AI IN THE LAB:

ALPHAFOLD2'S IMPACTS ON HUMAN-PRODUCED KNOWLEDGE*

Jiarui (Jerry) Qian †

March 22, 2026

Abstract

This paper studies how artificial intelligence affects knowledge production by scientists, using AlphaFold2 (AF2) as a case. AF2 predicts protein structures with near-experimental accuracy at a fraction of the cost, raising concern that it may substitute for a large share of scientists' work in solving protein structures. I find that overall experimental output continues to grow after AF2's release, but the composition shifts. Exploiting variation in AF2's prediction confidence in a difference-in-differences design, I find that experimental work on well-predicted proteins falls by about 35% relative to less-well-predicted ones. The decline is larger for proteins that are costly to experiment on or have limited downstream demand. This reallocation is also heterogeneous across experimental methods, concentrated in labor-intensive X-ray crystallography, while capital-intensive cryo-electron microscopy is largely unaffected. Meanwhile, AF2 has little effect on the quality of execution or team size within-project. One potential mechanism is that AF2 reduces the value of experimental validation on well-predicted targets, weakening incentives to pursue such work. Overall, AF2 shifts experimental effort toward areas where human comparative advantage persists.

Keywords: Artificial Intelligence; Scientific Labor; Productivity; Technological Change
JEL Codes: O31; O33; J24; D24; L65

*I am grateful to my advisors, John Pepper, Sandip Sukhtankar, Tim Layton and Anton Korinek for their invaluable guidance and support. I also thank Yutong Chen, Jonathan Colmer, Leora Friedberg, Emma Harrington, Enrique Ide, Po-Hsuan Lin, Lee Lockwood, Alexander MacKay, David McKenzie, Paul Niehaus, Pascual Restrepo, Andrew Simon, Alexander T. Whalley, Chen Xi, Zeyang Yu, and participants in NBER Digital Economics and AI Tutorial & Meeting, IESE Economics of AI Conference, UVA Labor/Public Student Workshop and Development Workshop for helpful comments. Special thanks to Huan Bao, Roman Barth, Christopher Deppmann, Alev Erisir and especially, Ahmad Jomaa for their valuable insights and explanation on the relevant scientific background. This work was supported by the Schmidt Sciences AI at Work program. All remaining errors are my own. Total word count: 8,987.

†Department of Economics, University of Virginia, 248 McCormick Road, Charlottesville, VA 22903; Email: jqian21@virginia.edu; Tel: +1 (434) 882-3425.

1 Introduction

The development of artificial intelligence (AI) has not only transformed the production of goods and services but also begun to reshape the innovation process itself (Aghion, Jones, and Jones, 2018; Cockburn, Henderson, and Stern, 2019; Trammell and Korinek, 2023). Over the past decade, AI tools have been increasingly integrated in almost all scientific fields (Gao and Wang, 2023; Chen et al., 2025). A defining characteristic that makes AI particularly salient for scientific discovery is its exceptional ability to generate data-driven predictions at low cost (Agrawal, Gans, and Goldfarb, 2019), thereby lowering barriers to domain knowledge and complementing the expertise of scientists. Yet as algorithms improve, AI-generated predictions may increasingly rival the outputs of lab experiments, raising the possibility that AI could substitute for a large share of scientists' work. Recent work also highlight broader concerns: AI may steer inquiry toward domains where it performs best (a "streetlight effect"), potentially overlooking superior alternatives and narrowing the diversity of inquiry (Hoelzemann et al., 2024); It may also incentivize knowledge workers to be more inclined to withhold newly generated knowledge when AI can expropriate and replicate their expertise with little compensation (Cullen, Li, and Li, 2025). Understanding whether, and under what conditions, AI complements or substitutes for scientists' work and how it shapes the direction of knowledge production is therefore central to the future development of innovation.

The introduction of AlphaFold2 (AF2) provides an ideal setting to empirically evaluate how AI affects scientific work. Developed by DeepMind, a subsidiary of Google, AF2 is the first AI system capable of predicting the three-dimensional (3D) structure of proteins with an accuracy comparable to the outcome of leading experimental methods. In addition to its precision, AF2 also offers extraordinary breadth: DeepMind open-sourced the model and code in 2021 Q3 (my time of treatment), and the accompanying database of precomputed predictions scaled up in subsequent waves, ultimately covering over 214 million proteins by mid-2022, making a large volume of high-quality predictions freely accessible to researchers worldwide. This groundbreaking achievement has been viewed as one of the most successful AI applications in the scientific discovery process. As recognition, the main developers were awarded the 2024 Nobel Prize in Chemistry.

In this paper, I will empirically examine how the release of AF2's predictions has affected the quantity, direction, quality and labor organization of knowledge production in structural biology. Structural biology focuses on determining the 3D structures of biological molecules, primarily proteins, through experiments, providing a foundation for understanding biological functions and downstream applications. For example, solving the structures of SARS-CoV-2 proteins has been essential for developing vaccines and therapeutics during the COVID-19 pandemic, by revealing key drug targets (Gordon et al., 2020; Wrapp et al., 2020). AF2's ability to "solve" protein structures via prediction, at an accuracy comparable to many experimentally determined structures, directly overlaps with a central task of structural biologists.

Solving protein structures through experiments has historically been a slow, resource-intensive, and uncertain process (Edwards et al., 2011). Current experimental methods each have limitations: mainstream X-ray crystallography requires growing high-quality crystals, a trial-and-error process that can take months or even years (Spence, 2017). Modern cryo–electron microscopy (Cryo-EM) bypasses the crystallization bottleneck (Shoemaker and Ando, 2018), but the installation and annual operating costs can be expensive (Crow, 2021), and the technique remains less suitable for smaller macromolecules (Wu and Rapoport, 2021). As a result, progress in mapping the "protein structure universe" has been incremental. By 2020, the Protein Data Bank (PDB) contained roughly 180,000 deposited 3D structures, corresponding to around 55,000 distinct proteins—less than 0.1% of all known protein sequences by that time (Varadi et al., 2022).

In conceptual terms, solving proteins' structure can be modeled as a production process in which scientists' expertise and experimental capital are combined to generate knowledge outputs. The introduction of AF2 serves as a large, target-specific productivity shock to this field. For proteins with high-quality AF2 predictions, the marginal cost of obtaining the structural information may be substantially reduced to a large extent. This raises two opposing possibilities. On the one hand, lower costs may attract scientists toward these well-predicted targets, leading to an influx of effort and a surge in follow-up experimental work. On the other hand, if AF2 already supplies much of the relevant structural information, the expected returns to additional experimental

work may fall as well, leading scientists exit from these areas, reallocating effort toward regions where human experimentation remains essential. The empirical analysis that follows tests these hypotheses by tracing how experimental effort, target selection, and scientific rewards respond to this heterogeneous productivity shock across proteins and methods.

My empirical analysis builds on two unique data sources: the Protein Data Bank (PDB) and the AlphaFold Protein Structure Database (AFDB). The PDB is a comprehensive repository of experimentally determined protein structures submitted by scientists worldwide. Since the early 1990s, major journals and funding agencies have required structural biologists to deposit the 3D structures they solved into the PDB prior to publication or grant review (Berman et al., 2000, 2016). As a result, the PDB provides a near-universe of completed experimental projects in structural biology and has been widely used in prior work (Hill and Stein, 2025a,b; Kim, 2025). Through the specific protein target being studied, I match each human-solved structure in the PDB to the corresponding prediction in the AFDB, which stores over 200 million protein structures predicted by AF2's algorithm.

I leverage variation in AlphaFold's prediction confidence to define treatment and control groups. For each protein, AF2 reports a confidence score indicating how closely the predicted 3D structure is expected to match an experimental one. Higher scores indicate that the predicted 3D structure is more likely to be very close to what an experiment would find, while low scores indicate a limited performance. I partition the entire structure "search space" into regions by prediction confidence. Structures targeting proteins in the highest-confidence region form the treated group; those targeting all other regions serve as controls. I then use a difference-in-differences design to compare how the quantity and quality of experimentally solved structures evolve for well-predicted versus less- or unpredicted regions before and after AF2's release. Event-study estimates show limited systematic pre-treatment differences between these groups, supporting the assumption that absent AF2, experimental effort would have followed similar trajectories across regions.

Starting with quantity, I show that the overall structure deposition has been growing steadily

both before and after AF2's release. On average, the total number of structures deposited per quarter rose by about 21.1% after AF2's introduction. But there's a pronounced reallocation in the composition of deposited structures. Regions with very high confidence predictions, interpreted by DeepMind as roughly on par with experimental results, experienced a decline about 35% in total structure depositions relative to regions where predictions are weaker or absent. The contraction is especially pronounced for certain types of proteins such as "first-attempt" structures working on previously experimentally unattempted proteins, which are more exploratory and ex ante uncertain; and non-human proteins that are further from immediate clinical demand. These patterns suggest that scientists reallocate effort away from validating structures that AF2 can already predict accurately. Interestingly, I also find that this reallocation occurs without changing the typical experimental team size for structures continued to be pursued across prediction regions.

I also find that the impact of AF2 differs sharply across experimental methods. The overall growth in experimental output is driven primarily by the expansion of Cryo-EM, the capital-intensive frontier method. Yet the method-specific analysis shows that most of the compositional shift away from experimental work on well-predicted proteins comes from X-ray crystallography, the traditional labor-intensive and more widely accessible approach. In contrast, Cryo-EM appears largely insensitive to AF2's prediction confidence, with no significant evidence of reallocation across structure types, except for the "first-attempt" deposits. This divergence suggests that the underlying experimental method is a key determinant of how scientists adapt to AI in the discovery process: AF2 may complement labor-intensive X-ray crystallography method on the intensive margin by helping scientists redirect effort toward less well-predicted regions, but the extensive-margin growth of knowledge production is driven mainly by the capital-intensive Cryo-EM method that appears less exposed to substitution by AF2.

Finally, I examine how AF2 affects the quality of structures that scientists continue to pursue, along both the execution and innovation margins. To measure execution precision, I use experimental resolution and find limited evidence that AF2's predictions improve execution quality for either method. The more striking changes appear on the innovation margin: structures targeting

well-predicted proteins face a "publication penalty," being 4.9 percentage points less likely to appear in top-decile journals than comparable structures on less well-predicted targets. This penalty is especially severe for "first-attempt" structures on well-predicted proteins, where the probability of publication in a top-decile journal falls by 11.2 percentage points, suggesting that AF2 shifts such work from exploration toward validation and thereby lowers the academic credit it receives.

Taken together, my findings suggest the opposite of the "streetlight" worry that scientific efforts may disproportionately flock to problems where AI performs best. In this case, AF2 operates more like a "floodlight" that illuminates the entire search space, sharply reducing uncertainty about many protein structures and encouraging scientists to move toward the large set of underexplored areas where their comparative advantage over AI remains strongest. At the same time, the reallocation response is tool-specific: concentrated in labor-intensive X-ray crystallography, while capital-intensive Cryo-EM adjusts less and continues to expand. This suggests that AF2 may also reshape the distribution of scientific opportunity, widening gaps across labs with unequal access to frontier equipment.

This paper contributes to several threads of literature. First, it joins the large body of literature on how automation technology may affect labor productivity and the organization of work (Schumpeter, 1942; Autor, Katz, and Krueger, 1998; Bresnahan, Brynjolfsson, and Hitt, 2002; Autor, Levy, and Murnane, 2003; Autor, 2015; Graetz and Michaels, 2018; Acemoglu and Restrepo, 2019, 2020) and more recent literature on the impact of AI at work (Brynjolfsson, Mitchell, and Rock, 2018; Webb, 2019; Zolas et al., 2021; Acemoglu et al., 2022; Noy and Zhang, 2023; Dell'Acqua et al., 2023; Brynjolfsson, Chandar, and Chen, 2025; Brynjolfsson, Li, and Raymond, 2025; Cui et al., 2025). Existing measures of "AI exposure" typically construct occupation-level indices by inferring affected tasks from occupation descriptions (Frey and Osborne, 2017; Felten, Raj, and Seamans, 2018; Webb, 2019; Eloundou et al., 2024). While this approach fits the general-purpose nature of generative AI (Goldfarb, Taska, and Teodoridis, 2023), it is inherently indirect and masks how different types of work within the same occupation are differentially affected. In this paper, I build a unique new protein-level measure that directly reflects AI's exposure over the

full idea search space of structural biology and provides clear evidence of a substantial reallocation of work caused by AI.

This paper also contributes to the emerging literature on how AI affects innovation and the knowledge production process. Relative to the rich set of theoretical frameworks (Cockburn, Henderson, and Stern, 2019; Agrawal, McHale, and Oettl, 2024; Ide and Talamàs, 2025; Jones, 2025; Mullainathan and Rambachan, 2025; Agrawal, McHale, and Oettl, 2026), the body of empirical evidence remains relatively limited. Existing empirical work primarily examines AI's impact on product innovation within firms (Lou and Wu, 2021; Bessen et al., 2022; Babina et al., 2024) or on knowledge created in online communities such as StackOverflow and Wikipedia (Burtch, Lee, and Chen, 2024; Lyu et al., 2025; Shan and Qiu, 2025). By contrast, this paper focuses the production of scientific knowledge, the upstream input for product innovation, with a strong public good property as well. Whereas some previous studies have raised concerns that prediction technologies could induce inefficient research herding toward targets where predictions are most accurate (Khurana, 2023; Hoelzemann et al., 2024; Kim, 2025), my results point to the opposite direction: research effort shifts away from the best predicted regions and towards not well predicted ones. Moreover, this paper also contributes to the literature on how access to research tools impacts knowledge production (Furman and Stern, 2011; Murray et al., 2016; Helmers and Overman, 2017; Teodoridis, 2018; Furman and Teodoridis, 2020), by showing that AI-based predictions may shift scientists' target choice differently based on what experimental method is used.

Lastly, although AlphaFold's impact has garnered substantial attention within the natural science community (Bertoline et al., 2023; Varadi and Velankar, 2023; Kovalevskiy et al., 2024), most discussions focus on evaluating its reliability in different contexts and its potential integration into future scientific discovery. However, less attention has been given to its effect on knowledge production from human labs in structural biology, an aspect that may be of greater interest to economists and may guide policymakers in designing better science policy. To my knowledge, only a handful of studies have attempted to address this gap. Cavalli (2024) studies the change in

lab organization after the emergence of AlphaFold1[1] and Yu (2026) studies how AF2's introduction affect scientists' productivity and find highly cited scientists are more likely to adopt AF2 and leads a widening inequality gap. The paper most closely related to mine is Hill and Stein (2026), which also studies how AF2 reshapes the direction of knowledge production. However, the two papers differ in both emphasis and identification. Their analysis centers on AF2's downstream spillovers to subsequent biology and pharmaceutical research, whereas mine focuses on the upstream experimental choices on which proteins to study and the constraints shaping those choices. The empirical strategies also differ: I define exposure based on AI capability, measured by how well AF2 can predict a given protein, while they define it based on prior human knowledge, namely whether a protein has been experimentally solved before. Taken together, these two papers offer a more complete view of how AF2 redirects knowledge production at different stages of the research pipeline.

The remainder of the paper is organized as follows. Section 2 provides relevant background knowledge on structural biology and introduces how AF2 revolutionized this field of study. Section 3 describes my data sources and the construction of key measures used in analysis. Section 4 provides descriptive statistics on the trend and composition of aggregate knowledge production in structural biology. Section 5 presents the empirical strategy and main findings. Section 6 concludes with a discussion on potential directions for future work.

2 Background

2.1 Development in Structural Biology: Experiment and Prediction

Structural biology is the study of the three-dimensional (3D) structures of biological macromolecules, including deoxyribonucleic acid (DNA), ribonucleic acid (RNA), and, most commonly, proteins. Proteins are often described as the "building blocks of life" because they perform essential roles

[1] AlphaFold1 was the first version of AlphaFold that was introduced in 2018 as DeepMind's entry to the CASP13 protein-structure prediction competition. Unlike AlphaFold2, only the competition results and a high-level description of this early version was made public. Meanwhile, its accuracy still fell well short of experimental methods.

at every stage of biological processes—providing nutrients for embryonic development (e.g., albumin), catalyzing chemical reactions in living organisms (e.g., enzymes), and defending against foreign pathogens (e.g., antibodies). As these diverse functions are fundamentally determined by the unique 3D structures of proteins, structural biology has emerged as a cornerstone of the life sciences, providing the essential foundation for research that connects molecular functions to practical applications. Since 1901, dozens of Nobel Prizes have been awarded to studies related to structural biology (PDB-101, 2024).

Several different experimental methods have been developed over the history to study the 3D structure since proteins are too small to be directly observed. Among these, X-ray crystallography has been widely accepted as the mainstream. This method involves three main steps: crystallization (growing a suitable protein crystal), data collection (recording X-ray diffraction patterns), and structure solution (analyzing the diffraction data to determine atomic positions in 3D). However, solving protein's structure following this method can be a laborious process with lots of uncertainty. First, growing protein crystals suitable for X-ray diffraction is a slow and highly unpredictable, trial-and-error process. Some structures may take years before crystals can be obtained, if at all, (Gorrec, 2021) and the overall success rate from target selection to solved structure is only around 5% (Slabinski et al., 2007). Even after the crystal is generated, the refinement on structure can also be a "tedious" process (Strasser, 2019). Scientists may go back-and-forth fitting the structural model with experimental data until they decide the result is "good enough" (Brown and Ramaswamy, 2007).

In recent years, Cryo-electron microscopy (Cryo-EM) has become a popular alternative to X-ray crystallography. Cryo-EM flash-freezes biological samples in a thin layer of vitreous ice, which preserves their native structures without the need for crystallization (Doerr, 2016).[2] As Figure I illustrates, the remaining steps are largely similar to X-ray crystallography (Goodsell, 2019) though specimen preparation and data analysis still requires several days to months of computation and expert curation (Grassucci, Taylor, and Frank, 2007; Thompson et al., 2019). Moreover,

[2]Cryo-EM is particularly useful for proteins that are difficult to crystallize, such as large complexes and membrane proteins (Wang and Wang, 2017; Shoemaker and Ando, 2018).

Cryo-EM is far more capital intensive: instrument installation can cost over $5 million with a continuing expensive maintenance fee (Fernandez-Leiro and Scheres, 2016). By contrast, installation of an academic-style X-ray diffraction system is typically quoted in the $180,000–$350,000 range, and per-sample charges at major university core facilities are often only $200–$500.[3] This cost gap makes Cryo-EM prohibitively expensive for many labs, while X-ray crystallography remains widely accessible.

Alongside advances in experimental methods, scientists have also pursued an alternative path: developing automation tools to predict a protein's 3D structure directly from its one-dimensional amino acid sequence. Since sequencing methods had been standardized by the 1970s (Sanger and Thompson, 1953; Edman and Begg, 1967), such tools in principle could allow scientists to bypass the costly and time-consuming experimental process altogether. However, prior to 2018, even the most advanced models remained far from achieving accuracy comparable to experimental methods (Kryshtafovych et al., 2019).

The introduction of AF2 marks a major breakthrough in this direction. Leveraging deep learning algorithms, AF2 is the first method that achieves prediction accuracy comparable to laboratory experiments for a substantial share of proteins (Callaway, 2020). In July 2021, DeepMind published a paper in *Nature* that open-sourced AF2's algorithm, enabling scientists worldwide freely access to this powerful method (Jumper et al., 2021). That paper quickly became one of the most cited publications, accumulating over 30,000 citations in less than 3 years. In October 2024, the main developers of AlphaFold, Demis Hassabis and John M. Jumper were awarded the Nobel Prize in Chemistry for their contributions in protein structure prediction.

[3]The installation quote is drawn from Rigaku, a leading manufacturer and distributor of scientific, analytical and industrial instrumentation specializing in X-ray related technologies, including X-ray crystallography. The sample cost is estimated based on posted fee schedules from several representative U.S. universities including California Institute of Technology, Johns Hopkins University, and The University of Kansas.

2.2 Unique Characteristics of Structural Biology

Structural biology has emerged as a particularly informative setting for studying innovation and the economics of science recently (Hill and Stein, 2025a,b; Kim, 2025; Zhuo, 2025). This field offers several unique advantages. First, experimental projects in this field are highly comparable: scientists worldwide share a common objective—determining molecular structures through standardized experimental procedures, with the structure itself serving as the project's final output. Second, this field is exceptionally well archived: rich, standardized data are stored in repositories such as the Protein Data Bank (PDB), where each structure's target, timing, method, quality, and outcome are recorded in a consistent and publicly accessible format. Moreover, these structural records can also be linked to complementary datasets, such as the publication records and drug development applications, that provide deeper insight into the direction and progress of scientific innovation.

Combined with the comprehensive coverage of experimental projects, the introduction of AF2 in structural biology offers a unique context to study the impact of AI on scientific knowledge production following a difference-in-differences analysis. The public release of AF2 constitutes a sharp, exogenous technological shock with clearly defined timing. Moreover, variation in the confidence of AF2's predictions across proteins creates natural heterogeneity in exposure: some proteins are predicted with near-perfect accuracy, comparable to the most precise laboratory results,[4] while others remain relatively poorly predicted. This variation enables the construction of treatment and control groups.

[4] According to Tunyasuvunakool et al. (2021), among protein structure prediction for the human proteome, around 36% meet this standard.

3 Data

3.1 Data Sources

The Protein Data Bank (PDB) Established in 1971, the Protein Data Bank (PDB) is a comprehensive repository in structural biology that archives experimental projects worldwide. These projects primarily focus on determining the structures of large biological molecules, 92% of which are proteins.[5] When the exploration is completed, depositing the structure in the PDB has become a necessary step, as the scientific community has required this practice for anyone before submitting publications or grant applications since the 1990s. In this process, a unique PDB ID is assigned to a certain structure deposited and a wide range of information about the structure is recorded. I mainly leverage the following aspects: target, which includes specific proteins being studied and the corresponding biological features; execution, which includes the timeline (e.g., deposition/release dates), method (e.g., X-ray crystallography/Cryo-EM), quality (e.g., resolution), and personnel (e.g., structure authors/citation authors) involved in the experimental discovery of the target; and citation, which links the structure to the primary academic paper that cites it, if available.

The AlphaFold Protein Structure Database (AFDB) When DeepMind open-sourced AF2 in July 2021, they also launched the AlphaFold Protein Structure Database (AFDB) to make the predicted protein structures publicly accessible.[6] The database aims to lower technical barriers and accelerate scientific discovery by providing ready-to-use predictions without the need for complex software setup.[7] Each prediction has been assigned a unique AFDB ID. AFDB also provides a mean pLDDT score for each prediction, which serves as a global metric indicating its confidence

[5] PDB Data Distributions, by Polymer Entity Type

[6] Case study: AlphaFold uses open data and AI to discover the 3D protein universe

[7] Since the initial launch in July 2021, AFDB experienced several rounds of expansion and by July 2022 it includes over 214 million predicted structures, covering all the known unique protein sequences as of April 2021 (Varadi et al., 2022). See Figure A.1 for an illustration.

in the average accuracy across all local regions within that predicted structure.[8] I leverage this variation in prediction quality across proteins to construct the treatment indicator. Further details are discussed in Section 3.2.

The UniProt Knowledgebase (UniProtKB) The UniProtKB provides a comprehensive catalog of nearly all publicly known protein sequences and their functional annotations.[9] It is a superset to AFDB, covering both predicted proteins and newly discovered ones not covered by the AFDB.[10] It also has broader coverage than the PDB, which only records experimentally determined structures. Meanwhile, each UniProtKB entry lists all corresponding PDB IDs, as well as the related AFDB IDs if available. This cross-referencing enables direct linkage between experimental structures and predictions through the same protein being studied.

Other Data Sources To evaluate the downstream scientific and commercial influence of protein structures, I supplement the sample with data from the Web of Science and DrugBank, following Hill and Stein (2025b). The Web of Science provides bibliographic records that allow measurement of publication outcomes if a specific structure is published in a paper that is recorded in the PubMed database.[11] DrugBank further lists PDB IDs associated with drug targets, enzymes, transporters, or carriers, allowing identification of structures with potential commercial applications.

[8] The rule of thumb is based on local pLDDT score as each prediction of protein may have certain regions predicted with high-confidence or low-confidence. I follow Jumper et al. (2021) to use the same rule of thumb for global metrics. More details on rule of thumb and how to interpret it can be found in the FAQ of AFDB.

[9] Some sequences are excluded, such as synthetic constructs, patent submissions, or redundant proteomes. Nonetheless, UniProtKB remains the best available proxy for the universe of known proteins. There are two sections in the UniProtKB: a reviewed section called Swiss-Prot, which includes over 570,000 high-quality, non-redundant protein sequences manually annotated by biological experts; and an unreviewed section called TrEMBL, which documents all other protein sequences not yet manually checked by experts. Once a protein has gone through the manually check process, it will be passed from TrEMBL to Swiss-Prot. In this paper, I consider both sections to fully capture the entire idea search space for structural biology. See details at UniProtKB Coverage.

[10] AFDB only covers around 214 million proteins known as of April 2021. while the number of proteins recorded in the UniProtKB had expanded to roughly 248 million by May 2024.

[11] The linkage between the Web of Science and the PDB is established through PubMed identifiers (PMIDs), unique IDs assigned by the U.S. National Library of Medicine. Each PDB entry is manually linked to the paper that first reports the structure and includes its PMID, while the Web of Science provides a PMID crosswalk.

3.2 Measuring and Mapping Key Variables

Treatment Indicator The treatment indicator is built upon the prediction confidence score reported in the AFDB. Following the rule of thumb provided by DeepMind (Jumper et al., 2021; Fleming et al., 2025), predictions are classified into four regions with different levels of prediction confidence based on their mean pLDDT scores:

- Very High (VH) confidence if $pLDDT \geq 90$. This means the prediction is expected to have high accuracy, typically consistent with high-resolution experimental data.
- High (H) confidence if $70 \leq pLDDT < 90$. This means the prediction is expected to be good with some misplacement chains.
- Low (L) confidence if pLDDT $50 \leq pLDDT < 70$. This means the prediction should be treated with caution.
- Very Low (VL) confidence if $pLDDT < 50$. This means the prediction should not be interpreted.

When linking predictions to experimental structures in the PDB, it is important to note that the mapping occurs at a more granular level, called "entity". A protein structure in the PDB may represent either a small protein with single entity or a protein complex composed of multiple entities (See Figure III). To obtain a structure-level measure of prediction quality, I calculate the average prediction confidence $pLDDT_s$ score across all entities within a structure, weighted by each entity's amino acid sequence length. This provides a consistent proxy for how well a structure, on average, is predicted by AF2. For entities without a corresponding prediction, I assign a confidence score of zero, reflecting the absence of predictive information.[12]

In the baseline analysis, a structure is classified as "treated" if its average confidence score falls within the "VH" region ($pLDDT_s \geq 90$), where AF2 predictions are nearly indistinguishable

[12]Such cases include proteins outside AF2's prediction coverage, including isoforms, sequences with non-standard or unknown amino acids, and proteins shorter than 16 amino acids or too long to be included as full-length entries in Varadi et al. (2022). They may also include proteins that were newly discovered after April 2021.

from experimental accuracy. In the sensitivity analysis, I include a second treatment group for those structures in the high-confidence ("H") prediction region ($70 \leq pLDDT_s < 90$). The remaining structures, those with lower-confidence predictions or outside AF2 coverage, are grouped into a residual category ("R"). The baseline specification pools the H and R groups into a single "Non-VH" control group.

Quantity of Knowledge Production Quantity measures are constructed to capture the volume and composition of human-produced knowledge. I begin by counting the total number of structure deposits into the PDB each quarter. This provides a baseline measure of total quantity of knowledge produced by structural biologists. To further characterize the composition of experimental effort, I construct several subgroup measures. First, I identify "first-attempt" and "non–first-attempt" structures—those that experimentally determine a protein's structure for the first time versus follow-up studies on previously attempted proteins.[13] This comparison reveals whether the release of AF2 shifted the balance between exploratory projects (first attempts) and exploitative projects (non–first attempts). Second, I distinguish structures targeting human versus non-human proteins to assess whether effort increasingly concentrates on human proteins, which are more likely to have high downstream demand. Third, I also document the composition of these different types of structures varies by experimental method. Together, these measures provide a comprehensive view of how AI has reshaped experimental priorities and research efforts of different types of studies in structural biology.

[13]To measure experimental "attempts", I use the clustering algorithm provided by the PDB, which groups entities into clusters based on amino-acid sequence similarity. Following Brown and Ramaswamy (2007) and Hill and Stein (2025b), I treat entities within the same 50% sequence-similarity cluster as representing the same underlying protein. Within each cluster, I rank entities by release date and define an entity as a "first-attempt" entity if it is the earliest-released entity in that cluster. I then map this entity-level measure to the structure level by classifying a structure as a "first-attempt" structure if it contains at least one "first-attempt" entity. Consequently, for multi-entity complexes, a structure in which all entities are novel and a structure in which only one among several entities is novel are both coded as "first-attempt" under this baseline definition. In the robustness check, I implement several alternative definitions and samples, including restricting the analysis to single-entity structures only and redefining "attempts" using alternative sequence-similarity thresholds.

Quality of Knowledge Production Building on the analysis of quantity, I construct quality measures to evaluate the quality of structures deposited, including informational value and innovation contribution. To measure informational value, I use resolution reported in the PDB which reflects structure's execution precision, where lower values indicate more detailed and reliable determinations. To assess innovation contribution, I examine both the scientific influence and commercial relevance. To measure scientific influence, I rely on each structure's primary citation record in the PDB. I first identify whether the corresponding paper is published and indexed in the PubMed. Conditional on publication, I construct two separate indicators of scientific influence: one for whether the article appears in a journal in the top decile of Journal Impact Factor (JIF), and another for whether it is published in Cell, Nature, or Science (CNS), which are viewed as the pinnacle of scientific publication.[14] Because AF2's release is too recent for citation measures to mature at the paper level, the quality of the publishing journal is used as a short-term proxy to measure a structure's academic influence. Finally, I assess commercial relevance by identifying whether a structure is cross-listed in DrugBank, indicating its potential in drug discovery.

Group Structures by Prediction and Class To account for systematic differences in prediction confidence across protein types, I construct a grouping variable ("bin") that combines two time-invariant characteristics of each structure: AF2 prediction confidence (VH, H, R) and 85 functional classification reported in the PDB (e.g., antibiotic, binding, membrane, transport). This grouping aligns the definition of treatment intensity with biologically meaningful comparisons.

In the following analysis on quantity, I construct a balanced panel with bin as the unit of analysis, which represents a specific domain with a distinct level of AI prediction. In the analysis on quality, which is conducted at the structure level, I include bin fixed effects to absorb unobserved, time-invariant heterogeneity across biological domains that may systematically influence scientific difficulty, demand, or publication outcomes. Together, this design ensures that identification accounts for both time-based trends and time-invariant pre-existing group differences.

[14] JIF measures the average number of citations received by articles published in a particular journal over a certain period. I use the most recent 2024 JIF for the analysis.

3.3 Sample Construction

The sample construction begins with 87,276 structures deposited in the PDB between January 1, 2018, and June 30, 2024 executed either by X-ray crystallography (63,978; 73.3%) or Cryo-EM method (23,298; 26.7%).[15] This window yields a reasonably balanced pre- and post-period around AF2's July 2021 release. I end the sample in 2024 Q2 to isolate the effects of AF2 from those of AlphaFold3, which was released in the end of that quarter.[16] The cutoff also reduces missing-data concerns generated by the PDB's release policy, under which deposited structures become public only when the associated paper is published or after one year has elapsed since deposition.[17] I drop 1,172 structures with no protein entities and a few hundred structures for which the author field lists only an institution. When constructing bins, I only keep classes of proteins that have some structures deposited from each of the three prediction regions (VH, H, R) to make sure each bin had at least one deposit within this period of time. Finally, I construct a balanced panel with 255 bins (85 classes $\times$ 3 prediction regions) in 26 quarters (2018 Q1-2024 Q2) containing 84,154 unique structures.

I provide the baseline summary statistics with involvement in the "VH" prediction region as treated and those in the "H" or "R" ("Non-VH") as control below. Panel A of Table I presents quantity outcomes at the (prediction region $\times$ structure class) bin level. We observe that the magnitude of knowledge produced, either measured by total number of structures deposited or papers published per cyear-quarter are comparable, suggesting they may share similar level of academic attention. The fraction of "first-attempt"s is similar across groups and the fraction of human structures is slightly higher in the control group. Panel B of Table I presents quality outcomes at the structure level by different methods. We see structures in the treatment group are better executed while structures in the control group have better publication records if published.

[15] Over this period, 97.3% of all deposited structures follow one of these two methods.

[16] In May 2024, DeepMind released AlphaFold 3, which expanded its predictive capabilities to include the structures and interactions of all of life's molecules—not only proteins, but also DNA, RNA, ligands, and more are included (Abramson et al., 2024).

[17] Once released, the PDB records both the release date and the original deposit date, which corresponds to when the research team judged the structure solved and submitted it. The raw data were downloaded from the PDB on August 26, 2025, and therefore capture all structures released by that date.

4 Descriptive Statistics

This section provides a descriptive overview of how the aggregate knowledge production in structural biology evolved around the introduction of AF2. From the empirical trends and cross-sectional distributions, four facts are summarized.

Fact 1: Total deposits keep rising, with a shift away from well-predicted proteins.

Figure IV illustrates both the overall increase in structural depositions and the compositional shift in prediction confidence following AF2's release. As Figure IVa shows, the number of structures deposited per quarter has been rising consistently over this period, from an average of 2,949 in the pre-AF2 period to 3,572 afterward, representing a 21.1% increase. Figure IVb illustrates a compositional shift away from depositing structures in regions with (VH) prediction confidence. This figure compares the distribution of prediction confidence scores ($pLDDT_s$) for experimentally deposited structures before and after AF2's release in my sample. The VH share of total depositions declines by 9.5 percentage points, from 38.0% to 28.5%, after AF2's introduction. To interpret where the research activity reallocates, Figure IVc provides a benchmark: the global distribution of AF2 pLDDT scores across the roughly 214 million proteins within AF2's coverage. Relative to this global profile, the reallocation of effort in Figure IVb is not driven by large shifts towards regions where AF2 performs poorly (L and VL) or where predictions are moderately good (H but not VH). Instead, the largest increase is concentrated in proteins outside AF2's scope: the out-of-coverage share rises by 5.5 percentage points, from 14.8% to 20.3%.

Fact 2: A rise in Cryo-EM use and heterogeneous impact across experimental methods.

Figure V presents trends in structure depositions by experimental method. Figure Va shows that the overall rise in total depositions is mainly driven by the Cryo-EM adoption takeoff: average quarterly Cryo-EM depositions has more than doubled, increasing from 504 to 1,318, whereas X-ray crystallography depositions declines slightly, from 2,445 to 2,255 (-7.8%). Figure Vb and Vc indicate that AF2's introduction affects research activities unevenly across experimental methods.

Before AF2's release, the target distributions already differed systematically: X-ray crystallography was more concentrated in regions that AF2 would later predict with very high confidence, while Cryo-EM was relatively more concentrated in harder-to-predict regions. This pre-period pattern suggests a form of sorting into different regions of the idea search space across methods, with the capital-intensive Cryo-EM technology being used on targets harder to predict. Following AF2's release, the labor-intensive X-ray crystallography exhibits a clear shift away from the very high-confidence region, where AF2 is closest to a direct substitute for experimentation.[18] Cryo-EM, in contrast, shows little change in its confidence distribution, suggesting that prediction confidence exerts little influence on scientists' choices of where to exert experimental efforts.

These patterns suggest that AF2 may be labor-augmenting for X-ray crystallography by complementing experimental workflows on projects that remain worth pursuing. At the same time, because X-ray is also more exposed to substitution on well-predicted targets, its aggregate growth remains limited. Cryo-EM, by contrast, is less directly displaced by AF2 and continues to be valuable for targets that are harder to predict. As a result, AI may reallocate X-ray effort toward more difficult problems while aggregate expansion in the field is primarily driven by the capital-intensive Cryo-EM method.

Fact 3: The exploratory deposit remains steady, while the exploitative deposit grows.

Figure VI compares the composition of exploratory versus exploitative research, measured by whether a structure is identified as "first-attempt". As shown in Figure VIa, the trends in "first-attempt" and "non–first-attempt" structures are largely parallel prior to AF2's introduction. After the release, the production of "first-attempt" structures remains relatively stable, while the number of "non–first-attempt" structures rises: average quarterly depositions increase from 2,304 to 2,858 (+24%), indicating an expansion in exploitative work without a corresponding decline in exploratory work. Figures VIb and VIc further show a shift toward less well-predicted areas, with this reorientation more pronounced among "first-attempt" projects: the share of depositions from

[18] Hill and Stein (2026) provide a possible explanation for why AF2 may complement X-ray crystallography more strongly: AF2 predictions can be directly incorporated into X-ray workflows through molecular replacement, where they serve as starting templates for structure solution.

moderately well-predicted regions (H but not VH) also declines.

Fact 4: The typical experimental team size is stable within continuing projects.

In PDB, each structure lists the authors credited with determining and depositing it, and I count the number of authors per structure to measure team size. Because the distribution of team size is highly skewed—with a few structures involving unusually large teams—the interquartile range (IQR) and median of team size in each quarter is used for comparison. Figure VII shows that structures across prediction regions exhibit similar ranges and temporal patterns in team size. This pattern suggests that AF2 does not meaningfully alter the typical organizational scale of the teams producing structures, regardless of the experimental method used or whether the target had been attempted previously. Across all panels, the median and IQR remain broadly stable within each prediction-confidence region.

These descriptive patterns suggest that AF2's introduction is associated with higher overall research output, a reallocation of effort toward regions predicted less confidently by AF2—especially targets outside its coverage—and substantial heterogeneity across experimental methods. At the same time, AF2 appears to affect marginal project selection much more than inframarginal project organization. Taken together, these observations motivate the empirical analysis that follows, in which I formally estimate AF2's causal impact on the quantity, composition, and quality of human-produced knowledge in structural biology.

5 Empirical Strategy and Results

5.1 Quantity and Composition of Knowledge Production

This section aims to evaluate how AF2 affects the quantity and composition of human-produced knowledge in structural biology. The estimation equation follows a generalized difference-in-differences model:

$$y_{bt} = \beta(Post_t \times VH_b) + \gamma_b + \delta_t + \varepsilon_{bt} \tag{5.1}$$

The outcome variable y_{bt} is the count of certain types of projects deposited from (prediction region $\times$ structure class) bin b at year-quarter t of deposit. The independent variable of interest is the interaction term between two binary indicators: time indicator $Post_t$ that turns 1 since 2021 Q3 when AF2 is publicly released, and treatment indicator VH_b that turns 1 if the prediction branch of a certain bin belongs to the very high confidence region. Bin fixed effects γ_b and year-quarter fixed effects δ_t are included in all the specifications. ε_{bt} is the error term with a conditional mean of 0.

The coefficient of interest is β. If $\beta > 0$, that implies more deposits are produced working on proteins with predictions on par with the experimental results, suggesting AF2 shifts the direction of knowledge production towards areas that are best predicted. However, if $\beta < 0$, that implies scientists have shifted away from producing knowledge in regions that AF2 has the highest prediction confidence.

Table II reports the estimates of Equation 5.1 with total count of structure deposits as the outcome. Column (1) reports the result with a Poisson pseudo-maximum likelihood (PPML) estimator. This is used as the main specification since the outcome is a count variable and the conditional mean is modeled as $E[y_{bt}|\cdot] = exp(\beta(Post_t \times VH_b) + \gamma_b + \delta_t)$. Column (2) reports the OLS estimator with the common linear index in levels. Estimation results from Column (1)-(2) indicate that bins exposed to the very high confidence prediction region of AF2 experienced a substantial decline in total structure depositions, approximately 34.8% to 42.8% relative to those that were not. Columns (3)-(4) repeat the estimation under the OLS specification using inverse hyperbolic sine transformed total count as the outcome variable. In column (4), the regression is also weighted by the total number of structures deposited over the pre-treatment period in each class in order to make sure the finding is not driven by classes that account for a small share of total knowledge production. The coefficient of interest is still negative and significant with similar magnitude. Results across columns suggest a reallocation of scientific effort away from regions that AF2 predicts very well towards those it does not.

Moving forward, I evaluate how the composition of human-produced knowledge changed, reflected by the deposit counts of different types of structures. Table III reports the results from the

PPML specification. The impact on total deposit is listed in Column (1) as a benchmark. Columns (2)-(3) compare deposits of "first-attempt" (exploratory) versus "non–first-attempt" (exploitative) structures. Deposits in the VH region declined for both categories, but the drop is larger for "first-attempt" structures (–39.8% versus -33.7%). This suggests that when AF2 already offers highly accurate predictions, scientists are less likely to experimentally explore unattempted structures in this region. Columns (4)-(5) compare deposits of structures targeting human versus non-human proteins. Deposits for non-human proteins in the VH region fall more steeply, indicating that researchers scaled back on less central, well-predicted proteins while maintaining effort on human targets. Columns (6)–(7) further split human proteins into those that were previously "stuck", historically difficult to resolve, and those that were not.[19] Deposits for "stuck" human proteins fall by 29.4% in the VH region, compared with a 14% decline for non-stuck proteins, but both estimates are imprecise with large standard errors. These results are best read as at most weakly suggestive, rather than informative evidence, that AF2 may reduce experimental incentives for the most challenging targets when high-confidence predictions are available.

Beyond studying how AF2 affects the types of structures scientists study, I further evaluate whether it changes the typical labor intensity of experimental work for the structures that continue to be pursued. I measure labor intensity as the median number of authors involved in solving a given type of structure within each bin–deposit-quarter cell. If AF2 also reduced the labor required for experimentation, we would expect smaller teams for structures with predictions in the VH region. However, the results in Table IV show only marginal differences in median team size between VH and non-VH bins, with all coefficients close to zero and statistically insignificant.[20]

In order to check the pre-existing trends in structure deposit and explore the dynamic relationship between AF2's prediction and knowledge production, I also estimate the following dynamic

[19]To identify a set of "stuck" proteins that were poorly resolved experimentally before AF2's release, I follow benchmark standards established by structural biologists: for X-ray crystallography, resolutions above 2.5 Å are generally unsuitable for drug design (Anderson (2003)), while for Cryo-EM, resolutions above 3.5 Å are difficult to interpret (Casanal et al. (2019)). I compute the median pre-AlphaFold resolution for each class–method pair in my sample. If this median exceeds the corresponding benchmark, structures under these class–method pairs are labeled as "stuck".

[20]For each structure type, the analysis is restricted to bins with at least one deposits each year-quarter over the sample period to ensure consistent within-bin comparisons.

difference-in-differences model:

$$y_{bt} = \sum_{\tau \in \Gamma^{Pre}} \beta_\tau (\mathbf{1}\{t = \tau\} \times VH_b) + \sum_{\tau \in \Gamma^{Post}} \beta_\tau (\mathbf{1}\{t = \tau\} \times VH_b) + \gamma_b + \delta_t + \varepsilon_{bt} \tag{5.2}$$

In this equation, set Γ^{Pre} includes all the time periods in our sample from the pre-treatment period, except for 2021 Q2, the quarter right before AF2's release. Set Γ^{Post} includes all the time periods in our sample since AF2's release. If the β_τs are not statistically distinguishable from zero when $\tau \in \Gamma^{Pre}$, it supports the parallel trends assumption for our main estimation. When $\tau \in \Gamma^{Post}$ the β_τ identify the dynamic impact of AF2 in year-quarter τ.

Figure VIII presents coefficient estimates from Equation 5.2, under the PPML specification. Overall, pre-treatment coefficients are close to zero and statistically flat for the "first-attempt"s (Figure VIIIa), lending support to the parallel-trends assumption that production of such types of structures exposed to the VH region or not were on similar trends. For the "non-first-attempt"s and total deposits (Figure VIIIb, VIIIc), the pre-treatment coefficients display a modest upward trend prior to AF2's release in the earlier periods, suggesting proteins later classified as with very high confidence predictions were experiencing a slightly faster growth in experimental activity about two years before AF2's release, but not in the more recent pre-treatment periods. Post-release estimates are all negative for the "first-attempt"s, suggesting a persistent decline in experimental output among the treated group.

To sum up, results above suggest AF2 significantly substituted for experimental efforts in regions where prediction confidence is very high. Such substitution is more pronounced, around a 40% decline, for targets with lower net returns of experimentation, either because the expected cost of experiment is too high ("first-attempt"s) or because there's limited downstream demand (non-human targets). Meanwhile, AF2's impact operates primarily through reallocating efforts away from studying these types of structures experimentally, with little evidence of labor-saving effects within the experiments that scientists continue to pursue.

5.2 Heterogeneity by Experimental Methods

A natural concern with the baseline estimates is that structures solved by different experimental method are pooled together. Because X-ray crystallography and Cryo-EM differ in their workflows, capacity constraints, and the types of proteins they typically tackle, a single aggregate coefficient may mask important heterogeneity across methods. To assess this, I reconstruct the dataset at the method-bin-deposit-quarter level and estimate the PPML specification adding a triple interaction between the post-AlphaFold indicator, the very-high confidence group, and a Cryo-EM method indicator:

$$E[y_{mbt}|\cdot] = exp(\beta_1(Post_t \times VH_b) + \beta_2(Post_t \times VH_b \times Cryo_m) + \gamma_{mb} + \delta_{mt}) \tag{5.3}$$

Here y_{mbt} is the count of structure deposit following method m in bin b at deposit quarter t. γ_{mb} are the method $\times$ bin fixed effect. Because Cryo-EM adoption evolves over time, I also allow time shocks to vary by method through including δ_{mt}, the method $\times$ deposit quarter fixed effect. Under this parameterization, β_1 captures the impact of AF2 within the X-ray crystallography group (the omitted method), while β_2 captures the differential impact for Cryo-EM relative to X-ray crystallography. Accordingly, the impact within Cryo-EM group is captured by $\beta_1 + \beta_2$.

Table V reports the heterogeneous effects of AF2 on the composition of experimental knowledge production across methods. Two patterns stand out. First, the decline in "first-attempt" structures is broadly consistent across methods, as Column (2) indicates. Second, within X-ray crystallography there is a comparable overall contraction in total deposits (Column (1)), accompanied by a reallocation away from non-human proteins and from "stuck" targets (Columns (5)–(6)). This suggests that X-ray crystallography disproportionately scales back work on lower-demand or intrinsically difficult targets once high-quality predictions become available. In contrast, Cryo-EM output appears largely insensitive to AF2's prediction confidence: the treatment coefficients are small and imprecise, and there is little evidence of systematic reallocation across structure types within Cryo-EM, aside from "first-attempt" deposits.

These patterns are suggestive from an inequality perspective: AF2 appears to reallocate efforts most strongly in X-ray crystallography, which is relatively more labor-intensive, while leaving high-end, capital-intensive Cryo-EM capacity largely unchanged.[21]

5.3 Quality of Knowledge Production

In this section, I evaluate the impact of AF2 on the quality of experimentally solved structures. The estimation follows a variant of Equation 5.1 under an OLS specification:

$$y_{ibt} = \beta(Post_t \times VH_b) + \gamma_b + \delta_t + \varepsilon_{ibt} \tag{5.4}$$

where y_{ibt} denotes the quality measure of structure i belonging to bin b that deposited in year-quarter t. All other terms are defined analogously to those in Equation 5.1, but the unit of analysis is now at the individual structure level. Because each structure is observed only once, the dataset analyzed in this Section is repeated cross-sectional and results in this section should be interpreted with caution. As Section 5.1 documents the compositional shifts in the types of projects deposited after AF2's release, the estimated coefficient here reflects a combination of (i) the causal effect of AF2 on the quality of inframarginal structures—those that continue to be experimentally solved despite receiving accurate predictions—and (ii) selection differences in quality between marginal structures that are no longer attempted and those that remain.

Table VI reports the estimated impact of AF2 on the execution precision of experimental structures, measured by their resolution (lower values indicate higher quality). Because resolution is not directly comparable across methods, I estimate the model separately for structures solved by X-ray crystallography and Cryo-EM. Panel A shows limited effect of AF2 on structures determined by X-ray crystallography. Among different types of structures, there's only a modest improvement about 4.6% is observed for treated structures working on human proteins and 4.7% for human pro-

[21] Meanwhile, the method-specific coefficients are all smaller in magnitude than the baseline pooled estimates. This is expected since the aggregate specification captures the combined response across all methods as well as the expansion of Cryo-EM adoption. More Cryo-EM structures are classified to the control group.

teins that are "non-stuck". For non-human proteins, the quality of resolution decreases by 3.3%. Meanwhile, Panel B shows no significant change in resolution for Cryo-EM structures, suggesting that AF2 had little direct effect on the precision of Cryo-EM based experimental outcomes.

I proceed to examine how AF2 affects a structure's innovation outcome, including both the academic influence and the commercial relevance. For academic influence, it is first measured whether a structure's primary citation is published in an academic paper cross-listed in the PubMed database. Then conditional on publication, I evaluate the corresponding journal's academic influence based on the scientometrics indices introduced earlier. As one paper may include multiple structures, when building the structure-paper link, I only include pairs with papers containing structures originating from a same single bin to ensure that the mapped publication reflects a consistent treatment exposure, avoiding ambiguity from papers that span structures belonging to different prediction-confidence regions or functional classes. 81.8% of published papers in my sample satisfy this filter. For the commercial relevance, it is measured by whether a structure is cross-listed in the DrugBank.

Table VII reports the estimation results. Panel A includes all the structures and I observe the introduction of AF2 shows only marginal effects on whether a structure is cross-listed in PubMed, DrugBank, or published in the most prestigious CNS journals. However, Columns (3)-(4) indicate that structures linked to proteins with very high prediction confidence are, on average, 4.9 percentage points less likely to publish in journals within the top decile of the Journal Impact Factor (JIF) distribution. To better understand the source of this publication penalty, I re-estimate the specification separately for "first-attempt" and "non–first-attempt" structures which may have different objectives: the "first-attempt"s represent exploratory efforts, aimed at solving previously unknown structures, whereas the "non-first-attempt"s reflect follow-up or exploitative research focused on functional characterization or application. Clear heterogeneity is observed: in Panel B, which only includes the "first-attempt"s, all estimated effects for publication-related outcomes are negative and larger in magnitude, with the probability of publication in a top-decile JIF journal decreases by 11.2 percentage points. By contrast, Panel C shows no comparable decline among the "non-

first-attempt"s.

This pattern suggests that a potential mechanism to explain the shift-away from the well-predicted region: AF2 transforms the nature of "first-attempt" projects in this area from exploration to validation. Under the current reward structure, such validation work receives less academic credit: for structures with highly accurate predictions, the publication premium for being the first to experimentally solve a previously unknown structure largely disappears.

5.4 Robustness Checks

In this section, I check the robustness of our main result, the compositional change in knowledge production. First, I evaluate if the impact is driven by other concurrent shocks around the release of AF2 or a small number of protein classes that attract disproportionate attention. Second, I alter how treatment and "first-attempt"s are defined and see if we can still observe the similar result. Finally, I incorporate an analysis with multiple treatment intensities.

Alternative Drivers Around the release of AF2 in 2021 Q3, there's also a surge in biomedical research related with COVID-19. In Panel A of Table A.1, I identified the protein entities originating from COVID-19 using the taxonomy ID and drop all the structures including any COVID-19 protein entities. The results are marginally smaller in magnitude as most of COVID-19 protein entities are beyond AF2's coverage and correspondingly structures involved are mainly assigned to the control group. Still a 31.8% decline is observed for total structure depositions and such shrink is more substantial for the "first-attempt"s (-39.6%). Therefore, the surge of COVID-19 studies is ruled out as the main driver of my result. Then, I evaluate if the results are driven by the shift of research interest in a small set of heavily studied protein classes. In Panel B of Table A.1, I report the results after dropping structures from the top five mostly studied classes, each contain more than 5,000 structures.[22] Still, a contraction of experimental activity in the VH region is observed after the most intensively studied classes are removed. This indicates the treatment effects estimated are

[22]Hydrolases, transferases, viral proteins, oxidoreductases, and membrane proteins are excluded.

not mainly driven by the shift of research interest in a few blockbuster classes, but instead reflect a broader reorganization of experimental work across all fields in structural biology.

Verifying Treatment and Outcome Measurement A second major concern is that the treatment variable is not properly measured at the structure level. Because the existence of structures consists of multiple protein entities, the main specification aggregates entity-level pLDDT scores into a structure-level measure using a weighted average with entities' amino acid sequence length as the weight. If alternative aggregation rules lead to materially different classifications of "treated" structures, the estimated effects may reflect measurement choices rather than the differences in how well each structure is predicted by AF2.

To assess whether the main findings are sensitive to how treatment is defined, I implement three complementary robustness exercises and report the results in Table A.2. Panel A restricts the sample to single-entity structures, where treatment assignment is directly observed and no aggregation is required. The estimated effects remain largely unchanged. The main deviation is in Column (6), where the substitution away from "stuck" human proteins becomes weaker. This likely reflects the fact that most structures classified as "stuck" are multi-entity complexes, which are excluded in this exercise. Panel B evaluates an alternative aggregation rule by replacing the baseline length-weighted average pLDDT with the median, giving equal weight to all entities within a complex protein structure. The magnitude of the effects becomes slightly smaller, but the qualitative conclusions remain the same. Finally, Panel C assigns treatment based on the maximum entity-level pLDDT within each structure. This is the most broad definition: structures are considered "treated" if they include any entity with very high prediction confidence. Even under this classification, I continue to observe an overall contraction in deposits, especially for "first-attempt" structures.

I also re-evalute the measurement of "first-attempt" structures. In the PDB, proteins are grouped into "clusters" based on sequence similarity, and in the baseline I classify a structure as "first-attempt" if it contains any entity that is the first release within a 50% similarity cluster. To test whether this definition drives the results, I re-create the indicator using two alternative

thresholds: a more permissive 30% similarity cutoff and a stricter 100% identity cutoff. As the similarity threshold increases, clusters become narrower, so more structures are mechanically labeled as "first-attempt". Table A.3 shows that the main result persists: the decline in "first-attempt" structures in the VH region remains large and statistically significant even under the 100% identity threshold. This confirms that the baseline results are not an artifact of missing near-duplicate structures or marginal cases around the similarity cutoff.

Treatment Intensity Finally, I extend the analysis for multiple treatment intensities. In the baseline specification, only structures in the VH region are viewed as "treated", while all others serve as controls. This binary definition may raise two concerns: first, pooling moderately predicted (H) regions with poorly predicted (R) ones as the control group may underestimate the impact if AF2 meaningfully changes scientists' behavior even with intermediate confidence; second, including regions beyond AF2's prediction coverage in the control group may overestimate effects if AF2 induces a positive spillover of experimental effort toward proteins where predictions are unavailable.

To address both concerns, I re-estimate the model with a multi-category treatment design. In addition to the VH group, I introduce (i) an indicator for high-confidence region (H) and (ii) a separate "ZERO" group for deposits outside AF2's prediction scope. Under this case, the low (L)/very-low (VL) confidence regions that covered by AF2 are used as the control group. The resulting specification is shown below.

$$E[y_{bt}|\cdot] = exp(\beta_1(Post_t \times VH_b) + \beta_2(Post_t \times H_b) + \beta_3(Post_t \times ZERO_b) + \gamma_b + \delta_t) \tag{5.5}$$

Table A.4 reports the results. Under this design, the substitution effect remains concentrated in the most intensively treated VH region and is slightly larger than in the baseline (about -39.1%), consistent with very-high-confidence AF2 predictions being viewed as close substitutes for experimentally resolved structures. I also find suggestive evidence of a confidence gradient: outcomes decline in the H region along several margins (especially "first-attempt" and non-human), but the magnitudes are smaller than in VH, indicating that AF2's impact is not driven by an arbitrary VH

threshold. Finally, coefficients for the ZERO region are positive but statistically indistinguishable from zero across outcomes, offering little evidence that the baseline VH effect is mainly driven by spillovers toward targets outside AF2's predictive scope. Taken together, these results reinforce the interpretation that the baseline pattern reflects a genuine substitution response that scales with predictive confidence, consistent with effort being reallocated away from targets where confidently predicted structures are readily available.

6 Conclusion

This paper studies how AI affects the quantity, quality, and composition of human-produced knowledge, using the introduction of AlphaFold2 in structural biology as a setting. I find that overall experimental output continues to grow after AF2's release, but its composition changes sharply: experimental effort shifts away from proteins that AF2 predicts with very high confidence, with depositions falling by roughly 35% relative to less well-predicted proteins. This reallocation is strongest for lower-return targets, including proteins that are costly to solve from scratch and non-human proteins with more limited downstream demand. By contrast, AF2 appears to have little effect on the productivity of the projects scientists continue to pursue: conditional on selecting a target, team size and structure resolution change little. We may conclude that, at least in the short run, AI reshapes experimental science primarily through target selection rather than broad within-project efficiency gains.

I also find that adaptation to AI is highly tool-specific: AI may augment labor-intensive methods on the intensive margin while shifting the extensive margin of field growth toward capital-intensive methods. At the same time, AF2 may change the nature of some projects from exploration to validation, thereby reducing the academic credit attached to this work. In conclusion, AI-based prediction "floodlights" such as AF2 may redirect effort away from tasks where prediction can substitute for experiment and toward domains where experiment retains distinct value. Yet who can fully exploit this newly illuminated search space may depend critically on access to

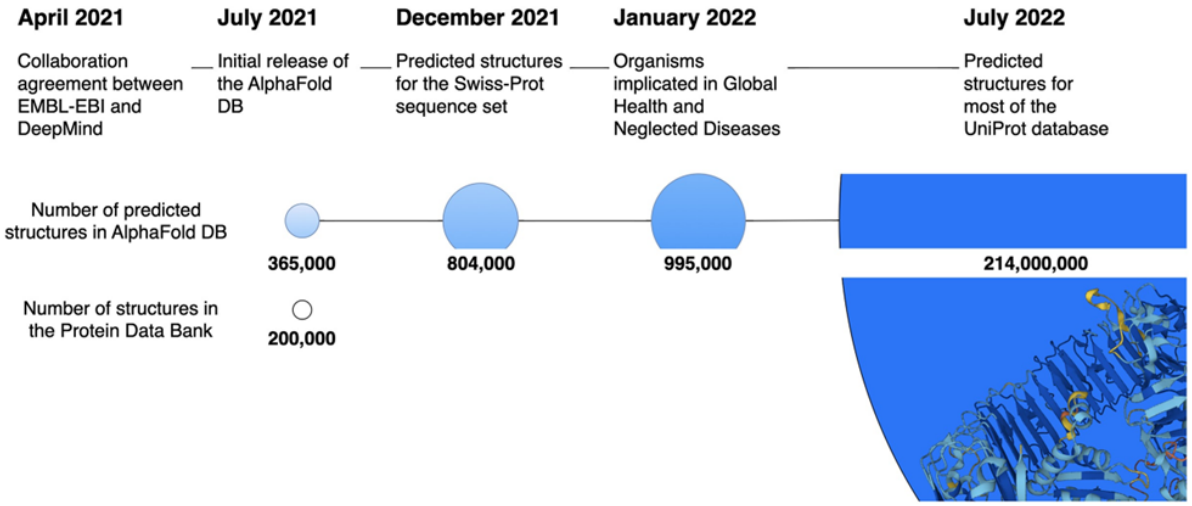
complementary experimental tools, and better incentive design for validation work may be needed to reduce the risk that scientists overlook false positives in areas where AI suggests the predictions are highly reliable.

AlphaFold's success underscores the value of freely accessible, large-scale scientific mapping efforts for innovation, such as the earlier Human Genome Project (Williams, 2013). Yet its longer-run implications for scientific discovery remain uncertain. While this paper shows that AF2 reshapes the allocation of experimental effort, important questions remain about who drives these adjustments, how it affects the organization of scientists with different expertise and access to experimental tools, and how universities, funding agencies, and downstream firms respond to AI breakthroughs in science. More broadly, AF2 represents an early generation of AI that automates a specific scientific task, whereas newer generative AI systems may affect many more stages of the research process, from hypothesis generation to experimental design. Understanding when such systems expand genuinely novel discovery, when they instead redirect effort toward more commercially attractive but less frontier-advancing and less publicly accessible problems, and how institutions can preserve broad access and incentives for validation remains an important agenda for future research.

Author Affiliation: Jiarui (Jerry) Qian: Department of Economics, University of Virginia.

Appendix

Figure A.1: The Expansion of AlphaFold DataBase



Source: (Varadi, Bertoni, Magana, Paramval, Pidruchna, Radhakrishnan, Tsenkov, Nair, Mirdita, Yeo, et al., 2024)

Table A.1: Robustness: Potential Alternative Drivers

Panel A. Drop Structures Related with COVID-19

	Total Deposits	Attempt Status		Origin Status		Stuck Status	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Outcome: Count	-	First	Non First	Human	Non Human	Stuck	Non Stuck
POST × VH	-0.382***	-0.505***	-0.356***	-0.221	-0.451***	-0.332	-0.119
	(0.078)	(0.087)	(0.090)	(0.156)	(0.069)	(0.331)	(0.176)
% Δ (VH vs Non VH)	-31.8	-39.6	-29.9	-19.9	-36.3	-28.2	-11.2
Deposit Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bin FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N of Bins	255	224	250	198	249	126	174
N of Observations	6630	5824	6500	5148	6474	3276	4524

Panel B. Drop Structures from the Top Five Most Studied Classes

	Total Deposits	Attempt Status		Origin Status		Stuck Status	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Outcome: Count	-	First	Non First	Human	Non Human	Stuck	Non Stuck
POST × VH	-0.364***	-0.616***	-0.306***	-0.225	-0.416***	-0.275	-0.066
	(0.093)	(0.120)	(0.115)	(0.231)	(0.071)	(0.252)	(0.251)
% Δ (VH vs Non VH)	-30.5	-46.0	-26.3	-20.2	-34.0	-24.0	-6.4
Deposit Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bin FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N of Bins	240	209	235	183	234	116	160
N of Observations	6240	5434	6110	4758	6084	3016	4160

The unit of observation is bin-deposit-quarter. Panel A excludes all structures that include any protein entity of COVID-19 origin, identified by taxonomy ID. Panel B excludes structures from the five most-studied protein classes: hydrolase, transferase, viral protein, oxidoreductase, and membrane protein. Column (1) reports the total number of structures deposited. Columns (2)–(3) report counts of "first-attempt" and "non–first-attempt" structures, respectively. Columns (4)–(5) report counts of structures from human and non-human proteins, respectively. Columns (6)–(7) restrict to human proteins and report counts for structures from "stuck" protein classes versus those that are not. Bins with no deposits of certain type of structures are dropped in each regression. The specification follows Poisson pseudo- maximum likelihood and %Δ is computed as $100 * (e^{\hat{\beta}} - 1)$. Clustered standard errors at the bin level are reported in parentheses. Statistical significance is indicated as: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A.2: Robustness: Treatment Measurement

Panel A. Structures with Single Entity Only

	Total Deposits	Attempt Status		Origin Status		Stuck Status	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Outcome: Count	-	First	Non First	Human	Non Human	Stuck	Non Stuck
POST × VH	-0.332***	-0.343***	-0.329***	-0.111	-0.439***	-0.070	-0.108
	(0.110)	(0.084)	(0.128)	(0.180)	(0.131)	(0.751)	(0.191)
% Δ (VH vs Non VH)	-28.2	-29.0	-28.1	-10.5	-35.5	-6.7	-10.3
Deposit Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bin FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N of Bins	237	198	229	171	233	63	163
N of Observations	6162	5148	5954	4446	6058	1638	4238

Panel B. Aggregate by Structure's Median pLDDT

	Total Deposits	Attempt Status		Origin Status		Stuck Status	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Outcome: Count	-	First	Non First	Human	Non Human	Stuck	Non Stuck
POST × VH	-0.362***	-0.375***	-0.364***	-0.137	-0.452***	-0.172	-0.092
	(0.091)	(0.084)	(0.105)	(0.151)	(0.095)	(0.354)	(0.172)
% Δ (VH vs Non VH)	-30.3	-31.3	-30.5	-12.8	-36.4	-15.8	-8.8
Deposit Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bin FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N of Bins	246	220	240	198	240	115	177
N of Observations	6396	5720	6240	5148	6240	2990	4602

Panel C. Aggregate by Structure's Max pLDDT

	Total Deposits	Attempt Status		Origin Status		Stuck Status	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Outcome: Count	-	First	Non First	Human	Non Human	Stuck	Non Stuck
POST × VH	-0.185*	-0.246**	-0.167	0.109	-0.316***	0.178	0.047
	(0.109)	(0.096)	(0.123)	(0.150)	(0.116)	(0.273)	(0.174)
% Δ (VH vs Non VH)	-16.9	-21.8	-15.4	11.5	-27.1	19.5	4.8
Deposit Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bin FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N of Bins	246	222	242	197	240	108	178
N of Observations	6396	5772	6292	5122	6240	2808	4628

The unit of observation is bin-deposit-quarter. Panel A restricts the sample to single-entity structures. Panel B aggregates entity-level pLDDT to the structure level using the within-structure median. Panel C aggregates using the within-structure maximum. Column (1) reports the total number of structures deposited. Columns (2)–(3) report counts of "first-attempt" and "non–first-attempt" structures, respectively. Columns (4)–(5) report counts of structures from human and non-human proteins, respectively. Columns (6)–(7) restrict to human proteins and report counts for structures from "stuck" protein classes versus those that are not. Bins with no deposits of certain type of structures are dropped in each regression. The specification follows Poisson pseudo- maximum likelihood and %Δ is computed as $100 * (e^{\hat{\beta}} - 1)$. Clustered standard errors at the bin level are reported in parentheses. Statistical significance is indicated as * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A.3: Robustness: "First-Attempt" Measurement

	First-Attempts			Non First-Attempts		
	(1)	(2)	(3)	(4)	(5)	(6)
Outcome: Count						
Cluster Standard:	30 %	50 % (Main)	100 %	30 %	50 % (Main)	100 %
POST × VH	-0.621***	-0.508***	-0.439***	-0.410***	-0.412***	-0.417***
	(0.102)	(0.086)	(0.086)	(0.105)	(0.111)	(0.121)
% Δ (VH vs Non VH)	-46.3	-39.8	-35.6	-33.6	-33.7	-34.1
Deposit Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Bin FE	Yes	Yes	Yes	Yes	Yes	Yes
N of Bins	217	224	245	250	250	246
N of Observations	5642	5824	6370	6500	6500	6396

The unit of observation is bin-deposit-quarter. Columns (1)–(3) use as the outcome the count of "first-attempt" structures, defined using alternative sequence-similarity clustering thresholds. Columns (4)–(6) use as the outcome the count of "non–first-attempt" structures, defined analogously. Bins with no deposits of certain type of structures are dropped in each regression. The specification follows Poisson pseudo- maximum likelihood and %Δ is computed as $100 * (e^{\hat{\beta}} - 1)$. Clustered standard errors at the bin level are reported in parentheses. Statistical significance is indicated as: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A.4: Robustness: Deposit Composition under Multiple Treatment Intensity

	Total Deposits	Attempt Status		Origin Status		Stuck Status	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Outcome: Count	-	First	Non First	Human	Non Human	Stuck	Non Stuck
POST × VH	-0.495***	-0.648***	-0.460***	-0.307*	-0.630***	-0.362	-0.275
	(0.083)	(0.090)	(0.100)	(0.170)	(0.060)	(0.361)	(0.199)
POST × H	-0.178	-0.291***	-0.143	-0.097	-0.277***	-0.020	-0.183
	(0.113)	(0.107)	(0.129)	(0.169)	(0.085)	(0.229)	(0.209)
POST × ZERO	0.106	0.174	0.093	-0.086	0.018	0.029	-0.130
	(0.180)	(0.127)	(0.208)	(0.238)	(0.171)	(0.549)	(0.230)
% Δ (VH vs VL, L)	-39.1	-47.7	-36.9	-26.4	-46.8	-30.4	-24.0
% Δ (H vs VL, L)	-16.3	-25.3	-13.3	-9.3	-24.2	-2.0	-16.7
% Δ (ZERO vs VL, L)	11.1	19.0	9.7	-8.3	1.8	3.0	-12.2
Deposit Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bin FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N of Bins	327	283	320	239	319	143	212
N of Observations	8502	7358	8320	6214	8294	3718	5512

The unit of observation is bin-deposit-quarter. Column (1) reports the total number of structures deposited. Columns (2)–(3) report counts of "first-attempt" and "non–first-attempt" structures, respectively. Columns (4)–(5) report counts of structures from human and non-human proteins, respectively. Columns (6)–(7) restrict to human proteins and report counts for structures from "stuck" protein classes versus those that are not. Prediction regions VH, H and ZERO are included as three separate treatment groups and region VL and L are the control group. Bins with no deposits of certain type of structures are dropped in each regression. The specification follows Poisson pseudo- maximum likelihood and %Δ is computed as $100 * (e^{\hat{\beta}} - 1)$. Clustered standard errors at the bin level are reported in parentheses. Statistical significance is indicated as: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

References

Abramson, J., J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A. J. Ballard, J. Bambrick, et al. (2024). Accurate Structure Prediction of Biomolecular Interactions with AlphaFold 3. *Nature*, 1–3.

Acemoglu, D. et al. (2022). Artificial Intelligence and Jobs: Evidence from Online Vacancies. *Journal of Labor Economics 40*(S1), S293–S340.

Acemoglu, D. and P. Restrepo (2019). Automation and New Tasks: How Technology Displaces and Reinstates Labor. *Journal of Economic Perspectives 33*(2), 3–30.

Acemoglu, D. and P. Restrepo (2020). Robots and Jobs: Evidence from US Labor Markets. *Journal of Political Economy 128*(6), 2188–2244.

Aghion, P., B. F. Jones, and C. I. Jones (2018). Artificial Intelligence and Economic Growth. In *NBER Chapters*, pp. 237–282.

Agrawal, A., J. S. Gans, and A. Goldfarb (2019). Exploring the Impact of Artificial Intelligence: Prediction Versus Judgment. *Information Economics and Policy 47*, 1–6.

Agrawal, A., J. McHale, and A. Oettl (2024). Artificial Intelligence and Scientific Discovery: A Model of Prioritized Search. *Research Policy 53*(5), 104989.

Agrawal, A. K., J. McHale, and A. Oettl (2026). AI in Science. Technical report, National Bureau of Economic Research.

Anderson, A. C. (2003). The Process of Structure-Based Drug Design. *Chemistry Biology 10*(9), 787–797.

Autor, D. H. (2015). Why Are There Still So Many Jobs? The History and Future of Workplace Automation. *Journal of Economic Perspectives 29*(3), 3–30.

Autor, D. H., L. F. Katz, and A. B. Krueger (1998). Computing Inequality: Have Computers Changed the Labor Market? *The Quarterly Journal of Economics 113*(4), 1169–1213.

Autor, D. H., F. Levy, and R. J. Murnane (2003). The Skill Content of Recent Technological Change: An Empirical Exploration. *The Quarterly Journal of Economics 118*(4), 1279–1333.

Babina, T. et al. (2024). Artificial Intelligence, Firm Growth, and Product Innovation. *Journal of Financial Economics 151*, 103745.

Berman, H. M., S. K. Burley, G. J. Kleywegt, J. L. Markley, H. Nakamura, and S. Velankar (2016). The Archiving and Dissemination of Biological Structure Data. *Current Opinion in Structural Biology 40*, 17–22.

Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne (2000). The Protein Data Bank. *Nucleic Acids Research 28*(1), 235–242.

Bertoline, L. M. F. et al. (2023). Before and After AlphaFold2: An Overview of Protein Structure Prediction. *Frontiers in Bioinformatics 3*, 1120370.

Bessen, J. et al. (2022). The Role of Data for AI Startup Growth. *Research Policy 51*(5), 104513.

Boby, M. L., D. Fearon, M. Ferla, M. Filep, L. Koekemoer, M. C. Robinson, C. M. Consortium‡, J. D. Chodera, A. A. Lee, N. London, et al. (2023). Open science discovery of potent noncovalent SARS-CoV-2 main protease inhibitors. *Science 382*(6671), eabo7201.

Bresnahan, T. F., E. Brynjolfsson, and L. M. Hitt (2002). Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence. *The Quarterly Journal of Economics 117*(1), 339–376.

Brown, E. N. and S. Ramaswamy (2007). Quality of Protein Crystal Structures. *Biological Crystallography 63*(9), 941–950.

Brynjolfsson, E., B. Chandar, and R. Chen (2025). Canaries in the Coal Mine? Six Facts About the Recent Employment Effects of Artificial Intelligence. *Digital Economy*.

Brynjolfsson, E., D. Li, and L. Raymond (2025). Generative AI at Work. *The Quarterly Journal of Economics 140*(2), 889–942.

Brynjolfsson, E., T. Mitchell, and D. Rock (2018). What Can Machines Learn and What Does It Mean for Occupations and the Economy? In *AEA Papers and Proceedings*, Volume 108, pp. 43–47. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Burtch, G., D. Lee, and Z. Chen (2024). The Consequences of Generative AI for Online Knowledge Communities. *Scientific Reports 14*(1), 10413.

Callaway, E. (2020). 'It Will Change Everything': DeepMind's AI Makes Gigantic Leap in Solving Protein Structures. *Nature 588*(7837), 203–205.

Casanal, A., S. Shakeel, and L. A. Passmore (2019). Interpretation of Medium Resolution CryoEM Maps of Multi-Protein Complexes. *Current Opinion in Structural Biology 58*, 166–174.

Cavalli, G. (2024). How Scientific Organizations Adapt to Advances in Artificial Intelligence: The Impact of AlphaFold1. In *Academy of Management Proceedings*, Volume 2024, pp. 21333. Academy of Management Valhalla, NY 10595.

Chen, Q., M. Yang, L. Qin, J. Liu, Z. Yan, J. Guan, D. Peng, Y. Ji, H. Li, M. Hu, et al. (2025). AI 4Research: A Survey of Artificial Intelligence for Scientific Research. *arXiv preprint arXiv:2507.01903*.

Cockburn, I. M., R. Henderson, and S. Stern (2019). The Impact of Artificial Intelligence on Innovation. *The Economics of Artificial Intelligence: An Agenda*, 115.

Crow, J. M. (2021). The Must-Have Multimillion-dollar Microscopy Machine.

Cui, Z. K. et al. (2025). The Effects of Generative AI on High-Skilled Work: Evidence from Three Field Experiments with Software Developers. Technical Report 4945566, SSRN.

Cullen, Z., D. Li, and S. Li (2025). Labor as Capital: AI and the Ownership of Expertise. Technical report, Working Paper.

Dell'Acqua, F. et al. (2023). Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality. *Harvard Business School Technology & Operations Mgt. Unit Working Paper* (24-013).

Doerr, A. (2016). Single-Particle Cryo-Electron Microscopy: A Brief Overview of How to Solve a Macromolecular Structure Using Single-Particle Cryo-Electron Microscopy (Cryo-EM). *Nature Methods 13*(1), 23–24.

Edman, P. and G. Begg (1967). A Protein Sequenator. *European Journal of Biochemistry 1*(1), 80–91.

Edwards, A. M. et al. (2011). Too Many Roads Not Taken. *Nature 470*(7333), 163–165.

Eloundou, T. et al. (2024). GPTs Are GPTs: Labor Market Impact Potential of LLMs. *Science 384*(6702), 1306–1308.

Felten, E. W., M. Raj, and R. Seamans (2018). A Method to Link Advances in Artificial Intelligence to Occupational Abilities. In *AEA Papers and proceedings*, Volume 108, pp. 54–57. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Fernandez-Leiro, R. and S. H. Scheres (2016). Unravelling the Structures of Biological Macromolecules by Cryo-EM. *Nature 537*(7620), 339.

Fleming, J., P. Magana, S. Nair, M. Tsenkov, D. Bertoni, I. Pidruchna, M. Q. L. Afonso, A. Midlik, U. Paramval, A. Žídek, et al. (2025). AlphaFold Protein Structure Database and 3D-Beacons: New Data and Capabilities. *Journal of Molecular Biology*, 168967.

Frey, C. B. and M. A. Osborne (2017). The Future of Employment: How Susceptible Are Jobs to Computerisation? *Technological Forecasting and Social Change 114*, 254–280.

Furman, J. L. and S. Stern (2011). Climbing Atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research. *American Economic Review 101*(5), 1933–1963.

Furman, J. L. and F. Teodoridis (2020). Automation, Research Technology, and Researchers' Trajectories: Evidence from Computer Science and Electrical Engineering. *Organization Science 31*(2), 330–354.

Gao, J. and D. Wang (2023). Quantifying the Benefit of Artificial Intelligence for Scientific Research. *arXiv preprint arXiv:2304.10578*.

Goldfarb, A., B. Taska, and F. Teodoridis (2023). Could Machine Learning Be a General Purpose Technology? A Comparison of Emerging Technologies Using Data from Online Job Postings. *Research Policy 52*(1), 104653.

Goodsell, D. S. (2019). Methods for Determining Atomic Structures. Technical report, Protein Data Bank: PDB-101 2019.

Gordon, D. E., G. M. Jang, M. Bouhaddou, J. Xu, K. Obernier, K. M. White, M. J. O'Meara, V. V. Rezelj, J. Z. Guo, D. L. Swaney, et al. (2020). A SARS-CoV-2 Protein Interaction Map Reveals Targets for Drug Repurposing. *Nature 583*(7816), 459–468.

Gorrec, F. (2021). A Beginner's Guide to Macromolecular Crystallization. *The Biochemist 43*(1), 36–43.

Graetz, G. and G. Michaels (2018). Robots at Work. *Review of Economics and Statistics 100*(5), 753–768.

Grassucci, R. A., D. J. Taylor, and J. Frank (2007). Preparation of Macromolecular Complexes for Cryo-Electron Microscopy. *Nature Protocols 2*(12), 3239–3246.

Helmers, C. and H. G. Overman (2017). My Precious! The Location and Diffusion of Scientific Research: Evidence from the Synchrotron Diamond Light Source. *The Economic Journal 127*(604), 2006–2040.

Hill, R. and C. Stein (2025a). Race to the Bottom: Competition and Quality in Science. *The Quarterly Journal of Economics 140*(2), 1111–1185.

Hill, R. and C. Stein (2025b). Scooped! Estimating Rewards for Priority in Science. *Journal of Political Economy 133*(3), 793–845.

Hill, R. and C. Stein (2026). How Artificial Intelligence Shapes Science: Evidence from AlphaFold. Accessed through GitHub March 21, 2026.

Hoelzemann, J. et al. (2024). The Streetlight Effect in Data-Driven Exploration. Technical report, National Bureau of Economic Research.

Ide, E. and E. Talamàs (2025). Artificial Intelligence in the Knowledge Economy. *Journal of Political Economy 133*(12), 3762–3800.

Jones, B. (2025). Artificial Intelligence in Research and Development. Technical report, National Bureau of Economic Research.

Jumper, J., R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, et al. (2021). Highly Accurate Protein Structure Prediction with AlphaFold. *Nature 596*(7873), 583–589.

Khurana, M. P. (2023). *Streetlight Effects: How Does the Streetlight Influence What We Choose to Study?* Cambridge University Press.

Kim, S. (2025). Navigating the Rugged Data Landscape: The Impact of Data-Extrapolation Technologies on Knowledge Production. *Columbia Business School*.

Kovalevskiy, O. et al. (2024). AlphaFold Two Years on: Validation and Impact. *Proceedings of the National Academy of Sciences 121*(34), e2315002121.

Kryshtafovych, A. et al. (2019). Critical Assessment of Methods of Protein Structure Prediction (CASP)—Round XIII. *Proteins: Structure, Function, and Bioinformatics 87*(12), 1011–1020.

Lou, B. and L. Wu (2021). AI on Drugs: Can Artificial Intelligence Accelerate Drug Development? Evidence from a Large-Scale Examination of Bio-Pharma Firms. *MIS Quarterly 45*(3), 1451–1482.

Lyu, L. et al. (2025). Wikipedia Contributions in the Wake of ChatGPT. In *Companion Proceedings of the ACM on Web Conference 2025*, pp. 1176–1179.

Mullainathan, S. and A. Rambachan (2025). Science in the Age of Algorithms. *The Economics of Transformative AI*.

Murray, F. et al. (2016). Of Mice and Academics: Examining the Effect of Openness on Innovation. *American Economic Journal: Economic Policy 8*(1), 212–252.

Noy, S. and W. Zhang (2023). Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence. *Science 381*(6654), 187–192.

PDB-101 (2024). Structural Biology and Nobel Prizes. https://pdb101.rcsb.org/learn/other-resources/structural-biology-and-nobel-prizes.

Sanger, F. and E. O. P. Thompson (1953). The Amino-Acid Sequence in the Glycyl Chain of Insulin. 1. The Identification of Lower Peptides from Partial Hydrolysates. *Biochemical Journal 53*(3), 353.

Schumpeter, J. A. (1942). *Capitalism, Socialism and Democracy*. Routledge.

Shan, G. and L. Qiu (2025). Examining the Impact of Generative AI on Users' Voluntary Knowledge Contribution: Evidence from a Natural Experiment on Stack Overflow. *Information Systems Research*.

Shoemaker, S. C. and N. Ando (2018). X-Rays in the Cryo-Electron Microscopy Era: Structural Biology's Dynamic Future. *Biochemistry 57*(3), 277–285.

Slabinski, L. et al. (2007). XtalPred: A Web Server for Prediction of Protein Crystallizability. *Bioinformatics 23*(24), 3403–3405.

Spence, J. C. H. (2017). XFELs for Structure and Dynamics in Biology. *IUCrJ 4*(4), 322–339.

Strasser, B. J. (2019). *Collecting Experiments: Making Big Data Biology*. University of Chicago Press.

Teodoridis, F. (2018). Understanding Team Knowledge Production: The Interrelated Roles of Technology and Expertise. *Management Science 64*(8), 3625–3648.

Thompson, R. F. et al. (2019). Collection, Pre-Processing and On-the-Fly Analysis of Data for High-Resolution, Single-Particle Cryo-Electron Microscopy. *Nature Protocols 14*(1), 100–118.

Trammell, P. and A. Korinek (2023). Economic Growth Under Transformative AI. Technical report, National Bureau of Economic Research.

Tunyasuvunakool, K. et al. (2021). Highly Accurate Protein Structure Prediction for the Human Proteome. *Nature 596*(7873), 590–596.

Varadi, M., S. Anyango, M. Deshpande, S. Nair, C. Natassia, G. Yordanova, D. Yuan, O. Stroe, G. Wood, A. Laydon, et al. (2022). AlphaFold Protein Structure Database: Massively Expanding the Structural Coverage of Protein-Sequence Space with High-Accuracy Models. *Nucleic Acids Research 50*(D1), D439–D444.

Varadi, M., D. Bertoni, P. Magana, U. Paramval, I. Pidruchna, M. Radhakrishnan, M. Tsenkov, S. Nair, M. Mirdita, J. Yeo, et al. (2024). AlphaFold Protein Structure Database in 2024: Providing Structure Coverage for Over 214 Million Protein Sequences. *Nucleic Acids Research 52*(D1), D368–D375.

Varadi, M. and S. Velankar (2023). The Impact of AlphaFold Protein Structure Database on the Fields of Life Sciences. *Proteomics 23*(17), 2200128.

Wang, H.-W. and J.-W. Wang (2017). How Cryo-Electron Microscopy and X-Ray Crystallography Complement Each Other. *Protein Science 26*(1), 32–39.

Webb, M. (2019). The Impact of Artificial Intelligence on the Labor Market. Technical Report 3482150, SSRN.

Williams, H. L. (2013). Intellectual Property Rights and Innovation: Evidence from the Human Genome. *Journal of Political Economy 121*(1), 1–27.

Wrapp, D., N. Wang, K. S. Corbett, J. A. Goldsmith, C.-L. Hsieh, O. Abiona, B. S. Graham, and J. S. McLellan (2020). Cryo-EM Structure of the 2019-nCoV Spike in the Prefusion Conformation. *Science 367*(6483), 1260–1263.

Wu, X. and T. A. Rapoport (2021). Cryo-EM Structure Determination of Small Proteins by Nanobody-Binding Scaffolds (Legobodies). *Proceedings of the National Academy of Sciences 118*(41), e2115001118.

Yu, Z. (2026). The Impacts of AI at Scale: Evidence from Research Scientists. Technical report, CESifo.

Zhuo, R. (2025). Navigating the Exploitation-Exploration Tradeoff: An Empirical Study of Resource Allocation in Research Labs. *Available at SSRN 5428035*.

Zolas, N. et al. (2021). Advanced Technologies Adoption and Use by US Firms: Evidence from the Annual Business Survey. Technical report, National Bureau of Economic Research.

Tables

Table I: Baseline Summary Statistics

Panel A. Quantity

	Treatment Group VH		Control Group Non VH		Two-sided T-Test
	Mean	Std. Dev.	Mean	Std. Dev.	$Pr(\|T\| > t)$
N of Deposit per Year-Quarter	13.18	37.95	10.76	24.20	0.537
N of Papers Published per Year-Quarter	10.53	29.33	9.26	20.82	0.690
(%) First-Attempts Fraction	22.43	22.53	27.25	21.64	0.099
(%) Human Structures Fraction	21.95	27.25	30.64	29.56	0.024*
N	85		170		255

Panel B. Quality

	Treatment Group VH		Control Group Non VH		Two-sided T-Test
	Mean	Std. Dev.	Mean	Std. Dev.	$Pr(\|T\| > t)$
Method: X-ray Crystallography					
Resolution (Lower is Better)	2.00	0.53	2.15	0.65	0.000***
(%) PubMed Listed	0.02	0.93	0.02	0.41	0.941
Primary Citation's 2024 JIF Score	6.75	8.21	8.16	9.47	0.000***
(%) DrugBank Listed	0.00	0.01	0.02	0.40	0.000***
N	14911		19325		34236
Method: Cryo-EM					
Resolution (Lower is Better)	3.61	1.46	3.95	2.16	0.000***
(%) PubMed Listed	0.04	0.57	0.01	0.00	0.000***
Primary Citation's 2024 JIF Score	13.65	13.97	19.32	16.40	0.000***
(%) DrugBank Listed	0.00	0.00	0.00	0.00	0.000***
N	774		6278		7052

The unit of observation is at (prediction region $\times$ structure class) bin level in Panel A and structure level in Panel B. Structures deposited between 2018 Q1 and 2021 Q2 are included. Statistical significance is indicated as: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table II: Impact of AlphaFold2 on the Total Amount of Deposit

	PPML (Count)	OLS (Count)	OLS (Asinh)	
	(1)	(2)	(3)	(4)
POST × VH	-0.427***	-5.439***	-0.242***	-0.334***
	(0.095)	(1.676)	(0.048)	(0.073)
% Δ (VH vs Non VH)	-34.8	-42.8	-24.3	-33.5
Control Mean	12.72	12.72	12.72	12.72
Deposit Quarter FE	Yes	Yes	Yes	Yes
Bin FE	Yes	Yes	Yes	Yes
Weighting	None	None	None	PreAF2 Size
R-squared	-	0.805	0.811	0.895
N of Bins	255	255	255	255
N of Observations	6630	6630	6630	6630

The unit of observation is bin-deposit-quarter. In Columns (1)-(2), the outcome variable is the count of structures deposited in each bin-deposit-quarter, estimated using Poisson pseudo-maximum likelihood and OLS, respectively. In Column (3)-(4), OLS is estimated with the outcome count reported in inverse hyperbolic sine transformation. %Δ is computed as $100 * (e^{\hat{\beta}} - 1)$ in column (1), $100 * (\frac{\hat{\beta}}{\bar{y}})$ in column (2) and $100 * (\hat{\beta} * \sqrt{1 + \frac{1}{\bar{y}^2}})$ in column (3), (4) where $\bar{y}$ is the control mean. Column (4) is weighted by the total amount of structures deposited under the same class in the pretreatment period of our sample. Clustered standard errors at the bin level are reported in parentheses. Statistical significance is indicated as: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table III: Impact of AlphaFold2 on the Deposit Composition

	Total Deposits	Attempt Status		Origin Status		Stuck Status	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Outcome: Count	-	First	Non First	Human	Non Human	Stuck	Non Stuck
POST × VH	-0.427***	-0.508***	-0.412***	-0.239	-0.509***	-0.349	-0.151
	(0.095)	(0.086)	(0.111)	(0.156)	(0.098)	(0.329)	(0.177)
% Δ (VH vs Non VH)	-34.8	-39.8	-33.7	-21.2	-39.9	-29.4	-14.0
Deposit Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bin FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N of Bins	255	224	250	198	249	125	175
N of Observations	6630	5824	6500	5148	6474	3250	4550

The unit of observation is bin-deposit-quarter. Column (1) reports the total number of structures deposited. Columns (2)–(3) report counts of "first-attempt" and "non–first-attempt" structures, respectively. Columns (4)–(5) report counts of structures from human and non-human proteins, respectively. Columns (6)–(7) restrict to human proteins and report counts for structures from "stuck" protein classes versus those that are not. Bins with no deposits of certain type of structures are dropped in each regression. The specification follows Poisson pseudo-maximum likelihood and $\%\Delta$ is computed as $100 * (e^{\hat{\beta}} - 1)$. Clustered standard errors at the bin level are reported in parentheses. Statistical significance is indicated as: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table IV: Impact of AlphaFold2 on the Size of Experiment Team

	Total Deposits	Attempt Status		Origin Status	
	(1)	(2)	(3)	(4)	(5)
Outcome: Size of Team	-	First	Non First	Human	Non Human
POST × VH	0.018	-0.019	-0.005	0.044	0.016
	(0.075)	(0.056)	(0.077)	(0.088)	(0.080)
% Δ (VH vs Non VH)	1.8	-1.9	-0.5	4.5	1.6
Deposit Quarter FE	Yes	Yes	Yes	Yes	Yes
Bin FE	Yes	Yes	Yes	Yes	Yes
N of Bins	71	28	57	31	57
N of Observations	1846	728	1482	806	1482

The unit of observation is bin-deposit-quarter. Column (1) reports the total number of structures deposited. Columns (2)–(3) report counts of "first-attempt" and "non–first-attempt" structures, respectively. Columns (4)–(5) report counts of structures from human and non-human proteins, respectively. $\%\Delta$ is computed as $100 * (e^{\hat{\beta}} - 1)$. Each column includes only bins with at least one deposit of the corresponding type of structure each deposit quarter over the sample period. Clustered standard errors at the bin level are reported in parentheses. Statistical significance is indicated as: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table V: Heterogeneous Impacts by Experimental Methods

	Total Deposits	Attempt Status		Origin Status		Stuck Status	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Outcome: Count	-	First	Non First	Human	Non Human	Stuck	Non Stuck
POST × VH	-0.196**	-0.242***	-0.186	0.002	-0.312**	-0.361	0.013
	(0.095)	(0.083)	(0.113)	(0.146)	(0.123)	(0.287)	(0.151)
POST × VH × Cryo-EM	0.198	-0.001	0.223	0.079	0.314	0.452	0.049
	(0.168)	(0.161)	(0.194)	(0.226)	(0.200)	(0.337)	(0.380)
% Δ*Xray* (VH vs Non VH)	-17.8	-21.5	-17.0	0.2	-26.8	-30.3	1.3
% Δ*CryoEM* (VH vs Non VH)	0.2	-21.6	3.7	8.4	0.2	9.6	6.4
Method × Deposit Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Method × Bin FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N of Method-Bins	456	377	437	345	433	133	212
N of Observations	11856	9802	11362	8970	11258	3458	5512

The unit of observation is method-bin-deposit-quarter. Column (1) reports the total number of structures deposited. Columns (2)–(3) report counts of "first-attempt" and "non–first-attempt" structures, respectively. Columns (4)–(5) report counts of structures from human and non-human proteins, respectively. Columns (6)–(7) restrict to human proteins and report counts for structures from "stuck" protein classes versus those that are not. Bins with no deposits of certain type of structures are dropped in each regression. Specifications of both panels follow the Poisson pseudo- maximum likelihood. %ΔXray is computed as $100 * (e^{\hat{\beta}_1} - 1)$, %ΔCryoEM is computed as $100 * (e^{\hat{\beta}_1 + \hat{\beta}_2} - 1)$. Clustered standard errors at the method-bin level are reported in parentheses. Statistical significance is indicated as: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table VI: Impact of AlphaFold2 on the Resolution of Structures

Panel A. Structures by X-ray Crystallography

	Total Deposits	Attempt Status		Origin Status		Stuck Status	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Outcome: Resolution	-	First	Non First	Human	Non Human	Stuck	Non Stuck
POST × VH	0.020	0.010	0.018	-0.098**	0.070**	-0.056	-0.100**
	(0.030)	(0.021)	(0.034)	(0.046)	(0.034)	(0.140)	(0.047)
% Δ (VH vs Non VH)	-0.9	-0.4	-0.9	4.6	-3.3	2.2	4.7
Control Mean	2.13	2.24	2.10	2.12	2.14	2.59	2.11
Deposit Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bin FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-Squared	0.1155	0.0927	0.1341	0.1814	0.1251	0.2440	0.1658
N of Bins	236	189	227	168	229	21	147
N of Observations	61276	12084	49167	20154	41107	574	19580

Panel B. Structures by Cyro-EM

	Total Deposits	Attempt Status		Origin Status		Stuck Status	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Outcome: Resolution	-	First	Non First	Human	Non Human	Stuck	Non Stuck
POST × VH	0.080	-0.002	0.122	0.180	0.008	0.155	0.359
	(0.124)	(0.123)	(0.153)	(0.183)	(0.143)	(0.334)	(0.227)
% Δ (VH vs Non VH)	-2.2	0.1	-3.4	-5.0	-0.2	-4.1	-10.9
Control Mean	3.62	3.59	3.62	3.57	3.64	3.82	3.31
Deposit Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bin FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-Squared	0.0954	0.1291	0.1129	0.1395	0.1101	0.1397	0.1348
N of Bins	192	144	182	137	176	93	44
N of Observations	22847	5466	17362	6748	16074	3584	3164

The unit of observation is each structure deposited. Column (1) includes all structures. Columns (2)–(3) report results for "first-attempt" and "non–first-attempt" structures, respectively. Columns (4)–(5) focus on human and non-human proteins, respectively. Columns (6)–(7) restrict the sample to human proteins, distinguishing between stuck and non-stuck proteins, respectively. Specifications of both panels follow the OLS regression. As less the resolution is, higher the quality of execution, %Δ is computed as $-100 * (\frac{\hat{\beta}}{\bar{y}})$ where $\bar{y}$ is the control mean. Clustered standard errors at the bin level are reported in parentheses. Statistical significance is indicated as: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table VII: Impact of AlphaFold2 on the Innovation Outcomes of Structures

Panel A. All Structures

	PubMed		Top JIF Decile		CNS		DrugBank	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Outcome: [Event] (0/1)								
POST × VH	-0.012	-0.011	-0.049**	-0.046**	-0.017	-0.014	0.018	0.019
	(0.021)	(0.021)	(0.022)	(0.022)	(0.012)	(0.012)	(0.018)	(0.018)
Control Mean	0.83	0.83	0.50	0.50	0.11	0.11	0.23	0.23
Deposit Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bin FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes	No	Yes
R-squared	0.065	0.067	0.092	0.099	0.094	0.114	0.152	0.153
N of Bins	249	249	227	227	227	227	249	249
N of Observations	84148	84148	55224	55224	55224	55224	84148	84148

Panel B. First-Attempts

	PubMed		Top JIF Decile		CNS		DrugBank	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Outcome: [Event] (0/1)								
POST × VH	-0.038*	-0.038*	-0.112***	-0.110***	-0.027	-0.025	-0.014	-0.014
	(0.020)	(0.020)	(0.032)	(0.033)	(0.018)	(0.018)	(0.015)	(0.015)
Control Mean	0.87	0.87	0.51	0.51	0.13	0.13	0.14	0.14
Deposit Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bin FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes	No	Yes
R-squared	0.082	0.084	0.129	0.136	0.106	0.134	0.180	0.185
N of Bins	205	205	187	187	187	187	205	205
N of Observations	17577	17577	11664	11664	11664	11664	17577	17577

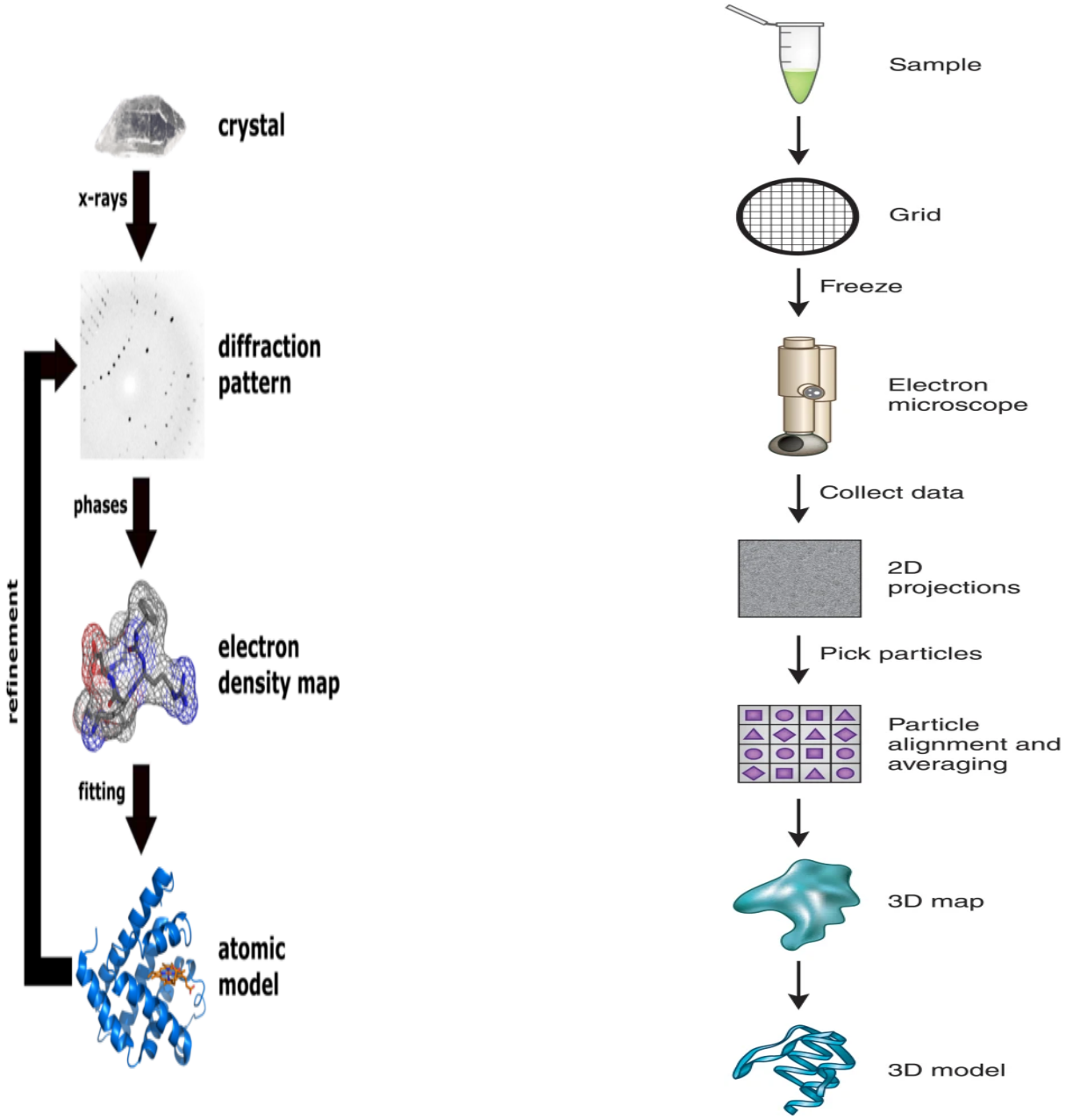
Panel C. Non-First-Attempts

	PubMed		Top JIF Decile		CNS		DrugBank	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Outcome: [Event] (0/1)								
POST × VH	-0.006	-0.004	-0.029	-0.025	-0.013	-0.009	0.029	0.030
	(0.024)	(0.024)	(0.025)	(0.025)	(0.012)	(0.012)	(0.020)	(0.020)
Control Mean	0.82	0.82	0.50	0.50	0.10	0.10	0.25	0.25
Deposit Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bin FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes	No	Yes
R-squared	0.073	0.075	0.097	0.104	0.112	0.128	0.174	0.174
N of Bins	239	239	221	221	221	221	239	239
N of Observations	66547	66547	43543	43543	43543	43543	66547	66547

The unit of observation is each structure deposited. The specification follows a linear probability model. Outcomes in cols. (1)–(2) are primary citations from PubMed. Cols. (3)–(4) indicate publication in top JIF-decile journals. Cols. (5)–(6) indicate publication in Cell, Nature, or Science. Cols. (7)–(8) indicate structures listed in DrugBank. Controls included are biological features measuring the complexity of structures including the number of entities, molecular weight, total number of polymer residues and number of non-hydrogen atoms. Clustered standard errors at the bin level are reported in parentheses. Statistical significance is indicated as: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

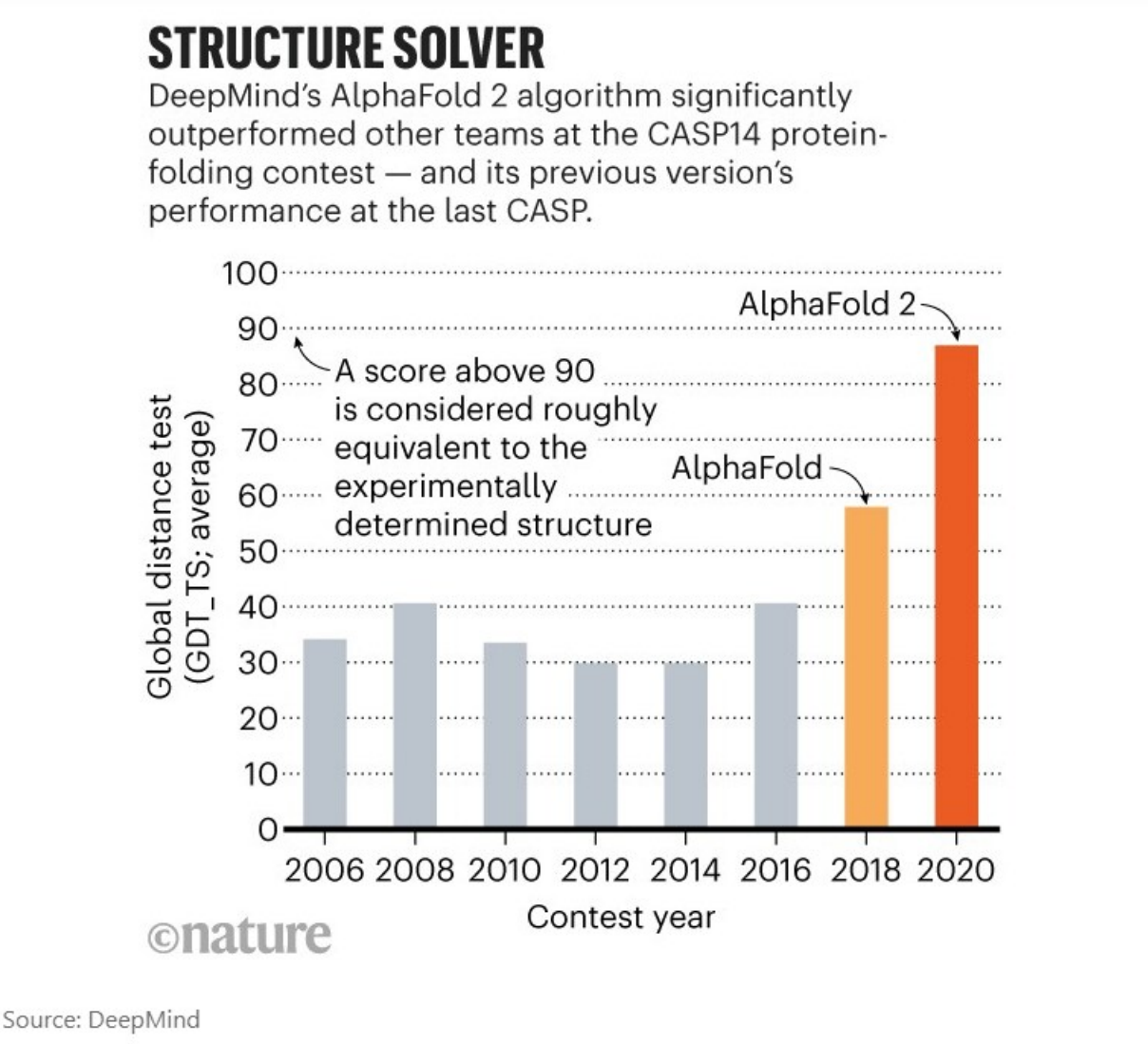
Figures

Figure I: Experimental Methods: X-ray Crystallography vs. Cryo Electron Microscopy



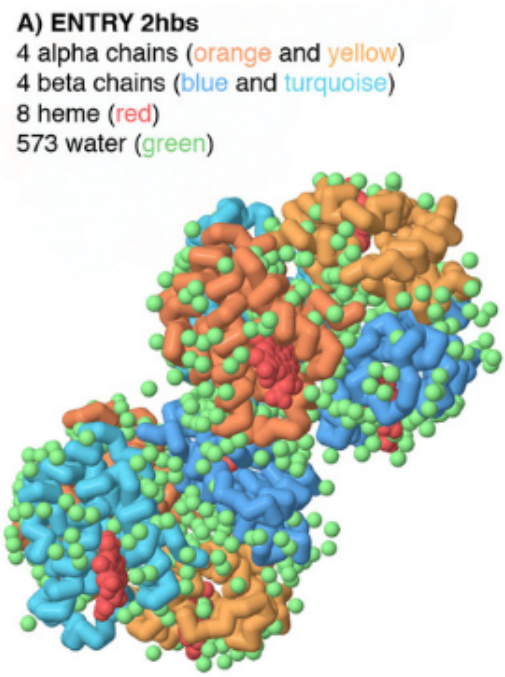
Source: Left: Wikipedia: X-ray crystallography; Right: (Doerr, 2016)

Figure II: Average Global Test Score Across Target Proteins in CASP Contests



Source: (Callaway, 2020)

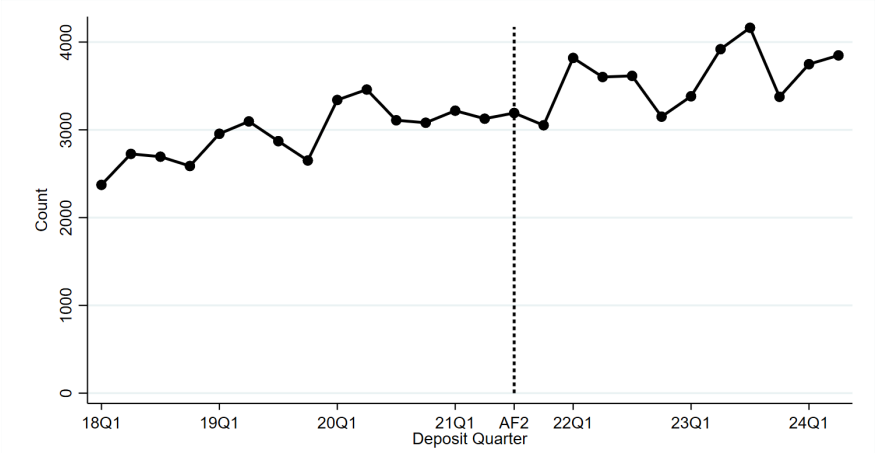
Figure III: An Example of Protein Structure with Multiple Entities



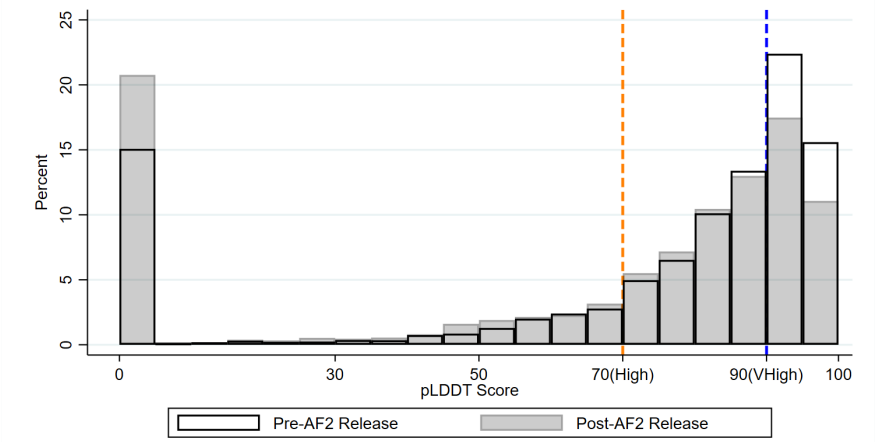
Source: Organization of 3D Structures in the Protein Data Bank

Figure IV: The Evolution of Aggregate Structure Deposit by Prediction Confidence

(a) Quarterly Count of Structure Deposits in the PDB



(b) Histogram of Structures' Average pLDDT Score



(c) Histogram of Proteins' Average pLDDT Score Covered by AF2

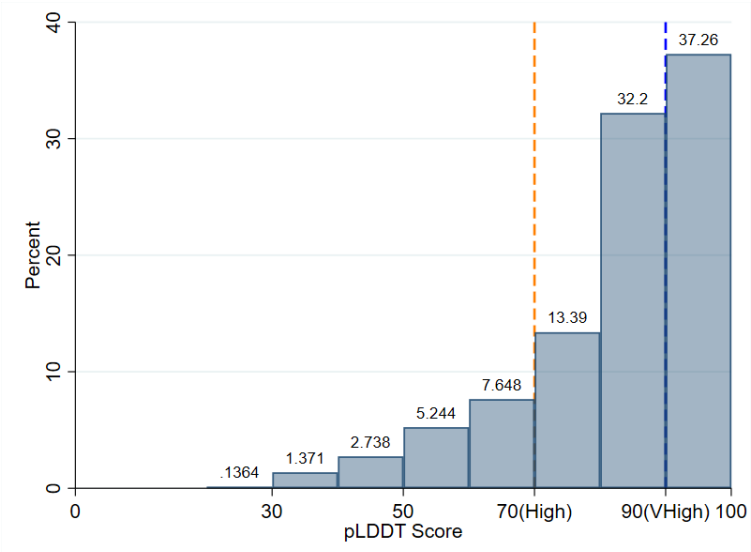
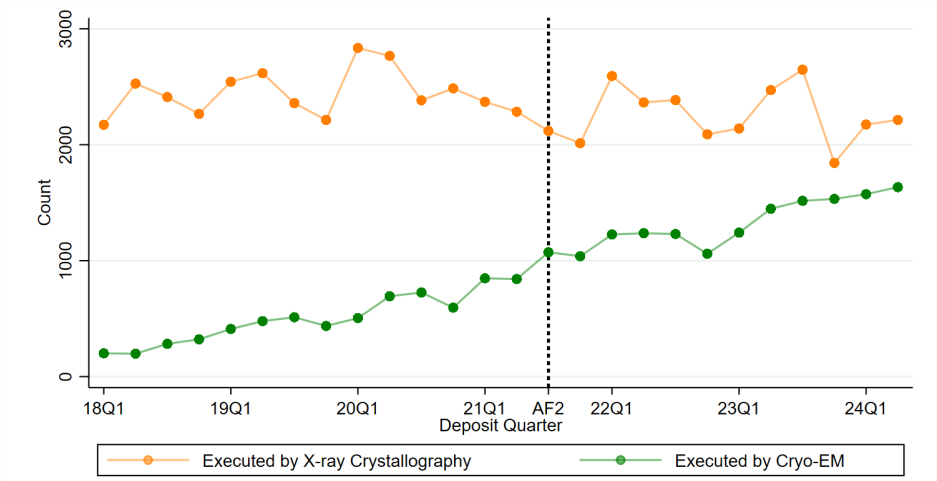
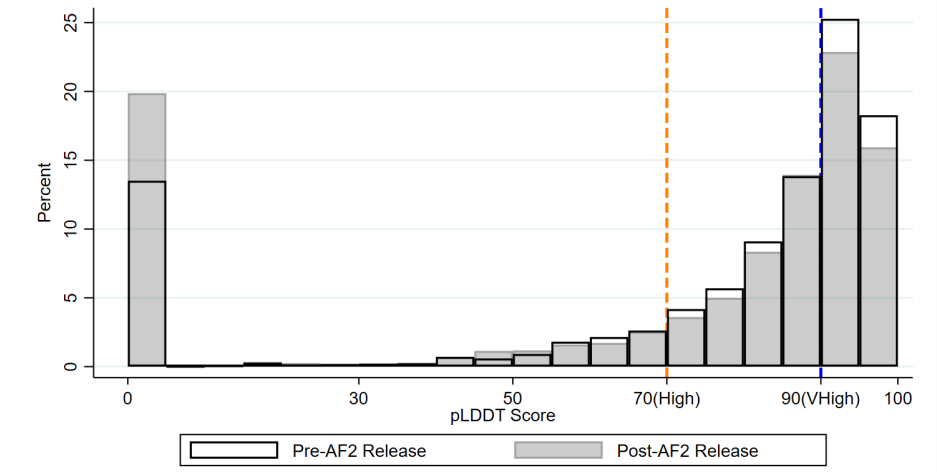


Figure V: The Evolution of Aggregate Structure Deposit by Experimental Methods

(a) Quarterly Count of Structure Deposits in the PDB



(b) Histogram of Structures' Average pLDDT Score (By X-ray Crystallography)



(c) Histogram of Structures' Average pLDDT Score (By Cryo-EM)

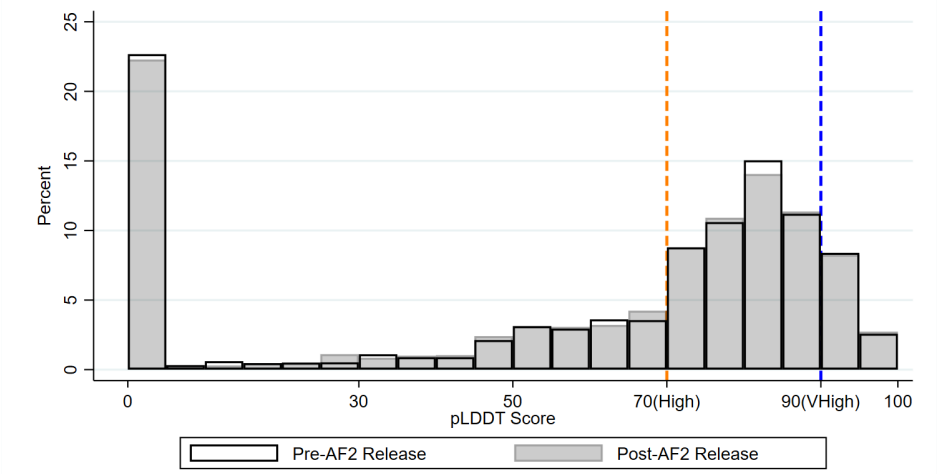
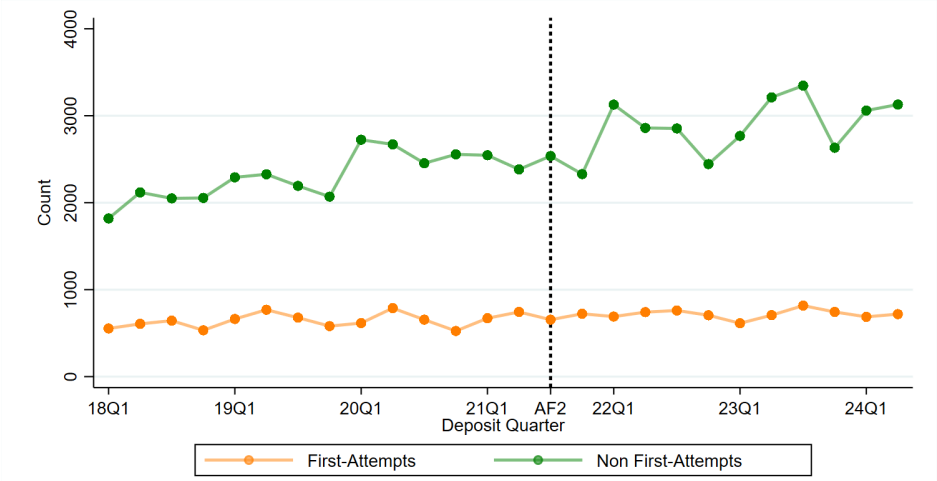
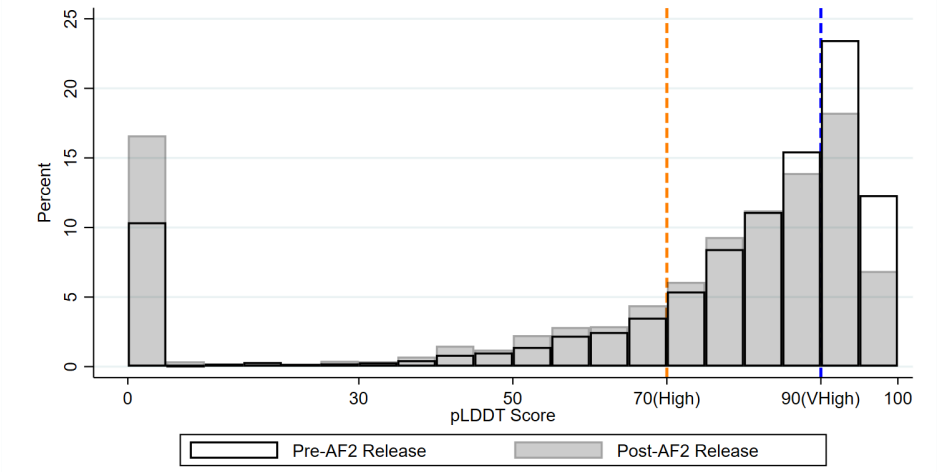


Figure VI: The Evolution of Exploratory Versus Exploitative Structure Deposit

(a) Deposit Count Trend: First-Attempts vs. Non-First-Attempts



(b) Histogram of Structures' Average pLDDT Score (First-Attempts)



(c) Histogram of Structures' Average pLDDT Score (Non-First-Attempts)

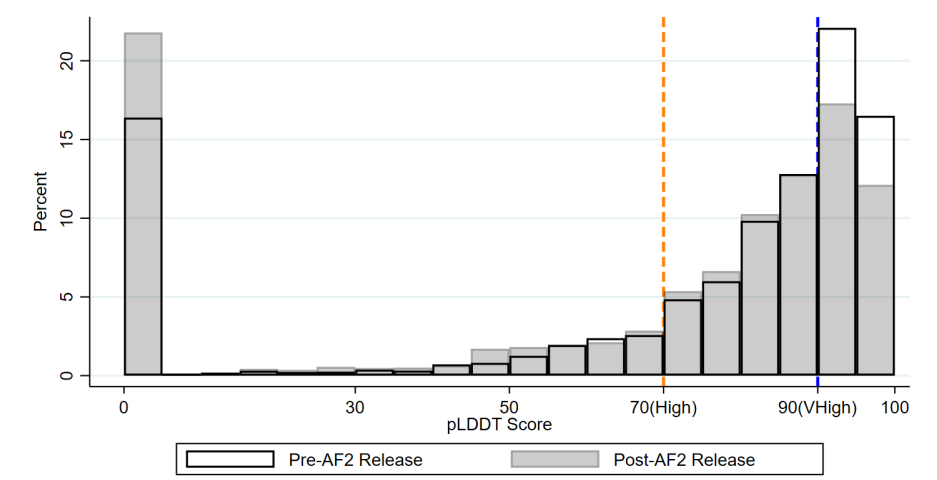
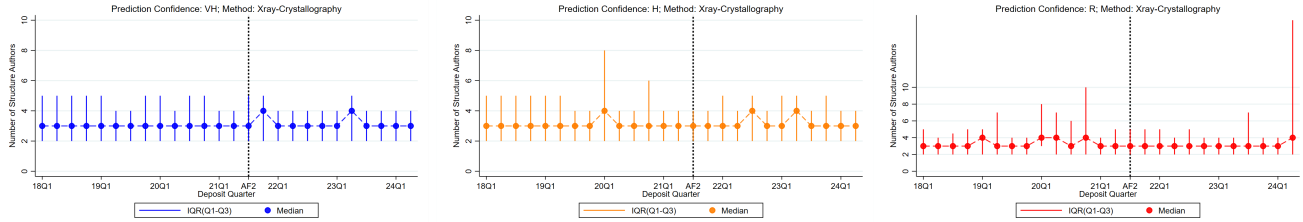
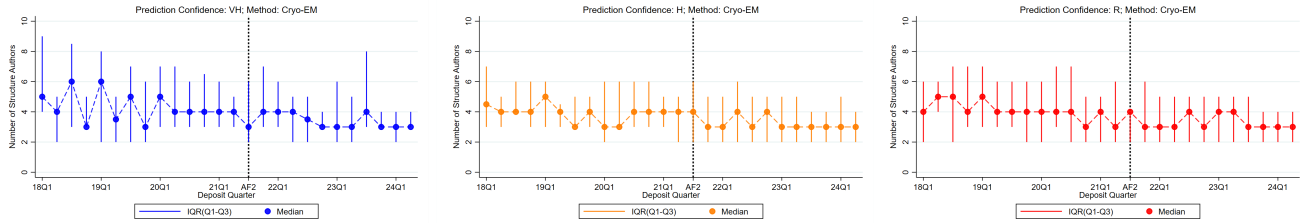


Figure VII: Experimental Team Size Across Prediction Confidence Regions

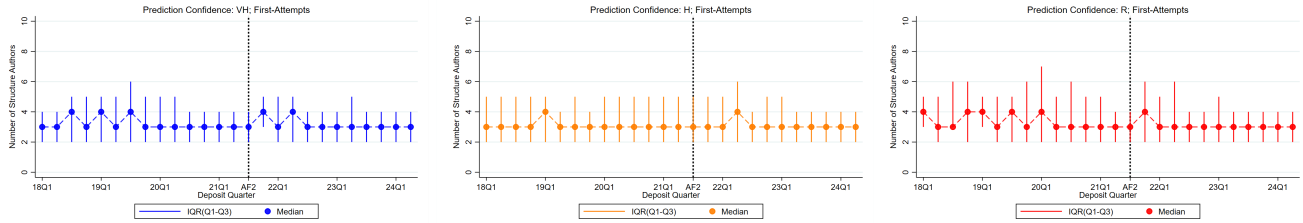
(a) Structures executed by X-ray crystallography



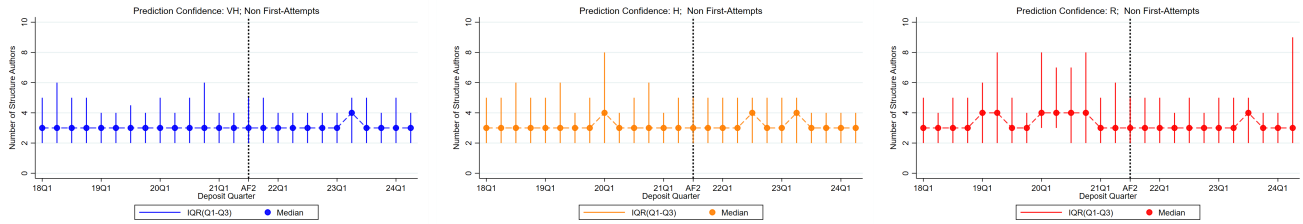
(b) Structures executed by Cyro-EM



(c) First-Attempts



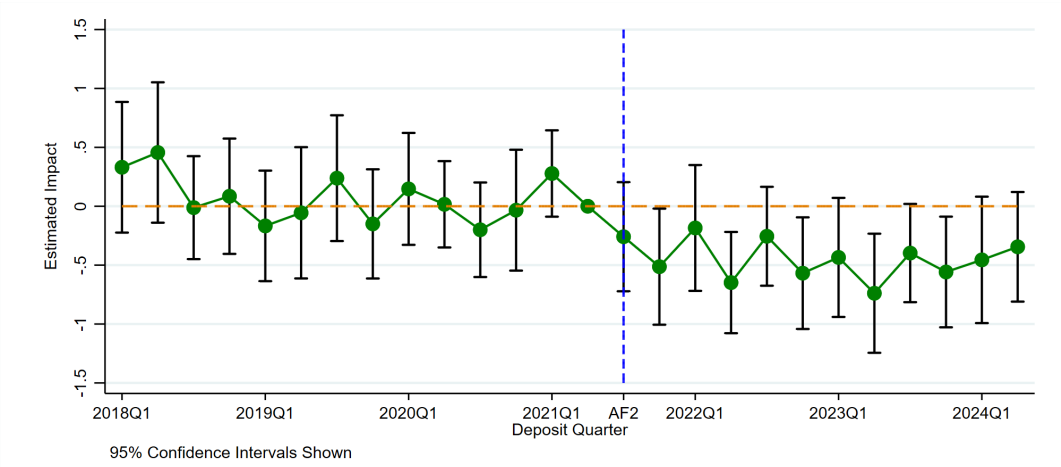
(d) Non-First-Attempts



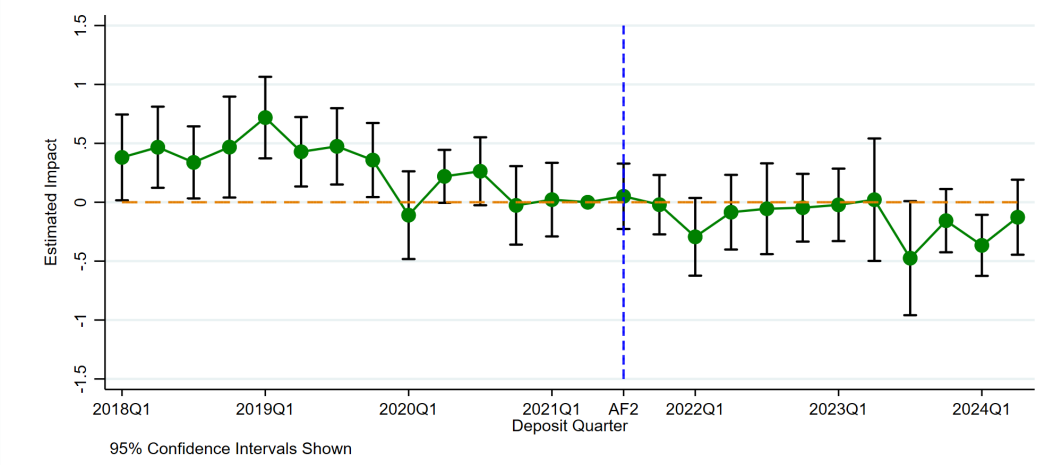
Note: This figure does not include Boby et al. (2023), an outlier in team size, which reports the COVID Moonshot project, a large open-science consortium that combined many labs and contributors across different fields.

Figure VIII: Event Study: Impact of AF2 on Structure Deposit Count (PPML)

(a) First-Attempts



(b) Non-First-Attempts



(c) Total Deposit

