NBER WORKING PAPER SERIES

HOW HEAVILY ARE PHD STUDENTS OFFLOADING THEIR SCIENTIFIC WRITING TO AI? EVIDENCE FROM US STEM PHD DISSERTATIONS

Daniel P. Gross
Dror Shvadron
Hansen Zhang

Working Paper 35859
http://www.nber.org/papers/w35859

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2026

We are grateful to Lee Fleming for his collaboration in related research, Kevin Bryan for helpful comments, and Ben Goihman for outstanding research assistance. We thank Pangram for API credits and discounted pricing, and the Duke University Fuqua School of Business, Alfred P. Sloan Foundation (Grant No. G-2023-21064), and National Science Foundation (Grant No. 2420824) for financial support. Claude Fable 5.1 was used in the initial drafting of the discussion of the empirical findings, and for a subset of the data appendix, though all AI-generated text has been substantially manually revised. All errors are our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2026 by Daniel P. Gross, Dror Shvadron, and Hansen Zhang. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

How Heavily Are PhD Students Offloading Their Scientific Writing to AI? Evidence from US STEM PhD Dissertations

Daniel P. Gross, Dror Shvadron, and Hansen Zhang

NBER Working Paper No. 35859
October 2026
JEL No. I23, J24, O31, O33

ABSTRACT

Large language models are diffusing rapidly into science, raising concerns that delegating research writing to AI may undermine the development of expertise that scientific training is supposed to build. We measure AI-generated writing in the dissertations of a near-census of US STEM PhD graduates from 2019 to 2026 and link graduates to their subsequent employment. AI writing is absent before 2023, present in 29% of dissertations filed in 2026, and rapidly growing. Use is more common among students from non-English-speaking countries and at lower-ranked programs. Among graduates in the same program and cohort, those whose dissertations contain AI writing are less likely to enter academic research careers and more likely to enter non-tenure-track roles or industry, with this pattern strongest among US-origin students. At the same time as AI may increase research productivity, it thus appears to also be displacing a potentially important portion of the work through which scientific expertise is developed.

Daniel P. Gross
Duke University
Fuqua School of Business
and NBER
daniel.gross@duke.edu

Hansen Zhang
Duke University
Fuqua School of Business
hansen.zhang@duke.edu

Dror Shvadron
University of Toronto
Rotman School of Management
dror.shvadron@rotman.utoronto.ca

Introduction

Artificial intelligence (AI) technologies are rapidly diffusing across science and society (Chatterji et al. 2025, Handa et al. 2025) and increasingly reshaping how science is done (Gao and Wang 2024, Hao et al. 2026, Codreanu et al. 2026). Although AI takes many forms that are being used for different scientific ends—from predicting protein structures (Jumper et al. 2021, Varadi et al. 2022) to resolving mathematical conjectures (Alon et al. 2026)—the most common (and increasingly ubiquitous) manifestation of AI in science is the use of large language models (LLMs), as a growing share of research papers are conceived, executed, and partially or fully written by LLMs (see, e.g., Boiko et al. 2023, Liang et al. 2025, Kobak et al. 2025, Kusumegi et al. 2025, Holzwarth et al. 2026, Lu et al. 2026, Siler 2026, Gartenberg et al. 2026, among others).

Enthusiasm about the potential for AI to accelerate the pace of discovery is substantial (e.g., Cockburn et al. 2019, Bianchini et al. 2022, Wang et al. 2023, Amodei 2024, Gil and Moler 2025, Dolgin 2026). However, as LLM use in science spreads, concerns are mounting over the pressures it puts on the scientific enterprise. Much of this concern centers around research evaluation and quality control vis-à-vis peer review (e.g., Liang et al. 2024, Naddaf 2025, Gibney 2026, Thorp 2026a,b, Nature 2026). Other observers have questioned the future of the research paper as a scientific unit (e.g., Mollick 2024, Cowen 2026a,b, Nature 2026), and some have even begun to reconsider who or what research is for if far more knowledge is being produced than can be consumed (Jones 2009, Chu and Evans 2021, Messeri and Crockett 2024, Liu et al. 2026).

A more fundamental threat, however, may be to scientific expertise and human-embodied knowledge (Acemoglu et al. 2026). This is particularly the case for current and future doctoral students, who represent the next generation of scientists. Much of an individual scientist's research ability is tacit and learned through experience (Ma et al. 2020), with the core of this knowledge cultivated in graduate school, where students have traditionally developed skills and taste through slow, arduous, hands-on work pitted with failures and continual iteration. Like research technologies of the past (e.g., computers), AI makes it possible to operate at a higher level of abstraction than previously possible, potentially cutting out much of this work and putting at risk the expertise this experience generates (Bastani et al. 2025, Strömberg et al. 2026, Autor et al. 2026). LLMs' cost and speed make them an especially attractive substitute to this effort. Intensifying competition at conferences and journals—itself a symptom of AI—merely reinforces these incentives, creating a feedback loop that only further encourages using AI to stay competitive on output and other evaluative metrics on which academic scientists' careers are built.

This paper examines how heavily AI is being used by PhD students, who is using it, and how it relates to students' professional outcomes—a potential proxy for new scientists' quality of training and depth of expertise. Following recent literature, we focus on AI-generated writing, evaluating the AI content of PhD dissertations in a nearly-complete sample of US STEM PhD graduates since 2019, and comparing career trajectories of students who were more versus less reliant on AI in their dissertation work. We argue that AI-generated writing is a particularly important endpoint for evaluating training effort, as the writing stage of research is where ideas must be stated precisely and in turn where failures in logic and method are often discovered and worked out. The experience of consolidating and communicating research may thus be one way in which students become more effective critical evaluators and improve their understanding of their own work. Delegating this effort to AI risks increasing short-term productivity at the expense of early career scientists' skill acquisition, scientific capabilities, and long-run research capacity.

Methods and Data

To evaluate these questions, we collected data on US-trained PhD graduates since 2019, applied AI detection tools to dissertation abstracts to assess AI use, and linked individuals to their LinkedIn profiles to evaluate post-PhD outcomes. While a small number of prior articles have assessed LLM writing in dissertation research in small samples—typically specific universities or fields, often assessed through surveys and interviews (e.g., Hoomanfard and Shamsi 2025, Akbar 2025, Sarofian-Butin 2026)—we undertake this analysis at a national scale.

Our sample consists of STEM PhD graduates from US universities whose dissertations were filed in ProQuest Dissertations & Theses Global (PQDT) between 2019 and May 2026, which we extend by collecting data from university repositories for universities not represented in PQDT over this period. The ≈226,000 graduates in this combined sample align closely with totals in the National Science Foundation's Survey of Earned Doctorates (an annual census of US PhD graduates) by year, institution, and field, with approximately complete coverage of US PhD graduates in the life, physical, and mathematical sciences and 85% coverage of engineering. We evaluate the AI content of dissertations using Pangram 3.3.2 (Emi and Spero 2024), a validated commercial AI detection service which is increasingly used in research for AI detection (see, e.g., Jabarian and Imas 2025 for validation, and Gartenberg et al. 2026 for an application). We process all abstracts and a stratified random sample of complete dissertations (stratifying over fields and years), which we use to verify that AI-generated abstracts are a surrogate for AI-generated writing in the full text—consistent with prior research which has found the same (e.g., Gartenberg et al. 2026).

We identify graduates' post-PhD employment outcomes by linking them to LinkedIn profiles (via Revelio Labs data), leveraging a linking infrastructure developed in prior work (Shvadron et al. 2025). We also measure high school and baccalaureate country for a subset of graduates, which we use as a proxy for likely country of origin and to identify potential non-native English-speakers. Using metadata on the graduates in our sample from PQDT and repositories, we also identify degree characteristics such as institution and degree subjects, which we group up to major fields. We use data from the Integrated Postsecondary Education Data System (IPEDS) to measure university characteristics and from National Research Council (2011) to measure individual program rankings, which we merge into our graduate sample on university and field.

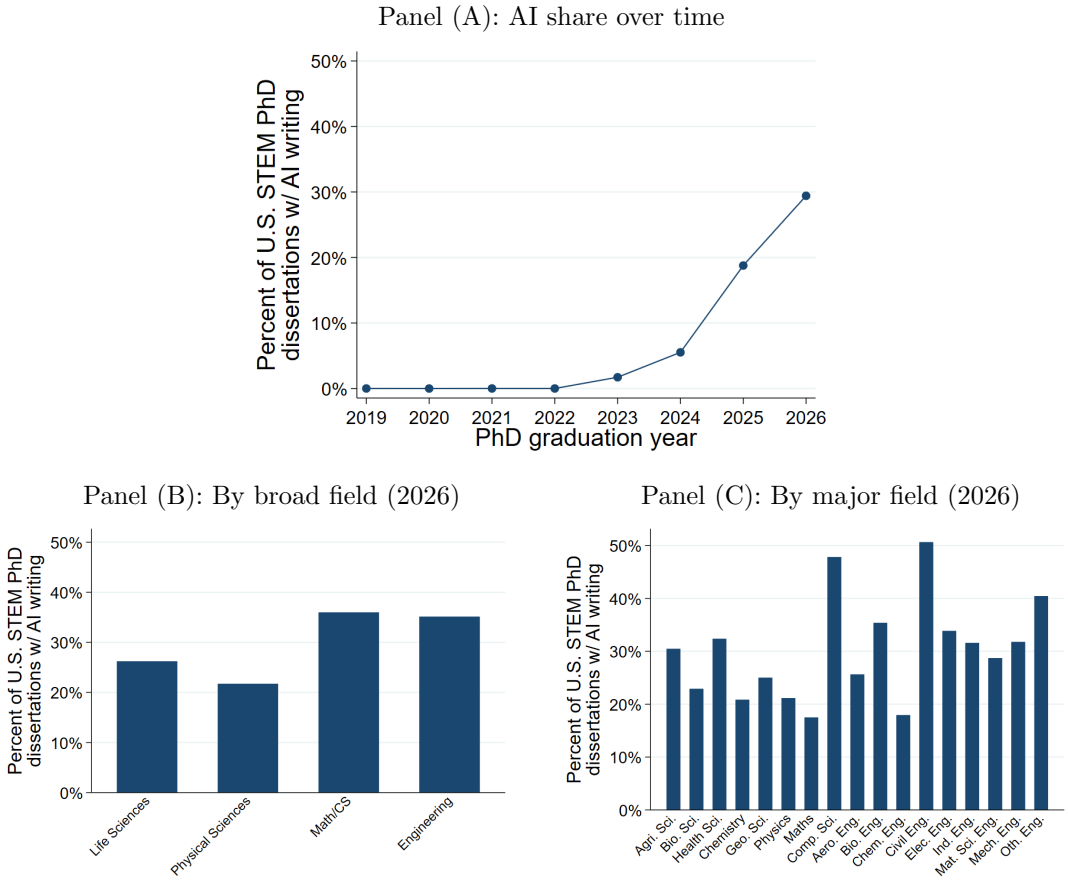
Nearly 30% of dissertations have AI-generated content

Figure 1 reports the share of US STEM dissertations with AI-generated abstracts, by year of degree (Panel A) and by field (Panels B and C, for 2026 graduates only). Through 2022, the AI share is zero: across roughly 126,000 dissertations filed between 2019 and 2022, Pangram flags a total of six abstracts as AI-generated. Because these cohorts composed their dissertations before LLMs were publicly available, these pre-2023 values provide a validation test of the detection method and indicate a false positive rate of less than 0.01%. AI writing first appears among 2023 graduates (1.7% of dissertations), then grows to 5.5% in 2024, 18.8% in 2025, and 29.4% in the 2026 cohort, where our data include dissertations filed through May 2026.

Benchmarked against recent research on AI use in scientific writing (which primarily examines peer-reviewed journal articles and pre-prints), students' PhD research appears to be lagging wider scientific activity in AI usage by 1-2 years: Liang et al. (2025) and Kobak et al. (2025) previously estimated that 13-22% of 2024 abstracts in computer science and biomedicine (respectively) were LLM-modified, whereas our corresponding estimate for 2024 dissertations is ≈5%. Reweighting our measures of AI use in abstracts (across the full sample) to predict use in full text (using conditional probabilities of Pr(AI in full text | AI in abstract) in our full text subsample) yields only modest changes, but suggests these estimates should be viewed as lower bounds: we project that ≈8% of dissertations were likely to have AI writing in 2024 and 31% in 2026.

Panels B and C show that adoption is uneven across fields. In the 2026 cohort, AI-generated abstracts are most common in the mathematical and computer sciences (36%) and engineering (35%), followed by the life sciences (26%) and physical sciences (22%). These averages mask a wider spread across specific fields, from 17% in mathematics and statistics and 18% in chemical engineering to 48% in computer science and 51% in civil engineering, with other engineering disciplines, health

Figure 1: How heavily is AI being used in dissertation research?



Notes: Panel (A) shows the annual share of dissertations whose abstracts were AI generated, according to Pangram detection. The evaluated AI share was 0 through 2022 but has been rapidly accelerating since then, and is nearly 30% in our 2026 sample (covering dissertations filed with ProQuest through May 2026). Panels (B) and (C) show the share of dissertations in our 2026 sample with AI abstracts, by broad field and major field.

sciences, and agricultural sciences in between (29-40%). The fields with the most intensive AI use in dissertation research are those in which LLM adoption in published research has also been fastest (Liang et al. 2025, Kousha and Thelwall 2026); those at the bottom (e.g., chemistry and physics) are fields whose abstracts are more likely to include specific notation, compounds, and quantities. The same rank-order holds among 2025 graduates (engineering 24%, math/CS 24%, life sciences 16%, physical sciences 14%), suggesting that this heterogeneity reflects field-specific differences in the pace of AI adoption rather than its likely eventual reach.

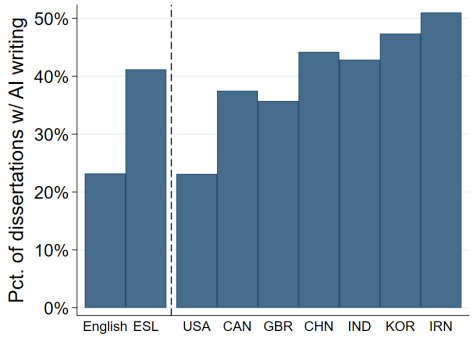
Dissertation research remains human generated at the top

Use of AI in dissertation research writing varies across students and institutions. Figure 2 further disaggregates AI use by likely country of origin and native language, the research intensity of their

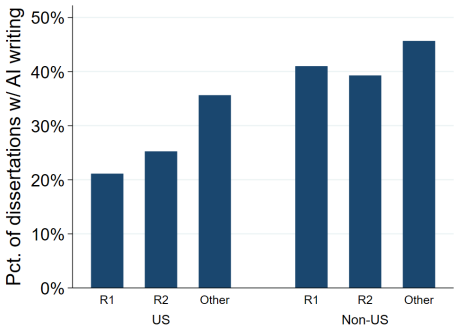
PhD institution, and the National Research Council rank of their doctoral program. We proxy for origin country with the country of a graduate's high school or baccalaureate degree, observed in linked LinkedIn education histories for roughly 60% of the sample.

Figure 2: Which PhD students are using AI to write their research?

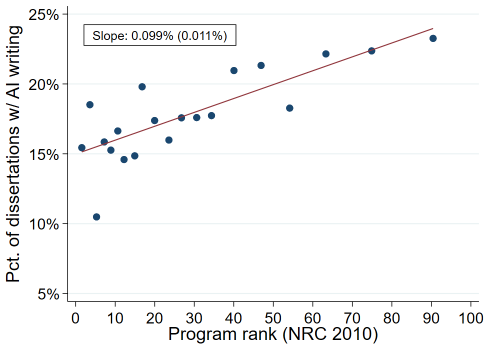
Panel (A): AI use by language and country



Panel (B): AI use by Carnegie Class



Panel (C): AI use by program rank



Notes: Panel (A) shows graduates' tendency to use AI to draft dissertation research by whether they are from an English-speaking country, and by country for select high-volume countries. Panel (B) shows this tendency by PhD institution quality (measured by its 2021 Carnegie Classification), separately for US and foreign origin students. Panel (C) shows this tendency by program rank (measured using National Research Council 2010 rankings; see appendix for details). Sample in Panels (A) and (B) consists of 2026 dissertations available through May 2026; sample in Panel (C) consists of 2025-2026 dissertations. English as a second language students and students at lower research intensity institutions and lower-ranked programs are significantly more likely to substitute AI writing.

Non-native English-speaking students use AI far more intensively than native speakers.[1] Among 2026 graduates, 41% of dissertations by students educated in non-English-speaking countries have AI-generated abstracts, against 23% for students from English-speaking countries (Panel A). The gap is present for every major sending country: 44% for students from China, 43% from India, 49%

[1]Note that although early research on use of LLMs (especially ChatGPT) in scientific writing suggested that AI classifiers are more likely to flag non-native, human-generated English prose as machine-generated (Liang et al. 2023), this does not appear to be a concern for modern detectors (like Pangram) or in our context. Among 2019-2022 graduates (pre-LLMs), abstracts by students from non-English-speaking countries are flagged as AI-generated at a 0% rate (identical to those by English-speaking peers), on samples of roughly 18,000 graduates per year. The gap we observe emerges only after 2022 and grows with adoption.

from South Korea, and 51% from Iran, versus 23% for those with US lower degrees—a benchmark which already is potentially biased upwards by the inclusion of foreign-born students with US baccalaureates. The gap has also widened each year as adoption spreads: non-English-speaking students' rate of AI writing was 2.8% in 2023 against 0.9% for English-speaking students, 8.3% against 3.7% in 2024, and 28% against 14% in 2025. In regressions with university-field fixed effects, which effectively compare students in the same doctoral program, the difference is 1.5, 3.7, and 11.7 percentage points (p.p.) in the three periods (respectively), and the gap remains largest for students from Iran (23 p.p. in 2025-26), South Korea (15 p.p.), China (13 p.p.), and India (11 p.p.), whereas students from Canada and the UK do not differ from US-origin students in AI use. These patterns are consistent with evidence from studies on published research, where LLM-assisted writing is most prevalent among authors in non-English-speaking countries (Kobak et al. 2025, Liu et al. 2025, Kusumegi et al. 2025), and it is consistent with the presumption of LLMs being most valuable to students for whom English writing is most costly (Amano et al. 2023).

AI use also increases at institutions with lower research intensity and programs of lower rank. In the 2026 cohort, 28% of dissertations at R1 universities have AI writing, compared with 31% at R2 universities and 37% at other doctoral institutions, and this gradient is particularly strong among US-origin students (foreign-origin students write with AI at high rates at all institutions; Panel B). Within R1 universities, the relationship with program quality is roughly monotone: among 2025-26 graduates of programs ranked in the top ten of their field, 15% have AI-generated abstracts, rising to 22% for programs ranked 40-50 and 26% for programs ranked 90-100, a gradient of about one p.p. per ten rank positions (Panel C). Roughly a third of this gradient reflects the composition of top programs, which enroll more international students; conditioning on origin, the slope falls from 0.10 to 0.07 p.p. per rank position but remains precisely estimated.

These differences are stark at the extremes. Among R1 universities with at least 50 graduates since 2025, the AI share is 9% at MIT, 10% at Caltech, and 13% at Princeton, versus 30%+ at the ten universities with the highest rates (Appendix Table B.1). Several top-ranked programs—e.g., mathematics at Berkeley, bioengineering at Northwestern, chemistry at Ohio State—have no AI-generated abstracts at all, while the highest university-field cells exceed 50% (Appendix Table B.2). AI use also clusters within programs and advisors: a 10 p.p. higher rate of AI writing among the previous year's graduates of the same doctoral program is associated with a 3 p.p. higher probability that a student's own writing is AI-generated—patterns which may reflect peer effects, common access, shared norms, or other shared features. Similar correlations hold among students with the

same primary advisor or committee chair (Appendix Table B.3).

Together these patterns indicate that AI writing has diffused fastest where the pressures it relieves are greatest—for students writing in a second language and at lower-ranked programs which are likely to have less prepared students—and slowest at the programs that train the largest share of future faculty (Clauset et al. 2015, Wapman et al. 2022). At the top of the training hierarchy, dissertation research remains, at least for now, largely human-written.

Graduates with AI-written dissertations tend to enter industry

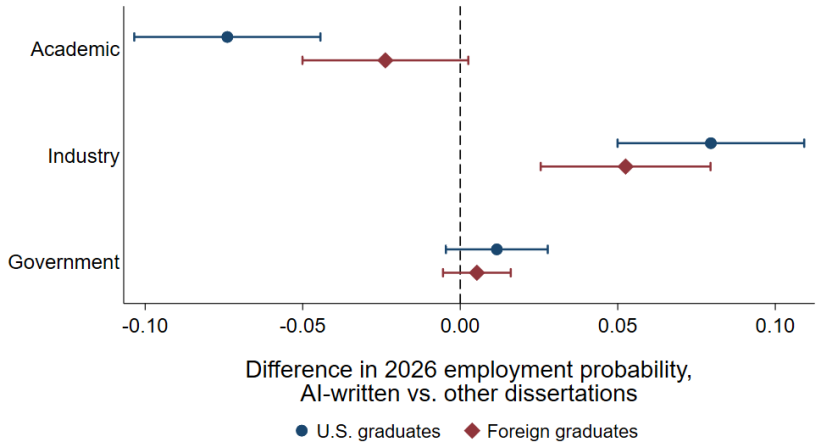
Assessing training quality, and how it varies with AI, remains difficult. One potential surrogate endpoint is a graduate's post-PhD employment outcome. Using data on graduates' LinkedIn employment histories, we evaluate differences in the current (2026) employment of 2023-2025 PhD graduates of R1 universities, allowing at least half a year after graduation for LinkedIn profiles to update, and compare graduates with AI-written dissertations to others of the same university and field. Because AI use is concentrated among foreign-origin students, who on average also differ in their post-PhD employment opportunities, we estimate the relationship separately for US- and foreign-origin graduates.[2] The results are presented in Figure 3.

Among US-origin graduates, those with AI-written dissertations are 7.4 p.p. less likely to be employed in academia in 2026 (17% over the 44% base rate) and 8.0 p.p. more likely to be employed in industry (16% over a 50% base), with no difference in government employment. These differences attenuate substantially for likely foreign-origin graduates. We also find differences within academia: conditional on academic employment, graduates with AI-written dissertations are 5.4 p.p. more likely to hold non-tenure-track positions (on a base of 12%), with no significant difference in the likelihood of tenure-track or postdoctoral positions. AI writing thus appears disproportionately common among graduates who exit an academic research career.

Importantly, this analysis is correlational (not causal), and it admits multiple interpretations. Interpreted causally, it suggests that graduates who write with AI may have a training deficit that leaves them unprepared for a research career. On the other hand, a plausible interpretation is reverse causal: late-stage students who do not seek a research career may write lower-effort dissertations. If so, we might expect other markers of low effort, including shorter or more poorly written dissertations.

[2]Specifically, we regress each employment outcome on an indicator for an AI-generated abstract, an indicator for a foreign lower degree, and their interaction, with university-field and graduation year fixed effects and standard errors clustered by university-field. The figure plots the marginal effect of AI writing for each origin group. Sectors are not mutually exclusive: a graduate who changed jobs during 2026 can appear in two, and "academic" employment includes universities, teaching hospitals, and federally funded research centers.

Figure 3: Differences in employment patterns among graduates with AI-written dissertations



Notes: Figure shows the differential tendency for graduates with AI-written dissertations to be employed in academia, industry, or government in 2026. Differences estimated over graduates who obtained their PhD at R1 universities between 2023 and 2025 (allowing time for LinkedIn profiles to update post-graduation). Estimates obtained from a regression of the employment outcome on indicators for (i) AI writing, (ii) foreign lower degree (a proxy for non-US origin), and (iii) their interaction. All regressions include university-field and year fixed effects, with standard errors clustered by university-field; results are similar in magnitude and precision with university-field-year (i.e., program cohort) fixed effects. Error bars represent 95% confidence intervals.

Empirically, however, dissertation length (measured in pages) does not correlate with AI use, and the employment differences in Figure 3 persist controlling for length. We thus far view the evidence as suggesting an adverse impact of AI use on research capability.

There are two reasons why the differences in Figure 3 may be conservative. First, our abstract-based AI detection captures only one part of an AI-written dissertation (Appendix Tables A.1 and A.2), and this measurement error may bias magnitudes downwards. The true relationship between full-text AI use and employment outcomes may be materially larger. Second, the sample is restricted to R1 graduates, who comprise a large majority of our sample but among whom both (i) research careers are more common and (ii) AI use is comparatively lower.

Discussion

Taken together, these patterns suggest the US scientific training system is further stratifying on dimensions which predict AI use: students at lower-tier institutions, and non-native English speakers, are disproportionately delegating to AI the stage of research at which understanding is consolidated. If expertise is built through experience, then the distribution of AI use we document suggests a growing disparity in expertise across these categories.

A natural counterargument, however, is that the benefits of AI may flow in the other direction, narrowing performance gaps. Non-native English speakers face large, well-documented costs to writing science in English (Amano et al. 2023), and the experimental literature finds that generative AI helps more disadvantaged writers most (Noy and Zhang 2023, Brynjolfsson et al. 2025), narrowing gaps in readability and citations for scientists at non-English institutions (Kaniel et al. 2025, Lin et al. 2025). It may also be the case that both patterns are true: AI may close gaps in writing quality or comprehension while also removing the impetus for the careful thinking which writing requires. The evidence in this paper is consistent with both, as PhD graduates with AI writing in their dissertations are less likely to remain in academic research, but this difference is attenuated for those ostensibly from non-native English-speaking countries.

Though the evidence in this paper is suggestive of a cost, the endpoints we study ultimately cannot fully resolve whether PhD students' use of AI in scientific writing is in fact degrading their preparation for a research career. Observers who see the act of writing as forcing a precise understanding are likely to believe so, while those who view scientific communication as an instrumentality fully separable from research itself may disagree. Empirically, it is likely too soon to know for sure: indications of degraded research capability—such as evidence that these students are less likely to publish independently, or win competitive research grants—will take more time to accumulate, and causal evaluative designs will also be needed to close the question.

More generally, we are still in the early days of AI, and much may yet change. Adoption is clearly growing, and at its current growth rate, a majority of US STEM dissertations will have AI-generated content by 2028. AI technology is also rapidly improving, and institutions are evolving with them, albeit more slowly. What is clear to us, however, is that even as AI holds the potential to unlock new discoveries and dramatically accelerate progress, it is changing the returns to human capital investment and challenging the case for developing deep scientific expertise through slow, careful scholarship (Gartenberg et al. 2026). Economic theory suggests sustaining students' incentive to do the hard work themselves will require either changes in how scientists are evaluated and what is professionally rewarded, or verification technologies to measure and enforce human-generated effort—or both. Unlike peer-reviewed journals, however, doctoral degrees already have a verification institution in the form of a live defense, whose importance as a test of expertise and understanding—and as a condition for degree conferral—may grow. Ultimately, whether the next generation of scientists is trained or merely credentialed remains to be seen. The answer will reveal itself over time, but research in the interim may offer leading indications.

References

Acemoglu, Daron, Dingwen Kong, and Asuman Ozdaglar. 2026. "AI, Human Cognition and Knowledge Collapse." NBER Working Paper 34910, National Bureau of Economic Research.

Akbar, Muhammad Naveed. 2025. "Use of artificial intelligence tools by doctoral students: a mixed-methods explanatory-sequential investigation." *Journal of Further and Higher Education* **49**(7):995-1013.

Alon, Noga, Thomas F. Bloom, W. Timothy Gowers, Daniel Litt, Will Sawin, Arul Shankar, Jacob Tsimerman, Victor Wang, and Melanie Matchett Wood. 2026. "Remarks on the disproof of the unit distance conjecture." *arXiv preprint arXiv:2605.20695* .

Amano, Tatsuya, Valeria Ramírez-Castañeda, Violeta Berdejo-Espinola, Israel Borokini, Shawan Chowdhury, Marina Golivets, Juan David González-Trujillo, Flavia Montaño-Centellas, Kumar Paudel, Rachel Lynn White, and Diogo Veríssimo. 2023. "The manifold costs of being a non-native English speaker in science." *PLOS Biology* **21**(7):e3002184.

Amodei, Dario. 2024. "Machines of Loving Grace: How AI Could Transform the World for the Better." Essay, https://darioamodei.com/essay/machines-of-loving-grace.

Autor, David, Tanya Rodchenko, Josh Martin, Zanna Iscenko, Scott Strand, David Pearl, and Melissa Ferere. 2026. "Does AI Assistance Enhance or Erode Expertise? Evidence from a Three-Month Field Experiment in Patent Drafting." NBER Working Paper 35720, National Bureau of Economic Research.

Bastani, Hamsa, Osbert Bastani, Alp Sungu, Haosen Ge, Özge Kabakcı, and Rei Mariman. 2025. "Generative AI without guardrails can harm learning: Evidence from high school mathematics." *Proceedings of the National Academy of Sciences* **122**(26):e2422633122.

Bianchini, Stefano, Moritz Müller, and Pierre Pelletier. 2022. "Artificial intelligence in science: An emerging general method of invention." *Research Policy* **51**(10):104604.

Boiko, Daniil A., Robert MacKnight, Ben Kline, and Gabe Gomes. 2023. "Autonomous chemical research with large language models." *Nature* **624**:570-578.

Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond. 2025. "Generative AI at Work." *Quarterly Journal of Economics* **140**(2):889-942.

Chatterji, Aaron, Thomas Cunningham, David J. Deming, Zoë Hitzig, Christopher Ong, Carl Yan Shan, and Kevin Wadman. 2025. "How People Use ChatGPT." NBER Working Paper 34255, National Bureau of Economic Research.

Chu, Johan SG and James A Evans. 2021. "Slowed canonical progress in large fields of science." *Proceedings of the National Academy of Sciences* **118**(41):e2021636118.

Clauset, Aaron, Samuel Arbesman, and Daniel B. Larremore. 2015. "Systematic inequality and hierarchy in faculty hiring networks." *Science Advances* **1**(1):e1400005.

Cockburn, Iain M., Rebecca Henderson, and Scott Stern. 2019. "The Impact of Artificial Intelligence on Innovation: An Exploratory Analysis." In *The Economics of Artificial Intelligence: An Agenda*, edited by Ajay Agrawal, Joshua Gans, and Avi Goldfarb, pp. 115-146. University of Chicago Press.

Codreanu, Mihai, Alex Imas, Juan Mateos-Garcia, Joseph Emmens, Evalyne Muiruri, Arthur Turrell, Julian Jacobs, Atoosa Kasirzadeh, Ana Trišović, Yiyuan Chen, Tanya Rodchenko, Catherine Pollard, Scott Strand, Daniel Rock, Zanna Iscenko, Fabien Curto Millet, Neil Thompson, and James Manyika. 2026. "AI in Science: Early Insights." Working Paper 7465898, SSRN.

Cowen, Tyler. 2026a. "When will "the research paper" disappear in economics?" Marginal Revolution, https://marginalrevolution.com/marginalrevolution/2026/03/when-will-the-research-paper-disappear-in-economics.html.

Cowen, Tyler. 2026b. "Will AI kill the research paper?" Marginal Revolution, https://marginalrevolution.com/marginalrevolution/2026/05/will-ai-kill-the-research-paper.html.

Dolgin, Elie. 2026. "AI co-scientists are revolutionizing how research is done." *Nature* **657**:108-110.

Emi, Bradley and Max Spero. 2024. "Technical Report on the Pangram AI-Generated Text Classifier." *arXiv preprint arXiv:2402.14873* .

Gao, Jian and Dashun Wang. 2024. "Quantifying the use and potential benefits of artificial intelligence in scientific research." *Nature Human Behaviour* **8**:2281-2292.

Gartenberg, Claudine, Sharique Hasan, Alex Murray, and Lamar Pierce. 2026. "More Versus Better: Artificial Intelligence, Incentives, and the Emerging Crisis in Peer Review." *Organization Science* **37**(3):795-812.

Gibney, Elizabeth. 2026. "How AI slop is causing a crisis in computer science." *Nature* .

Gil, Darío and Kathryn A. Moler. 2025. "Accelerating science with AI." *Science* **390**(6777):965.

Handa, Kunal, Alex Tamkin, Miles McCain, Saffron Huang, et al. 2025. "Which Economic Tasks are Performed with AI? Evidence from Millions of Claude Conversations." *arXiv preprint arXiv:2503.04761* .

Hao, Qianyue, Fengli Xu, Yong Li, and James Evans. 2026. "Artificial intelligence tools expand scientists' impact but contract science's focus." *Nature* **649**:1237-1243.

Holzwarth, Lena, Rita González-Márquez, and Dmitry Kobak. 2026. "Most biomedical publications show signs of LLM-assisted writing." *arXiv preprint arXiv:2608.10715* .

Hoomanfard, MohammadHamed and Yaser Shamsi. 2025. "Generative AI in dissertation writing: L2 doctoral students' self-reported use, AI-giarism, and perceived training needs." *Journal of English for Academic Purposes* **78**.

Jabarian, Brian and Alex Imas. 2025. "Artificial Writing and Automated Detection." NBER Working Paper 34223, National Bureau of Economic Research.

Jones, Benjamin F. 2009. "The Burden of Knowledge and the "Death of the Renaissance Man": Is innovation getting harder?" *The Review of Economic Studies* **76**(1):283-317.

Jumper, John, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. 2021. "Highly accurate protein structure prediction with AlphaFold." *Nature* **596**:583-589.

Kaniel, Ron, Huaxia Rui, Shujing Sun, and Pingle Wang. 2025. "Writing Matters: Generative AI as an Academic Impact Equalizer." Working Paper 5361925, SSRN.

Kobak, Dmitry, Rita González-Márquez, Emőke-Ágnes Horvát, and Jan Lause. 2025. "Delving into LLM-assisted writing in biomedical publications through excess vocabulary." *Science Advances* **11**(27):eadt3813.

Kousha, Kayvan and Mike Thelwall. 2026. "How much are LLMs changing the language of academic papers after ChatGPT? A multi-database and full text analysis." *Scientometrics* .

Kusumegi, Keigo, Xinyu Yang, Paul Ginsparg, Mathijs de Vaan, Toby Stuart, and Yian Yin. 2025. "Scientific production in the era of large language models." *Science* **390**(6779):1240-1243.

Liang, Weixin, Zachary Izzo, Yaohui Zhang, Haley Lepp, Hancheng Cao, Xuandong Zhao, Lingjiao Chen, Haotian Ye, Sheng Liu, Zhi Huang, Daniel A. McFarland, and James Y. Zou. 2024. "Monitoring AI-Modified Content at Scale: A Case Study on the Impact of ChatGPT on AI Conference Peer Reviews." In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 29575-29620.

Liang, Weixin, Mert Yuksekgonul, Yining Mao, Eric Wu, and James Zou. 2023. "GPT detectors are biased against non-native English writers." *Patterns* **4**(7):100779.

Liang, Weixin, Yaohui Zhang, Zhengxuan Wu, Haley Lepp, Wenlong Ji, Xuandong Zhao, Hancheng Cao, Sheng Liu, Siyu He, Zhi Huang, Diyi Yang, Christopher Potts, Christopher D. Manning, and James Y. Zou. 2025. "Quantifying large language model usage in scientific papers." *Nature Human Behaviour* **9**:2599-2609.

Lin, Dingkang, Naixuan Zhao, Dan Tian, and Jiang Li. 2025. "Large language models reshape the language of science." *arXiv preprint arXiv:2504.12317* .

Liu, Jialin, Yongyuan He, Zhihan Zheng, Yi Bu, and Chaoqun Ni. 2025. "AI-Assisted Writing Is Growing Fastest Among Non-English-Speaking and Less Established Scientists." *arXiv preprint arXiv:2511.15872* .

Liu, Jiachen, Jiaxin Pei, Jintao Huang, Chenglei Si, Ao Qu, Xiangru Tang, Runyu Lu, Lichang Chen, Xiaoyan Bai, Haizhong Zheng, Carl Chen, Zhiyang Chen, Haojie Ye, Yujuan Fu, Zexue He, Zijian Jin, Zhenyu Zhang, Shangquan Sun, Maestro Harmon, John Dianzhuo Wang, Jianqiao Zeng, Jiachen Sun, Mingyuan Wu, Baoyu Zhou, Chenyu You, Shijian Lu, Yiming Qiu, Fan Lai, Yuan Yuan, Yao Li, Junyuan Hong, Ruihao Zhu, Beidi Chen, Alex Pentland, Ang Chen, Mosharaf Chowdhury, and Zechen Zhang. 2026. "The Last Human-Written Paper: Agent-Native Research Artifacts." *arXiv preprint arXiv:2604.24658* .

Lu, Chris, Cong Lu, Robert Tjarko Lange, Yutaro Yamada, Shengran Hu, Jakob Foerster, David Ha, and Jeff Clune. 2026. "Towards end-to-end automation of AI research." *Nature* **651**:914-919.

Ma, Yifang, Satyam Mukherjee, and Brian Uzzi. 2020. "Mentorship and protégé success in STEM fields." *Proceedings of the National Academy of Sciences* **117**(25):14077-14083.

Messeri, Lisa and M. J. Crockett. 2024. "Artificial intelligence and illusions of understanding in scientific research." *Nature* **627**:49-58.

Mollick, Ethan. 2024. "Four Singularities for Research." One Useful Thing, https://www.oneusefulthing.org/p/four-singularities-for-research.

Naddaf, Miryam. 2025. "AI is transforming peer review—and many scientists are worried." *Nature* **639**(8056):852-854.

National Research Council. 2011. *A Data-Based Assessment of Research-Doctorate Programs in the United States*. Washington: The National Academies Press.

Nature. 2026. "AI scientists are changing research—institutions, funders and publishers must respond." *Nature* Editorial, 25 March 2026.

Noy, Shakked and Whitney Zhang. 2023. "Experimental evidence on the productivity effects of generative artificial intelligence." *Science* **381**(6654):187-192.

Sarofian-Butin, Dan. 2026. "Detecting AI in Doctoral Dissertations: An Exploratory Empirical Analysis of EdD Dissertation Integrity and Implications for Doctoral Program Design." Working Paper 6772501, SSRN.

Shvadron, Dror, Hansen Zhang, Lee Fleming, and Daniel P Gross. 2025. "A Quarter of US-Trained Scientists Eventually Leave. Is the US Giving Away Its Edge?" *arXiv preprint arXiv:2512.11146* .

Siler, Kyle. 2026. "The diffusion of large language models in published academic articles." *Proceedings of the National Academy of Sciences* **123**(22):e2605754123.

Strömberg, David, Victor Lei, and Yanhui Wu. 2026. "The Generative AI Learning Penalty: Evidence from Chinese Secondary Education." CEPR Discussion Paper 21577, Centre for Economic Policy Research.

Thorp, H. Holden. 2026a. "AI in scientific publishing: Slower, worse, and more expensive." *Science* **393**(6808):225.

Thorp, H. Holden. 2026b. "Resisting AI slop." *Science* **391**(6780):5.

Varadi, Mihaly, Stephen Anyango, Mandar Deshpande, Sreenath Nair, Cindy Natassia, Galabina Yordanova, David Yuan, Oana Stroe, Gemma Wood, Agata 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.

Wang, Hanchen, Tianfan Fu, Yuanqi Du, Wenhao Gao, Kexin Huang, Ziming Liu, Payal Chandak, Shengchao Liu, Peter Van Katwyk, Andreea Deac, et al. 2023. "Scientific discovery in the age of artificial intelligence." *Nature* **620**:47-60.

Wapman, K. Hunter, Sam Zhang, Aaron Clauset, and Daniel B. Larremore. 2022. "Quantifying hierarchy and dynamics in US faculty hiring and retention." *Nature* **610**(7930):120-127.

Online Appendix

How heavily are PhD students offloading their scientific writing to AI? Evidence from US STEM PhD dissertations

Daniel P. Gross, Dror Shvadron, and Hansen Zhang

Table of Contents

A Data Appendix **2**

A.1 Base sample: ProQuest dissertations + university repositories 2

A.2 Measuring professional outcomes 8

B Supplementary Results **13**

B.1 AI use by university and program 13

B.2 Peer AI use as a predictor of individual use 14

Appendix References **15**

A Data Appendix

This appendix describes our data sources, measurement, and analytical methods. Note that portions of this appendix—particularly those documenting the ProQuest sample and our approach to linking graduates to their post-PhD employment—overlap with the supplementary material to Shvadron et al. (2025a) and Shvadron et al. (2025b), where we use a partially overlapping sample and partially overlapping measures to study different questions.

A.1 Base sample: ProQuest dissertations + university repositories

The sample for this paper consists of U.S. STEM (Science, Technology, Engineering, and Mathematics) PhD graduates who received their degrees between January 2019 and May 2026 (the time of data collection), providing us roughly four graduate cohorts prior to the public release of ChatGPT (in November 2022) and four post-ChatGPT cohorts.

Our core data source for this sample is ProQuest Dissertations & Theses Global (PQDT), a commercial database of higher degree graduates and their theses, from which we obtain graduate metadata.[1] At the time of acquisition, the PQDT database contained over six million global master's theses and doctoral dissertations, including in the natural sciences, engineering, humanities, social sciences, and other domains. PQDT data report several characteristics of the graduates in its sample, including their name, degree year, degree name, university, dissertation characteristics (title, abstract, and subjects), advisor(s), and more. Following our own prior work, we filter the PQDT corpus to U.S. STEM PhD graduates as follows. We first removed (i) non-U.S. dissertations and (ii) non-doctoral theses and dissertations. We also manually reviewed degree names to identify and restrict to those corresponding to research degrees in the natural sciences and engineering.[2] We then crosswalk these graduates' self-reported subjects (as provided in the PQDT data) to "major fields" defined by the U.S. National Center for Science and Engineering Statistics (NCSES) and filter our sample to dissertations in 17 major fields which are typically considered STEM subjects.[3] We extend the PQDT sample by retrieving analogous data from digital thesis repositories of universities which are not represented in PQDT in our target years, for as many such universities as we could crawl online, compiling a similar record (name, degree year and name, dissertation characteristics, advisors) and filtering and mapping subjects in analogous ways.

[1] ProQuest is the leading distributor of U.S. dissertations, which it re-publishes under license, and its data have been used in numerous prior studies of scientific trainees (e.g., Toole and Czarnitzki 2010, Bikard et al. 2015, Buffington et al. 2016, Jiang et al. 2023, Antman et al. 2023, Arora et al. 2026, among others).

[2] Though most doctoral degrees are PhDs, PQDT also includes other doctorates such as DScs or field-specific doctorates (e.g., D.CS. (Doctor of Computer Science) or Chem. D. (Doctor of Chemistry)), as well as certain doctoral professional degrees (e.g., PsyD (Doctor of Psychology) or D.N.P. (Doctor of Nursing Practice)).

[3] Our focal major fields are: Agricultural sciences and natural resources, Biological and biomedical sciences, and Health sciences (belonging to the broad field "Life Sciences"); Physics and astronomy, Geosciences, atmospheric sciences, and ocean sciences, and Chemistry (belonging to "Physical Sciences"); Computer and information sciences and Mathematics and statistics (belonging to "Mathematical and Computer Sciences"); and Aerospace, aeronautical, and astronautical engineering, Bioengineering and biomedical engineering, Chemical engineering, Civil engineering, Electrical and electronics engineering, Industrial and manufacturing engineering, Materials science engineering, Mechanical engineering, and Other engineering (belonging to "Engineering").

A.1.1 Additional person characteristics

Following the linking procedures described in the next section, we link graduates in this sample to LinkedIn profiles, using data from Revelio Labs and matching on name and PhD degree details. We use these links to obtain individuals' complete education histories, identify reported high school and baccalaureate degrees, and retrieve the country of each degree. We in turn use these lower degree countries to measure whether a graduate is likely to be from a foreign (non-US) country, and to categorize those countries into native English speaking countries (USA, Canada, UK, Australia, New Zealand, Ireland) versus others. Though imperfect, this approach is conservative, prioritizing precision over recall: in a U.S. PhD student sample, non-US high school and baccalaureate degrees are overwhelmingly obtained in students' native countries (such that when observed, we can precisely and confidently infer the origin country), but we may fail to measure as foreign those PhD students who completed their baccalaureate degree in the U.S. (reducing recall). This mismeasurement will tend to attenuate English/non-English and US/non-US comparisons, which implies any such differences are likely a lower bound.

A.1.2 Additional program characteristics

We manually crosswalk degree institutions to Integrated Postsecondary Education Data System (IPEDS) UnitIDs, the institutional identifiers which SED, IPEDS, and other U.S. higher education datasets are identified on. We use IPEDS to retrieve institutions' 2021 Carnegie Classification, grouping them into R1 ("Doctoral Universities: Very High Research Activity"), R2 ("Doctoral Universities: High Research Activity"), and others (a catch-all which consolidates professional doctoral universities, PhD-granting research institutes, and any others).

We supplement these data with measures of university and program rank. We obtain general university rankings from U.S. News & World Report (where available; excludes specialty institutions) and doctoral program rankings from the National Research Council (2011) publication "A Data-Based Assessment of Research-Doctorate Programs in the United States" and accompanying rankings tables.[4] This project produced rankings of over 5,000 doctoral programs in 62 fields at 212 U.S. universities, based on a 2005-2006 program characteristics. The data include both survey-based rankings and statistical rankings with field-specific estimated weights on different program characteristics, with each ranking bootstrapped and the accompanying files reporting the bottom 5% and top 95% rank of each program across bootstrap iterations. We calculate the average regression-based rank for each program (averaging between its 5th and 95th percentile bootstrap ranks) and use this as our measure of program quality. To link the NRC data to our sample, we crosswalk universities to IPEDS, NRC fields to our canonical major fields, and link on university-field; where a university has multiple programs in the same major field, we measure its rank as the best rank reported among any of the associated programs.

[4]See https://www.nationalacademies.org/projects/BHEW-Q-03-02-A/publication/12994 for the report and data.

A.1.3 Completeness/representativeness of sample

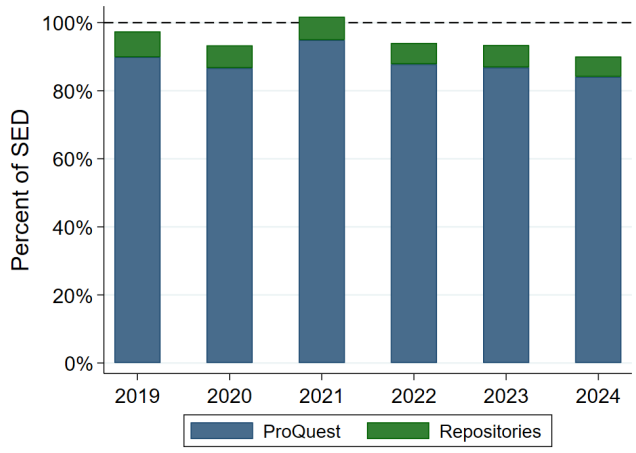
To evaluate the completeness and representativeness of our data, we compare counts of PQDT + Repository (henceforth, "PQDT+") graduates to counts of U.S. doctoral graduates from the Survey of Earned Doctorates (SED), an annual census of research doctorate recipients from U.S. universities which has been administered by NSF since 1958, with >90% completion rates. Although modest differences in counts may arise due to differences in field assignment—SED similarly assigns graduates to fields based on self-reported subjects, but we do not have access to their mapping technique—it provides an administrative benchmark for our sample.

We make several comparisons between PQDT+ and SED, including on annual counts, totals by broad field for our study period, and correlations at different units of aggregation. Figure A.1 shows that our PQDT+ sample represents 90-100% of SED STEM graduates in every year for which we have data (the most recently-reported SED year is 2024), with PQDT comprising 85-95% of the SED population and repository data filling in another ≈5%. That the PQDT+ sample slightly exceeds SED in one year (2021) can be attributed to one of three possibilities: that (i) PQDT+ was more complete than SED this year (as SED response rates are very high but less than 100%), (ii) differences in timing (PQDT+ reports when dissertations were made available, SED when students graduate), or (iii) modest deviations in how we map subjects to fields versus how SED performs this mapping which can lead to modest sample differences.

Figure A.2 evaluates completeness by broad field, and indicates that the PQDT+ sample returns approximately complete sampling of life science, physical science, and mathematical/computer science graduates, and very high (85%) sampling of engineering graduates; that Math/CS slightly exceeds 100% is potentially attributable to the same reasons, as well as fungibility in the coding of some subjects to computer science versus electrical engineering.

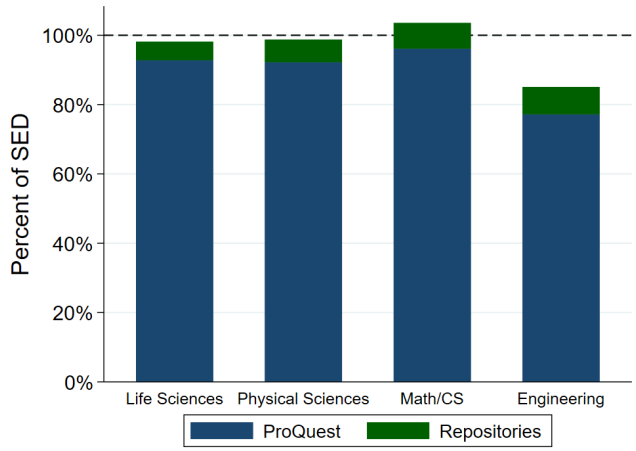
Figure A.3 provides binned scatterplots of SED vs. PQDT+ graduates at the university level, university-year level, field-year level, and university-field-year level (Panels A to D, respectively), including bivariate regression coefficients. At all levels of aggregation, PQDT+ and SED counts closely track each other and cluster around the 45-degree line.

Figure A.1: Completeness of sample, by year (benchmark: SED)



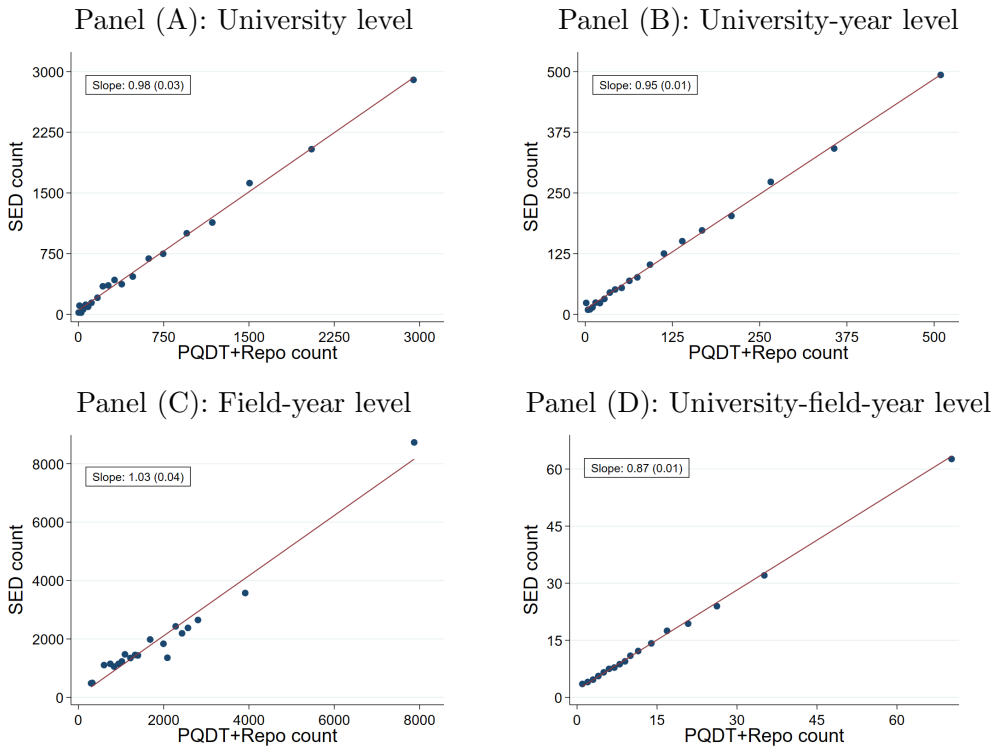
Notes: Figure evaluates the completeness of our sample across years, using the SED as a benchmark. Each bar reports the number of U.S. STEM PhD graduates in our PQDT sample (blue) and university repository backfill (green) relative to the SED total. Graduates in each of these data sources (PQDT, repositories, SED) are associated to fields as described in text. SED data currently available through 2024.

Figure A.2: Completeness of sample, by broad field (benchmark: SED)



Notes: Figure evaluates the completeness of our sample across broad fields, using the SED as a benchmark. Each bar reports the number of U.S. STEM PhD graduates in our PQDT sample (blue) and university repository backfill (green) relative to the SED total. Graduates in each of these data sources (PQDT, repositories, SED) are associated to fields as described in text. SED data currently available through 2024.

Figure A.3: Correlation of PQDT+ vs. SED graduate counts at varying units of analysis



Notes: Figure shows a binned scatterplot of PQDT + Repository vs. SED graduate counts in our sample at each of four levels: university (Panel A), university-year (Panel B), field-year (Panel C), university-field-year (Panel D). Graduates are associated to fields as described in text. Sample covers STEM graduates of universities in the continental U.S. between 2019 and 2024 (the last year of available SED data).

A.1.4 Measuring dissertation contents

We retrieve dissertation abstracts for all of the 226,911 graduates in our sample from PQDT and university repositories, and full dissertation text for a subsample of PQDT-listed graduates, randomly sampling 5 dissertations per major field (17 fields) per year between 2023 and 2026 (4 years). We then use Pangram (validated commercial AI detection service which is used frequently in similar research) to assess the AI content of these inputs, applying the Pangram v3.3.2 model through its bulk API, with discounted research pricing. Around 300 dissertations did not have abstracts or could not be scored by Pangram, leaving us with 226,591 abstracts and 340 randomly sampled full text dissertations that we assess for AI writing.

Table A.1 uses the full text sample to validate the abstract-level measures. We estimate a simple bivariate regression of full text AI use against abstract AI use:

$$\text{Full text AI}_i = \beta \cdot \text{Abstract AI}_i + \varepsilon_i \tag{A.1}$$

We evaluate this relationship at the individual level (Columns 1 to 3), then at more aggregated levels: university-year (Columns 4 to 5) and field-year (Columns 6 to 7). The table shows that a

dissertation with an AI-generated abstract is 55 percentage points (p.p.) more likely to have AI detected in the full text (quadrupling the base rate), has 11% more AI content (nearly quintupling the base rate), and is 20 p.p. more likely to have at least 20% of its text be AI-written (6x the base rate). Similar patterns emerge when evaluated at higher levels of aggregation.

In Table A.2 we switch the left- and right-hand side variables, such that the parameters indicate the likelihood of an AI abstract given a certain amount of AI content in the full text (conveying the same correlation as Table A.1 in different form). Accordingly, the same patterns emerge but now can be interpreted as follows. A dissertation with any AI in its text is 56% more likely to have an AI-generated abstract; a dissertation with 10 p.p. more of its content AI-generated is 25 p.p. more likely to have an AI-generated abstract, and a dissertation where >20% of the full text is AI written has over 90% probability of having an AI-generated abstract.

Table A.1: Abstract AI writing predicts full dissertation AI writing

	Individual dissertations			Univ-year dissertations		Field-year dissertations	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Has AI	Pct. AI	>20% AI	Pct. w/ AI	Mean AI	Pct. w/ AI	Mean AI
Has AI in abstract	0.556***	0.111***	0.204***				
	(0.068)	(0.021)	(0.056)				
Pct. w/ AI in abstract				0.513***	0.118***	0.709***	0.129***
				(0.089)	(0.031)	(0.085)	(0.026)
N	339	339	339	215	215	68	68
R^2	0.31	0.28	0.16	0.28	0.25	0.60	0.54
Y mean	0.153	0.023	0.035	0.156	0.025	0.153	0.023

Notes: Table shows that (i) AI writing in a dissertation's abstract significantly predicts whether it has AI writing in the full dissertation, and (ii) the AI-generated share of the abstract predicts the AI-generated share of the full dissertation. The sample for this analysis is a stratified random sample of 340 dissertations filed between 2023-2026, with 5 per major field and year, for which we retrieved the full text and evaluated its AI content using Pangram, as we did for all abstracts (note: one dissertation is associated with an institution we could not match to IPEDS and excluded from the estimation). The table presents analysis at multiple levels: at the individual level (Columns 1-3), university-year level (Columns 4-5), and field-year level (Columns 6-7). Outcomes are as follows: Column (1), whether a dissertation has AI in the full text; Column (2), the share of the full text that is AI-generated; Column (3), whether >20% of a dissertation's text is AI-generated; Columns (4) and (6), the share of dissertations in the cell that have AI in the full text; Columns (5) and (7), the average AI share of full text among dissertations in the cell. *, **, *** represent significance at the 0.1, 0.05, and 0.01 levels, respectively. Heteroskedasticity-robust SEs in parentheses.

Table A.2: Full text AI writing predicts abstract AI writing

	Individual abstracts			Univ-year abstracts		Field-year abstracts	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Has AI	Has AI	Has AI	Pct. w/ AI	Pct. w/ AI	Pct. w/ AI	Pct. w/ AI
Has AI in full text	0.565***						
	(0.069)						
AI pct. of full text		2.514***					
		(0.339)					
>20% AI in full text			0.788***				
			(0.082)				
Pct. w/ AI in full text				0.552***		0.853***	
				(0.090)		(0.081)	
Mean AI pct. of full text					2.141***		4.155***
					(0.299)		(0.479)
N	339	339	339	215	215	68	68
R^2	0.31	0.28	0.16	0.28	0.25	0.60	0.54
Y mean	0.156	0.156	0.156	0.164	0.164	0.156	0.156

Notes: Table shows that (i) AI writing in a dissertation's full text significantly predicts whether it has AI writing in the abstract, and (ii) the AI-generated share of the full text predicts an AI-generated abstract. The sample for this analysis is a stratified random sample of 340 dissertations filed between 2023-2026, with 5 per major field and year, for which we retrieved the full text and evaluated its AI content using Pangram, as we did for all abstracts (note: one dissertation is associated with an institution we could not match to IPEDS and excluded from the estimation). The table presents analysis at multiple levels: at the individual level (Columns 1-3), university-year level (Columns 4-5), and field-year level (Columns 6-7). The outcome at the individual level is an indicator for whether an abstract is AI-composed; at the group level, the share of abstracts that are AI composed. *, **, *** represent significance at the 0.1, 0.05, and 0.01 levels, respectively. Heteroskedasticity-robust SEs in parentheses.

A.2 Measuring professional outcomes

We link graduates in our sample to career histories assembled by Revelio Labs from public professional profiles, principally LinkedIn, to observe their post-PhD employment. Revelio provides us what is effectively a mirror of the structured relational database behind LinkedIn profiles, periodically re-scraping the website to retrieve new users and profile updates. Revelio standardizes each profile in its database into a person record, a set of education records (institution, degree, field, and dates), and a set of position records (employer, title, start and end dates, and location), and resolves employers to a company identifier with associated metadata (web domain, industry code, parent company, and, for listed firms, securities identifiers). We access these data under license through Wharton Research Data Services, in the June 25, 2026 release.

A.2.1 Linking dissertations to career profiles

The dissertation data provide several pieces of fixed identifying information which can be used to link to professional profiles—chief among them are name and education history (degrees, institutions, and years). We use this information to make an initial set of candidate links between our sample and the Revelio data. This procedure begins with the manual construction of a crosswalk mapping PhD institutions in Revelio and PQDT to common identifiers (IPEDS UnitIDs), and then filtering Revelio profiles to those reporting a doctoral degree (PhD or equivalent) at an institution also present in PQDT. Doing so yields 2,762,178 candidate profiles for the 1,375,136 dissertations in the PQDT base sample, prior to matching them on name.

We then split names in both sources into given, middle, and family names. Because raw name strings can have inconsistent order, punctuation, or inclusion of middle names and credentials, this processing is performed with batched LLM review followed by manual correction (10,345 corrections on the PQDT side/5,902 on the Revelio side), with the same normalization applied to both sides.[5] A candidate pair is formed when the school matches, the degree year is consistent with the profile's education dates, and the names agree under one of seventeen matching rules. These range from exact agreement of all components, through structural rules (agreement on initials, middle-name subsequences, and token-boundary shifts such as a hyphenated given name split into two), to reviewed equivalence tables for nicknames and for common romanization variants of Chinese family names (five pairs, such as Lee/Li and Chang/Zhang), and finally to bounded fuzzy similarity. This process yields 642,699 candidate dissertation-profile pairs.

Every candidate pair receives an explicit accept or reject decision. Pairs matched by a deterministic rule are reviewed once; pairs matched by a fuzzy or initials-only rule pass through repeated language-model name reviews under prompts of varying strictness, with disagreements escalated. The 620,798 pairs that survive name review are then subjected to a single automatic check of contextual consistency (degree field, year, and institutional context). This check functions as a veto for a small set of documented contradictions rather than as a similarity threshold. Through this check we reject 1,384 pairs (0.22%), leaving 619,414 accepted pairs.

Because pairs are reviewed independently, a dissertation can be accepted to multiple profiles and a profile to multiple dissertations. Of the accepted pairs, 546,229 are clean one-to-one assignments. The remaining 73,185 form 34,815 connected conflict clusters, which we classify into nine shapes—most commonly one dissertation linked to two profiles that appear to be duplicates of the same person (11,389 clusters), or one dissertation contested by two distinct users (15,221)—and adjudicate cluster by cluster on profile, education, and position evidence, assigning each dissertation to at most one profile. The final match assigns 584,178 dissertations (42.5% of the PQDT base) to 581,318 unique Revelio users. Among the 2019-2026 graduates studied in this paper, 159,297 of 226,911 (70.2%) are linked to a Revelio profile. We use the linked profiles for two purposes. The first, described above, is to recover PQDT graduates' lower-degree institutions and countries as a proxy for country of origin. The second is to measure employment in 2026.

A.2.2 Employment in 2026

For every linked graduate with a PhD year of 2019 or later, we extract all positions held during calendar 2026, defined as positions that began on or before December 31, 2026 and either ended on or after January 1, 2026 or have no recorded end date.[6] We treat graduates who are linked to a profile but have no reported 2026 position as having uncertain employment, rather than as unemployed, and retain these graduates so that selection into the outcome sample can be examined

[5] Name parsing and name review use OpenAI models through the Batch API, with ambiguous cases escalated to a stronger model and, where needed, to manual review. All prompts, requests, and responses have been retained so that the workflow can be reproduced offline without further model calls.

[6] On LinkedIn, open-ended positions are common, and 95.7% of the positions in our extract have no end date.

directly. The extract contains 191,639 positions in 2026 for 145,043 graduates, and 14,254 linked graduates with no 2026 position. Our Revelio data run through June 23, 2026, such that "2026 employment" refers to positions held in the first half of the year.

We next classify each position in this sample along two dimensions: (i) the sector of the employer, and (ii) for academic employers, the type/track of the position.

Employer sector. Positions are grouped by employer, using Revelio's company identifier where present (94.6% of positions) and a normalized employer name otherwise, yielding 44,971 distinct employers. Each employer is assigned to one of eleven sectors: university; hospital or academic medical center; federally funded research and development center (FFRDC) or national laboratory; independent research institute; industry; government; non-profit; professional society; K–12 education; self-employed, independent, or founder; and unclassifiable. Classification is performed by a large language model ensemble in which two models independently classify every employer and a third adjudicates the cases on which they disagree, given the employer's name, web domain, industry (NAICS) code, parent company, and self-description as evidence.[7] The two primary models agree on 93.6% of employers, leaving 2,867 to adjudication. Against a blind hand-labeled sample of 50 employers, the ensemble's final label matches in 49 cases.

Academic position type. For positions held by PhDs at universities, hospitals (typically academic medical centers), and FFRDCs (namely, national labs)—which we collectively consider quasi-academic employment—we classify the raw job title into eight roles: tenure-track faculty, non-tenure-track faculty, postdoctoral researcher, research staff, clinical, administrative, graduate student or doctoral trainee, and other. We classify titles using the same LLM ensemble, providing it distinct (title, sector) pairs (23,453 pairs), with instructions to pay particular attention to the modifiers that indicate a tenure-track split (e.g., "Assistant Professor" is tenure-track, whereas "Research Assistant Professor," "Teaching Assistant Professor," "Clinical Assistant Professor," and "Assistant Professor of the Practice" are not). Against a blind hand-labeled sample of 50 titles, the ensemble is correct in 44 cases; of the six disagreements, five concern distinctions among administrative, research staff, and other roles that do not enter the outcomes we study.

Stale doctoral positions. Graduates commonly list their doctoral program itself as a position ("Graduate Research Assistant," "PhD Candidate"), typically with no end date, such that it can appear as ongoing university employment in 2026. We remove such entries with two rules. First, any position whose title is classified as graduate student or doctoral trainee is dropped, at any employer (13,888 positions). Second, any position at the graduate's own PhD institution that began before the year of the degree is dropped (17,020 positions). Positions at the graduate's own institution that began at or after graduation are retained, since staying on as a postdoctoral researcher or faculty member can be a real employment outcome. Together these rules remove 18,789 positions, leaving 135,911 graduates with at least one surviving 2026 position.

[7]The primary classifiers are Claude Sonnet 5 (Anthropic) and GPT-5.6 Terra (OpenAI), with Gemini 3.5 Flash (Google) adjudicating disagreements. Prompts available on request.

Outcome variables. From the surviving positions, we construct, for each graduate, indicators for holding any 2026 position at an employer in each sector, with "academic" employment defined as any of university, hospital or academic medical center, or FFRDC employment. The resulting sector indicators are not mutually exclusive: graduates who changed employers in 2026, or with concurrent appointments, can be in more than one sector, such that the indicators can add to more than one. Conditional on academic employment, we also create dummies for holding a tenure-track faculty, non-tenure-track faculty, or postdoctoral position. For graduates employed at a U.S. university, we additionally record whether the employer is an R1 institution and the employer's U.S. News & World Report and NRC program ranks, and where a graduate holds more than one university position, we take the best value across them—though these placement-quality measures are not currently being used in the analyses reported in this paper.

A.2.3 Coverage and selection

Coverage of LinkedIn is inevitably not random. If students who write with AI differ in their propensity to maintain a profile, the estimation sample for analyses of professional outcomes could be selected on characteristics correlated with the treatment.

To assess this possibility, we examine selection at each step of our measurement funnel: whether a graduate is linked to a Revelio profile at all, whether the linked profile reports any 2026 position (conditional on having a profile), and whether any position survives the stale-entry rules and contributes an outcome (conditional on having a position). Table A.3 reports regressions of each indicator on the AI-abstract indicator, with university-field and year fixed effects and standard errors clustered by university-field, over the 2023-2025 cohorts used in the paper's employment analysis. Graduates with AI-written abstracts are about three percentage points more likely to be linked to a profile (on a base of 70%). Conditional on having a linked profile, however, they are no more or less likely to have a reported 2026 position, and conditional on a position only marginally (and statistically weakly) more likely to have a position that survives our cleaning. The main difference is thus that graduates with AI-written abstracts are more likely to be on LinkedIn—a pattern which is consistent with their increased tendency to enter industry. Selection of this size cannot account for the employment differences in Figure 3: trimming the AI-abstract group by the excess link rate (per Lee 2009) bounds the industry difference between 6 and 11 p.p. and the academic difference between -10 and -6 p.p., both of which are far from zero.

Two further limitations should be noted. First, a position's employer is self-reported and updated at the profile owner's discretion—such that a graduate who has changed jobs but not updated their profile will be recorded at the earlier employer. This is the main reason we exclude 2026 graduates from the employment analysis: restricting this analysis to 2023-2025 cohorts permits at least half a year for profile updates, which limits but does not eliminate this problem. Second, because the analysis is restricted to R1 graduates, among whom research careers are more common and AI use is lower than in the sample as a whole, the estimated employment differences may not generalize to graduates of other institutions. That said, over 80% of graduates in our data are from

R1 universities, and we consider them our primary population of interest.

Table A.3: Selection into the employment sample

	(1) Linked to Revelio	(2) Reports a 2026 position	(3) Has a counted outcome
Has AI in abstract	0.030***	0.007	0.010*
	(0.006)	(0.005)	(0.005)
N	90297	62674	55859
R^2	0.12	0.06	0.08
Y mean	0.696	0.892	0.914

Notes: Table estimates differences in Revelio coverage among 2023-2025 graduates by whether the graduate has an AI-generated dissertation abstract. Column (1) estimates differences in whether the graduate is linked to a Revelio (LinkedIn) profile; Column (2), whether the linked profile reports at least one position during 2026 (conditional on a linked profile); and Column (3), whether at least one 2026 position survives the exclusion of stale doctoral-program entries and contributes to our measurement of employment outcomes (conditional on having a 2026 position). All specifications include university-field and year fixed effects. *, **, *** represent significance at the 0.1, 0.05, and 0.01 levels, respectively. Standard errors clustered by university-field in parentheses.

B Supplementary Results

B.1 AI use by university and program

Tables B.1 and B.2 list the universities and university-fields with the fewest and most students with AI writing. Both tables are restricted to R1 universities and 2025-2026 graduates, and report the number of such graduates and share with AI-generated abstracts. Table B.1 lists the ten universities with the lowest and highest shares among those with at least 50 graduates over this period; Table B.2 does the same for university-fields with at least 20 graduates. The lowest rates are concentrated at a small number of highly selective institutions and in mathematics, physics, and chemistry, whereas the highest are spread across a wider set of universities and fields, with computer science and engineering appearing most often.

Table B.1: Top/bottom universities by share of AI dissertations, 2025- (R1 only)

Lowest share			Highest share		
University	N	Percent AI	University	N	Percent AI
U. of Massachusetts-Amherst	51	4%	U. of Idaho	90	28%
U. of Oregon	123	9%	Lehigh U.	131	28%
Massachusetts Inst. of Tech.	389	9%	Arizona State U.	529	29%
Syracuse U.	84	10%	Saint Louis U.	57	30%
Boston Coll.	61	10%	U. of Nevada-Reno	117	30%
California Inst. of Tech.	151	10%	U. of Southern California	73	30%
Ohio State U.	285	12%	U. of Memphis	82	30%
Georgetown U.	67	12%	U. of North Texas	81	33%
U. of Houston	66	12%	U. of Wyoming	75	35%
Princeton U.	444	13%	Virginia Commonwealth U.	50	36%

Notes: Table lists universities with the lowest and highest shares of their post-2025 dissertations with AI abstracts. Analysis restricted to R1 universities with at least 50 graduates over this period. Table reports, for each university, the number of graduates and the AI share.

Table B.2: Top/bottom university-fields by share of AI dissertations, 2025- (R1 only)

Lowest share				Highest share			
University	Field	N	Percent AI	University	Field	N	Percent AI
U. of California-Berkeley	Maths	33	0%	U. of Florida	Agri. Sci.	21	48%
Ohio State U.	Chemistry	20	0%	Temple U.	Comp. Sci.	21	48%
Rice U.	Maths	20	0%	Rutgers U.	Comp. Sci.	48	50%
Northwestern U.	Bio. Eng.	29	0%	U. of Illinois Chicago	Oth. Eng.	20	50%
Massachusetts Inst. of Tech.	Physics	57	2%	U. of California-Davis	Comp. Sci.	21	52%
U. of Minnesota-Twin Cities	Health Sci.	34	3%	Northeastern U.	Health Sci.	66	53%
U. of California-Santa Barbara	Physics	31	3%	Indiana U.-Bloomington	Comp. Sci.	28	54%
Virginia Tech	Bio. Sci.	32	3%	Iowa State U.	Comp. Sci.	20	55%
Massachusetts Inst. of Tech.	Health Sci.	29	3%	U. of California-Santa Cruz	Comp. Sci.	26	58%
U. of Colorado Boulder	Maths	28	4%	U. of California-Irvine	Oth. Eng.	26	58%

Notes: Table lists university-fields with the lowest and highest shares of their post-2025 dissertations with AI abstracts. Analysis restricted to R1 universities and to university-fields with at least 20 graduates over this period. Table reports, for each university-field, the number of graduates and the AI share.

B.2 Peer AI use as a predictor of individual use

The patterns in Figure 2 raise the question of whether AI use is a property of individual students or of the environments in which they train. To evaluate this question, we relate a graduate's own AI use to that of their immediate predecessors in the same program or under the same primary advisor (committee chair). We estimate the following specification:

$$AI_{it} = \beta \cdot PeerAI_{p,t-1} + \delta_t + \varepsilon_{it}. \tag{B.1}$$

where i indexes graduates, p indexes peer groups (a doctoral program, defined as a university-field, or an advisor), and t indexes graduation years; AI_{it} is an indicator for an AI-generated abstract; $PeerAI_{p,t-1}$ is the average AI share of abstracts among the previous year's graduates of the same peer group; and δ_t are year fixed effects, with an estimation sample of 2023 to 2025 graduates with at least one same-group graduate in the preceding year.

Table B.3 reports estimates with the program-based measure (Columns 1 to 3) and the advisor-based measure (Columns 4 to 6), in each case sequentially narrowing the sample from all institutions to R1 and R2 institutions to R1 institutions only. The estimates are stable across samples: a 10 p.p. higher AI share among the previous year's graduates of the same program is associated with a roughly 3 p.p. higher probability that a graduate's own abstract is AI-generated, and the corresponding figure for the same advisor is roughly 2 p.p. We interpret these results as indicating that AI use clusters within training environments—whether due to peer influence, shared norms, common access, or advisor practices—though not as a source of exogenous variation in AI use per se, since these same environments could affect graduates' outcomes directly.

Table B.3: Same-program and same-advisor peer AI use predicts own AI use

	Program-based			Advisor-based		
	(1)	(2)	(3)	(4)	(5)	(6)
	All	R1/R2	R1 only	All	R1/R2	R1 only
Program AI share	0.305***	0.274***	0.277***			
	(0.036)	(0.038)	(0.044)			
Advisor AI share				0.183***	0.176***	0.172***
				(0.024)	(0.026)	(0.029)
N	88838	82141	73705	33667	31195	28400
R^2	0.07	0.07	0.07	0.08	0.08	0.08
Year FEs	Y	Y	Y	Y	Y	Y

Notes: Table shows that an individual's use of AI can be predicted from (i) the rate at which prior-year students in the same doctoral program (measured by university-field) use AI (Columns 1 to 3) and (ii) the rate at which prior-year students with the same advisor use AI (Columns 4 to 6). Within each prediction set, we sequentially narrow the sample from all institutions, to R1 and R2 institutions, to R1 institutions only; patterns are similar across them. Sample restricted to 2023-2025 graduates with at least one prior-cohort peer in the same program or with the same advisor. *, **, *** represent significance at the 0.1, 0.05, and 0.01 levels, respectively. Standard errors clustered by doctoral program in parentheses.

Appendix References

Antman, Francisca M, Xuechao Qian, Kirk Doran, and Bruce A Weinberg. 2023. *Innovation Nation: Evidence from Broadening Access to Ph.D. Training in the U.S.* Working Paper.

Arora, Ashish, Sharon Belenzon, Larisa Cioaca, Lia Sheer, and Hansen Zhang. 2026. "The Effect of University Science on Corporate Innovation." *Management Science* .

Bikard, Michaël, Fiona Murray, and Joshua S Gans. 2015. "Exploring trade-offs in the organization of scientific work: Collaboration and scientific reward." *Management Science* **61**(7):1473-1495.

Buffington, Catherine, Benjamin Cerf, Christina Jones, and Bruce A Weinberg. 2016. "STEM training and early career outcomes of female and male graduate students: Evidence from UMETRICS data linked to the 2010 census." *American Economic Review* **106**(5):333-338.

Jiang, Xuan, Joseph Staudt, and Bruce A Weinberg. 2023. *A tale of two fields? STEM career outcomes*. NBER Working Paper No. 31835.

Lee, David S. 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *Review of Economic Studies* **76**(3):1071-1102.

National Research Council. 2011. *A Data-Based Assessment of Research-Doctorate Programs in the United States*. Washington: The National Academies Press.

Shvadron, Dror, Hansen Zhang, Lee Fleming, and Daniel P Gross. 2025a. *Measuring the Production of Scientific Human Capital: New Data, Methods, and Evidence on the U.S. Scientific Training Ecosystem*. NBER working paper no. 33944.

Shvadron, Dror, Hansen Zhang, Lee Fleming, and Daniel P Gross. 2025b. "A Quarter of US-Trained Scientists Eventually Leave. Is the US Giving Away Its Edge?" *arXiv preprint arXiv:2512.11146* .

Toole, Andrew A and Dirk Czarnitzki. 2010. "Commercializing science: Is there a university "brain drain" from academic entrepreneurship?" *Management Science* **56**(9):1599-1614.