12th International Conference on Computational Social Science IC^{2}S^2
July 28-31, 2026, Burlington, VT, United States

PromptHelper: A Prompt Recommender System for Encouraging Creativity in AI Chatbot Interactions

Keywords: LLM, Recommender System, Prompting, Creativity, Personalization

Extended Abstract

Prompting has become the primary interface through which end users interact with large language models (LLMs) [11, 17, 16], yet effective prompting remains challenging. While generative AI systems appear accessible through a simple chatbox interface, prior research shows that users often struggle to articulate intent, systematically explore alternatives, and understand how variations in phrasing shape model outputs [19, 14]. As a result, interaction with LLMs frequently narrows to a small region of the prompt space, limiting both output quality [2] and creative exploration [18]. In many current systems, the burden of iteration, comparison, and refinement falls almost entirely on the user: individuals must generate variations manually, remember previous attempts, and reason about how prompt modifications influence responses [19, 14]. This dynamic creates cognitive overhead and may lead to fixation on a narrow range of ideas, especially in creative and knowledge-intensive tasks such as writing [19].

These challenges have motivated a wide range of AI support tools aimed at assisting users in constructing, refining, and managing prompts, including templates, guidelines, visualizations, and automated prompt rewriting or optimization techniques [1, 6, 10, 12, 9, 13]. Each approach varies in their degree of autonomy and objective towards optimization or exploration. While existing approaches are effective in different contexts, few support user-controlled exploration of the prompt space during ongoing AI workflows. Such support is critical in this context, where autonomy can enhance creative outcomes, while ongoing exploratory guidance may mitigate idea fixation [15].

In this work, we propose **prompt recommender systems (PRS)** as an interaction paradigm that treats prompts themselves as recommender-eligible items. Rather than optimizing toward a single "best" prompt, PRS are designed to expand users' awareness of possible next steps by surfacing contextually relevant, semantically diverse follow-up prompts. We instantiate this concept through PromptHelper, a PRS prototype integrated into an AI-assisted writing environment (Figure 1). PromptHelper generates structured prompt recommendations grounded in the user's most recent query and the model's response, and presents them as selectable, modifiable alternatives that reflect distinct creative directions. Recommendations are intentionally concise, diverse across categories, and designed to preserve user agency: users may ignore, modify, or directly reuse suggested prompts. **We release open-source resources** (`https://anonymous.4open.science/r/Prompt_Recommender-6B80`) including system code, prompt data, and analysis scripts for further research on prompt assistance.

We evaluate PromptHelper in a fully within-subjects 2×2 study ($N = 32$) crossing Task (Creative vs. Academic writing) with System State (PromptHelper ON vs. OFF). Participants completed four 10-minute writing tasks using a baseline chatbot, with and without PromptHelper support. After each task, participants completed validated measures derived from the Creativity Support Index (CSI) [4] and NASA Task Load Index (NASA-TLX) [7] , capturing perceived exploration, expressiveness, results-worth-effort, usability, and workload

(Table 1 and Figure 2). Full quantitative results can be found in Table 2 and Table 3 and qualitative results in Table 4.

▷ **F1: PROMPTHELPER supports exploration of the conversational space.** Quantitative results indicate that enabling PROMPTHELPER significantly increased perceived exploration ($p = .001, \eta_p^2 = .293$), with follow-up tests confirming improvements in both academic ($p = .0110$, $d_z = 0.6$) and creative tasks ($p = .0242$, $d_z = .45$). Participants described the system as expanding the range of possible next steps, surfacing alternative directions they would not have generated independently, and helping them move past moments of uncertainty. Across contexts, recommendations were characterized as varied and diverse, suggesting that continuous, in-context prompt suggestions support sustained exploration.

▷ **F2: PROMPTHELPER scaffolds user expressiveness in constrained writing contexts.** PROMPTHELPER also had a significant main effect on expressiveness ($p = .011, \eta_p^2 = .193$). Follow-up analyses show that this effect was significant for academic writing ($p = .0018$, $d_z = 0.57$), but not for creative writing ($p = .2043$, $d_z = 0.25$). These results suggest that the system more strongly supports expressiveness in domains characterized by established jargon, which can make it difficult for users to articulate effective prompts without guidance [8]. In this context, participants reported that recommendations provided concrete examples and templates for framing requests, reducing uncertainty about how to articulate intent.

▷ **F3: PROMPTHELPER shifts cognitive effort from prompt generation to prompt evaluation.** Across workload and usability measures, no significant differences were observed between conditions ($p > .05$), indicating that the benefits of PROMPTHELPER were not accompanied by increased mental demand, frustration, temporal demand, or overall workload. Usability ratings likewise did not differ significantly. Qualitative feedback suggests that rather than reducing total effort, the system redistributed it. Participants reported using recommendations as starting points, modifying them or drawing inspiration rather than copying verbatim. This pattern indicates a shift in cognitive effort away from generating prompts from scratch and toward evaluating, comparing, and refining suggested directions, supporting a more iterative and judgment-focused interaction process.

Together, these findings position prompt recommender systems as a lightweight yet effective interaction technique for supporting exploratory human–AI workflows. By framing prompts as first-class interaction objects that can be surfaced, compared, and iteratively refined, PRS expand the design space of LLM interfaces beyond the single blank chatbox. More broadly, this work highlights the potential of recommender-system principles to shape interactive AI design, opening future directions for personalization, adaptive ranking, support for AI literacy, and responsible integration of recommendations into conversational systems.

Beyond interface-level effects, prompt recommender systems introduce a new layer of algorithmic mediation in human–AI knowledge production. By shaping which directions are encouraged during interaction, PRS may alter search trajectories within semantic spaces, reallocate attention across competing argumentative frames, and influence how ideas converge or diverge across interaction sequences. At scale, such systems could systematically structure these patterns, with implications for collective creativity, epistemic diversity, and the distribution of attention across topics and perspectives. **Framing prompts as recommender-eligible items therefore opens a broader computational social science agenda**: modeling prompt-space navigation as a form of guided cognitive search, examining how recommendation strategies alter discourse trajectories, and investigating how personalization and ranking mechanisms may shape knowledge production in AI-augmented environments.

Note: This work has been accepted to the 2026 ACM CHI Conference on Human Factors in Computing Systems (CHI 2026).

References

[1] Lakshya A Agrawal et al. "Gepa: Reflective prompt evolution can outperform reinforcement learning". In: *arXiv preprint arXiv:2507.19457* (2025).

[2] Liam Barkley and Brink van der Merwe. *Investigating the Role of Prompting and External Tools in Hallucination Rates of Large Language Models*. 2024. arXiv: 2410.19385 [cs.CL]. URL: https://arxiv.org/abs/2410.19385.

[3] Janet A Boekhorst, Michael Halinski, and Jessica RL Good. "Fun, friends, and creativity: A social capital perspective". In: *The Journal of Creative Behavior* 55.4 (2021), pp. 970–983.

[4] Erin Cherry and Celine Latulipe. "Quantifying the creativity support of digital tools through the creativity support index". In: *ACM Transactions on Computer-Human Interaction (TOCHI)* 21.4 (2014), pp. 1–25.

[5] Glorianna Davenport, Lars Erik Holmquist, and Maureen Thomas. "Fun: A condition of creative research". In: *IEEE MultiMedia* 5.03 (1998), pp. 10–15.

[6] Chrisantha Fernando et al. "Promptbreeder: Self-referential self-improvement via prompt evolution". In: *arXiv preprint arXiv:2309.16797* (2023).

[7] Sandra G Hart and Lowell E Staveland. "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research". In: *Advances in psychology*. Vol. 52. Elsevier, 1988, pp. 139–183.

[8] Imose Itua et al. "Exploring barriers and solutions to academic writing: Perspectives from students, higher education and further education tutors". In: *Journal of further and Higher Education* 38.3 (2014), pp. 305–326.

[9] Anjali Khurana, Hariharan Subramonyam, and Parmit K Chilana. "Why and When LLM-Based Assistants Can Go Wrong: Investigating the Effectiveness of Prompt-Based Interactions for Software Help-Seeking". In: *Proceedings of the 29th International Conference on Intelligent User Interfaces*. IUI '24. ACM, Mar. 2024, pp. 288–303. DOI: 10.1145/3640543.3645200. URL: http://dx.doi.org/10.1145/3640543.3645200.

[10] Weize Kong et al. "Prewrite: Prompt rewriting with reinforcement learning". In: *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. 2024, pp. 594–601.

[11] Pengfei Liu et al. "Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing". In: *ACM computing surveys* 55.9 (2023), pp. 1–35.

[12] Qianou Ma et al. "What should we engineer in prompts? training humans in requirement-driven llm use". In: *ACM Transactions on Computer-Human Interaction* 32.4 (2025), pp. 1–27.

[13] Aditi Mishra et al. "PromptAid: Visual prompt exploration, perturbation, testing and iteration for large language models". In: *IEEE Transactions on Visualization and Computer Graphics* 31.10 (2025), pp. 6946–6962.

[14] Sydney Nguyen et al. "How beginning programmers and code llms (mis) read each other". In: *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 2024, pp. 1–26.

[15] Peinuan Qin et al. "Timing Matters: How Using LLMs at Different Timings Influences Writers' Perceptions and Ideation Outcomes in AI-Assisted Ideation". In: *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. CHI '25. Association for Computing Machinery, 2025. ISBN: 9798400713941. DOI: 10.1145/3706598.3713146. URL: https://doi.org/10.1145/3706598.3713146.

[16] Pranab Sahoo et al. "A systematic survey of prompt engineering in large language models: Techniques and applications". In: *arXiv preprint arXiv:2402.07927* (2024).

[17] Sander Schulhoff et al. "The prompt report: a systematic survey of prompt engineering techniques". In: *arXiv preprint arXiv:2406.06608* (2024).

[18] Abhinav Sood, Maria Teresa Llano, and Jon McCormack. *Do Conversational Interfaces Limit Creativity? Exploring Visual Graph Systems for Creative Writing*. 2025. arXiv: 2507.08260 [cs.HC]. URL: https://arxiv.org/abs/2507.08260.

[19] Hari Subramonyam et al. "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*. 2024, pp. 1–19.

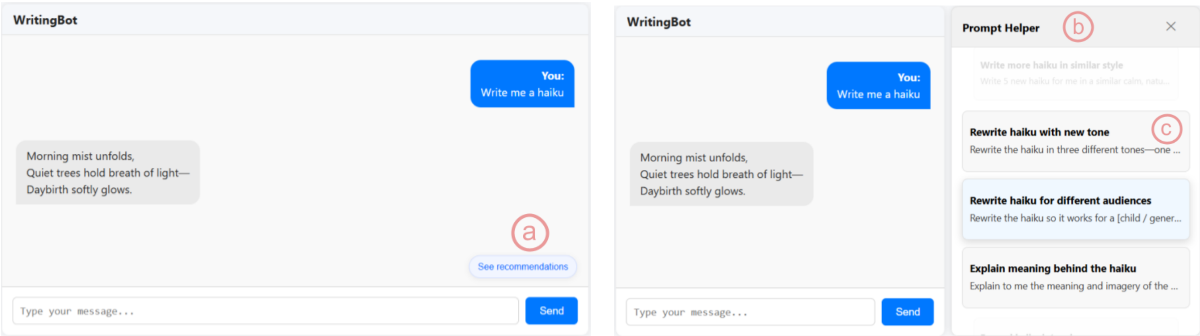

Figure 1: **Compared to a standard chatbot (left), PROMPTHELPER (right) supports user interactions by encouraging creativity and diversity:** The user interacts with a baseline writing chatbot as normal. Clicking (a) surfaces PROMPTHELPER, (b), which displays contextually relevant follow-up prompt suggestions, based on the user's last prompt and the model's response. Suggested prompts, (c), highlight different creative directions (e.g., rewriting, audience adaptation, explanation) and can be copied and modified to support iterative exploration during writing.

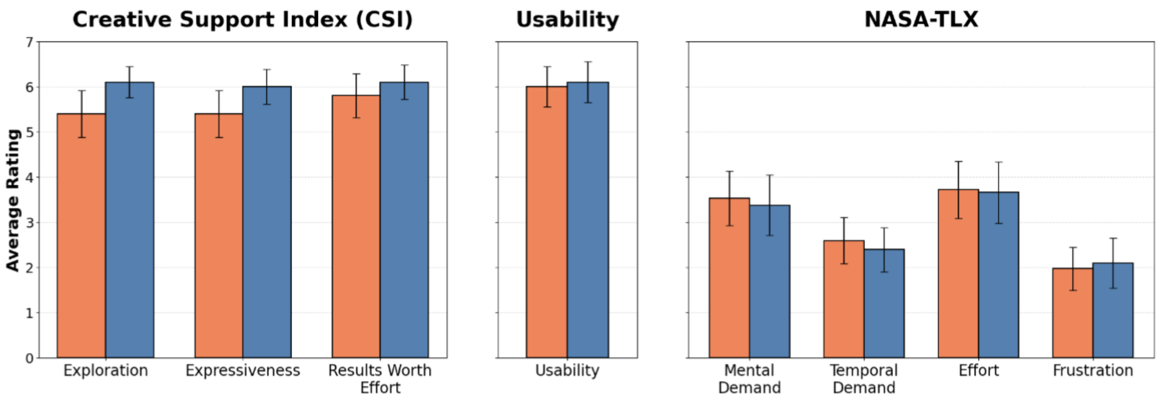

Figure 2: **Across tasks, PROMPTHELPER improved users' perceived ability to explore and express ideas, while leaving workload and usability unchanged**, suggesting that prompt recommender systems can scaffold exploratory interaction without imposing additional cognitive burden. Orange indicates ratings for the baseline chatbot, and blue indicates ratings for the baseline chatbot with PROMPTHELPER enabled. We report the mean and standard error of all Likert scales using a 95% confidence interval.

Q	Type	Source	Composite	Question
(A) Pre-Survey Questions – Performed once before the survey.				
QA1	Open-Ended	-	-	What experience do you have with using chatbots (e.g. ChatGPT, Gemini, Copilot, Claude, etc.)?
QA2	Likert	-	-	How often do you use such tools for writing academic or creative pieces?
QA3	Likert	-	-	How comfortable are you with using these tools in writing?
QA4	Open-Ended	-	-	What do you know about prompting?
(B) Survey Questions – Performed four times after each 10-minute writing task in the survey.				
QB1	Likert	CSI	Exploration	It was easy for me to explore many different ideas, options, designs, or outcomes, using this system.
QB2	Likert	CSI	Exploration	The system was helpful in allowing me to track different ideas, outcomes, or possibilities.
QB3	Likert	CSI	Expressiveness	I was able to be very creative while doing the activity inside this system.
QB4	Likert	CSI	Expressiveness	The system allowed me to be very expressive.
QB5	Likert	CSI	Results-Worth-Effort	I was satisfied with what I got out of the system.
QB6	Likert	CSI	Results-Worth-Effort	What I was able to produce was worth the effort I had to exert to produce it.
QB7	Likert	NASA-TLX	Mental Demand	How mentally demanding was the task?
QB8	Likert	NASA-TLX	Results-Worth-Effort	How hurried or rushed was the pace of the task?
QB9	Likert	NASA-TLX	Mental Demand	How hard did you have to work to accomplish your level of performance?
QB10	Likert	NASA-TLX	Usability	How insecure, discouraged, irritated, stressed, and annoyed were you?
QB11	Likert			Self-rate your written response on an (A-F) grading scale.
QB12	Likert	NASA-TLX	Usability	Rate the overall usability of the system.
(C) Post-Survey Questions – Performed once after the survey.				
QC1	Open-Ended	-	-	How comfortable were you with using Prompt Helper and WritingBot? Were there any moments where you felt uncertain about what it was doing or suggesting?
QC2	Open-Ended	-	-	How did your approach to writing using WritingBot compare between tasks with and without Prompt Helper?
QC3	Open-Ended	-	-	How would you describe the range or variety of prompts suggested by Prompt Helper (e.g., varied, repetitive, surprising, helpful)?
QC4	Open-Ended	-	-	In what ways did Prompt Helper affect how you generated or developed your ideas? Did it lead you to explore directions you wouldn't have taken on your own?
QC5	Open-Ended	-	-	How did you use the prompts being recommended by Prompt Helper (e.g. ignored, inspired, modified, or copied verbatim)?
QC6	Open-Ended	-	-	Did interacting with Prompt Helper make it easier or harder to decide what to ask next? Please describe why.
QC7	Open-Ended	-	-	Is there anything else you'd like to share about your experience with WritingBot and/or Prompt Helper?

Table 1: **Study questionnaire design**: Likert (1–7) and open-ended items mapped to CSI [4] and NASA-TLX NASA-TLX [7] constructs and aggregated into composite measures for analysis.

Source	SS	df	MS	F	p	η_p^2	Sig.
Exploration							
Task (condition)	0.5000	1,31	0.5000	0.51	.479	.016	✗
System (state)	14.4453	1,31	14.4453	12.82	.001	.293	✓
Task × System	0.3828	1,31	0.3828	0.44	.512	.014	✗
Expressiveness							
Task (condition)	0.0957	1,31	0.0957	0.14	.710	.005	✗
System (state)	8.7676	1,31	8.7676	7.41	.011	.193	✓
Task × System	1.0332	1,31	1.0332	1.95	.172	.059	✗
Results Worth Effort							
Task (condition)	4.1843	1,31	4.1843	5.99	.020	.162	✓
System (state)	3.1698	1,31	3.1698	2.48	.126	.074	✗
Task × System	0.0057	1,31	0.0057	0.01	.922	.000	✗
Usability							
Task (condition)	2.8203	1,31	2.8203	3.58	.068	.103	✗
System (state)	0.3828	1,31	0.3828	0.25	.618	.008	✗
Task × System	1.7578	1,31	1.7578	1.91	.177	.058	✗
Frustration							
Task (condition)	0.0313	1,31	0.0313	0.02	.882	.001	✗
System (state)	0.5000	1,31	0.5000	0.29	.593	.009	✗
Task × System	0.7813	1,31	0.7813	0.91	.348	.028	✗
Mental Demand							
Task (condition)	0.5000	1,31	0.5000	0.41	.525	.013	✗
System (state)	0.7813	1,31	0.7813	0.29	.596	.009	✗
Task × System	0.1250	1,31	0.1250	0.23	.635	.007	✗
Temporal Demand							
Task (condition)	0.1953	1,31	0.1953	0.14	.715	.004	✗
System (state)	1.3203	1,31	1.3203	1.49	.231	.046	✗
Task × System	1.3203	1,31	1.3203	1.91	.177	.058	✗
Effort							
Task (condition)	0.7813	1,31	0.7813	0.34	.563	.011	✗
System (state)	0.1250	1,31	0.1250	0.04	.841	.001	✗
Task × System	2.5313	1,31	2.5313	1.96	.171	.060	✗

Table 2: **Repeated-measures ANOVA results** for within-subjects effects of task (creative vs. academic), system state (PromptHelper ON vs. OFF), and their interaction across all dependent measures. Reported statistics include sums of squares (SS), degrees of freedom (df), mean squares (MS), F-values, p-values, and partial eta-squared (η_p^2).

Comparison	t	p_{unc}	p_{corr}	Sig.	d_z
Exploration					
Academic Task: ON vs. OFF	2.98	.0055	.0110	✓	0.60
Creative Task: ON vs. OFF	2.37	.0242	.0242	✓	0.45
Expressiveness					
Academic Task: ON vs. OFF	3.67	.0009	.0018	✓	0.57
Creative Task: ON vs. OFF	1.30	.2043	.2043	✗	0.25
Results Worth Effort					
Academic Task vs. Creative Task in the ON state	1.92	.0636	.1272	✗	0.32
Academic Task vs. Creative Task in the OFF state	1.72	.0949	.1272	✗	0.27

Table 3: **For significant composites from the ANOVA (shown in Table 2), we conduct follow-up pairwise t-tests with Holm-Bonferroni corrections for multiple comparisons**, and show the results here.

Participant	Question	Quote
On **saving time/efficiency**:		
P31	Q6	*"It made it easy by saving me the time I would have used to come up with the follow-up prompts"*
P10	QC1	*"The recommendations provided by Prompt Helper were comprehensible and understandable in most cases. The instances of some suggestions being unpredictable or slightly off topic were few, and meant that I had to think hard about whether to implement them or not, but on the whole, these were few and did not interfere with the workflow."*
P2	Q8	*"Interacting with Prompt Helper made it easier to decide what to ask next because it reduced the uncertainty around how to phrase a prompt. Seeing examples helped me quickly identify what kind of information the system could use, and that made the next step feel more obvious. It also saved time by giving me a clearer direction instead of having to come up with every prompt from scratch."*
On **considering new perspectives** and **navigation of the prompt space**:		
P29	Q4	*"I was able to gain a new perspective that will help my story. I wouldn't have discovered it on my own if it hadn't been shown to me."*
P17	Q4	*"I let the Prompt Helper design the road map for the conversation. I was curious to see how close it could get to mirroring my thoughts. I found it had interesting perspectives that went in different directions form my own and made me reconsider a few of my views."*
On **usability**:		
P29	Q1	*"I was fairly comfortable using the writing bot. However, I had some uncertain moments with the prompt helper because I couldn't figure out how to use it at that time."*
P23	Q7	*"Everything worked fine. Just the "when do I stop" part was the only confusing thing."*
P7	Q7	*"I copied verbatim, I didn't want to confuse the AI with what I was asking."*
On **fun**, a precondition for creative knowledge work [5, 3]:		
P30	Q7	*"This was a fun task."*

Table 4: **Representative participant quotes** illustrating perceived efficiency gains, exploration of new perspectives, usability considerations, and enjoyment when using PROMPTHELPER.