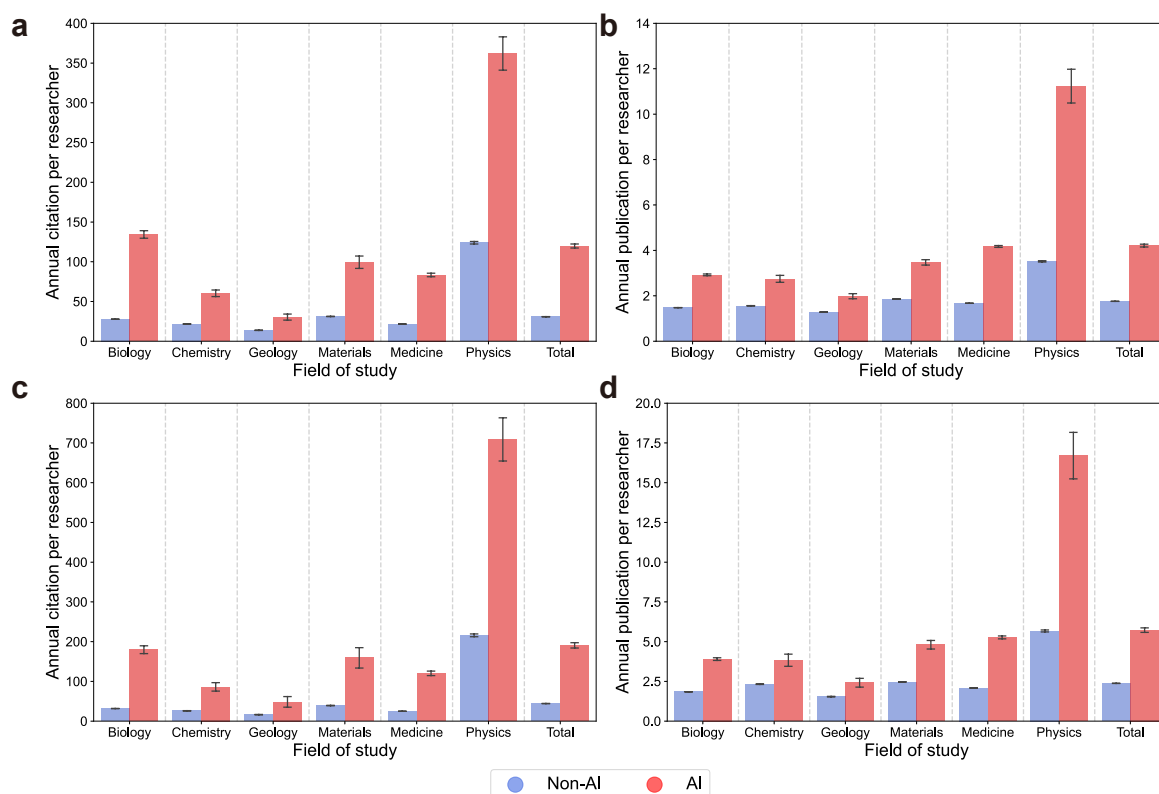


Figure 18. The impact of AI on the productivity and citation of researchers with multi-year continuous publication records. (a) Comparison of annual citations between researchers adopting AI and their counterparts without AI among those with at least five consecutive years of publications ($p < 0.001, n = 1,495,265$). (b) Comparison of annual publications between researchers adopting AI and their counterparts without AI among those with at least five consecutive years of publications ($p < 0.001, n = 1,495,265$). (c) Comparison of annual citations between researchers adopting AI and their counterparts without AI among those with at least ten consecutive years of publications ($p < 0.001, n = 525,716$). (d) Comparison of annual publications between researchers adopting AI and their counterparts without AI among those with at least ten consecutive years of publications ($p < 0.001, n = 525,716$). These results confirm the positive impact of AI adoption on the career progression of continuously publishing core scientists. For all panels, 99% CIs are shown as error bars centred at the mean. All statistical tests use a two-sided t-test.

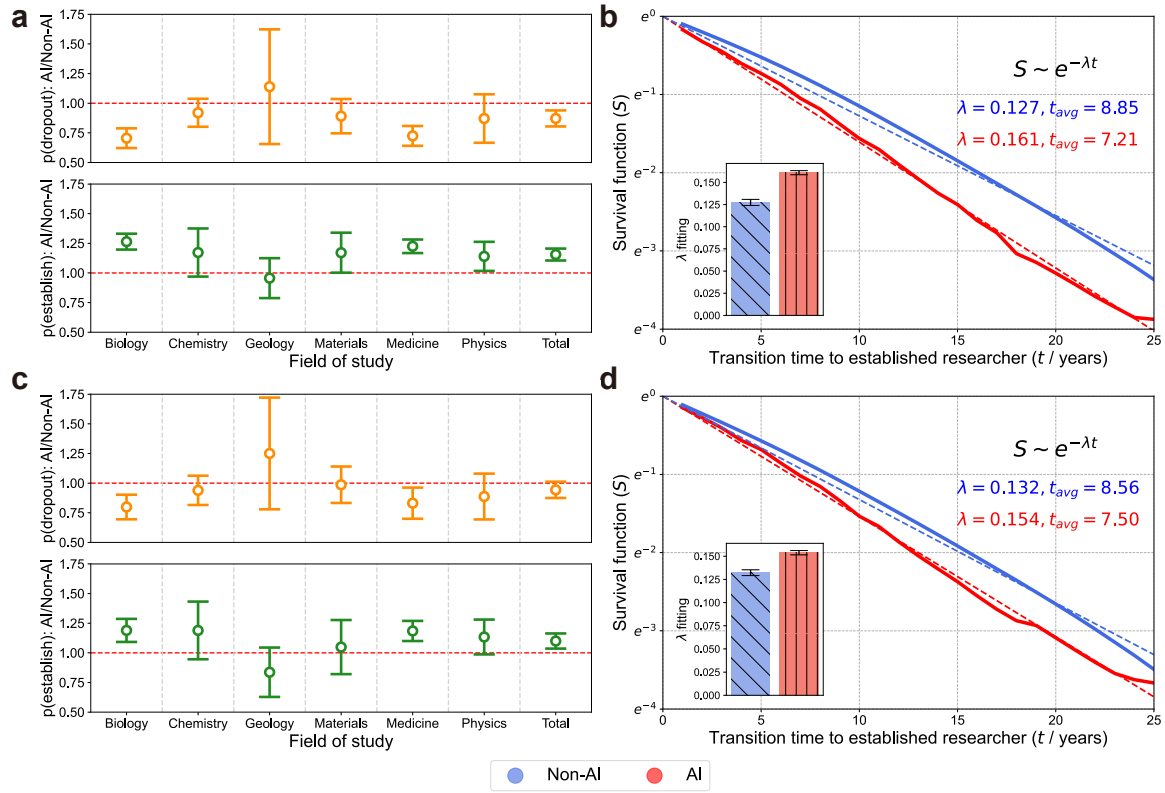


Figure 19. Impact of AI adoption on researchers' careers with different thresholds in detecting the dropout of researchers. (a) (c) The probability of two role transitions between junior scientists adopting AI and their counterparts without AI ($n = 46$ year observations for each field). (b) (d) Survival functions for the transition from junior to established researcher (b, $p < 0.001, n = 2,858,901$; d, $p < 0.001, n = 1,947,315$). In (a) and (b), the threshold for detecting researcher dropout is 2 years. In (c) and (d), the threshold for detecting researcher dropout is 4 years. Results are consistent for both shorter and longer thresholds. For all panels, 99% CIs are shown as error bars centred at the mean. All statistical tests use a two-sided t-test.

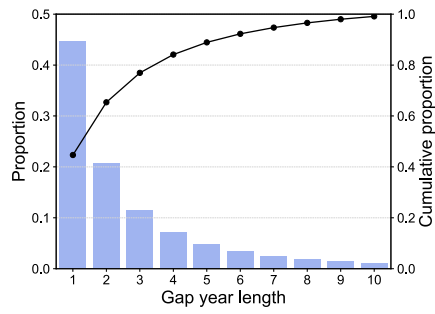


Figure 20. The distribution of gap year durations in researcher's careers. A period of gap years refers to a temporary interruption in a researcher's publication activity, followed by a subsequent resumption of publishing. Results show that 44.67% of all gap periods lasted only one year, and 76.94% lasted no more than three years.

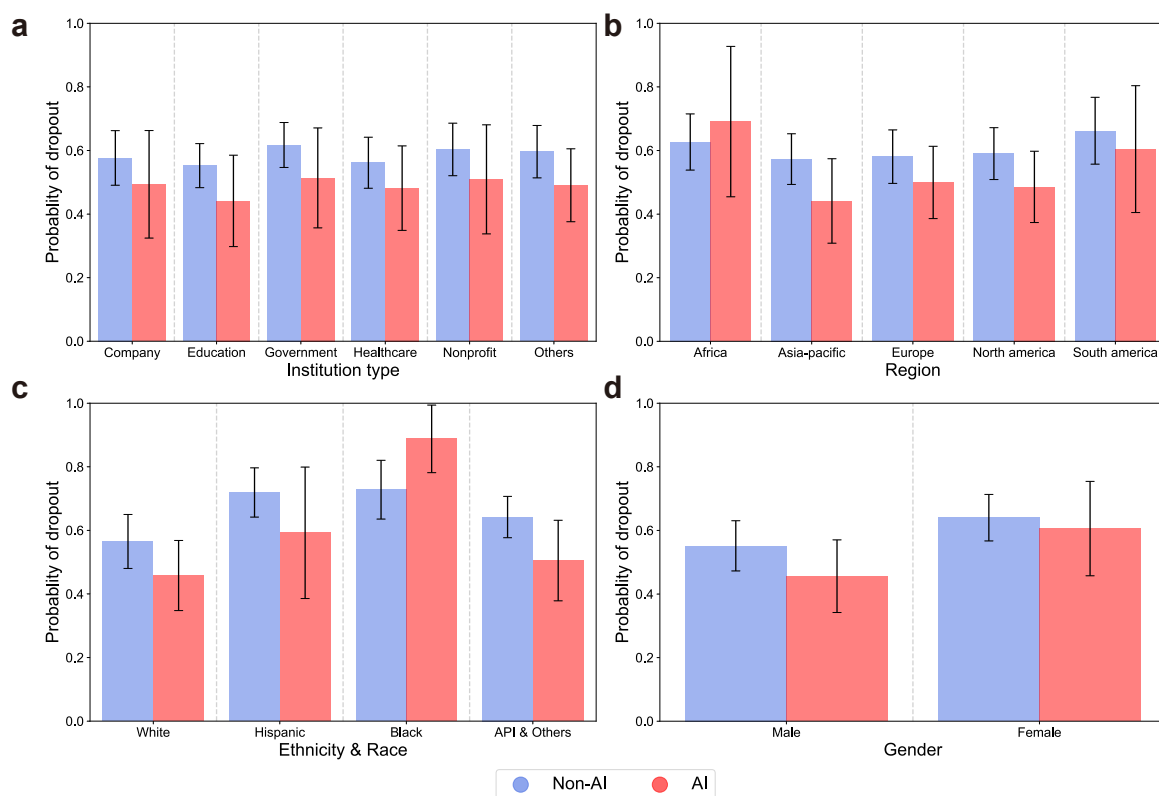


Figure 21. Impact of AI adoption on research careers across different demographic groups. (a) Researchers affiliated with institutions of different types ($n = 2,140,845$). (b) Researchers affiliated with institutions within different geographic regions ($n = 2,140,845$). (c) Researchers of different ethnicity and race ($n = 1,438,544$). (d) Researchers of different genders ($n = 502,336$). Results show the heterogeneous impact of AI on scientific careers across different demographic groups. For all panels, 99% CIs are shown as error bars centred at the mean.

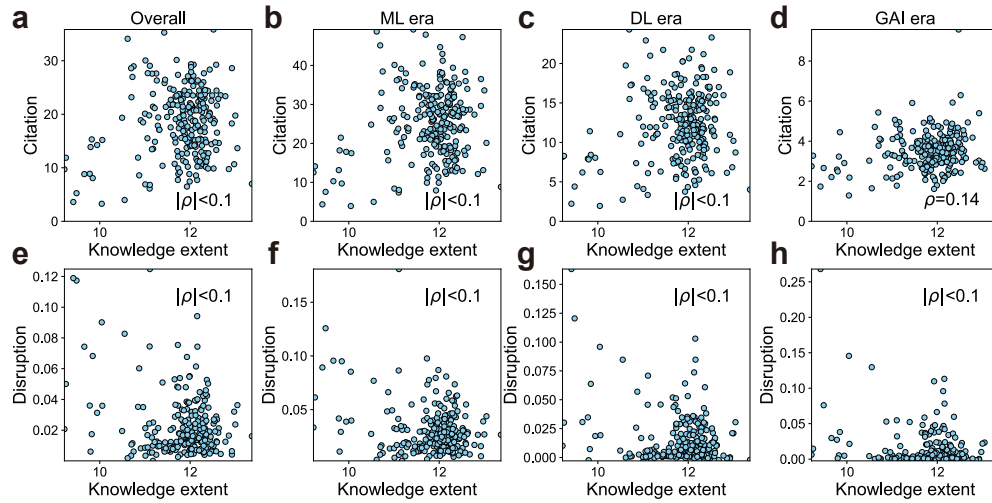


Figure 22. Correlation analysis between knowledge extent and citation or disruption of sub-fields. For all panels, each scatter represents a sub-field, and the corresponding Pearson’s r is indicated correspondingly. Results show that, across sub-fields and eras of AI development, neither citation impact (as a proxy for impact) nor disruption (as a proxy for novelty) is obviously correlated with the sub-field’s knowledge extent.

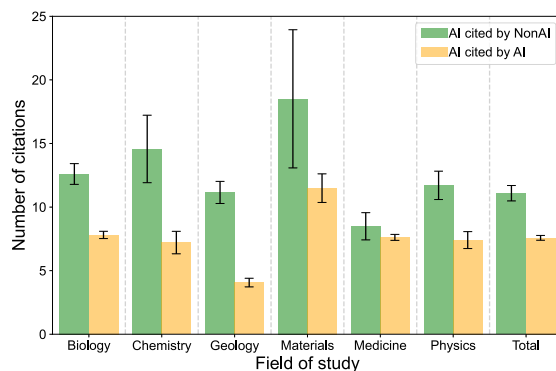


Figure 23. Selectivity of AI adoption in different topics regarding topicality itself. The citation relationships between AI and non-AI papers show that AI research across various fields is more frequently cited by non-AI studies than by AI studies themselves ($p < 0.001, n = 27,405,011$). This suggests that topics in AI research influence non-AI research rather than forming isolated, self-referential clusters of AI literature. There tends to be no obvious difference in topic selection between AI and non-AI research, indicating that inherent topicality does not account for the observed more narrow focus of AI-augmented research. 99% CIs are shown as error bars centred at the mean, and the statistical tests use a two-sided t-test.

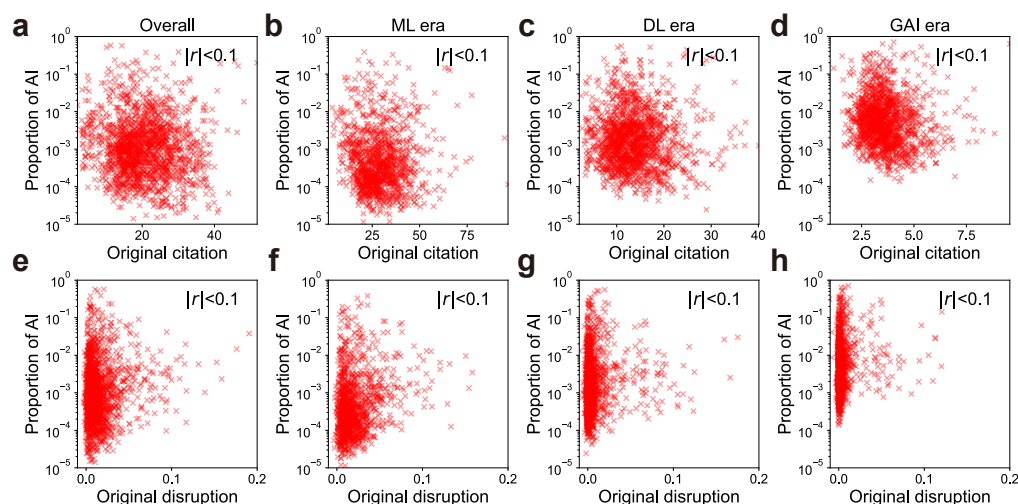


Figure 24. Selectivity of AI adoption in different topics associated with the topics' original impact. In all panels, each scatter shows the degree of AI penetration in a topic and the original citation count or disruption score within that topic. The corresponding Pearson's r is indicated in each panel. Results show that across topics and AI eras, neither the original citation nor the original disruption is obviously correlated with the proportion of AI adoption, suggesting that AI adoption does not selectively favor topics based on differential original impact.

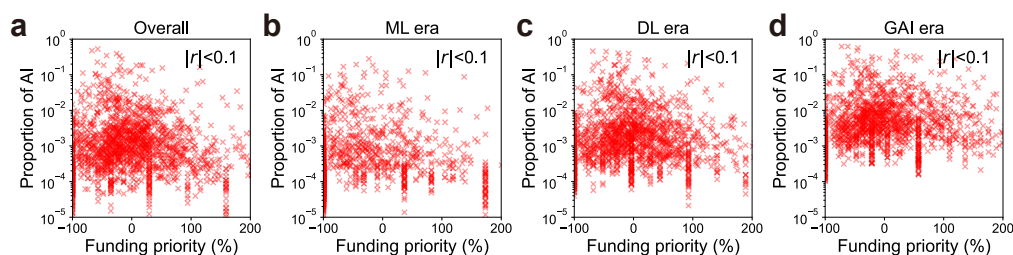


Figure 25. Selectivity of AI adoption in different topics regarding funding priority across the topics. In all panels, each scatter shows the degree of AI penetration in a topic and funding priority to that topic. The corresponding Pearson's r is indicated correspondingly in each panel. Results show that across topics and AI eras, there are heterogeneous funding priorities for different topics, but funding priority is not obviously correlated with the proportion of AI adoption, suggesting that AI adoption does not selectively favor topics prioritized by funding agencies.

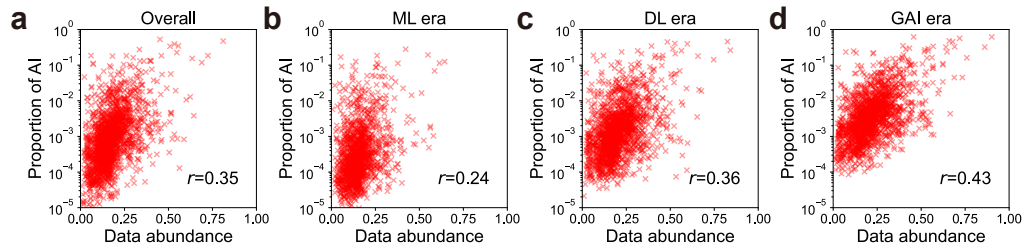


Figure 26. Selectivity of AI adoption in different topics regarding data abundance in the topics. In all panels, each scatter shows the degree of AI penetration in a topic and the data abundance in that topic. The corresponding Pearson's r is indicated correspondingly in each panel. The results show that across topics and AI eras, the data abundance is positively correlated with the proportion of AI adoption, suggesting that AI is more likely to be adopted in topics with abundant data resources. Notably, the correlation of AI-use with frequency of mentioned data resources rises with the size and intensity of AI-models.

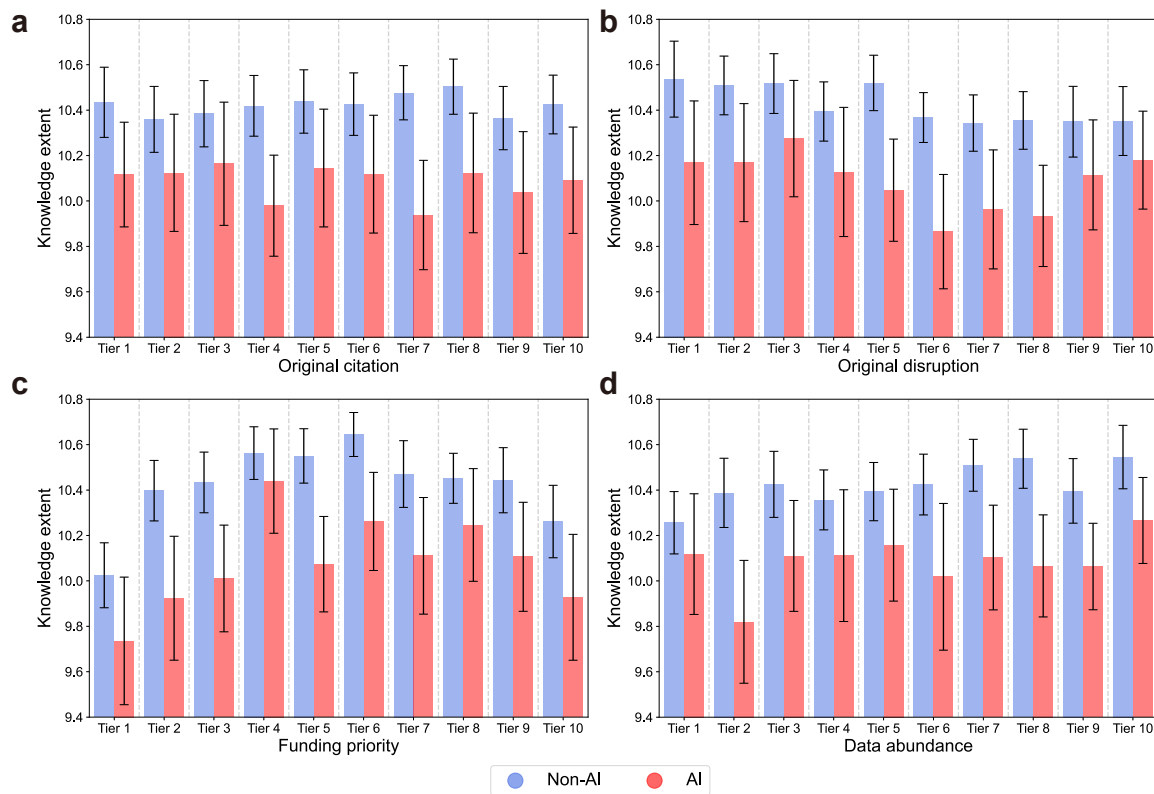


Figure 27. Stratified analyses controlling for the effects of potential confounding variables on the contracted knowledge extent of AI-augmented research. The results show that when dividing each external factor into 10 tiers based on magnitude, the contraction in knowledge extent for AI-augmented research remains evident within each tier ($p < 0.001, n = 17,529,094$). Therefore, regardless of whether a topic is more or less likely to be selected for AI adoption, existing AI-augmented research within the topic consistently covers a contracted knowledge space compared to non-AI research. For all panels, 99% CIs are shown as error bars centred at the mean. All statistical tests use a two-sided t-test.

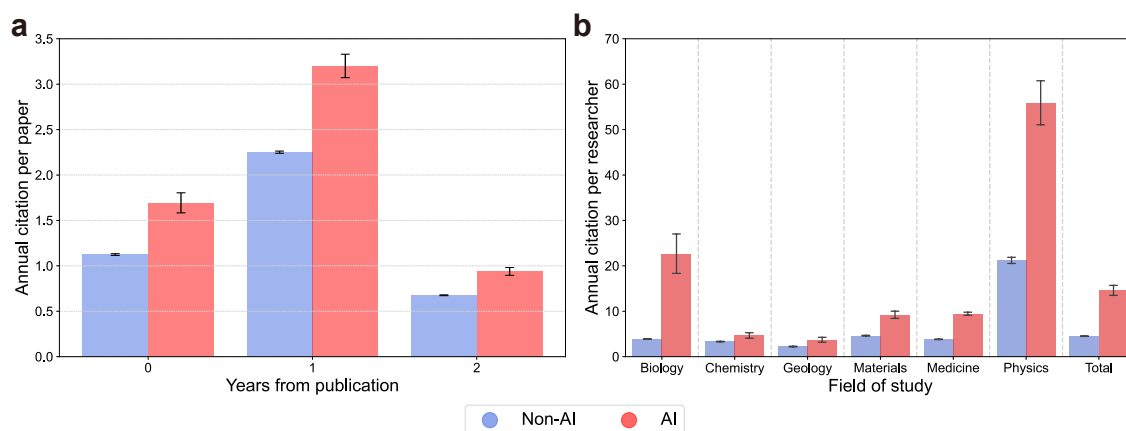


Figure 28. Replication of “AI enlarges the impact of papers” with only the subset of papers published during the GAI era. (a) Annual citations after publication of AI and non-AI papers ($n = 1,888,694$). **(b)** Average annual citations of researchers adopting AI and their counterparts without AI ($p < 0.001, n = 2,888,737$). For all panels, 99% CIs are shown as error bars centred at the mean. All statistical tests use a two-sided t-test. Results in the GAI era are consistent with those obtained over the full time span featured in the main text.

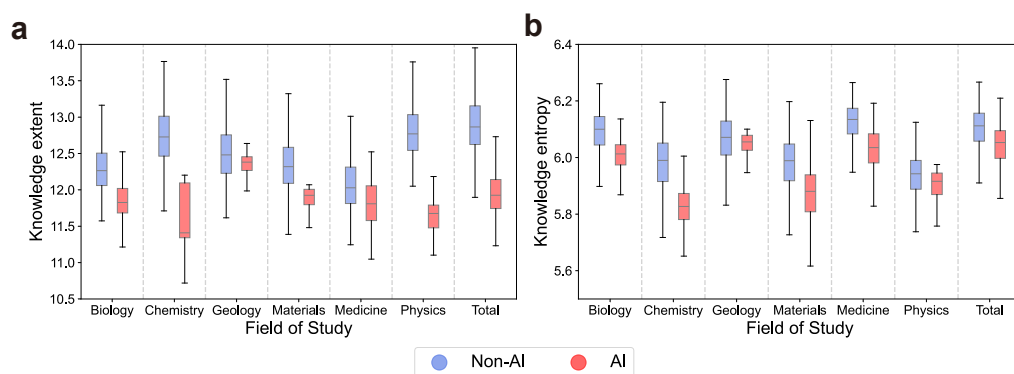


Figure 29. Replication of “AI usage is associated with a contraction in knowledge extent within and across scientific fields” with only the subset of papers published during the GAI era. (a) Knowledge extent of AI and non-AI papers in each field ($n = 1,000$ samples in each field). **(b)** Knowledge entropy of AI and non-AI papers in each field ($n = 1,000$ samples in each field). Boxplots are centred at the median and bounded at the first and third quartile (Q1 and Q3), with 1.5 times of the inter-quartile range (IQR) shown as whiskers from the box. Results in the GAI era are consistent with those obtained over the full time span featured in the main text.

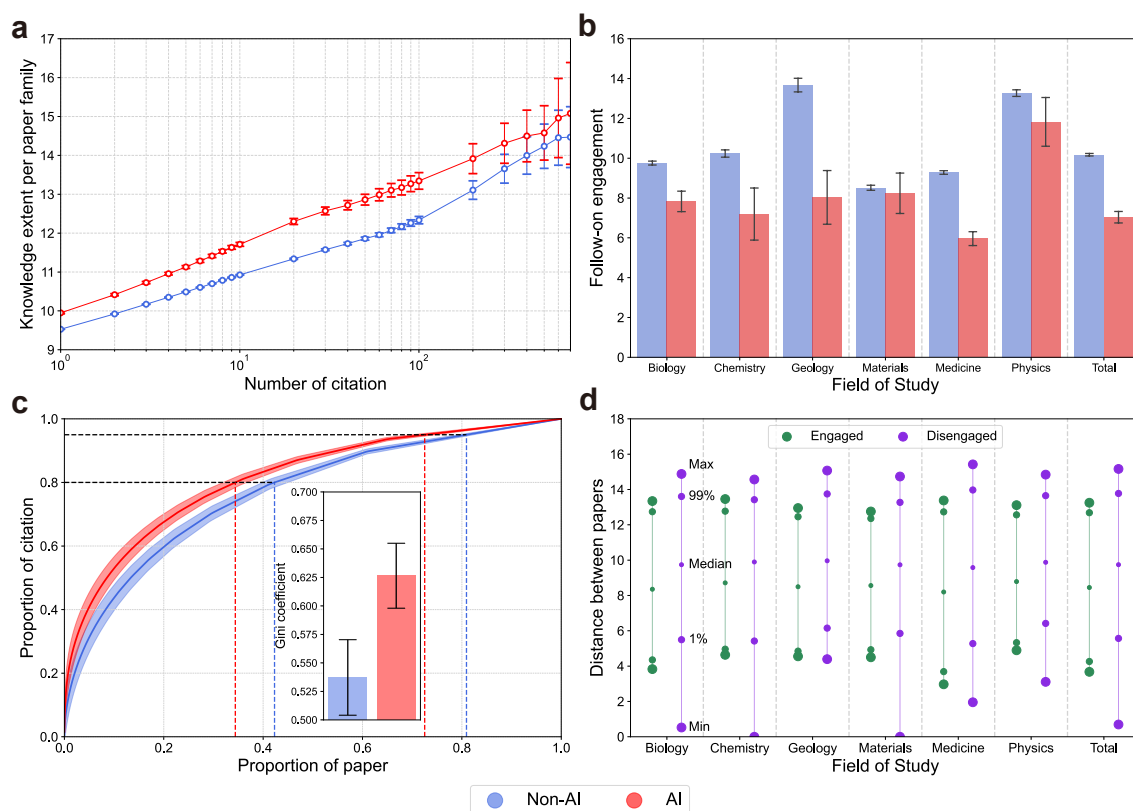


Figure 30. Replication of “reduced follow-on engagement and more overlapped works in AI research” with only the subset of papers published during the GAI era. (a) Knowledge extent of individual AI and non-AI paper families ($n = 1,888,694$). **(b)** Engagement among papers that cite AI vs. non-AI papers ($p < 0.001, n = 518,156$). **(c)** Distribution of citations to AI vs. non-AI papers ($p < 0.001, n = 100$ sampled paper groups). **(d)** Distribution of distances between paper pairs that cite the same papers, with or without citing each other—engaged versus disengaged ($n = 258,529$ sampled paper pairs). 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. Results in the GAI era are consistent with those on the full time span featured in the main text.

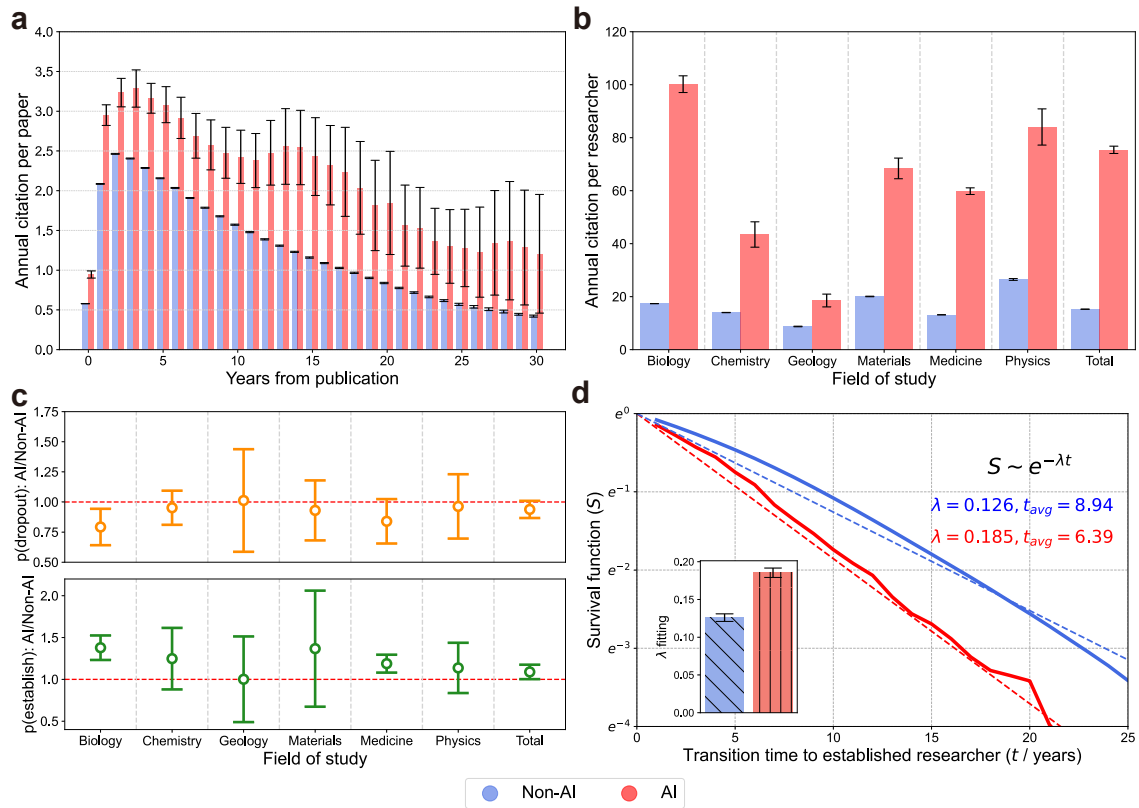


Figure 31. Replication of “AI enlarges the impact of papers and enhances the career of researchers” on the WOS dataset. (a) Annual citations after publication of AI and non-AI papers ($n = 16,706,988$). (b) Average annual citations of researchers adopting AI and their counterparts without AI ($p < 0.001, n = 3,620,795$). (c) The probability of two role transitions between junior scientists adopting AI and their counterparts without AI ($n = 46$ year observations for each field). (d) Survival functions for the transition from junior to established researcher ($p < 0.001, n = 1,556,338$). For all panels, 99% CIs are shown as error bars centred at the mean. All statistical tests use a two-sided t-test. Results on the WOS dataset are consistent with those from the OpenAlex dataset featured in the main text.

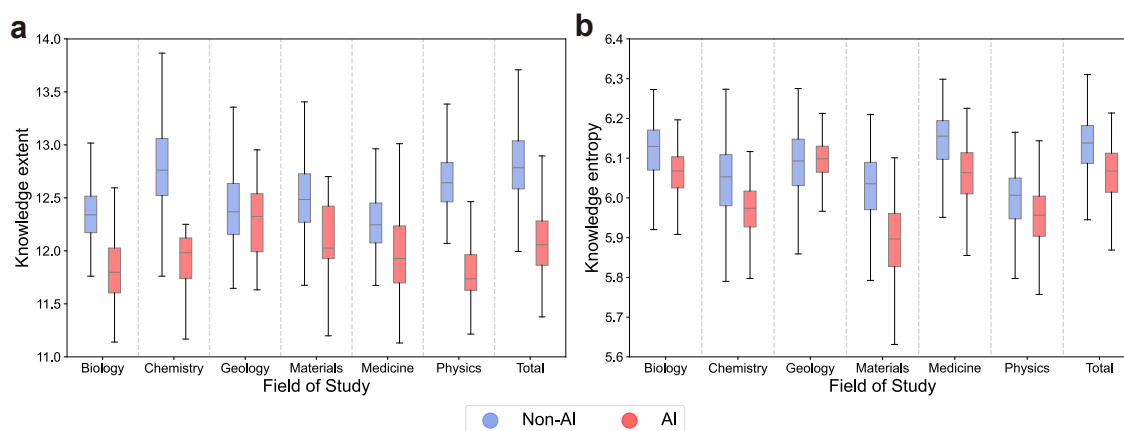


Figure 32. Replication of “AI usage is associated with a contraction in knowledge extent within and across scientific fields” on the WOS dataset. (a) Knowledge extent of AI and non-AI papers in each field ($n = 1,000$ samples in each field). **(b)** Knowledge entropy of AI and non-AI papers in each field ($n = 1,000$ samples in each field). Boxplots are centred at the median and bounded at the first and third quartile (Q1 and Q3), with 1.5 times of the inter-quartile range (IQR) shown as whiskers from the box. Results on the WOS dataset are consistent with the ones on OpenAlex featured in the main text.

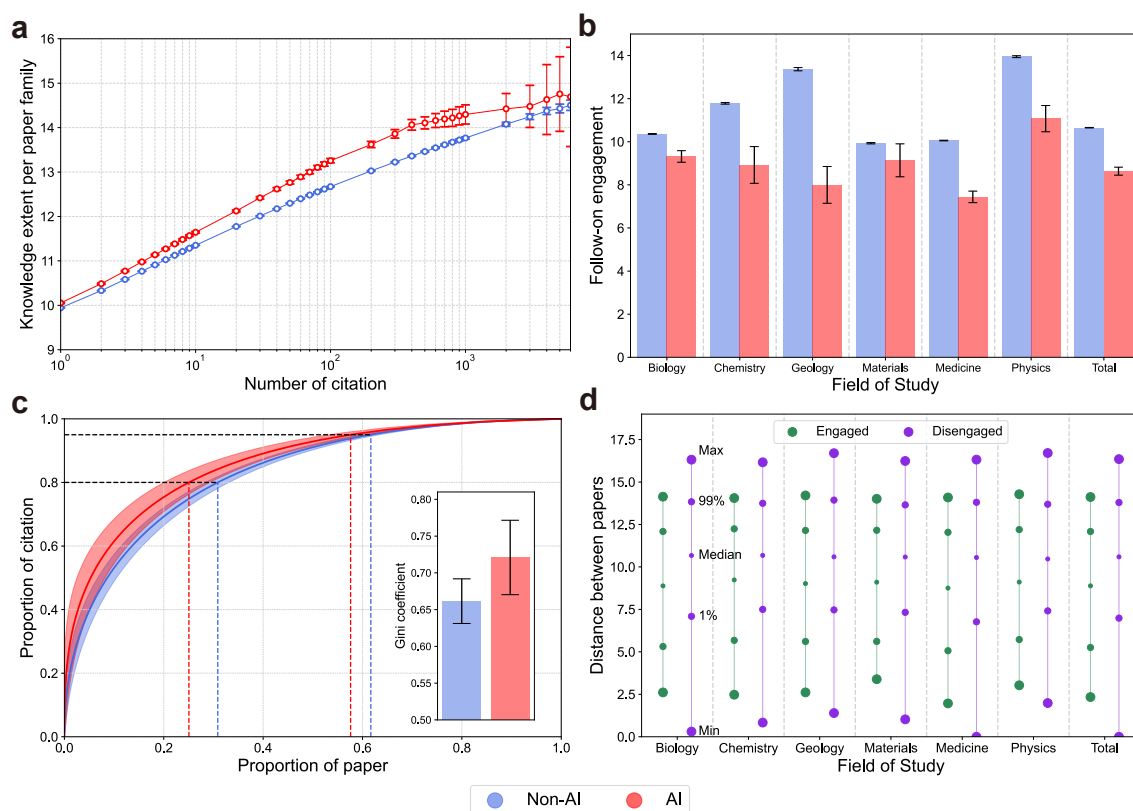


Figure 33. Replication of “reduced follow-on engagement and more overlapped works in AI research” on the WOS dataset. (a) Knowledge extent of individual AI and non-AI paper families ($n = 16,706,988$). (b) Engagement among papers that cite AI vs. non-AI papers ($p < 0.001, n = 15,048,254$). (c) The distribution of citations to AI vs. non-AI papers ($p < 0.001, n = 100$ sampled paper groups). (d) The distribution of distances between paper pairs that cite the same papers, with or without citing each other—engaged versus disengaged ($n = 674,132,352$ sampled paper pairs). 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. Results from the WOS dataset are consistent with those from the OpenAlex dataset featured in the main text.

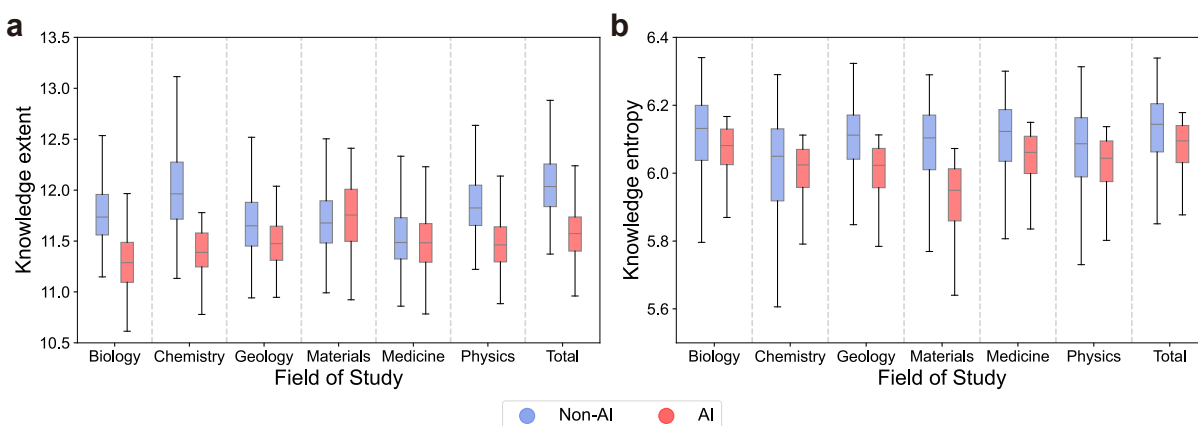


Figure 34. Replication of “AI usage is associated with a contraction in knowledge extent within and across scientific fields” in an embedding space after dimensional reduction using PCA. (a) Knowledge extent of AI and non-AI papers in each field ($n = 1,000$ samples in each field). **(b)** Knowledge entropy of AI and non-AI papers in each field ($n = 1,000$ samples in each field). Boxplots are centred at the median and bounded at the first and third quartile (Q1 and Q3), with 1.5 times of the inter-quartile range (IQR) shown as whiskers from the box. Results in the reduced-dimensional space are consistent with those from the original embedding space reported in the main text.

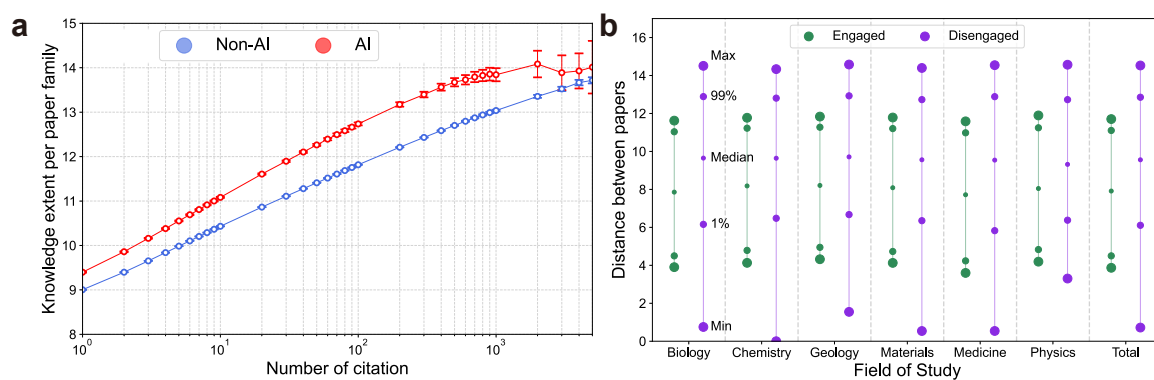


Figure 35. Replication of “reduced follow-on engagement and more overlapped works in AI research” in an embedding space after dimensional reduction using PCA. (a) Knowledge extent of individual AI and non-AI paper families ($n = 27,405,011$). 99% CIs are shown as error bars centred at the mean. (b) The distribution of distances between paper pairs that cite the same papers, with or without citing each other—engaged versus disengaged ($n = 590,325,130$ sampled paper pairs). Results in the reduced-dimensional space are consistent with those from the original embedding space reported in the main text.

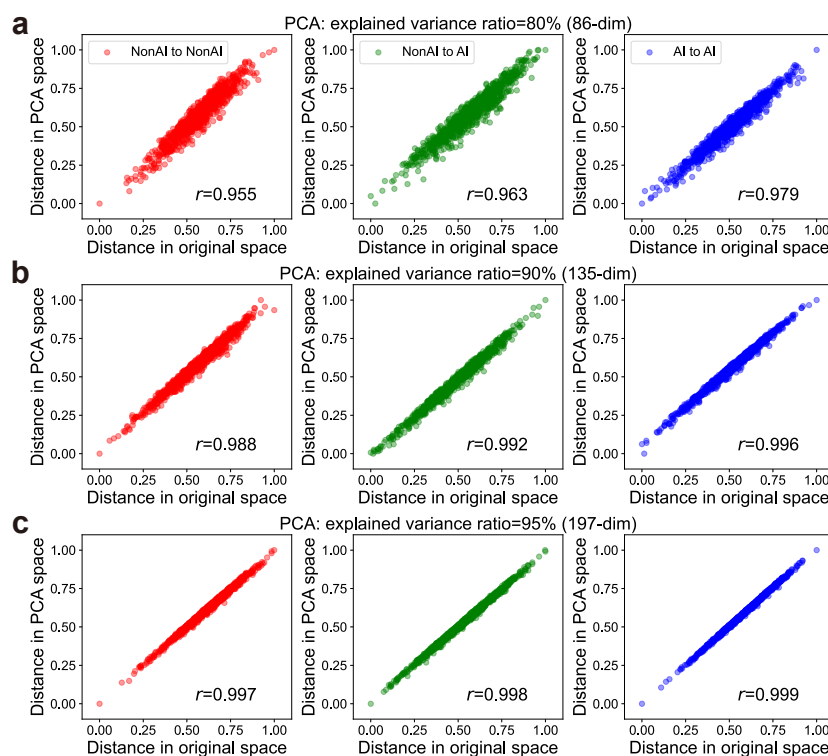


Figure 36. General sensitivity check on the distance calculations in high-dimensional paper embedding space. We reduce the original embeddings to 86, 135, and 197 dimensions with PCA, which correspond to the principal components that explain (a) 80%, (b) 90%, and (c) 95% of the total variance, respectively. We then randomly sampled 1,000 NonAI-NonAI, 1,000 AI-AI, and 1,000 NonAI-AI paper pairs and calculated the distance for each pair within each of the reduced-dimensional spaces. Results show that the distances in the various reduced-dimensional spaces are highly correlated with those in the original space, indicating the reliability of our paper embeddings for distance calculations.

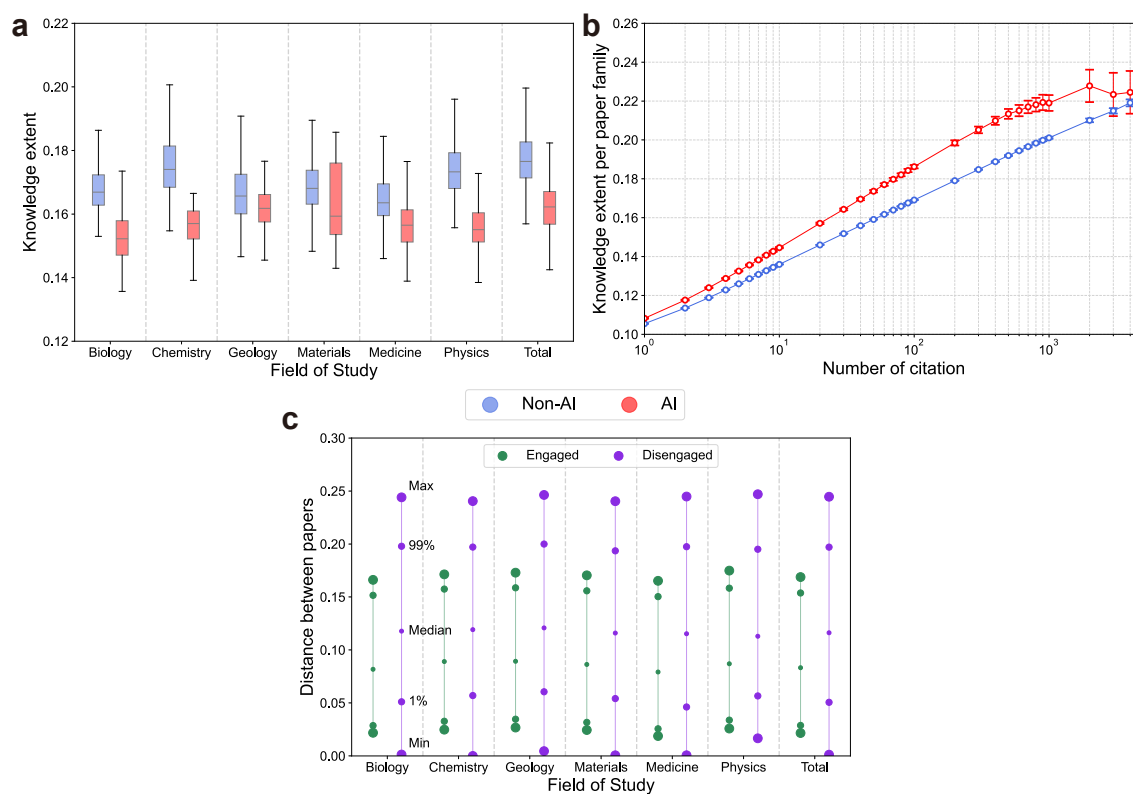


Figure 37. Replication of results related to high-dimensional distance calculations with Cosine distance. (a) Knowledge extent of AI and non-AI papers in each field ($n = 1,000$ samples in each field). Boxplots are centred at the median and bounded at the first and third quartile (Q1 and Q3), with 1.5 times of the inter-quartile range (IQR) shown as whiskers from the box. (b) Knowledge extent of individual AI and non-AI paper families ($n = 27,405,011$). 99% CIs are shown as error bars centred at the mean. (c) The distribution of distances between paper pairs that cite the same papers, with or without citing each other-engaged versus disengaged ($n = 590,325,130$ sampled paper pairs). Results with Cosine distance are consistent with those with Euclidean distance reported in the main text.

Supplementary Tables

Table 1. Information of experts involved in scrutinizing our identification results

ID	Majority	Educational background	Number of publications	AI experience/y
1	EECS	Ph.D.	31	8
2	EECS	Ph.D. student	4	2
3	EECS	Ph.D.	7	6
4	EECS	Ph.D.	11	5
5	EECS	Ph.D.	21	6
6	Physics, EECS	Ph.D.	6	3
7	EECS	Ph.D.	38	7
8	EECS	Ph.D.	16	5
9	EECS	Ph.D.	12	6
10	Biology, EECS	Ph.D.	15	5
11	EECS	Ph.D.	40	8
12	EECS	Ph.D.	21	7

Table 2. Number of papers sampled for expert annotation in identification accuracy evaluation

Group	ML era	DL era	GAI era	Overall
Biology 1	59	33	18	110
Biology 2	57	39	14	110
Chemistry 1	60	32	18	110
Chemistry 2	69	28	13	110
Geology 1	69	25	16	110
Geology 2	55	40	15	110
Materials science 1	65	29	16	110
Materials science 2	63	32	15	110
Medicine 1	60	33	17	110
Medicine 2	54	37	19	110
Physics 1	66	29	15	110
Physics 2	60	36	14	110
Overall	737	393	190	1320

Table 3. Fleiss’ Kappa of expert annotation results in identification accuracy evaluation

Group	ML era	DL era	GAI era	Overall
Biology 1	0.955	0.943	0.926	0.950
Biology 2	0.951	0.951	1.000	0.963
Chemistry 1	1.000	0.952	0.924	0.976
Chemistry 2	0.981	0.940	0.889	0.963
Geology 1	0.921	0.941	1.000	0.939
Geology 2	0.976	1.000	0.909	0.975
Materials science 1	1.000	0.893	1.000	0.976
Materials science 2	1.000	1.000	0.909	0.988
Medicine 1	0.976	0.898	0.921	0.952
Medicine 2	0.914	0.961	0.921	0.939
Physics 1	0.960	1.000	0.909	0.964
Physics 2	1.000	0.962	0.899	0.976
Overall	0.971	0.957	0.935	0.964

Table 4. F1-score of BERT identification results against expert labels in identification accuracy evaluation

Group	ML era	DL era	GAI era	Overall
Biology 1	0.923	0.920	0.889	0.917
Biology 2	0.878	0.897	0.857	0.885
Chemistry 1	0.809	0.857	0.900	0.844
Chemistry 2	0.892	0.923	0.857	0.898
Geology 1	0.800	0.812	0.941	0.826
Geology 2	0.783	0.912	0.875	0.857
Materials science 1	0.857	0.895	0.941	0.881
Materials science 2	0.800	0.923	0.875	0.855
Medicine 1	0.978	0.913	0.875	0.935
Medicine 2	0.882	0.917	0.917	0.906
Physics 1	0.825	0.833	0.875	0.835
Physics 2	0.816	0.909	0.875	0.862
Overall	0.852	0.896	0.892	0.875

Table 5. The most frequently adopted AI methods in all selected disciplines in different AI development eras

Era	Year	Top AI methods									
		1	2	3	4	5	6	7	8	9	10
ML	[1980,1990]	Expert System	Artificial Neural Network (ANN)	Knowledge Base	Connectionist Network (CN)	Principal Component Analysis (PCA)	Hebbian Learning	Maximum Likelihood Classifier	Logistic Regression (LR)	Generalized Additive Model	Nonlinear Regression
	[1991,2000]	Artificial Neural Network (ANN)	Principal Component Analysis (PCA)	Expert System	Knowledge Base	Radial Basis Functions (RBF)	Hidden Markov Model (HMM)	Unsupervised Learning	Linear Discriminant Analysis (LDA)	Maximum Likelihood Classifier	Markov Chain Monte Carlo (MCMC)
	[2001,2005]	Artificial Neural Network (ANN)	Support Vector Machine (SVM)	Principal Component Analysis (PCA)	Hidden Markov Model (HMM)	Linear Discriminant Analysis (LDA)	Markov Chain Monte Carlo (MCMC)	Radial Basis Functions (RBF)	Unsupervised Learning	Knowledge Base	Byesian Network
	[2006,2010]	Support Vector Machine (SVM)	Artificial Neural Network (ANN)	Principal Component Analysis (PCA)	Linear Discriminant Analysis (LDA)	Hidden Markov Model (HMM)	Markov Chain Monte Carlo (MCMC)	Radial Basis Functions (RBF)	Non-negative Matrix Factorization (NMF)	Unsupervised Learning	Conditional Random Field (CRF)
	[2011,2015]	Support Vector Machine (SVM)	Artificial Neural Network (ANN)	Principal Component Analysis (PCA)	Markov Chain Monte Carlo (MCMC)	Linear Discriminant Analysis (LDA)	Hidden Markov Model (HMM)	Conditional Random Field (CRF)	Unsupervised Learning	Non-negative Matrix Factorization (NMF)	Radial Basis Functions (RBF)
DL	[2016,2020]	Convolutional Neural Network (CNN)	Support Vector Machine (SVM)	Random Forest (RF)	Principal Component Analysis (PCA)	Generative Adversarial Network (GAN)	Long Short-term Memory (LSTM)	Decision Tree (DT)	Unsupervised Learning	Reinforcement Learning (RL)	Linear Discriminant Analysis (LDA)
	[2021,2022]	Convolutional Neural Network (CNN)	Support Vector Machine (SVM)	Random Forest (RF)	Transfer Learning (TL)	Long Short-term Memory (LSTM)	Generative Adversarial Network (GAN)	Gradient Boosting	K-nearest Neighbor (KNN)	Graph Convolutional Network (GCN)	Principal Component Analysis (PCA)
GAI	2023	Convolutional Neural Network (CNN)	Support Vector Machine (SVM)	Random Forest (RF)	Long Short-term Memory (LSTM)	Gradient Boosting	Large Language Model (LLM)	Generative Adversarial Network (GAN)	Transfer Learning (TL)	K-nearest Neighbor (KNN)	Graph Convolutional Network (GCN)
	2024	Convolutional Neural Network (CNN)	Support Vector Machine (SVM)	Random Forest (RF)	Large Language Model (LLM)	Gradient Boosting	Long Short-term Memory (LSTM)	Generative Adversarial Network (GAN)	K-nearest Neighbor (KNN)	Shapley Additive Explanations (SHAP)	Logistic Regression (LR)
	2025	Convolutional Neural Network (CNN)	Support Vector Machine (SVM)	Large Language Model (LLM)	Random Forest (RF)	Shapley Additive Explanations (SHAP)	Long Short-term Memory (LSTM)	Gradient Boosting	Generative Adversarial Network (GAN)	Graph Neural Network (GNN)	Logistic Regression (LR)

Table 6. The most frequently adopted AI methods in biology in different AI development eras

Era	Year	Top AI methods									
		1	2	3	4	5	6	7	8	9	10
ML	[1980,1990]	Artificial Neural Network (ANN)	Expert System	Knowledge Base	Connectionist Network (CN)	Hebbian Learning					
	[1991,2000]	Artificial Neural Network (ANN)	Principal Component Analysis (PCA)	Expert System	Knowledge Representation	Hidden Markov Model (HMM)	Genetic Programming (GP)	Unsupervised Learning	Linear Discriminant Analysis (LDA)		
	[2001,2005]	Artificial Neural Network (ANN)	Support Vector Machine (SVM)	Principal Component Analysis (PCA)	Hidden Markov Model (HMM)	Markov Chain Monte Carlo (MCMC)	Unsupervised Learning	Reinforcement Learning (RL)	Linear Discriminant Analysis (LDA)	Bsyesian Network	
	[2006,2010]	Support Vector Machine (SVM)	Artificial Neural Network (ANN)	Principal Component Analysis (PCA)	Hidden Markov Model (HMM)	Linear Discriminant Analysis (LDA)	Markov Chain Monte Carlo (MCMC)	Bsyesian Network	Conditional Random Field (CRF)	Unsupervised Learning	Reinforcement Learning (RL)
	[2011,2015]	Support Vector Machine (SVM)	Artificial Neural Network (ANN)	Principal Component Analysis (PCA)	Markov Chain Monte Carlo (MCMC)	Hidden Markov Model (HMM)	Linear Discriminant Analysis (LDA)	Unsupervised Learning	Conditional Random Field (CRF)	Non-negative Matrix Factorization (NMF)	Reinforcement Learning (RL)
DL	[2016,2020]	Convolutional Neural Network (CNN)	Support Vector Machine (SVM)	Long Short-term Memory (LSTM)	Random Forest (RF)	Principal Component Analysis (PCA)	Generative Adversarial Network (GAN)	Reinforcement Learning (RL)	Linear Discriminant Analysis (LDA)	Unsupervised Learning	K-nearest Neighbor (KNN)
	[2021,2022]	Convolutional Neural Network (CNN)	Support Vector Machine (SVM)	Long Short-term Memory (LSTM)	Random Forest (RF)	Generative Adversarial Network (GAN)	Transfer Learning (TL)	Principal Component Analysis (PCA)	Reinforcement Learning (RL)	K-nearest Neighbor (KNN)	Decision Tree (DT)
GAI	2023	Convolutional Neural Network (CNN)	Support Vector Machine (SVM)	Transfer Learning (TL)	Random Forest (RF)	Long Short-term Memory (LSTM)	Generative Adversarial Network (GAN)	Gradient Boosting	K-nearest Neighbor (KNN)	Decision Tree (DT)	Reinforcement Learning (RL)
	2024	Convolutional Neural Network (CNN)	Support Vector Machine (SVM)	Large Language Model (LLM)	Random Forest (RF)	Transfer Learning (TL)	Long Short-term Memory (LSTM)	Generative Adversarial Network (GAN)	Gradient Boosting	Attention mechanism	K-nearest Neighbor (KNN)
	2025	Convolutional Neural Network (CNN)	Support Vector Machine (SVM)	Large Language Model (LLM)	Graph Neural Network (GNN)	Long Short-term Memory (LSTM)	Random Forest (RF)	Generative Adversarial Network (GAN)	Gradient Boosting	Spiking Neural Network	Shapley Additive Explanations (SHAP)

Table 7. The most frequently adopted AI methods in chemistry in different AI development eras

Era	Year	Top AI methods									
		1	2	3	4	5	6	7	8	9	10
ML	[1980,1990]	Logistic Regression (LR)	Expert System	Principal Component Analysis (PCA)	Generalized Additive Model	Knowledge Base	Maximum Likelihood Classifier	Nonlinear Regression			
	[1991,2000]	Principal Component Analysis (PCA)	Artificial Neural Network (ANN)	Linear Discriminant Analysis (LDA)	Multiple Linear Regression	Support Vector Machine (SVM)	Expert System	Gaussian Mixture Model (GMM)			
	[2001,2005]	Principal Component Analysis (PCA)	Support Vector Machine (SVM)	Artificial Neural Network (ANN)	Linear Discriminant Analysis (LDA)	Gaussian Mixture Model (GMM)	Singular Value Decomposition (SVD)	Hidden Markov Model (HMM)	Linear Regression		
	[2006,2010]	Support Vector Machine (SVM)	Principal Component Analysis (PCA)	Linear Discriminant Analysis (LDA)	Artificial Neural Network (ANN)	Canonical Correlation Analysis (CCA)	Gaussian Mixture Model (GMM)	Non-negative Matrix Factorization (NMF)	Singular Value Decomposition (SVD)	Kernel Function	Locally Linear Embedding (LLE)
	[2011,2015]	Support Vector Machine (SVM)	Principal Component Analysis (PCA)	Artificial Neural Network (ANN)	Linear Discriminant Analysis (LDA)	Canonical Correlation Analysis (CCA)	Markov Chain Monte Carlo (MCMC)	Gaussian Mixture Model (GMM)	Singular Value Decomposition (SVD)	Unsupervised Learning	Multiple Linear Regression
DL	[2016,2020]	Convolutional Neural Network (CNN)	Support Vector Machine (SVM)	Principal Component Analysis (PCA)	Random Forest (RF)	Gaussian Process Regression	K-nearest Neighbor (KNN)	Markov Chain Monte Carlo (MCMC)	Decision Tree (DT)	Gaussian Mixture Model (GMM)	Long Short-term Memory (LSTM)
	[2021,2022]	Convolutional Neural Network (CNN)	Support Vector Machine (SVM)	Principal Component Analysis (PCA)	Transfer Learning (TL)	Random Forest (RF)	Linear Discriminant Analysis (LDA)	Gaussian Process Regression	Long Short-term Memory (LSTM)	K-nearest Neighbor (KNN)	Decision Tree (DT)
GAI	2023	Convolutional Neural Network (CNN)	Support Vector Machine (SVM)	Transfer Learning (TL)	Principal Component Analysis (PCA)	Random Forest (RF)	Gradient Boosting	K-nearest Neighbor (KNN)	Gaussian Process Regression	Shapley Additive Explanations (SHAP)	Long Short-term Memory (LSTM)
	2024	Convolutional Neural Network (CNN)	Support Vector Machine (SVM)	Transfer Learning (TL)	Principal Component Analysis (PCA)	Gradient Boosting	Long Short-term Memory (LSTM)	Decision Tree (DT)	K-nearest Neighbor (KNN)	Attention mechanism	Gaussian Process Regression
	2025	Convolutional Neural Network (CNN)	Principal Component Analysis (PCA)	Support Vector Machine (SVM)	Random Forest (RF)	Graph Neural Network (GNN)	Large Language Model (LLM)	Generative Adversarial Network (GAN)	Long Short-term Memory (LSTM)	Linear Discriminant Analysis (LDA)	Decision Tree (DT)

Table 8. The most frequently adopted AI methods in geology in different AI development eras

Era	Year	Top AI methods									
		1	2	3	4	5	6	7	8	9	10
ML	[1980,1990]	Expert System	Knowledge Base								
	[1991,2000]	Artificial Neural Network (ANN)	Expert System	Knowledge Base	Hidden Markov Model (HMM)	Markov Chain Monte Carlo (MCMC)					
	[2001,2005]	Support Vector Machine (SVM)	Artificial Neural Network (ANN)	Principal Component Analysis (PCA)	Linear Discriminant Analysis (LDA)	Hidden Markov Model (HMM)					
	[2006,2010]	Support Vector Machine (SVM)	Principal Component Analysis (PCA)	Linear Discriminant Analysis (LDA)	Artificial Neural Network (ANN)	Conditional Random Field (CRF)	Locally Linear Embedding (LLE)	Gaussian Mixture Model (GMM)			
	[2011,2015]	Support Vector Machine (SVM)	Principal Component Analysis (PCA)	Linear Discriminant Analysis (LDA)	Artificial Neural Network (ANN)	Locally Linear Embedding (LLE)	Hidden Markov Model (HMM)	Markov Chain Monte Carlo (MCMC)	Gaussian Mixture Model (GMM)	K-nearest Neighbor (KNN)	
DL	[2016,2020]	Convolutional Neural Network (CNN)	Support Vector Machine (SVM)	Generative Adversarial Network (GAN)	Long Short-term Memory (LSTM)	Random Forest (RF)	Principal Component Analysis (PCA)	Unsupervised Learning	Transfer Learning (TL)	Conditional Random Field (CRF)	
	[2021,2022]	Convolutional Neural Network (CNN)	Long Short-term Memory (LSTM)	Generative Adversarial Network (GAN)	Support Vector Machine (SVM)	Random Forest (RF)	Attention mechanism	Graph Convolutional Network (GCN)	Unsupervised Learning	Transfer Learning (TL)	
GAI	2023	Convolutional Neural Network (CNN)	Attention mechanism	Long Short-term Memory (LSTM)	Support Vector Machine (SVM)	Generative Adversarial Network (GAN)	Random Forest (RF)	Unsupervised Learning	Graph Convolutional Network (GCN)	Gradient Boosting	Gated Recurrent Unit (GRU)
	2024	Convolutional Neural Network (CNN)	Long Short-term Memory (LSTM)	Generative Adversarial Network (GAN)	Support Vector Machine (SVM)	Random Forest (RF)	Attention mechanism	Gradient Boosting	Graph Convolutional Network (GCN)	Gated Recurrent Unit (GRU)	Unsupervised Learning
	2025	Convolutional Neural Network (CNN)	Long Short-term Memory (LSTM)	Generative Adversarial Network (GAN)	Support Vector Machine (SVM)	Artificial Neural Network (ANN)	Large Language Model (LLM)	Random Forest (RF)	Linear Regression	Attention mechanism	Graph Convolutional Network (GCN)

Table 9. The most frequently adopted AI methods in materials science in different AI development eras

Era	Year	Top AI methods									
		1	2	3	4	5	6	7	8	9	10
ML	[1980,1990]	Expert System	Causal Reasoning								
	[1991,2000]	Artificial Neural Network (ANN)	Expert System								
	[2001,2005]	Artificial Neural Network (ANN)	Support Vector Machine (SVM)	Markov Chain Monte Carlo (MCMC)							
	[2006,2010]	Support Vector Machine (SVM)	Artificial Neural Network (ANN)	Reinforcement Learning (RL)	Markov Chain Monte Carlo (MCMC)	Conditional Random Field (CRF)	Hidden Markov Model (HMM)				
	[2011,2015]	Artificial Neural Network (ANN)	Support Vector Machine (SVM)	Principal Component Analysis (PCA)	Random Forest (RF)	Markov Chain Monte Carlo (MCMC)	Hidden Markov Model (HMM)	Gaussian Process Regression			
DL	[2016,2020]	Convolutional Neural Network (CNN)	Support Vector Machine (SVM)	Gaussian Process Regression	Random Forest (RF)	Ridge Regression	Generative Adversarial Network (GAN)	Principal Component Analysis (PCA)	Long Short-term Memory (LSTM)	Unsupervised Learning	Reinforcement Learning (RL)
	[2021,2022]	Convolutional Neural Network (CNN)	Transfer Learning (TL)	Support Vector Machine (SVM)	Generative Adversarial Network (GAN)	Gaussian Process Regression	Long Short-term Memory (LSTM)	Random Forest (RF)	Reinforcement Learning (RL)	Graph Neural Network (GNN)	Unsupervised Learning
GAI	2023	Graph Neural Network (GNN)	Transfer Learning (TL)	Large Language Model (LLM)	Support Vector Machine (SVM)	Gaussian Process Regression	Generative Adversarial Network (GAN)	Shapley Additive Explanations (SHAP)	Gradient Boosting	Random Forest (RF)	Active Learning (AL)
	2024	Graph Neural Network (GNN)	Large Language Model (LLM)	Support Vector Machine (SVM)	Active Learning (AL)	Gradient Boosting	Generative Adversarial Network (GAN)	Random Forest (RF)	Gaussian Process Regression	Reinforcement Learning (RL)	Attention mechanism
	2025	Graph Neural Network (GNN)	Large Language Model (LLM)	Convolutional Neural Network (CNN)	Principal Component Analysis (PCA)	Generative Adversarial Network (GAN)	Gaussian Process Regression	Gradient Boosting	Artificial Neural Network (ANN)	Shapley Additive Explanations (SHAP)	Long Short-term Memory (LSTM)

Table 10. The most frequently adopted AI methods in medicine in different AI development eras

Era	Year	Top AI methods									
		1	2	3	4	5	6	7	8	9	10
ML	[1980,1990]	Expert System	Knowledge Base								
	[1991,2000]	Expert System	Artificial Neural Network (ANN)	Knowledge Base	Logistic Regression (LR)	Markov Chain Monte Carlo (MCMC)					
	[2001,2005]	Artificial Neural Network (ANN)	Support Vector Machine (SVM)	Expert System	Principal Component Analysis (PCA)	Knowledge Base	Hidden Markov Model (HMM)	Linear Discriminant Analysis (LDA)	Decision Tree (DT)	Markov Chain Monte Carlo (MCMC)	
	[2006,2010]	Support Vector Machine (SVM)	Artificial Neural Network (ANN)	Principal Component Analysis (PCA)	Linear Discriminant Analysis (LDA)	Hidden Markov Model (HMM)	Decision Tree (DT)	Expert System	Knowledge Base	Markov Chain Monte Carlo (MCMC)	
	[2011,2015]	Support Vector Machine (SVM)	Artificial Neural Network (ANN)	Principal Component Analysis (PCA)	Decision Tree (DT)	Random Forest (RF)	Linear Discriminant Analysis (LDA)	Hidden Markov Model (HMM)	Expert System	Long Short-term Memory (LSTM)	
DL	[2016,2020]	Convolutional Neural Network (CNN)	Support Vector Machine (SVM)	Random Forest (RF)	Decision Tree (DT)	Long Short-term Memory (LSTM)	Principal Component Analysis (PCA)	K-nearest Neighbor (KNN)	Generative Adversarial Network (GAN)	Logistic Regression (LR)	
	[2021,2022]	Convolutional Neural Network (CNN)	Support Vector Machine (SVM)	Transfer Learning (TL)	Random Forest (RF)	Gradient Boosting	Long Short-term Memory (LSTM)	K-nearest Neighbor (KNN)	Generative Adversarial Network (GAN)	Naïve Bayes (NB)	
GAI	2023	Convolutional Neural Network (CNN)	Support Vector Machine (SVM)	Gradient Boosting	Large Language Model (LLM)	Long Short-term Memory (LSTM)	Shapley Additive Explanations (SHAP)	K-nearest Neighbor (KNN)	Transfer Learning (TL)	Generative Adversarial Network (GAN)	
	2024	Convolutional Neural Network (CNN)	Support Vector Machine (SVM)	Large Language Model (LLM)	Gradient Boosting	Shapley Additive Explanations (SHAP)	Long Short-term Memory (LSTM)	K-nearest Neighbor (KNN)	Generative Adversarial Network (GAN)	Attention mechanism	Transfer Learning (TL)
	2025	Convolutional Neural Network (CNN)	Support Vector Machine (SVM)	Large Language Model (LLM)	Random Forest (RF)	Shapley Additive Explanations (SHAP)	Gradient Boosting	Logistic Regression (LR)	Artificial Neural Network (ANN)	Long Short-term Memory (LSTM)	Generative Adversarial Network (GAN)

Table 11. The most frequently adopted AI methods in physics in different AI development eras

Era	Year	Top AI methods									
		1	2	3	4	5	6	7	8	9	10
ML	[1980,1990]	Expert System	Radial Basis Functions (RBF)	Logistic Regression (LR)							
	[1991,2000]	Artificial Neural Network (ANN)	Radial Basis Functions (RBF)	Unsupervised Learning	Boltzmann Machine	Kernel Density Estimation (KDE)	Recurrent Neural Network (RNN)	Principal Component Analysis (PCA)			
	[2001,2005]	Artificial Neural Network (ANN)	Radial Basis Functions (RBF)	Principal Component Analysis (PCA)	Support Vector Machine (SVM)	Markov Chain Monte Carlo (MCMC)	Hebbian Learning	Hidden Markov Model (HMM)	Singular Value Decomposition (SVD)		
	[2006,2010]	Recurrent Neural Network (RNN)	Support Vector Machine (SVM)	Markov Chain Monte Carlo (MCMC)	Radom Graph Model	Principal Component Analysis (PCA)	Radial Basis Functions (RBF)	Non-negative Matrix Factorization (NMF)	Random Walk		
	[2011,2015]	Artificial Neural Network (ANN)	Recurrent Neural Network (RNN)	Support Vector Machine (SVM)	Radom Graph Model	Markov Chain Monte Carlo (MCMC)	Latent Dirichlet Allocation	Stochastic Block Model	Random Walk		
DL	[2016,2020]	Convolutional Neural Network (CNN)	Generative Adversarial Network (GAN)	Support Vector Machine (SVM)	Stochastic Block Model	Reinforcement Learning (RL)	Long Short-term Memory (LSTM)	Graph Convolutional Network (GCN)	Markov Chain Monte Carlo (MCMC)	Radom Graph Model	Latent Dirichlet Allocation
	[2021,2022]	Convolutional Neural Network (CNN)	Generative Adversarial Network (GAN)	Reinforcement Learning (RL)	Graph Neural Network (GNN)	Long Short-term Memory (LSTM)	Support Vector Machine (SVM)	Non-negative Matrix Factorization (NMF)	Unsupervised Learning	Markov Chain Monte Carlo (MCMC)	Principal Component Analysis (PCA)
GAI	2023	Convolutional Neural Network (CNN)	Generative Adversarial Network (GAN)	Long Short-term Memory (LSTM)	Reinforcement Learning (RL)	Support Vector Machine (SVM)	Graph Neural Network (GNN)	Random Forest (RF)	Markov Chain Monte Carlo (MCMC)	Unsupervised Learning	Large Language Model (LLM)
	2024	Convolutional Neural Network (CNN)	Long Short-term Memory (LSTM)	Support Vector Machine (SVM)	Reinforcement Learning (RL)	Graph Convolutional Network (GCN)	Generative Adversarial Network (GAN)	Random Forest (RF)	Gradient Boosting	Gaussian Process Regression	Large Language Model (LLM)
	2025	Convolutional Neural Network (CNN)	Graph Neural Network (GNN)	Long Short-term Memory (LSTM)	Artificial Neural Network (ANN)	Support Vector Machine (SVM)	Generative Adversarial Network (GAN)	Graph Convolutional Network (GCN)	Large Language Model (LLM)	Gated Recurrent Unit (GRU)	Reinforcement Learning (RL)

Table 12. Fraction of papers with alphabetically listed authors

Discipline	OpenAlex author/%	Random author (σ)/%	Actual fraction/%
Biology	16.59	15.03 (0.01)	1.57
Chemistry	17.80	16.70 (0.02)	1.10
Geology	22.15	20.28 (0.03)	1.87
Materials science	14.87	13.82 (0.02)	1.05
Medicine	16.07	14.38 (0.01)	1.69
Physics	20.56	18.67 (0.01)	1.89
Total	16.89	15.31 (0.01)	1.58

References

- [1] Hanchen Wang, Tianfan Fu, Yuanqi Du, Wenhao Gao, Kexin Huang, Ziming Liu, Payal Chandak, Shengchao Liu, Peter Van Katwyk, Andreea Deac, et al. “Scientific discovery in the age of artificial intelligence”. In: *Nature* 620.7972 (2023), pp. 47–60.
- [2] John J Hopfield. “Neural networks and physical systems with emergent collective computational abilities.” In: *Proceedings of the national academy of sciences* 79.8 (1982), pp. 2554–2558.
- [3] John J Hopfield. “Neurons with graded response have collective computational properties like those of two-state neurons.” In: *Proceedings of the national academy of sciences* 81.10 (1984), pp. 3088–3092.
- [4] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. “Deep learning”. In: *nature* 521.7553 (2015), pp. 436–444.
- [5] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. “Imagenet classification with deep convolutional neural networks”. In: *Advances in neural information processing systems* 25 (2012).
- [6] Geoffrey E Hinton and Ruslan R Salakhutdinov. “Reducing the dimensionality of data with neural networks”. In: *science* 313.5786 (2006), pp. 504–507.
- [7] Geoffrey E Hinton. “Training products of experts by minimizing contrastive divergence”. In: *Neural computation* 14.8 (2002), pp. 1771–1800.
- [8] Brian Kuhlman, Gautam Dantas, Gregory C Ireton, Gabriele Varani, Barry L Stoddard, and David Baker. “Design of a novel globular protein fold with atomic-level accuracy”. In: *science* 302.5649 (2003), pp. 1364–1368.
- [9] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Židek, Anna Potapenko, et al. “Highly accurate protein structure prediction with AlphaFold”. In: *nature* 596.7873 (2021), pp. 583–589.
- [10] Jian Gao and Dashun Wang. “Quantifying the use and potential benefits of artificial intelligence in scientific research”. In: *Nature Human Behaviour* (2024), pp. 1–12.
- [11] James A Evans. “Electronic publication and the narrowing of science and scholarship”. In: *science* 321.5887 (2008), pp. 395–399.
- [12] Tufan Adıgüzel, Mehmet Haldun Kaya, and Fatih Kürşat Cansu. “Revolutionizing education with AI: Exploring the transformative potential of ChatGPT”. In: *Contemporary Educational Technology* (2023).
- [13] Selin Akgun and Christine Greenhow. “Artificial intelligence in education: Addressing ethical challenges in K-12 settings”. In: *AI and Ethics* 2.3 (2022), pp. 431–440.
- [14] Bertalan Meskó and Eric J Topol. “The imperative for regulatory oversight of large language models (or generative AI) in healthcare”. In: *NPJ digital medicine* 6.1 (2023), p. 120.
- [15] Hui Wen Loh, Chui Ping Ooi, Silvia Seoni, Prabal Datta Barua, Filippo Molinari, and U Rajendra Acharya. “Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022)”. In: *Computer Methods and Programs in Biomedicine* (2022), p. 107161.
- [16] Imran Ahmed, Gwanggil Jeon, and Francesco Piccialli. “From artificial intelligence to explainable artificial intelligence in industry 4.0: a survey on what, how, and where”. In: *IEEE Transactions on Industrial Informatics* 18.8 (2022), pp. 5031–5042.
- [17] Mihaly Varadi, Stephen Anyango, Mandar Deshpande, Sreenath Nair, Cindy Natassia, Galabina Yordanova, David Yuan, Oana Stroe, Gemma Wood, Agata Laydon, et al. “AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models”. In: *Nucleic acids research* 50.D1 (2022), pp. D439–D444.
- [18] Jonas Degraeve, Federico Felici, Jonas Buchli, Michael Neunert, Brendan Tracey, Francesco Carpanese, Timo Ewalds, Roland Hafner, Abbas Abdolmaleki, Diego de Las Casas, et al. “Magnetic control of tokamak plasmas through deep reinforcement learning”. In: *Nature* 602.7897 (2022), pp. 414–419.
- [19] Alhussein Fawzi, Matej Balog, Aja Huang, Thomas Hubert, Bernardino Romera-Paredes, Mohammadamin Barekatin, Alexander Novikov, Francisco J R Ruiz, Julian Schrittwieser, Grzegorz Swirszcz, et al. “Discovering faster matrix multiplication algorithms with reinforcement learning”. In: *Nature* 610.7930 (2022), pp. 47–53.
- [20] Daniil A Boiko, Robert MacKnight, Ben Kline, and Gabe Gomes. “Autonomous chemical research with large language models”. In: *Nature* 624.7992 (2023), pp. 570–578.

- [21] Chris Stokel-Walker and Richard Van Noorden. “What ChatGPT and generative AI mean for science”. In: *Nature* 614.7947 (2023), pp. 214–216.
- [22] Aidan Gilson, Conrad W Safranek, Thomas Huang, Vimig Socrates, Ling Chi, Richard Andrew Taylor, David Chartash, 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”. In: *JMIR Medical Education* 9.1 (2023), e45312.
- [23] Ali Salimi and Hady Saheb. “Large language models in ophthalmology scientific writing: ethical considerations blurred lines or not at all?” In: *American journal of ophthalmology* 254 (2023), pp. 177–181.
- [24] Weixin Liang, Yaohui Zhang, Zhengxuan Wu, Haley Lepp, Wenlong Ji, Xuandong Zhao, Hancheng Cao, Sheng Liu, Siyu He, Zhi Huang, Diyi Yang, Christopher Potts, Christopher D Manning, and James Y. Zou. “Mapping the Increasing Use of LLMs in Scientific Papers”. In: *First Conference on Language Modeling*. 2024.
- [25] Taesoon Hwang, Nishant Aggarwal, Pir Zarak Khan, Thomas Roberts, Amir Mahmood, Madlen M Griffiths, Nick Parsons, and Saboor Khan. “Can ChatGPT assist authors with abstract writing in medical journals? Evaluating the quality of scientific abstracts generated by ChatGPT and original abstracts”. In: *PLoS One* 19.2 (2024), e0297701.
- [26] Dmitry Kobak, Rita González-Márquez, Emőke-Ágnes Horvát, and Jan Lause. “Delving into LLM-assisted writing in biomedical publications through excess vocabulary”. In: *Science Advances* 11.27 (2025), eadt3813.
- [27] Zachary Wojtowicz and Simon DeDeo. “Undermining mental proof: How ai can make cooperation harder by making thinking easier”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 39. 2. 2025, pp. 1592–1600.
- [28] Morgan R Frank, Dashun Wang, Manuel Cebrian, and Iyad Rahwan. “The evolution of citation graphs in artificial intelligence research”. In: *Nature Machine Intelligence* 1.2 (2019), pp. 79–85.
- [29] OpenAlex. *OpenAlex*. <https://openalex.org/>. 2025.
- [30] Clarivate. *Web of Science*. <https://www.webofscience.com>. 2025.
- [31] Philippe Mongeon and Adèle Paul-Hus. “The journal coverage of Web of Science and Scopus: a comparative analysis”. In: *Scientometrics* 106 (2016), pp. 213–228.
- [32] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*. Ed. by Jill Burstein, Christy Doran, and Tamar Solorio. Association for Computational Linguistics, 2019, pp. 4171–4186.
- [33] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. “Transformers: State-of-the-Art Natural Language Processing”. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. Online: Association for Computational Linguistics, Oct. 2020, pp. 38–45.
- [34] Iz Beltagy, Kyle Lo, and Arman Cohan. “SciBERT: A Pretrained Language Model for Scientific Text”. In: *EMNLP/IJCNLP (1)*. Association for Computational Linguistics, 2019, pp. 3613–3618.
- [35] Arman Cohan, Sergey Feldman, Iz Beltagy, Doug Downey, and Daniel S. Weld. “SPECTER: Document-level Representation Learning using Citation-informed Transformers”. In: *ACL*. Association for Computational Linguistics, 2020, pp. 2270–2282.
- [36] Amanpreet Singh, Mike D’Arcy, Arman Cohan, Doug Downey, and Sergey Feldman. “SciRepEval: A Multi-Format Benchmark for Scientific Document Representations”. In: *EMNLP*. Association for Computational Linguistics, 2023, pp. 5548–5566.
- [37] J Richard Landis and Gary G Koch. “The measurement of observer agreement for categorical data”. In: *biometrics* (1977), pp. 159–174.
- [38] Joseph L Fleiss. “Measuring nominal scale agreement among many raters.” In: *Psychological bulletin* 76.5 (1971), p. 378.
- [39] Johan SG Chu and James A Evans. “Slowed canonical progress in large fields of science”. In: *Proceedings of the National Academy of Sciences* 118.41 (2021), e2021636118.
- [40] Clarivate. <https://jcr.clarivate.com/jcr/home>, title = Journal Citation Reports. 2021.

- [41] John PA Ioannidis, Kevin W Boyack, and Richard Klavans. “Estimates of the continuously publishing core in the scientific workforce”. In: *PloS one* 9.7 (2014), e101698.
- [42] David G Kendall. “Birth-and-death processes, and the theory of carcinogenesis”. In: *Biometrika* 47.1/2 (1960), pp. 13–21.
- [43] Santo Fortunato, Carl T Bergstrom, Katy Börner, James A Evans, Dirk Helbing, Staša Milojević, Alexander M Petersen, Filippo Radicchi, Roberta Sinatra, Brian Uzzi, et al. “Science of science”. In: *Science* 359.6379 (2018), eaao0185.
- [44] Staša Milojević. “Quantifying the cognitive extent of science”. In: *Journal of Informetrics* 9.4 (2015), pp. 962–973.
- [45] Peter McMahan and James Evans. “Ambiguity and engagement”. In: *American Journal of Sociology* 124.3 (2018), pp. 860–912.
- [46] Robert K Merton. “The Matthew effect in science: The reward and communication systems of science are considered.” In: *Science* 159.3810 (1968), pp. 56–63.
- [47] Jessica G Borger, Ashley P Ng, Holly Anderton, George W Ashdown, Megan Auld, Marnie E Blewitt, Daniel V Brown, Melissa J Call, Peter Collins, Saskia Freytag, et al. “Artificial intelligence takes center stage: exploring the capabilities and implications of ChatGPT and other AI-assisted technologies in scientific research and education”. In: *Immunology and cell biology* 101.10 (2023), pp. 923–935.
- [48] Neil D Lawrence and Jessica Montgomery. “Accelerating AI for science: open data science for science”. In: *Royal Society Open Science* 11.8 (2024), p. 231130.
- [49] Ross D King, Jem Rowland, Stephen G Oliver, Michael Young, Wayne Aubrey, Emma Byrne, Maria Liakata, Magdalena Markham, Pinar Pir, Larisa N Soldatova, et al. “The automation of science”. In: *Science* 324.5923 (2009), pp. 85–89.
- [50] Benjamin Burger, Phillip M Maffettone, Vladimir V Gusev, Catherine M Aitchison, Yang Bai, Xiaoyan Wang, Xiaobo Li, Ben M Alston, Buyi Li, Rob Clowes, et al. “A mobile robotic chemist”. In: *Nature* 583.7815 (2020), pp. 237–241.
- [51] Alexander Krauss. “Debunking revolutionary paradigm shifts: evidence of cumulative scientific progress across science”. In: *Proceedings A*. Vol. 480. 2302. The Royal Society. 2024, p. 20240141.
- [52] Microsoft. *Microsoft Academic Graph*. <https://www.microsoft.com/en-us/research/project/microsoft-academic-graph>. 2015.
- [53] Aminer. *Open Academic Graph*. <https://www.aminer.cn/oag-2-1>. 2020.
- [54] Alan Porter and Ismael Rafols. “Is science becoming more interdisciplinary? Measuring and mapping six research fields over time”. In: *Scientometrics* 81.3 (2009), pp. 719–745.
- [55] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. “Learning representations by back-propagating errors”. In: *nature* 323.6088 (1986), pp. 533–536.
- [56] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. “Backpropagation applied to handwritten zip code recognition”. In: *Neural computation* 1.4 (1989), pp. 541–551.
- [57] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. “Deep residual learning for image recognition”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.
- [58] Hugging Face. *BERT pretrain model from Hugging Face*. https://huggingface.co/docs/transformers/model_doc/bert#transformers.BertForSequenceClassification.
- [59] Vedran Sekara, Pierre Deville, Sebastian E Ahnert, Albert-László Barabási, Roberta Sinatra, and Sune Lehmann. “The chaperone effect in scientific publishing”. In: *Proceedings of the National Academy of Sciences* 115.50 (2018), pp. 12603–12607.
- [60] Tianqi Chen and Carlos Guestrin. “Xgboost: A scalable tree boosting system”. In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 2016, pp. 785–794.
- [61] Ryan Hill, Yian Yin, Carolyn Stein, Xizhao Wang, Dashun Wang, and Benjamin F Jones. “The pivot penalty in research”. In: *Nature* (2025), pp. 1–8.
- [62] Staša Milojević, Filippo Radicchi, and John P Walsh. “Changing demographics of scientific careers: The rise of the temporary workforce”. In: *Proceedings of the National Academy of Sciences* 115.50 (2018), pp. 12616–12623.
- [63] Fengli Xu, Lingfei Wu, and James Evans. “Flat teams drive scientific innovation”. In: *Proceedings of the National Academy of Sciences* 119.23 (2022), e2200927119.
- [64] Yiling Lin, Carl Benedikt Frey, and Lingfei Wu. “Remote collaboration fuses fewer breakthrough ideas”. In: *Nature* 623.7989 (2023), pp. 987–991.

- [65] John Frank Charles Kingman. *Poisson processes*. Vol. 3. Clarendon Press, 1992.
- [66] Torben Meisling. “Discrete-time queuing theory”. In: *Operations Research* 6.1 (1958), pp. 96–105.
- [67] Jessica S Damoiseaux, Serge ARB Rombouts, Frederik Barkhof, Philip Scheltens, Cornelis J Stam, Stephen M Smith, and Christian F Beckmann. “Consistent resting-state networks across healthy subjects”. In: *Proceedings of the national academy of sciences* 103.37 (2006), pp. 13848–13853.
- [68] Marwin HS Segler, Mike Preuss, and Mark P Waller. “Planning chemical syntheses with deep neural networks and symbolic AI”. In: *Nature* 555.7698 (2018), pp. 604–610.
- [69] Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, et al. “Evolutionary-scale prediction of atomic-level protein structure with a language model”. In: *Science* 379.6637 (2023), pp. 1123–1130.
- [70] Milind Kandlikar, James Risbey, and Suraje Dessai. “Representing and communicating deep uncertainty in climate-change assessments”. In: *Comptes Rendus. Géoscience* 337.4 (2005), pp. 443–455.
- [71] Sidney Redner. “How popular is your paper? An empirical study of the citation distribution”. In: *The European Physical Journal B-Condensed Matter and Complex Systems* 4.2 (1998), pp. 131–134.
- [72] Per O Seglen. “The skewness of science”. In: *Journal of the American society for information science* 43.9 (1992), pp. 628–638.
- [73] Qianyu Hao, Jingyang Fan, Fengli Xu, Jian Yuan, and Yong Li. “HLM-Cite: Hybrid Language Model Workflow for Text-based Scientific Citation Prediction”. In: *NeurIPS*. 2024.
- [74] Md Kamrul Hasan, Md Ashraful Alam, Dola Das, Eklas Hossain, and Mahmudul Hasan. “Diabetes prediction using ensembling of different machine learning classifiers”. In: *IEEE Access* 8 (2020), pp. 76516–76531.
- [75] Saloni Kumari, Deepika Kumar, and Mamta Mittal. “An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier”. In: *International Journal of Cognitive Computing in Engineering* 2 (2021), pp. 40–46.
- [76] Jack W Smith, James E Everhart, William C Dickson, William C Knowler, and Robert Scott Johannes. “Using the ADAP learning algorithm to forecast the onset of diabetes mellitus”. In: *Proceedings of the annual symposium on computer application in medical care*. 1988, p. 261.
- [77] Lingfei Wu, Dashun Wang, and James A Evans. “Large teams develop and small teams disrupt science and technology”. In: *Nature* 566.7744 (2019), pp. 378–382.
- [78] Michael Park, Erin Leahey, and Russell J Funk. “Papers and patents are becoming less disruptive over time”. In: *Nature* 613.7942 (2023), pp. 138–144.
- [79] Jeffrey W Lockhart, Molly M King, and Christin Munsch. “Name-based demographic inference and the unequal distribution of misrecognition”. In: *Nature Human Behaviour* 7.7 (2023), pp. 1084–1095.

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 (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>.

Code availability

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>.

Author contributions

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