The Rise of Large Language Models and the Direction and Impact of US Federal Research Funding

Yifan Qian[1,2,3,4], Zhe Wen[1,2,3,4], Alexander C. Furnas[1,2,3,4], Yue Bai[1,2,3,4], Erzhuo Shao[1,2,3,5], and Dashun Wang[1,2,3,4,5,*]

[1]Center for Science of Science and Innovation, Northwestern University, Evanston, IL, USA
[2]Ryan Institute on Complexity, Northwestern University, Evanston, IL, USA
[3]Northwestern Innovation Institute, Northwestern University, Evanston, IL, USA
[4]Kellogg School of Management, Northwestern University, Evanston, IL, USA
[5]McCormick School of Engineering, Northwestern University, Evanston, IL, USA
[*]Correspondence to: dashun.wang@kellogg.northwestern.edu

Abstract

Federal research funding shapes the direction, diversity, and impact of the US scientific enterprise. Large language models (LLMs) are rapidly diffusing into scientific practice, holding substantial promise while raising widespread concerns. Despite growing attention to AI use in scientific writing and evaluation, little is known about how the rise of LLMs is reshaping the public funding landscape. Here, we examine LLM involvement at key stages of the federal funding pipeline by combining two complementary data sources: confidential National Science Foundation (NSF) and National Institutes of Health (NIH) proposal submissions from two large US R1 universities, including funded, unfunded, and pending proposals, and the full population of publicly released NSF and NIH awards. We find that LLM use rises sharply beginning in 2023 and exhibits a bimodal distribution, indicating a clear split between minimal and substantive use. Across both private submissions and public awards, higher LLM involvement is consistently associated with lower semantic distinctiveness, positioning projects closer to recently funded work within the same agency. The consequences of this shift are agency-dependent. LLM use is positively associated with proposal success and higher subsequent publication output at NIH, whereas no comparable associations are observed at NSF. Notably, the productivity gains at NIH are concentrated in non-hit papers rather than the most highly cited work. Together, these findings provide large-scale evidence that the rise of LLMs is reshaping how scientific ideas are positioned, selected, and translated into publicly funded research, with implications for portfolio governance, research diversity, and the long-run impact of science.

Introduction

Federal research funding is the primary mechanism through which the United States converts public resources into scientific knowledge[1,2]. Through competitive grant programs at agencies such as the National Science Foundation (NSF) and the National Institutes of Health (NIH), decisions about the allocation of taxpayer dollars influence which scientific ideas receive support, which fields advance, and which investigators are able to initiate and sustain research programs[3–10]. Thus, the federal funding process plays a central role in shaping the direction, diversity, and long-run impact of the US scientific enterprise and—given America's scientific leadership—the global research landscape as well[3–10]. Understanding forces that influence this process is therefore essential not only for science policy and the rate and direction of scientific progress, but also for the stewardship and accountability of public investment in research[11–14].

Despite growing attention to the use of AI in science[15–26], little is known about how the rise of large language models (LLMs) is reshaping the public funding landscape. This gap is consequential, as funding decisions operate upstream of publication, hiring, and recognition, shaping which scientific ideas are pursued long before other forms of evaluation occur. Yet proposal texts, especially unfunded submissions, are typically confidential, making this stage of the funding pipeline difficult to study systematically at scale. As a result, we lack systematic evidence on how LLM adoption enters the funding pipeline, how it relates to proposal evaluation, and how it shapes the research that ultimately receives public support.

At the same time, LLMs represent a potentially transformative development for scientific work[15–21]. Recent large-scale evidence shows that the diffusion of LLMs has already been accompanied by sharp increases in scientific production across major preprint platforms, underscoring their potential to substantially raise output at scale[16]. By dramatically reducing the cost of drafting, synthesizing, and revising complex text, LLMs may help scientists articulate ideas more clearly, navigate the growing burden of knowledge, and explore unfamiliar domains more efficiently[27,28]. In principle, such tools could accelerate the pace of discovery, lower barriers to entry, facilitate interdisciplinary recombination, and support more effective pivoting across research directions[22,28–31]. From this perspective, LLMs hold the promise of expanding scientific exploration and improving the efficiency with which ideas are generated, refined, and communicated[32,33].

Yet these same features also raise critical concerns[32]. Because LLMs are trained on large corpora of existing scientific text, including published papers and public records of funded projects, they may tend to produce high-probability, discipline-typical language. When used in proposal preparation, such tools may enhance fluency and alignment with prevailing norms while reducing rhetorical and topical variance, making proposed projects more closely resemble those that have recently succeeded[23,26,34]. In this way, LLM use could shift the balance between exploration and exploitation in scientific funding, pulling proposals toward the center of existing funding patterns and potentially narrowing the diversity of ideas that are submitted for or receive public support, even as it may lower the barriers to producing competitive applications[29,35]. Recent work highlights this tension more broadly, showing that AI can expand individual scientists' productivity and impact while simultaneously contracting the collective focus of science, underscoring the possibility that gains at the individual level may coexist with reduced diversity at the system level[26].

Beyond these concerns, a bottlenecks perspective further suggests caution in extrapolating writing-related productivity gains to downstream scientific outputs[36]. Discovery is constrained by multiple frictions, from data collection and experimental throughput to coordination, implementation, and more, many of which LLM involve-

ment does not directly resolve. Because funding outputs depend on both communication and execution, there are reasons to believe that productivity gains observed in manuscript writing may not translate to the impact of funded projects. This distinction makes federal research funding a particularly important context in which to evaluate how LLM adoption reshapes scientific activity.

Reflecting the policy relevance of this uncertainty, federal agencies have begun to issue guidance on generative AI in proposal preparation. In July 2025, NIH clarified that applications "substantially developed by AI" will not be considered applicants' original work (see NOT-OD-25-132), while NSF has emphasized investigator responsibility for accuracy and originality and encouraged disclosure of generative AI use (see Notice to research community). These actions underscore both the growing salience of LLMs in public funding and the need for empirical evidence on how they are being used and with what consequences.

Here, we examine LLM use at key stages in the federal funding pipeline by employing two complementary levels of analysis. First, we collect novel datasets containing the complete set of NSF (D1) and NIH (D2) proposal submissions (hereafter private NSF and NIH data) from two large US R1 universities with request start dates from 2021 to 2025, including funded, unfunded, and pending proposals. These data offer us a unique opportunity to observe proposal language at the point of submission. Second, we analyze the full population of publicly released NSF (D3) and NIH (D4) award abstracts (hereafter public NSF and NIH data) over the same period, together with publications supported by these awards. These data allow us to systematically analyze the funded awards (see Supplementary Note S1 for details on datasets). Overall, these data enable us to characterize the evolution of LLM use in federal funding and to examine how it relates to the positioning of scientific ideas, their selection for funding, and their translation into publicly supported research output.

Rapid rise and bimodal distribution of LLM use in US federal research funding

To estimate LLM use in federal grant proposals and awards, we analyze textual traces in their abstracts. Specifically, we leverage an established detection method proposed by Liang *et al.*[15]. Using public grant abstracts with start dates in 2021, which preceded the widespread use of LLMs, we estimate the word distribution of human-written text. We also estimate a corresponding word distribution of LLM-generated text by prompting OpenAI's GPT-3.5-turbo-0125 model to rewrite the same abstracts, simulating LLM involvement in proposal language. Comparing these two distributions allows us to identify systematic differences between human and LLM-modified writing (see Supplementary Note S2.1 for details on model training).

We then apply the detection method at both the corpus and individual grant levels. At the corpus level, we pool grant abstracts for each focal month and its two adjacent months, and we estimate the fraction of LLM-modified sentences (α) within this combined set in each dataset. At the individual grant level, we estimate α separately for each grant abstract in each dataset. These two levels represent complementary estimates, as the corpus-level estimates leverage large volumes of text to capture aggregate LLM use over time, whereas grant-level estimates enable grant-level analyses that capture heterogeneity across grants, allowing us to examine how LLM use relates to semantic distinctiveness and other grant outputs.

Figures 1a-d trace the evolution of LLM involvement (α) in NSF and NIH grants across two stages of the funding pipeline, proposal submission and award funding, drawing on confidential proposal submission abstracts (private data) and publicly released award abstracts (public data), respectively. Across all four datasets, α is relatively stable prior to late 2022, and then rises sharply beginning in 2023, coinciding with the widespread availability of ChatGPT.

This increase is visible both at the point of submission, when ideas first enter the system, and among funded awards that have passed peer review, indicating that LLM use is not confined to a single stage of the pipeline. Over this period, NIH exhibits the same temporal pattern as NSF but with overall lower average levels of LLM involvement.

At the level of individual grants, Figs. 1e-h show that α is not smoothly distributed. Among grants starting in 2023-2025, it features a bimodal distribution, with one mode near zero and a second centered around roughly 10-15%. This pattern reveals a clear split: one group of grant abstracts that largely avoids LLM use and another that incorporates LLMs more substantively into proposal preparation. These results indicate that LLM adoption has rapidly become a detectable feature of federal research proposals and awards, but it remains highly uneven across them. This raises the question of whether the heterogeneity is systematically related to the kinds of ideas proposed and funded, or to downstream scientific outputs.

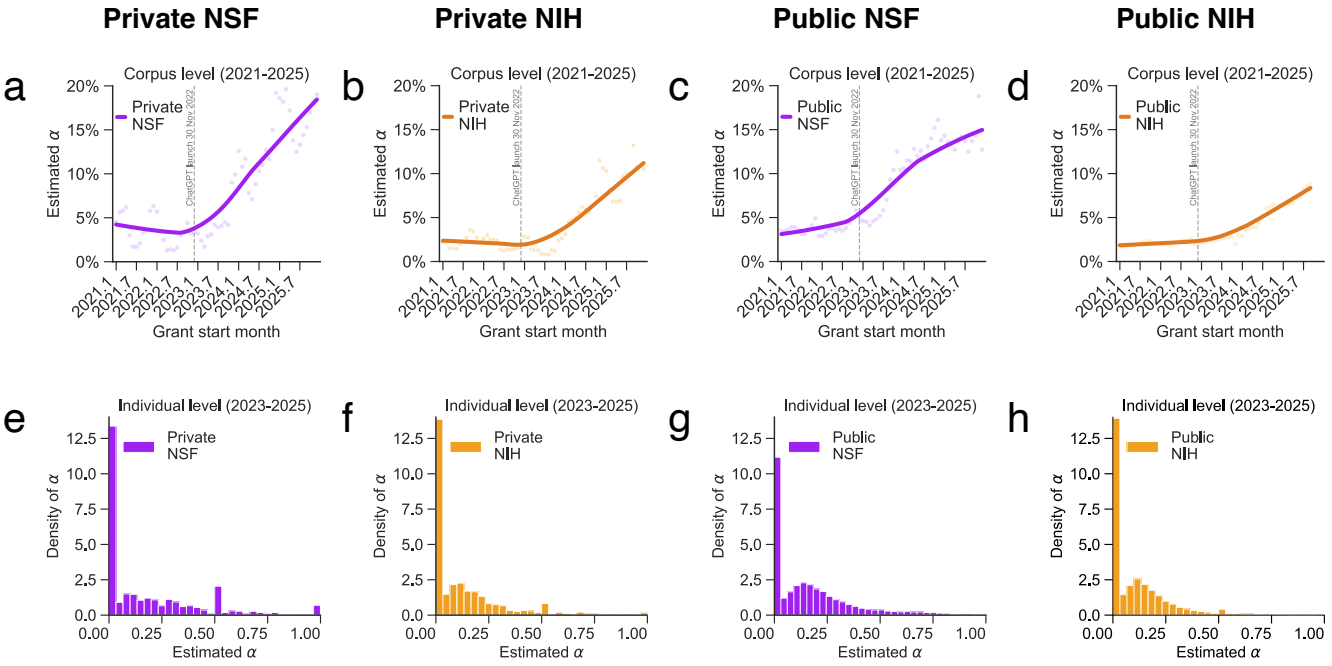


Fig. 1: **Rapid rise and bimodal distribution of LLM use in US federal research funding.** (**a**-**d**) Corpus-level estimates of LLM use (α) for private and public NSF and NIH grants from 2021 to 2025, computed using rolling three-month windows (points). Solid lines show locally weighted regressions. The vertical dashed line marks November 30, 2022, corresponding to the public release of ChatGPT. (**e**-**h**) Distributions of individual-grant α for private and public NSF and NIH proposals and awards with start dates between 2023 and 2025, showing a bimodal pattern consistent with a split between minimal and substantive LLM use across grants.

LLM use and semantic distinctiveness in US federal research funding

The rapid and uneven adoption of LLMs prompts another critical line of inquiry: Is greater LLM involvement associated with a shift in the kinds of ideas that enter, and ultimately shape, the federal research portfolio? To quantify the positioning of ideas, we measure the semantic distinctiveness of each proposal or award abstract (D1-D4) relative to the agency's recent funding frontier, building on recent work using transformer-based embeddings[26,37]. Specifically, for grants starting between 2023 and 2025, we compute each abstract's average cosine distance (using SPECTER2[38] embeddings, an established embedding method for scientific texts) to all abstracts funded in the previous year within the same agency. We then convert these distances into within-year percentiles, where higher

values indicate more distinctive ideas relative to other grants in that agency-year.

We relate distinctiveness to LLM involvement (α), estimating regressions that include grant start year, field, and investigator (PI and co-PI) fixed effects, along with controls for requested or awarded funding amount (Supplementary Note S2.2). These fixed effects absorb time-invariant heterogeneity at the field and investigator levels while accounting for common shocks across grant start years. Across all four datasets, including private submissions and public awards for both NSF and NIH, higher α is consistently associated with lower distinctiveness percentiles (Fig. 2, Supplementary Note S2.2, Supplementary Tables S1-S4), indicating that proposals and awards with greater LLM involvement are positioned closer to the grants that the agency funded most recently. For example, for NSF awards, moving from low LLM involvement (25th percentile) to high LLM involvement (75th percentile) corresponds to a ˜5-point decrease in the distinctiveness percentile; for NIH awards, the decrease is ˜4 points. Importantly, the investigator fixed effects compare each investigator only to themselves across proposals and awards rather than comparing different investigators to one another. We see that for the same investigator's proposals and awards, when the LLM involvement is higher, their proposals and awards become systematically less distinctive. These findings suggest that as LLM use rises, the proposal and award portfolio shifts toward the center of recently funded work, providing empirical evidence of idea homogenization in federal research funding that is already occurring across both agencies.

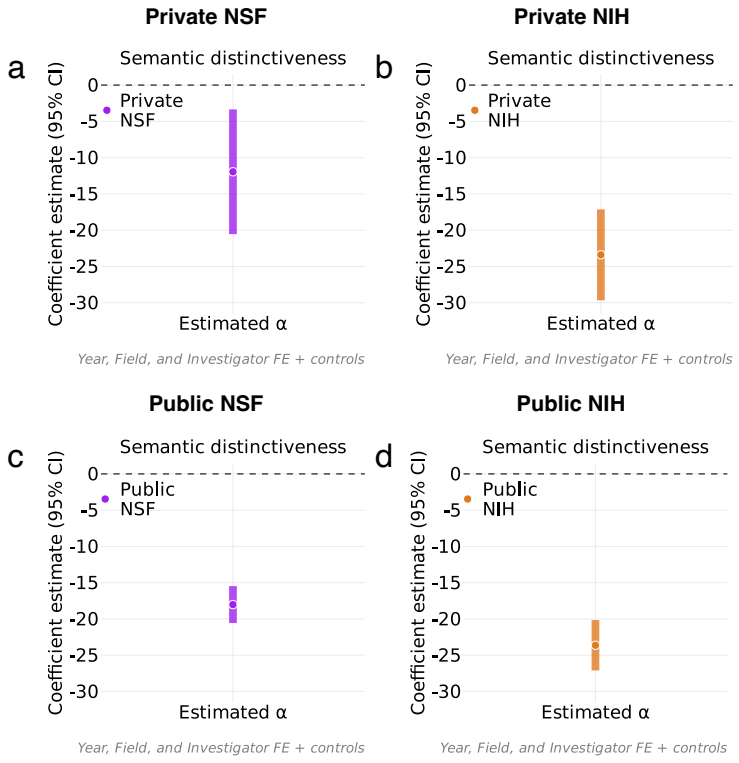


Fig. 2: **LLM use and semantic distinctiveness in US federal research funding.** (**a**-**d**) Regression estimates relating grant-level LLM use (α) to semantic distance from abstracts funded in the prior year within the same agency, expressed as within-year percentiles. Panels show results separately for private NSF (**a**), private NIH (**b**), public NSF (**c**), and public NIH (**d**) grants. All regressions include grant start year, field, and investigator fixed effects, as well as controls for funding amount. Points indicate coefficient estimates, and bars denote 95% confidence intervals. Negative coefficients correspond to proposals and awards that are positioned closer, in semantic space, to recently funded work within the same agency.

LLM use and federal research proposal success

We next ask whether LLM adoption is associated with a first-order consequence in the funding pipeline: which proposals actually receive funding. Using confidential submissions that include both funded and unfunded proposals (D1-D2), we regress proposal success on proposal-level LLM involvement (α) for request start years 2023-2025, estimating models with request start year, field, and investigator fixed effects and controlling for requested funding (Fig. 3, Supplementary Note S2.2, Supplementary Tables S5-S6). These fixed effects absorb time-invariant heterogeneity across research areas and investigators, while accounting for common shocks across years. For example, the investigator fixed effects hold constant time-invariant differences in investigators' ability, style, and topic choice, asking whether an investigator's proposals are more likely to be funded when their LLM involvement is higher.

We find no statistically significant association between α and proposal success for NSF proposals. In contrast, for NIH submissions, α is positively and significantly associated with proposal success. Specifically, moving from low LLM involvement (25th percentile) to high LLM involvement (75th percentile) corresponds to a ˜4-percentage-point jump in NIH funding probability. Overall, these results indicate that the consequences of LLM adoption for proposal selection are strongly agency-dependent: while LLM use does not systematically advantage proposals at NSF, it is associated with a higher likelihood of being funded at NIH, even for the same investigators. This agency-dependent selection pattern raises the question of translation: conditional on being funded, is LLM adoption associated with downstream outputs from these funded projects?

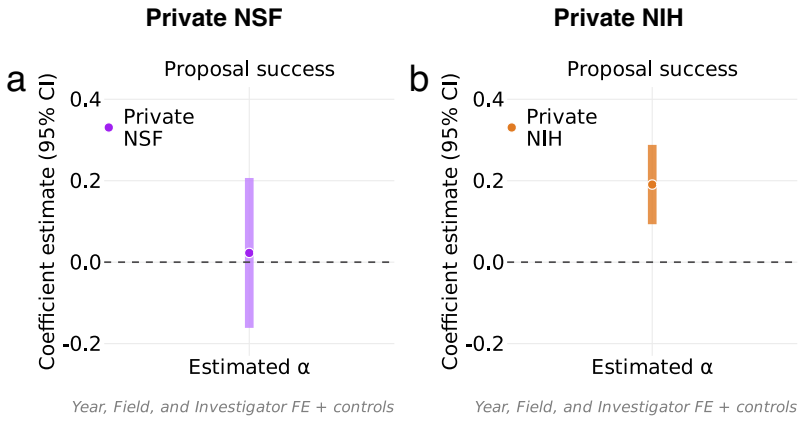


Fig. 3: **LLM use and federal research proposal success.** Based on private NSF and NIH proposal submissions from two large US R1 universities, this figure examines the relationship between LLM use at submission (α) and proposal success. (**a**) Regression estimates for NSF submissions. (**b**) Corresponding estimates for NIH submissions. All regressions include proposal request start year, field, and investigator fixed effects, as well as controls for requested funding amount. Points indicate coefficient estimates, and bars denote 95% confidence intervals.

LLM use and federal research funding outputs

Using the public NSF and NIH award data (D3-D4), we next relate grant-level LLM involvement (α) to subsequent publication output for grants starting in 2023-2024, estimating regression models that control for funding amount and include grant start year, field, and investigator fixed effects (Fig. 4, Supplementary Note S2.2, Supplementary Tables S7-S12).

We again find no statistically significant relationship between α and publication output for NSF awards (Fig. 4a). However, for NIH awards, higher α is strongly associated with more resulting publications (Fig. 4b). Moving from low LLM involvement (25th percentile) to high LLM involvement (75th percentile) predicts ˜5% more publications for NIH grants. Moreover, this NIH productivity premium appears mostly concentrated in publication volume rather than high-impact outputs: when we restrict our analyses to highly-cited papers (e.g., top 5% papers in Figs. 4c-d; top 1% in Supplementary Fig. S1), the relationship attenuates and is no longer statistically significant at conventional levels. Taken together, these results suggest that the downstream consequences of LLM adoption are also agency-dependent. At NIH, higher α is associated with more follow-on publications, but the additional output is concentrated in non-hit papers rather than the most highly-cited work.

These productivity patterns both align with and qualify recent evidence of large output gains following LLM diffusion in manuscript production[16]. While we observe a positive association between LLM involvement and publication output for NIH-funded grants, the magnitude of this association is smaller than the large increases documented in preprint production and is strongly agency-dependent, with no detectable relationship at NSF. In the federal funding context, where projects are selected through peer review and outputs reflect execution constraints and portfolio objectives rather than writing speed alone, LLM adoption appears to translate into additional publications primarily at NIH, and predominantly among non-hit papers, while showing no detectable relationship with publication output at NSF.

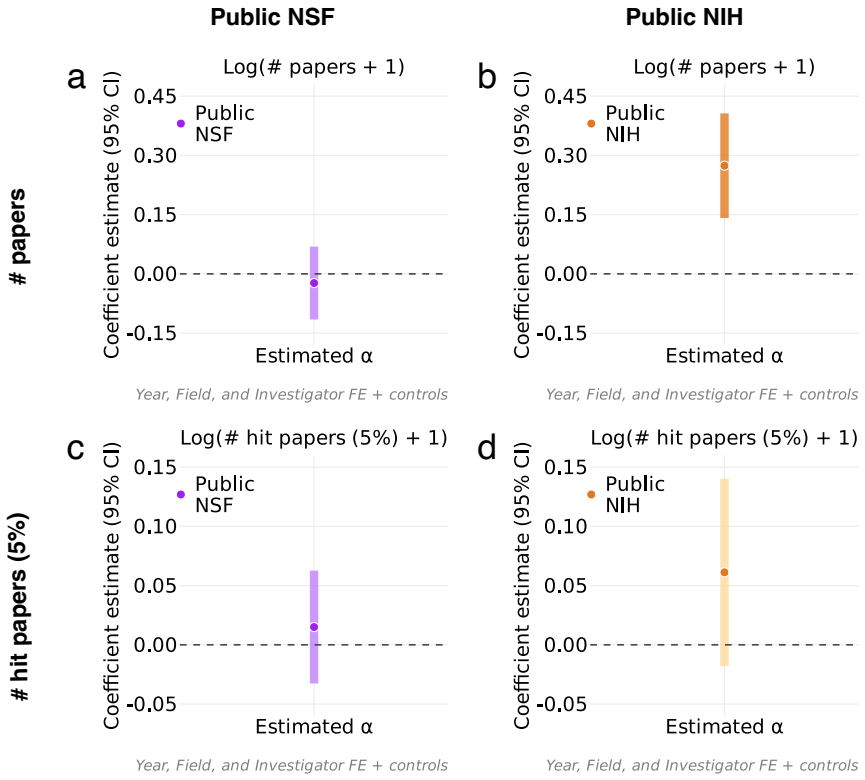


Fig. 4: **LLM use and federal research funding outputs.** (**a**-**b**) Regression estimates relating grant-level LLM use (α) to the total number of resulting publications for NSF (**a**) and NIH (**b**) grants. (**c**-**d**) Corresponding estimates for high-impact outputs, where a "hit" paper is defined as one whose citations fall within the top 5% of all papers published worldwide in the same year and field. All regressions include grant start year, field, and investigator fixed effects, as well as controls for funding amount. Points indicate coefficient estimates, and bars denote 95% confidence intervals.

Robustness and sensitivity checks

We conduct a series of robustness and sensitivity checks across numerous dimensions. First, to better characterize proposals and awards with higher LLM involvement, we follow prior work[16] and compute the Flesch Reading Ease score for each abstract as a measure of writing complexity (Supplementary Note S3.1). Consistent with prior work[16], we find that higher LLM involvement is positively associated with greater writing complexity in both proposals and awards (Supplementary Fig. S2).

Second, recent work has shown that the use of promotional language in grant proposals is positively associated with funding outputs and related proposal characteristics[14]. To assess the extent to which our results may be driven by promotional words, we conduct a robustness analysis that explicitly removes promotional words at all stages of our text processing pipeline. The resulting LLM-use measures remain highly similar to the original estimates across all datasets, indicating that our results are robust to the exclusion of promotional words (Supplementary Note S3.2, Supplementary Fig. S3).

Third, in the main analysis of semantic distinctiveness, we define the comparison set using grants funded in the preceding year. We redefine the comparison set using funded grants starting in the prior two years, finding the results are robust (Supplementary Note S3.3, Supplementary Fig. S4).

Fourth, while our primary measure of semantic distinctiveness relies on cosine distances between embedding vectors, we also compute distances using L2 metrics and obtain consistent results (Supplementary Note S3.4, Supplementary Fig. S5).

Fifth, in our primary specifications, we use ordinary least squares regressions. We re-estimate proposal success using logistic regressions and publication outputs using negative binomial regressions and continue to observe the same associations between α and outcomes (Supplementary Note S3.5, Supplementary Figs. S6-S7).

We further examine robustness by field using public award data where sample sizes are sufficiently large to support field-level analyses and find broadly consistent results (Supplementary Note S3.6, Supplementary Figs. S8-S13). Lastly, we conduct analogous robustness checks by subagency using the public award data and again find broadly consistent patterns (Supplementary Note S3.7, Supplementary Figs. S14-S19).

Discussion

Overall, our findings reveal that the rapid diffusion of LLMs is already reshaping key stages of the US federal research funding pipeline, from proposal submission to downstream scientific output. By combining confidential proposal submissions with the full population of NSF and NIH awards, we provide systematic evidence that LLM adoption is associated with (i) how scientific ideas are positioned relative to recent funded work, (ii) whether proposals clear the funding threshold, and (iii) the publication output that funded projects go on to produce. These results show that generative AI is not only changing how scientific text is produced, but also beginning to reshape the composition and consequences of public investment in research.

A central concern in science policy is maintaining a diverse and exploratory research portfolio, one that supports both cumulative progress and the pursuit of unconventional ideas. Across both agencies and both stages (proposal submission and award funding), higher LLM involvement is consistently associated with lower semantic distinctiveness: proposals with greater LLM traces are positioned closer to recently funded work within the same agency. This pattern echoes recent evidence that AI tools can expand individual scientists' productivity and impact

while simultaneously contracting the collective focus of science[26], extending that tension to the upstream context of public research funding and proposal selection. A portfolio that is closer to recent funding patterns may reflect improved clarity, tighter alignment with reviewer expectations, or lower transaction costs in articulating a fundable project. But it also implies reduced exploration in the idea landscape, which matters for public funders explicitly tasked with sustaining high-variance discovery, with implications for long-run impact and sustainability of science.

Importantly, the downstream consequences of LLM adoption are agency-dependent. In the confidential submission data, LLM involvement appears unrelated to proposal success at NSF but positively associated with funding at NIH. In the public award data, we similarly detect no relationship between LLM involvement and publication output at NSF, whereas at NIH, higher LLM involvement predicts greater publication output, though concentrated in non-hit papers rather than the most highly-cited work. One possible interpretation is that NIH funding and review norms more strongly reward incremental, executable projects that yield multiple publications, and LLM-assisted drafting may help proposals conform to those established templates. In addition, publication counts capture a larger share of NIH outputs than NSF outputs, which often include software, datasets, instrumentation, and training. It is also possible that high-impact work unfolds over longer horizons than our follow-up window. Disentangling these possibilities and, more generally, explaining why LLM adoption translates into selection and output effects in some contexts but not others, is a key direction for future research.

This context dependence is especially striking when contrasted with recent large-scale evidence that LLM diffusion has been accompanied by sharp increases in manuscript production[16]. Taken together, the manuscript setting and the funding-to-output setting suggest a simple but consequential hypothesis: a meaningful share of the productivity gains associated with LLM adoption may operate through accelerating communication—the speed and ease with which researchers articulate, package, and iterate on ideas—rather than accelerating execution, which remains constrained by project design, coordination, data collection, and other bottlenecks that LLM-enabled writing does not directly resolve. This interpretation is necessarily tentative, given limitations of publication-based output measures and the time required for high-impact work to accrue. Nonetheless, the contrast highlights an important implication for science policy: the benefits of generative AI may accrue first and most strongly where writing is the binding constraint, whereas their effects on the composition and ultimate impact of publicly funded research also hinge on institutional incentives and execution constraints.

Our study also helps address a persistent blind spot in the literature: how generative AI enters the funding pipeline at the moment when ideas are first articulated for review. Because unfunded proposals are rarely observable at scale, most prior work has had to infer behavior from public records of funded projects or published papers. By directly observing confidential submissions alongside the population of funded awards, we can separate proposal preparation from post-award editing and examine selection outcomes, not just characteristics of funded projects. It is important to note that our measurement strategy—textual traces of LLM involvement in proposal abstracts—captures only one dimension of how LLMs may be used in proposal preparation. Investigators may also use LLMs for ideation, literature synthesis, experimental planning, or other tasks that leave fewer detectable footprints in the abstract. Accordingly, our estimates should be interpreted as measuring how LLMs shape the articulation and positioning of ideas in proposal narratives, not the full scope of LLM use in research development.

Finally, because federal grants deploy public funds, the rise of LLMs in proposals has implications not only for scientific efficiency but also for accountability and trust. If LLM adoption systematically pulls proposals toward safer, more easily justifiable ideas, it could quietly tilt public portfolios toward exploitation even as agencies emphasize the value of high-risk, high-reward research. At the same time, heavy reliance on LLMs in documents

intended to reflect investigators' original ideas raises questions about authorship, responsibility, and the integrity of the funding process. By quantifying LLM involvement at scale and documenting its associations with idea positioning, selection, and downstream output, our study provides an empirical foundation for evidence-informed policy, ranging from disclosure norms and reviewer training to mechanisms that protect high-variance exploration and preserve portfolio diversity as generative AI becomes increasingly embedded in scientific practice.

Acknowledgments

We thank Yian Yin for thoughtful comments. We thank all members of the Center for Science of Science and Innovation (CSSI) at Northwestern University for helpful discussions, and Alyse Freilich for her careful editing and valuable feedback. This work is supported by the National Science Foundation under Award Number 2404035. The data collection protocol on private NSF and NIH proposals from two US universities was reviewed and approved by Institutional Review Board of Northwestern University (IRB no. STU00222200).

Author contributions

Y.Q. and D.W. jointly conceived the project, Y.Q., A.F., and D.W. designed the experiments, Y.Q. collected data, Y.Q. and Z.W. performed empirical analyses with the help from A.F., Y.B., and E.S., all authors collaboratively interpreted results, Y.Q., A.F., and D.W. wrote the manuscript, and all authors edited the manuscript.

Competing interests

The authors declare no competing interests.

Data availability

Data necessary to reproduce all plots will be made freely available.

Code availability

Code necessary to reproduce all plots will be made freely available.

Supplementary Note for
The Rise of Large Language Models and the Direction and Impact of US Federal Research Funding

S1 Datasets

This study integrates four complementary datasets capturing distinct stages of the US federal research funding pipeline at the National Science Foundation (NSF) and the National Institutes of Health (NIH). Two datasets consist of confidential proposal submissions obtained directly from universities (D1-D2), while two other datasets are constructed from publicly released award records (D3-D4). Together, these data enable us to characterize the evolution of LLM use in federal funding and to examine how it relates to the positioning of scientific ideas, their selection for funding, and their translation into publicly supported research output.

S1.1 Private NSF and NIH proposal data (D1-D2)

D1 and D2 comprise confidential NSF and NIH proposal submissions collected from two large US R1 research universities, under approved Institutional Review Board (IRB) protocols. Data from the first university were obtained in May 2025, and data from the second university were obtained in November 2025. These data include the full population of proposals submitted to NSF (D1) and NIH (D2) with request start dates between 2021 and 2025, covering funded, unfunded, and pending proposals. The inclusion of unfunded proposals enables analyses that are not possible using public data alone. In total, D1 contains 1.6K NSF proposal submissions, and D2 contains 4.1K NIH proposal submissions over the study period (2021-2025).

For each proposal, the datasets include the proposal abstract text, which serves as the primary textual input for estimating LLM involvement. In addition, the data include proposal-level metadata used in the empirical analyses, including the funding agency (NSF or NIH), request start year, requested funding amount, funding outcome (funded, unfunded, or pending), and investigator names. We link investigators to Dimensions using names and affiliations and use Dimensions researcher IDs as unique investigator identifiers. One university provides principal investigator (PI) names only, whereas the other provides both PI and co-PI names. We assign scientific fields to each proposal by identifying, among grants funded by the same agency, the grant whose abstract has the smallest cosine distance to the proposal abstract based on SPECTER2 embeddings, and assigning the corresponding grant fields from Dimensions[39] to the proposal. Field assignments use Dimensions' main (level-0) field classification (e.g., Information and Computing Sciences, Biological Sciences), which defines 22 fields. These variables allow us to estimate proposal-level LLM involvement, relate it to semantic distinctiveness, and examine proposal success while controlling for funding amount and investigator, field, and year fixed effects.

S1.2 Public NSF and NIH award data (D3-D4)

D3 and D4 consist of the full population of publicly released NSF and NIH funded awards constructed from Dimensions. Dimensions (https://www.dimensions.ai/) is a large-scale, integrated research information platform that uniquely enables systematic linking of research grants to their resulting publications through funding acknowledgment data, making it particularly well suited for studying downstream scientific outputs of funded grants. D3 includes NSF awards with start dates between 2021 and 2025, and D4 includes NIH awards with start dates over

the same period, aligning the observation window with the private proposal data. The award data were downloaded from Dimensions in November 2025. In total, D3 contains 57K NSF awards, and D4 contains 74K NIH awards.

For each funded award, the datasets include the award abstract text, which serves as the primary textual input for estimating LLM involvement at the award stage. In addition, the data include award-level metadata used in the empirical analyses, including the funding agency (NSF or NIH), award start year, awarded funding amount, scientific fields, and investigator names and identifiers. Both PI and co-PI identifiers are provided directly by Dimensions. Scientific fields are also directly obtained from Dimensions using its main (level-0) field classification system (e.g., Information and Computing Sciences, Biological Sciences), in which Dimensions defines 22 fields.

To examine downstream scientific output, we link funded awards in D3 and D4 to their resulting publications in Dimensions, including both preprints and published articles. When a paper has both a preprint and a published version, we retain only the published version to avoid double counting. Using these linked records, we construct measures of total publication output and high-impact output based on field- and year-normalized citation percentiles, where citations are measured as cumulative citations through November 2025. These measures allow us to relate award-level LLM involvement to semantic distinctiveness relative to recently funded work within each agency and to assess downstream research outputs while controlling for investigator, field, and year fixed effects.

The combination of private proposal submissions (D1-D2) and public award records (D3-D4) enables a unified analysis across multiple stages of the funding process. The private data capture idea articulation and selection at submission, including proposals that were never funded, while the public data capture post-selection outcomes, including funded project descriptions and subsequent publications.

S2 Methods

S2.1 LLM detection

Following recent work [15,16,20], we quantify the extent of large language model (LLM) use in grant proposal and award abstracts using the distributional LLM-quantification framework introduced by Liang *et al.* [15]. We estimate separate detection models for NSF and NIH grants to account for agency-specific writing styles and topical distributions. Below we describe the NSF pipeline; the NIH pipeline is constructed analogously.

We begin by assembling a reference corpus of 6,000 NSF grant abstracts, a random subset of publicly available NSF grant abstracts with start dates in 2021, prior to the widespread adoption of LLMs. These abstracts are used to estimate the word distribution of human-written text and to construct a corresponding LLM-generated distribution by prompting OpenAI's GPT-3.5-turbo-0125 model to rewrite the same abstracts, simulating LLM involvement in proposal language. The prompts used, which were adapted from those in Liang *et al.* [15], are shown below. Comparing these two distributions allows us to identify systematic differences between human and LLM-modified writing.

More specifically, the framework consists of the following steps applied to the NSF datasets (D1 and D3):

1. **Problem formulation:** denote X as each single sentence, $\mathcal{P}$ and $\mathcal{Q}$ as the population-level distributions of human-written and LLM-modified sentences, respectively. The mixture distribution is given by

$$\mathcal{D}_\alpha(X) = (1-\alpha)\mathcal{P}(x) + \alpha\mathcal{Q}(x), \tag{1}$$

```
The aim here is to reverse-engineer the author's writing process by taking a grant
abstract and compressing it into a more concise form.  This process simulates how an
author might distill their thoughts and key points into a structured, yet not overly
condensed form.  Now as a first step, given a grant abstract, reverse-engineer it into
a list of bullet points.
```

```
Following the initial step of reverse-engineering the author's writing process by
compressing a grant abstract, you now enter the second phase.  Here, your objective is to
expand upon the concise version previously crafted.  This stage simulates how an author
elaborates on the distilled thoughts and key points, enriching them into a detailed,
structured narrative similar in length to the original abstract.  Given the concise output
from the previous step, your task is to develop it into a fully fleshed-out text.
```

where α is the mixture weight of the LLM-modified sentence distribution in the observed data. The goal is to estimate α based on observed sentences $\{X_i\}_{i=1}^{N} \sim \mathcal{D}_\alpha(X)$, where i is an integer index of observed sentences.

2. **Parameterization:** to make α identifiable, the framework models the distributions of token occurrences in human-written and LLM-modified sentences, denoted as $\mathcal{P}_T$ and $\mathcal{Q}_T$, respectively, for a chosen list of tokens $T = \{t_i\}_{i=1}^{M}$. The occurrence probabilities of each token in human-written and LLM-modified sentences, p_t and q_t, are used to parameterize $\mathcal{P}_T$ and $\mathcal{Q}_T$:

$$\mathcal{P}_T(X) = \prod_{t \in T} p_t^{\mathbb{1}\{t \in X\}} (1 - p_t)^{\mathbb{1}\{t \notin X\}}, \qquad \mathcal{Q}_T(X) = \prod_{t \in T} q_t^{\mathbb{1}\{t \in X\}} (1 - q_t)^{\mathbb{1}\{t \notin X\}}. \tag{2}$$

3. **Estimation:** the occurrence probabilities p_t and q_t are estimated using collections of known human-written and LLM-modified sentences, $\{X_j^P\}_{j=1}^{n_P}$ and $\{X_j^Q\}_{j=1}^{n_Q}$, respectively, here j is an integer index of documents with known sources

$$\hat{p}_t = \frac{1}{n_P} \sum_{j=1}^{n_P} \mathbb{1}\{t \in X_j^P\}, \qquad \hat{q}_t = \frac{1}{n_Q} \sum_{j=1}^{n_Q} \mathbb{1}\{t \in X_j^Q\}. \tag{3}$$

4. **Inference:** the fraction α is estimated by the maximum likelihood estimator (MLE) on the observed sentences under the mixture distribution $\hat{\mathcal{D}}_{\alpha,T}(X) = (1 - \alpha)\hat{\mathcal{P}}_T(X) + \alpha\hat{\mathcal{Q}}_T(X)$:

$$\hat{\alpha}_T^{\text{MLE}} = \arg\max_{\alpha \in [0,1]} \sum_{i=1}^{N} \log\left((1 - \alpha)\hat{\mathcal{P}}_T(X_i) + \alpha\hat{\mathcal{Q}}_T(X_i)\right). \tag{4}$$

S2.2 Regression models

To examine the relationship between LLM involvement in grants and semantic distinctiveness, proposal success, and downstream scientific outputs, we estimate a series of ordinary least squares (OLS) linear regression models separately for NSF and NIH grants. Across all model specifications, we use a consistent set of explanatory and control variables, varying only the dependent variable to reflect the outcome of interest.

Dependent variables: We consider three classes of outcome variables, denoted y_{ijft}, where i indexes individual

proposals or awards, j indexes investigators, f denotes the scientific field, and t indicates the proposal request start year or award start year. The outcomes are defined as follows:

- **Semantic distinctiveness**: Defined as the percentile rank (ranging from 0 to 100, with higher values indicating greater distinctiveness) of the average cosine distance between the SPECTER2[40] embedding of a grant abstract and the embeddings of all grants funded by the same agency in the preceding year. This measure captures how semantically distinct a grant is relative to recently funded work and is available for both public and private datasets (D1-D4).

- **Proposal success**: Defined as a binary indicator equal to one if a proposal is funded and zero if unfunded. This outcome is observed only in the private proposal dataset (D1-D2).

- **Federal research funding outputs**: Defined as the logarithm of the number of publications resulting from a funded grant plus one, $\log(\#\text{ papers} + 1)$. In additional specifications, we examine high-impact output, where a "hit" paper is defined as a publication whose citation count falls within the top 5% of all papers published worldwide in the same field and year; the corresponding outcome is $\log(\#\text{ hit papers} + 1)$. Measures of downstream output are constructed using public award data only (D3-D4).

Explanatory variables: The key explanatory variable, α_i, measures the extent of LLM use in the grant abstract associated with proposal or award i and ranges from 0 to 1.

Control variables: To account for systematic differences across investigators, research fields, and time, all specifications include investigator fixed effects η_j, field fixed effects ϕ_f, and grant start year fixed effects μ_t. We additionally control for $\log(\text{Funding}_i)$, defined as the logarithm of the total requested funding amount for proposals or the awarded funding amount for funded grants.

Depending on the choice of outcome variable, we estimate the following specification:

$$y_{ijft} = \beta\,\alpha_i + \delta\,\log(Funding_i) + \mu_t + \phi_f + \eta_j + \varepsilon_{ijft} \tag{5}$$

The dependent variable y_{ijft} alternatively corresponds to semantic distinctiveness, proposal success, or downstream research output. Standard errors are clustered at the investigator level.

S3 Robustness Checks

S3.1 Writing complexity

To illustrate the differences between the abstracts with higher and lower levels of LLM involvement (i.e., α)[16], we compute the Flesch Reading Ease score for each abstract as a measure of writing complexity[41]. The Flesch score is a readability metric that captures syntactic complexity through sentence length, and lexical complexity through the average number of syllables per word. Since higher values of the original Flesch Reading Ease score correspond to lower complexity, we multiply the score by -1 so that higher values consistently indicate higher writing complexity (Equation 6). This transformation facilitates a more intuitive interpretation and aligns the direction of the measure with our analysis. Consistent with recent findings on preprints[16], we find that higher LLM involvement is positively associated with greater writing complexity in both proposals and awards (Fig. S2).

$$\text{Writing complexity} = 1.015 \cdot \frac{\#\,\text{total words}}{\#\,\text{total sentences}} + 84.6 \cdot \frac{\#\,\text{total syllables}}{\#\,\text{total words}} - 206.835 \tag{6}$$

S3.2 Promotional words

Recent work has shown that the use of promotional language in grant proposals is positively associated with funding outcomes and related proposal characteristics [14,42]. To assess the extent to which our results may be driven by promotional wording, we conduct a robustness analysis that explicitly removes promotional terms (139 words identified in prior studies [14,42]) at all stages of our estimation pipeline.

Specifically, we re-estimate individual-level LLM use (α) after excluding promotional words throughout the entire pipeline. Promotional terms are removed from the original abstracts prior to LLM rewriting and from the LLM-rewritten abstracts themselves.

Figure S3 compares the original α estimates with those obtained after promotional-word removal for private NSF, private NIH, public NSF, and public NIH samples. Across all settings, the recalculated α remains highly correlated with the original measure, with Pearson correlation coefficients ranging from $r = 0.95$ to $r = 0.97$.

The consistently high correlations indicate that removing promotional language has little effect on the resulting LLM-use estimates. Overall, this analysis shows that our measure of LLM use remains stable when promotional wording is excluded, supporting the robustness of our results to this alternative text preprocessing choice.

S3.3 Different prior windows for measuring semantic distinctiveness

In the main analysis of semantic distinctiveness, we define the comparison set using grants funded in the preceding one year. Our results are robust to alternative choices of the comparison window. Specifically, when we construct the reference set using grants that started in the prior two years, the estimated relationships remain qualitatively and quantitatively similar (Fig. S4).

S3.4 Different distance metrics for measuring semantic distinctiveness

In the main analysis of semantic distinctiveness, we use cosine distance to measure distance between embedding vectors. We also compute distances using L2 metrics and obtain consistent results (Fig. S5).

S3.5 Alternative regression models

While our primary specifications use ordinary least squares regressions, we re-estimate proposal success using logistic regressions (Fig. S6) and publication outputs using negative binomial regressions, and we continue to observe the same associations between α and outcomes (Fig. S7).

S3.6 Robustness by field for public data

To examine whether our main findings are consistent across award fields in the public datasets, we focus on fields where sample sizes are sufficiently large to support field-level analysis and re-estimate all main results separately by field. We focus on fields with adequate sample sizes to ensure robust estimation. Specifically, for awards with start dates between 2023 and 2025, we calculate the number of grants in each field and retain those with more than 2,000 awards. All remaining fields are merged into a single category labeled “Others.“

For NSF, in addition to "Others," the fields (ordered by the number of grants) are Information and Computing Sciences, Engineering, Education, Biological Sciences, Earth Sciences, Physical Sciences, and Chemical Sciences. For NIH, in addition to "Others," the fields (ordered by the number of grants) are Biomedical and Clinical Sciences, Biological Sciences, Health Sciences, and Psychology.

Using this grouping, we re-run the related analyses reported in the main text separately by field. We find broadly consistent results and the corresponding results are shown in Figs. S8-S13.

S3.7 Robustness by subagency for public data

To assess whether our main findings are consistent across award subagencies in the public datasets, we focus on fields where sample sizes are sufficiently large to permit subagency-level analysis and re-estimate all main results separately by subagency. We restrict attention to subagencies with adequate sample sizes to ensure robust estimation. Specifically, for awards with start dates between 2023 and 2025, we compute the number of grants in each subagency. For NSF, all subagencies are retained. For NIH, we retain subagencies with more than 2,000 awards and merge all remaining subagencies into a single category labeled "Others."

For NSF, the subagencies (ordered by the number of grants) are Directorate for Mathematical & Physical Sciences (MPS), Directorate for Computer & Information Science & Engineering (CISE), Directorate for Engineering (ENG), Directorate for Geosciences (GEO), Directorate for STEM Education (EDU), Directorate for Biological Sciences (BIO), Directorate for Technology, Innovation and Partnerships (TIP), Directorate for Social, Behavioral & Economic Sciences (SBE), and Office of the Director (OD). For NIH, in addition to "Others," the subagencies (ordered by the number of grants) are National Cancer Institute (NCI), National Institute of Allergy and Infectious Diseases (NIAID), National Heart Lung and Blood Institute (NHLBI), National Institute of General Medical Sciences (NIGMS), National Institute of Neurological Disorders and Stroke (NINDS), National Institute on Aging (NIA), National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), and Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD).

Using this grouping, we re-run the related analyses reported in the main text separately by subagency. We find broadly consistent results and the corresponding results are shown in Figs. S14-S19.

Supplementary Figures

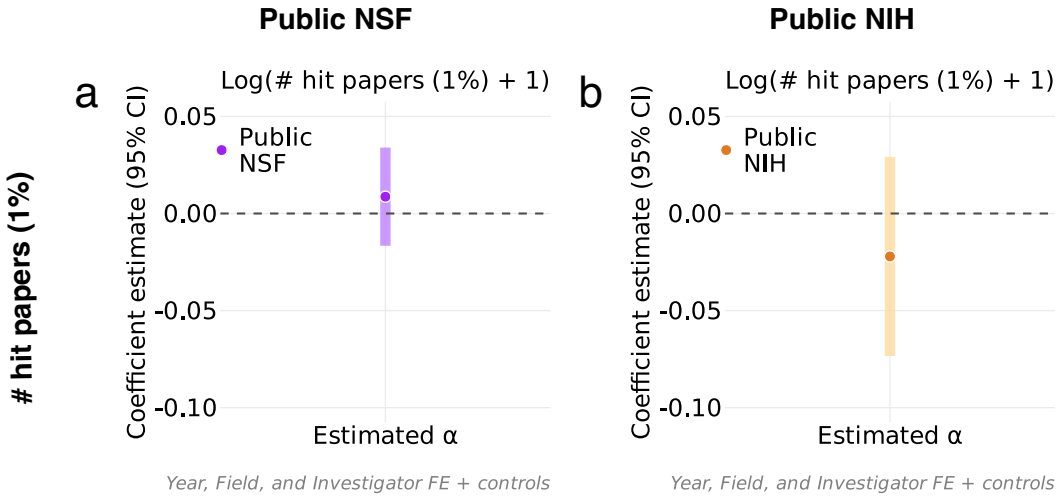


Fig. S1: **LLM use and downstream output in US federal research funding.** (**a**-**b**) Regression estimates relating grant-level LLM use (α) to high-impact outputs, where a "hit" paper is defined as one whose citations fall within the top 1% of all papers published worldwide in the same year and field. All regressions include grant start year, field, and investigator fixed effects, as well as controls for funding amount. Points indicate coefficient estimates, and bars denote 95% confidence intervals.

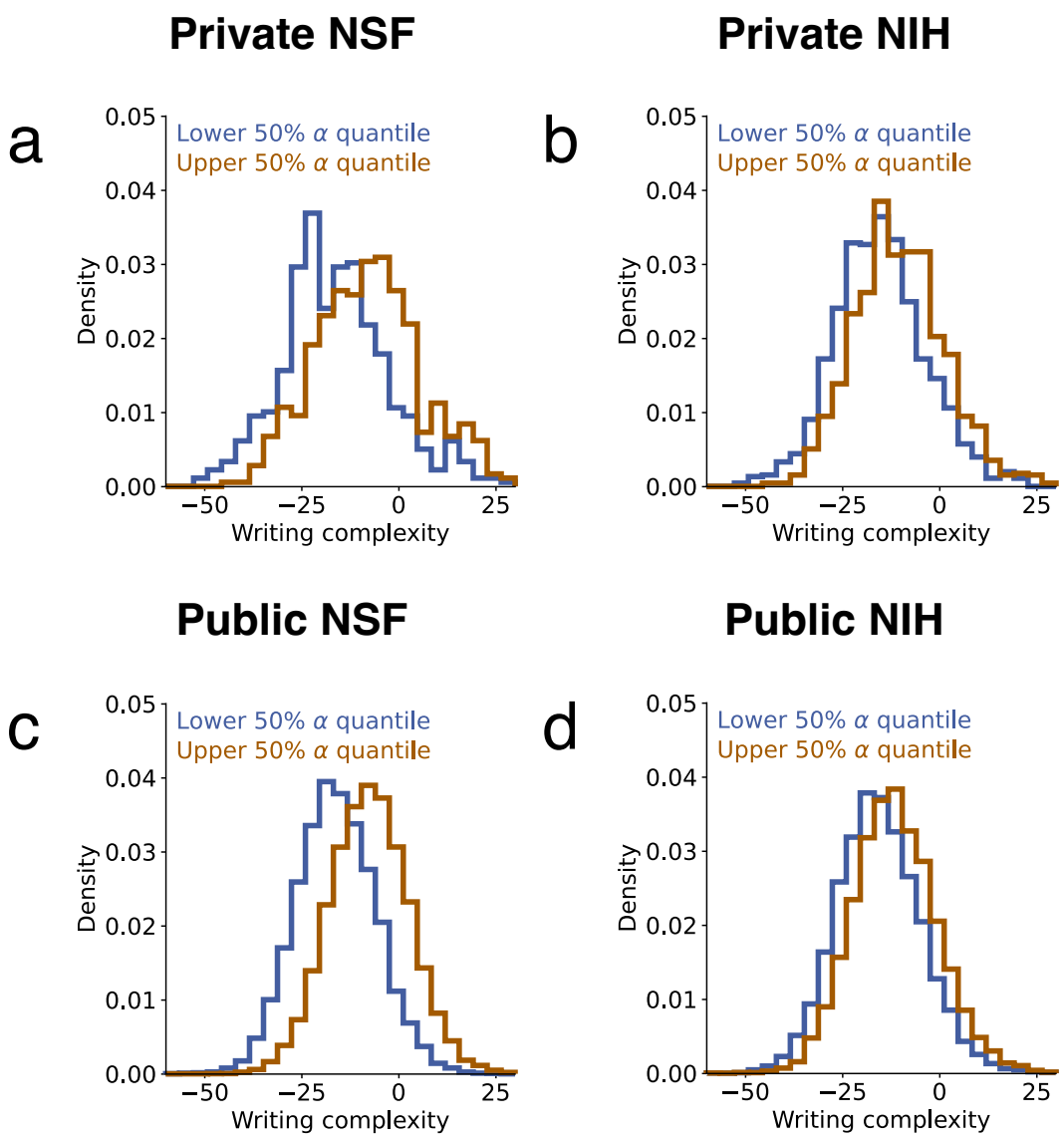


Fig. S2: **Distributions of writing complexity by LLM use intensity.** This figure uses grants with start years between 2023 and 2025 across four datasets. Writing complexity is measured based on the Flesch Reading Ease score (see Section S3.1). Grants are split by the median α value into lower 50% (blue) and upper 50% (brown) quantiles. Panels show results separately for private NSF (**a**), private NIH (**b**), public NSF (**c**), and public NIH (**d**) grants. Across all four datasets, grants with higher α values show systematically higher writing complexity. In addition, within each panel, the differences between the two groups are statistically significant (two-sided Welch's t-test, p ¡ 0.001 in all cases).

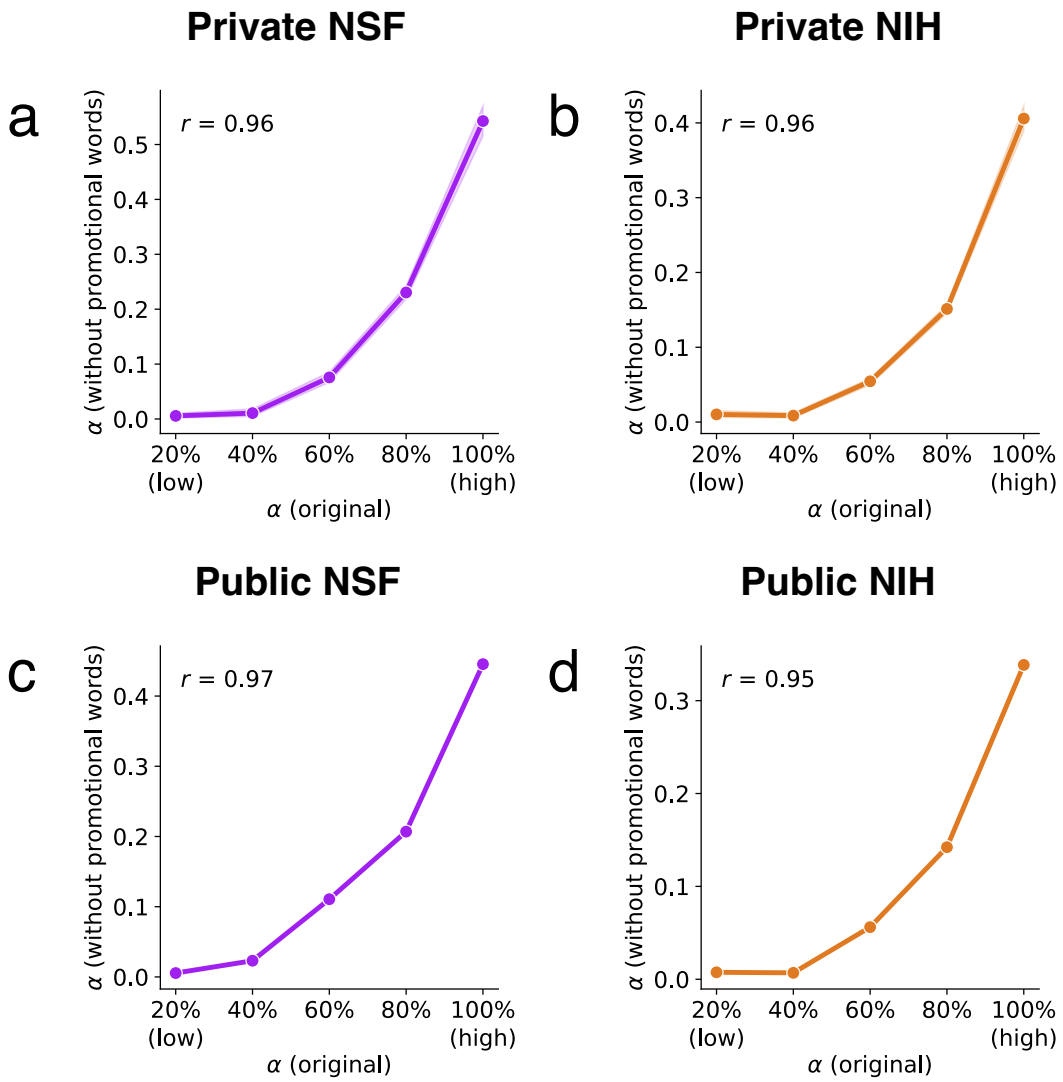


Fig. S3: **LLM-use estimates after removing promotional words.** This figure assesses whether our measure of individual-level LLM use (α) is driven by promotional language. Panels report the relationship between the original α and a recalculated α after removing promotional words for private NSF (**a**), private NIH (**b**), public NSF (**c**), and public NIH (**d**) samples. Points denote estimates at different α quantiles, with lines connecting adjacent quantiles. Across all samples, the recalculated measure remains highly correlated with the original estimate (r = 0.95-0.97, $p < 0.001$ in all cases), where r denotes the Pearson correlation coefficient between the original and recalculated α across quantiles.

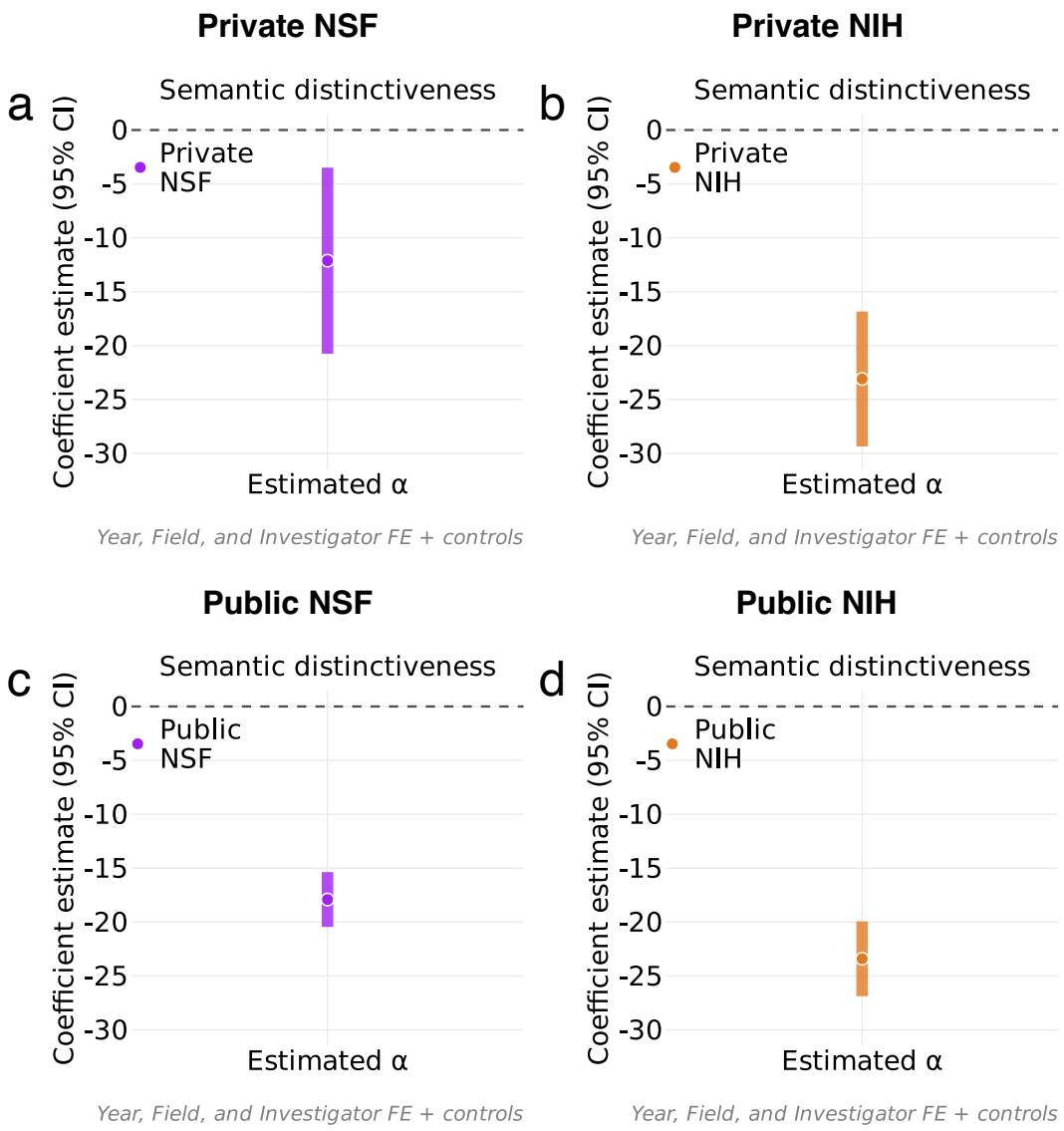


Fig. S4: **LLM use and semantic distinctiveness in US federal research funding.** Regression estimates relating grant-level LLM use (α) to semantic distance from abstracts funded in the prior two years within the same agency, expressed as within-year percentiles. Panels show results separately for private NSF (**a**), private NIH (**b**), public NSF (**c**), and public NIH (**d**) grants. All regressions include grant start year, field, and investigator fixed effects, as well as controls for funding amount. Points indicate coefficient estimates, and bars denote 95% confidence intervals. Negative coefficients correspond to proposals and awards that are positioned closer, in semantic space, to recently funded work within the same agency.

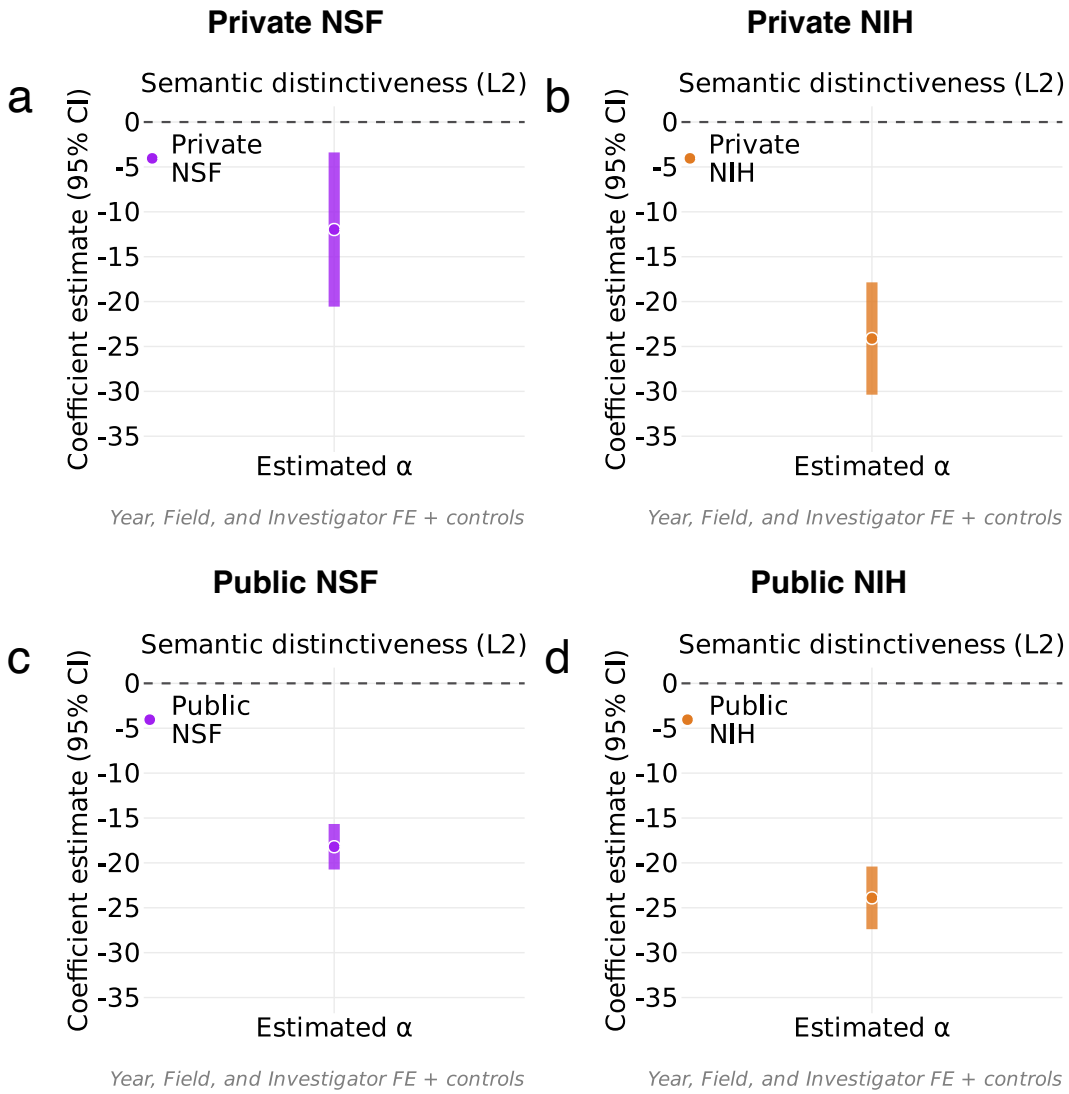


Fig. S5: **LLM use and semantic distinctiveness in US federal research funding using L2 distance.** (**a-d**) Regression estimates relating grant-level LLM use (α) to semantic distance from abstracts funded in the prior year within the same agency, expressed as within-year percentiles. Panels show results separately for private NSF (**a**), private NIH (**b**), public NSF (**c**), and public NIH (**d**) grants. All regressions include grant start year, field, and investigator fixed effects, as well as controls for funding amount. Points indicate coefficient estimates, and bars denote 95% confidence intervals. Negative coefficients correspond to proposals and awards that are positioned closer, in semantic space, to recently funded work within the same agency.

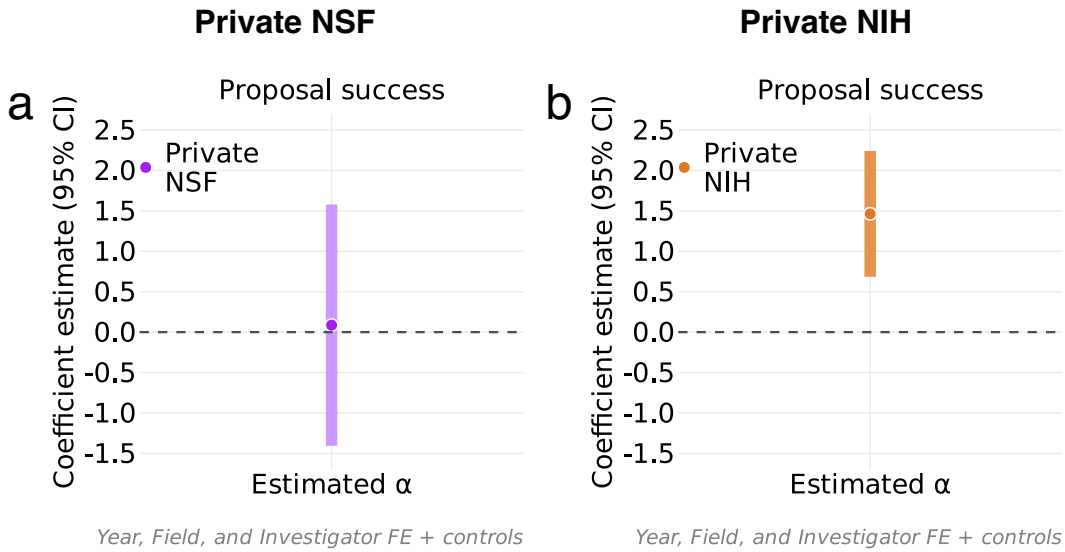


Fig. S6: **LLM use and proposal success in US federal research funding.** Using private NSF and NIH proposal submissions from two large US R1 universities, this figure examines the relationship between LLM use at submission (α) and funding success. (**a**) Logistic regression estimates for NSF submissions. (**b**) Corresponding estimates for NIH submissions. All regressions include proposal request start year, field, and investigator fixed effects, as well as controls for requested funding amount. Points indicate coefficient estimates, and bars denote 95% confidence intervals.

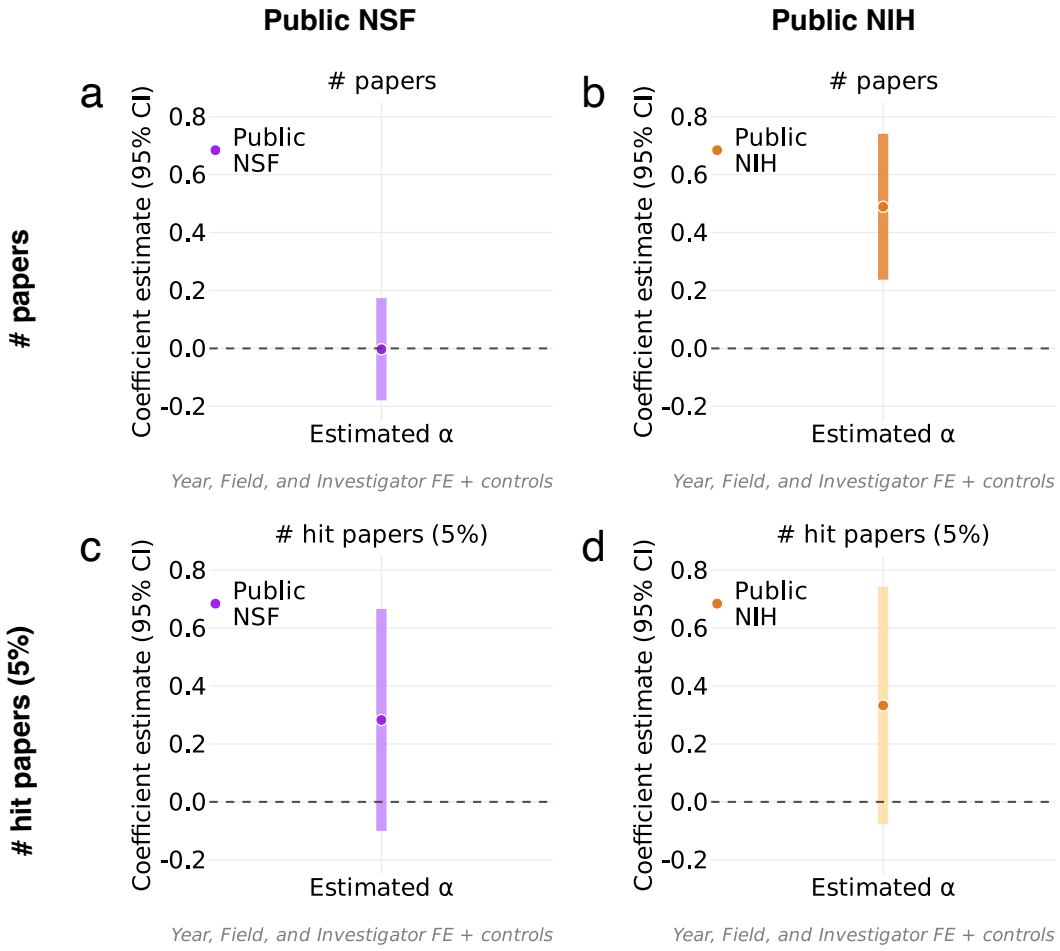


Fig. S7: **LLM use and downstream output in US federal research funding.** (**a**-**b**) Negative binomial regression estimates relating grant-level LLM use (α) to the total number of resulting publications for NSF (**a**) and NIH (**b**) grants. (**c**-**d**) Corresponding estimates for high-impact outputs, where a "hit" paper is defined as one whose citations fall within the top 5% of all papers published worldwide in the same year and field. All regressions include grant start year, field, and investigator fixed effects, as well as controls for funding amount. Points indicate coefficient estimates, and bars denote 95% confidence intervals.

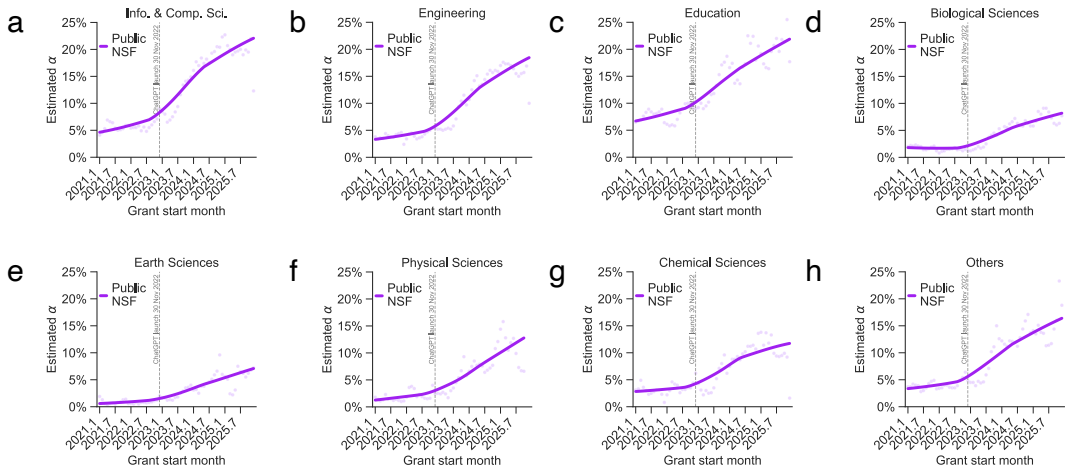


Fig. S8: **Rapid rise of LLM use in US federal research funding by field in public NSF data.** (**a**-**h**) Corpus-level estimates of LLM use (α) for NSF grants from 2021 to 2025, computed separately by field. The estimates are computed using rolling three-month windows. Solid lines show locally weighted regressions within each field. The vertical dashed line marks November 30, 2022, corresponding to the public release of ChatGPT.

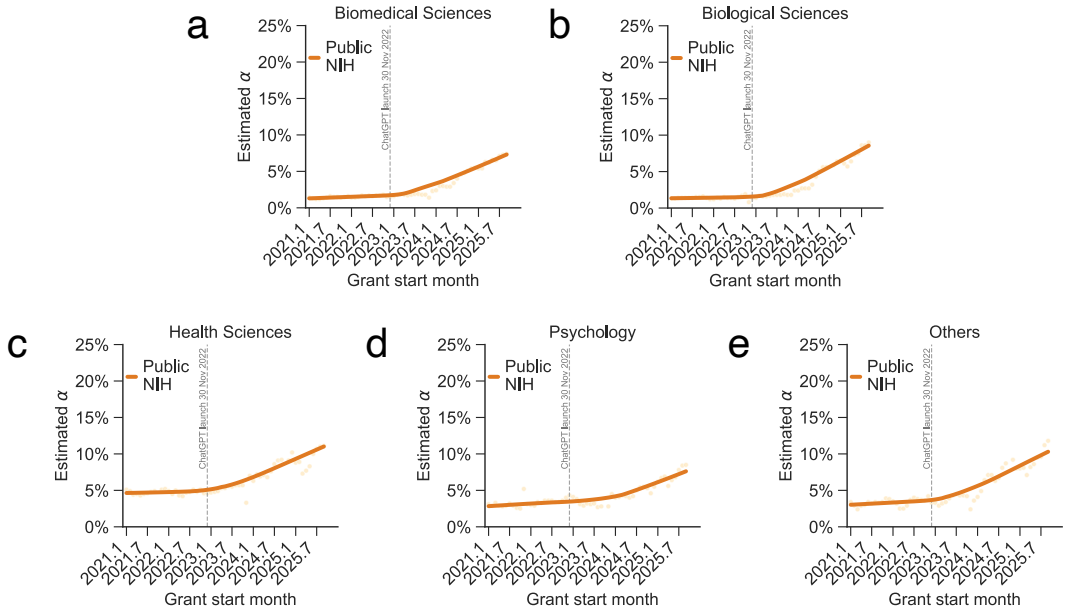


Fig. S9: **Rapid rise of LLM use in US federal research funding by field in public NIH.** (**a**-**e**) Corpus-level estimates of LLM use (α) for public NIH grants from 2021 to 2025, computed separately by field. The estimates are computed using rolling three-month windows. Solid lines show locally weighted regressions within each field. The vertical dashed line marks November 30, 2022, corresponding to the public release of ChatGPT.

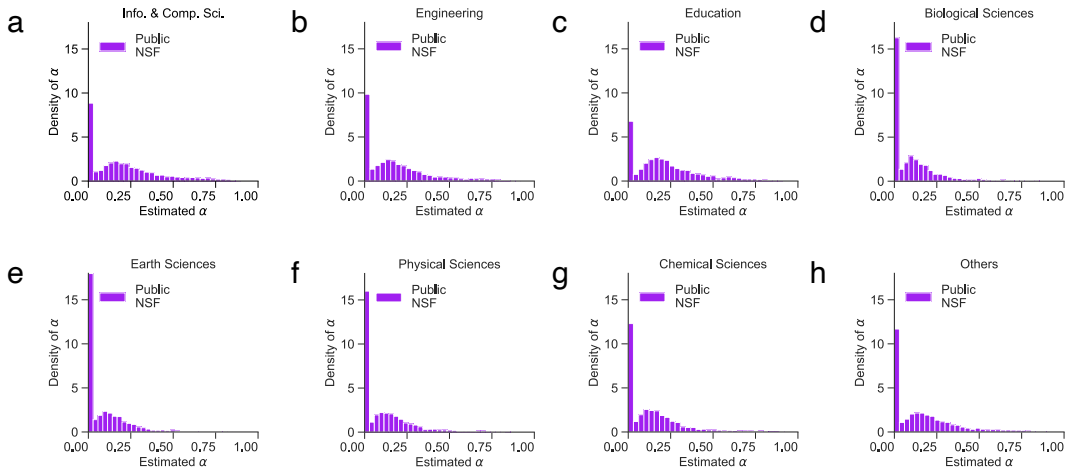


Fig. S10: **Bimodal distribution of LLM use in US federal research funding by field in public NSF data.** (**a**-**h**) Distributions of individual-grant α for NSF awards with start dates between 2023 and 2025, shown separately by field. Each field reveals a bimodal pattern consistent with a split between minimal and substantive LLM use across grants.

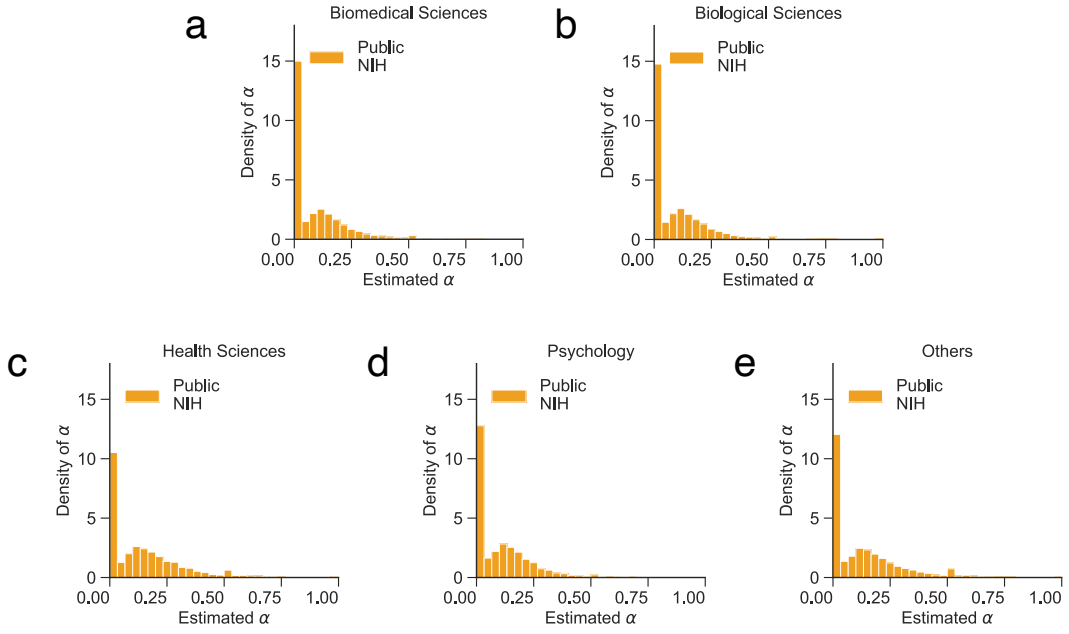


Fig. S11: **Bimodal distribution of LLM use in US federal research funding by field in public NIH data.** (**a**-**e**) Distributions of individual-grant α for NIH awards with start dates between 2023 and 2025, shown separately by field. Each field reveals a bimodal pattern consistent with a split between minimal and substantive LLM use across grants.

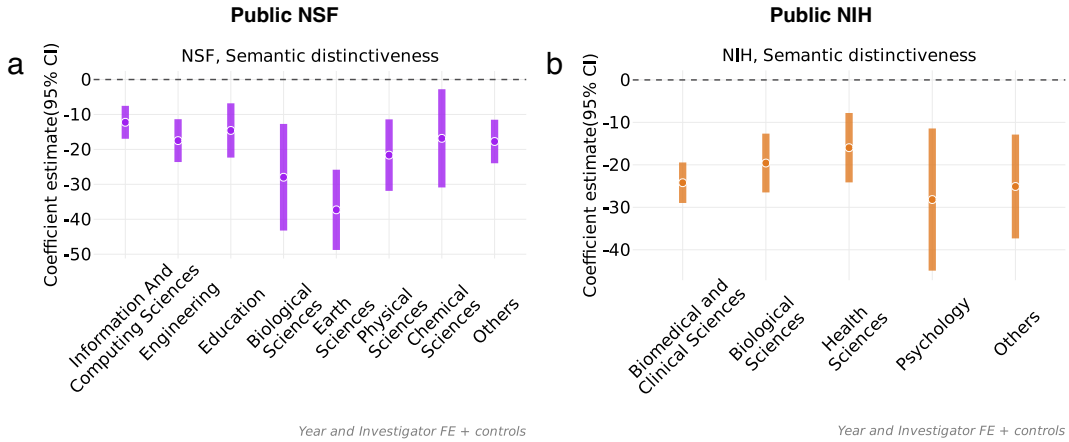


Fig. S12: **LLM use and semantic distinctiveness in US federal research funding by field.** (**a**-**b**) Regression estimates relating grant-level LLM use (α) to semantic distance from abstracts funded in the prior year within the same agency. Panels show results separately for public NSF and public NIH grants. All regressions include grant start year and investigator fixed effects, as well as controls for funding amount. Points indicate coefficient estimates, and bars denote 95% confidence intervals.

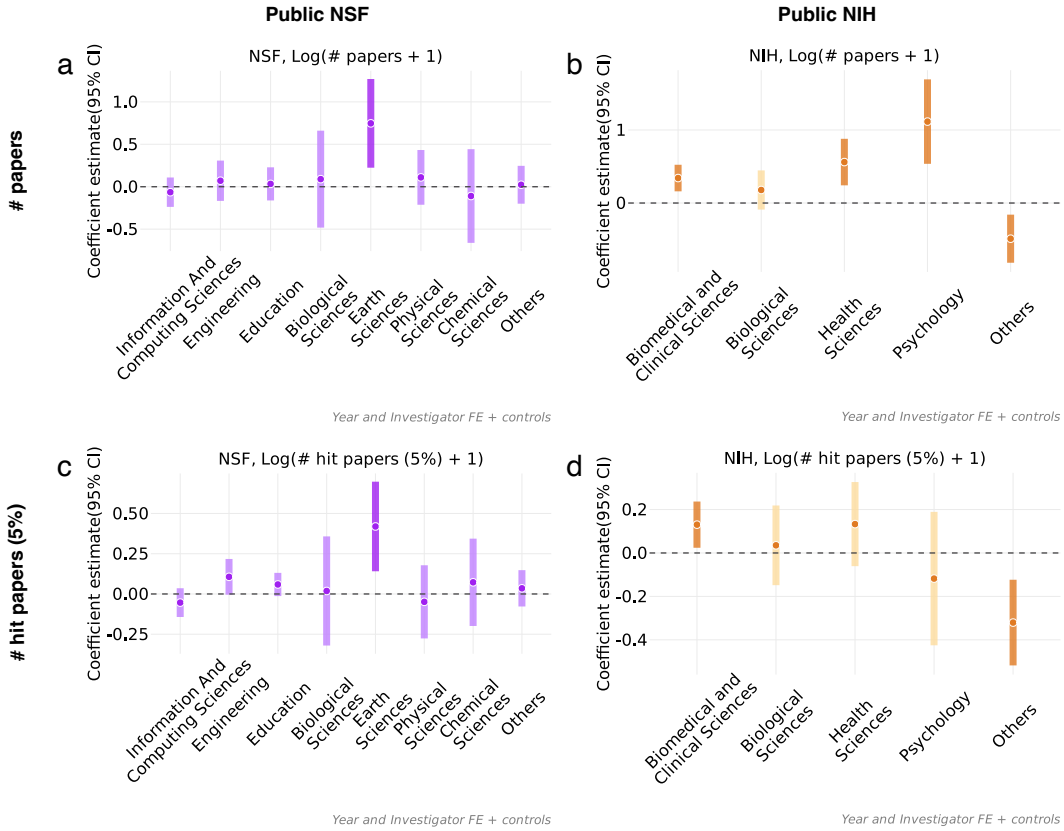


Fig. S13: **LLM use and federal research funding outputs by field.** (**a-b**) Regression estimates relating grant-level LLM use (α) to the total number of resulting publications for NSF and NIH grants. (**c-d**) Corresponding estimates for high-impact outputs, where a "hit" paper is defined as one whose citations fall within the top 5% of all papers published worldwide in the same year and field. All regressions include grant start year and investigator fixed effects, as well as controls for funding amount. Points indicate coefficient estimates, and bars denote 95% confidence intervals.

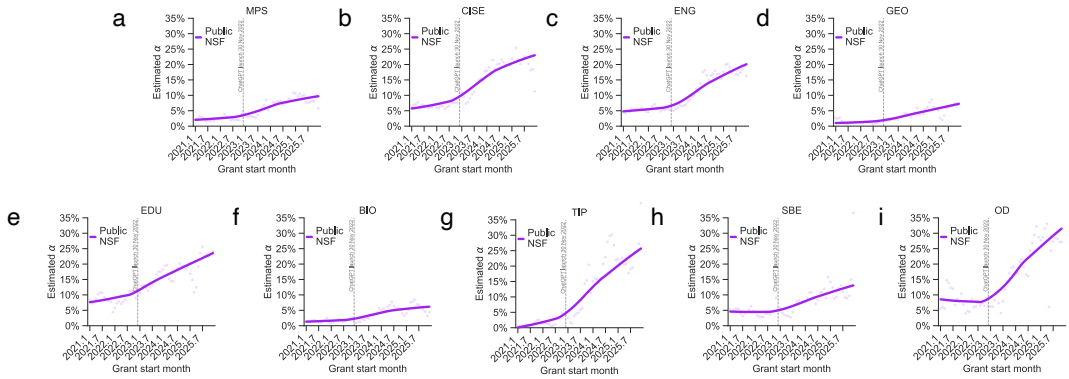


Fig. S14: **Rapid rise of LLM use in US federal research funding by subagency in public NSF data.** (**a-i**) Corpus-level estimates of LLM use (α) for NSF grants from 2021 to 2025, computed separately by subagency. The estimates are computed using rolling three-month windows. Solid lines show locally weighted regressions within each field. The vertical dashed line marks November 30, 2022, corresponding to the public release of ChatGPT.

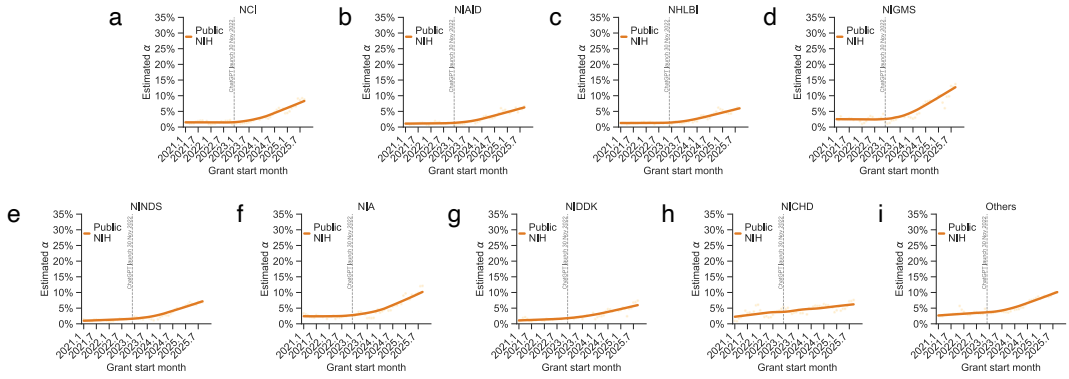


Fig. S15: **Rapid rise of LLM use in US federal research funding by subagency in public NIH data.** (**a-i**) Corpus-level estimates of LLM use (α) for NIH grants from 2021 to 2025, computed separately by subagency. The estimates are computed using rolling three-month windows. Solid lines show locally weighted regressions within each field. The vertical dashed line marks November 30, 2022, corresponding to the public release of ChatGPT.

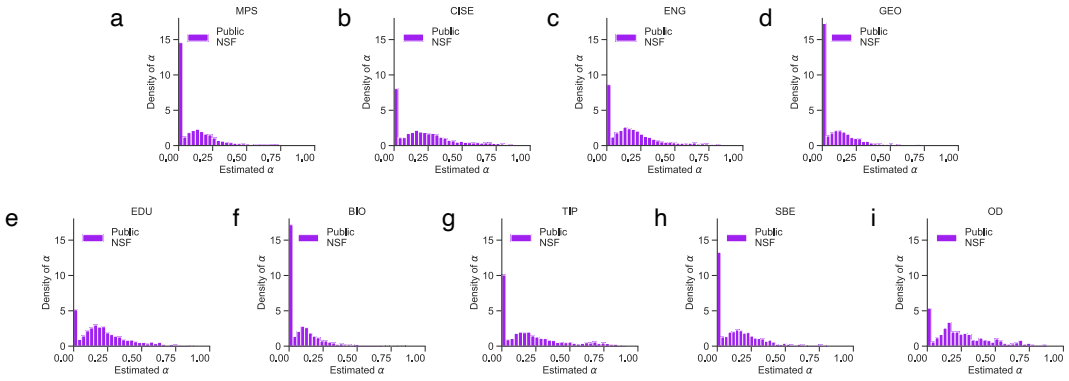


Fig. S16: **Bimodal distribution of LLM use in US federal research funding by subagency in public NSF data.** (**a**-**i**) Distributions of individual grant α for NSF awards with start dates between 2023 and 2025, shown separately by subagency. Each subagency reveals a bimodal pattern consistent with a split between minimal and substantive LLM use across grants.

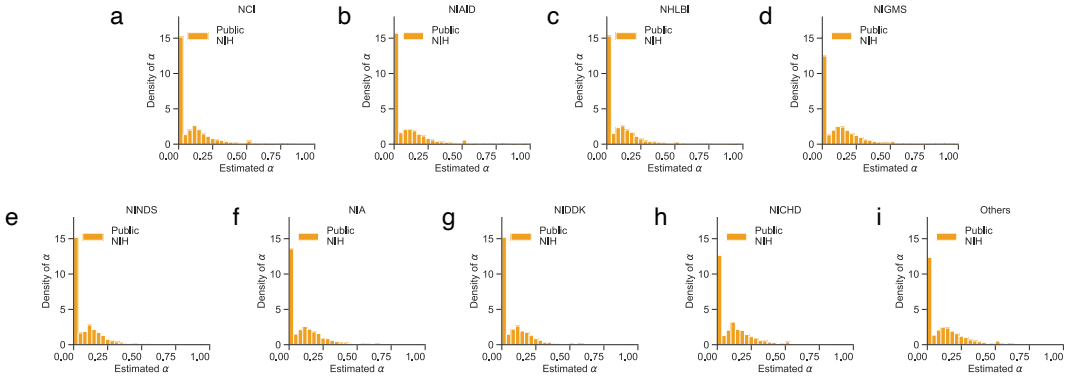


Fig. S17: **Bimodal distribution of LLM use in US federal research funding by subagency in public NIH data.** (**a**-**i**) Distributions of individual-grant α for NIH awards with start dates between 2023 and 2025, shown separately by subagency. Each subagency reveals a bimodal pattern consistent with a split between minimal and substantive LLM use across grants.

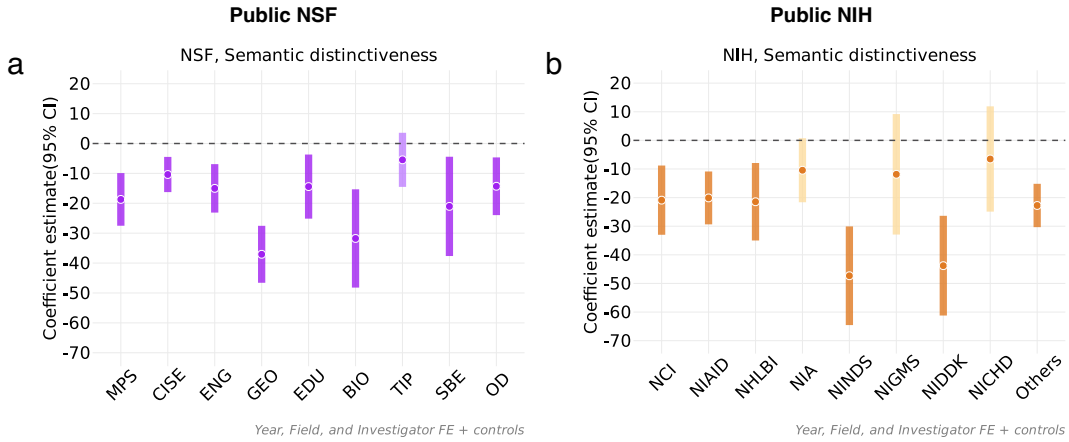


Fig. S18: **LLM use and semantic distinctiveness in US federal research funding by subagency.** (**a**-**b**) Regression estimates relating grant-level LLM use (α) to semantic distance from abstracts funded in the prior year within the same agency. Panels show results separately for public NSF and public NIH grants. All regressions include grant start year, field, and investigator fixed effects, as well as controls for funding amount. Points indicate coefficient estimates, and bars denote 95% confidence intervals.

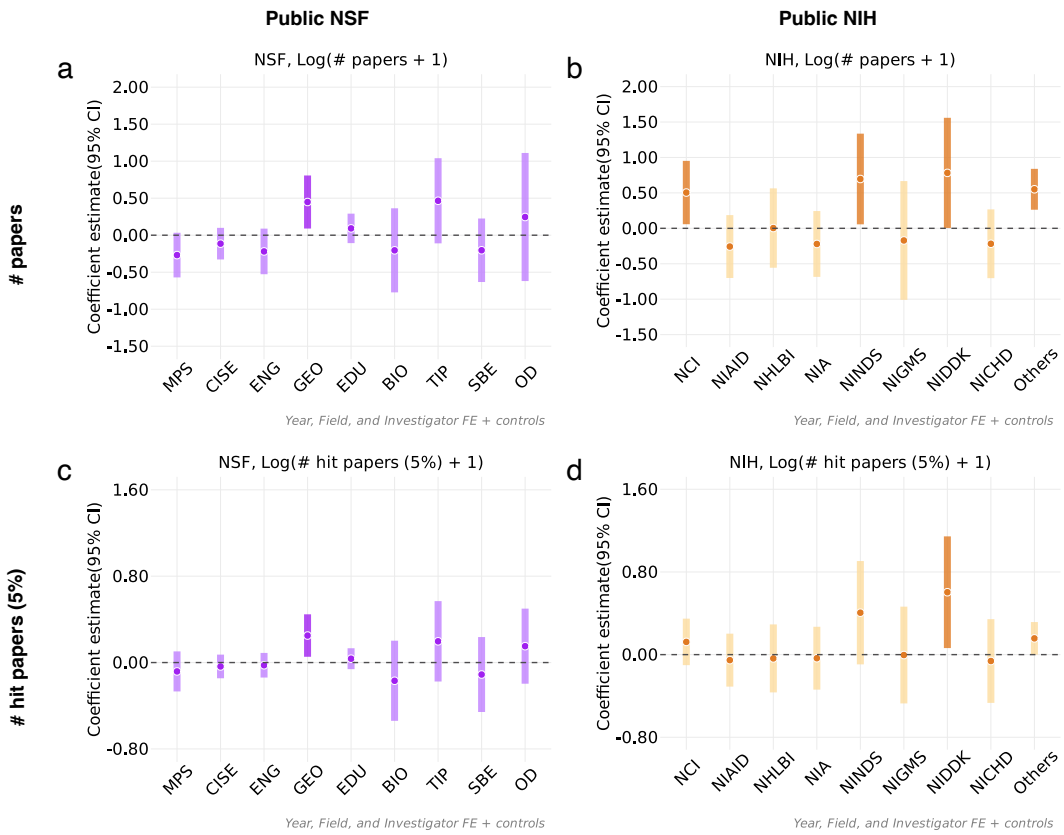


Fig. S19: **LLM use and federal research funding outputs by subagency.** (**a-b**) Regression estimates relating grant-level LLM use (α) to the total number of resulting publications for NSF and NIH grants. (**c-d**) Corresponding estimates for high-impact outputs, where a "hit" paper is defined as one whose citations fall within the top 5% of all papers published worldwide in the same year and field. All regressions include grant start year, field, and investigator fixed effects, as well as controls for funding amount. Points indicate coefficient estimates, and bars denote 95% confidence intervals.

Supplementary Tables

Table S1: **LLM use and semantic distinctiveness (prior 1 year) in private NSF data.** Columns progressively add fixed effects for start year, scientific field, and investigators.

	Private NSF		
	Semantic distinctiveness (prior 1 year)		
	Model 1	Model 2	Model 3
α	-16.456***	-11.2727**	-11.9472**
	(4.455)	(3.9927)	(4.3752)
log(funding)	-0.4146	-0.3281	-1.7646*
	(0.9348)	(0.9539)	(0.8016)
start year	No	Yes	Yes
field	No	Yes	Yes
investigator	No	No	Yes
N. of Observations	1001	1412	1886
R^2	0.01886	0.11690	0.72441

Table S2: **LLM use and semantic distinctiveness (prior 1 year) in private NIH data.** Columns progressively add fixed effects for start year, scientific field, and investigators.

	Private NIH		
	Semantic distinctiveness (prior 1 year)		
	Model 1	Model 2	Model 3
α	-7.974*	-10.6152**	-23.3958***
	(3.7419)	(3.6906)	(3.1939)
log(funding)	0.0972	0.2624	-0.4638
	(0.4898)	(0.4666)	(0.4191)
start year	No	Yes	Yes
field	No	Yes	Yes
investigator	No	No	Yes
N. of Observations	2532	3686	8198
R^2	0.00252	0.06034	0.53894

Table S3: **LLM use and semantic distinctiveness (prior 1 year) in public NSF data.** Columns progressively add fixed effects for start year, scientific field, and investigators.

	Public NSF		
	Semantic distinctiveness (prior 1 year)		
	Model 1	Model 2	Model 3
α	-24.539***	-18.6069***	-18.0163***
	(0.871)	(0.8992)	(1.2958)
log(funding)	1.2935***	1.4454***	0.7485***
	(0.1517)	(0.1487)	(0.184)
start year	No	Yes	Yes
field	No	Yes	Yes
investigator	No	No	Yes
N. of Observations	32278	46504	60283
R^2	0.02805	0.07757	0.78893

Table S4: **LLM use and semantic distinctiveness (prior 1 year) in public NIH data.** Columns progressively add fixed effects for start year, scientific field, and investigators.

	Public NIH		
	Semantic distinctiveness (prior 1 year)		
	Model 1	Model 2	Model 3
α	-11.093***	-17.6547***	-23.6247***
	(1.1549)	(1.1611)	(1.7726)
log(funding)	0.9335***	0.7318***	0.7554**
	(0.2288)	(0.2048)	(0.2471)
start year	No	Yes	Yes
field	No	Yes	Yes
investigator	No	No	Yes
N. of Observations	36617	54033	56413
R^2	0.00507	0.10213	0.86886

Table S5: **LLM use and proposal success in private NSF data.** Columns progressively add fixed effects for start year, scientific field, and investigators.

	Private NSF		
	Proposal success		
	Model 1	Model 2	Model 3
α	0.0323	-0.0086	0.0227
	(0.0674)	(0.0728)	(0.0936)
log(funding)	-0.0728**	-0.0846***	-0.0292
	(0.0225)	(0.0214)	(0.0205)
start year	No	Yes	Yes
field	No	Yes	Yes
investigator	No	No	Yes
N. of Observations	816	1145	1464
R^2	0.03969	0.06128	0.56523

Table S6: **LLM use and proposal success in private NIH data.** Columns progressively add fixed effects for start year, scientific field, and investigators.

	Private NIH		
	Proposal success		
	Model 1	Model 2	Model 3
α	0.039	0.0599	0.1906***
	(0.0575)	(0.0595)	(0.0496)
log(funding)	-0.0522***	-0.0502***	-0.0443***
	(0.0091)	(0.0086)	(0.0096)
start year	No	Yes	Yes
field	No	Yes	Yes
investigator	No	No	Yes
N. of Observations	1783	2608	5632
R^2	0.03124	0.04564	0.37319

Table S7: **LLM use and downstream output (# papers) in public NSF data.** Columns progressively add fixed effects for start year, scientific field, and investigators.

	Public NSF		
	Log(# papers + 1)		
	Model 1	Model 2	Model 3
α	-0.0981***	0.1357***	-0.0232
	(0.0294)	(0.0287)	(0.0471)
log(funding)	0.2024***	0.1984***	0.2636***
	(0.0046)	(0.0044)	(0.0075)
start year	No	Yes	Yes
field	No	Yes	Yes
investigator	No	No	Yes
N. of Observations	23741	34193	43415
R^2	0.08056	0.20724	0.84442

Table S8: **LLM use and downstream output (# papers) in public NIH data.** Columns progressively add fixed effects for start year, scientific field, and investigators.

	Public NIH		
	Log(# papers + 1)		
	Model 1	Model 2	Model 3
α	-0.2395***	-0.0369	0.2739***
	(0.0397)	(0.0406)	(0.0676)
log(funding)	0.1965***	0.1749***	0.2957***
	(0.0047)	(0.0046)	(0.0105)
start year	No	Yes	Yes
field	No	Yes	Yes
investigator	No	No	Yes
N. of Observations	30548	45185	46577
R^2	0.07682	0.12805	0.86251

Table S9: **LLM use and downstream output (# top 5% hit papers) in public NSF data.** Columns progressively add fixed effects for start year, scientific field, and investigators.

	Public NSF		
	Log(# hit papers(5%) + 1)		
	Model 1	Model 2	Model 3
α	-0.0202	0.0514***	0.015
	(0.0124)	(0.0127)	(0.0243)
log(funding)	0.0537***	0.053***	0.074***
	(0.0025)	(0.0025)	(0.0043)
start year	No	Yes	Yes
field	No	Yes	Yes
investigator	No	No	Yes
N. of Observations	23741	34193	43415
R^2	0.03012	0.07571	0.77834

Table S10: **LLM use and downstream output (# top 5% hit papers) in public NIH data.** Columns progressively add fixed effects for start year, scientific field, and investigators.

	Public NIH		
	Log(# hit papers(5%) + 1)		
	Model 1	Model 2	Model 3
α	-0.0848***	-0.0088	0.0612
	(0.0188)	(0.0194)	(0.0403)
log(funding)	0.0833***	0.0756***	0.1161***
	(0.0026)	(0.0025)	(0.0061)
start year	No	Yes	Yes
field	No	Yes	Yes
investigator	No	No	Yes
N. of Observations	30548	45185	46577
R^2	0.04699	0.07676	0.83363

Table S11: **LLM use and downstream output (# top 1% hit papers) in public NSF data.** Columns progressively add fixed effects for start year, scientific field, and investigators.

	Public NSF		
	Log(# hit papers(1%) + 1)		
	Model 1	Model 2	Model 3
α	-0.0048	0.0195**	0.0087
	(0.0063)	(0.0064)	(0.0129)
log(funding)	0.02***	0.0197***	0.0275***
	(0.0015)	(0.0015)	(0.0025)
start year	No	Yes	Yes
field	No	Yes	Yes
investigator	No	No	Yes
N. of Observations	23741	34193	43415
R^2	0.01641	0.03586	0.73823

Table S12: **LLM use and downstream output (# top 1% hit papers) in public NIH data.** Columns progressively add fixed effects for start year, scientific field, and investigators.

	Public NIH		
	Log(# hit papers(1%) + 1)		
	Model 1	Model 2	Model 3
α	-0.0307**	-0.0039	-0.0221
	(0.0107)	(0.0112)	(0.0262)
log(funding)	0.0342***	0.031***	0.0435***
	(0.0015)	(0.0015)	(0.0036)
start year	No	Yes	Yes
field	No	Yes	Yes
investigator	No	No	Yes
N. of Observations	30548	45185	46577
R^2	0.02381	0.03764	0.81908

References

[1] Bush, V. Science, the endless frontier: A report to the president on a program for postwar scientific research (1945).

[2] Stephan, P. *How Economics Shapes Science* (Harvard University Press, 2012).

[3] Li, D., Azoulay, P. & Sampat, B. N. The applied value of public investments in biomedical research. *Science* **356**, 78–81 (2017).

[4] Galkina Cleary, E., Beierlein, J. M., Khanuja, N. S., McNamee, L. M. & Ledley, F. D. Contribution of NIH funding to new drug approvals 2010–2016. *Proceedings of the National Academy of Sciences* **115**, 2329–2334 (2018).

[5] Azoulay, P., Graff Zivin, J. S., Li, D. & Sampat, B. N. Public R&D investments and private-sector patenting: Evidence from NIH funding rules. *The Review of Economic Studies* **86**, 117–152 (2019).

[6] Fleming, L., Greene, H., Li, G., Marx, M. & Yao, D. Government-funded research increasingly fuels innovation. *Science* **364**, 1139–1141 (2019).

[7] Yin, Y., Dong, Y., Wang, K., Wang, D. & Jones, B. F. Public use and public funding of science. *Nature Human Behaviour* **6**, 1344–1350 (2022).

[8] Azoulay, P., Clancy, M., Li, D. & Sampat, B. N. What if NIH had been 40% smaller? *Science* **389**, 1303–1305 (2025).

[9] Furnas, A. C., Fishman, N., Rosenstiel, L. & Wang, D. Partisan disparities in the funding of science in the united states. *Science* **389**, 1195–1200 (2025).

[10] Wang, Y. *et al.* Funding the frontier: Visualizing the broad impact of science and science funding. *arXiv preprint arXiv:2509.16323* (2025).

[11] Yin, Y., Wang, Y., Evans, J. A. & Wang, D. Quantifying the dynamics of failure across science, startups and security. *Nature* **575**, 190–194 (2019).

[12] Wang, Y., Jones, B. F. & Wang, D. Early-career setback and future career impact. *Nature Communications* **10**, 4331 (2019).

[13] Li, D. & Agha, L. Big names or big ideas: Do peer-review panels select the best science proposals? *Science* **348**, 434–438 (2015).

[14] Peng, H., Qiu, H. S., Fosse, H. B. & Uzzi, B. Promotional language and the adoption of innovative ideas in science. *Proceedings of the National Academy of Sciences* **121**, e2320066121 (2024).

[15] Liang, W. *et al.* Quantifying large language model usage in scientific papers. *Nature Human Behaviour* 1–11 (2025).

[16] Kusumegi, K. *et al.* Scientific production in the era of large language models. *Science* **390**, 1240–1243 (2025).

[17] Liang, W. *et al.* Monitoring AI-modified content at scale: A case study on the impact of chatgpt on AI conference peer reviews. In *International Conference on Machine Learning (ICML)* (2024).

[18] Liang, W. *et al.* The widespread adoption of large language model-assisted writing across society. *Patterns* (2025).

[19] Kobak, D., González-Márquez, R., Horvát, E.-Á. & Lause, J. Delving into LLM-assisted writing in biomedical publications through excess vocabulary. *Science Advances* **11**, eadt3813 (2025).

[20] Liu, J., He, Y., Zheng, Z., Bu, Y. & Ni, C. AI-assisted writing is growing fastest among non-english-speaking and less established scientists. *arXiv preprint arXiv:2511.15872* (2025).

[21] Bao, H., Sun, M. & Teplitskiy, M. Where there's a will there's a way: Chatgpt is used more for science in countries where it is prohibited. *Quantitative Science Studies* 1–16 (2025).

[22] Wang, H. *et al.* Scientific discovery in the age of artificial intelligence. *Nature* **620**, 47–60 (2023).

[23] Gao, J. & Wang, D. Quantifying the use and potential benefits of artificial intelligence in scientific research. *Nature Human Behaviour* **8**, 2281–2292 (2024).

[24] Jabarian, B. & Imas, A. Artificial writing and automated detection. *National Bureau of Economic Research* (2025).

[25] Swanson, K., Wu, W., Bulaong, N. L., Pak, J. E. & Zou, J. The virtual lab of AI agents designs new SARS-CoV-2 nanobodies. *Nature* **646**, 716–723 (2025).

[26] Hao, Q., Xu, F., Li, Y. & Evans, J. Artificial intelligence tools expand scientists' impact but contract science's focus. *Nature* 1–7 (2026).

[27] Jones, B. F. The burden of knowledge and the "death of the renaissance man": Is innovation getting harder? *The Review of Economic Studies* **76**, 283–317 (2009).

[28] Hill, R. *et al.* The pivot penalty in research. *Nature* 1–8 (2025).

[29] Liu, L., Dehmamy, N., Chown, J., Giles, C. L. & Wang, D. Understanding the onset of hot streaks across artistic, cultural, and scientific careers. *Nature Communications* **12**, 5392 (2021).

[30] Tripodi, G. *et al.* Tenure and research trajectories. *Proceedings of the National Academy of Sciences* **122**, e2500322122 (2025).

[31] Shao, E. *et al.* Sciscigpt: advancing human–AI collaboration in the science of science. *Nature Computational Science* 1–15 (2025).

[32] Bail, C. A. Can generative AI improve social science? *Proceedings of the National Academy of Sciences* **121**, e2314021121 (2024).

[33] Musslick, S. *et al.* Automating the practice of science: Opportunities, challenges, and implications. *Proceedings of the National Academy of Sciences* **122**, e2401238121 (2025).

[34] Doshi, A. R. & Hauser, O. P. Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances* **10**, eadn5290 (2024).

[35] March, J. G. Exploration and exploitation in organizational learning. *Organization Science* **2**, 71–87 (1991).

[36] Jones, B. Artificial intelligence in research and development. *National Bureau of Economic Research* (2025).

[37] Scharfmann, E., Marx, M. & Fleming, L. Pasteur's quadrant researchers bring novelty, impact to publishing, and patenting. *Science* **390**, 891–893 (2025).

[38] Singh, A., D'Arcy, M., Cohan, A., Downey, D. & Feldman, S. Scirepeval: A multi-format benchmark for scientific document representations. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, 5548–5566 (2023).

[39] Herzog, C., Hook, D. & Konkiel, S. Dimensions: Bringing down barriers between scientometricians and data. *Quantitative Science Studies* **1**, 387–395 (2020).

[40] Singh, A., D'Arcy, M., Cohan, A., Downey, D. & Feldman, S. Scirepeval: A multi-format benchmark for scientific document representations. In *Conference on Empirical Methods in Natural Language Processing* (2022).

[41] Flesch, R. A new readability yardstick. *Journal of Applied Psychology* **32**, 221 (1948).

[42] Millar, N., Batalo, B. & Budgell, B. Trends in the use of promotional language (hype) in abstracts of successful national institutes of health grant applications, 1985-2020. *JAMA Network Open* **5**, e2228676–e2228676 (2022).