EXECUTIVE BRIEFING

Generative AI at Work

Evidence from 5,172 Customer-Service Agents on the Real Productivity Impact of GenAI

5,172 agents	**+15%**	**Causal DiD**	**Fortune 500 firm**
Sample size	Avg. productivity gain	Method	Setting

Brynjolfsson, Li & Raymond · NBER WP 31161 / arXiv:2304.11771v2 · 2024

CONTEXT

Why This Study Matters

First rigorous, large-scale, causal evidence of how GenAI changes real workplaces.

BEFORE THIS PAPER

- Lab experiments (GPT writing code/emails)
- Self-reported survey data (McKinsey, Accenture)
- Anecdotes from early adopters
- No causal identification

WHAT THIS PAPER ADDS

- 5,172 real customer-service agents
- Staggered rollout → quasi-random assignment
- Two-way fixed effects DiD
- Outage natural experiment for learning persistence

5,172
Agents observed

12 mo.
Panel window

Multi-country
Incl. PH, IN offshoring sites

Source reference: §1 Introduction; §3 Empirical Setting

What This Deck Covers

WHAT THE DATA SHOWS

01 Findings & Trends

- Average +15% productivity
- But the gain is wildly uneven
- Tenure premium collapses

The Headline Numbers

AI lifts average productivity by 15%, but the gains are deeply uneven.

+15%

Avg. resolutions per hour (RPH) across 5,172 agents

From 1.97 to 2.27 RPH baseline

§Abstract; Table 2

+36%

RPH gain for the lowest-skill quintile

Vs. AI-naive baseline performance

§5.2; Table 3

≈ 0%

RPH gain for the highest-skill quintile

With small declines in quality metrics

§5.2; Table 3

−40%

Reduction in new-hire attrition (≤6 mo. tenure)

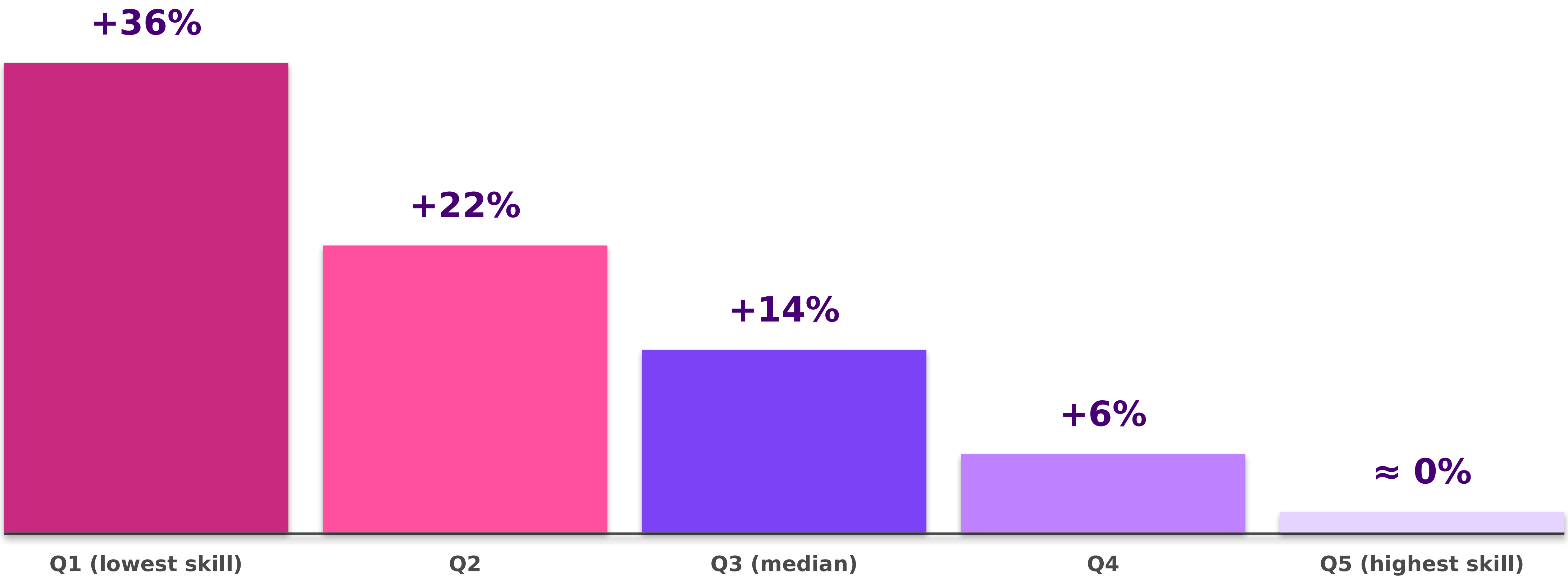
From 25% baseline down to ~15%

§7.2

> *Average gain masks a striking distribution: junior agents leap forward, senior agents barely move. The same intervention, very different outcomes — and that asymmetry is the entire story.*

The Productivity Lift Is Not Evenly Shared

Gains concentrate at the bottom; the top barely moves.



← Largest learning headroom

Already at AI-trained peer level →

Source: §5.2; Table 3 — RPH gain by pre-treatment skill quintile (illustrative; intermediate quintile values approximated from Figures 4–5).

WHAT THE AUTHORS ARE ARGUING

02 Core Views

- GenAI is skill-equalizing
- Tacit knowledge gets encoded
- Top performers train the system

DIRECT FROM THE PAPER

Thesis: GenAI Is Skill-Equalizing

> "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.
>
> **— §Abstract**

Treated agents with two months of tenure perform just as well as untreated agents with more than six months of tenure.

— §6.1 Tenure analysis

WHY THIS MATTERS

This is the first major IT wave that does not behave as skill-biased technical change. The historical pattern — where new tech rewarded the strong and widened inequality — is reversed.

Three Mechanisms Behind the Gains

The lift isn't magic — it's encoded best practice flowing top → bottom.

① Encoded Tacit Knowledge

AI learns top performers' patterns and broadcasts them to everyone — durable learning, not real-time dependency.

EVIDENCE

During AI service outages, previously-treated agents still outperform their pre-AI baseline. Learning persists.

§6.2 outage natural experiment

② Language Uplift

Conversation analysis shows non-native English speakers' fluency converging toward top-performer style.

EVIDENCE

Linguistic features of low-skill agents shift measurably toward those of high-skill peers post-AI exposure.

§6.3 text analysis

③ Rare-Problem Coverage

The largest gains are on rare issues — where the human baseline is weakest, AI net contribution is highest.

EVIDENCE

Counter-intuitive: AI has less training data on rare topics, but human baseline is even worse, so net help is greatest.

§6.4 topic stratification

HOW THIS CHANGES WORK ITSELF

03 Themes

- Compressed learning curves
- Better customer experience
- Lower new-hire attrition

THEME

Theme A — The Learning Curve Is Compressed

Ramp-up time collapses; tenure-based seniority loses its signal value.

WITHOUT AI

- 8–10 months to reach 2.5 RPH
- 25% new-hire attrition within 6 mo.
- Tenure strongly predicts performance
- Senior agents 2× as productive as new hires

WITH AI

- 5 months to reach 3.0 RPH
- ~15% new-hire attrition (−40%)
- Tenure premium weakens substantially
- 2-mo. new hires ≈ 6+ mo. veterans

STRATEGIC IMPLICATION

Traditional 'pay-for-experience' compensation curves were designed for a world where tenure = capability. AI breaks that proxy.

Source: §6.1 Tenure dynamics; §7.2 Attrition analysis

THEME

Theme B — Customer Experience Improves Too

AI assistance doesn't make agents feel robotic — customers respond better.

CUSTOMER BEHAVIOR

- Fewer aggressive / negative messages
- Manager-escalation requests −25%
- From 6.0% baseline to ~4.5%
- Politeness-coded language up materially

ADOPTION PATTERN

- Avg. agent accepts 35% of suggestions
- Adoption rises with exposure time
- Initial skeptics (seniors) warm up most
- Higher adoption → larger productivity gain

WHY THIS IS NOVEL

First large-scale evidence that AI augmentation can improve the experience of work itself — not just throughput.

Source: §7.1 Customer sentiment; §6.5 Adoption dynamics

WHAT THIS MEANS STRATEGICALLY

04 Insights

- Today's win, tomorrow's risk
- Top talent gets flattened
- Training data quietly decays

Strategic Insight: Two Long-Term Risks

Today's productivity win quietly creates two structural problems.

RISK 1 — Top Talent Flattening

If senior agents increasingly defer to AI's median-quality suggestions, their original creative contribution declines.

SIGNAL IN THE DATA

Top quintile shows small declines in quality metrics (resolution rate, NPS) even as speed marginally rises.

→ The very signal future AI models depend on starts to weaken at the source.

RISK 2 — Training Data Decay

Agent conversations are simultaneously customer service AND model training data — but employees are not compensated for the latter.

AUTHORS' FRAMING

"Whether and how workers should be compensated for the data they provide for training AI systems" is left as an open policy question.

→ Without attribution mechanisms, the source of intelligence erodes over time.

Source: §8 Discussion; §Conclusion

WHAT ENTERPRISE LEADERS SHOULD DO NOW

05 Recommended Actions

- Reset HR economics
- Rewire ROI targets
- Treat conversations as capital

RECOMMENDED ACTIONS

Five Moves for Enterprise Leaders

Treat this paper as an operating playbook, not a forecast.

#	ACTION	WHAT TO DO	EXPECTED OUTCOME
1	**Reset HR Economics**	Re-evaluate seniority-based pay curves — AI compresses the experience premium.	*Pay-for-skill replaces pay-for-tenure; reduces overpaying for time served.*
2	**Rewire AI ROI Targets**	Make new-hire retention a hard KPI for AI deployments, not just speed metrics.	*−40% attrition has more durable ROI than +15% throughput.*
3	**Treat Conversations as Capital**	Establish attribution & incentive systems for employees whose work trains AI models.	*Protects long-term training data quality; aligns incentives with model health.*
4	**Compress Onboarding**	Use AI to shrink customer-service ramp-up from 6–9 months to 2–3 months.	*Faster time-to-productivity; lower cost-to-train per agent.*
5	**Protect Top-Talent Originality**	Reserve high-difficulty, AI-free tasks for senior staff to preserve creative output.	*Prevents flattening; preserves the 'human signal' future models need.*

SYNTHESIS

The Big Picture

GenAI is the first IT wave that lifts the bottom — but it asks new questions of the top.

01

What changed

AI shifted from skill-amplifier to skill-equalizer. The 5,172-agent dataset is the strongest causal evidence to date.

02

What it means

Tenure-based seniority loses signal value. New-hire ramp-up compresses. Customer experience gets better, not worse.

03

What to watch

Top-talent flattening and training-data decay are slow-moving risks. They don't show up in this year's KPIs.

> *The window to redesign HR economics, ROI targets, and data-attribution mechanisms is now — while the gains are still accruing.*

CITATION MAP

References & Source

PRIMARY SOURCE

- Brynjolfsson, E., Li, D., & Raymond, L. (2024). Generative AI at Work. NBER Working Paper No. 31161 / arXiv:2304.11771v2 (Nov 7, 2024).
- Published as: Quarterly Journal of Economics, Vol. 140, No. 2, 2025, pp. 889–942.
- URL: https://arxiv.org/abs/2304.11771

METHODOLOGY

- Sample: 5,172 customer-service agents at a Fortune 500 business-process software firm.
- Identification: Two-way fixed effects DiD; Sun-Abraham robust estimators.
- Natural experiment: AI service outages used to test learning persistence.

WHERE EACH STATISTIC APPEARS

- +15% RPH lift — §Abstract; Table 2
- +36% / 0% quintile dispersion — §5.2; Table 3
- −25% manager-escalation — §7.1
- −40% new-hire attrition — §7.2
- 35% AI suggestion adherence — §6.5
- 12% worker-hour savings — Conclusion §8

Thank You

AI Daily Reads — Curated by aidigest.club

Read the full paper

https://arxiv.org/abs/2304.11771