Writing with AI Lowers Psychological Ownership, but Longer Prompts Can Help

Nikhita Joshi
Cheriton School of Computer Science
University of Waterloo
Waterloo, Ontario, Canada
nvjoshi@uwaterloo.ca

Daniel Vogel
Cheriton School of Computer Science
University of Waterloo
Waterloo, Ontario, Canada
dvogel@uwaterloo.ca

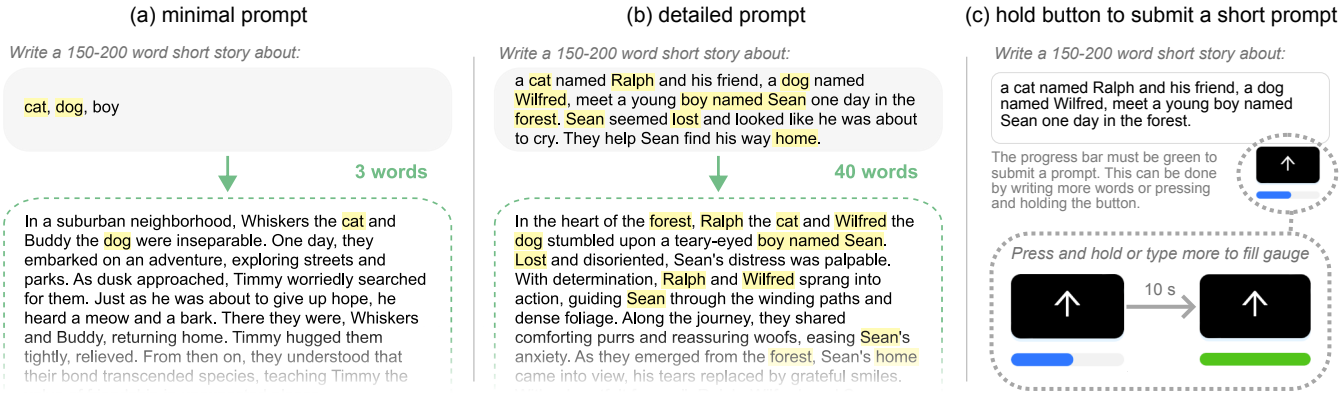

Figure 1: Writing a short story with (a) a minimal prompt with only 3 words and (b) a detailed prompt with 40 words. Yellow highlight shows similarities between the original prompt and the resulting story. Our work shows that longer prompts increase psychological ownership, likely due in part to this increased similarity. To encourage users to write longer prompts, we propose (c) a time delay experienced when pressing the button to submit a short prompt.

ABSTRACT

The feeling of something belonging to someone is called "psychological ownership." A common assumption is that writing with generative AI lowers psychological ownership, but the extent to which this occurs and the role of prompt length are unclear. We report on two experiments to examine the relationship between psychological ownership and prompt length. Participants wrote short stories either completely by themselves or wrote prompts of varying lengths. Results show that when participants wrote longer prompts, they had higher levels of psychological ownership. Their comments suggest they thought more about their prompts, often adding more details about the plot. However, benefits plateaued when prompt length was 75-100% of the target story length. To encourage users to write longer prompts, we propose augmenting the prompt submission button so it must be held down a long time if the prompt is short. Results show that this technique is effective at increasing prompt length.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; **Interaction techniques**.

KEYWORDS

generative AI, controlled experiments, interaction techniques

ACM Reference Format:
Nikhita Joshi and Daniel Vogel. 2025. Writing with AI Lowers Psychological Ownership, but Longer Prompts Can Help. In *Proceedings of the 7th ACM Conference on Conversational User Interfaces (CUI '25), July 8–10, 2025, Waterloo, ON, Canada.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3719160.3736608

CUI '25, July 8–10, 2025, Waterloo, ON, Canada

1 INTRODUCTION

People often write collaboratively, which can lead to many benefits, like improved text quality [38, 57]. However, it can also lead to negative outcomes, such as a loss of *psychological ownership* [3, 4, 6]. This is a concept about feelings of the text belonging to the writer, regardless of legal ownership [41]. Preserving feelings of psychological ownership is important when writing collaboratively, as prior work suggests that losing too much psychological ownership can deter people from writing together in the future [6].

People are increasingly writing "collaboratively" with generative AI services like ChatGPT. For example, author Rie Kudan, winner of the 170th Akutagawa Prize in Japan, used ChatGPT to write her award-winning book [8]. Given this new form of collaborative writing, understanding how psychological ownership is affected by generative AI writing assistants is crucial. It seems intuitive that writing with a generative AI assistant will result in lower feelings of psychological ownership than writing something independently, which has been corroborated by early work investigating the roles

and designs of AI writing assistants (e.g., [12, 28, 30]). However, we argue that more nuance of understanding is needed. Unlike collaboratively writing with another person, whose actions cannot be controlled, someone collaborating with a generative AI assistant can exert more control over the output through their prompts, which can vary in *length*. Someone may write a short prompt that lacks detail about the content or style, or may write a long prompt to include these details. Writing longer detailed prompts should result in generated text that is more representative of these additional details, and including these details requires an investment of time, energy, and sense of self; all of which are important for psychological ownership [40, 41, 51]. Therefore, someone who invests more by writing longer prompts may have increased feelings of psychological ownership than someone who writes less, and it is possible that writing longer prompts may even lead to similar levels of psychological ownership as independent writing. Despite the potential influence of prompt length on psychological ownership, this has not been empirically investigated.

We conducted two experiments to answer the following research question: *how much psychological ownership do people feel over writing produced by generative AI, and how does prompt length affect it?* During the first experiment, participants wrote 150-200-word short stories either completely by themselves or by writing prompts of varying lengths enforced through word minimums and maximums. Results showed that as participants wrote more to meet higher word minimums, they had higher levels of psychological ownership. Specifically, participants included more details about the story plot in their prompts, which resulted in more similarities between their prompts and their stories, and required more thought. The second experiment used even higher word minimums, yet psychological ownership plateaued.

These experiments required users to write longer prompts using word minimums. Though valuable for increased experimental control, this approach may not work well in practice, where users desire some flexibility. As such, we examine an important follow-up research question: *how can we encourage users to write longer prompts?* We propose a simple change to the prompt entry interface, introducing a *time delay* based on prompt length. Specifically, the user must press and hold the prompt submission button for some time before their prompt is submitted. The time delay is a function of prompt length, meaning if someone writes a longer prompt, they experience less delay. Using a similar creative writing task, we conducted an experiment with 20- and 60-second delays, and found prompt length increased compared to having no delays.

The effects of writing with a generative AI assistant on psychological ownership is important for the HCI community to understand, as the community is developing writing systems with generative AI support that offload work from the user. While this may be beneficial for productivity, users may come to feel negatively about such systems and stop using them over time, especially for tasks where the end product should be "theirs." Our work contributes empirical and quantitative findings on the effects of writing with a generative AI assistant on psychological ownership and how this is impacted by prompt length, and a simple yet effective approach for increasing prompt length through time delays that can be integrated into current chat-based AI interfaces.

2 BACKGROUND AND RELATED WORK

It has been long understood that people feel possessive towards objects, ideas, places, and creations. In psychology, these possessive feelings are referred to as *psychological ownership* [40, 41]. Many factors contribute towards increased feelings of psychological ownership, like the investment of time, energy, and sense of self, which may explain why people feel possessive towards creative work [40, 41, 51]. Feeling psychological ownership can have positive outcomes, like an increased sense of responsibility and citizenship behaviour [41], but can negatively impact collaborative tasks [40]. People often write collaboratively, which can improve text quality [38, 57], however, writers often feel psychological ownership over jointly-written text [4, 6, 18, 50]. Although some types of collaborative writing can improve psychological ownership, such as synchronous collaborations with more communication [2], fears of losing psychological ownership can often lead to territorial behaviours [25] or avoiding future collaboration [6].

Psychological ownership has been identified as important to the design [26] and evaluation [47] of AI writing systems. Biermann et al. [1] conducted interviews with hobbyist and professional writers to better understand how they write and the challenges they face while writing. Participants commented on mock-up user interfaces designed for different AI assistant roles, such as using the AI to receive suggestions for new text or to generate dialog for different characters. Psychological ownership was an important factor for these writers, and some designs were poorly received because they *"overstep the boundaries"* of the human writer. Draxler et al. [12] asked participants to write postcards entirely by themselves or to edit or choose postcards pre-written by AI before "signing" the postcard. Participants did not feel psychological ownership towards text that was pre-written by AI, however, they still did not list "AI" as an author, suggesting the presence of an *AI Ghostwriter Effect*. Li et al. [30] asked participants to write persuasive essays and stories with an AI assistant, where the human and the AI could be the primary writer or the editor. Having the AI as the primary writer lowered psychological ownership, however, participants still valued AI assistance as they were willing to forego some payment to receive help. These early evaluations suggest that psychological ownership is lowered when collaborating with AI, however, they focused on the role of the AI and did not explore other factors that contribute towards lower psychological ownership in depth.

There are many writing systems that leverage AI and LLMs (e.g., [20, 58]), but we focus on those that evaluated psychological ownership. Dramatron [33] helps writers create theatre screenplays and scripts, but writers reported low feelings of psychological ownership over the resulting pieces. GhostWriter [55] allows writers to refine writing style preferences through highlights, but psychological ownership was the most varied and lowest scoring metric. Metaphoria [17] suggests metaphors while writing poems, but writers felt less psychological ownership when using it, especially when *"the suggestions were particularly good."* The type of suggestion may also play a role. For example, Dhillon et al. [10] found that psychological ownership decreased when writers received entire paragraphs or sentences as suggestions from an AI writing assistant. Lee et al. [27] suggested that as writers receive more suggestions from AI, they write less and feel less psychological ownership. However,

they note that these results were preliminary and more research is needed to further validate these findings.

It has been hypothesized that increased control is desirable for human-AI collaborations (e.g., [37, 43, 44]), and that having more control can increase feelings of psychological ownership while writing. For example, Lehmann et al. [28] compared AI-generated suggestions that the user had to manually accept to suggestions that were automatically inserted into the text body. They found that manually-accepted suggestions led to higher feelings of psychological ownership, which may be due to the increased control at the composition stage.

To summarize, prior work has shown that psychological ownership is important to writers who write collaboratively with humans and with AI. Although some AI writing systems have included features to give writers more control to increase psychological ownership, the effects of more fundamental properties, such as the AI prompt length, have yet to be explored.

3 EXPERIMENT 1: PROMPT LENGTH

The goal of this experiment is to understand the impact of writing with generative AI on psychological ownership and how prompt length affects it. Through a within-subjects experimental design, participants wrote short stories with generative AI assistance by writing prompts of varying word lengths, which were enforced through word minimums and maximums for increased experimental control: 3 words, 50-100 words, and 150-200 words. As a baseline, participants also wrote a 150-200 word short story without any AI assistance.

3.1 Participants

We recruited 41 participants through the crowdsourcing experiment service, Prolific.[1] Participants were restricted to the United States and Canada, those who had completed at least 2,500 tasks, and those who had an approval rating greater than 98%. We manually examined all open-ended responses and user input to identify fraudulent responses [45] and filtered out participants who experienced technical issues with our interface and those who cheated by padding out their prompts or stories with dummy text to meet the word minimums and those who did not make enough keystrokes to generate their user input, suggesting they copied and pasted content from another source. In total, 10 participants (24%) were excluded, leaving 31 valid responses (Table A.1). All self-reported being proficient at reading and writing in English (all ≥ 5 on a 1-7 scale). Participants received $15 as remuneration.

3.2 Apparatus and Task

The experiment software was a Node.js and React web application (Figure 2). The main writing interface had a toolbar on the top and a side-by-side view underneath. The top toolbar contained instructions for the trial and three nouns that the story had to be about (i.e., a unique and randomly assigned "triad" from a set of nine triads adapted from Foley et al.'s text composition task [16]), which is useful for encouraging creative thinking on the spot during experiments [13]. Writing independent short stories is an ideal task for this type of experiment as creative writing requires a high personal investment and expression of self, which are important for psychological ownership, yet it provides increased control for an experiment.

A text box on the left provided a space for participants to write a prompt for the GPT-3.5 Turbo model. In a baseline condition, they wrote a complete story. Copying and pasting within the text box were disabled to prevent cheating, but this may not work across all browsers, so keystroke-level activity was also logged to verify response validity. A word counter above the text box displayed the number of words in the text box. Once the word count satisfied the requirements for the condition (i.e., the word count was between the minimum and maximum), a *Finalize* button could be pressed, causing the story to "type out" on the right side of the screen, following the behaviour of LLM AI systems like ChatGPT. The *Finalize* button could only be pressed once. The story was either generated by the model by including the participant's prompt within an engineered prompt,[2] or it was the same text written by the participant in the baseline. The same typing effect was used in both cases, and this provided an opportunity for participants to read the story. After the story was fully typed, a *Continue* button could be pressed to end the trial.

3.3 Procedure

Participants received a link to the web application through Prolific and the experiment was restricted to laptop and desktop devices. After entering basic demographic information and reading instructions, they were shown the writing interface and were asked to write a short story or a prompt about the three presented nouns. After the writing stage, they answered questions about their story and the writing experience. They repeated this for all conditions before answering questions about their overall thoughts and preferences. The entire experiment took approximately one hour.

3.4 Choice of Prompt Lengths

We were interested in examining the effects of prompt length, believing that being required to write longer prompts will increase the prompt level of detail and improve psychological ownership. As such, we needed to test a range of prompt lengths representing key cases, like a story generated with very little guidance from the writer, a story heavily guided by the writer, and a story written solely by the writer. At the same time, we had to ensure that the experiment duration was not too long for participants. Therefore, our choice of word minimums (and word maximums) was guided by the duration of conditions requiring more writing from the participant (e.g., writing a story without any AI assistance). We ultimately decided that the final stories for all conditions had to be 150-200 words, so that the experiment was not too time-consuming but the stories were still long enough to be fully developed. With a 150-200 word final story in mind, we chose conditions to examine the effects of prompt length and a baseline:

- **3 words with GPT**, representing 1% of the maximum final story word count. This best approximates the case where the story is fully generated by an AI assistant.

[1] https://www.prolific.com

[2] The prompt used to generate the story is included in the supplementary materials.

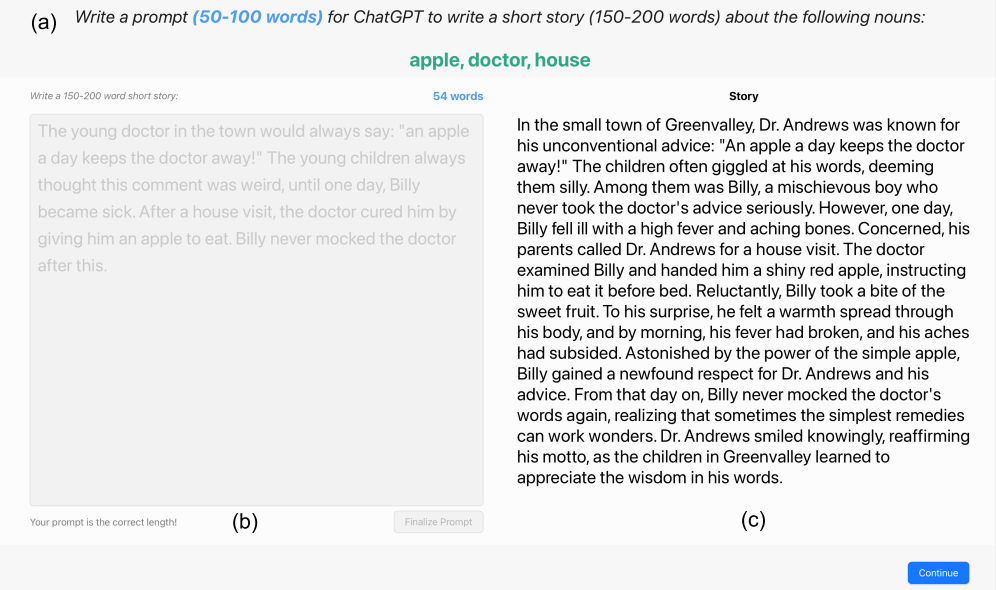

Figure 2: Experimental interface (shown with AI assistance): (a) top toolbar showing the instructions and three nouns the story had to be about; (b) a text box where the participant wrote their prompt or short story; (c) the story written by themselves or with AI assistance.

- **50-100 words with GPT**, representing 25-50% of the maximum final story word count. The story is generated by an AI assistant but has some guidance from the participant.
- **150-200 words with GPT**, representing 75-100% of the maximum final story word count. The story is generated by an AI assistant, but with a lot of guidance from the participant.
- **150-200 words without GPT**, a baseline representing the case where the story is fully written by the participant.

3.5 Design

This is a within-subjects design with one primary independent variable, CONDITION (levels: AI-3, AI-50, AI-150, NO-AI; the number representing the word minimum). CONDITION was randomly assigned. For each CONDITION, the primary measures were 10 subjective question scores, all interval numeric scores within a 1-7 range.[3] Four questions represented metrics related to psychological ownership and the phrasing was similar to prior work [6, 35]: *Personal Ownership*, *Responsibility*, *Personal Connection*, *Emotional Connection*. We average these four scores to create a composite measure, *Psychological Ownership*, which is a common technique in psychology papers. The internal consistency reliability was calculated using Cronbach's alpha, and the reliability was very high (α = 0.96), indicating that *Psychological Ownership* was a reliable composite measure for data analysis. Six questions represented metrics from the NASA-TLX: *Mental Demand*, *Physical Demand*, *Temporal Demand*, *Performance*, *Effort*, and *Frustration*. In addition, objective metrics were calculated from logs: the time taken in minutes to write the prompt or story (*Duration*); the number of words in each (*Word Count*); and the semantic text similarity between the final story and the participant's prompt (*Text Similarity*), which was calculated using Google's Universal Sentence Encoder (0-1 range; 1 being identical texts) [7].

[3]See the supplementary materials for question wording and data from the experiment.

4 EXPERIMENT 1: RESULTS

Where applicable, we use a Friedman test and Wilcoxon signed-rank tests, with Holm's corrections for multiple comparisons; and Spearman's correlations. *To streamline the presentation of results, all statistical test details are shown in Appendix A in Table A.2.* Where applicable, graphs show the mean and individual participant scores, and all error bars represent 95% confidence intervals (bootstrapped with 10,000 resamples).

4.1 Psychological Ownership

Overall, writing more words improved feelings of psychological ownership (Figure 3). A significant effect of CONDITION on *Psychological Ownership* and post hoc tests revealed that AI-3 led to the lowest scores ($M = 1.80$, $SD = 1.13$), followed by AI-50 ($M = 3.90$, $SD = 1.61$), AI-150 ($M = 4.57$, $SD = 1.70$), and NO-AI ($M = 6.29$, $SD = 0.80$).

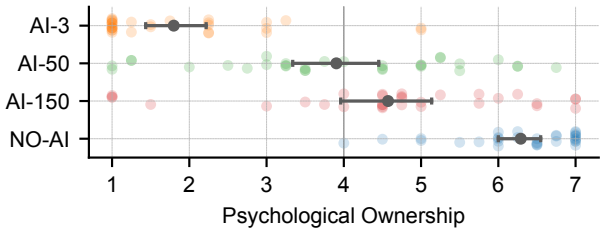

Figure 3: *Psychological Ownership* by CONDITION. Higher scores correspond to higher feelings of *Psychological Ownership*.

4.2 Writing Activities

To better understand factors that may have contributed to these results, we examine participants' writing activities and task workload. Open-ended responses were grouped by the first author as the data was straightforward given the number of participants and the short response length (2-3 sentences each).

4.2.1 Duration. We anticipated that spending more time writing may lead to higher *Psychological Ownership*, but they were only moderately positively correlated (r_s = .45; p < .001). A significant effect of CONDITION on *Duration* and post hoc tests revealed that AI-3 (M = 5.26, SD = 8.08) and AI-50 (M = 3.88, SD = 2.60) were faster than AI-150 (M = 8.37, SD = 4.91) and NO-AI (M = 9.54, SD = 6.83; Figure 4).

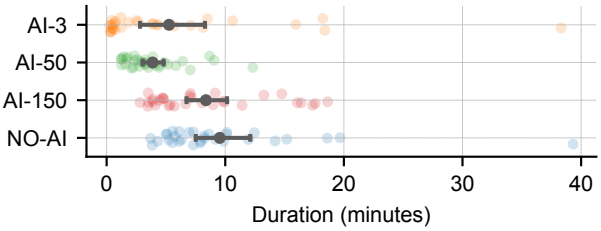

Figure 4: *Duration* by CONDITION.

4.2.2 Word Count. We observed a strong positive correlation between *Word Count* and *Psychological Ownership* (r_s = .72; p < .001), suggesting that as participants wrote more, they felt more psychological ownership over their writing. As expected, we observed a significant effect of CONDITION on *Word Count* with post hoc tests showing participants wrote fewer words with AI-3 and AI-50 (M = 66.94, SD = 17.91) than they did with AI-150 (M = 159.52, SD = 9.86) and NO-AI (M = 170.74, SD = 16.63; Figure 5). Despite having the same word limits, we also observed a significant difference between AI-150 and NO-AI. One possibility is that participants encountered a mental block with AI-150, as some mentioned how it was *"more taxing to write [a] longer prompt"* (P31) and how they *"couldn't think of what [...] to write about"* (P21).

4.2.3 Text Similarity. Overall, writing longer prompts led to more semantic text similarity with the generated story. There was a strong positive correlation between *Word Count* and *Text Similarity* (r_s = .82, p < .001), and a significant effect of CONDITION on *Text Similarity*, revealing that prompts written with AI-3 were the least similar to the final story (M = 0.27, SD = 0.10), followed by AI-50 (M = 0.43, SD = 0.12), AI-150 (M = 0.59, SD = 0.16), and NO-AI (Figure 6). Much like *Word Count*, having increased *Text Similarity* was strongly positively correlated with higher *Psychological Ownership* (r_s = .75, p < .001).

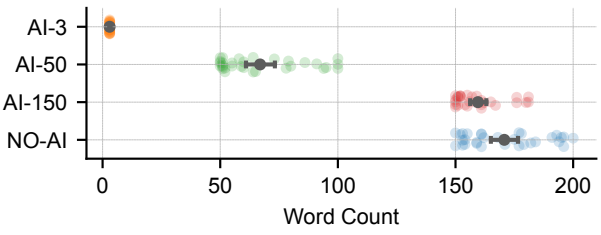

Figure 5: *Word Count* by CONDITION.

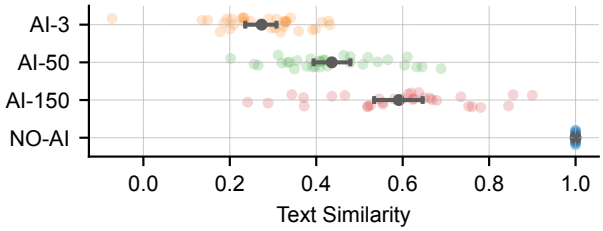

Figure 6: *Text Similarity* by CONDITION. Higher scores refer to more similarities between the input and output (1 = identical texts).

4.2.4 Prompt Strategies. We observed participants adopting several strategies with their prompts (Table 1). As expected, most participants wrote the three specified nouns for AI-3, so we only consider AI-50 and AI-150. Only 4 prompts (6%) purely consisted of instructions without any details about the story plot; 25 (40%) purely consisted of story plot details and resembled partially complete stories; and majority (33, 53%) were a mix of instructions and story details. As would be expected, it seems like requiring participants to write longer prompts encouraged greater levels of details about the story, something that 11 (35%) participants noted, for example: *"when I provide more words, I see that ChatGPT produces a story that is almost the same as mine"* (P1). Of these participants, 8 (25%) noted how this increased feelings of psychological ownership, for example: *"with longer prompts, I could influence what was on display and felt more ownership [towards] the work"* (P13).

4.2.5 Task Workload. The longest prompt and writing the full story were more mentally demanding and the shortest prompt required the least effort (Figure 7), but we did not observe meaningful trends for other task workload factors. For *Mental Demand*, a significant effect of CONDITION and post hoc tests revealed that AI-3 (M = 1.61, SD = 0.37) and AI-50 (M = 2.68, SD = 1.62) were less mentally demanding than AI-150 (M = 3.90, SD = 1.90) and NO-AI (M = 4.10, SD = 1.81; Figure 7a). There were no significant differences between AI-150 and NO-AI. *Mental Demand* was strongly positively correlated with *Psychological Ownership* (r_s = .65, p < .001).

There was a significant effect of CONDITION on *Effort*. Unsurprisingly, post hoc tests showed that AI-3 required the least *Effort* (M = 1.39, SD = 0.99). Although AI-50 (M = 3.39, SD = 1.63) required less effort than NO-AI (M = 4.35, SD = 1.52), there were no significant differences between AI-50 and AI-150 (M = 3.94, SD = 1.88) or between AI-150 and NO-AI (Figure 7b). *Effort* was strongly positively correlated with *Psychological Ownership* (r_s = .66, p < .001). The relationship between *Mental Demand* and *Effort* and *Psychological Ownership* makes sense; it has been long understood that investing more energy and self improves psychological ownership [40, 41, 51], which requires more thought and effort.

Table 1: Example prompt types submitted by participants (examples from AI-50 condition).

Plot	Mix	Instructions
"The girl asked her grandma for a dog every time she saw her. The grandma kept saying no, until finally she surprised her granddaughter with a dog for Christmas that year. The girl had tears of joy that she finally got a dog and was in disbelief that it was happening." (P18)	*"Create an average word paragraph using the following nouns: House, grandpa and mug. Start with a setting of grandpa sitting in a rocking chair on the porch. Have grandpa tell a quick story to his 8 year old grand daughter named Lisa. End story with grandpa going into the house."* (P15)	*"Write a short story about these three nouns: Boy, boat, cat. The story should be 150 - 200 words in length. The whole story should be about these three nouns. Write a story that can be read to children, so the words should be as simple as possible so children could understand."* (P26)

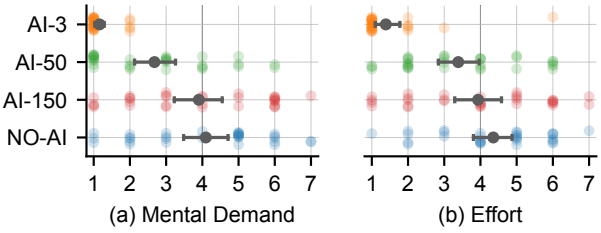

Figure 7: (a) *Mental Demand* and (b) *Effort* by CONDITION. Lower scores correspond to lower *Mental Demand* and *Effort*.

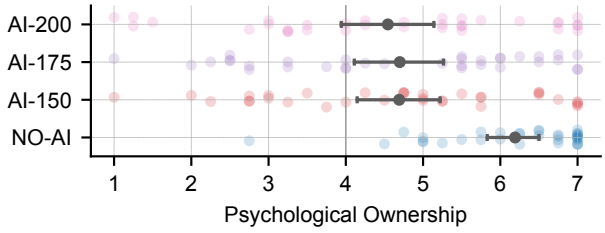

Figure 8: *Psychological Ownership* by CONDITION. Higher scores correspond to higher feelings of *Psychological Ownership.*

4.3 Summary

To summarize, our results suggest that requiring people to write longer prompts increases psychological ownership for their writing. By writing more, prompts became more similar to the resulting stories, likely because writers were encouraged to include more details about the story and its plot, which required more mental demand. An interesting question is: does this effect plateau with even longer prompts, perhaps even longer than the generated text?

5 EXPERIMENT 2: PLATEAU

To better understand this, we conducted a follow-up experiment with 39 new participants. Five (13%) were removed from the analysis for cheating or because they experienced technical issues, leaving 34 participants (Table B.1). All participants self-reported being proficient at reading and writing in English (all ≥ 5 on a 1-7 scale). The protocol and measures were the same as Experiment 1 (α = 0.92 for *Psychological Ownership*), but participants had to write prompts that were 150-200, 175-200, and 200-250 words long. The experiment was slightly longer at roughly 90 minutes and participants received $22.50 as remuneration.

5.1 Choice of Prompt Lengths

Two conditions were identical (AI-150 and NO-AI), and two new word minimums and maximums pushed participants to write even more:

- **175-200 words with GPT**, representing 87.5-100% of the maximum final story word count. 175 words roughly corresponds to the mean word count of solo-writing in Experiment 1 (AI-175), which encourages participants to write slightly more words than what was anticipated for NO-AI.
- **200-250 words with GPT**, representing 100-150% of the maximum final story word count. This condition requires the participant to write more words than the baseline solo-writing condition and the generated story, which may lead to higher feelings of psychological ownership (AI-200).

5.2 Results

We use the same analysis as Experiment 1 with *detailed statistical test results in Table B.2 in Appendix B.*

5.2.1 Psychological Ownership. Overall, psychological ownership when writing with AI seems to plateau beyond a 150 word minimum (Figure 8). There was a significant effect of CONDITION on *Psychological Ownership*, and post hoc tests revealed that people felt less *Psychological Ownership* for all conditions that involved AI than NO-AI ($M = 6.19$, $SD = 0.99$), and there were no significant differences between AI-150 ($M = 4.69$, $SD = 1.64$), AI-175 ($M = 4.70$, $SD = 1.73$), and AI-200 ($M = 4.54$, $SD = 1.81$).

5.2.2 Writing Activities. This plateauing effect is further supported by participants' writing activities. Although *Word Count*, *Text Similarity*, *Mental Demand*, and *Effort* were all strongly positively correlated with *Psychological Ownership* in Experiment 1, none of these correlations held in this experiment ($0.22 \leq r_s \leq 0.48$ for significant correlations). Furthermore, there were no meaningful significant effects of CONDITION on *Text Similarity*, *Mental Demand*, and *Effort*, despite significant differences in *Word Count*.

5.3 Summary

Overall, our results suggest a plateau point around a 150 word minimum for the 150-200 word stories generated in our experiment. Even when participants wrote even longer prompts with AI, they still felt less psychological ownership than writing alone. For this writing task, a prompt length similar to the target word length appears to be a "sweet spot" [39] for trying to maximize psychological ownership when writing with AI, as writing with higher word minimums did not improve psychological ownership.

6 ENCOURAGING LONGER PROMPTS IN A REAL INTERFACE

In our experiments, we had to enforce strict prompt length minimums for the purpose of internal validity and experimental control. But this kind of "hard" constraint may not work well in real-world applications, where users expect more flexibility. How can we encourage users to write more within a typical chat-style AI user interface?

To better understand how users can be nudged towards writing longer prompts without strict word minimums, we designed alternative interface designs and interaction techniques that modify the presentation of the output and the prompt entry interface while remaining highly compatible with current chat-based generative AI tools (Figure C.1; see Appendix C for detailed descriptions). We ran a pilot experiment (n=90, 11-18 per condition) to better understand which modifications were the most promising for increasing prompt length, and our preliminary results[4] suggested that integrating a *time delay* into the prompt submission button could be effective (Figure C.2).

With the time delay modification, the user must press and hold the button for some time to fill up a "gauge," only after which the prompt is submitted. However, this gauge can also be filled by writing more words in the prompt entry text box. These two methods of filling the gauge provide the user with a choice: they can either write a very short prompt and press and hold the submission button for several seconds, or they can write a longer prompt and press and hold the button for a shorter amount of time, or for no time at all if their prompt length met or exceeded a system-defined "word minimum" that leads to higher levels of psychological ownership.

The potential of time delays for increasing prompt length makes sense, as prior work shows that purposely slowing user interactions can effectively nudge users. Specifically, they can cause people to shift their attention and focus [36] and think more analytically [5, 52, 53], which can lead to positive outcomes. For example, time delays can encourage users to reflect on social media posts before they are shared with the public [52]; help users evaluate the validity of output generated by an AI assistant [5]; and even help users learn expert menu selection techniques associated with marking menus [24, 29].

Given these findings, we investigate the use of time delays in more depth to see whether they could increase prompt length. Based on prior work, we suspect that the act of pressing and holding the prompt submission button for several seconds will deter users from writing shorter prompts as they may lose attention and focus, and encourage users to reflect on their prompts and expand them.

7 EXPERIMENT 3: DELAYED SUBMISSION

The goal of this experiment is to understand the effectiveness of a time delay experienced by pressing and holding the prompt submission button. Participants completed a similar creative writing task as the previous two experiments (writing prompts to generate a 150-200 word short story), but were assigned a random time delay condition for the entire experiment. Two baselines included: no modifications, and a strict word minimum like the previous experiments.

[4]Data from these participants is also included in the main results of Experiment 3.

7.1 Participants

We recruited 198 new participants with 42 (21%) removed from the analysis for cheating or experiencing technical issues, leaving 156 valid responses (Table D.1). All reported being proficient at reading and writing in English (all ≥ 4 on a 1-7 scale). Participants received $15 as remuneration.

7.2 Apparatus

The experiment software was similar to that used in the prior experiments, but with a few cosmetic modifications to make the interface more closely resemble existing chat-based AI user interfaces like ChatGPT (Figure 9). The prompt entry and output interfaces were stacked on top of each other. The prompt entry text box, which has a default height of 46 pixels, grows by 46 pixels with every new line typed, until a maximum height of 300 pixels.

With the exception of the baseline condition with no modifications, a brief message underneath the prompt entry text box provides instructions for prompt submission (Figure 9d). To mitigate risks of bias, the message simply explained how the delayed button worked. Inline with the previous experiments, the time delay mechanics used a 150-word minimum for optimal psychological ownership. The baseline condition had to indicate 150 words as the minimum, but the delayed button conditions made no mention of 150 words.

7.2.1 Delayed Button Interaction and Mechanics. A "gauge" that must be filled is visualized as a progress bar below the prompt submission button (Figure 1c). When the button is not pressed, the length of the bar is the ratio of the current prompt word count to the *word minimum* (150 words in this experiment). When the button is held down, the progress bar gradually increases (updated every 500 ms) according to an *experienced delay*. This time delay period is proportional to the unfilled bar ratio to the *maximum delay*. For example, consider a 150-word minimum and a maximum delay of 20 s. If the user writes a 15-word prompt, the experienced delay (the time they must hold the button down) will be 18 s. If the user enters 135 words, the experienced delay will only be 2 s. If the user enters at least 150 words, the experienced delay will be 0 s (i.e., the button works with a normal click). The progress bar is blue when not full, turning green when enough words have been typed or when enough delay has been experienced by pressing and holding the button. Note that the user must press and hold the button in one single act; releasing it before the progress bar turns green resets the progress bar to its original size based on the number of words in the prompt.

7.3 Procedure

The procedure was nearly identical to the prior experiments. However, rather than experiencing all conditions once, participants experienced a single condition four times as we suspected a learning curve for the experimental conditions. As the focus of this study was not psychological ownership, they did not answer questions after every story, instead, they answered questions about their experience after all four trials. The entire experiment took 20 to 60 minutes, depending on the condition.

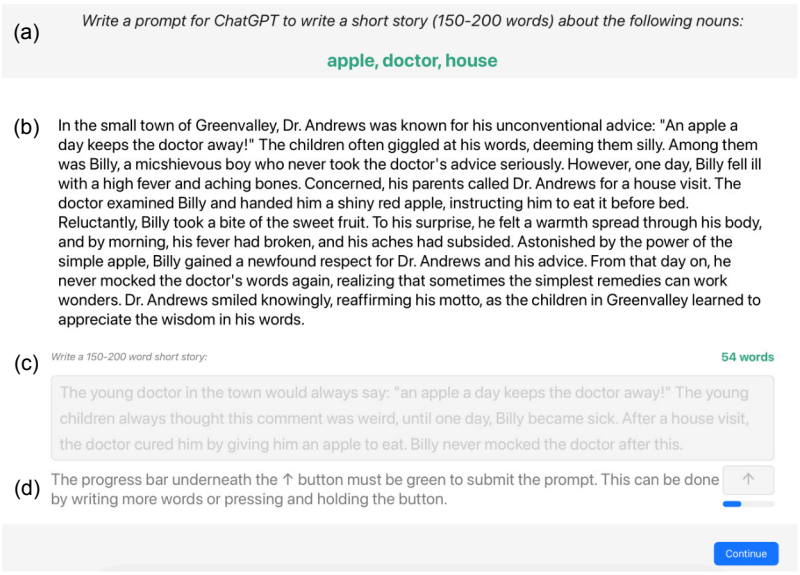

Figure 9: Experimental interface with a time delay: (a) top toolbar showing the instructions and three nouns the story had to be about; (b) the story written with AI assistance; (c) a text box where the participant wrote their prompt; (d) instructions about the technique.

7.4 Choice of Time Delays

Prior work suggests that delays longer than 10 seconds will cause users to shift their attention [36], which is desirable as participants may opt to write a longer prompt instead of pressing and holding the prompt submission button. Delays of 20 seconds are tolerable [46], so the initial pilot experiment described above tested a 20-second delay. The results suggested a bimodal distribution of prompt word length, participants either wrote very short prompts, or very long prompts (Figure C.2).

Based on the pilot, we introduce two additional time delay conditions: an even longer, 60-second delay [46], to increase frustration and encourage more participants to write longer prompts; and a 0-second delay, where the same progress bar visualization was shown underneath the prompt submission button, but the participant did not have to press and hold the button. The latter was included to ensure that the effects were indeed caused by a time delay, rather than confounding factors such as minor changes to the user interface. Therefore, we test the following five conditions:

- **No Modifications** (n=32), a baseline where participants did not see or experience any modifications to the user interface. This best approximates current chat-based AI interfaces.
- **0-Second Delay** (n=32), where the participant sees a progress bar indicating how many words out of 150 they had typed, but do not experience any delay.
- **20-Second Delay** (n=31), where the participant sees the same progress bar indicator and can experience a delay up to 20 seconds long, depending on how many words they write out of 150.
- **60-Second Delay** (n=32), which is similar to the 20-second delay condition, but with a longer, 60-second delay.
- **150-Word Minimum** (n=29), a baseline representing the case where the participant must write a prompt that is at least 150 words long. This condition is guaranteed to increase prompt length, which allows us to make relative comparisons to other conditions.

7.5 Design

This is a mixed-design with two primary independent variables: CONDITION (levels: NONE, 0-SEC, 20-SEC, 60-SEC, WORDS), which was between-subjects and randomly assigned, and TRIAL (levels: 1, 2, 3, 4), which was within-subjects.

The primary measures were objective metrics calculated from logs: the word count of the submitted prompt (*Final Word Count*), the number of times the participant tried to submit a prompt by pressing the prompt submission button (*Submission Attempts*), the word count of the prompt for each submission attempt (*Attempted Word Count*), and the number of seconds they held the button for each submission attempt (*Attempted Duration*). Note that *Submission Attempts*, *Attempted Word Count* and *Attempted Duration* are only applicable for the 20-SEC and 60-SEC conditions, as the prompt submission button could only be pressed once for all other conditions due to the lack of delay (i.e., *Attempted Word Count* is equal to *Final Word Count* for NONE, 0-SEC, and WORDS).

8 EXPERIMENT 3: RESULTS

Where applicable, we use the Aligned Rank Transform (ART) [54] and post hoc contrast tests (ART-C) [15], Kruskal-Wallis tests and Mann-Whitney U tests, Holm's corrections for multiple comparisons, and Spearman's correlations. *Detailed statistical test results are in Table D.2.*

8.1 Final Word Count

Overall, experiencing a time delay when submitting a prompt increased prompt length (Figure 10). A significant main effect of CONDITION on *Final Word Count* and post hoc tests revealed that

20-SEC (M = 76.91, SD = 68.13) and 60-SEC (M = 74.57, SD = 63.80) led to significantly higher *Final Word Count* than NONE (M=34.66, SD=46.56) and 0-SEC (M=28.09, SD=32.69). This was a result of the time delay, rather than changes to the user interface, as there were no significant differences between NONE and 0-SEC. However, as expected, a time delay is not as effective as a strict word minimum, as WORDS had the highest *Final Word Count* (M=164.65, SD=19.54).

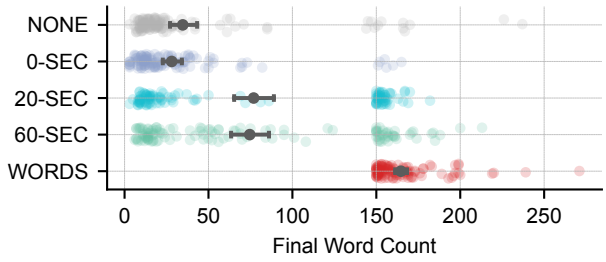

Figure 10: *Final Word Count* by CONDITION.

8.2 Learning Effect

We expected that participants would learn how to use the delayed prompt submission across multiple trials, but this did not occur. Rather, most participants discovered their preferred way of interacting with the button in the first trial, and continued this for all four trials. Specifically, significant main effects of TRIAL on *Submission Attempts* and *Attempted Word Count* and post hoc tests only revealed significant differences between TRIAL 1 and all other trials. Therefore, we focus our discussion of the learning effect on the results from the first trial.

8.2.1 Submission Attempts. Participants made significantly more *Submission Attempts* in the first trial for 20-SEC (M=4.35, SD=4.45) and 60-SEC (M=4.28, SD=4.70) than subsequent trials ($1.06 \leq M \leq 1.84$; $0.24 \leq SD \leq 4.78$). However, the median *Attempted Duration* was only 0.1 seconds for both 20-SEC and 60-SEC, suggesting that participants were initially trying to press the button normally, without holding it down.

8.2.2 Attempted Word Count versus Final Word Count. Examining prompt submission attempts in the first trial, *Attempted Word Count* for 20-SEC (M=32.86, SD=42.55) and 60-SEC (M=36.91, SD=43.96) was on par with the *Final Word Count* of NONE (M=31.59, SD=42.97) and 0-SEC (M = 30.34, SD = 40.81), with no significant differences observed. But 20-SEC and 60-SEC led to a significantly higher *Final Word Count*, suggesting that participants eventually wrote longer prompts after attempting to write shorter prompts. Specifically, in the first trial for 20-SEC and 60-SEC, participants first tried to submit prompts that were 26.81 and 39.16 words long on average, before finally submitting prompts that were 78.10 and 69.03 words long on average, respectively. In other words, *Final Word Count* nearly tripled and doubled in length from the first *Attempted Word Count*.

8.3 Strategies

Overall, there were no significant differences between 20-SEC and 60-SEC for *Final Word Count*, and both techniques invoked similar behaviours from participants (Figure 10): they either wrote short prompts (≤ 50 words), or long prompts (≥ 150 words). Examining individual prompt submission attempts for the first trial for 20-SEC and 60-SEC provides insights into these diverging strategies. Some participants are consistent and do not change their behaviour, with 29 participants (46%) showing no differences between their first *Attempted Word Count* and *Final Word Count*, but 32 participants (51%) wrote more, with 17 of them (27%) writing at least 50 more words, which consisted of more details about the story plot and writing style. These differences suggest three primary types of users and strategies (Figure 11): people who do not need to be nudged as they always write long prompts; people who are more resistant to being nudged and always write shorter prompts; and people who can be nudged to write longer prompts.

Feedback from participants suggested that, as expected, those who do not need to be nudged because they wrote longer prompts did not notice any delay, with comments like *"it's nothing that caught my attention"* (P9). Participants who prefer to write shorter prompts noted that pressing and holding the button was a worthwhile trade-off, for example, *"it required far less time and effort to hold down the button than to write the stories myself"* (P46). This may be related to typing speed. A participant who wrote short prompts said *"[someone] who types fast may not mind [writing more]"* (P136) while some participants said the opposite, for example: *"it was easier to type more into the prompt rather than to hold down the button itself"* (P156).

Some felt like they *"hadn't written enough prompt words"* (P109) when they experienced a delay. For others, the time delay helped them to *"reflect on the prompt"* (P116) and *"make adjustments"* (P105) before it was submitted. This reflection may have encouraged creative thinking, supported by comments like: *"it got me to think more of how I wanted the story to be and pushed me to be more creative"* (P40).

8.4 Summary

Overall, our results suggest that a time delay can nudge people to write longer prompts. Although participants initially tried to submit shorter prompts in the first trial, they eventually submitted prompts that were significantly longer, and repeated this behaviour for subsequent trials. Some participants cannot or do not need to be nudged, but many who might normally write shorter prompts can be encouraged to write more.

9 DISCUSSION

Our first two experiments show that writing with a chat-based generative AI assistant lowers feelings of psychological ownership. These results extend and support prior work studying psychological ownership when writing with other human collaborators and early work examining writing with other forms of AI. The most profound finding from our work is a strong indication that when writing with a chat-based generative AI assistant, psychological ownership can be improved by requiring or encouraging users to write longer prompts. Specifically, writing longer prompts encouraged users to provide more details about the story plot, exerting more control over the output and making it more similar to their original prompt. This often required more mental demand and effort, which are

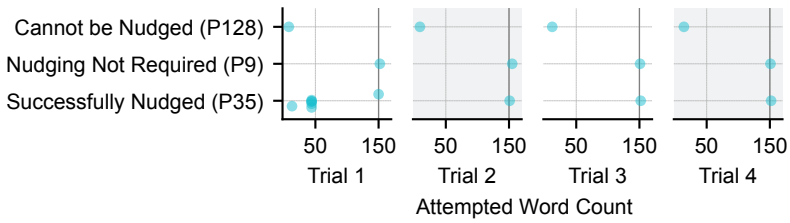

Figure 11: Prompt word count each time the submit button was pressed (i.e., *Attempted Word Count*) for these participants. Examples from 20-sec condition, each *Submission Attempt* is a blue dot.

important contributing factors to psychological ownership. To encourage users to write longer prompts for increased psychological ownership, we propose a simple change to the prompt submission button: the user must press and hold the button for a delay time if their prompt is too short. Our results show that this simple idea can increase prompt length. We believe that our work has the potential to change how we work and collaborate with generative AI assistants for increased psychological ownership.

9.1 Applicability to Other Contexts

Our experimental setup focused on a creative writing context, but our findings and interaction technique could be adapted to suit other writing contexts and even be beneficial for other outcomes beyond psychological ownership.

9.1.1 Other Writing Contexts. People may feel differently about owning different types of text. For example, Nicholes [35] found that undergraduate students feel more psychological ownership towards creative writing while graduate students feel more psychological ownership towards academic writing. Writers may also have different priorities and goals, with some valuing productivity and financial gain over creative fulfillment [1, 30]. Similarly, prior work shows that writers benefit from receiving generated suggestions upon request to overcome writer's block [49] and from receiving feedback from collaborators earlier in the writing process [23]. Generative AI writing assistants can be used as a tool to optimize for productivity and to receive suggestions and feedback, so depending on the goal, reduced psychological ownership may be a worthwhile trade-off for writers. Repeating a similar experiment but for a wider range of document types and writer goals would give insights into these trade-offs. In cases where psychological ownership is less important to writers, shorter prompts may suffice. One possibility is to further modify the interface to elicit goals from writers [23] before using the generative AI assistant, and use these goals to determine whether the system should invoke techniques to encourage longer prompts from the writer. As the writer's goals change, such techniques could dynamically adapt to better suit them. For example, if the user's goal is to write an abstract, the system could require a longer prompt, but this requirement could be lifted or adjusted if the user's goal changes to receive suggestions for a specific sentence.

9.1.2 Beyond Psychological Ownership. Encouraging or requiring longer prompts could be beneficial to encourage other positive outcomes for users. For example, there are growing concerns that *"effortlessly generated information"* provided by generative AI assistants will reduce critical thinking and problem-solving skills among students [21]. Encouraging or enforcing longer prompts that increase the mental demand and effort could potentially mitigate such risks. Specifically, requiring students to write longer prompts could encourage them to put more thought and effort into core learning outcomes before receiving output from a generative AI assistant (e.g., thinking about the arguments and structure of an essay, or thinking about a coding algorithm). Studying the effect of prompt length on other outcomes, like student learning outcomes, is an interesting avenue for future work.

9.2 Implementation Approaches and Challenges

We show that longer prompts can be encouraged through a simple user interface adjustment: adding a submission button time delay. However, there are related design considerations and possibilities for other approaches.

9.2.1 Identifying Optimal "Word Minimums". Our piloted interaction techniques all attempted to nudge users towards a known word minimum of 150 words as our first two experiments suggested that psychological ownership plateaued at 150 words for the given task. This 150-word minimum likely does not apply to other types of tasks and user goals, necessitating additional research and methods for identifying prompt lengths that optimize for psychological ownership. Another option is to create interaction techniques and interface augmentations that do not require any knowledge of optimal word minimums. For example, as the user writes, the generative AI assistant could dynamically append a sentence fragment to the prompt to elicit more information (e.g., "Here is some more background information:").

9.2.2 Other Nudging Techniques. We piloted several modifications to the user interface (Appendix C) before investigating the use of time delays to nudge users to write longer prompts. This approach requires minimal changes and can easily be integrated into current chat-based generative AI interfaces. However, other approaches may also encourage longer prompts and may be more effective than time delays. For example, the prompt entry interface could enforce a different prompt structure through scaffolding techniques [31] (e.g., by breaking the prompt into multiple but smaller responses or requiring the user to "fill in the blanks" of a pre-defined prompt).

9.2.3 Non-Chat-Based Interfaces. Encouraging users to write longer prompts is immediately relevant to chat-based generative AI services like ChatGPT, but as AI becomes more integrated through

techniques like direct manipulation [32], factors other than prompt length may become more dominant. For example, sketches convey more information than text [19], and sketching can be an effective prompting technique [56]. Future work could explore the role of sketching to elicit more details from users for increased psychological ownership, which could benefit writing and even image generation tasks. An important implication of our work is that designers should think of ways to encourage users to invest more when interacting with generative AI assistants for increased psychological ownership.

9.3 Limitations

We chose to explore the role of prompt length within a single prompt, where the user mimics the role of a writing "consultant" [42]. In practice, users likely refine generated text over multiple prompts. A natural extension of our work is to explore the effect of word count across multiple messages within a conversation. This is representative of a broader class of user interactions with systems like ChatGPT, where users iterate over the output through multiple prompts.

One possible reason for the plateauing psychological ownership identified in Experiment 2 may be plateauing semantic text similarity. As models continue to improve to incorporate all details specified in prompts, this plateauing effect may disappear, necessitating further investigations of this effect for newer models. We did test if newer GPT-4 Turbo and 4-o models generated more similar output using the same prompts entered by our participants, but we did not observe any differences in *Text Similarity* between the older and newer models.

Another approach to improve psychological ownership is to design systems that incorporate all details indicated by users and that encourage humans to exert more control over AI-generated output (e.g., by manually accepting suggestions [28]). We chose to implement and investigate systems that work like existing, widely-used commercial tools such as ChatGPT for increased ecological validity. However, as other generative AI systems that are designed to increase psychological ownership become more common, our results could serve as a valuable baseline for future studies, or could even be integrated into new systems for even higher feelings of psychological ownership.

The results from Experiment 3 may be lacking in ecological validity. Our experiment did not provide additional rewards to participants who wrote longer prompts, however, the effects of a time delay on prompt length may have been more pronounced due to the use of crowdworkers on Prolific, who are known to be highly conscientious participants [11]. As such, they may have been more willing to write longer prompts to do the task "well" and less likely to abandon using such systems than the general public. Similarly, the 20- and 60-second time delays worked well in the context of this experiment, but they may not transfer well to real-world systems. It is possible that even shorter time delays, such as 2-second delays studied in prior work [29, 34, 48], may also increase prompt length while mitigating risks of users abandoning the system [22]. Providing users with feedback and detailed explanations on why a delay exists can also increase their willingness to wait [14, 34]. Deploying variations of this nudging technique in real-world systems would give more insights into optimal design parameters that strike a balance between increasing prompt length and preventing system abandonment.

10 CONCLUSION

Our work shows that writing with an AI assistant reduces feelings of psychological ownership, but requiring or encouraging users to write longer prompts can improve it. This is no replacement for writing something independently, but it is a simple way to encourage writers to include more details about the content and put more thought and effort into their prompts. A time delay experienced when submitting a prompt can effectively increase prompt length along with psychological ownership, a simple interface change to positively impact human-AI collaborations.

ACKNOWLEDGMENTS

This work was made possible by NSERC Discovery Grant 2024-03827.

REFERENCES

[1] 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 *Proceedings of the 2022 ACM Designing Interactive Systems Conference* (Virtual Event, Australia) *(DIS '22)*. Association for Computing Machinery, New York, NY, USA, 1209–1227. https://doi.org/10.1145/3532106.3533506

[2] Jeremy Birnholtz, Stephanie Steinhardt, and Antonella Pavese. 2013. Write Here, Write Now!: An Experimental Study of Group Maintenance in Collaborative Writing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. Association for Computing Machinery, New York, NY, USA, 961–970. https://doi.org/10.1145/2470654.2466123

[3] Ina Blau and Avner Caspi. 2009. Sharing and Collaborating with Google Docs: The Influence of Psychological Ownership, Responsibility, and Student's Attitudes on Outcome Quality. In *E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*. Association for the Advancement of Computing in Education (AACE), 3329–3335.

[4] Ina Blau and Avner Caspi. 2009. What Type of Collaboration Helps? Psychological Ownership, Perceived Learning and Outcome Quality of Collaboration using Google Docs. In *Proceedings of the Chais Conference on Instructional Technologies Research*, Vol. 12. 48–55.

[5] Zana Buçinca, Maja Barbara Malaya, and Krzysztof Z. Gajos. 2021. To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 188 (April 2021), 21 pages. https://doi.org/10.1145/3449287

[6] Avner Caspi and Ina Blau. 2011. Collaboration and Psychological Ownership: How does the Tension between the Two Influence Perceived Learning? *Social Psychology of Education* 14 (2011), 283–298.

[7] Daniel Cer, Yinfei Yang, Sheng-yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St John, Noah Constant, Mario Guajardo-Cespedes, Steve Yuan, Chris Tar, et al. 2018. Universal Sentence Encoder for English. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. 169–174.

[8] Christy Choi and Francesca Annio. 2024. The Winner of a Prestigious Japanese Literary Award has Confirmed AI Helped Write her Book. https://www.ctvnews.ca/sci-tech/the-winner-of-a-prestigious-japanese-literary-award-has-confirmed-ai-helped-write-her-book-1.6734205

[9] Andy Cockburn, Per Ola Kristensson, Jason Alexander, and Shumin Zhai. 2007. Hard Lessons: Effort-Inducing Interfaces Benefit Spatial Learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '07)*. Association for Computing Machinery, New York, NY, USA, 1571–1580. https://doi.org/10.1145/1240624.1240863

[10] Paramveer S. Dhillon, Somayeh Molaei, Jiaqi Li, Maximilian Golub, Shaochun Zheng, and Lionel Peter Robert. 2024. Shaping Human-AI Collaboration: Varied Scaffolding Levels in Co-writing with Language Models. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 1044, 18 pages. https://doi.org/10.1145/3613904.3642134

[11] Benjamin D Douglas, Patrick J Ewell, and Markus Brauer. 2023. Data Quality in Online Human-subjects Research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. *Plos one* 18, 3 (2023), e0279720.

[12] Fiona Draxler, Anna Werner, Florian Lehmann, Matthias Hoppe, Albrecht Schmidt, Daniel Buschek, and Robin Welsch. 2024. The AI Ghostwriter Effect: When Users do not Perceive Ownership of AI-Generated Text but Self-Declare as Authors. *ACM Trans. Comput.-Hum. Interact.* 31, 2, Article 25 (feb 2024), 40 pages. https://doi.org/10.1145/3637875

[13] Mark D Dunlop, Emma Nicol, Andreas Komninos, Prima Dona, and Naveen Durga. 2016. Measuring Inviscid Text Entry using Image Description Tasks. In *Proceedings of the 2016 CHI Workshop on Inviscid Text Entry and Beyond.*

[14] Serge Egelman, David Molnar, Nicolas Christin, Alessandro Acquisti, Cormac Herley, and Shriram Krishnamurthi. 2010. Please Continue to Hold. In *Ninth Workshop on the Economics of Information Security.*

[15] Lisa A. Elkin, Matthew Kay, James J. Higgins, and Jacob O. Wobbrock. 2021. An Aligned Rank Transform Procedure for Multifactor Contrast Tests. In *The 34th Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '21).* Association for Computing Machinery, New York, NY, USA, 754–768. https://doi.org/10.1145/3472749.3474784

[16] Margaret Foley, Géry Casiez, and Daniel Vogel. 2020. Comparing Smartphone Speech Recognition and Touchscreen Typing for Composition and Transcription. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20).* Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3313831.3376861

[17] Katy Ilonka Gero and Lydia B. Chilton. 2019. Metaphoria: An Algorithmic Companion for Metaphor Creation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19).* Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300526

[18] Aaron Halfaker, Aniket Kittur, Robert Kraut, and John Riedl. 2009. A Jury of your Peers: Quality, Experience and Ownership in Wikipedia. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration* (Orlando, Florida) *(WikiSym '09).* Association for Computing Machinery, New York, NY, USA, Article 15, 10 pages. https://doi.org/10.1145/1641309.1641332

[19] Catrinel Haught and Philip N Johnson-Laird. 2003. Creativity and Constraints: The Production of Novel Sentences. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 25.

[20] Chieh-Yang Huang, Sanjana Gautam, Shannon McClellan Brooks, Ya-Fang Lin, Tiffany Knearem, and Ting-Hao 'Kenneth' Huang. 2024. Inspo: Writing with Crowds Alongside AI. arXiv:2311.16521 [cs.HC] https://arxiv.org/abs/2311.16521

[21] Enkelejda Kasneci, Kathrin Sessler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günnemann, Eyke Hüllermeier, Stephan Krusche, Gitta Kutyniok, Tilman Michaeli, Claudia Nerdel, Jürgen Pfeffer, Oleksandra Poquet, Michael Sailer, Albrecht Schmidt, Tina Seidel, Matthias Stadler, Jochen Weller, Jochen Kuhn, and Gjergji Kasneci. 2023. ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education. *Learning and Individual Differences* 103 (2023), 102274. https://doi.org/10.1016/j.lindif.2023.102274

[22] Stefan Köpsell. 2006. Low Latency Anonymous Communication – How Long Are Users Willing to Wait?. In *Emerging Trends in Information and Communication Security*, Günter Müller (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 221–237.

[23] Sneha R. Krishna Kumaran, Yue Yin, and Brian P. Bailey. 2021. Plan Early, Revise More: Effects of Goal Setting and Perceived Role of the Feedback Provider on Feedback Seeking Behavior. 5, CSCW1, Article 24 (April 2021), 22 pages. https://doi.org/10.1145/3449098

[24] Gordon Kurtenbach and William Buxton. 1994. User Learning and Performance with Marking Menus. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* 258–264.

[25] Ida Larsen-Ledet and Henrik Korsgaard. 2019. Territorial Functioning in Collaborative Writing: Fragmented Exchanges and Common Outcomes. *Comput. Supported Coop. Work* 28, 3–4 (Jun 2019), 391–433. https://doi.org/10.1007/s10606-019-09359-8

[26] Mina Lee, Katy Ilonka Gero, John Joon Young Chung, Simon Buckingham Shum, Vipul Raheja, Hua Shen, Subhashini Venugopalan, Thiemo Wambsganss, David Zhou, Emad A. Alghamdi, Tal August, Avinash Bhat, Madiha Zahrah Choksi, Senjuti Dutta, Jin L.C. Guo, Md Naimul Hoque, Yewon Kim, Simon Knight, Seyed Parsa Neshaei, Antonette Shibani, Disha Shrivastava, Lila Shroff, Agnia Sergeyuk, Jessi Stark, Sarah Sterman, Sitong Wang, Antoine Bosselut, Daniel Buschek, Joseph Chee Chang, Sherol Chen, Max Kreminski, Joonsuk Park, Roy Pea, Eugenia Ha Rim Rho, Zejiang Shen, and Pao Siangliulue. 2024. A Design Space for Intelligent and Interactive Writing Assistants. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24).* Association for Computing Machinery, New York, NY, USA, Article 1054, 35 pages. https://doi.org/10.1145/3613904.3642697

[27] 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* (New Orleans, LA, USA) *(CHI '22).* Association for Computing Machinery, New York, NY, USA, Article 388, 19 pages. https://doi.org/10.1145/3491102.3502030

[28] Florian Lehmann, Niklas Markert, Hai Dang, and Daniel Buschek. 2022. Suggestion Lists vs. Continuous Generation: Interaction Design for Writing with Generative Models on Mobile Devices Affect Text Length, Wording and Perceived Authorship. In *Proceedings of Mensch Und Computer 2022* (Darmstadt, Germany) *(MuC '22).* Association for Computing Machinery, New York, NY, USA, 192–208. https://doi.org/10.1145/3543758.3543947

[29] Blaine Lewis and Daniel Vogel. 2020. Longer Delays in Rehearsal-based Interfaces Increase Expert Use. *ACM Trans. Comput.-Hum. Interact.* 27, 6, Article 45 (nov 2020), 41 pages. https://doi.org/10.1145/3418196

[30] 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 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24).* Association for Computing Machinery, New York, NY, USA, Article 1048, 25 pages. https://doi.org/10.1145/3613904.3642625

[31] Kurt Luther, Jari-Lee Tolentino, Wei Wu, Amy Pavel, Brian P. Bailey, Maneesh Agrawala, Björn Hartmann, and Steven P. Dow. 2015. Structuring, Aggregating, and Evaluating Crowdsourced Design Critique. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Vancouver, BC, Canada) *(CSCW '15).* Association for Computing Machinery, New York, NY, USA, 473–485. https://doi.org/10.1145/2675133.2675283

[32] Damien Masson, Sylvain Malacria, Géry Casiez, and Daniel Vogel. 2024. DirectGPT: A Direct Manipulation Interface to Interact with Large Language Models. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24).* Association for Computing Machinery, New York, NY, USA, Article 975, 16 pages. https://doi.org/10.1145/3613904.3642462

[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* (Hamburg, Germany) *(CHI '23).* Association for Computing Machinery, New York, NY, USA, Article 355, 34 pages. https://doi.org/10.1145/3544548.3581225

[34] Fiona Fui-Hoon Nah. 2004. A Study on Tolerable Waiting Time: How Long are Web Users Willing to Wait? *Behaviour & Information Technology* 23, 3 (2004), 153–163.

[35] Justin Nicholes. 2017. Measuring Ownership of Creative Versus Academic Writing: Implications for Interdisciplinary Praxis.

[36] Jakob Nielsen. 1993. Response Times: The 3 Important Limits. https://www.nngroup.com/articles/response-times-3-important-limits/

[37] Changhoon Oh, Jungwoo Song, Jinhan Choi, Seonghyeon Kim, Sungwoo Lee, and Bongwon Suh. 2018. I Lead, You Help but Only with Enough Details: Understanding User Experience of Co-Creation with Artificial Intelligence. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal, QC, Canada) *(CHI '18).* Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3174223

[38] Judith S. Olson, Dakuo Wang, Gary M. Olson, and Jingwen Zhang. 2017. How People Write Together Now: Beginning the Investigation with Advanced Undergraduates in a Project Course. *ACM Trans. Comput.-Hum. Interact.* 24, 1, Article 4 (mar 2017), 40 pages. https://doi.org/10.1145/3038919

[39] Balder Onarheim and Michael Mose Biskjaer. 2015. Balancing Constraints and the Sweet Spot as Coming Topics for Creativity Research. *Creativity in Design: Understanding, Capturing, Supporting* 1 (2015), 1–18.

[40] Jon L. Pierce, Tatiana Kostova, and Kurt T. Dirks. 2001. Toward a Theory of Psychological Ownership in Organizations. *The Academy of Management Review* 26, 2 (2001), 298–310. https://doi.org/10.2307/259124

[41] Jon L. Pierce, Tatiana Kostova, and Kurt T. Dirks. 2003. The State of Psychological Ownership: Integrating and Extending a Century of Research. *Review of General Psychology* 7, 1 (2003), 84–107. https://doi.org/10.1037/1089-2680.7.1.84

[42] Ilona R. Posner and Ronald M. Baecker. 1992. How People Write Together (Groupware). In *Proceedings of the Twenty-Fifth Hawaii International Conference on System Sciences*, Vol. iv. 127–138 vol.4. https://doi.org/10.1109/HICSS.1992.183420

[43] Jeba Rezwana and Mary Lou Maher. 2023. User Perspectives on Ethical Challenges in Human-AI Co-Creativity: A Design Fiction Study. In *Proceedings of the 15th Conference on Creativity and Cognition* (Virtual Event, USA) *(C&C '23).* Association for Computing Machinery, New York, NY, USA, 62–74. https://doi.org/10.1145/3591196.3593364

[44] Quentin Roy, Futian Zhang, and Daniel Vogel. 2019. Automation Accuracy Is Good, but High Controllability May Be Better. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19).* Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3290605.3300750

[45] Timothy Ryan. 2020. Fraudulent Responses on Amazon Mechanical Turk: A Fresh Cautionary Tale. https://timryan.web.unc.edu/2020/12/22/fraudulent-responses-on-amazon-mechanical-turk-a-fresh-cautionary-tale/

[46] Paula Selvidge. 1999. How Long is Too Long to Wait for a Website to Load. *Usability News* 1, 2 (1999), 1–3.

[47] Hua Shen and Tongshuang Wu. 2023. Parachute: Evaluating Interactive Human-LM Co-writing Systems. arXiv:2303.06333 [cs.HC] https://arxiv.org/abs/2303.06333

[48] Ben Shneiderman. 1984. Response Time and Display Rate in Human Performance with Computers. *ACM Comput. Surv.* 16, 3 (Sept. 1984), 265–285. https://doi.org/10.1145/2514.2517

[49] Pao Siangliulue, Joel Chan, Krzysztof Z. Gajos, and Steven P. Dow. 2015. Providing Timely Examples Improves the Quantity and Quality of Generated Ideas. In *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition* (Glasgow, United Kingdom) *(C&C '15)*. Association for Computing Machinery, New York, NY, USA, 83–92. https://doi.org/10.1145/2757226.2757230

[50] Jennifer Thom-Santelli, Dan R. Cosley, and Geri Gay. 2009. What's Mine is Mine: Territoriality in Collaborative Authoring. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 1481–1484. https://doi.org/10.1145/1518701.1518925

[51] QianYing Wang, Alberto Battocchi, Ilenia Graziola, Fabio Pianesi, Daniel Tomasini, Massimo Zancanaro, and Clifford Nass. 2006. The Role of Psychological Ownership and Ownership Markers in Collaborative Working Environment *(ICMI '06)*. Association for Computing Machinery, New York, NY, USA, 225–232. https://doi.org/10.1145/1180995.1181041

[52] Yang Wang, Pedro Giovanni Leon, Alessandro Acquisti, Lorrie Faith Cranor, Alain Forget, and Norman Sadeh. 2014. A Field Trial of Privacy Nudges for Facebook. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 2367–2376. https://doi.org/10.1145/2556288.2557413

[53] P.C. Wason and J.ST.B.T. Evans. 1974. Dual Processes in Reasoning? *Cognition* 3, 2 (1974), 141–154. https://doi.org/10.1016/0010-0277(74)90017-1

[54] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The Aligned Rank Transform for Nonparametric Factorial Analyses using only Anova Procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 143–146. https://doi.org/10.1145/1978942.1978963

[55] Catherine Yeh, Gonzalo Ramos, Rachel Ng, Andy Huntington, and Richard Banks. 2025. GhostWriter: Augmenting Collaborative Human-AI Writing Experiences Through Personalization and Agency. arXiv:2402.08855 [cs.HC] https://arxiv.org/abs/2402.08855

[56] Ryan Yen, Jian Zhao, and Daniel Vogel. 2025. Code Shaping: Iterative Code Editing with Free-form AI-Interpreted Sketching. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 872, 17 pages. https://doi.org/10.1145/3706598.3713822

[57] Soobin Yim, Dakuo Wang, Judith Olson, Viet Vu, and Mark Warschauer. 2017. Synchronous Collaborative Writing in the Classroom: Undergraduates' Collaboration Practices and their Impact on Writing Style, Quality, and Quantity. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) *(CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 468–479. https://doi.org/10.1145/2998181.2998356

[58] 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* (Helsinki, Finland) *(IUI '22)*. Association for Computing Machinery, New York, NY, USA, 841–852. https://doi.org/10.1145/3490099.3511105

A EXPERIMENT 1: ANALYSIS DETAILS

Table A.1: Demographic information for Experiment 1, adapted from Masson et al. [32].

Gender		Age		Education		Creative Writing Frequency		Prompt Engineering Familiarity	
Men	14	25-34	10	Some University (no credit)	4	Daily	1	Extremely	3
Women	16	35-44	7	Technical Degree	1	Weekly	7	Moderately	3
Non-binary	1	45-54	6	Bachelor's Degree	21	Monthly	7	Somewhat	3
		55-64	6	Master's Degree	5	Less than Monthly	11	Slightly	7
		65-74	2			Never	5	Not at All	15

ChatGPT Frequency		Weekly ChatGPT Use		Other AI Services				AI Usage			
Daily	3	0 times	7	Nothing	3	Gemini	9	Nothing	1	Generate Images	12
Weekly	14	1-5 times	13	GPT-2	6	Grammarly	9	Write Emails	11	Get a Definition	9
Monthly	4	5-10 times	8	GPT-3	12	Stable Diffusion	2	Write Papers/Essays	2	Explore a Topic	17
Less than Monthly	9	10-30 times	2	GPT-4	6	Dall-E	4	Write Stories	6	Brainstorming	25
Never	1	30+ times	1	Copilot	2	Midjourney	5	Edit Text	14	Find References	8
				Bing	12	Other	6	Generate Code	3	Clarification	7
								Edit Code	3	Translate Text	5
								Debug Code	3	Other	4

Table A.2: Omnibus and post hoc statistical tests for Experiment 1: (a) *Psychological Ownership*, (b) *Duration*, (c) *Word Count*, (d) *Text Similarity*, (e) *Mental Demand*, and (f) *Effort*.

(a) *Psychological Ownership*

$\chi^2_{3,N=31} = 78.01$, $p < .001$, $W = .84$

comparisons		*p-value*	
AI-3	AI-50	< .001	***
AI-3	AI-150	< .001	***
AI-3	NO-AI	< .001	***
AI-50	AI-150	.01	**
AI-50	NO-AI	< .001	***
AI-150	NO-AI	< .001	***

(b) *Duration*

$\chi^2_{3,N=31} = 35.86$, $p < .001$, $W = .38$

comparisons		*p-value*	
AI-3	AI-50	.87	*n.s.*
AI-3	AI-150	.02	*
AI-3	NO-AI	.01	**
AI-50	AI-150	< .001	***
AI-50	NO-AI	< .001	***
AI-150	NO-AI	.24	*n.s.*

(c) *Word Count*

$\chi^2_{3,N=31} = 85.87$, $p < .001$, $W = .92$

comparisons		*p-value*	
AI-3	AI-50	< .001	***
AI-3	AI-150	< .001	***
AI-3	NO-AI	< .001	***
AI-50	AI-150	< .001	***
AI-50	NO-AI	< .001	***
AI-150	NO-AI	.005	**

(d) *Text Similarity*

$\chi^2_{3,N=31} = 99.00$, $p < .001$, $W = 1.00$

comparisons		*p-value*	
AI-3	AI-50	< .001	***
AI-3	AI-150	< .001	***
AI-3	NO-AI	< .001	***
AI-50	AI-150	< .001	***
AI-50	NO-AI	< .001	***
AI-150	NO-AI	< .001	***

(e) *Mental Demand*

$\chi^2_{3,N=31} = 60.21$, $p < .001$, $W = .65$

comparisons		*p-value*	
AI-3	AI-50	< .001	***
AI-3	AI-150	< .001	***
AI-3	NO-AI	< .001	***
AI-50	AI-150	< .001	**
AI-50	NO-AI	< .001	***
AI-150	NO-AI	.33	*n.s.*

(f) *Effort*

$\chi^2_{3,N=31} = 49.99$, $p < .001$, $W = .54$

comparisons		*p-value*	
AI-3	AI-50	< .001	***
AI-3	AI-150	< .001	***
AI-3	NO-AI	< .001	***
AI-50	AI-150	.16	*n.s.*
AI-50	NO-AI	.02	*
AI-150	NO-AI	.13	*n.s.*

B EXPERIMENT 2: ANALYSIS DETAILS

Table B.1: Demographic information for Experiment 2.

Gender		Age		Education		Creative Writing Frequency		Prompt Engineering Familiarity	
Men	12	18-24	3	High School Diploma	3	Daily	2	Extremely	1
Women	19	25-34	11	Some University (no credit)	8	Weekly	3	Moderately	3
Non-binary	1	35-44	4	Technical Degree	2	Monthly	7	Somewhat	7
Other	1	45-54	11	Bachelor's Degree	18	Less than Monthly	12	Slightly	12
Unknown	1	55-64	4	Master's Degree	2	Never	10	Not at All	11
		Unknown	1	Doctorate Degree	1				

ChatGPT Frequency		Weekly ChatGPT Use		Other AI Services				AI Usage			
Daily	10	0 times	8	Nothing	6	Gemini	7	Nothing	2	Generate Images	14
Weekly	8	1-5 times	12	GPT-2	6	Grammarly	8	Write Emails	14	Get a Definition	17
Monthly	6	5-10 times	4	GPT-3	12	Stable Diffusion	3	Write Papers/Essays	4	Explore a Topic	22
Less than Monthly	6	10-30 times	6	GPT-4	6	Dall-E	9	Write Stories	5	Brainstorming	19
Never	4	30+ times	4	Copilot	0	Midjourney	6	Edit Text	15	Find References	5
				Bing	10	Other	4	Generate Code	3	Clarification	16
								Edit Code	3	Translate Text	5
								Debug Code	3	Other	3

Table B.2: Omnibus and post hoc statistical tests for Experiment 2: (a) *Psychological Ownership*, (b) *Duration*, (c) *Word Count*, and (d) *Text Similarity*. Note that *Mental Demand* and *Effort* are excluded as these omnibus tests were not significant.

(a) *Psychological Ownership*: $\chi^2_{3,N=34} = 43.63$, $p < .001$, $W = .43$

comparisons		*p-value*	
AI-150	AI-175	.69	*n.s.*
AI-150	AI-200	.69	*n.s.*
AI-150	NO-AI	< .001	***
AI-175	AI-200	.69	*n.s.*
AI-175	NO-AI	< .001	***
AI-200	NO-AI	< .001	***

(b) *Duration*: $\chi^2_{3,N=34} = 11.32$, $p < .01$, $W = .11$

comparisons		*p-value*	
AI-150	AI-175	1.0	*n.s.*
AI-150	AI-200	.03	*
AI-150	NO-AI	1.0	*n.s.*
AI-175	AI-200	.14	*n.s.*
AI-175	NO-AI	1.0	*n.s.*
AI-200	NO-AI	.13	*n.s.*

(c) *Word Count*: $\chi^2_{3,N=34} = 71.27$, $p < .001$, $W = .70$

comparisons		*p-value*	
AI-150	AI-175	< .001	***
AI-150	AI-200	< .001	***
AI-150	NO-AI	.16	*n.s.*
AI-175	AI-200	< .001	***
AI-175	NO-AI	.003	**
AI-200	NO-AI	< .001	***

(d) *Text Similarity*: $\chi^2_{3,N=34} = 61.94$, $p < .001$, $W = .61$

comparisons		*p-value*	
AI-150	AI-175	.46	*n.s.*
AI-150	AI-200	.46	*n.s.*
AI-150	NO-AI	< .001	***
AI-175	AI-200	.77	*n.s.*
AI-175	NO-AI	< .001	***
AI-200	NO-AI	< .001	***

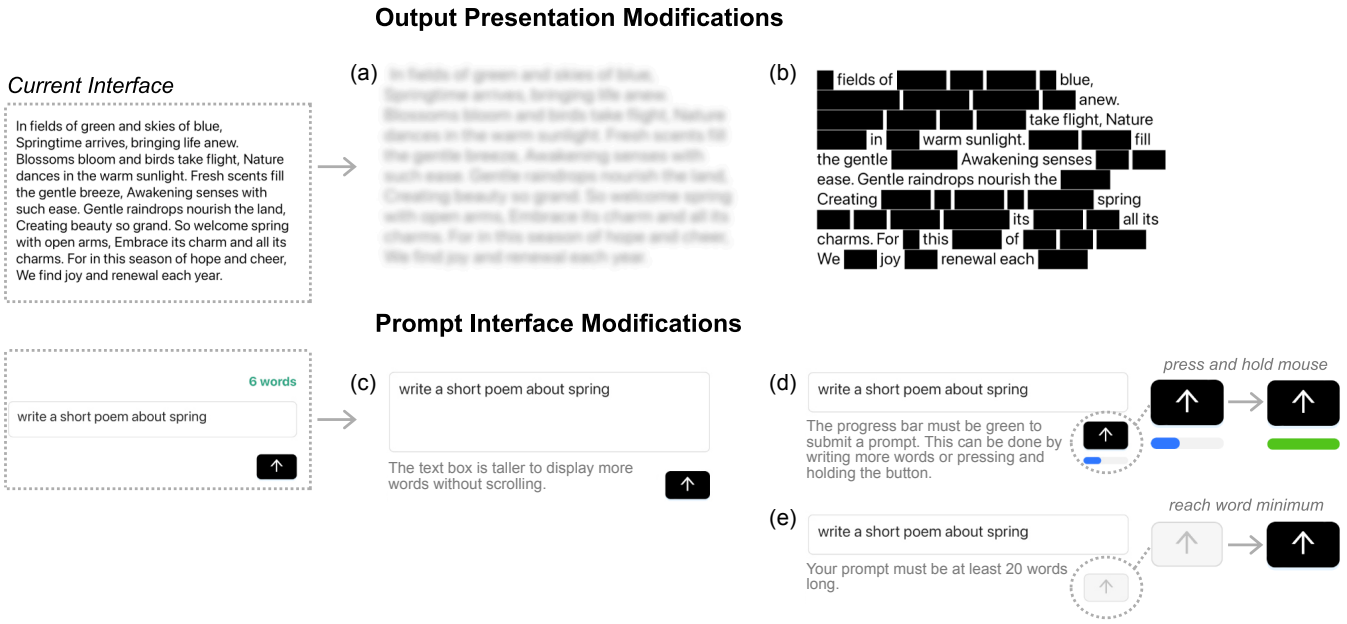

Figure C.1: Prompt entry interface designs that augment the prompt entry area and the response area: (a) blurring or (b) redacting the text more if the prompt contained less words; or nudging users to write longer prompts by (c) making the prompt entry text box taller or (d) requiring the user to press and hold the submit button. We compared these to the baseline "hard" constraint of (e) preventing prompt submission if the prompt is not long enough.

C PRELIMINARY INTERFACE MODIFICATIONS AND INTERACTION TECHNIQUES

Before deciding to explore the use of time delays to increase prompt length in more depth, we also considered the following modifications to both the output presentation and the prompt entry interfaces (Figure C.1). These modifications were the result of brainstorming sessions between the authors, and were selected as they are compatible with current chat-based generative AI tools like ChatGPT.

- **Increasingly blurring the generated output** the shorter their prompt is, requiring the user to press and hold it make the output clearer and easier to read (Figure C.1a). Much like the time delay technique, the time spent pressing and holding the output is mapped to prompt length. This technique was inspired by Cockburn et al.'s frost-brushing technique for teaching users about the spatial information of user interfaces [9].
- **Increasingly redacting parts of the generated output** the shorter their prompt is, "punishing" the user if the prompt length was too short (Figure C.1b). If the user writes a shorter prompt, each chunk of generated output has a higher chance of being redacted. Having a few chunks redacted unlikely hinders overall comprehension as the user can easily "fill in the blanks," but if a majority of the text is redacted and the output is illegible, the user has to increase the length of their existing prompt to see more of the generated text next time.
- **Increasing the height of the prompt entry text box** so it can fully hold the word minimum (150 words).

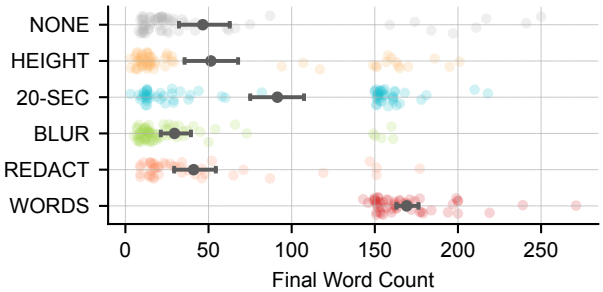

Figure C.2: *Final Word Count* by CONDITION for the initial pilot experiment with all interface modifications.

D EXPERIMENT 3: ANALYSIS DETAILS

Table D.1: Demographic information for Experiment 3.

Gender		Age		Education		Creative Writing Frequency		Prompt Engineering Familiarity	
Men	93	18-24	5	Less than a High School Diploma	2	Daily	2	Extremely	8
Women	61	25-34	47	High School Diploma	22	Weekly	16	Moderately	26
Non-binary	2	35-44	51	Some University (no credit)	37	Monthly	31	Somewhat	28
		45-54	25	Technical Degree	9	Less than Monthly	73	Slightly	44
		55-64	16	Bachelor's Degree	65	Never	34	Not at All	50
		65-74	11	Professional Degree Beyond Bachelor's Degree	1				
		75+	1	Master's Degree	20				

ChatGPT Frequency		Weekly ChatGPT Use		Other AI Services				AI Usage			
Daily	19	0 times	34	Nothing	12	Gemini	65	Nothing	4	Generate Images	68
Weekly	76	1-5 times	70	GPT-2	37	Grammarly	54	Write Emails	52	Get a Definition	59
Monthly	27	5-10 times	32	GPT-3	79	Stable Diffusion	12	Write Papers/Essays	21	Explore a Topic	111
Less than Monthly	26	10-30 times	16	GPT-4	60	Dall-E	36	Write Stories	35	Brainstorming	100
Never	8	30+ times	4	Copilot	14	Midjourney	19	Edit Text	65	Find References	35
				Bing	82	Other	13	Generate Code	26	Clarification	67
								Edit Code	19	Translate Text	33
								Debug Code	15	Other	11

Table D.2: Omnibus and post hoc statistical tests for Experiment 3: (a) *Final Word Count*, (b) *Submission Attempts*, (c) *Attempted Word Count* for TRIAL = 1.

(a) *Final Word Count*

CONDITION ($F_{4,151} = 26.45$, $p < .001$, $\eta_P^2 = 0.41$) TRIAL ($F_{3,453} = 2.87$, $p < .05$, $\eta_P^2 = 0.02$)

comparisons		*p-value*					
NONE	0-SEC	1.00	*n.s.*	1	2	.11	*n.s.*
NONE	20-SEC	.042	*	1	3	.13	*n.s.*
NONE	60-SEC	.013	*	1	4	.049	*
NONE	WORDS	< .001	***	2	3	1.00	*n.s.*
0-SEC	20-SEC	.013	*	2	4	1.00	*n.s.*
0-SEC	60-SEC	.0024	**	3	4	1.00	*n.s.*
0-SEC	WORDS	< .001	***				
20-SEC	60-SEC	1.00	*n.s.*				
20-SEC	WORDS	< .001	***				
60-SEC	WORDS	< .001	***				

(b) *Submission Attempts*

TRIAL ($F_{3,183} = 96.95$, $p < .001$, $\eta_P^2 = 0.61$)

comparisons		*p-value*	
1	2	< .001	***
1	3	< .001	***
1	4	< .001	***
2	3	.15	*n.s.*
2	4	.84	*n.s.*
3	4	.16	*n.s.*

(c) *Attempted Word Count*

CONDITION ($\chi^2_{4,N=156} = 71.11$, $p < .001$, $\eta^2 = .17$)

comparisons		*p-value*	
NONE	0-SEC	1.00	*n.s.*
NONE	20-SEC	1.00	*n.s.*
NONE	60-SEC	1.00	*n.s.*
NONE	WORDS	< .001	***
0-SEC	20-SEC	1.00	*n.s.*
0-SEC	60-SEC	1.00	*n.s.*
0-SEC	WORDS	< .001	***
20-SEC	60-SEC	1.00	*n.s.*
20-SEC	WORDS	< .001	***
60-SEC	WORDS	< .001	***