nature human behaviour

Article

https://doi.org/10.1038/s41562-024-02020-5

Quantifying the use and potential benefits of artificial intelligence in scientific research

Received: 26 April 2023

Accepted: 12 September 2024

Published online: 11 October 2024

Check for updates

Jian Gao [1,2,3,4] & Dashun Wang [1,2,3,5]

The rapid advancement of artificial intelligence (AI) is poised to reshape almost every line of work. Despite enormous efforts devoted to understanding AI's economic impacts, we lack a systematic understanding of the benefits to scientific research associated with the use of AI. Here we develop a measurement framework to estimate the direct use of AI and associated benefits in science. We find that the use and benefits of AI appear widespread throughout the sciences, growing especially rapidly since 2015. However, there is a substantial gap between AI education and its application in research, highlighting a misalignment between AI expertise supply and demand. Our analysis also reveals demographic disparities, with disciplines with higher proportions of women or Black scientists reaping fewer benefits from AI, potentially exacerbating existing inequalities in science. These findings have implications for the equity and sustainability of the research enterprise, especially as the integration of AI with science continues to deepen.

The rapid advances in artificial intelligence (AI) may lead to massive value creation and capture across many facets of human society[1–5], creating a wealth of social and economic opportunities[6–8], and just as many challenges[9–17]. While extensive efforts have been devoted to understanding the impact of AI on the labour market and the economy[18–22], the impact of AI on the growing research enterprise remains unclear. Indeed, recent AI advances have shown promise to achieve and, in some cases, exceed expert-level performance across many economically valuable tasks[23–30]. As society prepares for the moment when AI may outperform or even replace human recruiters, bankers, doctors, lawyers, composers and drivers, an important question arises: what are the benefits associated with the use of AI in advancing scientific research across different disciplines and fields?

A better understanding of the extent of AI use in science and its potential benefits may not only help guide AI development, bridging AI innovations more closely with scientific research, but also hold implications for science and innovation policy. Understanding the links between AI use and scientific advances is both timely and important given AI's recent remarkable success in advancing research frontiers across several fields[31–44], from predicting the structure of proteins in biology[45–47] to designing new drug candidates in medicine[48–51], from discovering natural laws in physics[52–54] to solving complicated equations and discovering new conjectures in mathematics[55–57], from controlling nuclear fusion[58] to predicting new material properties[59–62], from designing taxation policy[63] to suggesting democratic social mechanisms[64], and many more[65–70]. These advances raise the possibility that as AI continues to improve in accuracy, robustness and reach[71–76], it may bring meaningful benefits to science, propelling scientific progress across a range of research areas while substantially augmenting researchers' innovation capabilities.

Yet, despite the rapid progress of AI and its broad applications in several domains, there is substantial scepticism about whether today's AI is capable or substantial enough to advance scientific research in every discipline and field. Indeed, most current AI applications belong to the category of 'narrow AI'[77–79], which tackles specifically defined problems, and hence may not be suitable to fulfil the broad range of tasks that scientific research demands[2,18,80]. Further, to the extent that AI may provide automated solutions to an existing problem, science is about not only solving well-defined problems but also spotting new frontiers and generating novel hypotheses[81]. These views paint a

[1]Center for Science of Science and Innovation, Northwestern University, Evanston, IL, USA. [2]Kellogg School of Management, Northwestern University, Evanston, IL, USA. [3]Ryan Institute on Complexity, Northwestern University, Evanston, IL, USA. [4]Faculty of Social Sciences, The University of Hong Kong, Hong Kong SAR, China. [5]McCormick School of Engineering, Northwestern University, Evanston, IL, USA. e-mail: dashun.wang@northwestern.edu

more nuanced picture of AI's applicability to advancing science, suggesting that AI may be better suited to perform some research tasks than others[2,10,82].

Building on the growing literature on the future of work[83–87] and the science of science[88–96], here we develop a quantitative framework for estimating the direct use of AI in science, as well as the potential benefits to science that are associated with the use of AI in scientific research (see Methods for details). Our primary dataset contains 74.6 million publications from 1960 to 2019 from the Microsoft Academic Graph (MAG) dataset[97], spanning 19 disciplines and 292 fields (see Supplementary Note 1.1 for details). We integrate this dataset with 7.1 million patents granted between 1976 and 2019 by the US Patent and Trademark Office (USPTO) (Supplementary Note 1.2). We then follow previous studies to identify AI publications and AI patents using a keyword-based approach (see Supplementary Notes 2.1 and 3.1 for details)[91,97,98], allowing us to measure AI use in scientific research and its potential associated benefits at two levels. First, we quantify the direct use of AI using an 'AI n-gram framework' (Fig. 1a), which estimates the relative frequency of the use of AI in a field (Supplementary Note 2.3). Specifically, we extract AI n-grams (bigrams and trigrams; for example, 'deep learning' and 'convolutional neural network') from both the titles and abstracts of AI publications and calculate the frequency of their occurrences to approximate AI advances[91,96]. We then repeat this n-gram measurement for publications in each field and year, allowing us to calculate the weighted frequency of AI n-grams appearing in a paper to approximate the direct use of AI in each field and year. Second, motivated by the future of work literature[82–84], we use an 'AI capability–field task framework' (Fig. 2a) to measure the alignment between AI capabilities and the tasks of a field (see Supplementary Note 3.3). In particular, we infer the capabilities of AI (that is, what AI can do) by extracting verb–noun pairs (for example, 'learn representation') from the titles of AI publications and AI patents using natural language processing (NLP) techniques and calculating their relative frequency[99–101]. Here, following previous work[82], we rely only on titles as they have a higher signal-to-noise ratio than abstracts. We then estimate the tasks of each field (that is, what a field does) by calculating the relative frequency of verb–noun pairs extracted from the titles of publications in each field and year. Calculating the overlap between the prevalent tasks in a field and the inferred AI capabilities allows us to approximate the potential benefits associated with AI use in each field and year (see Methods for more details).

Results

Widespread use of AI across the sciences

Overall, AI research presents a dynamically evolving landscape (Fig. 1b,c). While the frequency of certain dominant AI n-grams in 2019 (for example, 'machine learning', 'convolutional neural network', 'deep learning', 'deep neural network' and 'artificial intelligence') shows an overall upward trend (Fig. 1b), some AI n-grams emerged only recently (for example, 'generative adversarial network'), some rose to prominence after a long period of dormancy (for example, 'deep learning'), and some were popular a decade ago but have become less prevalent in recent years (for example, 'support vector machine'; Fig. 1c). Amidst this rapidly evolving AI research landscape, there has been a precipitous rise in the use of AI by many disciplines, as proxied by the mention of AI n-grams in the titles and abstracts of publications (Fig. 1d; see Supplementary Fig. 2 for details).

This increase in the use of AI by different disciplines raises an interesting question: how do the citations of papers that use AI compare to those of other papers in the same field? To answer this question, we define hit papers as those in the top 5% by total citations in the same field and year and calculate the likelihood that a paper is a hit paper. We find that for a majority of disciplines, papers that mention AI n-grams tend to be associated with a higher probability of being hit papers within their disciplines (the ratio of AI over non-AI regarding the hit rate of papers is larger than 1 in 18 out of 19 disciplines; mean ratio = 1.816, standard error (s.e.) = 0.138 and 95% confidence interval (CI) = (1.547, 2.086); Fig. 1e), and they also receive more citations from other disciplines compared with papers that do not mention AI n-grams (the ratio of AI over non-AI regarding the share of outside-field citations is larger than 1 in 11 out of 19 disciplines; mean ratio = 1.069, s.e. = 0.028 and 95% CI = (1.015, 1.124); Fig. 1f; see Supplementary Note 2.4 for details). This citation premium of papers that mention AI n-grams appears to be stronger in disciplines with a lower propensity to use AI (two-sided Pearson's correlation test for the negative relationship between the hit rate of papers that mention AI n-grams and the direct AI use score at the field level, Pearson's $r = -0.378$, $P < 0.001$ and 95% CI = (−0.472, −0.275); two-sided Pearson's correlation test for the negative relationship between the outside-citation-ratio for papers that mention AI n-grams and the direct AI use score at the field level, Pearson's $r = -0.215$, $P < 0.001$ and 95% CI = (−0.322, −0.102); see detailed correlation analysis in Supplementary Note 2.4 and Supplementary Fig. 4). These results suggest that disciplines that seem distant from AI may see substantial benefits from using AI to advance their research.

The dynamic AI landscape also prompts us to explore changes in the direct use of AI in scientific research over time. Specifically, we calculate the direct AI use score using the 'AI n-gram framework' for each discipline between 2000 and 2019 (see Supplementary Fig. 3 for the results for 1960–2019), extracting AI n-grams from the titles and abstracts of AI publications and calculating their frequency of occurrence in the publications in each discipline (see Methods and Supplementary Note 2 for details). We find that disciplines overall used more AI in their research over the past two decades (for example, the direct AI use score for computer science (CS) increased from 0.5% in 2000 to 1.3% in 2019; statistical test for the increasing trend, slope $b = 0.00031$, $P < 0.001$ and 95% CI = (0.00025, 0.00037); Fig. 1g, solid lines). This increase occurs not only in CS but also in a wide range of other disciplines (Fig. 1h), including, for example, physics, biology and economics. Moreover, this increase has not been linear; there has been a notably sharp increase in the use of AI since 2015 across many disciplines (Fig. 1g).

To better understand whether the recent rise in the direct use of AI on science is associated with changes in AI capabilities or field-specific shifts in research direction, we calculate an alternative measure for AI use scores by keeping AI capabilities fixed in 2015 and apply this alternative measure to each discipline and year between 2015 and 2019 (see Supplementary Note 4.2). More specifically, we use (1) AI n-grams extracted from AI publications before 2015 without changing either their terms or their frequency for 2015–2019, and (2) AI n-grams extracted from publications in each discipline and year during the period 2015–2019 (see dashed lines in Fig. 1g). We find that the new scores deviate substantially from the original scores (for example, the original score for CS is 37% larger than the new score; discipline level statistics, mean = 22%, s.e. = 1.9% and 95% CI = (18%, 26%); see solid lines in Fig. 1g and Supplementary Fig. 7), which indicates that across disciplines, sciences benefit more from cutting-edge AI advances. Overall, these results suggest that new AI capabilities are associated with the recent, sharp increase in the use of AI across disciplines (see Supplementary Note 4 for results related to AI's growing use in science).

Potential benefits associated with AI use

While explicit mentions of AI n-grams by publications signal the direct use of AI in research, AI may also be associated with other benefits in scientific research beyond these direct uses. In particular, the growing AI capabilities may help perform some core tasks that a research field demands. Here we build on the future of work literature, which suggests that AI capabilities and field tasks can be captured by verb–noun pairs (for example, 'learn representation')[82–84], prompting us to develop an 'AI capability–field task framework' to quantify the potential benefits associated with AI use in scientific research (Fig. 2a). We apply NLP

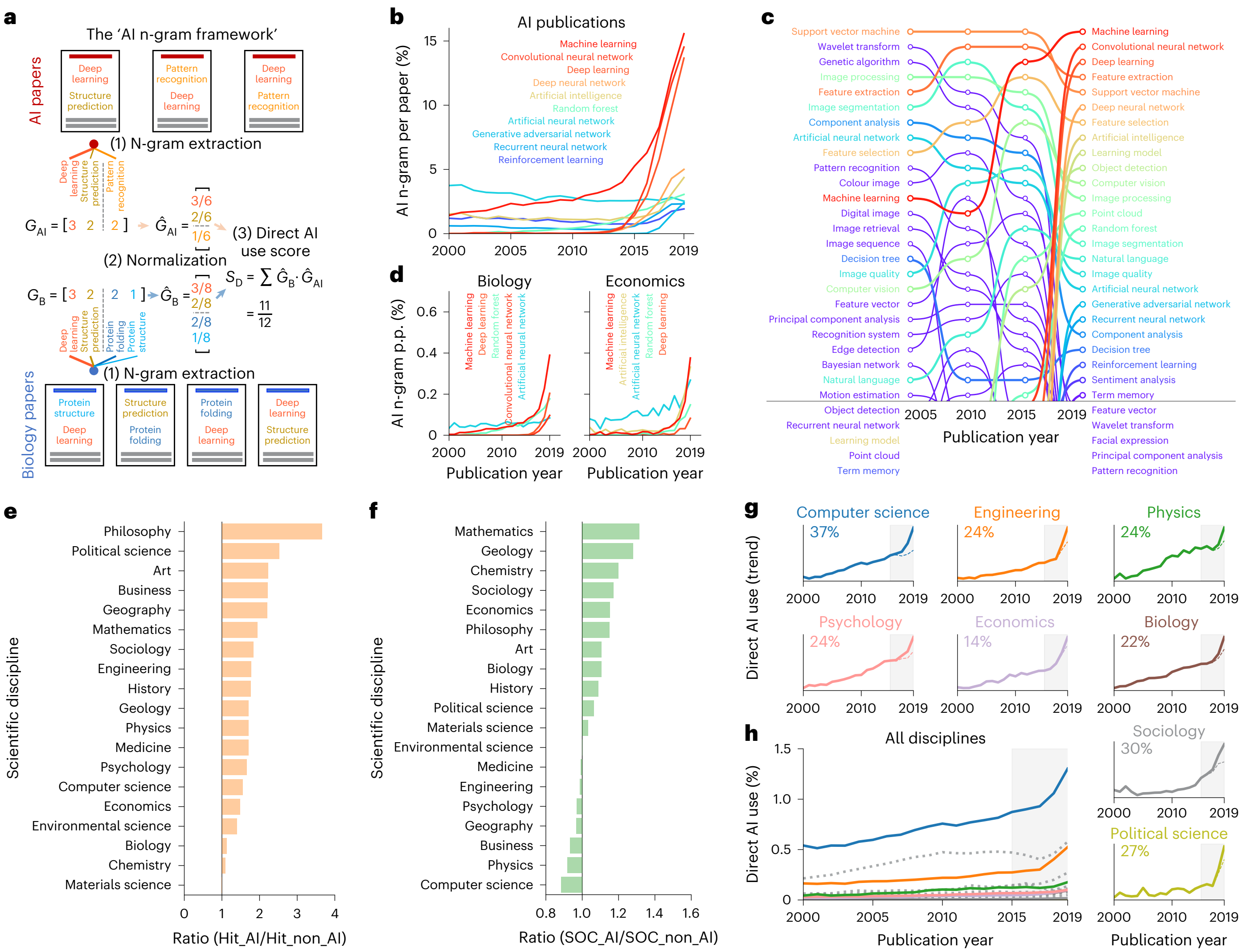

Fig. 1 | Measuring the direct use of AI in scientific research. a, The 'AI n-gram framework' for estimating the direct use of AI. First, AI-related publications are identified by the MAG five AI fields. Then, n-grams are extracted from the titles and abstracts of AI publications. Next, the frequency of AI n-grams per paper is calculated after normalization. Similarly, n-grams are extracted for publications in each field, and the frequency of n-grams per paper (n-gram p.p.) is calculated. Finally, a field's direct AI use score for a year is calculated by the dot product of the frequency of AI n-grams cumulated up to the year and the field's n-grams at the year. **b**, The frequency of ten AI n-grams in 2019 and the trend in use of these n-grams over the past two decades. **c**, Temporal changes in the rankings of the top 30 AI n-grams in 2019. AI n-grams are presented in rainbow colour order, according to their ranking in 2019. **d**, The frequency over the 2000–2019 period of the top five AI n-grams in biology and economics in 2019. **e**, The ratio of the hit rate of AI-using papers over non-AI-using papers. Here AI-using papers are identified as those that mention at least one AI n-gram, and the hit rate of papers (Hit) is defined as the likelihood that a paper is in the top 5% by total citations among papers in the same field and year. **f**, The ratio of the share of outside-field citations (SOC) for AI-using papers over that for non-AI-using papers. **g**, Temporal trends in the direct AI use scores of disciplines as shown by solid colour lines. The dashed colour line shows the score calculated using each discipline's yearly n-grams and AI n-grams fixed in 2015. The percentage change comparing the two scores in 2019 is shown. Each plot uses its y-axis scale to illustrate the relative change best. **h**, The direct AI use scores of disciplines using the same y-axis scale. Coloured lines correspond to disciplines in **g**, and grey dotted lines represent other disciplines (see Supplementary Fig. 7 for detailed results).

algorithms to extract verb–noun pairs from the titles of AI papers and AI patents to estimate AI capabilities[99–101] (Fig. 2b; see Methods for details and Supplementary Note 8.3 for validations of the approach). Applied systematically to all disciplines and fields, this framework allows us to estimate which subfields within a discipline may benefit most from AI. For instance, we find that the subfield of biology that features large potential AI benefits is 'biological system' (Fig. 2c, curve in red), as many of its basic tasks appear aligned with inferred AI capabilities (for example, 'extract feature', 'detect object' and 'improve prediction'). Interestingly, the 'biological system' field, ranked seventh among all non-CS fields by the potential AI benefit score (Fig. 2d), also happens to be the field for the AlphaFold paper[45], which *Science* called the 2021 Breakthrough of the Year[102].

While there are considerable differences in the direct use of AI across scientific disciplines (Fig. 1h), the differences in the potential AI benefit scores are relatively small across disciplines (Supplementary Fig. 7), suggesting the potentially widespread applicability of AI in science. We further study within-discipline heterogeneity by examining the percentiles of direct AI use scores and potential AI benefit scores for each discipline's subfields (see Supplementary Note 4.1 for details). We find that the two percentiles in each field are highly correlated with each other (Fig. 2e; two-sided Pearson's correlation test, Pearson's $r = 0.891$, $P < 0.001$ and 95% CI = (0.865, 0.913)). Moreover, the top three subfields within each discipline according to the two percentiles are entirely overlapped in almost half of the 19 disciplines (Fig. 2f), indicating that the two measurements are consistent in identifying fields most

https://doi.org/10.1038/s41562-024-02020-5

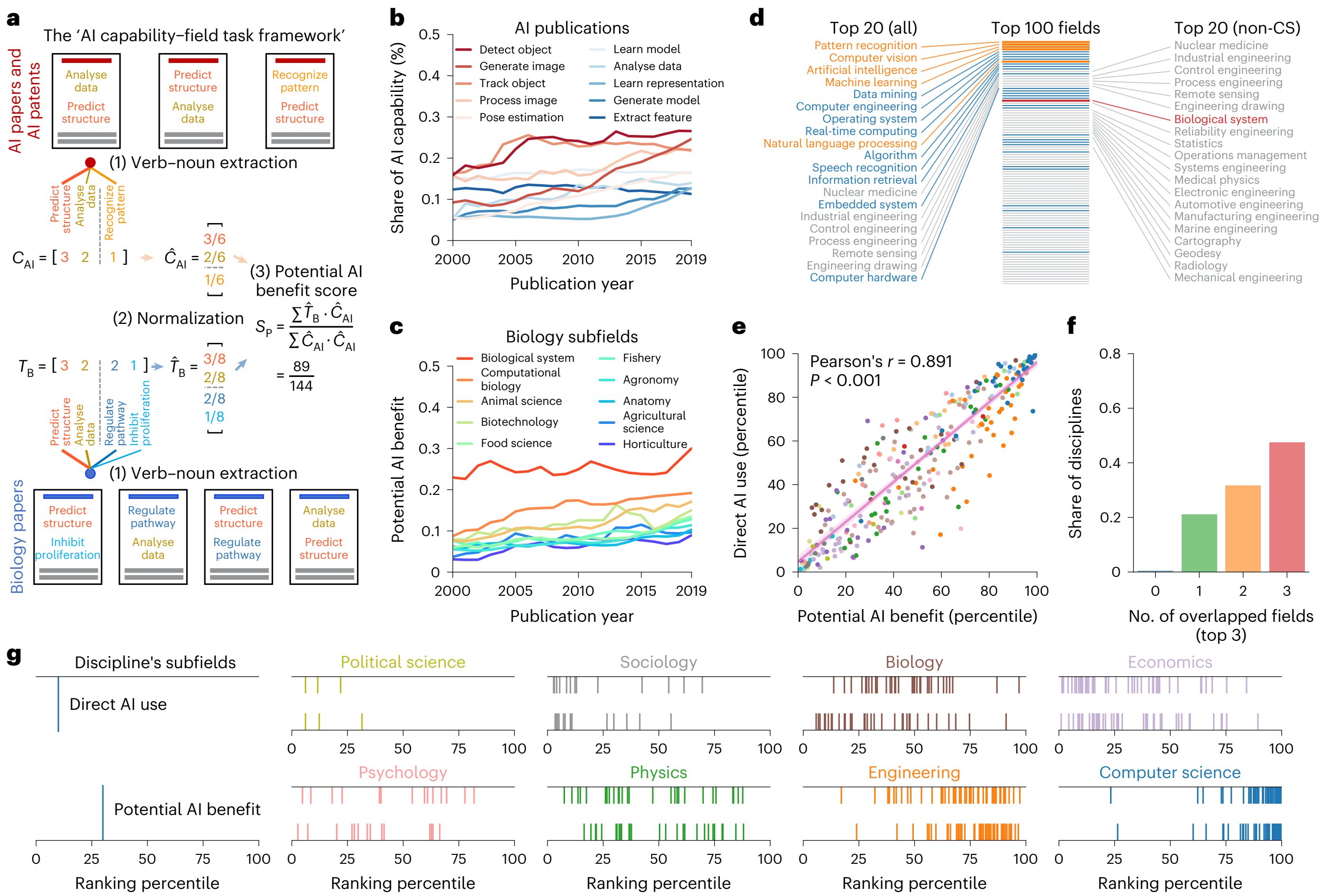

Fig. 2 | Measuring the potential benefits of AI and discipline heterogeneity. **a**, The 'AI capability–field task framework' for estimating the potential benefits of AI. First, AI capabilities are inferred by extracting verb–noun pairs from the titles of AI publications and AI patents using a dependency parsing algorithm. Then, field tasks are inferred from publications in each field using the same method. Next, the potential AI benefit score is determined by calculating the overlap between field tasks and AI capabilities, after discounting the frequency of commonly appearing verb–noun pairs. **b**, The frequency of ten AI verb–noun pairs in 2019 and their temporal trends over the past two decades. **c**, The top ten subfields of biology according to the potential AI benefit score in 2019 and their temporal trend over the past two decades. The 'biological system' field is consistently ranked first among all subfields. **d**, The top 100 fields according to the potential AI benefit score in 2019 are shown with colour-coded lines. The top 20 fields are listed on the left, with the five AI fields in orange, CS fields in dark blue and others in grey. The top 20 non-CS fields are listed on the right, with 'biological system' in red and ranked seventh. **e**, The strong correlation between direct AI use scores and potential AI benefit scores of research fields based on their percentiles. The correlation was determined using a two-sided Pearson's correlation test. Linear fit (centre line) with 95% confidence intervals (error bands) is shown. **f**, A large overlap between the top three subfields for each discipline according to direct AI use score and potential AI benefit score. Most disciplines exhibit three overlapped subfields (see Supplementary Table 1 for details). **g**, The substantial heterogeneity of AI's use and potential benefits within scientific disciplines. As illustrated by the legend on the left, each plot shows the percentiles of a discipline's subfields, where the percentiles based on the direct AI use score are in the upper row and those based on the potential AI benefit score are in the lower row. Eight disciplines are presented for illustration; all other disciplines are shown in Supplementary Fig. 8.

benefit from AI (see Supplementary Table 1 for the list of three subfields within each discipline that have the highest direct AI use score and the highest potential AI benefit score, respectively). Nevertheless, almost every discipline has some subfields with substantial AI use, which holds robust even for disciplines with low AI use overall, such as sociology and economics (Fig. 2g; see Supplementary Fig. 8 for the results for all disciplines). Taken together, these results suggest that the direct use of AI in research is pervasive across disciplines and fields, and its potential benefits to research may extend beyond its current uses in science.

Growing knowledge demands for AI

The rapidly expanding AI frontier and its increasing use in science may lead to growing demands for AI expertise from domain experts, raising the question of whether the current education and training on AI skills are commensurate with AI use. To answer this question, we analyse 4.2 million university course syllabi from the Open Syllabus Project (OSP) database[103] and estimate the level of AI education in each discipline (see Methods and Supplementary Note 5 for details). We find that, excluding the top three computational disciplines (that is, CS, mathematics and engineering), the correlation between the AI education level in a discipline and the use of AI in the discipline decreases, as well as its significance (Fig. 3a,b; two-sided Pearson's correlation test, Pearson's $r = 0.493$, $P = 0.074$ and 95% CI = (−0.051, 0.811) for the direct AI use score and Pearson's $r = 0.263$, $P = 0.363$ and 95% CI = (−0.310, 0.697) for the potential AI benefit score). The results suggest that the supply of AI talent and knowledge in most disciplines appears to be incommensurate with the benefits that these disciplines may extract from AI capabilities, highlighting a substantial AI use–AI training gap. This result is robust under some alternative measures of AI education levels (Supplementary Fig. 10).

To meet the growing knowledge demands on AI, domain experts may rely on cross-discipline collaborations to access AI capabilities.

https://doi.org/10.1038/s41562-024-02020-5

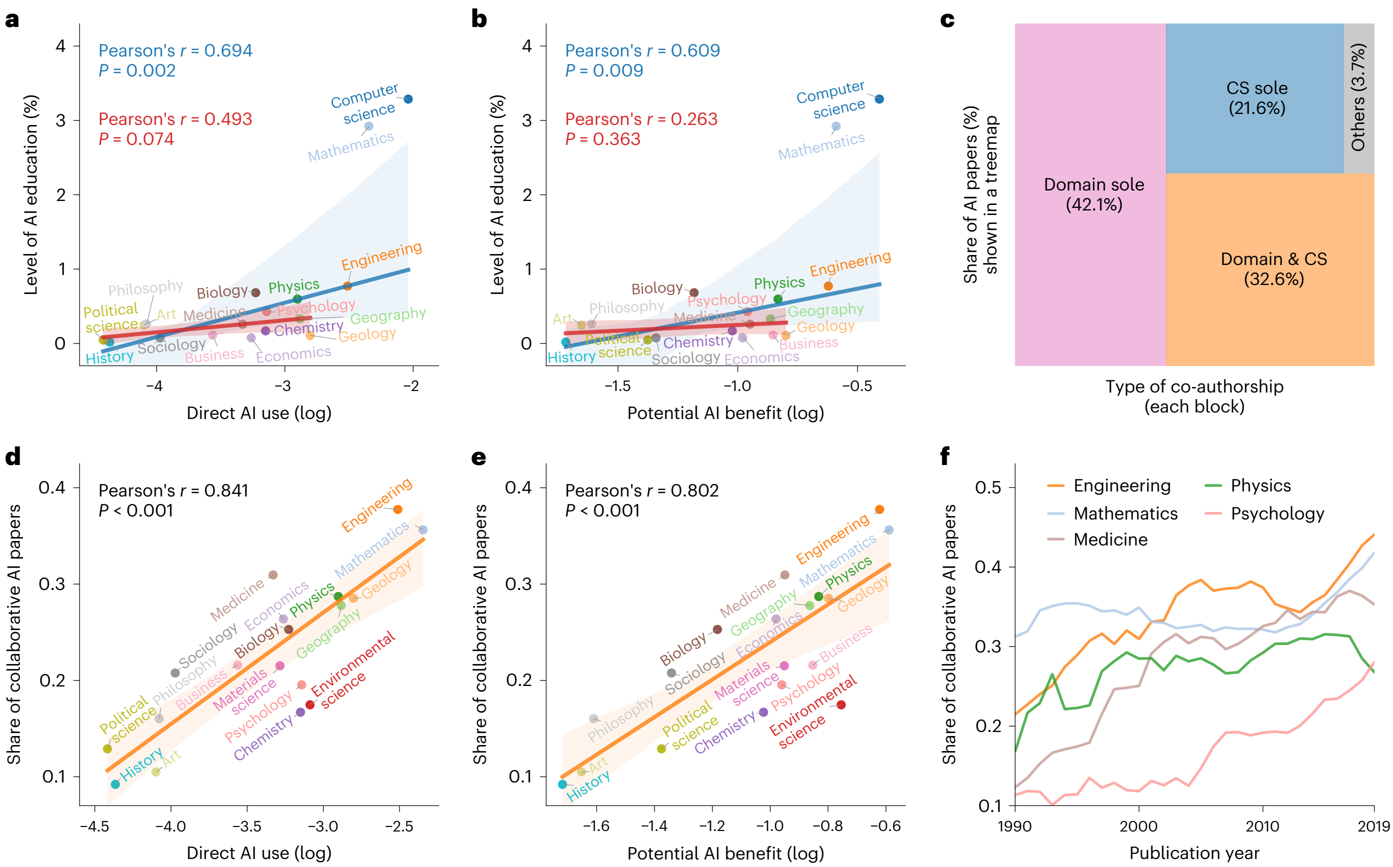

Fig. 3 | Misalignment between AI education and AI use and benefits, but growing knowledge demand for AI. a, The correlation between the direct AI use score and the AI education level that is estimated by the share of syllabus references to AI publications. Linear fits (centre lines) with 95% confidence intervals (error bands) are shown. The red line shows that the correlation loses significance when excluding the three disciplines with the largest AI use scores: CS, engineering and mathematics. **b**, The correlation between the potential AI benefit score and the AI education level. Linear fits (centre lines) with 95% confidence intervals are shown. **c**, The treemap chart shows the share of AI publications by four co-authorship types, where 'domain & CS' represents collaborative AI publications by domain experts and computer scientists, 'domain sole' represents AI publications by domain experts only, 'CS sole' represents AI publications by computer scientists only, and 'others' represents AI publications that are neither by domain experts nor by computer scientists. Here only AI publications in disciplines other than CS with at least two authors are considered. **d**, The positive correlation between the direct AI use score and the share of collaborative ('domain & CS') AI publications in each discipline. Linear fit (centre line) with 95% confidence intervals (error bands) is shown. **e**, The positive correlation between the potential AI benefit score and the share of collaborative ('domain & CS') AI publications in each discipline. Linear fit (centre line) with 95% confidence intervals (error bands) is shown. **f**, The share of collaborative AI publications ('domain & CS') in five disciplines across the period 1990–2019. Curves are smoothed by taking a 3-year moving average. Results for other disciplines are shown in Supplementary Fig. 12. All correlations were determined using a two-sided Pearson's correlation test.

We analyse collaboration patterns for AI publications in domains other than CS that are co-authored by domain experts and/or computer scientists (as a proxy for AI researchers; see Methods and Supplementary Notes 6.1 and 6.2 for detailed methods and alternative proxies for AI researchers). We find that, in aggregate, about 40% of AI publications are published by domain experts and about one-third are collaborative works (Fig. 3c). In disciplines where AI has more direct uses and potential benefits, we see a larger propensity for domain experts to collaborate with computer scientists (Fig. 3d,e; two-sided Pearson's correlation test, Pearson's $r = 0.841$, $P < 0.001$ and 95% CI = (0.616, 0.939) for direct AI use; Pearson's $r = 0.802$, $P < 0.001$ and 95% CI = (0.535, 0.923) for potential AI benefits), and the share of collaborative AI publications is increasing over time (for example, the share for engineering increased from 0.21 in 1990 to 0.44 in 2019; statistical test for the increasing trend, slope $b = 0.0057$, $P < 0.001$ and 95% CI = (0.0047, 0.0068); Fig. 3f; see Supplementary Fig. 12 for the results for other disciplines), suggesting that domain experts' reliance on AI expertise is growing. These results are robust when we use an alternative method of determining AI researchers (see Supplementary Note 6.3 for detailed methods and results). Taken together, these findings highlight the importance of teamwork and cross-domain collaborations amidst AI's potentially increasing use and benefits for scientific research and the narrowing of individual domain expertise across the sciences[104–106].

Demographic disparities

As the connection between AI and scientific research deepens, it is important to understand who benefits from AI, which has implications for the equity and sustainability of the research enterprise. Here we study the gender and racial/ethnic composition of each discipline and further examine potential differences in the distribution of AI use and benefits across demographic groups. Specifically, we leverage the de-identified Survey of Doctorate Recipients (SDR) data to solicit demographic information on US-trained doctoral scientists and engineers by the discipline of doctorate, sex and race/ethnicity. We then crosswalk the SDR disciplines of doctorate to the disciplines in the MAG data to estimate the share of women scientists and underrepresented minorities (URM) scientists in each discipline (see Methods and Supplementary Note 7 for details).

We find a negative correlation between the share of women scientists within each discipline and its AI scores for both direct use

https://doi.org/10.1038/s41562-024-02020-5

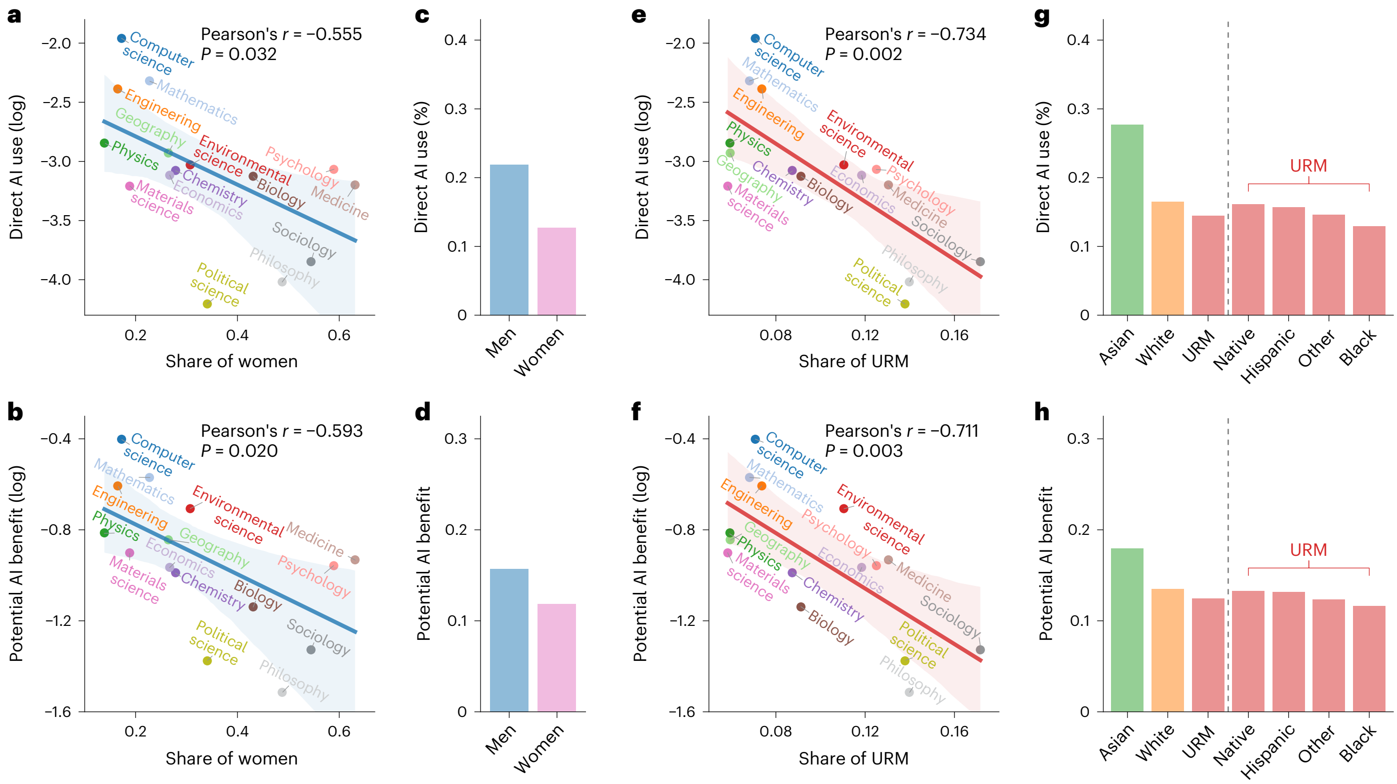

Fig. 4 | Gender and racial disparities in the use and benefits of AI across disciplines. a, The negative correlation between the direct AI use score and the share of women scientists in each discipline. Linear fit (centre line) with 95% confidence intervals (error bands) is shown. **b**, The negative correlation between the potential AI benefit score and the share of women scientists in each discipline. Linear fit (centre line) with 95% confidence intervals (error bands) is shown. **c**, The average direct AI use scores for women and men scientists. The average score for men/women is calculated by weighting the direct AI use score of each discipline by the share of men/women in the discipline. **d**, The average potential AI benefit scores for women and men scientists. **e**, The negative correlation between the direct AI use score and the share of URM scientists. The URM category includes 'African American or Black', 'American Indian or Alaska Native', 'Hispanic or Latino' and 'Native Hawaiians or other Pacific Islanders'. Linear fit (centre line) with 95% confidence intervals (error bands) is shown. **f**, The negative correlation between the potential AI benefit score and the share of URM scientists. Linear fit (centre line) with 95% confidence intervals (error bands) is shown. **g**, The average direct AI use score for each racial and ethnic group. The average score for each group is calculated by weighting the direct AI use score of each discipline by its share of the particular racial and ethnic group in the discipline. The average score for each racial and ethnic group under the URM category is shown separately on the right. **h**, The average potential AI benefit score for each racial and ethnic group. All correlations were determined using a two-sided Pearson's correlation test.

(two-sided Pearson's correlation test, Pearson's $r = -0.555$, $P = 0.032$ and 95% CI = (−0.831, −0.059); Fig. 4a) and potential benefits (two-sided Pearson's correlation test, Pearson's $r = -0.593$, $P = 0.020$ and 95% CI = (−0.848, −0.116); Fig. 4b). Aggregating the AI scores of all disciplines by their gender composition (see Methods for details), we find that women scientists tend to be associated with a smaller score and thus less benefits from AI (Fig. 4c,d). Studying the racial and ethnic composition across disciplines, we find another negative relationship between the share of URM scientists in each discipline and its AI scores, a pattern that is again robust for both direct use (two-sided Pearson's correlation test, Pearson's $r = -0.734$, $P = 0.002$ and 95% CI = (−0.906, −0.355); Fig. 4e) and potential benefits (two-sided Pearson's correlation test, Pearson's $r = -0.711$, $P = 0.003$ and 95% CI = (−0.897, −0.312); Fig. 4f). This pattern appears strong for Black scientists within the URM group, where the score of Black is 78% and 86% less than that of white for direct AI use and potential AI benefits, respectively (Fig. 4g,h). On average, women and URM researchers benefit less from AI. We further performed career-level analysis, looking at what happens when one starts to engage in AI research (see Supplementary Note 9 for details). We find that while the average hit rate of a researcher's papers tends to increase immediately after engaging in AI research, this citation premium is less concentrated among underrepresented groups, with women and URM researchers appearing to profit less from AI engagement compared with their counterparts. Together, these results suggest that while AI has the potential to bring benefits to all disciplines, the benefits may be distributed unequally across demographic groups. Hence, as the use of AI in science continues to grow, these unequal career effects may further amplify existing disparities in science[107,108].

Discussion

In this study, we develop a measurement framework to estimate the direct use and potential benefits of AI across a range of scientific disciplines and research fields. We find that scientific disciplines are increasingly using AI, as proxied by the mention of AI-related terms in publication titles and abstracts, with especially sharp growth in recent years. Publications that use AI tend to see a citation premium, as they are more likely to be cited both within and outside their disciplines. While there is substantial heterogeneity in the direct use and potential benefits of AI across different disciplines, almost every discipline includes some subfields that see great benefits from AI. For example, the medicine discipline as a whole is not ranked among the highest in terms of AI benefits, but some of its subfields (for example, 'nuclear medicine', 'optometry' and 'medical physics') show substantial AI benefits (Supplementary Fig. 8). Overall, these results suggest that the benefits that AI may bring to scientific research are widespread across a range of disciplines and fields, potentially extending beyond the current uses of AI in science.

A systematic understanding of the use and potential benefits of AI for scientific research may better inform science and education policy. Our research suggests that the use of AI in scientific disciplines has raced ahead across science, facilitated in part by cross-discipline collaborations, while the educational focus on AI to upskill future scientists within each discipline has lagged. This misalignment between AI use and AI education (that is, the AI use–AI training gap) has important implications for best practices in preparing next-generation scientists to fully leverage AI advances. While these analyses are correlational by nature, they support the hypothesis that collaboration between domain experts and AI researchers may represent an important way to facilitate the use of AI across science. They also suggest a further benefit of increasing AI training across disciplines, which would likely help the disciplines to develop domain-specific AI expertise, allowing them to enjoy greater and timelier benefits from AI advances.

It is also important to recognize that as AI becomes increasingly capable of performing research tasks, it may have an unequal impact on the research workforce. There are long-standing concerns about demographic disparities in science[109–111]. Our results suggest that the groups that have been historically underrepresented in science are also the groups that may benefit less from AI in scientific research. These results are somewhat expected, given that gender disparities tend to correlate with technical fields, which tend to be dominated by men[112,113]. Nevertheless, our analysis highlights that as AI plays more important roles in accelerating science, it may exacerbate existing disparities in science, with implications for building a diverse, equitable and inclusive research workforce. It thus underscores the importance of expanding the AI-related professoriate by broadening participation and opportunities in AI research and increasing funding and educational programmes targeted towards women and underrepresented groups in AI-related fields[114].

While this study takes an initial step towards quantifying the use and potential benefits of AI for scientific research, it has several limitations that are important to consider when interpreting the results. First, our analyses build on the future of work literature and rely on publication and patent data. Given its multidimensional nature, however, the potential benefits of AI for science may go beyond the advantages that can be estimated from such datasets. These frameworks, in fact, may underestimate the full range of benefits that AI may bring to scientific research. AI may, for example, optimize the research process by powering new tools and systems that improve the efficiency of doing science, including improving access to information, reducing the knowledge burden, guiding human intuition, automating routine research tasks and more[115,116]. Second, AI research evolves rapidly, suggesting the need for continuous monitoring and updates to the estimates of its benefits to science. As our datasets trace publications and patents to the end of 2019, they cannot capture newer developments, such as the recent rise of foundation models in AI research[117–120]. Given that these foundation models, such as large language models, can be adapted to a wide range of downstream tasks through fine-tuning, they may play an important role in augmenting research. Third, as a general-purpose technology[71,121], AI may generate downstream spillover effects, with indirect impacts on various domains. For example, by discovering faster matrix multiplication algorithms[122], AI may have indirect impact on disciplines that would benefit from such advances. Fourth, although the direct mention of AI n-grams in publication titles and abstracts is suggestive of the use of AI in research, the same n-gram may have different meanings in different contexts. Also, the same AI capability may bring different benefits to different fields, amidst alternative ways to define AI terms[63] (see Supplementary Note 8.1 for details), suggesting fruitful future directions to further improve our frameworks for understanding AI capabilities and their uses in scientific research. Lastly, as AI's capabilities and its benefits to science continue to grow, it will become ever more crucial to understand the impact of AI on fairness and equity in research[16,123]. Equally important is to understand how AI may introduce potential biases or otherwise create unintended consequences in the genesis of scientific knowledge, especially given the 'black box' nature of many leading AI tools[124–127].

Overall, these findings based on large-scale quantitative analyses may prove useful to the AI research community, helping us better understand the AI capabilities that may be most fruitful for scientific research. At the same time, the misalignment between the level of AI education and its use in research suggests that collaborations between domain experts and AI researchers may be especially productive, bridging deep domain expertise and new AI advances. Given that tomorrow's technological developments often begin upstream from basic scientific research[128–130], a more robust understanding of the impact of AI on science may further inform a range of important policy considerations for the future of education, research and innovation[2–4].

Methods

The study protocol was reviewed by the Institutional Review Board (IRB) of Northwestern University. The study was determined as Not Human Research and exempt from formal ethics review (IRB no. STU00221828).

Data sources

To estimate the use and potential benefits of AI for science, we use a variety of datasets that include information regarding scientific publications, patents, course syllabi and the demographics of researchers (see Supplementary Note 1 for details). We introduce two primary datasets. (1) We use the MAG database for publication data. We collect information on 74.6 million publications between 1960 and 2019 of various types ('journal', 'conference', 'book' or 'book chapter'). These publications are categorized into 19 disciplines (for example, 'computer science') and 292 fields (for example, 'machine learning') under the MAG 'field of study' taxonomy, in which one discipline contains several child fields (see Supplementary Note 1.1 for details). For each publication, we collect the title, abstract, year, discipline and field information. (2) We use PatentsView for patent data. We collect information on 7.1 million patents granted between 1976 and 2019 from PatentsView, a data platform based on bulk data from the USPTO. Each patent is associated with a list of patent classification codes and keywords. Using these codes and keywords, we identify AI-related patents (see Supplementary Note 1.2 for details). Together, the MAG publication data and USPTO patent data allow us to estimate the direct use and potential benefits of AI for each discipline and field.

We supplement the analysis with two more datasets to examine the alignment of AI use and benefits with the level of AI education and to study the gender, racial and ethnic composition in science. (1) We use syllabus data that are sourced from the OSP, the world's first large-scale database of university course syllabus documents. Our syllabus dataset contains 4.2 million English-language syllabi published between 2000 and 2018 (see Supplementary Note 1.3 for details). Each syllabus is associated with a list of Classification of Instructional Programs (CIP) codes representing its academic fields and a list of referenced publications[103]. We manually crosswalk CIP codes to MAG fields, and we link syllabus references to MAG publications (see Supplementary Note 5.1 for details). (2) We use the SDR for de-identified demographic data regarding individuals with a US research doctoral degree in a science, engineering or health field. We use the 2017 SDR data on scientists and engineers, including the discipline of their doctorate, their sex, and their race and ethnicity. We manually crosswalk the SDR doctorate disciplines to the MAG disciplines (see Supplementary Note 7.1 for details).

Calculation of AI use and benefit scores

We estimate the direct use of AI by implementing the 'AI n-gram framework'. Specifically, following previous studies[91], we identify AI-related publications using the five MAG field categories ('machine learning', 'artificial intelligence', 'computer vision', 'natural language processing' and 'pattern recognition'). Because MAG used a topic modelling approach to label each paper's field categories, the AI-related publications identified here go beyond the explicit mention of these five keywords. Identifying AI research from large-scale publication databases remains a challenging task, but the simple approach that we use balances precision and recall in determining AI publications (see Supplementary Note 8.1 for details). There are also other ways to identify AI research (see Supplementary Note 2.1 for details), and our results are robust under these alternative approaches (see Supplementary Note 8 for details).

From the titles and abstracts of AI publications, we extract n-grams (bigrams and trigrams; for example, 'deep learning' and 'deep neural network') and normalize them by standardizing words. From these normalized n-grams, we filter AI n-grams using a list of topics under the five AI field categories in the MAG 'field of study' taxonomy. This taxonomy is constructed primarily based on Wikipedia topics (see Supplementary Note 1.1 for details). We calculate the frequency of AI n-grams per paper to approximate cumulative AI advances. Formally, the AI n-gram frequency vector at year t is $\hat{G}^t_{\mathrm{AI}} = G^t_{\mathrm{AI}}/N^t_{\mathrm{AI}}$, where G^t_{AI} is the vector that summarizes the counts of AI n-grams extracted from AI publications before year t, and N^t_{AI} is the number of AI publications. We repeat this process for publications in each field to extract n-grams (both AI n-grams and non-AI n-grams), and we calculate their frequency to approximate current field development. For example, the biology n-gram frequency vector at year t is $\hat{G}^t_{\mathrm{B}} = G^t_{\mathrm{B}}/N^t_{\mathrm{B}}$, where G^t_{B} is the vector that summarizes the count of n-grams extracted from biology publications at year t, and N^t_{B} is the number of these biology publications. The coordinate of the same AI n-gram in the biology frequency vector and the AI frequency vector is the same. In other words, each coordinate of $\hat{G}^t_{\mathrm{B}}$ represents one n-gram, where AI n-grams have the same coordinates as those in $\hat{G}^t_{\mathrm{AI}}$. Finally, we calculate the direct AI use score for biology at year t based on the frequency of AI n-grams:

$$S^t_{\mathrm{D}} = \sum \hat{G}^t_{\mathrm{B}} \cdot \hat{G}^t_{\mathrm{AI}} \tag{1}$$

where the symbol '$\sum \cdot$' represents the dot product of the biology frequency vector and the AI frequency vector of the same AI n-grams. In this calculation, only their common n-grams in these two vectors are considered, and the same n-gram has the same coordinate in these two vectors. There are other ways to calculate the direct AI use score, and our results are largely robust under some alternative calculations (see Supplementary Note 8.2 for details). A larger direct AI use score indicates that AI is being used more extensively in the field.

We estimate the potential benefits of AI by implementing the 'AI capability–field task framework', which is built on the future of work literature[82–84]. It assumes that research fields may potentially benefit from AI if their basic tasks are aligned with AI capabilities (see Supplementary Note 3 for details on the underlying assumptions). We predict the capabilities of AI (that is, what AI can do) by extracting verb–noun pairs (for example, 'learn representation') from the titles of AI publications and AI patents using a dependency parsing algorithm developed in NLP[99–101] (see Supplementary Note 3.2 for details). Here, following the previous work[82], we only use titles because they have a higher signal-to-noise ratio than the other text fields. After normalizing verb–noun pairs through standardization, we calculate their relative frequency to approximate AI capabilities. Specifically, the AI capability frequency vector for AI papers at year t is $\mathrm{Paper}(\hat{C}^t_{\mathrm{AI}}) = C^t_{\mathrm{AI}}/\sum C^t_{\mathrm{AI}}$, where C^t_{AI} is the vector that summarizes the counts of verb–noun pairs extracted from AI publications before year t. We repeat this process for AI patents and calculate the vector $\mathrm{Patent}(\hat{C}^t_{\mathrm{AI}})$. By taking an average of common verb–noun pairs in these two vectors, we calculate the AI capability frequency vector $\hat{C}^t_{\mathrm{AI}}$ to approximate cumulative AI capabilities at year t:

$$\hat{C}^t_{\mathrm{AI}} = \left[\mathrm{Paper}\left(\hat{C}^t_{\mathrm{AI}}\right) + \mathrm{Patent}\left(\hat{C}^t_{\mathrm{AI}}\right)\right]/2 \tag{2}$$

where the symbol '+' represents summing up the frequencies of the same verb–noun pair in the two vectors. Analogously, we predict the basic tasks of a research field (that is, what the field does) by extracting verb–noun pairs from the titles of publications in the field. Taking the biology field as an example, the field task frequency vector at year t is given by $\hat{T}^t_{\mathrm{B}} = T^t_{\mathrm{B}}/\sum T^t_{\mathrm{B}}$, where T^t_{B} is the vector that summarizes the counts of verb–noun pairs extracted from biology publications at year t. In the calculation, we apply the term frequency-inverse document frequency (tf-idf) to discount the weight of commonly appearing verb–noun pairs in both AI capability and field task vectors (see Supplementary Note 3.3 for details). Finally, we calculate the potential AI benefit score of biology at year t based on the alignment between its tasks and AI capabilities:

$$S^t_{\mathrm{P}} = \frac{\sum \hat{T}^t_{\mathrm{B}} \cdot \hat{C}^t_{\mathrm{AI}}}{\sum \hat{C}^t_{\mathrm{AI}} \cdot \hat{C}^t_{\mathrm{AI}}} \tag{3}$$

where the symbol '$\sum \cdot$' represents the dot product of the AI vector and the biology vector, and the denominator is applied to normalize the score for comparison across time. In the calculation, only common verb–noun pairs in the AI vector and the biology vector are considered, and the same verb–noun pair has the same coordinate in the two vectors. A larger potential AI benefit score means that AI is predicted to have greater benefits for the field. There are other ways to calculate the potential AI benefit score, and our findings are largely robust under some alternative calculations (see Supplementary Note 8.3 for details).

Estimation of AI education levels

We measure the level of AI education in each discipline by leveraging OSP syllabus data and MAG publication data. This measure assumes that a discipline has a higher AI education level if a larger fraction of publications referenced by syllabi in the discipline are AI publications. The OSP dataset categorizes course syllabi by educational fields and provides a link from syllabi to their referenced publications. As syllabi with more references more likely correspond to graduate-level or research-oriented courses, we only use syllabi with at least five references and those in the recent period 2014–2018. First, we crosswalk the taxonomies of educational disciplines and academic disciplines by mapping OSP fields to MAG disciplines, and we match syllabi-referenced publications to MAG publications using the digital object identifier (DOI), title and year (see Supplementary Note 5.1 for details). From these publications, we identify AI publications based on the MAG five AI field categories (see Supplementary Note 2.1 for details). We then estimate a discipline's AI education level by calculating the fraction of citations in the discipline's syllabi that are citations to AI publications (see Supplementary Note 5.2 for details). As robustness checks, we also use syllabi with at least ten references, calculate an alternative measure for the level of AI education defined as the fraction of a discipline's syllabi that cites at least one AI publication, and repeat the analysis for different time periods between 2000 and 2018 (see Supplementary Note 5.2 for detailed methods and additional results).

Calculation of cross-discipline collaborations on AI

We estimate the level of cross-discipline collaborations on AI research between domain experts and AI researchers using AI publications in each discipline other than CS. Specifically, we first assign a primary discipline to each researcher based on the discipline in which they published most frequently in the period 1960–2019 and treat authors whose primary discipline is CS as AI researchers (see Supplementary Note 6 for more details). We then categorize each AI publication in a discipline into one of the four co-authorship types based on its authors' primary disciplines: (1) 'domain & CS', which involves both domain experts and computer scientists; (2) 'domain sole', which involves only domain experts; (3) 'CS sole', which involves only computer scientists; and (4) 'others', which involves neither domain experts nor computer scientists. Next, we calculate the share of collaborative AI publications (that is, those in the 'domain & CS' type) for each discipline (see Supplementary Note 6.1 for detailed methods). Here the calculation only considers AI publications with at least two authors that were published in the period 1980–2019. As robustness checks, we also use an alternative approach to identify primary AI researchers (see Supplementary Note 6.3 for detailed methods and results).

Reporting summary

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

Data availability

The MAG data are available at https://doi.org/10.5281/zenodo.6511057 (ref. 131) and ref. 132. The USPTO patent data are available at https://patentsview.org. The OSP dataset is available from the paper at https://doi.org/10.1073/pnas.1804247115. The SDR data are available at https://www.nsf.gov/statistics/srvydoctoratework, and the datasets used in this study are de-identified, containing only summary statistics for each discipline. The data met the assumption of tests in the analysis. The data necessary to reproduce all main plots in this paper are available at https://kellogg-cssi.github.io/ai4science.

Code availability

Data are linked and analysed with customized code in Python 3 using standard software packages within these programmes, including pandas 1.3.5, numpy 1.21.5, scipy 1.7.3, matplotlib 3.5.1, seaborn 0.11.2, spacy 3.7.2, nomquamgender 0.1.0, demographicx 0.0.1 and others. The code necessary to reproduce all main plots and statistical analyses is available at https://kellogg-cssi.github.io/ai4science.

References

1. Herbert, A. S. *The Sciences of the Artificial* (MIT Press, 1969).
2. Brynjolfsson, E. & Mitchell, T. What can machine learning do? Workforce implications. *Science* **358**, 1530–1534 (2017).
3. Agrawal, A., Gans, J. & Goldfarb, A. *The Economics of Artificial Intelligence: An Agenda* (Univ. Chicago Press, 2019).
4. Autor, D., Mindell, D. A. & Reynolds, E. B. *The Work of the Future: Shaping Technology and Institutions* (MIT Task Force, 2019).
5. Acemoglu, D., Autor, D., Hazell, J. & Restrepo, P. Artificial intelligence and jobs: evidence from online vacancies. *J. Labor Econ.* **40**, S293–S340 (2022).
6. Aghion, P., Jones, B. F. & Jones, C. I. in *The Economics of Artificial Intelligence: An Agenda* Ch. 9, 237–290 (Univ. Chicago Press, 2019).
7. Cockburn, I. M., Henderson, R. & Stern, S. in *The Economics of Artificial Intelligence: An Agenda* Ch. 4, 115–148 (Univ. Chicago Press, 2019).
8. Tomasev, N. et al. AI for social good: unlocking the opportunity for positive impact. *Nat. Commun.* **11**, 2468 (2020).
9. Dwivedi, Y. K. et al. Artificial intelligence (AI): multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *Int. J. Inf. Manag.* **57**, 101994 (2021).
10. Frey, C. B. & Osborne, M. A. The future of employment: how susceptible are jobs to computerisation? *Technol. Forecast. Soc. Change* **114**, 254–280 (2017).
11. Acemoglu, D. & Restrepo, P. The race between man and machine: implications of technology for growth, factor shares, and employment. *Am. Econ. Rev.* **108**, 1488–1542 (2018).
12. Khan, H. N., Hounshell, D. A. & Fuchs, E. R. H. Science and research policy at the end of Moore's law. *Nat. Electron.* **1**, 14–21 (2018).
13. Iansiti, M. & Lakhani, K. R. *Competing in the Age of AI: Strategy and Leadership When Algorithms and Networks Run the World* (Harvard Business Press, 2020).
14. Eshraghian, J. K. Human ownership of artificial creativity. *Nat. Mach. Intell.* **2**, 157–160 (2020).
15. Marcus, G. & Davis, E. *Rebooting AI: Building Artificial Intelligence We Can Trust* (Pantheon Books, 2019).
16. Liang, W. et al. Advances, challenges and opportunities in creating data for trustworthy AI. *Nat. Mach. Intell.* **4**, 669–677 (2022).
17. Bengio, Y. et al. Managing extreme AI risks amid rapid progress. *Science* **384**, 842–845 (2024).
18. Frank, M. R. et al. Toward understanding the impact of artificial intelligence on labor. *Proc. Natl Acad. Sci. USA* **116**, 6531–6539 (2019).
19. Agrawal, A., Gans, J. S. & Goldfarb, A. Artificial intelligence: the ambiguous labor market impact of automating prediction. *J. Econ. Perspect.* **33**, 31–50 (2019).
20. Koebis, N., Starke, C. & Rahwan, I. The promise and perils of using artificial intelligence to fight corruption. *Nat. Mach. Intell.* **4**, 418–424 (2022).
21. Brynjolfsson, E., Li, D. & Raymond, L. R. *Generative AI at Work* NBER Working Paper No. w31161 (National Bureau of Economic Research, 2023).
22. Noy, S. & Zhang, W. Experimental evidence on the productivity effects of generative artificial intelligence. *Science* **381**, 187–192 (2023).
23. Silver, D. et al. Mastering the game of Go with deep neural networks and tree search. *Nature* **529**, 484–489 (2016).
24. Geirhos, R. et al. Generalisation in humans and deep neural networks. In *Proc. Advances in Neural Information Processing Systems* 7538–7550 (MIT Press, 2018).
25. Grace, K., Salvatier, J., Dafoe, A., Zhang, B. & Evans, O. When will AI exceed human performance? Evidence from AI experts. *J. Artif. Intell. Res.* **62**, 729–754 (2018).
26. Liu, X. et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digit. Health* **1**, e271–e297 (2019).
27. Ishowo-Oloko, F. et al. Behavioural evidence for a transparency-efficiency tradeoff in human-machine cooperation. *Nat. Mach. Intell.* **1**, 517–521 (2019).
28. Yang, Y., Youyou, W. & Uzzi, B. Estimating the deep replicability of scientific findings using human and artificial intelligence. *Proc. Natl Acad. Sci. USA* **117**, 10762–10768 (2020).
29. Wurman, P. R. et al. Outracing champion Gran Turismo drivers with deep reinforcement learning. *Nature* **602**, 223–228 (2022).
30. Maslej, N. et al. *The AI Index 2024 Annual Report* (AI Index Steering Committee, Institute for Human-Centered AI, Stanford Univ., 2024).
31. Gil, Y., Greaves, M., Hendler, J. & Hirsh, H. Amplify scientific discovery with artificial intelligence. *Science* **346**, 171–172 (2014).
32. Wang, H. et al. Scientific discovery in the age of artificial intelligence. *Nature* **620**, 47–60 (2023).
33. Carleo, G. et al. Machine learning and the physical sciences. *Rev. Mod. Phys.* **91**, 045002 (2019).
34. Rahwan, I. et al. Machine behaviour. *Nature* **568**, 477–486 (2019).
35. Jimenez-Luna, J., Grisoni, F. & Schneider, G. Drug discovery with explainable artificial intelligence. *Nat. Mach. Intell.* **2**, 573–584 (2020).
36. Xu, Y. et al. Artificial intelligence: a powerful paradigm for scientific research. *Innovation* **2**, 100179 (2021).
37. Davies, A. et al. Advancing mathematics by guiding human intuition with AI. *Nature* **600**, 70–74 (2021).

38. Peng, H., Ke, Q., Budak, C., Romero, D. M. & Ahn, Y.-Y. Neural embeddings of scholarly periodicals reveal complex disciplinary organizations. *Sci. Adv.* **7**, eabb9004 (2021).
39. Krenn, M. et al. On scientific understanding with artificial intelligence. *Nat. Rev. Phys.* **4**, 761–769 (2022).
40. Belikov, A. V., Rzhetsky, A. & Evans, J. A. Prediction of robust scientific facts from literature. *Nat. Mach. Intell.* **4**, 445–454 (2022).
41. Grossmann, I. et al. AI and the transformation of social science research. *Science* **380**, 1108–1109 (2023).
42. Groh, M. et al. Deep learning-aided decision support for diagnosis of skin disease across skin tones. *Nat. Med.* **30**, 573–583 (2024).
43. Bail, C. A. Can generative AI improve social science? *Proc. Natl Acad. Sci. USA* **121**, e2314021121 (2024).
44. Alvarez, A. et al. Science communication with generative AI. *Nat. Hum. Behav.* **8**, 625–627 (2024).
45. Senior, A. W. et al. Improved protein structure prediction using potentials from deep learning. *Nature* **577**, 706–710 (2020).
46. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
47. Baek, M. et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).
48. Popova, M., Isayev, O. & Tropsha, A. Deep reinforcement learning for de novo drug design. *Sci. Adv.* **4**, aap7885 (2018).
49. Zhavoronkov, A. et al. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nat. Biotechnol.* **37**, 1038–1040 (2019).
50. Schneider, P. et al. Rethinking drug design in the artificial intelligence era. *Nat. Rev. Drug Discov.* **19**, 353–364 (2020).
51. Sadybekov, A. V. & Katritch, V. Computational approaches streamlining drug discovery. *Nature* **616**, 673–685 (2023).
52. Iten, R., Metger, T., Wilming, H., Del Rio, L. D. & Renner, R. Discovering physical concepts with neural networks. *Phys. Rev. Lett.* **124**, 010508 (2020).
53. Seif, A., Hafezi, M. & Jarzynski, C. Machine learning the thermodynamic arrow of time. *Nat. Phys.* **17**, 105–113 (2021).
54. Wu, T. & Tegmark, M. Toward an artificial intelligence physicist for unsupervised learning. *Phys. Rev. E* **100**, 033311 (2019).
55. Lu, L., Jin, P., Pang, G., Zhang, Z. & Karniadakis, G. E. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nat. Mach. Intell.* **3**, 218–229 (2021).
56. Han, J., Jentzen, A. & Weinan, E. Solving high-dimensional partial differential equations using deep learning. *Proc. Natl Acad. Sci. USA* **115**, 8505–8510 (2018).
57. Raayoni, G. et al. Generating conjectures on fundamental constants with the Ramanujan Machine. *Nature* **590**, 67–73 (2021).
58. Degrave, J. et al. Magnetic control of tokamak plasmas through deep reinforcement learning. *Nature* **602**, 414–419 (2022).
59. Tshitoyan, V. et al. Unsupervised word embeddings capture latent knowledge from materials science literature. *Nature* **571**, 95–98 (2019).
60. Sanchez-Lengeling, B. & Aspuru-Guzik, A. Inverse molecular design using machine learning: generative models for matter engineering. *Science* **361**, 360–365 (2018).
61. Chen, C. et al. A critical review of machine learning of energy materials. *Adv. Energy Mater.* **10**, 1903242 (2020).
62. Merchant, A. et al. Scaling deep learning for materials discovery. *Nature* **624**, 80–85 (2023).
63. Zheng, S. et al. The AI Economist: taxation policy design via two-level deep multiagent reinforcement learning. *Sci. Adv.* **8**, eabk2607 (2022).
64. Koster, R. et al. Human-centred mechanism design with Democratic AI. *Nat. Hum. Behav.* **6**, 1398–1407 (2022).
65. Dunjko, V. & Briegel, H. J. Machine learning & artificial intelligence in the quantum domain: a review of recent progress. *Rep. Prog. Phys.* **81**, 074001 (2018).
66. Sturm, B. L. et al. Machine learning research that matters for music creation: a case study. *J. N. Music Res.* **48**, 36–55 (2019).
67. Rajkomar, A., Dean, J. & Kohane, I. Machine learning in medicine. *N. Engl. J. Med.* **380**, 1347–1358 (2019).
68. Ramesh, A. et al. Zero-shot text-to-image generation. In *Proc. 38th International Conference on Machine Learning* 8821–8831 (ICML, 2021).
69. Epstein, Z., Hertzmann, A. & the Investigators of Human Creativity. Art and the science of generative AI. *Science* **380**, 1110–1111 (2023).
70. Swanson, K. et al. Generative AI for designing and validating easily synthesizable and structurally novel antibiotics. *Nat. Mach. Intell.* **6**, 338–353 (2024).
71. Crafts, N. Artificial intelligence as a general-purpose technology: an historical perspective. *Oxf. Rev. Econ. Policy* **37**, 521–536 (2021).
72. Bloom, N., Hassan, T. A., Kalyani, A., Lerner, J. & Tahoun, A. *The Diffusion of New Technologies* NBER Working Paper No. w28999 (National Bureau of Economic Research, 2021).
73. Caselli, F. & Coleman, W. J. Cross-country technology diffusion: the case of computers. *Am. Econ. Rev.* **91**, 328–335 (2001).
74. Comin, D. & Hobijn, B. An exploration of technology diffusion. *Am. Econ. Rev.* **100**, 2031–2059 (2010).
75. Zenil, H. et al. The future of fundamental science led by generative closed-loop artificial intelligence. Preprint at https://arxiv.org/abs/2307.07522 (2023).
76. Topol, E. J. High-performance medicine: the convergence of human and artificial intelligence. *Nat. Med.* **25**, 44–56 (2019).
77. Hidalgo, C. A., Orghian, D., Albo Canals, J., de Almeida, F. & Martín Cantero, N. *How Humans Judge Machines* (MIT Press, 2021).
78. Raisch, S. & Krakowski, S. Artificial intelligence and management: the automation–augmentation paradox. *Acad. Manag. Rev.* **46**, 192–210 (2021).
79. Fjelland, R. Why general artificial intelligence will not be realized. *Humanit. Soc. Sci. Commun.* **7**, 10 (2020).
80. Messeri, L. & Crockett, M. J. Artificial intelligence and illusions of understanding in scientific research. *Nature* **627**, 49–58 (2024).
81. Kanitscheider, I. & Fiete, I. Training recurrent networks to generate hypotheses about how the brain solves hard navigation problems. In *Proc. Advances in Neural Information Processing Systems* 4529–4538 (MIT Press, 2017).
82. Webb, M. *The Impact of Artificial Intelligence on the Labor Market* SSRN 3482150 (Social Science Research Network, 2019).
83. Kogan, L., Papanikolaou, D., Schmidt, L. D. & Seegmiller, B. *Technology, Vintage-Specific Human Capital, and Labor Displacement: Evidence from Linking Patents with Occupations* NBER Working Paper No. w29552 (National Bureau of Economic Research, 2022).
84. Atalay, E., Phongthiengtham, P., Sotelo, S. & Tannenbaum, D. The evolution of work in the United States. *Am. Econ. J. Appl. Econ.* **12**, 1–34 (2020).
85. Felten, E. W., Raj, M. & Seamans, R. A method to link advances in artificial intelligence to occupational abilities. *AEA Pap. Proc.* **108**, 54–57 (2018).
86. Wu, L., Hitt, L. & Lou, B. Data analytics, innovation, and firm productivity. *Manag. Sci.* **66**, 2017–2039 (2020).
87. Brynjolfsson, E., Mitchell, T. & Rock, D. What can machines learn, and what does it mean for occupations and the economy? *AEA Pap. Proc.* **108**, 43–47 (2018).
88. Wang, D. & Barabási, A.-L. *The Science of Science* (Cambridge Univ. Press, 2021).

89. Fortunato, S. et al. Science of science. *Science* **359**, eaao0185 (2018).
90. Zeng, A. et al. The science of science: from the perspective of complex systems. *Phys. Rep.* **714**, 1–73 (2017).
91. Frank, M. R., Wang, D., Cebrian, M. & Rahwan, I. The evolution of citation graphs in artificial intelligence research. *Nat. Mach. Intell.* **1**, 79–85 (2019).
92. Miao, L. et al. The latent structure of global scientific development. *Nat. Hum. Behav.* **6**, 1206–1217 (2022).
93. Liu, L., Jones, B. F., Uzzi, B. & Wang, D. Data, measurement and empirical methods in the science of science. *Nat. Hum. Behav.* **7**, 1046–1058 (2023).
94. Sourati, J. & Evans, J. A. Accelerating science with human-aware artificial intelligence. *Nat. Hum. Behav.* **7**, 1682–1696 (2023).
95. Murray, D. et al. Unsupervised embedding of trajectories captures the latent structure of scientific migration. *Proc. Natl Acad. Sci. USA* **120**, e2305414120 (2023).
96. Krenn, M. et al. Forecasting the future of artificial intelligence with machine learning-based link prediction in an exponentially growing knowledge network. *Nat. Mach. Intell.* **5**, 1326–1335 (2023).
97. Sinha, A. et al. An overview of Microsoft Academic Service (MAS) and applications. In *Proc. 24th International Conference on World Wide Web* 243–246 (WWW, 2015).
98. World Intellectual Property Organization (WIPO). *WIPO Technology Trends 2019—Artificial Intelligence* (WIPO, 2019).
99. Nivre, J. & Nilsson, J. Pseudo-projective dependency parsing. In *Proc. 43rd Annual Meeting of the Association for Computational Linguistics* 99–106 (ACL, 2005).
100. Honnibal, M. & Johnson, M. An improved non-monotonic transition system for dependency parsing. In *Proc. 2015 Conference on Empirical Methods in Natural Language Processing* 1373–1378 (ACL, 2015).
101. Benetka, J. R., Krumm, J. & Bennett, P. N. Understanding context for tasks and activities. In *Proc. 2019 Conference on Human Information Interaction and Retrieval* 133–142 (ACM, 2019).
102. Service, R. Science's 2021 Breakthrough of the Year: protein structures for all. *Science* https://www.science.org/content/article/breakthrough-2021 (2021).
103. Börner, K. et al. Skill discrepancies between research, education, and jobs reveal the critical need to supply soft skills for the data economy. *Proc. Natl Acad. Sci. USA* **115**, 12630–12637 (2018).
104. Wuchty, S., Jones, B. F. & Uzzi, B. The increasing dominance of teams in production of knowledge. *Science* **316**, 1036–1039 (2007).
105. Wu, L., Wang, D. & Evans, J. A. Large teams develop and small teams disrupt science and technology. *Nature* **566**, 378–382 (2019).
106. Littmann, M. et al. Validity of machine learning in biology and medicine increased through collaborations across fields of expertise. *Nat. Mach. Intell.* **2**, 18–24 (2020).
107. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K. & Galstyan, A. A survey on bias and fairness in machine learning. *ACM Comput. Surv.* **54**, 1–35 (2021).
108. Young, E., Wajcman, J. & Sprejer, L. *Where Are the Women? Mapping the Gender Job Gap in AI* (The Alan Turing Institute, 2021).
109. Xie, Y. & Shauman, K. A. *Women in Science: Career Processes and Outcomes* (Harvard Univ. Press, 2003).
110. Hoppe, T. A. et al. Topic choice contributes to the lower rate of NIH awards to African-American/black scientists. *Sci. Adv.* **5**, eaaw7238 (2019).
111. Ginther, D. K. et al. Race, ethnicity, and NIH research awards. *Science* **333**, 1015–1019 (2011).
112. Larivière, V., Ni, C., Gingras, Y., Cronin, B. & Sugimoto, C. R. Bibliometrics: global gender disparities in science. *Nature* **504**, 211–213 (2013).
113. Huang, J., Gates, A. J., Sinatra, R. & Barabási, A.-L. Historical comparison of gender inequality in scientific careers across countries and disciplines. *Proc. Natl Acad. Sci. USA* **117**, 4609–4616 (2020).
114. The National Network for Critical Technology Assessment (NNCTA). *Securing America's Future: A Framework for Critical Technology Assessment* (NNCTA, 2023).
115. Cachola, I., Lo, K., Cohan, A. & Weld, D. S. TLDR: extreme summarization of scientific documents. In *Proc. 2020 Conference on Empirical Methods in Natural Language Processing* 4766–4777 (ACL, 2020).
116. Lew, A., Agrawal, M., Sontag, D. & Mansinghka, V. PClean: Bayesian data cleaning at scale with domain-specific probabilistic programming. In *Proc. 24th International Conference on Artificial Intelligence and Statistics* 1927–1935 (JMLR, 2021).
117. Bommasani, R. et al. On the opportunities and risks of foundation models. Preprint at https://arxiv.org/abs/2108.07258 (2021).
118. Wei, J. et al. Emergent abilities of large language models. Preprint at https://arxiv.org/abs/2206.07682 (2022).
119. Moor, M. et al. Foundation models for generalist medical artificial intelligence. *Nature* **616**, 259–265 (2023).
120. Huang, Z., Bianchi, F., Yuksekgonul, M., Montine, T. J. & Zou, J. A visual-language foundation model for pathology image analysis using medical Twitter. *Nat. Med.* **29**, 2307–2316 (2023).
121. Goldfarb, A., Taska, B. & Teodoridis, F. Could machine learning be a general purpose technology? A comparison of emerging technologies using data from online job postings. *Res. Policy* **52**, 104653 (2023).
122. Fawzi, A. et al. Discovering faster matrix multiplication algorithms with reinforcement learning. *Nature* **610**, 47–53 (2022).
123. Jobin, A., Ienca, M. & Vayena, E. The global landscape of AI ethics guidelines. *Nat. Mach. Intell.* **1**, 389–399 (2019).
124. Arrieta, A. B. et al. Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* **58**, 82–115 (2020).
125. Lenharo, M. An AI revolution is brewing in medicine. What will it look like? *Nature* **622**, 686–688 (2023).
126. Bockting, C. L., van Dis, E. A. M., van Rooij, R., Zuidema, W. & Bollen, J. Living guidelines for generative AI—why scientists must oversee its use. *Nature* **622**, 693–696 (2023).
127. Schwartz, I. S., Link, K. E., Daneshjou, R. & Cortes-Penfield, N. Black box warning: large language models and the future of infectious diseases consultation. *Clin. Infect. Dis.* **78**, 860–866 (2024).
128. Ahmadpoor, M. & Jones, B. F. The dual frontier: patented inventions and prior scientific advance. *Science* **357**, 583–587 (2017).
129. Mukherjee, S., Romero, D. M., Jones, B. & Uzzi, B. The nearly universal link between the age of past knowledge and tomorrow's breakthroughs in science and technology: the hotspot. *Sci. Adv.* **3**, e1601315 (2017).
130. Yin, Y., Dong, Y., Wang, K., Wang, D. & Jones, B. F. Public use and public funding of science. *Nat. Hum. Behav.* **6**, 1344–1350 (2022).
131. Microsoft Academic. Microsoft Academic Graph. *Zenodo* https://doi.org/10.5281/zenodo.6511057 (2022).
132. Lin, Z., Yin, Y., Liu, L. & Wang, D. SciSciNet: a large-scale open data lake for the science of science research. *Sci. Data* **10**, 315 (2023).

Acknowledgements

We thank Y. Yin, Y. Qian, B. Wang, N. Dehmamy, L. Varshney, L. Miao, L. Wu, A. Freilich and all members of the Center for Science of

Science and Innovation (CSSI) at Northwestern University for helpful discussions. This work is supported by the Air Force Office of Scientific Research FA9550-19-1-0354 (D.W.), the National Science Foundation SBE 1829344, TIP 1123649-464363//2241237, and TIP 2404035 (D.W.), the Alfred P. Sloan Foundation G-2019-12485 (D.W.), and the Peter G. Peterson Foundation 21048 (D.W.). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the paper.

Author contributions

J.G. and D.W. conceived the idea. D.W. supervised the project. J.G. collected data and performed analyses. J.G. and D.W. analysed the results, interpreted the findings and wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-024-02020-5.

Correspondence and requests for materials should be addressed to Dashun Wang.

Peer review information *Nature Human Behaviour* thanks Alexander Gates and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

nature portfolio

Corresponding author(s): Dashun Wang

Last updated by author(s): Sep 6, 2024

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection	Publication, patent, and syllabus data was directly downloaded from databases; Researcher demographic summary statistics data was downloaded from the National Science Foundation website.
Data analysis	Data is analyzed with customized code in Python 3 using standard software packages within these programs, including pandas 1.3.5, numpy 1.21.5, scipy 1.7.3, matplotlib 3.5.1, seaborn 0.11.2, spacy 3.7.2, nomquamgender 0.1.0, demographicx 0.0.1, and others. The code necessary to reproduce all main plots and statistical analyses has been made available at https://kellogg-cssi.github.io/ai4science.

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

Data

Policy information about availability of data

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

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

The MAG data are available at https://zenodo.org/record/6511057. The USPTO patent data are available at https://patentsview.org. The OSP dataset is available from the paper at https://www.pnas.org/doi/10.1073/pnas.1804247115. The SDR data are available at https://www.nsf.gov/statistics/srvydoctoratework. The data necessary to reproduce all main plots in this paper has been made available at https://kellogg-cssi.github.io/ai4science.

Research involving human participants, their data, or biological material

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

Reporting on sex and gender	Sex of researchers was originally collected and reported in the Survey of Doctorate Recipients (SDR), which is sponsored by the National Center for Science and Engineering Statistics (NCSES) within the National Science Foundation (NSF) and by the National Institutes of Health (NIH). We use the number of female and male researchers in each scientific discipline and research field. We report results for both female and male.
Reporting on race, ethnicity, or other socially relevant groupings	Race and ethnicity of researchers were originally collected and reported in the Survey of Doctorate Recipients (SDR). We use the number of White, Asian, African American or Black, American Indian or Alaska Native, Hispanic or Latino, and Native Hawaiians or other Pacific Islanders researchers in each scientific discipline and field. We report results for all these race and ethnicity groups.
Population characteristics	n/a
Recruitment	n/a
Ethics oversight	n/a

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

Field-specific reporting

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

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

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

Behavioural & social sciences study design

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

Study description	A study to quantify the use and benefits of artificial intelligence for scientific research
Research sample	We use 74.6 million publications from the Microsoft Academic Graph (MAG) database, 7.1 million patents from the USPTO PatentsView database, 4.2 million course syllabi from the Open Syllabus Project (OSP), and demographic statistical data of researchers from the Survey of Doctorate Recipients (SDR). The data represent the most representative samples of publications, patents, syllabi, and researcher composition to date. The study samples were chosen to study all disciplines in science and education.
Sampling strategy	We consider MAG publications of various types (“journal,” “conference,” “book,” or “book chapter”) in 1960-2019, USPTO patents approved in 1976-2019, OSP syllabus documents published in 2000-2018, and the 2017 SDR data on scientists and engineers. There was no sampling produce to select the data, and no sample size calculation was performed. We use all data for each domain, which is the most comprehensive sample at the time of analysis.
Data collection	Publication data was downloaded from the Microsoft Academic Graph (MAG) database; Patent data was downloaded from PatentsView, a data platform based on bulk data from the U.S. Patent and Trademark Office (USPTO); Course syllabus data was sourced from the Open Syllabus Project (OSP); Demographic data of researchers was collected from the Survey of Doctorate Recipients (SDR), which is public available from the NSF website.
Timing	Data was collected in the period of 2019-2021.
Data exclusions	We only consider MAG publications with four types (“journal,” “conference,” “book,” or “book chapter”) in 1960-2019, USPTO patents approved in 1976-2019, OSP syllabus documents published in 2000-2018, and the 2017 SDR data on scientists and engineers.
Non-participation	No participants were involved in this study since all data were from existing public databases.
Randomization	Randomization was not applicable, given the observational nature of the study.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

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

Methods

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