AI-qualizing Science

Anantha Divakaruni,[1*] Francois Bares,[2] Ludovic Phalippou[2]

[1]Department of Economics, University of Bergen

[2]Saïd Business School, University of Oxford, UK

[*]Correspondence to: anantha.divakaruni@uib.no

February 10, 2025

Researchers face significant disparities in accessing resources for high-impact research. Artificial Intelligence (AI) promises to bridge these gaps by offering capabilities previously unavailable to many institutions. This paper examines the effects on protein research of AlphaFold, an AI tool that won the 2024 Nobel Prize in Chemistry for accurately predicting protein structures. Using comprehensive publication data, we show that AlphaFold benefits researchers at lower-ranked universities as their share of top-journal publications increases significantly following its release. These findings suggest that AI tools can lower barriers to entry in resource-intensive scientific fields and challenge established knowledge production hierarchies. AI can lead to a more equitable distribution of opportunities, with broader implications for innovation, scientific discovery, and research policy.

Introduction

Although several articles explore the influence of AI on low and medium-skill activities such as image recognition [1], programming [2], writing [3], and customer support [4], there is limited evidence on the impact of AI on productivity within highly specialized sectors [5]. Filling this gap is crucial due to AI rapidly advancing from automating relatively straightforward tasks (e.g., chatbots, code generators) to engaging in complex, high-skill areas that demand substantial expertise and research infrastructure.

The most profound impact of AI in a specialized field has arguably been observed recently within the life sciences. An AI tool called AlphaFold, developed by Google DeepMind, has achieved very high accuracy levels in predicting three-dimensional protein structures that closely match those of experimental methods [6–8]. As precise protein structures are crucial for understanding protein functions, designing therapeutics, and investigating biological pathways, AlphaFold is widely recognized as a significant advance in structural biology and the broader domain of protein research. Consequently, its creators were honored with the 2023 Breakthrough Prize in Life Sciences and the 2024 Nobel Prize in Chemistry.

In the domain of protein research, where groundbreaking discoveries often depend on access to state-of-the-art resources and specialized knowledge, the Matthew effect is particularly evident [9–12]. Before the release of AlphaFold, researchers in the top 10% universities published 55% articles in leading journals and secured 50% of research grants [13, 14]. In addition, top scientists often serve as editors of journals and grant reviewers, which allows them to favor research conducted by peers who use similar methods and are affiliated with similarly prestigious institutions [15–17]. Therefore, one may expect that new tools such as AlphaFold would either not be adopted at all or be adopted by the leading universities first, thus reinforcing existing hierarchies. Alternatively, by lowering technical and computational barriers to protein structure prediction, AlphaFold has the potential to broaden access to advanced research capabilities. This increased accessibility may allow researchers from lower-ranked institutions to contribute competitively to top journals, thereby altering the traditional stratification of scientific publication.

This research examines the extent to which AlphaFold has modified the pre-existing hierarchical structure among universities. Universities are categorized into quartiles according to their overall publication shares in the top journals prior to AlphaFold's release. By analyzing the changes in the top publications in each quartile, we evaluate whether AlphaFold's innovative features and open access have democratized high-impact research or perpetuated existing scientific inequalities.

We find that within five years of AlphaFold's introduction, universities in the bottom quartile (ranked between 137^{th} and 500^{th} before AlphaFold) exhibit a persistent increase in top publications. The effect is monotonic, with second and third quartile universities also showing similar gains, although to a lesser extent than bottom-quartile peers. In contrast, top-quartile universities experience a decline in their top publication market shares following AlphaFold's release. Overall, these results imply that AI can reduce entry barriers in resource-intensive scientific domains and can disrupt traditional knowledge production hierarchies, fostering a more balanced allocation of scientific opportunities.

Sample Construction

We examine the proportion of publications in leading journals for the top 500 universities over time. Publications are attributed to universities on the basis of the institutional affiliation of the authors. To identify top journals in protein research, we compile a list of 2,170 "biochemistry, genetics, and molecular biology" journals and 175 interdisciplinary journals publishing articles related to protein research. We rank the journals by their average SCImago Journal Rank (SJR) – a journal reputation metric that accounts for citation counts of articles published in a given journal and the prestige of the journals where these citations appear. To account for annual fluctuations in SJR, we use the average SJR over the six-year period (2013–18) preceding AlphaFold's release. There are 17 journals with an average SJR of 10 or above (see Table S2 for details), and we classify these journals as top journals.

We obtain data on all articles published in these 17 journals from OpenAlex (the successor to Microsoft Academic Graph). OpenAlex provides detailed article metadata including titles, abstracts, covered topics, citations, author affiliations, and the publishing journal. Each article

is classified into one of three categories:

- *Structural biology*: if at least one author is a structural biologist or the paper covers any of 238 identified structural biology topics.
- *Non-structural protein*: if none of the authors is a structural biologist and the paper's topics fall under one or more of 1,066 protein research topics outside structural biology.
- *Non-protein*: for all other publications.

We rank universities based on their average total normalized citation scores (TNCS) between 2013 and 2018 in the category "biomedical and health sciences" (computed by Leiden University). The score for university u is defined as:

$$Score_u = \frac{1}{6} \sum_{t=2013}^{2018} TNCS_{u,t} \tag{1}$$

Where $TNCS_{u,t}$ denotes the total normalized citation scores in year t. The top 12 universities are Harvard, Johns Hopkins, UCSF, Toronto, UPenn, Washington, UCL, Stanford, Duke, Michigan, UCLA and Oxford (Table S4).

Each author is assigned to a single university (highest ranked university for authors with multiple affiliations). The research output of each university u in year t is measured as the annual count of articles published in top journals where at least one author is affiliated with that university. We denote this output $P^{all}_{u,t}$ when we take into account all authors listed in a publication, and $P^{main}_{u,t}$ when we only account for the two main authors (listed first and last)[1]. The corresponding shares of top journal publications of university u in year t, which we refer to as "market share" is computed as follows:

$$MktS^{all}_{u,t} = \frac{P^{all}_{u,t}}{\sum_{i=1}^{500} P^{all}_{i,t}} \tag{2}$$

$$MktS^{main}_{u,t} = \frac{P^{main}_{u,t}}{\sum_{i=1}^{500} P^{main}_{i,t}} \tag{3}$$

[1] See section S2.2 for a formal definition of $P^{all}_{u,t}$ and $P^{main}_{u,t}$

The cumulative density function (CDF) of $MktS_{u,t}^{all}$ is quasi-monotone and concave (Figure 1). The top 10 universities generate 18% of all the top publications. The top 50 generate nearly half of all publications. The top 150 universities produce almost three times as many top publications as the next 350 universities (Figure 1(A)). Figure 1(B) shows that the distribution is similar if we only take into account the main authors (first and last authors) of each publication.

We classify universities by their market share and divide them into quartiles, with each quartile accounting for 25% of the top publications from 2013 to 2018. The top quartile includes 12 universities, the second quartile contains the subsequent 36 universities, the third quartile includes 88 universities, and the bottom quartile comprises the remaining 364 universities.

Publication rates around the AlphaFold shock

Changes in Aggregate Market Shares

We plot a time series of the aggregate market shares of each university quartile in Figure 2. In both panels 2(A) (structural biology publications) and 2(B) (non-structural biology, but within protein research), each quartile maintains a constant market share before the release of AlphaFold. The graphs also highlight three key milestones in the development of AlphaFold: (i) in 2018, AlphaFold 1 (AF1) wins the 13^{th} Critical Assessment of Protein Structure Prediction Competition (CASP); (ii) in 2020, AlphaFold 2 (AF2) wins CASP 14; and (iii) in 2021, millions of AlphaFold 2 predicted protein structures are made publicly available. Section S1.3 provides further details.

We observe a notable shift within structural biology publications after the release of AF1 (Figure 2(A)). Universities in the bottom quartile expanded their market share at the expense of those in the top two quartiles. In just two years, the market share of bottom-quartile universities rose from slightly above 24% to nearly 29%. This upward trend persisted with the release of AF2, and by 2023, the market share of the bottom quartile reached 30%, while the market share of the top two quartiles both fell to around 22%. A similar picture emerges when we compute market share only taking into account main -first and last- authors (Equation 3). Universities in the bottom quartile increased their share from 25% to over 30%. Meanwhile, the share of the

top quartile dropped from 25% to 20% (Figures 2(C) and 2(D)).

The pattern observed in non-structural protein publications is also remarkable (panel 2(B)). No changes occur after AF1, but post AF2, there is a jump in top publications from institutions in the bottom quartile. The market share of universities in the bottom quartile increased from 26% to 30% in just two years. A similar, but less significant, increase is observed for universities in the third quartile. In contrast, institutions in the top two quartiles experienced a considerable reduction in market share, with the top quartile experiencing the largest decline.

We perform two sets of counterfactual analysis. We examine aggregate publication trends for i) publications in protein research journals with an average SJR between 1 and 10; ii) publications in top-journals that do not focus on protein research. If the changes in the distribution of market shares observed in Figure 2 are driven by factors not related to AlphaFold, we would observe similar trends outside the top journals (results shown in Figure S2) or outside protein research (results shown in Figure S3). There is no pre-trend in either figure, and there is no effect post-AlphaFold either. The market shares remain the same in the period around AlphaFold's release, showing that AlphaFold has only impacted protein research in the top journals.

Regression Analysis

To formally test whether the changes in market shares observed in Figure 2 can be attributed to the introduction of AlphaFold, we employ a two-way fixed effects regression framework. Our analysis is conducted at the university level using a yearly panel of universities from 2013 to 2023. Each university is assigned to its pre-AlphaFold quartile as in Figure 2. This quartile-based approach allows us to estimate the average post-AlphaFold effect within each group, capturing shifts in publication dynamics while accounting for both time-invariant university characteristics and common yearly shocks. Specifically, for each university u in year t, we estimate the following equation:

$$y_{ut} = \beta_0 + \alpha_u + \gamma_t + \sum_{q \in \{2,3,4\}} Quartile_{q,u} * PostAF1_t * \beta_q + \mathbf{X}'_{ut}\delta + \epsilon_{ut} \tag{4}$$

where $Quartile_q$ is a binary variable equal to 1 if university u belongs to the quartile u (zero

otherwise). $PostAF1$ is a binary variable set to 1 for all years from 2019 onward, capturing the period after the introduction of AlphaFold. The vector $\mathbf{X}_{u,t}$ includes university-level control variables, with δ representing the corresponding coefficients. Fixed effects α_u control for time-invariant university characteristics, while γ_t capture common shocks affecting all universities in a given year.

We estimate the impact of AlphaFold on university research output using three variations of the outcome variable y_{ut}. Each variation captures university publication activity in a different way. First, we define y_{ut} as the number of top-journal publications (in log) by all authors at university u in year t, to assess absolute changes in the number of publications. Second, we assess relative changes by defining y_{ut} as the university's annual market share using equations 2 and 3, respectively.

We first estimate these models on the full sample of protein research publications. Panel 1(A) of Table 1 shows estimates for the three outcome variables. In each of these specifications, the interaction terms for the second, third, and fourth quartiles (relative to the top quartile) are positive and statistically significant at the 1% level. Ordinary least squares (OLS) estimates in model (1) indicate that the average fourth quartile university nearly doubled its number of yearly publications post–AlphaFold (increase by $e^{0.47} - 1 = 0.60$ from a 0.80 average). The effect is monotonic. The second and third quartiles exhibit the same upward trend, but with smaller magnitudes.

As university market shares are bounded between 0 and 1, models (2) and (3) use a generalized linear model with a logit link, which is essentially a logistic regression. This method transforms the probability of an event occurring into its log-odds ratio before relating it to a linear combination of predictor variables. Model (2) shows that following AlphaFold, fourth quartile universities show an average increase of 0.21 in the log-odds of the market share for top protein publications compared to top quartile peers. Similarly, second and third quartile universities experience an increase in log-odds, though with smaller magnitudes. Finally, estimates in model (3) confirm that our results hold even when market shares are based solely on the main authors.

In addition, we separately analyze publications on structural and non-structural proteins (panels

1(B) and 1(C)). The AlphaFold effect is equally strong in both structural and non-structural protein research, suggesting that its influence extends beyond structural biology.

Discussion

The concentration of scientific research in elite institutions has long been attributed to differences in resources, infrastructure, and institutional advantages rather than differences in innate ability. Our findings provide evidence that AI can disrupt these entrenched disparities. The introduction of AlphaFold represents an exogenous technological shock that altered the competitive landscape of protein research, particularly in structural biology, by lowering the expertise and resource thresholds required for high-quality research in this field. The resulting shifts in publication outcomes indicate that AI can serve as a scientific equalizer, enabling researchers from lower-ranked universities to gain a greater foothold in high-impact journals.

These results align with previous studies on technological shocks in science, showing that when barriers to entry are lowered, whether through reduced experimental costs [18] or enhanced access to computational tools [19], the institutional distribution of research can change. However, while previous disruptions were often gradual and temporary, AlphaFold's impact has been rapid and persistent, highlighting the unique role of AI in accelerating changes in research.

Our findings contribute to long-standing debates about the Matthew effect in science. Although elite universities have historically maintained their dominance by leveraging cumulative advantages in funding, network effects, and editorial influence, our results suggest that AI can counteract these forces. Crucially, the open access features of AlphaFold distinguishes it from previous scientific innovations. Whether future AI breakthroughs will maintain this open-access model or shift toward proprietary restrictions remains an open question, with implications for the long-term trajectory of scientific equity.

Although AlphaFold reduces the cost of structure prediction, many downstream applications (experimental validation and drug development) remain resource intensive and therefore concentrated in elite institutions. Moreover, the widespread adoption of AlphaFold raises new concerns about the centralization of knowledge production in AI-driven research. If the de-

velopment of AI-driven scientific tools remains concentrated within a few private entities such as Google DeepMind, which developed AlphaFold – future iterations of these models may become increasingly reliant on proprietary datasets, computationally intensive architectures, or restricted access policies. This could create new disparities in research capabilities, favoring institutions with the financial and technical resources to engage with these rapidly evolving technologies, potentially reversing some of the democratizing effects observed in this study.

Future research should investigate whether the AF shock translates into a lasting reconfiguration of scientific leadership or whether these effects are transient. More work is also needed to assess whether AI-induced research gains extend beyond publication success to broader metrics of scientific influence. As AI continues to reshape scientific discovery, understanding its role in reinforcing or dismantling institutional hierarchies will be critical for the future of research policy and innovation.

Acknowledgments We are grateful to Savvas Savvides, Rob van der Kant, Emiel Michiels, Lucrezia Vittoria Viti, and Hans Hvide for their invaluable feedback and support. We also thank seminar participants at the University of Bergen.

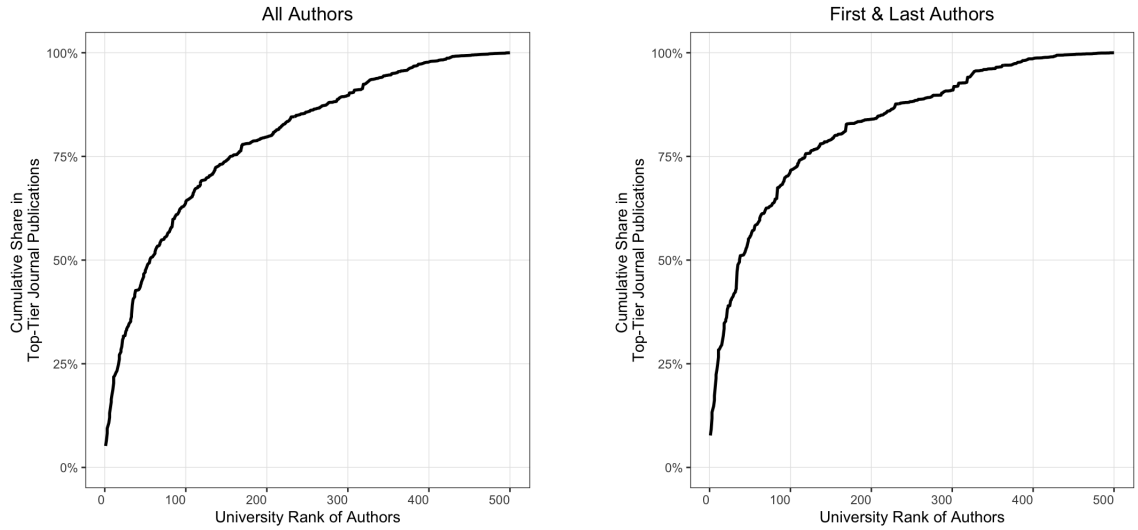
References

1. Andrew Caplin et al. *The ABC's of Who Benefits from Working with AI: Ability, Beliefs, and Calibration*. Working Paper. Oct. 2024.
2. Zheyuan (Kevin) Cui et al. *The Effects of Generative AI on High Skilled Work: Evidence from Three Field Experiments with Software Developers*. en. SSRN Scholarly Paper. Rochester, NY, Sept. 2024.
3. Shakked Noy and Whitney Zhang. "Experimental evidence on the productivity effects of generative artificial intelligence". In: *Science* 381.6654 (July 2023), pp. 187–192.
4. Erik Brynjolfsson, Danielle Li, and Lindsey R. Raymond. *Generative AI at Work*. Working Paper. Apr. 2023.
5. Aidan Toner-Rodgers. *Artificial Intelligence, Scientific Discovery, and Product Innovation*. Dec. 2024.
6. Mohammed AlQuraishi. "AlphaFold at CASP13". en. In: *Bioinformatics* 35.22 (Nov. 2019). Ed. by Alfonso Valencia, pp. 4862–4865.
7. John Jumper et al. "Highly accurate protein structure prediction with AlphaFold". In: *Nature* 596.7873 (2021), pp. 583–589.
8. Josh Abramson et al. "Accurate structure prediction of biomolecular interactions with AlphaFold 3". en. In: *Nature* (May 2024).
9. Robert K. Merton. "The Matthew Effect in Science". In: *Science* 159.3810 (Jan. 1968), pp. 56–63.
10. Pierre Azoulay, Toby Stuart, and Yanbo Wang. "Matthew: Effect or Fable?" In: *Management Science* 60.1 (Jan. 2014), pp. 92–109.
11. Thijs Bol, Mathijs de Vaan, and Arnout van de Rijt. "The Matthew effect in science funding". In: *Proceedings of the National Academy of Sciences* 115.19 (May 2018), pp. 4887–4890.
12. Jonathan Brogaard et al. "On the Causal Effect of Fame on Citations". In: *Management Science* (Dec. 2023).
13. Emil Bargmann Madsen and Kaare Aagaard. "Concentration of Danish research funding on individual researchers and research topics: Patterns and potential drivers". In: *Quantitative Science Studies* 1.3 (Aug. 2020), pp. 1159–1181.
14. Athen Ma, Raúl J. Mondragón, and Vito Latora. "Anatomy of funded research in science". In: *Proceedings of the National Academy of Sciences* 112.48 (Dec. 2015), pp. 14760–14765.
15. Pierre Azoulay, Christian Fons-Rosen, and Joshua S. Graff Zivin. "Does Science Advance One Funeral at a Time?" en. In: *American Economic Review* 109.8 (Aug. 2019), pp. 2889–2920.
16. Joshua M. Nicholson and John P. A. Ioannidis. "Conform and be funded". en. In: *Nature* 492.7427 (Dec. 2012), pp. 34–36.
17. Mikko Packalen and Jay Bhattacharya. *Does the NIH Fund Edge Science?* Working Paper. July 2018.

18. Fiona Murray et al. “Of Mice and Academics: Examining the Effect of Openness on Innovation”. en. In: *American Economic Journal: Economic Policy* 8.1 (Feb. 2016), pp. 212–252.

19. Jeffrey L. Furman and Florenta Teodoridis. “Automation, Research Technology, and Researchers’ Trajectories: Evidence from Computer Science and Electrical Engineering”. In: *Organization Science* 31.2 (Mar. 2020), pp. 330–354.

Figure 1: **Market shares of top 500 universities before release of AlphaFold**
University rankings are determined by the average *total normalized citation score* (TNCS) of each university in the "biomedical and health sciences" category between 2013 and 2018.



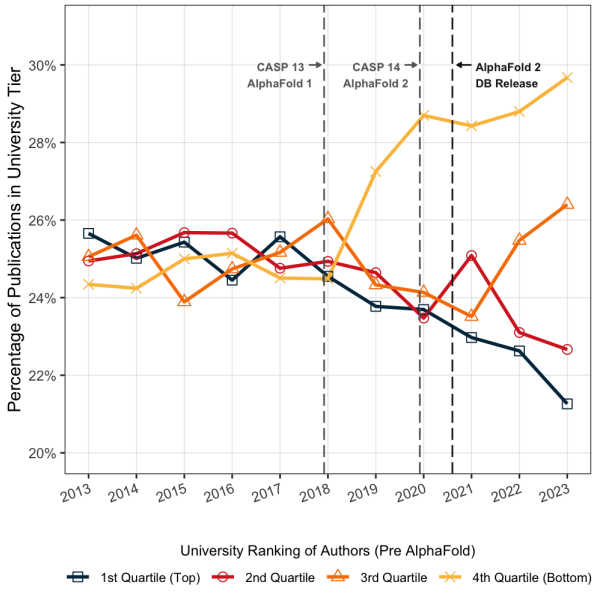
(A) Counting all authors

(B) Counting only first and last author

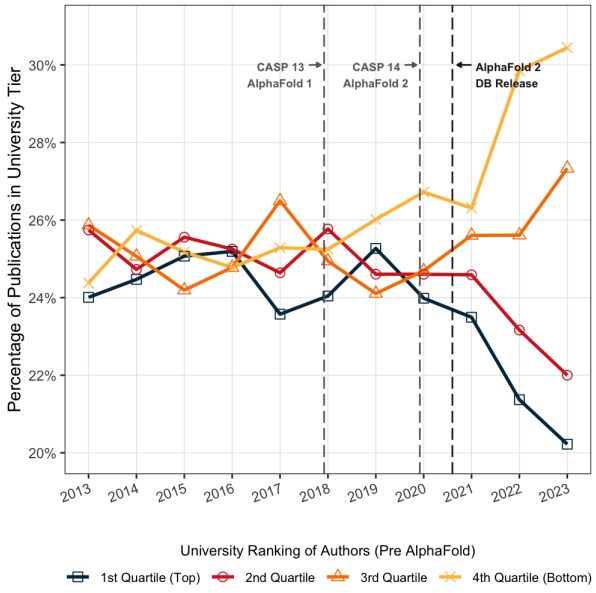
Figure 2: **Protein publications in top journals by university quartiles**

Figure illustrates the percentage of protein research publications in top journals (with an SJR $\geq$ 10 between 2013 and 2018, before the release of AlphaFold 1). The analysis is limited to authors affiliated with the top 500 universities, which are divided into four quartiles based on their share of publications in these top journals during the same period. Panel A and B show market share computed with all authors and panel C and D show market share computed with main authors only.

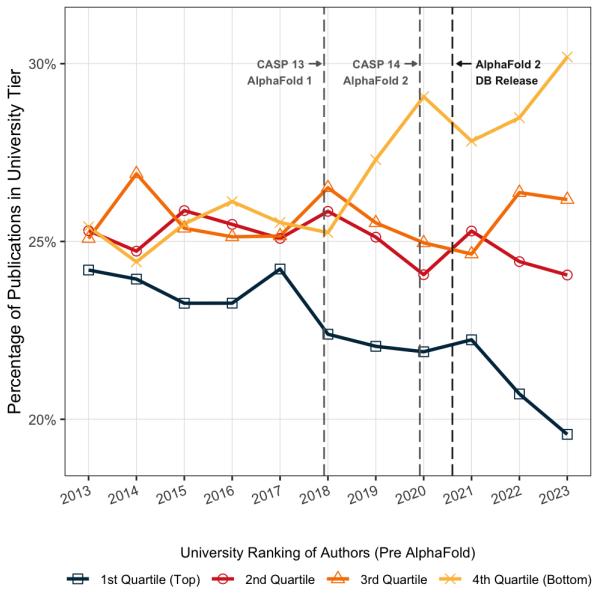
(A) Structural biology- All Authors



(B) Non-structural protein - All Authors



(C) Structural Biology - Main Authors



(D) Non-Structural Protein - Main Authors

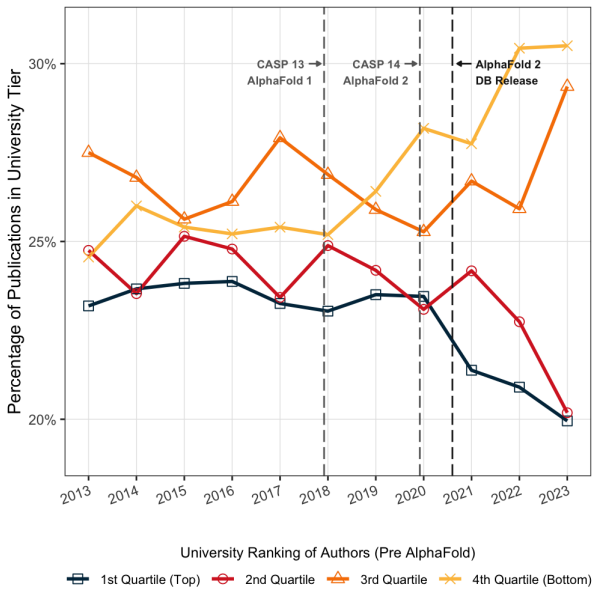


Table 1: **Protein publications in top journals around AlphaFold**

PostAF1 is a binary variable set to one for papers published after the release of AlphaFold V1 in December 2018. The sample is limited to all protein research papers (both structural and non structural biology) published in top journals with an SJR score of 10 or higher and to authors affiliated with the top 500 universities prior to the release of AlphaFold 1. Universities are ranked based on their average TNCS in the period preceding the release of AlphaFold 1. The highest ranked universities placed in the top quartile (*University Tier = 1st Quartile*) are used as the reference group. Standard errors are double-clustered by university and year, and reported in parantheses. ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively.

(A) All protein publications

Dependent Variables:	Log(1 + No. of Publications)	Market Share (% Publications, All Authors)	Market Share (% Publications, Main Authors)
Model:	(1)	(2)	(3)
University Tier = 2nd Quartile × PostAF1	0.19***	0.06***	0.10***
	(0.02)	(0.00)	(0.00)
University Tier = 3rd Quartile × PostAF1	0.31***	0.13**	0.14**
	(0.03)	(0.02)	(0.02)
University Tier = 4th Quartile × PostAF1	0.47***	0.21***	0.23***
	(0.04)	(0.02)	(0.02)
Controls	✓	✓	✓
Year FE	✓	✓	✓
University FE	✓	✓	✓
Observations	5,500	5,500	5,500
Adjusted/Pseudo R^2	0.82	0.92	0.93

(B) Structural biology publications

Dependent Variables:	Log(No. of Publications)	Market Share (% Publications, All Authors)	Market Share (% Publications, Main Authors)
Model:	(1)	(2)	(3)
University Tier = 2nd Quartile × PostAF1	0.23***	0.13***	0.15***
	(0.02)	(0.00)	(0.00)
University Tier = 3rd Quartile × PostAF1	0.36***	0.16***	0.16***
	(0.03)	(0.02)	(0.01)
University Tier = 4th Quartile × PostAF1	0.54***	0.25***	0.24***
	(0.04)	(0.03)	(0.02)
Controls	✓	✓	✓
Year FE	✓	✓	✓
University FE	✓	✓	✓
Observations	5,500	5,500	5,500
Adjusted/Pseudo R^2	0.81	0.91	0.92

(C) Non-structural protein publications

Dependent Variables:	Log(No. of Publications)	Market Share (% Publications, All Authors)	Market Share (% Publications, Main Authors)
Model:	(1)	(2)	(3)
University Tier = 2nd Quartile × PostAF1	0.28***	0.09***	0.12***
	(0.03)	(0.01)	(0.01)
University Tier = 3rd Quartile × PostAF1	0.40***	0.17***	0.17***
	(0.03)	(0.03)	(0.03)
University Tier = 4th Quartile × PostAF1	0.54***	0.25***	0.30***
	(0.03)	(0.02)	(0.03)
Controls	✓	✓	✓
Year FE	✓	✓	✓
University FE	✓	✓	✓
Observations	5,500	5,500	5,500
Adjusted/Pseudo R^2	0.79	0.90	0.91

Supplementary Information for

AI-qualizing Science

Anantha Divakaruni,[1*] Francois Bares,[2] Ludovic Phalippou[2]

[1]Department of Economics, University of Bergen
[2]Saïd Business School, University of Oxford, UK

*Correspondence to: anantha.divakaruni@uib.no.

S1 Literature & Background

S1.1 Market Concentration in Science

Scientific research is highly concentrated, with the top 1% of scientists obtaining 20% of the citations [1] and 50% of funding [2, 3]. The pros and cons of this concentration are actively debated [4]. Advocates for a more even distribution argue that concentrating funds stifles scientists' creativity.For example, superstar scientists may act as gatekeepers and hinder the emergence of new paradigms [5]. Similarly, in grant reviews, it has been found that homophily among senior scientists can penalize the more novel research proposals [6–8]. In contrast, critics point to economies of scale and efficient resource use by more productive scientists.

Research in this area has focused on explaining this concentration. A prevailing explanation is that scientific concentration arises partly because of the winner take all reward system in science: renowned scientists receive a disproportionate share of funding and credit for their research. This boosts their productivity and fame, thus further entrenching their position at the top. A phenomenon known as the "Matthew effect" [9]. This theory has been empirically tested in various contexts [3, 10, 11].

Related to our work are papers examining the effect of external shocks on competition in science. The hiring of stars in lower-ranked institutions boosts the productivity of future recruits without affecting existing researchers [12]. The deaths of stars' scientists negatively impact their collaborators but benefit others in the same field [5, 13]. The influx of Soviet mathematicians decreased career opportunities for prominent American mathematicians post-USSR [14]. Bitnet improved collaboration and publication results for mid-tier researchers located near top scientists [15]. The introduction of CRISPR, a novel gene editing tool, first improved the prospect of institutions that were already at the cutting edge of genetic research [16]. The free availability of Microsoft's Kinect democratized motion sensing technology, increasing the competitiveness of outsider researchers and the diversity of ideas [17]. The reduced costs for accessing genetically modified mice led to new researchers entering the field and to more diverse research being produced [18].

S1.2 Structural Biology

Structural biology is a branch of protein research, alongside genomics, drug discovery etc. Its primary objective is to identify the three-dimensional shapes of proteins. Knowing the structure of proteins is key to understand how they interact with various molecules [19] and is fundamental for drug development [20]. Before 2018, most protein structures were determined by X-ray crystallography [21]. The cost of this technique varies between $250,000 for soluble human proteins (25% success rate) and $1,000,000 for membrane proteins (10% success rate) [22].

Proteins are crystallized by preparing a concentrated protein solution to induce crystal formation. This crystallization process is complex and highly variable, often involving extensive trial and error with conditions such as temperature, pH, and specific chemical additives [23]. Once suitable crystals are obtained, they are exposed to high-intensity X-ray beams at a synchrotron, which produces a diffraction pattern that reveals the atomic arrangement. The diffraction data are transformed into an electron density map that shows the atomic positions within the protein. The scientists then test and refine their model to align with the data. This iterative process can take months, requiring multiple adjustments to accurately capture structural details [24].

An alternative to X-ray crystallography is cryo-electron microscopy (cryo-EM). Cryo-EM freezes proteins at cryogenic temperatures, maintaining their natural structures. An electron beam captures 2D images that are computationally assembled into a 3D model. Cryo-EM is popular for resolving large, complex, and membrane protein structures [25, 26].

S1.3 The Advent of AlphaFold

AlphaFold (AF) was developed by London-based DeepMind AI lab (part of Google). Their first model used a neural network trained on experimentally determined 3D structures of proteins from the Protein Data Bank. The DeepMind team was not the first to attempt protein structure prediction using deep learning but was the first to succeed. They used a much larger, and more computationally efficient neural network than previously [27]. AF processed amino acid relationships and inferred structural data faster and more accurately than traditional methods [28], and won the 13^{th} Critical Assessment of Structure Prediction (CASP) competition.

The results of CASP13 validated the deep learning approach after years of limited success. It sparked the development of various deep learning models to predict structures. The associated research paper was subsequently published in the journal Nature in January 2020.

AF precision was still insufficient for reliable predictions of complex proteins, and in 2020, AlphaFold 2 (AF2) was introduced. It combined evolutionary data, structural templates, and a revised deep learning infrastructure that featured an attention mechanism. Compared to AF1, AF2 better understood the spatial dependencies between amino acids and achieved precision comparable to experimental methods [29]. On 30 November 2020, it won the CASP 14 competition by a record margin.[2]

On 15 July 2021, the paper introducing AlphaFold 2 was published in Nature as an advance access article [30]. The publication came with an open source software and a searchable database with predicted structures for more than 214 million proteins, covering most known natural sequences, including the entire human proteome. In contrast, pre-AF, only 17% of the human protein structures were known [31]. The paper received more than 20,000 citations in two years, and its main authors won the 2024 Nobel Prize in Chemistry.

Since the release of AF2, several deep learning tools emerged to address structural protein challenges (Table S1 provides examples), and on 8 May 2024, DeepMind announced the release of an enhanced version (AlphaFold 3). AF3 is capable of predicting highly complex protein structures and protein-protein interactions. The development of AlphaFold and other models is summarized in figure S1.

S1.4 Role of AlphaFold in Structural Biology

AlphaFold can serve both as a complement and a substitute to experimental methods in structural biology. AlphaFold predictions can be used either to bypass experimental work entirely or

[2] AF2 scored above 90 for around two-thirds of the proteins. This score represents the degree to which a predicted structure (such as using a computational tool like AlphaFold) is similar to the experimentally determined structure, with 100 denoting a complete match between the two. AF2 achieved a world record-breaking overall score of 92.4, representing exceptional precision in protein structure prediction See https://www.guinnessworldrecords.com/world-records/642132-highest-score-at-the-casp-competition.

to automate some aspect of the experimental process.

As a substitute, AlphaFold enables researchers to obtain predicted structures without costly experiments [31].The availability of predicted structures surged with the AF2 database, allowing easy access to structures without needing to run a model like AlphaFold. However, exclusive reliance on AlphaFold without experimental validation may jeopardize accuracy. AlphaFold 2 can encounter difficulties with novel or large protein complexes in particular, leading to inaccurate predictions even when prediction confidence intervals are low [32].

As a complement, AlphaFold makes experimental structure discovery faster and cheaper. AlphaFold predictions provide more complete starting hypotheses to build models that better fit the experimental data. AlphaFold predictions also enable faster and cheaper trial and error over large sets of potential structures. More iterations lead to finer interpretations of experimental data, thus improving structure determination [33].

AlphaFold does not only improve productivity for existing tasks, it also enables the study of larger protein systems. For instance, the nuclear pore complex, crucial for nuclear transport, presented considerable modeling challenges. A detailed structure was constructed using various advanced experimental techniques to integrate individual segments into a unified model [34]. Subsequently, this model was refined using predictions derived from AlphaFold[35]. Such advanced models would be unattainable without AlphaFold.

S2 Data and Sample Construction

Our main data source is OpenAlex. OpenAlex was developed by OurResearch, a nonprofit organization, to facilitate access to the Microsoft Academic Graph after Microsoft discontinued the project in 2021.[3] Each publication is assigned a unique identifier, and its metadata include the title of the article, the abstract, citation counts, the list of authors with their institutional affiliations and departmental affiliation, and the range of topics covered.

[3]Microsoft Academic Graph is less comprehensive than Google Scholar but more comprehensive than other providers such as Scopus or Web of Science [36]. Unlike Google Scholar, however, Microsoft Academic Graph enables large-scale downloads, and OpenAlex offers API access. The OpenAlex platform aggregates data from Crossref, ORCID, and PubMed, as well as open-access research repositories such as arXiv and Zenodo. OpenAlex also covers papers released as preprints.

S2.1 Journal Selection

We select journals in 'biochemistry, genetics, molecular biology' (2,170 journals) and 'Multidisciplinary' (175 journals), then label those with an average SJR greater than 10 as 'top journals' (17 in total, listed in Table S2). We split the articles published in these journals into three groups labeled as follows: i) 'Structural biology', ii) 'Non-structural protein', and iii) 'Non-protein.' A paper is classified as structural biology if it has an author who is a structural biologist or if the paper list of topics includes one of the 238 topics we identify as structural biology. If a paper does not satisfy this condition but has one of the 1,066 topics of protein research, then it falls in the second group. All other papers are assigned to the third group. Figure S4 shows category trends over time. In 2013, "non-protein" research accounted for 50% of publications, with protein-related studies split equally (25% each). By 2024, protein-related research dropped to 20% (10% each for structural and non-structural), while non-protein research grew. Overall publications declined from 4,000 in 2013 to 2,800 in 2024, independent of COVID-19 research.

We measure the quality of the journal in which the research is published with the SCImago Journal Rank (SJR). SCImago is an online platform that provides a comprehensive list of academic journals in various disciplines, together with statistics such as citation counts and the proportion of female authors. To generate its SJR, SCImago employs an algorithm that uses both the number of citations of articles published in a given journal and the prestige of the journals where the citations appear. As the SJR of each journal fluctuates from year to year, we rank the journals using the average SJR during the six-year period preceding the release of AlphaFold (i.e., 2013–2018).

S2.2 University Ranking

The Centre for Science and Technology Studies (CWTS) at Leiden University produces the *Total Normalized Citation Score* (TNCS) of about one thousand universities around the world. We compute the average TNCS score of each university between 2013 and 2018, in the category "biomedical and health sciences". TNCS is computed as follows. First, each publication is assigned to a biomedical subfield. Then, the citation count for each publication is compared to

the average in the subfield in that year, resulting in a normalized citation score, where 1 represents the overall average. The Mean Normalized Citation Score (MNCS) is then calculated by averaging these normalized scores across all the publications in the subfield. Finally, the TNCS is obtained by multiplying the MNCS by the total number of publications, providing a size-dependent indicator of the university's overall citations. This method thus accounts for differences in citation practices between subfields and across years.We match authors to universities based on their affiliation. In our data some authors are not affiliated with universities, but instead with research centers, private companies or hospitals. These researchers are dropped from the data.

We compute the yearly publication count of university u as follows:

Let

$$\mathscr{P}_t = \{i : \text{article } i \text{ is published in a top journal in year } t\},$$

and for each article i, let

$$A_i = \{a_{i,1}, a_{i,2}, \ldots, a_{i,n_i}\}$$

denote its ordered set of authors. Denote by $u(a)$ the university assigned to author a (i.e., the highest ranked among any multiple affiliations).

Then, the publication counts are defined as follows:

1. All authors count:

$$P_{u,t}^{all} = \sum_{i \in \mathscr{P}_t} \mathbf{1}\Big\{\exists\, k \in \{1, \ldots, n_i\} : u(a_{i,k}) = u\Big\},$$

2. Main authors count:

$$P_{u,t}^{main} = \sum_{i \in \mathscr{P}_t} \mathbf{1}\Big\{u(a_{i,1}) = u \quad \text{or} \quad u(a_{i,n_i}) = u\Big\}.$$

These definitions ensure that each article contributes at most one count per university, regardless of the number of affiliated authors.

For robustness, we also construct a measure of market share that only takes into account the two main contributors on a publication (first and last author in the author list).

$$MktS_{u,t}^{main} = \frac{P_{u,t}^{main}}{\sum_{i=1}^{500} P_{i,t}^{main}} \tag{5}$$

Market concentration is slightly stronger when we only account for the main authors. The top 50 universities then generate more than half of all top publications.

S2.3 Additional Data

High-performance computing (HPC) plays a pivotal role in modern protein research, enabling complex simulations, large-scale data analyses, and computationally intensive methods such as those employed by AlphaFold. Access to substantial HPC resources is therefore essential for advancing the field, as it directly supports the development and application of sophisticated computational tools in structural biology. To empirically assess the influence of HPC on protein research outputs, we use data from Top500.org as a proxy for the computational power available at academic institutions. Since 1993, Top500.org has systematically tracked HPC facilities at universities worldwide on a biannual basis, providing detailed metrics on total processing power (quantified by the number of processor cores) and computational speed (measured in petaflops per second). This dataset allows us to assess the link between a university's HPC capabilities and its structural biology output. We compare research volume and quality against available computational power over time. This analysis reveals whether advanced HPC drives high-quality protein research, especially in studies requiring structural biology expertise.

S2.4 Descriptive statistics

A range of descriptive statistics is presented in table S3. They highlight significant disparities in research production across university quartiles prior to AlphaFold. Universities in the first quartile produce more in both structural and nonstructural protein research. They average 8.58 publications on structural biology (per university annually), compared to 2.39, 1.75, and 0.8 for the second, third, and fourth quartiles, respectively. The numbers are similar in non-structural protein research. It follows that first-quartile universities also hold a larger market share, both in terms of the share of publication and the share of authorship. The average first quartile university produces 0.37% of all the top structural biology publications, in contrast to 0.1%,

0.08% and 0.03% for the second, third and fourth quartile, respectively. The numbers are similar for non-structural protein publication and authorship share.

Table S3 indicates that top-quartile universities have more resources at their disposal. The average faculty size is nearly double that of the second quartile and more than triple that of the fourth quartile. Access to HPC resources varies greatly. The first quartile universities have a higher rate of HPC access (50.9% versus 9% of fourth quartile universities), and when they do, have more computing power (351.11 petaFlops versus 54 petaFlops).

In terms of geographical distribution, American universities are the most prevalent in quartile 1, European universities peak in quartile 2, and Asian universities in quartile 3. Additionally, the top-quartile universities have more researchers affiliated with industry but fewer with governmental labs. There are few differences between quartiles in other dimensions. For example, the racial and gender distribution of the average university is similar across quartiles: around one third female, half Asian, one quarter white.

S3 Empirical Design

We split the publication distribution into four quartiles, each representing approximately 25% of all top journal publications and categorize universities accordingly. The first quartile includes the top 12 universities. The second quartile includes the next 36 universities, the third quartile the next 88, and the fourth quartile the next 364.

OpenAlex does not provide demographic indications for authors. We determine author ethnicity using the rethnicity package in R. This package predicts ethnicity from first names using probabilistic models trained on demographic data. Gender is identified by the genderizeR package in R, using first names and a large online database. Researchers in departments specializing in structural or computational biology or similar fields are classified as structural biologists.

In the main regression specifications, we include as control all of the variables mentioned in Table S3. to control for changes in faculty composition. The dependent variable y_{ut} is one of three variables:

1. The log number of publications in top journals from university u at time t: $y_{ut} = log(P^{all}_{u,t})$
2. The market share of university u at time t: $y_{ut} = MktS^{all}_{ut}$
3. The market share of university u at time t taking into account only main authors is: $y_{ut} = MktS^{main}_{u,t}$

S4 Additional Results

We classify a publication as using AlphaFold if i) it explicitly mentions AlphaFold or an analogous AI tool in their abstract, or ii) it cites [37], the main paper that introduced AlphaFold . Note that this method for identifying AlphaFold usage is neither a lower-bound nor an upper bound of real adoption: one paper might be mentioning AlphaFold without using it, while another paper might be using AlphaFold without mentioning it. Note, however, that we can only search for AlphaFold mentions in the abstract, and thus most likely underestimate its use.

Figure S5 shows the proportion of publications using AlphaFold in both structural and non-structural protein publications. We observe a relatively modest adoption in 2018, which remains stable for three years, before a large increase follows the release of the AF2 database. Subsequently, we see a rapid adoption in structural biology, reaching 30% in 2023, as shown in figure S5(A). The results are similar for nonstructural protein research, but the magnitudes are smaller. If we restrict the sample to papers on protein structure prediction and macromolecular crystallography, the adoption rate reaches 60%[4].

S4.1 AlphaFold Adoption

We approximate the use of AlphaFold or other equivalent AI systems at the publication level. We classify a publication as using AlphaFold if i) it explicitly mentions AlphaFold or an analogous AI tool in their abstract, or ii) it cites the original AlphaFold paper [37], the main paper that introduced AlphaFold . Note that this method for identifying AlphaFold usage is neither a lower-bound nor an upper bound of real adoption: one paper might be mentioning or citing AlphaFold without using it, while another paper might be using AlphaFold without mentioning

[4]See Figure S5(B) in the appendix.

it or citing it. With that being said, we do expect our measure to underestimate the use of AI more than it overestimates it: We can only search for AlphaFold mentions in the abstract, and we expect that many publications mention AlphaFold only in the body of the text.

Figure S5 shows the proportion of publications using AlphaFold in both structural and non-structural protein publications, as well as in different subfields of structural biology. We observe a relatively modest adoption in 2018, which remains stable for three years, before a large increase follows the release of the AF2 database. Subsequently, we see a rapid adoption in structural biology, reaching 30% in 2023, as shown in figure S5(A). The results are similar for nonstructural protein research, but the magnitudes are smaller. If we restrict the sample to papers on protein structure prediction and macromolecular crystallography, the adoption rate reaches 60% (Figure S5(B)).

Adoption rates also vary between university quartiles. The panel S6(A) shows the AlphaFold share of publications for each university quartile. In structural biology, the two lower quartiles produce the majority of AlphaFold articles (36% and 31%, respectively). The picture is more nuanced in nonstructural protein research. Panel S6(B) indicates that all quartiles had a similar AlphaFold paper output between 2020 and 2022, with a stronger adoption by the bottom quartile pre-2020. In the last year of our sample we see a shift where bottom quartile shares reach 30% (quartile three) and 34% (bottom quartile), while the shares of quartile one and quartile fall to 16% and 20% respectively.

S5 Supplementary Figures

Figure S1: **Timeline of AI development in structural biology**
The development of AI systems in structural biology took place in two distinct phase. An early period in green where the adoption of AI likely requires substantial domain knowledge and ICT skills. Then a later phase in orange with the introduction of the AF2 database where adoption becomes much easier and likely touch a much wider group.

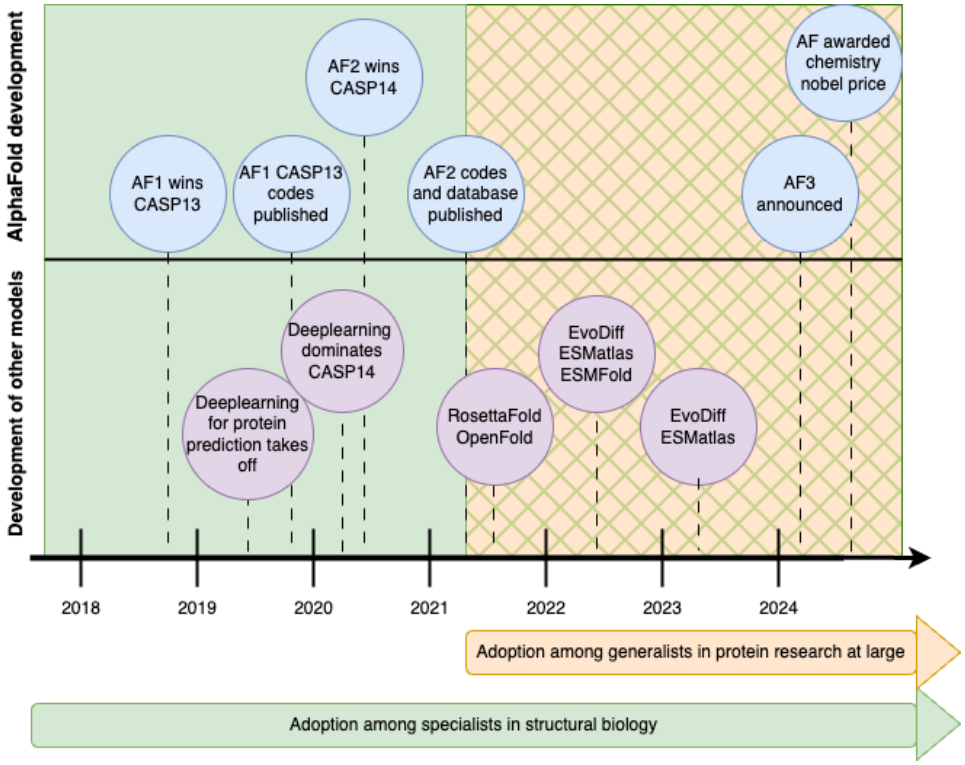
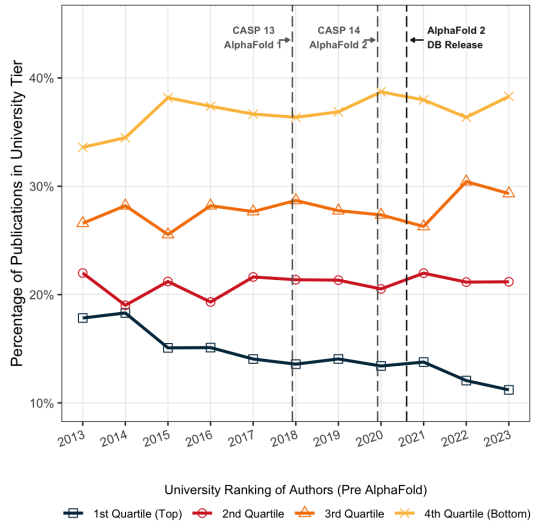


Figure S2: **Publications in lower-tier journals by university quartiles**

Figure illustrates the percentage of protein research publications in lower-tier journals with an SJR $\in$ [1, 10) between 2013 and 2018, before the release of AlphaFold 1. The analysis is limited to authors affiliated with the top 500 universities, which are divided into four quartiles based on their aggregate share of publications in high-impact journals during the same period.

(A) Structural biology publications



(B) Non-structural protein publications

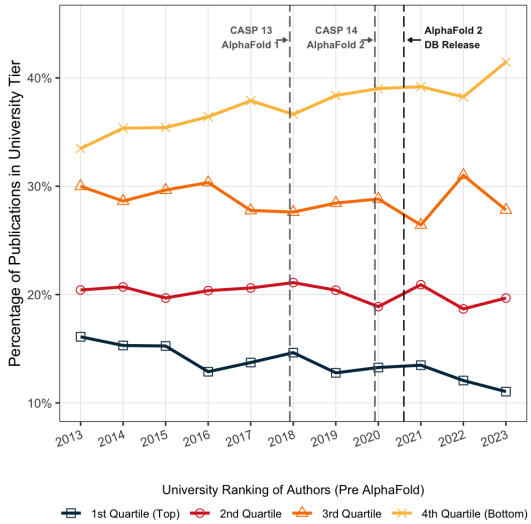


Figure S3: **Publications outside protein research in top journals by university quartiles**
Figure illustrates the percentage of publications outside protein research in top journals (with an SJR $\geq$ 10 between 2013 and 2018, before the release of AlphaFold 1). The analysis is limited to authors affiliated with the top 500 universities, which are divided into four quartiles based on their share of publications in these top journals during the same period. University rankings are determined across all authors on a publication.

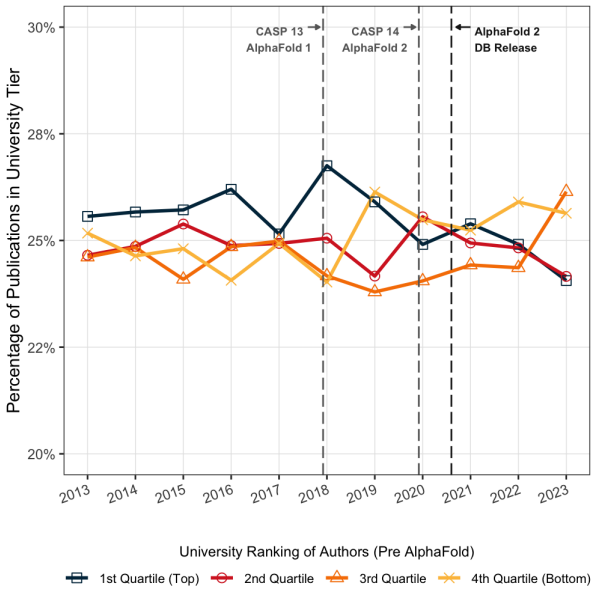


Figure S4: **Distribution of top journal publications over time**

Figure illustrates the annual number of research articles published in top-tier journals (SJR $\geq$ 10) between 2013 and 2023, covering three broad areas: structural biology, non-structural biology protein research, and non-protein research.

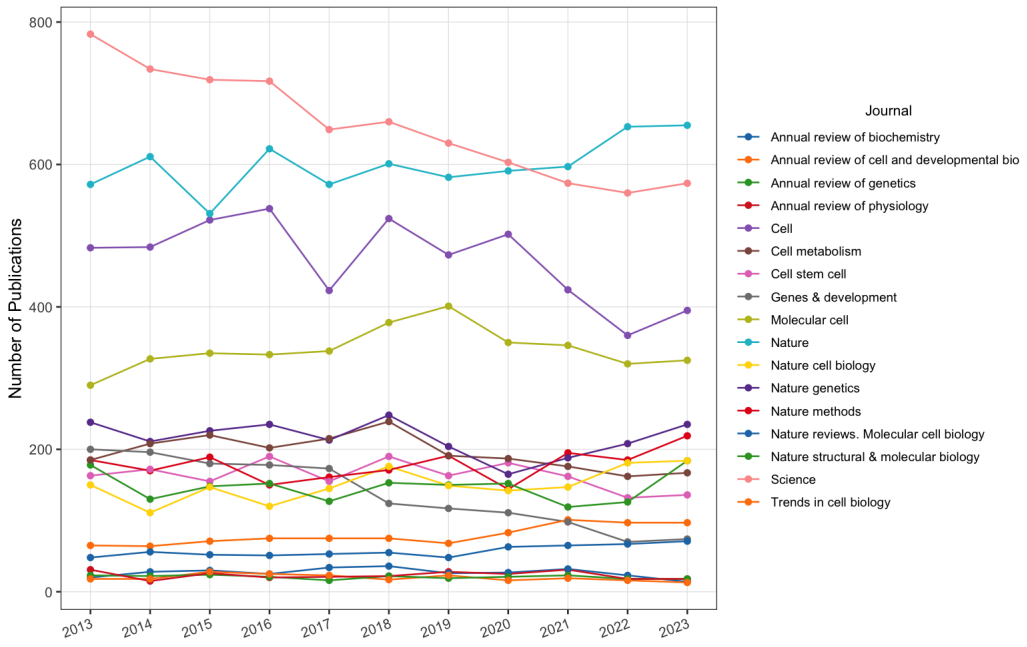
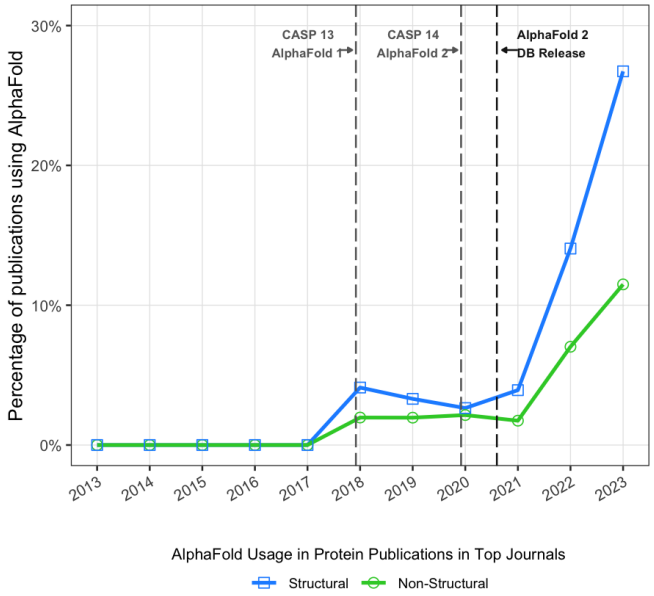


Figure S5: **AlphaFold mentions in top protein publications**

Figure shows details on AlphaFold adoption in top journals (SJR $\geq$ 10). Figure shows the share of protein research papers published that mention explicit use of AlphaFoldin their abstract or related AI tools, such as RoseTTAfold. Panel (A) shows this share for structuran and non-structural protein research. Panel (B) breaks down the adoption rate by subfield of structural biology.

(A) All protein publications



(B) AlphaFold mentions in top subfields within structural biology

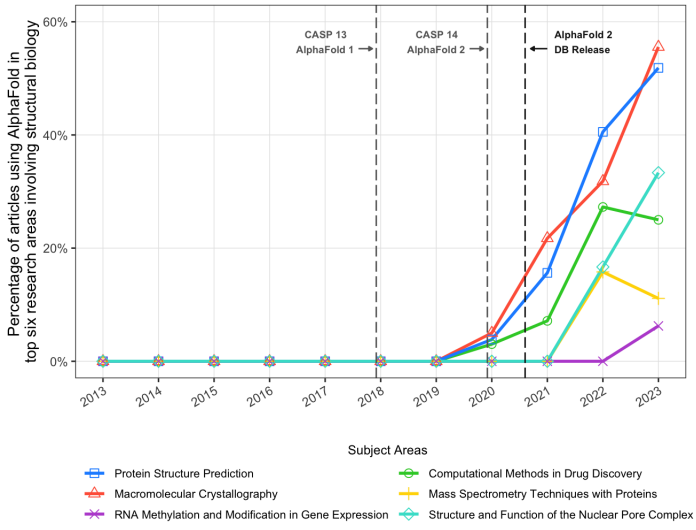
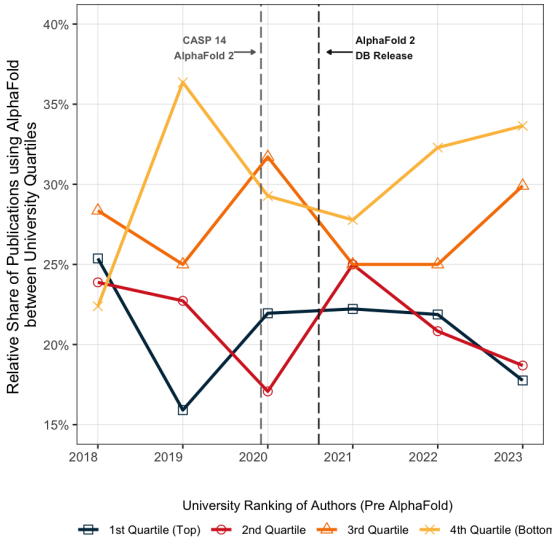
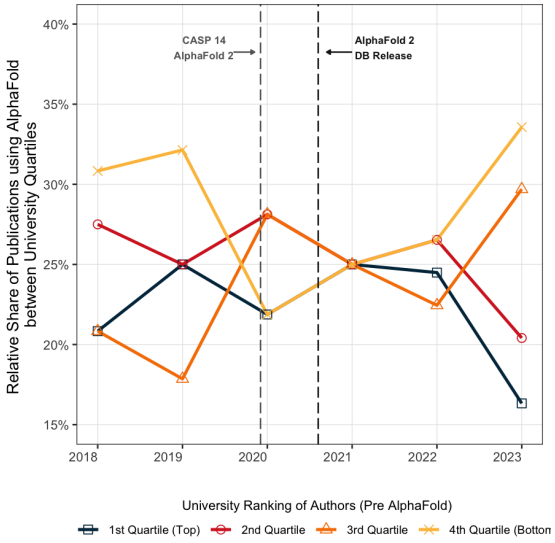


Figure S6: **AlphaFold mentions in top protein publications by university quartile**
Figure shows the relative share of AlphaFold usage among university quartiles in structural biology (Panel A) and non-structural protein publications (Panel B).

(A) Structural biology publications



(B) Non-structural protein publications



S6 Supplementary Tables

Table S1: **Other AI tools developed for protein structure prediction**
Table lists some of the most prominent AI tools that were developed following AlphaFold success. Though most of them are based on the same core transformer models like AlphaFold many of them focuses on different tasks and/or use different training data.

AI Product	Release Date	Open-Source	Creator(s)	Key Feature
RoseTTAFold	July 2021	Yes	University of Washington	Predicts Protein-nucleic acid complexes
EvoDiff	September 2022	Yes	Microsoft	Applies diffusion models to get protein design from evolutionary information
EsmFold	November 2022	No	Meta AI	Uses amino acid sequence only for faster predictions. ESM atlas is a database with 617 million predicted protein structures
ProGen	January 2023	No	Salesforce AI Research	Synthesizes new protein sequences with desired properties from natural language inputs
BaseFold	February 2023	No	Basecamp Research	Uses proprietary protein samples in training data
RFDiffusion	July 2023	Yes	University of Washington	Applies diffusion models to focus on protein design

Table S2: **Top journals in protein research**
Table lists all journals with an SJR above ten which are used to constitute the sample of publication in the analysis throughout.

Journal Name	ISSN	Mean SJR (2013–18)	Mean Impact Factor (2013–18)	Publisher
Nature reviews. Molecular cell biology	1471-0072	28.16	39.74	Nature Portfolio
Cell	0092-8674	27.21	32.02	Cell Press
Nature genetics	1061-4036	23.01	28.02	Nature Portfolio
Annual rev. of biochemistry	1545-4509	22.12	28.44	Annual Reviews
Nature	1476-4687	18.46	41.24	Nature Portfolio
Nature methods	1548-7091	18.14	27.13	Nature Portfolio
Annual review of cell and dev. biology	1530-8995	14.97	14.94	Annual Reviews
Nature cell biology	1476-4679	14.63	18.37	Nature Portfolio
Cell stem cell	1934-5909	13.70	22.49	Elsevier BV
Molecular cell	1097-4164	13.47	14.24	Elsevier BV
Science	1095-9203	13.08	39.10	AAAS
Annual review of genetics	1545-2948	12.95	11.98	Annual Reviews
Cell metabolism	1550-4131	11.21	18.79	Cell Press
Genes & development	1549-5477	11.06	10.84	Cold Spring Harbor
Nature structural & molecular biology	1545-9993	10.82	12.72	Nature Portfolio
Trends in cell biology	0962-8924	10.41	14.39	Elsevier BV
Annual review of physiology	0066-4278	10.04	17.22	Annual Reviews

Table S3: **Descriptive statistics: top protein publications and university characteristics before AlphaFold**

Table presents summary statistics at the university-year level for the period 2013–18, representing the years before the release of AlphaFold 1. The sample is restricted to protein-related publications in top journals with an SJR score of 10 or higher, authored by researchers affiliated with the top 500 universities prior to AlphaFold's release. Each cell reports the mean of the respective variable, measured annually during the pre-AlphaFold period, with universities grouped into quartiles.

Variable (*Pre AlphaFold estimates*)	1^{st} (reference group)	2^{nd}	3^{rd}	4^{th}
Structural biology publications				
Number of Publications per Year per University	8.58	2.39	1.75	0.80
Market Share (% publications)	0.37	0.10	0.08	0.03
Market Share (% of authors on publication)	0.39	0.08	0.06	0.02
Non-structural protein publications				
Number of Publications per Year per University	7.88	2.58	1.59	0.88
Market Share (% publications)	0.36	0.12	0.07	0.04
Market Share (% of authors on publication)	0.37	0.10	0.05	0.03
University characteristics				
Faculty Size	614	308	267	178
Share Structural Biologists (%)	5.98	4.90	5.19	3.93
Share Female (%)	31.04	32.96	31.46	34.23
Share White (%)	28.91	26.11	23.62	22.61
Share Asian (%)	51.09	50.15	55.80	54.69
Share Black (%)	12.68	12.92	8.96	9.43
Share affiliated in US (%)	45.01	22.60	16.93	21.87
Share affiliated in Europe (%)	38.04	54.28	48.36	44.56
Share affiliated in Asia (%)	20.68	19.77	33.26	26.14
Share affiliated with Govt (%)	5.76	9.58	6.59	7.18
Share affiliated with Research Lab (%)	14.26	16.73	12.06	13.31
Share affiliated with Industry (%)	3.22	2.70	2.70	1.99
Share affiliated with Medical Inst (%)	16.97	16.92	16.46	14.46
Share HPC Access (%)	50.90	32.26	31.00	9.07
HPC power (petaFlop/s)	351.11	205.55	231.27	54.18
Number of Universities	12	36	88	364

Table S4: **List of universities in the top quartile**

These 12 universities together make up approximately 25% of top journal publications before AlphaFold (year 2013-2018)

Name	Country
Harvard University	USA
Johns Hopkins University	USA
University of California, San Francisco	USA
University of Toronto	Canada
University of Pennsylvania	USA
University of Washington	USA
University College London	UK
Stanford University	USA
Duke University	USA
University of Michigan–Ann Arbor	USA
University of California, Los Angeles	USA
University of Oxford	UK

Table S5: **Descriptive statistics: top non-protein publications and university characteristics before AlphaFold**

Table presents summary statistics at the university-year level for the period 2013–18, representing the years before the release of AlphaFold 1. The sample is restricted to non-protein publications in top journals with an SJR score of 10 or higher, authored by researchers affiliated with the top 500 universities prior to AlphaFold's release. Each cell reports the mean and standard deviation (in brackets) of the respective variable, measured annually during the pre-AlphaFold period, with universities grouped into quartiles. ***, **, and * denote statistical significance in the difference in means between a variable in the given quartile and the first quartile (reference group) at the 1%, 5%, and 10% levels, respectively.

Variable (*Pre AlphaFold estimates*)	University Quartile: mean (SD)			
	1^{st}	2^{nd}	3^{rd}	4^{th}
Number of publications per year per university	3.19(12.13)	3.55(11)	3.7(7.45)∗	2.89(4.86)
Market Share (% publications)	0.06(0.25)	0.07(0.22)	0.07(0.15)∗	0.06(0.1)
Market share (% of authors on publication)	0.06(0.28)	0.08(0.27)	0.07(0.16)∗	0.05(0.1)
Number of Universities	12	36	88	364

Table S6: **Structural biology top publications around AlphaFold: robustness checks**

PostAF1 is a binary variable set to one for papers published after the release of AlphaFold V1 in December 2018. The sample is limited to structural biology papers published in top journals with an SJR score of 10 or higher and to authors affiliated with the top 500 universities prior to the release of AlphaFold 1. Universities are ranked based on their average TNCS in the period preceding the release of AlphaFold 1. The highest ranked universities placed in the top quartile (*University Tier = 1st Quartile*) are used as the reference group. Standard errors are double-clustered by university and year, and reported in parantheses. ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively.

(A) Publications

Dependent Variable:	Log(1 + No. of Publications)				
Model:	(1)	(2)	(3)	(4)	(5)
University Tier = 2nd Quartile × PostAF1	0.2444***	0.2443***	0.2351***	0.2418***	0.2309***
	(0.0172)	(0.0173)	(0.0145)	(0.0197)	(0.0166)
University Tier = 3rd Quartile × PostAF1	0.3555***	0.3557***	0.3764***	0.3425***	0.3605***
	(0.0211)	(0.0210)	(0.0262)	(0.0272)	(0.0322)
University Tier = 4th Quartile × PostAF1	0.5849***	0.5845***	0.5598***	0.5678***	0.5392***
	(0.0321)	(0.0320)	(0.0334)	(0.0379)	(0.0412)
Controls: Share Structural Biologists		✓			✓
Controls: HPC Power			✓		✓
Controls: Author Characteristics				✓	✓
Year FE	✓	✓	✓	✓	✓
University FE	✓	✓	✓	✓	✓
Observations	5,500	5,500	5,500	5,500	5,500
Adjusted R^2	0.780	0.781	0.805	0.785	0.809

(B) Market share (% publications, all authors)

Dependent Variable:	Market Share (% Publications, All Authors)				
Model:	(1)	(2)	(3)	(4)	(5)
University Tier = 2nd Quartile × PostAF1	0.1387***	0.1394***	0.1406***	0.1322***	0.1342***
	(0.0019)	(0.0018)	(0.0014)	(0.0049)	(0.0037)
University Tier = 3rd Quartile × PostAF1	0.1433***	0.1445***	0.1738***	0.1404***	0.1644***
	(0.0028)	(0.0025)	(0.0219)	(0.0059)	(0.0216)
University Tier = 4th Quartile × PostAF1	0.2827***	0.2813***	0.2837***	0.2584***	0.2497***
	(0.0043)	(0.0050)	(0.0232)	(0.0105)	(0.0271)
Controls: Share Structural Biologists		✓			✓
Controls: HPC Power			✓		✓
Controls: Author Characteristics				✓	✓
Year FE	✓	✓	✓	✓	✓
University FE	✓	✓	✓	✓	✓
Observations	5,500	5,500	5,500	5,500	5,500
Pseudo R^2	0.904	0.904	0.906	0.917	0.921

(C) Market share (% publications, main authors)

Dependent Variable:	Market Share (% Publications, Main Authors)				
Model:	(1)	(2)	(3)	(4)	(5)
University Tier = 2nd Quartile × PostAF1	0.1482***	0.1480***	0.1291***	0.1683***	0.1480***
	(0.0020)	(0.0024)	(0.0029)	(0.0036)	(0.0043)
University Tier = 3rd Quartile × PostAF1	0.1409***	0.1405***	0.1543***	0.1578***	0.1623***
	(0.0032)	(0.0036)	(0.0160)	(0.0063)	(0.0145)
University Tier = 4th Quartile × PostAF1	0.2700***	0.2664***	0.2688***	0.2560***	0.2420***
	(0.0044)	(0.0058)	(0.0209)	(0.0111)	(0.0242)
Controls: Share Structural Biologists		✓			✓
Controls: HPC Power			✓		✓
Controls: Author Characteristics				✓	✓
Year FE	✓	✓	✓	✓	✓
University FE	✓	✓	✓	✓	✓
Observations	5,500	5,500	5,500	5,500	5,500
Pseudo R^2	0.899	0.901	0.903	0.915	0.919

Table S7: **Non-structural protein top publications around AlphaFold: robustness checks**
PostAF1 is a binary variable set to one for papers published after the release of AlphaFold 1 in December 2018. The sample is limited to non-structural biology protein papers published in top journals with an SJR score of 10 or higher and to authors affiliated with the top 500 universities prior to the release of AlphaFold 1. Universities are ranked based on their average TNCS in the period preceding the release of AlphaFold 1. The highest ranked universities placed in the top quartile (*University Tier = 1st Quartile*) are used as the reference group. Standard errors are double-clustered by university and year, and reported in parantheses. ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively.

(A) Publications

Dependent Variable:	Log(1 + No. of Publications)				
Model:	(1)	(2)	(3)	(4)	(5)
University Tier = 2nd Quartile × PostAF1	0.2912***	0.2909***	0.2931***	0.2757***	0.2767***
	(0.0201)	(0.0201)	(0.0194)	(0.0281)	(0.0276)
University Tier = 3rd Quartile × PostAF1	0.4052***	0.4043***	0.4183***	0.3913***	0.4018***
	(0.0185)	(0.0185)	(0.0215)	(0.0240)	(0.0289)
University Tier = 4th Quartile × PostAF1	0.5963***	0.5957***	0.5572***	0.5766***	0.5361***
	(0.0279)	(0.0279)	(0.0259)	(0.0347)	(0.0340)
Controls: Share Structural Biologists		✓			✓
Controls: HPC Power			✓		✓
Controls: Author Characteristics				✓	✓
Year FE	✓	✓	✓	✓	✓
University FE	✓	✓	✓	✓	✓
Observations	5,500	5,500	5,500	5,500	5,500
Adjusted R^2	0.760	0.761	0.787	0.765	0.791

(B) Market share (% publications, all authors)

Dependent Variable:	Market Share (% Publications, all authors)				
Model:	(1)	(2)	(3)	(4)	(5)
University Tier = 2nd Quartile × PostAF1	0.0876***	0.0869***	0.0943***	0.0840***	0.0904***
	(0.0028)	(0.0023)	(0.0061)	(0.0050)	(0.0083)
University Tier = 3rd Quartile × PostAF1	0.1482***	0.1488***	0.1834***	0.1392***	0.1715***
	(0.0030)	(0.0032)	(0.0195)	(0.0084)	(0.0280)
University Tier = 4th Quartile × PostAF1	0.2505***	0.2501***	0.2628***	0.2387***	0.2470***
	(0.0044)	(0.0044)	(0.0143)	(0.0074)	(0.0227)
Controls: Share Structural Biologists		✓			✓
Controls: HPC Power			✓		✓
Controls: Author Characteristics				✓	✓
Year FE	✓	✓	✓	✓	✓
University FE	✓	✓	✓	✓	✓
Observations	5,500	5,500	5,500	5,500	5,500
Pseudo R^2	0.884	0.883	0.888	0.899	0.902

(C) Market share (% publications, main authors)

Dependent Variable:	Market Share (% Publications, main authors)				
Model:	(1)	(2)	(3)	(4)	(5)
University Tier = 2nd Quartile × PostAF1	0.1246***	0.1242***	0.1288***	0.1142***	0.1164***
	(0.0024)	(0.0017)	(0.0052)	(0.0070)	(0.0096)
University Tier = 3rd Quartile × PostAF1	0.1532***	0.1529***	0.1893***	0.1398***	0.1707**
	(0.0024)	(0.0019)	(0.0195)	(0.0131)	(0.0312)
University Tier = 4th Quartile × PostAF1	0.2840***	0.2832***	0.3096***	0.2767***	0.2962***
	(0.0041)	(0.0038)	(0.0185)	(0.0117)	(0.0300)
Controls: Share Structural Biologists		✓			✓
Controls: HPC Power			✓		✓
Controls: Author Characteristics				✓	✓
Year FE	✓	✓	✓	✓	✓
University FE	✓	✓	✓	✓	✓
Observations	5,500	5,500	5,500	5,500	5,500
Pseudo R^2	0.881	0.880	0.885	0.897	0.901

References

1. Mathias Wullum Nielsen and Jens Peter Andersen. "Global citation inequality is on the rise". In: *Proceedings of the National Academy of Sciences* 118.7 (Feb. 2021), e2012208118.
2. Emil Bargmann Madsen and Kaare Aagaard. "Concentration of Danish research funding on individual researchers and research topics: Patterns and potential drivers". In: *Quantitative Science Studies* 1.3 (Aug. 2020), pp. 1159–1181.
3. Thijs Bol, Mathijs de Vaan, and Arnout van de Rijt. "The Matthew effect in science funding". In: *Proceedings of the National Academy of Sciences* 115.19 (May 2018), pp. 4887–4890.
4. Kaare Aagaard, Alexander Kladakis, and Mathias W. Nielsen. "Concentration or dispersal of research funding?" In: *Quantitative Science Studies* 1.1 (Feb. 2020), pp. 117–149.
5. Pierre Azoulay, Christian Fons-Rosen, and Joshua S. Graff Zivin. "Does Science Advance One Funeral at a Time?" en. In: *American Economic Review* 109.8 (Aug. 2019), pp. 2889–2920.
6. Joshua M. Nicholson and John P. A. Ioannidis. "Conform and be funded". en. In: *Nature* 492.7427 (Dec. 2012), pp. 34–36.
7. Danielle Li. "Expertise versus Bias in Evaluation: Evidence from the NIH". en. In: *American Economic Journal: Applied Economics* 9.2 (Apr. 2017), pp. 60–92.
8. Mikko Packalen and Jay Bhattacharya. *Does the NIH Fund Edge Science?* Working Paper. July 2018.
9. Robert K. Merton. "The Matthew Effect in Science". In: *Science* 159.3810 (Jan. 1968), pp. 56–63.
10. Pierre Azoulay, Toby Stuart, and Yanbo Wang. "Matthew: Effect or Fable?" In: *Management Science* 60.1 (Jan. 2014), pp. 92–109.
11. Corinna Ghirelli et al. "The Long-Term Causal Effects of Winning an ERC Grant". en. In: (2023).
12. Ajay Agrawal, John McHale, and Alexander Oettl. "How stars matter: Recruiting and peer effects in evolutionary biology". In: *Research Policy* 46.4 (May 2017), pp. 853–867.
13. Pierre Azoulay, Joshua S. Graff Zivin, and Jialan Wang. "Superstar Extinction *". en. In: *Quarterly Journal of Economics* 125.2 (May 2010), pp. 549–589.
14. George J. Borjas and Kirk B. Doran. "The Collapse of the Soviet Union and the Productivity of American Mathematicians*". In: *The Quarterly Journal of Economics* 127.3 (Aug. 2012), pp. 1143–1203.
15. Ajay Agrawal and Avi Goldfarb. "Restructuring Research: Communication Costs and the Democratization of University Innovation". en. In: *American Economic Review* 98.4 (Sept. 2008), pp. 1578–1590.
16. Neil Thompson and Samantha Zyontz. *Decomposing the "Tacit Knowledge Problem:" Codification of Knowledge and Access in CRISPR Gene-Editing*. en. SSRN Scholarly Paper. Rochester, NY, Nov. 2017.
17. Jeffrey L. Furman and Florenta Teodoridis. "Automation, Research Technology, and Researchers' Trajectories: Evidence from Computer Science and Electrical Engineering". In: *Organization Science* 31.2 (Mar. 2020), pp. 330–354.

18. Fiona Murray et al. “Of Mice and Academics: Examining the Effect of Openness on Innovation”. en. In: *American Economic Journal: Economic Policy* 8.1 (Feb. 2016), pp. 212–252.

19. Jeremy M Berg, John L Tymoczko, and Lubert Stryer. *Biochemistry*. WH Freeman, 2002.

20. Amy C Anderson. “Structure-based functional design of drugs: from target to lead compound”. In: *Methods in molecular biology* 230 (2003), pp. 375–384.

21. Yigong Shi. “A milestone in protein crystallography: The 100th anniversary of X-ray diffraction”. In: *Science China Life Sciences* 57.9 (2014), pp. 856–857.

22. M Grabowski et al. “The cost and value of three-dimensional protein structure”. In: *Drug Discovery World* 7.3 (2016), pp. 64–72.

23. Naomi E Chayen and Emmanuel Saridakis. “Protein crystallization strategies for structural genomics”. In: *Nature Methods* 5.2 (2008), pp. 147–153.

24. Elspeth F Garman. “Developments in x-ray crystallographic structure determination of biological macromolecules”. In: *Science* 343.6175 (2014), pp. 1102–1108.

25. Yifan Cheng et al. “Single particle cryo-EM—how did it get here and where will it go”. In: *Cell* 161.3 (2015), pp. 438–449.

26. Eva Nogales. “The development of cryo-EM into a mainstream structural biology technique”. In: *Nature methods* 13.1 (2016), pp. 24–27.

27. Mohammed AlQuraishi. “AlphaFold at CASP13”. en. In: *Bioinformatics* 35.22 (Nov. 2019). Ed. by Alfonso Valencia, pp. 4862–4865.

28. Andrew W Senior et al. “Improved protein structure prediction using potentials from deep learning”. In: *Nature* 577.7792 (2020), pp. 706–710.

29. John Jumper et al. “Highly accurate protein structure prediction with AlphaFold”. In: *Nature* 596.7873 (2021), pp. 583–589.

30. Zhenyu Yang et al. “AlphaFold2 and its applications in the fields of biology and medicine”. en. In: *Signal Transduction and Targeted Therapy* 8.1 (Mar. 2023), pp. 1–14.

31. Kathryn Tunyasuvunakool et al. “Highly accurate protein structure prediction for the human proteome”. In: *Nature* 596.7873 (2021), pp. 590–596.

32. Thomas C. Terwilliger et al. “AlphaFold predictions are valuable hypotheses and accelerate but do not replace experimental structure determination”. en. In: *Nature Methods* 21.1 (Jan. 2024), pp. 110–116.

33. James D Watson and Janet M Thornton. “AlphaFold and implications for intrinsically disordered proteins”. In: *Nature Reviews Molecular Cell Biology* 23.7 (2022), pp. 461–463.

34. Alexander von Appen et al. “In situ structural analysis of the human nuclear pore complex”. en. In: *Nature* 526.7571 (Oct. 2015), pp. 140–143.

35. Shyamal Mosalaganti et al. “AI-based structure prediction empowers integrative structural analysis of human nuclear pores”. In: *Science* 376.6598 (June 2022), eabm9506.

36. Martijn Visser, Nees Jan van Eck, and Ludo Waltman. “Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic”. en. In: *Quantitative Science Studies* 2.1 (Apr. 2021), pp. 20–41.

37. John Jumper et al. “Highly accurate protein structure prediction with AlphaFold”. en. In: *Nature* 596.7873 (Aug. 2021), pp. 583–589.