AI Adoption and Productivity in Healthcare*

Erkmen G. Aslim* Manuel Hoffmann‖ Haonan Yin‖
*University of Vermont
‖University of California, Irvine

This Extended Abstract Version: *February 22, 2026*

> "We are giving agency back to clinicians, creating clarity for patients, and establishing radical efficiencies to transform healthcare from the conversation up."
>
> — Shiv Rao, MD
>
> CEO and Founder, Abridge

Artificial Intelligence (AI) is predicted to substantially transform the economy, work processes, and productivity (Hoffmann et al., 2024; Brynjolfsson, Li, and Raymond, 2025; Dillon et al., 2025; Bick, Blandin, and Deming, 2025; Hoffmann et al., 2025). In healthcare, one particularly promising application of generative AI is the reduction of documentation burden. Ambient documentation systems (i.e., "AI scribes") aim to alleviate this burden by automatically generating structured clinical notes from audio recordings of patient–physician encounters (Afshar et al., 2025). These systems use speech recognition and large language models to transcribe conversations, extract clinically relevant information, and format it into standardized notes compatible with electronic health records (EHRs). By automating documentation tasks, AI scribes are expected to reduce administrative workload, mitigate physician burnout, allow clinicians to focus more fully on patients during visits, enhance patient–physician interaction, and support more targeted treatment decisions.

If AI scribes meaningfully reduce documentation burden, they should alter physicians' time allocation, workflow, and potentially their well-being. We study the implementation of a generative

*This study was reviewed and approved by the Institutional Review Board at the University of Vermont (IRB No. 00000485) and the University of California, Irvine (IRB No. 00000596).

AI clinical documentation tool, characterize physician adoption and usage intensity, and estimate its effects on documentation efficiency, measured by editing effort, and physician productivity. The tool we examine is Abridge, an ambient AI documentation system deployed within a large, multi-hospital academic health network integrated into its electronic health record platform. The network includes six operating hospitals, employs over 15,000 staff, and serves more than one million patients across two states, providing a sizable and geographically diverse clinical setting.

We leverage novel, high-frequency (daily) physician-level proprietary data on AI scribe adoption, usage intensity, and output quality, linked to detailed productivity metrics at the physician–encounter level. A central advantage of our setting is the ability to construct novel physician-level linkages between AI scribe activity and electronic health record (EHR) data, allowing us to measure objective productivity outcomes such as time spent in the system, note completion timing, appointment closure rates, and documentation turnaround time. This unusually granular integration of AI usage data with operational clinical metrics enables us to observe how generative AI adoption translates into measurable changes in workflow and performance.[1]

The staggered rollout of AI scribes across the health network generated plausibly exogenous variation in physician eligibility over time, enabling a difference-in-differences (DID) framework to estimate causal effects.[2] We observe the full roster of physicians practicing within the network through administrative EHR data, allowing us to construct a panel at the physician level and follow both adopters and non-adopters over time. This comprehensive coverage enables us to include physician fixed effects and estimate within-physician changes in productivity associated with eligibility and subsequent tool use.

Our primary analysis focuses on physicians in family medicine and general internal medicine, clinical settings in which documentation demands are substantial and ambient AI tools are particularly well suited to routine outpatient encounters. While the broader rollout spans additional

[1]For a subset of physicians, we additionally observe repeated institutional survey responses, which provide complementary measures of well-being (e.g., self-reported burnout) and time allocation across clinical and administrative tasks. These data allow us to explore mechanisms underlying the observed productivity effects.

[2]Importantly, physicians were not informed of the precise timing of eligibility in advance, reducing the likelihood of anticipatory behavioral responses prior to treatment.

specialties, we concentrate on primary care physicians to provide a focused evaluation in a setting where adoption incentives are strongest.

— Figure 1 about here —

Figure 1 presents event-time evidence on adoption and usage following eligibility within th etime frame from May 2024 to Nov 2025. The figure is based on a balanced panel of 293 physicians and 5,567 patient–physician encounters observed before and after eligibility. Adoption is immediate and widespread: 56 percent of physicians create their first AI-generated note within the first month of eligibility, and adoption rises steadily to near universal use within fourteen months.

Usage intensity increases in parallel. The number of AI-generated notes per physician rises quickly from approximately 20 in the first month to nearly 80 notes per month after one year. Audio recording time exhibits a similar pattern, increasing sharply upon eligibility and exceeding 1,700 minutes per physician per month by the end of the first year. Taken together, these patterns imply a remarkably stable documentation structure of roughly 20 minutes of audio per note, suggesting systematic rather than sporadic use.

Importantly, physicians accept approximately 80 to 90 percent of AI-generated text without modification throughout the post-eligibility period. This high and stable acceptance rate provides a first-order indication of substantial documentation efficiency gains and suggests that the tool is meaningfully integrated into routine clinical workflow rather than used experimentally or intermittently.[3] Such reductions in documentation effort create scope for time reallocation, potentially increasing physician productivity.

The rapid adoption of AI scribes raises a natural question: does experience with the tool further reshape clinical workflow? To examine dynamic effects, we estimate models with physician and year-month fixed effects that exploit variation in exposure duration following eligibility. We define physician experience relative to the first month post-eligibility and group physicians into 2–5 months, 6–12 months, and more than 12 months of AI exposure.

[3]These patterns are confirmed in staggered difference-in-differences models with physician and year-month fixed effects, where the static average treatment effect on the treated (ATT) is positive, large in magnitude, and precisely estimated.

— Table 1 about here —

Table 1 provides a comprehensive overview of how physician interaction with AI scribes evolves across six distinct dimensions of practice. By analyzing these metrics from left to right, we observe a transition from initial adoption surges to long term professional optimization.

The analysis begins with usage and technical engagement (columns 1–2). Upon becoming eligible, physicians demonstrate an immediate and sustained increase in documentation volume, with those in the 2-12 month cohorts generating approximately 30 additional notes compared to the baseline (column 1). Interestingly, column (2) reveals a significant initial increase in average audio duration of 2.44 minutes during the first five months. This likely represents an initial calibration phase where physicians provide more exhaustive input as they adjust to the system. However, as clinicians reach the one–year mark, they significantly reduce their average audio recording duration by 4.73 minutes per note. This suggests that "power users" develop more concise communication patterns, likely learning how to structure their patient interactions to be more efficiently processed by the AI system.

Following this technical adjustment, we observe the integration and adoption of AI content. Column (3) indicates that during the first year, physicians are 11.8% to 14.5% more likely to "keep" or utilize an AI–generated note relative to the baseline. While this specific propensity settles after one year, the sustained significance in column (1) suggests the tool remains a permanent fixture in the clinical workflow, even as the method of engagement shifts from more passive to more active engagement with AI scribes.

Finally, columns (4) through (6) illustrate the evaluation and curation phase, where the most complex behavioral adjustments occur. The AI acceptance rate (column 4) - the percentage of raw text accepted without editing - declines significantly as experience grows, reaching a reduction of 0.654 for the most senior users. However, this does not indicate a failure of the technology. When viewed alongside column (5), we see that while the propensity to provide a binary "note rating" decreases over time, the "average note rating" in column (6) shows that the initial significant dissatisfaction of -0.641 in the first few months disappears as users cross the one–year mark.

There are multiple plausible explanations for these observed dynamics. An efficiency story suggests that the tool effectively reduces the marginal cost of documentation, allowing for higher patient volume. A learning story is supported by the audio duration and acceptance data, indicating that physicians undergo a cognitive adjustment period before mastering the tool as a collaborative partner. Finally, a compositional change may be at play, where the types of encounters or the profile of physicians utilizing the tool shift as it moves from early adopters to the broader clinical population. Taken together, this progression suggests that the AI scribe journey begins with a focus on quantity and ends with sophisticated curation, where experienced physicians utilize efficiency gains to actively fine–tune notes to their professional standards.

This paper contributes to the rapidly-evolving literature on the impacts of AI on productivity and the changing nature of professional work. While previous research has focused on AI tools that reduce the cost of core tasks— such as software development (Hoffmann et al., 2024), we document a reverse but complementary effect. In our setting, the AI scribe targets "peripheral" or bureaucratic tasks to enable a focus on core clinical work. This finding serves as a vital complement to the idea that structural frictions within EHR systems act as a significant drag on productivity (Dix, Moran, and Nguyen, 2025). We demonstrate that by improving the inflow of information, AI scribes effectively reduce these administrative frictions, which is as significant to productivity as enhancing core technical outputs.

Second, we provide empirical evidence that there are remarkably few barriers to AI adoption when the tool is specifically designed to alleviate well–documented professional burdens. The rapid pace of adoption and the intensity of usage observed among the physician cohort in our sample serve as a powerful signal of the tool's value through revealed preferences. This suggests that the "implementation gap" often seen with new medical technologies can be bridged when the technology aligns directly with the clinician's desire to return to patient–centered care.

Third, our study offers a unique longitudinal view of the "AI–practitioner" relationship. By utilizing proprietary data on word acceptance and audio duration, we move beyond simple binary adoption metrics. We demonstrate that the value of these documentation tools does not just come

from a temporary growth in encounters, but from a sustained increase in documentation volume and a long–term improvement in dictation efficiency as physicians master the tool.

References

Afshar, Majid, Mary Ryan Baumann, Felice Resnik, Josie Hintzke, Anne Gravel Sullivan, Graham Wills, Kayla Lemmon, Jason Dambach, Leigh Ann Mrotek, Mariah Quinn et al. 2025. "A pragmatic randomized controlled trial of ambient artificial intelligence to improve health practitioner well-being." *NEJM AI* 2 (12):AIoa2500945.

Bick, Alexander, Adam Blandin, and David Deming. 2025. "The impact of generative ai on work productivity." Tech. rep., Federal Reserve Bank of St. Louis.

Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond. 2025. "Generative AI at Work." *The Quarterly Journal of Economics* 140 (2):889–942.

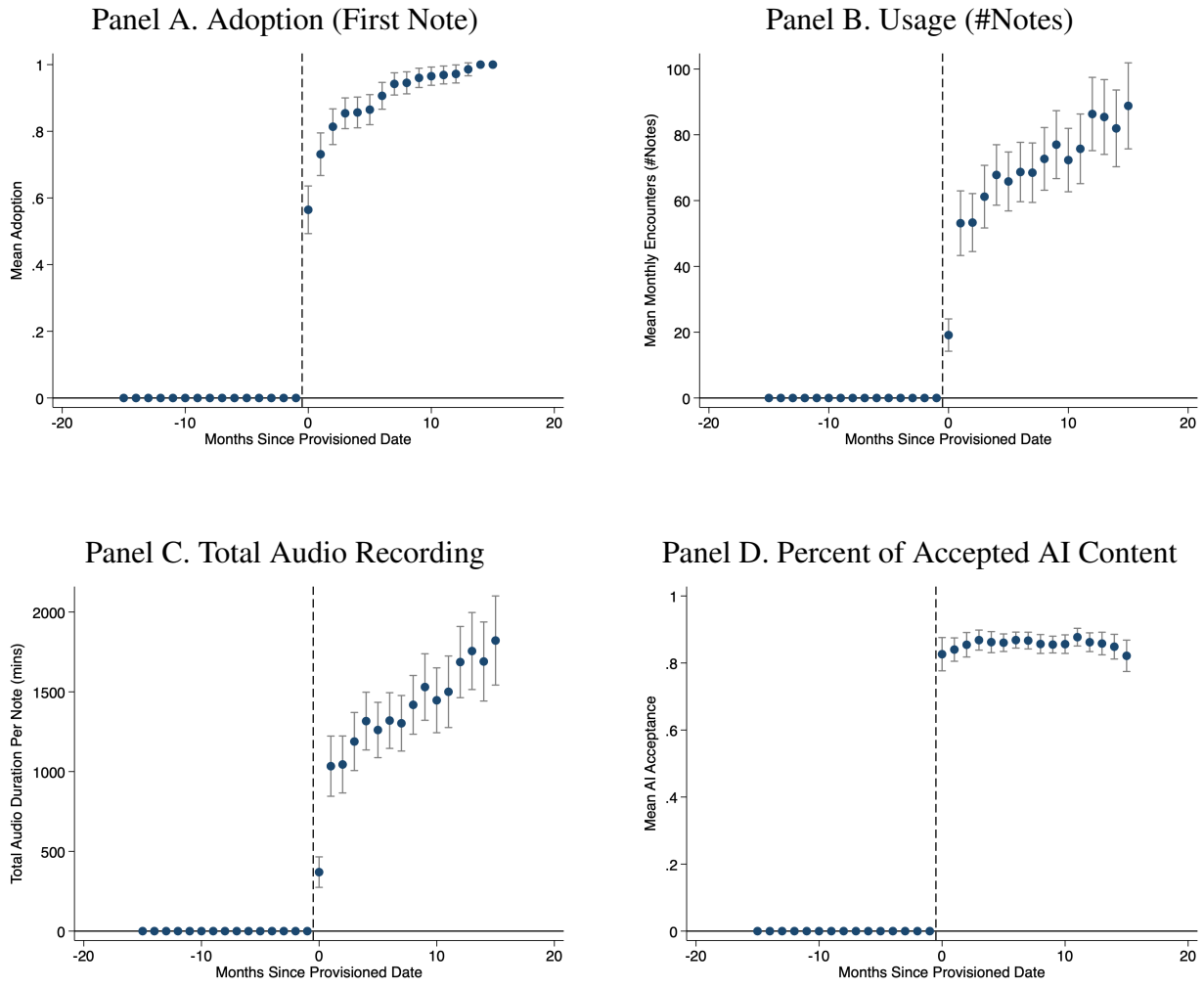
Dillon, Eleanor W, Sonia Jaffe, Nicole Immorlica, and Christopher T Stanton. 2025. "Shifting work patterns with generative ai." Tech. rep., National Bureau of Economic Research.

Dix, Rebekah, Kelsey Moran, and Thi Mai Anh Nguyen. 2025. "Costs of Technological Frictions: Evidence from EHR (Non-) Interoperability." Tech. rep., Working Paper.

Hoffmann, Manuel, Sam Boysel, Frank Nagle, Sida Peng, and Kevin Xu. 2024. "Generative AI and the Nature of Work." *Revise & Resubmit at Management Science* .

Hoffmann, Manuel, Sam Boysel, Frank Nagle, and Kevin Xu. 2025. "The Generative AI - Gender Puzzle." *Working Paper* .

Figure 1: Staggered Rollout: AI Scribes Adoption and Usage



Notes: Panel A illustrates initial adoption of AI scribes, defined as the first AI-generated note following physician eligibility. Panel B reports the number of clinical notes generated using the AI scribe. Panel C displays total audio recording duration per physician. Panel D presents the percentage of AI-generated text accepted without modification, measured as the share of AI-produced content incorporated directly into the final clinical note. This metric serves as a proxy for documentation efficiency. The figures are based on a time period of May 2024 to Nov 2025. 95 percent confidence intervals are shown, with standard errors clustered at the physician level.

Table 1: Heterogeneous Effect of Experience on Usage

	(1) Usage (#Notes)	(2) Avg Audio Duration	(3) Keep AI Notes	(4) AI Acceptance Rate	(5) Note Rating	(6) Avg Note Rating
Exp. 2–5 Months	28.63***	2.436***	0.118***	-0.115***	-0.0583**	-0.641***
	(3.381)	(0.770)	(0.0351)	(0.0265)	(0.0294)	(0.209)
Exp. 6–12 Months	30.62***	1.153	0.145***	-0.285***	-0.133***	-0.403
	(3.142)	(0.875)	(0.0337)	(0.0301)	(0.0300)	(0.344)
Exp. over 12 Months	13.15***	-4.733***	-0.0565*	-0.654***	-0.210***	-0.106
	(2.282)	(0.806)	(0.0300)	(0.0269)	(0.0298)	(0.499)
Constant	16.83***	10.80***	0.221***	0.845***	0.254***	3.770***
	(2.226)	(0.740)	(0.0285)	(0.0258)	(0.0276)	(0.277)
Observations	5,567	5,567	5,567	3,700	5,567	457
R-squared	0.713	0.700	0.555	0.904	0.357	0.596

Notes: This table contains usage intensity, audio efficiency, and note quality metrics. All specifications include physician and year–month fixed effects. The table is based on a time period of May 2024 to Nov 2025. Standard errors clustered at the physician level are reported in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$