Expertise and the productivity effects of AI

AI assistance, the value of expertise & learning on the job

Sergio Ocampo
ECON 3399 · AI and Robots at Work

Three papers on expertise, productivity and learning

Paper	Question
Autor & Thompson Expertise	Which tasks make expertise valuable?
Brynjolfsson, Li & Raymond Generative AI at Work	Who becomes more productive with an assistant?
Autor et al. Patent-lawyer experiment	Does AI-assisted work develop independent judgment?

> Expertise determines who can do the work.
> AI can change both access to that expertise and how it is acquired.

Expertise



EXPERTISE

David Autor
Massachusetts Institute of Technology and NBER, United States

Neil Thompson
Massachusetts Institute of Technology FutureTech, CSAIL and Initiative on the Digital Economy, United States

Abstract
When job tasks are automated, does this augment or diminish the value of labor in the tasks that remain? We argue the answer depends on whether removing tasks raises or reduces the expertise required for remaining non-automated tasks. Since the same task may be relatively expert in one occupation and inexpert in another, automation can simultaneously replace experts in some occupations while augmenting expertise in others. We propose a conceptual model of occupational task bundling that predicts that changing occupational expertise requirements have countervailing wage and employment effects: automation that decreases expertise requirements reduces wages but permits the entry of less expert workers; automation that increases expertise requirements increases wages but reduces the set of qualified workers. We develop a novel, content-agnostic method for measuring job-task expertise, and we use it to quantify changes in occupational expertise demands over four decades attributable to job task removal and addition. We document that automation has raised wages and decreased employment in occupations where it eliminated inexpert tasks, but reduced wages and increased employment in occupations where it eliminated expert tasks. These effects are distinct from—and in the case of employment, opposite to—the effects of changing task quantities. The expertise framework resolves the puzzle of why routine task automation has decreased employment but often increased wages in routine task-intensive occupations. It provides a general tool for analyzing how task automation and new task creation reshape the scarcity value of human expertise within and across occupations. (JEL: E24, J11, J23, J24)

The editor in charge of this paper was Matthias Doepke.

Acknowledgments: Joseph R. Schumpeter lecture, presented to the European Economic Association on August 29, 2024 in Rotterdam, with lecture title "Does Automation Replace Experts or Complement Expertise? The Answer is Yes." We thank Daron Acemoglu, Erik Brynjolfsson, Martin Fleming, Joshua Gans, Anton Korinek, Frank Levy, Sendhil Mullainathan, Ashesh Rambachan, Pascual Restrepo, Anna Salomons, Lawrence Schmidt and numerous seminar participants for valuable comments that improved the paper. We are deeply indebted to Lucy Hampton, Joanne Liang, Ragini Srinivasan, Christian Vogt, and Can Yeşildere for exceptional research assistance, and to Fredric Kong for expert review of the theoretical

Journal of the European Economic Association 2025 23(4):1203–1271
https://doi.org/10.1093/jeea/jvaf023
© The Author(s) 2025. Published by Oxford University Press on behalf of European Economic Association. All rights reserved. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site-for further information please contact journals.permissions@oup.com

Downloaded from https://academic.oup.com/jeea/article/23/4/1203/8175003 by guest on 15 August 2025

David Autor and Neil Thompson

Journal of the European Economic Association · 2025

Open the paper

Published article · First page

Automation changes the expertise needed in an occupation or job

- Occupations are bundles of tasks
- Tasks require *expertise* to be performed
- Automation takes away tasks and changes expertise

Accounting and inventory clerks after automation

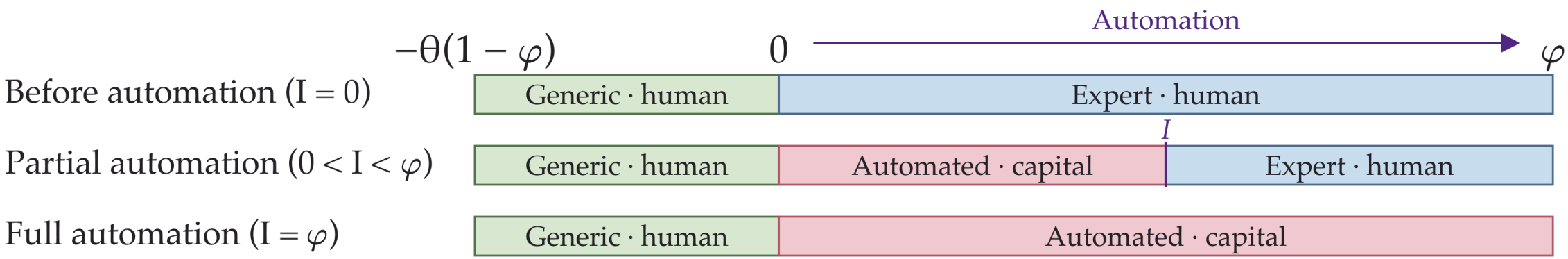
Occupation	**Routine work automated**	**Work left with people**
Accounting clerks	Record transactions; reconcile records	Resolve discrepancies; exercise accounting judgment
Inventory clerks	Calculate prices; manage inventory records	Count, stock & weigh goods

> Similar exposure can leave very different expertise requirements in the remaining job.

Expertise determines which tasks a worker can perform

Object	**Meaning in the model**
Worker expertise e	A capability level between 0 and 1; worker e can do task t if e $\geq$ t.
Task requirement t	Expert tasks have t $>$ 0; generic tasks have t $\leq$ 0.
Occupation φ	The highest task requirement in its bundle.
Automation frontier I	Capital can perform expert tasks with $0 < t \leq I$.
Generic work $\theta > 1$	Controls the amount of generic (**human!**) work.

> More expert occupations contain fewer generic tasks, but always some.
> Automation changes who performs the tasks; it does not redraw these bundles.



Each vertical slice is one occupation's task bundle

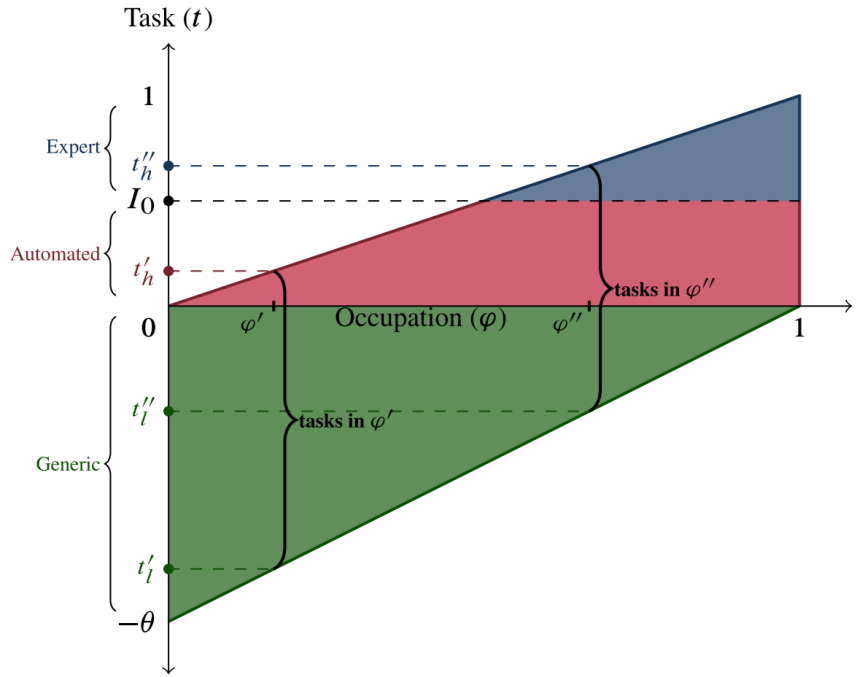


Figure 1. I_0: automation frontier; t_l/t_h: lower/upper endpoints of a bundle.
Green: generic tasks; red: automated tasks; blue: still-expert tasks.

> Left occupation: capital performs every expert task; people do generic work.
> Right occupation: its worker still performs both generic and expert tasks.

Follow one occupation as automation advances

Illustrative occupation: its most demanding task requires $\varphi = \mathbf{0.6}$.

Frontier	**Human work left**	**Who can enter?**
I = 0.3	Expert tasks from 0.3 to 0.6 + generic tasks	Workers with e $\geq$ 0.6
I = 0.6	Generic tasks only	Workers at any expertise level

Productivity: saving expert time can raise the expert wage.
Entry: opening the job to everyone can eliminate the expert premium.

The next two figures solve a sequence of equilibria

Move **I** to the right; hold the capital stock and worker skills fixed.

Object in the figures	**How to read it**
$w(\varphi)$: blue curves	Wage in an occupation still requiring expertise.
w_g: green; r: red	Generic wage; rental payment per unit of capital.
$L(\varphi)$: employment	Labour allocated to an occupation type; expert jobs start at 1.

> Worker expertise e is distributed uniformly between 0 and 1.
> Workers choose the highest-paying occupation they can perform.

The model predicts gains and losses across expertise levels

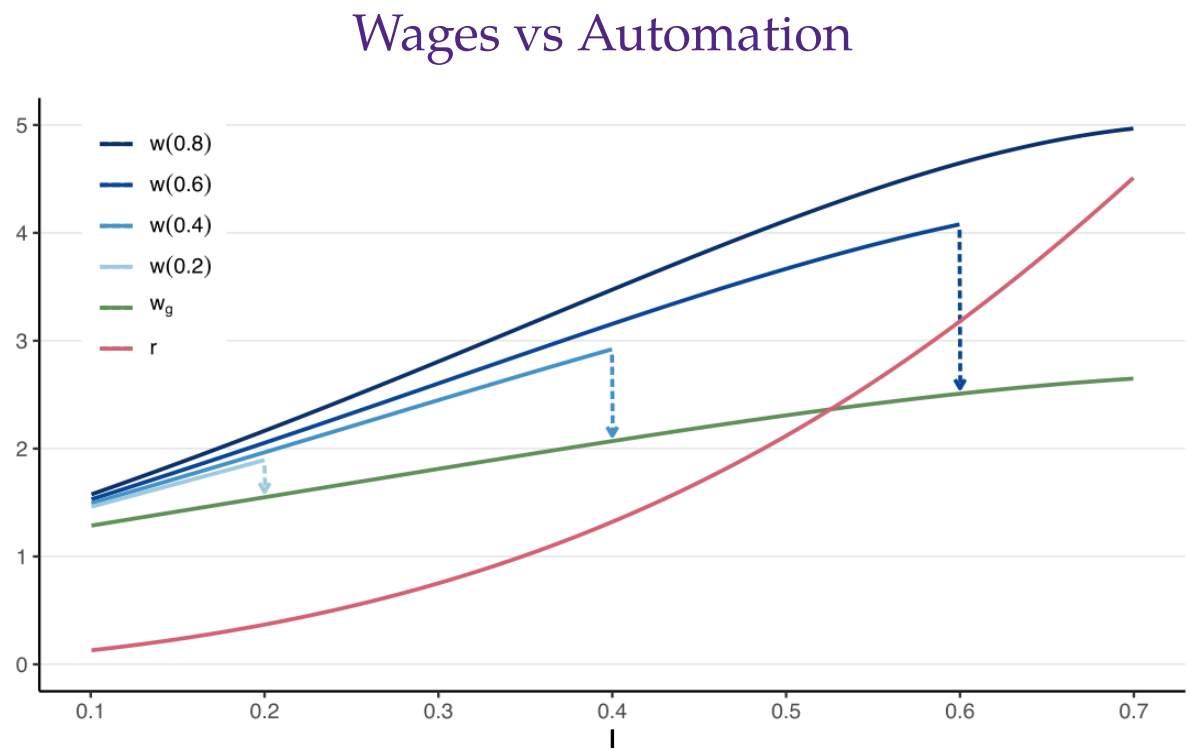


Figure 2. Horizontal: automation frontier I. Vertical: wages and capital rental in model units. Blue: w(φ); green: w_g; red: r.

Expertise benefits from labor saving automation, as long as $I < \varphi$.
Full automation removes expertise premium $\longrightarrow$ Sudden collapse in wages

Changing expertise requirements reallocates employment

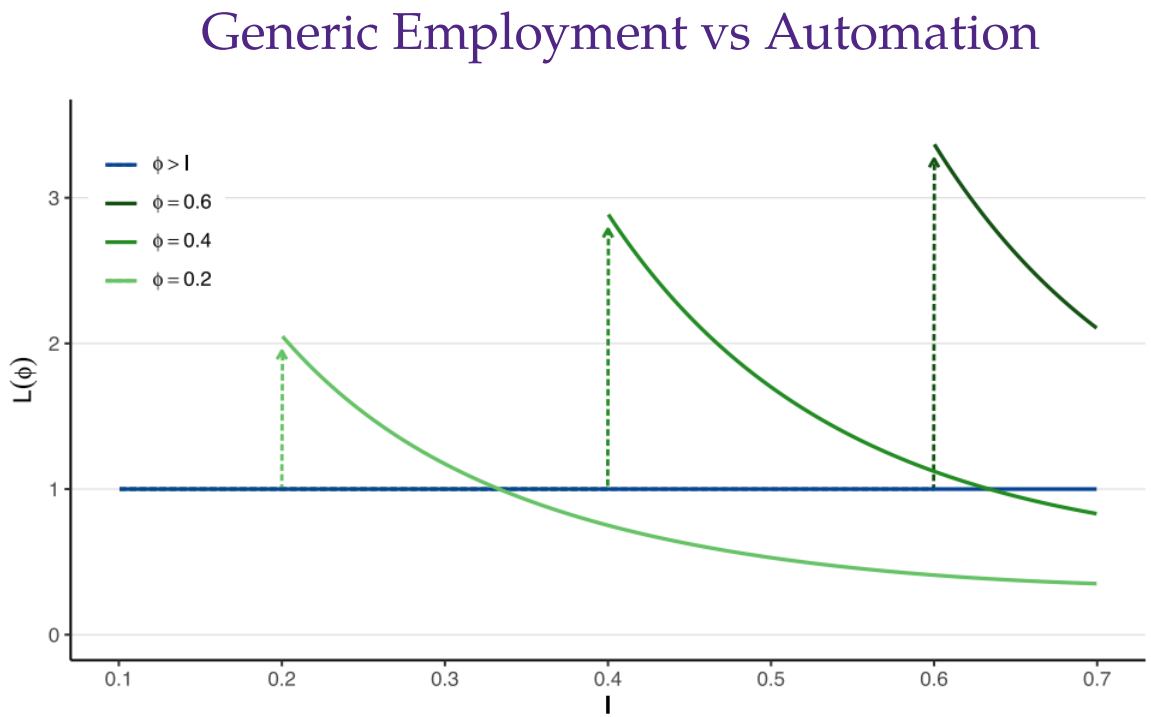


Figure 3. Horizontal: I. Vertical: employment density $L(\varphi)$, initially 1 for expert jobs. Blue: still-expert jobs; green: jobs whose expert tasks have been automated.

> At $I = \varphi$, entry expands even as the occupational wage drops.
> Later automation draws workers toward other newly accessible jobs.

How do the authors measure expertise in the data?

- 1977 Dictionary of Occupational Titles + 2018 O*NET task descriptions
- Score words from school and college texts: expertise = 100 - SFI
  - The **Standard Frequency Index (SFI)** combines frequency and dispersion across subject areas.
- Adjust with the Dale–Chall list of 3,000 words familiar to fourth graders:
  cap the highest 10% of expertise scores among familiar words;
  raise the lowest 10% among unfamiliar words.
- Average adjusted word scores into tasks; task scores into occupations
  (Comparing occupations over time)

Word examples: scores before adjustment

Word	SFI	100 - SFI
color	61.0	39.0
luminance	30.5	69.5

First identify what work stayed, disappeared, or appeared

Match tasks by **meaning within an occupation**, not identical wording.

File-clerk example	Classification
Sort incoming material ⟶ scan/read material to classify it	Retained: similar activity, new wording.
Operate a keypunch for tabulating cards	Removed: no close 2018 counterpart.
Input document information into computer systems	Added: no close 1977 counterpart.

Task removal: compare the remaining tasks with the old bundle

Objective: Did task removal increase or decrease expertise in the occupation?

$$\Delta XPT^{\text{removal}} = \overline{XPT}_{1977,\text{retained}} - \overline{XPT}_{1977,\text{all}}$$

Illustrative 1977 task scores	All-task mean	Retained mean	Change
0, 1, 2; remove task 0	1	1.5	+0.5
0, 1, 2; remove task 2	1	0.5	-0.5

> Removing a low-expertise task raises average expertise.
> Removing a high-expertise task lowers it—even with the same task-count loss.

Task addition: compare the new bundle with its retained tasks

Objective: Did task addition increase or decrease expertise in the occupation?

$$\Delta XPT^{\text{addition}} = \overline{XPT}_{2018,\text{all}} - \overline{XPT}_{2018,\text{retained}}$$

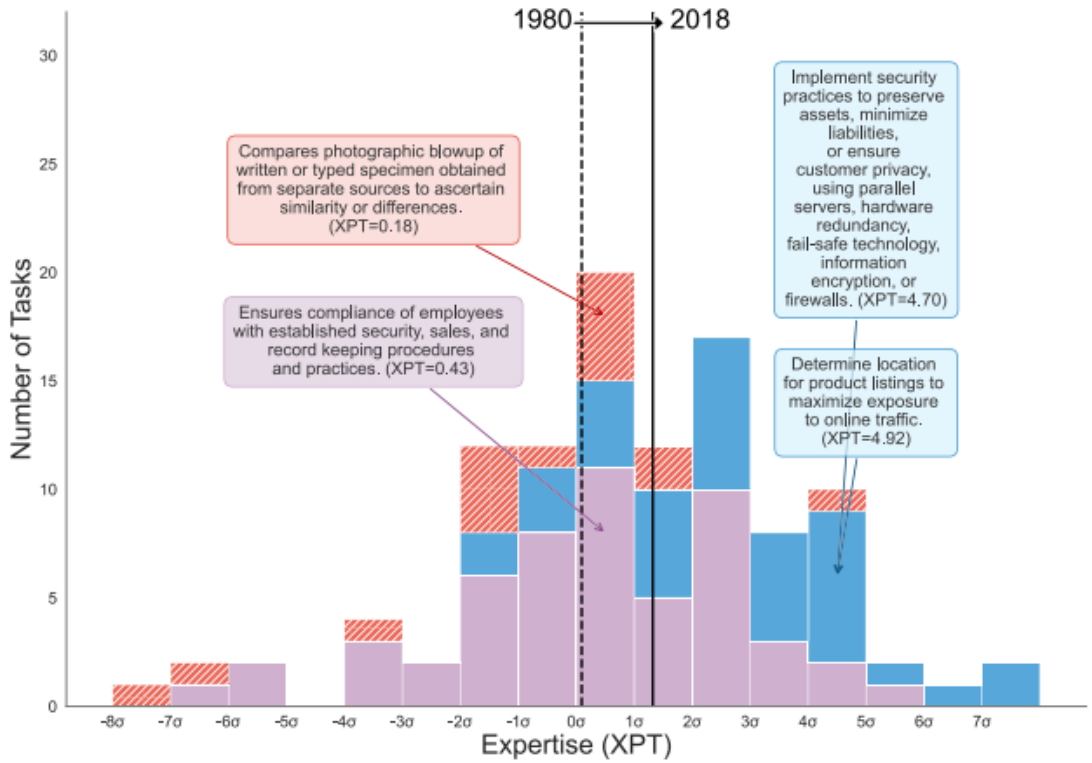
Illustrative retained scores: 1, 2	**New mean**	**Retained mean**	**Change**
Add task 3	2	1.5	+0.5
Add task 0	1	1.5	-0.5

Net Expertise Change:

$$\Delta XPT = \Delta XPT^{\text{removal}} + \Delta XPT^{\text{addition}}$$

Each comparison uses one date's wording, excluding rewritten retained tasks.

In the data: management support gains expertise



Task lists: 1977 & 2018

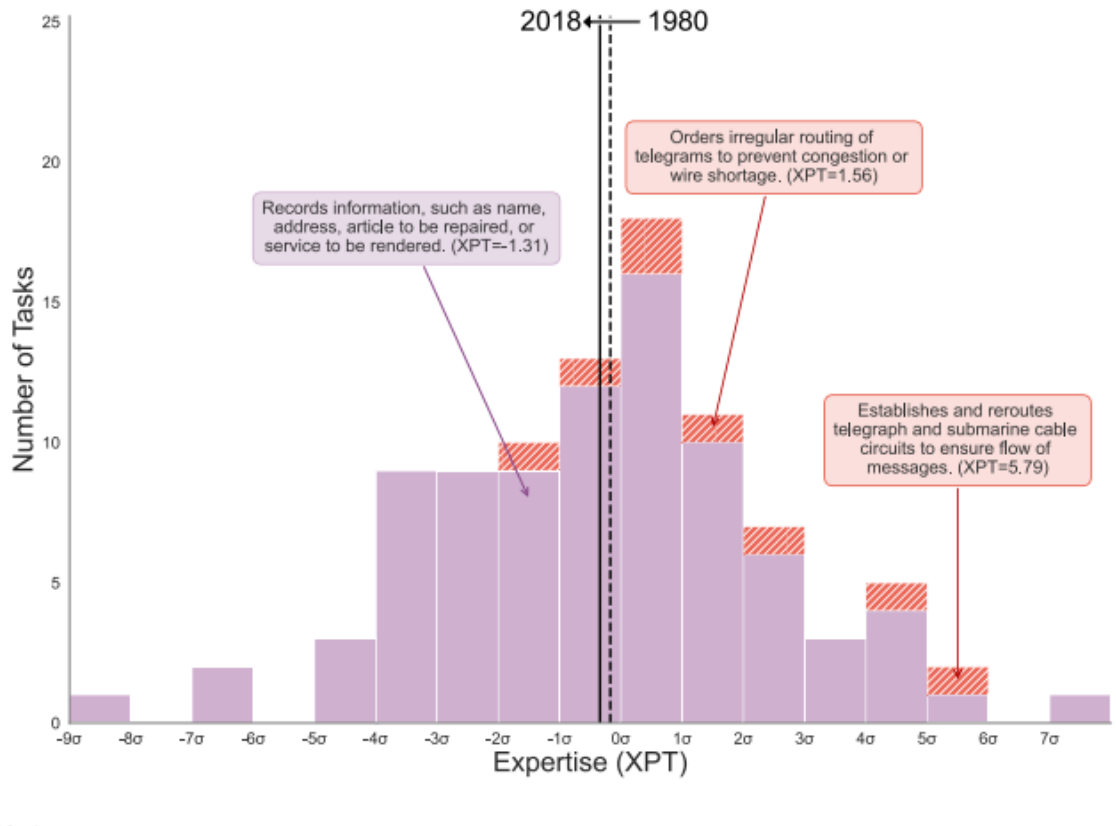
- **Retained:** in both years
- **Removed:** old tasks only
- **Added:** new tasks only

Horizontal: standardized expertise
Vertical: number of tasks

Occupation means:
dashed 1980; solid 2018

Removed tasks lie mainly below the old occupational mean.
Added tasks extend the expert end of the bundle; both changes raise expertise.

Dispatchers: task removal lowers expertise



Task lists: 1977 & 2018

- **Retained:** in both years
- **Removed:** old tasks only
- **Added:** new tasks (none shown)

Horizontal: standardized expertise
Vertical: number of tasks

Occupation means:
dashed 1980; solid 2018

The sign depends on which tasks disappear, not simply how many.
For dispatchers, removal of relatively expert tasks lowers measured expertise.

Link task changes to wages and employment

$$\Delta \log y_o = a + \beta \Delta XPT_o + \gamma \Delta Tasks_o + \delta_{g(o)} + u_o$$

Term	What is measured
y_o	Occupational wage or employment
ΔXPT_o	Net expertise change, in 1980 standard deviations
$\Delta Tasks_o$	Added share - removed share

o: occupation; β, γ: coefficients; δ: broad occupation-group effects; a: intercept; u: residual. All changes are per decade.

> Within broad occupation groups, separate expertise changes from changes in task quantity.

Expertise change and wage change

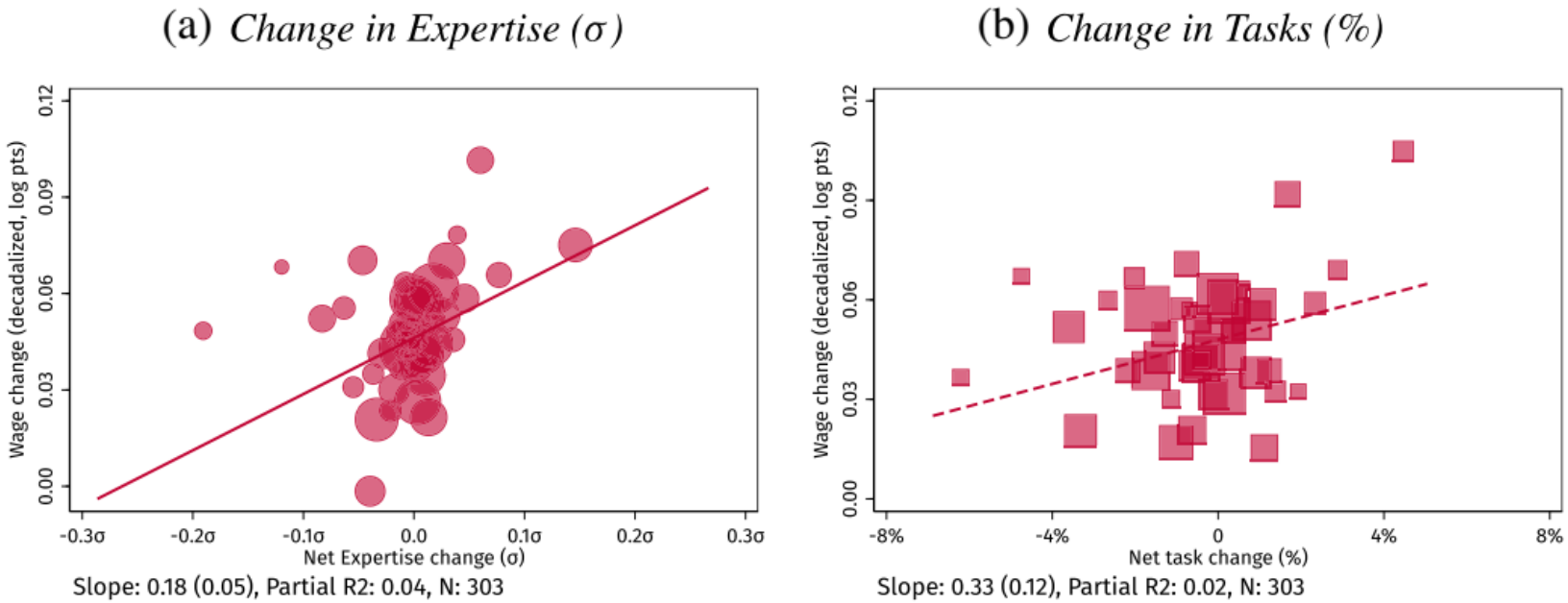


Figure 10: decadalized occupational wage change versus expertise change (left) and task-quantity change (right).

> Wages rise with expertise + task quantity, holding each other and occupation group fixed.

Expertise and task quantity have different employment implications

Explanatory variable	Δ log wage	Δ log employment
Net expertise change	0.175 (0.047)	-0.510 (0.199)
Net task change	0.334 (0.123)	1.430 (0.523)
Occupations	303	303

Tables 4A/5A, col. 5. Per-decade changes, 1980–2018; broad occupation effects; 1980 employment weights. SEs in parentheses.

> More expertise predicts higher wages but lower employment.
> More tasks predict higher wages and higher employment.

How large are the occupation-level associations?

Conditional change	Predicted wage change	Predicted employment change
+0.1 SD in expertise	+1.75 log points	-5.10 log points
+10% in the task set	+3.34 log points	+14.30 log points

Multiply each published coefficient by 0.1. Changes are per decade; the other regressor is held fixed.

> Expertise and task quantity both predict wages.
> Their employment associations point in opposite directions.

Removing tasks does not have a single wage implication

Change in the occupation	Coefficient in wage regression
Expertise change due to task removal	0.147 (0.051)
Expertise change due to task addition	0.255 (0.088)
Share of tasks removed	-0.230 (0.141)
Share of tasks added	0.549 (0.206)

Table 4B, col. 5; 303 occupations. Conditional associations; standard errors in parentheses.

> Losing expert tasks and losing supporting tasks have different implications.
> Task counts alone cannot describe what remains valuable.

Generative AI at Work



GENERATIVE AI AT WORK*

ERIK BRYNJOLFSSON
DANIELLE LI
LINDSEY RAYMOND

We study the staggered introduction of a generative AI–based conversational assistant using data from 5,172 customer-support agents. Access to AI assistance increases worker productivity, as measured by issues resolved per hour, by 15% on average, with substantial heterogeneity across workers. The effects vary significantly across different agents. Less experienced and lower-skilled workers improve both the speed and quality of their output, while the most experienced and highest-skilled workers see small gains in speed and small declines in quality. We also find evidence that AI assistance facilitates worker learning and improves English fluency, particularly among international agents. While AI systems improve with more training data, we find that the gains from AI adoption are largest for moderately rare problems, where human agents have less baseline experience but the system still has adequate training data. Finally, we provide evidence that AI assistance improves the experience of work along several dimensions: customers are more polite and less likely to ask to speak to a manager. *JEL codes:* D80, J24, M15, M51, O33.

I. INTRODUCTION

The emergence of generative artificial intelligence (AI) has attracted significant attention for its potential economic impact.

* We gratefully acknowledge the editors, Lawrence Katz and Andrei Shleifer, and anonymous referees for many valuable insights and suggestions. We are grateful to Daron Acemoglu, David Autor, Amittai Axelrod, Eleanor Dillon, Zayd Enam, Luis Garicano, Alex Frankel, Sam Manning, Sendhil Mullainathan, Emma Pierson, Scott Stern, Ashesh Rambachan, John Van Reenen, Raffaella Sadun, Kathryn Shaw, Christopher Stanton, Sebastian Thrun, and various seminar participants for helpful comments and suggestions. We thank Max Feng and Esther Plotnick for providing excellent research assistance and the Stanford Digital Economy Lab for financial support. The content is solely the responsibility of the authors and does not necessarily represent the official views of Stanford University, MIT, or the NBER.

© The Author(s) 2025. Published by Oxford University Press on behalf of President and Fellows of Harvard College. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
The Quarterly Journal of Economics (2025), 889–942. https://doi.org/10.1093/qje/qjae044.
Advance Access publication on February 4, 2025.

Published article · First page

Erik Brynjolfsson, Danielle Li, and Lindsey Raymond

The Quarterly Journal of Economics · 2025

Open the paper

Can AI spread the practices of successful support agents?

Setting: a customer-support call center using live chats.
Design: nonrandom staggered rollout, following a small randomized pilot.

- 5,172 customer-support agents; about three million chats
- An assistant suggests responses + links to technical documentation
- Access begins at different dates across workers

> Compare performance before and after access with workers who do not yet have access.

What the researchers observe in the support center

Data	What it lets them measure
Chat text, timestamps & outcomes	Measure individual support interactions
Worker, team, location & tenure	Track the same agents over time
AI access dates & system interruptions	Compare access periods and outages

> The productivity measures come from work records.
> They are distinct from reported time savings.

- **Resolutions per hour:** issues resolved / hours worked
- **Speed:** average time per chat
- **Throughput:** chats handled per hour
- **Quality:** issue resolution rate + customer satisfaction

Difference in differences

$$\hat{\beta} = (\overline{y}_{T,after} - \overline{y}_{T,before}) - (\overline{y}_{C,after} - \overline{y}_{C,before})$$

Use the comparison group's change to account for what would have happened to the treated group without the intervention.

> Key assumption: parallel counterfactual changes.
> Initial productivity levels may differ.

Implementing the comparison with repeated observations

$$y_{jt} = a_j + \lambda_t + \beta\, AI_{jt} + g(\text{tenure}_{jt}) + u_{jt}$$

Term	**Role**
Worker effect: a_j	Stable differences across workers
Calendar-time effect: λ_t	Common changes at a given date
Tenure controls	Experience accumulated on the job
AI_{jt}	Whether the worker has access

Productivity rises when AI access begins

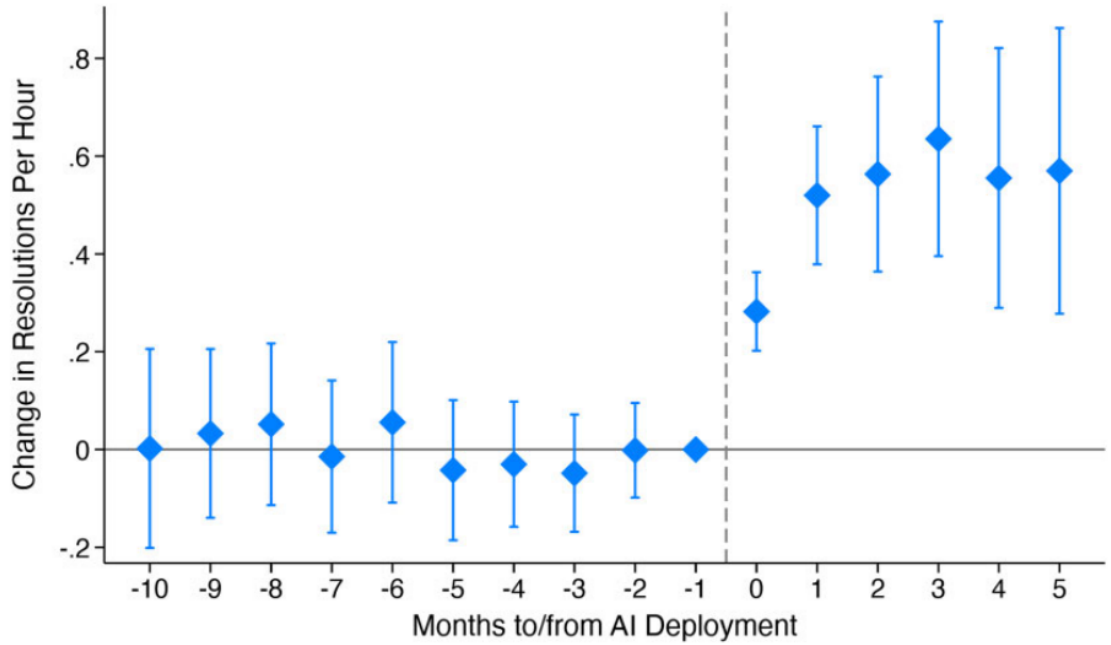


Figure II. Event time is measured relative to AI access; bars are 95% confidence intervals.

> The gain begins at access and persists.
> No comparable jump appears before treatment.

The productivity estimate remains positive with richer controls

Outcome: resolutions per hour	**(1)**	**(2)**	**(3)**
AI access	0.469	0.371	0.301
Standard error	(0.0325)	(0.0318)	(0.0329)
Worker effects	No	Yes	Yes
Worker-tenure effects	No	No	Yes
Agent–month observations	13,192	12,295	12,295

Calendar-month and location controls throughout. Standard errors clustered by agent. Table II.

> The preferred estimate is 0.301 additional resolutions per hour: about 15.2% of the treated workers' pre-access mean of 1.97.

Higher throughput, with no detectable average quality loss

Outcome	Estimated effect	Standard error
Chat duration, minutes	-3.746	0.369
Chats handled per hour	+0.365	0.0345
Resolution probability, percentage points	+1.32	0.882
Net promoter score, points	-0.119	0.524

Preferred controls as before; agent-clustered errors. Resolution coefficient and error multiplied by 100. Table III.

> Agents finish chats faster and handle more of them.
>
> Resolution and satisfaction estimates show no statistically clear average change.

The gains differ with expertise: highest for the lowest-skilled workers

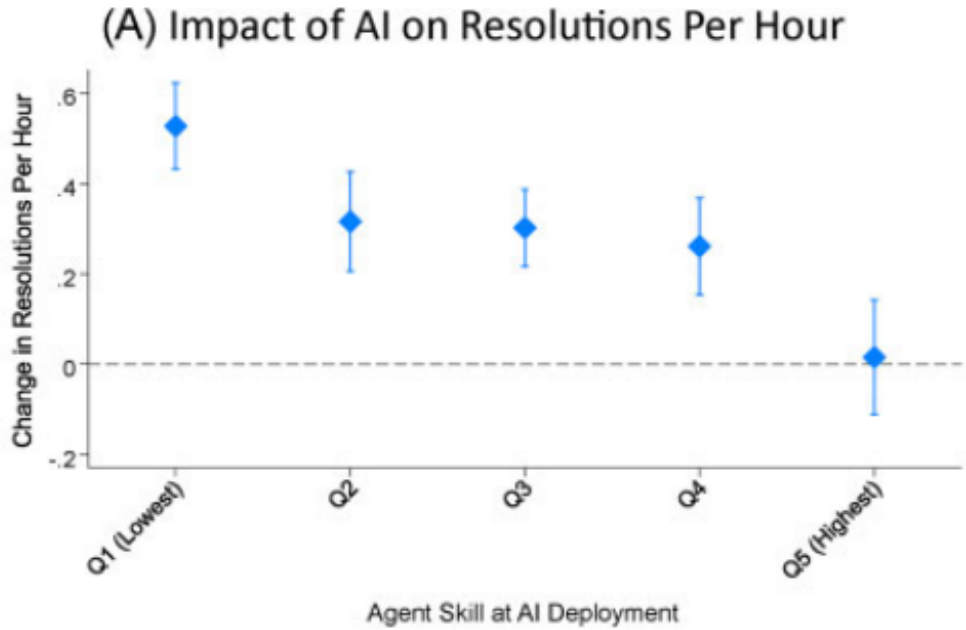


Figure III(a): change in resolutions/hour by prior performance quintile, controlling for tenure.

> The lowest-performance quintile gains about 0.5 resolutions/hour, or 36%.
> The highest quintile has no statistically clear productivity gain.

The best initial performers can lose quality

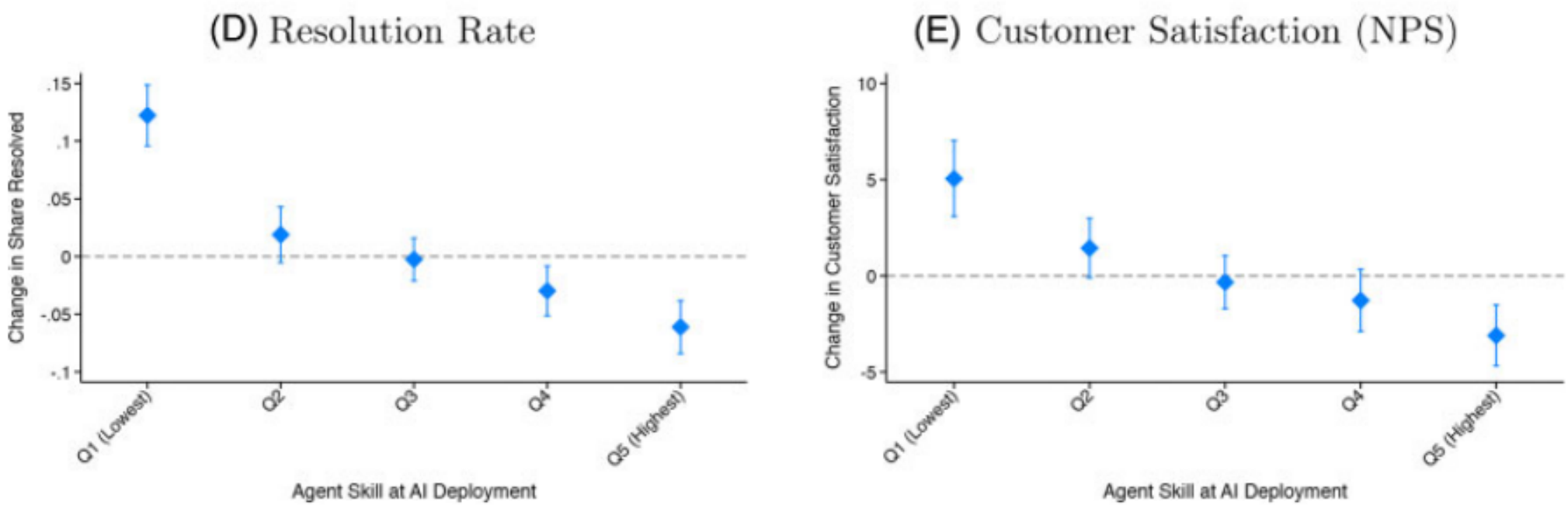


Figure III, panels D–E: resolution rate and customer satisfaction by initial performance quintile.

> Lower-performing agents improve quality; top performers show small declines.
> A common assistant can spread good practice while diluting exceptional practice.

AI assistance shortens the initial learning curve

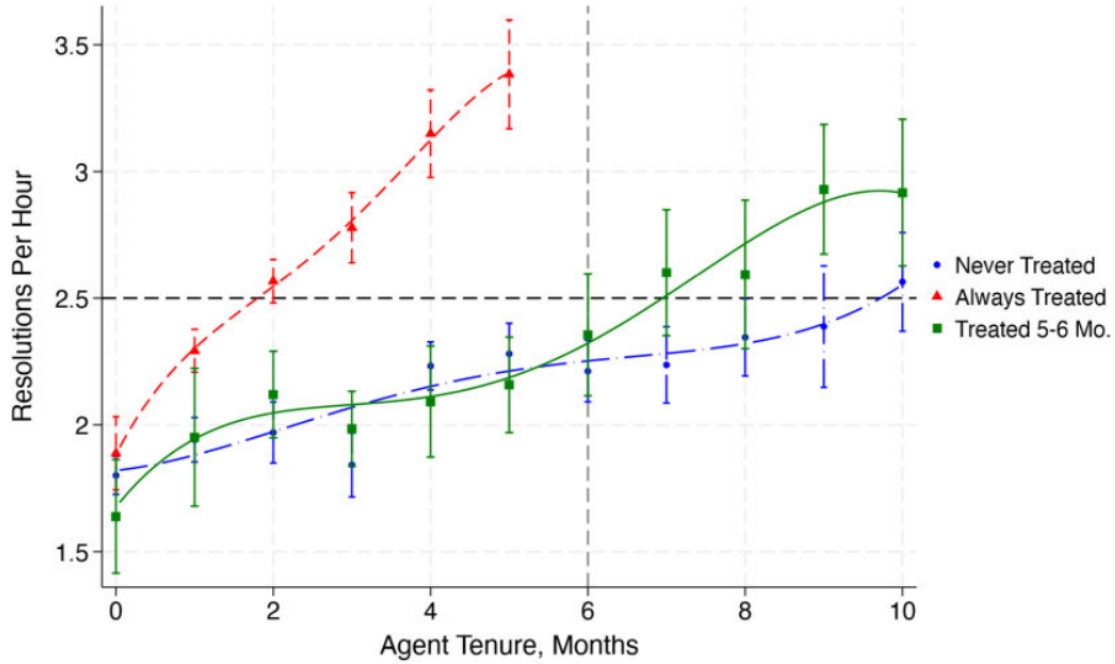


Figure V: resolutions/hour by tenure for never-treated, always-treated, and month-five adopters.

> Workers with access reach about 2.5 resolutions/hour after two months; unaided workers take roughly eight to ten months to reach that level.

Some improvements persist when assistance is unavailable

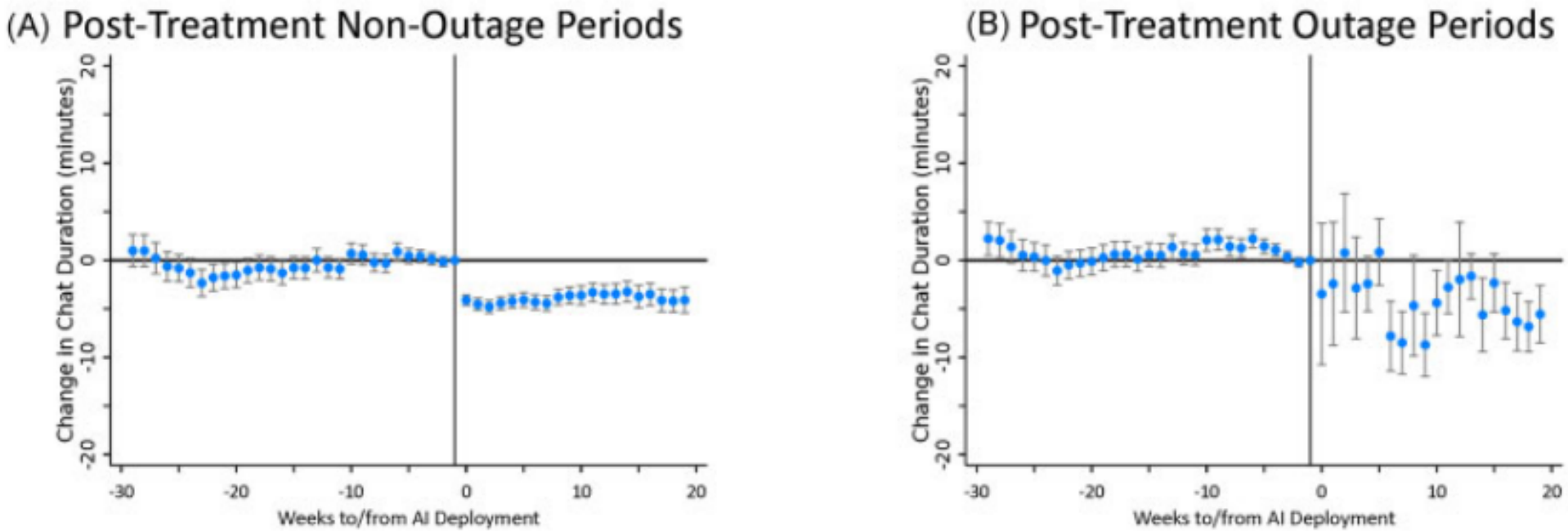


Figure VII, panels A–B: chat duration outside outages (left) and during major outages (right).

> Previously exposed agents still handle chats faster during outages, especially after longer exposure. This is evidence consistent with learning.

The interaction with customers also improves

Outcome	AI access effect	Standard error
Customer sentiment (-1 to +1 scale)	+0.177	0.0116
Agent sentiment (same scale)	+0.0198	0.00599
Requests for a manager, percentage points	-0.875	0.201

Table IV; preferred controls and agent-clustered errors. Manager-request coefficient and error multiplied by 100.

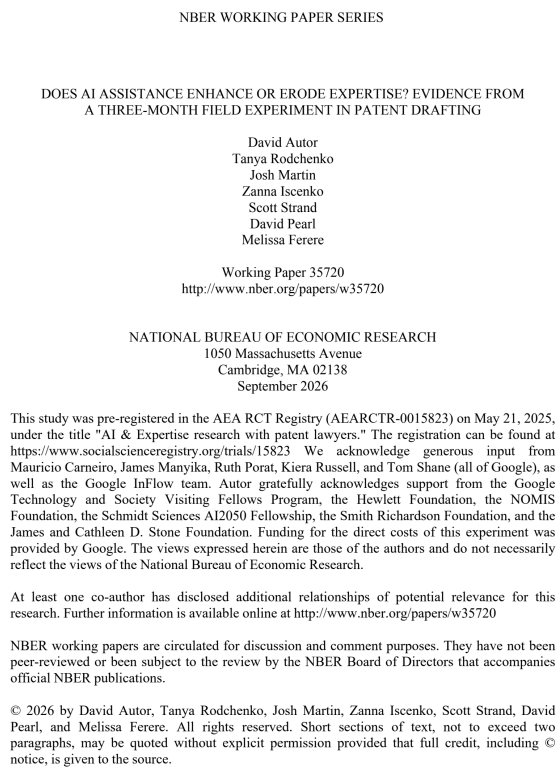
> Customers' messages become more positive and escalations decline.
> The gain includes a less difficult interaction, as well as faster work.

Assistance and learning in customer support

- About 15% more resolved issues per hour
- Larger gains for novices & initially low-performing agents
- Faster learning curve + persistent gains during outages
- Quality declines for some top performers despite the average gain

> Some useful practice carries over without the assistant.
> Does that also happen in work requiring years of professional training?

Does AI assistance enhance or erode expertise?



NBER WORKING PAPER SERIES

DOES AI ASSISTANCE ENHANCE OR ERODE EXPERTISE? EVIDENCE FROM A THREE-MONTH FIELD EXPERIMENT IN PATENT DRAFTING

David Autor
Tanya Rodchenko
Josh Martin
Zanna Iscenko
Scott Strand
David Pearl
Melissa Ferere

Working Paper 35720
http://www.nber.org/papers/w35720

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

This study was pre-registered in the AEA RCT Registry (AEARCTR-0015823) on May 21, 2025, under the title "AI & Expertise research with patent lawyers." The registration can be found at https://www.socialscienceregistry.org/trials/15823 We acknowledge generous input from Mauricio Carneiro, James Manyika, Ruth Porat, Kiera Russell, and Tom Shane (all of Google), as well as the Google InFlow team. Autor gratefully acknowledges support from the Google Technology and Society Visiting Fellows Program, the Hewlett Foundation, the NOMIS Foundation, the Schmidt Sciences AI2050 Fellowship, the Smith Richardson Foundation, and the James and Cathleen D. Stone Foundation. Funding for the direct costs of this experiment was provided by Google. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w35720

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

© 2026 by David Autor, Tanya Rodchenko, Josh Martin, Zanna Iscenko, Scott Strand, David Pearl, and Melissa Ferere. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

David Autor, Tanya Rodchenko, Josh Martin,
Zanna Iscenko, Scott Strand,
David Pearl & Melissa Ferere

A three-month experiment with practicing patent lawyers

Open the paper · Google Research summary

NBER Working Paper 35720 · September 2026

Better drafting and better judgment are separate outcomes

With an assistant	Without an assistant
Draft a patent from invention materials	Find and correct flaws in an existing draft
Produce a stronger document today	Use professional judgment after three months

> An assistant can improve our work without developing our independent judgment.

Random assignment within participating law firms

- 133 practicing patent lawyers at 11 U.S. law firms
- 90 assigned AI access; 43 assigned to the control group
- Randomize within firm, stratifying by legal experience
- Both groups offered AI basics; treated lawyers also receive tool training

> Treatment is access to the assistant + its training.
> The comparison estimates the effect of assignment to that package.

What the AI assistant does

- Draft or rewrite selected text using a prompt
- Insert text using the surrounding document as context
- Give feedback as an editor, expert or intended audience
- Identify inconsistencies against a set of review criteria

> The lawyer supplies the invention + drafting goals.
> The lawyer also decides what to accept, change or reject.

Three assessments: drafting and independent review

When	Task	AI during the test?	Completed
Day 10	Draft dependent claims + description	Treated only	105 lawyers
Day 90	Draft a different, harder patent	Treated only	98 lawyers
Day 90	Redline a flawed patent draft	Neither group	91 lawyers

Blinded experts evaluate five dimensions of quality

Dimension	What a stronger patent does
Enforceability	Avoid wording that weakens legal protection
Technical accuracy	Describe the invention correctly
Strategic ambiguity	Keep useful breadth without creating confusion
Completeness & alignment	Include the elements needed to support the claims
Clarity	Communicate the invention and claims clearly

Two independent patent-lawyer raters; 1–5 rubric scores; treatment status hidden.

The regression compares assigned groups within firms

$$z_{irk} = a + \underbrace{\tau}_{\text{Coefficient of Interest}} T_i + \underbrace{\lambda_{f(i)} + \delta_k}_{\text{Firm + Quality effects}} + u_{irk}$$

Outcome: standardized drafting quality	**Day 10**	**Day 90**
Assigned AI access	0.34 (0.16)	0.38 (0.15)
Lawyers completing the task	105	98
Rating × dimension observations	1,050	995
Firm + quality-dimension effects	Yes	Yes

Human expert ratings. Standard errors clustered by lawyer, in parentheses. Tables 3–4, col. 2.

Drafting gains extend beyond clearer prose

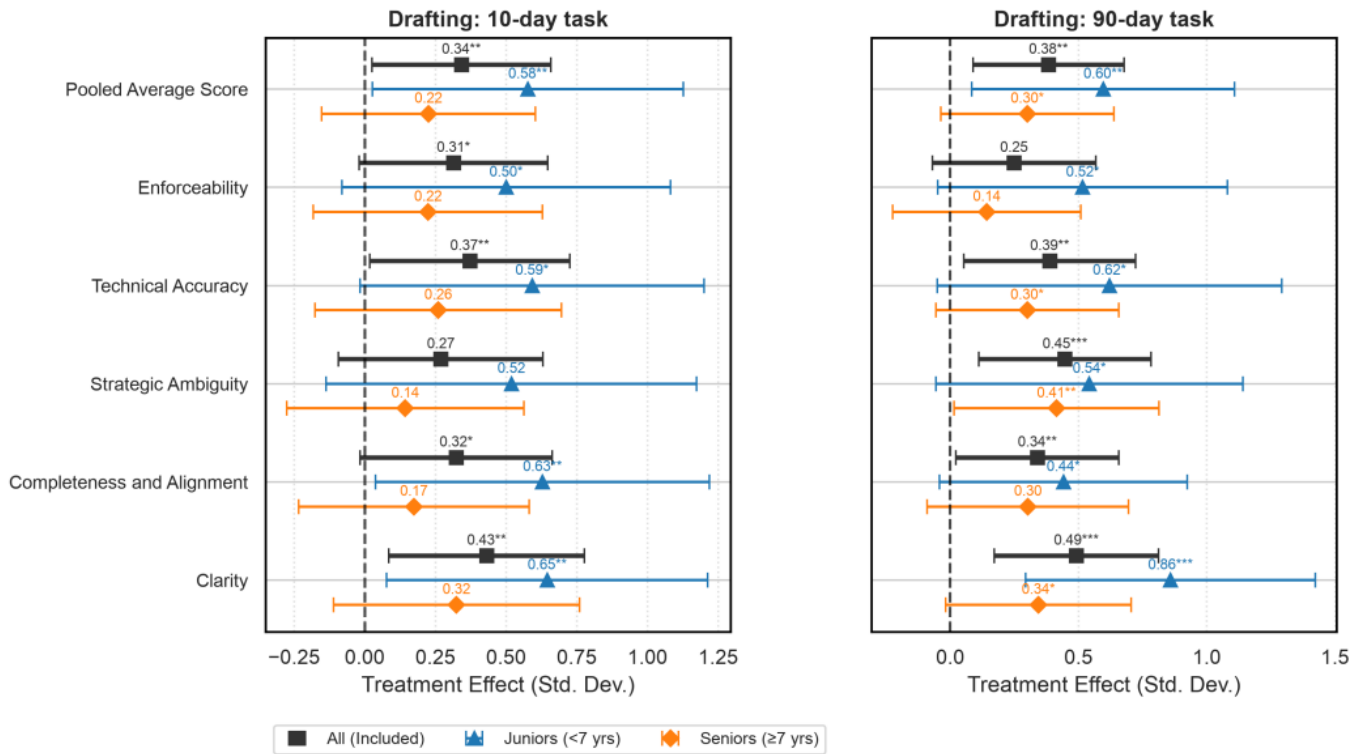


Figure 1, drafting panels · Black: all; blue: junior; orange: senior. Bars show 95% confidence intervals.

> Point estimates improve across all five dimensions.
> Junior lawyers have larger estimated gains in both drafting tests.

Junior lawyers also save time on the first drafting task; Seniors do not!

Effect on minutes spent	Day 10	Day 90
All lawyers	-9.97 (5.11)	-9.61 (8.65)
Junior lawyers	-17.91 (9.13)	-21.63 (18.24)
Senior lawyers	-4.91 (5.22)	-4.54 (9.59)

Table 8, cols. 1–4. Self-reported timed work; robust SEs. N = 102 and 96.

> Early time savings are concentrated among juniors.
> The day-90 estimates are too imprecise to establish time savings.

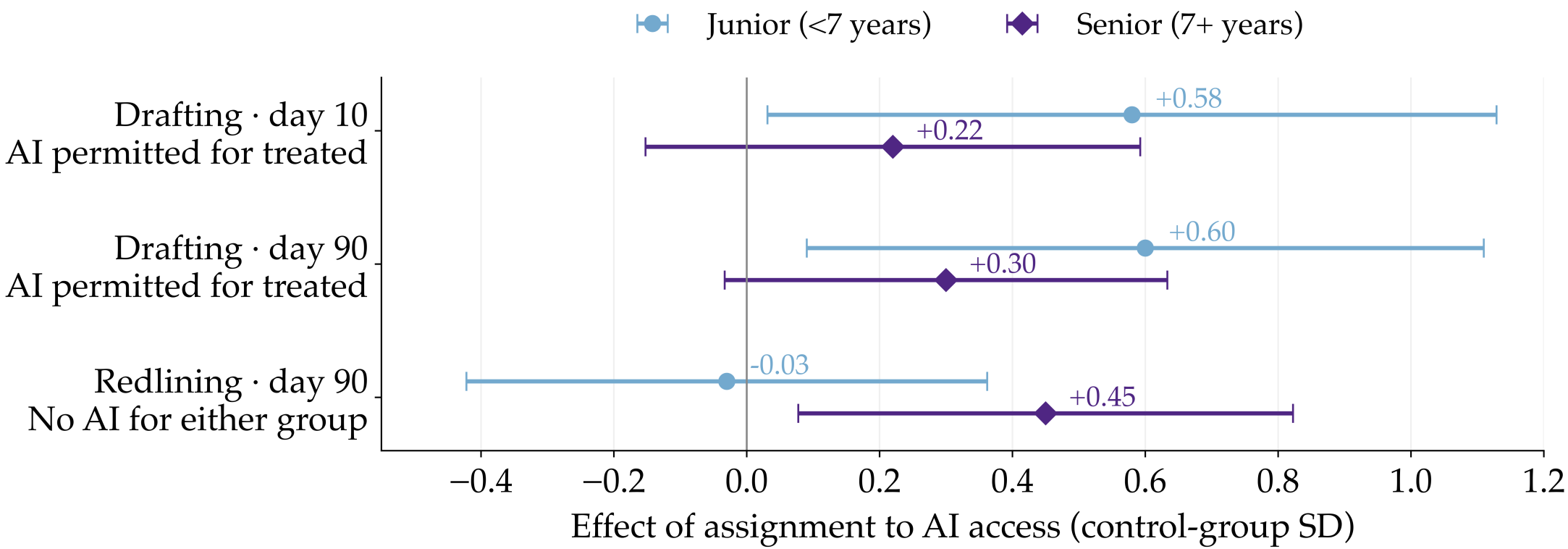
The unaided judgment gain is concentrated among seniors

Effect on day-90 redlining quality	All lawyers	Junior	Senior
Earlier assignment to AI access	0.32 (0.15)	-0.03 (0.20)	0.45 (0.19)
Experience group	All	Under 7 years	7+ years

Table 6, cols. 2 and 4. No AI permitted for either group. 91 lawyers; 925 rating × dimension observations.

> Seniors given earlier access perform better without AI.
> For juniors, the estimate is close to zero and imprecise.

Who benefits depends on which outcome we measure



Published human-rating coefficients; course plot. Intervals: estimate ± 1.96 × rounded SE.

> Juniors' larger drafting gains do not carry over to the unaided test.
> Seniors show the reverse pattern: a clear gain in independent judgment.

Senior gains span the dimensions of judgment

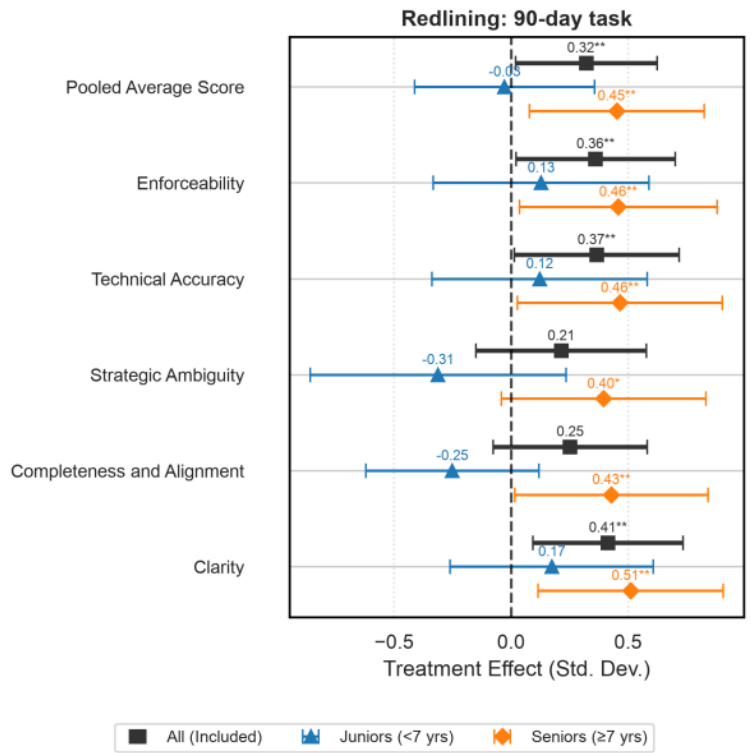


Figure 1, redlining panel · All participants instructed to work without AI.

> Senior estimates are positive across all five dimensions.
> Junior estimates are small or uncertain, with some below zero.

The junior average hides a wider spread of scores

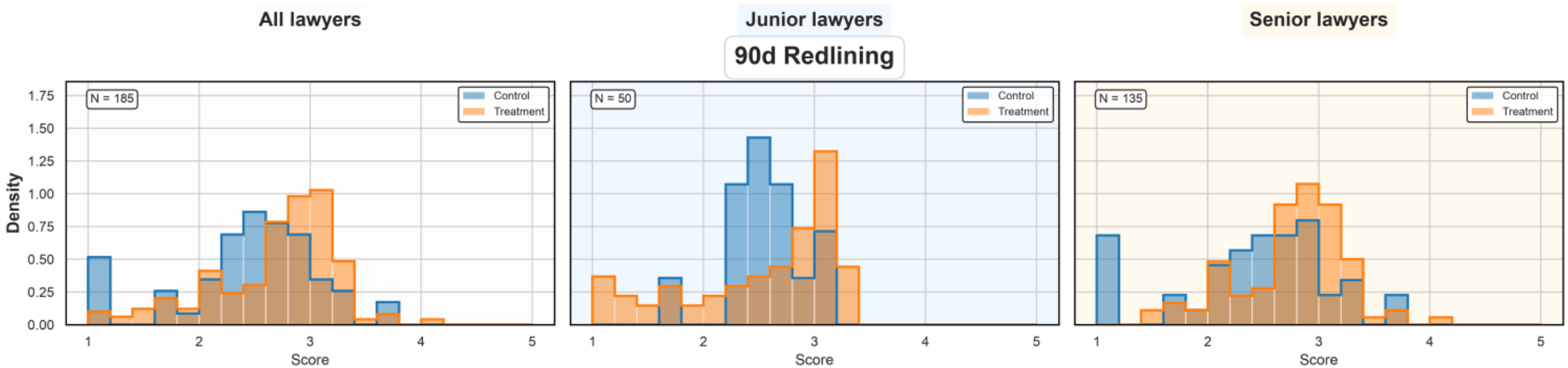


Figure 3 · Blue: control; orange: earlier AI access. Firm-adjusted average ratings, 1–5 scale.
Probability changes by score fifth: juniors vs seniors (appendix)

> Junior scores spread toward both poor and good performance.
> Senior scores shift toward stronger performance overall.

How far can we take the learning result?

- Three months of access; one unaided patent-review exercise
- Completion falls from 133 assigned lawyers to 91 redlining submissions
- No-AI instructions were clear; individual compliance could not be monitored
- Excluding suspected AI use preserves the pattern but reduces precision

> Evidence for improved unaided judgment among seniors.
> No average junior gain is not proof of permanent skill loss.

AI changes output, expertise and the value of work

Paper	What its results establish
Expertise	The tasks left in a job matter for wages and entry
Generative AI at Work	Assistance raises output most for novices and low performers
Patent-lawyer experiment	Better assisted drafting need not produce better unaided judgment

> Who gains depends on the task, the worker and the outcome.
> Current productivity gains alone do not tell us how expertise develops.

Appendix: Employment and wages can move in opposite directions

Change in remaining expertise	Entry and allocation mechanism	Possible wage implication
Lower requirement	More workers qualify for the remaining job	Lower scarcity value for incumbents
Greater importance of expert work	Production relies more on scarce expertise	Higher reward to that expertise

Task quantities and product demand also affect employment.

Appendix: From a mechanism to a workplace

Can AI transmit useful expertise to workers who initially have less of it?

- Observed access to a specific tool
- Consistent outcomes before + after access
- Credible comparison without access

Appendix: Why a before-and-after comparison is incomplete

Productivity could rise even without AI access.

- Worker experience
- Customer problem mix
- Management, staffing & demand

Workers who receive access may also differ from those who never receive it.

Appendix: An illustrative difference-in-differences calculation

Illustrative classroom example · Issues resolved per hour

Group	Before	After	Change
Later receives AI	2.00	2.40	+0.40
Comparison	1.80	1.90	+0.10

Difference in differences: 0.30

Identification requires the comparison-group change to describe the treated group's change without AI. These numbers are invented for teaching.

Appendix: What a “20% productivity gain” would and would not establish

Required evidence	Conclusion it supports
Output with quality held accountable	A gain in useful production, not merely faster responses.
A credible untreated comparison	An effect attributable to access in the studied setting.
Effects across workers and tasks	Who benefits, who does not, and where training helps.
Prices, demand, entry, and pay	Whether productivity gains become jobs or higher wages.

Appendix: Drafting score distributions

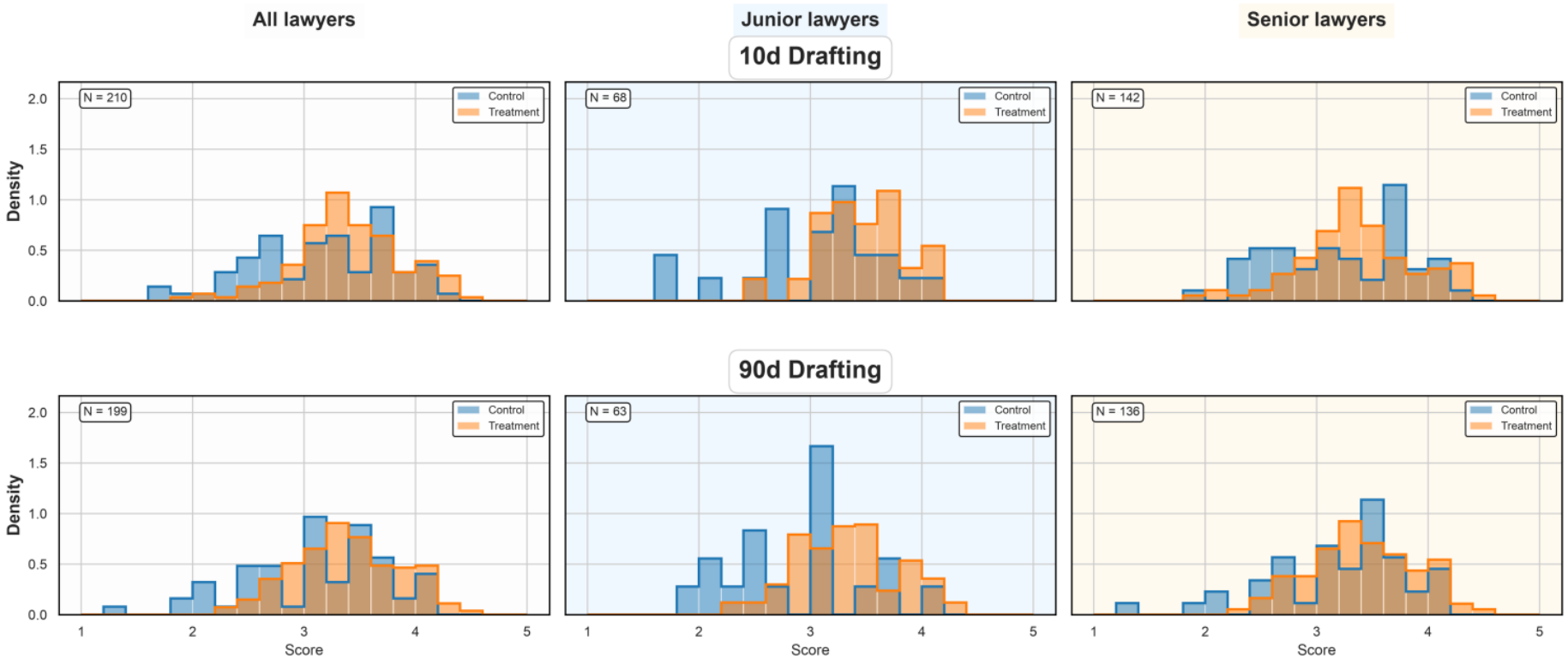


Figure 2 · Blue: control; orange: AI access. Each column is a different experience group.

> The distribution generally moves away from poor drafting scores.
> The amount and precision of this shift differ by task and group.

Appendix: The original redlining regression table

Table 6: Impact of AI Access on Performance in Non-AI-Assisted Redlining Task

	Expert Ratings				LLM Ratings		Pooled Ratings	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	0.32**	0.32**	0.32**		0.56***		0.42***	
	(0.16)	(0.15)	(0.16)		(0.19)		(0.13)	
Treat × Junior				−0.03		0.29		0.12
				(0.20)		(0.36)		(0.20)
Treat × Senior				0.45**		0.66***		0.54***
				(0.19)		(0.22)		(0.16)
Junior				0.21		0.07		−0.05
				(0.12)		(0.30)		(0.15)
Senior				0.09		−0.02		−0.16
				(0.15)		(0.17)		(0.13)
LLM Rater							0.64***	0.64***
							(0.14)	(0.14)
Constant	−0.28	−0.20	−0.19		0.29		−0.19	
	(0.17)	(0.17)	(0.17)		(0.20)		(0.14)	
$H_0 : \alpha_{\text{junior}} < \alpha_{\text{senior}}$				0.74		0.60		0.70
$H_0 : \beta_{\text{junior}} > \beta_{\text{senior}}$				0.96		0.81		0.95
$H_0 : \alpha_{\text{junior}} + \beta_{\text{junior}} > \alpha_{\text{senior}}$				0.35		0.10		0.13
$H_0 : \alpha_{\text{junior}} + \beta_{\text{junior}} < \alpha_{\text{senior}} + \beta_{\text{senior}}$				0.05*		0.13		0.03**
Subject-level Average	Yes	No	No	No	No	No	No	No
Rating-level Disaggregated	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Sub-indicator FE	No	Yes	No	Yes	Yes	Yes	Yes	Yes
Firm × Sub-ind. FE	No	No	Yes	No	No	No	No	No
Observations	91	925	925	925	455	455	1,380	1,380
R^2	0.16	0.10	0.11	0.12	0.26	0.27	0.17	0.18

Notes: Table reports intent-to-treat estimates using OLS regressions. Model (1) reports effects on the subject-level average of subcomponents using robust (HC1) standard errors. Models (2)-(8) report rating-level disaggregated regressions using standard errors clustered at the individual level. Pooled models (7)–(8) control for rater type. All outcome subcomponents are standardized to mean zero and unit variance using the control group calculated locally within the full active sample. Models (1), (2), and (4)-(8) incorporate firm fixed effects, with Models (2) and (4)-(8) also including sub-indicator fixed effects. Model (3) incorporates firm × sub-indicator fixed effects. Models are weighted by $w_i = 1/n_i$, where n_i is the number of ratings individual i receives so that each individual carries equal total weight. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 6 · Human expert ratings in cols. 1–4; LLM and combined ratings in cols. 5–8.

> The lecture uses human ratings: pooled effect in col. 2 and experience-specific effects in col. 4.

Appendix: Excluding suspected AI use reduces precision

Unaided redlining effect	All lawyers	Senior lawyers
Main human-rating estimate	0.32 SD; p = .04	0.45 SD; p = .02
Exclude 15 flagged submissions	0.25 SD; p = .14	0.39 SD; p = .06

> The direction survives this check; the evidence becomes less precise.

Appendix: Do benchmark gains show up in ordinary patent work?

Self-reported work outcome	Effect	Standard error
Weeks to complete a patent	-0.76	0.81
Reviewer input, SD	-0.39	0.28
Edits and corrections, SD	-0.15	0.28

Table 9, cols. 5, 7 and 9. N = 67, 66 and 66; firm effects; robust SEs.

> These estimates are too imprecise to establish improvements in ordinary client work.

Appendix: Comparing occupations over time

Step	Construction
Occupational title	Mean of its task scores.
303 Census occupations	Employment-weighted mean of the underlying titles.
Common scale	Subtract the 1980 occupation mean; divide by its standard deviation.

A task score of **+1** is one baseline occupation standard deviation above the mean. Apply the same scaling to both dates.

> This lets us compare expertise across occupations and over time.
> Vague descriptions or expertise expressed in common words can be missed.

Appendix: Measuring changes in task quantity

Quantity measure	Denominator	Illustrative example
Share of tasks removed	All 1977 tasks	1 of 3 = 33.3%
Share of tasks added	All 2018 tasks	1 of 4 = 25.0%
Net task change	Added share - removed share	25.0% - 33.3% = -8.3 points

> The regressions include expertise change and task-quantity change.
> "Net task change" is a difference of shares, not growth in the raw task count.

Appendix: Changes in redlining scores by fifth

Change in probability of a score in...	Junior	Senior
Highest fifth	+6 (4)	+3 (7)
Second-highest fifth	+19 (15)	+22 (7)
Middle fifth	-5 (10)	-4 (7)
Second-lowest fifth	-42 (16)	-8 (8)
Lowest fifth	+22 (11)	-14 (8)

Percentage points; robust SEs in parentheses. Table 7. Fifths defined from the full valid sample.

> Junior scores leave the second-lowest fifth for both tails.
> For seniors, the clearest increase is in the second-highest fifth.