

## **AI predictions and the expansion of scientific frontiers: Evidence from structural biology**

Mengyi Sun<sup>1</sup>, Sukwoong Choi<sup>2</sup>, Yian Yin<sup>3\*</sup>

<sup>1</sup>Simons Center for Quantitative Biology, Cold Spring Harbor Laboratory, NY

5 <sup>2</sup>Department of Information Systems and Business Analytics, University at Albany, NY

<sup>3</sup>Department of Information Science, Cornell University, NY

\*Correspondence to: Yian Yin ([yian.yin@cornell.edu](mailto:yian.yin@cornell.edu))

10 **Artificial intelligence holds the potential to expand the frontier of scientific research, yet recent work has raised concern that it may instead narrow scientific attention to well-established areas. Here, leveraging the 2021 release of AlphaFold2 as a quasi-experimental opportunity, we provide field-level evidence that AI can redirect collective attention toward more novel research targets. Tracking 245,396 experimental structures in the Protein Data Bank, we show that a long-running decline in the study of novel proteins halted after**

15 **AlphaFold2's release, with the shift concentrated among studies citing AlphaFold2 and targets with high-confidence predictions. This pattern extends to 248,191 downstream papers that consume structural knowledge, where engagement with genes lacking experimental structures and with understudied human genes increased since 2021. Amid rising concern that AI may reinforce scientific canons, our findings offer an early field-level**

20 **case where AI predictions expand scientific frontiers, consistent with the idea that the real-world consequences of AI on science depend on where their informational gains are greatest. These results suggest AI can complement human knowledge and redirect collective attention in science, with broad implications for emerging AI for science models.**

Science progresses through sustained exploration of the unknown (1–4). Yet as the burden of knowledge accumulates, the frontier becomes prohibitively costly to expand (5, 6). Growing concerns have emerged that scientific attention is increasingly concentrated on familiar questions and well-explored paradigms (7, 8), leading to a documented decline in scientific novelty and disruptiveness across disciplines (9, 10). This contraction, if continued, may jeopardize the recognition and nurturing of emerging scientists (11), undermine the long-term vitality of the R&D ecosystem (12), and threaten the pipeline of breakthroughs that fuels economic growth and human prosperity (13).

Artificial intelligence offers a potentially powerful, yet contested, mechanism (14) to counter this drift. By generating predictions at scale, AI systems may drastically lower the cost of exploring unknown areas and expand the horizon of viable targets. Recent models (15–17) increasingly learn transferable regularities that can generalize beyond observed samples, raising the prospect that their predictions can prove informative in the broader space and empower researchers to venture beyond the established canon (18–23). However, recent empirical evidence warns of a darker possibility: while AI can boost individual productivity, it may simultaneously narrow the collective diversity of ideas (24–28). Driven by institutional incentives, competitive environments, and resource constraints, scientists may strategically converge on data-rich areas where models perform best, further reinforcing existing paradigms (29, 30).

Despite these promises and perils, there has been limited large-scale evidence documenting whether and how AI can expand the scientific frontier in the wild. Although AI models have grown increasingly capable, a translation gap from computational predictions into real-world discovery persists, and adoption rates for frontier models—particularly in experimental wet labs—have remained too low to demonstrate macro-level shifts. We address this empirical challenge by leveraging the 2021 release of AlphaFold2 (16, 31) (AF2) as a unique quasi-experimental design. As one of the first widely adopted AI-for-science tools, AF2 provides structure predictions for over 200 million proteins, bringing predictive power at unprecedented scale into a field historically dominated by the intensive exploitation of familiar targets and a vast expanse of structural dark matter. Moving beyond recent studies of individual adopters (32–34), we examine whether AF2 altered research directions in structural biology and beyond.

## Results

### AI adoption in structural determination

How rapidly has AI been adopted in determining protein structures? We collected and analyzed large-scale data on all experimental structures deposited in the Protein Data Bank (35), including 245,396 PDB entries and 87,656 associated papers (SM S1.1). We further use citations to key AlphaFold papers as a proxy for AI use in experimental structure determination (SM S1.4, S2.1). Figure 1A, B document rapid uptake shortly after the tool became available. Although only four years have passed since AF2’s debut, 20.5% of new structures deposited in 2024 are now associated with papers citing AlphaFold2. Adoption is markedly higher among cryo-electron microscopy (cryo-EM) entries (38.7%), nearly doubling the rate for X-ray crystallography (20.2%). Notably, in both modalities, adoption follows highly nonlinear dynamics: initial acceleration followed by slower growth and eventual saturation. The S-shaped curve is well captured by a logistic diffusion model; extrapolating the curve from empirical data up to late 2022 (18 months post-AF2) is sufficient to predict subsequent saturation (SM S2.1). Despite the unprecedented speed of AlphaFold’s uptake, its longer-term diffusion appears to follow the same trajectory as many earlier technological tools (36).

The rapid adoption prompts us to ask: does the growing volume of AI-assisted research come at a cost to experimental quality? The quality effect of AI for science remains debated in the literature (37–39), largely because of the challenges in measuring scientific quality at scale. Here we leverage a unique quality measure for PDB structures: R-free (40), which quantifies the agreement between the diffraction data and the refined model. Fig. 1C plots the distribution of R-free for structures citing AI algorithms alongside contemporaneous non-citing entries and pre-AF2 entries. AF2-citing entries exhibit higher (worse) R-free values, suggesting an apparent net loss in average quality.

Why are AI-powered discoveries associated with lower quality? We further explore this apparent puzzle by examining a second quality indicator: resolution, which measures the precision of initial diffraction data. Resolution is determined primarily by crystallization and imaging, and thus unlikely to be substantially affected by predictive algorithms. Figure 1D examines the distribution of resolution across the three groups and shows that AF2-using structures also link to lower-quality initial data ( $P < 0.001$ ). More importantly, when we condition on resolution, all three groups collapse onto a single curve (Fig. 1E). Controlling for initial data quality (resolution), AlphaFold use is associated with neither gains nor losses in outcome quality ( $P = 0.980$ ). The apparent quality deficit therefore suggests a compositional shift: AI use is not uniform in the protein space. Rather, it appears preferentially deployed on challenging cases with inferior diffraction data, which is sufficient to account for the R-free difference. These results highlight that AI use is more closely tied to changes in research focus than to quality. A sequence embedding-based visualization of the protein space is consistent with this interpretation (Fig. 1F, SM S2.2,2.3): AlphaFold-citing entries enrich regions of sequence space largely distinct from those densely populated by pre-AF2 PDB entries, indicating that AF2 redirected attention toward proteins beyond the established structural canon.

### Directional shifts in experimental structures

To systematically measure changes in research focus, we construct entity-level measures of novelty by grouping PDB entities into sequence-similarity clusters (70% identity, alignment-based) (41, 42) (SM S2.4). We then classify a PDB structure deposited in year  $t$  as novel if no prior entities targeted the same cluster before  $t - \Delta t$ , where  $\Delta t = 1$  year to account for the lag between project initiation and deposition (SM S2.6). The relationship between research projects and PDB entities is often not one-to-one, as a single project may yield multiple PDB entities. To this end, we group all entities deposited by the same team on the same day into a single project, assigning equal fractional weights within each project (SM S2.7, Fig. S1).

Figure S2 plots novelty, defined as the average fraction of novel structures targeted per project, over time. Consistent with existing analysis (43), we find a historical decline through the early 2020s. Yet in more recent years, the decline halted, where the inflection appears to align well with the public release of AF2 and the availability of its large-scale prediction. To test this relationship more rigorously, we estimate a segmented linear regression examining changes in slope around AF2's release (Fig. 2A, SM S2.10). The results suggest a significant slope change: prior to AlphaFold the novelty rate declines by 1.2 percentage points per year ( $P < 0.001$ ), a trend that halts and becomes statistically indistinguishable from zero ( $P = 0.590$ ) in the post-AF2 period (Fig. 2A).

Further, using a linear approximation of the pre-existing trend (Fig. 2B), we estimate that by late 2024, the rate of novel PDB structures is 4.4 percentage points higher than expected. This link between AI use and novel research is striking, as most AI algorithms are trained on existing data derived from prior research and often struggle where training examples are sparse. If scientists

leveraging AI predictions were to prioritize targets with accurate predictions, one might expect research to increasingly converge to well-explored areas (7, 28, 44, 45). Yet we observe the opposite pattern: AI predictions have enabled exploration into previously uncharacterized regions of the protein space, directing experimental effort toward proteins that lacked prior structural data.

5 Several lines of further analyses support this interpretation. First, the effect is more pronounced when the algorithm yields high-quality predictions. To this end, we exploit a key feature of AF2: the predicted Local Distance Difference Test (pLDDT) score (16, 46), a residue-level confidence estimate that is intrinsic to the prediction. On a 0–100 scale, scores above 70 are often considered comparable to experimental accuracy. Indeed, for clusters that can be directly mapped to  
10 AlphaFold predictions, the arrest of novelty decline is concentrated in those with high-pLDDT predictions (Fig. 2C) and largely absent in those with low-pLDDT predictions (Fig. 2D).

Further, the elevated novelty is predominantly contributed by entries whose corresponding papers explicitly cite AlphaFold (Fig. 2E). PDB papers citing AlphaFold engage 38.3% more novel protein sequences than non-citing papers from the same period, again pointing to a close link  
15 between AI use and expansion into unexplored sequence space. Yet comparing non-citing papers before and after AF2 release, we find no statistical evidence for the nonlinear transition ( $P = 0.639$ ), suggesting that the macro-level shift cannot be simply explained by alternative temporal factors.

Overall, the increase in research novelty is not spread evenly across the structural dark matter, but  
20 concentrated when scientists actively deploy AI tools and where AI provides confident predictions. Amid ongoing discussion about whether predictions might substitute for experiments, our results suggest the opposite: structural biologists selectively use AlphaFold predictions to pursue novel proteins precisely where the AI is confident, indicating a complementary rather than substitutive relationship (42, 47). SM S4 documents further robustness across different analytical choices and  
25 thresholds, showing our results remain remarkably consistent (Fig. S4-S7).

The novelty change documented in Fig. 2A-E also implies a broader redistribution of structural research effort: as researchers increasingly examine proteins beyond existing literature, the research ecosystem may see an increase in target diversity. To this end, we calculate, for each 6-  
30 month time period, the Gini coefficient of PDB research intensity (SM S2.9). Consistent with this hypothesis (Fig. 2F), while the Gini coefficient had grown over time ( $P = 0.009$ ), it has largely stalled post-AF2 ( $P = 0.718$ ).

Together, these results point to a directional shift that goes beyond the productivity and quality effects of AI: as AF2 converts a substantial fraction of unknowns into knowns, scientists start  
redirecting attention toward a diverse set of promising directions.

### 35 **The directional shift propagates to downstream research**

The directional shift among experimental producers raises a further question: can AF2 further impact the broader community that consumes structural knowledge in downstream research? Protein structures inform a wide range of biomedical tasks, such as drug design (48), functional  
40 annotation (49), and mechanistic modeling (50). Thus, the population of experimental *producers* examined above represents only one side of the ecosystem, and reallocation of structural effort toward novel targets may have limited impact if downstream *consumers* continued to draw selectively on well-characterized proteins.

To answer this question, we used NCBI PubTator3 (51), a text-mining system that maps biomedical concepts to entity identifiers in PubMed titles and abstracts, linking 2.89M papers with  
45 132K genes from 2015 to 2025 (SM S1.3). For each paper, we used GPT-4o-mini to classify

whether it likely consumed structural information (see SM S2.5 and Fig. S3 for classification details and validations against a hand-labeled subset). Figure 3A plots the fraction of papers classified as consuming structural knowledge, which follows a modest increasing trend before AF2's release (0.10 percentage points per year) but accelerates markedly thereafter (0.46 percentage points per year). Tracing citations within these papers, the disproportionate post-2021 increase is largely associated with papers citing AlphaFold, suggesting an extensive margin expansion in AF2-related downstream research (Fig. S8A).

On the intensive margin, we further examine which genes scientists engage with. Paralleling the novelty definition in Fig. 2, we focus on an extreme case: genes without any experimental structure released by the year of the consumer paper's publication (SM S2.6,2.8). At the individual level, papers citing AF2 study proteins without prior experimental structures at 3.11 times the rate of non-citing papers (Fig. S8B). Consistent with this substantial contrast, Fig. 3B shows an observable system-level shift: within structure-consuming papers, the engagement of structurally unsolved genes follows a persistent pre-2021 decline that has been interrupted sharply thereafter. An event study adjusting for pre-trend (Fig. 3C) estimates an increase of 5.58 percentage points by 2025, representing 21.7% relative growth over the pre-AF2 baseline. To examine whether this can be explained by alternative temporal confounds unrelated to structural knowledge, we further implement a difference-in-differences design, using gene-mentioning papers that do not consume structural information as controls (Fig. 3D). The two groups exhibit parallel trends through 2020, after which structural consumers start targeting structurally unsolved genes at a significantly higher rate, with a magnitude comparable to the event study estimate (5.22 percentage points by 2025).

The growing focus on structurally unsolved genes has important implications along biologically meaningful dimensions. As an illustrative example, we examine whether AF2 redirected attention toward valuable but historically underserved targets in the human genome. Drawing on a machine learning model that predicts expected research intensity from gene-specific chemical, physical, and biological features (52), 6,981 out of 15,056 human genes have been identified as potentially understudied in the early 2010s. In our data, the fraction of consumer papers engaging these genes continued to decline until 2021 but stabilized thereafter (Fig. 3E). Compared with the 2020 baseline, we estimate a 29.0% relative increase in research on understudied genes (3.24 percentage points, Fig. 3F), which suggests a substantial diversification effect within the human genome. Together, these findings provide initial evidence consistent with AF2's directional impact across the broader biomedical literature, highlighting an expansion in the volume of structural engagement and the scope of research targets.

### **A minimal model of AI predictions and research directions**

Our results may seem paradoxical. AI models are trained on existing data, and a large body of recent work warns that such systems tend to reinforce attention on well-studied targets, concentrating effort where training examples are densest rather than expanding into the unknown (18, 28, 53). Why, then, does AF2 produce the opposite pattern? The answer, we argue, lies not in whether a tool uses AI, but in where it delivers its greatest informational contribution.

To formalize this idea, we develop a simple framework (Figure S9A) in which a researcher chooses among projects (e.g., protein sequences) with varying levels of distance from the existing knowledge space, denoted by  $z$ , where higher  $z$  corresponds to thinner prior evidence and greater baseline ambiguity (e.g., fewer close homologs). Each researcher selects a single project (or target), maximizing the expected payoff from discovery—the probability of success times the value upon discovery—minus the cost of experimental pursuit. Because only one project can be

pursued, choosing a more attractive project raises the opportunity cost of all others. The model suggests that the impact of AI predictions on research directions depends not only on their own predictive power, but also on how that power is distributed across the knowledge landscape. We define the tool's information gain as  $r(z) = h^{post}(z) / h^{pre}(z)$ , the ratio of (perceived) success probabilities with and without AI. When  $r(z)$  declines with  $z$ , gains concentrate in data-rich regions and raise the opportunity cost of frontier work—a pattern we term *data-reinforcing*. When  $r(z)$  rises with  $z$ , gains accrue disproportionately to hard, data-sparse targets, reducing the risk premium on novel work and shifting effort outward—a pattern we term *frontier-expanding*.

This framework shifts attention from whether a tool uses AI to where its informational lift is concentrated across the knowledge landscape. AF2 appears broadly consistent with the frontier-expanding pattern: its architecture appears to capture transferable physical and geometric constraints that generalize beyond close homologs, delivering greater relative gains where prior evidence is thinnest (Fig. S9B). An interesting contrasting case is molecular replacement (MR) (54), a widely used technology that uses known structures as templates to solve the crystallographic phase problem (29). MR is particularly powerful for proteins with close homologs but adds little value at higher  $z$ , which appears consistent with the data-reinforcing pattern (Fig. S9C). Full model derivation and connections to empirical proxies are documented in Supplementary Material S3.

## Discussion

Our findings provide field-level evidence that AI predictions can redirect scientific effort toward new frontiers. In structural biology, a persistent decline in the targeting of novel proteins halted after the release of AF2. The shift is concentrated where AF2 should matter most—among studies that use it, for targets with high-confidence predictions, and in downstream research that consumes structural knowledge. While AI is often treated as a productivity tool that accelerates existing workflows, our results suggest a broader function: it can reorganize which problems a field regards as tractable and worth pursuing.

This paper complements a growing literature that has mainly examined how AI affects individual scientists or labs, comparing research behaviors between adopters and non-adopters<sup>32,33,55–57</sup>. By treating the field as the unit of analysis, we identify a collective effect: once predictive tools diffuse broadly enough, they can reshape the distribution of attention across a field. AF2 offers a unique empirical opportunity, combining a quasi-experimental design, rapid diffusion, and observable downstream impact. But the same patterns will likely emerge as AI models across domains become increasingly integrated into scientific workflows and capable of learning and generalizing transferable scientific regularities.

The directional change documented here has broad implications for how effects of AI in science should be evaluated. AF2-linked structures appear to underperform on experimental quality measures, which reflects a compositional shift in the problems being attempted: scientists use AI on problems that appear intrinsically harder. More generally, when a technology induces selection into more challenging problems, standard metrics such as quantity and quality can mislead if interpreted without regard to that selection. The same logic applies to novelty. In a setting where novelty has been declining, arresting that decline is itself a meaningful shift, and analyses that focus only on the post-period levels may risk understating the true impact<sup>58</sup>.

These findings further suggest AI need not simply reinforce attention on scientific canons, even when trained on data generated by past research. Our minimal modeling framework is useful in integrating these seemingly contradictory perspectives (Fig. S9): for scientific progress, it may

matter at least as much where a tool's informational gain is concentrated as how well it performs on average. An AI system can score highly on benchmarks while concentrating its gains in already-tractable regions, thereby reinforcing reliance on well-established areas. Conversely, a tool with modest average performance can be consequential if it extends useful guidance into data-sparse domains. In our case, AF2 appears more consistent with a frontier-expanding pattern (Fig. S9B), whereas molecular replacement more closely reflects a data-reinforcing pattern (Fig. S9C, also see Ref (29)).

There are several limitations that readers should keep in mind. We do not claim a perfectly isolated causal estimate, as AF2 emerged alongside other changes in biomedical research. While we have excluded COVID-related research<sup>59</sup> in our analysis, we cannot fully rule out the role of the post-COVID resumption of laboratory research. Further, other relevant AI technologies, including protein language models<sup>60,61</sup>, also emerged in the same period. However, the null effect among low-pLDDT targets, the high novelty of AF2-citing studies, and the divergence between papers that do and do not consume structural knowledge, appear difficult to reconcile with a simple background temporal trend.

Even with these caveats, the broader lesson remains: beyond boosting quantity and quality, AI can matter for scientific progress through where it makes exploration more feasible. With human scientists in the loop, AI predictions need not fully replace experiment to be important. Rather, it may be sufficient that they complement human efforts<sup>62</sup> by providing guidance where prior knowledge is sparse and the perceived risks of exploration are high. AF2 showcases an early and clear view of this broader dynamic: it helped arrest a long-running novelty decline, even without matching experimental standards across all targets<sup>63,64</sup>. As foundation models and AI agents proliferate<sup>65</sup> across mathematical<sup>66,67</sup>, physical<sup>68</sup>, and life sciences<sup>69</sup>, their long-run impact on science may depend not only on how accurately they predict, but also on how far beyond existing concentrations of knowledge they remain useful.

## References and Notes

1. Fortunato, S. *et al.* Science of science. *Science* **359**, (2018).
2. Uzzi, B., Mukherjee, S., Stringer, M. & Jones, B. Atypical combinations and scientific impact. *Science* **342**, 468–472 (2013).
3. Stephan, P. How economics shapes science. (2015).
4. Kuhn, T. S. & Hacking, I. *The Structure of Scientific Revolutions*. vol. 2 (University of Chicago press Chicago, 1970).
5. Jones, B. F. The burden of knowledge and the “death of the renaissance man”: Is innovation getting harder? *The Review of Economic Studies* **76**, 283–317 (2009).
6. Milojević, S. Quantifying the cognitive extent of science. *J. Informetr.* **9**, 962–973 (2015).

7. Foster, J. G., Rzhetsky, A. & Evans, J. A. Tradition and innovation in scientists' research strategies. *Am. Sociol. Rev.* **80**, 875–908 (2015).
8. Bhattacharya, J. & Packalen, M. Stagnation and scientific incentives. (2020).
9. Park, M., Leahey, E. & Funk, R. J. Papers and patents are becoming less disruptive over time. *Nature* **613**, 138–144 (2023).
10. Chu, J. S. G. & Evans, J. A. Slowed canonical progress in large fields of science. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2021636118 (2021).
11. Cui, H., Lin, Y., Wu, L. & Evans, J. A. Aging and the narrowing of scientific innovation. *arXiv [cs.DL]* (2025) doi:10.48550/arXiv.2202.04044.
12. Bloom, N., Jones, C. I., Van Reenen, J. & Webb, M. Are ideas getting harder to find? *Am. Econ. Rev.* **110**, 1104–1144 (2020).
13. Bush, V. *The Endless Frontier*. (National Science Foundation, 2020).
14. Wang, H. *et al.* Scientific discovery in the age of artificial intelligence. *Nature* **620**, 47–60 (2023).
15. Kashinath, K. *et al.* Physics-informed machine learning: case studies for weather and climate modelling. *Philos. Trans. A Math. Phys. Eng. Sci.* **379**, 20200093 (2021).
16. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
17. Maslej, N. *et al.* Artificial Intelligence Index Report 2025. *arXiv [cs.AI]* (2025) doi:10.48550/arXiv.2504.07139.
18. Agrawal, A., McHale, J. & Oettl, A. Artificial intelligence and scientific discovery: a model of prioritized search. *Res. Policy* **53**, 104989 (2024).
19. Agrawal, A. K., McHale, J. & Oettl, A. AI in Science. (2026).

20. Wu, L., Hitt, L. & Lou, B. Data analytics, innovation, and firm productivity. *Manage. Sci.* **66**, 2017–2039 (2020).
21. Agarwal, D. *et al.* AutoDiscovery: Open-ended scientific discovery via Bayesian surprise. *arXiv [cs.LG]* (2026) doi:10.48550/arXiv.2507.00310.
- 5 22. Ludwig, J. & Mullainathan, S. Machine learning as a tool for hypothesis generation. *Q. J. Econ.* **139**, 751–827 (2024).
23. Liu, H., Zhou, Y., Li, M., Yuan, C. & Tan, C. Literature meets data: A synergistic approach to hypothesis generation. 245–281 (2025).
24. Doshi, A. R. & Hauser, O. P. Generative AI enhances individual creativity but reduces the  
10 collective diversity of novel content. *Sci. Adv.* **10**, eadn5290 (2024).
25. Hao, Q., Xu, F., Li, Y. & Evans, J. Artificial intelligence tools expand scientists’ impact but contract science’s focus. *Nature* **649**, 1237–1243 (2026).
26. Boussioux, L., Lane, J. N., Zhang, M., Jacimovic, V. & Lakhani, K. R. The crowdless future? Generative AI and creative problem-solving. *Organ. Sci.* (2024) doi:10.1287/orsc.2023.18430.
- 15 27. Choi, S., Kang, H., Kim, N. & Kim, J. The dual edges of AI: Advancing knowledge while reducing diversity. *PNAS Nexus* **4**, pgaf138 (2025).
28. Hoelzemann, J., Manso, G., Nagaraj, A. & Tranchero, M. *The Streetlight Effect in Data-Driven Exploration*. <http://dx.doi.org/10.3386/w32401> (2024) doi:10.3386/w32401.
29. Kim, S. *Navigating the Rugged Data Landscape: The Impact of Data-Extrapolation Technologies on Knowledge Production*. [https://soomi-kim.com/assets/Soomi\\_Kim\\_JMP\\_Latest.pdf](https://soomi-kim.com/assets/Soomi_Kim_JMP_Latest.pdf) (2025).
- 20 30. Zhou, E. B., Lee, D. & Gu, B. Who expands the human creative frontier with generative AI: Hive minds or masterminds? *Sci. Adv.* **11**, eadu5800 (2025).

31. Varadi, M. *et al.* AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444 (2022).
32. Stein, C. & Hill, R. How artificial intelligence shapes science: Evidence from AlphaFold. *Available at SSRN 6395041* (2026).
- 5 33. Yu, Z. The Impacts of AI at Scale: Evidence from Research Scientists. (2024).
34. Vicente, D. A. & Richardson, G. *AI in Science: Emerging Evidence from AlphaFold 2*. [https://www.innovationgrowthlab.org/wp-content/uploads/2025/11/ai\\_in\\_science\\_af2\\_igl\\_full.pdf](https://www.innovationgrowthlab.org/wp-content/uploads/2025/11/ai_in_science_af2_igl_full.pdf) (2025).
- 10 35. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
36. Rogers, E. M., Singhal, A. & Quinlan, M. M. Diffusion of innovations. 432–448 (2014).
37. Si, C., Yang, D. & Hashimoto, T. Can LLMs generate novel research ideas? A large-scale human study with 100+ NLP researchers. *arXiv [cs.CL]* (2024) doi:10.48550/arXiv.2409.04109.
- 15 38. Messeri, L. & Crockett, M. J. Artificial intelligence and illusions of understanding in scientific research. *Nature* **627**, 49–58 (2024).
39. Binz, M. *et al.* How should the advancement of large language models affect the practice of science? *Proc. Natl. Acad. Sci. U. S. A.* **122**, e2401227121 (2025).
40. Brünger, A. T. Free R value: a novel statistical quantity for assessing the accuracy of crystal  
20 structures. *Nature* **355**, 472–475 (1992).
41. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).
42. Hill, R. & Stein, C. Scooped! Estimating rewards for priority in science. *J. Polit. Econ.* **133**, 793–845 (2025).

43. Levitt, M. Growth of novel protein structural data. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 3183–3188 (2007).
44. He, X. & Zhang, J. On the growth of scientific knowledge: yeast biology as a case study. *PLoS Comput. Biol.* **5**, e1000320 (2009).
- 5 45. Pfeiffer, T. & Hoffmann, R. Temporal patterns of genes in scientific publications. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 12052–12056 (2007).
46. Mariani, V., Biasini, M., Barbato, A. & Schwede, T. IDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics* **29**, 2722–2728 (2013).
- 10 47. Brynjolfsson, E., Li, D. & Raymond, L. Generative AI at work. *Q. J. Econ.* **140**, 889–942 (2025).
48. Anderson, A. C. The process of structure-based drug design. *Chem. Biol.* **10**, 787–797 (2003).
49. Lee, D., Redfern, O. & Orengo, C. Predicting protein function from sequence and structure. *Nat. Rev. Mol. Cell Biol.* **8**, 995–1005 (2007).
- 15 50. Xu, P. *et al.* Structural insights into the diversity and DNA cleavage mechanism of Fanzor. *Cell* **187**, 5238–5252.e20 (2024).
51. Wei, C.-H. *et al.* PubTator 3.0: an AI-powered literature resource for unlocking biomedical knowledge. *Nucleic Acids Res.* **52**, W540–W546 (2024).
52. Stoeger, T., Gerlach, M., Morimoto, R. I. & Nunes Amaral, L. A. Large-scale investigation of  
20 the reasons why potentially important genes are ignored. *PLoS Biol.* **16**, e2006643 (2018).
53. Gans, J. *The Microeconomics of Artificial Intelligence*. (MIT Press, London, England, 2025).
54. Rossmann, M. G. The molecular replacement method. *Acta Crystallogr. A* **46 ( Pt 2)**, 73–82 (1990).
55. Ampudia, D. & Richardson, G. *AI in Science: Emerging Evidence from AlphaFold 2*. (2025).

56. Kusumegi, K. *et al.* Scientific production in the era of large language models. *Science* **390**, 1240–1243 (2025).
57. Bao, H., Sun, M. & Teplitskiy, M. Where there's a will there's a way: ChatGPT is used more for science in countries where it is prohibited. *Quant. Sci. Stud.* **6**, 716–731 (2025).
58. Frank, M. R., Sabet, A. J., Simon, L., Bana, S. H. & Yu, R. AI-exposed jobs deteriorated before ChatGPT. *arXiv [econ.GN]* (2026) doi:10.48550/arXiv.2601.02554.
59. Chen, Q., Allot, A. & Lu, Z. LitCovid: an open database of COVID-19 literature. *Nucleic Acids Res.* **49**, D1534–D1540 (2021).
60. Lin, Z. *et al.* Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **379**, 1123–1130 (2023).
61. Xiao, Y. *et al.* Protein large language models: A comprehensive Survey. *arXiv [q-bio.BM]* (2025) doi:10.48550/arXiv.2502.17504.
62. Wu, S., Yang, D., Chang, J., Hearst, M. A. & Lo, K. Human-AI collaboration: How AIs augment human teammates. in *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 5: Tutorial Abstracts)* (eds. Arase, Y., Jurgens, D. & Xia, F.) 5–6 (Association for Computational Linguistics, Stroudsburg, PA, USA, 2025).
63. Terwilliger, T. C. *et al.* AlphaFold predictions are valuable hypotheses and accelerate but do not replace experimental structure determination. *Nat. Methods* **21**, 110–116 (2024).
64. Agarwal, V. & McShan, A. C. The power and pitfalls of AlphaFold2 for structure prediction beyond rigid globular proteins. *Nat. Chem. Biol.* **20**, 950–959 (2024).
65. Gao, J. & Wang, D. Quantifying the use and potential benefits of artificial intelligence in scientific research. *Nat. Hum. Behav.* **8**, 2281–2292 (2024).
66. Hubert, T. *et al.* Olympiad-level formal mathematical reasoning with reinforcement learning. *Nature* **651**, 607–613 (2026).

67. Davies, A. *et al.* Advancing mathematics by guiding human intuition with AI. *Nature* **600**, 70–74 (2021).
68. Merchant, A. *et al.* Scaling deep learning for materials discovery. *Nature* **624**, 80–85 (2023).
69. Avsec, Ž. *et al.* Advancing regulatory variant effect prediction with AlphaGenome. *Nature* **649**, 1206–1218 (2026).

**Acknowledgments:** We thank Juan Mateos Garcia, George Richardson, Ryan Hill, and Carolyn Stein, as well as seminar participants at Google DeepMind, Cold Spring Harbor Laboratory, and the ICML AI4Science workshop for helpful correspondence and discussions.

**Author contributions:**

Conceptualization: MS, SC, YY

Methodology: MS, SC, YY

Investigation: MS, SC, YY

Visualization: MS, YY

Funding acquisition: YY

Project administration: YY

Supervision: YY

Writing – original draft: MS, SC, YY

Writing – review & editing: MS, SC, YY

**Competing interests:** Authors declare that they have no competing interests.

**Data, code, and materials availability:** Data and code for reproduce the analysis would be made available upon publication.

**Supplementary Materials**

Materials and Methods

Supplementary Text

Figs. S1 to S9

Tables S1 to S2

References (1–20)

**Fig. 1. Adoption of AlphaFold2 in structural biology.** **A–B**, Fraction of new PDB structures associated with papers citing AlphaFold2, shown separately for X-ray crystallography (**A**) and cryo-electron microscopy (**B**) by deposition quarter. Points show empirical data; curves show fitted logistic diffusion models estimated on data through late 2022. **C**, Distribution of R-free values for structures deposited before AF2 release (pre-AF2; red), and post-AF2 structures in papers citing AlphaFold2 (post-AF2, citing; blue) or not citing AlphaFold2 (post-AF2, non-citing; orange). **D**,

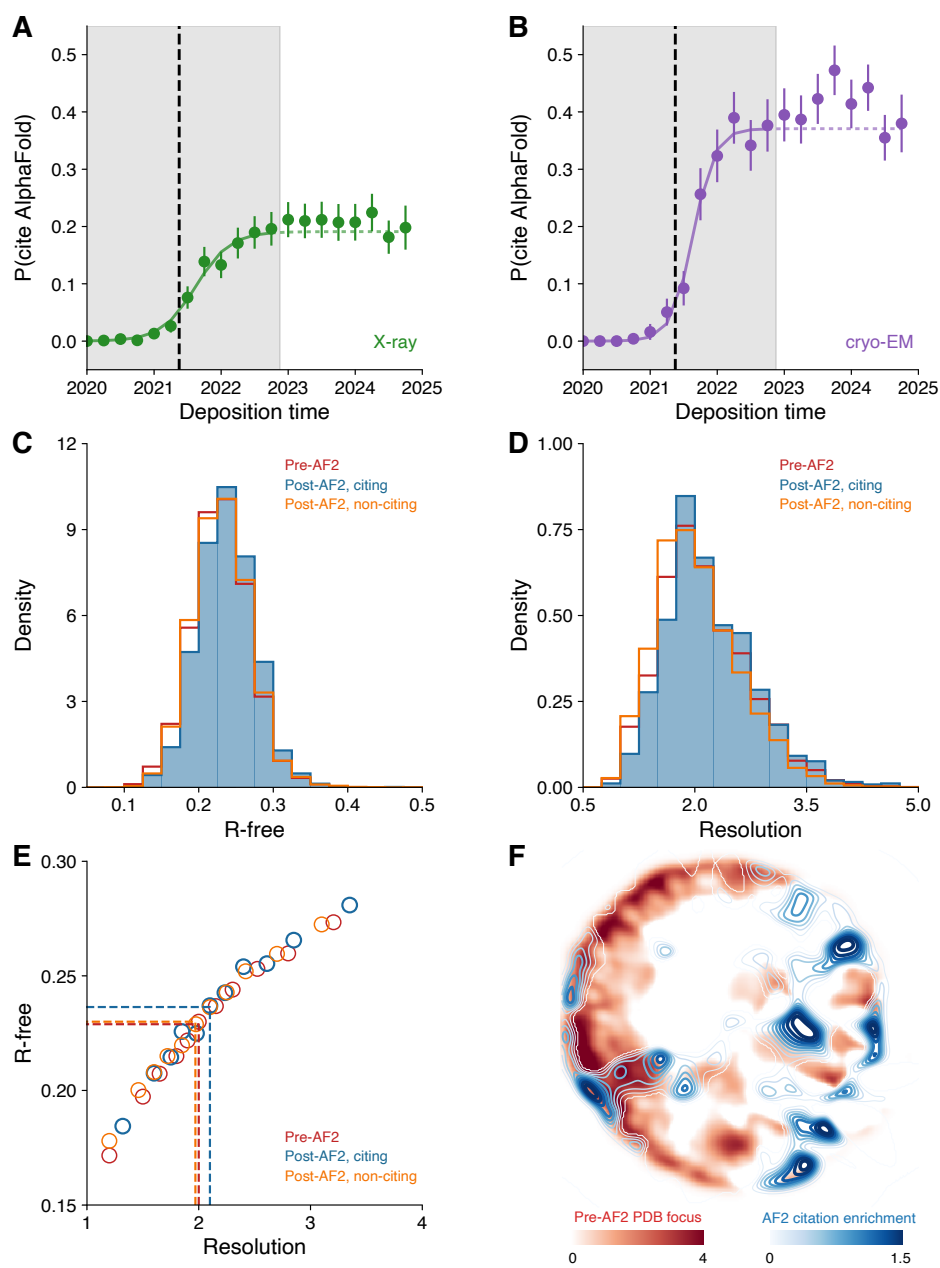
Distribution of crystallographic resolution across the same three groups. **E**, Mean R-free as a function of resolution, binned by resolution quantile. Dashed lines indicate group medians. When conditioned on resolution, the three groups collapse onto a single curve, indicating that the apparent quality deficit in AF2-citing structures is fully explained by a compositional shift toward harder targets. **F**, Two-dimensional projection of protein sequence embedding space (ProtTrans T5 XL U50), with PaCMAP applied to PCA-reduced AlphaFold embeddings as the reference manifold. Red shading indicates density enrichment of pre-AF2 PDB entries relative to the AlphaFold universe; blue shading indicates density enrichment of post-AF2 AF2-citing PDB entries relative to post-AF2 non-citing entries. AF2-citing entries occupy regions of sequence space largely distinct from the pre-AF2 structural canon. Error bars in all panels represent 95% confidence intervals.

**Fig. 2. AlphaFold2 arrests a long-running decline in the novelty of experimental structural targets.**

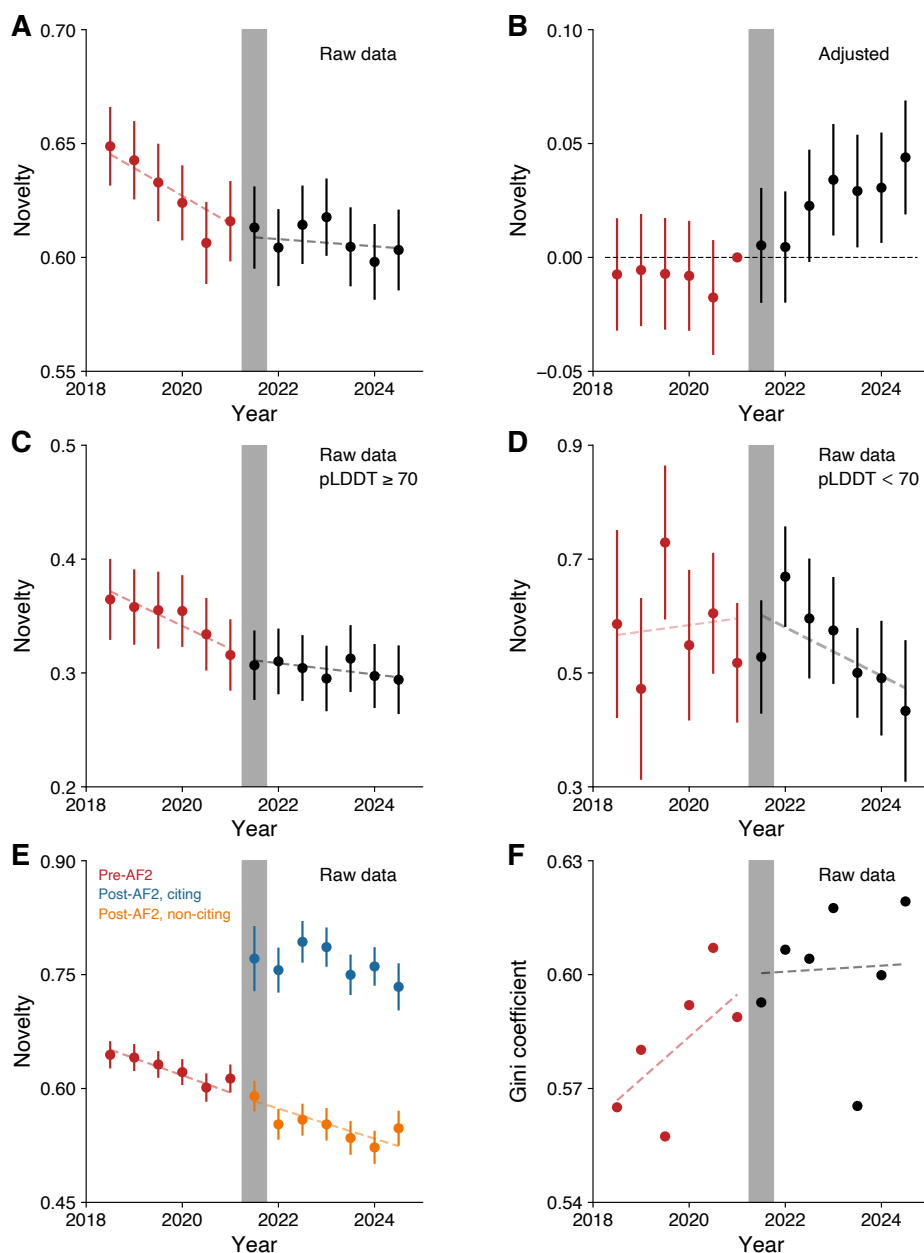
**A**, Fraction of novel protein clusters targeted per PDB project, by deposition time (6-month bin). A project is classified as targeting a novel cluster if no prior PDB entry mapped to the same 70% sequence-identity cluster before the preceding year. Red points indicate pre-AF2 period; black points indicate post-AF2 period. Dashed lines show segmented linear regression fits. Prior to AF2, novelty declined by 1.2 percentage points per year ( $P < 0.001$ ); post-AF2, the slope is statistically indistinguishable from zero ( $P = 0.590$ ). **B**, Event-study representation of novelty deviations relative to the extrapolated pre-AF2 linear trend. **C–D**, As in (A), stratified by protein cluster with AlphaFold2 prediction pLDDT  $\geq 70$  (**C**) or pLDDT  $< 70$  (**D**). The novelty inflection appears concentrated in the high-confidence regime, yet absent in the low-confidence regime. **E**, As in (A), stratified by whether the depositing paper cites AlphaFold2. AF2-citing papers target 38.3% more novel sequences than non-citing contemporaneous papers; no significant inflection is observed among non-citing papers. **F**, Gini coefficient of PDB research intensity across 70% sequence-identity clusters, calculated in 6-month bins. The Gini coefficient grew significantly in the pre-AF2 period ( $P = 0.009$ ) but has been statistically stable thereafter ( $P = 0.718$ ), indicating an arrest of increasing concentration. Error bars in all panels represent 95% confidence intervals.

**Fig.3. The directional shift propagates to broader biomedical papers consuming structural knowledge.**

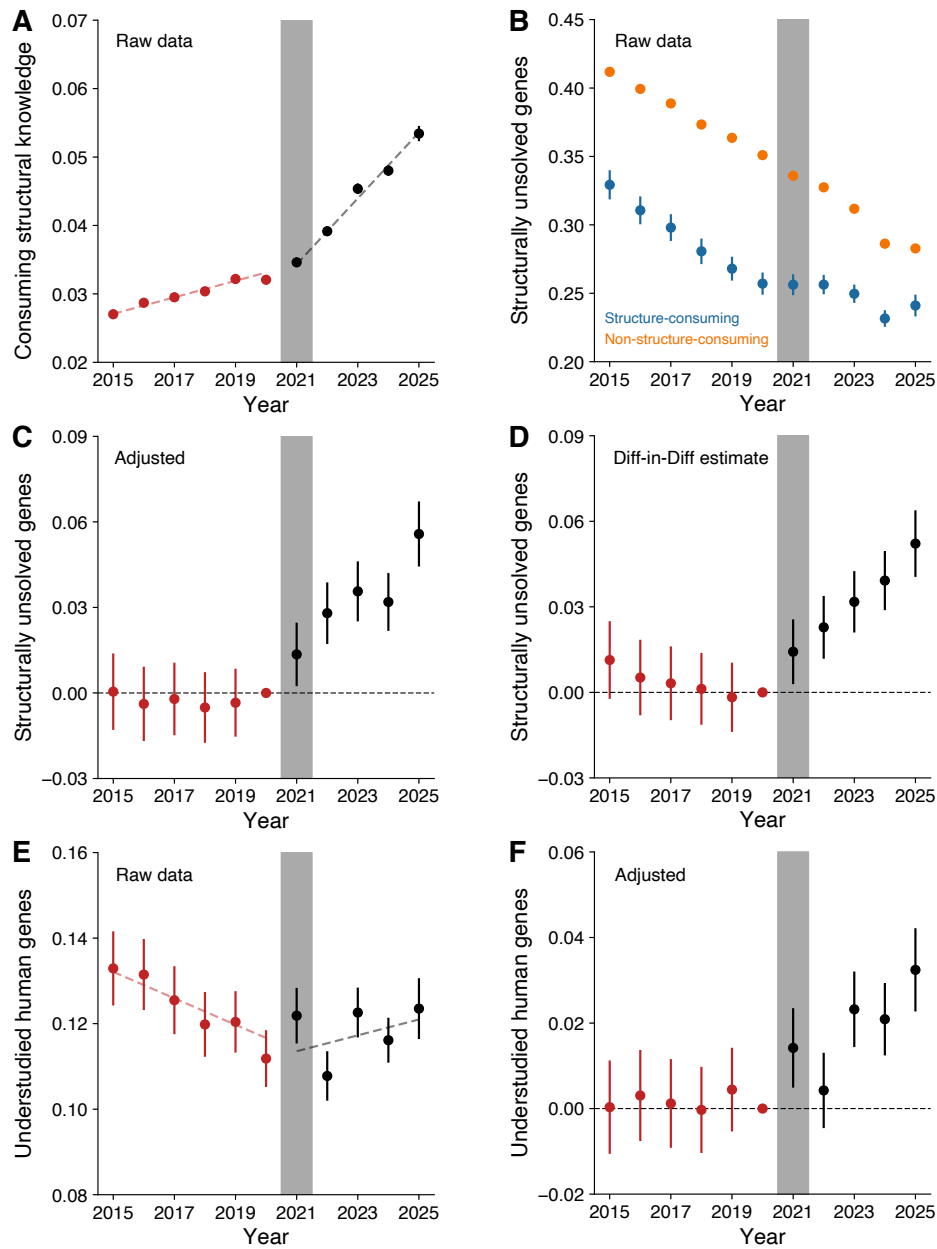
**A**, Fraction of PubTator3-annotated gene-mentioning papers classified as structural knowledge consuming by publication year, following a marked acceleration post-AF2 (0.46 percentage points per year). **B**, Among structure-consuming (blue) and non-consuming (orange) papers, the fraction engaging genes without any experimentally released structure at the time of publication. Structural consumers show a sustained pre-2021 decline that reverses after AF2's release, while non-consumers continue to decline. **C**, Event-study estimates of deviation from pre-AF2 trend in the fraction of unsolved genes targeted by structure-consuming papers. By 2025, the rate is 5.58 percentage points above the pre-trend extrapolation. **D**, Difference-in-differences estimates using non-structure-consuming gene-mentioning papers as controls. The two groups exhibit parallel pre-trends through 2020; structure-consuming papers diverge significantly thereafter (5.22 percentage points higher by 2025). **E**, Fraction of structure-consuming papers engaging genes identified as potentially understudied in the human genome. **F**, Event-study estimates of deviation from pre-AF2 trend in understudied gene engagement, corresponding to a 3.24 percentage point increase. Error bars in all panels represent 95% confidence intervals.



**Fig. 1. Adoption of AlphaFold2 in structural biology.**



**Fig. 2. AlphaFold2 arrests a long-running decline in the novelty of experimental structural targets.**



**Fig. 3. The directional shift propagates to broader biomedical papers consuming structural knowledge.**