TECHNOLOGY POLICY INSTITUTE

RESEARCH PAPER

NOVEMBER 2026

Is AI Replacing Human Research Assistants?

Evidence from NBER Working Papers

Scott Wallsten

President and Senior Fellow, Technology Policy Institute

swallsten@techpolicyinstitute.org

techpolicyinstitute.org

2001 L Street NW, Suite 500, Washington, DC 20036

Is AI Replacing Human Research Assistants? Evidence from NBER Working Papers

Scott Wallsten*
November 1, 2026

Abstract

Is generative AI replacing research assistants in economics? I cannot observe the assistants, so I measure the printed record. Language models under my direction read the acknowledgments of all 11,718 NBER working papers issued from 2018 through October 2026 and recorded whether each thanks a human for research assistance and whether it says the authors used AI. A credit reflects norms about whom to thank as well as whom the authors hired. The share of papers thanking a research assistant fell from 40 percent before ChatGPT to 37 percent after, and by two to two and a half points, depending on the series) against 2018 and 2019 alone. Authors who published at NBER in both periods thank assistants three to five points less often than before. Four large programs, Asset Pricing, Monetary Economics, Economic Fluctuations and Growth, and Public Economics, fell by statistically significant amounts on their own, and Asset Pricing's decline survives a simultaneous test of all 20 programs. The decline is entirely in whether a paper credits anyone. Papers that credit an assistant name as many as before, and the assistants still credited do not appear on more papers each. Statements that the authors used AI rose from none to about one paper in twenty, but outside NBER's methods series the decline in credits is not where the disclosed AI use is. The sharpest change is in NBER's Technical Working Papers, its econometrics and methods series, where in 2026, 2 of 28 thanked a research assistant and 8 said the authors used AI, mostly for drafting and coding. I found that cell after inspecting every program, and it is small. Because papers appear after the research is done, and because users take years to learn what a new tool can do, these data may not yet show the effects of the newest tools.

---

* President and Senior Fellow, Technology Policy Institute. I thank NBER, and especially Kevin Tasley and James Poterba, for permission to have AI agents read the public pages of its working paper series. Claude, an AI system from Anthropic, carried out this study under my direction, and ChatGPT ran an independent replication (Sections 5 and 6). I am solely responsible for any errors.

1. Introduction

Research assistants (RAs) are the junior partners of empirical economics. They clean data, write code, and run the first estimates. In return they get training, a letter, modest pay, and sometimes a line in the acknowledgments. When a tool arrives that can clean data, write code, and run the first estimates, the natural question is whether it replaces the human RA.

This paper asks a narrow version of that question. Across 11,718 NBER working papers issued from 2018 through October 2026, I measure the change in the share of papers that credit a human for research assistance and in the share that say the authors used an AI tool.

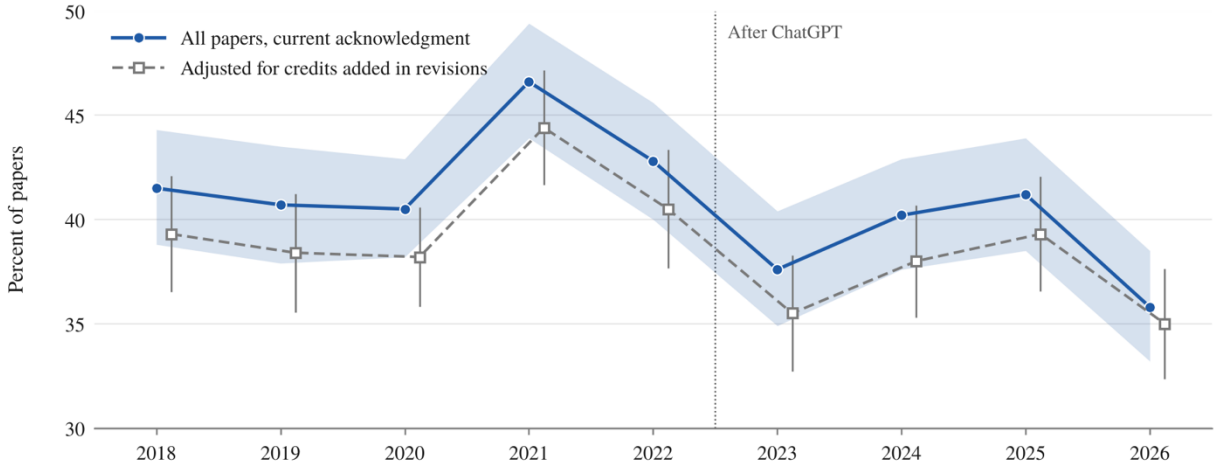
I also estimate regressions that relate a credit, the number of assistants named, and an AI-use statement to a paper's field, team size, and revision status and to characteristics of its authors, including whether any author works at a business school, how long the most senior author has been publishing, and how many economics graduate students the authors' institution enrolls.

The analysis yields two main results.

First, the share of NBER papers crediting a research assistant has fallen modestly. When I count every paper and adjust for credits that authors add in later revisions, 40 percent of papers issued before ChatGPT's release in November 2022 credited a research assistant, against 37 percent of those issued after it (Figure 1). Against that 2018 to 2022 baseline the drop is statistically significant, but 2021 and 2022 were the highest years in the series, and pandemic-era output may have raised assistant use in those years, so part of the decline may be a return to the 2018 and 2019 level. Measured against 2018 and 2019 alone, the decline across all papers is about two and a half points ($p = 0.04$), and about two points after the revision correction ($p = 0.08$). Authors who published at NBER in both periods credit assistants three to five points less often after ChatGPT than before, against every baseline. The year-by-year trend is weaker, and whether it differs from zero depends on the measure. In regressions that hold team size, revision, and field fixed, the drop is 3.3 points ($p < 0.001$), and 4.1 points within authors ($p < 0.001$). The decline is entirely in whether a paper credits anyone. Papers that still credit an assistant name as many as before, and the assistants still credited appear on no more papers each. Outside the methods series, the decline is not where the disclosed AI use is. Credits fell no more, and across programs less, where statements of AI use rose. Finance papers, a grouping I formed after seeing the program results and then tested (Section 4.1), account for a disproportionate share of the decline. On a wider definition that also counts predoctoral and research fellows, the decline is 3.6 points against 2018 through 2022 and 2.1 against 2018 and 2019 ($p = 0.08$), so assistants moving onto the author line account for little of it.[1]

[1] A parallel study in a different AI system, which began first and coded every paper, finds the same pattern (Section 6).

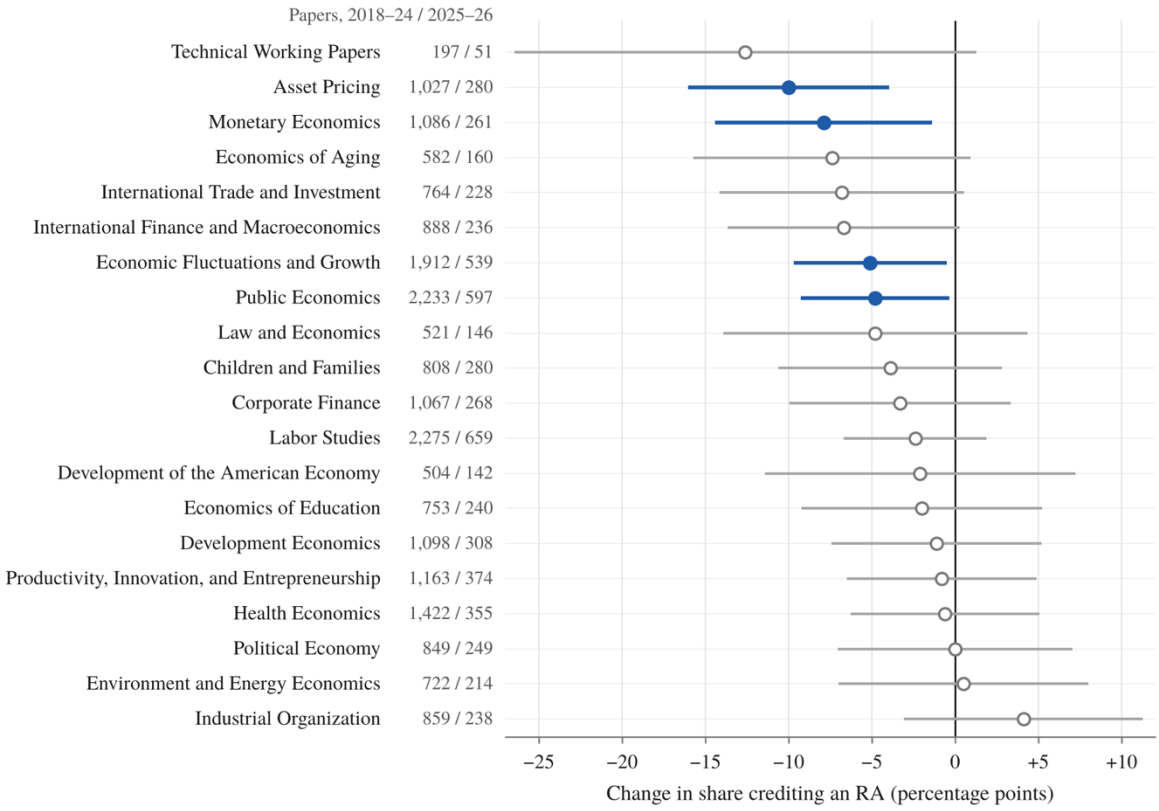
Figure 1. Share of NBER working papers crediting a research assistant



Notes: All 11,718 NBER working papers. 95 percent intervals: band for all papers; bars for the adjusted series, including the error in the revision correction. 2026 runs through October 13. Dotted line: ChatGPT release, November 30, 2022.

Second, the decline differs across fields, with the largest in NBER's Technical Working Papers, a series of mostly econometrics and methods papers (Figure 2).

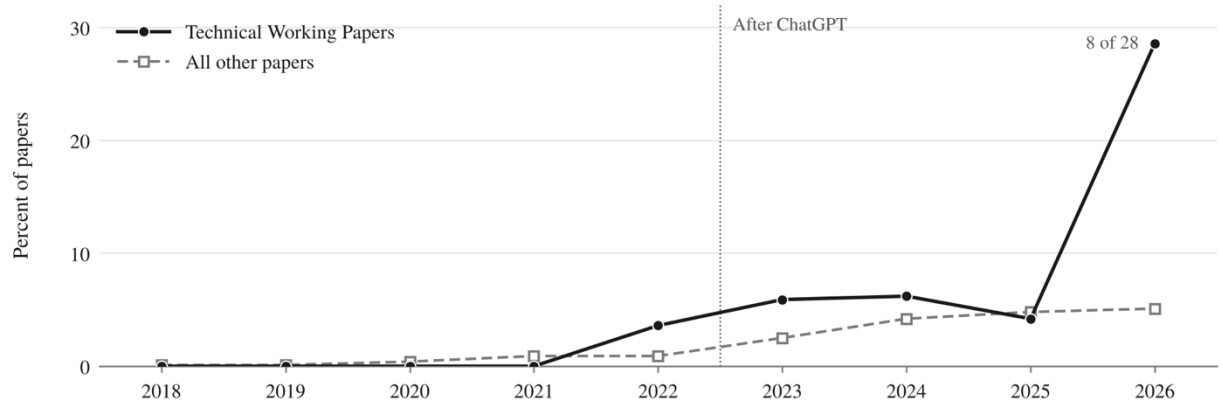
Figure 2. Change in the share of papers crediting a research assistant, by NBER program, 2025 and 2026 against 2018 to 2024



Notes: Change in the share of papers crediting a research assistant, 2025 and 2026 pooled minus 2018 to 2024, all papers. Normal 95 percent intervals, unadjusted for 20 comparisons; filled markers exclude zero. Papers count in every program they list. Technical Working Papers is NBER's econometrics and methods series. 2026 runs through October 13.

From 2018 through 2025, between 29 and 52 percent of these papers credited a research assistant each year. In 2026, 2 of 28 did, and 8 of the 28 stated that the authors used AI, mostly for tasks such as drafting and coding (Figure 3).

Figure 3. Share of papers stating that the authors used AI



Notes: Share of papers stating that the authors used AI. Technical Working Papers is NBER's econometrics and methods series; all other papers are every paper outside it. Sources are the acknowledgment NBER displays plus full text where retrieved; full text covers every paper through 2024 but few in 2025 and 2026, so those years are floors. 2026 runs through October 13. Dotted line: ChatGPT release, November 30, 2022.

The authors of those 28 papers had not stopped crediting assistants. They did so on 48 percent of their 264 earlier NBER papers and on 39 percent of their 18 papers outside the series in 2026. Papers outside the series that carry an econometric-methods JEL code moved the same way, but by less, from 38 percent in 2018 through 2022 to 32 percent in 2023 through 2026.

The Technical Working Papers series is small, and four large programs moved in the same direction on far more papers. Across all NBER papers, the share crediting an assistant in 2025 and 2026 was 10.0 points below its 2018 to 2024 level in Asset Pricing and 7.9 points below in Monetary Economics. Economic Fluctuations and Growth fell 5.1 points and Public Economics 4.8. Each of these declines is statistically significant by itself, and Asset Pricing's stays significant in a simultaneous test of all 20 programs. None of the four shows an unusual rate of AI-use statements in 2026 (none, 4, 3, and 3 percent), so whatever is happening there does not show up in what authors say about their tools.

The regressions find no difference in the decline between business-school authors and others. Papers whose most senior author has been publishing longer credit assistants less often, and their decline is smaller, though not significantly so.

Publication lag affects how to interpret the results. NBER issues papers well after the research ends. The median revision in this frame comes 13 months after issue, but I do not observe the lag between the start of the work and the working paper. A second lag sits behind the first. The history of computing shows that the binding delay in a general-purpose technology is the co-invention that makes it useful, as users learn what to do with it, reorganize work around it, and build the complements it needs. That took years for mainframes and for client-server computing (Bresnahan and Trajtenberg 1995; Bresnahan and Greenstein 1996). An economist who replaces an assistant with a model must learn to specify, check, and supervise the model's work, and that learning is slow and costly. Bresnahan, Greenstein, and Yin (2025) call this incremental co-invention, the low-cost, low-risk adaptation that extends what adopters already do and leaves the product, here the paper, unchanged. Incremental co-invention should show up in whom authors hire before it shows up in what they publish. Tools that run code on uploaded data existed by

mid-2023; tools that carry out a multi-step empirical task with little supervision arrived in 2025. On the co-invention reading, a series that does not steepen through October 2026 is what this stage should look like, and the record so far speaks against fast task-level substitution without speaking to slow reorganization. These data do not measure either lag. The record through October 2026 shows no large fall in printed research-assistant credits.

I designed this study and had large language models carry it out under a frozen protocol, from retrieval through coding to drafting the text. Section 5 explains how, what that implies for the counts, and why a study of this kind is itself evidence of what the tools could do in late 2026. I froze the protocol on October 15, 2026, before the study coded any paper. Each later change is an amendment, numbered A1 to A40, logged with its date and with what the study had seen when it made the change. The paper cites amendments by number, and Appendix D lists them all and gives the coding and validation detail.

2. Related work

2.1 Acknowledgments as data

Economists have used acknowledgments as data before, mostly to study informal collaboration. Laband and Tollison (2000) coded the acknowledgments of economics journal articles to measure the exchange of comments among economists, and Rose and Georg (2021) built a network from the acknowledgments of more than 5,000 financial economics papers. The bibliometric literature has used automated extractions to study acknowledgments at larger scale (Giles and Councill 2004; Cronin, Shaw, and La Barre 2003; Desrochers, Paul-Hus, and Pecoskie 2017) and documented their limits. Web of Science records acknowledgment text only when the acknowledgment names a funder. That covers about 28 percent of social science articles (Smirnova and Mayr 2023) and about 39 percent of top-five economics articles (Baccini and Petrovich 2022). Automated extraction also misses some of what authors print (Alvarez-Bornstein, Morillo, and Bordons 2017), and Giles and Councill (2004) report a precision of about 78 percent. Printed credit understates labor (Ross et al. 2022; Salandra, Miraldo, and Criscuolo 2026) and carries status content (Baccini and Petrovich 2022; Paul-Hus et al. 2020).

The closest work to this study is Caro (2024), who coded research-assistance credits in NBER working papers from 1973 to 2023. He identified about 14,400 research assistants. The share of papers crediting one rose from roughly 10 to 15 percent in the 1970s and 1980s to about 35 percent over 2013 to 2023, with a peak near 40 percent in 2021 (Figure 5). Over the same period, the number of assistants per crediting paper rose from about 1.5 to over 3. His series ends in 2023. This paper extends the count through October 2026, adds author statements of AI use, and compares each paper's first and current versions, which no earlier work does.

2.2 Research assistants in organizing empirical work

Economics has moved toward team production and data-intensive work (Wuchty, Jones, and Uzzi 2007; Jones 2021; Hamermesh 2013; Angrist et al. 2017; Einav and Levin 2014), which plausibly raises demand for the tasks research assistants do. Technical tasks fall to junior scholars (Larivière et al. 2016), and research-assistant and predoctoral positions are a stepping stone into economics PhD programs (Broady 2026). Korinek (2023) lists dozens of research

tasks that language models can perform for economists, several of which research assistants traditionally do.

2.3 AI and entry-level work

Experimental studies find the largest productivity gains from generative AI for less experienced workers (Noy and Zhang 2023; Brynjolfsson, Li, and Raymond 2023; Dell'Acqua et al. 2023; Cui et al. 2026). Employment evidence is observational. Brynjolfsson, Chandar, and Chen (2025, revised 2026) track payroll records and find that employment of 22- to 25-year-olds in the occupations most exposed to AI has fallen about 19 percent below what it would have been had it kept pace with similar workers in less exposed occupations. They find no comparable gap for experienced workers, no economy-wide displacement, and a gap that has widened with each release of the data. The decline operates through reduced hiring rather than separations, concentrates where observed AI use automates tasks rather than complements them, and falls on occupations that rely on codified knowledge, the kind taught in courses and manuals. Employment in occupations that rely on tacit knowledge has risen among experienced workers. Hosseini Maasoum and Lichtinger (2025) find that junior hiring fell at adopting firms; Humlum and Vestergaard (2025) find precise null effects on earnings and hours in Denmark. A research assistant is an early-career worker hired afresh for each project to do codified work, including cleaning data, writing code, and running first estimates, so the research-assistant market is a small, well-documented case of the group in which Brynjolfsson, Chandar, and Chen find the effect. None of these studies observes research assistants in academic economics, and all face the timing problem described in Section 1.

2.4 Measuring AI use in research text

Corpus-level estimators infer language-model involvement from vocabulary (Liang et al. 2025; Kobak et al. 2025) but do not classify individual papers, and document-level detectors are unreliable (Liang et al. 2023; Weber-Wulff et al. 2023; Sadasivan et al. 2023; Saha et al. 2026). Disclosure counts are low relative to surveyed use (He and Bu 2025; Ans et al. 2025; Liao et al. 2024; Linardon et al. 2025) and do not distinguish a model used as a method from an assistant used for tasks. This paper counts author statements, splits them by role, and treats them as a floor.

3. Data and Methods

This section describes the data, how the study identifies research-assistant credits and AI use, and the methods.

3.1 Data sources

The paper relies on two sources of data.

The first source is the "Acknowledgements and Disclosures" block on each paper's NBER abstract page, which I obtained for all 11,718 papers. This block is the cover-page acknowledgment of the paper's current version.

This source has two limitations. First, it omits first-page footnotes and the body of the paper. That matters little for research-assistant credits, which authors almost always print on the cover.[2] Authors tend to put AI-use statements in the body or footnotes, so the abstract page misses many of them. The cover block shows at least one in thirteen of those statements (Section 4.3). Second, it reflects the current version, and revised papers acquire credits as assistants turn over. Older papers have had more time for revision, so a series built from current text alone overstates credits in earlier years and slopes down for that reason. Section 3.3 describes how the paper estimates and subtracts that effect, and Appendix B gives the measurement.

The second source is the PDF of each paper's first version. It shows what the authors wrote before any revision and includes the full text. It lets the study measure how much revision adds, check the corrected series against text that needs no correction, and find AI-use statements anywhere in the paper. I obtained the first version of every paper NBER would serve, 9,644 in all, including every paper issued from 2018 through 2024. NBER serves PDFs of papers issued from May 2025 onward only to subscribers, which I am not, so the first-version sample holds only 449 papers for 2025 and 54 for 2026. That leaves the AI-use rates for those two years understated. Appendix A gives the retrieval detail.

The regressions also use author and institution characteristics. Author affiliations, h-indexes, and first publication years come from OpenAlex, which matches 10,203 of the 11,718 papers to at least one institution. Economics graduate enrollment by institution and year comes from NSF's Survey of Graduate Students and Postdoctorates.

3.2 Identifying RA credits and AI use

A paper has a research-assistant credit when the acknowledgment text explicitly credits one or more non-author humans with research assistance using the words "research assistance," "research assistant," "RA," or a direct variant. Named people, unnamed people, groups, and institutional staff all count. Comments, discussants, referees, data access granted by an agency, funding described as research support, and the authors' descriptions of their own roles do not count. Wording that might be a credit but is not explicit, such as "research support" where one cannot tell people from money, or thanks to a person with no role stated, counts as ambiguous. Ambiguous and unreadable cases stay out of the point estimates and enter only bounds that count them all as credits or all as none. After the coding was complete I read the ambiguous cover blocks and decided that any named assistance with the research counts as a credit; the series below use that definition, and Section 5 reports its effect (amendment A35). A wider definition adds predoctoral fellows, research fellows and associates credited in that role, and "research support" clearly attributed to named people. For papers with a credit the study records the names as printed, the number of distinct named persons where the text makes the allocation clear, and any task the text states. Every share in the tables below is the share of papers coded as crediting an assistant among papers coded as crediting an assistant or crediting none.

A paper has an AI-use disclosure when the authors state that they used an AI tool, language model, or named assistant for some part of the research or writing. A paper that studies AI, cites AI research, or names a grant with AI in its title does not disclose use. I split disclosures into

[2] About 1 percent of papers credit an assistant only outside the cover block.

analytic-method use, where the authors apply a pretrained or third-party model or service (a language model, an embedding model, a translation or face-recognition service) to data as part of the design, and assistant-task use, where a tool helped with coding, editing, drafting, or search. Self-trained machine learning is not a disclosure. For the first-version sample the AI screen ran on the full text. For the other papers the study had only the cover block, so a cover-block paper counts as disclosing only when that block says so, and the cover-block coders recorded whether the paper disclosed use and named the tool, not the kind of use. An AI tool thanked "for research assistance" counts as an AI disclosure and not as a human credit; the two fields are independent.

3.3 Methods

I built four series to check the robustness of each result. The all-paper census reads the acknowledgment NBER currently displays for each of the 11,718 papers. It covers every year fully, but because revised papers acquire credits, older papers carry more credits than they did when first released. The version-corrected census subtracts an estimate of those extra credits, each cohort's revision rate times a pooled gain, which assumes the gain is the same across cohorts. The first-version sample, the 9,644 papers whose original PDF I obtained, supplies the measurement behind that correction and an uncorrected check on it. The same-authors comparison uses the all-paper census and holds the set of authors fixed, which removes changes in who publishes at NBER. Table 1 describes each series.

The version-corrected census subtracts each cohort's expected revision-added credits.[3]

[3] $corrected(t) = census(t) - r(t) \times g$, where r(t) is the share of year-t papers that NBER lists as revised and g is the net gain from revision, the share of revised papers that credit an assistant in their current version minus the share that did in their first version. Coding both versions of 2,964 revised papers gives g = 7.0 percentage points (SE 0.55). In 2018, 31 percent of papers had been revised, so the correction is 2.2 points, and in 2026, 12.5 percent had, so it is 0.9. Appendix B reports the measurement.

Table 1. The series of research-assistant credits used in this paper

Series	How it is built	What the construction implies
All-paper census	The current cover block on the NBER abstract page for all 11,718 papers.	One instrument, every paper, every year, including full 2025 and 2026 cohorts. Most outside readers of NBER, SSRN, or any text corpus would use this series.
Version-corrected census	The all-paper census, adjusted to remove credits added by revision (Section 3.3).	Removes the mechanical accumulation of credits. It lowers the 2018 to 2024 cohorts by about two points and the 2026 cohort by less than one.
First-version sample	The 9,644 papers whose PDF NBER served, coded from the earliest version available, including every paper for 2018 through 2024.	The measurement behind the version correction. It needs no correction because it reads text written before any revision. It holds only 449 papers for 2025 and 54 for 2026 because NBER serves newer PDFs only to subscribers, so those two years depend on which papers NBER happened to serve.
Same-authors comparison	Author-paper pairs for the 3,662 NBER authors with papers both before and after November 30, 2022, coded from the current cover block, with author fixed effects.	Holds the set of authors fixed, so a change in which authors publish at NBER cannot drive it, though their topics, coauthors, and mix of papers can still change. Multi-author papers count once per author, and standard errors are clustered by author and by paper. Weighting by one over the number of authors gives nearly the same answer.

Comparisons before and after ChatGPT split papers by issue date at November 30, 2022, the tool's public release, and a paper issued after that date counts as after. The protocol fixed this split in advance. It marks only the release date and says nothing about when authors began using the tool. Because of the publication lag (Section 1), the paper also reports shares by year and, where it groups years, compares 2025 and 2026 with 2018 through 2024. Section 4.1 also gives the before-and-after difference with 2023 omitted. Trends are least-squares slopes on issue year in percentage points per year.

Unless the text says otherwise, I report 95 percent confidence intervals and heteroskedasticity-robust standard errors. The protocol fixed five primary tests in advance, all on the first-version sample. They are the trend in the share of papers crediting a research assistant, the change in that share after November 30, 2022, the trend in named assistants per paper, the trend in a broader measure of research support, and the trend in AI-use statements. The protocol made the first-version sample primary because it needs no revision correction. The main text leads with the all-paper census because the first-version sample covers few 2025 and 2026 papers.

I adjust the five primary tests with the Romano-Wolf stepdown procedure (Romano and Wolf 2005), which accounts for the correlation among them.[4] I adjust the prespecified program-by-

[4] The protocol specified a Holm adjustment for these five tests. The Claude session that ran the study drafted it, and I approved it before the study coded any paper. I did not know the Holm method when I approved the protocol. Romano-Wolf is the adjustment I would have chosen. I made the switch after seeing every outcome, which departs from the protocol, and the repository records it as amendment A32. It changes no conclusion. The Holm-adjusted p-values are 0.42 for the first-version trend, 0.011 for the before-and-after difference, 0.039 for the named-count trend, 0.21 for the research-support measure, and below 0.0001 for the AI-use trend. The Romano-Wolf values come from 2,000 bootstrap draws of papers within each issue year. The episode is a small example of a risk in delegating research to a model. An author who approves a complete, competent-looking protocol can sign off on choices he would not have made and could not defend, and this one came to light only when I asked the session why it had

program trends with the Benjamini-Hochberg procedure, as the protocol specified. The post hoc program comparisons in Section 4.4 are unadjusted, and where I say a decline survives a Bonferroni correction I mean that its p-value is below 0.05 divided by 20. The other sensitivity analyses are unadjusted.

Regressions ask what predicts a credit and whether the post-ChatGPT change differs across kinds of papers and authors. They use the all-paper census, which covers every paper. The first models whether a paper credits any research assistant, with a logit whose estimates I report as changes in probability; the second models the number of assistants named, with Poisson pseudo-maximum likelihood; and the third models whether the cover block states that the authors used AI, with a linear probability model. Every regression includes an indicator for issue after November 30, 2022; the log of the number of authors, because larger teams have more people who might hire an assistant; an indicator for a revised paper, because revision adds credits; and indicators for the paper's NBER programs, because fields differ in how much data work they require.

Three author characteristics come from OpenAlex. An indicator for any author at a business school stands in for research budgets, because business schools tend to spend more on research support and may have paid for AI tools sooner than economics departments and research grants did.[5] The career age of the paper's most senior author, the years since that author's first publication, measures seniority, which may affect both whether an author can pay for assistants and how quickly the author adopts new tools. The highest h-index among the authors measures standing in the field. A fourth variable, full-time enrollment in the economics graduate program at the authors' largest-program institution, measures the local supply of potential assistants. Interactions of the business-school indicator and career age with the post-ChatGPT indicator test whether the decline differs along those lines. Author fixed effects compare the same authors' papers before and after November 2022, and institution fixed effects identify the enrollment coefficient from changes within an institution over time. The protocol did not include these regressions, and the repository records them as amendment A34.

4. Results

4.1 Papers that credit any human research assistant

In the first-version sample, the share of papers crediting a research assistant falls from 40.0 percent before November 30, 2022, to 36.8 percent after, a drop of 3.2 points ($p = 0.003$) that matches the version-corrected census. The trend is minus 0.2 points a year but is not statistically different from zero ($p = 0.42$).

Table 2 reports the share by issue year. The 2025 and 2026 cells hold only the papers whose PDFs NBER served, 449 from 2025 and 54 from 2026, of which the study could code 445 and 53 as crediting or not crediting an assistant. If I treat the papers NBER refused as unknown, the share could lie anywhere from 14.3 to 79.8 percent in 2025 and from 1.3 to 97.2 percent in 2026.

chosen that method.

[5] I thank Shane Greenstein for suggesting this comparison.

On the all-paper census the share for 2025 and 2026 is 38.5 percent against 41.4 percent for 2018 to 2024, and 35.8 percent for 2026 alone.

Table 2. Share of papers crediting a research assistant, first-version sample (PDF, earliest version served)

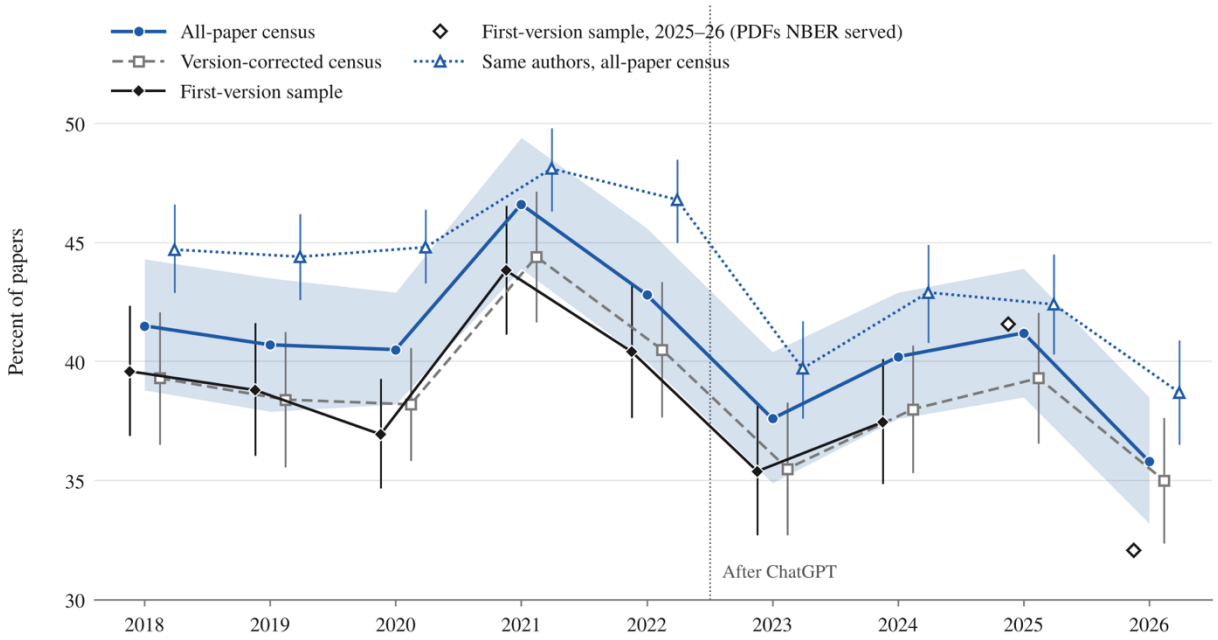
Year	n	Positive	Share (%)	95% CI
2018	1,220	483	39.6	36.9 to 42.4
2019	1,165	452	38.8	36.0 to 41.6
2020	1,689	624	36.9	34.7 to 39.3
2021	1,287	564	43.8	41.1 to 46.5
2022	1,173	474	40.4	37.6 to 43.2
2023	1,181	418	35.4	32.7 to 38.2
2024	1,292	484	37.5	34.9 to 40.1
2025†	445	185	41.6	37.1 to 46.2
2026†	53	17	32.1	21.1 to 45.5

n counts papers coded as crediting a research assistant or not. The table leaves out the 139 papers whose acknowledgments are ambiguous or unreadable. Counting all of them as credits, or all as no credit, changes no year's share by more than about a point. Rows marked † cover only the 2025 and 2026 papers whose PDFs NBER served (Appendix A).

In all four series, 2023 and 2026 have the lowest shares, so the size of the before-and-after difference depends on where 2023 falls. With 2023 omitted and the post period starting in 2024, as the parallel study in Section 6 defines it, the all-paper census falls by 3.5 points (SE 1.0) and the version-corrected census by 2.9 (SE 1.0); with 2023 in the post period, as the protocol specified, they fall by 4.0 and 3.5, each with SE 0.9.

Removing the credits that revision adds flattens the all-paper trend from minus 0.50 to minus 0.40 points a year, and its p-value rises from 0.004 to 0.03.

Figure 4. Share of papers crediting a research assistant, by issue year and series



Notes: Share of papers crediting a research assistant, with 95 percent intervals. The first-version sample holds only the 449 papers from 2025 and 54 from 2026 whose PDFs NBER served, so those points (open diamonds) have no intervals or connecting lines. The same-authors series covers the 3,632 authors with NBER papers both before and after November 30, 2022, and plots year effects from a regression with author fixed effects. It sits higher than the other series because it counts each multi-author paper once per author. Points are offset slightly within each year. 2026 runs through October 13. Dotted line: ChatGPT release, November 30, 2022.

The decline also appears within authors. The papers of the 3,662 NBER authors with papers in both periods are 5.1 points less likely to credit an assistant after November 2022 than before (SE 0.9), and 4.6 points less likely after the revision correction. Each paper counts once per author, so multi-author papers carry more weight, and I cluster standard errors by both author and paper; counting each paper once gives nearly the same answer. The same authors' rate also rose 2.8 points between two windows before ChatGPT, from 2018 through 2020 to 2021 and 2022, so this comparison alone does not date the change to November 2022. Figure 4 plots the same-authors series by year.

Regressions agree (Table 3). With controls for team size, revision, and program, the share crediting an assistant falls 3.3 points after November 2022 ($p < 0.001$), and with author fixed effects it falls 4.1 points ($p < 0.001$). Across 85 variants of the series, the baseline, and the controls, every estimate is negative, and every estimate against the prespecified baseline of January 2018 to November 2022 is significant, between −4.1 and −2.8 points. With 2018 and 2019 as the baseline and 2024 through 2026 as the post period, none is. The revision indicator is partly an outcome, since authors have had less time to revise recent papers. Without it, the census falls 4.1 points ($p < 0.001$), and on the first-version sample, which revision cannot touch, 3.4 points ($p = 0.001$).

Table 3. Correlates of crediting any research assistant, all-paper census

	(1)	(2)	(3)	(4)	(5)	(6)
Model	Logit	Logit	Logit	Logit, no finance	Linear	Linear
Post-ChatGPT	−3.3 (<0.001)	−3.1 (<0.001)	−2.2 (0.02)	−1.3 (0.23)	−4.1 (<0.001)	−2.8 (0.03)
Business school		3.9 (<0.001)	3.8 (<0.001)	3.9 (0.001)		
Business school × post		−0.8 (0.67)	−0.4 (0.82)	1.2 (0.60)		
Senior author's career, per ten years			−2.9 (<0.001)	−3.4 (<0.001)		
Career × post			0.6 (0.40)	0.7 (0.40)		
Log(1 + highest h-index)			1.2 (0.17)	2.3 (0.01)		
Log economics graduate enrollment						0.5 (0.83)
Log number of authors	2.0 (0.04)	1.5 (0.13)	1.8 (0.08)	2.5 (0.03)	−2.8 (0.02)	−0.7 (0.63)
Revised	11.4 (<0.001)	11.3 (<0.001)	10.5 (<0.001)	10.7 (<0.001)	8.8 (<0.001)	11.0 (<0.001)
Program indicators	Yes	Yes	Yes	Yes	No	Yes
Fixed effects					Author	Institution
Observations	11,547	11,519	10,696	8,731	22,914 pairs	8,935

Notes: Estimates in percentage points, with p-values in parentheses. Columns 1 to 4 are logits reported as changes in the probability of a credit: the post, business-school, and career rows are contrasts with all interactions recomputed, the two × post rows are differences in differences, and the log rows are average marginal effects. Columns 5 and 6 are linear probability models. P-values come from robust standard errors, clustered by author and by paper in column 5 and by institution in column 6. Columns 2 to 4 use papers with OpenAlex affiliations; column 4 drops papers listing Asset Pricing or Corporate Finance; column 5 uses author-paper pairs for authors with papers on both sides of November 30, 2022; column 6 uses papers with an author at a US institution with an economics graduate program, with enrollment the year before issue. Business school means at least one author at a business school. Career is the senior author's years since first publication; the h-index is as of October 2026.

Finance accounts for a disproportionate share of the decline. Papers listing Asset Pricing or Corporate Finance, 18 percent of the total, fall 6.5 points (p = 0.002) in the column 1 regression, and all other papers fall 2.5 points (p = 0.01).

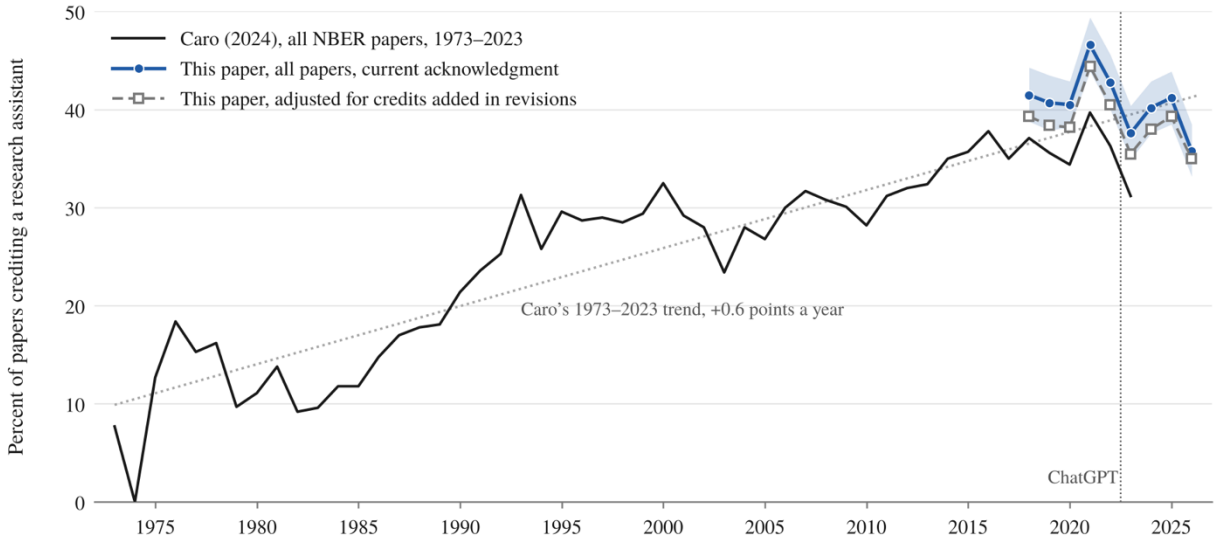
I formed the finance grouping after seeing the program results, so two checks follow, under rules written before they ran (amendment A38). The finance decline is not a return from a high level. Across the 20 programs, the 2025 to 2026 change does not fall with the 2018 to 2022 level (slope +0.17, p = 0.24), and Asset Pricing sits 5.4 points below the fitted line, Corporate Finance 0.3 below. Nor does it look like a business-school budget effect. Among the 2,111 finance papers with an affiliation match, papers with no business-school author fell 5.1 points (SE 4.1) and papers with one fell 6.8 (SE 2.5), a difference of 1.7 (p = 0.72). A budget effect would appear only in the second group. A field effect appears in both, as here, though the first group is small.

A further control does not change the estimate. A theory indicator built from JEL codes and title words flags 11 percent of papers, which credit assistants five points less often, and adding the indicator leaves the post-ChatGPT change at −3.3 points. Theory papers fell more than empirical ones (−7.5 against −2.6), which does not fit a story in which the tools replaced data work first.

Two factors limit how much these results can show. First, the decline does not steepen. In the all-paper census the low years are 2023 and 2026, and 2025 returns to 41 percent before 2026 falls to 36. Most of the research behind a paper issued in 2023 predates ChatGPT, so the 2023 low is hard to attribute to the tools, and 2024 and 2025 sit at the 2018 and 2019 level. Read that way, the case for displacement rests on 2026. A steadily improving tool that displaced research assistants might produce a decline that grows over time, and this one does not. Second, the pandemic disrupted hiring and campus access in 2020 and 2021, so a pre-period that includes those years is not a clean baseline. Appendix Table E2 repeats the before-and-after comparison with other baselines. Dropping 2020 and 2021 changes little. The decline shrinks as the baseline moves away from 2021 and 2022, the highest years in the series. Against 2018 and 2019 alone, the all-paper census falls 2.5 points ($p = 0.04$) and the version-corrected census 2.1 ($p = 0.08$). The first versions fall 3.1 points ($p = 0.02$), and the same authors' rate falls 4.0 points and stays significant against every baseline. The regression with author fixed effects and controls (Table 3, column 5) falls 2.7 points against 2018 and 2019 ($p = 0.03$). Without 2020 and 2021, the all-paper trend is minus 0.43 points a year ($p = 0.025$).

Three facts put the change in perspective. First, against the 2018 to 2022 baseline, the before-and-after drop of three and a half to four points is about 9 percent of the earlier level. Second, the levels match earlier work. Caro (2024) finds that about 35 percent of NBER papers credited an assistant over 2013 to 2023, and this paper's shares for 2018 to 2023 run 38 to 47 percent and rise and fall in the same years as his (Figure 5). His series fell by similar amounts around 1980 and 2003 before rising again, so a few years of data cannot tell whether the long rise has ended. Third, the fitted 2018 to 2026 trend rules out a steep average decline. Its interval excludes a fall faster than 0.85 points a year, about seven points from 2018 to 2026.

Figure 5. Share of NBER working papers crediting a research assistant since 1973, Caro (2024) and this paper



Notes: Caro's series is read from his Figure 1, Panel (a) (Caro 2024), which reports no annual values; digitized values are accurate to about a quarter of a point. Caro counts papers whose acknowledgments credit someone as a research assistant. This paper's series cover January 2018 to October 13, 2026; the band is a 95 percent interval. In the six years both cover, the two move together (correlation 0.94) and this paper's level is 3.5 to 6.1 points higher. Dotted vertical line: ChatGPT release, November 30, 2022.

The broader measure of research support, which counts any non-author credited for research work beyond comments, was about 48 percent of papers in 2018 and 2019 and between 43 and 49 percent in 2023 through 2025 (trend minus 0.4 points a year, $p = 0.11$, or 0.15 after adjustment). The trend in research-assistant credits is smaller, so part of the movement comes from other kinds of support.[6]

4.2 Number of human research assistants credited

Table 4 reports the mean number of distinct named research assistants per paper, counting papers with no credit, and credits that name no one, as zero. The mean ran between 1.08 and 1.27 from 2018 through 2023 and rose to 1.29 in 2024. Among papers with a credit the median is two in every year. The 2024 rise comes from 26 papers that name ten or more assistants, one of which names 85, and without those papers the 2024 mean is 0.91. The trend is plus 0.035 named assistants a year ($p = 0.013$, or 0.033 after adjustment), but winsorizing at the 99th percentile halves it to 0.018 ($p = 0.078$), so the small rise rests on the few papers that name many assistants.

[6] The first-version trend changes little across definitions and samples. The wider definition of a research-assistant credit, which adds predoctoral fellows and research associates credited in that role, gives minus 0.1 points a year ($p = 0.64$). Counting ambiguous cases as credits gives minus 0.2 ($p = 0.37$), and using only the first four months of each year gives plus 0.2 ($p = 0.69$). Adjusting for the number of authors, clustering standard errors on the first author, and restricting each year to January 1 through October 13 leave the trend between minus 0.2 and minus 0.3 points a year, with p between 0.29 and 0.43. The 2026 cell for the broader support measure is 27 of 52 papers, all from the PDFs NBER served.

Authors rarely say what their assistants did; in every year from 2018 through 2025, 95 to 98 percent of credits leave at least one assistant's task unstated.

Table 4. Mean named research assistants per paper, first-version sample

Year	n	Mean, all papers	SE	Mean, papers with a credit	Median, papers with a credit
2018	1,220	1.09	0.06	2.75	2
2019	1,165	1.08	0.06	2.79	2
2020	1,689	1.12	0.07	3.02	2
2021	1,287	1.27	0.07	2.90	2
2022	1,171	1.21	0.07	3.00	2
2023	1,181	1.12	0.07	3.18	2
2024	1,290	1.29	0.11	3.46	2
2025†	445	1.47	0.15	3.53	2
2026†	53	1.02	0.27	3.18	2

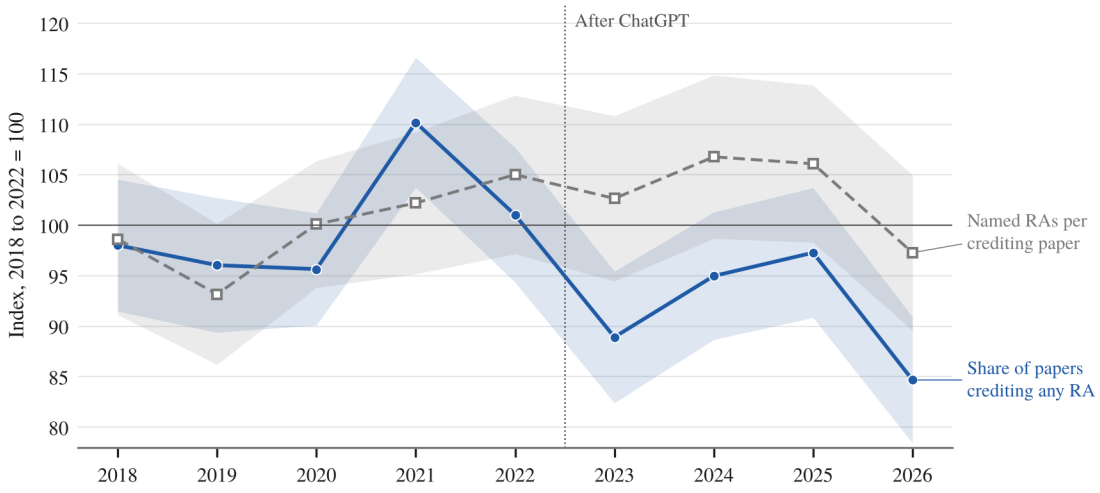
Rows marked † cover only the 2025 and 2026 papers whose PDFs NBER served (Appendix A).

The decline comes from fewer papers crediting any assistant. Papers that still credit one name as many as before (Figure 6). Relative to its 2018 to 2022 average, the share crediting an assistant fell to 89 percent of that average in 2023 and 85 percent in 2026, and the mean number named by crediting papers stayed between 97 and 107 percent of its own average in every year after ChatGPT.

That pattern matches the hiring margin in Brynjolfsson, Chandar, and Chen (2025). Among young workers in exposed occupations, employment fell through reduced hiring rather than separations, and here the change is in whether a paper credits an assistant at all rather than in how many it names. It cuts against the version of the displacement story in which one assistant with a tool does the work of three, a version the experimental evidence of large gains for the least experienced workers might suggest (Brynjolfsson, Li, and Raymond 2023). It differs from their series in one respect. Their gap widened with each release of the data, from 13 to 15 to 19 percent, while this series does not steepen through 2026.

The assistants still credited do not appear on more papers each. When I match the 15,087 names on credited cover blocks within each issue year, a named assistant appears on 1.10 NBER papers a year in 2018 through 2022 and 1.08 in 2023 through 2026, and the share named on two or more papers in a year falls from 8.3 to 6.8 percent. In the finance, macroeconomics, and methods programs, where the tools would raise an assistant's output first, papers per assistant fall from 1.10 to 1.06. If assistants were using the tools to do more, the same names would spread across more papers, and they do not. I match names only within NBER papers, so an assistant's other work is invisible, and 2025 and 2026 cover blocks will gain names as authors revise them.

Figure 6. Share of papers crediting a research assistant and named assistants per crediting paper, indexed to 2018 to 2022



Notes: All-paper census, each series divided by its 2018 to 2022 average. The share crediting any assistant and the mean number of named assistants among crediting papers (winsorized at the 99th percentile); their product is named assistants per paper. Bands are 95 percent intervals. Counts exclude credits that name no assistant or whose allocation is unclear. 2026 runs through October 13.

Table 5 repeats the regressions in Table 3 with the number of named assistants as the outcome. No column shows a significant post-ChatGPT change. Larger teams and revised papers name more assistants, and papers whose most senior author has been publishing longer name fewer, about 10 percent fewer for each ten years of career ($p < 0.001$).

Table 5. Correlates of the number of research assistants named, all-paper census

	(1)	(2)	(3)	(4)	(5)	(6)
Model	Poisson	Poisson	Poisson	Poisson, no finance	Poisson	Poisson
Post-ChatGPT	−2.7 (0.47)	−1.2 (0.81)	−0.5 (0.95)	0.2 (0.98)	−5.1 (0.15)	0.3 (0.94)
Business school		12.0 (0.02)	8.6 (0.12)	12.0 (0.05)		
Business school × post		−2.1 (0.79)	0.2 (0.98)	0.7 (0.94)		
Senior author's career, per ten years			−10.0 (<0.001)	−12.1 (<0.001)		
Career × post			−0.2 (0.94)	0.5 (0.89)		
Log(1 + highest h-index)			−0.7 (0.85)	4.7 (0.25)		
Log economics graduate enrollment						−5.4 (0.67)
Log number of authors	37.1 (<0.001)	35.6 (<0.001)	40.1 (<0.001)	42.3 (<0.001)	19.7 (<0.001)	27.0 (<0.001)
Revised	48.5 (<0.001)	48.3 (<0.001)	45.5 (<0.001)	48.0 (<0.001)	45.4 (<0.001)	45.5 (<0.001)
Program indicators	Yes	Yes	Yes	Yes	No	Yes
Fixed effects					Author	Institution
Observations	11,284	11,257	10,450	8,509	19,374 pairs	8,695

Notes: Poisson pseudo-maximum likelihood coefficients times 100 (approximately percent), with p-values in parentheses. The outcome is the number of distinct named research assistants, zero for papers with no credit; papers whose credit names no clear number are excluded. Samples, controls, and clustering follow Table 3. Added at readers' request (amendment A34).

4.3 AI statements

Two instruments measure AI statements. The first searches the full text of first versions and counts a statement anywhere in the paper; the second reads only the cover block of every paper. Of the 158 statements in first-version full text, at least 12, about one in thirteen, also appear in the cover block, which bounds the overlap from below. The two sources also differ in vintage, since the full text is the first version and the cover block the current one.[7]

Table 6 and Figure 7 report AI-use statements by year. In the full text of first versions, statements were rare through 2022 and then rose to about 8 percent of the 2025 papers NBER

[7] The study had full text for every paper through 2024 but for only 449 papers in 2025 and 54 in 2026, because NBER serves recent PDFs only to subscribers (Appendix A). For 2025 and 2026 the study therefore observes AI statements mostly through the cover block, so the AI-use rates it reports for those two years are understated relative to earlier years. The 2026 methods-series count (8 of 28) is a cover-block count and a floor for the same reason. Table 6 reports both instruments.

served. The upward trend survives the multiple-testing adjustment. Most statements describe a model used to analyze data, such as an embedding model or a language model that classifies text. Only 24 of the 158 describe an assistant used for tasks such as coding or editing, the use that bears on substituting for a research assistant, and that kind of statement never reaches 2 percent of papers in any year.

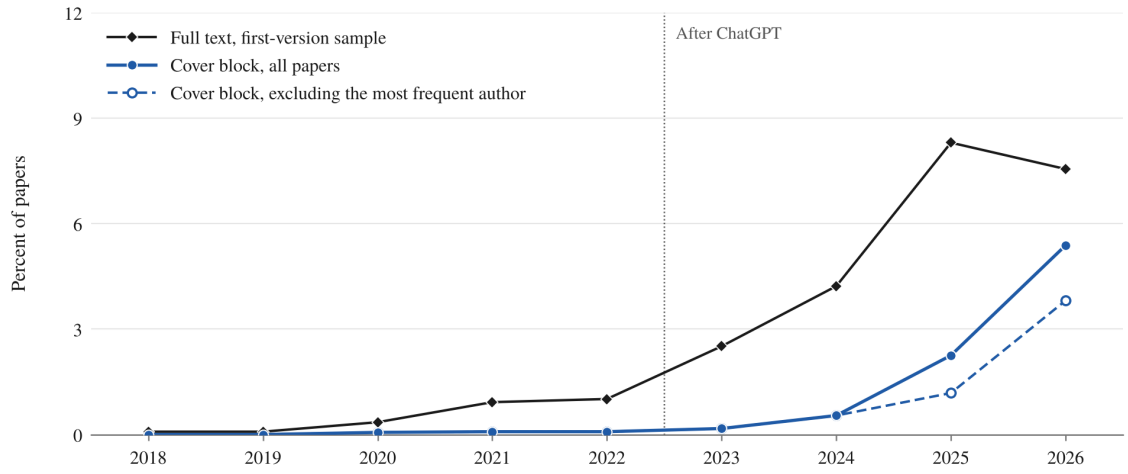
Cover blocks show the same rise at a lower level, reaching about 5 percent of 2026 papers. One author, Joshua Gans, appears on about a third of the 2026 cover-block statements. Without his papers the rise is smaller but still present. He accounts for 34 of the 37 statements that thank an AI tool "for research assistance," the wording authors otherwise use for people, so that phrasing is one author's practice, and his practice bears on the credit series too. He credits a research assistant on 4 percent of his 83 papers in the frame, and 38 of them appeared in 2025 and 2026, so a prolific author who rarely credits an assistant and posts more papers lowers the share crediting an assistant without displacing anyone. He has described 2025 as a year of AI-first research, in which he moved a backlog of ideas to posted working papers as quickly as he could, without seminars or other peer discussion along the way (Gans 2026). Without his papers the controlled post-ChatGPT change in credits is −3.1 points instead of −3.3, and the within-author change stays at −4.1. These AI-statement rates count only authors who chose to say they used AI.

Table 6. Share of papers with an author-reported AI-use statement, by instrument

Year	Full text: n	Disclosures	Share (%)	Cover block: n	Disclosures	Share (%)	Cover block, without most frequent disclosing author (%)
2018	1,233	1	0.1	1,236	0	0.0	0.0
2019	1,182	1	0.1	1,185	0	0.0	0.0
2020	1,708	6	0.4	1,713	1	0.1	0.1
2021	1,308	12	0.9	1,311	1	0.1	0.1
2022	1,187	12	1.0	1,191	1	0.1	0.1
2023	1,195	30	2.5	1,197	2	0.2	0.2
2024	1,304	55	4.2	1,307	7	0.5	0.5
2025	446	37	8.3	1,289	29	2.2	1.2
2026	53	4	7.5	1,282	69	5.4	3.8

Full text is the first-version sample (n counts papers with a clear AI code, so it differs slightly from Table 2's n); cover block is the current cover block of every paper with a clear code. The last column drops every paper by the one NBER author who appears on the most AI-use statements (36 of the 250 papers with an AI-use statement in either source), from both the count and the denominator. The cover-block columns include the first-version papers, so the two instruments overlap, and adding them would double-count. Figure 7 plots the two instruments.

Figure 7. Author AI-use statements by instrument



Notes: Share of papers with a statement that the authors used an AI tool. Full text sees statements anywhere in the paper but covers only 446 papers in 2025 and 53 in 2026; the cover block covers every paper and captures a fraction of what the full text shows, so recent years are floors. 2026 runs through October 13. Dotted line: ChatGPT release, November 30, 2022.

Adding an indicator for an AI-use statement to the column 1 regression in Table 3 leaves the post-ChatGPT estimate unchanged, and its own coefficient is small and imprecise (2.9 points, $p = 0.35$). That coefficient is a correlation within one choice. Authors decide in the same acknowledgment whether to thank an assistant and whether to disclose a tool, and norms about both are changing together.

Table 7 relates an AI-use statement in the cover block to the variables in Table 3. It uses linear probability models, because statements are almost absent before ChatGPT and a logit with program indicators cannot be estimated. The share of papers with a statement rises 2.1 points after November 2022 ($p < 0.001$), from 0.05 percent before, and 1.5 points within authors ($p < 0.001$). Papers in the methods series are 2.8 points more likely than others to carry one ($p = 0.02$). Business-school papers show a larger rise only without finance papers (1.9 points, $p = 0.003$; otherwise 0.7 points, $p = 0.10$). The career interactions depend on one author. With every paper included, statements rose more on papers with more senior authors (0.6 points per ten years of the senior author's career, $p = 0.001$); without the papers of the author on the most statements, they rose less (−0.2 points, $p = 0.08$).

Table 7. Correlates of an AI-use statement in the cover block, all-paper census

	(1)	(2)	(3)	(4)	(5)	(6)
Sample	All	All	All	No finance	Same authors	Without top discloser
Post-ChatGPT	2.11 (<0.001)	1.88 (<0.001)	0.06 (0.92)	0.09 (0.89)	1.49 (<0.001)	2.20 (<0.001)
Business school		0.19 (0.03)	0.24 (0.010)	−0.12 (0.26)		0.12 (0.15)
Business school × post		0.71 (0.10)	0.75 (0.10)	1.87 (0.003)		−0.36 (0.30)
Senior author's career, per ten years			0.18 (0.002)	0.23 (<0.001)		0.08 (0.14)
Career × post			0.64 (0.001)	0.65 (0.003)		−0.24 (0.08)
Log(1 + highest h-index)			−0.44 (0.03)	−0.55 (0.02)		−0.17 (0.38)
Technical Working Papers	2.78 (0.02)	2.87 (0.01)	2.76 (0.02)	2.97 (0.02)		2.86 (0.02)
Log number of authors	−1.40 (<0.001)	−1.46 (<0.001)	−1.60 (<0.001)	−1.81 (<0.001)	−0.43 (0.21)	−0.18 (0.40)
Revised	0.66 (0.002)	0.64 (0.002)	0.67 (0.003)	0.83 (0.003)	0.33 (0.18)	0.41 (0.03)
Share before ChatGPT (%)	0.05	0.05	0.03	0.04	0.07	0.03
Program indicators	Yes	Yes	Yes	Yes	No	Yes
Fixed effects					Author	
Observations	11,711	11,683	10,843	8,853	23,318 pairs	10,764

Notes: Linear probability models; estimates in percentage points, with p-values in parentheses. The outcome is a statement in the cover block that the authors used an AI tool, which every paper has. Columns 1 to 6 follow Table 3, except that column 6 drops the papers of the author on the most AI-use statements in place of using institution fixed effects. P-values come from robust standard errors, clustered by author and by paper in column 5. The share before ChatGPT is the sample's mean before December 2022.

4.4 Fields

The program comparisons are descriptive. Papers count in every program they list, and the tests do not adjust for the 20 comparisons. Figure 2 and Appendix Table E1 compare each program's share in 2025 and 2026 with its 2018 to 2024 level. The largest declines are in the methods series (Section 4.5), Asset Pricing, and Monetary Economics. Asset Pricing, Monetary Economics, Economic Fluctuations and Growth, and Public Economics each fell significantly on their own, and Asset Pricing's stays significant after a Bonferroni correction for testing 20 programs, in this study's codes and in the ChatGPT study's (Appendix C). None of the four shows an unusual rate of AI-use statements in 2026.

The first-version sample cannot confirm these declines because it holds few 2025 and 2026 papers.[8]

Regressions with program × post interactions and controls for team size and revision point to the same programs over the whole period after November 2022 (Table 8). Corporate Finance (−8.4 points, $p = 0.002$), Monetary Economics (−7.6, $p = 0.004$), Asset Pricing (−6.6, $p = 0.01$), and Public Economics (−4.9, $p = 0.009$) fall by significant amounts on their own; Corporate Finance passes a Bonferroni threshold for 20 tests ($p < 0.0025$), and Monetary Economics comes closest. AI-use statements rose in nearly every program. Apart from the methods series, the programs with the largest declines in credits are not the ones with the largest rises in statements.

[8] None of the first-version sample's 20 program trends survives the adjustment for multiple comparisons, and programs with large changes on one measure, such as Corporate Finance and International Trade, show little change on the other. Industrial Organization, Political Economy, and Health Economics are flat or rising on every measure. Programs that started high did not fall more than others, so the declines are not reversion toward the average. The study's repository reports the same comparison for 2018 to 2022 against 2023 to 2026 and by JEL letter.

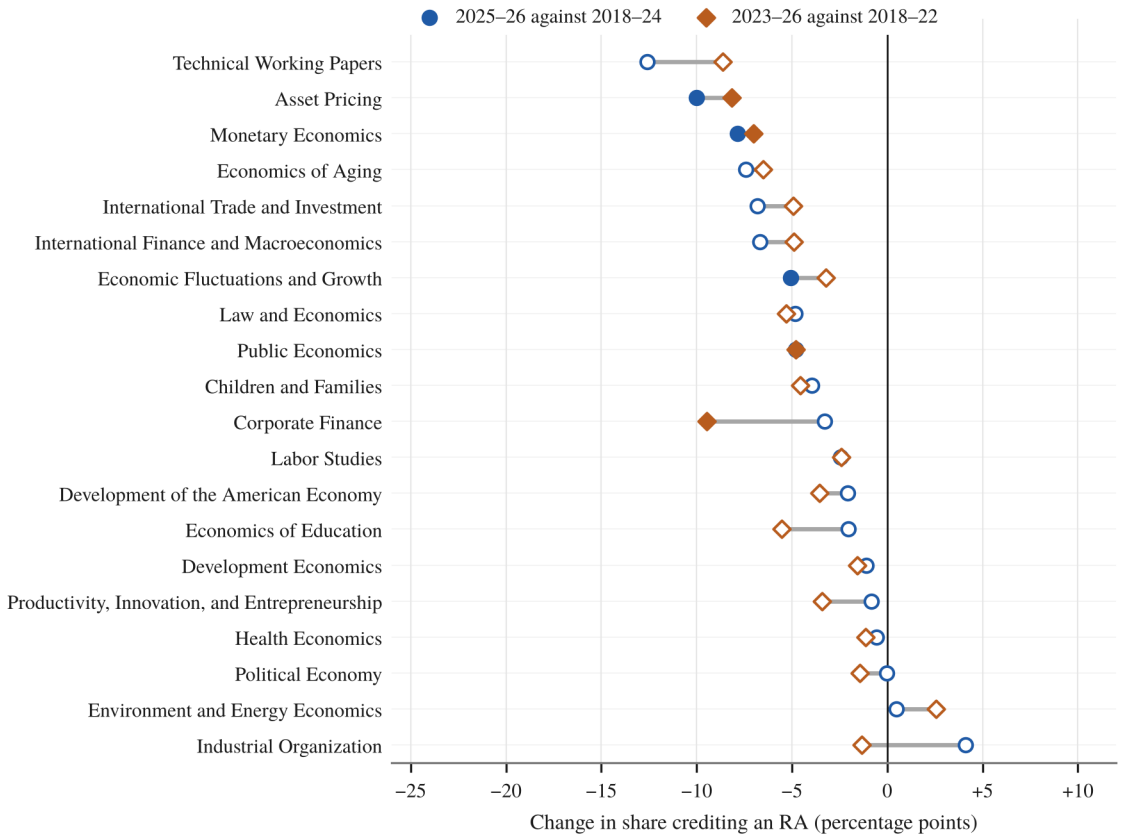
Table 8. Change after ChatGPT by NBER program, research-assistant credits and AI-use statements

Program	Papers	Research-assistant credit	AI-use statement
Corporate Finance	1,335	−8.4 (0.002)	0.7 (0.05)
Monetary Economics	1,347	−7.6 (0.004)	1.0 (0.02)
Asset Pricing	1,307	−6.6 (0.01)	0.4 (0.13)
Technical Working Papers	248	−6.3 (0.28)	9.1 (0.001)
Economics of Education	993	−5.3 (0.09)	1.0 (0.05)
Aging	742	−5.1 (0.15)	1.9 (0.01)
Public Economics	2,830	−4.9 (0.009)	0.9 (<0.001)
International Finance and Macroeconomics	1,124	−4.7 (0.11)	0.5 (0.10)
International Trade and Investment	992	−4.4 (0.17)	2.0 (0.001)
Development of the American Economy	646	−3.9 (0.31)	0.7 (0.13)
Children	1,088	−3.5 (0.24)	0.5 (0.20)
Law and Economics	667	−3.1 (0.42)	0.5 (0.12)
Economic Fluctuations and Growth	2,451	−2.7 (0.17)	0.8 (0.004)
Labor Studies	2,934	−2.0 (0.27)	1.6 (<0.001)
Productivity, Innovation, and Entrepreneurship	1,537	−1.6 (0.51)	1.9 (<0.001)
Industrial Organization	1,097	−1.3 (0.66)	2.1 (0.001)
Health Economics	1,777	−1.0 (0.68)	0.8 (0.01)
Development Economics	1,406	−0.3 (0.92)	3.2 (<0.001)
Political Economy	1,098	0.2 (0.96)	1.0 (0.02)
Environment and Energy	936	2.2 (0.48)	1.2 (0.04)

Notes: Each cell is the average change after November 30, 2022, for the program's papers, in percentage points with p-values in parentheses, from one regression with indicators for all 20 programs, their interactions with the post-ChatGPT indicator, the log number of authors, and revision. Papers count in every program they list, and a paper's program interactions switch together. Research-assistant credits use a logit and the all-paper census. AI-use statements use a linear probability model and the cover block, without the papers of the author on the most statements. P-values are unadjusted; the Bonferroni threshold for 20 programs is 0.0025.

The program comparisons hold under a second window (Figure 8). Asset Pricing falls 10.0 points comparing 2025 and 2026 with 2018 through 2024 and 8.2 points comparing 2023 through 2026 with 2018 through 2022, and Monetary Economics falls 7.9 and 7.0. Corporate Finance is the exception, falling 3.3 points in the first comparison and 9.5 in the second.

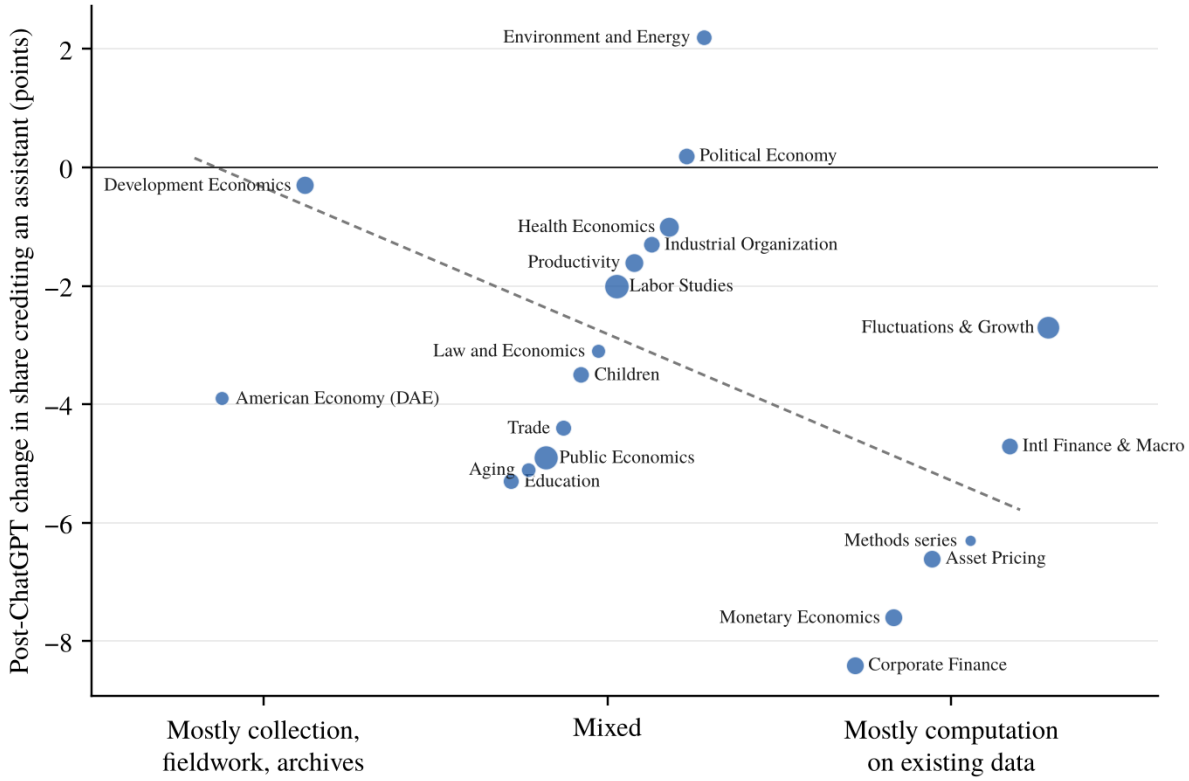
Figure 8. Change in the share of papers crediting a research assistant by NBER program under two windows



Notes: All-paper census. Each row shows a program's change in the share of papers crediting a research assistant under two windows; a long bar means the result depends on the window. Filled markers have 95 percent intervals that exclude zero, unadjusted for 20 programs. Papers count in every program they list. Technical Working Papers is NBER's econometrics and methods series. Post hoc.

If the tools displace the coding and data-handling work an assistant does against existing data, the decline should concentrate in programs where that is most of the work and be smallest where assistants run surveys, field projects, or archival collection, work no model does. Whether automating a task displaces the person who did it depends on whether the task is the whole job or one input to a judgment that person still makes (Agrawal, Gans, and Goldfarb 2019). I score each program from 1 (mostly collection, fieldwork, or archives) to 3 (mostly computation on existing data) and plot the Table 8 changes against the score (Figure 9). Programs scored 3, the finance and macroeconomics programs and the methods series, fell 5.6 points on average, weighted by papers; programs scored 2 fell 2.5; and the two scored 1, Development Economics and Development of the American Economy, fell 1.4. The weighted slope is −2.5 points per step ($p = 0.007$). The scores are a judgment I made after seeing the program results, and programs share papers, so the gradient is a lead rather than a test, and it does not depend on the methods series alone.

Figure 9. Post-ChatGPT change in the share of papers crediting a research assistant, by how much of a program's assistant work is computation on existing data



Each point is an NBER program (size = papers); change from the Table 8 regression. Dashed line: paper-weighted fit, slope -2.5 points per step (p = 0.007). Scores assigned after the program results were known; see text.

Paper-level estimates weaken the gradient and show that credits and AI-use statements do not move together. When I treat the two as responses to the same shock and interact the post-ChatGPT period with each paper's exposure score, with program indicators and standard errors clustered by first author, the credit gradient is −2.2 points per step (SE 1.9, p = 0.26), with no pre-trend (2018 to 2021 interactions jointly p = 0.88). The program-level slope gets its precision from treating 20 overlapping programs as independent. AI-use statements rose less in exposed programs (−1.5 points per step, p = 0.01), because most statements describe a model applied to data, which is commonest in development, labor, and political economy; statements of help with coding, drafting, or editing show no gradient. Across the 19 programs other than the methods series, credits fell least where statements rose most. Across the 3,659 authors with papers of known AI-statement status in both periods, an author's change in AI statements predicts nothing about the author's change in credits (p = 0.93), and the 101 authors whose statements rose cut credits by 4.4 points against 5.2 for everyone else. Outside the methods series, the decline in credits is not where the disclosed AI use is.

4.5 Technical Working Papers

NBER's "Technical Working Papers" cover econometric, statistical, and computational methods, although I could not find a published definition. NBER issued technical papers as a separate numbered series from 1979 to 2007 (about 345 papers) and now publishes all papers in its main

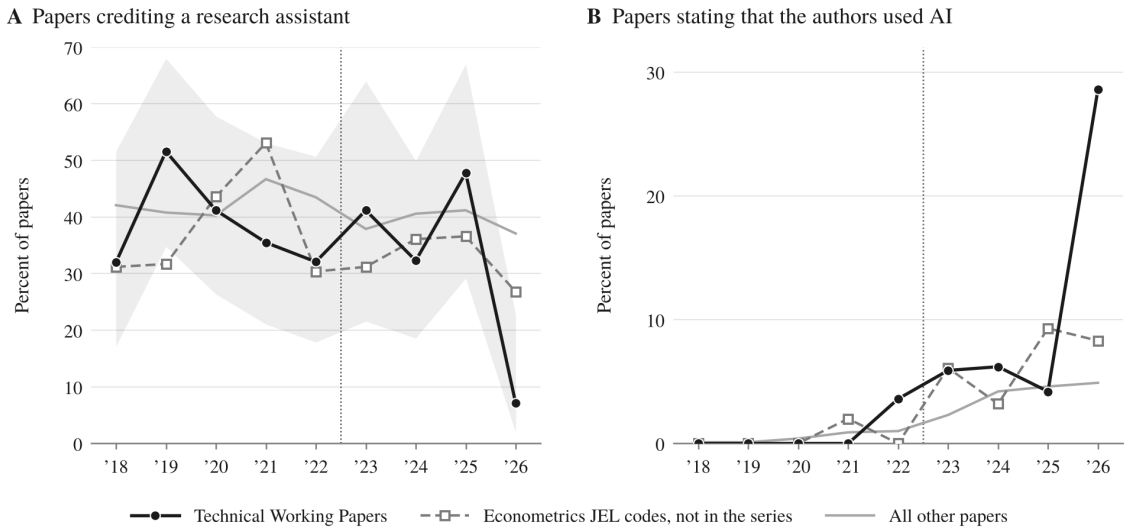
series (NBER n.d.; CitEc 2022). The label survives as a program tag on 250 of the 11,718 papers in the frame (248 with a clear credit code), 17 to 34 a year, and 199 of the 250 also list a substantive program. Methods papers generally do not require fieldwork, surveys, or new data collection, so if AI were displacing research assistants anywhere first, it would plausibly be here. The acknowledgments rarely state the assistant's task (76 of the 78 credited methods papers in the first-version sample leave at least one assistant's task unstated), so the paper cannot say what these assistants did.

The methods series credited a research assistant on 29 to 52 percent of its papers every year from 2018 through 2025, and on 2 of 28 papers in 2026. That is far below the 36.5 percent of other 2026 papers. Three cautions apply, and I state them here once. The cell holds 28 papers. The 2026 cohort is incomplete and largely unrevised, and later revisions would add about one credit. And I found the cell by looking across all 180 program-by-year cells; the gap from the other 2026 papers (Fisher $p = 0.001$) is not significant after that search (0.18), although under a program-plus-year null the expected number of cells this extreme among 180 is about 0.03. The authors of these papers have not stopped crediting assistants elsewhere. They credited assistants on about half of their earlier papers and on 39 percent of their other 2026 papers.

The same 28 papers also have the highest rate of AI-use statements of any program in any year, 8 of 28. They come from five groups of overlapping authors, none of which includes the author who accounts for most of the statements thanking a tool for research assistance. All eight statements appear in the cover block, where the codebook did not record the kind of use. Read afterward, six describe help with tasks such as suggestions, drafting, coding, or checking for consistency, one says the authors used models in the analysis and writing, and one does not say. None describes a model applied to data as part of the design, the use that dominates the full-text statements in Section 4.3. The link holds only at the group level, since the papers that state AI use are no more likely than the others to have dropped an assistant credit.

The regressions show the same contrast. Papers in the series are 2.8 points more likely than other papers to carry an AI-use statement ($p = 0.02$; Table 7), and over the whole period after November 2022 their statements rose 9.1 points ($p = 0.001$) while their research-assistant credits fell 6.3 points ($p = 0.28$; Table 8). With standard errors clustered by first author instead of by paper, those p-values are 0.06, 0.009, and 0.29.

Figure 10. The methods series against methods-adjacent and other papers: research-assistant credits and AI-use statements by year



Notes: Current cover-block text, all 11,718 papers; AI statements add full text where retrieved, so 2025 and 2026 are floors. Shaded band: 95 percent interval for the Technical Working Papers, which number 16 to 34 a year. Dotted line: ChatGPT release, November 30, 2022.

The evidence does not yet show whether the methods series is unusual or just first. Econometrics papers outside the series moved in the same direction, but by less, and all other papers moved less still (Figure 10). The difference between the series and all other papers, 2025 and 2026 against 2018 to 2024, is −9.9 points, with a 95 percent interval from −24.0 to +4.2. On one reading, the methods series is a canary in the coal mine. It may be the first place the tools could do an assistant's whole job, with the rest of economics to follow as the tools reach less purely computational data work. On another, methods work differs from the rest of empirical economics, where assistants gather data, run surveys, and manage field projects that no model does, and the series may remain an exception. On a third, the cell is an unusual draw, and the next cohort will look ordinary. Later cohorts, whose research could have used capable agentic tools, may separate these readings.

Two further checks bear on these readings. Outside the methods label, papers that are code-intensive by JEL code (C1, C6, or C8) or by method words in the title, 479 of 11,299, fell no more than other papers in 2025 and 2026 (interaction +1.6 points, SE 5.4; rule written before the test, amendment A36), so the pattern did not reach methods-heavy work elsewhere. Amendment A37 also records, before the papers exist, what the methods series should show from October 14, 2026 through June 2027 under each reading. Displacement implies a credit share below 20 percent with an AI-statement share above 15 percent, and an unusual draw implies a credit share of 30 percent or more with statements at 10 percent or less.

4.6 Affiliation and seniority

Business-school authors credit assistants more often, and their decline after ChatGPT does not differ detectably from other authors'. Papers with at least one business-school author, 36 percent of those with a known affiliation, credit an assistant 3.9 points more often ($p < 0.001$; Table 3, column 2) and name about 12 percent more assistants ($p = 0.02$; Table 5, column 2). The

difference in differences is −0.8 points (p = 0.67), −0.4 with controls for seniority, and +1.2 (p = 0.60) without finance papers, which column 4 drops because most finance faculty work in business schools. With author fixed effects it is −2.6 points (p = 0.17), which points toward a larger decline at business schools but is not distinguishable from zero. A classifier on the printed affiliation strings identifies business schools; a hand check of 25 flagged papers found one error, and the measure leaves out institutions that combine an economics department with a business school, such as Bocconi.

OpenAlex covers 68 percent of 2022 papers, against 85 to 95 percent in other years, and the missing 2022 papers credit assistants at a high rate. I fill a missing paper's business-school status from its authors' other papers, which brings coverage to 11,690 papers; without the fill, the post-ChatGPT estimate in column 2 is −2.0 points. Dropping finance papers in column 4 shrinks the estimate to −1.3 points (p = 0.23).

Grants offer a second budget measure. A third of cover blocks name a funder, including a US federal agency on 18 percent, a private foundation on 12 percent, and a non-US public funder on 9 percent. Funded papers credit assistants more often; before ChatGPT, 52 percent of foundation-funded papers did, against 41 percent of the rest. Funded papers did not fall more after ChatGPT (relative to unfunded papers, +1.6 points for federal funding, p = 0.51, and +4.3 for foundations, p = 0.14). Federal acknowledgments themselves fell, from 21 percent of papers in 2018 to 13 percent in 2026, most of the fall in 2025 and 2026, but federally funded papers held up slightly better than others in those years (+2.8 points, p = 0.36), so cuts in federal research funding do not appear to explain the decline in credits.

Papers whose most senior author has been publishing longer credit assistants less often. Before ChatGPT, papers whose most senior author had been publishing for under ten years credited an assistant 50 percent of the time, against 37 percent for papers whose senior author had been publishing for 30 years or more. Each ten years of the senior author's career lowers the probability of a credit by 2.9 points (p < 0.001; Table 3, column 3) and the number named by about 10 percent (p < 0.001; Table 5, column 3). I measure career age as the years from an author's first publication in OpenAlex to the paper's issue year. OpenAlex assigns new author IDs to recent papers it has not yet merged with an author's record, so on 2026 papers many established authors at first appeared new. I link IDs across an author's papers where the surname matches, treat an ID with five or fewer works whose first publication is within a year of the paper as unknown rather than young in every year, and drop ages above 60 as misattributed records; 14 percent of 2026 author records are unknown, against 2 percent in earlier years. The h-index, which OpenAlex reports as of October 2026 and which the same problem understates for recent papers, adds nothing once career age is in the model (p = 0.17).

The decline is larger on papers with younger senior authors, but not significantly. Each ten years of the senior author's career shrinks the post-ChatGPT decline by 0.6 points (p = 0.40; Table 3, column 3), and the number named shows no such pattern (Table 5, column 3).

The decline concentrates among younger senior authors working alone, a pattern I found in the table of career age by team size. Among papers whose senior author has been publishing for under 20 years, single-authored papers fell 14 points (SE 6, 327 papers, p = 0.02), two- and three-author papers fell 4, and papers with four or more authors did not fall. Among more senior authors, single-authored papers did not change. A young author working alone is the case in

which a tool most plausibly replaces an assistant, but the cell is small and later cohorts will have to confirm the pattern.

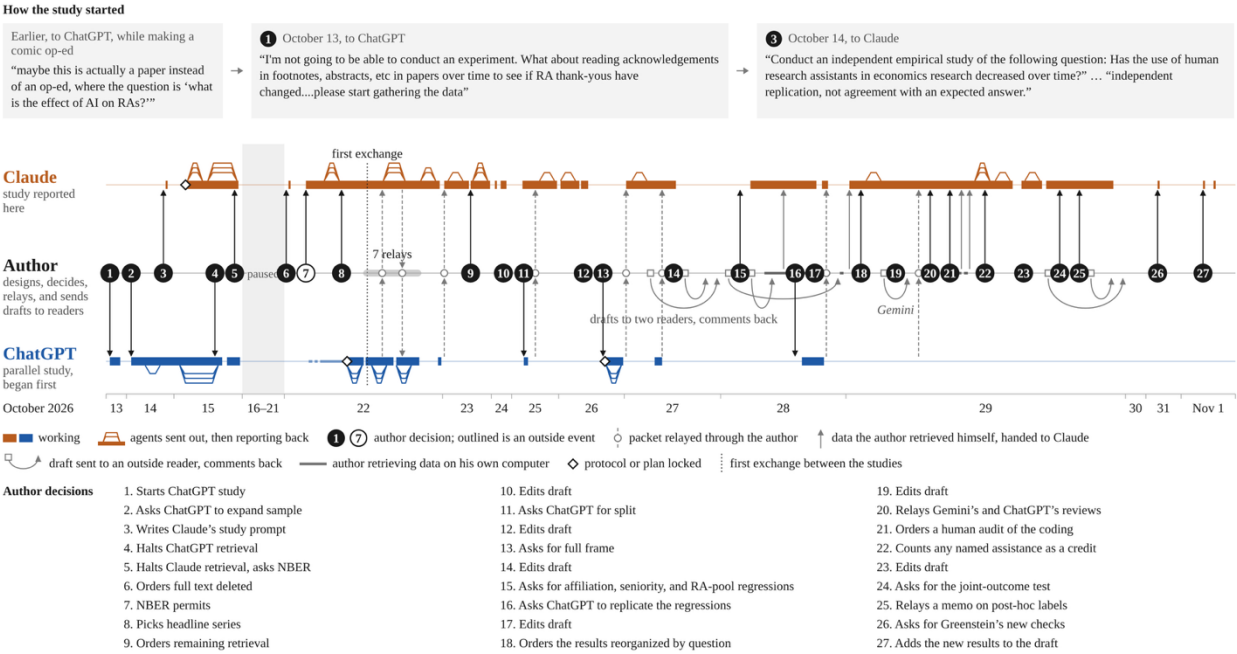
4.7 The pool of potential assistants

The enrollment measure offers no explanation for the decline, though the within-institution estimate is too imprecise to rule one out. National full-time enrollment in economics graduate programs rose from 11,391 in 2018 to 12,699 in 2024. For the 8,935 papers with a clear credit code and an author at a US institution that has an economics graduate program, I measure the pool as full-time economics enrollment at the largest such program in the year before the paper appeared. Across institutions, a paper from a program twice as large is about five points more likely to credit an assistant ($p < 0.001$), a difference that may reflect richer departments more than the supply of students. Within an institution over time, changes in enrollment predict neither a credit (0.5 points per log point, $p = 0.83$; Table 3, column 6) nor the number named ($p = 0.67$; Table 5, column 6). With institution fixed effects the post-ChatGPT estimate in this sample is −2.8 points ($p = 0.03$). Enrollment at an institution changes by only a few percent a year, so the within-institution estimate is imprecise, and the measure misses assistants who work outside universities, at central banks, research centers, or the NBER itself.

5. How this study was produced

The project began in ChatGPT, which was drawing the comic for an op-ed I was writing. Partway through, I decided that the op-ed's question, what AI is doing to research assistants, deserved a paper. On October 13 I asked ChatGPT to start gathering acknowledgment data, and it built the frame of 11,718 NBER papers and began coding a sample. On October 14 I asked Claude for an independent replication, giving it only the frame's paper IDs and selection metadata and instructing it not to consult the other model's notes, code, codes, or results. The Claude study coded every paper and became the analysis this paper reports, and the ChatGPT study became the parallel study in Section 6. Figure 11 traces both studies and my part in them through October 29, including the drafts I sent to two outside readers on October 27, whose comments led to the regressions in Sections 4.6 and 4.7. Appendix Figure D1 gives the full version, with the role and count of every group of agents.

Figure 11. How the two studies were produced



Notes: The horizontal axis keeps the order of events but not their spacing, and October 16 to 21, when work paused while NBER considered the permission request, is collapsed into one band. Each line that leaves a lane stands for agents sent out by that lead session, and it rejoins the lane when they report. Diamonds mark when each study fixed its protocol or analysis plan, and the dotted line marks the first packet passed between them. Arcs on the author's lane show drafts sent to two outside readers and their comments coming back, after which the author asked both studies for the regressions in Sections 4.6 and 4.7 and retrieved the OpenAlex data for the Claude study himself. On October 29 he sent a draft to Gemini for one review, audited 100 cover blocks and adjudicated the codes on which the two studies disagreed, passed both back to the Claude study, and set the further tests the study ran through November 1 as readers' comments came in. The two AI systems never communicated directly; every exchange passed through the author. Appendix Figure D1 gives the agents' roles and counts.

I designed the study, wrote and froze the protocol with the lead session (the Claude session that ran the study under my direction), decided every amendment, and directed the analysis and the writing. Language model systems did the rest, including retrieval under NBER's stated permission, text extraction, locating acknowledgment passages, all coding, blind second coding, negative audits, the literature search and verification, the analysis scripts, the comparison with the parallel ChatGPT study, the referee reports that produced this revision, and the drafting of this text. This is the disclosure the paper counts in other papers, made in the form the paper counts it.

Model coders read every acknowledgment under a codebook frozen before the study coded any paper, and each coder had to match reference codes on calibration items before the study accepted its work. Blind second coding by a different model agreed with the first codes on 97 to 100 percent of research-assistant codes, depending on the sample, and audits that re-read 356 papers coded as having no credit found no missed credit. Every credit and AI statement the study counts carries a verbatim quote checked against the stored text. Appendix D gives the coding and validation detail.

The division of labor has two implications. First, the method is itself a data point about what the tools could do in late 2026. Reading 11,718 acknowledgment blocks and 9,644 full texts, coding them under a codebook, second-coding samples, auditing the negatives, and writing up the result is work that would have occupied several research assistants for months; here it took days, with my time going to design, permission, and judgment.

Second, the design limited the coders' opportunity to favor a result. The coders saw only a paper ID and its acknowledgment text, coded human credits and AI statements under the frozen

codebook, had no access to any outcome share, and produced codes that agree closely with the parallel ChatGPT study's (Section 6). On October 29 I read 100 cover blocks drawn at random from the census, 40 coded as crediting an assistant and 60 as not, against the stored codes. One coded negative credits an assistant ("provided outstanding research assistance"), and the other 99 codes were right. The parallel ChatGPT study, which coded every cover block from the same text, disagrees with this study on 14 papers, 11 of which it codes as credits that this study missed, including the one my audit found, and 3 the other way. It also codes as no credit most of a further 149 papers that this study coded ambiguous. I then read all 163 of those cover blocks (amendment A35). The 14 disagreements all resolved in the ChatGPT study's favor on the 11 and in this study's favor on the 3, so all 14 are credits. The corrected codes enter every number in this paper; they move the share crediting an assistant by at most 0.3 points in any year and the regression estimates by at most 0.1 points. Of the 149 ambiguous papers, I judged 104 to credit an assistant, almost all for help with data, coding, or a specific task described without the words "research assistance," 20 not to, and 24 to remain ambiguous. The protocol defined a credit by its wording; after reading those cover blocks I decided that any named assistance with the research counts, and the series in this paper use that definition (amendment A35). The decision raises the before-and-after decline on the all-paper census from 3.7 points to 4.0, the regression estimate with controls from −2.9 points to −3.3, and the within-author estimate from −3.7 to −4.1, because the ambiguous cases are concentrated in 2018 to 2021 (26 to 52 a year, against 12 in 2025 and 2 in 2026). Under the protocol's wording rule, with only the 14 disagreements corrected, every conclusion in the paper holds with those smaller numbers. Two independent coders and a human reading therefore bracket the error in the credit code. The miss rate among coded negatives is about 1 percent in the random audit and 0.2 percent between the two studies, which shifts a 40 percent share by less than a point and cannot move the trend unless it changes over time. Table 2 and Table 4, which come from the first-version codes, are unaffected. The same point applies to the counts of AI-use statements, the one field in which one might think a model has a stake. Those counts agree between the two systems, and a referee model found the one concentration the paper reports (Section 4.3), which the study then verified against the stored evidence quotes.

6. The parallel ChatGPT study

The ChatGPT study began a day before this one and built the frame of papers both studies use (Section 5). It ran independently, with its own coding rules and analysis plan, and at my request it coded the current cover block of all 11,708 papers it could read. The two studies disagree on 14 of the 11,271 papers that both coded as crediting or not crediting an assistant, and on 1 of 11,663 AI codes. Its shares are within half a point of this study's in every year. With 2023 in the post period it finds a before-and-after decline of 3.3 points (SE 0.9), against 4.0 here, and on its own window, which omits 2023, a decline of 2.7 points (SE 1.0), against 3.5. It ranks the same programs at the bottom, led by Technical Working Papers, where 2 of 28 papers in 2026 credit an assistant. Because both studies read every paper, sampling error plays no part in the comparison, and the small remaining differences come from coding rules. Agreement between two model systems shows that the coding is reproducible. Only a human audit can show whether it is accurate (Section 5). Appendix C reports the comparison in detail, including agreement rates before and after the ChatGPT study received this study's codes.

The ChatGPT study replicated the regressions in Section 4 independently, from its own codes, its own OpenAlex and NSF downloads, and its own classifier and institution matching. Its estimates are close to this study's. Its logit post-ChatGPT change is −2.6 points (SE 0.9), against −3.3 here on the adjudicated series and −2.9 under the protocol's wording rule; with author fixed effects it is −3.6 (SE 1.0), against −4.1 and −3.7. Its business-school difference in differences is +0.2 points (SE 2.1), against −0.8, and −2.3 (SE 2.2) with author fixed effects, against −2.6. Its within-institution pool coefficient is 0.9 (SE 2.7), against 0.5. It also finds no fall in the number of named assistants per crediting paper, and its career-age estimates show smaller declines on papers with more senior authors, none significant.

7. Limitations

The outcome measure is a printed sentence. It responds to norms about whom to thank, to NBER's acknowledgment template, and to whether authors remember the assistant when they write, as well as to the assistance itself. A flat series is consistent with stable assistance and with falling assistance offset by a rising propensity to credit it. A small decline is consistent with falling assistance and with a falling propensity to credit it.

The frame is the NBER series, which selects on affiliation. The first-version sample covers every paper from 2018 through 2024 but only 449 of the 1,290 papers from 2025 and 54 of the 1,287 from 2026, because NBER's access limits stopped PDF retrieval, and the papers it does cover for those two years are a seasonal slice. The study could not audit coded negatives in full text for the 5,559 papers retrieved last, because it deleted their full text at retrieval. The cover-block source covers every paper but misses about 2 percent of credits and up to about nine tenths of AI statements.

The AI series counts authors who chose to say they used a tool, in a period when norms about saying so were forming, and it pools pretrained models used as methods with assistants used for tasks; the assistant-task series is a few papers a year. I did not use detector-based measures because their validation does not support treating them as ground truth.

8. Conclusion

Printed research-assistant credits in NBER working papers fell by about three and a half percentage points after November 2022, from about 40 percent of papers to about 37, measured against 2018 through 2022 and after correcting for the credits that revision adds. That is a 9 percent relative decline, which is not large. Against 2018 and 2019 alone, the decline is two to two and a half points. The parallel ChatGPT study, which coded every paper, finds the same shares and a decline of 3.3 points. Authors who published at NBER in both periods credit assistants three to five points less often than before, against every baseline. The median number of assistants named per crediting paper is two in every year. Credits accumulate as authors revise papers and assistants turn over, so a study that reads only current versions would count part of the papers' age as a decline, about seven points for a cohort in which every paper has been revised and about two points for a typical older cohort. Two readings compete with displacement. Credits may have peaked with pandemic-era output and returned to their 2018 and

2019 level, and some assistants may have become coauthors, though counting predoctoral and research fellows as assistants shrinks the decline by less than half a point.

Author statements of AI use rose, and in full text most describe a pretrained model applied to data; the practice of thanking a tool for research assistance belongs to one author, Joshua Gans. Outside the methods series, the decline in credits is not where the disclosed AI use is, and the assistants still credited appear on no more papers each. This record cannot say whether any of the decline is displacement, and because NBER papers appear after the research is done, the record may largely predate the tools that could do a research assistant's work.

Most programs fell in 2025 and 2026 against 2018 through 2024, and four large ones by a statistically significant amount on their own, Asset Pricing by 10.0 points, Monetary Economics by 7.9, Economic Fluctuations and Growth by 5.1, and Public Economics by 4.8, without an unusually high rate of AI-use statements. The sharpest change is in the methods series, econometrics and methods papers that rarely require fieldwork, surveys, or new data collection, where displacement would plausibly appear first. In 2026 the research-assistant sentence nearly vanished from its acknowledgments, while statements of AI use, most describing help with drafting and coding, became common in the same cohort. The same authors credited assistants on 48 percent of their 264 earlier papers and 39 percent of their 18 non-methods 2026 papers, against 7 percent of their 28 methods papers in 2026. Twenty-eight papers cannot settle whether this field is the first to show what the others will show or a field unlike the others; later cohorts may.

Appendix A. PDF retrieval and access

The full processing order yielded 9,644 first-version files, covering every paper for 2018 through 2024, 449 of the 1,290 papers for 2025 (405 of them issued January through April, 401 of which the study could code), and 54 of the 1,287 for 2026. NBER refused 2,071 attempted papers in all (882 in the first retrieval and 1,189 in the completion), every one issued in 2025 or 2026, and three more failed for other reasons. NBER served 98 papers issued May 2025 or later; the study did not identify what distinguished them. The study coded and kept two further files retrieved outside the prefix (one pre-protocol format check disclosed in the protocol, one out-of-order fetch), against the letter of amendment A20. Amendment A27 discloses this. Dropping them moved the trend on the 4,085-paper sample from minus 0.58 to minus 0.57 points per year, and the A30 retrieval covers their positions in the order. The study stopped PDF retrieval on October 15, 2026, after noticing NBER's prohibition on automated access, asked NBER for permission, and on October 22 received permission to use the acknowledgment passages already held and to read abstract pages at no more than one request per second. It never retained PDFs and deleted its full-text extractions, including from version history, on October 22.

Under amendment A30 the study retrieved the remaining first versions on October 23 and 24, at my instruction and under my reading of NBER's October 22 permission. It stored the located acknowledgment passages and deleted each PDF and its full text in the same step.

Appendix B. The version effect

B.1 The version effect, measured directly

Credits accumulate across versions for a mechanical reason. Research assistants graduate and new ones replace them while a project runs, and authors add the new name without dropping the old. The effect matters here only because the all-paper census in Section B.2 would otherwise compare old papers that have accumulated credits with new papers that have not. The study expected the two sources to agree, since the cover block on the abstract page is the cover-page acknowledgment. A prespecified check on 200 papers found that they agree on positives but not on negatives, and that in 23 of the 29 disagreements where the abstract page had a credit that the first version lacked, the paper had a later revision. This appendix measures the effect directly. After the referee reports, the study coded the current cover blocks of 4,085 first-version papers with the cover-block pipeline (amendment A28; every batch of 150 scored nine or ten of ten on hidden calibration items). The other 5,559 first-version papers carry the cover-block codes from the earlier abstract-page coding (A23, with A24 second coding and the A26 correction). Together these give a paired code, first version against current cover block, for every paper in the first-version sample. The referee reports asked for this measurement before it was run (amendment A28).

Table B1 gives the result. Among the 2,964 revised papers with a clear code in both versions, 246 have a credit on the current cover that the first version lacked anywhere in its text (8.3 percent of revised papers, interval 7.4 to 9.3; 14.7 percent of those with no first-version credit, interval 13.1 to 16.5), and 38 have a first-version credit the current cover lacks (1.3 percent). Among the 6,491 unrevised papers, where the two texts are the same release, 74 gained and 52 lost (2.2 percent of credited papers). The losses measure the source difference; the gains are almost all papers whose current cover block I adjudicated as a credit under amendment A35 while the first-version code keeps the protocol's wording rule. About 2 percent of credits sit in footnotes or body text and not on the cover. The loss rate among revised papers is somewhat higher (2.9 against 2.2 percent of credited papers, Fisher $p = 0.18$), so revision adds about 246 credits and may remove a few. Within the first-version sample the current-cover share exceeds the first-version share by 1.8 to 3.2 points in every cohort from 2018 through 2024 and by 1.1 in 2025, and falls short of it in 2026, where authors have revised few papers.

Table B1. Research-assistant credit in the first version and on the current cover block, first-version sample, by revision status

	Current cover: none	Current cover: credit	Total
Revised papers (n = 2,964)			
First version: none	1,422	246	1,668
First version: credit	38	1,258	1,296
Unrevised papers (n = 6,491)			
First version: none	4,016	74	4,090
First version: credit	52	2,349	2,401

Papers with an ambiguous or unreadable code in either version are excluded (303 of 9,644).

B.2 A version-corrected census

Table B2 reports the all-paper census, the current cover-block text of all 11,718 papers, which is what any reader of the NBER abstract pages sees, and the same series with each cohort's expected revision-added credits removed (Section 3.3). The uncorrected all-paper census has a trend of minus 0.50 points per year (interval minus 0.85 to minus 0.16, p = 0.004) and a before-and-after difference of minus 4.0 points.

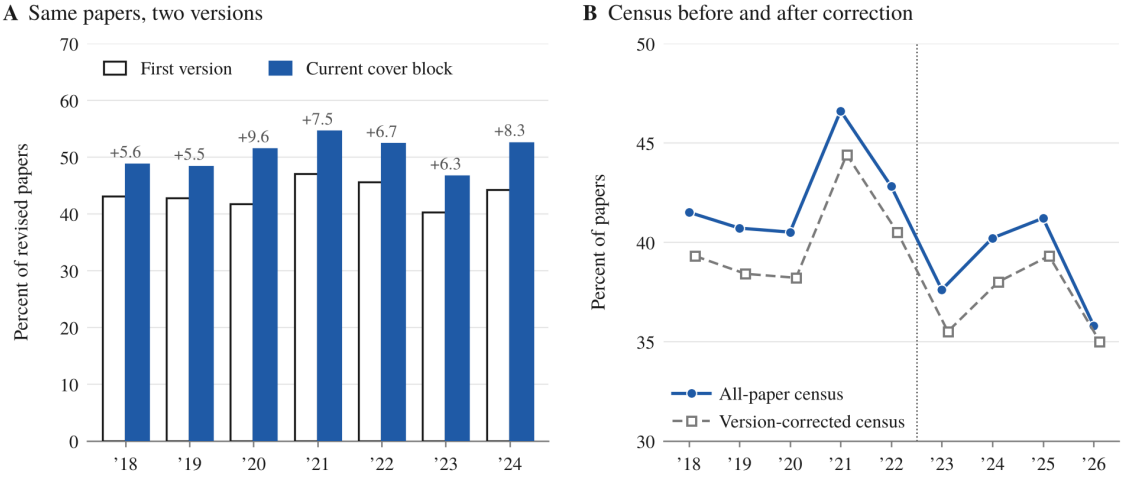
The per-year gains for 2018 through 2024 (5.5 to 9.6 points) are statistically indistinguishable, so the correction uses the pooled value. The version-corrected census has a fitted slope of minus 0.40 points per year (bootstrap SE 0.17) and averages 40.2 percent for 2018 to 2022 against 36.9 for 2023 to 2026 (difference minus 3.2, SE 0.9); at the paper level with the prespecified split it falls from 40.3 to 36.8. The version effect is real and runs in the direction of a spurious decline, but its size depends on the comparison. It shifts a comparison of older, heavily revised cohorts against the 2026 cohort by about one point. It shifts the prespecified before-and-after comparison by only a few tenths, because the 2023 and 2024 cohorts are as heavily revised as the pre-period cohorts. The net gain includes the 0.3-point net gain that unrevised papers show in Table B1, which comes from the A35 adjudication and not from revision; the difference in differences is 6.7 points, and using it changes the corrected shares by a tenth of a point.

Table B2. All-paper census and version-corrected census (11,547 papers with a clear code)

Year	n	Census share (%)	SE	Revised share of cohort (%)	Correction (points)	Corrected share (%)	SE
2018	1,222	41.5	1.4	31.3	2.2	39.3	1.4
2019	1,166	40.7	1.4	31.5	2.2	38.4	1.4
2020	1,682	40.5	1.2	32.9	2.3	38.2	1.2
2021	1,293	46.6	1.4	32.2	2.3	44.4	1.4
2022	1,181	42.8	1.4	31.7	2.2	40.5	1.4
2023	1,185	37.6	1.4	30.7	2.2	35.5	1.4
2024	1,296	40.2	1.4	30.8	2.2	38.0	1.4
2025	1,258	41.2	1.4	26.4	1.9	39.3	1.4
2026	1,264	35.8	1.3	12.5	0.9	35.0	1.4

The all-paper census is the current cover-block text of every paper, including the 9,644 first-version papers, whose current cover blocks the study coded separately. The correction is r(t) × g, defined in footnote 4. The SE combines the census sampling error with the gain's. Figure B1 shows both halves of the measurement.

Figure B1. The version effect, measured on the same papers



Notes: A: the 2,937 revised first-version papers issued 2018 to 2024, coded from both versions; labels give the net gain in percentage points. B: the correction subtracts each cohort's revision rate times the pooled net gain of 6.2 points (Equation 1). Dotted line: ChatGPT release, November 30, 2022.

B.3 Revised papers differ at baseline

Papers that authors eventually revise have more credits to begin with. In the first-version sample, papers with a later revision have a first-version share of 43.5 percent (interval 41.7 to 45.3) against 36.8 percent (35.7 to 38.0) for papers with none. The series of papers with no revision

listed runs 38.1, 37.0, 34.8, 42.3, 38.1, 33.2, 34.5, 37.8, and 35.7 percent for 2018 through 2026 (trend minus 0.26, interval minus 0.65 to 0.14). For 2025 and 2026 “no revision listed” mostly means “not yet,” so that group still contains most of the papers authors will later revise, and its expected rate under no behavioral change lies near the cohort mean.

Both effects concern the text rather than the assistants. A series built from current versions measures something that drifts with a paper’s age.

Appendix C. Comparison with the parallel ChatGPT study

The ChatGPT study began on October 13, a day before this one, and built the frozen frame of 11,718 NBER papers; this study received only the frame's paper IDs and selection metadata. The ChatGPT study used its own retrieval, coding rules, and analysis plan. It captured the current cover block for 11,708 of the 11,718 papers between October 13 and 22. It first coded a sample of 100 papers a year chosen by hash order, and later analyzed the 289 a year it had coded by October 15. That sample could detect only a decline of about five points, so at my request the ChatGPT study coded every eligible paper on October 26 under its own frozen specification. It recorded its analysis plan before coding began and froze its codes before computing any result.

Table C1 compares the two studies year by year. Of the 11,271 papers that both studies coded as crediting or not crediting an assistant, they disagree on 14 (agreement 99.88 percent, kappa 0.997); this study coded 3 of those as credits and the ChatGPT study 11. Each study left about 300 papers ambiguous or unreadable, and the two sets overlap on 115. On AI-use statements the studies disagree on 1 of 11,663 papers. The ChatGPT study's own blind second coding of 1,175 papers agreed on 99.0 percent of human-credit codes (kappa 0.98), and its audit of 350 papers coded as having no credit found no confirmed missed credit.

Table C1. Share of papers crediting a research assistant by year, this study and the ChatGPT study

Year	This study: n	Share (%)	ChatGPT study: n	Share (%)	Papers coded differently
2018	1,199	40.7	1,214	40.4	3
2019	1,153	40.2	1,161	39.9	0
2020	1,661	39.9	1,668	39.6	2
2021	1,271	45.7	1,275	45.3	2
2022	1,165	42.0	1,164	41.9	1
2023	1,168	37.1	1,169	37.0	1
2024	1,281	39.6	1,275	39.8	3
2025	1,253	40.9	1,262	40.8	2
2026	1,263	35.9	1,262	35.8	0

Papers eligible in both studies, each study's own codes, with ambiguous or unreadable codes excluded, so this study's n differs slightly from Table B2. Papers coded differently counts papers that one study coded as a credit and the other as no credit. Both studies read the current cover block; 2026 runs through October 13.

The ChatGPT study used its own rules and coders, and it was not blind to this study. Its coordinator had seen this study's drafts and results, and on October 23 this study had sent it paper-level codes for all 11,718 papers, so the full-frame coding ran with those codes in its environment. Its coders worked in isolated contexts that did not receive them. The 2,602 papers the ChatGPT study coded on October 14 and 15, before the two studies exchanged any codes,

agree with this study at the same rate as the 9,106 it coded afterward (99.84 against 99.87 percent), so the exchange did not raise agreement. One amendment to the ChatGPT study's rules, recorded after its first calibration round and before it coded the new papers, clarified that a block crediting only an operational task, such as data access or statistical help, counts as no credit rather than as unknown.

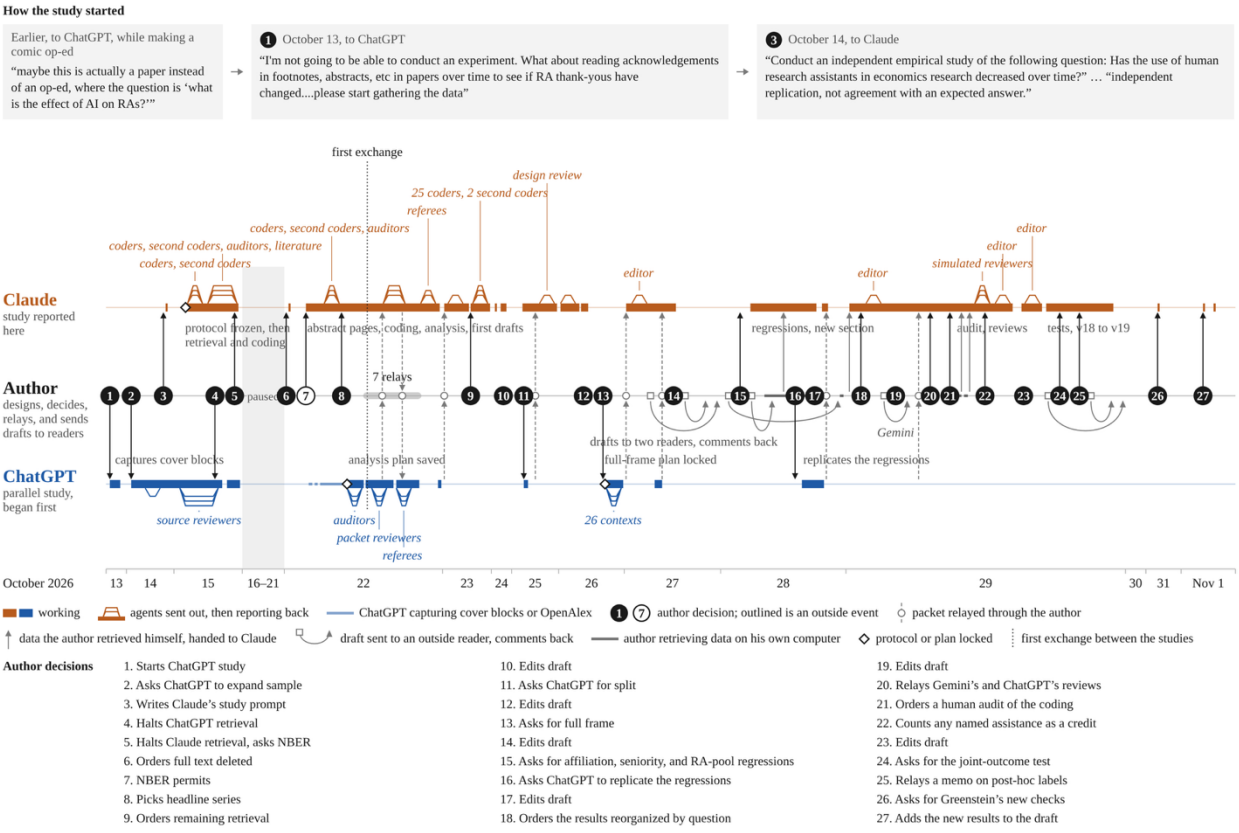
The ChatGPT study's annual shares are within 0.4 points of this study's all-paper census in every year. With 2023 in the post period its before-and-after change is minus 3.3 points (SE 0.9, interval minus 5.1 to minus 1.5), against minus 4.0 here (minus 3.7 under the protocol's wording rule). On its own registered window, which omits 2023, the change is minus 2.7 (SE 1.0, interval minus 4.7 to minus 0.8), against minus 3.5. Against 2018 and 2019 alone it finds minus 2.0 ($p = 0.11$), and without 2020 and 2021 minus 2.9 ($p = 0.008$). Its trend is minus 0.39 points a year ($p = 0.03$), against minus 0.50. Its largest program declines are in the same three programs, Technical Working Papers (minus 12.9 points, with 2 of 28 papers crediting an assistant in 2026), Asset Pricing (minus 9.6), and Monetary Economics (minus 7.3). In both studies' codes the Asset Pricing decline survives a Bonferroni correction for 20 programs (simultaneous interval minus 18.9 to minus 0.3 in the ChatGPT study's, minus 18.7 to minus 1.3 in this study's). By the ChatGPT study's own decomposition, its earlier sample estimates, minus 2.2 points on the split at November 30, 2022, and minus 0.1 on its own window, differ from the full-frame results mainly because of which papers the sample drew.

Section 5 reports my audit of 100 cover blocks and the 14 disagreements between the studies' credit codes. The ChatGPT study's earlier review of this study found the coding-rule inconsistency corrected in A26, an asymmetric date cutoff in this study's first program comparison, and an error in the audit miss-rate formula, all since corrected. The ChatGPT study's code file, quality reports, and analysis output, and this comparison, are in the study's repository.

Appendix D. Coding, validation, and protocol amendments

Figure D1 is the full version of Figure 11. It labels each group of agents with its role and count, shows when ChatGPT was capturing cover blocks or OpenAlex records, and marks the OpenAlex retrieval I ran on my own computer for the Claude study when its workspace could not reach OpenAlex.

Figure D1. How the two studies were produced, full timeline



Author decisions

1. Starts ChatGPT study
2. Asks ChatGPT to expand sample
3. Writes Claude's study prompt
4. Halts ChatGPT retrieval
5. Halts Claude retrieval, asks NBER
6. Orders full text deleted
7. NBER permits
8. Picks headline series
9. Orders remaining retrieval
10. Edits draft
11. Asks ChatGPT for split
12. Edits draft
13. Asks for full frame
14. Edits draft
15. Asks for affiliation, seniority, and RA-pool regressions
16. Asks ChatGPT to replicate the regressions
17. Edits draft
18. Orders the results reorganized by question
19. Edits draft
20. Relays Gemini's and ChatGPT's reviews
21. Orders a human audit of the coding
22. Counts any named assistance as a credit
23. Edits draft
24. Asks for the joint-outcome test
25. Relays a memo on post-hoc labels
26. Asks for Greenstein's new checks
27. Adds the new results to the draft

Notes: The horizontal axis keeps the order of events but not their spacing. Periods of activity are widened, longer ones more than shorter ones, idle hours are compressed, and October 16 to 21, when work paused while NBER considered the permission request, is collapsed into one band. Each line that leaves a lane stands for agents sent out by that lead session; it rejoins the lane when they report. At most three lines are drawn per burst, and labels give the agents' roles and counts. ChatGPT's 26 contexts on October 26 were calibration runs, coders, second coders, auditors, adjudicators, and a verifier. Its bar for October 14 and 15 spans intermittent work. Diamonds mark when each study fixed its protocol or analysis plan, and the dotted line marks the first packet passed between them. Arcs on the author's lane show drafts sent to two outside readers on October 27 and 28 and their comments coming back. The author then asked both studies for the regressions in Sections 4.6 and 4.7, retrieved the OpenAlex author records for the Claude study on his own computer when the workspace could not reach OpenAlex, and passed them on (solid gray arrows); ChatGPT retrieved its own. On October 29 the author read 100 randomly drawn cover blocks against the stored codes and adjudicated the 163 papers on which the studies' codes disagreed or this study's was ambiguous, and passed both back (solid gray arrows). The small arc that day is a draft sent to Gemini for one review; four simulated reviewers then read the draft. From the afternoon of October 29 through November 1 the author's memo on the "post hoc" label set further tests, two readers sent more comments, and the study ran the joint-outcome, papers-per-assistant, finance, code-intensity, and funder checks and produced drafts v17 to v19. The two AI systems never communicated directly; every exchange passed through the author. Sources: the study repository's commit log and run log, the author's mail, and ChatGPT's conversation, context, and coding logs.

D.1 Protocol, coding, and validation

The study froze the protocol and codebook on October 15, 2026, before coding any paper. The protocol set five primary tests and eight sensitivity analyses. Amendments added four more sensitivity analyses after coding began and before the study computed any year-by-year share. One restricts the sample to papers issued January through April, because NBER's access limits left 2025 incomplete. Two pool abstract-page text with full text, a source NBER permitted later.

The fourth uses only text that revision had not changed; the study added it after the source-agreement check showed the version difference. The study had seen aggregate counts of codes, and no annual share, when it wrote the last three. Appendix Table E3 reports the two series not used in the main text, and Section D.2 lists every amendment.

Coding subagents of the Claude model family coded the first 4,085 first-version papers from the located acknowledgment passages, after each coder reached at least 90 percent field-level agreement with a 42-paper gold set coded by the lead session. An independent run of a different model in the same family blind second-coded 400 of those papers from the same passages; research-assistant codes agreed on all 400, and the weakest fields were the support-type labels. Because the second coder saw the same located passages, that agreement measures coding consistency given the locator. A locator validation on 220 papers measured recall (96.2 percent on a 130-paper pilot and 100 percent on 90 fresh papers after amendment A1), and two negative audits, in which auditors re-read in full 176 papers coded as having no credit, found no missed credit. A third session coded the 5,559 papers retrieved under A30 from the located passages alone, because the study had deleted their full text at retrieval. Each of its 25 coders (13 runs of claude-fable-5-1 and 12 of claude-opus-5-5) matched the gold set on every field other than the type labels before the study accepted its work. Two Sonnet coders blind second-coded 556 of these papers (10 percent, stratified by year). They agreed with the first codes on 99.5 percent of research-assistant codes (kappa 0.989) and 99.8 percent of AI codes (kappa 0.952), and the lead adjudicated the four disagreements on those fields, changing one code.

Sonnet runs coded all 11,718 cover blocks (7,633 under amendment A23 and the other 4,085 under A28) from the text alone, in batches of 150 with ten hidden calibration items each, and the study accepted a batch at nine of ten correct. For the 2,074 papers with no first version, the cover block is the only text. Opus runs blind second-coded 763 cover blocks (research-assistant agreement 97.1 percent, kappa 0.945; named counts in exact agreement on all 286 papers both called positive with an integer count, of 289 both called positive) and audited 180 papers coded as having no credit (no misses). The lead adjudicated 22 disagreements, changing four codes, none of them a research-assistant credit. The cover blocks of the first 4,085 first-version papers, coded later under amendment A28, had no second coding or audit beyond the calibration items. Every positive, ambiguous, or disclosed code carries a verbatim quote that a validator checks against the stored text.

Zero misses in an audit of 176 or 180 coded negatives gives a 95 percent upper bound of 2.1 percent on the share of coded negatives that are true positives in each source. Applied to a share p, the corrected upper share is p plus 2.1 percent of (1 minus p), about 1.3 points at p of 40 percent; under a constant miss rate the corrected trend is 98 percent of the observed trend. A pooled audit of 20 papers per year cannot rule out a miss rate that varies by year, and the cover-block audit covered 4.0 percent of negatives, short of the protocol's 5 percent. The full-text bound covers only the 4,085 papers coded first (Section 7). The cover-block audit cannot find credits printed only in footnotes or body text, which Appendix B measures.

The full-text and cover-block coding applied one rule differently. Amendment A6 (October 15) counts a named person credited only with "research support" under the wider definition and not as a research-assistant credit; the cover-block instructions coded the same wording as ambiguous. An external review found the inconsistency, and the study corrected 52 cover-block records on October 23 (amendment A26), after it had viewed every year-by-year outcome. No

research-assistant credit, no code under the wider definition, and no first-version result changed. Amendment A31 records the lead's rulings on about 60 rule questions that the third session's coders raised and normalizes 28 named-count fields. It also discloses a trial run. Before the coders finished the last 165 papers and before the second coding of the added papers, the lead ran the analysis scripts once on a trial copy of the data to find code errors, and the run printed the first-version shares by year. The lead wrote the A31 rulings after that run, and they changed no research-assistant or AI code; the one research-assistant code changed afterward came from the second-coding adjudication.

The prespecified primary tests use the first-version sample (Section 4.1). Everything else came after the protocol froze, and the amendment log records when and why. The additions are of two kinds. The study added most without choosing among outcomes. The referees asked for the version-effect measurement, the within-author estimates, and the author-concentration analysis before any was run, and readers proposed the regressors in Section 3.3 before I estimated any regression, with the specification curve reporting every combination. I chose two analyses after seeing outcomes, and the text says so where they appear. I found the methods-series cell in Section 4.5 by inspecting all 180 program-by-year cells, and I formed the finance grouping in Section 4.1 after the program results were known and then tested it under rules written in advance (amendments A36 to A38). I also set the exposure scores in Section 4.4 and found the career-by-team pattern in Section 4.6 after seeing results. After ChatGPT reviewed this draft on October 27, the study classified the eight 2026 methods-series AI statements by reading their stored quotes and recomputed the same-authors standard errors with clustering by author and by paper (amendment A33).

D.2 Amendments to the protocol

Each amendment changed the frozen protocol or codebook or recorded a decision about it. The repository log gives each one in full, with its reason and the files it affected. The last column says what outcomes the study had seen when it made the change; through A25 the study had computed no year-by-year share. Three further log entries revised reference codes in the gold set used to calibrate coders.

Amendment	Date (2026)	What changed	Outcomes seen when made
A1	Oct 15	Locator captures trial-registry lines and acknowledgment sections that continue onto the next page.	None
A2	Oct 15	PDFs NBER refuses are recorded as unretrieved; no login or other workaround.	None
A3	Oct 15	AI use means a pretrained or third-party model or service; models the authors train themselves do not count.	None
A4	Oct 15	Locator captures AI declaration headings; AI screen adds pretrained-model and translation patterns.	None
A5	Oct 15	Named-assistant counts cover research-assistant credits only; credits under the wider definition are recorded separately.	None
A6	Oct 15	Coding clarifications after calibration: wording that counts only under the wider definition, variants of research-assistant wording, mixed roles, groups.	None
A7	Oct 15	Clarifications after blind second coding: research-team mentions, review scope.	None
A8	Oct 15	Full-text search for author statements of AI use in papers with many AI hits.	None
A9	Oct 15	Clarifications: task-specific non-use of AI, ambiguous credits, research managers.	None
A10	Oct 15	Clarifications: help with data, credits with stated tasks; wider AI search promised.	None
A11	Oct 15	Partners thanked with no stated task are ambiguous support.	None
A12	Oct 15	Bare “help” or “assistance” to a named person is an ambiguous credit; consistency pass required.	None
A13	Oct 15	Bare “support” from named people is ambiguous support.	None
A14	Oct 15	Wider AI-statement search; refused PDFs not re-requested; consistency pass defined.	None
A15	Oct 15	Ambiguous components inside credits already coded positive.	None
A16	Oct 15	AI screen adds vision, OCR, and speech models; rule for named OCR services.	Counts of coded negatives by year, to set audit quotas; no share
A17	Oct 15	Thanks with no activity named is not a credit; rule for unnamed NLP methods.	None
A18	Oct 15	Help using data counts as support and access alone does not; “research help” is a research-assistant credit.	None
A19	Oct 15	Named-assistant measure includes every paper with a whole-number count.	None
A20	Oct 15	Analysis sample after retrieval stopped; January-to-April sensitivity added; definitions fixed.	None

Amendment	Date (2026)	What changed	Outcomes seen when made
A21	Oct 22	NBER permission recorded; abstract-page text added for papers without a PDF; pooling test set in advance.	None
A22	Oct 22	Abstract pages show the current version, not the first; abstract-only papers kept as sensitivity series.	Aggregate code counts, all years
A23	Oct 22	Cover-block coding of 7,633 abstract-only papers completed, with calibration items.	Aggregate code counts, all years
A24	Oct 22	Second coding and negative audit of cover-block codes; adjudication.	Agreement tables
A25	Oct 22	Reporting decision on headline series; an AI tool thanked for research assistance counts as an AI disclosure.	None beyond A21 to A24
A26	Oct 23	Fifty-two cover-block codes corrected to follow A6; second analysis run; reporting corrections.	All year-by-year outcomes
A27	Oct 23	Two papers retrieved outside the planned order disclosed; audit-bound formula corrected.	All outcomes
A28	Oct 23	Current cover blocks of 4,085 first-version papers coded, to measure the version effect directly. Requested by the referees before it ran.	All outcomes
A29	Oct 23	Planned retrieval of 2025 and 2026 PDFs through a signed-in session; withdrawn before any retrieval.	All outcomes
A30	Oct 23	Remaining first versions retrieved (5,559 papers) and coded; coder model changed after a usage limit.	All outcomes on the earlier sample
A31	Oct 24	Rulings on coders' open questions; 28 count fields normalized; a trial run disclosed.	Trial-run yearly first-version shares
A32	Oct 26	Romano-Wolf replaces Holm as the adjustment for the five primary tests.	All outcomes
A33	Oct 27	Stored quotes of the eight 2026 methods-series AI statements classified by kind of use; same-authors standard errors clustered by author and by paper. Post hoc.	All outcomes
A34	Oct 28	Regressions with OpenAlex affiliations, author h-index, career age, and NSF economics graduate enrollment; two-margin and two-window figures (Sections 4.1 to 4.7). Added at readers' request; regressors proposed by readers, not chosen from the data.	All outcomes
A35	Oct 29	Author adjudicated the 14 papers on which the two studies' credit codes disagree (all 14 are credits; 11 corrections) and the 149 ambiguous codes (104 credits, 20 not, 24 still ambiguous, reported as a sensitivity). Author audit of 100 random cover blocks: one miss. Author decision.	All outcomes
A36	Oct 29	Code-intensity rule (JEL C1, C6, C8, or method words in the title) written before the test; 2025 to 2026 change by code intensity outside the TWP label.	Section 4.5
A37	Oct 29	Dated prediction for TWP papers issued October 14, 2026 to June 30, 2027, under the displacement and unusual-draw readings.	Section 4.5

Amendment	Date (2026)	What changed	Outcomes seen when made
A38	Oct 29	Finance tests with rules written before they ran: mean reversion across programs; business-school against other authors within finance.	Section 4.1
A39	Oct 29	Author decision: “post hoc” reserved for analyses chosen after seeing outcomes; timing of the rest recorded here without hedging in the text; A34 entry revised.	Text
A40	Nov 1	Papers per named assistant; credits and AI statements as joint responses to post-ChatGPT exposure; funder acknowledgments; career age by team size (Sections 4.2, 4.4, 4.6).	Sections 4.2, 4.4, 4.6

Appendix E. Additional tables

The analysis files, the protocol and codebook, the amendment log, the comparison with the ChatGPT study, and the referee reports are in the study's repository. Table E1 gives the program numbers behind Figure 2. Table E2 gives the before-and-after comparison under other baselines, and Table E3 gives two prespecified series not used in the main text.

Table E1. Share of papers crediting a research assistant, by program, 2025 and 2026 against 2018 to 2024 (numbers behind Figure 2)

Program	2018 to 2024 (%)	n	2025 (%)	n	2026 (%)	n	Change (points)	SE	AI use, 2026 (%)
Technical Working Papers	38.1	197	47.8	23	7.1	28	−12.6	7.1	28.6
Asset Pricing	37.9	1,027	32.5	114	24.7	166	−10.0	3.1	0.0
Monetary Economics	42.7	1,086	34.4	125	35.3	136	−7.9	3.3	4.4
Economics of Aging	40.5	582	42.7	75	24.7	85	−7.4	4.2	7.1
International Trade and Investment	49.3	764	42.2	135	43.0	93	−6.8	3.8	4.3
International Finance and Macroeconomics	43.1	888	34.7	121	38.3	115	−6.7	3.5	0.0
Economic Fluctuations and Growth	39.7	1,912	35.9	290	33.3	249	−5.1	2.3	2.8
Public Economics	46.8	2,233	47.0	313	36.6	284	−4.8	2.3	2.8
Law and Economics	48.0	521	54.3	70	32.9	76	−4.8	4.7	2.6
Children and Families	44.3	808	41.0	156	39.5	124	−3.9	3.4	3.2
Corporate Finance	45.5	1,067	43.9	139	40.3	129	−3.3	3.4	0.8
Labor Studies	44.9	2,275	43.7	357	41.1	302	−2.4	2.2	4.0
Development of the American Economy	51.4	504	46.8	77	52.3	65	−2.1	4.8	6.2
Economics of Education	46.2	753	49.6	133	37.4	107	−2.0	3.7	4.7
Development Economics	55.0	1,098	57.9	159	49.7	149	−1.1	3.2	8.1
Productivity, Innovation, and Entrepreneurship	40.4	1,163	47.5	183	31.9	191	−0.8	2.9	13.2
Health Economics	38.6	1,422	42.3	182	33.5	173	−0.6	2.9	2.9
Political Economy	53.8	849	58.4	125	49.2	124	−0.0	3.6	4.9
Environment and Energy Economics	40.2	722	43.2	111	37.9	103	+0.5	3.8	4.9
Industrial Organization	45.1	859	51.6	122	46.6	116	+4.1	3.7	6.0

Change is 2025 and 2026 pooled minus 2018 to 2024, in points, with its normal standard error, unadjusted for 20 comparisons. AI use, 2026 is the share of the program's 2026 papers with an AI-use statement, read from full text where the study retrieved the first version and from the cover block otherwise.

Table E2. Before-and-after difference in the share crediting a research assistant, alternative baselines (percentage points)

Baseline	All papers	Version-corrected	First versions, to 2024	Same authors
As prespecified: January 2018 to November 2022	−4.0 (0.9)	−3.5 (0.9)	−3.9 (1.1)	−5.1 (0.9)
Without 2020 and 2021	−3.5 (1.1)	−3.0 (1.1)	−3.9 (1.3)	−5.1 (1.1)
2018 and 2019 only	−2.5 (1.2)	−2.1 (1.2)	−3.1 (1.4)	−4.0 (1.3)
Without 2020, 2021, and 2023	−3.2 (1.1)	−2.6 (1.1)	−3.4 (1.5)	−4.7 (1.2)
2018 and 2019 against 2024 to 2026	−2.0 (1.3)	−1.4 (1.3)	−1.7 (1.7)	−3.4 (1.4)

Standard errors in parentheses. The before period is the stated baseline, and the after period runs from December 1, 2022, to October 13, 2026, except in the last two rows. First versions exclude the 2025 and 2026 cells, which depend on which PDFs NBER served, so the first row differs from the 3.2-point difference in Section 4.1. Same-authors estimates use author fixed effects on author-paper pairs for authors with papers on both sides of each split, with standard errors clustered by author and by paper.

Table E3. Two prespecified series not used in the main text

Series	Papers	Before (%)	After (%)	Change (points)	Trend per year (p)
Text unchanged by revision	11,170	40.0	36.7	−3.3	−0.35 (0.057)
Full text and cover block pooled	11,534	40.0	37.2	−2.8	−0.23 (0.32)

Before and after split at November 30, 2022. The pooled series' trend includes an indicator for text source.

References

Agrawal, Ajay, Joshua Gans, and Avi Goldfarb. 2019. "Artificial Intelligence: The Ambiguous Labor Market Impact of Automating Prediction." Journal of Economic Perspectives 33 (2): 31-50. doi:10.1257/jep.33.2.31.

Alvarez-Bornstein, Belen, Fernanda Morillo, and Maria Bordons. 2017. "Funding Acknowledgments in the Web of Science: Completeness and Accuracy of Collected Data." Scientometrics 112 (3): 1793-1812. doi:10.1007/s11192-017-2453-4.

Angrist, Joshua, Pierre Azoulay, Glenn Ellison, Ryan Hill, and Susan Feng Lu. 2017. "Economic Research Evolves: Fields and Styles." American Economic Review 107 (5): 293 to 297. doi:10.1257/aer.p20171117.

Ans, M., L. Maggio, H. Algodi, J. Costello, E. Driessen, K. Oswald, and L. Lingard. 2025. "The Presence and Nature of AI-Use Disclosure Statements in Medical Education Journals: A bibliometric study." medRxiv preprint, posted 13 November 2025. DOI: 10.1101/2025.11.11.25340015.

Baccini, Alberto, and Eugenio Petrovich. 2022. "Normative versus Strategic Accounts of Acknowledgment Data: The Case of the Top-Five Journals of Economics." Scientometrics 127: 603-635. doi:10.1007/s11192-021-04185-6.

Bresnahan, Timothy F., and Shane Greenstein. 1996. "Technical Progress and Co-invention in Computing and in the Uses of Computers." Brookings Papers on Economic Activity: Microeconomics 1996: 1-83.

Bresnahan, Timothy F., Shane Greenstein, and Pai-Ling Yin. 2025. "New Economic Forces Behind the Value Distribution of Innovation." NBER Working Paper 34090. https://www.nber.org/papers/w34090.

Bresnahan, Timothy F., and Manuel Trajtenberg. 1995. "General Purpose Technologies: 'Engines of Growth'?" Journal of Econometrics 65 (1): 83-108.

Broady, Kristen. 2026. "Who's Earning Doctorates in Economics and What Factors Contribute to Success in This Field?" Federal Reserve Bank of Chicago Working Paper 2026-13, July 2026. doi:10.21033/wp-2026-13.

Brynjolfsson, Erik, Bharat Chandar, and Ruyu Chen. 2025. "Canaries in the Coal Mine? Six Facts about the Recent Employment Effects of Artificial Intelligence." Working paper, Stanford Digital Economy Lab, August 2025; revised August 2026 with data through June 2026.

Brynjolfsson, Erik, Danielle Li, and Lindsey R. Raymond. 2023. "Generative AI at Work." NBER Working Paper 31161. doi:10.3386/w31161. Published as Quarterly Journal of Economics 140 (2): 889 to 942, 2025.

Caro, Florian. 2024. “Research Assistants, Sorting, and Career Outcomes: Evidence from the NBER Working Paper Series.” Working paper, SSRN 4944297, dated September 2, 2024, posted October 7, 2024.

CitEc. 2022. “Citation Analysis for NBER Technical Working Papers.” RePEc. https://citec.repec.org/s/2022/nbrnberte.html.

Cronin, Blaise, Debora Shaw, and Kathryn La Barre. 2003. “A Cast of Thousands: Coauthorship and Subauthorship Collaboration in the 20th Century as Manifested in the Scholarly Journal Literature of Psychology and Philosophy.” Journal of the American Society for Information Science and Technology 54 (9): 855-871.

Cui, Kevin Zheyuan, Mert Demirer, Sonia Jaffe, Leon Musolff, Sida Peng, and Tobias Salz. 2026. “The Effects of Generative AI on High-Skilled Work: Evidence from Three Field Experiments with Software Developers.” Management Science, published online February 27, 2026, ahead of print. doi:10.1287/mnsc.2025.00535.

Dell’Acqua, Fabrizio, Edward McFowland III, Ethan Mollick, Hila Lifshitz-Assaf, Katherine C. Kellogg, Saran Rajendran, Lisa Krayer, Francois Candelon, and Karim R. Lakhani. 2023. “Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of Artificial Intelligence on Knowledge Worker Productivity and Quality.” Harvard Business School Working Paper 24-013.

Desrochers, Nadine, Adele Paul-Hus, and Jen Pecoskie. 2017. “Five Decades of Gratitude: A Meta-Synthesis of Acknowledgments Research.” Journal of the Association for Information Science and Technology 68 (12): 2821-2833. doi:10.1002/asi.23903.

Einav, Liran, and Jonathan Levin. 2014. “The Data Revolution and Economic Analysis.” In Innovation Policy and the Economy, Volume 14, edited by Josh Lerner and Scott Stern. Also NBER Working Paper 19035.

Gans, Joshua. 2026. “Reflections on Vibe Researching.” Substack, January 3. https://joshuagans.substack.com/p/reflections-on-vibe-researching.

Giles, C. Lee, and Isaac G. Councill. 2004. “Who Gets Acknowledged: Measuring Scientific Contributions through Automatic Acknowledgment Indexing.” Proceedings of the National Academy of Sciences 101 (51): 17599-17604. doi:10.1073/pnas.0407743101.

Hamermesh, Daniel S. 2013. “Six Decades of Top Economics Publishing: Who and How?” Journal of Economic Literature 51 (1): 162 to 172. doi:10.1257/jel.51.1.162.

He, Yongyuan, and Yi Bu. 2025. “Academic journals’ AI policies fail to curb the surge in AI-assisted academic writing.” arXiv:2512.06705 (submitted 7 December 2025).

Hosseini Maasoum, Seyed Mahdi, and Guy Lichtinger. 2025. “Generative AI as Seniority-Biased Technological Change: Evidence from U.S. Resume and Job Posting Data.” SSRN Working Paper 5425555, first version August 31, 2025; version reviewed dated October 27, 2025.

Humlum, Anders, and Emilie Vestergaard. 2025. “Large Language Models, Small Labor Market Effects.” NBER Working Paper 33777, May 2025. Revised March 2026 under the title “Still

Waters, Rapid Currents: Early Labor Market Transformation under Generative AI." doi:10.3386/w33777.

Jones, Benjamin F. 2021. "The Rise of Research Teams: Benefits and Costs in Economics." Journal of Economic Perspectives 35 (2): 191 to 216. doi:10.1257/jep.35.2.191.

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. DOI: 10.1126/sciadv.adt3813. Preprint arXiv:2406.07016.

Korinek, Anton. 2023. "Generative AI for Economic Research: Use Cases and Implications for Economists." Journal of Economic Literature 61 (4): 1281 to 1317. doi:10.1257/jel.20231736.

Laband, David N., and Robert D. Tollison. 2000. "Intellectual Collaboration." Journal of Political Economy 108 (3): 632 to 662. doi:10.1086/262132.

Larivière, Vincent, Nadine Desrochers, Benoît Macaluso, Philippe Mongeon, Adèle Paul-Hus, and Cassidy R. Sugimoto. 2016. "Contributorship and Division of Labor in Knowledge Production." Social Studies of Science 46 (3): 417 to 435.

Liang, Weixin, Mert Yuksekgonul, Yining Mao, Eric Wu, and James Zou. 2023. "GPT detectors are biased against non-native English writers." Patterns. Preprint arXiv:2304.02819.

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 Zou. 2025. "Quantifying large language model usage in scientific papers." Nature Human Behaviour. DOI: 10.1038/s41562-025-02273-8. Published 4 August 2025. Earlier version: "Mapping the Increasing Use of LLMs in Scientific Papers," arXiv:2404.01268, submitted 1 April 2024.

Liao, Zhehui, Maria Antoniak, Inyoung Cheong, Evie Yu-Yen Cheng, Ai-Heng Lee, Kyle Lo, Joseph Chee Chang, and Amy X. Zhang. 2024. "LLMs as Research Tools: A Large Scale Survey of Researchers' Usage and Perceptions." arXiv:2411.05025 (submitted 30 October 2024).

Linardon, Jake, Jennifer J. Thomas, Scott J. Crow, Ata Ghaderi, Anja Hilbert, Kelly L. Klump, Tracey D. Wade, B. Timothy Walsh, and Ruth Weissman. 2025. "Conducting Eating Disorder Research in the Era of Generative AI: Researcher Perspectives and Guidelines From the International Journal of Eating Disorders." International Journal of Eating Disorders 58(12): 2306-2316. DOI: 10.1002/eat.24543.

NBER (National Bureau of Economic Research). n.d. "NBER Working Papers Metadata." https://www.nber.org/research/data/nber-working-papers-metadata.

Noy, Shakked, and Whitney Zhang. 2023. "Experimental evidence on the productivity effects of generative artificial intelligence." Science 381 (6654): 187 to 192. doi:10.1126/science.adh2586.

Paul-Hus, Adele, Philippe Mongeon, Maxime Sainte-Marie, and Vincent Lariviere. 2020. "Who Are the Acknowledgees? An Analysis of Gender and Academic Status." Quantitative Science Studies 1 (2): 582-598. doi:10.1162/qss_a_00036.

Romano, Joseph P., and Michael Wolf. 2005. “Stepwise Multiple Testing as Formalized Data Snooping.” Econometrica 73 (4): 1237-1282. doi:10.1111/j.1468-0262.2005.00615.x.

Rose, Michael E., and Co-Pierre Georg. 2021. “What 5,000 Acknowledgements Tell Us About Informal Collaboration in Financial Economics.” Research Policy 50 (6): 104236. doi:10.1016/j.respol.2021.104236.

Ross, Matthew B., Britta M. Glennon, Raviv Murciano-Goroff, Enrico G. Berkes, Bruce A. Weinberg, and Julia I. Lane. 2022. “Women Are Credited Less in Science than Men.” Nature 608: 135-145. doi:10.1038/s41586-022-04966-w.

Sadasivan, Vinu Sankar, Aounon Kumar, Sriram Balasubramanian, Wenxiao Wang, and Soheil Feizi. 2023. “Can AI-Generated Text be Reliably Detected?” arXiv:2303.11156 (v1, 17 March 2023; v4, 17 January 2025).

Saha, Rounak, Gurusha Juneja, Dayita Chaudhuri, Naveeja Sajeevan, Nihar B. Shah, and Danish Pruthi. 2026. “Policies Permitting LLM Use for Polishing Peer Reviews Are Currently Not Enforceable.” arXiv:2603.20450 (submitted 20 March 2026).

Salandra, Rossella, Marisa Miraldo, and Paola Criscuolo. 2026. “Attribution of Credit in Acknowledgements: The Case of Systematic Reviews in Medicine.” PLOS ONE 21 (1): e0338714. doi:10.1371/journal.pone.0338714.

Smirnova, Nina, and Philipp Mayr. 2023. “A Comprehensive Analysis of Acknowledgement Texts in Web of Science: A Case Study on Four Scientific Domains.” Scientometrics 128 (1): 709-734. doi:10.1007/s11192-022-04554-9.

Weber-Wulff, Debora, Alla Anohina-Naumeca, Sonja Bjelobaba, Tomáš Foltýnek, Jean Guerrero-Dib, Olumide Popoola, Petr Šigut, and Lorna Waddington. 2023. “Testing of detection tools for AI-generated text.” International Journal for Educational Integrity 19(1): 26. DOI: 10.1007/s40979-023-00146-z.

Wuchty, Stefan, Benjamin F. Jones, and Brian Uzzi. 2007. “The Increasing Dominance of Teams in Production of Knowledge.” Science 316 (5827): 1036 to 1039. doi:10.1126/science.1136099.