
Supplementary information

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

In the format provided by the
authors and unedited

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

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

This supplementary information (SI) file includes Supplementary Notes Section 1-4, Supplementary Figures 1-37, and Supplementary Tables 1-12. These provide further detail and background information, and numerous extended analyses and robustness tests to the main results are provided.

Contents

1	Identifying artificial intelligence in scientific research	2
1.1	Identification model and method	2
1.2	Identification accuracy evaluation	2
1.3	Profile of the identified AI papers	2
1.4	Increasing prevalence of AI adoption in science	3
2	Extended analyses on how AI impacts individual science	3
2.1	The effect of AI on individual papers	3
2.2	The effect of AI on individual scientists' careers	4
3	Extended analyses on how AI impacts the overall science	5
3.1	The benefit of topic diversity	6
3.2	Selectivity of AI adoption across different topics	6
4	Robustness analyses	7
4.1	Robustness of results in the generative AI era	7
4.2	Robustness of results on the Web of Science dataset	7
4.3	Robustness of distance calculations in the high-dimensional paper embedding space	7
5	Supplementary figures	9
6	Supplementary tables	46

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 coefficient¹⁰. 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* 2006³, 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* 2018¹⁹, and our second is a biology paper with AI published in *Nature* 2021⁷, 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* 2023¹¹, 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*⁸, 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^{17,18}, 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 dataset¹⁵, 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 citations⁴, 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^{5,9}, both cite a 1988 paper on the ADAP learning algorithm²⁰ 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) quantiles¹ (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^{21,16} (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 researchers⁶. 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 research¹³. 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 names¹². 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^{21,16} 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^{21,16} 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^{2,14}, 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^{2,14}. 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.

5 Supplementary figures

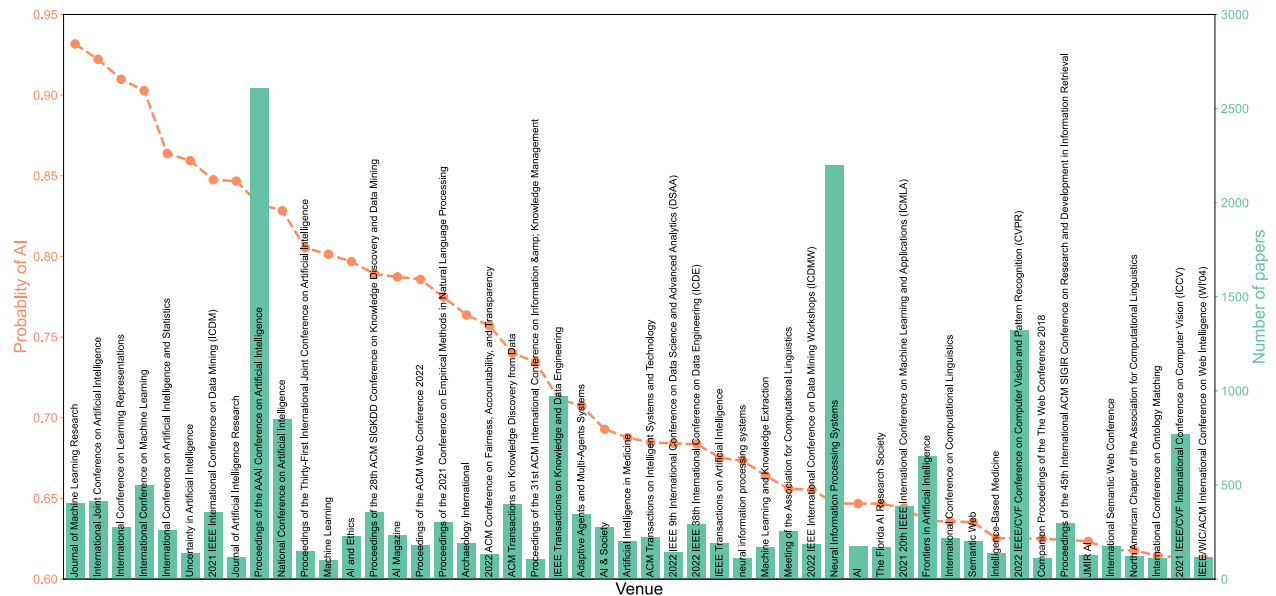


Figure S1. Probability of AI (orange) and number of papers (green) for selected venues. Final identification results combine the best models in both stages of fine-tuning. Venues are ordered according to the probability of AI papers within them.

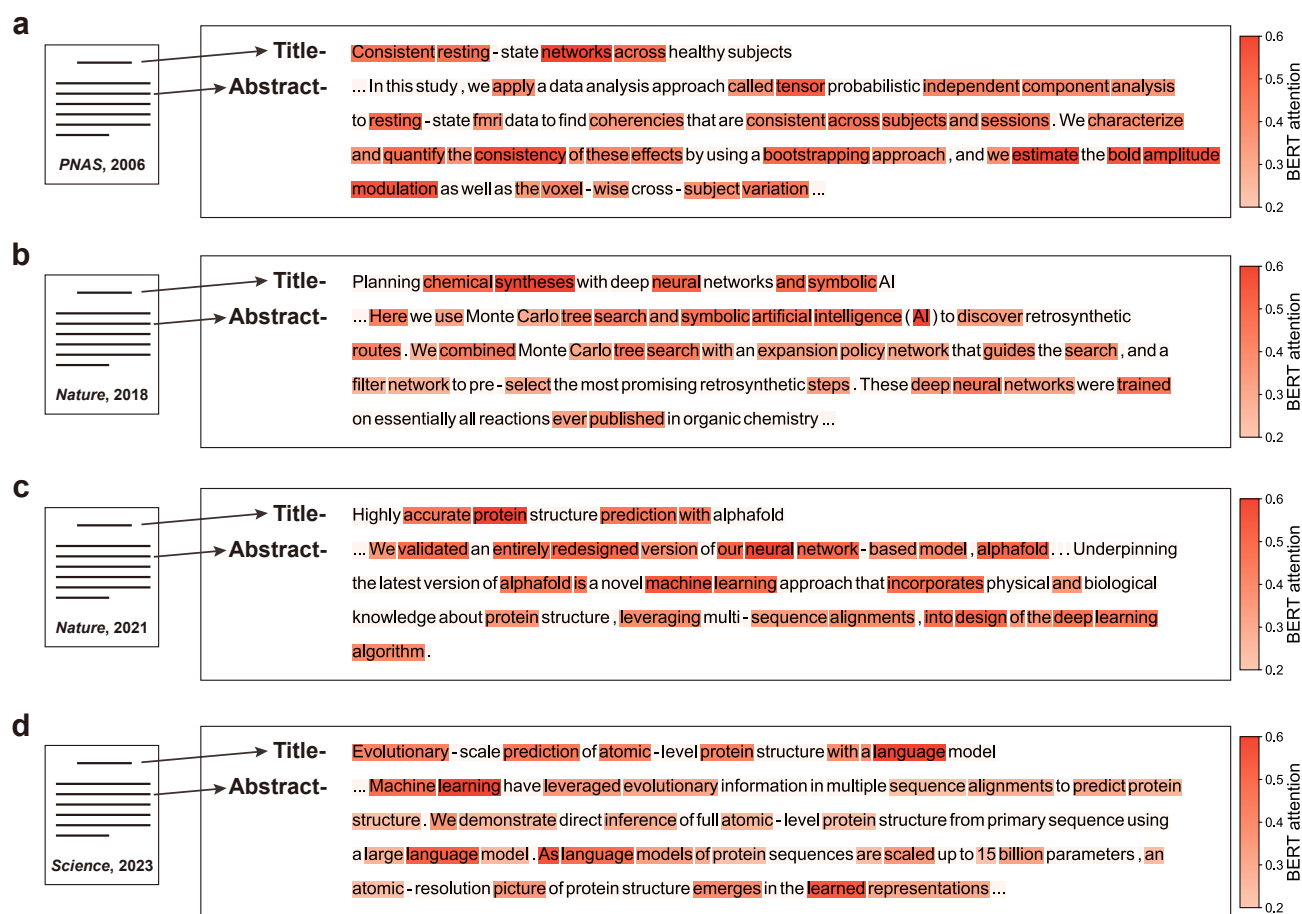


Figure S2. Correct examples of AI paper identification across different eras of AI development. (a) Example from the ML era, a medicine paper with AI published in *PNAS* 2006. (b) Example from the DL era, a chemistry paper with AI published in *Nature* 2018. (c) Example from the DL era, a biology paper with AI published in *Nature* 2021. (d) Example from the GAI era, a biology paper with AI published in *Science* 2023.

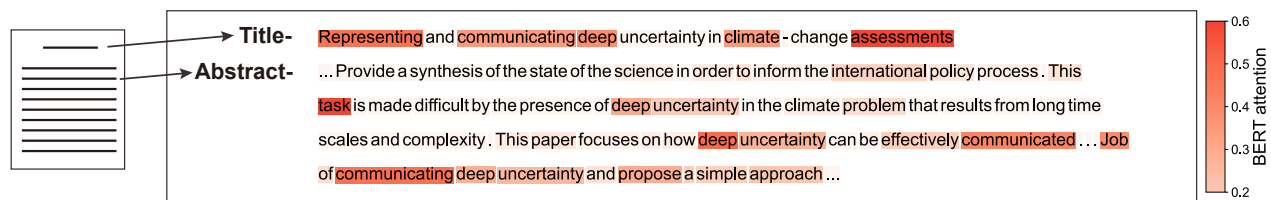


Figure S3. An incorrect example of AI paper identification. The example is a geology paper without AI published in 2006, which is mistaken to be AI-related by the identification model.

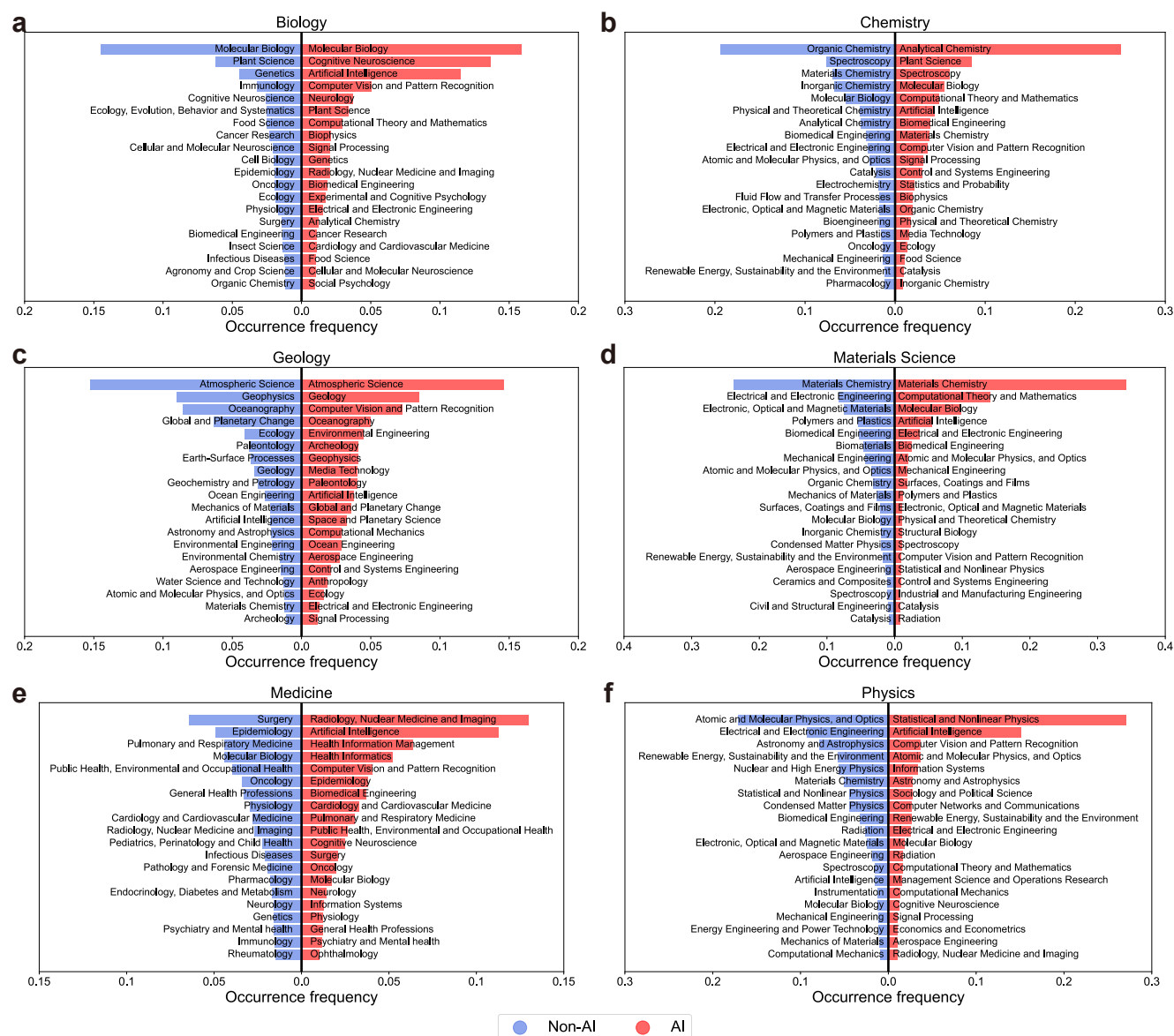


Figure S4. Topics with top occurrence frequency in AI and non-AI papers. Results show that AI’s primary contribution to conventional research fields is around computer science and machine learning algorithms. Across disciplines, identified AI papers turn out to be selected combinations of conventional research topics and AI-related techniques, including “Artificial Intelligence”, “Computer Vision and Pattern Recognition”, and “Computational Theory and Mathematics”.

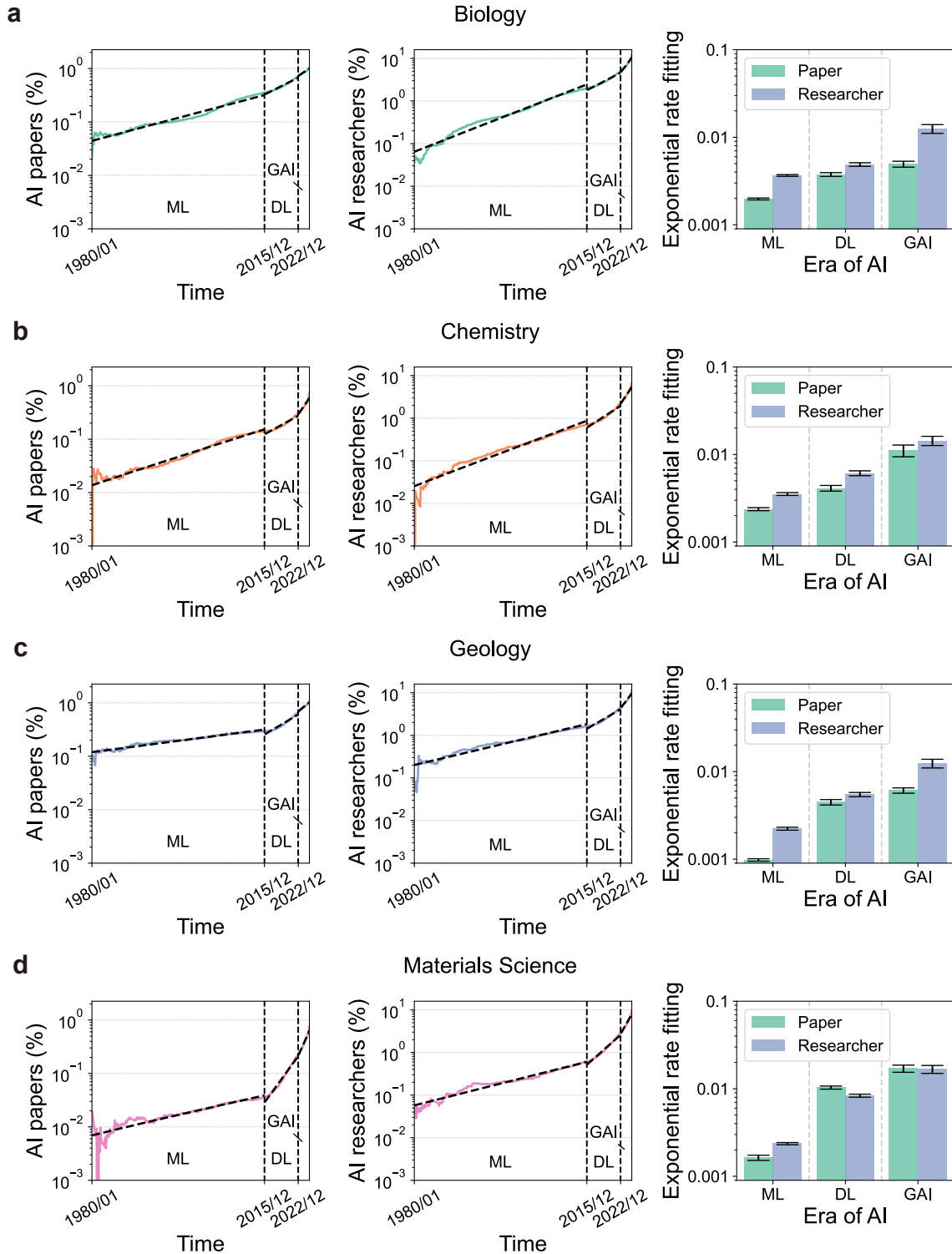


Figure S5. Proportion of papers and researchers adopting AI in each discipline with log-scale y-axis and estimated growth rate with exponential fitting for each era. Proportion of papers and researchers adopting AI fit to straight regions in the log-scale plot, indicating high precision of the exponential 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.

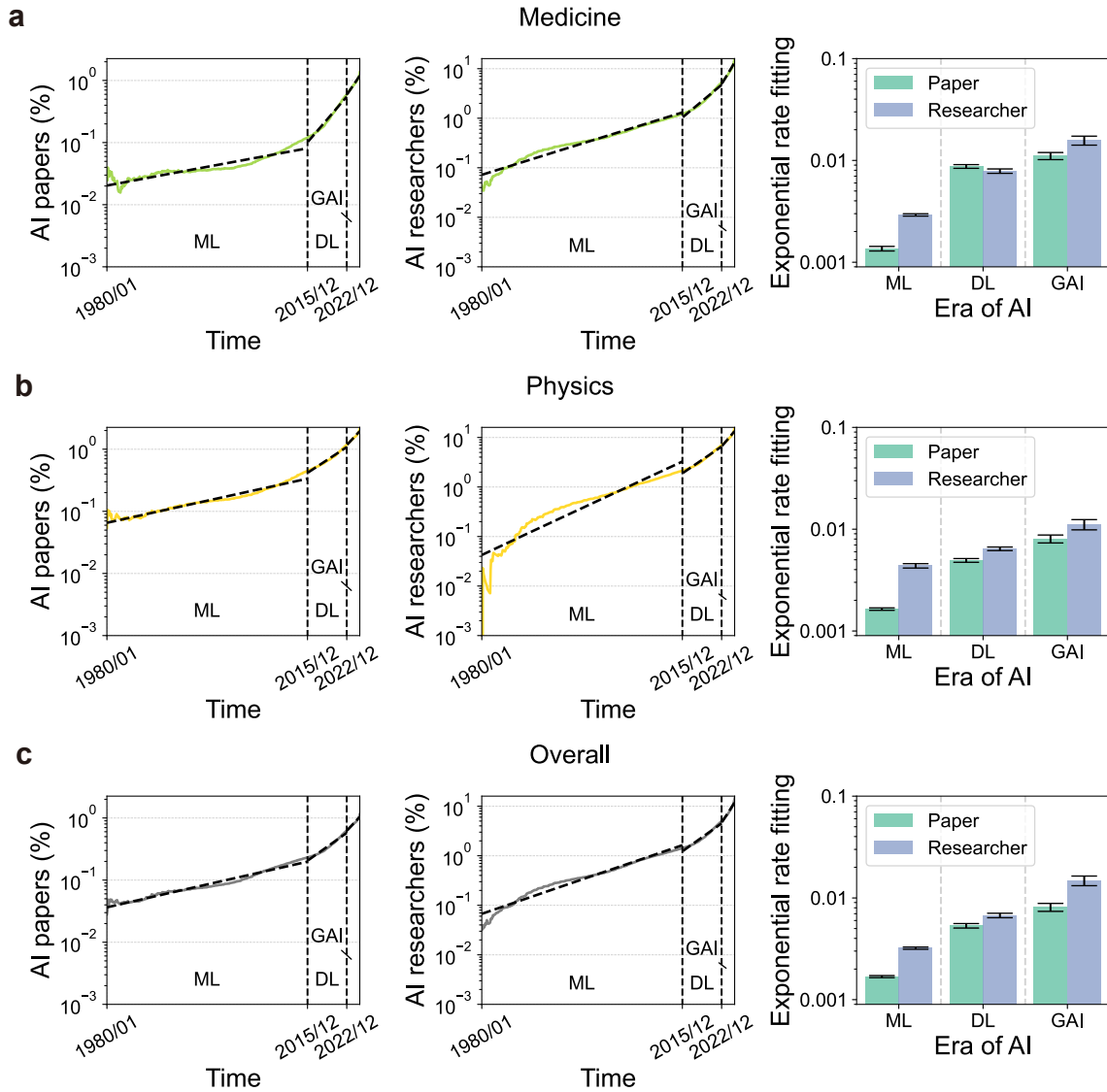


Figure S6. (Continued from Supplementary Fig. S5) The proportion of papers and researchers adopting AI in each discipline with log-scale y-axis and estimate the growth rate with exponential fitting in each era. The proportion of papers and researchers adopting AI fit to straight regions in the log-scale 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.

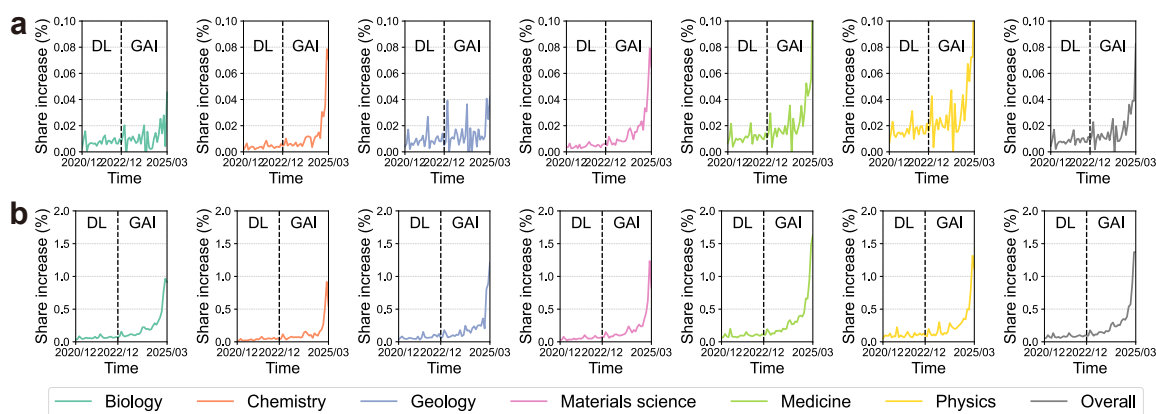


Figure S7. The monthly increase in the proportion of AI (a) papers and (b) researchers across disciplines in recent years. Following the release of ChatGPT in December 2022, growth rates in the proportion of papers and researchers initially remain consistent with prior trends. A period of time later, these growth rates begin to exhibit a marked acceleration across disciplines, providing evidence for the impact of generative AI advancement and use.

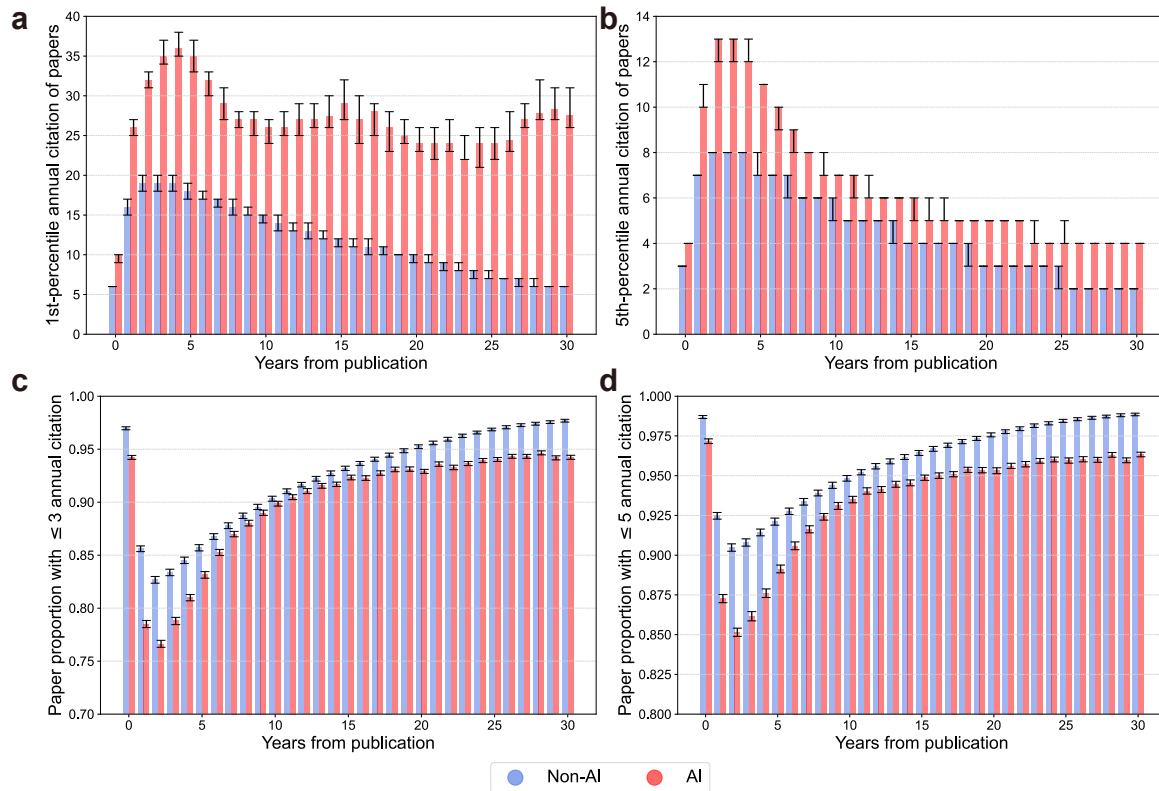


Figure S8. Alternative statistical indicators for the annual citation comparison between AI and non-AI papers. (a) Top 1st-percentile annual citation for AI and non-AI papers from year of publication ($n = 27,405,011$). (b) Top 5th-percentile annual citation for AI and non-AI papers from year of publication ($n = 27,405,011$). (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.

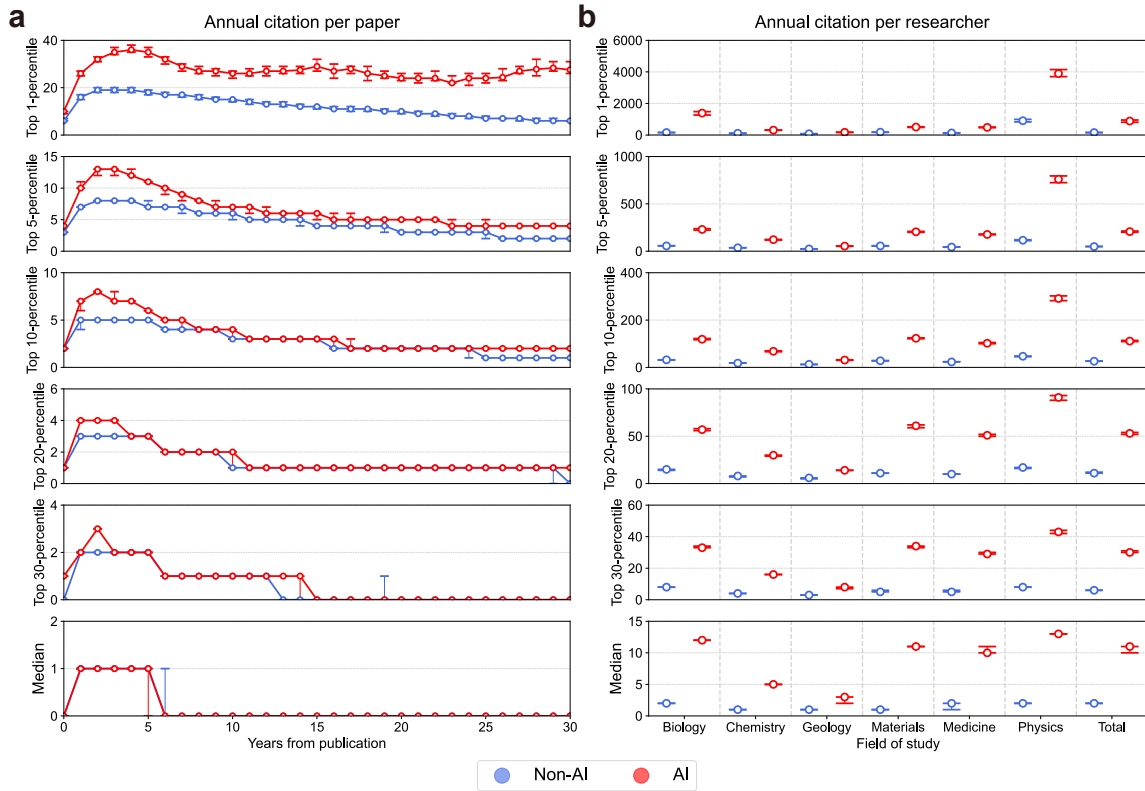


Figure S9. Different percentile statistics for the annual citation comparison between AI and non-AI papers. (a) Annual citations after publication of AI and non-AI papers ($n = 27,405,011$). **(b)** Annual citations for researchers adopting AI and their counterparts without AI ($n = 5,377,346$). Consistently 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.

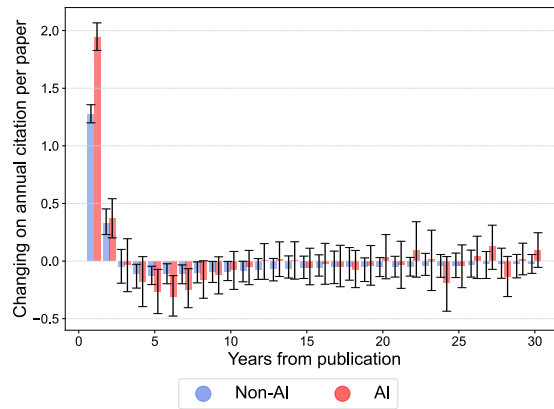


Figure S10. Comparison of “acceleration” of citation, namely year-over-year change in annual citation counts, between AI and non-AI papers. AI papers experience a faster acceleration in citation growth during the initial post-publication years, and exhibit a more rapid deceleration in annual citations 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.

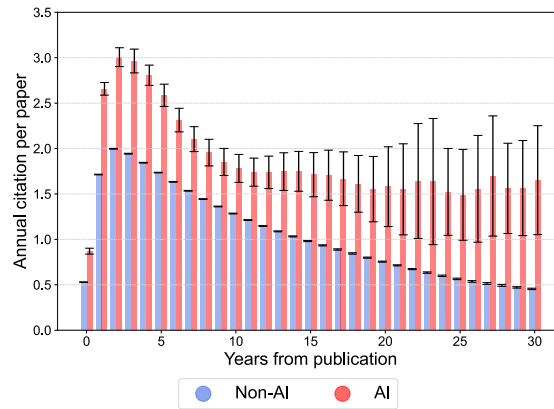


Figure S11. Annual citations after the publication of AI and non-AI original research papers, excluding review articles, editorial pieces and other special publications. Results show that AI papers attract more citations, indicating higher academic impact than papers without AI ($n = 24,867,012$). Results are consistent with the overall statistics in the main text. 99% CIs are shown as error bars centred at the mean.

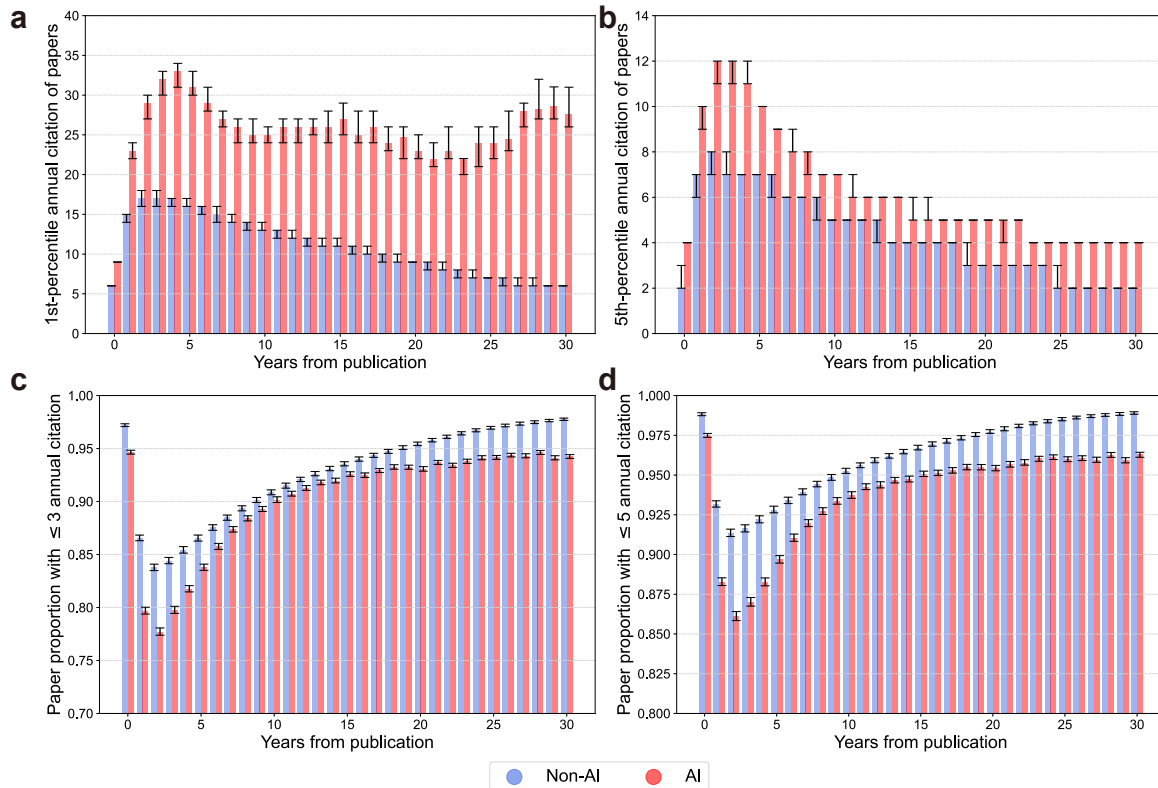


Figure S12. Alternative statistical indicators for the annual citation comparison between AI and non-AI original research papers, excluding review articles, editorial pieces and other special publications. (a) Top 1st-percentile annual citation for AI and non-AI papers from year of publication ($n = 24,867,012$). (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.

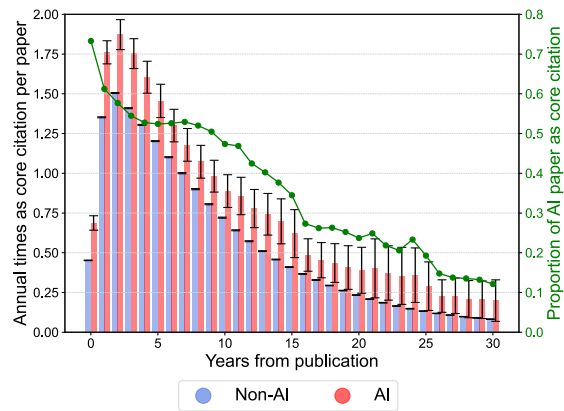


Figure S13. Distinguishing between core and superficial citations. Results show that the proportion of AI papers cited as core citations tends to decrease over time (green line), while foundational influence can still be observed in decades-old papers ($n = 27,405,011$). In terms of the absolute count 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.

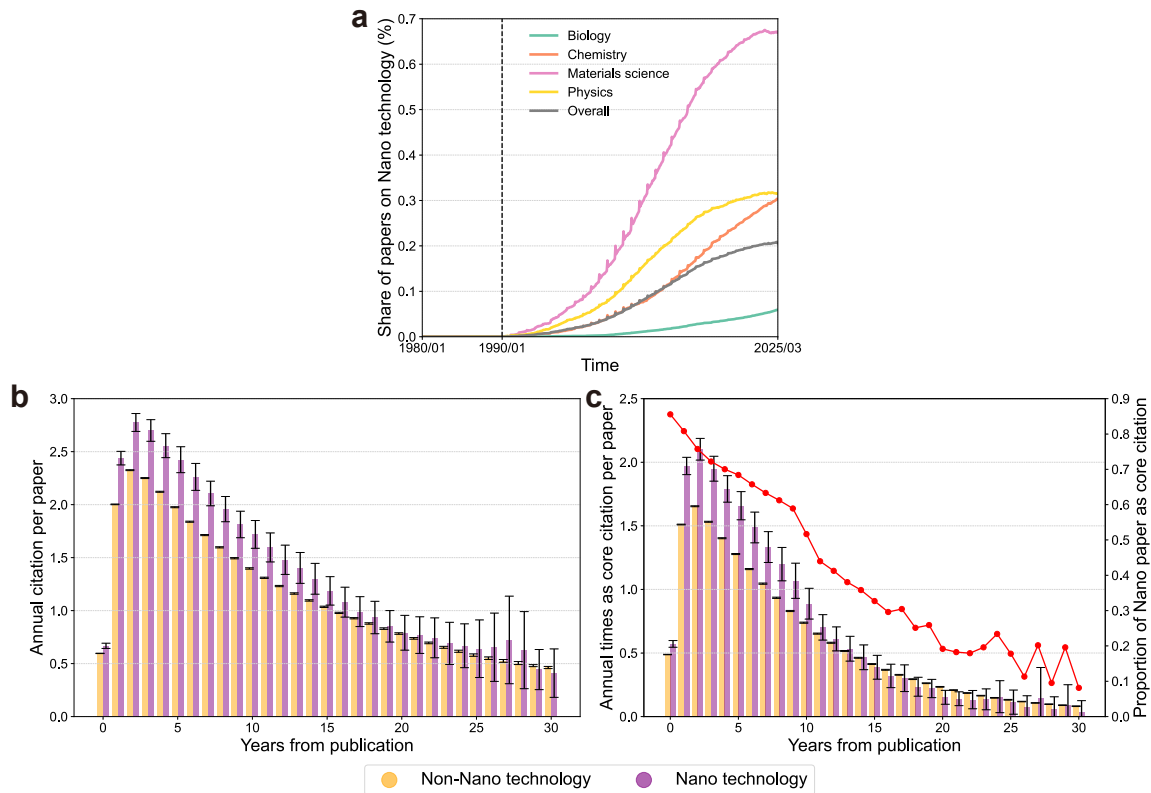


Figure S14. Comparison between citation patterns of Nanotechnology and non-Nanotechnology papers. (a) Growth of Nanotechnology over time in different disciplines. (b) Annual citations after publication of Nanotechnology and non-Nanotechnology papers ($n = 17,976,303$). (c) Annual times of being core 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.

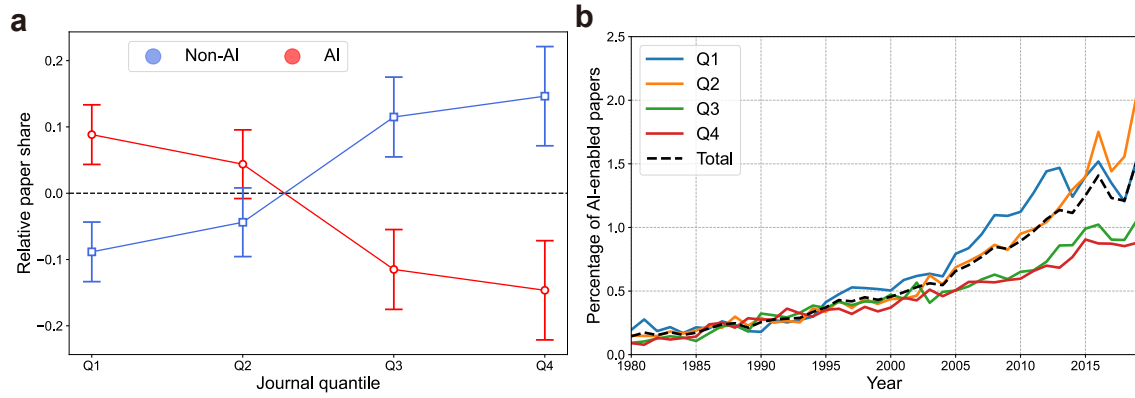


Figure S15. Distribution of AI papers across journals of varying Journal Citation Report (JCR) quantiles. (a)

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.

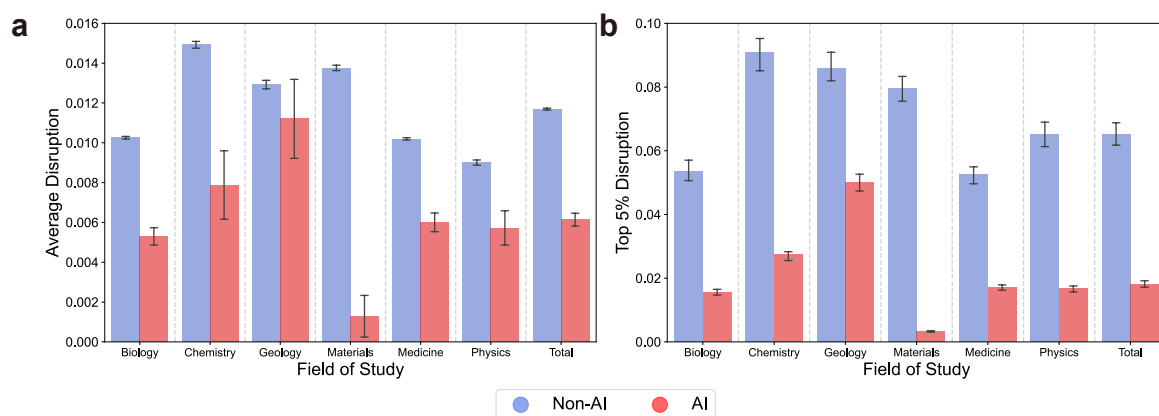


Figure S16. Comparison of disruption between AI and non-AI papers. (a) Average disruption of AI and non-AI papers in each field ($p < 0.001$, $n = 23,199,583$). (b) Top 5th-percentile disruption of AI and non-AI papers in each field ($p < 0.001$, $n = 23,199,583$). 99% CIs are shown as error bars centred at the mean or the 5% percentile. All statistical tests use a two-sided t-test.

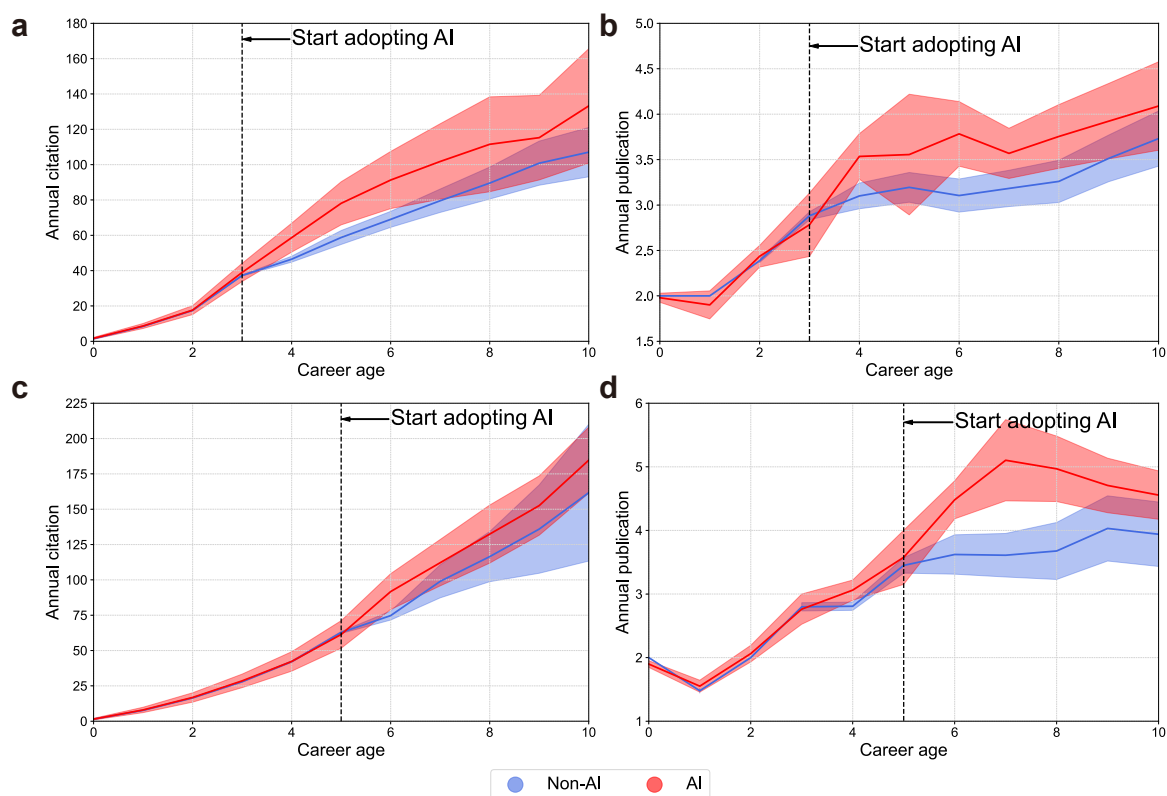


Figure S17. Match-and-comparison analysis on the role of AI in driving the advantages of scientific in productivity and scholarly impact. (a) Comparison between scientists who began adopting AI in the third year of their careers and matched counterparts who exhibited comparable annual citation counts 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.

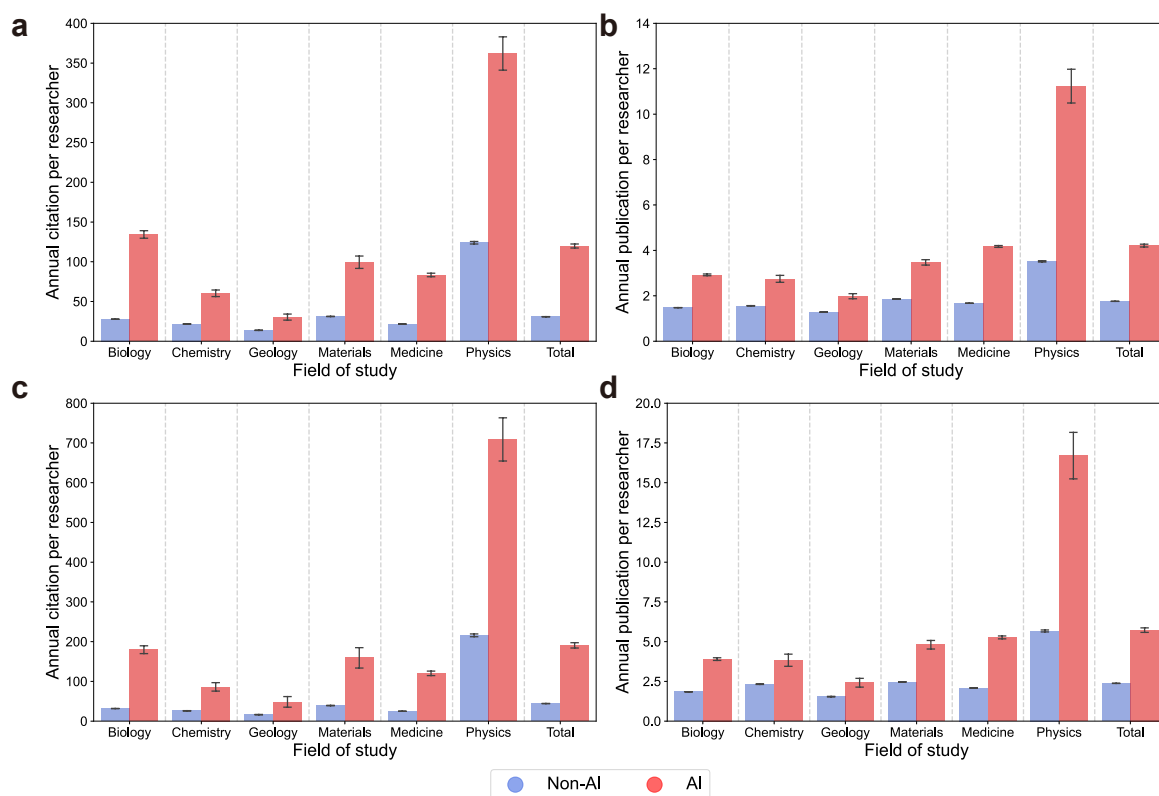


Figure S18. The impact of AI on the productivity and citation of researchers with multi-year continuous publication records. (a) Comparison of annual citations between researchers adopting AI and their counterparts without AI among those with at least five consecutive years of publications ($p < 0.001, n = 1,495,265$). (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.

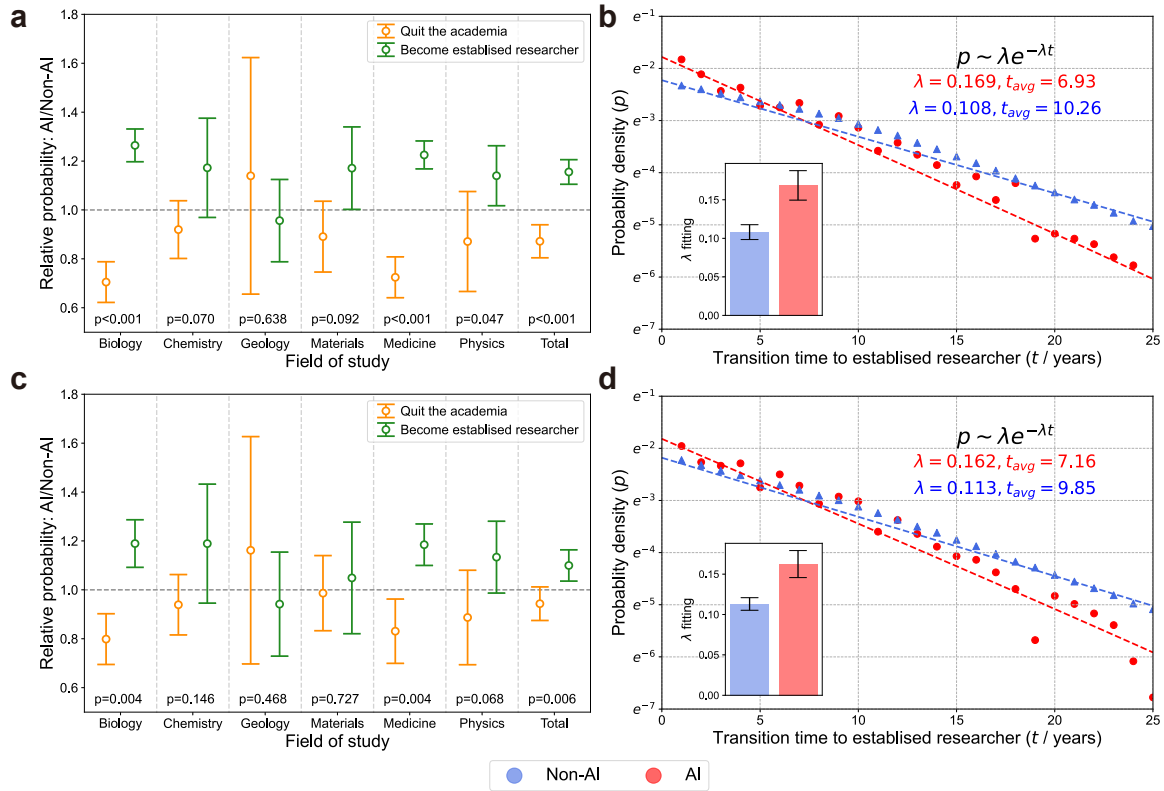


Figure S19. Impact of AI adoption on researchers' careers with different thresholds in detecting the dropout of researchers. (a) (c) The probability of two role transitions between junior scientists adopting AI and their counterparts without AI ($n = 46$ year observations for each field). (b) (d) Survival 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.

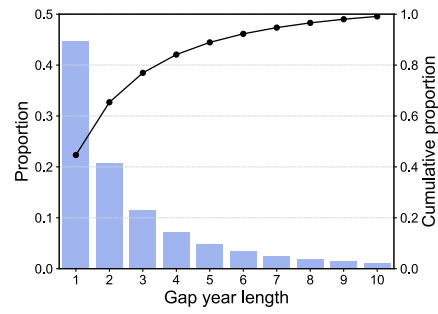


Figure S20. The distribution of gap year durations in researcher's careers. A period of gap years refers to a temporary interruption in a researcher's publication activity, followed by a subsequent resumption of publishing. Results show that 44.67% of all gap periods lasted only one year, and 76.94% lasted no more than three years.

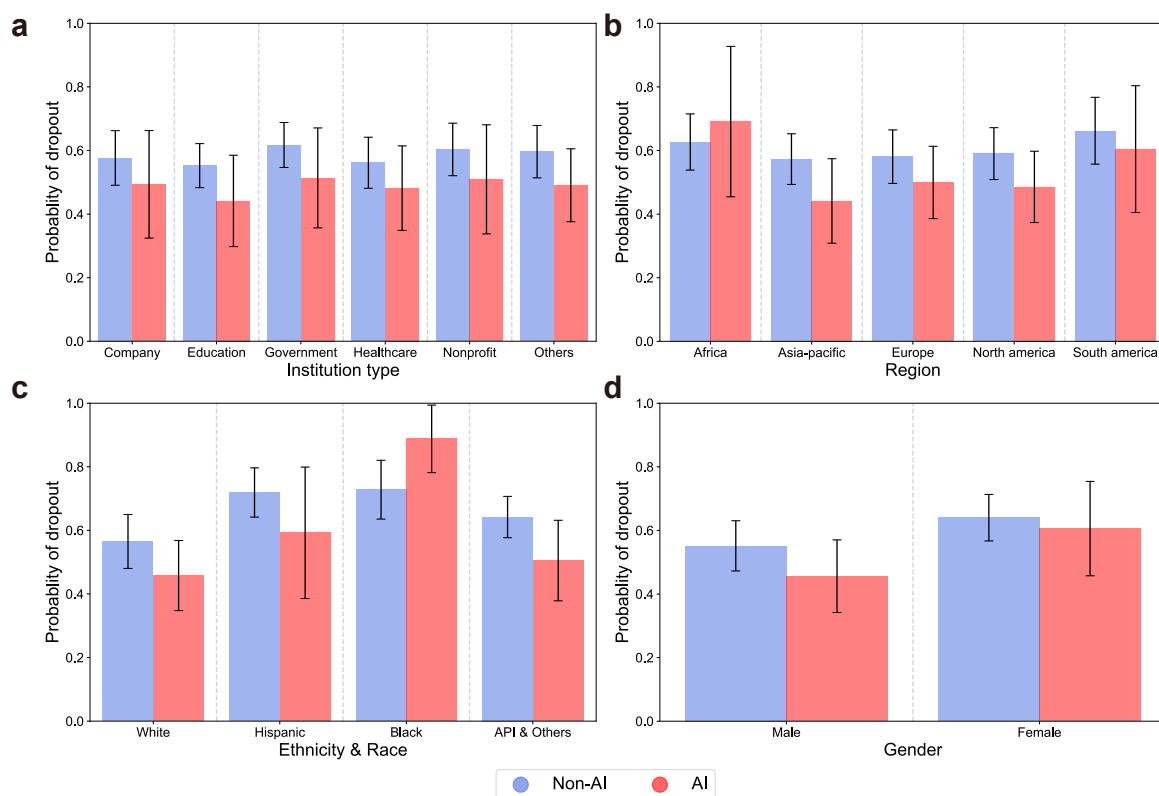


Figure S21. Impact of AI adoption on research careers across different demographic groups. (a) Researchers affiliated with institutions of different types ($n = 2,140,845$). (b) Researchers affiliated with institutions within different geographic regions ($n = 2,140,845$). (c) Researchers of different 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.

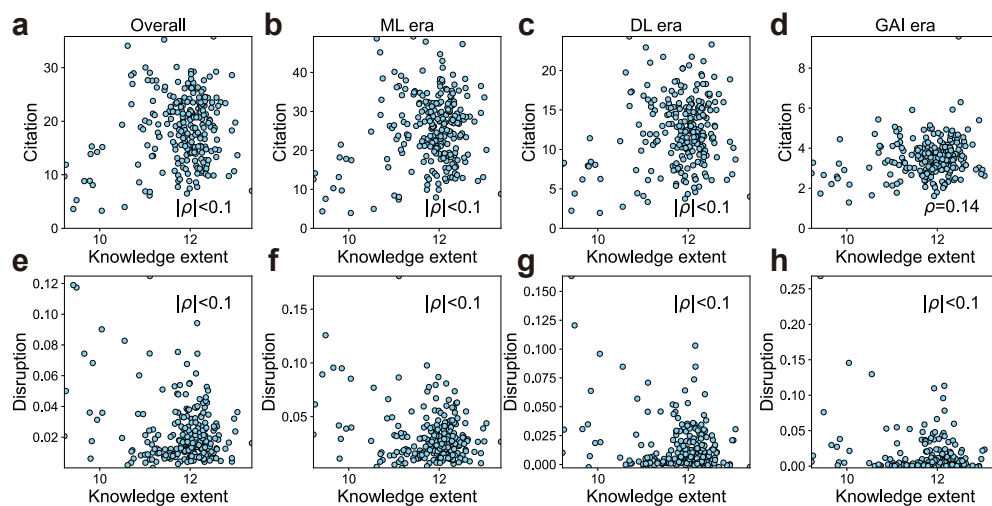


Figure S22. Correlation analysis between knowledge extent and citation or disruption of sub-fields. For all panels, each scatter represents a sub-field, and the corresponding Pearson's r is indicated correspondingly. Results show that, across sub-fields and eras of AI development, neither citation impact (as a proxy for impact) nor disruption (as a proxy for novelty) is obviously correlated with the sub-field's knowledge extent.

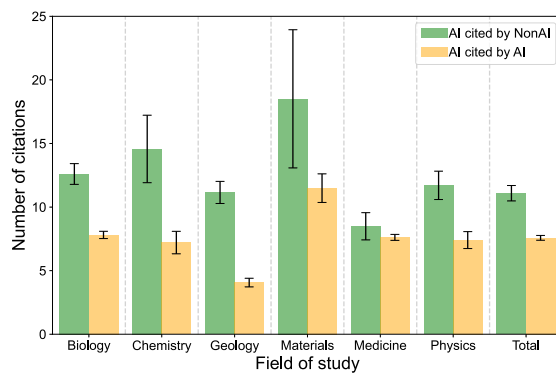


Figure S23. Selectivity of AI adoption in different topics regarding topicality itself. The citation relationships between AI and non-AI papers show that AI research across various fields is more frequently cited by non-AI studies than by AI studies themselves ($p < 0.001, n = 27,405,011$). This suggests 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.

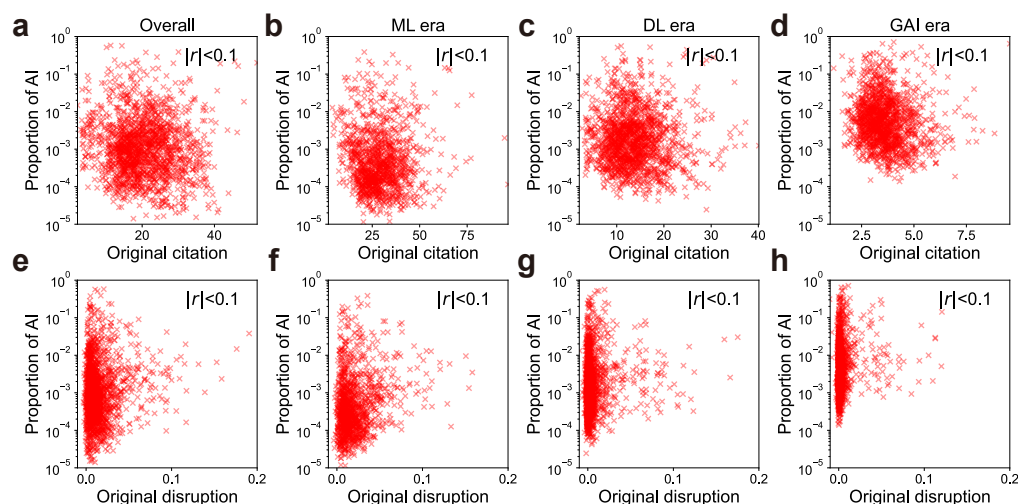


Figure S24. Selectivity of AI adoption in different topics associated with the topics' original impact. In all panels, each scatter shows the degree of AI penetration in a topic and the original citation count or disruption score within that topic. The corresponding Pearson's r is indicated in each panel. 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.

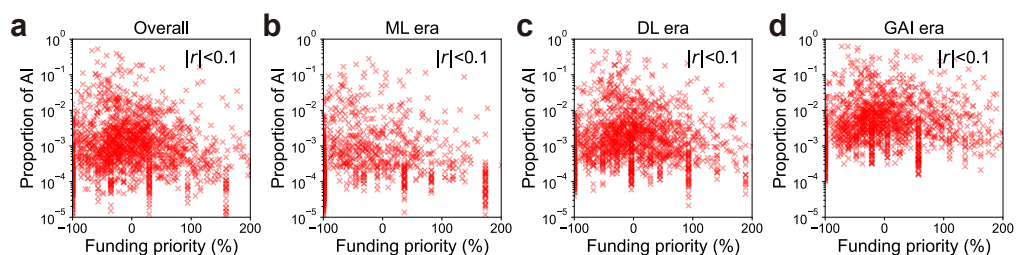


Figure S25. Selectivity of AI adoption in different topics regarding funding priority across the topics. In all panels, each scatter shows the degree of AI penetration in a topic and funding priority to that topic. The corresponding Pearson's r is indicated correspondingly in each panel. Results show 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.

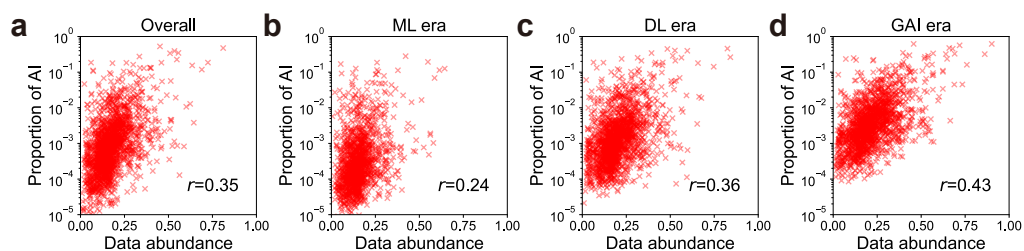


Figure S26. Selectivity of AI adoption in different topics regarding data abundance in the topics. In all panels, each scatter shows the degree of AI penetration in a topic and the data abundance in that topic. The corresponding Pearson's r is indicated correspondingly in each panel. The results show 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.

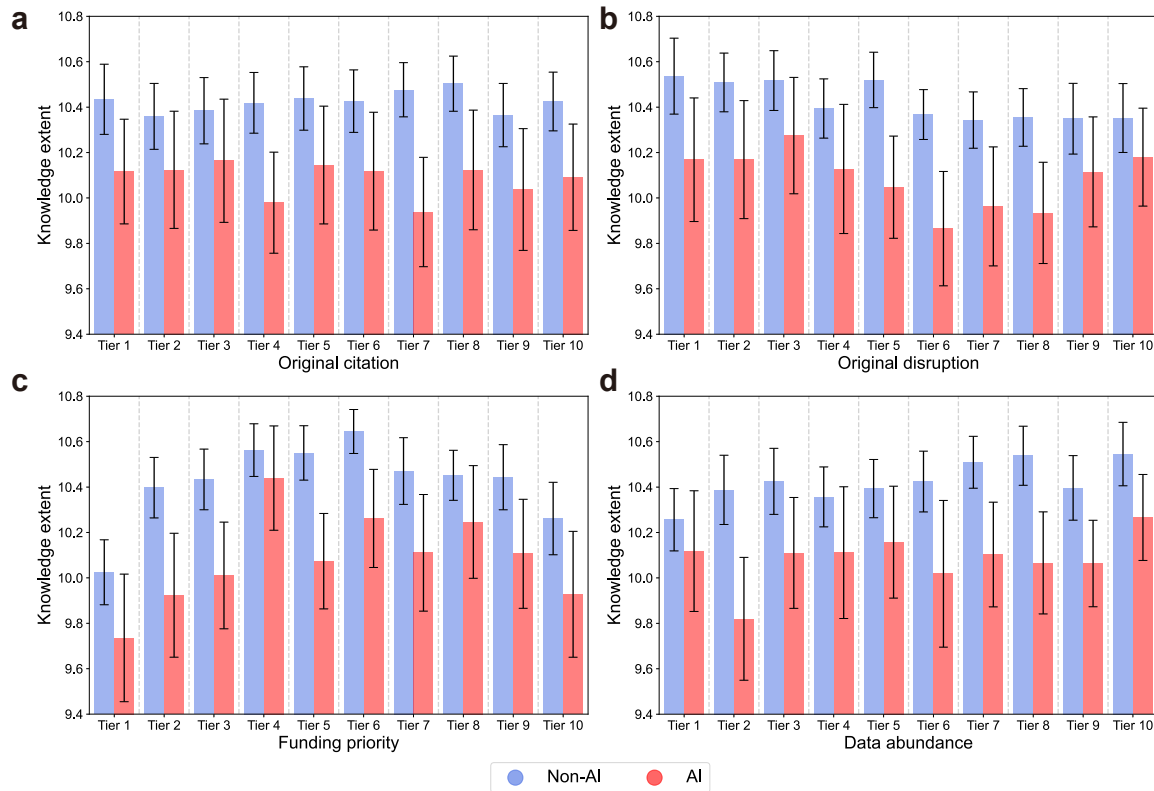


Figure S27. Stratified analyses controlling for the effects of potential confounding variables on the contracted knowledge extent of AI-augmented research. The results show that when dividing each external factor into 10 tiers based on magnitude, the contraction in knowledge extent for AI-augmented research 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.

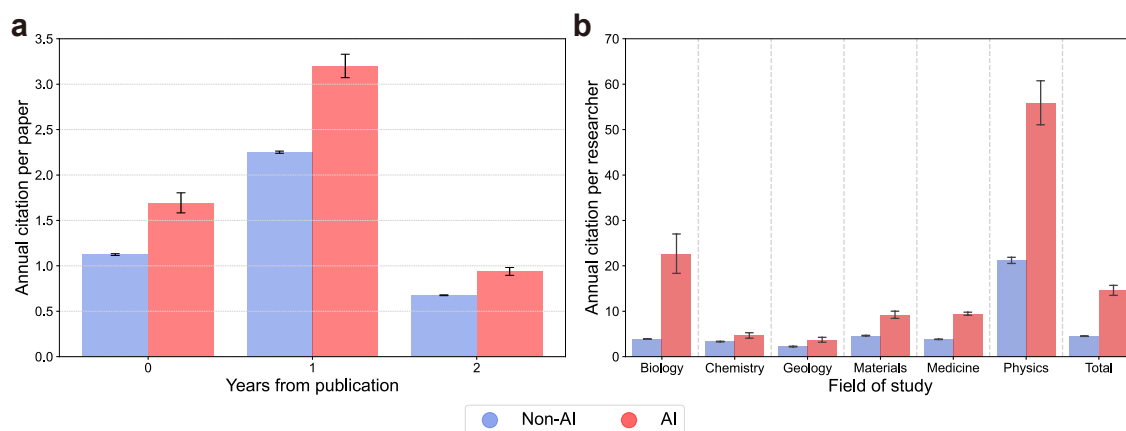


Figure S28. Replication of “AI enlarges the impact of papers” with only the subset of papers published during the GAI era. (a) Annual citations after publication of AI and non-AI papers ($n = 1,888,694$). **(b)** Average annual citations of researchers adopting AI and their counterparts without AI ($p < 0.001, n = 2,888,737$). 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.

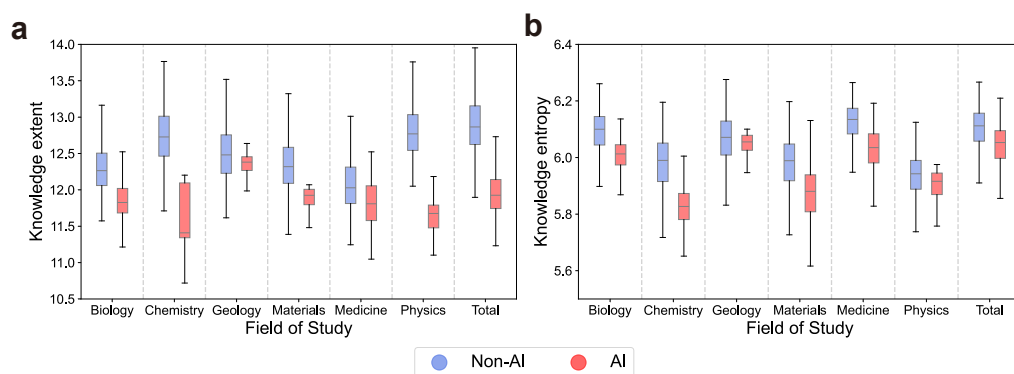


Figure S29. Replication of “AI usage is associated with a contraction in knowledge extent within and across scientific fields” with only the subset of papers published during the GAI era. (a) Knowledge extent of AI and non-AI papers in each field ($n = 1,000$ samples in each field). **(b)** Knowledge 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.

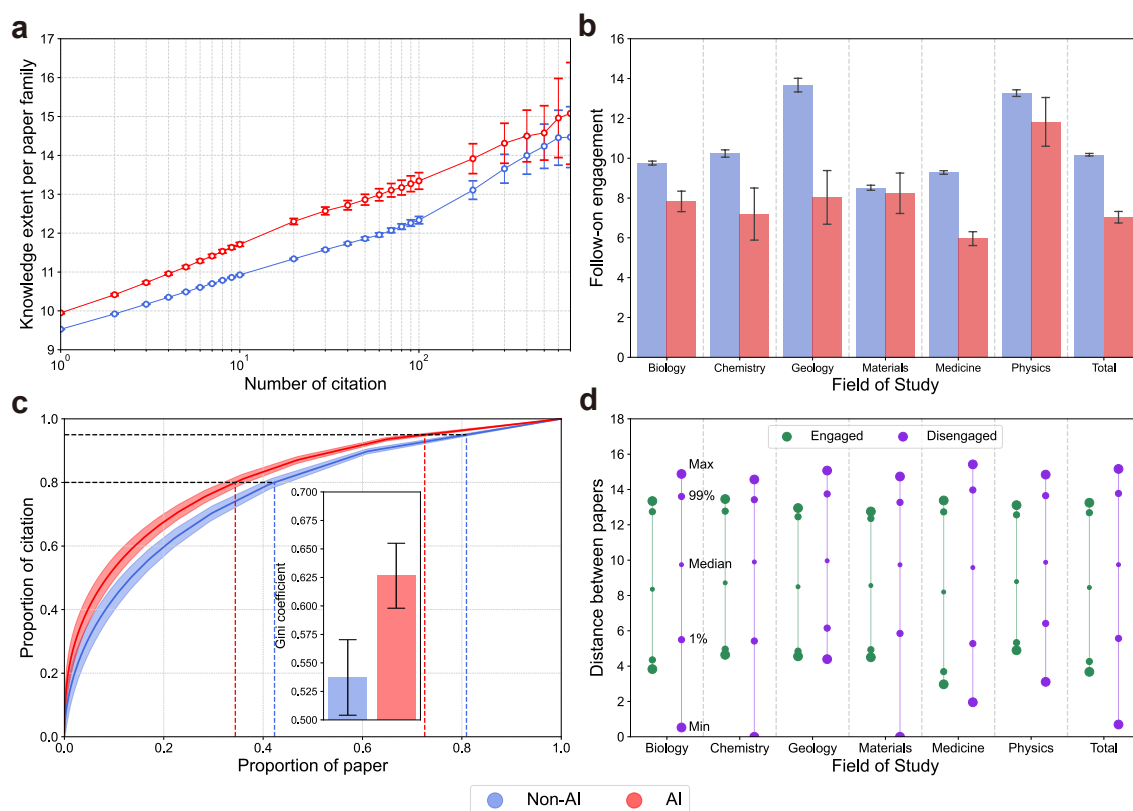


Figure S30. Replication of “reduced follow-on engagement and more overlapped works in AI research” with only the subset of papers published during the GAI era. (a) Knowledge extent of individual AI and non-AI paper families ($n = 1,888,694$). **(b)** Engagement among papers that cite AI vs. non-AI papers ($p < 0.001, n = 518,156$). **(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.

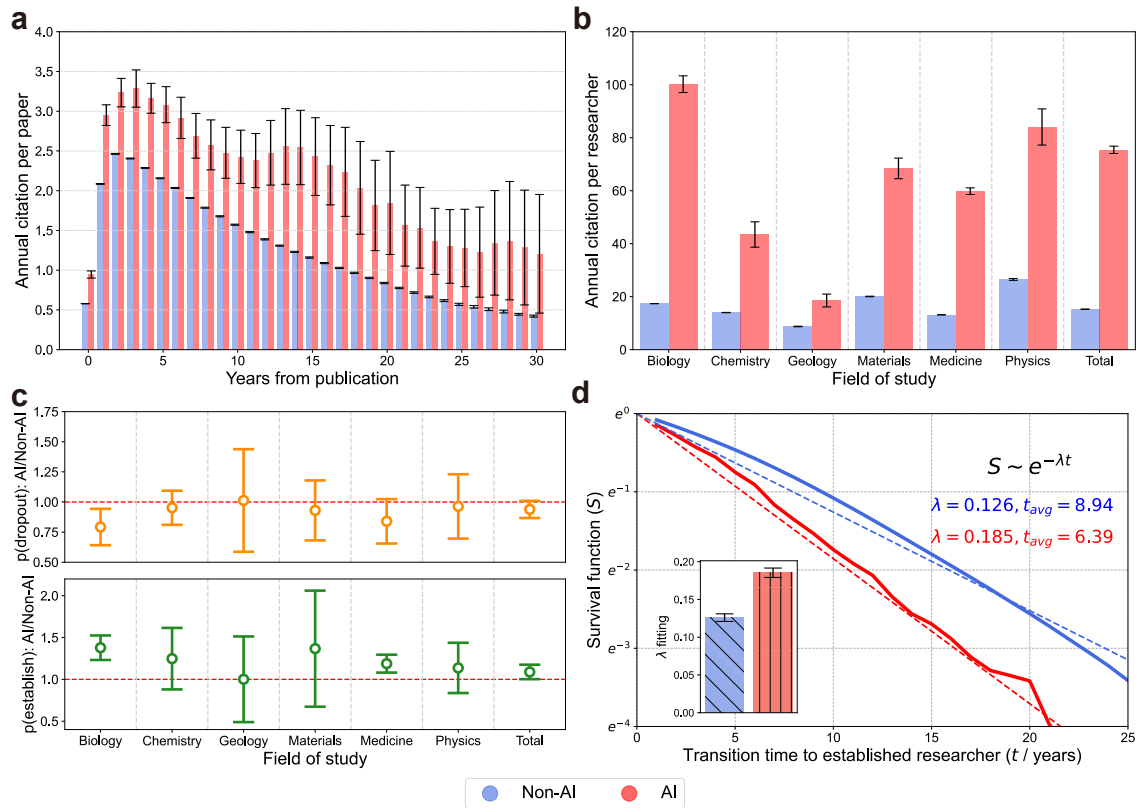


Figure S31. Replication of “AI enlarges the impact of papers and enhances the career of researchers” on the WOS dataset. (a) Annual citations after publication of AI and non-AI papers ($n = 16,706,988$). (b) Average annual citations of researchers adopting AI and their counterparts without AI ($p < 0.001, n = 3,620,795$). (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.

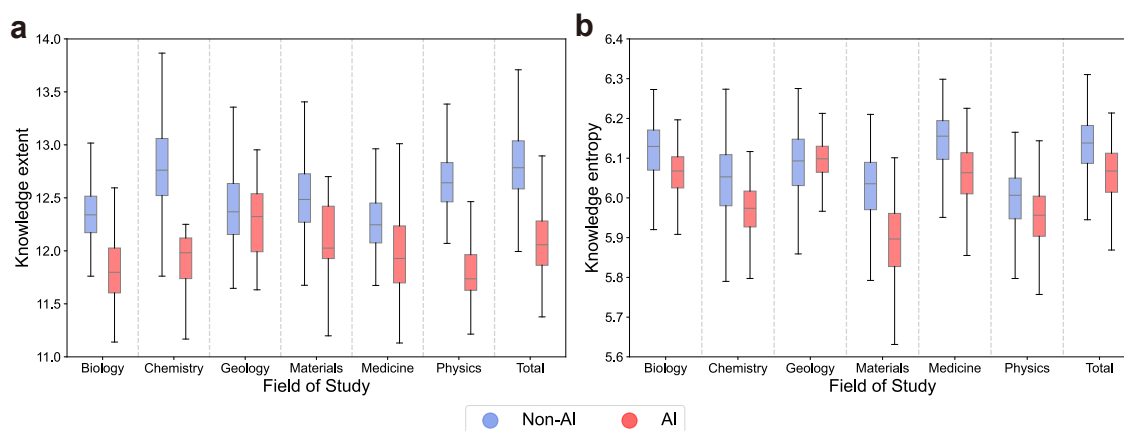


Figure S32. Replication of “AI usage is associated with a contraction in knowledge extent within and across scientific fields” on the WOS dataset. (a) Knowledge extent of AI and non-AI papers in each field ($n = 1,000$ samples in each field). (b) Knowledge entropy of AI and non-AI papers in each field ($n = 1,000$ 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.

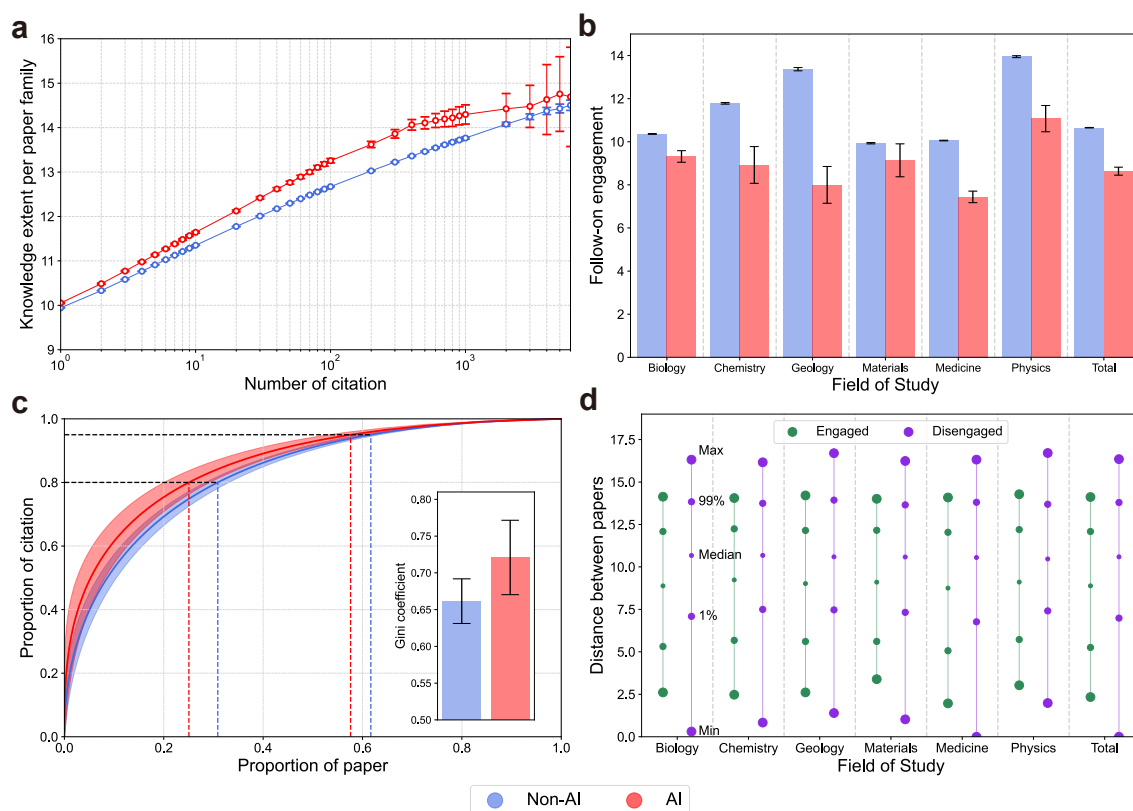


Figure S33. Replication of “reduced follow-on engagement and more overlapped works in AI research” on the WOS dataset. (a) Knowledge extent of individual AI and non-AI paper families ($n = 16,706,988$). (b) Engagement among papers that cite AI vs. non-AI papers ($p < 0.001, n = 15,048,254$). (c) The 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.

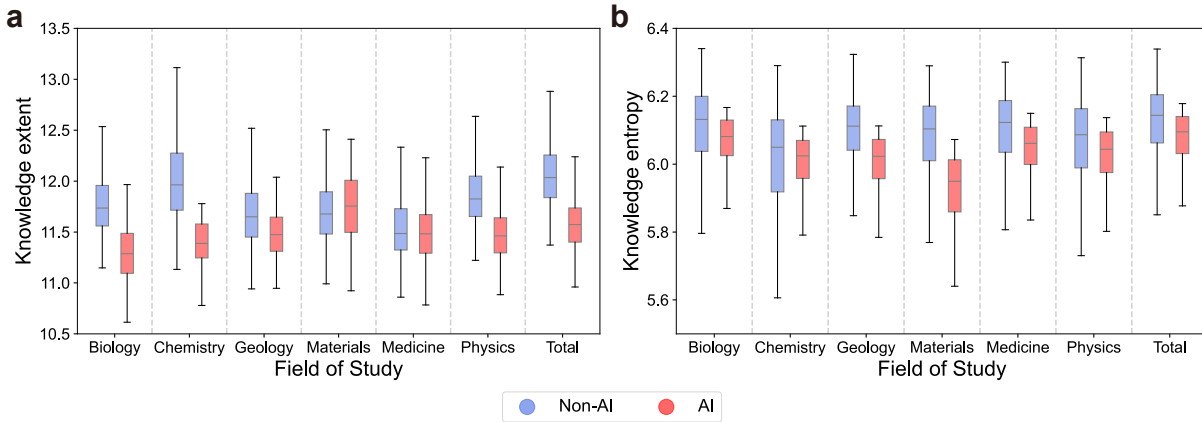


Figure S34. Replication of “AI usage is associated with a contraction in knowledge extent within and across scientific fields” in an embedding space after dimensional reduction using PCA. (a) Knowledge extent of AI and non-AI papers in each field ($n = 1,000$ samples in each field). **(b)** Knowledge 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.

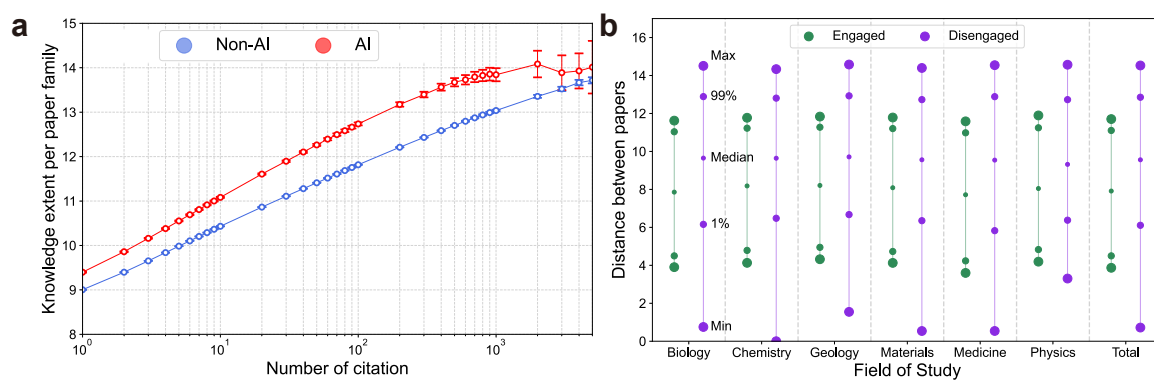


Figure S35. Replication of “reduced follow-on engagement and more overlapped works in AI research” in an embedding space after dimensional reduction using PCA. (a) Knowledge extent of individual AI and non-AI paper families ($n = 27,405,011$). 99% CIs are shown as error bars centred at the mean. (b) 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.

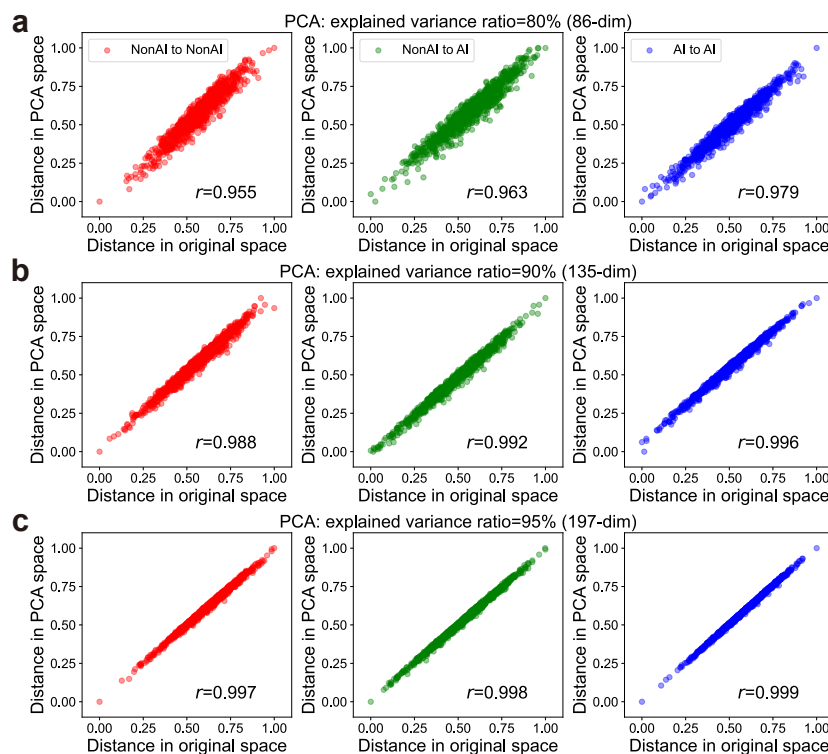


Figure S36. General sensitivity check on the distance calculations in high-dimensional paper embedding space. We reduce the original embeddings to 86, 135, and 197 dimensions with PCA, which correspond to the principal components that explain (a) 80%, (b) 90%, and (c) 95% of the total variance, respectively. 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.

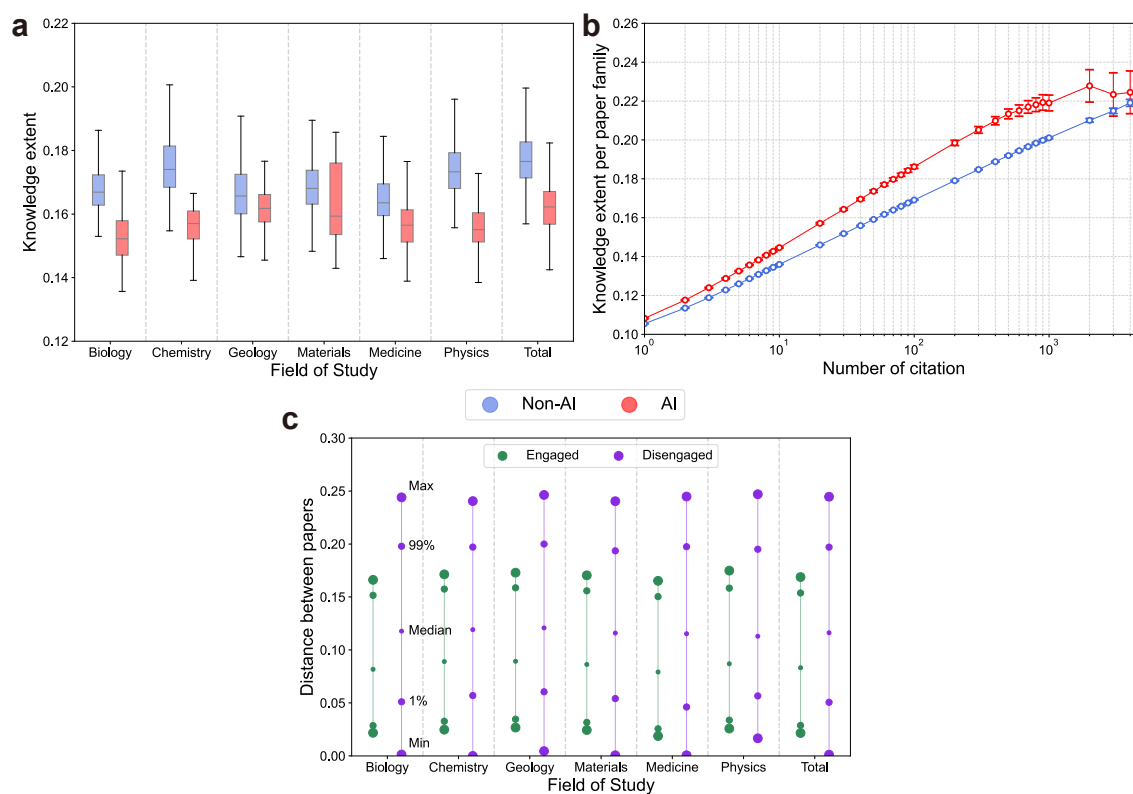


Figure S37. Replication of results related to high-dimensional distance calculations with Cosine distance. (a) Knowledge extent of AI and non-AI papers in each field ($n = 1,000$ samples in each field). Boxplots are centred at the median and bounded at the first and third quartile (Q1 and Q3), with 1.5 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.

6 Supplementary tables

Table S1. 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 S2. 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 S3. 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 S4. 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 S5. 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 S6. 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 S7. 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 S8. 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 S9. 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 S10. 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 S11. 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 S12. 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] Clarivate. <https://jcr.clarivate.com/jcr/home>, title = Journal Citation Reports. 2021.
- [2] Clarivate. *Web of Science*. <https://www.webofscience.com>. 2025.
- [3] 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.
- [4] 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.
- [5] Md Kamrul Hasan, Md Ashraf 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.
- [6] 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.
- [7] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. “Highly accurate protein structure prediction with AlphaFold”. In: *nature* 596.7873 (2021), pp. 583–589.
- [8] 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.
- [9] 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.
- [10] J Richard Landis and Gary G Koch. “The measurement of observer agreement for categorical data”. In: *biometrics* (1977), pp. 159–174.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] OpenAlex. *OpenAlex*. <https://openalex.org/>. 2025.
- [16] 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.
- [17] 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.
- [18] Per O Seglen. “The skewness of science”. In: *Journal of the American society for information science* 43.9 (1992), pp. 628–638.
- [19] 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.
- [20] 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.
- [21] 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.