Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

When AI Saves Time but Creates More Work: Digital Task Expansion and Employee Well-Being

Sarfaraz Abu Abdullah

rajon3430@gmail.com

University of Rajshahi https://orcid.org/0009-0000-3826-1626

Research Article

Keywords: Digital Task Expansion, Generative AI, Work Intensification, Employee Well-Being, Job Demands–Resources, Burnout, Productivity Paradox, Human–AI Work

Posted Date: November 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11818440/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

When AI Saves Time but Creates More Work: Digital Task Expansion and Employee Well-Being

[Sarfaraz Abu Abdullah]
[Department of Management, Rajshahi University, Rajshahi, Bangladesh] · Email: [rajon3430@gmail.com]

Abstract

Generative AI is sold as a time-saving technology, yet a growing body of workplace evidence describes people who use it and end up with more to do. This paper examines that contradiction through the idea of digital task expansion: the net growth in the tasks, scope and pace of a job that follows AI adoption, arising through verification and repair, learning and tool management, wider job scope and rising output expectations. It assembles published experiments, national surveys, a register-linked Danish study, an eight-month ethnography and the Gallup workforce panel (n = 21,648 in August 2026), and it applies secondary statistical analysis to the published tables. Time savings measured in the three positive controlled studies (12% to 40%) are roughly five times those measured in population surveys (2.8% to 5.4%), and one randomized trial found experienced developers 19% slower. Among frequent users in the Gallup data, time falls sharply for information gathering (net balance −44 points) and writing (−40) but not for reviewing and correcting work (−2, 95% CI −4.2 to 0.2) or learning new tools (+3, 1.0 to 5.0). Frequent users had nearly three times the odds of saying output expectations have risen (OR 2.95, 95% CI 2.60 to 3.36). In a 6,000-worker survey, frequent verification work rose from 35% of light users to 74% of heavy users. A break-even analysis shows that if overhead equals 58% of each hour saved and half of what remains is reabsorbed by wider scope, only 16% of users keep an hour a week free. The evidence is mostly self-reported and cross-sectional, so task expansion is treated here as a well-supported pattern and harm to well-being as a plausible, partly evidenced pathway. Organizations that measure freed time instead of tool use, and that protect context, quality standards and recovery, are best placed to avoid it.

Keywords

Digital Task Expansion, Generative AI, Work Intensification, Employee Well-Being, Job Demands–Resources, Burnout, Productivity Paradox, Human–AI Work

1. Introduction

Ask almost anyone who has adopted a generative AI tool at work what it does for them and the first answer is that it saves time. The second answer, if you keep asking, is often less comfortable. The draft arrives in seconds but has to be checked line by line. The summary is tidy and misses the one point the client cares about. A colleague forwards a polished memo that has to be rewritten. And the manager, noticing that reports now take half as long, asks for twice as many.

This paper is about the gap between those two answers. It uses the term digital task expansion for the growth in the number, scope and pace of tasks in a job that follows the adoption of AI tools, and it asks whether that growth cancels the time saved and, if it does, what it costs the people doing the work.

The puzzle has a numerical form. Controlled experiments report that AI cuts the time of a task by a quarter or more (Noy & Zhang, 2023; Dell'Acqua et al., 2023). When economists link surveys of Danish workers to administrative records, however, the average time saving is 2.8% of working hours, and there is no detectable effect on earnings or recorded hours (Humlum & Vestergaard, 2025). A national U.S. survey puts the saving at 5.4% of hours among users (Bick et al., 2025). A randomized trial found that experienced open-source developers took 19% longer with AI tools while believing they had been faster (Becker et al., 2025). And an eight-month field study of a technology company of about 200 employees reported that AI intensified work and did not reduce it (Ranganathan & Ye, 2026).

1.1. Statement of the Problem

Organizations are adopting AI on the premise that it frees time. In a 2024 survey by the Upwork Research Institute, 96% of C-suite executives expected AI to raise productivity, while 77% of employees who used it said it had added to their workload in at least one way (Upwork Research Institute, 2024). If the time saved is absorbed by new tasks, or exceeded by them, the business case rests on a benefit that employees do not experience, and the cost falls on them in the form of demands without matching resources. That is the situation the job demands–resources literature associates with exhaustion (Bakker & Demerouti, 2017).

The evidence on the matter is scattered. It sits in randomized trials, working papers, vendor surveys, consultancy reports and a handful of ethnographies, each with its own measure, sample and incentive. No study has laid the pieces next to each other, reduced them to comparable quantities and asked how much of the apparent contradiction is real and how much is an artifact of measurement. Managers who must decide how to roll out these tools, and occupational health bodies that must decide whether to treat AI as a psychosocial risk, are left with anecdotes and headline percentages.

1.2. Research Questions and Purpose

The paper asks five questions.

RQ1 How large is the gap between time savings measured in experiments and time savings measured in workplaces and populations?

RQ2 Which tasks lose time and which do not shrink or grow among frequent AI users?

RQ3 Do output expectations, pressure and hidden labor rise with the intensity of AI use?

RQ4 Under what conditions does an individual keep any of the time saved, once verification overhead and reabsorption into wider scope are taken into account?

RQ5 What does the evidence say about the connection to employee well-being, and which organizational resources appear to moderate it?

The purpose is to define digital task expansion, to test its signatures in the best available public data, and to give organizations a simple way to check whether the time saved by AI is reaching the people who use it.

1.3. Scope and Contribution

The study is a secondary analysis. It collects no new primary data, it cannot identify causal effects, and much of the evidence it uses is self-reported. Within those limits it makes three contributions. First, it separates four channels through which expansion occurs, which are usually discussed as a single lump of extra work. Second, it applies statistical analysis that the original reports did not carry out: design-adjusted tests on the Gallup tables, a trend analysis of adoption, relative-risk and trend analysis of a dose–response in a 6,000-worker survey, and a break-even analysis of freed time. Third, it states plainly which conclusions the evidence supports, which it only suggests, and where the sources disagree.

Section 2 reviews the literature and states six hypotheses. Section 3 describes the data and methods. Section 4 reports the results, Section 5 discusses them, Section 6 draws implications, Section 7 sets out limitations, and Section 8 concludes.

2. Literature Review and Hypotheses

2.1. The Promise and the Measured Gains

The early experimental literature gave good reasons for optimism. In an online experiment with college-educated professionals, ChatGPT cut the time taken on writing tasks by about 40% and raised rated quality by 18% (Noy & Zhang, 2023). In a customer-support firm, access to an AI assistant raised the number of issues resolved per hour by 14% on average and by 34% for novice workers (Brynjolfsson et al., 2025).

Consultants given GPT-4 completed tasks inside its capability frontier more quickly and with higher quality, but were 19 percentage points less likely to be correct on a task outside it (Dell'Acqua et al., 2023).

These are task-level results. A job is a bundle of tasks plus the work of coordinating them, and a saving on one task need not show up as a saving in the working week. The productivity J-curve argument makes the point at the level of the firm: general-purpose technologies need complementary investments in process and skills before measured gains appear, and in the meantime the costs of adjustment are real (Brynjolfsson et al., 2021). The population-level estimates fit that view. Workers who used generative AI in the United States reported saving 5.4% of their work hours, which corresponds to 1.4% of all hours in the economy (Bick et al., 2025). In Denmark, where about 25,000 workers in 11 exposed occupations were surveyed and matched to employer and tax records, users reported saving 2.8% of their time, yet the authors could rule out effects on earnings or hours larger than 1% (Humlum & Vestergaard, 2025).

2.2. Why Time Saved Does Not Become Time Free

Four channels can turn a saving on one task into more work overall. They differ in mechanism and in what would reduce them, so it is worth keeping them apart.

Verification and repair. Output from a generative model has to be checked, and the checking is done by a person who often knows less about the content than the person who would have written it. Bainbridge (1983) described the irony in older automation: the operator is left with the monitoring and the exceptions, which are harder than the routine work that was removed. Parasuraman and Manzey (2010) show that monitoring is also where complacency and automation bias creep in. Practitioners now have names for the new version. Glean's Work AI Institute calls supervising, debugging and re-prompting AI output botsitting, and finds that it takes more than a third of the time employees spend on AI (Hinds et al., 2026). Niederhoffer et al. (2025) call polished but substance-poor AI output passed between colleagues workslop and report that it shifts the cost of fixing to the person who receives it.

Learning and tool management. New tools need new skills, new prompts, new context and new routines for deciding when to use them. The time involved is a fixed cost for some users and a running one for others whose tools change every few months. In the Danish study, new tasks created by AI appeared for 8.4% of workers, including some who did not use the tools themselves (Humlum & Vestergaard, 2025).

Scope widening and rising expectations. Saved time is rarely left alone. Parkinson's (1955) old observation was that work expands to fill the time available. Jevons (1865) noted that more efficient use of coal increased its consumption. Applied to work, the point is that a faster process often raises the volume demanded of it. Green (2004) links rising work effort and falling discretion to a decline in well-being at work in Britain. The Berkeley field study finds the same pattern inside one firm: employees took on tasks that had previously belonged to others or that they would have deferred, because AI made them feel doable (Ranganathan & Ye, 2026).

Blurred boundaries and multitasking. The same study found that prompts were started in breaks and at the end of the day because they felt light, and that people ran several AI threads alongside their own work, so that attention was divided more often (Ranganathan & Ye, 2026). The tasks themselves may be small. The recovery time they erode is not.

2.3. Work Demands and Well-Being

The job demands–resources model holds that strain arises when demands are high and resources, including autonomy, support and feedback, are insufficient to meet them (Bakker & Demerouti, 2017). Conservation of resources theory adds that people try to protect and build resources, that loss is more powerful than gain, and that initial loss leads to further loss (Hobfoll, 1989). A meta-analysis of burnout correlates found that workload, role ambiguity and role conflict were positively related to emotional exhaustion while control and autonomy were negatively related to it (Alarcon, 2011). Longitudinal evidence indicates that the links

between job stressors and burnout run in both directions, which is the loss spiral that conservation of resources predicts (Guthier et al., 2020).

Not all demands harm equally. Challenge demands, which carry a prospect of growth or reward, tend to relate positively to engagement, while hindrance demands such as red tape and role conflict relate to strain (Crawford et al., 2010; LePine et al., 2005). Checking, repairing and re-doing machine output looks like a hindrance demand: it takes effort and delivers no development. Wider scope can be a challenge demand when it comes with autonomy and recognition and a hindrance demand when it arrives as an unmeasured increase in the load. Earlier work on technostress described similar effects of information technology on role overload and productivity (Tarafdar et al., 2007). Research on AI at work reaches related conclusions. Across four studies, employees who interacted more frequently with AI systems reported more loneliness, insomnia and after-work drinking (Tang et al., 2023).

2.4. Framework and Hypotheses

Figure 1 summarizes the framework. AI adoption and the intensity of use produce gross time savings on the tasks AI substitutes. The four channels of expansion add work. The net of the two is the time dividend, and the demand load that results interacts with the resources available to determine well-being. Six hypotheses follow.

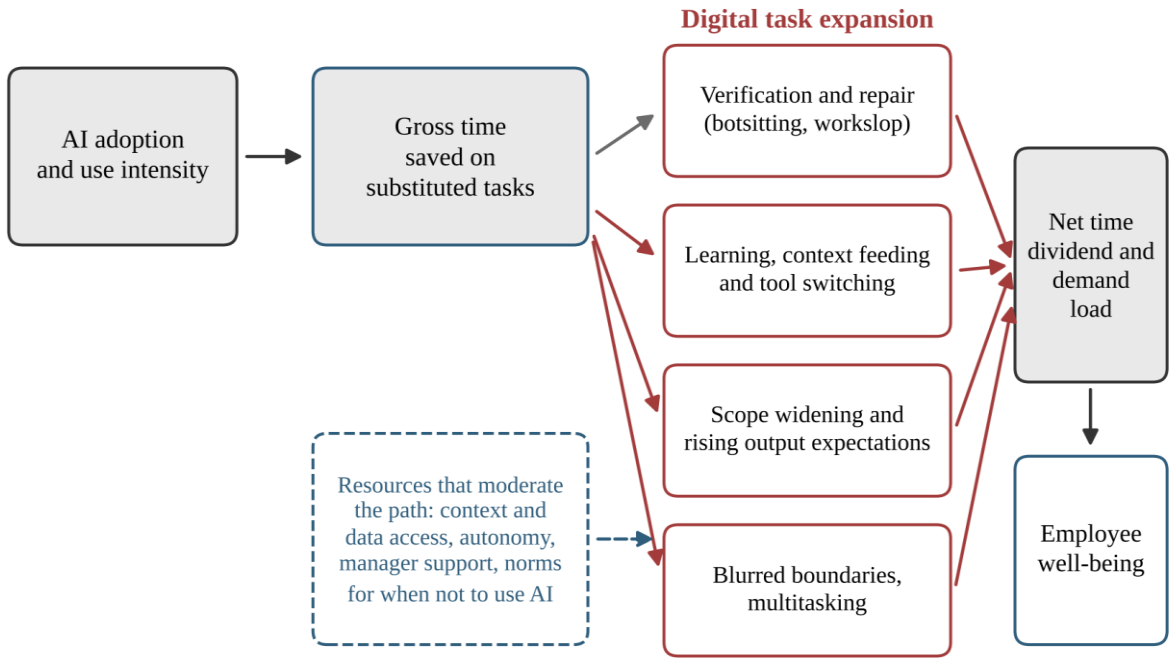

Figure 1. Conceptual framework: digital task expansion between AI use and well-being.

H1 Time savings measured in controlled experiments are much larger than those measured in population surveys of work hours.

H2 Among frequent users, time falls sharply on tasks AI substitutes but does not fall on review, correction and learning, so that the time is reallocated and not released.

H3 Expectations of higher output and pressure to use AI are higher among frequent than infrequent users.

H4 Hidden labor (verification, cleanup, inability to explain AI output) rises with the intensity of AI use.

H5 The probability that a user keeps at least an hour a week of freed time depends more on the overhead ratio and on reabsorption than on the size of gross savings.

H6 Workers with more organizational resources for using AI, such as context and data access, report less exhaustion than those with fewer.

3. Data and Methods

3.1. Research Design

The study is a structured secondary synthesis with three analytical components. The first is an evidence map that puts published time-saving estimates on a common scale and compares them by study design. The second is a secondary statistical analysis of published tables from two large surveys, using the sample sizes and design effects that the publishers report. The third is a break-even analysis that asks how much freed time a user keeps under different assumptions about verification overhead and reabsorption. The design is descriptive. It does not estimate causal effects and it does not claim to be exhaustive.

3.2. Sources and Selection

Sources were chosen because they report quantities that can be placed on a common scale and bear on at least one hypothesis, and because together they span experiments, field studies, population surveys, workplace surveys and qualitative work. Searches were run in November 2026 for published or publicly released studies on AI, time savings, workload, task change and well-being, and each study used here was compared with its original publication or official release where one could be opened. Vendor-affiliated sources (Upwork, Glean, BetterUp) are used because they contain measures found nowhere else, and they are flagged as such throughout. Table 1 lists each source with its design, sample and verification status.

Table 1. Evidence base and verification status.

Source	Design and sample	Used for	Verification
Gallup (2026)	Probability-based U.S. employee panel; *n* = 21,648; 5–19 Aug 2026; ±0.9 points; design effect 1.90	H1–H3	Read from Gallup's published tables (16 Nov 2026)
Hinds et al. (2026), Glean Work AI Index	6,000 full-time digital workers (U.S., U.K., Australia); Dec 2025–Jan 2026; vendor-affiliated with academic co-authors; tier sizes not reported	H4, H6	Read from the published report; self-reported
Upwork Research Institute (2024)	2,500 respondents (1,250 C-suite, 625 employees, 625 freelancers); U.S., U.K., Australia, Canada; Apr–May 2024; base for the 77% item is AI-using employees, *n* not reported	H3, H6	Report headline figures matched in several summaries
Niederhoffer et al. (2025)	1,150 U.S. full-time employees; BetterUp Labs with Stanford Social Media Lab	H4	Headline figures matched in several summaries
Ranganathan & Ye (2026)	Eight-month ethnography (Apr–Dec 2025), U.S. technology firm of about 200 employees; observation and 40 interviews	Mechanisms	Haas newsroom release and summaries of the HBR article
Humlum & Vestergaard (2025)	About 25,000 Danish workers, 7,000 workplaces, 11 occupations; surveys linked to registers; 2023–24	H1	Working-paper abstract and summaries
Bick et al. (2025)	Nationally representative U.S. surveys; time saved 5.4% of work hours	H1, H5	Fed working paper and St. Louis Fed publications; the 20.5% figure via a summary
Becker et al. (2025)	Randomized trial, 16 experienced open-source developers, 246 tasks, early-2025 tools	H1	Several summaries; arXiv record
Noy & Zhang (2023); Brynjolfsson et al. (2025); Dell'Acqua et al. (2023)	Online experiment; customer-support field study; consultant field experiment	H1	Abstracts and summaries
Tang et al. (2023)	Four studies (*n* = 166, 126, 214, 294), Taiwan, Indonesia, U.S., Malaysia	Well-being	APA press release and summaries

Source	Design and sample	Used for	Verification
EU-OSHA (2025)	OSH Pulse 2025, 25,688 EU27 interviews, Mar–Apr 2025	Section 6.2	Official infographic

Note. Verification was carried out in November 2026. "Summaries" are reports and news items that quote the original; the original was not opened for those figures.

3.3. Secondary Analysis of the Gallup Tables

The main quantitative source is Gallup's quarterly workforce survey of August 2026, a probability-based panel of 21,648 U.S. employees fielded from 5 to 19 August 2026, with a margin of error of ±0.9 percentage points and a design effect of 1.9045 (Gallup, 2026). The design effect inflates variances for weighting and clustering, so an effective sample size is $n_{\text{eff}} = n / 1.9045 = 11{,}367$. Gallup reports that 32% of employees use AI frequently (a few times a week or more) and 53% use it at least a few times a year; the infrequent group is therefore taken as 21% of employees. Subgroup effective sizes are approximated as 3,637 for frequent users and 2,387 for infrequent users. Because subgroup sizes are not published, all confidence intervals for subgroup statistics are approximate and are best read as indicating precision, not as exact coverage.

For each of twelve activities, frequent users report whether they spend less, about the same, or more time than before they used AI. Because the three categories come from one respondent, the net balance, the share reporting more minus the share reporting less, has a multinomial variance:

$$\operatorname{Var}(p_{\text{m}} - p_{\text{l}}) = [p_{\text{m}} + p_{\text{l}} - (p_{\text{m}} - p_{\text{l}})^2] / n_{\text{eff}} \tag{1}$$

Two-sided z tests are adjusted for the twelve comparisons with the Holm (1979) procedure. Differences between frequent and infrequent users in the shares agreeing with an item are expressed as odds ratios with Woolf (1955) intervals computed on the effective sizes. Adoption trends across the eight published waves, from May 2023 to August 2026, are fitted by ordinary least squares, with slopes reported in percentage points per year.

3.4. Dose–Response in the Work AI Index

The Glean Work AI Index is a survey of 6,000 full-time digital workers in the United States, United Kingdom and Australia, fielded between December 2025 and January 2026 (Hinds et al., 2026). It reports behaviors and outcomes by share of work time involving AI: light (1–19%), moderate (20–49%) and heavy (50% or more). The tier sizes are not published. To avoid understating uncertainty, relative risks of heavy against light use, with log-scale intervals, and Cochran–Armitage trend tests (Cochran, 1954; Armitage, 1955) are computed under the deliberately conservative assumption of 500 respondents per tier. Results that remain clear at that size would remain clear at larger ones.

3.5. Break-Even Analysis of Freed Time

Let G be the gross time a user saves per week on substituted tasks. Let ρ be the overhead ratio, the hours of verification, repair and learning created per hour saved, and let α be the share of the remaining saving that is reabsorbed by wider scope and higher expectations. Net free time is

$$F = G\,(1 - \rho)\,(1 - \alpha) \tag{2}$$

with $F \leq 0$ whenever $\rho \geq 1$. The mean of G is set at 2.16 hours per week, which is 5.4% of a 40-hour week (Bick et al., 2025). The same source, as summarized by the Federal Reserve Bank of St. Louis, indicates that 20.5% of users save four or more hours per week. A gamma or lognormal distribution with that mean can reach at most about 17% and 13% above four hours respectively, so neither reproduces both moments. A spliced distribution does: with probability 0.205 the saving is 4 hours plus an exponential excess with mean 3 hours, and otherwise it follows a truncated exponential on 0 to 4 hours with the mean needed to

match the overall average (0.91 hours). Mean excesses of 2 and 4 hours are used as sensitivity checks. Results come from 400,000 draws per distribution (Rubinstein & Kroese, 2017).

Anchors for ρ are taken from the Glean ledger, which reports both claimed savings (11 hours per week) and the time spent on botsitting (6.4 hours) and on learning and building (27% of AI time). Because the ratio is taken within one self-reported sample, it does not depend on that sample's heavy use. ρ is not estimated from a single dataset, and the grid in Section 4.6 should be read as a scenario analysis, not a forecast.

3.6. Verification and Reproducibility

All statistics were computed in Python with NumPy and SciPy from the published inputs in Tables 2 to 8, with fixed seeds for the simulations. Where a figure could be checked only through summaries of the original report, the table in which it appears says so.

4. Results

4.1. The Scale of Exposure

Whatever task expansion exists now reaches many more people than it did when most experiments were run. Gallup's series shows the share of U.S. employees using AI at work at least a few times a year rising from 21% in May 2023 to 53% in August 2026, frequent use (a few times a week or more) from 11% to 32%, and daily use from 4% to 16% (Table 3, Figure 2). A linear trend fitted to all eight waves gives 10.7 percentage points per year for any use (95% CI 9.3 to 12.2, R^2 = .98). Frequent use grew by 7.0 points per year over the whole period and by 10.1 points per year (7.9 to 12.2) since May 2025, and daily use by 4.0 and 6.4 points (5.4 to 7.4) respectively. The early slopes rest on three waves and their intervals are too wide to say whether growth accelerated (frequent use: 4.0, −18.0 to 26.0). What the data do show is that the share of frequent users rose by 13 points in the fifteen months to August 2026.

Table 3. Gallup series of AI use at work and trend estimates.

Panel A. Share of U.S. employees using AI at work (%)

Wave	Any use	Frequent use	Daily use
May 2023	21	11	4
May 2024	27	12	4
May 2025	40	19	8
Aug 2025	45	23	10
Nov 2025	46	26	12
Feb 2026	50	28	13
May 2026	52	30	15
Aug 2026	53	32	16

Panel B. Linear trend, percentage points per year (95% CI)

Period	Any use	Frequent use	Daily use
All eight waves	10.7 (9.3 to 12.2)	7.0 (5.1 to 9.0)	4.0 (2.7 to 5.3)
May 2025 to Aug 2026	10.3 (7.1 to 13.4)	10.1 (7.9 to 12.2)	6.4 (5.4 to 7.4)
May 2023 to May 2025 (3 waves)	9.5 (−16.2 to 35.2)	4.0 (−18.0 to 26.0)	2.0 (−12.7 to 16.7)

Note. Wave values as published by Gallup (2026); "frequent" is a few times a week or more. The three-wave estimates are uninformative and shown for completeness.

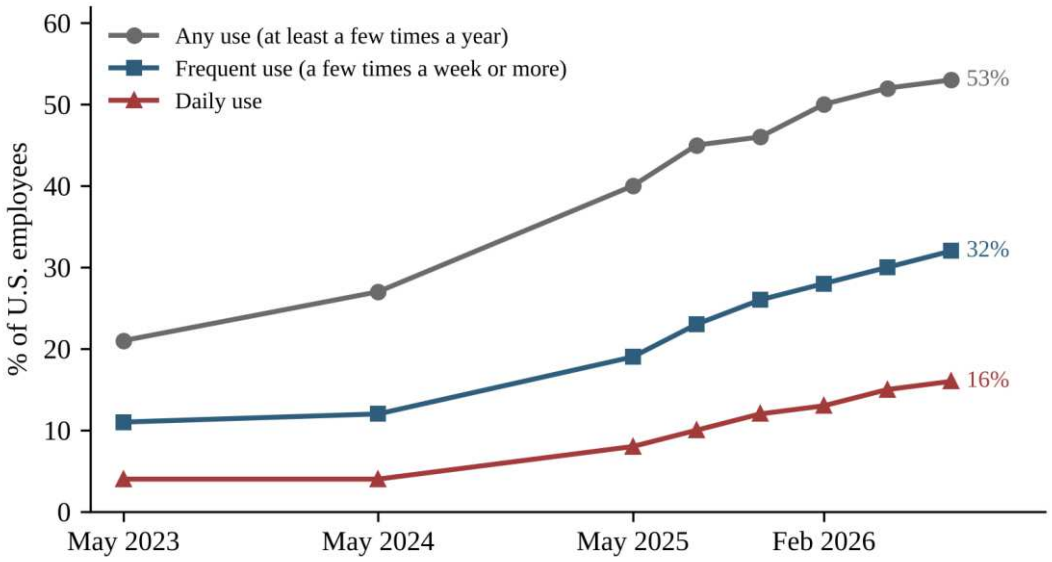

Figure 2. Share of U.S. employees using AI at work, May 2023 to August 2026.
Source: Gallup (2026).

4.2. The Savings Gap (H1)

Table 2 and Figure 3 put the published time-saving estimates on one scale, the percentage of baseline time saved. The three controlled studies with positive results give 40% (Noy & Zhang, 2023), 25.1% (Dell'Acqua et al., 2023, for tasks inside the frontier) and 12.3%, which is the 14% gain in resolutions per hour of Brynjolfsson et al. (2025) converted to time per resolution (1 − 1/1.14). Their median is 25.1%. The two population estimates are 5.4% (Bick et al., 2025) and 2.8% (Humlum & Vestergaard, 2025), with a median of 4.1%. The ratio of medians is 6.1. If the one negative trial is included, in which experienced developers took 19% longer, the experimental median falls to 18.7% and the ratio to 4.6. The savings measured in experiments are therefore roughly five times those measured in populations. H1 is supported, in the descriptive sense that is all seven heterogeneous studies allow. The ratio indexes magnitude; it is not an estimate with a sampling distribution.

Table 2. Published estimates of time saved from generative AI, by study design.

Study	Design and population	Outcome as reported	Setting type	Time saved, % of baseline
Noy & Zhang (2023)	Online experiment; college-educated professionals; writing tasks	Time taken about 40% lower; quality +18%	Controlled	40.0
Dell'Acqua et al. (2023)	Field experiment; 758 consultants; tasks inside the AI frontier	Tasks completed 25.1% faster; outside the frontier 19 points less likely correct	Controlled	25.1
Brynjolfsson et al. (2025)	Field study; customer-support agents	Issues resolved per hour +14% (novices +34%)	Controlled	12.3*
Becker et al. (2025)	Randomized trial; experienced developers; early-2025 tools	Completion time 19% longer; developers believed they were about 20% faster	Controlled	−19.0
Bick et al. (2025)	National surveys; U.S. workers using generative AI	Time saved 5.4% of work hours (about 2.2 hours a week)	Population survey	5.4
Humlum & Vestergaard (2025)	Danish surveys linked to registers	Time saved 2.8%; no effect on earnings or hours above 1%; new tasks for 8.4% of workers	Population survey	2.8

Study	Design and population	Outcome as reported	Setting type	Time saved, % of baseline
Hinds et al. (2026)	Vendor survey; digital workers	Self-estimated saving of 11 hours a week	Vendor survey	27.5†

Note. *Converted from throughput: 1 − 1/1.14. †11 hours of an assumed 40-hour week; self-reported. Median of the three positive controlled estimates 25.1; median of the two population estimates 4.1; ratio 6.1. Including the developer trial, the controlled median is 18.7 and the ratio 4.6. The ratio indexes magnitude and has no sampling distribution.

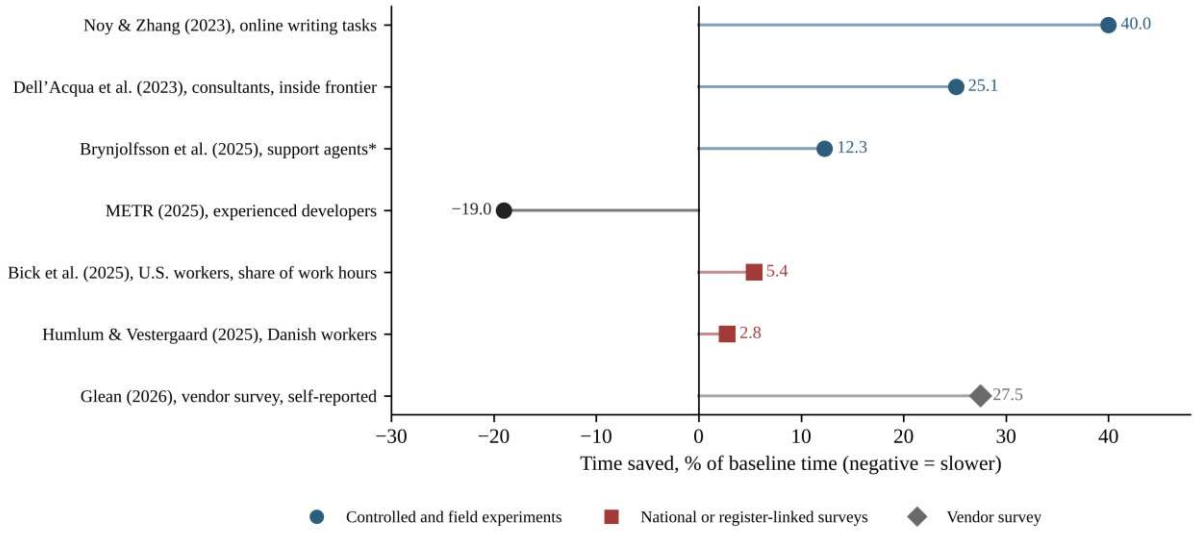

Figure 3. Published estimates of time saved by generative AI, by study design.
*Brynjolfsson et al. (2025), converted from a 14% throughput gain. Vendor estimate is self-reported (11 hours of a 40-hour week).

Self-perception is an unreliable guide to the true figure in either direction. The developers in the randomized trial expected AI to speed them up, and after the study still believed it had made them about 20% faster, a gap of 39 percentage points from the measured 19% slowdown (Becker et al., 2025). The vendor survey from Glean reports a self-estimated saving of 11 hours a week (Hinds et al., 2026), 5.1 times the 2.16 hours (5.4% of a 40-hour week) in the national survey. Four things plausibly account for the gaps. Experiments measure tasks chosen to suit the tool and finished within a session, whereas jobs include coordination and the tasks the tool cannot touch. Population surveys measure whole working weeks. Vendor samples are skewed to heavy users. And all of the household and workplace estimates are self-reports.

4.3. Reallocation, Not Release (H2)

Table 4 and Figure 4 give the change in time use reported by frequent users in the Gallup survey. The largest net reductions are in the tasks AI substitutes most directly: searching for or gathering information (net balance of more minus less −44.0 points, 95% CI −46.1 to −41.9), writing and drafting (−40.0), analyzing data (−27.0), manual or administrative tasks (−22.0) and creative or idea-generation work (−16.0). The mean net balance across those five activities is −29.8 points. The picture is different for the two activities that look like overhead. For reviewing and correcting work, 23% report less time and 21% more, a net balance of −2.0 points (−4.2 to 0.2; Holm-adjusted p = .069), which cannot be distinguished from zero. For learning new skills or tools, more outnumbers less by 3.0 points (1.0 to 5.0; Holm p = .015), and attending meetings shows a similar net rise of 2.0 (0.6 to 3.4; Holm p = .015). The mean net balance across reviewing and learning is +0.5 points, against −29.8 for the substituted tasks.

Table 4. Change in time spent by activity among frequent AI users (Gallup, August 2026).

Activity	Less %	Same %	More %	Net balance	95% CI	*z*	Holm *p*
Searching for or gathering information	52	40	8	−44.0	−46.1 to −41.9	−41.63	<.001
Writing and drafting	47	47	7	−40.0	−42.0 to −38.0	−39.13	<.001
Analyzing data	38	51	11	−27.0	−29.1 to −24.9	−25.21	<.001
Manual or administrative tasks	29	64	7	−22.0	−23.8 to −20.2	−23.77	<.001
Creative or idea-generation work	29	58	13	−16.0	−18.0 to −14.0	−15.37	<.001
Deep thinking or problem solving	22	64	13	−9.0	−10.9 to −7.1	−9.28	<.001
Planning and prioritizing	16	73	11	−5.0	−6.7 to −3.3	−5.83	<.001
Collaborating with coworkers	13	78	9	−4.0	−5.5 to −2.5	−5.16	<.001
Reviewing and correcting work	23	56	21	−2.0	−4.2 to 0.2	−1.82	.069
Working with customers or clients	10	82	8	−2.0	−3.4 to −0.6	−2.85	.015
Attending meetings	8	83	10	+2.0	0.6 to 3.4	2.85	.015
Learning new skills or tools	18	60	21	+3.0	1.0 to 5.0	2.90	.015

Note. Net balance = % more minus % less, in percentage points. Variance from Equation 1 with effective frequent-user sample 3,637 (21,648 × 0.32 / 1.9045). *p* values are two-sided and Holm-adjusted across 12 activities. Rows sum to 100 within rounding.

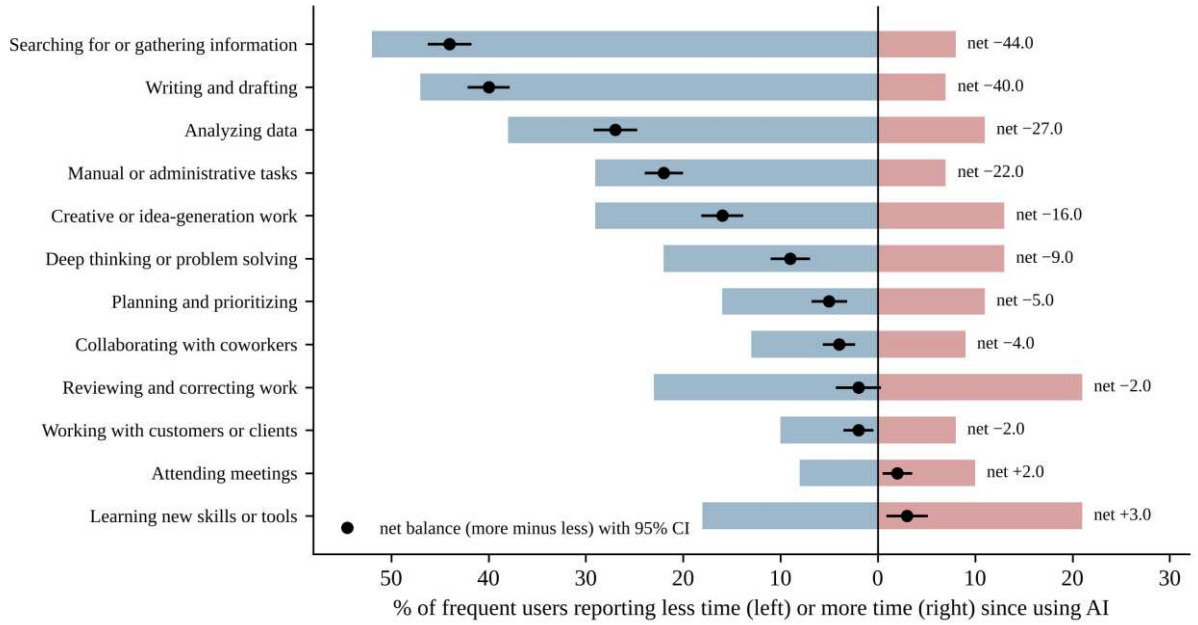

Figure 4. Change in time use among frequent AI users, by activity.

Source: Gallup (2026), analysis in Table 4. Bars show shares reporting less or more time; the remainder report about the same time.

A net balance hides movement in both directions. Of frequent users, 44% say their time on reviewing and correcting changed (23% down and 21% up) and 39% say the same of learning (18% down and 21% up). In other words, roughly one in five frequent users is spending more time checking and one in five more time learning, while the net across the group is flat. The same data show that AI does lighten some people's burdens: 21% of frequent users report that their mental effort is lower than before, against 14% who report it is higher. H2 is therefore supported in the form in which it was stated. Time saved on substituted tasks is not matched by savings on review and learning, so the work is reallocated and not released. It is not supported in the stronger form that review and learning grow for the average frequent user, which holds only for learning.

4.4. Expectations and Pressure (H3)

Table 5 and Figure 5 compare frequent and infrequent users. Among frequent users, 36% agree that expectations for their daily output have risen since AI was introduced, against 16% of infrequent users, an odds ratio of 2.95 (95% CI 2.60 to 3.36). Pressure to use AI in order to keep up is felt by 33% of frequent and 17% of infrequent users (OR 2.40, 2.12 to 2.73). Mental effort is reported to have risen by 14% of frequent users and 10% of infrequent ones (OR 1.47, 1.24 to 1.72) and to have fallen by 21% and 13% (OR 1.78, 1.54 to 2.05). Frequent use is thus associated with more polarized experience: more people feel relief and more feel strain. H3 is supported, with the proviso that these are bivariate comparisons that do not adjust for occupation, seniority or sector.

Table 5. Expectations, pressure and mental effort: frequent vs infrequent AI users.

Panel A. Gallup (August 2026)

Item	Frequent %	Infrequent %	Odds ratio (95% CI)	*p*
Expectations for daily output have risen	36	16	2.95 (2.60–3.36)	<.001
Pressure to use AI to keep up with expectations	33	17	2.40 (2.12–2.73)	<.001
Mental effort higher than before	14	10	1.47 (1.24–1.72)	<.001
Mental effort lower than before	21	13	1.78 (1.54–2.05)	<.001

Panel B. Upwork Research Institute (2024), as reported

Item	Value
C-suite executives who increased demands on workers in the past year	81%
AI-using employees asked to do more work because of AI	21%
AI-using employees spending more time reviewing or moderating AI content	39%
AI-using employees spending more time learning to use AI tools	23%
AI-using employees saying AI added to their workload in at least one way	77%

Note. Panel A: odds ratios with Woolf intervals on effective subgroup sizes of 3,637 (frequent) and 2,387 (infrequent); bivariate and unadjusted. Panel B: the base for the percentages is employees who use AI; the number is not reported.

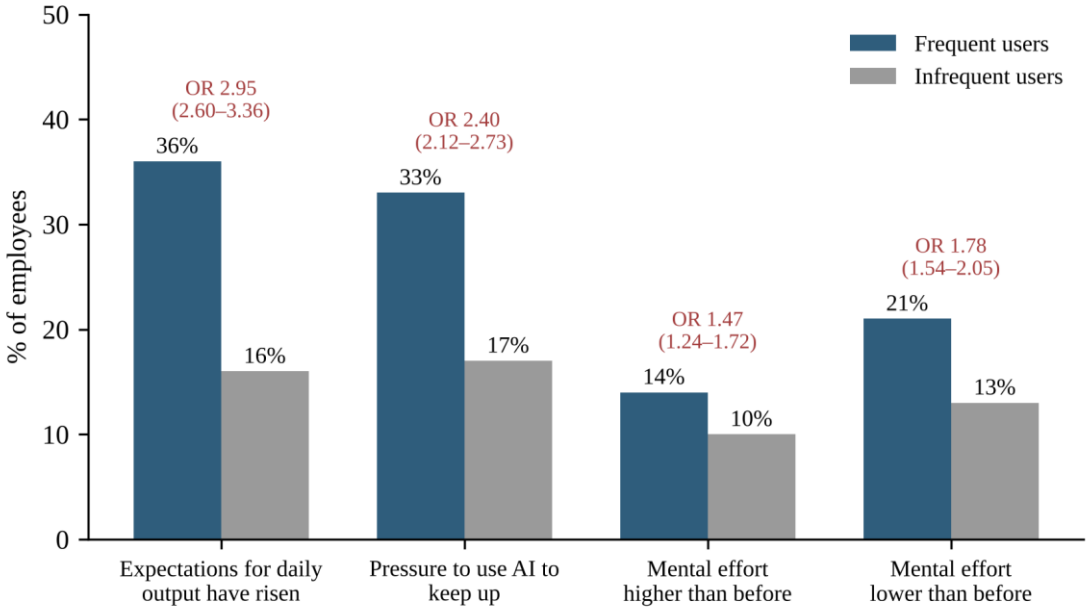

Figure 5. Expectations, pressure and mental effort among frequent and infrequent AI users.
Source: Gallup (2026); odds ratios with 95% CI from effective subgroup sizes.

Other surveys show the same direction with larger numbers, which should be read as a warning about wording. Gallup finds that about three in ten employees in AI-adopting organizations say that AI has raised expectations for daily output. In the Upwork survey, 81% of executives said they had increased demands on workers in the past year, 21% of employees who used AI said they had been asked to do more work because of it, 39% said they spent more time reviewing or moderating AI-generated content, and 23% said they spent more time learning the tools (Upwork Research Institute, 2024). The 77% of AI-using employees who said AI had added to their workload did so on an omnibus item, "in at least one way," which gives any one of several burdens a chance to qualify. Items that name a specific dimension return smaller shares. Both kinds of figure belong in the evidence. They answer different questions: how widespread is some burden, and how widespread is a given burden.

4.5. Hidden Labor and Its Dose–Response (H4)

The Glean Work AI Index allows a look at how hidden labor varies with the intensity of use. Table 6 and Figure 6 show that every measured burden rises steeply from light to heavy use. The share of workers who spend significant time on frequent botsitting rises from 35% of light users (1–19% of work time involving AI) to 65% of moderate users and 74% of heavy users (relative risk of heavy against light 2.11, 95% CI 1.86 to 2.41 under the conservative tier size; trend $z = 12.5$). Weekly cleanup of AI-assisted work rises from 20% to 39% (RR 1.95, 1.59 to 2.40), being unable to explain AI output from 24% to 54% (RR 2.25, 1.89 to 2.68), and blaming AI for one's own mistake from 12% to 41% (RR 3.42, 2.64 to 4.43). All trend tests remain clear at 500 respondents per tier. H4 is supported in the cross-section.

Table 6. Hidden labor by intensity of AI use (Glean Work AI Index).

Panel A. Share of workers by share of work time involving AI (%)

Outcome	Light	Moderate	Heavy	Diff. (points)	RR heavy/light (95% CI)	Trend *z*
Frequent botsitting	35	65	74	39	2.11 (1.86–2.41)	12.5
Weekly cleanup of AI work	20	30	39	19	1.95 (1.59–2.40)	6.6
Botshitting (any form)	50	73	82	32	1.64 (1.49–1.81)	10.9
Cannot explain AI output	24	43	54	30	2.25 (1.89–2.68)	9.7
Blame AI for own mistake	12	29	41	29	3.42 (2.64–4.43)	10.3
Fear AI will eliminate role	33	43	51	18	1.55 (1.33–1.80)	5.8
AI automated meaningful work	35	54	62	27	1.77 (1.54–2.03)	8.5

Panel B. Weekly time ledger, average respondent

Component	Hours per week	% of AI time	Reported exhaustion multiplier
Feeding context to AI	2.3	14	1.2
Supervising AI output	2.2	13	1.1
Debugging AI output	1.7	10	1.4
Other botsitting	0.2	2	1.1
Total botsitting	6.4	37	—
Learning and building (derived)	4.7	27	—
Using AI for the work itself (derived)	6.2	36	—
Total AI time (derived)	17.3	100	—

Note. Panel A: tiers are light (1–19% of work time), moderate (20–49%) and heavy (50% or more); intervals and trend tests assume 500 respondents per tier because tier sizes are not published (Section 3.4). Panel B: components as reported; hour totals derived from the reported 37% share (6.4 / 0.37); the reported component shares do not sum exactly to the 37% total.

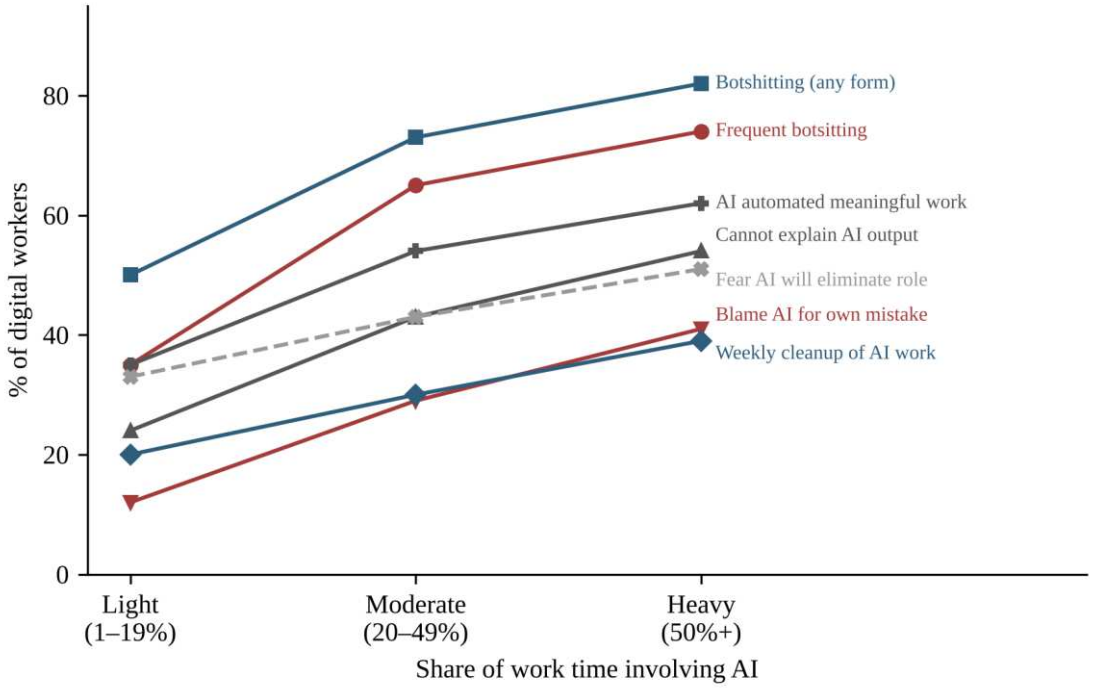

Figure 6. Hidden labor and related outcomes by share of work time involving AI.
Source: Hinds et al. (2026). Tier sizes not published; see Table 6.

The ledger in panel B of Table 6 puts hours on the burden. The average respondent spends 6.4 hours a week botsitting, which is 37% of an estimated 17.3 hours of AI time: 2.3 hours feeding context to the tool, 2.2 supervising outputs, 1.7 debugging, and 0.2 on other tasks. A further 27% of AI time, about 4.7 hours, goes to learning and building, and 36%, about 6.2 hours, to using AI for the work itself. The authors report that 77% of respondents had corrected or redone AI-assisted work in the past month and 30% did so weekly, and they estimate that each type of botsitting is associated with a 1.1 to 1.4 times higher likelihood of feeling worn out by AI. These are vendor-affiliated, self-reported and cross-sectional data, from a sample of digital workers in which 87% already use AI, so the levels should not be generalized. The gradient is the point: the more AI is used, the more of the work consists of managing it.

Workslop adds a burden borne by recipients. In a survey of 1,150 U.S. full-time employees, 40% said they had received AI-generated work that looked polished but lacked substance in the previous month, and they estimated that each instance took about 1 hour and 56 minutes to deal with (Niederhoffer et al., 2025). If recipients meet one to three such instances a month, the expected cost is 0.18 to 0.53 hours per week across the whole sample (0.40 × 1.93 hours × the number of instances ÷ 4.33 weeks). That is small beside botsitting. Its importance is that it is imposed on people who did not choose to use AI.

4.6. Break-Even Analysis (H5)

Under the spliced distribution matched to the national survey, the median user saves 0.94 hours per week. About 48% save at least one hour, 30% more than two hours, 23% more than three and 9% more than 6.4 hours (Table 7, panel C; Figure 7a). The anchors for the overhead ratio are in the Glean ledger. Counting only botsitting, 6.4 hours against 11 claimed hours gives $\rho = 0.58$. Adding the 4.7 hours of learning and building gives $\rho = 1.01$. The randomized trial of developers implies $\rho > 1$ for that setting, because the net effect was a slowdown.

Table 7. Break-even analysis of freed time (hours per week).

Panel A. Mean net free time, $F = G(1 - \rho)(1 - \alpha)$, hours per week

Reabsorption α	ρ = 0	ρ = .25	ρ = .5	ρ = .58	ρ = .75	ρ = 1.0	ρ = 1.25
0	2.16	1.62	1.08	0.91	0.54	0.00	−0.54
0.25	1.62	1.21	0.81	0.68	0.40	0.00	−0.40
0.5	1.08	0.81	0.54	0.45	0.27	0.00	−0.27
0.75	0.54	0.40	0.27	0.23	0.13	0.00	−0.13

Panel B. Percentage of users keeping at least one free hour per week

Reabsorption α	ρ = 0	ρ = .25	ρ = .5	ρ = .58	ρ = .75	ρ = 1.0	ρ = 1.25
0	48	40	30	26	20	0	0
0.25	40	32	24	22	13	0	0
0.5	30	24	20	16	5	0	0
0.75	20	13	5	3	0	0	0

Panel C. Gross savings distribution: % of users saving more than *h* hours per week

h (hours)	Mean excess 2 h	Mean excess 3 h (main)	Mean excess 4 h
1	57	48	38
2	36	30	24
3	26	23	21
4	20	21	20
6.4	6	9	11

Note. G has mean 2.16 hours (5.4% of 40) and 20.5% of users above 4 hours (spliced distribution, Section 3.5). Anchors for ρ: 0.58 (6.4 botsitting hours ÷ 11 claimed hours saved), 1.01 (adding 4.7 hours of learning and building), and above 1 in the developer trial. Scenario analysis; ρ and α are not estimated from one dataset.

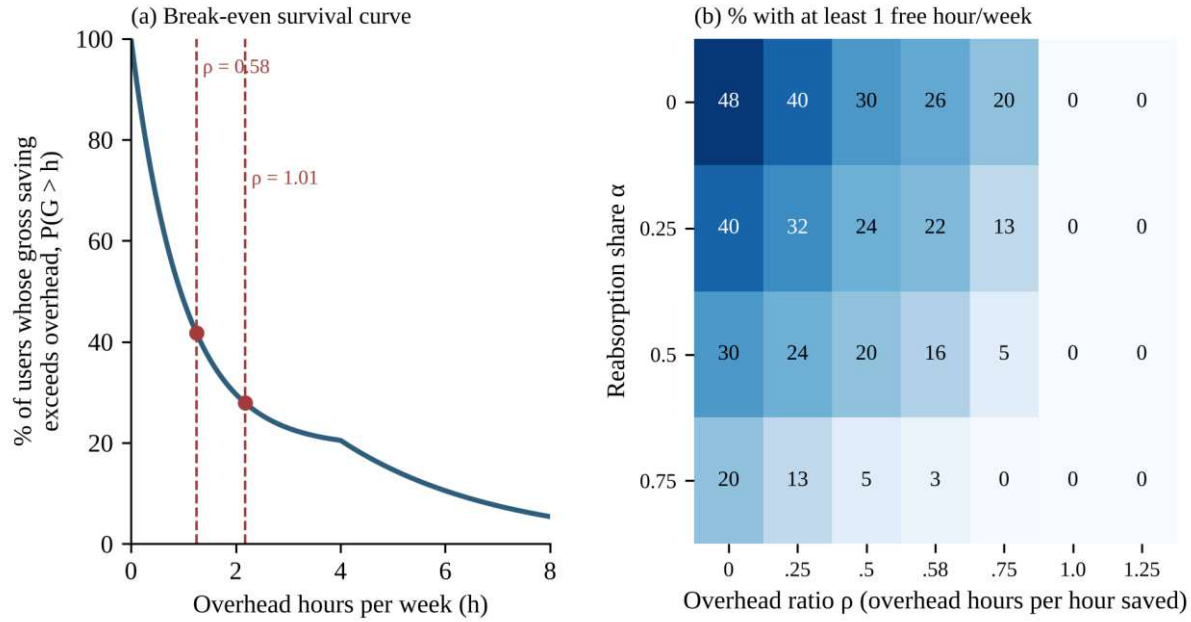

Figure 7. Break-even analysis: (a) share of users whose gross saving exceeds a given overhead; (b) share keeping at least one free hour a week, by overhead ratio ρ and reabsorption share α.

Spliced savings distribution calibrated to a mean of 2.16 hours a week and 20.5% of users above 4 hours (Section 3.5). Scenario analysis.

At $\rho = 0.58$ and no reabsorption ($\alpha = 0$), mean free time is 0.91 hours per week, and 26% of users keep at least an hour (22% to 31% across the alternative tail assumptions). If half of what remains is reabsorbed by wider scope ($\alpha = 0.5$), mean free time falls to 0.45 hours and the share keeping an hour falls to 16% (14% to 17%). At $\rho = 1$ or above, nobody keeps any time. The sign of the dividend is set by ρ and the size of the remainder by α, whereas gross savings enter only proportionally. Raising ρ from 0.58 to 0.68 reduces mean free time by 24% (from 0.91 to 0.69 hours at $\alpha = 0$), while a 10% increase in gross savings raises it by 10%. H5 is supported by the accounting. The verification and learning that accompany each hour saved matter more to the outcome than how many hours are saved.

The analysis is an accounting identity applied to calibrated inputs. It does not estimate ρ or α from a single dataset, and it treats the overhead as proportional to savings. What it contributes is a transparent map of the conditions under which a worker is better or worse off in time, and a reminder that the claim "AI saves time" and the claim "AI creates work" are not mutually exclusive: both can be true of the same person in the same week, and the net can be of either sign.

4.7. Well-Being and Organizational Resources (H6)

Table 8 collects the evidence on well-being. None of it is longitudinal, and almost none comes from probability samples. In the Glean survey, the share who say they feel worn out by AI is 33% in the United States, 30% in the United Kingdom and 43% in Australia. Workers who say they lack the context and data access needed to use AI well report feeling worn out in 50% of cases against 18% of those with the context, a relative risk of 2.78 (95% CI 2.26 to 3.42, assuming 500 respondents in each group). H6 is supported in this cross-section: resources are strongly associated with lower exhaustion. The association could reflect selection, since the workers with the best context may be those with the best jobs, and it cannot show that adding context would reduce exhaustion.

Table 8. Evidence on well-being and organizational resources.

Source	Measure	Finding	Design caveat
Hinds et al. (2026)	Feeling worn out by AI	U.S. 33%, U.K. 30%, Australia 43%	Vendor survey; self-report
Hinds et al. (2026)	Worn out, by access to context and data	50% with poor context vs 18% with rich context; RR 2.78 (2.26–3.42 at 500 per group)	Cross-sectional; selection possible
Upwork (2024), as summarized in press	Burnout; struggling with demands	71% of full-time employees reported burnout; 65% struggling with increased employer demands	Vendor survey; press summary
Gallup (2026)	Mental effort since using AI	Frequent users: 21% lower, 14% higher; infrequent: 13% lower, 10% higher	Self-report; bivariate
Tang et al. (2023)	Loneliness, insomnia, after-work drinking	Higher with more frequent interaction with AI systems in four studies	Mixed survey and experimental; not workload-specific
Ranganathan & Ye (2026)	Intensification, boundaries, multitasking	Task expansion, work in breaks and evenings, parallel AI threads	One firm; qualitative
EU-OSHA (2025)	Effects of digitalization on work	48% pace; 28% workload; 19% skills; 16% decision latitude	Digitalization in general, not AI

Note. All entries are associations or descriptions; none estimates a causal effect of AI use on well-being.

The other well-being evidence is consistent in direction and varied in quality. Press summaries of the Upwork survey report that 71% of full-time employees were burned out and that 65% were struggling with increased employer demands. In the Gallup data, frequent users were more likely to report both lower and higher mental effort (Section 4.4). Tang et al. (2023) found in four studies that frequent interaction with AI systems was associated with loneliness, insomnia and after-work alcohol use. The Berkeley ethnography records workers describing a sense of always having something to do and of work spilling into pauses

(Ranganathan & Ye, 2026). Against the background of the meta-analytic link between workload and emotional exhaustion (Alarcon, 2011), none of this is surprising, but none of it amounts to a causal estimate of what AI-driven expansion does to well-being.

5. Discussion

The evidence supports a more careful claim than either the enthusiasts or the skeptics make. AI does save time on the tasks it substitutes, and for the frequent users in the Gallup panel the saving on information gathering and writing is the dominant change in the working week. What the evidence does not support is the assumption that this saving arrives as free time. Review and correction do not shrink on average, learning grows, a fifth of frequent users spend more time checking, hidden labor climbs steeply with the intensity of use, and expectations and pressure are about three and two and a half times as likely, in odds terms, to have risen among frequent users. The break-even analysis shows how little it takes for the arithmetic to turn: with overhead at 58 cents per hour saved and half the remainder reabsorbed, one user in six keeps an hour a week.

The sources disagree about how common the problem is, and the disagreement is informative. Omnibus items that allow any one of several burdens to qualify give figures such as 77%. Items that name a single dimension give figures between 14% and 36%. Vendor samples that are heavy in AI use show gradients from 35% to 74% on a single burden. A fair summary is that task expansion is real, is concentrated among frequent and heavy users, and that the share of the workforce affected depends on how it is asked and who is asked. The Gallup data also show that a substantial minority of frequent users report lighter mental effort, so the technology is producing winners as well as strained workers. Bick et al. (2025) find the same unevenness in time savings, with about a fifth of users saving four hours a week or more while many save little.

It helps to separate two dividends that are usually run together. The first is an output dividend: more or better work for the same hours. The second is a time dividend: the same work in fewer hours, or hours returned to recovery. An organization may reasonably want the first. The trouble arises when the output dividend is claimed in the form of higher expectations without being budgeted for in the demands that come with it, and when the time dividend is assumed to exist because the first is visible. The job demands–resources model predicts what follows: demands rise, the resources to meet them (context, time, autonomy, support) do not, and exhaustion follows (Bakker & Demerouti, 2017). Hobfoll's (1989) account adds the spiral. A worker who is tired makes more errors, which creates more repair, which erodes the resources further, and the reciprocal effects found in longitudinal work (Guthier et al., 2020) are consistent with that.

Classification into challenge and hindrance demands (Crawford et al., 2010; LePine et al., 2005) suggests where the risk is concentrated. Checking and redoing machine output, chasing context and switching between tools look like hindrance demands. They take effort and give little back in growth or recognition. The wider scope that the Berkeley study describes could be a challenge demand if it came with autonomy and acknowledgment. The same study suggests it often does not, because it creeps in through small additions that nobody decides on. This is perhaps the most useful practical insight: the expansion is hard to manage because it is rarely authorized. Nobody gives the instruction to work through lunch, rewrite a colleague's output, or take on a neighboring task. Each step is a reasonable local choice made easy by the tool.

The measurement problems deserve as much attention as the findings. Self-reports of time saved overstate what independent measures find, as the developer trial and the comparison between the vendor survey and the national survey show. Experiments measure the tasks that suit the tool. Surveys count what respondents notice, and the burden most likely to go unnoticed is the one that has become routine. If workers cannot see their own overhead, managers certainly cannot, and organizations that rely on adoption dashboards and usage counts will not see it either.

6. Implications

6.1. For Organizations

Six practices follow. First, measure freed time and not tool use. A usage count says nothing about whether anyone has gained an hour. Even a simple quarterly question on net time saved and net time added, asked in the same form to the same people, would be an improvement on adoption dashboards. Second, budget for verification. If an organization expects AI output to be checked, the checking is work and belongs in the plan and the capacity. Third, define where not to use AI. The experimental literature shows large losses outside the frontier of what the tool can do (Dell'Acqua et al., 2023), and a rule of thumb that stops people using it for tasks where checking costs more than doing is worth more than another training session. Fourth, decide explicitly what happens to saved time. If the intent is to raise output, say so and adjust targets and staffing; if the intent is to give time back, protect it. Leaving it to drift is how it gets reabsorbed. Fifth, supply the resources the evidence links to lower exhaustion: context and data access, clear quality standards, and manager support for saying no. Sixth, watch the edges of the day. Where prompts and replies move into breaks and evenings, recovery is being spent, and a norm that AI work stays inside working hours costs nothing.

6.2. For Policy and Occupational Health

Psychosocial risk assessments should include the effects of AI tools on workload, pace and autonomy. European survey data already show how widely digitalization is felt. In the EU-OSHA OSH Pulse 2025, 48% of workers said digital technologies determine the pace of their work, 28% that they increase workload, 19% that they reduce opportunities to use their knowledge and skills, and 16% that they reduce their ability to decide how to do their work (EU-OSHA, 2025). Those figures cover digitalization in general and do not isolate generative AI. They indicate the kind of conditions under which expansion would turn into strain. Labor regulators and statistical agencies could usefully add items on AI-related tasks and time use to existing labor force and working conditions surveys, so that the question can be answered from probability samples and not from vendor reports.

6.3. For Research

Four priorities stand out. Longitudinal panels and diary studies that follow the same workers before and after adoption would separate expansion from selection. Administrative or telemetry time-use data, with consent, would measure overhead directly and replace the assumed ρ in the break-even analysis. Field experiments that vary what organizations do with freed time, for example protecting it, redeploying it or raising targets, would estimate α. And studies of well-being should measure the intermediate quantities, such as hours, recovery and perceived control, and not only a final burnout score.

7. Limitations

The limits are substantial. First, most of the evidence is self-reported and cross-sectional, so associations between use intensity, hidden labor, expectations and exhaustion may reflect selection, reverse causation or common-method bias. Second, the Gallup subgroup sizes had to be approximated from published percentages, and the Glean tier sizes were assumed to be 500 each, so the intervals indicate precision and not exact coverage; the group comparisons are not adjusted for occupation or seniority. Third, several sources are vendor-affiliated and some figures could be checked only through summaries of the original report, which Table 1 flags. Fourth, the evidence map compares studies that differ in task, population and metric, and the conversion of the throughput gain in Brynjolfsson et al. (2025) to time saved is approximate; the developer trial concerns tools available in early 2025. Fifth, the break-even analysis is stylized. It treats overhead as proportional to savings, takes its savings distribution from published moments that do not identify the shape, and anchors ρ on a self-reported sample. Sixth, the evidence is overwhelmingly from the

United States, Western Europe and Australia, and from knowledge-work settings. Finally, the technology is changing quickly, and tools that handle verification and context better may shift ρ in either direction.

8. Conclusion

This paper asked whether AI's time savings are cancelled by new work and what that does to the people involved. The answer from the best available public evidence is that savings are real but small once measured across a working week (2.8% to 5.4% in population surveys, against 12% to 40% in the favorable experiments), that they are concentrated in the tasks AI substitutes, and that review, correction and learning do not shrink in step. Frequent users report higher expectations (OR 2.95) and pressure (OR 2.40), hidden labor rises steeply with the intensity of use, and a break-even analysis shows that modest overhead and reabsorption leave most users with less than an hour a week of free time. Workers with more context and data access report much less exhaustion, though the data cannot show that supplying it would help.

The word "expansion" was chosen carefully. It names a drift and not a decision, and it is the lack of a decision that makes it hard to see and hard to stop. Organizations that count freed time as carefully as they count licenses, that put verification in the plan, and that decide in advance what the saved time is for, have a good chance of keeping the benefit the technology promises. Those that do not are likely to find that their most enthusiastic users are also their most stretched.

Author Contributions

[To be completed by the author. For a single-author paper: Conceptualization, methodology, formal analysis, writing—original draft preparation, and writing—review and editing were carried out by the author, who has read and agreed to the published version of the manuscript.]

Declaration of Generative AI and AI-Assisted Technologies in Manuscript Preparation

During preparation of this manuscript an AI assistant (Claude, Anthropic) was used to search for and check published figures against their sources, to run the secondary statistical analyses and to edit the language of the text. The author reviewed all output and takes full responsibility for the content. [Author to confirm or amend this statement to match the journal's policy and actual use.]

Data Availability

All data are from public sources listed in the reference list and in Table 1. The derived statistics can be reproduced from the published figures in Tables 2 to 8 and the equations in Section 3. No new primary data were collected.

Funding

This study received no external funding.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

Alarcon, G. M. (2011). A Meta-Analysis of Burnout with Job Demands, Resources, and Attitudes. *Journal of Vocational Behavior, 79*, 549-562. https://doi.org/10.1016/j.jvb.2011.03.007

Armitage, P. (1955). Tests for Linear Trends in Proportions and Frequencies. *Biometrics, 11*, 375-386. https://doi.org/10.2307/3001775

Bainbridge, L. (1983). Ironies of Automation. *Automatica, 19*, 775-779. https://doi.org/10.1016/0005-1098(83)90046-8

Bakker, A. B., & Demerouti, E. (2017). Job Demands–Resources Theory: Taking Stock and Looking Forward. *Journal of Occupational Health Psychology, 22*, 273-285. https://doi.org/10.1037/ocp0000056

Becker, J., Rush, N., Barnes, E., & Rein, D. (2025). *Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity* (arXiv:2507.09089). METR. https://arxiv.org/abs/2507.09089

Bick, A., Blandin, A., & Deming, D. J. (2025). *The Rapid Adoption of Generative AI* (Working Paper No. 2024-027). Federal Reserve Bank of St. Louis. https://doi.org/10.20955/wp.2024.027

Brynjolfsson, E., Li, D., & Raymond, L. (2025). Generative AI at Work. *Quarterly Journal of Economics, 140*, 889-938. https://doi.org/10.1093/qje/qjae044

Brynjolfsson, E., Rock, D., & Syverson, C. (2021). The Productivity J-Curve: How Intangibles Complement General Purpose Technologies. *American Economic Journal: Macroeconomics, 13*, 333-372. https://doi.org/10.1257/mac.20180386

Cochran, W. G. (1954). Some Methods for Strengthening the Common χ^2 Tests. *Biometrics, 10*, 417-451. https://doi.org/10.2307/3001616

Crawford, E. R., LePine, J. A., & Rich, B. L. (2010). Linking Job Demands and Resources to Employee Engagement and Burnout: A Theoretical Extension and Meta-Analytic Test. *Journal of Applied Psychology, 95*, 834-848. https://doi.org/10.1037/a0019364

Dell'Acqua, F., McFowland, E., III, Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, K. R. (2023). *Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality* (Harvard Business School Working Paper No. 24-013). https://ssrn.com/abstract=4573321

EU-OSHA (2025). *OSH Pulse 2025: Digitalisation* [Infographic]. European Agency for Safety and Health at Work. https://healthy-workplaces.osha.europa.eu/sites/hwc/files/hwc/infographic/OSH-pulse-2025-digitalisation_infographic_EN.pdf

Gallup (2026, November 16). *How AI Is Changing the Way Employees Spend Time at Work* [Gallup Workplace]. https://www.gallup.com/workplace/716585/ai-changing-employees-spend-time-work.aspx

Green, F. (2004). Work Intensification, Discretion, and the Decline in Well-Being at Work. *Eastern Economic Journal, 30*, 55-71.

Guthier, C., Dormann, C., & Voelkle, M. C. (2020). Reciprocal Effects between Job Stressors and Burnout: A Continuous Time Meta-Analysis of Longitudinal Studies. *Psychological Bulletin, 146*, 1146-1173. https://doi.org/10.1037/bul0000304

Hinds, R., Hoffman, M., Baladi, S., Cao, H., Lee, Y. S., Leonardi, P., Ranganathan, A., Rhymer, J., Rogelberg, S., Sutton, B., & Zhu, Y. (2026). *Work AI Index 2026*. Glean Work AI Institute. https://www.glean.com/work-ai-institute/reports/work-ai-index

Hobfoll, S. E. (1989). Conservation of Resources: A New Attempt at Conceptualizing Stress. *American Psychologist, 44*, 513-524. https://doi.org/10.1037/0003-066X.44.3.513

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics, 6*, 65-70.

Humlum, A., & Vestergaard, E. (2025). *Large Language Models, Small Labor Market Effects* (SSRN Working Paper No. 5219933). https://ssrn.com/abstract=5219933

Jevons, W. S. (1865). *The Coal Question: An Inquiry Concerning the Progress of the Nation, and the Probable Exhaustion of Our Coal-Mines*. Macmillan.

LePine, J. A., Podsakoff, N. P., & LePine, M. A. (2005). A Meta-Analytic Test of the Challenge Stressor–Hindrance Stressor Framework: An Explanation for Inconsistent Relationships among Stressors and Performance. *Academy of Management Journal, 48*, 764-775. https://doi.org/10.5465/amj.2005.18803921

Niederhoffer, K., Rosen Kellerman, G., Lee, A., Liebscher, A., Rapuano, K., & Hancock, J. T. (2025, September 22). AI-Generated "Workslop" Is Destroying Productivity. *Harvard Business Review*.

Noy, S., & Zhang, W. (2023). Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence. *Science, 381*, 187-192. https://doi.org/10.1126/science.adh2586

Parasuraman, R., & Manzey, D. H. (2010). Complacency and Bias in Human Use of Automation: An Attentional Integration. *Human Factors, 52*, 381-410. https://doi.org/10.1177/0018720810376055

Parkinson, C. N. (1955, November 19). Parkinson's Law. *The Economist*.

Ranganathan, A., & Ye, X. M. (2026, February). AI Doesn't Reduce Work—It Intensifies It. *Harvard Business Review*.

Rubinstein, R. Y., & Kroese, D. P. (2017). *Simulation and the Monte Carlo Method* (3rd ed.). Wiley.

Tang, P. M., Koopman, J., Mai, K. M., De Cremer, D., Zhang, J. H., Reynders, P., Ng, C. T. S., & Chen, I.-H. (2023). No Person Is an Island: Unpacking the Work and After-Work Consequences of Interacting with Artificial Intelligence. *Journal of Applied Psychology, 108*, 1766-1789. https://doi.org/10.1037/apl0001103

Tarafdar, M., Tu, Q., Ragu-Nathan, B. S., & Ragu-Nathan, T. S. (2007). The Impact of Technostress on Role Stress and Productivity. *Journal of Management Information Systems, 24*, 301-328. https://doi.org/10.2753/MIS0742-1222240109

Upwork Research Institute (2024). *From Burnout to Balance: AI-Enhanced Work Models for the Future of Work*. Upwork. https://www.upwork.com/research/ai-enhanced-work-models

Woolf, B. (1955). On Estimating the Relation between Blood Group and Disease. *Annals of Human Genetics, 19*, 251-253. https://doi.org/10.1111/j.1469-1809.1955.tb01348.x