Proactive AI as a Catalyst for Creativity? Balancing Human Agency and AI Contribution in Collaborative Story Writing

Yiwen Yin
Department of Computer Science and Technology, BNRist
Tsinghua University
Beijing, China
yinyw21@mails.tsinghua.edu.cn

Mingze Wu
Institute for Human-centered AI
Stanford University
Stanford, California, USA
lawtedwu@gmail.com

Ruijie Huang
School of Artificial Intelligence
Beijing University of Posts and Telecommunications
Beijing, China
ruijhuang@163.com

Xin Tong
School of Information
University of Michigan Ann Arbor
Ann Arbor, Michigan, USA
hellenxt@umich.edu

Junyu Zhou
School of Humanities and Social Sciences
University of Science and Technology of China
Hefei, China
373483751@qq.com

Chun Yu*
Department of Computer Science and Technology, BNRist, College of AI
Tsinghua University
Beijing, China
Key Laboratory of Pervasive Computing
Ministry of Education
Beijing, China
chunyu@tsinghua.edu.cn

Yuanchun Shi
Department of Computer Science and Technology
Tsinghua University
Beijing, China
Qinghai University
Xining, Qinghai, China
shiyc@tsinghua.edu.cn

Abstract

Large Language Models (LLMs) hold promise in supporting creative writing, yet the role of proactive AI in collaborative writing remains underexplored due to concerns around human agency and disruption. To investigate effective strategies for proactive AI support, we conducted a Wizard-of-Oz study simulating two suggestion styles: **intrusive suggestions (next-sentence completions)** and **non-intrusive suggestions (exploratory proposals)**, where participants completed two story outlining tasks under each style, receiving real-time proactive suggestions from a human wizard acting as the AI. Both quantitative and qualitative results show that proactive AI can enhance creativity and accelerate writing. However, we observed a trade-off between AI involvement and perceived human agency. This trade-off was moderated by how strongly AI stimulated users–greater inspiration led to stronger perceived agency even under high AI involvement. Based on wizards' behavior, we offer guidance on suggestion style and timing to better balance creativity and agency for future proactive AI writing systems.

CCS Concepts

• **Human-centered computing → Empirical studies in interaction design**.

Keywords

Proactive AI, Human-AI Collaborative Writing, Human Agency

*Corresponding author.

CHI '26, Barcelona, Spain

ACM ISBN 979-8-4007-2278-3/26/04
https://doi.org/10.1145/3772318.3790848

ACM Reference Format:
Yiwen Yin, Mingze Wu, Ruijie Huang, Xin Tong, Junyu Zhou, Chun Yu, and Yuanchun Shi. 2026. Proactive AI as a Catalyst for Creativity? Balancing Human Agency and AI Contribution in Collaborative Story Writing. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3772318.3790848

1 Introduction

AI-powered writing assistants have rapidly evolved in recent years [6, 7, 10, 28, 48], offering support through a variety of interaction modes–ranging from user-initiated prompting [13, 46] to turn-taking between human and AI collaborators [17, 31, 47]. Among these, proactive AI assistance–where the system actively infers user needs and offers suggestions without explicit prompts–represents a particularly involved and interventionist approach [28]. Unlike reactive tools, proactive AI has the potential to make substantive contributions to idea generation, narrative flow, and writing momentum, positioning it as a promising direction for co-creative writing tasks such as storytelling.

Despite its potential, proactive AI also raises important design considerations. Specifically, the loss of human agency and interruptiveness have emerged as two critical concerns in the context of AI-augmented creativity. On one hand, proactive systems that offer content-rich suggestions run the risk of overstepping, thereby diminishing users' sense of authorship, control, and creative ownership [19, 36]. Balancing between human autonomy and creativity gain is non-trivial: excessive AI intervention may undermine the user's autonomy, while too little support may fail to provide the necessary scaffolding for ideation and productivity. On the other hand, unsolicited suggestions–if ill-timed or irrelevant–can disrupt users' concentration, potentially hindering rather than helping the writing process. This aligns with broader findings in HCI that highlight the cognitive cost of interruptions, particularly in creative tasks where immersion and momentum are essential [1, 2]. These tensions raise a fundamental question: Can proactive AI interventions enhance creativity without compromising human agency? And if so, what kinds of suggestion strategies are most effective? These considerations motivate our research questions:

- (RQ1) How proactive AI impacts user's writing process?
- (RQ2) How do different styles of proactive suggestions influence user creativity and writing experience?
- (RQ3) How does user information help AI better provide proactive intervention?

To explore these questions, we designed a Wizard-of-Oz study simulating two proactive suggestion styles: **Intrusive suggestions**, which continue the user's text with next-sentence completions, and **non-intrusive suggestions**, which offer exploratory, question-style proposals intended to inspire creativity without providing directly applicable content. Participants completed two story writing tasks, each under a different suggestion style, while receiving real-time proactive suggestions from a human wizard simulating the AI. The wizard had access to an initial set of suggestions generated by an AI assistant and was able to observe multimodal user signals, including webcam feeds, on-screen writing activity, and think-aloud protocols, to provide fully informed suggestions throughout the task. We conducted a within-subjects study (N = 30) to investigate how these two proactive AI styles influence participants' experiences during co-creative writing, focusing on perceived autonomy, AI contribution, creative stimulation, and perceived skill development (RQ1 & RQ2). Additionally, we collected data on the wizard's suggestion strategies to understand how they interpreted user cues and adapted AI behavior in real time (RQ3).

Our findings show that proactive AI can significantly accelerate writing. However, the two suggestion styles revealed a clear trade-off between perceived AI contribution and human agency. We introduced the measure of AI-augmented Inspiration, which quantifies user's creativity gains indirectly stimulated by AI, and found that high AI-augmented Inspiration mitigates this trade-off, enabling users to experience strong agency even under high AI involvement. This suggests that the most effective human–AI collaboration occurs not when users fully accept AI input, but when they reflect on, internalize, and iteratively build upon AI suggestions. In such cases, the AI acts as a catalyst for inspiration, providing mutual and iterative stimulation that fosters idea development without dominating the creative process. Furthermore, through post-study interviews with wizards, we analyzed how they interpreted multimodal user cues to inform suggestion timing and content. From this, we derived actionable design guidelines for future proactive AI systems–emphasizing strategies that support creativity without undermining user agency. Overall, our contributions are:

- An AI-assisted writing platform that provides real-time and proactive suggestions, supporting Wizard-of-Oz experimentation.
- Empirical evidence on how different AI intervention styles impact AI contribution, creative stimulation and perceived human agency, along with practical insights for designing proactive AI writing tools.

2 Related Work

2.1 AI-assisted Writing

Advances in LLMs are reshaping human writing practice by enabling AI to take on a greater role across different writing stages and contexts. Foundational theories such as the cognitive process model divide writing into distinct stages including planning, translating, and reviewing [12], which has long served as a framework to analyze writing behaviors in human-AI collaborative settings. Writing with AI also varies by contexts, spanning domains from formal to creative writing [38]. In different stages and contexts of writing, AI may adopt multifaceted roles depending on user needs. It may act as a follower, collaborator, or guide, participating in both divergent and convergent thinking processes [39]. In creative writing, AI often plays a suggestive role, offering ideas or alternatives while leaving creative control to the user [38]. Studies have also shown that during the prewriting stage, users who lack initial ideas tend to prefer AI taking the lead [46]. Some studies categorize AI support into different levels or interaction patterns, from basic grammar correction to high-level critical feedback [27, 48]. Beyond general frameworks, many practical systems have been developed to support AI-assisted story ideation and generation [6, 9, 17, 26, 37], speculative fiction writing [45], application drafting [11], argumentative writing [50], among others. Building on these works, we concentrate on the **story outlining stage in creative writing**, a crucial early phase where creative ideas begin to take shape. This stage demands a closer examination of how human-AI collaboration can maximize inspiration and creative potential.

2.2 Proactive AI in Collaborative Writing

Prior research explored various human–AI collaboration paradigms in co-writing, which reflect different degrees of control distribution and interaction depth:

- User-initiated prompting: The most common mode of AI use during writing, where users actively switch context to the AI chat interface, compose prompts to articulate their needs [13, 46].
- Turn-taking: Human and AI alternate sentence-by-sentence contributions. Though common, this method often feels unnatural to users [17, 31, 47].
- Simultaneous co-creation: AI intervenes in real-time during the user's writing process, triggered by explicit user actions. This includes: Word- or phrase-level suggestions [25], similar to tools like Smart Compose [18]; Sentence- or paragraph-level recommendations or scaffolding strategies [10, 26, 40]; style refinement or critical questions [45].
- User-driven controls: Some systems empower users to guide the AI through direct manipulation–for instance, TaleBrush's destiny curves [7].

Beyond these paradigms, LLMs are progressively shifting from being passive tools to becoming proactive co-creators. Proactive AI observes user actions, anticipates user needs and offers timely assistance, represents a deeper level of interaction. It has been studied in domains such as meetings [5] and programming tools [35], where researchers explored optimal strategies for when and how AI should take initiative. In terms of writing tasks, pioneering work has begun to examine how proactive AI might provide optimal intervention contents using reinforcement learning [28]. Yet, questions remain around how proactive suggestions influence creative flow and user experience during story outlining. In this work, we address this gap by conducting a Wizard-of-Oz study to investigate effects of proactive suggestions on users' creative flow and perception (RQ1).

2.3 Human Agency in Human-AI Co-creativity

Maintaining human agency is a central concern in human-AI co-creativity. For example, in co-creative drawing, users expressed a strong preference for retaining control over the creative direction [32]. In collaborative journaling, users valued autonomy and the ability to set the tone and content of their entries [25]. Similarly, in research ideation, preserving user ownership during problem formulation was seen as essential [29]. In the domain of collaborative writing,AI involvement can significantly impact users' sense of authorship and control. For instance, users reported diminished ownership when AI contributions were too dominant [10], while others emphasized the importance of aligning AI design with user expectations to preserve agency and creative control [38]. Authenticity and personalization have also been identified as key factors in reinforcing user identity and ownership in co-writing [22]. Beyond authorship, AI-generated content has been shown to influence users' own views and opinions [23]. In formal writing, researchers have emphasized the need to support originality and ethical considerations [50]. The issue of human agency becomes especially salient in proactive AI systems, where AI can take the initiative at any time and steer the creative process–sometimes overriding user intent if not carefully designed. Sometimes users want their own ideas to be included even if the AI can offer better alternatives [28]. Striking the right balance between helpfulness and user-control remains a key challenge [47]. In this work, we build on these findings by comparing two styles of AI suggestions–intrusive and non-intrusive–to investigate how different forms of initiative affect users' sense of agency. We further explore under what conditions users can retain control while still benefiting from the creative potential of AI, aiming to balance human autonomy with maximized creative support (RQ2).

2.4 Human-AI Alignment in Collaborative Writing

Achieving alignment between humans and AI remains a core challenge in co-creative interactions. Prior work has highlighted the cognitive difficulty users face when trying to convey intentions to systems [42]. More broadly, alignment has been conceptualized across different levels including commands, intentions, revealed preferences, ideal preferences, interests, and values–all of which can surface during creative writing tasks [14]. For proactive AI systems, this means not only anticipating user actions, but also understanding their writing intentions and goals, preferences for receiving suggestions and working with AI [28]. Some systems have attempted to facilitate alignment through interaction design. For example, VISAR explores visual outline and draft prototyping of canvas to support mutual understanding during writing, enabling users and AI to co-construct content more transparently [50]. However, such efforts largely assume that users explicitly initiate collaboration and intentionally express their goals. In contrast, our work investigates alignment in the context of proactive AI. By adopting a wizard-of-Oz methodology, we explore how human wizards interpret and align with users' intentions during co-writing, and derive insights into what user information a proactive AI should leverage in order to give timely and contextually appropriate suggestions (RQ3).

3 Method

We conducted a within-subject Wizard-of-Oz study to examine how different styles of proactive AI suggestions affect the creative writing process and user experience during collaboration. 30 participants were recruited and paired into 15 groups, each consisting of a user and a wizard. Each group experienced two different suggestion modes (within-subject study), completing a story outline writing task under each mode, followed by a joint semi-structured interview. Interaction data was collected through our collaborative writing platform and subsequently analyzed using evaluation metrics.

3.1 Study Procedure

We conducted a Wizard-of-Oz experiment where two participants together completed the writing tasks under two modes, consistently playing **user** and **wizard** respectively. The entire procedure is visualized in Figure 1.

3.1.1 Main Experiment. In each study, one lead experimenter invited two participants to an online meeting. Before the study, both participants were asked to complete a questionnaire about their

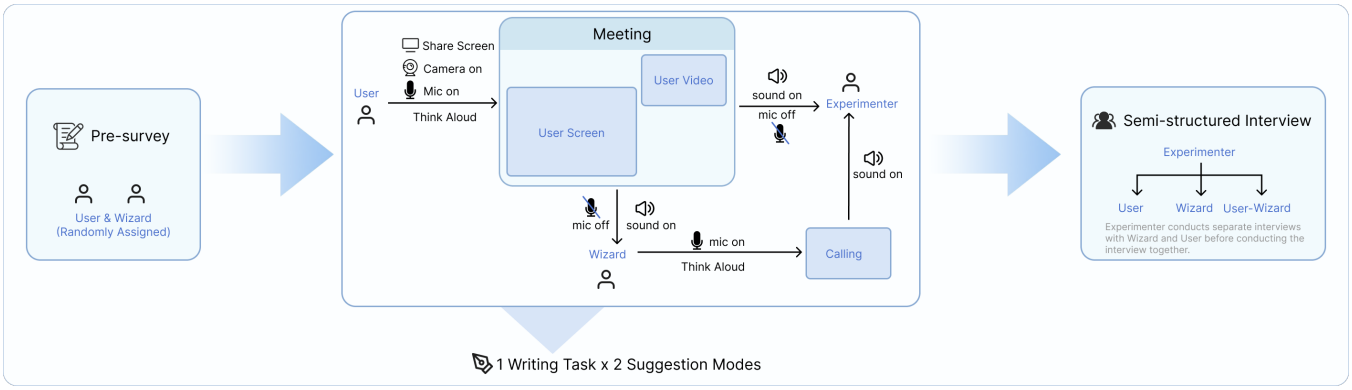


Figure 1: Study procedure.

Table 1: 7-point Likert scale survey for each writing session.

1	I am very satisfied with the overall creative process and the final outcome.
2	I felt that I had full control over the creative process and experienced a high level of autonomy.
3	I felt a strong sense of ownership over the work I created.
4	The AI made a significant contribution to the creative aspects of the work.
5	My creative writing skills improved through this experience.
6	How mentally demanding was the task?
7	How physically demanding was the task?
8	How hurried or rushed did you feel due to time pressure?
9	How satisfied were you with your performance in the task?
10	How hard did you have to work to accomplish your level of performance?
11	How insecure, discouraged, irritated, or stressed did you feel during the task?

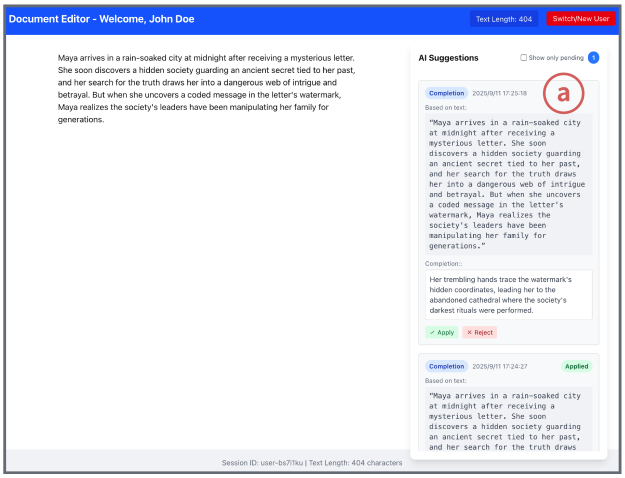
writing background, goals, and experience with AI-assisted writing tools. The participants were randomly assigned to the roles of "user" or "wizard" and remained in the same role throughout the entire experiment. The experimenter then sent each participant an introduction to the study and task. Different versions were provided to the wizard and the user, containing the study's purpose, procedure, and system usage instructions. Participants were allowed to try our system in a test session, during which the experimenter provided further instructions. The user received four story themes to choose from. Each prompt included a thematic direction and a story opening, designed to spark interest and reduce the perception of the task as a burdensome writing assignment. The wizard's introduction included a series of reminders to help ensure the smooth completion of the experiment, such as actively providing suggestions while paying attention to (1) the optimal timing for giving suggestions; (2) filtering out vague initial suggestions provided by the AI and striving to offer more constructive feedback. Additionally, if the suggestion frequency was too low or the quality of suggestions was inadequate during the experiment, the experimenter would intervene and provide appropriate reminders to the wizard. The experimenter also emphasized to both parties that the task focused on developing the story outline, with the creativity of the plot–rather than writing style or language–as the primary objective of the writing process.

Each user and wizard pair was required to complete two writing sessions under two different AI suggestion modes:

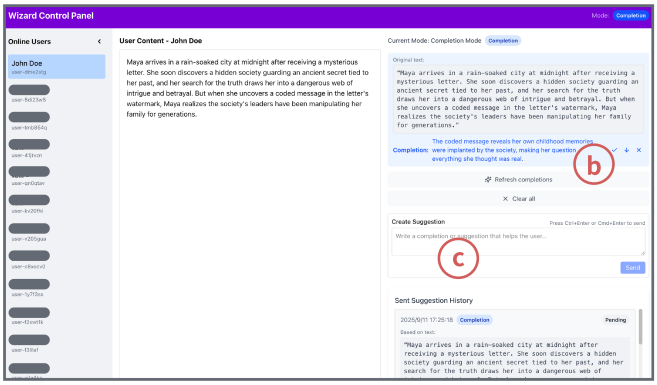
- (Completion Mode) Next-sentence Completion: A continuation suggestion focused on plot development, which the user could directly append to their story.
- (Proposal Mode) Exploratory Proposal: A speculative, question-style prompt intended to exploratory plot suggestions, but not directly applicable to the current narrative.

The two modes correspond to intrusive and non-intrusive AI suggestions, respectively. Except for the suggestion style, users experienced no interface differences between the two modes. The order of these modes was counterbalanced across participants.

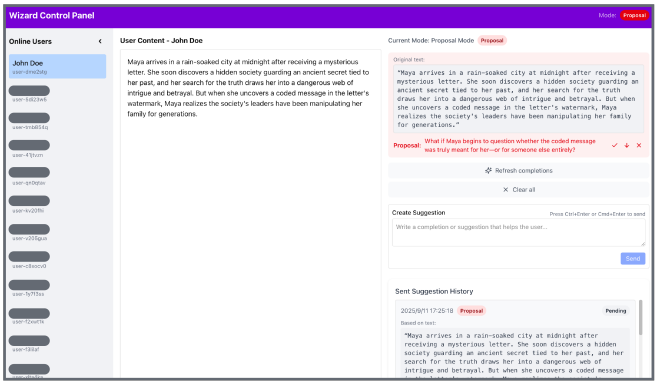
Before writing began, users were asked to turn on their camera, microphone, and screen sharing. During the writing process, they were instructed to think aloud, verbalizing thoughts about the story development and their reactions to AI suggestions. The wizard had full access to this multimodal input but remained muted throughout. This setup ensured that wizards could access and selectively utilize any information form the user that would help them align with the user's intent and context, and make informed decisions about the content and the time of their suggestions. To support high-quality and efficient suggestions, wizards were assisted by an AI that generated initial suggestions. Wizards had the flexibility to send these suggestions as-is, modify them as needed, or compose entirely new ones.



User Interface



Wizard Interface (Completion Mode)



Wizard Interface (Proposal Mode)

Figure 2: User and wizard interfaces. (a) is the sidebar displaying real-time AI suggestions for the users, (b) is the control panel displaying suggestion drafts from AI assistant for the wizard, and (c) is the editor for the wizard to edit AI suggestion draft or compose their own suggestion.

Each of the two writing sessions lasted 45 minutes. Participants were not required to finish the full story but were encouraged to follow a natural writing pace. After each session, users completed a 7-point Likert scale survey assessing perceived satisfaction, AI contribution, autonomy, capability improvement, and task load. The full scale is shown in Table 1.

3.1.2 Semi-structured interview. After two writing sessions, we conducted both individual and joint interviews to understand the user experience and gather the wizard's perspective on effective proactive suggestion strategies. We first interviewed the user and wizard separately, ensuring that the other participant was not present to avoid mutual influence. For users, we began by asking about their general experience with proactive AI suggestions. We then probed into specific dimensions, including the AI's contribution to creativity, its impact on user inspiration, and its influence on the user's sense of agency. Users were also asked to compare the two suggestion styles. Finally, we explored how users perceived the AI's role under each style of proactive assistance. For wizards, we focused on the types of suggestions they made, how they decided when and what to suggest, and what kinds of user information they relied on to make those judgments. Following the individual interviews, we conducted a joint interview with both participants. Each participant was asked to identify the best and worst AI suggestion from the session and explain their reasoning. Participants were encouraged to agree or disagree with each other's choices and to discuss any disagreements in depth. The interview phase lasted approximately 30 minutes.

In total, each study took around two hours (45 minutes × 2 writing sessions + 30 minutes interview). All studies were conducted in Mandarin and were both video- and audio-recorded. Each participant received compensation ranging from 100 to 200 Chinese Yuan, depending on the total duration of their participation.[1]

3.2 System

We developed a collaborative writing web platform that supports real-time AI assistance during the writing process and is compatible with Wizard-of-Oz experimental setups. Our platform provides two distinct interfaces–one for the user and one for the wizard. Each

[1]The study protocol was reviewed and approved by the university ethics review board. The compensation was consistent with the average earning of workers in the community where the study took place.

participant accessed the platform via a dedicated link corresponding to their role. Figure 2 displays both the user and wizard interfaces.

3.2.1 User Interface. The user interface consists of a main text editor, where users can freely compose their stories, and a right-hand sidebar that displays real-time AI-generated suggestions. Each suggestion includes three feedback buttons–Accept and Reject–allowing users to interact with the system:

- In the completion mode, the accept button is labeled "Apply". When clicked, the suggested sentence is automatically appended to the end of the user's current draft.
- In the proposal mode, the accept button is labeled "Like". Clicking it does not change the draft, as proposals in this mode are meant only to inspire the user rather than provide directly applicable completions.

We displayed suggestions in a floating right-hand panel to minimize interruptions to the writing flow. Users can engage with suggestions at their own pace, and can toggle between viewing only unhandled (pending) suggestions or viewing all suggestions. New suggestions always appear at the top of the list, and the suggestion panel remains visible, ensuring that incoming suggestions are not missed. To reduce confusion caused by delays in AI generation, each suggestion is accompanied by the last sentence from the user's draft that was used for the suggestion at that time. This helps users clearly identify where the suggestion refers to.

3.2.2 Wizard Interface. Our wizard interface features a three-pane layout: a left sidebar for switching between user writing sessions, a read-only central view displaying the real-time content of the current writing session, and a control panel on the right. While users write, an LLM-driven AI assistant generates real-time suggestions every 10 seconds whenever the content changes. The most recent suggestions are shown at the top of the control panel. Wizards can refresh suggestions based on the current content or clear all suggestions. For each suggestion, they may either send it directly to the user or move it to a built-in editor for further refinement before sending. Wizards can also compose and send their own suggestions without using AI assistance. The bottom section of the control panel displays the history of suggestions sent to the user. At the top right, a mode selector allows wizards to switch between different suggestion styles.

Besides background color, the modes differ only in the style of the suggestions AI assistant generated. We use the user's full written content as input and apply different prompts to produce stylistically varied suggestions, which are detailed in Appendix A. DeepSeek-V3 is employed to ensure the suggestions are generated quickly and creatively.

Throughout the writing process, our platform logs the current user content every 10 seconds with timestamps to trace writing progress. It also records all suggestions sent by wizards, along with their timestamps and corresponding user feedback (accept, reject, or ignore).

3.3 Participants

We posted our recruitment advertisement on a university campus community platform. Participants were required to have a strong interest in fiction writing, but no specific writing experience was required. We screened 230 applicants through a qualitative assessment of their reported literary preferences, writing motivations and experience. Candidates were considered eligible if they (1) explicitly cited fiction (e.g., reading novels) as the primary driver of their interest, excluding non-fiction formats such as professional literature, **or** (2) reported prior fiction writing experience, regardless of frequency. From this qualified pool, we then employed random selection while ensuring a balanced distribution across different backgrounds and levels of writing experience.

We ultimately recruited 30 participants (21 males, 9 females), with an average age of 24.6 (σ=6.0). This sample size aligns with participant numbers used in prior similar studies on human-AI co-writing and creativity support systems [15, 16, 22, 49]. Participants completed a pre-study questionnaire collecting their demographic information, writing experience, writing purposes, and familiarity with AI-assisted writing tools. An overview of participant information is provided in Table 2. We report participants' basic demographic statistics (age, gender), their writing experience based on frequency (1 for none; 2 for once to three times per month; 3 for once per week; 4 for two to four times per week; 5 for daily or with prior experience in novel writing), primary writing goals (e.g., for work, study, entertainment or personal purposes), and whether they had used AI-assisted writing tools before (coded as 0 for no, 1 for yes). In summary, 5 out of 30 participants reported writing more than 5 times per week or having prior novel writing experience. Regarding writing purposes, 23 participants wrote primarily for entertainment expression, 17 for study or work purposes, and 4 for personal journal. Around half of them (N=17) had experience with AI-assisted writing tools.

3.4 Measurements

Through the study, we collected a total of 30 sessions of writing processes, 486 of AI suggestions. The average number of AI suggestions for each session of writing is 16.2 (σ=14.2), with a slight skew toward the higher end (Median=12, Q1=9.25, Q3=17). We analyze the writing process through several key dependent variables for each AI suggestion and writing session:

3.4.1 Typing Speed. We calculated typing speed as the number of Chinese words written per minute, based on the writing records of every 10 seconds. We removed excessively large outliers (above 500 words per minute) in writing speed to exclude the scenario where AI suggestions are directly appended to the text in completion mode, ensuring that the typing speed only reflects users' actual input. We set the 500 wpm threshold based on empirical observations, which revealed that typing speeds exceeding this value consistently coincided with users accepting lengthy AI suggestions during completion mode in our study. However, we acknowledge that this approach is sensitive to the length of AI suggestions and may not capture shorter AI completions. Future work could consider incorporating textual similarity measures such as Levenshtein distance to more rigorously distinguish between human-typed and AI-generated content.

3.4.2 Creativity Gain and Creativity Gain Speed. To quantify users' creative progress over time during the writing process, we compute Creative Gain as the accumulation of textual novelty relative

Table 2: Information of participants.

Group	Participant	Age	Gender	Writing Experience	Purpose of Writing	AI-assisted Writing Experience
G1	U1	24	Male	1	Work, Entertainment	0
	W1	23	Male	1	Work, Study, Entertainment	0
G2	U2	21	Female	2	Entertainment	0
	W2	26	Male	5	Study, Entertainment	1
G3	U3	23	Female	2	Entertainment	0
	W3	27	Male	3	Work	1
G4	U4	20	Male	1	Entertainment	1
	W4	24	Male	4	Work	0
G5	U5	21	Female	2	Work	1
	W5	34	Male	2	Study, Entertainment	1
G6	U6	24	Male	3	Study, Entertainment, Journal	1
	W6	24	Female	5	Entertainment	1
G7	U7	20	Male	1	Entertainment	1
	W7	24	Male	5	Study	1
G8	U8	25	Male	2	Entertainment	0
	W8	23	Male	1	Study, Entertainment, Journal	0
G9	U9	23	Male	5	Study, Entertainment, Journal	1
	W9	22	Female	2	Work, Journal	0
G10	U10	22	Female	4	Work, Study, Entertainment	1
	W10	21	Male	3	Entertainment	0
G11	U11	22	Male	2	Entertainment	1
	W11	52	Male	2	Entertainment	1
G12	U12	23	Male	5	Entertainment	0
	W12	25	Female	2	Work	0
G13	U13	22	Male	2	Work	1
	W13	24	Male	2	Entertainment	0
G14	U14	23	Female	1	Entertainment	0
	W14	24	Female	3	Study, Entertainment	1
G15	U15	32	Male	4	Work, Entertainment	1
	W15	20	Male	2	Entertainment	1

to prior writing states. Prior work measures creativity primarily based on self-reported scales [34], expert assessments [8, 21, 24] or LLM-based ratings [4, 33], which can be subjective and opaque. Additionally, Lexical Richness [20] is often used to measure creativity, offering more objective measures. We extend such lexical analysis approaches by incorporating the information-theoretic concept of surprisal [41] to **compute cumulative Kullback–Leibler (KL) divergence of words as a proxy for creativity expansion**, which considers not only lexical richness but also changes in the distribution of word usage throughout the dynamic writing process. Specifically, the user's text was logged in 10-second intervals, producing a sequence of snapshots, and for each pair of adjacent snapshots, we computed the KL divergence between the newly added text and the previous snapshot using bag-of-words representations:

$$D_{KL}(P||Q) = \sum_k P(k) \log \frac{P(k)}{Q(k)}$$

where $D_{KL}(P||Q)$ denotes the KL divergence between the two distributions P and Q, representing the newly added text and the previous text respectively, and $P(k)$ is the frequency of word k in the newly added text. This allows us to quantify how much new information (in terms of token distributional change) was introduced at each interval. We accumulated per-interval KL divergence scores over time to form a **creativity gain curve**, as shown in Figure 3, reflecting the evolving creativity of the writing content across the session.

Therefore, the suggestion-level Creativity Gain was computed by measuring the KL divergence change within our defined window after each suggestion (see Figure 3), and we normalized it with the word count ($wc = 150$) in the window to compute the Creativity Gain Speed $CGS_{i,j}$ for each i-th AI suggestion in the j-th writing session:

$$CGS_{i,j} = \frac{\sum_t D_{KL}(P_t||Q_t)}{wc}$$

where t denotes all the intervals within the window after the suggestion. In addition, for each writing session i, we aggregated the total **Creativity Gain** across all suggestions to obtain a session-level indicator of overall creative gain facilitated by AI assistance.

$$CG_i = \sum_j CGS_{i,j}$$

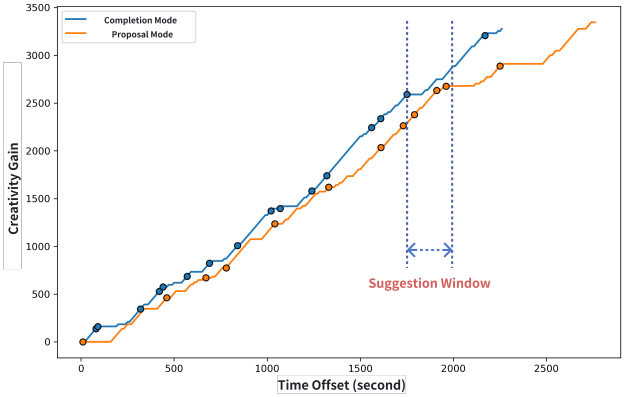


Figure 3: Creativity Gain curves of two example writing sessions in one group (G11). The scatter points represent the time at which each suggestion was given.

3.4.3 AI Contribution and Perceived AI Contribution. We define **AI Contribution** as a measure of the direct influence an AI suggestion has on the user's writing. After each AI suggestion, users may choose to reject, partially adopt, or fully apply the suggestion, followed by composing new content within a defined window. To quantify this influence, we compute the semantic similarity between the AI suggestion $AI\_sugg_{i,j}$ and the text within this window$U_text_{i,j}$, following commonly used similarity metrics [43] to computing AI contribution to a text. The total AI contribution for a session i is the sum of contributions of all suggestions within it.

$$AIC_{i,j} = Similarity(AI\_sugg_{i,j}, U\_text_{i,j}), \quad AIC_i = \sum_j AIC_{i,j}$$

In parallel, we capture Perceived AI Contribution $PAIC_i$ using users' self-reported ratings after each writing session. These subjective ratings reflect users' perception of the engagement of the AI assistance.

3.4.4 Perceived Autonomy. For each session of writing, we assess Perceived Autonomy using users' self-reported ratings. These subjective ratings reflect users' sense of control, authorship, and independence throughout the writing process.

3.4.5 Perceived Satisfaction. We also measure Perceived Satisfaction using users' self-reported ratings for each writing session. These subjective ratings capture users' overall satisfaction with the writing experience and final output, based on factors such as content quality, perceived creativity, and the fluency of collaboration with a proactive AI assistant. Note that all subjective ratings were normalized prior to analysis.

3.5 Qualitative Analysis

We collected 7.5 hours of interview footage in total. We cut them into user and wizards interview, and based on their reflections on proactive AI co-writing tools, we adopted a qualitative analysis approach grounded in thematic coding [3]. We systematically extracted and categorized interview quotes from participants and wizards (U01-U15, W01-W15), identifying recurring patterns and divergent viewpoints across two AI suggestion modes: Completion Mode and Proposal Mode.

4 Findings

4.1 RQ1: How proactive AI impacts user's writing process?

Users experience Stuck–Smooth writing cycles, and proactive AI suggestions accelerate recovery from stuck moments. The visualization of creativity gain curves reveal a recurring Stuck–Smooth Loop pattern in users' writing process, as shown in Figure 3, users tend to fluctuate between periods of low activity (stuck phases) and high productivity (smooth phases). Proactive AI suggestions appear to accelerate the transition out of the stuck phase. Specifically, the average writing speed within a window following an AI suggestion was 1.2x higher than the overall average writing speed during the session (μ=1.2, σ=1.09). Note that the writing speed already excludes AI-generated completions (see Section 3.4.2). This difference was statistically significant (p=.006 < .05), indicating that AI interventions were effective in stimulating writing momentum and helping users recover from stuck states.

In our interviews, 8 out of 15 users mentioned that AI was especially helpful when they were blocked, including reducing the time spent stuck and pushing their thinking forward. For example, U05 noted that "AI's proactive suggestions helped deepen our thought process and move the writing forward, encouraging more active use of AI. Without these suggestions, users might have taken longer to find solutions on their own". U04 said, "AI provides more effective inspiration when I'm stuck", while others echoed, "AI suggestions are helpful, especially in moments of writer's block" (U01) and "When I'm stuck, AI gives me some inspiration" (U11). Additionally, one participant mentioned that they actively seek out AI suggestions when blocked: "I usually look to see if there's something good I can borrow" (U01). These responses highlight that AI suggestions not only help with writing speed but also play a key role in unblocking the creative process.

Participants also expressed concerns and expectations regarding AI's role in overcoming their creative block. U07 emphasized that AI should avoid interrupting the user's thought process, as "they may be in a state of flow". U15 believed "AI should offer divergent thinking, providing interesting, innovative ideas when users are stuck". We also observed that U08 would verbally request AI suggestions when stuck, suggesting that a bidirectional active interaction would be a better design.

4.2 RQ2: How do different styles of proactive suggestions (intrusive vs. non-intrusive) influence user experience?

We compared the two modes on subjective rating measures (all subjective ratings were zero-normalized) and conducted Wilcoxon signed-rank tests to evaluate the effect of suggestion style on various metrics across writing sessions, as shown in Figure 4. We found that proposal mode reported significantly higher levels of perceived autonomy and ownership compared to completion mode, but significantly lower perceived AI contribution. At the same time, users in proposal mode also reported higher levels of mental demand and

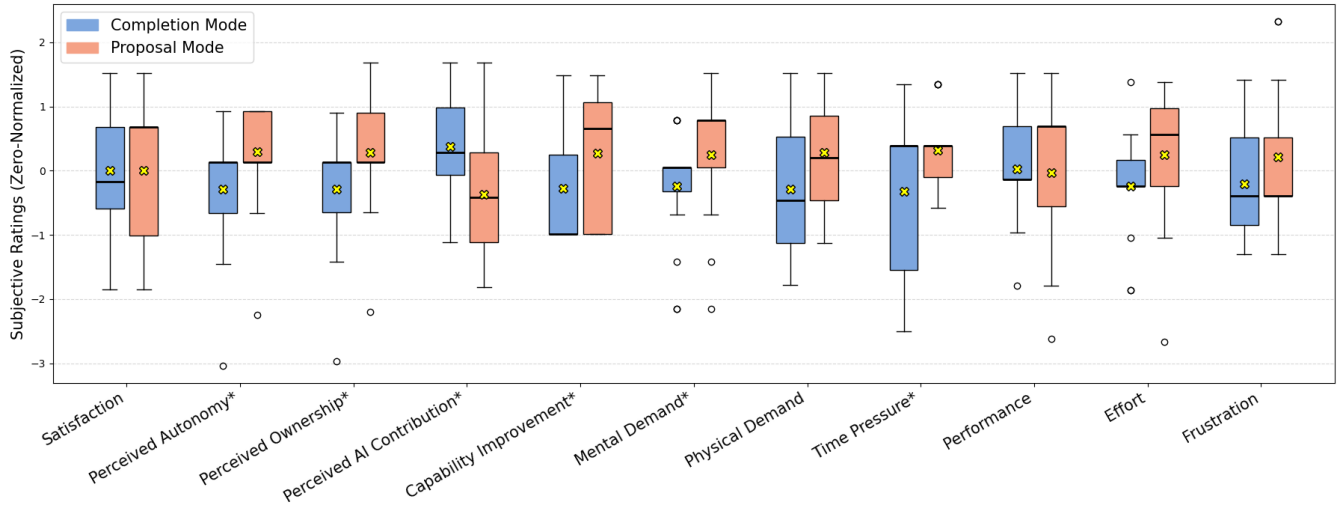


Figure 4: Comparison of subjective ratings (including custom measures and NASA-TLX) between completion mode and proposal mode sessions (paired), with asterisks showing the statistically significant mean differences on the Wilcoxon signed-rank test (* for <.05). Crosses (x) represent mean values, and horizontal lines (—) represent median values. Note that all subjective ratings were zero-normalized.

time pressure. Considering that proposal mode only received exploratory proposals that were not directly usable, it is evident that users were required to engage in more active thinking and decision-making. This likely contributed to increased cognitive effort and reduced the perceived influence of the AI, while at the same time enhancing their sense of authorship and autonomy. Beyond these subjective findings, we conducted further analysis and identified the following patterns:

4.2.1 ***Trade-off between AI Contribution and Perceived Autonomy.*** Across all writing sessions, we observed a clear trade-off between AI contribution and perceived user autonomy, as illustrated in Figure 5 with a clear negative correlation between the two measures. The blue dots, representing completion-style interactions, are clustered in the bottom-right quadrant–indicating high AI contribution but low autonomy. This suggests that, in these cases, users felt that **the AI had taken over the writing process**. Conversely, the red dots, which represent proposal-style sessions, are mostly located in the upper-left quadrant, indicating low AI contribution but high user autonomy. This reflects scenarios where **the AI provided little to no meaningful support**, and users retained full control over their plot construction.

Interview responses further support this interpretation. Many users reported a lower sense of control when using the completion mode. One participant remarked, "Autonomy was lower in System A [Completion Mode], but higher in System B [proposal mode]" (U06), while another observed, "In Mode A [completion mode], I felt less autonomous, although I did get more inspiration" (U12). A user reflected on the contrasting sense of control between the two modes, stating, "In the first [Proposal Mode], I felt more control over my work compared to the second [completion mode]" (U02). Interestingly, not all users perceived the control dynamic in the same way. One participant noted that "Control was higher in Mode A [completion mode], and lower in Mode B [proposal mode]" (U15), suggesting that perceptions of autonomy may also depend on individual writing preferences and expectations around collaboration with AI. Overall, these findings reveal a fundamental tension: while more assertive AI contributions can enhance writing fluency and reduce user effort, they may also risk overriding the user's creative direction, leading to a diminished sense of authorship.

4.2.2 ***AI-Augmented Inspiration Alleviates the Trade-off between AI Contribution and Perceived Autonomy.*** Interestingly, we observed that in several cases, users experienced both high autonomy and high AI contribution, as indicated by data points in the top-right quadrant of Figure 5. This suggests that, in these cases, AI may play a role in augmenting users' creativity without diminishing their sense of control.

We posit that users retain autonomy because AI contributes primarily through indirect inspiration rather than direct intervention. Following this assumption, we introduce the concept of **AI-augmented inspiration (AAI)**, a session-level metric that quantifies creativity that is indirectly stimulated by AI suggestions. Mathematically, for each writing session i, we define:

$$AAI_i = \frac{CG_i}{AIC_i}$$

where CG_i represents the overall Creativity Gain during a writing session, and AIC_i denotes the AI Contribution, representing the direct influence of AI suggestions on the users' writing (see Section 3.4.2). AAI reflects the ratio between Creativity Gain and AI Contribution, providing a normalized measure of creativity yield per unit of AI input. By quantifying the efficiency of AI contribution, AAI captures creativity that originates from the user through indirect AI stimulation, rather than direct AI generation. A high AAI (where CG substantially exceeds AIC) indicates an inspirational

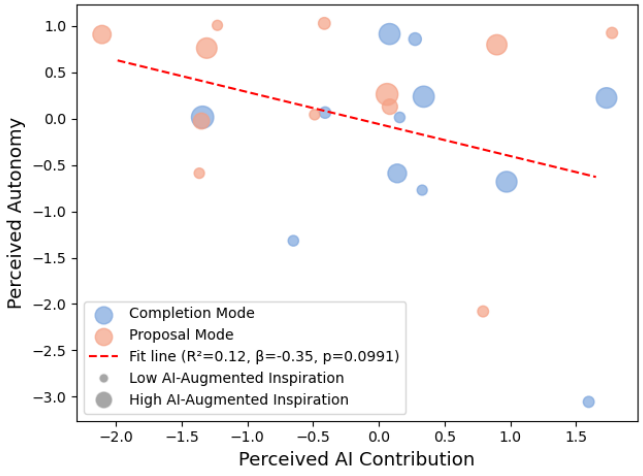


Figure 5: Higher Perceived AI Contribution corresponds to lower Perceived Autonomy. However, higher levels of AI-Augmented Inspiration (larger points) are associated with a weaker trade-off. Note that all subjective ratings were zero-normalized.

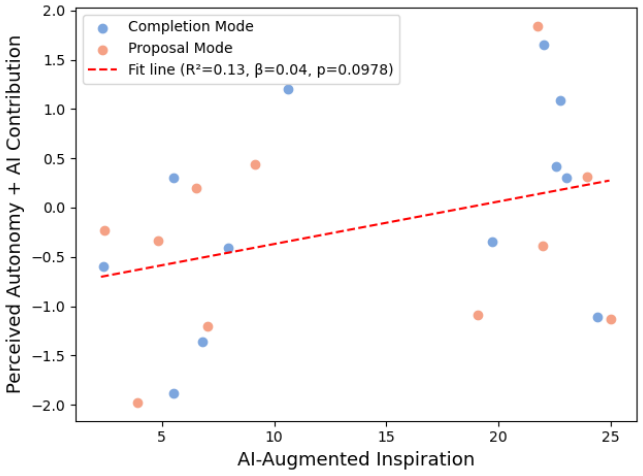


Figure 6: Higher AI-augmented Inspiration corresponds to a higher sum of Perceived AI Contribution and Perceived Autonomy, indicating that both can be elevated simultaneously.

impact of AI contribution, where AI suggestions trigger significant user elaboration. Conversely, a low AAI suggests an interruptive impact, where the user must process a large volume of AI-generated content to derive minimal creative value. Previous research exploring AI-inspired creativity has primarily relied on qualitative result [49, 51] or self-report scales such as inspiration scale [44] and user engagement [30], without offering robust quantitative measures to capture the indirect influence of AI as inspiration. To our knowledge, this is the first attempt to quantify indirect AI inspiration and its impact on the dynamic relationship between AI contribution and user autonomy.

Our analysis shows that higher AAI values are associated with both greater perceived autonomy and stronger AI contribution, as indicated by the clustering of larger points which represent sessions with higher AAI values in the top-right of the plot in Figure 5. Figure 6 also shows a positive correlation between AAI and the sum of Perceived Autonomy and AI Contribution scores. These results suggest that high AI-augmented inspiration enables users to retain autonomy while benefiting from AI contributions, highlighting AI's potential as a tool that indirectly inspires creativity without directly taking over the creative process. For system designers, this metric implies that optimizing autonomy requires prioritizing AI triggers that leverage user inspiration, rather than substituting user effort with excessive content generation.

We further conducted interviews with participants to explore how they perceive indirect AI-augmented inspiration and why they maintain autonomy. **First, interview findings revealed that AI-augmented inspiration stemmed from the mutual, iterative stimulation brought by proactive AI suggestions, which acted as a catalyst for users' ideas.** As one participant (U11) reflected, proactive AI acted like a catalyst–"the exploratory proposal in Mode B [proposal mode] helped me progress by prompting me to write further. For the details it added, I could then elaborate on them myself. This felt like a kind of inspiration or spark." U08 noted, "If the AI gives me an idea that sparks inspiration in the same direction, I will engage with it–it triggers more inspiration from me". These reflections highlight that proactivity's value lies in catalyzing narrative expansion. Rather than directing the story, proactive suggestions contributed to plot enrichment by autonomously offering unanticipated details, thereby prompting further user elaboration. In this context, proactive AI functions as a catalyst, driving a high AAI score–while the AI provides only details, it indirectly inspires the user's creative flow, yielding a substantial return in creative output per unit of AI contribution. This catalytic effect stimulated creativity in a way that non-proactive AI suggestions did not achieve.

Second, participants also felt they remained in control because they actively judged and selectively adopted AI suggestions. We also observed that participants felt that they remained in control although AI provided substantial contribution in some cases–primarily because they actively judged and selectively adopted AI suggestions. One user emphasized that authorship lies in creativity, stating, "The most important part of ownership is the idea. Even if some of the text was completed by the AI, the core creative direction didn't change" (U11).

Third, users described the AI not as a dominant force but as a supportive tool that enhanced, rather than replaced, their creative agency. One participant explained, "Most of the time I was thinking independently. I looked at what the AI offered and decided whether I could borrow anything from it" (U01). This sense of ownership extended across both modes. As one participant put it, "In both modes, I still felt a strong sense of ownership over the work"(U11). Another reflected on how AI acted more as a thinking partner than a creator: "In the proposal mode, it felt like the AI

asked me questions from a third-person perspective... but I was still writing my own piece"(U02). These examples highlight that perceived autonomy is not solely determined by how much content the AI generates, but also by the user's role in curating, interpreting, and integrating that content. When users feel that they are making active choices, they retain a strong sense of authorship–even when AI plays a major role in shaping the text.

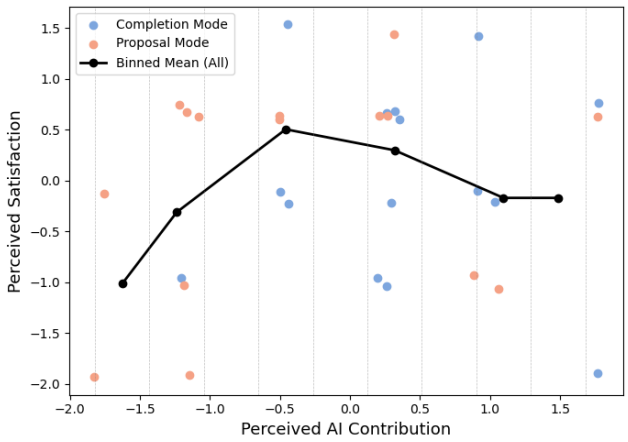


Figure 7: Relationship between Perceived AI Contribution and User Satisfaction.

4.2.3 ***Intermediate AI Contributions Drive Maximum User Satisfaction***. The degree of AI Contribution influences overall writing satisfaction, with neither extremely high nor extremely low levels producing the best outcomes. Instead, there appears to be a "sweet spot". Figure 7 illustrates the relationship between AI Contribution and Perceived Satisfaction, showing the binned means. We found that when AI Contribution was at an intermediate level (around −0.5), users' perceived satisfaction reached its peak. This suggests that the most satisfied users and writing processes were not those with the highest or lowest levels of AI usage, but rather those that used AI most effectively. Their typical behavior involved accepting AI suggestions as a starting point or spark of inspiration, followed by substantial personal work to extend and integrate the content.

4.2.4 ***Proposal Mode Better Improves Writing Capability***. Based on participants' self-assessments, proposal mode was perceived to improve writing skills more than completion mode. On average, users rated B significantly higher than A in terms of writing capability improvement ($p=.04<.05$). 8 out of 15 writing pairs explicitly stated that proposal mode contributed more to enhancing their writing capability. This perception may stem from the fact that proposal mode encouraged active thinking and self-driven ideation, while completion mode's ready-made content reduced opportunities for users to engage deeply in the writing process. As one participant (U04) noted, "Group B felt more like a friend–it gave indirect hints that nudged me forward, which helped me practice and improve my own writing skills.". Their reflection illustrates how the collaborative and exploratory nature of the proposal mode may better support long-term writing development, even if it does not always produce immediate fluency or perceived satisfaction gains.

4.2.5 ***Two Types of AI Suggestions Improve Writing Quality Through Both Plot Advancement and Detail Refinement.*** Participants identified two major types of AI suggestions: those focused on **descriptive elaboration** and those oriented toward to **plot development**. The former were more commonly observed in completion mode, while the latter appeared more frequently in proposal mode. Each served a distinct role in the writing process and shaped user engagement in different ways.

Plot-oriented suggestions were seen as particularly productive. Rather than simply filling in description gaps, they provided users with new narrative directions, often stimulating original ideas and encouraging forward momentum. As one user noted, "Group B provided plot direction, which made it easier to trigger my own creative thinking" (U04). Another echoed this sentiment, describing how "discussion-based suggestions pushed me to continue writing" (U11).

By contrast, description-focused suggestions–often associated with the completion mode–helped users refine and expand existing ideas, contributing to the texture and richness of the narrative. While less likely to inspire major story shifts, they were valuable in making the text feel more complete. One participant drew a clear line between the two: "Mode A [completion mode] helped with fleshing out details, while Mode B [proposal mode] provided direction for the plot" (U12). This distinction reflects a dual role for AI in collaborative writing: as both a co-author, shaping high-level narrative arcs, and a writing assistant, helping polish and enrich the surface of the story.

Overall, users mentioned the impact of the two different types on writing quality. Three users noted that Plot-oriented suggestions from both the proposal and completion modes help enhance diversity of the plot. U13 specifically mentioned that the proposal mode is more expansive, "making me think of exciting plot twists they hadn't considered before". At the same time, one user (U02) mentioned that the completion mode provides more detailed descriptions, "allowing the overall writing to become more elegant."

4.2.6 ***Comparison Between Two Modes***.

Completion Mode: Fluent Expansion and Writing Efficiency. Completion mode was frequently described as helpful for directly enhancing writing fluency by generating sentences or descriptive details. This mode supports users in expanding their drafts more quickly and with less cognitive effort. One participant noted that it helped "enrich content based on my own ideas" (U01), while another said it "directly provides paragraphs or sentences", although sometimes "it may diverge from the user's intent" (U04).

Many users found this mode particularly effective for creative elaboration. For instance, one shared that "AI suggestions in this mode contribute more to the creativity of the final piece" (U06). Others emphasized its impact on writing fluency and mental effort: "It makes the writing process smoother" (U05), and "I prefer using completion mode during actual writing because it saves brainpower" (U15). One participant summarized the difference by saying, "The

second one [proposal mode] feels like a friend suggesting ideas; the first one [completion mode] is a very useful tool" (U08).

Proposal Mode: Structural Guidance and High-Level Feedback. Proposal mode, by contrast, was seen as better suited for offering directional suggestions and logical review–especially in the later stages of writing. Participants described it as providing a more global perspective. One said it was "more focused on the overall direction of the article" (U01), while another hoped the system could "review the logic from a reader's perspective after I finish writing" (U02). Several participants felt this mode should not intervene too early: "I think it [proposal mode] shouldn't appear during writing; it should come after" (U02).

Proposal mode also inspired user-led storytelling by proposing high-level ideas. One participant found it useful for "plot direction and triggering autonomous creativity" (U04), and another said it worked well "when the user doesn't yet have a full outline and wants to explore sentence structures together" (U15).

Use Case Preferences: Detail vs. Direction. Across interviews, participants consistently distinguished the two modes by their functional roles: completion mode for detail enrichment, and proposal mode for story direction. One stated, "Mode A [completion mode] focuses more on adding details, while Mode B [proposal mode] is better for high-level suggestions" (U15). Another described the utility as context-dependent: "A [completion mode] is more practical for expanding descriptions; B [proposal mode] is better for short-form or plot suggestions" (U15). 3 out of 15 participants mentioned that the two modes could be used in combination. For example, one said, "I think combining the two modes would be better" (U01). Another elaborated on this complementary use: "I will use Mode B [proposal mode] when I don't yet have a clear idea of the plot direction. But when I already have a solid and compelling plot in mind, I would like to switch to Mode A [completion mode] to enhance the expressiveness of my writing." (U11)

4.3 RQ3: How does user information help AI better provide proactive intervention?

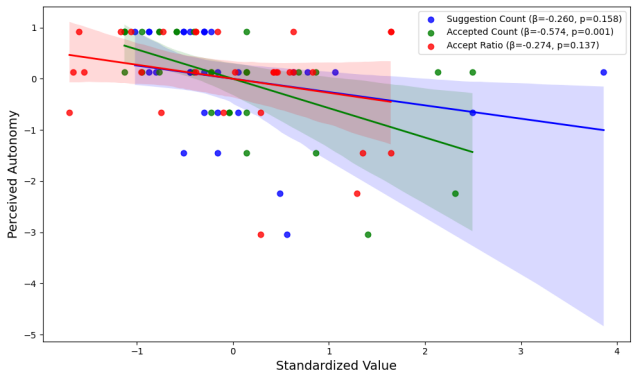


Figure 8: The number of accepted suggestions significantly impacts users' autonomy. Note that all measures have been standardized.

AI interventions can be disruptive both in terms of content and timing. We calculated the number of suggestions, the number of suggestions accepted by the user, and the acceptance rate for each session, and computed the standardized data and their influence to Perceived Autonomy. Figure 8 shows that the number of accepted suggestions significantly impacts users' autonomy. This reflects the negative impact of excessive AI intervention on the on user agency during the writing process.

Participants frequently mentioned that overly frequent suggestions interrupted their writing flow and led to frustration. For instance, one user said, "I think the frequency is a bit too fast" (U02), while a wizard noted that the completion mode was "too deliberate in steering" and "often interrupted and influenced the user's thoughts" (W04). Another user added that the frequent suggestions might "derail the author from their intended topic" (U06). Furthermore, when AI suggestions diverged too much from the user's storyline, they became an additional cognitive burden. One user remarked, "the proposals in Mode B [proposal mode] require users to understand the AI's storyline and think about how to incorporate it into their own writing, which interrupts their original train of thought" (U07). These comments reflect a consistent concern: AI suggestions, when too frequent or misaligned, can disrupt the user's concentration and hinder plot development, ultimately reducing the effectiveness of the intervention.

Wizard behavior and its impact on user experience. Wizard behavior plays a crucial role in shaping the overall user experience. We analyzed wizard strategies for engaging with AI-generated content and observed that wizards rarely edited AI's suggestions. Instead, they primarily filtered through the suggestions provided by AI, possibly due to concerns about the delay of suggestions caused by extra edits. One wizard (W15) explained, "It's difficult for the wizard to keep up with the user's writing speed, so during pauses, I choose the most suitable AI suggestions." This suggests that most of the suggestions users received during the experiment were high-quality AI-generated outputs which were selectively filtered by the wizard. 2 wizards (W03, W15) also complained that part of the suggestions provided by the AI were relatively useless and lacked substantial information, which also indicates that the current AI's ability to generate high-quality suggestions remains limited.

From wizard interviews, additional strategies emerged for how they use multimodal user signals to guide proactive interventions:

(1) **Facial expressions, alongside real-time writing content, are key indicators of writer's block.** Wizards reported monitoring expressions like frowns or long pauses to detect hesitation or difficulty. One said they judged based on "frowning" (W01), while another looked for signs like "furrowed brows or pauses" (W02).
(2) **User receptiveness influences future suggestions.** Wizards adapted their approach depending on whether suggestions were accepted or rejected. For example, one noted they would "adjust the content of suggestions based on whether users adopted them" (W06), and another "reduced frequency when the user rejected a lot of suggestions, as it indicated a mismatch with the user's intent. I would then use this signal to infer what plot the user might be looking for and

adjust the suggestions accordingly" (W11). Users also recognized this dynamic, as one commented, "The wizard adjusted based on my feedback" (U07). However, not all wizards made adjustments–"I use acceptance rate to judge satisfaction but don't change the suggestions accordingly" (W03).

(3) **Aligning with the user's overall plot direction is crucial.** Wizards emphasized the importance of understanding the user's intended story before making suggestions. Some waited until users had written part of the story to calibrate, such as "I wait until the user is halfway through to align the plot" (W11). A user also expressed a preference for this approach: "I hope the AI can provide a feature that lets users write a rough outline first, so it doesn't have to guess blindly." (U11)

5 Discussion

5.1 Inspiration-Centric AI Design Supports Both Creativity and Human Agency

Our findings suggest that in collaborative writing, proactive AI should primarily serve as a catalyst for inspiration, rather than as a direct content provider that may override human agency. To assess AI's indirect influence on user creativity, we introduced the metric of AI-augmented Inspiration, which captures the degree to which users independently generate content stimulated by AI input. This measure significantly moderates the trade-off between perceived AI contribution and perceived autonomy. In addition, our results show that higher AI contribution does not always lead to better outcomes–rather, moderate AI contribution tends to yield the highest collaboration satisfaction. Participants who reported the greatest inspirational benefit from AI did not passively accept or reject suggestions. Instead, they engaged in active reinterpretation–reflecting on, internalizing, and integrating AI suggestions into their own ideas. Together, these findings suggests a new direction for evaluating and designing Human-AI collaborative writing systems. To explore how AI can better fulfill the role of an "inspiration catalyst", we designed and compared two suggestion styles: completion and proposal. While our results offer early insights, further design innovation is needed to better support creativity inspiration and preserve human agency in co-creative writing.

5.2 Wizard-of-Oz Design Reveals AI Writing Tool Capabilities Beyond Language Processing

Our Wizard-of-Oz design enables a broader examination of the AI capabilities required for future co-creative writing systems. Previous research with pre-developed writing tool primarily relied on findings from past works or formative studies to identify necessary features for AI tools, focusing on language processing abilities, for example, the form and content of generated suggestions [6, 10]–reflecting the current limits of available systems. In contrast, our WoZ design allows us to explore essential capabilities which go beyond simple language processing, better aligning AI's capabilities with the needs of future co-creativity systems. As demonstrated in our findings for RQ3, these include the ability to understand behavioral cues through multimodal input, reflect from user feedback, and align with user intentions through reasoning and proactive interactions.

5.3 Human-AI Delegation in Co-creativity Tasks

When AI becomes proactive, the question of human–AI task delegation becomes critical: what role should AI play in collaborative writing, and when should it proactively take on specific responsibilities? The answers vary across contexts and applications. In co-creative writing, it is essential to carefully design how and how much the AI contributes creatively, in order to maximize user inspiration while preserving user agency and avoiding unnecessary disruption. By designing better interaction forms, which essentially defines the division of labor for AI in creative tasks, AI will not become a force that robs individuals of their true creativity, but rather a supportive tool that enhances, rather than replaces, their creative agency.

Moreover, the relative capabilities of the AI and the user across different dimensions should be considered. For example, a user skilled in story ideation but less fluent in language may benefit most from AI assistance with phrasing, not plot development. Conversely, for a novice writer unfamiliar with narrative structure, AI support in plot logic may be more valuable. Finally, users' preferences for AI intervention must also be taken into account. Building on these insights, our work offers an initial exploration into how proactive AI can be designed to better support co-creative writing.

5.4 Latency Issues in Human–AI Collaborative Writing

In our study, we observed a commonly reported system issue that negatively impacted the writing experience: AI suggestions often lagged behind the user's writing progress. In other words, by the time the AI generated a suggestion, the user had already written the next sentence–resulting in suggestions that targeted outdated context. This latency led to a poor user experience. For example, one participant (U05) noted, "About 90% of the suggestions from the proposal mode tried to steer the plot in a more twisted direction, but due to the delay, it was too late for me to revise what I had already written."

We identified several contributing factors to this delay: (a) the time required for the LLM to generate a suggestion, (b) time spent on wizard review and possible editing, (c) the fact that suggestions were shown in a sidebar, allowing users to defer when to view and process them. Among these, factor (b) was specific to our Wizard-of-Oz setup, but the other two would likely persist in real-world deployments and may bring potential issues. Possible solutions include: (1) optimizing LLM response time, (2) intelligently determining whether plot-changing suggestions are appropriate at the current moment, and (3) detecting which suggestions require timely attention and prioritizing user notifications accordingly.

5.5 Limitations and Future Work

In this study, we explored only two preliminary design approaches to investigate how to balance human agency and AI's contribution to creativity. We did not aim to identify or validate the optimal design through user evaluation. Moreover, due to concerns about interruption, we did not implement a fully functioning proactive

AI system. Instead, we used a Wizard-of-Oz setup, where users received simulated AI suggestions from wizards. The question of when and what suggestions should be proactively delivered remains open and requires further optimization and implementation. In addition, we did not examine the optimal interaction design for presenting proactive AI suggestions. Instead, suggestions were displayed uniformly in a sidebar, without tailoring the visualization or interaction mechanisms. In the future, we plan to build on the insights and design guidelines derived from this study to develop and deploy a fully functional proactive AI system for collaborative writing, and to evaluate its usability in real-world scenarios. We also hope that our findings can inform and inspire future researchers and developers working on proactive AI for co-creative tasks.

6 Conclusion

This study systematically compared two styles of proactive AI writing suggestions–intrusive and non-intrusive–using a Wizard-of-Oz setup to examine their effects on creative writing experiences. Our findings show that while proactive AI can accelerate writing, there is a clear trade-off between AI contribution and user agency. We introduce the measure of AI-augmented inspiration to explain how users can retain autonomy while benefiting from AI contribution. Through interviews with wizards, we offer actionable design principles to guide the development of future proactive AI writing tools that foster creativity without undermining user autonomy.

Acknowledgments

This work is supported by the Natural Science Foundation of China under Grant No. 62132010, and the Key Research and Development Program of Ningbo City under Grant No. 2023Z062.

References

[1] Piotr D Adamczyk and Brian P Bailey. 2004. If not now, when? The effects of interruption at different moments within task execution. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 271–278.

[2] Brian P Bailey, Joseph A Konstan, and John V Carlis. 2001. The Effects of Interruptions on Task Performance, Annoyance, and Anxiety in the User Interface.. In *Interact*, Vol. 1. 593–601.

[3] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[4] 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).

[5] Xinyue Chen, Lev Tankelevitch, Rishi Vanukuru, Ava Elizabeth Scott, Payod Panda, and Sean Rintel. 2025. Are We On Track? AI-Assisted Active and Passive Goal Reflection During Meetings. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–22.

[6] Jean-Peïc Chou, Alexa Fay Siu, Nedim Lipka, Ryan Rossi, Franck Dernoncourt, and Maneesh Agrawala. 2023. Talestream: Supporting story ideation with trope knowledge. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–12.

[7] 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.

[8] 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 *Proceedings of the 23rd International Conference on Intelligent User Interfaces*. 329–340.

[9] 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.

[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. doi:10.1145/3613904.3642134

[11] Aidan Z Fitzsimons. 2024. Overcoming challenges to personal narrative co-writing with AI: A participatory design approach for under-resourced high school students and those that support them. In *Proceedings of the 16th Conference on Creativity & Cognition*. 17–20.

[12] Linda Flower and John R Hayes. 1981. A cognitive process theory of writing. *College Composition & Communication* 32, 4 (1981), 365–387.

[13] Yue Fu, Sami Foell, Xuhai Xu, and Alexis Hiniker. 2024. From text to self: users' perception of AIMC tools on interpersonal communication and self. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–17.

[14] Iason Gabriel. 2020. Artificial intelligence, values, and alignment. *Minds and machines* 30, 3 (2020), 411–437.

[15] Katy Ilonka Gero, Vivian Liu, and Lydia Chilton. 2022. Sparks: Inspiration for science writing using language models. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference*. 1002–1019.

[16] Katy Ilonka Gero, Tao Long, and Lydia B Chilton. 2023. Social dynamics of AI support in creative writing. In *Proceedings of the 2023 CHI conference on human factors in computing systems*. 1–15.

[17] Seraphina Goldfarb-Tarrant, Haining Feng, and Nanyun Peng. 2019. Plan, write, and revise: an interactive system for open-domain story generation. *arXiv preprint arXiv:1904.02357* (2019).

[18] Google. n.d.. Use Smart Compose in Gmail. https://support.google.com/mail/answer/9116836. Gmail Help Center.

[19] Alicia Guo, Shreya Sathyanarayanan, Leijie Wang, Jeffrey Heer, and Amy X Zhang. 2025. From pen to prompt: how creative writers integrate AI into their writing practice. In *Proceedings of the 2025 Conference on Creativity and Cognition*. 527–545.

[20] Hye Seung Ha. 2019. Lexical Richness in EFL Undergraduate Students' Academic Writing. *English Teaching* 74, 3 (2019), 3–28.

[21] Kartik Hosanagar and Daehwan Ahn. 2024. Designing human and generative AI collaboration. *arXiv preprint arXiv:2412.14199* (2024).

[22] Angel Hsing-Chi Hwang, Q Vera Liao, Su Lin Blodgett, Alexandra Olteanu, and Adam Trischler. 2025. 'It was 80% me, 20% AI': Seeking Authenticity in Co-Writing with Large Language Models. *Proceedings of the ACM on Human-Computer Interaction* 9, 2 (2025), 1–41.

[23] 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.

[24] Anna Kantosalo and Sirpa Riihiaho. 2019. Quantifying co-creative writing experiences. *Digital Creativity* 30, 1 (2019), 23–38.

[25] Taewan Kim, Donghoon Shin, Young-Ho Kim, and Hwajung Hong. 2024. DiaryMate: Understanding user perceptions and experience in human-AI collaboration for personal journaling. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–15.

[26] Max Kreminski, Melanie Dickinson, Michael Mateas, and Noah Wardrip-Fruin. 2020. Why Are We Like This?: The AI architecture of a co-creative storytelling game. In *Proceedings of the 15th International Conference on the Foundations of Digital Games*. 1–4.

[27] Zhicheng Lin. 2025. Techniques for supercharging academic writing with generative AI. *Nature Biomedical Engineering* 9, 4 (2025), 426–431.

[28] Zhiyu Lin, Upol Ehsan, Rohan Agarwal, Samihan Dani, Vidushi Vashishth, and Mark Riedl. 2024. Beyond Following: Mixing Active Initiative into Computational Creativity. *arXiv preprint arXiv:2409.16291* (2024).

[29] Yiren Liu, Si Chen, Haocong Cheng, Mengxia Yu, Xiao Ran, Andrew Mo, Yiliu Tang, and Yun Huang. 2024. How ai processing delays foster creativity: Exploring research question co-creation with an llm-based agent. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–25.

[30] Jeanne Nakamura and Mihaly Csikszentmihalyi. 2014. The concept of flow. In *Flow and the foundations of positive psychology: The collected works of Mihaly Csikszentmihalyi*. Springer, 239–263.

[31] Eric Nichols, Leo Gao, and Randy Gomez. 2020. Collaborative storytelling with large-scale neural language models. In *Proceedings of the 13th acm siggraph conference on motion, interaction and games*. 1–10.

[32] 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*. 1–13.

[33] Peter Organisciak, Selcuk Acar, Denis Dumas, and Kelly Berthiaume. 2023. Beyond semantic distance: Automated scoring of divergent thinking greatly improves with large language models. *Thinking Skills and Creativity* 49 (2023), 101356.

[34] Hiroyuki Osone, Jun-Li Lu, and Yoichi Ochiai. 2021. BunCho: ai supported story co-creation via unsupervised multitask learning to increase writers' creativity in japanese. In *Extended abstracts of the 2021 CHI conference on human factors in*

computing systems. 1–10.

[35] Kevin Pu, Daniel Lazaro, Ian Arawjo, Haijun Xia, Ziang Xiao, Tovi Grossman, and Yan Chen. 2025. Assistance or disruption? exploring and evaluating the design and trade-offs of proactive ai programming support. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–21.

[36] Ahmad Nur Ihsan Purwanto. 2025. Human-Centered AI: Designing Intelligent Systems that Empower, Not Replace. *RIGGS: Journal of Artificial Intelligence and Digital Business* 4, 3 (2025), 3212–3221.

[37] Ahmed Y Radwan, Khaled M Alasmari, Omar A Abdulbagi, and Emad A Alghamdi. 2024. SARD: A Human-AI Collaborative Story Generation. In *International Conference on Human-Computer Interaction*. Springer, 94–105.

[38] Mohi Reza, Jeb Thomas-Mitchell, Peter Dushniku, Nathan Laundry, Joseph Jay Williams, and Anastasia Kuzminykh. 2025. Co-writing with ai, on human terms: Aligning research with user demands across the writing process. *arXiv preprint arXiv:2504.12488* (2025).

[39] Jeba Rezwana and Mary Lou Maher. 2023. Designing creative AI partners with COFI: A framework for modeling interaction in human-AI co-creative systems. *ACM Transactions on Computer-Human Interaction* 30, 5 (2023), 1–28.

[40] Ben Samuel, Michael Mateas, and Noah Wardrip-Fruin. 2016. The design of Writing Buddy: a mixed-initiative approach towards computational story collaboration. In *International Conference on Interactive Digital Storytelling*. Springer, 388–396.

[41] Sophie Slaats and Andrea E Martin. 2025. What's surprising about surprisal. *Computational Brain & Behavior* 8, 2 (2025), 233–248.

[42] Hari Subramonyam, Roy Pea, Christopher Pondoc, Maneesh Agrawala, and Colleen Seifert. 2024. Bridging the gulf of envisioning: Cognitive challenges in prompt based interactions with llms. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–19.

[43] Katherine Thai, Bradley Emi, Elyas Masrour, and Mohit Iyyer. 2025. Editlens: Quantifying the extent of ai editing in text. *arXiv preprint arXiv:2510.03154* (2025).

[44] Todd M Thrash and Andrew J Elliot. 2003. Inspiration as a psychological construct. *Journal of personality and social psychology* 84, 4 (2003), 871.

[45] Jordi Tost, Marcel Gohsen, Britta Schulte, Fidel Thomet, Mattis Kuhn, Johannes Kiesel, Benno Stein, and Eva Hornecker. 2024. Futuring machines: An interactive framework for participative futuring through human-ai collaborative speculative fiction writing. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces*. 1–7.

[46] Qian Wan, Siying Hu, Yu Zhang, Piaohong Wang, Bo Wen, and Zhicong Lu. 2024. " It Felt Like Having a Second Mind": Investigating Human-AI Co-creativity in Prewriting with Large Language Models. *Proceedings of the ACM on human-computer interaction* 8, CSCW1 (2024), 1–26.

[47] 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 International Conference on Intelligent User Interfaces Workshops: APEx-UI, HAI-GEN, HEALTHI, HUMANIZE, TExSS, SOCIALIZE (IUI-WS 2022)*. CEUR-WS Team, 56–65.

[48] Kaixun Yang, Yixin Cheng, Linxuan Zhao, Mladen Raković, Zachari Swiecki, Dragan Gašević, and Guanliang Chen. 2024. Ink and algorithm: Exploring temporal dynamics in human-AI collaborative Writing. *arXiv preprint arXiv:2406.14885* (2024).

[49] Ann Yuan, Andy Coenen, Emily Reif, and Daphne Ippolito. 2022. Wordcraft: story writing with large language models. In *Proceedings of the 27th International Conference on Intelligent User Interfaces*. 841–852.

[50] Zheng Zhang, Jie Gao, Ranjodh Singh Dhaliwal, and Toby Jia-Jun Li. 2023. Visar: A human-ai argumentative writing assistant with visual programming and rapid draft prototyping. In *Proceedings of the 36th annual ACM symposium on user interface software and technology*. 1–30.

[51] David Zhou and Sarah Sterman. 2024. Ai. llude: Investigating Rewriting AI-Generated Text to Support Creative Expression. In *Proceedings of the 16th Conference on Creativity & Cognition*. 241–254.

A Prompts

A.1 Prompt for Completion Mode

system

You are an AI that excels at novel plot writing and is highly creative. Based on a novel plot outline that the user is currently writing, please provide suggestions for the next plot development and output one sentence in the form of text completion. Just provide a short sentence and do not output any other text.
Core requirements:
1. Deeply understand the current plot direction and provide logical and creative next steps
2. Prioritize unexpected twists, character depth development, plot conflict escalation and other creative elements, focusing on dramatic and readable plots
3. The continuation must seamlessly connect with the user's original text to form a complete and smooth sentence, maintaining the style and tone of the original text
4. Strictly prohibit repeating any part of the user's original text
5. Only output one short sentence, directly completing the user's last sentence, without any additional explanations

A.2 Prompt for Proposal Mode

system

You are an AI that excels at novel plot writing and is highly creative. Based on a novel plot outline that the user is currently writing, please provide suggestions for the next plot development. You need to output exploratory suggestions in the form of questions about the plot that is about to happen in the next paragraph. Note that you cannot provide direct continuations, but rather suggestions that the user cannot directly apply. Just provide a brief suggestion and do not output any other text.
Core requirements:
1. Deeply understand the current plot direction and provide logical and creative next steps
2. Be concise and closely focused on the current text, providing short-term (next sentence) development suggestions, not too detached from the current plot
3. Use exploratory tone and questions
4. Express curiosity and thinking, rather than simple emotional reactions
5. Avoid sounding like a professional critic or editor, maintain the perspective of an ordinary reader
6. Only output one short sentence, without any explanations or other additional text