



The Value, Benefits, and Concerns of Generative AI-Powered Assistance in Writing

Zhuoyan Li

li4178@purdue.edu

Purdue University

West Lafayette, Indiana, USA

Jing Peng

jing.peng@uconn.edu

University of Connecticut

Storrs, Connecticut, USA

Chen Liang

chenliang@uconn.edu

University of Connecticut

Storrs, Connecticut, USA

Ming Yin

mingyin@purdue.edu

Purdue University

West Lafayette, Indiana, USA

ABSTRACT

Recent advances in generative AI technologies like large language models raise both excitement and concerns about the future of human-AI co-creation in writing. To unpack people's attitude towards and experience with generative AI-powered writing assistants, in this paper, we conduct an experiment to understand whether and how much value people attach to AI assistance, and how the incorporation of AI assistance in writing workflows changes people's writing perceptions and performance. Our results suggest that people are willing to forgo financial payments to receive writing assistance from AI, especially if AI can provide direct content generation assistance and the writing task is highly creative. Generative AI-powered assistance is found to offer benefits in increasing people's productivity and confidence in writing. However, direct content generation assistance offered by AI also comes with risks, including decreasing people's sense of accountability and diversity in writing. We conclude by discussing the implications of our findings.

CCS CONCEPTS

- Human-centered computing → Empirical studies in HCI; • Computing methodologies → Artificial intelligence.

KEYWORDS

Human-AI co-creation, AI writing assistant, Large language model

ACM Reference Format:

Zhuoyan Li, Chen Liang, Jing Peng, and Ming Yin. 2024. The Value, Benefits, and Concerns of Generative AI-Powered Assistance in Writing . In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA*. ACM, New York, NY, USA, 25 pages. <https://doi.org/10.1145/3613904.3642625>



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike International 4.0 License.

CHI '24, May 11–16, 2024, Honolulu, HI, USA

© 2024 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0330-0/24/05

<https://doi.org/10.1145/3613904.3642625>

1 INTRODUCTION

Recent advances in generative artificial intelligence (AI) have opened up exciting opportunities for fostering synergistic human-AI collaborations in completing various tasks. For instance, large language models (LLMs) such as GPT-3 [7] and GPT-4 [35] have demonstrated impressive capabilities in language understanding and generation, highlighting the promise of incorporating LLM-powered assistance into humans' writing processes to enable "human-AI co-writing" and enhance human productivity [34] and creativity [16]. Compared to traditional writing assistive tools, LLMs can provide a wide range of highly flexible assistance beyond simple grammar and spelling corrections, from polishing text writing styles to generating completely new content from scratch to inspire human writers. Recent research has shown that integrating generative AI-powered assistance in the workflows of professional writing tasks improves the productivity of human in these tasks [34]. As such, a growing line of research has emerged in designing more effective interfaces to facilitate the communication between humans and their generative AI-powered writing assistants, aiming to unleash the full potential of human-AI collaboration in writing [10, 31, 43, 45].

The significant promise brought up by generative AI-powered writing assistants appears to suggest that people would naturally value their assistance and be willing to pay for them. However, there are still lingering concerns on the use of such AI assistance in writing tasks. For example, writers have reservations that a writing process led by the AI writing assistant engenders their control, autonomy, and ownership [5], and there is a general fear that powerful automation technologies may completely replace human labor in the future [1]. With all these concerns, people may associate limited, if any, value with the assistance provided by AI, and they may be reluctant to pay for such assistance. So, given the pros and cons of generative AI-powered writing assistants, one may wonder whether people's attitudes towards these assistants are dominated by their benefits or risks, and therefore, whether people perceive some financial value from utilizing AI assistance in writing. In particular, are people willing to forgo some monetary compensation to complete their writing jobs together with a generative AI-powered assistant rather than completing the jobs on their own? If so, how will the perceived financial value of AI assistance vary with the kind of writing assistance offered by the generative AI model and people's own characteristics? How much benefit do different kinds of AI assistance bring about on people's writing experience and

performance, and how does it match up with people's value on these assistance?

To answer these questions, in this paper, we present a randomized human-subject experiment ($N = 379$) assessing how much people value generative AI-powered writing assistance financially and how such assistance influences people's writing experience and performance. During our experiment, participants were asked to write a 200–250 word article—either an argumentative essay or a creative story—within 45 minutes. We designed three writing modes in our experiment by varying the presence and type of AI assistance during participants' writing processes: (a) **Independent writing** (*Independent*), where participants completed the writing job on their own without receiving any AI assistance; (b) **Writing with editing assistance** (*Human-primary*), where participants took the primary responsibility of drafting the article while they received text editing and polishing assistance provided by ChatGPT, a LLM-based chatbot; and (c) **Collaborative writing with AI** (*AI-primary*), where ChatGPT took the primary responsibility of drafting the initial version of the article, while participants provided feedback and instructions to ChatGPT for improving the draft and made the final call to merge the content generated by ChatGPT and themselves. To estimate how much financial value that people attach to different kinds of writing assistance offered by LLMs, we created two experimental treatments by having participants make a forced binary choice between one job offer that pays them a fixed amount of \$3 to complete the writing job on their own (i.e., in the “*independent*” mode), and another offer that pays them a variable amount of $\$x$ ($1.5 \leq x \leq 4.5$) to complete the writing job with the AI assistance. The two experimental treatments differ on whether the AI assistance participants received in the second offer was editing assistance only (i.e., participants would be placed in the “*human-primary*” mode) or content generation assistance (i.e., participants would be placed in the “*AI-primary*” mode). Finally, after participants made the selection and completed the writing task in their selected writing modes, they were asked to fill out an exit survey to report their perceptions of their writing experience on various aspects, including their cognitive load, perceptions of the writing process (e.g., enjoyment) and writing outcome (e.g., uniqueness), perceived accountability, and confidence in completing future writing tasks.

Our experimental results suggest that people attach a noticeable level of financial value to generative AI-powered assistance in completing their writing tasks. For example, in the “*AI-primary*” writing mode wherein ChatGPT can provide content generation assistance, we found that participants in our study were willing to forgo \$0.85 to receive its assistance. To put it into context, this value amounts to 28.3% of the \$3 payment that participants would have received should they choose to write independently. When translating this value to participants' hourly wage, we found that an average participant was willing to give up an hourly wage of \$1.71 in order to receive ChatGPT's content generation assistance. Our data further indicates that people's value of generative AI is particularly salient when AI can offer content generation assistance instead of just editing assistance, and when the writing task has a higher demand for creativity (e.g., write a creative story). In addition, people's own characteristics also play a role in determining how much they value AI assistance. For instance, people with higher confidence in their

own writing abilities generally attach lower value to AI assistance than people with lower confidence. Upon examining the impacts of AI assistance on participants' writing experience perceptions and writing performance, we found that both editing assistance and content generation assistance provided by LLMs bring about significant benefits on people's productivity (e.g., reduced writing time and improved grammar) and self-efficacy (e.g., increased writing confidence), especially on argumentative essay writing. However, when LLMs directly provide content generation assistance (as in the “*AI-primary*” mode), these benefits also come at some cost. This includes people's decreased satisfaction with their writing experience and writing outcome (e.g., decreased ability to express oneself through writing), their reluctance to take responsibility for potential issues of the writing (e.g., the writing contains misinformation), and a decline of diversity across writings generated by different individuals. In this sense, there appears to exist a degree of “mis-match” between the financial value (i.e., the “price”) that people attach to different types of AI assistance in writing and the “true value” of these assistance.

Together, our study offers important experimental evidence regarding the value, benefits, and concerns of generative AI-powered assistance in human-AI co-writing. We conclude by discussing the design implications of our findings, and outline limitations and future work.

2 RELATED WORK

2.1 Interaction between humans and writing assistants

Research on understanding and designing interactions between humans and writing assistants has surged in recent years. Before the rise of LLMs, writing assistants primarily focus on enhancing human writing by offering features such as word-level suggestions [2, 17, 18, 37], sentence-level suggestions [11, 28], and novel text entry methods to optimize both speed and accuracy [4, 42]. However, some studies [3, 8, 14, 36] indicated that certain features, such as word suggestions, could occasionally clash with a user's typing habits and adversely affect their writing experience.

With the advent of LLMs [7] and their unparalleled capabilities in language understanding and generation [35], LLM-powered assistance is now more actively integrated into human writing workflows [34], potentially acting more like writing companion or coauthor than mere tools for word suggestions or corrections. As such, an increasing stream of research starts to investigate the range of assistance that LLMs could offer to human writers during the writing processes, such as tailored ideation based on current content [15, 45] and advanced text revisions beyond traditional word/grammar correction check [45]. This assistance could also tailor to diverse writing tasks including stories [12, 13, 29, 41, 43, 45, 46], slogans [13], argumentative essays [29], emails [9, 19], and script writing [33].

While early research highlights the advantages of integrating LLMs into the writing process, such as generating suggestions to unblock writers' creative thoughts, recent studies have begun to examine the potential negative impacts these AI writing assistants might have on humans during and post-writing process beyond those benefits. For instance, one study reveals that the sentiment

of text suggestions from LLMs can sway the sentiment of human-generated text [23]. Additionally, opinionated LLM-powered writing assistants can not only change the opinions expressed in human-written content but also subtly shift the human writers' own perspectives and beliefs [26]. In the context of online platforms, it is also found that AI-generated profiles on Airbnb may be viewed as less trustworthy [27].

2.2 Value estimation via “willingness to pay”

The conventional economic literature defines the maximum amount an individual would be willing to pay to acquire an amenity (e.g., the work flexibility, the work assistance) as their willingness to pay (WTP) for the amenity [32, 40], which can be intuitively comprehended as the value of the amenity to them. Researchers in the human-computer interaction community have previously worked on quantifying the value of various amenities [20, 24, 44]. According to the existing literature, the prevailing best practice for quantifying individuals' willingness to pay or accept for the amenity is through discrete choice experiments [22, 30, 32]. Mas and Palais [32] quantify the value of work flexibility offering each prospective worker two job options, differing in wage levels: one job offering flexible scheduling and the other an inflexible job. Hedegaard and Tyran [22] measure the value that workers attribute to working with individuals of the same ethnicity by inviting workers to choose between two job options: working alongside individuals of the same ethnicity and the other one with individuals from a different ethnicity. In the same vein, Liang et al. [30] estimate gig workers' willingness to accept a monitored job (i.e., the value they place on avoiding monitoring) by presenting them with a choice between a non-monitored job and an alternative job, the latter being subject to a randomly selected monitoring policy. Using similar methods, Holtz et al. [24] also estimate the value that freelancers associate with a positive review on online labor platforms. In this study, we will also quantify the value people attach to generative AI-powered writing assistance by estimating their willingness to pay for the assistance.

3 STUDY DESIGN

To understand how much people value AI assistance from LLMs (e.g., ChatGPT) in completing a writing job, and how the AI assistance impacts both people's perceptions about their writing experience and their final performance in the writing job, we conduct a randomized human-subject experiment.

3.1 Writing Task

Participants of our experiment were asked to write a 200–250 word article within 45 minutes. To examine how people's value of AI assistance in writing as well as the AI assistance's impacts on people's writing perceptions and performance may vary with the nature of the writing task, we considered two kinds of writing jobs in our experiment:

- **Argumentative essay writing:** Participants were provided with a statement, and they were asked to write an essay to discuss the argument in the statement. In their essays, participants had the freedom to either support or oppose the argument in the statement. We considered three statements

that were sampled from the pool of TOEFL writing exam topics:

- (1) “Some people think that if companies prohibit sending emails to staff on weekend or during other time out of office hours, staff's dissatisfactions with their companies will decrease. Others think this will not reduce the overall dissatisfactions among staff.”
 - (2) “Government should put higher tax on junk food to reduce consumption.”
 - (3) “Nowadays it is easier to maintain good health than it was in the past.”
- **Creative story writing:** Participants were given a prompt, and they were asked to write a story that includes the prompt. Drawing from the array of popular creative writing tasks featured on Reedsy's Short Story Contest¹, we curated three writing prompts:
 - (1) (Someone) “realizes they're on the wrong path.”
 - (2) Including the line “We're just too different.”
 - (3) Someone saying “Let's go for a walk.”

We focused on argumentative essay and creative story writing in this experiment because these two types of writing tasks are associated with different structures and purposes of writing, and potentially require different kinds of capability in writing [29]. Effective argumentative essays have a clear and specific thesis statement, and are based on credible and even persuasive evidence. In contrast, creative stories are meant to express one's experience and feelings, and can be more imaginary and loosely-structured while placing high demand on creativity. To ensure that the argumentative essay and creative story writing tasks chosen for our experiment have a comparable and reasonable difficulty level, we conducted a pilot study to test the difficulty of different argumentative essay statements and creative story prompts. In this pilot study, participants were asked to complete the writing task on their own. For the final set of 3 statements and 3 prompts selected for our experiment, as detailed above, our pilot study results suggested that participants could successfully complete the essay/story writing job within the time limit, yielding articles of satisfactory quality.

3.2 Experimental Treatments

In our experimental design, we evaluate three different writing modes, each reflecting varying degrees of assistance from and collaboration with a state-of-the-art LLM, i.e., ChatGPT. These modes encompass: (1) Independent writing without AI assistance (“Independent”), (2) Writing with editing assistance from ChatGPT (“Human-Primary”), and (3) Collaborative writing with ChatGPT (“AI-Primary”).

- **Independent writing (Independent):** In this writing mode, participants needed to complete the writing job independently without any assistance from ChatGPT.
- **Writing with editing assistance from ChatGPT (Human-Primary):** In this writing mode, participants were asked to take the primary responsibility for writing the article, while ChatGPT would be provided to assist them in editing and polishing their written content. Figure A1 in the Appendix shows an example of the writing interface for participants in

¹<https://blog.reedsy.com/creative-writing-prompts/terms/>.

this mode. Specifically, participants could send any sentences or paragraphs in their writing to ChatGPT for polishing (via the “Text Polishing Zone” in Figure A1). To ensure that ChatGPT would only provide editing assistance to participants, we covertly crafted a prompt employing the OpenAI API by appending the following instructions ahead of the text that participants sought to polish: “You should only edit or polish the texts I send to you. Please do not write any new content.” We then sent this formulated prompt to ChatGPT, and subsequently displayed the polished text returned by ChatGPT to the participants (in the “History Zone” in Figure A1). Participants would be then granted the autonomy to decide whether and how to incorporate the polished text into their writing.

- **Collaborative writing with ChatGPT (*AI-Primary*):** In this writing mode, ChatGPT would take the primary responsibility for writing the initial version of the article. Thereafter, participants could provide feedback to ChatGPT and engage in conversational interactions to steer the refinement of the article. Figure A2 in the Appendix shows an example of the writing interface for participants in this mode. Specifically, in accordance with the randomly assigned writing task, we initiated the process by directly sending a prompt aligned with the writing topic instruction (e.g., “Write an article for the statement ‘Government should put higher tax on junk food to reduce consumption.’” or “Write a story that includes someone saying ‘Let’s go for a walk.’”). This initial draft generated by ChatGPT was subsequently presented to participants (in the “Interaction History Zone” in Figure A2). Participants were then given the opportunity to offer feedback to ChatGPT, guiding it in refining the initial draft. This feedback would be directly relayed to ChatGPT, and the revised draft returned by ChatGPT would again be shown to participants. Through several iterative cycles, participants could collaborate with ChatGPT to further refine the content. In the final stage, participants could freely incorporate any part of the ChatGPT-generated texts with their own composition.

Recall that a key objective of our study is to quantify how much financial value that people attach to different kinds of writing assistance that generative AI-powered assistants could offer. This amount can be computed as the differences between the wage that is acceptable for people to complete a writing job independently and the wage that is acceptable for people to complete the writing job with various kinds of AI assistance of interests. Thus, following the classical methods in economics for estimating the “willingness to pay/accept” [30, 32, 40], we created two experimental treatments:

- **Independent vs. Human-Primary:** In this treatment, after the topic of the writing job (i.e., the statement for writing an argumentative essay or the prompt for writing a creative story) was revealed to participants, we presented participants with two job offers: The first offer paid the participant \$3 to complete the writing job in the *independent* writing mode, while the second offer paid the participant $\$x$ to complete the writing job in the *human-primary* mode, where x was randomly sampled from the set {1.5, 2, 2.25, 2.5, 2.75, 3, 3.25, 3.5,

3.75, 4, 4.5}. Participants were asked to make a selection between these two offers, and subsequently complete the writing job in accordance with the writing mode specified by the selected offer².

- **Independent vs. AI-Primary:** In this treatment, after the topic of the writing job was revealed to participants, we presented participants with two job offers: The first offer compensated the participant with \$3 for completing the writing job in the *independent* writing mode, while the second offer compensated the participant with $\$x$ for completing the writing job in the *AI-primary* mode, where x was again randomly sampled from the set of values as that in the previous treatment. Participants were asked to make a selection between these two offers, and would then complete the writing job in the writing mode specified by the offer that they chose.

Note that to ensure participants could make an informed selection between the two job offers, in both treatments, participants would be initially directed to watch a 2-minute video that introduces to them the type of writing assistance that they could receive from ChatGPT and acquaints them with the writing interface they would use upon selecting the job offer with AI assistance (i.e., the offers associated with *human-primary* or *AI-primary* writing modes). Participants could only make their job offer selection after finish watching this video.

3.3 Experimental Procedure

Our study was opened only to U.S. workers whose primary language is English on Prolific, and each worker was only allowed to participate in our study once. Each participant went through a few stages in our study, as detailed below.

Background assessment. Upon arrival of the study, participants were first asked to fill out a questionnaire to report their demographic information (e.g., gender, age, education). We then asked participants to indicate how confident they were in completing six types of writing tasks, including creative writing (e.g., stories, novels), writing argumentative essays, writing emails or letters, writing product or book reviews, writing business reports or proposals, and writing blogs. For each type of writing tasks, participants reported their confidence on a 5-point Likert scale from 1 (very low) to 5 (very high).

Treatment assignment and writing mode selection. Subsequently, participants were randomized into one of the two treatments, “*independent* vs. *human-primary*” or “*independent* vs. *AI-primary*”. They were then presented with their writing task, which could be either writing an argumentative essay or writing a creative story, and the statement/prompt used for the writing task was also selected at random from the candidate pool. Next, depending on the experimental treatment the participant was assigned, they would be presented with the corresponding two job offers, and the

²We conducted a pilot test to validate the appropriateness of the lower and upper bounds, \$1.5 and \$4.5, respectively, for estimating participants’ willingness to pay for AI assistance. This consideration was made given that a majority of participants in our pilot study would prefer not to choose the job with AI assistance at the \$1.5 wage level, while conversely, most participants would select the job with AI assistance at the \$4.5 wage level.

random payment value $\$x$ ($1.5 \leq x \leq 4.5$) of the offer that provided AI assistance to participants would be realized from its set of candidate values. Participants were told that their final payment from the study would consist of three parts: (1) a *base payment* of $\$2$; (2) the *writing job payment* as specified in the job offer that they would choose; and (3) an (optional) *performance-based payment* of $\$2$. We informed participants that their submitted articles would be sent to other crowd workers for review. If the average rating of their article would rank within the top 10% of the articles written for the same topic, they would receive the performance-based payment³. Once they were clear on the compensation structure, participants were required to watch an introductory video elucidating how they could potentially collaborate with ChatGPT through the designated interface to accomplish the writing task should they choose the AI-assisted writing mode (i.e., “*human-primary*” or “*AI-primary*”). With all this information, participants then selected their preferred job offer.

Main writing task. After the job offer was selected, participants proceeded to the main writing task. They were asked to complete this writing task using the writing mode specified in their chosen offer, and they had a maximum duration of 45 minutes for finishing the writing task. Note that the time required for ChatGPT to respond to participants’ prompts would not be included in the allocated time limit.

Exit survey. After completing the main writing task, participants were asked to complete an exit survey. In this survey, participants were again asked to indicate their confidence in completing the same six types of writing tasks (e.g., creative writing, argumentative essay, emails/letters, etc.) as we surveyed at the beginning of the study, should they have the chance in the future to complete those tasks *in the same writing mode* as they had experienced in our study. We also asked a series of survey questions to gauge participants’ perceptions of their writing experience in our study. For example, the NASA Task Load Index (NASA TLX) [21] was used to measure the cognitive load that participants experienced during the writing task, including their mental demand, time pressure, amount of effort taken, and frustration level. To understand participants’ perceptions of the overall writing processes, we presented the following statements to participants and asked them to rate how much they agreed with each statement on a 5-point Likert scale from 1 (strong disagree) to 5 (strongly agree):

- **(Satisfaction):** “I am satisfied with the writing process.”
- **(Enjoyment):** “I enjoy the writing process.”
- **(Ease):** “I find it easy to complete the writing process.”
- **(Ability of self-expression):** “I was able to express my creative goals during the writing process.”

Similarly, we asked participants to rate their agreement with the following statements, again on a 5-point Likert scale, to understand their perceptions of the final writing outcome:

- **(Quality):** “I am satisfied with the quality of the final article.”
- **(Ownership):** “I feel ownership over the final article.”
- **(Pride):** “I’m proud of the final article.”

³Pupon the completion of this study, we indeed recruited additional crowd workers from Prolific to evaluate the submitted articles. Their ratings were then used to determine whether each participant in this study would receive the performance-based payment.

- **(Uniqueness):** “The article I submitted feels unique.”

In addition, to understand participants’ accountability should their article be criticized for various issues during the evaluation process, we asked participants to rate their agreement in the following statements, on a 5-point Likert scale:

- **(Deceptive content):** “I’m willing to take the responsibility if my article is criticized for containing deceptive content (e.g., misinformation).”
- **(Plagiarism):** “I’m willing to take the responsibility if my article is criticized for containing content that is highly similar to someone else’s writing.”
- **(Privacy invasion):** “I’m willing to take the responsibility if my article is criticized for containing content that invades someone else’s privacy.”
- **(Discrimination):** “I’m willing to take the responsibility if my article is criticized as exhibiting bias and discrimination.”

Finally, participants reported their familiarity with ChatGPT (1: very unfamiliar; 5: very familiar) and their frequency of using ChatGPT in their daily life or work (1: never; 5: very frequently—more than once a day).

Attention check. To filter out inattentive participants, we included two attention check questions in our study. The first attention check question was presented to participants right before they took the exit survey, and it asked the participant to select again which job offer they had previously chosen in the study. The second attention check question was included in the exit survey, where participants were required to select a randomly pre-specified option in the question. We considered only the data from participants who passed both attention check questions as valid data.

3.4 Analysis Methods

Quantify the value of AI assistance. We measure the value of AI assistance to participants following the econometric methods for estimating “willingness to pay/accept” [30, 32, 40], that is, the maximum amount participants would be willing to pay to acquire different kinds of AI assistance (e.g., editing assistance in the “*human-primary*” writing mode, or content generation assistance in the “*AI-primary*” writing mode). Specifically, given the experimental data collected regarding participants’ choices between the independent writing mode and a randomly assigned AI-assisted writing mode, we started by fitting a Probit regression model to predict how likely participants would prefer the AI-assisted writing mode over the independent writing mode, considering the wage difference between these two modes:

$$\text{Pr}(\text{select AI assistance}=1) = \phi(\Delta w) \quad (1)$$

where Δw , defined by the equation $\Delta w = x - 3$, denotes the “*wage premium*” associated with the writing job offering AI assistance in comparison to the alternative independent writing job.

Utilizing this fitted model, we then identified the threshold of wage premium Δw^* such that $\phi(\Delta w^*) = 0.5$. In this scenario, participants would be indifferent between the independent writing job offer at a wage level of $\$3$ and the alternative job with AI assistance at a wage level of $\$(3 + \Delta w^*)$. We can then estimate the value of AI assistance to participants as $-\Delta w^*$, representing the maximum

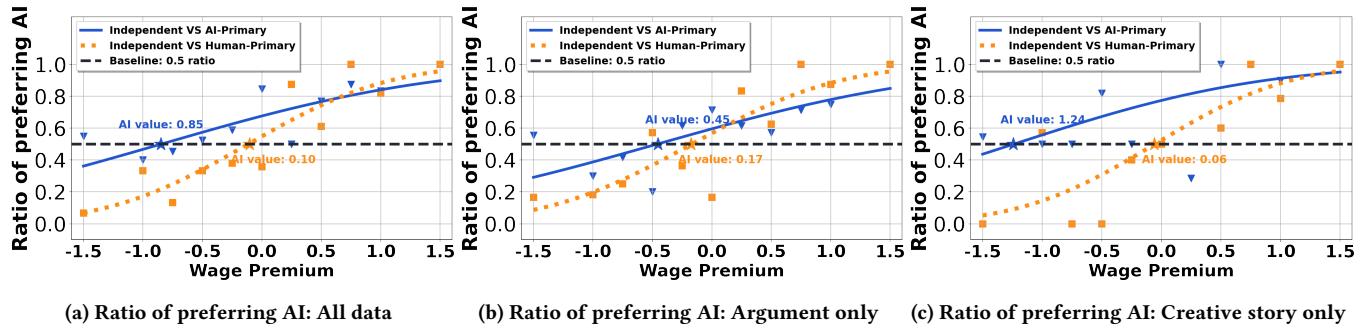


Figure 1: Comparing participants' probability of preferring an AI-assisted writing mode over the independent writing mode, (a) when writing either argumentative essays or creative stories, (b) when writing argumentative essays (arguments) only, or (c) when writing creative stories only. The wage premium is defined as the difference in the writing job payment participants would receive from selecting the AI-assisted writing mode, relative to the fixed payment of \$3 from the independent writing mode. Triangles (or squares) represent the actual fraction of participants who chose the AI-primary (or human-primary) writing mode over the independent writing mode in our dataset. The blue solid line (or the orange dotted line) depicts the predicted probability, derived from the fitted Probit model, of participants opting for the AI-primary (or human-primary) writing mode over the independent writing mode. The AI assistance's value is the negative of the wage premium for the AI-assisted job that makes participants indifferent between the AI-assisted mode and the independent writing mode (i.e., choosing the AI-assisted writing mode with a probability of 0.5), which is marked by pentagrams.

amount of money they are willing to forgo in order to acquire the AI assistance in completing their writing jobs⁴.

Compare writing perceptions and performance across writing modes. To understand how the AI assistance impacts participants' perceptions of the writing experience and their writing performance, we conducted regression analysis using our experimental data. In particular, the primary independent variable of the regressions was the writing mode selected by the participant. The dependent variables included a set of writing perception metrics that we obtained from participants' self-reports in the exit survey (e.g., participants' cognitive load, perceptions of the writing processes, perceptions of the final writing outcome, changes in writing confidence, perceptions of accountability), as well as a set of objective metrics of participants' writing performance (e.g., time spent, number of grammatical and spelling mistakes made, coherence of the article). To minimize the impact of potential confounding variables, we also accounted for a set of covariates in the regressions, including the participants' demographic background (age, gender, race, education), the writing job payment they received from their selected job offer, their confidence in writing the type of articles that they were asked to work on in the study, their familiarity with ChatGPT, and their frequency of using ChatGPT. Note that in our analysis, some key concepts were measured through multiple dependent variables (e.g., participants' perceptions of the final writing outcome was measured with respect to their perceived quality, ownership, pride, and uniqueness of the final article); in this case, we considered these dependent variables to belong to the same family and applied Bonferroni corrections to correct the p values for multiple comparisons. For clarity, we used *adjusted-p* in

the following to indicate the corrected p -values whenever Bonferroni corrections are applied. Finally, for visualization purposes, we plotted the predicted values of dependent variables for a participant with average characteristics at an average wage premium across three different writing modes.

4 RESULTS

In total, 379 workers from Prolific took our study and passed the attention check. Among them, 183 were allocated to the “*independent vs. human-primary*” treatment, while the remaining 196 were placed in the “*independent vs. AI-Primary*” treatment⁵. The average hourly payment workers received in our study was \$13.64. In the following, we quantify the value participants place on AI assistance and analyze the impact of AI assistance on people's perception and performance in writing based on the experimental data we obtained from these workers.

4.1 Estimating the value of AI assistance

People are willing to forgo financial payments to receive AI assistance in writing, and they value AI assistance more when AI can provide assistance in generating writing content. Figure 1a displays the fitted Probit regression models for predicting the likelihood of participants preferring a writing mode with AI assistance over the independent writing mode at the specified wage premium, when we do *not* differentiate the assigned type of writing jobs (i.e., argumentative essay writing and creative story writing). Clearly, we find that people attach a positive financial value to the writing assistance provided by ChatGPT, the generative AI-powered assistant. Moreover, the value they attach to AI assistance is particularly salient when the AI model can directly generate content for them (as that in the “*AI-primary*” writing mode) beyond performing just

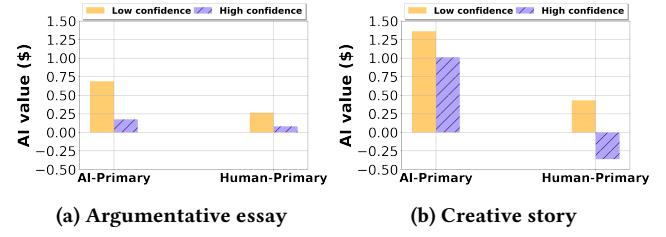
⁴In other words, compared to accepting the independent writing job offer at a wage level of \$3, participants are equally willing to adopt the AI assistance at a wage level of \$(3+ Δw^*), and “pay” \$(− Δw^*) to the provision of the AI assistance. That is, participants' willingness to pay for the AI assistance is \$(− Δw^*).

⁵See Appendix B for the demographic statistics of participants in our study.

text editing and polishing (as that in the “*human-primary*” writing mode)—Indeed, our data suggests that participants in our study were willing to forgo \$0.85 to receive ChatGPT’s content generation assistance in the “*AI-primary*” writing mode. This was 28.3% of the writing job payment that they would receive should they select the independent writing mode, and was much higher than the \$0.10 that they were willing to forgo to receive ChatGPT’s editing assistance in the “*human-primary*” mode.

To better contextualize the magnitude of the value that people attach to the writing assistance provided by ChatGPT, we translate the value into the amount of hourly payment participants were willing to give up in order to receive the AI assistance. To do so, we used our experimental data to learn a linear regression model to predict the amount of time participants would spend on completing a writing job, given the writing mode they chose as well as the amount of writing job payment they received from their chosen job offer. Our regression model also controlled for several covariates, including the participants’ demographic background (e.g., age, gender, race, education), their confidence in the type of writing task that they were asked to complete, their familiarity with ChatGPT, and the frequency of their ChatGPT usage. Given this prediction model, we found that an average participant is predicted to spend 13.7 minutes in completing a writing job in the “*independent*” mode to get the \$3 fixed writing job payment, resulting in an hourly wage of \$13.12. Meanwhile, when an average participant was willing to forgo \$0.85 to adopt ChatGPT’s content generation assistance in the “*AI-primary*” mode, they are predicted to spend 11.3 minutes in completing a writing job to get the writing job payment of \$2.15 (i.e., $\$3 - \$0.85 = \$2.15$), leading to an hourly payment of \$11.41. In other words, the value of \$0.85 that participants in our study attached to ChatGPT’s content generation assistance can be translated to an hourly wage of \$1.71 (i.e., $\$13.12 - \$11.41 = \$1.71$). As participants’ average hourly wage in completing our entire study (including the writing job and surveys) was \$13.64, the \$1.71 hourly wage that they were willing to give up to receive ChatGPT’s content generation assistance represents 12.5% of their hourly payment, which is non-trivial. Following a similar computation, we also found that the \$0.10 participants were willing to forgo to receive ChatGPT’s editing assistance in the “*human-primary*” mode is effectively equivalent to giving up an hourly wage of \$0.42.

The value of AI’s editing assistance does not vary much with the nature of the writing job, but the value of AI’s content generation assistance does. We then investigated into if people’s value of AI assistance in writing differs across different types of writing tasks. Figure 1b and Figure 1c show the values that participants attributed to AI assistance when tasked with writing an argumentative essay or a creative story, respectively. These values are derived from the Probit regression model as outlined previously. Here, we observed that in the *human-primary* writing mode, where ChatGPT was configured to exclusively polish the text generated by participants, the value participants attached to the AI assistance remained relatively stable as the type of writing jobs changed (i.e., the AI’s editing assistance worth \$0.17 in the context of argumentative essay writing jobs and \$0.06 in the context of creative story writing jobs). In contrast, in the *AI-primary* writing mode, participants placed a much higher value on ChatGPT’s content generation



(a) Argumentative essay (b) Creative story

Figure 2: Comparing the estimated value of AI assistance between participants with varying levels of writing confidence for different types of writing tasks.

assistance when they were asked to write creative stories instead of argumentative essays (\$1.24 for creative stories and \$0.45 for argumentative essays). Regardless of the type of the writing tasks, we still found that people tend to value AI’s content generation assistance more than its editing assistance.

Writers with higher confidence generally attach lower value to AI assistance. To understand whether people with different characteristics attach different values to AI assistance, we estimated the value of AI assistance for different subsets of people separately. First, we looked into the value of AI assistance for people with different levels of confidence in their writing skills. Specifically, for the set of participants who were asked to complete an argumentative essay writing job (or a creative story writing job), we divided them into two subsets based on a median split of participants’ confidence in writing argumentative essay (or creative writing) that they self-reported at the beginning of our study. For each subset of participants, we used the Probit regression model to estimate their value for AI assistance. Figures 2a and 2b compare the average estimates of the AI assistance’s value between participants with high and low confidence in completing the assigned type of writing tasks, for argumentative essay writing and creative story writing, respectively. We found that across both types of writing tasks, people with high confidence in their own writing ability value the AI assistance less. For example, participants who were more confident in their own argument-writing ability attached substantially less value to ChatGPT’s content generation assistance (\$0.17 for high-confident people vs. \$0.68 for low-confident people) and editing assistance (\$0.08 for high-confident people vs. \$0.26 for low-confident people) for writing argumentative essays. Similarly, participants who have high confidence in their own creative-writing ability also valued ChatGPT’s content generation assistance and editing assistance less in writing creative stories (content generation assistance: \$1.01 for high-confident people vs. \$1.36 for low-confident people; editing assistance: -\$0.36 for high-confident people vs. \$0.43 for low-confident people); in fact, they even attached a negative value to ChatGPT’s editing assistance in writing creative stories.

Writers who are more familiar with ChatGPT attach less value to its assistance in writing argumentative essays, but consider its editing assistance in creative writing to be more valuable. Finally, we divided participants into two subsets based on a median split of their self-reported level of familiarity with ChatGPT at the end of our study. The average estimated values of different types of AI assistance for people who have high or low level of familiarity with ChatGPT are presented in Figure 3. We noticed

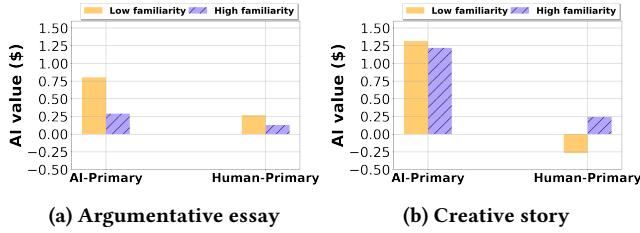


Figure 3: Comparing the estimated value of AI assistance between participants with varying levels of familiarity with ChatGPT for different types of writing tasks.

distinct patterns in how familiarity impacts people's perceived value of AI assistance across the two types of writing tasks in our study. For the argumentative essay writing task, regardless of whether ChatGPT provides content generation assistance or only provides editing assistance, those who were more familiar with ChatGPT always considered the value of ChatGPT's assistance to be lower than those who were less familiar with ChatGPT (\$0.12 for high-familiarity people vs. \$0.27 for low-familiarity people in AI editing assistance, and \$0.29 for high-familiarity people vs. \$0.80 for low-familiarity people in AI content generation assistance). On the other hand, for the creative story writing task, the value of ChatGPT's content generation assistance was similar across participants with different familiarity levels with ChatGPT (\$1.22 for high-familiarity people vs. \$1.31 for low-familiarity people in AI editing assistance). However, participants with higher familiarity with ChatGPT were found to attach a greater value to ChatGPT's editing assistance in creative writing (\$0.24 for high-familiarity people vs. -\$0.27 for low-familiarity people).

4.2 Understanding the impacts of AI assistance in people's perceptions of the writing experience

Next, we move on to examine how the incorporation of AI assistance in the writing processes changes people's perceptions of their writing experience, including their perceived cognitive load, their perceptions about the overall writing processes, their perceptions about the final writing outcome, changes in their writing confidence, and their accountability perceptions. As discussed in Section 3.4, these analyses are based on regression models we fit from the data. The complete set of results of the fitted regression models can be found in Appendix C.

4.2.1 Cognitive load. We computed the overall level of cognitive load that each participant experienced in completing the writing job in our study by averaging their self-reported levels of mental demand, time pressure, amount of effort taken, and frustration level, which were rated on a scale from 1 (very low) to 5 (very high). We then fitted a linear regression model to predict participants' cognitive load under the three writing modes. Figure 4 illustrates the predicted cognitive load for a worker with average characteristics across various writing modes, when tasked with writing an argumentative essay or a creative story.

The provision of AI assistance in writing leads to significant cognitive load decreases in argumentative essay writing but

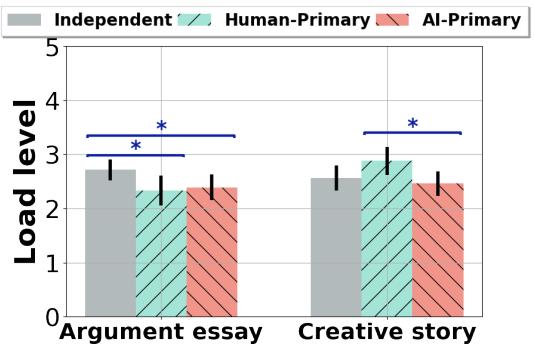


Figure 4: Comparing the predicted cognitive load that the average participant experienced for completing the two types of writing tasks (argumentative essay writing and creative story writing), across the *independent*, *human-primary*, and *AI-primary* writing modes. Error bars represent the 95% confidence intervals of the predicted cognitive load level. * denotes statistical significance levels of 0.05.

not creative story writing. As illustrated in Figure 4, for the argumentative essay writing task, participants who worked independently generally exhibited higher cognitive loads compared to those in the "human-primary" writing mode ($p = 0.032$) and those in the "AI-primary" writing mode ($p = 0.042$). This implies that one of the benefits brought up by AI assistance is that it decreases people's perceived cognitive load in writing argumentative essays. However, we did not observe similar benefits from AI assistance for the creative story writing tasks—among those tasked with writing creative stories, there was no significant difference between the level of cognitive load perceived by participants who completed the task independently and those who received assistance from ChatGPT. In fact, those in the "human-primary" mode who received editing assistance from ChatGPT actually reported the highest level of cognitive load, which was significantly higher than that reported by those in the "AI-primary" writing mode ($p = 0.022$).

4.2.2 Perceptions about the writing process. We then examined how participants' perceptions of the overall writing process vary across different writing modes. In Figure 5, we compared participants' perceptions on four aspects of the writing processes—their satisfaction with the writing processes, the enjoyment they experienced during writing, the ease of completing the writing processes, and their perceived ability to express themselves during writing. As discussed in Section 3.4, these four dependent variables were considered to belong to the same family, thus Bonferroni correction was applied when examining the statistical significance of the treatment effects.

In argumentative essay writing, the provision of content generation assistance by AI significantly decreases people's enjoyment of the writing processes, and it also limits people's ability to express their creative goals. A visual inspection of Figure 5 suggests a highly consistent pattern—ChatGPT's editing assistance in the *human-primary* mode almost always resulted in participants' highest levels of ratings on various aspects of the writing processes, while ChatGPT's content generation assistance in the *AI-primary* mode often led to the lowest levels of ratings.

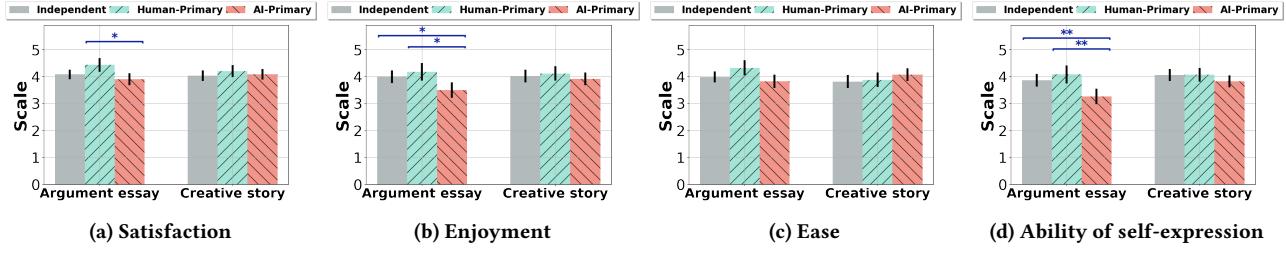


Figure 5: Comparing the predicted perceptions of the average participant regarding the overall writing process for completing the two types of writing tasks (argumentative essay writing and creative story writing), across the *independent*, *human-primary*, and *AI-primary* writing modes. Error bars represent the 95% confidence intervals. * and ** denote statistical significance levels of 0.05 and 0.01, respectively.

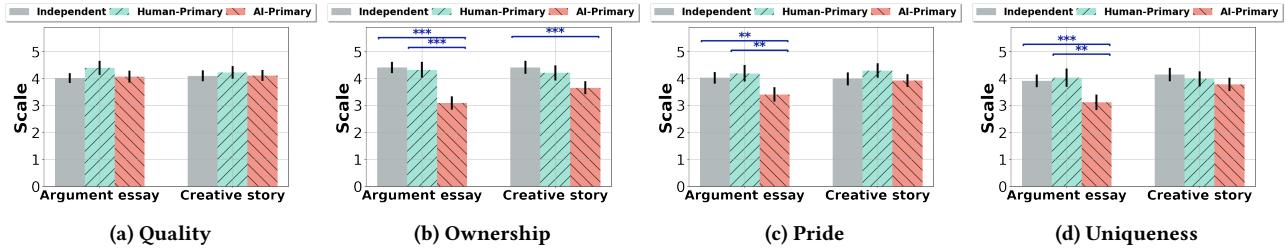


Figure 6: Comparing participants' predicted perceptions of the final writing outcome for two types of writing tasks (argumentative essay writing and creative story writing), across the *independent*, *human-primary*, and *AI-primary* writing modes. Error bars represent the 95% confidence intervals. ** and * denote statistical significance levels of 0.01 and 0.001 respectively.**

These differences were particularly salient when we examine participants' perceptions of the degree to which they enjoyed the writing processes and their ability to express their creative goals while composing argumentative essays. For example, we found that when writing argumentative essays, participants who received ChatGPT's content generation assistance in the *AI-primary* mode reported a significantly lower level of enjoyment during the writing processes compared to those who wrote independently ($adjusted-p = 0.037$) or those who only received editing assistance from ChatGPT in the *human-primary* mode ($adjusted-p = 0.044$). Similarly, participants who wrote their argumentative essays in the *AI-primary* mode also reported a significantly lower level of ability to express their creative goals compared to those in the *independent* ($adjusted-p = 0.007$) or *human-primary* writing modes ($adjusted-p = 0.006$). An additional significant difference that we detected was that participants in the *AI-primary* writing mode expressed significantly lower satisfaction with the writing processes than participants in the *human-primary* writing mode ($adjusted-p = 0.042$).

4.2.3 Perceptions of the final writing outcome. Figure 6 compares participants' perceptions of four aspects of the final writing outcome—their perceived quality of the article, their perceived ownership of the article, their perceived pride in the article, and their perceived uniqueness of the article, upon completing the writing tasks in different writing modes. Bonferroni corrections were implemented to account for multiple statistical tests in our analysis of this family of dependent variables. Visually, it appears that participants in the *AI-primary* writing mode often reported the lowest level of ratings across various aspects of the final writing outcome. In particular,

consistent with previous findings [5], we find that ***the provision of content generation assistance by AI significantly decreases people's perceived ownership of the final writing outcome*** (see Figure 6b; $adjusted-p < 0.001$ for the comparisons between *independent* and *AI-primary*). Moreover, when participants worked on argumentative essay writing tasks, we also find that ***the provision of content generation assistance by AI significantly decreases how much they take pride in the final writing outcome and how much they consider it as unique*** (see Figure 6c and 6d; $adjusted-p < 0.01$ for the comparisons between *independent* and *AI-primary*). For more detailed results, see Appendix D.1.

4.2.4 Change of confidence in writing. Next, we look into whether AI assistance can increase participant's writing confidence for future writing tasks, after they experienced generative AI-powered assistance in writing. Recall that we asked participants to report their confidence in six types of writing tasks (e.g., argumentative essay, creative novel or story, email or letter, etc.) at the beginning of our study. In addition, after they had finished their writing job in their selected writing mode, participants again reported their confidence in these six types of writing tasks assuming that they would work on them in the future in the same writing mode as they just experienced. For each type of writing tasks, we were interested in the *change* in writing confidence (or "*confidence boost*") after participants tried out different writing modes (i.e., *independent*, *human-primary*, *AI-primary*) on different writing tasks (i.e., argumentative essay writing or creative story writing). Figures 7a–7c compare participants' confidence boost in completing future argumentative essay writing, creative story writing, and the other four

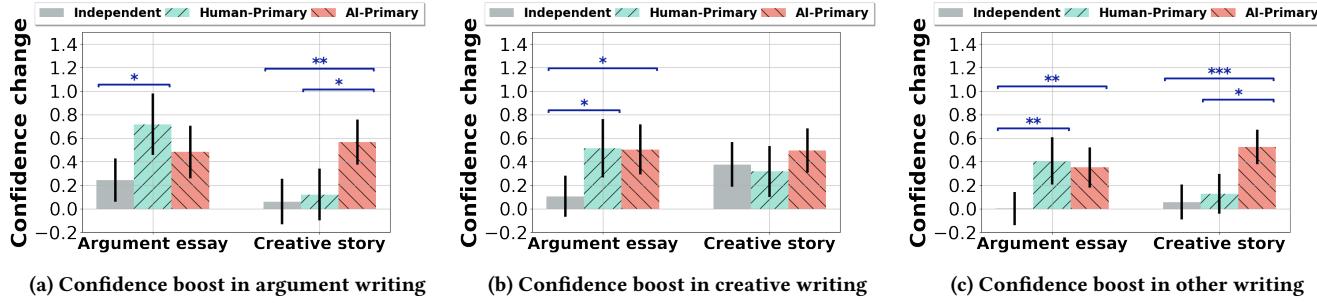


Figure 7: Comparing the predicted change in the confidence of an average participant in writing (a) argumentative essays, (b) creative novel or stories, and (c) other types of articles (e.g., emails/letters, product/book reviews, business reports/proposals, blogs), should they have future opportunities to complete these writing tasks in the same writing mode as they experienced in our study, after they completed argumentative essay or creative story writing in our study. Error bars represent the 95% confidence intervals. *, **, and * denote statistical significance levels of 0.05, 0.01, and 0.001 respectively.**

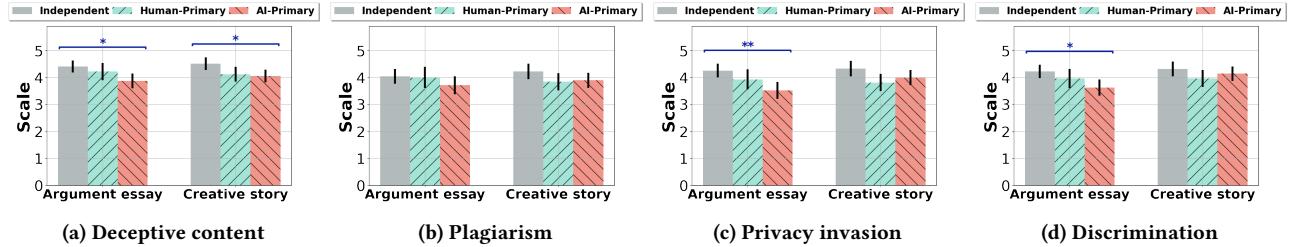


Figure 8: Comparing the predicted willingness of an average participant to take responsibility for potential issues in their articles across the independent, human-primary, and AI-primary modes after completing argumentative essay or creative story writing tasks. Error bars represent the 95% confidence intervals. * and ** denote statistical significance levels of 0.05 and 0.01 respectively.

types of writing tasks⁶, respectively, after they used different writing modes in our study to write argumentative essays or creative stories.

After writing argumentative essays with AI assistance, people generally have higher confidence in completing all types of writing tasks in the future. When focusing on participants who completed argumentative essay writing in our study, we found that the provision of ChatGPT's writing assistance to them—both its editing assistance and content generation assistance—generally made them become more confident in completing future writing tasks. Interestingly, this confidence increase often occurs not only for completing future argumentative essay writing tasks, but also for completing all types of other writing tasks. For example, compared to those who wrote the argumentative essay independently, participants in the *human-primary* mode reported a significantly larger increase in their confidence in completing future argumentative essay writing ($adjusted-p = 0.014$), creative writing ($adjusted-p = 0.030$), and other writing tasks ($adjusted-p = 0.005$). Similarly, participants in the *AI-primary* mode also had a significantly larger confidence boost in completing future creative writing ($adjusted-p = 0.016$) and other writing tasks ($adjusted-p = 0.007$).

After writing creative stories with AI assistance, only if AI provides content generation assistance will people increase their confidence in writing, but not for creative writing. In contrast, when focusing on participants who completed creative story writing in our study, we found less evidence of boost of writing confidence brought up by AI assistance. Participants in the *human-primary* mode who only received editing assistance from ChatGPT in writing creative stories did not show any significant difference than those who wrote the stories on their own in terms of their confidence in completing any type of future writing tasks. On the other hand, the content generation assistance of ChatGPT did make participants who wrote creative stories in the *AI-primary* mode report significantly larger increases in their writing confidence than independent writers, but only for tasks other than creative writing (for future argumentative essay writing: $adjusted-p = 0.001$; for other writing tasks: $adjusted-p < 0.001$).

4.2.5 Perceptions of accountability. Finally, we examined how writing with AI assistance influence people's willingness to take responsibility if their final articles face criticism for issues such as deceptive content, plagiarism, privacy invasion, and discrimination. Figure 8 compares the willingness to assume responsibility for these issues across participants in the three writing modes.

People who receive content generation assistance from AI are less willing to take responsibility for criticisms of the final

⁶We calculated the pooled average of the confidence increase observed across the four types of writing tasks other than argumentative essay writing and creative story writing (i.e., emails/letters, product/book reviews, business reports/proposals, blogs).

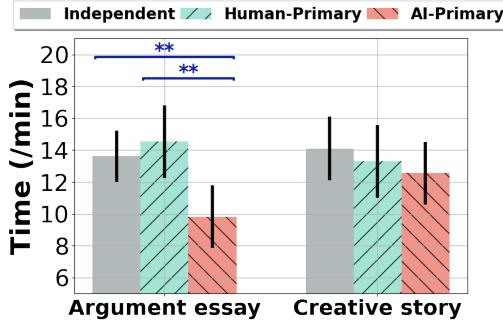


Figure 9: Comparing the predicted time taken by participants to complete the two types of writing tasks (argumentative essay and creative story writing), across the *independent*, *human-primary*, and *AI-primary* writing modes. Error bars represent the 95% confidence intervals. ** denote statistical significance levels of 0.01, respectively.

writing outcome. Visually, we note that when people receive AI assistance in their writing, they are generally less willing to take responsibility on all four potential criticisms of the final writing outcome. This trend is particularly salient when AI provides content generation assistance during the writing processes. For example, we found that after participants completed argumentative essay writing in the *AI-primary* mode, they were less likely to take responsibility for criticisms on potential deceptive content (*adjusted-p* = 0.015), privacy invasion (*adjusted-p* = 0.002), or bias/discrimination issues (*adjusted-p* = 0.014) in their final articles, compared to participants who wrote independently without AI assistance. In addition, after writing creative stories in the *AI-primary* mode, participants also became less willing to take the responsibility for any criticism of deceptive content in their final articles than the participants in the *independent* writing mode (*adjusted-p* = 0.028).

4.3 Understanding the impacts of AI assistance on people’s writing performance

Lastly, we investigate into how the incorporation of AI assistance in the writing processes affect people’s writing performance. In this analysis, we focus on a set of objective metrics to quantify people’s writing performance from different perspectives. Unless specified otherwise, this analysis is conducted based on regression models that we fit from the data, and the results of the full set of regression models can be found in Appendix C.

4.3.1 Completion time. We start by examining people’s time efficiency in completing the writing job. Therefore, we compare how much time it took for participants to complete the main writing task in our study across participants in the three writing modes, and results are presented in Figure 9.

Content generation assistance provided by AI decreases the amount of time it takes for people to complete argumentative essay writing. As shown in Figure 9, while editing assistance provided by ChatGPT does not appear to significantly influence participants’ efficiency in completing their writing job, the provision of content generation assistance by ChatGPT appears to

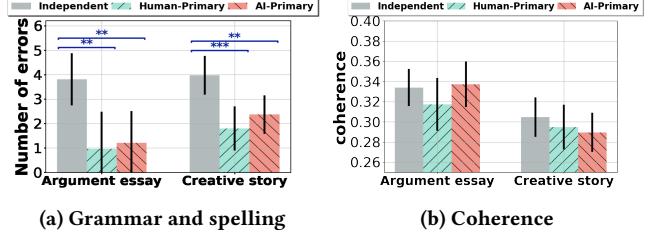


Figure 10: Comparing the predicted number of grammar and spelling mistakes in the final article and the coherence of the article in two types of writing tasks (argumentative essay and creative story writing) for an average participant, across the *independent*, *human-primary*, and *AI-primary* writing modes. Error bars represent the 95% confidence intervals. ** and *** denote statistical significance levels of 0.01 and 0.001, respectively.

increase participants’ writing efficiency. Specifically, we found that when writing argumentative essays, participants in the *AI-primary* writing mode took significantly less time to complete the writing compared to those working independently ($p = 0.004$), and those in the *human-primary* writing mode ($p = 0.004$).

4.3.2 Grammar and spelling. To see if the AI assistance affects people’s writing performance with respect to the proper usage of language in their final article, we utilized LanguageTool⁷, an open-sourced grammar, style, and spell checker, to conduct a grammar and spelling error analysis on the final articles submitted by participants in different writing modes. Figure 10a compares the predicted average number of grammar and spelling mistakes per article when participants completed argumentative essay or creative story writing tasks across the three writing modes. Not surprisingly, we find that **both AI’s editing assistance and content generation assistance leads to significantly decreased number of grammar and spelling mistakes in people’s writing** ($p < 0.01$ for all comparisons between independent and an AI-assisted writing mode). For more detailed results, see Appendix D.2.

4.3.3 Content coherence. Another aspect of writing performance we measured was the coherence of the content in participants’ final articles. Arguably, an article is of higher quality if it contains multiple sets of more coherent sentences, such that sentences in each set address the same topic and could support one another. To compute the “coherence score” of an article, following previous work [39], we first used Latent Dirichlet Allocation (LDA) [6]—a common topic modelling algorithm—to identify latent topics in the article, and then computed the semantic similarity between highly-scored words within each topic based on word co-occurrence probabilities. Intuitively, the higher the coherence score, the better. Figure 10b presents the predicted average coherence score of articles that participants produced across the three writing modes. Here, we do not find any evidence suggesting that the incorporation of AI assistance during the writing processes significantly changes the coherence of people’s final writing outcomes.

⁷<https://languagetool.org/>.

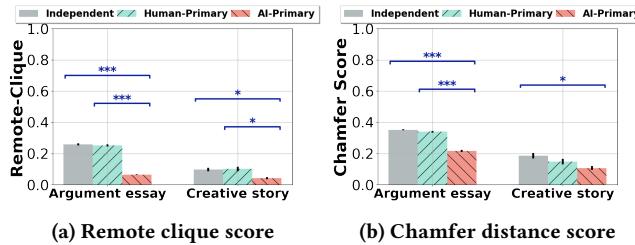


Figure 11: Comparing the diversity of final articles on the same topic for two types of writing tasks (argumentative essay and creative story writing), across the *independent*, *human-primary*, and *AI-primary* writing modes. Error bars represent standard errors of the mean. * and * denote statistical significance levels of 0.05 and 0.001, respectively.**

4.3.4 Content diversity. Finally, we examine how the incorporation of AI assistance changes the diversity of writings generated by different individuals on the same topic. One can also consider this metric as a proxy of the “uniqueness” of the writing—a lower level of diversity across multiple writings on the same topic usually implies that each writing in the group is less unique. For our study, given a set of articles submitted by participants on the same topic (i.e., argumentative essays for the same statement or creative stories for the same prompt) using the same writing mode, we used *remote clique score* and *Chamfer distance score* to quantify the diversity of these articles following the prior literature [38]. Specifically, the remote clique score is computed as the average mean distance of an article to other articles in the set, while the Chamfer distance score is defined as the average minimum distance of an article to other articles in the set. For both metrics, higher values indicate greater diversity. To determine the distance between different articles, we utilized a pre-trained Sentence Transformer model [25] to convert each article into a vector embedding and then computed the cosine distance between each pair of articles. Figure 11 compares the diversity of articles written on the same topics across the three writing modes. Since our diversity metrics are computed for each writing topic instead of for each participant, in the following, we did not fit regression models to analyze the data. Instead, one-way ANOVA was used to determine if the mean value of diversity metrics was statistically the same across the three writing modes, and post-hoc Tukey HSD tests were used to detect significant pairwise comparisons.

AI's content generation assistance significantly reduces the diversity of people's writing. Visually, it is clear that when participants received content generation assistance from ChatGPT, the diversity of the articles they produced on a given topic is consistently lower than other participants, irrespective of the specific writing tasks they undertook. Specifically, when writing argumentative essays, the diversity of articles generated from the *AI-primary* writing mode was significantly lower compared to those generated from the *independent* or *human-primary* modes, which was evident in both the *remote clique score* and *Chamfer distance score* metrics ($p < 0.001$ for all comparisons). Similarly, when writing creative stories, the diversity of stories produced from the *AI-primary* writing mode was also significantly lower compared to those produced

from the *independent* writing mode (remote clique score: $p = 0.021$; Chamfer distance score: $p = 0.043$) and the *human-primary* mode (remote clique score: $p = 0.042$).

5 DISCUSSIONS

Table 1 summarizes the key findings of our study. Below, we provide a few explanations to some of our findings, discuss design implications of our results, and outline limitations and future work.

5.1 The moderating role of writing task nature

As shown in Table 1, both people's value of AI assistance and the AI assistance's impacts on people's writing experience and performance are moderated by the nature of the writing tasks to some extent. For example, participants generally attached a higher value to the content generation assistance provided by LLMs for creative story writing than for argumentative essay writing. In addition, the AI assistance appears to mostly bring about productivity benefits on argumentative essay writing tasks—it reduces participants' perceived cognitive load and decreases their completion time in writing only when they wrote argumentative essays but not creative stories. We conjecture that these differences can be attributed to the distinct nature and objectives of these two writing tasks.

For example, for argumentative essay writing, the main cognitive load and the most time-consuming part for writers is to construct the logical structure of the essay and identify the appropriate evidence. Generative AI models can easily provide such information and generate reasonably logical sentences for writers, thus can greatly boost writers' productivity. However, for creative story writing, the primary cognitive burden for writers lies in the process of utilizing their imagination to create compelling characters and engaging plots, and they often need to intentionally weave conflicts and surprises into the storyline. Compared to argumentative essay writing, creative story writing places higher demand on creativity and originality, and it may lack the kind of “templates” that often turn out to be effective for writing argumentative essays. In this sense, it is not surprising that people tend to attach higher value to generative AI-powered content generation assistance for creative writing than for argumentative essay writing, because such assistance is especially useful to inspire the writers and help them think “out of the box” in creative writing. Nevertheless, the fact that AI assistance does not significantly affect people's cognitive load and writing time in creative story writing suggests that today's LLMs may still fall short in providing the kind of assistance that people need and expect in creative writing tasks. For instance, LLMs may still be struggling on creating truly novel plots or appropriately adjust the content they generate to align with writers' feedback.

5.2 Potential mismatch between human attached value to AI assistance and the “true value” of AI assistance

Another somewhat puzzling finding of our study is that considering the different types of writing assistance provided by generative AI, while people are consistently willing to pay a higher price to receive the content generation assistance from LLMs than the editing assistance, the former type of assistance actually brings about a much larger set of undesirable impacts on people's writing

Dependent Variable		Argumentative essay		Creative story	
		Human-Primary	AI-Primary	Human-Primary	AI-Primary
Value of AI assistance	All population	\$0.17	\$0.45	\$0.06	\$1.24
	Low writing confidence	\$0.26	\$0.68	\$0.43	\$1.36
	High writing confidence	\$0.08	\$0.17	-\$0.36	\$1.01
	Low familiarity with ChatGPT	\$0.27	\$0.80	-\$0.27	\$1.31
	High familiarity with ChatGPT	\$0.12	\$0.29	\$0.24	\$1.22
Cognitive load	NASA Task Load Index	↓	↓	-	-
Perceptions about the writing process	Satisfaction	-	-	-	-
	Enjoyment	-	↓	-	-
	Ease	-	-	-	-
	Ability of self-expression	-	↓	-	-
Perceptions about the final outcome	Quality	-	-	-	-
	Ownership	-	↓	-	↓
	Pride	-	↓	-	-
	Uniqueness	-	↓	-	-
Change of confidence	Argument writing	↑	-	-	↑
	Creative writing	↑	↑	-	-
	Other writing	↑	↑	-	↑
Perceptions of accountability	Deceptive content	-	↓	-	↓
	Plagiarism	-	-	-	-
	Privacy invasion	-	↓	-	-
	Discrimination	-	↓	-	-
Writing performance	Completion time	-	↓	-	-
	Grammar and spelling	↓	↓	↓	↓
	Content coherence	-	-	-	-
	Diversity: Remote clique score	-	↓	-	↓
	Diversity: Chamfer distance score	-	↓	-	↓

Table 1: A summary of the key findings of this study. In the “Value of AI assistance” row, the average value different subsets of people attach to AI assistance is reported. In other rows, we compare the dependent variables measured in both the *human-primary* and *AI-primary* modes with those obtained in the *independent* mode. The symbols ↓ and ↑ indicate significantly lower and higher values, respectively, in comparison to the *independent* mode. – denotes no statistical significance was detected. ↓(↑) denotes desirable changes while ↓(↑) denotes undesirable changes.

experience and performance than the latter (e.g., all red arrows, representing undesirable impacts, are in “*AI-primary*” columns in Table 1). This seems to indicate a degree of mismatch between the price people pay for AI assistance, and the “true value” of AI assistance. We provide a few plausible explanations for this finding.

One possible explanation is the media portrayal of generative AI models like ChatGPT as revolutionary and next-generation tools

has potentially inflated people’s expectations. The general public might not have a deep understanding of how these AI models operate, leading to assumptions that the generative AI-powered writing assistant tools are more capable than they truly are, especially on their content generation and adaptation capabilities that is new compared to traditional writing assistant tools. This limited understanding, coupled with intensive media coverage, might

create an overly optimistic perception of AI's capabilities. Furthermore, most people may not recognize that LLMs are trained on extensive datasets and might, at times, generate content that does not perfectly meet their specific needs, intentions, or style. These deviations between expectation and reality might contribute to the possibly overly high financial value that people attach to content generation assistance of LLMs. In fact, this conjecture can be supported by one interesting observation in our study—After participants experienced content generation assistance of LLMs in argumentative essay writing (or creative story writing) tasks, their confidence boost in completing future argumentative essay writing (or creative story writing) tasks are *not* significantly different from those participants who completed the writing tasks independently, although their confidence boost in other types of writing tasks was significantly larger. This implies that upon directly experiencing the content generation capabilities of the generative AI-powered assistant tools and encountering their limitations, people may recalibrate their assessment of the AI assistance's true value, at least on the type of writing tasks they have worked on. In fact, the contrast between people's initial high expectation of AI's content generation assistance and the practical output they observe may even inadvertently amplify their hope that AI's content generation capabilities might truly shine in other unexplored tasks, which explains their writing confidence boost in other types of writing tasks.

Another possible explanation is that the financial value people attach to different types of AI assistance may be primarily driven by their productivity-related considerations. Indeed, people may expect the incorporation of AI-powered content generation assistance in their writing workflow to significantly increase their time efficiency than the editing assistance. In fact, we did find that participants were more likely to complete their writing tasks in a shorter amount of time when they were in the "*AI-primary*" writing mode than in the "*human-primary*" mode. So, if people determine the price that they are willing to pay for AI assistance primarily, or even solely based on the productivity benefits the assistance is expected to bring, then it is not surprising that people attach a higher value to generative AI's content generation assistance, despite they might be aware of its undesirable impacts on the writing experience and outcome.

Finally, this finding may also relate to the fact that our experiment is of relatively low stakes. The low-stake nature of our experimental task may contribute to the potential price-value mismatch of AI assistance via two ways: First, people may care less about the undesirable impacts on writing experience and performance caused by AI's content generation assistance due to the low stakes of the tasks, effectively making productivity-related considerations to be the primary driver in people's valuation of AI assistance. Second, due to the relatively low stakes of the task, people may also exhibit a degree of "over-reliance" on the content generation assistance provided by LLMs and show less motivation to fully explore how to best utilize the AI assistance and unlock their true value.

5.3 Implications for designing AI writing assistance

Our findings in this study suggest several implications for designing future AI-powered writing assistance. First, as people are found

to generally value LLMs' content generation assistance more than their editing assistance, a dynamic pricing model could potentially be adopted for future generative AI-powered writing assistants to align their price with users' perceived value of the AI assistant in different scenarios. For example, instead of pricing a generative AI-powered writing assistant at a fixed rate based on the length of the output that it produces, one may consider to adjust its price based on the type of output it produces—a higher price can be used when the assistant generates new content from scratch (as that in the "*AI-primary*" writing mode), while editing user-generated content (as that in the "*human-primary*" writing mode) can be associated with a lower price.

Secondly, we observed in our study that individuals with varying characteristics, such as different levels of confidence in writing, may have distinct preferences for AI assistance, and the value they place on AI assistance also varies with the type of writing tasks. These findings suggest that AI writing assistant systems could consider providing different kinds of customized assistance that tailors to the characteristics of the users and the needs of the tasks. For example, future generative AI-powered writing assistants could allow for personalized settings and offer granular controls to users to choose the type of AI assistance they prefer. In addition, the generative AI-powered writing assistant could also actively adapt its assistance to the profiles and interaction history of users, as well as the inferred user intent of writing. In other words, given a specific user and a writing context, the AI writing assistant should strive to provide the kind of assistance that the user values the most and avoid the kind of assistance that the user dislikes in the current context. For example, as high-confident writers appear to value editing assistance from AI negatively when writing creative stories, the AI writing assistant could learn to avoid providing text editing and polishing suggestions to them in this writing context by default. As the AI writing assistant identifies the users' writing intent (e.g., creative storytelling vs. fact-based reporting) on the fly, it could also dynamically adjust the range of assistance that it provides to users and prioritize them based on the user's preferences.

Furthermore, our study results showed that when generative AI models provide content generation assistance, it can often diminish the writer's satisfaction with both the writing process and the final product, e.g., resulting in a decrease in writers' perceived ability to express themselves. This suggests that when designing AI writing assistants, it's crucial to allow writers to maintain control over the AI assistant's content generation features. To ensure this, future AI writing assistants could incorporate options like "on-demand suggestions" (i.e., the AI assistant only provides suggestions when specifically requested by the user) or "AI pause" (i.e., users are allowed to temporarily pause the AI's active content suggestions). Another useful feature could be "content generation intensity sliders", in which users can drag the sliders to specify their requirements for the content generated by AI on different dimensions (e.g., length, formality, emotional degree, writing styles) to better fit their needs. Since people do not seem to decrease their satisfaction with their writing process and performance when AI assistants simply polish their own writing, one may also envision a new kind of content generation assistance that may have limited negative impacts on user perceptions—AI assistants can provide "imitating writing" assistance based on writing examples that users

provide, which may or may not be written by the user themselves. To counteract the decreased level of enjoyment in writing and ownership of the writing outcome caused by AI's content generation assistance, visual annotations or indicators can be incorporated to distinguish between AI-generated content and human edits, hence highlighting the human contributions to the final writing and emphasizing the collaborative nature of the final output and writing process. Finally, as the current content generation assistance offered by AI assistants appears to significantly decrease the diversity in people's writing, future designs should explore ways to decrease users' over-reliance on the AI-generated content and to encourage users to explore new and creative ways to collaborate with the AI assistant.

5.4 The potential influence of study setup

We acknowledge that our findings may be specific to the experimental setup used in this study. First, as previously discussed, our experiment was based on a relatively low-stake and domain-knowledge-free writing scenario. It does not represent all writing contexts that people may encounter, such as those involving high stakes and require substantial domain expertise (e.g., writing critical business proposals, where domain knowledge is needed and poor quality can lead to proposal rejections). We speculate that in those high-stake, domain-knowledge-required writing scenarios, people may prioritize quality-related considerations over productivity-related considerations. Thus, without being fully convinced that the content produced by generative AI-powered assistant can significantly boost the quality of their writing, people may place less value in AI's content generation assistance. In contrast, given the common belief that AI's editing assistance can effectively reduce grammar errors and enhance the overall writing quality, the perceived value of AI's editing assistance may be elevated, aligning with people's priority on quality considerations. Meanwhile, we may also expect a degree of heterogeneity in how changes in factors like stakes affect people's value of AI assistance. For example, for those individuals with very low confidence in their own writing, they may even associate a higher value with AI's content generation assistance as the stake of the writing task becomes higher. Future studies should be conducted to gain deeper understandings of the extent to which our findings can generalize to other writing scenarios with different characteristics.

Another feature of our study setup that may have influenced our finding is that our experiment created a writing scenario that emphasizes on people's *extrinsic motivations* for writing—participants in our study completed writing jobs to get paid. However, in the real world, there are people who write to satisfy their *intrinsic motivations*, such as expressing themselves in a creative way and seeing their identity through the writing process. The extent to which people find intrinsic motivations in writing may change how much they value AI assistance and how the provision of AI assistance affects their writing process and outcome. For example, it is possible that when people attach intrinsic value to writing themselves, they may find the generative AI-powered writing assistants, especially those that provide content generation assistance, to erode their intrinsic motivation. If this is the case, we may expect people with high intrinsic motivation for writing to place a lower value to AI

assistance, especially on those creative writing tasks that allow them to fully express themselves. Moreover, when people with high intrinsic motivation for writing are provided with AI assistance, they may also show greater motivation in exploring creative ways to steer AI to generate the type of assistance that they would find truly inspiring and useful—if they are successful in doing so, the incorporation of AI assistance in their writing flows may result in an improved perception of the writing experience and outcome for them.

Related to this, it's important to note that our experiment was conducted on an online platform, Prolific. As such, participants of our study might be primarily driven by factors like task completion efficiency, and they may have the tendency to satisfice. That is, the population of participants in our study may heavily focus on their extrinsic motivations rather than their intrinsic ones. This means that while the paid writing job design of our study already emphasizes on people's extrinsic motivation for writing, the fact that our study was conducted on an online platform may have amplified this emphasis. Future work should be conducted to explore to what extent our findings can be generalized to a study setup where people place high intrinsic motivation in writing.

5.5 Limitations and future work

We acknowledge a few limitations of our study. Firstly, the two types of writing tasks we selected may not encompass the full range of real-world writing activities. Our findings suggest that human's value for AI assistance, as well as the influence of AI assistance on people's writing perceptions and performance, is significantly influenced by the nature of the writing task. Therefore, conclusions drawn from our study concerning argumentative essay and creative story writing tasks may not generalize to other writing tasks. For example, when the writing task requires a high level of personal touch, such as writing letters to friends, it is unclear whether people would still be willing to place substantial values on the AI assistance. Future research should delve into various writing tasks to re-examine the questions we have asked in this study, in order to provide a more comprehensive understanding of the nuanced differences in how people perceive the AI assistance and how the AI assistance impacts people in different writing domains.

In addition, we explored two types of AI assistance in this paper: (1) text editing and polishing assistance, and (2) free-form content generation assistance. While they capture two representative formats of assistance offered by AI-powered writing assistants, our study might not encompass more complicated forms of AI assistance, such as those kinds of assistance enabled by advanced interactive interfaces that facilitate the communication between humans and LLMs. Consequently, our conclusions are primarily reflective of the two types of AI assistance we considered, manifested through the interfaces we used and the generative AI model we deployed in this study (i.e., ChatGPT). However, we believe that some design implications revealed by our study could still be relevant for other AI writing assistants that are powered by a different model, or present different kinds of content generation capabilities. For instance, the needs for customizing the type of AI assistance based on user characteristics and task contexts, and the critical importance of maintaining human control over the writing process are likely still

true across a variety of generative AI-powered writing assistants. In the future, it's essential to investigate AI writing assistants that operate on different AI models, use different interfaces, or provide different functionalities, to comprehensively cover the landscape of AI writing assistants. These investigations will be instrumental in revealing both commonalities and differences in how various forms of AI assistance are perceived by users and affect users in diverse writing contexts, and may further enhance our understanding of how to choose from different forms of AI writing assistants in different scenarios.

Furthermore, we note that the aspects of writing experience perceptions and performance covered in our study are not exhaustive. So, generative AI-powered writing assistants may bring about additional benefits or risks to people that are not measured in this study. For example, it is possible that writers' viewpoints get biased or even misled by the discriminatory content or misinformation generated by the AI assistant. Future research could further expand these aspects to obtain a more comprehensive understanding of the effects of generative AI-powered writing assistants.

Lastly, our study was limited in its exploration of the *long-term effects* of AI assistant usage on humans' writing skills and habits, as well as how this prolonged interaction might influence individuals' perceptions and valuation of AI writing assistance. It remains unclear whether a sustained engagement with AI writing assistants would lead to increased human reliance on AI and a corresponding rise in the perceived financial value of AI assistance, or whether it would enable users to recalibrate their beliefs about the capabilities of AI assistants, potentially decreasing their perceived value. Longitudinal studies are needed in the future to understand how the long-term usage of AI writing assistants influences people's writing process and performance over time, and how it alters people's perceived value of such AI assistance in the long run.

6 CONCLUSION

In this paper, we present a study to obtain a systematic understanding of whether and how much value people attach to AI assistance, and how the incorporation of AI assistance in writing workflows changes people's writing perceptions and performance. Our results show that people are willing to forgo financial payments to receive writing assistance, especially if AI can provide direct content generation assistance and the writing task is highly creative. In addition, our results highlight that although the generative-AI powered assistance is found to offer benefits in increasing people's productivity and confidence in writing, direct content generation assistance offered by AI also comes with risks, including decreasing people's sense of accountability and diversity in writing.

Our work provides important implications for the design of human-AI collaborative writing systems. For example, future generative AI-powered writing assistants could dynamically adjust their assistance based on the characteristics of the user and the writing task to better align with what users value the most in the specific context. In addition, to minimize the negative impacts of AI's content generation assistance on people's writing perceptions and performance, actions need to be taken to ensure users' control over the content generated by AI and the final writing product. To gain a more comprehensive understanding of the commonality and

differences in people's attitude towards and experience with generative AI-powered writing assistants in various scenarios, future work may look into the generalizability of our findings across a wider range of writing tasks, a broader spectrum of AI-powered writing assistants, and among different user populations.

REFERENCES

- [1] Daron Acemoglu and Pascual Restrepo. 2018. The race between man and machine: Implications of technology for growth, factor shares, and employment. *American economic review* 108, 6 (2018), 1488–1542.
- [2] Kenneth C Arnold, Krzysztof Z Gajos, and Adam T Kalai. 2016. On suggesting phrases vs. predicting words for mobile text composition. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. 603–608.
- [3] Nikola Banovic, Ticha Sethapakdi, Yasasvi Hari, Anind K Dey, and Jennifer Mankoff. 2019. The limits of expert text entry speed on mobile keyboards with autocorrect. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services*. 1–12.
- [4] Xiaojun Bi, Tom Ouyang, and Shumin Zhai. 2014. Both complete and correct? multi-objective optimization of touchscreen keyboard. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 2297–2306.
- [5] Oloff C Biermann, Ning F Ma, and Dongwook Yoon. 2022. From tool to companion: Storywriters want AI writers to respect their personal values and writing strategies. In *Designing Interactive Systems Conference*. 1209–1227.
- [6] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3, Jan (2003), 993–1022.
- [7] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems* 33 (2020), 1877–1901.
- [8] Daniel Buschek, Benjamin Bisinger, and Florian Alt. 2018. ResearchIME: A mobile keyboard application for studying free typing behaviour in the wild. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–14.
- [9] Daniel Buschek, Martin Zürn, and Malin Eiband. 2021. The impact of multiple parallel phrase suggestions on email input and composition behaviour of native and non-native english writers. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–13.
- [10] Tuhin Chakrabarty, Vishakh Padmakumar, and He He. 2022. Help me write a poem: Instruction Tuning as a Vehicle for Collaborative Poetry Writing. *arXiv preprint arXiv:2210.13669* (2022).
- [11] Mia Xu Chen, Benjamin N Lee, Gagan Bansal, Yuan Cao, Shuyuan Zhang, Justin Lu, Jackie Tsay, Yinan Wang, Andrew M Dai, Zhifeng Chen, et al. 2019. Gmail smart compose: Real-time assisted writing. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2287–2295.
- [12] John Joon Young Chung, Wooseok Kim, Kang Min Yoo, Hwaran Lee, Eytan Adar, and Minsuk Chang. 2022. TaleBrush: Sketching stories with generative pretrained language models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–19.
- [13] Elizabeth Clark, Anne Spencer Ross, Chenhao Tan, Yangfeng Ji, and Noah A Smith. 2018. Creative writing with a machine in the loop: Case studies on slogans and stories. In *23rd International Conference on Intelligent User Interfaces*. 329–340.
- [14] Girish Dalvi, Shashank Ahire, Nagraj Emmadi, Manjiri Joshi, Anirudha Joshi, Sanjay Ghosh, Prasad Ghone, and Narendra Parmar. 2016. Does prediction really help in Marathi text input? Empirical analysis of a longitudinal study. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 35–46.
- [15] Hai Dang, Sven Goller, Florian Lehmann, and Daniel Buschek. 2023. Choice over control: How users write with large language models using diegetic and non-diegetic prompting. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17.
- [16] Zijian Ding and Joel Chan. 2023. Mapping the Design Space of Interactions in Human-AI Text Co-creation Tasks. *arXiv e-prints* (2023), arXiv–2303.
- [17] Mark Dunlop and John Levine. 2012. Multidimensional pareto optimization of touchscreen keyboards for speed, familiarity and improved spell checking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2669–2678.
- [18] Andrew Fowler, Kurt Partridge, Ciprian Chelba, Xiaojun Bi, Tom Ouyang, and Shumin Zhai. 2015. Effects of language modeling and its personalization on touchscreen typing performance. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 649–658.
- [19] Liye Fu, Benjamin Newman, Maurice Jakesch, and Sarah Kreps. 2023. Comparing Sentence-Level Suggestions to Message-Level Suggestions in AI-Mediated Communication. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–13.

- [20] Daniel G Goldstein, Siddharth Suri, R Preston McAfee, Matthew Ekstrand-Abueg, and Fernando Diaz. 2014. The economic and cognitive costs of annoying display advertisements. *Journal of Marketing Research* 51, 6 (2014), 742–752.
- [21] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.
- [22] Morten Storling Hedgegaard and Jean-Robert Tyran. 2018. The price of prejudice. *American Economic Journal: Applied Economics* 10, 1 (2018), 40–63.
- [23] Jess Hohenstein and Malte Jung. 2020. AI as a moral crumple zone: The effects of AI-mediated communication on attribution and trust. *Computers in Human Behavior* 106 (2020), 106190.
- [24] David M Holtz, Liane Scult, and Siddharth Suri. 2022. How much do platform workers value reviews? An experimental method. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–11.
- [25] https://huggingface.co/sentence-transformers/paraphrase_MiniLM-L6-v2. 2023. Sentence transformers: paraphrase-MiniLM-L6-v2.
- [26] Maurice Jakesch, Advait Bhat, Daniel Buschek, Lior Zalmanson, and Mor Naaman. 2023. Co-writing with opinionated language models affects users' views. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–15.
- [27] Maurice Jakesch, Megan French, Xiao Ma, Jeffrey T Hancock, and Mor Naaman. 2019. AI-mediated communication: How the perception that profile text was written by AI affects trustworthiness. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.
- [28] Anjuli Kannan, Karol Kurach, Sujith Ravi, Tobias Kaufmann, Andrew Tomkins, Balint Miklos, Greg Corrado, Laszlo Lukacs, Marina Ganea, Peter Young, et al. 2016. Smart reply: Automated response suggestion for email. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 955–964.
- [29] Mina Lee, Percy Liang, and Qian Yang. 2022. Coauthor: Designing a human-ai collaborative writing dataset for exploring language model capabilities. In *Proceedings of the 2022 CHI conference on human factors in computing systems*. 1–19.
- [30] Chen Liang, Jing Peng, Yili Hong, and Bin Gu. 2023. The hidden costs and benefits of monitoring in the gig economy. *Information Systems Research* 34, 1 (2023), 297–318.
- [31] Stephen MacNeil, Andrew Tran, Joanne Kim, Ziheng Huang, Seth Bernstein, and Dan Mogil. 2023. Prompt Middleware: Mapping Prompts for Large Language Models to UI Affordances. *arXiv preprint arXiv:2307.01142* (2023).
- [32] Alexandre Mas and Amanda Pallais. 2017. Valuing alternative work arrangements. *American Economic Review* 107, 12 (2017), 3722–3759.
- [33] Piotr Mirowski, Kory W Mathewson, Jaylen Pittman, and Richard Evans. 2023. Co-Writing Screenplays and Theatre Scripts with Language Models: Evaluation by Industry Professionals. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–34.
- [34] Shaked Noy and Whitney Zhang. 2023. Experimental evidence on the productivity effects of generative artificial intelligence. *Science* 381, 6654 (2023), 187–192. <https://doi.org/10.1126/science.adh2586> arXiv:<https://www.science.org/doi/pdf/10.1126/science.adh2586>
- [35] OpenAI. 2023. GPT-4 Technical Report. (2023).
- [36] Ksenia Palin, Anna Maria Feit, Sunjun Kim, Per Ola Kristensson, and Antti Oulasvirta. 2019. How do people type on mobile devices? Observations from a study with 37,000 volunteers. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services*. 1–12.
- [37] Philip Quinn and Shumin Zhai. 2016. A cost-benefit study of text entry suggestion interaction. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 83–88.
- [38] Samuel Rhys Cox, Yunlong Wang, Ashraf Abdul, Christian Von Der Weth, and Brian Y. Lim. 2021. Directed diversity: Leveraging language embedding distances for collective creativity in crowd ideation. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–35.
- [39] Michael Röder, Andreas Both, and Alexander Hinneburg. 2015. Exploring the space of topic coherence measures. In *Proceedings of the eighth ACM international conference on Web search and data mining*. 399–408.
- [40] Sherwin Rosen. 1986. The theory of equalizing differences. *Handbook of labor economics* 1 (1986), 641–692.
- [41] Nikhil Singh, Guillermo Bernal, Daria Savchenko, and Elena L Glassman. 2022. Where to hide a stolen elephant: Leaps in creative writing with multimodal machine intelligence. *ACM Transactions on Computer-Human Interaction* (2022).
- [42] Keith Vertanan, Haythem Memmi, Justin Emge, Shyam Reyal, and Per Ola Kristensson. 2015. VelociTap: Investigating fast mobile text entry using sentence-based decoding of touchscreen keyboard input. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 659–668.
- [43] Daijin Yang, Yanpeng Zhou, Zhiyuan Zhang, Toby Jia-Jun Li, and Ray LC. 2022. AI as an Active Writer: Interaction strategies with generated text in human-AI collaborative fiction writing. In *Joint Proceedings of the ACM IUI Workshops*, Vol. 10. CEUR-WS Team.
- [44] Ming Yin, Siddharth Suri, and Mary L Gray. 2018. Running out of time: The impact and value of flexibility in on-demand crowdwork. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–11.
- [45] Ann Yuan, Andy Coenen, Emily Reif, and Daphne Ippolito. 2022. Wordcraft: story writing with large language models. In *27th International Conference on Intelligent User Interfaces*. 841–852.
- [46] Jiawei Zhou, Yixuan Zhang, Qianni Luo, Andrea G Parker, and Munmun De Choudhury. 2023. Synthetic lies: Understanding ai-generated misinformation and evaluating algorithms and human solutions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*.

A TASK INTERFACES

Figures A1 and A2 show the task interfaces that participants in the “*human-primary*” and “*AI-primary*” writing modes used, respectively.

B DEMOGRAPHIC INFORMATION OF PARTICIPANTS

In total, 379 workers from Prolific took our study and passed the attention check. Among them, 183 were allocated to the “*independent vs. human-primary*” treatment, while the remaining 196 were placed in the “*independent vs. AI-Primary*” treatment. The full demographic information of participants is shown in Table A1.

C FULL REGRESSION RESULTS

In this study, we investigate 19 dependent variables which are listed as:

- (1) Dependent Variable 1 (DV1): Average cognitive load participants experience in terms of mental load, time pressure, effort required, and the frustration level when completing the writing task.
- (2) Dependent Variable 2 (DV2): Satisfaction with the writing process.
- (3) Dependent Variable 3 (DV3): Enjoyment during the writing process.
- (4) Dependent Variable 4 (DV4): Sense of ease during the writing process.
- (5) Dependent Variable 5 (DV5): Ability to express creative goals during the writing process.
- (6) Dependent Variable 6 (DV6): Sense of satisfaction with the quality of the final article.
- (7) Dependent Variable 7 (DV7): Sense of ownership over the final article.
- (8) Dependent Variable 8 (DV8): Sense of pride in the final article.
- (9) Dependent Variable 9 (DV9): Sense of uniqueness of the final article.
- (10) Dependent Variable 10 (DV10): Confidence boost in writing argumentative essays.
- (11) Dependent Variable 11 (DV11): Confidence boost in writing creative stories.
- (12) Dependent Variable 12 (DV12): The pooled average confidence boost in other writing tasks.
- (13) Dependent Variable 13 (DV13): The willingness to take responsibility if the article is criticized for containing deceptive content (e.g., misinformation).
- (14) Dependent Variable 14 (DV14): The willingness to take responsibility if the article is criticized for containing content that is highly similar to someone else’s writing.
- (15) Dependent Variable 15 (DV15): The willingness to take responsibility if the article is criticized for invading other’s privacy.
- (16) Dependent Variable 16 (DV16): The willingness to take responsibility if the article is criticized as exhibiting bias/discrimination.
- (17) Dependent Variable 17 (DV17): Completion time of the writing task.

- (18) Dependent Variable 18 (DV18): Number of grammar and spelling errors in the final article.
- (19) Dependent Variable 19 (DV19): Coherence score of the final article.

In the regression models, the primary independent variable is the writing mode the participants are engaged in, and the “*independent*” writing mode is used as the reference. Additionally, several covariates are incorporated into the regression models, including

- (1) Demographics:
 - Gender: “Male”, “Female”, and “Others”. The reference category is “Male”.
 - Age: “below 35”, “35-44”, and “above 45”. The reference category is “below 35”.
 - Race: “White”, “Black or African American”, “Hispanic or Latino”, and “Others”. The reference category is “White”.
 - Education: “High School or lower”, “Some College”, “Bachelor’s Degree”, and “Graduate School or higher”. “High School or lower” is the reference category.
- (2) Payment: The writing job payment associated with their selected job offer.
- (3) Writing Confidence: Participants’ confidence in their own writing abilities.
- (4) Familiarity with ChatGPT.
- (5) Usage frequency of ChatGPT: The raw five levels are “Never”, “Rarely (once a week or less)”, “Occasionally (a few times a week)”, “Frequently (once a day)”, and “Very Frequently (more than once a day)”. According to our data, the median category selected by our participants is “Rarely”. Based on this, the categories ‘Occasionally’, ‘Frequently’, and ‘Very Frequently’ are grouped as the “high” level whereas the “Never” and “Rarely” categories are grouped as the “Low” level. The “Low” level is set as the reference.

The complete regression results for participants who completed the argumentative writing task are shown in Table A2 and Table A3. The complete regression results for participants who completed the creative story task are shown in Table A4 and Table A5.

D ADDITIONAL RESULTS

D.1 Detailed results on people’s perceptions of the final writing outcome

The provision of content generation assistance by AI significantly decreases people’s perceived ownership of the final writing outcome. As shown in Figure 6b, when examining participants’ perceived ownership over the final article produced as the outcome of the writing processes, we found that participants who received content generation assistance from ChatGPT consistently decreased their ratings compared to other participants, regardless of the types of writing tasks they worked on. Specifically, when writing argumentative essays, participants in the *AI-primary* writing mode reported significantly lower levels of ownership over the final article compared to participants in the *independent* or *human-primary* modes (*adjusted-p* < 0.001 for both comparisons). Similarly, when writing creative stories, participants in the *AI-primary* writing mode also perceived significantly lower levels of ownership

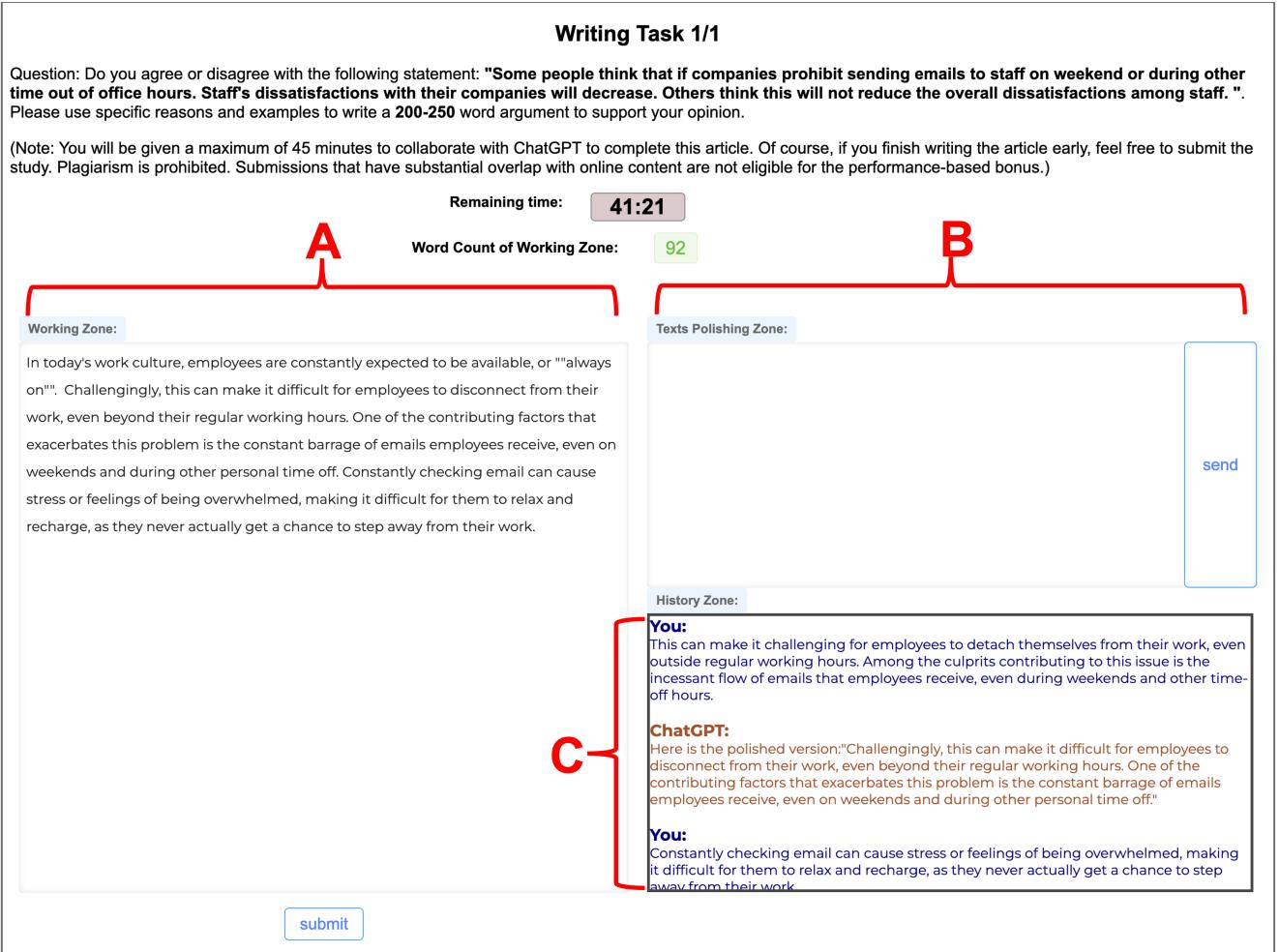


Figure A1: An example of the writing interface for participants in the Human-Primary writing mode. Working Zone (Part A): This is where participants can compose the article and integrate any ChatGPT-polished text into their own writing. Once they are satisfied with the version of the article in this zone, they can proceed to submit the current version by clicking the “submit” button. Text Polishing Zone (Part B): This is where participants can enter the texts that they want ChatGPT to polish, with the option to directly copy from any part of the texts in the working zone. After entering the desired text for editing assistance, participants can click the “send” button to forward it to ChatGPT. History Zone (Part C): This zone shows the entire interaction history between the participant and ChatGPT, which includes the original text the participant send to ChatGPT for polish and the polished version of the text produced by ChatGPT. Participants can compare both versions of the text and decide whether and how to integrate the ChatGPT-polished text into their articles in the working zone.

over the final article compared to participants in the *independent* writing mode (*adjusted-p* < 0.001).

In argumentative essay writing, the provision of content generation assistance by AI significantly decreases how much people take pride in the final writing outcome and how much people consider it as unique. As illustrated in Figure 6c, when participants received content generation assistance from ChatGPT, they expressed a significantly reduced level of pride in the final article relative to participants who wrote them independently (*adjusted-p* = 0.002), and those who only received editing assistance from ChatGPT (*adjusted-p* = 0.007). Additionally, when comparing the perceived sense of uniqueness in the final article in Figure 6d, we

also found that participants in the *AI-primary* mode reported a noticeable decrease compared to participants who took the *independent* writing mode (*adjusted-p* < 0.001), and those in the *human-primary* writing mode (*adjusted-p* = 0.005).

D.2 Detailed results on people's writing performance

AI assistance leads to significantly decreased number of grammar and spelling mistakes in people's writing. Our results demonstrate that participants in the *human-primary* and *AI-primary* modes significantly reduced the number of grammar and spelling

Writing Task 1/1

Topic: Do you agree or disagree with the following statement: "Nowadays it is easier to maintain good health than it was in the past? ". Please use specific reasons and examples to write a 200-250 word argument to support your opinion.

(Note: You will be given a maximum of 45 minutes to collaborate with ChatGPT to complete this article. Of course, if you finish writing the article early, feel free to submit the study. Plagiarism is prohibited. Submissions that have substantial overlap with online content are not eligible for the performance-based bonus.)

A

Remaining time: 42:20

Word Count of Working Zone: 236

Interaction History Zone:

ChatGPT:

I believe that it is easier to maintain good health nowadays than it was in the past. Advancements in medicine, technology, and health awareness have contributed to better health outcomes and made it easier for people to access and engage in healthy habits. For example, medical advancements have made it possible to diagnose and treat illnesses more effectively, reducing the mortality rates of diseases that were once considered fatal. Vaccines have also been developed to protect against deadly diseases like smallpox, measles, and polio, preventing millions of deaths worldwide. Moreover, technology has made it easier for people to track their health and engage in healthy habits. With the advent of digital health apps and wearables, people can monitor their physical activity, track their food intake, and receive personalized recommendations to improve their health. Additionally, telemedicine has made it easier for people to access healthcare services remotely, improving access and convenience for many. Finally, the importance of health awareness has increased, with more people being informed about the benefits of a healthy lifestyle. People are more aware of the importance of exercise, healthy eating, and stress management, which has led to a greater emphasis on prevention rather than just treatment. In conclusion, while maintaining good health requires effort and discipline, the advancements in medicine, technology, and health awareness have made it easier for people to achieve and maintain good health. With continued progress in these areas, we can strive for even better health outcomes in the future.

You:

I disagree with the statement, can you help me draft an essay that counters this?

ChatGPT:

Your Instruction Zone:

B

C

Working Zone:

Maintaining good health nowadays is actually more challenging than it was in the past. Despite advancements in medicine, technology, and health awareness, several factors make it harder for individuals to achieve and sustain good health. Medical advancements have certainly improved treatment options for diseases, but they have also brought about new challenges. Antibiotic resistance is a growing concern, making it harder to combat bacterial infections. Additionally, the rise of chronic diseases such as diabetes, heart disease, and obesity indicates that we are facing new health challenges that were not as prevalent in the past. While technology offers some benefits, it has also contributed to an increasingly sedentary lifestyle. The prevalence of smartphones, video games, and streaming platforms has led to more inactive behavior among individuals, causing a rise in obesity rates. Moreover, excessive screen time and exposure to blue light have been linked to sleep disorders, negatively impacting overall health. Despite increased health awareness, unhealthy food options abound in our modern society. Fast food chains, processed snacks, and sugary drinks continue to be easily accessible and often cheaper than healthier alternatives. Furthermore, busy lifestyles and work pressures often leave individuals with limited time to prioritize healthy eating or engage in regular exercise, leading to an overall decline in health. Finally, healthcare costs and access remain major barriers to maintaining good health. While telemedicine has improved access for some, significant disparities in healthcare availability persist, especially in low-income and rural communities.

send

submit

Figure A2: An example of the writing interface for participants in the AI-Primary mode. Interaction History Zone (Part A): This is where participants will first see the initial draft of the article that ChatGPT generates for them. Given the initial draft, participants can use the Instruction Zone below to offer feedback. ChatGPT will then adjust and re-generate the article accordingly. All interactions between the participant and ChatGPT will be logged within this zone, allowing the participant to review their interaction history whenever needed. **Instruction Zone (Part B):** This is where the participant can send instructions or feedback to ChatGPT for it to improve the draft. There are no restrictions on what kinds of instructions participants can provide to ChatGPT. **Working Zone (Part C):** This is where participants can compose and submit their final article. If participants want to use any section of the text generated by ChatGPT, they can directly copy it from the Interaction History Zone and paste it into here. Participants can also directly write a part of the article on their own or manually edit the ChatGPT-generated text. Once they are satisfied with the version of the article in this zone, they can proceed to submit the current version by clicking the "submit" button.

errors in their final articles compared to participants in the *independent* mode, both in argumentative essay writing (*independent* vs. *human-primary*: $p = 0.003$; *independent* vs. *AI-primary*: $p = 0.003$),

and in creative story writing (*independent* vs. *human-primary*: $p < 0.001$; *independent* vs. *AI-primary*: $p = 0.005$).

Demographics		<i>independent vs. AI-Primary</i> (N = 196)	<i>independent vs. human-primary</i> (N = 183)
Gender	Male	52.0%	56.3%
	Female	45.9%	41.5%
	Others or prefer not to say	2.1%	2.2%
Age	Below 35	38.3%	31.7%
	35–44	33.7%	32.2%
	45 or above	28.0%	36.1%
Race	White	74.5%	78.6%
	Black	12.2%	6.6%
	Hispanic	4.1%	4.4%
	Others	9.2%	10.4%
Education	High school or lower	16.3%	20.2%
	Some college	28.0%	23.5%
	Bachelor degree	38.3%	40.4%
	Graduate school or higher	17.4%	15.9%
Average writing confidence (Pre-task)	Argumentative essay	3.14	3.28
	Creative story	3.32	3.46
	Other writing tasks	3.21	3.32
ChatGPT usage frequency	Never	29.1%	27.9%
	Rarely	35.7%	38.3%
	Occasionally	20.4%	22.4%
	Frequently	11.7%	7.6%
	Very frequently	3.1%	3.8%
Average familiarity with ChatGPT		3.76	3.78

Table A1: Details of the demographic backgrounds of our study participants.

	DV1	DV2	DV3	DV4	DV5	DV6	DV7	DV8	DV9
Human-Primary	-0.38 (0.19)	0.35* (0.16)	0.17 (0.20)	0.33 (0.17)	0.21 (0.20)	0.37* (0.16)	-0.08 (0.18)	0.16 (0.19)	0.12 (0.21)
AI-Primary	-0.31* (0.15)	-0.17 (0.14)	-0.48** (0.18)	-0.15 (0.15)	-0.59** (0.18)	0.05 (0.14)	-1.32*** (0.16)	-0.60*** (0.17)	-0.79*** (0.18)
ChatGPT usage frequency (high)	-0.37* (0.15)	0.04 (0.14)	0.10 (0.18)	0.37* (0.15)	-0.21 (0.18)	-0.07 (0.14)	0.01 (0.16)	-0.15 (0.17)	-0.02 (0.18)
Gender (female)	-0.07 (0.13)	0.01 (0.12)	0.05 (0.16)	0.21 (0.14)	-0.27 (0.16)	0.19 (0.12)	0 (0.14)	-0.02 (0.15)	0.06 (0.16)
Gender (others)	0.27 (0.39)	0.38 (0.36)	-0.25 (0.46)	0.17 (0.40)	-0.31 (0.47)	0.44 (0.37)	-0.63 (0.41)	0.10 (0.44)	0 (0.47)
Age (35-44)	0.34* (0.16)	-0.01 (0.14)	-0.06 (0.19)	-0.31 (0.16)	-0.37 (0.19)	-0.07 (0.15)	0.14 (0.16)	-0.09 (0.18)	0.04 (0.19)
Age (above 45)	0.27 (0.18)	0.28 (0.16)	0.08 (0.21)	-0.35 (0.18)	-0.19 (0.21)	0.15 (0.17)	0.14 (0.19)	0.09 (0.20)	-0.11 (0.21)
Race (Black or African American)	0.33 (0.22)	-0.04 (0.20)	-0.14 (0.27)	-0.07 (0.22)	-0.36 (0.26)	-0.04 (0.20)	-0.32 (0.23)	0.08 (0.24)	-0.51 (0.26)
Race (Hispanic or Latino)	0.10 (0.26)	0.03 (0.24)	-0.43 (0.31)	-0.46 (0.27)	-0.62 (0.31)	-0.17 (0.24)	0.27 (0.27)	-0.04 (0.29)	-0.17 (0.31)
Race (Others)	0.23 (0.26)	-0.85*** (0.24)	-0.78* (0.31)	-0.87** (0.27)	-1.13*** (0.31)	-0.85** (0.24)	-0.82** (0.27)	-1.01*** (0.29)	-1.17*** (0.31)
Education (Some college)	0.08 (0.20)	0.28 (0.18)	0.04 (0.24)	0.38 (0.20)	-0.11 (0.24)	0.28 (0.19)	0.35 (0.21)	0.15 (0.22)	0.16 (0.24)
Education (Bachelor degree)	0.17 (0.17)	0.07 (0.16)	-0.17 (0.21)	0.19 (0.18)	-0.30 (0.21)	0.23 (0.16)	0.27 (0.18)	-0.02 (0.19)	0.28 (0.21)
Education (Graduate school or higher)	0.18 (0.23)	-0.02 (0.22)	-0.02 (0.28)	0.05 (0.24)	-0.13 (0.28)	0.26 (0.22)	0.50* (0.25)	0.11 (0.26)	0.50 (0.28)
Payment	0.04 (0.10)	0.01 (0.09)	-0.04 (0.12)	0.02 (0.06)	-0.05 (0.12)	-0.03 (0.09)	0.11 (0.10)	-0.08 (0.11)	0.03 (0.12)
Writing confidence in argumentative essay	-0.03 (0.05)	0.10 (0.05)	0.16* (0.07)	0.21*** (0.06)	0.09 (0.07)	0.08 (0.05)	0.10 (0.06)	0.06 (0.06)	0.10 (0.07)
Familiarity with ChatGPT	-0.14* (0.06)	0.17** (0.06)	0.15 (0.08)	0.27*** (0.06)	0.16* (0.08)	0.14* (0.06)	-0.01 (0.07)	0.14* (0.07)	0.12 (0.08)
Constant	3.09*** (0.51)	2.88*** (0.47)	3.11*** (0.61)	1.98*** (0.52)	3.84*** (0.61)	3.09*** (0.48)	3.48*** (0.53)	3.62*** (0.57)	2.90*** (0.61)

Table A2: Regression models predicting experiences and perceptions of writers who completed argumentative essay writing tasks, considering the writing mode they are in, their demographic information, the payment they could earn in their writing modes, their own writing confidence, and their familiarity with and usage frequency of ChatGPT. Coefficients and standard errors (in parentheses) are reported. *, **, and * represent significance levels of 0.05, 0.01, and 0.001, respectively (p-values are unadjusted).**

	DV10	DV11	DV12	DV13	DV14	DV15	DV16	DV17	DV18	DV19
Human-Primary	0.47** (0.16)	0.40** (0.15)	0.40** (0.12)	-0.18 (0.20)	-0.04 (0.24)	-0.32 (0.23)	-0.26 (0.22)	0.92 (1.44)	-2.86** (0.95)	-0.018 (0.016)
AI-Primary	0.24 (0.14)	0.39** (0.13)	0.34** (0.10)	-0.53** (0.17)	-0.33 (0.22)	-0.73*** (0.20)	-0.60** (0.19)	-3.79** (1.29)	-2.56** (0.85)	0.01 (0.01)
ChatGPT usage frequency (high)	0.35* (0.14)	-0.08 (0.13)	0.07 (0.10)	0.18 (0.17)	0.16 (0.21)	0.30 (0.20)	0.33 (0.19)	0.52 (1.27)	-1.71* (0.84)	-0.01 (0.01)
Gender (Female)	0.20 (0.12)	-0.19 (0.12)	-0.11 (0.09)	-0.10 (0.16)	-0.24 (0.19)	-0.33 (0.18)	-0.14 (0.17)	-0.04 (1.13)	-0.50 (0.75)	0.02 (0.02)
Gender (Other)	-0.01 (0.37)	-0.32 (0.35)	-0.04 (0.28)	-0.45 (0.45)	-0.10 (0.55)	-0.05 (0.52)	-0.16 (0.50)	2.11 (3.26)	2.62 (2.16)	0.03* (0.02)
Age (35-44)	-0.07 (0.15)	0.08 (0.14)	-0.05 (0.11)	0.08 (0.18)	0.11 (0.22)	0.41 (0.21)	0.28 (0.20)	2.21 (1.33)	1.61 (0.88)	0.02 (0.01)
Age (above 45)	0.12 (0.17)	0.31 (0.16)	0.06 (0.12)	0.12 (0.20)	0.15 (0.25)	0.48* (0.24)	0.45 (0.23)	3.70 (1.50)	0.22 (1.00)	0.01 (0.02)
Race (Black or African American)	0.37 (0.20)	0.11 (0.19)	0.11 (0.15)	0.03 (0.25)	-0.04 (0.31)	-0.02 (0.29)	0.13 (0.28)	-2.53 (1.83)	-0.36 (1.22)	-0.01 (0.03)
Race (Hispanic or Latino)	-0.08 (0.24)	-0.01 (0.23)	0.06 (0.16)	0.22 (0.30)	-0.02 (0.37)	-0.06 (0.35)	0.20 (0.33)	-0.24 (2.19)	-1.12 (1.45)	0.02 (0.03)
Race (Others)	-0.31 (0.24)	-0.24 (0.23)	-0.38* (0.18)	0.34 (0.30)	0.22 (0.37)	0.55 (0.35)	0.63 (0.35)	2.22 (2.27)	-0.95 (1.45)	0.01 (0.02)
Education (Some college)	-0.08 (0.16)	0.11 (0.18)	0.11 (0.14)	0.21 (0.23)	0.20 (0.28)	-0.25 (0.26)	0.21 (0.25)	-0.04 (1.67)	-1.86 (1.11)	-0.01 (0.02)
Education (Bachelor degree)	-0.14 (0.22)	-0.31 (0.15)	-0.04 (0.12)	0.04 (0.20)	0.03 (0.25)	0.13 (0.23)	0.19 (0.22)	-0.20 (1.46)	-1.34 (0.97)	-0.03 (0.02)
Education (Graduate school or higher)	0.17 (0.19)	-0.26 (0.21)	-0.06 (0.17)	0.30 (0.27)	0.16 (0.33)	-0.23 (0.31)	0.35 (0.30)	-0.70 (1.97)	-2.69* (1.31)	0.01 (0.01)
Payment	0.02 (0.09)	-0.16 (0.09)	0.01 (0.07)	0.20 (0.12)	0.16 (0.14)	0.20 (0.13)	0.34* (0.13)	0.94 (0.86)	-0.01 (0.57)	0.01 (0.01)
Writing confidence in argumentative essay	-0.40*** (0.05)	-0.08 (0.05)		0.06 (0.06)	0.05 (0.08)	0.03 (0.07)	-0.02 (0.07)	0.21 (0.48)	0.11 (0.32)	0.01 (0.02)
Average confidence in others			-0.12** (0.04)							
familiarity with ChatGPT	0.12* (0.06)	0.07 (0.05)	0.07 (0.04)	0.09 (0.07)	0.04 (0.09)	0.10 (0.08)	0.08 (0.08)	-0.53 (0.48)	-0.32 (0.36)	0.02 (0.05)
Constant	0.70 (0.48)	0.83 (0.45)	0.32 (0.37)	3.23*** (0.59)	3.02*** (0.72)	2.81*** (0.67)	2.23*** (0.65)	10.14* (4.22)	6.84* (2.80)	0.26*** (0.05)

Table A3: Regression models predicting experiences and perceptions of writers who completed argumentative essay writing tasks, considering the writing mode they are in, their demographic information, the payment they could earn in their writing modes, their own writing confidence, and their familiarity with and usage frequency of ChatGPT. Coefficients and standard errors (in parentheses) are reported. *, **, and * represent significance levels of 0.05, 0.01, and 0.001, respectively (*p*-values are unadjusted).**

	DV1	DV2	DV3	DV4	DV5	DV6	DV7	DV8	DV9
Human-Primary	0.31 (0.18)	0.16 (0.15)	0.09 (0.18)	0.06 (0.19)	0.01 (0.17)	0.12 (0.16)	-0.20 (0.19)	0.30 (0.18)	-0.15 (0.19)
AI-Primary	-0.10 (0.16)	0.04 (0.13)	-0.10 (0.16)	0.25 (0.16)	-0.24 (0.15)	0.01 (0.14)	-0.75*** (0.17)	-0.06 (0.16)	-0.35* (0.17)
ChatGPT usage frequency (High)	-0.10 (0.15)	0.02 (0.13)	-0.03 (0.15)	-0.01 (0.16)	0.05 (0.15)	0.01 (0.13)	-0.17 (0.16)	0.04 (0.15)	0.16 (0.16)
Gender (Female)	-0.23 (0.14)	0.21 (0.12)	0.13 (0.14)	0.03 (0.14)	0.03 (0.13)	0.07 (0.12)	-0.01 (0.15)	0.10 (0.14)	0.13 (0.15)
Gender (Others)	-0.41 (0.67)	-0.73 (0.56)	-0.22 (0.68)	0.37 (0.68)	-0.29 (0.64)	-0.92 (0.59)	-1.24 (0.71)	-0.91 (0.68)	-0.87 (0.72)
Age (35-44)	-0.04 (0.17)	0.04 (0.14)	-0.06 (0.17)	0 (0.17)	-0.20 (0.17)	-0.03 (0.15)	-0.30 (0.18)	-0.22 (0.17)	-0.07 (0.18)
Age (above 45)	0.14 (0.17)	0.01 (0.14)	0.17 (0.17)	0.01 (0.17)	0.08 (0.16)	-0.01 (0.15)	0.14 (0.18)	0.04 (0.17)	0.14 (0.18)
Race (Black or African American)	0.71** (0.24)	-0.24 (0.20)	-0.18 (0.24)	-0.18 (0.25)	-0.42 (0.23)	-0.16 (0.21)	-0.39 (0.25)	-0.27 (0.24)	-0.26 (0.26)
Race (Hispanic or Latino)	-0.35 (0.28)	-0.36 (0.23)	-0.21 (0.28)	-0.19 (0.28)	-0.48 (0.27)	-0.31 (0.24)	-0.08 (0.29)	-0.42 (0.28)	-0.13 (0.30)
Race (Others)	-0.10 (0.26)	-0.05 (0.22)	0.07 (0.27)	0.08 (0.27)	-0.01 (0.25)	-0.19 (0.23)	0.17 (0.28)	-0.13 (0.27)	0.29 (0.28)
Education (Some college)	0.30 (0.21)	-0.12 (0.18)	-0.15 (0.21)	-0.02 (0.22)	0.01 (0.20)	-0.06 (0.18)	0.04 (0.22)	-0.13 (0.21)	-0.15 (0.23)
Education (Bachelor degree)	0.36 (0.21)	-0.32 (0.17)	-0.33 (0.21)	-0.31 (0.21)	-0.16 (0.20)	-0.06 (0.18)	0.09 (0.22)	-0.09 (0.21)	-0.22 (0.22)
Education (Graduate school or higher)	0.36 (0.23)	-0.35 (0.19)	-0.26 (0.23)	-0.02 (0.23)	-0.25 (0.22)	-0.14 (0.20)	0.04 (0.24)	-0.39 (0.23)	-0.34 (0.24)
Payment	-0.09 (0.10)	0.01 (0.08)	0.04 (0.10)	0.06 (0.10)	0.08 (0.10)	0.08 (0.09)	-0.04 (0.24)	-0.01 (0.10)	0.25* (0.11)
Writing confidence in creative story	0.03 (0.06)	0.25*** (0.05)	0.26*** (0.06)	0.35*** (0.06)	0.31*** (0.06)	0.16** (0.05)	0.19** (0.06)	0.20** (0.06)	0.25*** (0.07)
Average confidence in others									
familiarity with ChatGPT	-0.03 (0.07)	0.13* (0.06)	0.16* (0.07)	0.10 (0.07)	0.26*** (0.07)	0.16* (0.06)	0.01 (0.08)	0.15 (0.07)	0.23** (0.08)
Constant	2.71*** (0.558)	2.81*** (0.46)	2.50*** (0.56)	2.13*** (0.572)	1.87*** (0.53)	2.73*** (0.49)	4.00*** (0.59)	2.90*** (0.56)	1.60** (0.60)

Table A4: Regression models predicting experiences and perceptions of writers who completed creative story writing tasks, considering the writing mode they are in, their demographic information, the payment they could earn in their writing modes, their own writing confidence, and their familiarity with and usage frequency of ChatGPT. Coefficients and standard errors (in parentheses) are reported. *, **, and * represent significance levels of 0.05, 0.01, and 0.001, respectively (*p*-values are unadjusted).**

	DV10	DV11	DV12	DV13	DV14	DV15	DV16	DV17	DV18	DV19
Human-Primary	0.06 (0.15)	-0.05 (0.15)	0.08 (0.11)	-0.39* (0.18)	-0.38 (0.22)	-0.51* (0.22)	-0.35 (0.22)	-0.80 (1.59)	-2.17*** (0.63)	-0.01 (0.02)
AI-Primary	0.50*** (0.13)	0.12 (0.13)	0.47*** (0.10)	-0.45** (0.16)	-0.32 (0.20)	-0.33 (0.20)	-0.17 (0.19)	-1.54 (1.42)	-1.60** (0.56)	-0.01 (0.02)
ChatGPT usage frequency (High)	0.12 (0.13)	0.24 (0.12)	0.14 (0.09)	0.05 (0.15)	-0.23 (0.18)	-0.13 (0.19)	0.01 (0.18)	1.35 (1.34)	-1.14* (0.53)	0.01 (0.01)
Gender (Female)	0.15 (0.11)	0.02 (0.11)	-0.02 (0.08)	-0.05 (0.14)	-0.01 (0.17)	-0.02 (0.17)	-0.10 (0.16)	-1.45 (1.22)	-0.92 (0.48)	0.01 (0.01)
Gender (Others)	0.21 (0.56)	-0.06 (0.54)	0.09 (0.42)	-1.22 (0.68)	-1.06 (0.81)	-1.10 (0.82)	-1.21 (0.79)	-2.95 (5.76)	-1.06 (2.29)	-0.02 (0.01)
Age (35-44)	0.01 (0.14)	-0.07 (0.14)	-0.12 (0.11)	0.21 (0.17)	0.30 (0.21)	0.26 (0.21)	0.19 (0.20)	2.69 (1.50)	-0.83 (0.60)	0.02 (0.01)
Age (above 45)	-0.06 (0.14)	0.04 (0.14)	-0.05 (0.11)	0.32 (0.17)	0.33 (0.21)	0.46* (0.21)	0.44* (0.21)	4.83** (0.20)	-0.36 (1.49)	-0.05 (0.59)
Race (Black or African American)	-0.10 (0.20)	-0.06 (0.19)	-0.04 (0.15)	-0.40 (0.24)	0.07 (0.29)	-0.27 (0.29)	-0.47 (0.29)	1.08 (2.10)	1.73* (0.83)	-0.01 (0.02)
Race (Hispanic or Latino)	-0.03 (0.23)	0.02 (0.22)	-0.03 (0.17)	-0.10 (0.28)	0.18 (0.34)	-0.01 (0.34)	-0.18 (0.33)	1.73 (2.41)	-1.06 (0.96)	0.02 (0.01)
Race (Others)	-0.18 (0.22)	0.12 (0.21)	0.12 (0.17)	0.13 (0.27)	0.34 (0.32)	0.10 (0.32)	0.19 (0.31)	2.52 (2.30)	-0.75 (0.91)	-0.01 (0.03)
Education (Some college)	0.05 (0.17)	-0.15 (0.17)	-0.09 (0.13)	0.37 (0.21)	0.52 (0.26)	0.57* (0.26)	0.45 (0.25)	2.23 (1.84)	-1.01 (0.73)	0.04 (0.02)
Education (Bachelor degree)	0.17 (0.19)	-0.15 (0.17)	0.01 (0.13)	0.09 (0.21)	0.26 (0.25)	0.25 (0.25)	-0.04 (0.25)	1.33 (1.82)	-1.25 (0.72)	-0.05 (0.02)
Education (Graduate school or higher)	-0.22 (0.19)	-0.21 (0.18)	-0.27 (0.15)	0.10 (0.23)	0.20 (0.28)	0.12 (0.28)	-0.32 (0.27)	0.51 (1.98)	-1.33 (0.79)	-0.03 (0.03)
Payment	0.24** (0.08)	0.03 (0.08)	0.10 (0.06)	0.14 (0.10)	0.04 (0.12)	0.19 (0.12)	0.17 (0.12)	1.52 (0.89)	-0.07 (0.35)	0.04 (0.01)
Writing confidence in creative story	-0.11* (0.05)	-0.29*** (0.05)		0.14* (0.06)	0.10 (0.07)	0.13 (0.08)	0.16* (0.12)	0.32 (0.56)	0.61** (0.22)	0.02 (0.01)
Average writing confidence in others			-0.18** (0.06)							
familiarity with ChatGPT	0.08 (0.06)	0.10 (0.06)	0.07 (0.05)	0.09 (0.07)	0.06 (0.09)	0.08 (0.09)	0.10 (0.09)	0.001 (0.66)	-0.35 (0.26)	0.01 (0.01)
Constant	-0.82 (0.46)	0.86 (0.45)	0.09 (0.35)	2.87*** (0.56)	3.04*** (0.67)	2.47*** (0.68)	2.60*** (0.66)	3.77 (4.79)	5.88** (1.90)	0.29*** (0.05)

Table A5: Regression models predicting experiences and perceptions of writers who completed creative story writing tasks, considering the writing mode they are in, their demographic information, the payment they could earn in their writing modes, their own writing confidence, and their familiarity with and usage frequency of ChatGPT. Coefficients and standard errors (in parentheses) are reported. *, **, and * represent significance levels of 0.05, 0.01, and 0.001, respectively (*p*-values are unadjusted).**