Better Technology, Worse Motivation: GenAI's Mediocrity Trap*

Yvonne Jie Chen† Jie Gong‡ Jin Li‡ Zibo Zhao‡

July 20, 2025

Abstract

While generative AI (GenAI) promises productive efficiency, it can paradoxically lead to lower-quality work. We conducted an experiment with professional illustrators and found that AI assistance flattens the quality curve—it accelerates initial gains but sharply diminishes the returns on sustained effort. Faced with this, a significant number of professionals made a strategic choice: they sacrificed the final quality to save time. Our finding highlights a critical challenge for GenAI, which can weaken the motivation required for creative excellence and innovation.

---

*We are grateful to Jen Brown, Jiahua Che, Wouter Dessein, Florian Englmaier, Daniel Ferreira, Guido Friebel, Alfonso Gambardella, Luis Garicano, Juan Pantano, Uta Schonberg, Marshall Van Alstyne, Yanhui Wu, Jubo Yan, Songfa Zhong, participants at ACOE, OESS and various seminars for helpful comments and suggestions. We thank Peixuan Huang, Yele Ma, Zhiwei Shang, Xinjue Yao for providing excellent research assistance. The content is solely the responsibility of the authors.

†School of Entrepreneurship and Management, ShanghaiTech University, China.

‡Faculty of Business and Economics, The University of Hong Kong.

1 Introduction

Generative AI (GenAI) is a transformative technology that impacts productivity across numerous sectors. Many recent studies consistently highlight substantial productivity gains attributable to the adoption of GenAI tools (Acemoglu et al., 2022b; Noy and Zhang, 2023; Peng et al., 2023; Jia et al., 2024; Brynjolfsson et al., 2025; Dell'Acqua et al., 2023; Kanazawa et al., 2022; Bono and Xu, 2024; Boussioux et al., 2024; Haslberger et al., 2024; Kreitmeir and Raschky, 2024; Lee and Chung, 2024; Dell'Acqua et al., 2025). However, viewing GenAI solely as a productivity-enhancing instrument is an oversimplification. Technology also impacts human attitudes and motivation, factors crucial in determining the quality of work.

This influence on motivation can be especially important in fields such as knowledge and artistic creation, where professionals enjoy significant autonomy over their work. In these areas, some spend long hours striving for meticulously crafted work, while others settle for pieces of mediocre quality. Given that motivation plays a key role in determining output quality, how does the introduction of GenAI affect motivation within these fields and, consequently, how does it impact the quality of the resulting work?

We investigate these questions by conducting an experiment involving professional illustrators, examining their performance with and without the assistance of GenAI. A distinctive feature of our experimental design is to video record the entire work process. This allows us to trace out the level of quality achieved against the time spent on the task and thereby construct a *quality curve*.

Analyzing the quality curve provides a framework for distinguishing two types of GenAI effects on quality. First, GenAI can have a direct productivity effect, that is, it can change *quality level* achieved for any given time spent. Second, GenAI can have an indirect effect on motivation. This is reflected in changes in *quality curve's slope*, which measures the return to effort. By affecting the return for each additional unit of time, the slope of the quality curve influences when the individuals will stop working and, therefore, their motivation for quality.

Empirically, we find that GenAI has a positive direct productivity effect by lifting the entire quality curve upward: at each point in time, the quality achieved using GenAI exceeds that without AI. More importantly, however, we identify a negative indirect effect on motivation for quality. While GenAI enables the quality curve to rise rapidly in the initial stages of work, it also causes the curve to flatten more quickly. In other words, the return to effort diminishes faster with GenAI.

The negative effect on motivation results in shorter work sessions for 77% of illustrators. Furthermore, it qualitatively changes the distribution of time spent by the illustrators. Without GenAI, illustrators are most likely to work until the time limit (at the 150-minute mark), creating a single peak in the time distribution. The negative motivation effect makes this distribution bimodal. The new peak around the 60-minute mark is precisely where the quality curve associated with GenAI assistance starts to flatten, indicating that additional time yields minimal further improvement.

For 36% of the illustrators, the detrimental impact of GenAI on motivation outweighed its direct productivity benefits. Even though GenAI enabled these participants to achieve better quality for any specific amount of time spent, their final quality was lower. Further analysis reveals that most of these creators (30% of 36%) also shortened their time spent, pointing to a strategic decision to stop earlier in response to the flattening of the quality curve.

Our findings serve as a cautionary note about the use of GenAI in organizations: while GenAI can directly enhance productivity, it may paradoxically harm employee motivation and lower the quality of the final output. The adverse effect of AI adoption has been observed in several studies (Dell'Acqua, 2022; Doshi and Hauser, 2024; Wang et al., 2024; Zhou and Lee, 2024). Our paper shows that the quality drop is associated with a reduced return on effort. As the reduced return is an inherent feature of GenAI assisted knowledge work, firms need to adjust their product market strategy and organizational design accordingly to fully realize the productivity and innovation potential offered by GenAI (Milgrom and Roberts, 1990; Ichniowski et al., 1997; Brynjolfsson and Milgrom, 2013; Cowgill et al., 2024; Acemoglu et al., 2022a).

2 Experimental Design

We conducted an experiment involving 219 professional and student artists who completed two illustrations based on novel excerpts, a creative task familiar from prior academic or professional experience.[1] We employed a within-subject design where the treatment group (N=168) completed their first illustration task without AI and their second task with free access to an AI tool. A randomly assigned control group (N=51) completed both tasks without AI to account for potential order effects between tasks.

The experiment took place remotely from August 7 to September 12, 2024, with participants using their own environments (e.g., art studios or classrooms) and equipment. We imposed a time limit of 150 minutes per illustration for comparability.[2] The AI tool is Tiamat, a text-to-image model comparable to Midjourney/DALL-E, optimized for Chinese (participants' primary language). The treated participants received access and instructions 30 minutes before their second task, facilitating familiarization given their high baseline AI literacy (80% prior experience with LLM or text-to-image tools).

To motivate high-quality work, incentives included a 100 RMB base payment plus performance bonuses. For each task, illustrations from control and treatment groups were pooled and evaluated for prizes: one top (10,000 RMB), two second-tier (5,000 RMB each) and five third-tier (2,000 RMB each). Participants were not aware of the total number of contestants,others' identities, or their outputs, ensuring comparable perceived competition.

Data collection involved screen or workspace recordings and quality evaluations. We

---

[1] The task required creating an illustration based on a short textual excerpt from a well-known novel. We predefined 12 illustration topics based on paragraphs extracted from four well-known novels. Each participant was randomly assigned two different topics from this set, one for each illustration task.

[2] Some participants exceeded the time limit due to overlooking the end notification, technical difficulties accessing instructions, or extra time spent uploading files. We retained data reflecting this extra time, as it was not systematically related to the treatment assignment.

extracted snapshots at 15 minute intervals and had experts assess the quality of the intermediate and final outputs. Output quality was evaluated by expert raters (≥ 5 per image) blinded to experimental conditions (treatment/control, task order, AI use) and participant identity, using a 0-100 scale. Each illustration was evaluated by at least five independent raters. Raw scores were normalized within rater-topic combinations before averaging to mitigate rater and topic effects.

A pre-task survey collected baseline information, including demographic characteristics, professional experience, level of education, and prior usage of AI. Table 1 provides summary statistics and balance checks, showing general comparability between treatment and control groups, with only a minor difference in average years of education. Within the treatment group, 79% utilized the AI tool during their second illustration task. Although adoption likelihood correlates with factors such as age and academic majors, these AI adopters closely resemble the overall sample in other measured characteristics.

Of the 219 participants, 146 provided usable video data suitable for recovering the quality curve. Data from the remaining 73 participants were excluded due to technical difficulties with video recording, failure to record the session, or corrupted files. Those excluded due to technical problems were comparable in most of the baseline characteristics (Figure C.1 in the appendix). Our analysis proceeds in two parts: first, we examine quality curves using the sample of AI-adopters who have valid video data; second, we estimate the average effect of AI *access* on time spent and quality using difference-in-differences (DID) comparison between the entire treatment and control groups.

3 Results

3.1 AI Flattens the Quality Curve

AI assistance fundamentally changes the relationship between the time spent and the quality of the output. Figure 1A plots average quality scores at 15-minute intervals for creators who adopted AI, comparing their performance without AI (Task 1, blue curve) and with AI (Task 2, red curve). Without AI, quality increases with time spent. However, the quality gain appears to slow down over time, suggesting diminishing returns to time spent on the task.

With AI, the quality curve shifts significantly upward, indicating that AI enables creators to achieve higher quality levels for any given amount of time spent. More importantly, AI changes the shape of the curve. There is a rapid, almost immediate surge in quality: AI users reached an average quality score of 53 points in just 15 minutes, a level that requires 60 minutes without AI. However, after the initial rapid improvement, AI reduces the subsequent rate of quality gain considerably, resulting in a substantially flatter curve after the initial stage compared to the non-AI condition. Consequently, the quality gap between AI-assisted and non-AI-assisted illustrations narrows over time. Conditional on working the full 150 minutes, the quality differences are no longer statistically significant.

This change in quality curve reflects AI's role in assisting the creative process. Specifically, AI enables creators to quickly generate detailed illustrations from prompts, leading to a rapid initial increase in quality. However, creators reach a quality plateau sooner, as subsequent improvements require fine-tuning of details and exploration of ideas, tasks that do not necessarily improve quality. In contrast, without AI, quality improvements arise primarily from manual implementation of details, leading to a steady and sustained increase in quality throughout most of the production process.

As creators stopped working at different times, we conducted robustness checks to ensure that these patterns were not driven by compositional changes. Figure 1B presents the change in output quality within a creator after AI access, estimated using a DID approach with creator fixed effects. The coefficients confirm that AI access significantly increased quality during the first 60 minutes; thereafter, the magnitude of the improvement decreases and becomes statistically insignificant after 90 minutes. Furthermore, separate analyses of early versus late stoppers yield similar quality curve patterns. Placebo tests using non-AI adopters and the control group confirm stable quality trajectories across both tasks in the absence of AI use (see Figures A.2 and A.3 in Appendix). Collectively, these findings suggest that the observed flattening of the quality curve represents a genuine effect of AI on the creative workflow, rather than an artifact of participant attrition.

3.2 Creators Strategically Reduce Time

Faced with the flattening of the quality curve, creators strategically adjusted their time spent on the task. Comparing the full treatment and control groups using a DID approach, we find that access to AI leads to a significant reduction in the average time used by 24 minutes (0.63 SD decrease, $p < 0.001$; Figure 2A). This average reduction reflects creators stopping earlier when faced with diminishing marginal returns to their time spent.

However, the average effect masks important underlying changes in the full distribution of time used. Figure 2B highlights these distributional changes among the treated participants. Without AI (Task 1), the distribution is unimodal, with most creators working up to the maximum allowed time of 150 minutes. With AI (Task 2), the distribution becomes distinctly bimodal, with a new peak emerging around the 60-minute mark. This bimodal pattern suggests that a significant share of creators chose to complete their work much earlier when assisted by AI. In contrast, the time distribution for non-adopters and the control group remains unimodal across both tasks (Figure A.4 in Appendix).

The emergence of the 60-minute peak corresponds closely to the point where incremental quality gains diminish substantially under AI assistance. We calculate the average quality improvement achieved during each 15-minute interval for AI adopters (Figure 3A). Without AI, incremental quality gains remain strictly positive until the 120-minute mark. With AI, quality improvements are consistently lower and become statistically indistinguishable from zero by the 60-minute mark.

We formally test the link between stopping behavior and marginal return to time spent. Figure 3B shows that creators are significantly more likely to stop working when the incremental quality change in the preceding 15 minutes is lower; most stopping occurs when the incremental quality gain drops below 10 points.[3] Regression analysis confirms this negative relationship: controlling for creator fixed effects, a 10-point decrease in the prior interval's quality gain is associated with a 3-percentage-point increase in the probability of stopping ($p < 0.001$). Strikingly, the estimated coefficients that relate quality gains to stopping probability are similar across treatment conditions and illustration tasks, suggesting that the creators' willingness to trade time for quality remains stable (Table A.1 in Appendix). They apply the same decision rule, but AI-induced change in the quality curve leads to different stopping points.

3.3 Combined Effects on Quality

We now examine how the changes in the quality curve and the resulting adjustment in creators' time spent jointly shape the distribution of output quality. Figure 4A shows that the average quality slightly increases by 0.15 standard deviations, although statistically insignificant ($p = 0.25$). The change in quality reflects two opposing effects: better quality for any given time spent (positive productivity effect) and less time spent because of the flattened quality curve (negative motivation effect). This modest net improvement in quality suggests that on average, the direct productivity benefits appear to slightly outweigh the adverse impact on motivation.

In addition to the changes in average quality, we also examine changes in the entire quality distribution. Figure 4B shows that AI assistance reduces the share of low-quality outputs while simultaneously increasing the proportion of high-quality work. Furthermore, output similarity does not decline with AI usage, suggesting that collective diversity is preserved despite the reduction in time spent (Figure A.5 in Appendix).

Notably, the overall distribution of the final quality remains largely unimodal. This

---

[3] Although creators were, on average, more likely to stop when returns diminished, a small subset (6%; rightmost bin in Figure 3B) behaved differently, stopping shortly after achieving high incremental gains (15 minutes). This suggests some may act as "target earners," stopping once a desired quality threshold is met, a behavior observed primarily in the AI-assisted task.

may appear surprising because the distribution of time spent is bi-modal. One explanation is that AI has weakened the link between time spent and quality achieved. Without AI, additional time spent is positively correlated with final quality (approx. 0.3 points per 10 minutes, $p = 0.03$). With AI, this correlation drops to near zero ($r < 0.001, p = 0.8$), suggesting that the final quality of the output depends mainly on the initial idea generation and conceptualization of the creators rather than on the implementation.

3.4 Joint Changes in Time and Quality

While the aggregate analyses reveal overall shifts in time use and quality distribution, these dimensions are inherently linked through the creators' choices along the quality curves. To understand how creators rebalance the tradeoff after AI changes the marginal return to time, we now examine the joint changes in time and quality.

Figure 4C shows that 77% creators reduced the time they spent compared to their earlier work without AI, and approximately 64% achieved higher output quality. More importantly, the distributions of the changes in time and quality are not independent. The bottom right quadrant is sparse, reflecting that very few creators simultaneously spend more time, yet achieve lower quality. Among creators who experienced lower quality, the vast majority (30% out of 36%) also reduced their time spent on the task. A formal comparison confirmed that this joint pattern of changes in time and quality within the treatment group was significantly different from the pattern observed in the control group (Chi-square test, $p = 0.005$).

This pattern suggests a deliberate trade-off between time and quality. If creators were randomly deciding how long to spend on the task, around 50% of the creators would reduce their time input, and the other 50% would increase. In this case, there would probably be more than 64% of creators who would obtain higher quality: after all, the quality curve with AI lies entirely above the curve without AI. In addition, if creators were solely targeting a specific quality level independent of time considerations, one would expect that half of the creators would experience a quality decline after adopting AI. This is not what is observed.

3.5 Implications on Inequality

Whether AI equalizes human performance or exacerbates existing gaps remains an open empirical question with mixed findings in the literature. We provide two pieces of evidence suggesting that, in our context, AI adoption may reduce inequality between high- and low-ability creators, defined by their baseline illustration performance without AI assistance.

First, we find that creators in the lowest baseline performance quartile experience the largest quality gains with AI, although the differences across quartiles are not statistically significant. This finding does not appear to be driven by heterogeneous responses in time adjustment; in fact, the magnitude of time reduction is similar across skill segments (Figures A.6A and A.6B in Appendix).

Second, AI assistance substantially weakens the persistence of individual performance differences. In the control group, creators who perform better on the first illustration task also tend to perform better on the second, with a correlation of 0.43 ($p < 0.01$). In contrast, among AI adopters, this correlation falls to just 0.13 and is statistically insignificant ($p = 0.16$) (Figures A.6C and A.6D in Appendix).

One explanation for this finding is that AI reduces the role of manual implementation skills, mitigating skill-based inequality. Without AI, creators differ significantly in implementation abilities, such as drawing technique, and these individual differences strongly correlate with performance across tasks. With AI assuming responsibility for much of the implementation process, these specialized technical skills become less relevant, weakening the performance correlation between tasks. Our findings point toward an increased importance of creative skills, along with a reduction in inequality in technical implementation skills.

4 Conclusion and Discussion

This paper investigates how GenAI affects the motivation of professional illustrators and its impact on the quality of the resulting work. We find that GenAI lifts the entire quality curve upwards, creating a direct positive effect on productivity. However, GenAI also accelerated the drop in marginal return to time spent, causing the quality curve to flatten earlier. This leads to a negative indirect effect on motivation.

The pattern we observe—a steep initial rise in quality followed by a rapid flattening—appears to be a general principle of GenAI-assisted knowledge work. This pattern is a mathematical inevitability: If the time required to achieve true excellence remains fixed, faster initial progress will necessarily result in a more sharply diminishing rate of return. While we focus on artwork, this principle is evident in many fields, from writing academic papers to drafting business reports. Individuals can produce good enough outcomes with remarkable speed, only to labor extensively for the additional improvement that marks true quality. This frustrating gap between initial ease and final effort is the essence of the GenAI Paradox: better technology, worse motivation.

Our findings have significant strategic implications for companies. The transformation of the quality curve suggests a hollowing out of the middle, forcing a choice between two distinct product-market strategies. Firms either aim for the highest echelon of quality or standardize their output at a "good enough" level. Intermediate levels of quality may become an inefficient choice.

Each strategic choice brings distinct organizational challenges. Firms pursuing top-tier quality face a fundamental motivation problem. When the quality curve flattens, workers must sustain high effort despite seeing only marginal improvements in output. These organizations need to discover new ways to keep workers motivated.

Firms choosing standardization have a different challenge: the need for comprehensive organizational redesign. Without restructuring and redesigning jobs, workers naturally drift toward the flattened portion of the quality curve, wasting significant time and effort for minimal gains. To prevent this inefficiency, these organizations must fundamentally reimagine their workflows, keeping workers focused on the steep portion of the curve where GenAI delivers maximum value.

Ultimately, our findings raise questions about the trajectory of technological innovation itself. If we can influence the direction of technology, we should aspire to develop tools that augment human motivation rather than undermine it. Only by creating an environment where technology empowers and inspires can we build a future where humans and AI thrive together.

References

Acemoglu, Daron, David Autor, Jonathon Hazell, and Pascual Restrepo, "Artificial Intelligence and Jobs: Evidence from Online Vacancies," *Journal of Labor Economics*, 2022, *40* (S1), S293–S340.

_ , **Gary W. Anderson, David N. Beede, Cathy Buffington, Eric E. Childress, Emin Dinlersoz, and Lucia S. Foster**, "Automation and the workforce: A firm-level view from the 2019 annual business survey," Working Paper w30659, National Bureau of Economic Research 2022.

Bono, James and Alec Xu, "Randomized Controlled Trials for Security Copilot for IT Administrators," 2024. arXiv preprint arXiv:2411.01067.

Boussioux, Léonard, Jacqueline N Lane, Miaomiao Zhang, Vladimir Jacimovic, and Karim R Lakhani, "The Crowdless Future? Generative AI and Creative Problem-Solving," *Organization Science*, 2024, *35* (5), 1589–1607.

Brynjolfsson, Erik and Paul Milgrom, "Complementarity in Organizations," in Robert Gibbons and John Roberts, eds., *The Handbook for Organization Economics*, Princeton University Press, 2013.

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

Cowgill, Bo, Pablo Hernandez-Lagos, and Nataliya Wright, "Does AI Cheapen Talk? Theory and Evidence From Global Entrepreneurship and Hiring," July 2024. Columbia Business School Research Paper No. 4702114.

Dell'Acqua, Fabrizio, Charles Ayoubi, Hila Lifshitz-Assaf, Raffaella Sadun, Ethan R. Mollick, Lilach Mollick, Yi Han, Jeff Goldman, Hari Nair, Stew

Taub, and Karim R. Lakhani, "The Cybernetic Teammate: A Field Experiment on Generative AI Reshaping Teamwork and Expertise," 2025. SSRN Working Paper 5188231.

_ , Edward McFowland III, Ethan R. Mollick, Hila Lifshitz-Assaf, Katherine Kellogg, Saran Rajendran, Lisa Krayer, François Candelon, and Karim R. Lakhani, "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality," Working Paper 24-013, Harvard Business School Technology & Operations Management Unit 2023.

Dell'Acqua, F., "Falling asleep at the wheel: Human/AI collaboration in a field experiment on HR recruiters," Technical Report, Harvard Business School 2022. Working paper.

Doshi, Anil R. and Oliver P. Hauser, "Generative AI Enhances Individual Creativity but Reduces the Collective Diversity of Novel Content," *Science Advances*, 2024, *10* (28), eadn5290.

Haslberger, Matthias, Jane Gingrich, and Jasmine Bhatia, "No Great Equalizer: Experimental Evidence on AI in the UK Labor Market," Technical Report, SSRN Working Paper 4594466 2024.

Ichniowski, Casey, Kathryn Shaw, and Giovanna Prennushi, "The Effects of Human Resource Management Practices on Productivity: A Study of Steel Finishing Lines," *The American Economic Review*, 1997, *87* (3), 291–313.

Jia, Nan, Xueming Luo, Zheng Fang, and Chengcheng Liao, "When and How Artificial Intelligence Augments Employee Creativity," *Academy of Management Journal*, 5 2024, *67* (1), 5–32.

Kanazawa, Kyogo, Daiji Kawaguchi, Hitoshi Shigeoka, and Yasutora Watanabe, "AI, Skill, and Productivity: The Case of Taxi Drivers," Working Paper w30612, National Bureau of Economic Research 2022.

Kreitmeir, David and Paul Raschky, “The Heterogeneous Productivity Effects of Generative AI,” Technical Report, SSRN Working Paper 4745624 2024.

Lee, Byung Cheol and Jaeyeon Chung, “An Empirical Investigation of the Impact of ChatGPT on Creativity,” *Nature Human Behaviour*, 2024, *8* (10), 1906–1914.

Milgrom, Paul and John Roberts, “The Economics of Modern Manufacturing: Technology, Strategy, and Organization,” *The American Economic Review*, 1990, *80* (3), 511–528.

Noy, Shakked and Whitney Zhang, “Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence,” *Science*, 2023, *381* (6654), 187–192.

Peng, Sida, Eirini Kalliamvakou, Peter Cihon, and Mert Demirer, “The Impact of AI on Developer Productivity: Evidence from GitHub Copilot,” 2023. arXiv preprint arXiv:2302.06590.

Wang, Dawei, Difang Huang, Haipeng Shen, and Brian Uzzi, “A Large-scale Evaluation of the Collaborative Potential of Human and Machine Creativity,” PsyArXiv September 28 2024.

Zhou, Eric and Dokyun Lee, “Generative Artificial Intelligence, Human Creativity, and Art,” *PNAS Nexus*, 2024, *3* (3), page052.

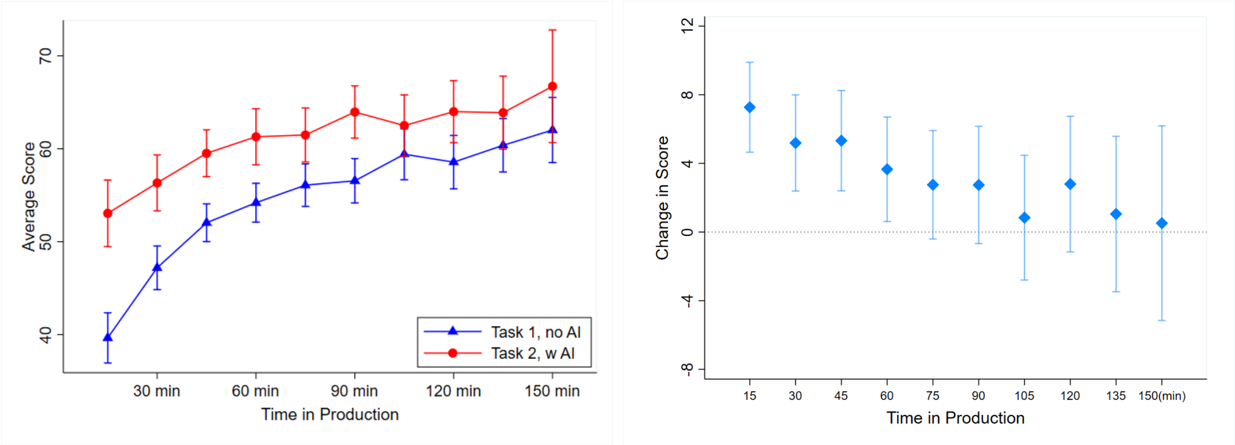
Tables and Figures

Table 1: Summary Statistics

VARIABLES	Treatment (N=168)	Control (N=51)	Difference	p-value
Female	0.76	0.76	0.00	0.95
Age	23.38	22.78	-0.60	0.51
Education (yrs)	16.37	16.02	-0.35	0.02**
Employed	0.26	0.33	-0.37	0.31
Majored in Art	0.68	0.73	0.05	0.54
Took art exam	0.62	0.61	-0.01	0.92
Working experience (yrs)	2.25	2.35	-0.20	0.87
Ever full-time in Art/Design	0.27	0.25	-0.10	0.59
Used Graphic AI	0.82	0.78	-0.61	0.91
Used LLM	0.84	0.84	0.00	0.96

Notes. This table reports the raw means for the treatment and control groups, along with the differences in means and corresponding p-values. "Employed" equals 1 if a participant has ever worked either full-time or part-time, and 0 otherwise. "Took art exam for college" equals 1 if the participant took the art track entrance exam for college, and 0 otherwise. "Ever full-time in Art/Design" indicates whether the respondent has ever held a full-time job in an art or design-related field. "Used Graphic AI" and "Used LLM" refer to self-reported use of generative AI tools for visual and language-based tasks, respectively.

Figure 1: Impact of AI on Time–Quality Relationship

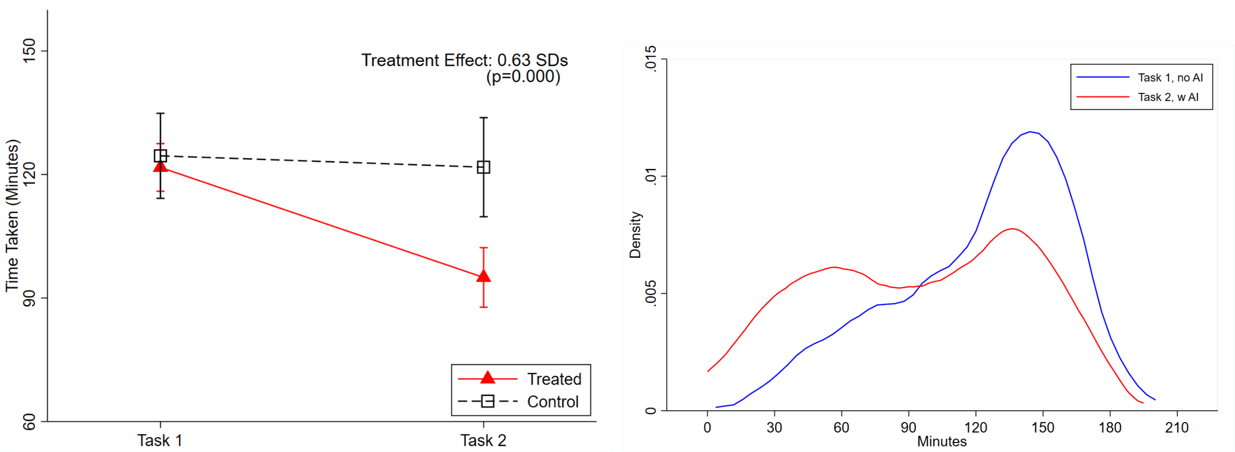


A Time-quality relationship among AI compliers

B Within-person quality gain from AI

Notes: Panels (A) and (B) use process data from participants. In (A), the curves plot average illustration quality (with 95% confidence intervals) every 15 minutes for treatment-group participants who actually adopted AI in Task 2. Red denotes Task 2 (with AI); blue denotes Task 1 (without AI). AI use is identified via self-report or video verification. Panel (B) shows point estimates (diamonds) and 95% confidence intervals of the causal effect of AI access on illustration quality at each 15-minute interval, estimated with individual and interval fixed effects.

Figure 2: Strategic Time Reduction

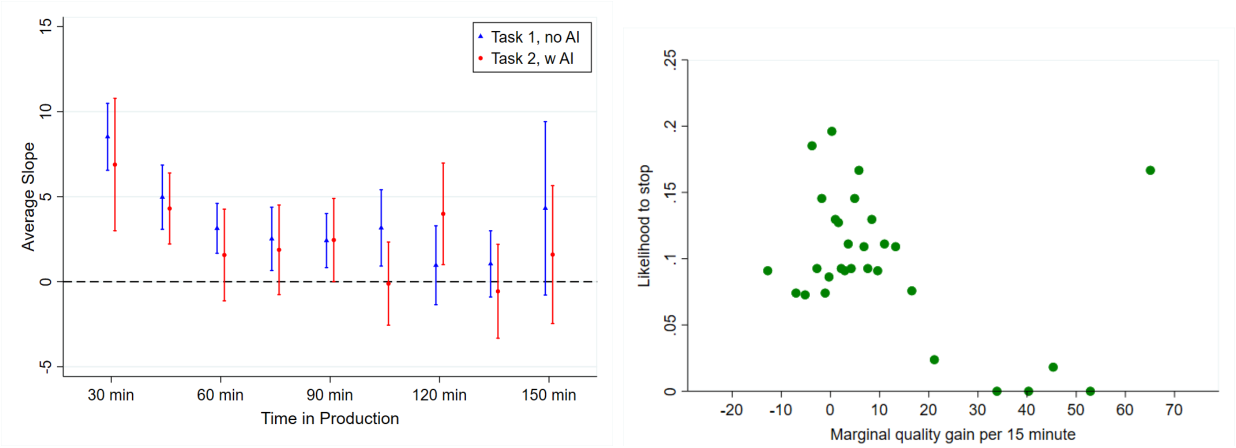


A Average effects on time spent

B Distribution of time spent (treatment group)

Notes: Panel (A) reports the mean minutes required to finish each task, with 95% confidence intervals, for treatment and control groups. The treatment effect is estimated from a regression of the within-participant change in time (Task 2 –Task 1) on a treatment-group indicator, using robust standard errors. Panel (B) displays the distribution of minutes spent on Task 1 (without AI, blue) versus Task 2 (AI available, red) for all participants in the treatment group. Note that only the treatment group received access to AI in Task 2; both groups completed Task 1 without AI.

Figure 3: Quality Gain and Stopping Decisions

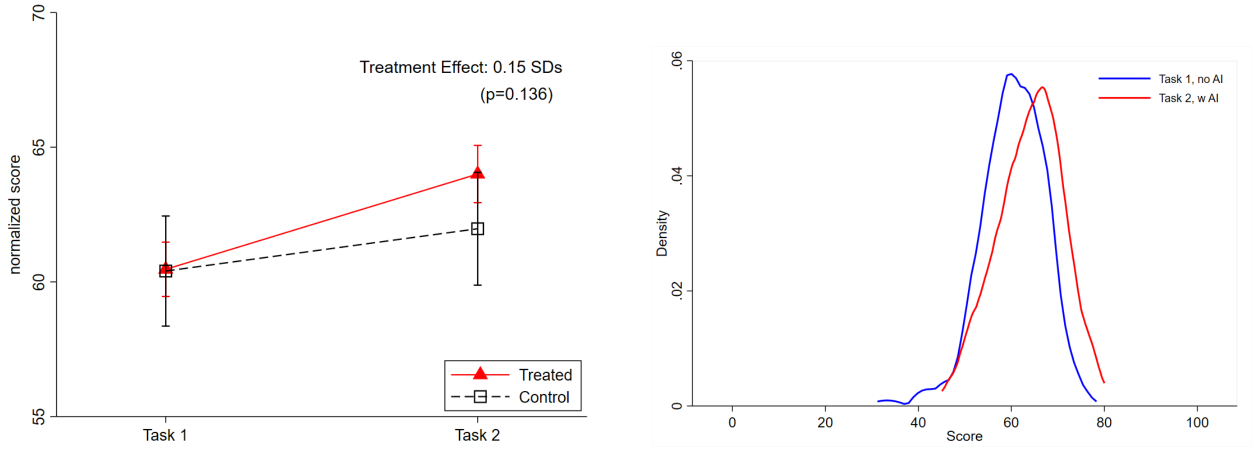


A Quality change per 15-min interval

B Probability of stopping vs. marginal quality gain

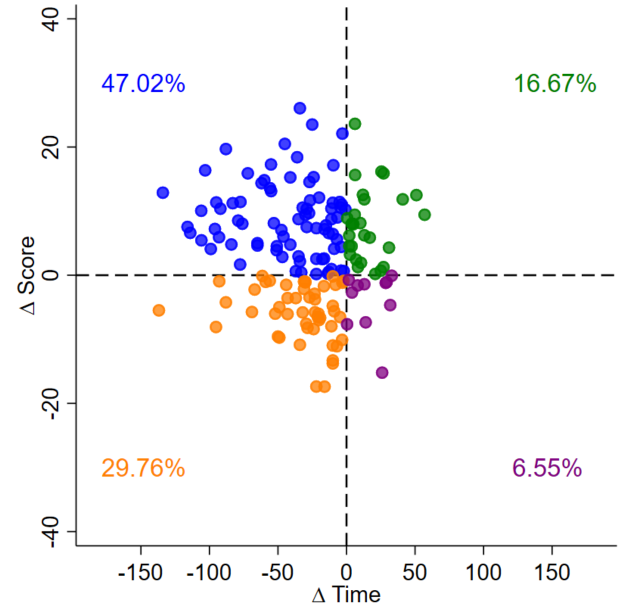
Notes: Panel (A) graphs the mean change in illustration quality during each 15-minute segment, with 95% confidence intervals, among treatment-group participants who actually adopted AI in Task 2. Red shows Task 2 (with AI) and blue shows Task 1 (without AI). Panel (B) relates the marginal quality improvement over the preceding 15 minutes to the likelihood that a participant stops working. Creator-time observations are binned by marginal quality gain; the sample includes all participants.

Figure 4: Final Output Quality and Joint Time–Quality Change



A Average effects on final output quality

B Distribution of final output quality (treatment group)



C Joint change in time and quality after using AI

Notes: Panel (A) reports mean illustration quality for each task, with 95% confidence intervals, for treatment and control groups. The treatment effect is estimated from a regression of the within-participant change in quality (Task 2 − Task 1) on a treatment-group indicator, using robust standard errors. Panel (B) shows the distribution of final-output quality for Task 1 (without AI, blue) and Task 2 (AI available, red) among treatment-group participants. Panel (C) plots the change in time spent (x-axis) and change in output quality (y-axis) between Task 1 and Task 2 for each treatment-group participant; percentages denote the share of observations in each quadrant. Note that only the treatment group received access to AI in Task 2; both groups completed Task 1 without AI.

A Appendix Tables and Figures

Figure A.1: Experimental flow chart

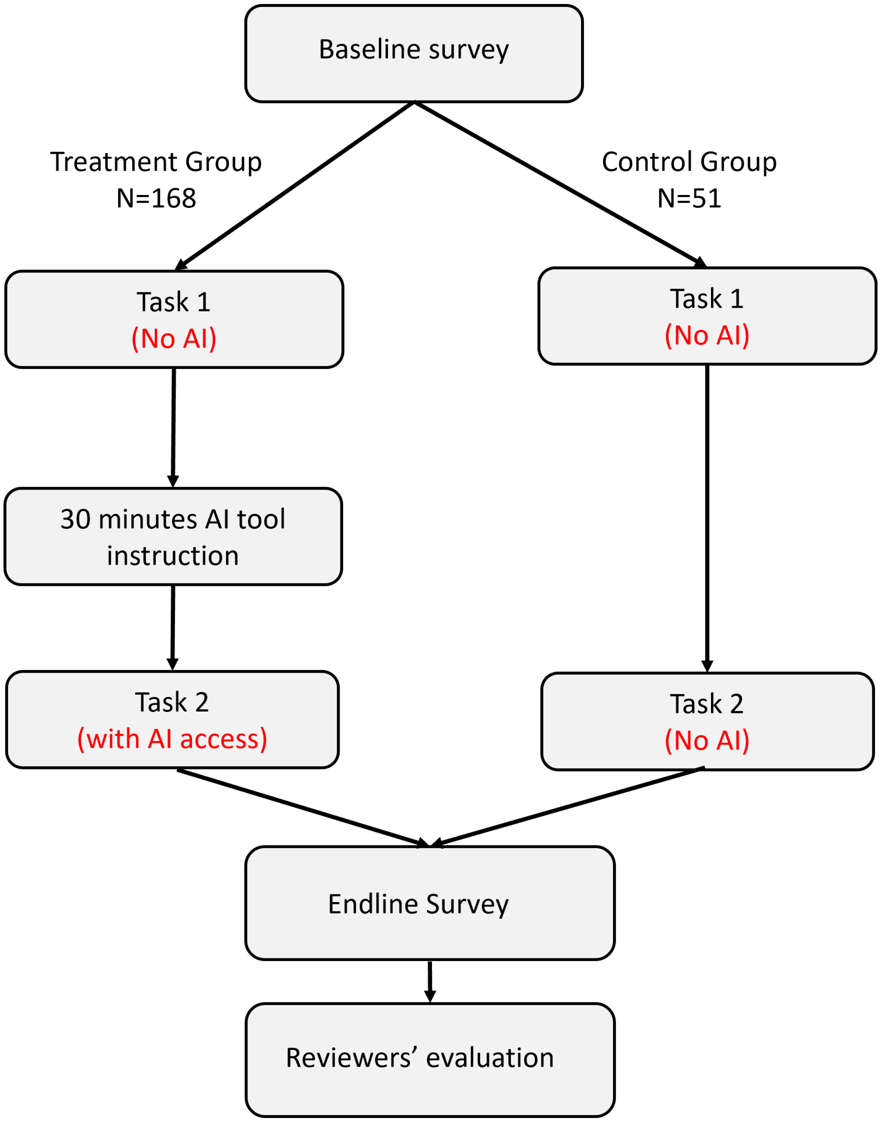
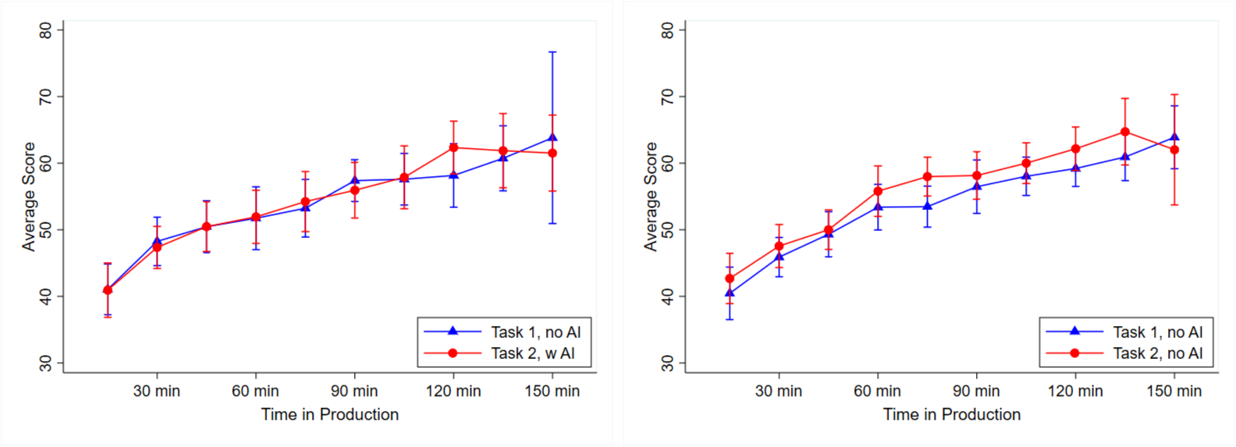


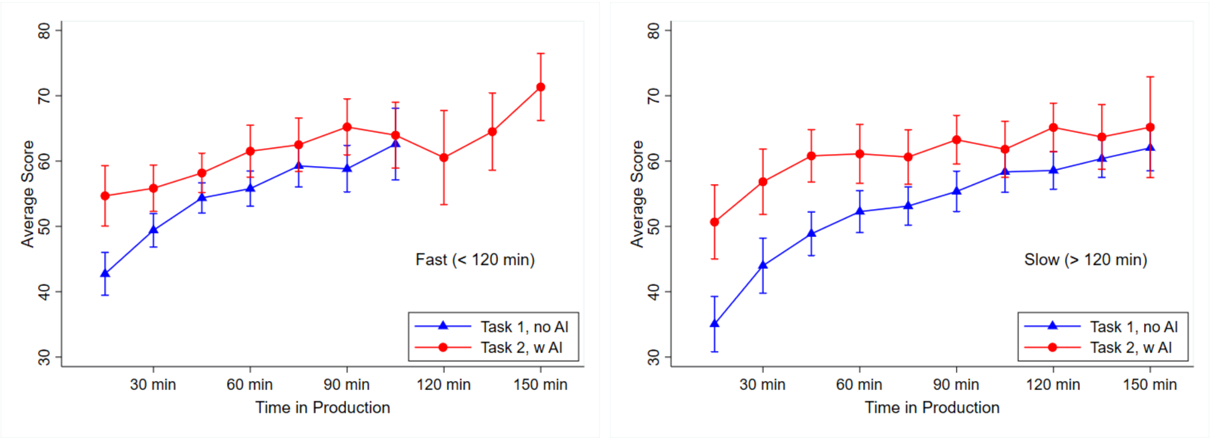
Figure A.2: Time–Quality Relationship for Control Group and Non-Adopters



A Non-adopters (treatment group) **B** Control group

Notes: Each point represents the average image quality at a given time interval, with 95% confidence intervals. Panel (A) plots non-adopters in the treatment group; Panel (B) plots the control group.

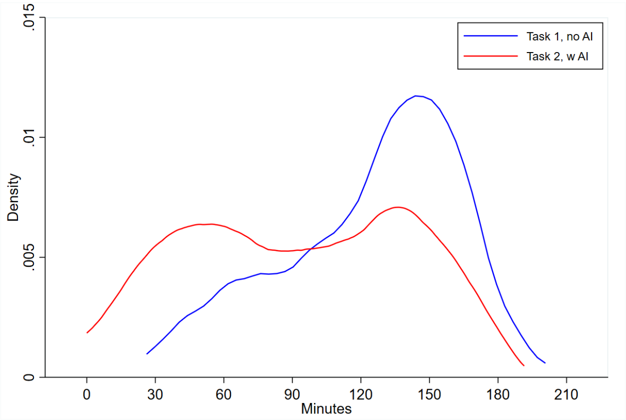
Figure A.3: Time–Quality Relationship by Time Spent in Task 1



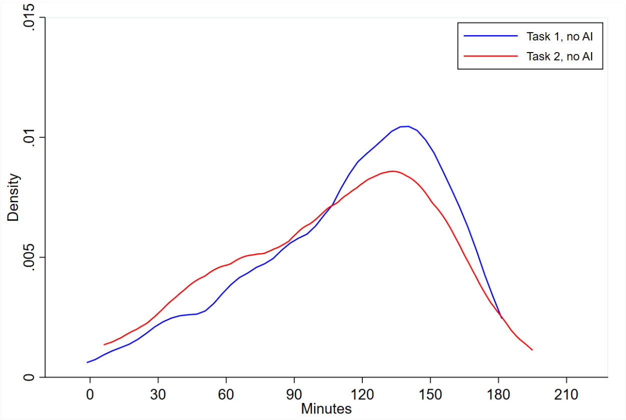
A Stop early in Task 1 **B** Stop late in Task 1

Notes: The sample is split at the median time spent on the first illustration. Panel (A) plots the time–quality trajectory for creators who finished Task 1 early; Panel (B) plots those who worked longer. Each point is the mean image quality at the corresponding time interval, with 95% confidence intervals.

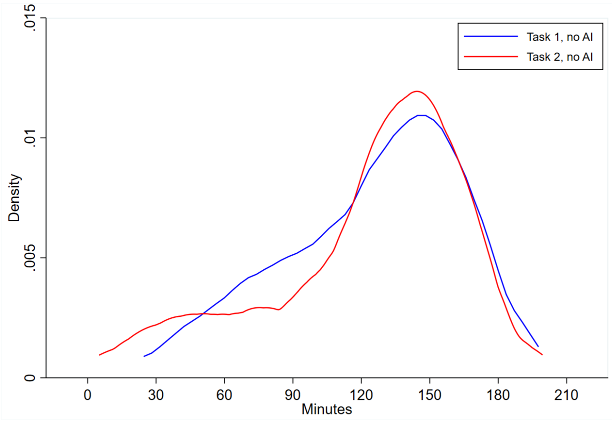
Figure A.4: Distribution of Time Taken by AI Adoption



A AI adopters



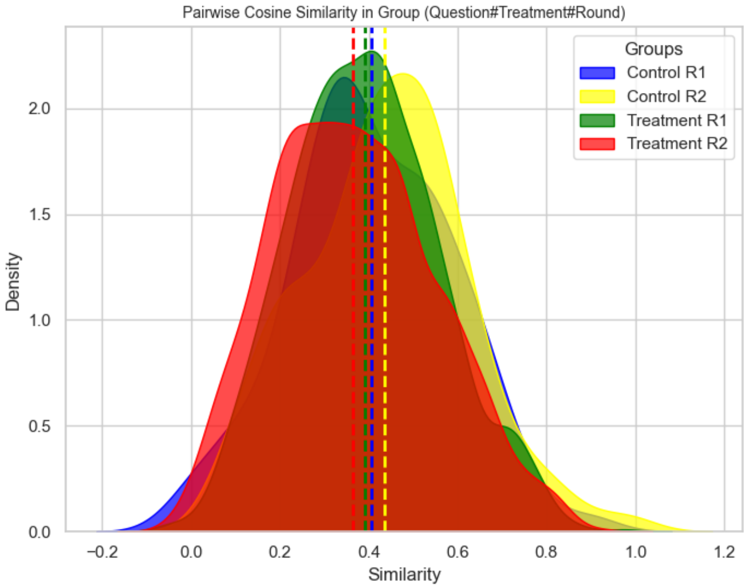
B Non-adopters



C Control group

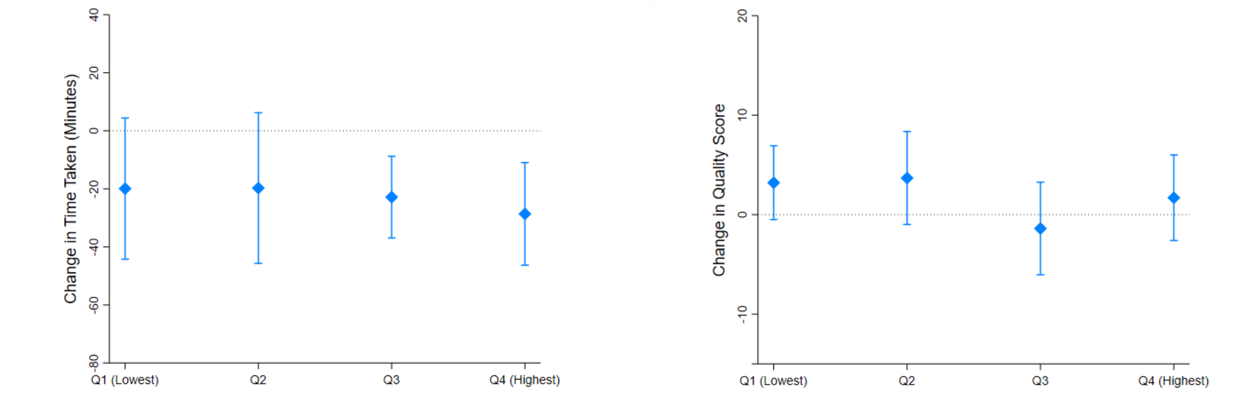
Notes: The panels plot the distribution of minutes spent on Task 1 (without AI, blue curve) and Task 2 (AI access only for the treatment group, red curve). Panel (A) shows AI adopters within the treatment group; Panel (B) shows non-adopters; Panel (C) shows the control group.

Figure A.5: Similarity among output

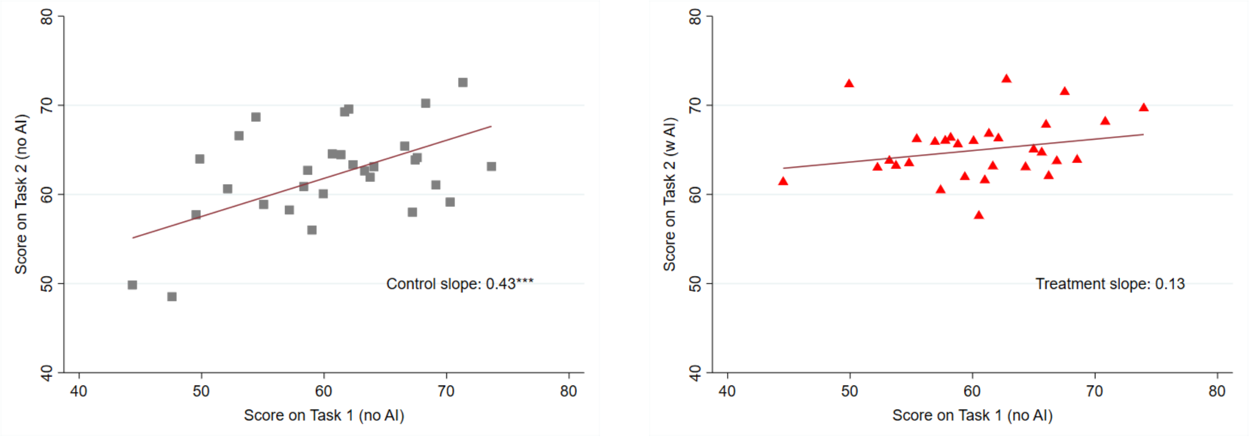


Notes: This figure shows the distribution of pairwise cosine similarity scores among illustrations within each experimental condition. In Task 1 (R1), neither the control nor the treatment group had access to the AI tool. In Task 2 (R2), the treatment group was given access to AI, while the control group continued without it. Cosine similarity measures how thematically similar participants' illustrations are. For each task topic (the text excerpts presented to participants as prompts for their illustrations), we used the BLIP-2 vision–language model to generate one-sentence captions for each illustration. These captions were then converted into numerical text embeddings using Sentence-BERT (paraphrase-multilingual-MiniLM-L12-v2), and cosine similarity was calculated between every pair of captions within the same topic. The distributions shown here pool these similarity scores across all topics for each group. Curves farther to the right indicate higher similarity. Dashed lines indicate the mean similarity for each group.

Figure A.6: Heterogeneity Analyses in Time and Quality Based on Baseline Skills



A Heterogeneous effect on time use **B** Heterogeneous effect on output quality



C Quality correlation — control group **D** Quality correlation — AI adopters

Notes: Panels (A) and (B) plot heterogeneous treatment effects on time use and final-output quality, respectively, across workers' baseline skills—proxied by their quality ratings in the first (pre-AI) illustration task. Panels (C) and (D) bin observations by the first-task quality score and show average second-task quality for the control group (C) and AI adopters (D); slopes derive from creator-level regressions of second-task quality on first-task quality.

Table A.1: Regression Results for Stopping Decision

	Control Group		Treatment Group	
	Task 1	Task 2	Task 1	Task 2
Slope	-0.003**	-0.003**	-0.003**	-0.003**
	(0.001)	(0.001)	(0.001)	(0.001)
Constant	0.107***	0.123***	0.101***	0.172***
	(0.024)	(0.027)	(0.017)	(0.034)
ID Fixed Effects	✓	✓	✓	✓
Observations	195	187	691	559
R-squared	0.115	0.118	0.147	0.242

Notes. This table reports estimates of the probability of stopping drawing as a function of marginal quality gain over the preceding 15-minute interval. The dependent variable is an indicator equal to 1 if the creator stopped drawing before the next 15-minute window. Marginal quality is defined as the change in output quality between two adjacent intervals. All regressions include creator fixed effects. Robust standard errors are in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.10$

B Methods

B.1 Experimental Procedure

The experiment was conducted remotely from August 7 to September 12, 2024. Participants were recruited from several sources, including advertisements on major Chinese social media platforms (WeChat, Weibo, and Little Red Book), posts on the freelancer platform (Zhima Work) and targeted outreach to faculty and student groups at art and design schools. Eligibility was determined by a screening survey. To be included, participants were required to have a college degree in art-related fields or to have full-time professional experience in the visual art or design industries. We recruited 221 individuals, from which two were excluded for being under 18 years of age, resulting in a final sample of 219 participants. This research received ethical approval from the Human Research Ethics Committee at the University of Hong Kong (EA240360). Experimental procedures are outlined below.

Experimental Flow. Figure A.1 in Appendix summarizes the experimental procedure. After signaling their readiness online, participants first completed a baseline survey to capture demographic information, professional background, and prior AI experience. Participants were then randomly assigned to either a control group (N=51) or a treatment group (N=168).

Participants completed two illustration tasks sequentially. The first task, completed by all participants without AI access, served as a baseline. For the second task, the treatment group received free access to an AI tool, while the control group continued without it. We employed an unequal assignment ratio (1:3) to enable more detailed analysis within the treatment condition; the control group served to account for potential order effects, such as fatigue or learning, between the two tasks. Each task had a 2.5-hour time limit, and participants were required to complete both tasks within a single day.

Before starting the second task, participants in the treatment group received a 30-minute instructional session on the AI tool, which included a detailed manual and a hands-on practice period to ensure familiarity. The control group received explicit instructions to continue without AI. After completing both illustrations, all participants filled out an exit survey.

Video Recording. Participants were instructed to record their entire production process for each task and submit the recordings. Most participants used screen recording software for digital tools (e.g. Photoshop, Procreate), while those using physical media recorded with a camera. The recordings captured only the artwork and process, with no personal identifiers visible.

B.2 Dependent Measures

Output Quality Assessment. The primary outcome, output quality, was assessed by twelve expert judges (university professors or industry professionals with at least five years of experience) on a 0–100 scale. All evaluations were blinded to treatment condition and participant identity. The materials assessed included the final illustrations and intermediate snapshots extracted from the video recordings at 15-minute intervals.

Each illustration was evaluated by at least five independent judges. To prevent bias from observing a participant's artistic development, no judge evaluated more than one image (i.e., a final piece or a snapshot) from the same participant's creative process for a given task. In addition to rating overall quality, judges also rated four subdimensions: originality, logic, technical skill, and emotional expression. For the final analysis, all quality scores were normalized by judge-topic before being averaged for each illustration.

B.3 Econometric Specification

To estimate the average effect of AI access on time use and final output quality, we used a difference-in-differences(DID) approach as follows:

$$Y_{is} = \beta AI_{is} + \pi_i + \gamma_s + u_{is} \tag{1}$$

where Y_{is} is the outcome for individual i in task s (task 1 or task 2). The outcome variables include time used for production and output quality based on professional evaluation scores. AI_{is} is the treatment variable that equals one if the individual i in task s received AI access and zero otherwise. π_i is the individual fixed effects, γ_s is the task fixed effects, and u_{is} is the random error term. β is the coefficient of interest. When the parallel trend assumption holds, the estimate of β captures the causal effect of having access to AI on time use and output quality (Figure 2A and 4A).

To examine whether the effect of AI access on the outcome differs by time interval during the production process, we replace AI_{is} with the interactions between AI_{is} and dummy indicators of time intervals from 15 minutes. All fixed effects are the same as before, and we additionally control for time interval fixed effects λ_t. Equation (1) is hence revised as follows:

$$\begin{aligned} Y_{ist} =& \tilde{\beta}_{15} AI_{is} \times \mathbf{1}[t = 15] + \tilde{\beta}_{30} AI_{is} \times \mathbf{1}[t = 30] + \cdots + \tilde{\beta}_{150} AI_{is} \times \mathbf{1}[t = 150] \\ & + \pi_i + \gamma_s + \lambda_t + u_{is} \qquad (2) \\ =& \Sigma_{l=15}^{135} \tilde{\beta}_l \, AI_{is} \times \mathbf{1}\{t = l\} + \pi_i + \gamma_s + \lambda_t + u_{is}. \end{aligned}$$

The estimates for $\tilde{\beta}_t$ in the above equation reflect the causal effect of AI access at each given time interval t (Figure 1B).

Selective attrition may pose a threat to our identification strategy. As participants freely choose when to stop drawing, creating missing intermediate outcome data for those who exit early. The additional results in Appendix show that our results are not likely driven by selective attrition in the sample.

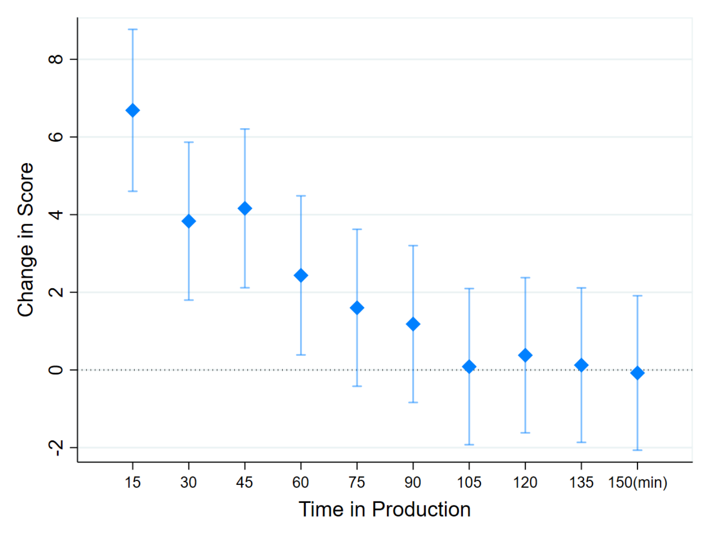
C Other Analyses: Sample Attrition

In this section, we address selective attrition in the our difference-in-differences analyses. Attrition may occur if some participants exit the experiment early, resulting in fewer observations for the process data. This can bias the estimated heterogeneous effects of the time interval treatment in two ways. First, if dropout in the baseline task (Task 1) is systematically related to treatment assignment—for example, treated participants with certain characteristics leave earlier than those in the control—baseline randomization is compromised and parallel trends become suspect from the outset. Second, even when baseline balance is intact, selective exit during the second task might lead to nonrandom missing observations, which would bias the estimated coefficients.

To address these issues, we perform two additional robustness checks. First, we verify that attrition in the baseline task (Task 1)—before introducing AI access—is balanced across treatment and control groups. Table C.2 shows that participants' time spent in the first task is uncorrelated with treatment assignment or any other individual characteristics. This ensures that our randomization holds, supporting the parallel trends assumption required for difference-in-differences identification.

In the second task, participants in the treatment group tend to spend less time, resulting in a higher likelihood of missing scores at later intervals. However, this pattern of missingness differs from classic attrition problems, where no outcomes are observed for participants who leave the sample. In our setting, participants who leave early still contribute observed outcomes before exiting the task; what remains unobserved is how their scores would have evolved had they continued drawing. Given that participants tend to stop when their incremental improvement becomes negligible, the Last Observation Carried Forward (LOCF) method offers a reasonable way to impute the missing outcomes for subsequent intervals. Figure C.1 presents the results when we reproduce Figure 1B using this approach. The results confirm that the overall pattern remains robust to this imputation strategy.

Figure C.1: Quality gain from AI with imputed missing data



Notes: This figure shows point estimates (diamonds) and 95% confidence intervals of the causal effect of AI assignment on illustration quality at each 15-minute interval, estimated with individual and interval fixed effects. We impute the missing observations using the last observed observation in a given round for each participant.

Table C.1: Balance Tests: Demographic Comparison by Video-Data Availability

Variable	w/ video data	w/o video data	Difference	p-value
Female	0.691	0.802	–0.110	0.085*
Age	24.863	22.486	2.377	0.003**
Education (yrs)	16.438	16.247	0.192	0.155
Employed	0.329	0.247	0.082	0.200
Majored in Art	0.726	0.678	0.048	0.470
Took art exam	0.616	0.616	0.000	1.000
Working experience (yrs)	2.356	2.233	0.123	0.477
Ever full-time in Art/Design	0.221	0.163	0.057	0.315
Used Graphic AI	0.863	0.788	0.075	0.179
Used LLM	0.822	0.849	–0.027	0.604
N	146	73		

Notes: $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.10$ “Difference” = mean(w/ video data) – mean(w/o video data).

Table C.2: Baseline Attrition Balance: Minutes Spent in Task 1

	Time Spent in Task 1 (minutes)	
	(1) Treatment only	(2) Control for Covariates
Treatment Group	-2.195	-3.465
	(8.053)	(8.473)
Age		0.153
		(1.328)
Majored in Art		1.181
		(9.794)
Employed		8.032
		(11.806)
Took Art Exam		-3.099
		(9.445)
Used Graphic AI		-14.958
		(10.005)
Used LLM		-13.960
		(11.460)
Education (years)		6.082
		(3.968)
Male		7.738
		(9.638)
Work Exp.		0.695
		(4.535)
Observations	146	141
R-squared	0.001	0.090

Notes: This table tests whether attrition in the *baseline* drawing task (Task 1) is balanced across treatment and control groups. The dependent variable is the maximum number of minutes a participant remained in Task 1 before exiting. Column (1) includes only the treatment indicator; column (2) adds demographic and background controls. Standard errors are in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.10$. Gender information was missing for 20 participants; we inferred their likely gender using auxiliary publicly available data. Results are robust to treating the gender of these participants as missing.