

Technological Breakthroughs and the Progress of Science: Evidence from Early Computers

Supplemental Online Appendix

Pedro Aldighieri* Franco Malpassi†

[Link to main paper](#)

Contents

A	Computer Installations Database: Variables, Sources, and Universities	1
B	Computer Database: Additional Descriptives	8
C	Computer Usage Across Disciplines and Diffusion	11
D	Computer Use Classification Details	15
E	Paper-Level Outcomes: Embedding Controls and Patent Citations	17
F	Citation Roles: Bucket Premium Scores and Examples	19
G	Reference Age and Sleeping Beauties: Embedding Controls	21
H	Author-Level Results: Early versus Late Adopters	22
I	Subfield-to-Department Crosswalk	23
J	Results for Health and Life Sciences	25
K	Research Methodology Classification Details	28

*Department of Economics, Northwestern University. pedro.aldighieri@u.northwestern.edu

†Department of Economics, Northwestern University. francomalpassi@u.northwestern.edu

A Computer Installations Database: Variables, Sources, and Universities

A.1 Computer Dataset Variables

The computer installations dataset contains the following variables:

1. **University:** Research Organization Registry (ROR)¹ and university name;
2. **Computer Model:**
 - (a) Department/unit of university where computer was installed;
 - (b) Year and month of installation;
 - (c) Year and month of decommission;
 - (d) Average hours of usage per month;
 - (e) Flag of whether computer is analog;
 - (f) Flag of whether computer was built in-house;
 - (g) List of survey-years that report the computer.
3. **Miscellaneous:** we also include additional information that is not systematically reported:
 - (a) Flag of whether computer was commissioned by the military;
 - (b) Flag of whether the computer was a donation.

Other relevant but unstructured information is provided in comments.

A.2 List of Sources

As reported in Section 3 of the main paper, we collect data from 24 survey sources comprising 82 survey-year pairs. Figure A.1 shows the timeline coverage of our survey sources.

Our sources are:

¹The Research Organization Registry (ROR) is a global, community-led registry that provides open persistent identifiers for research organizations. These identifiers facilitate the disambiguation of institutional affiliations and the accurate linking of the computer installations dataset to the publication data. For more information, check <https://ror.org/about/>

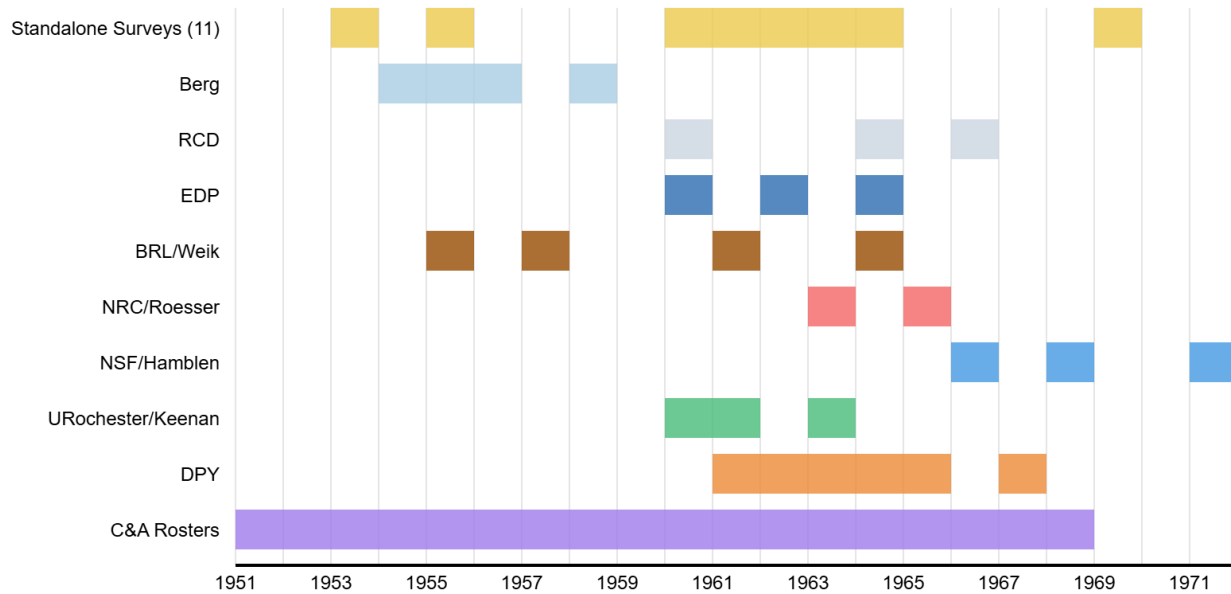


Figure A.1: Yearly coverage of surveys in the computer-installations database, 1950–1971. Each row is a distinct survey source; shaded cells indicate the year(s) in which that source published a computer census or inventory used to extract installations. Database draws on 82 survey-year pairs from 24 distinct sources. The full list of sources and acronyms is given below.

1. Computers and Automation – **Rosters of Organizations in the Computer Field** (1951-1968). Computers and Automation was a general interest magazine about computing technology. Rosters were published in July for each year. The magazine started a separate roster of school, college and university computer centers in 1961. Before that, only a handful of universities were mentioned in the general roster of organizations in the computer machinery field. Source acronym: *c&a*
2. Inventory of Computers in U.S. Higher Education - **Computers in higher education**: report by the John W. Hamblen of the Southern Regional Education Board commissioned by the NSF (1966; 1968; 1971). This is the only source that universally reports computer installation dates. Source acronym: *hamblen*
3. Thomas A. Keenan Surveys – **University of Rochester Annual Survey of University Computing Centers** (1960; 1961; 1963). Keenan conducted six surveys of university computer centers with detailed data for each, starting in 1957. We were able to recover the last three surveys thanks to Melissa Mead at the University of Rochester archives, but the first three seem to have been lost to time. Source acronym: *keenan*

4. National Research Council's Roesser Report - **Digital Needs in Universities and Colleges** (1966; covers 1963 and 1965). This is a survey within the context of a thorough NRC report about computer usage in higher education. Source acronym: *nrc*
5. American Mathematical Society Notices Survey - **Survey of High Speed Computers in Universities** (1962). Source acronym: *amsn*
6. **Survey of Automatic Digital Computers** by the Office of Naval Research (1953). Source acronym: *onr*
7. **Survey of Domestic Electronic Digital Computing Systems** by the Martin Weik of the Ballistic Research Labs (1955; 1957; 1961; 1964). Source acronym: *weik*
8. A study of administrative uses of computers in colleges and universities by **Florida State University** (1962). Source acronym: *fsu*
9. **IBM 650s installation data** from IBM leaflets and website (circa 1956). We thank Max Campbell at the IBM corporate archives for retrieving some of this information. Source acronym: *ibmarchives*
10. **Mathematics in Education**, 1961. Survey by the US Department of Health, Education, and Welfare. Source acronym: *mie*
11. **Datamation Magazine** May, 1962 survey - reporting results from the AMS notices survey and one survey by **Ohio State University**, which we were unable to recover. Source acronym: *datamation*
12. **Business Automation Magazine** survey, August, 1962. A survey of universities on computers and courses. Source acronym: *ba*
13. **Educational Programs and Facilities in Nuclear Science and Engineering** - three surveys (1960; 1962; 1964). One of the best surveys for analog computer usage. Source acronym: *edp*
14. **Data Processing for Management**: a general interest magazine from American Data Processing, Inc. with a section on **installation news**. We recover and process snippets mentioning installations of computers at universities from the news section. Source acronym: *dfpmm*

15. **Data Processing Yearbooks** - surveys of university facilities and courses (1961, 1962, 1963, 1964, 1965, 1967), with certain editions titled “**Computer Yearbook and Directory**” (1965-66 and 1967-68). Source acronym: *dpy*
16. **Research Centers Directory** (1960, 1964, 1966). Focus in research centers in general, reports computer centers and installations as well. Source acronym: *rcd*
17. **Business Electronics Reference Guide** (1954, 1955, 1956, 1958). Early surveys that report installations at firms and universities. Source acronym: *berg*
18. **Survey of Numerical Weather Prediction** by the University of Chicago (1955). Reports computers in use for numerical weather prediction in of universities. Source acronym: *snwp*
19. **Hearings on H.R. 4845** by the U.S. House of Representatives (1965; covers 1964). The hearings detail computers acquired in partnership with the federal government at universities or housed at federal institutions. Source acronym: *adpeh*
20. **Hearing on Use of Electronic Data Processing Equipment** by the Subcommittee on Census and Government Statistics (1963; covers 1963). Source acronym: *uedpeh*
21. Report from Task Force on **Hydrologic Computer Programs** (1963; covers 1963). Source acronym: *hcp*
22. Hearings on **AEC Authorizing Legislation Fiscal Year 1973** by the Joint Committee on Atomic Energy (1972; covers 1969). The hearings report computers acquired by the AEC and sometimes installed in universities. Source acronym: *cfuhef*
23. **Digital Computer Newsletter**, by the Office of Naval Research (1949-1968), as a supplementary source. We search the magazine for scattered textual mentions of installations at universities. Source acronym: *dcn*
24. **Datamation Magazine** (1957-1965). We searched the magazine for scattered mentions of computer installations at universities as a supplementary source. Sometimes in the news section, computer installation at universities were reported. Source acronym: *dmm*

A.3 List of Universities

The 184 universities in our sample are listed in [Table A.1](#) and [Table A.2](#).

Table A.1: Universities In Sample

Universities	
Abilene Christian College	Ohio University
American University	Oklahoma State University
Amherst College	Oregon State University
Arizona State University	Pennsylvania State University
Auburn University	Pomona College
CUNY, Baruch College	Princeton University
Baylor University	Providence College
Boston College	Purdue University
Boston University	Queensborough Community College
Brandeis University	Rensselaer Polytechnic Institute
Brigham Young University	Rice University
Brown University	Rose Polytechnic Institute
Bryn Mawr College	Rutgers University
California Institute Of Technology	Saint Louis University
California State University, Los Angeles	San Diego State University
Carnegie Institute Of Technology	Smith College
Carnegie Mellon University	South Dakota State University
Case Institute Of Technology	Southern Illinois University
Case Western Reserve University	Southern Methodist University
CUNY, City College	Stanford University
Clark University	SUNY At Buffalo
Clemson University	Stephen F. Austin State College
College Of William And Mary	Stevens Institute Of Technology
Colorado School Of Mines	Swarthmore College
Colorado State University	Syracuse University
Columbia University	Temple Junior College
Cornell University	Texas A&M University
Dartmouth College	Texas College Of Arts And Industries
Duke University	Texas Tech
Emory University	The King's College - Pennsylvania
Fairleigh Dickinson University	The Ohio State University
Florida State University	Tufts University
Foothill College	Tulane University
Fordham University	University Of Akron
Franklin Institute	University Of Alabama
George Washington University	University Of Alaska
Georgetown University	University Of Arizona
Georgia Institute Of Technology	University Of Arkansas
Georgia State University	UC, Berkeley
Harvard University	UC, Davis
Harvey Mudd College	UC, Irvine
Haverford College	UC, Los Angeles
Howard University	UC, Riverside
Illinois Institute Of Technology	UC, San Diego
Indiana Institute Of Technology	UC, San Francisco
Indiana University, Bloomington	UC, Santa Barbara
Iowa State University	UC, Santa Cruz
Jackson State College	University Of Chicago
Johns Hopkins University	University Of Cincinnati
Kansas State University	University Of Colorado Boulder
Kent State University	University Of Connecticut
Lehigh University	University Of Delaware
Long Island University	University Of Denver

(Continued on next page)

Table A.2: Universities In Sample (continued)

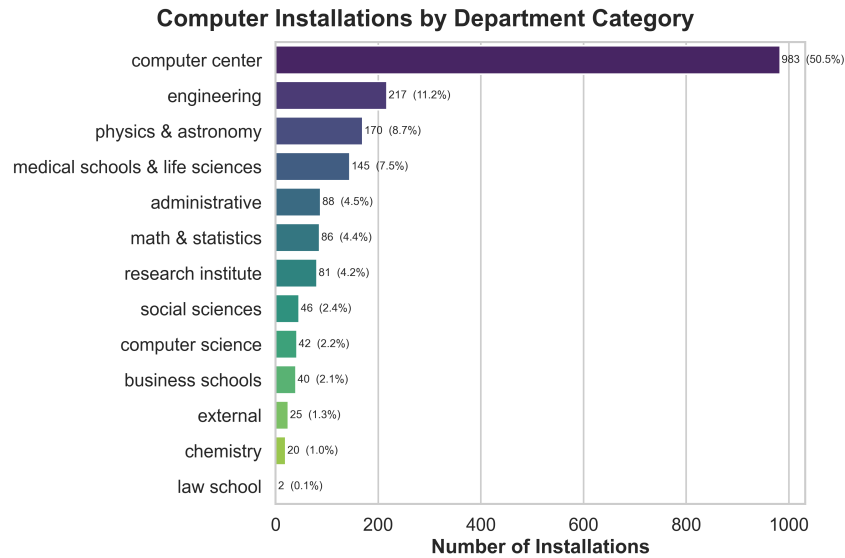
Universities	
Louisiana State University	University Of Florida
Lowell Technological Institute	University Of Georgia
Marquette University	University Of Hawaii
Massachusetts Institute Of Technology	University Of Houston
Michigan State University	University Of Idaho
Mississippi State University	University of Ill., Urbana-Champaign
Missouri Univ Of Science And Tech	University Of Iowa
Montana State University	University Of Kansas
New Mexico Inst Of Mining And Tech	University Of Kentucky
New Mexico State University	University Of Louisville
New School For Social Research	University Of Maine
NY St Col Of Agriculture At Cornell U	University Of Maryland
New York University	University Of Massachusetts, Amherst
North Carolina State University	University Of Miami
North Dakota State University	University Of Michigan, Ann Arbor
Northeastern University	University Of Minnesota
Northern Illinois University	University Of Mississippi
Northwestern University	University Of Missouri
University Of Nebraska	University Of Pennsylvania
University Of Nebraska, Omaha	University Of Pittsburgh
University Of Nevada	University Of Puerto Rico, Mayagüez
University Of New Hampshire	University Of Puerto Rico, Río Piedras
University Of New Mexico	University Of Puget Sound
UNC, Chapel Hill	University Of Rhode Island
University Of North Dakota	University Of Rochester
University Of Notre Dame	University Of South Carolina
University Of Oklahoma	University Of South Florida
University Of Oregon	University Of Southern California
Vanderbilt University	University Of Southwestern Louisiana
Vassar College	University Of Tennessee
Virginia Polytechnic Institute	University Of Texas, Austin
Washington And Lee University	University Of Utah
Washington State University	University Of Vermont
Washington University Of Saint Louis	University Of Virginia
Wayne State University	University Of Washington
Wesleyan University	University Of Wisconsin, Madison
West Virginia University	University Of Wisconsin-Milwaukee
Western Michigan University	University Of Wyoming
Western Reserve University	Utah State University
Wichita State University	Yale University
Williams College	

B Computer Database: Additional Descriptives

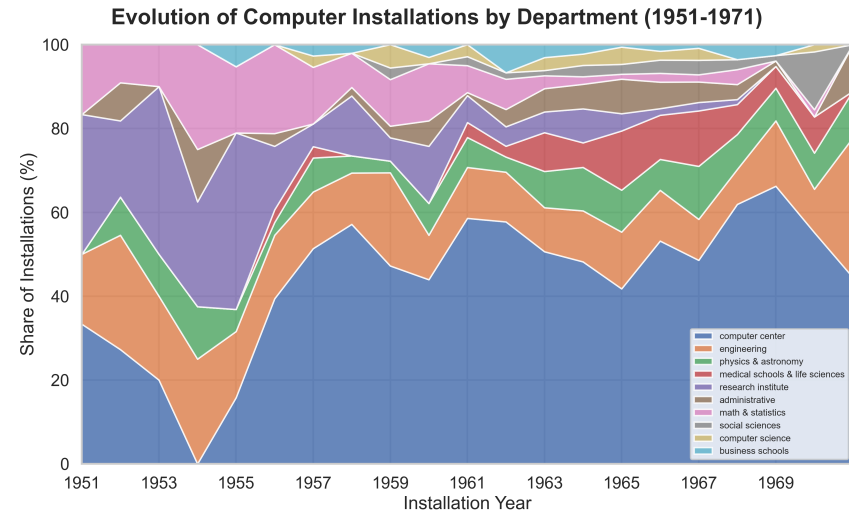
This section reports additional descriptive detail on the computer-installations database, complementing the Computer Database appendix of the main paper, Section B.

To parse raw installation strings into consistent location categories, we use an LLM to create a mapping between unique raw strings and location categories and then manually validate the mapping. In Figure [B.2a](#), we show the distribution of installation by location type. In Figure [B.2b](#), we display the location categories of computer installations over time, using the same data from Figure [B.2a](#).

Tables [B.3](#) and [B.4](#) complement the funding records summarized in that appendix: the former reports the mechanism by which universities acquired their first installation, the latter the funding types across all installations.



(a) Number of installations by department category



(b) Evolution of installations by department category over time

Figure B.2: Department/unit categories of computer installations in the 1950–1971 US university sample. Panel (a): bars show the number of installations assigned to each category; shares computed over installations with an identifiable department (omitting the \approx 10% of records lacking location information). Panel (b): yearly installations by category over 1950–1970, normalized to 100% within each year so colors sum to the year's total. Categories include computing centers, engineering departments, physical-science departments, business/administration, and other academic units.

Table B.3: Mechanism by which universities acquired their first installation, $N = 184$.

First installation	Number of universities with first installation	Share
Purchase	33	17.9%
Lease	30	16.3%
Built in-house	11	6.0%
Donation in kind	6	3.3%
Gratis (vendor)	4	2.2%
Unknown (surveyed, undetermined)	12	6.5%
No funding record found	88	47.8%
Total	184	100%

Notes: The first installation is defined as each university's first general-purpose digital computer, the same definition used to set the treatment date in the panel analyses of the main paper, Section 5. "No funding record found" identifies universities for which we were unable to recover an acquisition mechanism for the first installation.

Table B.4: Funding type across all installations, $N = 184$ universities, 2,138 installation rows.

Funding type	Number of universities with ≥ 1 such-funded installation	Share of all installations
NSF	100 (54.3%)	6.8%
University budget	92 (50.0%)	7.0%
Manufacturer	91 (49.5%)	7.3%
Federal civilian, non-NSF	66 (35.9%)	2.9%
Military (incl. ARPA/DARPA)	42 (22.8%)	2.7%
Industry / consortium	30 (16.3%)	0.9%
Foundation	22 (12.0%)	1.0%
Other	16 (8.7%)	0.6%
Alumni / named gift	13 (7.1%)	0.7%
No funding record at any installation	39 (21.2%)	79.4%

Notes: Columns do not sum because a university or installation can carry multiple funding records. "Military" includes any DoD or military-branch funding (Army, Navy, Air Force, ONR, Signal Corps, Aberdeen, ARPA/DARPA, FFRDC contracts). The first numeric column counts universities with at least one installation in that category; the second is the share of all 2,138 installation rows tagged with that funding type.

C Computer Usage Across Disciplines and Diffusion

This appendix presents robustness checks and further results for the descriptive diffusion analysis in Section 4 of the main paper.

Figure C.3 complements the main paper’s time series view in Section 4.2 with the cross-sectional distribution of computer papers across OpenAlex domains for the full period. Physical Sciences account for over 70% of computer-related papers, against roughly 40% of all searchable papers, confirming the leading role of that domain.

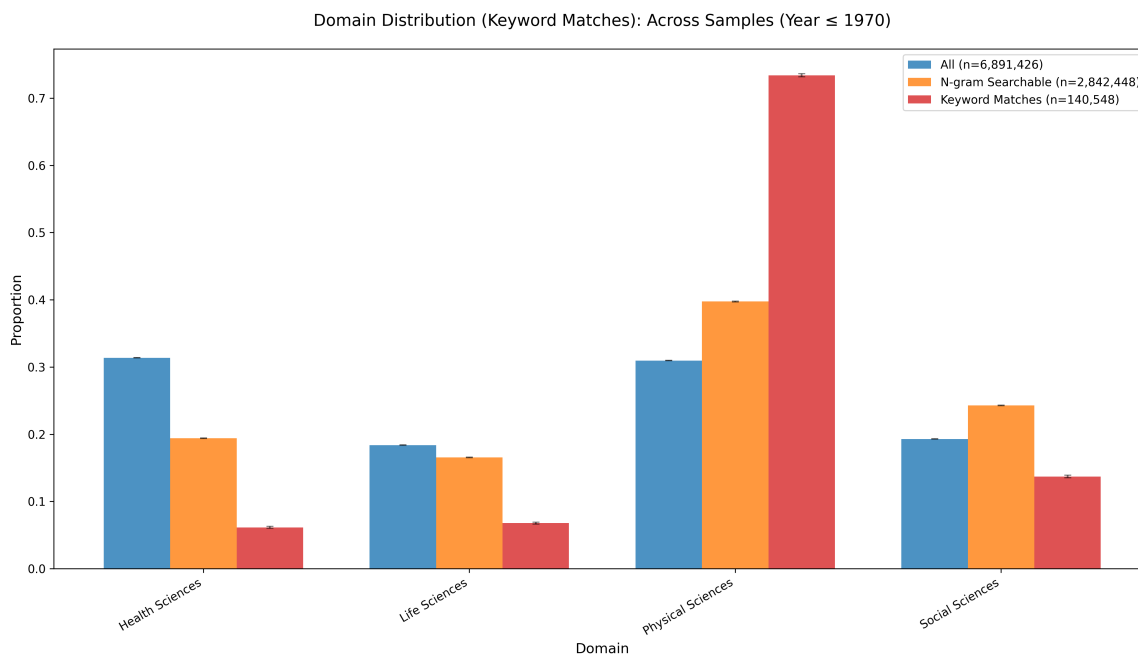


Figure C.3: Distribution of computer papers across OpenAlex domains. The plot compares three slices: (a) all papers; (b) papers we can search full-text n-grams for in OpenAlex; and (c) papers flagged as using or mentioning computers from keywords. Bars of the same color correspond to each slice and add to 1. Plot includes “computer” as a keyword.

In Figure C.4, we show that qualitative results of Figure C.3 do not change if we exclude “computer” as a standalone keyword or if we use LLMs to classify computer use or mention from the full text of papers instead.

In Figure C.5, we plot the distributions by field within each domain, using keyword matches including “computer.”² In relative terms, Engineering, Physics & Astronomy, and Computer Science, are the more computer intensive field within the Physical Sciences. Within

²For conciseness, we do not plot results excluding computers or using LLM classifications, but results are close and qualitatively unchanged.

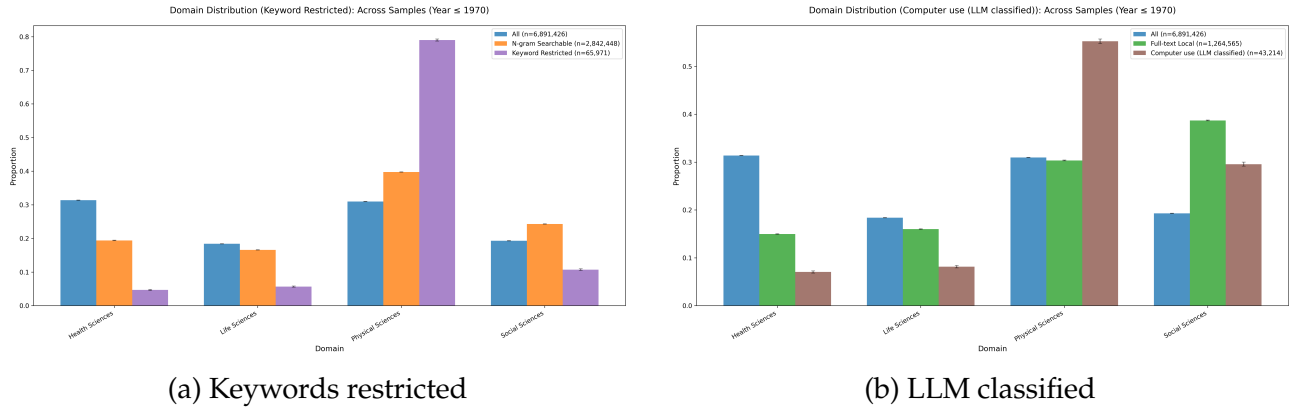
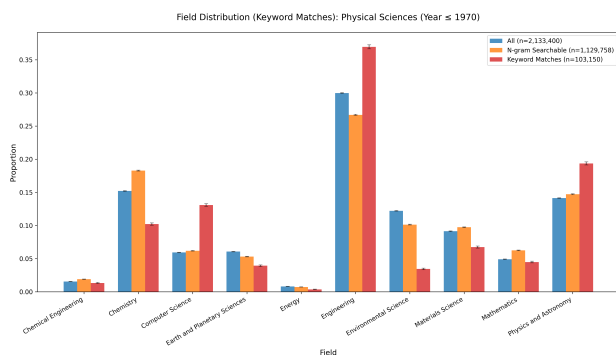


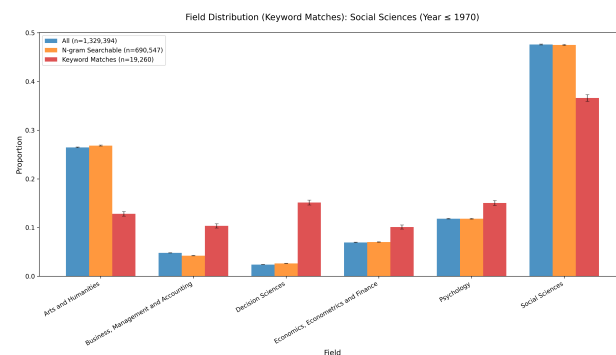
Figure C.4: Robustness variants of the domain-level distribution of computer papers shown in Figure C.3. Panel (a): restricted keyword set (excludes the word “computer” itself). Panel (b): LLM-classified computer papers from the local full-text sample. Each panel compares three slices: (i) all searchable papers, (ii) papers flagged as computer papers by the given method, (iii) baseline share of each domain in searchable OpenAlex. Bars of the same color (slice) add to 1 across domains within each panel.

the Social Sciences, Decision Sciences and Psychology lead the pack.

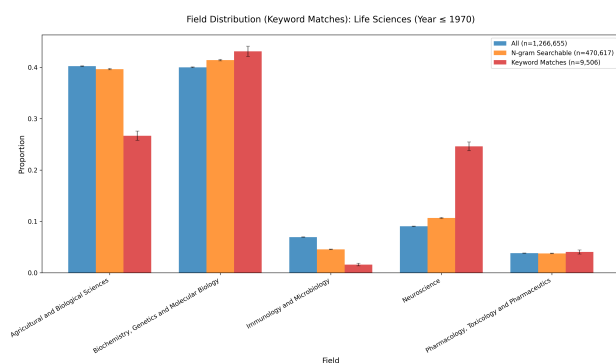
Zooming in selected subjects (Figure C.6), we see some subfields are very computer-intensive: Numerical Analysis and Statistics & Probability within Mathematics; Aerospace Engineering within Engineering; Nuclear & High Energy Physics within Physics & Astronomy; Management & Operations Research within Decision Sciences.



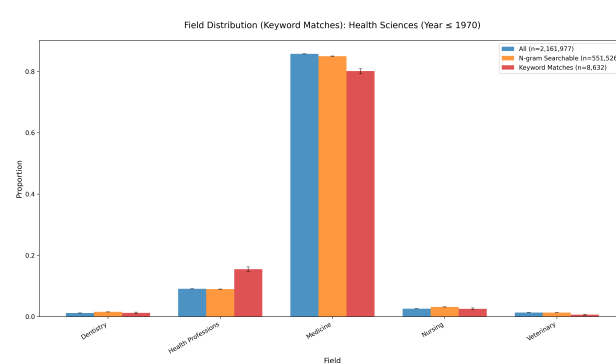
(a) Physical Sciences



(b) Social Sciences

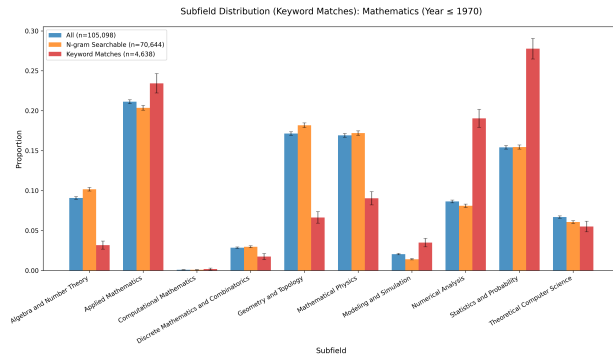


(c) Life Sciences

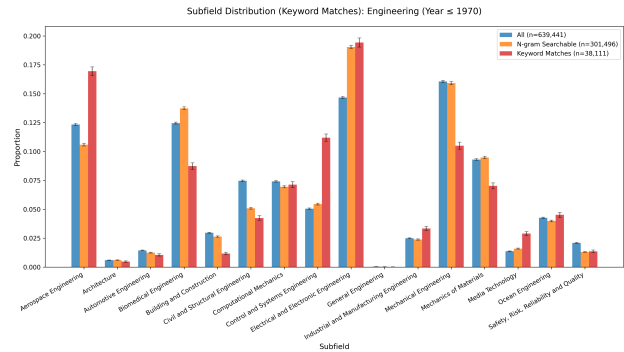


(d) Health Sciences

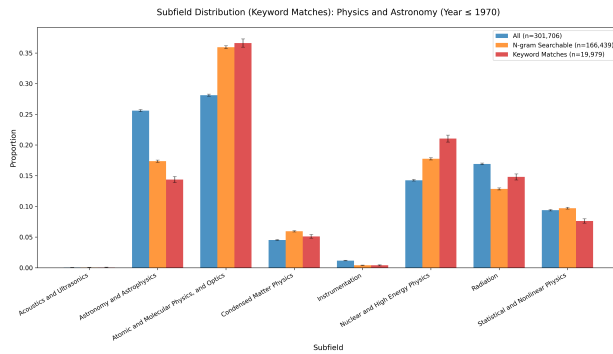
Figure C.5: Distribution of computer papers within each OpenAlex domain (Physical, Social, Life, Health Sciences). Within each domain panel, bars compare the field-level composition across three slices: (a) all papers, (b) searchable full-text papers, (c) papers flagged for computer keywords. Within-panel bars of the same color add to 1, so these are within-domain compositional shares and not cross-domain comparisons. Plots include “computer” as a keyword.



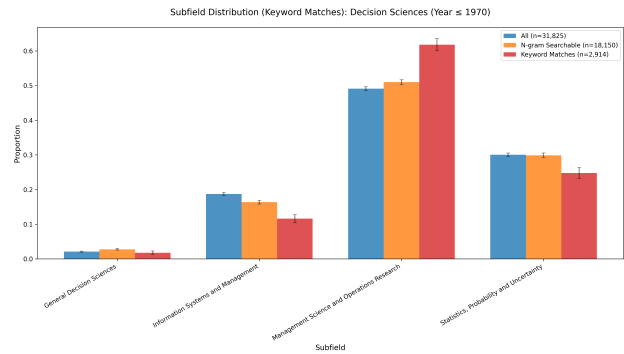
(a) Mathematics



(b) Engineering



(c) Physics & Astronomy



(d) Decision Sciences

Figure C.6: Distribution of computer papers across subfields of four selected fields: Mathematics, Engineering, Physics & Astronomy, and Decision Sciences. Within each field panel, bars compare subfield-level composition across three slices: (a) all papers, (b) searchable full-text papers, (c) papers flagged for computer keywords. Within-panel bars of the same color add to 1, so values are within-field compositional shares. Plots include “computer” as a keyword. Fields selected for visibly sharp pre-digital numerical-intensity gradients.

D Computer Use Classification Details

We assign the 60 lower-level computer-use clusters to 11 high-level clusters as follows:

1. Numerical Computation & Mathematical Modeling

- 0: Solve quantum mechanical equations
- 1: Solve equations for stress and stability
- 8: Refine crystal structures
- 16: Numerically evaluate mathematical functions
- 19: Implement algorithms for matrix algebra
- 22: Develop numerical methods for differential equations
- 26: Solve theoretical equations for nuclear structure
- 41: Numerical analysis for statistical methods
- 49: Verify conjectures in number theory
- 60: Analyze biochemical kinetics and macromolecular properties

2. Statistical Analysis & Data Processing

- 5: Calculate radiation dose
- 12: Statistical analysis for medical studies
- 13: Process and analyze earth-surface data
- 15: Statistical analysis for agricultural experiments
- 17: Analyze research data with statistical software
- 25: Automate scoring and analysis of educational data
- 33: Econometric modeling
- 34: Analyze biomedical data
- 39: Factor analysis on psychological data
- 42: Statistical analysis for archaeology
- 43: Statistical analysis for ecology
- 44: Statistical analysis of survey data
- 45: Statistical analysis for psychology

3. System Simulation & Dynamic Modeling

- 14: Model evolutionary dynamics
- 27: Analog and hybrid simulation
- 29: Model atmospheric dynamics
- 31: Model chemical kinetics and thermodynamics
- 32: Numerical modeling of geophysical systems
- 36: Numerically model physical phenomena
- 40: Model flow for water resources
- 46: Simulate complex systems
- 48: Numerical modeling of plasma
- 53: Numerically model fluid systems
- 54: Simulate physiological systems

4. Optimization & Operations Planning

- 2: Implement optimization algorithms
- 28: Model financial instruments
- 51: Analysis and planning in agriculture

5. Data Acquisition, Processing & Signal Enhancement

- 3: Model and visualize geographic systems
- 6: Process data from analytical instruments
- 9: Process astronomical data and model stellar systems
- 18: Perform signal averaging on physiological recordings
- 38: Analyze particle-detector data with Monte Carlo simulation

6. Data Management & Administrative Processing

- 7: Automate business data processing
- 11: Punch-card medical record tabulation
- 21: Automate library operations
- 52: Process administrative data with punch-cards

7. Real-time Control & Instrumentation

- 10: Automate real-time control for experiments
- 24: Simulate and control space systems
- 59: Integrated systems for simulation, control, and analysis

8. Core Computing: Hardware, Software & Theory

- 4: Foundational programming and system software
- 50: Theoretical foundations of computation and logic
- 56: Design and analyze early computer hardware
- 58: Algorithms for computational statistics and data management

9. AI, Language, Sound & Media Processing

- 20: Computational linguistics and machine translation
- 23: Analyze and generate musical sound

10. Conceptual, Philosophical & Societal Analysis

- 30: Computers referenced incidentally or metaphorically
- 35: Conceptualizing intelligence and cybernetic systems
- 47: Computerization of gov't and soc policy

11. Educational Technology & Digital Applications

- 37: Computer-assisted instruction
- 57: Applying digital technologies in education

In Figure D.7, we plot the evolution of computer-usage categories over time by domain. Numerical and modeling uses dominate uses in the Physical Sciences, whereas on other domains Statistical Analysis & Data Processing dominates and increases clearly over time.

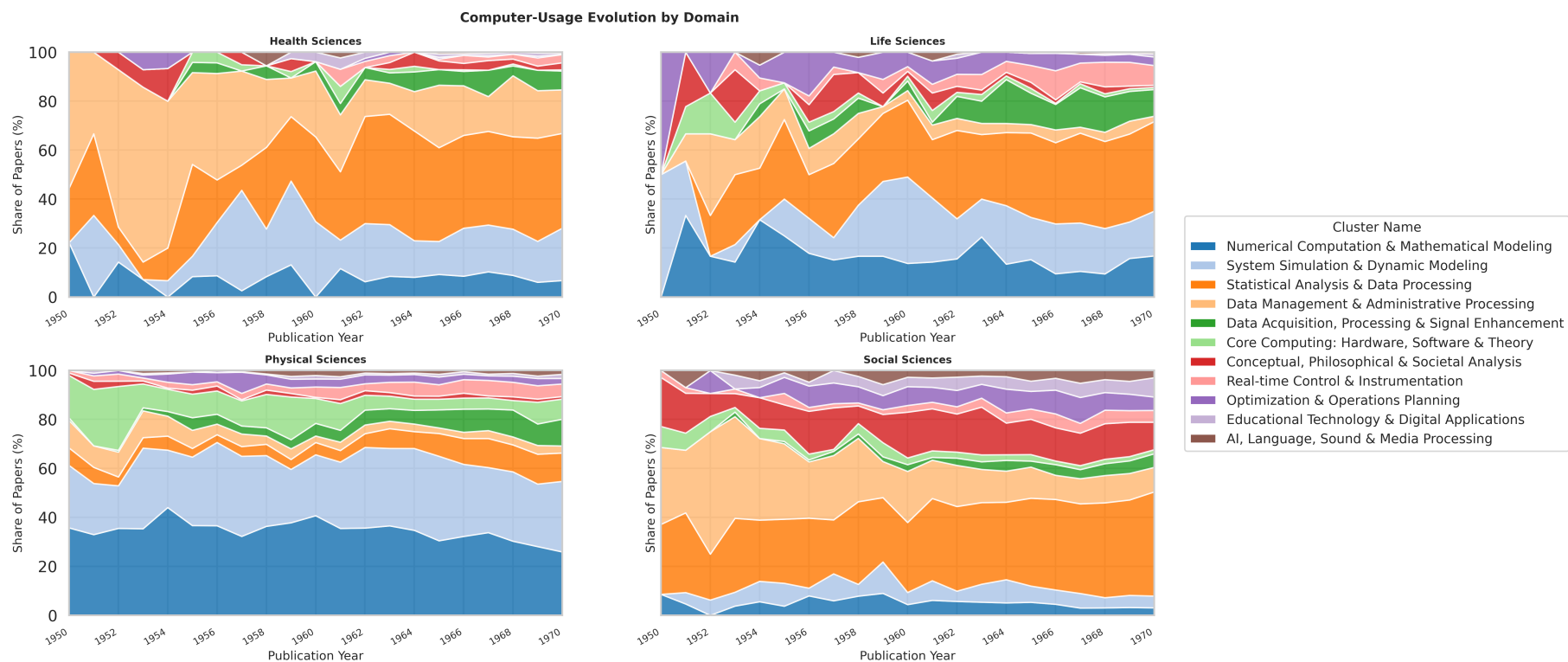


Figure D.7: Evolution of computer usage category shares over time by OpenAlex domain (Physical, Life, Health, Social Sciences), 1950–1970, clustered research-article sample ($N = 30,959$). Categories are the 11 high-level clusters listed above. Each panel shows yearly shares within the domain-specific subset, normalized to 100%.

E Paper-Level Outcomes: Embedding Controls and Patent Citations

This section reports additional paper-level results complementing the Paper Characteristics appendix of the main paper, Section D: a robustness check that additionally partials out semantic embeddings from paper titles and abstracts together with SPECTER embeddings that capture citation-neighborhood position, and regressions on patent citations and other outcomes.³

Table E.5 revisits the citation outcomes from Section 4.5 of the main paper under this richer control set. The estimated citation premia remain positive and statistically significant across all reported outcomes, including top-tail citation status and citations accrued within five years of publication.

Table E.5: Citation Outcomes with Embedding Controls

	Log cites	Top 10%	Top 1%	FWCI	C5	Cite pct.
Computer-Keyword Flag	0.196*** (0.005)	0.034*** (0.001)	0.005*** (0.001)	0.337*** (0.019)	0.530*** (0.042)	0.031*** (0.001)
Observations	1,882,763	1,882,654	1,882,654	1,882,004	1,882,763	1,882,654
R^2	0.002	0.000	0.000	0.000	0.000	0.001
Mean of Dep Var	-0.016	0.046	0.001	0.171	0.644	0.003

Notes: Columns report second-stage OLS coefficients from a double-machine-learning robustness specification. In a first stage, both the computer-keyword indicator and each outcome are residualized with respect to author, publication year, university, and primary-topic fixed effects; controls for the number of authors and NSF grants awarded to the paper; dense title and abstract embeddings; and SPECTER paper embeddings. Residuals are estimated with 3-fold cross-fitting and SGD, after which the residualized outcome is regressed on the residualized treatment. Paper-author observations are weighted by the inverse number of authors, and second-stage standard errors are clustered at the paper level (OpenAlex Work ID). Sample restricted to the 1947–1975 searchable-full-text analysis sample used in the paper-level regressions. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table E.6 shows regressions on patent citations, and other outcomes. Its main result is that papers mentioning computers are not more likely to be cited by patents.

³The title-and-abstract embeddings are 1024-dimensional vectors from the stella_en_1.5B_v5 model, computed on each OpenAlex work’s title and abstract (January 2025 snapshot); we use the pre-computed, openly-licensed release available on [Hugging Face](#). The SPECTER embeddings are drawn from SciSciNet.

Table E.6: Other Outcomes

	# Patent Cites	# Countries	# Institutions	# Refs
Computer-Keyword Flag	0.244 (0.197)	0.003*** (0.001)	0.013*** (0.002)	1.967*** (0.042)
# Authors	0.013 (0.012)	0.039*** (0.001)	0.140*** (0.002)	0.080*** (0.010)
NSF grants (paper)	-0.084 (0.154)	-0.006 (0.004)	0.009 (0.013)	4.132*** (0.297)
Observations	3,903,691	3,903,691	3,903,691	3,903,691
R^2	0.289	0.490	0.537	0.454
Mean of Dep Var	0.259	0.605	0.643	5.492
Author/Year/Univ FE	Yes	Yes	Yes	Yes
Subject FE	Topic	Topic	Topic	Topic

Notes: Paper-author observations, weighted by the inverse number of authors. Treatment is an indicator for the presence of computer-related keywords in the paper's full text. All models control for the number of authors and the number of NSF grants awarded to the paper, plus author, publication year, university, and primary topic fixed effects. Standard errors are clustered at the paper level (OpenAlex Work ID). Sample restricted to papers with at least one in-sample university affiliation and searchable full text in OpenAlex. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

F Citation Roles: Bucket Premium Scores and Examples

F.1 Regression Specification and Bucket Scores

The classification procedure described in Section 4.6 of the main paper assigns a role or bucket to each citing-cited paper pair. We then estimate the bucket premium scores at the cited-paper level. Let n_{pk} be the number of classified incoming citations to paper p that fall in citation role bucket k , let $N_p = \sum_{k=0}^4 n_{pk}$, and define $s_{pk} = n_{pk}/N_p$. Omitting bucket 0 (*Technical evidence and protocols*) in estimation and re-centering the fitted coefficients so that the five reported bucket scores average to zero, we estimate

$$FWCI_p = \alpha + X_p' \gamma + \lambda_{\text{year}(p)} + \delta_{\text{domain}(p)} + \sum_{k=1}^4 \theta_k s_{pk} + u_p,$$

where $FWCI_p$ is the paper's *Field-Weighted Citation Index*, and X_p includes number of authors and NSF grants. Figure F.8 reports these re-centered coefficients with and without a control for $\log(1 + \text{classified citation count})$. Canonical sources remain the highest premium bucket and established findings the lowest, even after partialling out citation volume.

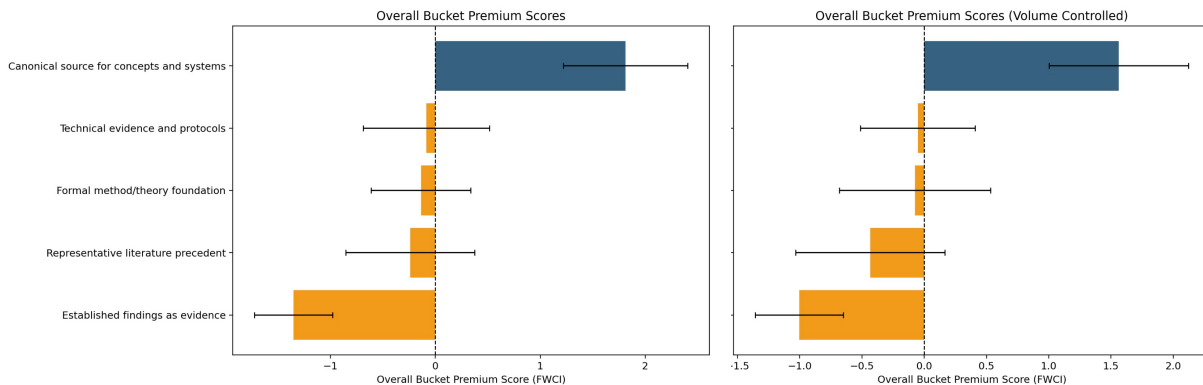


Figure F.8: Citation-function bucket premium scores. Cited-paper regression of FWCI on citation-function shares, controlling for number of authors, NSF grants, cited-paper year fixed effects, and domain fixed effects. $N = 15,938$ cited papers with at least one classified citation. The left panel omits citation volume, so coefficients summarize each bucket's overall association with citation impact; the right panel adds $\log(1 + \text{classified citation count})$ to partial out citation volume. Scores are re-centered so all five buckets are directly comparable; whiskers show 95% CIs.

F.2 Citation Function Examples

Table F.7 summarizes the substantive distinction between the five citation-function buckets. These examples are condensed from the full cluster example brief generated from centroid-nearest papers, nearby outside examples, and matched computer-versus-control pairs.

Table F.7: Illustrative Citation-Function Buckets

Bucket	What the citation is doing	Illustrative example
Technical evidence and protocols	Supplies a protocol, benchmark, calibration, or measurement input.	“As previously described” citations to earlier activation or purification procedures, or to benchmark values used to calibrate the current analysis.
Established findings as evidence	Provides factual support for a substantive claim.	Waterman and Horch (1966) cited as evidence on polarization sensitivity, or Daley et al. (1971) cited for measured radar backscatter results.
Representative literature precedent	Shows that a phenomenon, debate, or method already exists in the literature.	Siegel (1974) or Posner, Nissen, and Klein (1976) cited as one example in a broader list of prior work.
Formal method or theory foundation	Gives the estimator, derivation, or analytical framework the paper actually uses.	The fast Fourier transform of Cooley and Tukey, or the Kalman filter, cited for the algorithm the citing paper’s analysis actually runs.
Canonical source for concepts and systems	Establishes the recognized origin of a concept, language, system, or benchmark.	Dijkstra’s “Go To Statement Considered Harmful” cited as the origin of structured programming, for provenance rather than as direct empirical evidence.

Notes: Examples are shortened descriptions of representative citation contexts from the classified citation-link sample. They are intended to clarify the meaning of each bucket rather than to estimate any additional coefficients.

G Reference Age and Sleeping Beauties: Embedding Controls

This section reports a robustness check for the reference-age and dormancy results in Section 4.7 of the main paper that applies the same embedding controls used in Table E.5: semantic embeddings from titles and abstracts together with SPECTER embeddings that capture citation-neighborhood position. The overall patterns remain similar to the baseline specifications. Computer papers still cite younger references at the minimum and median, older references in the right tail, and references with higher Sleeping Beauty scores among the oldest cited works.

Table G.8: Reference Age and Dormancy Outcomes with Embedding Controls

<i>Panel A: Reference age</i>				
	Mean	Min	Median	Max
Computer-Keyword Flag	-0.048*	-0.121***	-0.086***	0.851***
	(0.023)	(0.015)	(0.022)	(0.068)
Observations	1,267,437	1,267,437	1,267,437	1,267,437
R^2	0.000	0.000	0.000	0.000
Mean of Dep Var	-0.405	0.062	0.220	-0.255
<i>Panel B: Reference dormancy (Sleeping Beauty scores)</i>				
	Mean	Median Ref.	Oldest	Age90 Mean
Computer-Keyword Flag	19494.863***	16492.462***	48145.559***	38288.376***
	(3049.162)	(3818.570)	(9517.431)	(8363.198)
Observations	1,266,675	1,262,071	1,263,557	1,264,247
R^2	0.000	0.000	0.000	0.000
Mean of Dep Var	10413.798	-61481.215	52900.751	36815.254

Notes: Columns report second-stage OLS coefficients from the same double-machine-learning robustness specification used in Table E.5. In a first stage, both the computer-keyword indicator and each outcome are residualized with respect to author, publication year, university, and primary-topic fixed effects; controls for the number of authors and NSF grants awarded to the paper; dense title and abstract embeddings; and SPECTER paper embeddings. Residuals are estimated with 3-fold cross-fitting and SGD, after which the residualized outcome is regressed on the residualized treatment. Paper-author observations are weighted by the inverse number of authors, and second-stage standard errors are clustered at the paper level (OpenAlex Work ID). Sample restricted to the 1947–1975 searchable-full-text analysis sample used in the paper-level regressions. Panel A reports reference-age outcomes. Panel B reports cited-reference Sleeping Beauty scores, including the score of the median-age cited reference. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

H Author-Level Results: Early versus Late Adopters

Table H.9 restricts the sample to computer adopters and compares early and late adopters. We use adoption lag – the number of years elapsed between the first computer publication and the acquisition of a computer – as a proxy for speed of adoption. Results suggest no significant difference between early and late adopters. Note, however, that our sample ends in 1970, so we are comparing authors within the set of relatively early adopters.

Table H.9: Author-Level Regressions: Early vs. Late Adopters

	(1)	(2)	(3)	(4)	(5)
	Log Works	Log Cites	H-Index	# Topics	# Affiliations
Adoption Lag (Freq)	0.00817 (0.00440)	0.0109 (0.00628)	-0.0391 (0.0368)	0.0713*** (0.0179)	-0.0405* (0.0177)
Number of Works		0.00515*** (0.000364)	0.0782*** (0.00427)	0.00283*** (0.000380)	0.0118*** (0.000983)
Observations	6,141	6,141	6,141	6,141	6,141
R^2	0.373	0.601	0.759	0.341	0.524
Mean of Dep Var	4.358	7.271	23.65	24.01	6.636
Affiliation/Cohort/Topic FE	Yes	Yes	Yes	Yes	Yes

Notes: Sample restricted to computer adopters with publications both before and after the computer installation year at their affiliations. Regressor: *Adoption Lag (Freq)* is the number of years between the author’s first computer-flagged publication and the installation year at their affiliations, where the installation year is a publication-count-weighted mean across the author’s affiliations; larger values indicate slower adoption relative to local availability. Outcomes are measured over the author’s entire career and defined as in the author-level regressions in Section 4.8 of the main paper. Columns (2)–(5) control for number of works. Affiliation, cohort, and topic fixed effects are likewise defined as in the main paper’s author-level regressions. Standard errors clustered at the affiliation level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

I Subfield-to-Department Crosswalk

This section lists the baseline many-to-many mapping from OpenAlex subfields to plausible academic departments, used in the department-level exposure diagnostics in Appendix H of the main paper. The sample covers the 136 Physical and Social Sciences subfields in the analysis; computer centers, administrative data-processing units, research institutes, and computer science are not treated as academic departments. A subfield mapping to m_s departments enters each with crosswalk weight $1/m_s$; 84 of the 136 subfields map to more than one department (at most three).

Physical Sciences

Chemical Engineering

- Bioengineering → bioengineering
- Catalysis → chemical engineering, chemistry
- Chemical Health and Safety → chemical engineering, industrial engineering
- Filtration and Separation → chemical engineering
- Fluid Flow and Transfer Processes → chemical engineering, mechanical engineering
- Process Chemistry and Technology → chemical engineering, chemistry

Chemistry

- Analytical Chemistry → chemistry
- Electrochemistry → chemistry, chemical engineering
- Inorganic Chemistry → chemistry
- Organic Chemistry → chemistry
- Physical and Theoretical Chemistry → chemistry, physics
- Spectroscopy → chemistry, physics

Earth and Planetary Sciences

- Atmospheric Science → meteorology, earth science
- Earth-Surface Processes → earth science, geography
- Geochemistry and Petrology → earth science, chemistry
- Geology → earth science
- Geophysics → earth science, physics
- Oceanography → oceanography, earth science
- Paleontology → earth science, biology
- Space and Planetary Science → astronomy, physics, earth science

Energy

- Energy Engineering and Power Technology → electrical engineering, mechanical engineering
- Fuel Technology → chemical engineering, mechanical engineering
- General Energy → mechanical engineering, electrical engineering

ing, electrical engineering

- Nuclear Energy and Engineering → nuclear engineering, physics
- Renewable Energy, Sustainability and the Environment → environmental engineering, mechanical engineering

Engineering

- Aerospace Engineering → aerospace engineering, mechanical engineering
- Architecture → architecture
- Automotive Engineering → mechanical engineering
- Biomedical Engineering → bioengineering, electrical engineering
- Building and Construction → civil engineering, architecture
- Civil and Structural Engineering → civil engineering
- Computational Mechanics → mechanical engineering, applied mathematics
- Control and Systems Engineering → electrical engineering, mechanical engineering
- Electrical and Electronic Engineering → electrical engineering
- General Engineering → engineering
- Industrial and Manufacturing Engineering → industrial engineering
- Mechanical Engineering → mechanical engineering
- Mechanics of Materials → mechanical engineering, materials science
- Media Technology → electrical engineering, communication
- Ocean Engineering → civil engineering, mechanical engineering, oceanography
- Safety, Risk, Reliability and Quality → industrial engineering

Environmental Science

- Ecological Modeling → ecology, applied mathematics
- Ecology → ecology, biology
- Environmental Chemistry → chemistry, environmental science
- Environmental Engineering → environmental engineering, civil engineering
- Global and Planetary Change → earth sci-

ence, geography

- Health, Toxicology and Mutagenesis → public health, biology
- Industrial and Manufacturing Engineering → industrial engineering
- Management, Monitoring, Policy and Law → public policy, law, environmental science
- Nature and Landscape Conservation → ecology, environmental science
- Pollution → environmental engineering, chemistry
- Water Science and Technology → civil engineering, environmental engineering

Materials Science

- Biomaterials → materials science, bioengineering
- Ceramics and Composites → materials science
- Electronic, Optical and Magnetic Materials → materials science, physics, electrical engineering
- General Materials Science → materials science
- Materials Chemistry → materials science, chemistry
- Metals and Alloys → materials science, metallurgy
- Polymers and Plastics → materials science, chemistry, chemical engineering
- Surfaces, Coatings and Films → materials science, chemistry

Mathematics

- Algebra and Number Theory → mathematics
- Applied Mathematics → applied mathematics, mathematics
- Computational Mathematics → applied mathematics, mathematics
- Discrete Mathematics and Combinatorics → mathematics
- Geometry and Topology → mathematics
- Mathematical Physics → mathematics, physics
- Modeling and Simulation → applied mathematics, statistics
- Numerical Analysis → applied mathe-

- matics, mathematics
- Statistics and Probability → statistics, mathematics
- Theoretical Computer Science → mathematics

Physics and Astronomy

- Acoustics and Ultrasonics → physics, mechanical engineering
- Astronomy and Astrophysics → astronomy, physics
- Atomic and Molecular Physics, and Optics → physics
- Condensed Matter Physics → physics, materials science
- Instrumentation → physics, electrical engineering
- Nuclear and High Energy Physics → physics
- Radiation → physics, public health
- Statistical and Nonlinear Physics → physics, applied mathematics

Social Sciences

Arts and Humanities

- Archeology → anthropology, classics
- Classics → classics
- Conservation → art history, environmental science
- General Arts and Humanities → humanities
- History → history
- History and Philosophy of Science → history, philosophy
- Language and Linguistics → linguistics
- Literature and Literary Theory → english literature
- Museology → art history
- Music → music
- Philosophy → philosophy
- Religious studies → religious studies
- Visual Arts and Performing Arts → art history, music

Business, Management and Accounting

- Accounting → accounting
- Business and International Management → management
- Industrial relations → industrial relations, economics
- Management Information Systems → management, information science
- Management of Technology and Innovation → management, industrial engineering
- Marketing → marketing
- Organizational Behavior and Human Resource Management → management, psychology
- Strategy and Management → management
- Tourism, Leisure and Hospitality Management → management

Decision Sciences

- General Decision Sciences → operations research, statistics
- Information Systems and Management → information science, management
- Management Science and Operations Research → operations research, industrial engineering
- Statistics, Probability and Uncertainty → statistics

Economics, Econometrics and Finance

- Economics and Econometrics → economics
- Finance → finance, economics
- General Economics, Econometrics and Finance → economics, finance

Psychology

- Applied Psychology → psychology
- Clinical Psychology → psychology
- Developmental and Educational Psychology → psychology, education

- Experimental and Cognitive Psychology → psychology
- General Psychology → psychology
- Neuropsychology and Physiological Psychology → psychology, biology
- Social Psychology → psychology, sociology

Social Sciences

- Anthropology → anthropology
- Archeology → anthropology, classics
- Communication → communication
- Cultural Studies → sociology, anthropology
- Demography → sociology, economics
- Development → economics, political science
- Education → education
- Gender Studies → sociology
- General Social Sciences → sociology, political science
- Geography, Planning and Development → geography, urban planning
- Health → public health
- Human Factors and Ergonomics → psychology, industrial engineering
- Law → law
- Library and Information Sciences → library and information science
- Life-span and Life-course Studies → sociology, psychology
- Linguistics and Language → linguistics
- Political Science and International Relations → political science
- Public Administration → public administration, political science
- Safety Research → industrial engineering, public health
- Sociology and Political Science → sociology, political science
- Transportation → civil engineering, urban planning
- Urban Studies → urban planning, sociology

J Results for Health and Life Sciences

As discussed in Section 6.2 of the main text, the triple-difference estimates for the Life and Health Sciences are, for most of our main outcomes, either null or noisy around zero. This is consistent with the weaker pre-digital numerical-intensity gradient in these domains: where the 1940–1944 measure is only weakly predictive of later computer penetration within a domain, first-stage power is limited and DDD effects are muted. In the Life Sciences a few outcomes are additionally hard to interpret because of negative placebo coefficients or pre-trends. Figures J.9 and J.10 present the Life- and Health-Sciences counterparts of the outcomes shown in the main text; the estimator (Callaway and Sant’Anna, 2021), sample, and construction of the gap outcome Y_{ut}^* are identical to those in Section 5 of the main text.

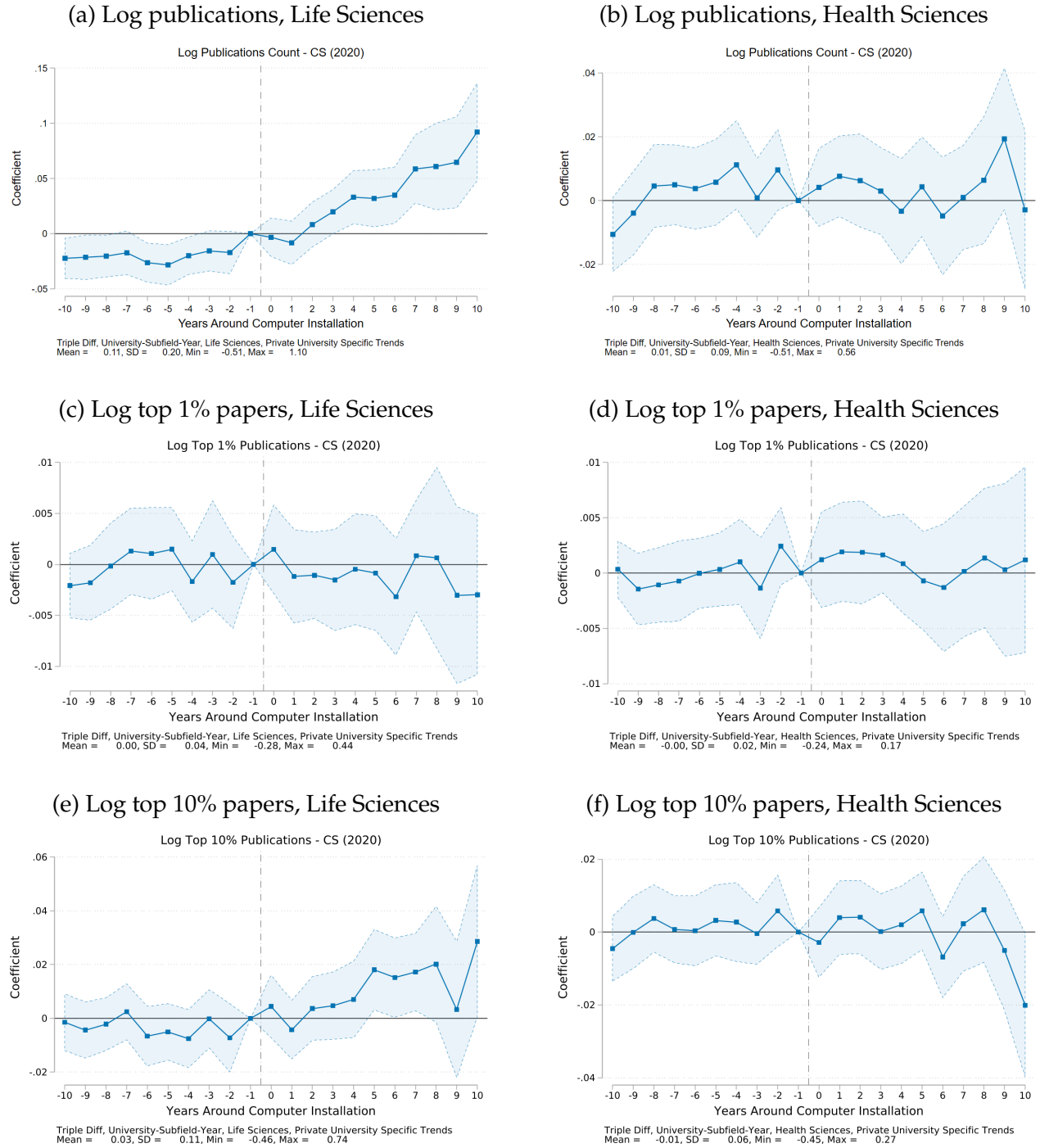
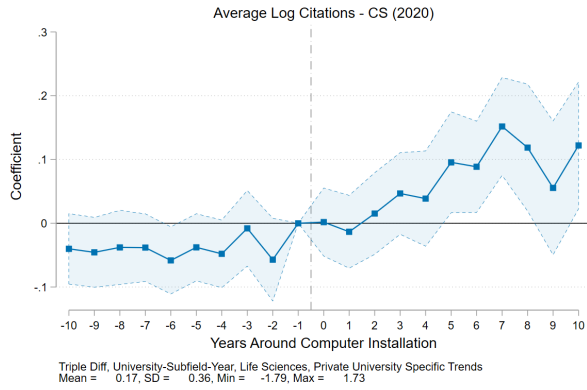
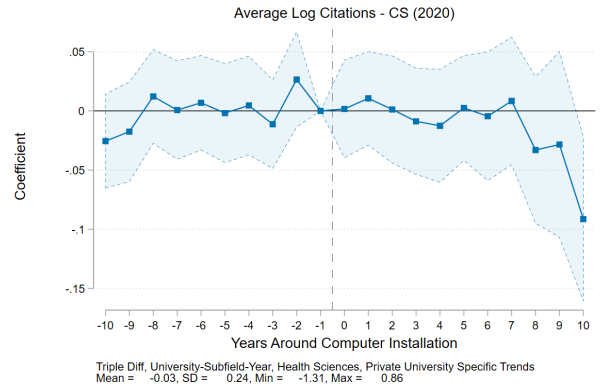


Figure J.9: DDD event studies for *log publication counts* (panels a–b), the *log count of top 1% papers* (panels c–d), and the *log count of top 10% papers* (panels e–f), Life and Health Sciences. Y_{ut}^* is the within-university gap between compute-amenable and less-amenable subfields (defined by the median 1940–1944 numerical-intensity share within each domain); Callaway and Sant’Anna (2021) estimator; $N = 184$ treated universities. Points are aggregated cohort ATTs with 95% CIs; the dashed vertical line marks first installation. These are the Life- and Health-Sciences counterparts of the main-text publication-count and top-cited-paper figures for the Physical and Social Sciences. Estimates include separate linear time trends for private and public universities.

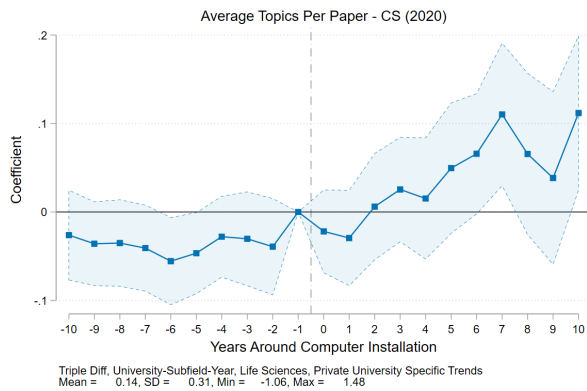
(a) Log avg. citations per paper, Life Sciences



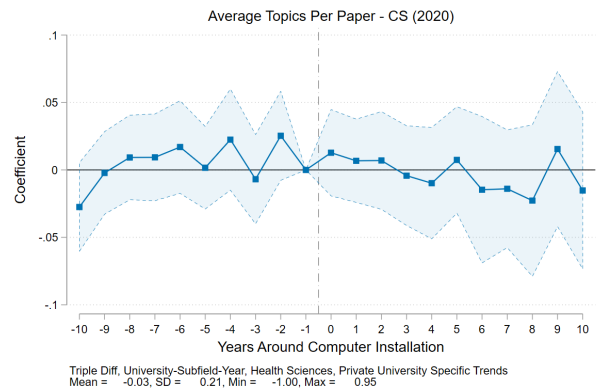
(b) Log avg. citations per paper, Health Sciences



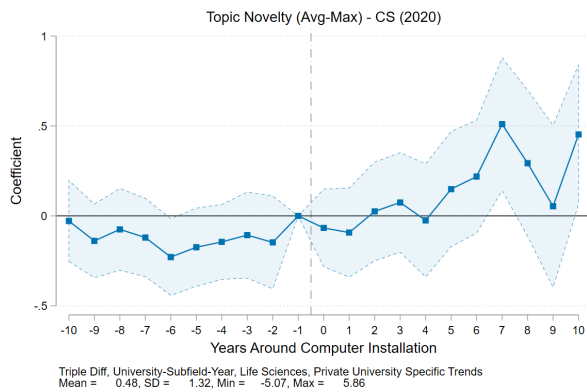
(c) Avg. topics per paper, Life Sciences



(d) Avg. topics per paper, Health Sciences



(e) Topic-combination novelty, Life Sciences



(f) Topic-combination novelty, Health Sciences

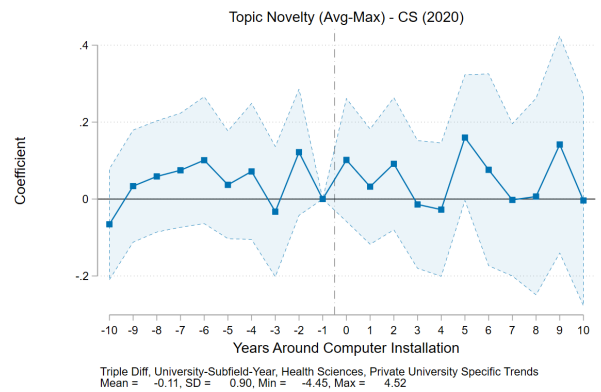


Figure J.10: DDD event studies for log average citations per paper (panels a–b), average OpenAlex topics per paper (panels c–d), and topic-combination novelty (panels e–f), Life and Health Sciences. For each outcome, Y_{ut}^* is the within-university gap between compute-amenable and less-amenable subfields; Callaway and Sant’Anna (2021) estimator; $N = 184$ treated universities. Points are aggregated cohort ATTs with 95% CIs; the dashed vertical line marks first installation. These are the counterparts of the main-text figures for average citations, topics per paper, and topic-combination novelty. Estimates include separate linear time trends for private and public universities.

K Research Methodology Classification Details

The classification proceeds in two stages. First, we use Gemini 3.0 Flash to label 177,243 papers with usable local full text into seven classes (Table K.11), treating these assignments as pseudo-ground truth. Second, we train a LightGBM classifier to reproduce those labels from metadata, keyword n -grams, and dense embeddings, yielding a model we can apply to papers for which local full text is unavailable. The two-stage design is necessary because local full text covers only a modest share of the universe we can reach with metadata and embeddings, and because running Gemini across that larger universe would in any case be prohibitively expensive. We evaluate the LightGBM model on the held-out validation split and on a separate 10,000-paper full-text benchmark; Table K.10 reports these benchmarks and the resulting analysis universe.

Table K.10: Methodology Classifier Benchmarks and Sample Accounting

Benchmark	Accuracy	Weighted F1	Macro F1	Universe	N
Main LightGBM (validation)	0.752	0.749	0.657	Unique papers in 1951–1969 panel slice	470,395
10k full-text benchmark	0.824	0.825	0.760	Classified unique papers	268,887
Fallback no-lexicon model	0.814	0.814	0.751	Keyword-matched computer papers in panel slice	25,665
				Keyword-matched computer papers among classified	17,667
				Classified papers with local full text	100,378
				Already in Gemini first-pass set	56,650

Notes: The main production model is LightGBM. The benchmark table keeps only the validation score we use in the paper, the 10k full-text benchmark, and the no-lexicon fallback. Universe counts refer to the 1951–1969 panel slice used for the methodology diagnostics.

The raw classifier predicts seven classes, which we collapse into the five buckets used in Section 6.4 of the main text. Computational or simulation papers are rare in this period. The *other* category is mostly reviews, bibliographies, and similar reference material rather than a generic model-failure residual.

Table K.11: Seven Class Base Taxonomy and Mapping into Main Paper Buckets

Raw class	Mass share	Maps to
Quantitative empirical	31.1%	Empirical
Qualitative empirical	17.2%	Empirical
Formal theoretical	12.9%	Theory
Discursive theoretical	5.4%	Theory
Methods	11.1%	Methods
Computational / Simulation	1.1%	Simulation
Other	21.3%	Other

Notes: Shares are average predicted mass across the 268,887 classified papers in the 1951–1969 panel slice. The paper’s main methodology shares use the collapsed five-bucket taxonomy and aggregate probability mass rather than argmax assignments.

K.1 Title-Based Validation: Representative Terms

To validate that the methodology shifts correspond to plausible computer-related content, the paper compares title language before and after adoption within each methodology category, contrasting document-level TF-IDF for treated-post versus treated-pre papers in the 1951–1969 classified slice. The exercise spans the full universe of title vocabulary, whereas the lexicon family exercise just looks at changes within the keyword-family terms we select. Table K.12 condenses those rankings into representative post-treatment title terms for each category.

Table K.12: Representative Post-Treatment Title Terms by Methodology Category

Category	Representative post > pre title terms
Empirical	spectral analysis, measurements, detection, prediction, correlation
Methods	computer, algorithm, programming, digital, automated
Theory	model, systems, control, optimal, plasma
Simulation	simulation, computer simulation, electronic structure, dynamics, wavefunctions

Notes: Representative terms drawn from the post > pre title contrasts in the 1951–1969 classified slice (268,887 papers; 10-year pre/post window around first installation). Within each methodology category, title tokens are ranked by the difference between the post-treatment and pre-treatment probability-weighted mean TF-IDF, using the paper’s methodology category probability as the weight.

K.2 Full-Text Validation and Manual Examples

Here, we build keyword families from manual reading of full-text packets and then apply them to the local full-text base. We have usable local snippets for 111,842 paper-university rows in the 1951–1969 classified slice. These families are *not* intended to partition the full vocabulary. Instead, they are used to interpret the kinds of content that rise after adoption.

Table K.13 reports the main family definitions. Figure K.11 applies these dictionaries to the classified panel slice, reporting pre/post differences in probability-weighted family hit rates within each methodology category: methods move toward computer-systems and algorithmic or numerical content, theory toward control, optimization, and matrix methods, and simulation toward direct simulation language. Table K.14 gives representative manual examples. The empirical examples reinforce a key point from Section 6.4 of the main text: empirical computer use in this period mostly reflects measurement, monitoring, signal analysis, diagnosis, and numerical processing of observed systems, rather than the later regression-heavy style familiar from modern social science.

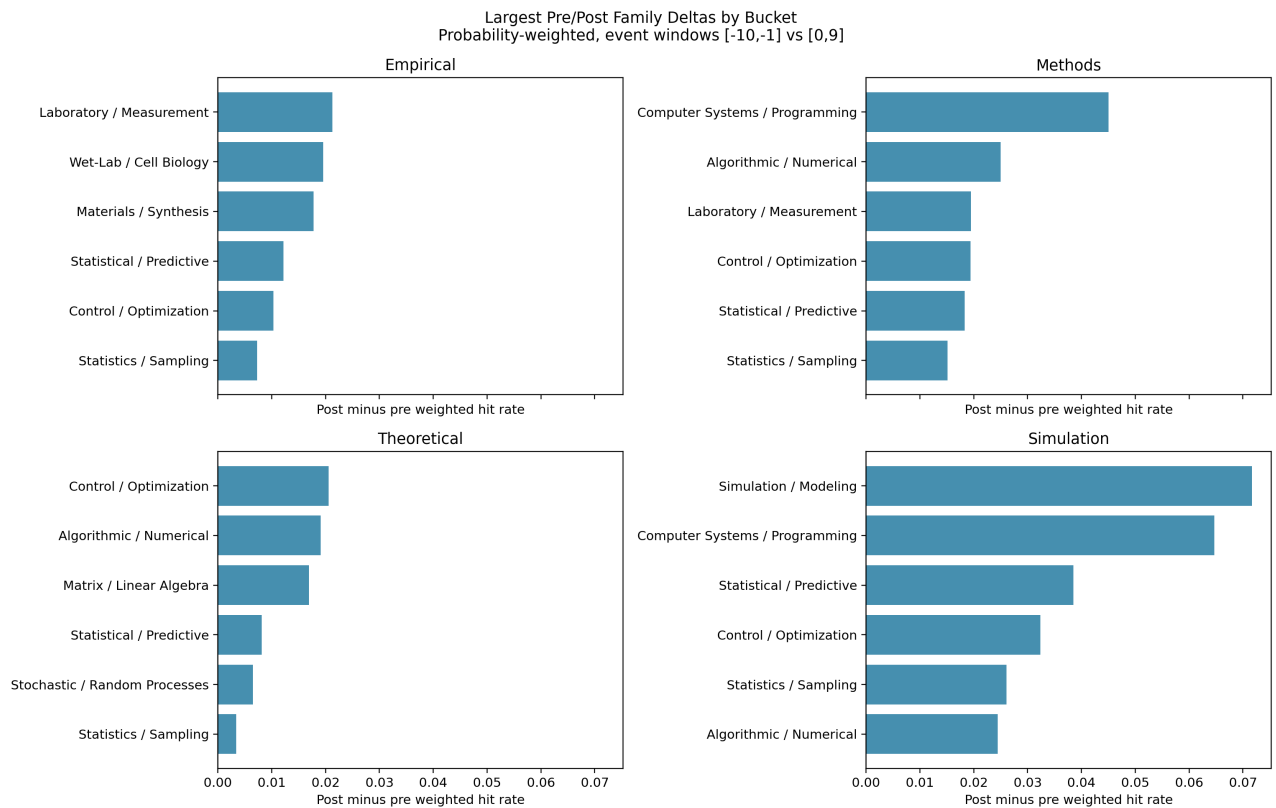


Figure K.11: Pre/post differences in probability-weighted keyword-family hit rates within each methodology category (1951–1969 classified slice, 268,887 papers, 10-year window around first installation). Rows are methodology categories (empirical, theory, methods, simulation); columns are keyword families (e.g., computer-systems, algorithmic/numerical, control/optimization; full definitions in Table K.13). Results shown are simple comparisons, not difference-in-differences estimates.

Table K.13: Family Dictionaries Used in the Validation Exercise

Family	Definition and example language
Computer systems / programming	Explicit computer, compiler, programming-language, assembler, time-sharing, or computer-center language. Example terms: computer, compiler, ALGOL, program, time-sharing.
Algorithmic / numerical	Algorithms, numerical procedures, matrix computation, finite-difference, and finite-element schemes. Example terms: algorithm, numerical, matrix, finite element.
Simulation / modeling	Simulation, simulation models, numerical forecasting, or model-based experimentation. Example terms: simulation, model, modeling, dynamics.
Control / optimization	Feedback, control, queueing, encoding, and related systems or optimization problems. Example terms: control, optimal, feedback, queueing, encoding.
Laboratory / measurement	Experimental apparatus, assay, protocol, or measurement language that often accompanies computer-assisted recording and monitoring. Example terms: apparatus, assay, procedure, perfusion, measurement.
Statistical / predictive	Quantitative processing of observed data. Example terms: regression, statistical, variance, correlation, prediction.
Biomedical signal / detection	Empirical signal-processing and diagnostic language. Example terms: electrocardiogram, heart sounds, body-surface potential, detection, spectral analysis.
Statistics / sampling	Statistical inference, regression, variance analysis, or survey/sampling design. Example terms: statistical, regression, sampling, inference.
Differential equations	PDE/ODE and boundary-value or continuum-model problems. Example terms: differential equation, boundary value, PDE, ODE.
Materials / synthesis	Chemistry and materials language used as a placebo-style comparison family rather than a claim about the main computer mechanism. Example terms: chemical synthesis, enzymatic synthesis, polymer, alloy.

Notes: Families are transparent dictionaries used for descriptive interpretation only. They were built from manual reading of full-text papers in the methods, theory, simulation, and empirical-computer subsets, then applied to the refreshed local full-text base.

Table K.14: Illustrative Manual Examples behind the Methodology Labels

Empirical computer use

- *Study of High Frequency Components in Electrocardiogram by Power Spectrum Analysis* ([W2003343840](#)): measured biomedical signals processed with spectral analysis.
- *The Detection of Heart Disease in Children* ([W2974997846](#)): tape-recorded heart sounds used for computer-assisted screening and classification.
- *The Prediction of Flow Patterns, Liquid Holdup and Pressure Losses Occurring During Continuous Two-Phase Flow In Horizontal Pipelines* ([W1975029220](#)): numerical prediction built on observed engineering measurements.

Methods

- *The structure of yet another ALGOL compiler* ([W2050572856](#)): compiler architecture and implementation.
- *Programming Technique: An improved hash code for scatter storage* ([W2003248512](#)): programming-method paper on data storage and retrieval.
- *The simplex method of linear programming using LU decomposition* ([W2071877138](#)): numerical optimization method built for computation.

Computer-adjacent theory

- *Feedback Queueing Models for Time-Shared Systems* ([W2074065133](#)): queueing/control theory aimed at time-sharing computer systems.
- *Levels of computer systems* ([W1983161084](#)): formal systems-theory discussion explicitly about computer architecture.
- *Source encoding in the presence of random disturbance* ([W1584278176](#)): information/encoding theory that sits on the computation-communications interface.

Simulation

- *The simulation of time sharing systems* ([W2113442017](#)): explicit simulation of computer-system performance.
- *Experimental and Numerical Simulation of Two-Phase Flow with Interphase Mass Transfer in One and Two Dimensions* ([W2149866374](#)): explicit numerical simulation paper centered on finite-difference solution of a physical system.

Notes: These examples come from manual full-text packet reads and are used only to anchor interpretation. The empirical examples show how computers entered measured and observational work through diagnosis, spectral analysis, monitoring, and numerical processing.

References

Callaway, B. and Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of econometrics*, 225(2):200–230.