AI in Science: Closing the Experience Gap?

The Impact of AlphaFold on Life Scientists' Productivity

Introduction

Production of knowledge, particularly scientific research, is a crucial driver of innovation and economic growth. The link between scientific advancements, innovation, and firm performance is well recognized in strategy literature (Cohen & Levinthal, 1990; Fleming & Sorenson, 2004). As such, understanding the process and direction of scientific research and their determining factors is of great importance and interest in the field (Wuchty, Jones & Uzzi, 2007; Bikard, Murray & Gans, 2015; Wu, Wang, & Evans, 2019; Park, Leahey & Funk,2023).

Among these factors, technological change has long been recognized as a significant force shaping the landscape of scientific research. New technologies have the potential to change the nature of scientific research by providing novel and enhanced methodologies for inquiry and transforming how research is organized (Griliches, 1957; Cockburn, Henderson & Stern, 2019). Historically, technological advancements such as the advent of the microscope (Gest, 2004), gene sequencing (Heather & Chain, 2016) and information technology (Ding et al., 2010) enabled enhanced productivity and novel discoveries in scientific research. Most recently, artificial intelligence (AI) is being seen as the next big technology to transform the practice of scientific research.

AI represents one of the most significant technological changes of our time. Over the last decade, AI has demonstrated remarkable capabilities, from beating the world champion at the game of Go (AlphaGo, 2016), to solving the protein folding problem (AlphaFold, 2020), and exhibiting human-like language capabilities (ChatGPT, 2022). Most recently, AI achieved near-human levels of performance in complex cognitive tasks, including Google DeepMind's success at the 2024 International Math Olympiad (DeepMind, 2024). Given the history of groundbreaking technologies driving scientific progress, AI's ability to tackle complex problems suggests that it, too, will play a pivotal role in scientific discovery. Researchers across fields are already integrating AI into their work (Bhardwaj, Kishore & Pandey, 2022; Wang et al., 2023), amplifying expectations for the future.

Despite the increasing integration of AI in the process of scientific discovery, its effects on research outcomes, processes, and implications for researchers remain underexplored. Understanding whether and how AI affects the performance and productivity of scientific research and what the implication for the researchers is an important question to be answered in the literature on the management of science and innovation. Not only would AI affect the pace and direction of scientific advancements, but it also has the potential to change the nature of the process and organization of innovative activities (Agrawal, McHale & Oettl, 2024; Cockburn, Henderson & Stern, 2019). Such changes could have important implications on firms' innovation strategy as they strive to identify, acquire, and utilize the right type of resources suitable for AI-driven innovation.

A key question in this regard is whether AI affects the research productivity of scientists, and if so, whether the effects are heterogeneous depending on researcher characteristics such as skill or experience. While several studies examine AI's impact on task performance outside the context of science (Brynjolfsson, Li & Raymond, 2023; Peng, Kalliamvakou, Cihon & Demirer, 2023; Noy & Zhang, 2023; Doshi & Hauser, 2023; Dell'Acqua et al., 2023, Kanazawa, Kawaguchi, Shigeoka & Watanabe, 2022), most studies are set on relatively simple, routine tasks where AI can provide a reliable, high-quality solution. Given AI's varied capabilities across domains (Dell'Acqua et al., 2023) and the complexity of scientific research, its impact on scientific research may differ from simpler contexts and requires further investigation.

This paper explores how adoption of AI in research affects the productivity of individual researchers. Specifically, I focus on the use of AlphaFold, a deep-learning AI system developed by Google DeepMind for protein 3D structure prediction (Jumper et al., 2021a), tracking its impact on the publication records of life scientists who incorporate this tool into their research. I explore the question of whether AI benefits highly experienced or inexperienced researchers more in terms of research productivity. My results suggest that AlphaFold plays a leveling role in scientific research, benefitting researchers with lower experience and productivity more than highly experienced researchers.

Background

AI in Science

Due to rapid advances in computing technology and developments of novel machine learning and deep learning algorithms, AI technology has shown remarkable growth in the past decade and is increasingly being integrated into creative and innovative activities. Science research is no exception with AI being used in each step of scientific inquiry. For example, in drug discovery studies, AI models are used to propose better drug candidates with higher rates of leading to a successful drug development. This new approach to drug discovery, dubbed computer-aided drug design (CADD), has radically increased the hit rate of drug screening compared to traditional methods (Liu, He, Luo, Zhang, Jiang, 2019). AI is also used to analyze large and complex data, such as identifying and classifying collision events and detecting anomalies from data generated by the Large Hadron Collider for particle physics research. Recent advances in large language models (LLMs) also accelerate literature review and optimize workflow. Overall, these use cases demonstrate the potential of AI to enhance and accelerate the process of science research to increase productivity and growth.

Literature on AI offers a more systematic documentation of how the technology is being diffused in the sciences. Cockburn, Henderson & Stern (2019) and Bianchini, Müller & Pelletier (2022) both report a recent boom of AI application in a wide variety of scientific outlets and disciplines, suggesting that AI could be considered an emerging general method of invention. Wang et al. (2023) reviews how AI is being integrated into different steps of scientific discovery—hypothesis generation, experiment design, data collection and analysis—across disciplines, noting the technology's potential to not only enhance scientific understanding but also transform the way science research is done.

As AI continues to gain attention and diffuse within the scientific community, many have high hopes that it will accelerate and even automate parts of scientific research. It is expected that AI can help boost productivity by analyzing data, uncovering new insights, and making complex tasks more efficient. However, despite these optimistic views, there remains a lack of studies that directly assess whether AI actually improves research productivity. While studies in management that link AI with firm-level outcomes—such as product innovation and firm growth (Babina et al., 2024), technological innovation (Liu et al., 2020), explorative activities (Johnson et al., 2022), and novel and process innovation (Rammer et al., 2022)—suggest that AI can enhance the effectiveness of innovative activities, AI's effect on scientific research productivity remains relatively unexamined. Agrawal, McHale & Oettl (2024) puts forward a theoretical framework on the use of AI in scientific discovery, focusing on AI's potential to enhance our ability to better distinguish good hypotheses—with higher probability of leading to a discovery—from bad ones. The model posits that having access to these high-fidelity predictions provided by AI can accelerate the process of scientific discovery. However, the predictions of this model have yet to be supported by empirical evidence. These shortcomings in the literature highlight the need for more research to understand AI's real impact on advancing scientific work and increasing research productivity.

AI and productivity

The relationship between AI and productivity has been a major focus in recent literature outside the context of scientific research, with most studies agreeing that AI contributes significantly to increased productivity (Seamans & Raj, 2018; Damioli, Van Roy & Vertesy, 2021; Czarnitzki, Fernández & Rammer, 2023). These studies mostly analyze the effects of AI adoption and development at the economy level or the firm level, showing how the technology contributes to overall productivity growth. However, questions remain regarding the mechanisms through which AI drives productivity gains and which groups benefit most from its use, requiring analysis at a more granular level.

In recent years, a number of studies have focused on the impact of AI at the individual level. Overall, these studies suggest that the benefits of using AI vary depending on the task and the user. One finding shared among many studies is the heterogeneous effect of AI on the productivity of workers with different levels of skills or experience, mostly favoring those with less experience and skill. For example, Brynjolfsson, Li & Raymond (2023) finds that lower-skilled and inexperienced service workers benefit more from using generative AI in their work than do higher-skilled and experienced workers. Similar ‘skill-leveling’ or ‘experience-leveling’ effects of AI have been found in various contexts such as software programming (Peng, Kalliamvakou, Cihon & Demirer, 2023), professional and creative writing (Noy & Zhang, 2023; Doshi & Hauser, 2023), business consulting (Dell’Acqua et al., 2023), and taxi driving (Kanazawa, Kawaguchi, Shigeoka & Watanabe, 2022).

A key question emerging from these findings is why less experienced users seem to derive greater benefits from using AI. One explanation is that AI may serve as an equalizing force by providing users of all experience levels access to a benchmark solution for relatively narrow and standardized tasks. The studies cited above primarily examine straightforward, well-defined tasks where AI can generate solutions whose quality remains fairly consistent regardless of the user's prior experience. In such task contexts, workers' performance would tend to converge toward the capability level established by the AI. Consequently, those with limited experience who initially performed below this threshold would see the greatest productivity improvements, while more experienced workers see smaller marginal gains since their performance already approaches the AI's output level. The access to high-performing solutions can also accelerate the learning process for inexperienced users (Brynjolfsson, Li & Raymond, 2023), allowing them to progress along the learning curve more rapidly and narrow the gap between themselves and their highly experienced counterparts.

However, it remains uncertain whether the same dynamics apply in more complex task environments, such as scientific research. In such environments, AI may not be able to fully substitute for human expertise due to its limited scope of tasks it can perform (Dell’Acqua et al., 2023) and the inherent complexity and creativity involved in the tasks. In this case, people with better ability to carry out the task portion untouched by AI may benefit more by optimizing their workflow. For example, Jia, Luo, Fang & Liao (2024) examines the effect of AI where the work is characterized as having a routine task component and a more complex one. When the routine task was automated by AI but not the complex portion, higher-skilled workers were able to better leverage the freed-up cognitive resources to be more creative in their complex tasks while lower-skilled workers faced increased pressure from the complex task, receiving diminished benefit from working with AI.

Additionally, in complex task environments, there could be greater latitude in how AI tools may be applied. Productivity in this case would depend not only on the decision to use AI, but also on how the AI would be used for which part of the task. Users with more experience with the task may help make such decisions to utilize AI in a way more suited for the task environment, resulting in larger productivity increase.

As a result, it remains unclear whether AI will play the same equalizing role as documented in the existing literature when applied to more complex and open-ended tasks like scientific research. How researchers with different levels of expertise may benefit from AI is an important question to understand the potential of the new technology in shaping scientific progress and its implication for the organization of science and the labor market for science. In this paper, I use AlphaFold as an example of an AI tool for science and explore how using the tool may have different effects on the productivity of researchers with different levels of experience.

AlphaFold

AlphaFold is an AI system developed by DeepMind, an AI research subsidiary of Google. Based on deep learning algorithms, it is designed to predict the 3-dimensional structure of a protein molecule from its sequence. Proteins are made up of long chains of amino acids which fold and bond with each other to create the protein's unique 3D structure. Identifying this structure is crucial to understanding a protein's function and its interactions with other molecules, with important implications for fields such as bioscience and drug development. Historically, researchers relied on experimental methods such as X-ray crystallography, Nuclear Magnetic Resonance (NMR) spectroscopy, or Cyro Electron Microscopy (Cyro-EM), which could be costly and time-consuming. According to Demis Hassabis, the founder and CEO of DeepMind, "… it takes one whole Ph.D., their whole Ph.D. time, … to crystallize one protein, and then using X-ray crystallography or electron microscopes, … to basically image these incredibly small, complex structures." (Klein, 2023)

To tackle this painstaking task, scientists for decades have worked on methods to predict a protein's 3D structure just from its 1D amino acid sequence, known as the protein-folding problem. Early computational methods such as homology modeling (inferring structures based on similar known structures) and physics-based simulations have been developed (Dill & MacCallum, 2012). A significant breakthrough was achieved when deep learning algorithms were incorporated to solve the protein-folding problem. In 2018, at CASP13, a biennial competition for protein structure models, AlphaFold came in first with significant margins from the runner-up. Two years later at CASP14, the improved AlphaFold2 model not only won the competition but also achieved accuracy levels comparable to those obtained from experimental methods. With some caveats (Al-Janabi, 2022; Jumper et al., 2021b), many, including the organizers of CASP14, saw this breakthrough as essentially solving the protein folding problem.

In addition to its unprecedented performance in computationally predicting protein structures, the public release of the source code for AlphaFold and its predictions over a large number of proteins as a database (Varadi et al., 2024) made AlphaFold a rapidly adopted tool in related fields of research. The foundational paper that describes AlphaFold (Jumper et al., 2021a), which one is required to cite when using the released model or its predictions, has over 15,000 citations according to Clarivate Web of Science. While not all citations indicate the use of the model, the extremely high count reflects the popularity of AlphaFold as a research tool.

AlphaFold and Research Productivity

To theorize how AlphaFold might affect research productivity, I model the process of scientific research as a search over a solution space—the set of all potential solutions—to a problem (Nickerson and Zenger, 2004; Sorenson et al., 2006). For example, the 'discovery' of a new chemical compound with potential drug applications is done by navigating the universe of over 10^{60} small molecules (Virshup et al., 2013). In organic chemistry, discovery is made in the form of identifying a reaction setting that gives a high yield of a target chemical. The yield of a chemical reaction can be affected by numerous factors, such as input materials, catalysts, temperature, pressure, and so on. The possible combinations of each factor make up the solution space, which can be as big as millions of different settings.

To discover a solution to a given problem, a researcher would 1) identify promising candidates from the solution space and 2) test the identified candidates through experiments. A scientist's ability to produce scientific discoveries would thus be determined by his/her ability to utilize his/her theoretical knowledge and experience to come up with accurate predictions, or hypotheses, about which candidates are the most likely to be a real solution, and the capacity to run the experiments. The former will reduce the number of trials and errors one must go through before making a discovery, and the latter will reduce the time needed to run a given number of trials.

The value of AlphaFold in life science research is that it provides low-cost, highly accurate predictions for most protein molecules. While it is still the norm in the field that AlphaFold prediction by itself is usually not accepted as direct evidence of the protein's structure, a large portion of AlphaFold's value lies in the quick and easy access it provides researchers to get low-cost, highly accurate predictions for a large number of protein molecules before committing resources into experiments. Because the structure of a protein is highly informative of the protein's characteristic and function, researchers can use AlphaFold predictions to gain better insights into whether a protein is likely to lead to a scientific discovery (development of a new drug, identification of a biological mechanism, etc.). Thus, while AlphaFold may not accelerate the speed of testing a hypothesis, it could still enhance research productivity by reducing the number of trials needed to make a discovery by generating better hypotheses with higher hit rates.

For the narrow task of generating hypotheses about promising solutions for a problem, an argument can be made that similar to findings from the literature, less experienced and lower-skilled researchers would benefit more from adopting AlphaFold. To the extent that the AI provides standardized, high-quality predictions about promising candidates, those who were less able to generate such hypotheses on their own would benefit more from the standard set by AlphaFold. Thus, confined to the task of generating hypotheses for a given research question, AlphaFold may enhance the performance of inexperienced researchers more than experienced ones.

However, the task complexity of science research makes it difficult to make the same prediction about the productivity in overall research in which many more processes are involved than hypotheses generation. On the contrary, AlphaFold may favor researchers with high skill and expertise for a number of reasons. For example, researchers can choose from numerous research questions to pursue, and experienced researchers may be better positioned to recognize research questions that could benefit the most from AlphaFold, maximizing the benefits from using the AI. Also, the researcher's knowledge and expertise may complement AlphaFold's predictions to further process the information and narrow down highly promising candidates. As such, because research productivity still largely relies on efficiency in task portions not affected by AlphaFold, and also because researchers may differ in their ability to utilize AlphaFold, it could be argued that highly experienced researchers may benefit more from adopting AlphaFold in terms of overall research productivity. Thus, this paper examines the correlation between the adoption of AlphaFold and subsequent research productivity among researchers with varying levels of expertise, without proposing a specific hypothesis.

Data

To investigate the research output of scientists, I use researcher-level publication data from the OpenAlex database (Priem, Piwowar & Orr, 2022), an open-source catalog of scholarly works maintained by OurResearch. I accessed and downloaded data from the database from July 24 through 31, 2024.

To identify researchers using AlphaFold in their research, I collect all 14,209 papers that cite the foundational AlphaFold paper (Jumper et al., 2021a) as listed in the database. Of these, I retain the 9,233 papers classified within the life sciences domain, authored by a total of 56,628 unique scientists. I collect publication history data for these scientists to construct a monthly panel dataset. I treat the month of an author's first AlphaFold-citing publication as his/her month of AlphaFold adoption and compile the records for 12 months preceding and following adoption. I exclude researchers with publication records shorter than 12 months preceding and following adoption. After filtering out erroneous data points (e.g., papers citing the AlphaFold paper before its publication date, papers with publication dates after the database access date, authors with implausibly high publication rates[1], long publication gaps[2], or numerous associated names[3]), the final panel dataset consists of 16,089 unique authors 386,136 unique author-month pairs.

Model

Two-Way Fixed Effects Regression

I use a two-way fixed effects (TWFE) regression model to estimate the effect of AlphaFold adoption on researcher outputs. The model accounts for variation in time of treatment across units with unit and time fixed effects. The model is specified as follows:

$$Y_{it} = \tau D_{it} + \beta X_{it} + \alpha_i + \lambda_t + \epsilon_{it}$$

Where Y_{it} is the monthly publication count of unit i at time t, τ is the treatment effect, D_{it} is the treatment indicator for unit i at time t, X_{it} are the covariates, and α_i and λ_t are unit and time

[1] More than 20 publications per month.
[2] More than 50 years between first and last publication.
[3] I perform a fuzzy string match on all the names listed under the same researcher ID and drop all records with 3 or more unique names.

fixed effects, respectively. I use the *xtreg* command in STATA 18 to run the regression. I cluster the standard errors on the unit (researcher) level.

Variables

Dependent Variables: Productivity Measures

Monthly Publication Count: As a measure of researcher productivity, I count the number of monthly publications for each researcher. While elementary, the number of publications within a given timeframe is a widely used metric for researcher productivity (Carpenter, Cone & Sarli, 2014).

Quality-Adjusted Monthly Publication Count: To account for the possibility that AlphaFold adoption affects publication quality rather than just quantity, I use an alternative dependent variable that weights monthly publication counts by the Journal Impact Factor (JIF) of the journal each work was published in. The JIF data is sourced from Clarivate's 2023 Journal Citation Reports. While JIF has known limitations as a measure of individual paper quality (Seglen, 1997), it serves as a useful proxy for the selectivity and visibility of publication venues. Higher JIF journals typically maintain more rigorous peer review processes and attract submissions with potentially higher scientific impact. Moreover, JIF continues to play a significant role in academic career advancement, with scientists facing institutional incentives to publish in high-impact journals (McKiernan et al., 2019; Tregoning, 2018). For these reasons, JIF of the publication outlet can be a practical proxy for paper quality, particularly when analyzing recent publications that have not had sufficient time to accumulate citations as an alternative quality measure.

Explanatory Variables

AlphaFold Adoption: For each researcher, I treat his/her first publication that cites the AlphaFold paper (Jumper et al., 2021a) as the time of adoption. I introduce a dummy variable for post-adoption period which is 0 before the adoption date and 1 after the adoption date. It should be noted that 1) the first publication that cites AlphaFold may not accurately capture the researcher's first time using AlphaFold, and that 2) the adoption is a non-random assignment depending on the researcher's own decision to use the tool or not. For these reasons, the difference in productivity between before and after the measured adoption should not be interpreted as a causal effect of AlphaFold adoption. I discuss this limitation in more detail in the discussion section.

Months since first publication: In addition to unit (author) and time fixed effects, I control for the number of months since each author's first publication to control for the potential time-dependent trends unrelated to the adoption of AlphaFold. For example, a researcher's productivity may naturally increase over time as they accumulate experience, skills, and resources. Without controlling for career duration, this increase in productivity may be mistakenly attributed to AlphaFold adoption.

Experience and skill levels at before adoption: To explore differences between high-skilled and low-skilled researchers, I introduce 3 moderator variables: total number of publications before adoption, months since first publication at the time of adoption, and productivity over 12 months before adoption. The total number of publications before adoption counts all publications authored by a researcher before their AlphaFold adoption month. Months since first publication at the time of adoption measures the time gap (in months) between a researcher's first publication month and their AlphaFold adoption month. Productivity over 12 months before adoption measures the researcher's average monthly publications during the 12-month period immediately preceding adoption.

Table 1 reports the descriptive statistics and correlation between each variable. The explanatory variables were centered around the mean for the regression analysis.

Analysis and Results

Baseline level change after adoption

Column 1 in Table 2 reports the regression results for the overall change in monthly publication counts after adopting AlphaFold for the full sample. The coefficient for the post-adoption period is positive, but insignificant ($\beta = 0.00430, SE = 0.00558$). Table 3 also reports the regression using JIF-weighted publication counts as the dependent variable. The coefficient for the post-adoption period is positive and significant ($\beta = 0.191, SE = 0.0597$).

Moderation by Experience Level

To explore whether the change in research productivity after adopting AlphaFold differ between researchers with different levels of experience, I use a number of proxy variables for researcher experience to interact with the post-adoption dummy. As my main result, I only report the results with the total number of publications before adoption as the moderator variable. I also conduct the same analysis using experience in months since first publication at the time of adoption and the average monthly publication for 1 year before adoption. I find similar results and report them in the appendix (Appendix Table 1).

Column 2 in Table 2 reports the regression results with the moderator variable. The interaction term between post-adoption period and the total number of publications before adoption has a negative and significant coefficient ($\beta = -0.000864, SE = 0.000151$). For researchers with a one standard deviation (41.35) higher total pre-adoption publication count, the association between adoption and yearly publications is 0.429 publications lower (7.4% of the mean or 0.037 SD). Overall, the results suggest that researchers with less experience and lower productivity experience a larger increase in productivity after adopting AlphaFold. However, as reported in Column 2 of Table 3, the interaction term is not significant when using JIF-weighted count as the dependent variable.

To compare the effect of AlphaFold on low- and high-experience researchers, I also split my sample into low experience and high experience groups, using a median split on the total number of publications before adoption. Tables 4 and 5 reports the descriptive statistics and correlation between variables for each group. On average, the high-experience group demonstrates higher productivity levels compared to their less experienced counterparts. Table 6 reports the regression results for the two groups. For the low experience group, the coefficient for the post-adoption period is positive and statistically significant ($\beta = 0.0162, SE = 0.00454$). For the high experience group, the coefficient is insignificant ($\beta = -0.00420, SE = 0.0104$), suggesting that AlphaFold benefits low-experience researchers but not highly experienced ones. Figure 1 plot the predicted monthly publication count for each group.

Similar results are reported in the appendix (Appendix Table 2) using JIF-weighted publication count as the dependent variable.

Robustness Checks

Transformation of variables

To account for potential multiplicative relationship between experience and productivity, and for the right-skewed distribution of some variables (the skewness of total publication count before adoption is 3.06, for example), I use a log transformation on the experience measures (total number of publications before adoption, months since first adoption), and an inverse hyperbolic sine transformation on the monthly publication count. The regression results are reported in Appendix Table 3. The results are similar, but with significantly positive coefficients for the post-adoption period.

Callaway & Sant'Anna (2021) estimator

To account for potential biases in estimating treatment effect parameters using two-way fixed effects regressions (De Chaisemartin & d'Haultfoeuille, 2020), I estimate the effects using the method provided by Callaway & Sant'Anna (2021) using the *csdid* command in STATA 18. I report the estimated average treatment effect on treated (ATT) on the monthly publication count and the JIF-weighted publication count, for the full sample, the low experience group, and the high experience group. While the results for the full sample are similar to the two-way fixed effects regression, the ATT estimated for the subgroups with different experience levels suggest no evidence of between-group difference.

Discussion

This paper explores the implication of AlphaFold as an AI research tool on improving research productivity. First, while the literature on AI and productivity finds a positive relationship between AI use and productivity, I find no significant increase in monthly publication counts after the adoption of AlphaFold among researchers overall. This could be because researchers face learning curves when incorporating the new tool into their workflows, or because users may decide to utilize the improved productivity afforded by AlphaFold not to publish faster but to improve the quality of the publication. The significantly positive increase in publication counts weighted by JIF provides partial support for the latter idea.

Second, the results show a higher correlation between AlphaFold adoption and research productivity for less experienced researchers than for highly experienced ones. Although the context of scientific research differs from the simple, standardized task environments usually studied in the AI-productivity literature, AI still seems to play an equalizing role in the context of scientific research. This could be for a number of reasons. First, it could be that the task portion that AlphaFold affects—generating hypotheses about promising candidates to run experiments on—is where skill and experience is most important, and the remaining tasks, such as running the experiments and interpreting the results, are not a significant source of variance in productivity. In this case, the leveling effect of AlphaFold on the narrow task of hypotheses generation will be preserved in the wider task of research overall. Or, since research experience would likely be positively correlated with researcher age, inexperienced researchers may come from a younger generation with more familiarity with new technologies such as AI, allowing them to better progress along the learning curve and benefit more from using AlphaFold. Lastly, it could be that research using AlphaFold is more likely to lead to a high-quality publication that grants the author recognition and access to resources. Such a growth in one's career would lead to increased productivity. However, the effect of such a boost could be larger for inexperienced early-stage researchers who are more likely to face resource constraints.

Examining and testing these speculations about the specific mechanisms through which AI affects research productivity would require supplementary data on the researchers, such as their demographic characteristics and educational backgrounds. Future directions would involve merging the OpenAlex data with researcher-level dataset such as LinkedIn profiles, ProQuest Dissertations & Theses, and grant data from the National Science Foundation. Using a richer dataset, it would be possible to tease out alternative explanations and also explore the impact of AI on researchers' career outcomes other than publications, such as grants, awards, and promotions.

It should be reiterated that the regression results are not to be interpreted as indicating any causal relationship between AlphaFold adoption and research productivity. The choice to adopt a new tool is endogenous and could be driven by many factors, such as motivation, technical knowledge, and access to complementary machinery. Thus, the results should be understood as reflecting a combination of the tool's impact and the characteristics of the researchers who choose to adopt it. Future studies would need to utilize a better identification of the causal link between AI use and research performance.

Another limitation of this study is the potential errors in inferring AlphaFold adoption by the publication of a paper that cites AlphaFold. A researcher may use the tool for research and not be able to publish anything from it, or he/she may not use the tool and still be on the author list of a paper that uses the tool. In the case of AlphaFold, researchers could be citing the AlphaFold paper for various reasons than using it for research: simply mentioning it as a background, building an AI model that improves on AlphaFold, etc. A more rigorous analysis should be taken to identify actual adoption and use of AlphaFold. For example, future studies could analyze the texts of the publications to determine if the tool was actually used in research. While this would be a data-intensive task involving sifting through a large number of publications, it could also potentially open up new areas of research, such as examining the different ways in which an AI tool is used in the research process and how they have different implications for the users.

Lastly, in the context of AI in scientific research, many interesting questions exist other than how AI affects productivity. For example, future work should also look at changes in the topics and research questions researchers pursue after adopting AI. With AI, would researchers explore a diverse set of topics using the new tool, or would they focus their research scope to exploit topics where AI can boost productivity the most? Another interesting topic could be on AI's effect on scientific collaboration networks. Does AI replace the need to collaborate with researchers with certain specializations? For example, after the public release of AlphaFold, did scientists specialized in the protein folding problem decrease in collaboration with other scientists? Or was their expertise in even more demand as more people were attracted to the topic? All of these are important research questions in understanding the direction and the dynamics of scientific progress in the age of AI.

References

Agrawal, A., McHale, J., & Oettl, A. (2024). Artificial intelligence and scientific discovery: A model of prioritized search. *Research Policy*, *53*(5), 104989.

Al-Janabi, A. (2022). Has DeepMind's AlphaFold solved the protein folding problem?.

Babina, T., Fedyk, A., He, A., & Hodson, J. (2024). Artificial intelligence, firm growth, and product innovation. *Journal of Financial Economics*, *151*, 103745.

Bhardwaj, A., Kishore, S., & Pandey, D. K. (2022). Artificial intelligence in biological sciences. *Life*, *12*(9), 1430.

Bianchini, S., Müller, M., & Pelletier, P. (2022). Artificial intelligence in science: An emerging general method of invention. *Research Policy*, *51*(10), 104604.

Bikard, M., Murray, F., & Gans, J. S. (2015). Exploring trade-offs in the organization of scientific work: Collaboration and scientific reward. *Management science*, *61*(7), 1473-1495.

Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). *Generative AI at work* (No. w31161). National Bureau of Economic Research.

Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of econometrics*, *225*(2), 200-230.

Carpenter, C. R., Cone, D. C., & Sarli, C. C. (2014). Using publication metrics to highlight academic productivity and research impact. *Academic emergency medicine*, *21*(10), 1160-1172.

Clarivate (2023). 2023 Journal Impact Factor, *Journal Citation Reports*

Cockburn, I. M., Henderson, R., & Stern, S. (2019). 4. The impact of artificial intelligence on innovation: An exploratory analysis. In *The economics of artificial intelligence* (pp. 115-148). University of Chicago Press.

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative science quarterly*, *35*(1), 128-152.

Czarnitzki, D., Fernández, G. P., & Rammer, C. (2023). Artificial intelligence and firm-level productivity. *Journal of Economic Behavior & Organization*, *211*, 188-205.

Damioli, G., Van Roy, V., & Vertesy, D. (2021). The impact of artificial intelligence on labor productivity. *Eurasian Business Review*, *11*, 1-25.

De Chaisemartin, C., & d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American economic review*, *110*(9), 2964-2996.

DeepMind. (2024, July 25). *AI achieves silver-medal standard solving International Mathematical Olympiad problems*. DeepMind. https://deepmind.google/discover/blog/ai-solves-imo-problems-at-silver-medal-level/

Dell'Acqua, F., McFowland III, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., ... & Lakhani, K. R. (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. Harvard Business School Technology & Operations Mgt. Unit Working Paper, (24-013).

Dill, K. A., & MacCallum, J. L. (2012). The protein-folding problem, 50 years on. *science*, *338*(6110), 1042-1046.

Ding, W. W., Levin, S. G., Stephan, P. E., & Winkler, A. E. (2010). The impact of information technology on academic scientists' productivity and collaboration patterns. *Management Science*, *56*(9), 1439-1461.

Doshi, A. R., & Hauser, O. (2023). Generative artificial intelligence enhances creativity. *Available at SSRN*.

Fleming, L., & Sorenson, O. (2004). Science as a map in technological search. *Strategic management journal*, *25*(8-9), 909-928.

Gest, H. (2004). The discovery of microorganisms by Robert Hooke and Antoni Van Leeuwenhoek, fellows of the Royal Society. *Notes and records of the Royal Society of London*, *58*(2), 187-201.

Griliches, Z. (1957). *Hybrid corn: An exploration in economics of technological change* (Doctoral dissertation, The University of Chicago).

Heather, J. M., & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics*, *107*(1), 1-8.

Jia, N., Luo, X., Fang, Z., & Liao, C. (2024). When and how artificial intelligence augments employee creativity. *Academy of Management Journal*, *67*(1), 5-32.

Johnson, P. C., Laurell, C., Ots, M., & Sandström, C. (2022). Digital innovation and the effects of artificial intelligence on firms' research and development–Automation or augmentation, exploration or exploitation?. *Technological Forecasting and Social Change*, *179*, 121636.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., ... & Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *nature*, *596*(7873), 583-589.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., ... & Hassabis, D. (2021). Applying and improving AlphaFold at CASP14. *Proteins: Structure, Function, and Bioinformatics*, *89*(12), 1711-1721.

Kanazawa, K., Kawaguchi, D., Shigeoka, H., & Watanabe, Y. (2022). *AI, skill, and productivity: The case of taxi drivers* (No. w30612). National Bureau of Economic Research.

Klein, E. (2023, July 11). Opinion | A.I. Could Solve Some of Humanity's Hardest Problems. It Already Has. *The New York Times*. https://www.nytimes.com/2023/07/11/opinion/ezra-klein-podcast-demis-hassabis.html

Liu, J., Chang, H., Forrest, J. Y. L., & Yang, B. (2020). Influence of artificial intelligence on technological innovation: Evidence from the panel data of china's manufacturing sectors. *Technological Forecasting and Social Change*, *158*, 120142.

Liu, B., He, H., Luo, H., Zhang, T., & Jiang, J. (2019). Artificial intelligence and big data facilitated targeted drug discovery. *Stroke and vascular neurology*, *4*(4).

McKiernan, E. C., Schimanski, L. A., Muñoz Nieves, C., Matthias, L., Niles, M. T., & Alperin, J. P. (2019). Use of the Journal Impact Factor in academic review, promotion, and tenure evaluations. *elife*, *8*, e47338.

Nickerson, J. A., & Zenger, T. R. (2004). A knowledge-based theory of the firm—The problem-solving perspective. *Organization science*, *15*(6), 617-632.

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, *381*(6654), 187-192.

Park, M., Leahey, E., & Funk, R. J. (2023). Papers and patents are becoming less disruptive over time. *Nature*, *613*(7942), 138-144.

Peng, S., Kalliamvakou, E., Cihon, P., & Demirer, M. (2023). The impact of ai on developer productivity: Evidence from github copilot. *arXiv preprint arXiv:2302.06590*.

Priem, J., Piwowar, H., & Orr, R. (2022). OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. *arXiv preprint arXiv:2205.01833*.

Rammer, C., Fernández, G. P., & Czarnitzki, D. (2022). Artificial intelligence and industrial innovation: Evidence from German firm-level data. *Research Policy*, *51*(7), 104555.

Seamans, R., & Raj, M. (2018). *AI, labor, productivity and the need for firm-level data* (No. w24239). National Bureau of Economic Research.

Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *Bmj*, *314*(7079), 497.

Sorenson, O., Rivkin, J. W., & Fleming, L. (2006). Complexity, networks and knowledge flow. *Research policy*, *35*(7), 994-1017.

Tregoning, J. (2018). How will you judge me if not by impact factor?. *Nature*, *558*(7710), 345-346.

Varadi, M., Bertoni, D., Magana, P., Paramval, U., Pidruchna, I., Radhakrishnan, M., ... &

Wang, H., Fu, T., Du, Y., Gao, W., Huang, K., Liu, Z., ... & Zitnik, M. (2023). Scientific discovery in the age of artificial intelligence. *Nature*, *620*(7972), 47-60.

Virshup, A. M., Contreras-García, J., Wipf, P., Yang, W., & Beratan, D. N. (2013). Stochastic voyages into uncharted chemical space produce a representative library of all possible drug-like compounds. *Journal of the American Chemical Society*, *135*(19), 7296-7303.

Wu, L., Wang, D., & Evans, J. A. (2019). Large teams develop and small teams disrupt science and technology. *Nature*, *566*(7744), 378-382.

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, *316*(5827), 1036-1039.

Table 1 Descriptive statistics and pairwise correlations between variables

	Mean	SD	Min	Max	(1)	(2)	(3)	(4)	(5)
(1) Monthly Publication Count	0.49	0.96	0.0	20.00					
(2) JIF-Weighted Monthly Publication Count	3.17	9.32	0.0	610.40	0.6497				
(3) Months since First Publication	244.15	139.39	0.0	609.00	0.1661	0.1107			
(4) Experience (Total Publications) before Adoption	35.82	41.35	1.0	505.00	0.5214	0.3533	0.3934		
(5) Experience (Month) before Adoption	244.15	139.19	12.0	597.0	0.1658	0.1099	0.9986	0.3940	
(6) Productivity (Publications per Month)	0.48	0.60	0.0	9.25	0.5778	0.3811	0.2736	0.8591	0.2739

Table 2 TWFE regression results (DV: Monthly Publication Count)

DV: Monthly Publication Count	(1)	(2)
Post Adoption	0.00430	0.00462
	(0.00558)	(0.00556)
Post × Total Publications before Adoption		-0.000864***
		(0.000151)
Months since First Publication	0.0176***	0.0176***
	(0.00175)	(0.00177)
Constant	0.752***	0.761***
	(0.0409)	(0.0412)
Publication Month Fixed Effects	Yes	Yes
Researcher Fixed Effects	Yes	Yes
Observations	386,136	386,136
Number of Units	16,089	16,089
R-squared	0.019	0.020

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3 TWFE regression results (DV: JIF-Weighted Monthly Publication Count)

DV: JIF-Weighted Monthly Publication Count	(1)	(2)
Post Adoption	0.191***	0.192***
	(0.0597)	(0.0595)
Post × Total Publications before Adoption		-0.00142
		(0.00169)
Months since First Publication	0.150***	0.150***
	(0.0201)	(0.0201)
Constant	4.976***	4.991***
	(0.475)	(0.476)
Publication Month Fixed Effects	Yes	Yes
Researcher Fixed Effects	Yes	Yes
Observations	386,136	386,136
Number of Units	16,089	16,089
R-squared	0.003	0.003

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4 Descriptive statistics and pairwise correlations between variables for the low experience group

Low Experience Group	Mean	SD	Min	Max	(1)	(2)	(3)	(4)	(5)
(1) Monthly Publication Count	0.19	0.49	0.0	20.00					
(2) JIF-Weighted Monthly Publication Count	1.31	5.36	0.0	502.70	0.5937				
(3) Months since First Publication	186.08	134.99	0.0	608.00	0.0226	0.0098			
(4) Experience (Total Publications) before Adoption	10.91	6.37	1.0	23.00	0.1556	0.0884	0.4027		
(5) Experience (Month) before Adoption	186.08	134.79	12.0	596.0	0.0204	0.0081	0.9985	0.4033	
(6) Productivity (Publications per Month)	0.17	0.16	0.0	1.5	0.2581	0.1387	0.0584	0.5017	0.0584

Table 5 Descriptive statistics and pairwise correlations between variables for the high experience group

High Experience Group	Mean	SD	Min	Max	(1)	(2)	(3)	(4)	(5)
(1) Monthly Publication Count	0.80	1.21	0.0	20.00					
(2) JIF-Weighted Monthly Publication Count	5.15	11.88	0.0	610.40	0.6374				
(3) Months since First Publication	305.75	115.50	7.0	609.00	0.0487	0.0380			
(4) Experience (Total Publications) before Adoption	62.25	46.09	24.0	505.00	0.4683	0.3215	0.2149		
(5) Experience (Month) before Adoption	305.75	115.26	19.0	597.0	0.0488	0.0374	0.9980	0.2153	
(6) Productivity (Publications per Month)	0.80	0.72	0.0	9.25	0.5416	0.3562	0.0858	0.8131	0.0860

Table 6 TWFE regression results by experience group (DV: Monthly Publication Count)

DV: Monthly Publication Count	(1) Low Experience	(2) High Experience
Post Adoption	0.0162***	-0.00420
	(0.00454)	(0.0104)
Months since First Publication	0.0218***	0.0207***
	(0.00149)	(0.00281)
Constant	0.642***	0.829**
	(0.0344)	(0.0644)
Publication Month Fixed Effects	Yes	Yes
Researcher Fixed Effects	Yes	Yes
Observations	198,768	187,368
Number of Units	8,282	7,807
R-squared	0.014	0.030

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

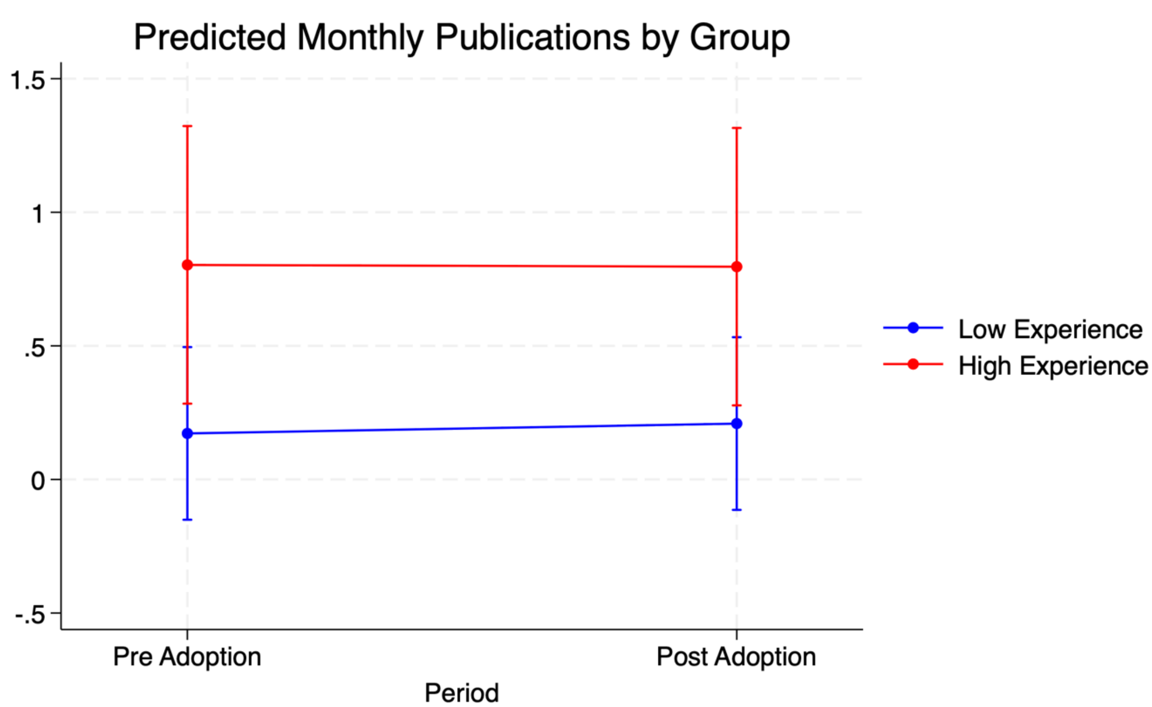

Figure 1 Predicted monthly publication count by group

Appendix

Appendix Table 1 TWFE results with alternative measures of experience/productivity

DV: Monthly Publication Count	(1)	(2)	(3)
Post Adoption	0.00430	0.00431	0.00541
	(0.00558)	(0.00558)	(0.00550)
Post × Months before Adoption		-9.09e-05***	
		(1.87e-05)	
Post × Productivity for 12 Months before Adoption			-0.159***
			(0.0109)
Months since First Publication	0.0176***	0.0176***	0.0174***
	(0.00175)	(0.00176)	(0.00176)
Constant	-3.553***	-3.543***	-3.466***
	(0.438)	(0.440)	(0.442)
Publication Month Fixed Effects	Yes	Yes	Yes
Researcher Fixed Effects	Yes	Yes	Yes
Observations	386,136	386,136	386,136
Number of Units	16,089	16,089	16,089
R-squared	0.019	0.019	0.023

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix Table 2 TWFE regression results by experience group (DV: JIF-Weighted Monthly Publication Count)

DV: JIF-Weighted Monthly Publication Count	(1) Low Experience	(2) High Experience
Post Adoption	0.217***	0.178
	(0.0499)	(0.111)
Months since First Publication	0.166***	0.210***
	(0.0208)	(0.0472)
Constant	4.537***	4.571***
	(0.515)	(1.046)
Publication Month Fixed Effects	Yes	Yes
Researcher Fixed Effects	Yes	Yes
Observations	198,768	187,368
Number of Units	8,282	7,807
R-squared	0.0047	0.0042

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix Table 3 TWFE regression results with IHS-transformed monthly publication count as DV and with logged covariates

DV: IHS-Transformed Monthly Publication Count	(1)	(2)
Post Adoption	0.00691**	0.00682**
	(0.00341)	(0.00341)
Post × Log Total Publications before Adoption		-0.0246***
		(0.00166)
Log Months since First Publication	0.0240***	-0.0464***
	(0.00798)	(0.00842)
Constant	0.214***	0.204***
	(0.0547)	(0.0555)
Publication Month Fixed Effects	Yes	Yes
Researcher Fixed Effects	Yes	Yes
Observations	386,136	386,136
Number of Units	16,089	16,089
R-squared	0.017	0.018

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix Table 4 Estimated ATT coefficients using Callaway & Sant'Anna (2021) estimator

DV	Sample	Coefficient	Standard Error
Monthly Publication Count	Full sample	0.0131216	0.0098061
	Low experience group	0.0068203	0.0087441
	High experience group	0.012506	0.0177423
JIF-Weighted Monthly Publication Count	Full sample	0.2801577**	0.1094086
	Low experience group	0.1776299**	0.0899598
	High experience group	0.3840123**	0.1915738

*** p<0.01, ** p<0.05, *p<0.1