Brainstorming with a Generative Language Model: Effect of Exposure to AI Ideas on Brainstorming Performance and Cognitive Load

Lucas Memmert, Izabel Cvetkovic, Navid Tavanapour, Eva Bittner

Business & Information Systems Engineering (2025)

Appendix (available online via http://link.springer.com)

Appendix A: Information Flow Overview

This screenshot (see Fig. A1) contains an overview of the interaction between the human and the brainstorming app as well as the information flow within the brainstorming app and towards the external generative language model (GLM) provider. Ideas entered by the user are displayed on the left. If the user clicks the button “Generate Ideas” (indicated by a green dot), a request is initiated to the external GLM provider, here, OpenAI’s GPT-3.5. The entire canvas contents of the participant, i.e., brainstorming question, list of ideas of the participant (if not empty), and list of previous GLM ideas for this participant (if not empty), are used to populate a predefined prompt template (see Appendix B). The GLM response to the prompt is split up into three separate ideas, and these are added as GLM ideas on top of the list of the “list of AI suggestions” on the right.

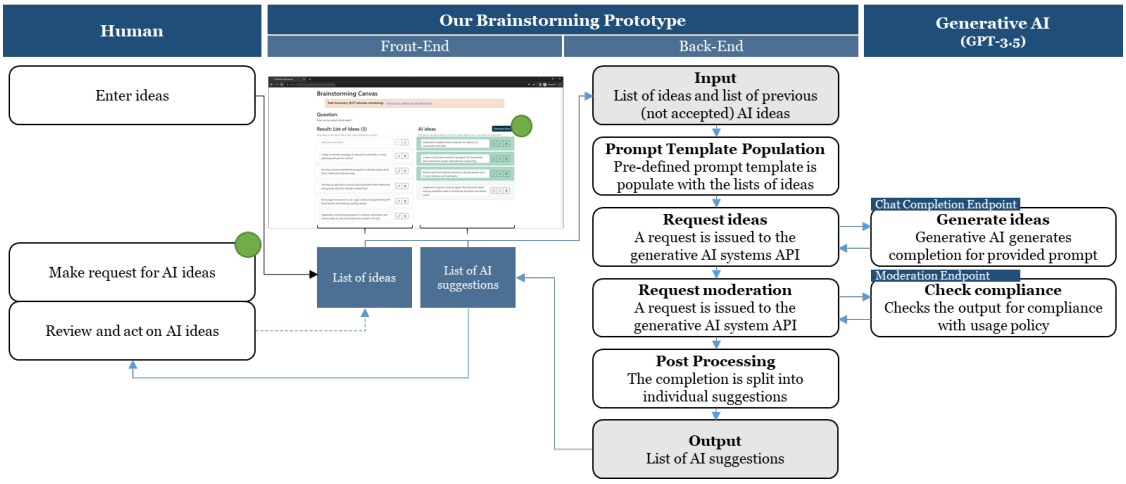

Fig. A.1 User interaction with and information flow within the brainstorming app

Appendix B: Prompt Template Development

The output of the GLM (in our study, the system-generated idea suggestions) is dependent on the prompt, making the prompt a critical component for the app (Schneider et al. 2024). The figure below contains the prompt template. The template is defined at development time and populated dynamically during runtime (see Appendix A for details). In this prompt template, curly brackets indicate placeholders, all caps placeholder labels (blue) indicate parameters fixed for this study, small caps placeholder labels (green) indicate parameters changing based on the current state of the canvas, and portions in italic font are only included if ideas are already present in the respective lists for the specific participant. We have adapted the prompt template from previous studies (Memmert and Tavanapour 2023; Memmert et al. 2024) and used the itemization prompt engineering technique (Mishra et al. 2022) when creating this template. This template is flexible to accept other brainstorming questions (see respective placeholder). The parameter values fixed for this study were:

- BRAINSTORMING_QUESTION= How can we reduce food waste?
- NUMBER_OF_IDEAS=3


```
You are part of a brainstorming team. Your goal is to come up with novel
and valuable ideas for the following question: {BRAINSTORMING_QUESTION}
Discarded ideas so far:
{list_glm_ideas from this human+GLM session}
Novel and valuable ideas so far:
{long_list_ideas from this human+GLM session}
Please come up with {NUMBER_OF_IDEAS} additional novel and valuable ideas
for the question: "{BRAINSTORMING_QUESTION}". Please provide the
{NUMBER_OF_IDEAS} additional ideas as enumerated, ordered list. Each idea
should be limited to a maximum of 20 words.
```

Fig. B.1 Prompt template with placeholders

Appendix C - Survey Instruments

Concept	Reference	Asked to	Question/Text	Scale	Scale_Detail
Demographics	---	ALL	How old are you?	numeric	
	---	ALL	What is your gender?	single select	
	---	ALL	What is you current study program?	single select	
Cognitive Load	Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In Advances in Psychology. Human Mental Workload (Vol. 52, pp. 139–183). Elsevier.	ALL	Mental Demand: How mentally demanding was the task?	21-point	Very low-Very high
		ALL	Physical Demand: How physically demanding was the task?	21-point	Very low-Very high
		ALL	Temporal Demand: How hurried or rushed was the pace of the task?	21-point	Very low-Very high
		ALL	Performance: How successful were you in accomplishing what you were asked to do?	21-point	Perfect-Failure
		ALL	Effort: How hard did you have to work to accomplish your level of performance?	21-point	Very low-Very high
		ALL	Frustration: How insecure, discouraged, irritated, stressed, and annoyed were you?	21-point	Very low-Very high
Cognitive Stimulation	Adapted from Gozzo, M., Woldendorp, M. K., & Rooij, A. de. (2022). Creative Collaboration with the "Brain" of a Search Engine: Effects on Cognitive Stimulation and Evaluation Apprehension. In M. Wölfel, J. Bernhardt, & S. Thiel (Eds.), Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering. ArtsIT, Interactivity and Game Creation (Vol. 422, pp. 209–223).	HUMAN+AI	Please indicate to what extent the following statements apply to you (there is no right or wrong answers here).	-	
		HUMAN+AI	I felt inspired during the brainstorming task	5-point-likert	Strongly Agree-Strongly Disagree
		HUMAN+AI	The input of the AI did not help me	5-point-likert	Strongly Agree-Strongly Disagree
		HUMAN+AI	The input of the AI helped me to think creatively	5-point-likert	Strongly Agree-Strongly Disagree
		HUMAN+AI	Working together with the AI at this moment was uninspiring	5-point-likert	Strongly Agree-Strongly Disagree
		HUMAN+AI	I could easily come up with associations	5-point-likert	Strongly Agree-Strongly Disagree
		HUMAN+AI	It was difficult to relate the input of the AI to my own thinking	5-point-likert	Strongly Agree-Strongly Disagree
Felt Responsibility	---	HUMAN+AI	For each result type described below, please indicate how you feel about your share of responsibility. Note: '0' means the AI was fully responsible...'100' means I feel I was fully responsible	-	
	---	HUMAN+AI	For the 'list of ideas' (prior to final idea selection) my share of responsibility was [%]:	numeric	0-100
	---	HUMAN+AI	For the 'list of 4 best ideas' (after the final idea selection) my share of responsibility was [%]:	numeric	0-100
	---	HUMAN+AI	Please briefly explain your reasoning for the previous answer for the share of responsiblity regarding the result type 'list of ideas' (prior to final idea selection):	long free-text	
	---	HUMAN+AI	Please briefly explain your reasoning for the previous answer for the share of responsiblity regarding the result type 'list of 4 best ideas' (after the final idea selection):	long free-text	

AI Team Role	---	HUMAN+AI	To achieve best performance when working with the tool with AI suggestions in a brainstorming task, I should...	5-point-likert	Focus on selecting and adjusting AI suggestions-Focus on generating ideas myself
GLM Effect	---	HUMAN+AI	How did the task change for you due to the presence of the AI (as compared to brainstorming alone)?	long free-text	
	---	HUMAN+AI	How did the AI (ideas/suggestions) help you in this brainstorming task?	long free-text	
	---	HUMAN+AI	How did the AI (ideas/suggestions) hinder you in this brainstorming task (i.e., made it difficult to perform the brainstorming task)?	long free-text	

Appendix D: Variable Operationalization

Table D.1 Operationalization of constructs (calculated per brainstorming session, considering only ideas of this session)

Construct	Variable name	List Basis	Operationalization	Rated by/ Data source	Reference(s)
Fluency	Fluency	long list	Number of ideas produced within the 10-minute session, according to the system log	system log	Paulus (2000)
Flexibility	Flexibility	long list	Number of distinct categories with ≥ 1 idea in the category system of Specht and Buck (2019)	blind student coder	Nijstad et al. (2010), Althuizen and Reichel (2016)
Novelty	Novelty	short list	Mean novelty on a 7-point Likert scale	5 blind-to-condition human judges	Siangliulue (2015b), Siangliulue et al. (2015a), Zhu et al. (2021)
	$\text{Novelty}_{\text{GPT-rated}}$	short list	Mean novelty on a 7-point Likert scale	GPT-4	Dell'Acqua et al. (2023), Haase and Hanel (2023), Organisciak et al. (2023)
	Good-Idea-Count-Novelty	long list	Number of long-list ideas with GPT-4 novelty rating > avg (avg. novelty calculated across long lists from all sessions)	GPT-4	Reinig et al. (2007)
Value	Value	short list	Mean value-rating on a 7-point Likert scale	5 blind-to-condition human judges	Siangliulue (2015b), Siangliulue et al. (2015a), Zhu et al. (2021)
	$\text{Value}_{\text{GPT-rated}}$	short list	Mean value-rating on a 7-point Likert scale	GPT-4	Dell'Acqua et al. (2023), Haase and Hanel (2023), Organisciak et al. (2023)
	Good-Idea-Count-Value	long list	Number of long-list ideas with GPT-4 value-rating > avg (avg. value calculated across long lists from all sessions)	GPT-4	Reinig et al. (2007)
Cognitive Load	NASA-TLX	—	NASA-TLX overall score (6-dim self-report)	participant self-report	Hart and Staveland (1988)

All idea-related variables (i.e., all variables except for cognitive load) were calculated on the level of each separate brainstorming session. Each variable was calculated on two performance levels (also see Fig. 2 in the main manuscript): human-level (considering only ideas of human origin) and human-GLM dyad-level (considering ideas of both human and GLM origin).

Appendix E: Coding Schema

Table E.1 Coding schema for analyzing the open-ended survey responses

Category	Code	Explanation & Example
Cognitive stimulation/inertia		Participant describes how the GLM affected their idea generation, e.g., stimulating or constraining it.
	stimulation	GLM helped initiate or expand ideas (e.g., new perspectives, entry support). *"The ideas on the side which came from the AI were a great inspiration for developping new ideas on my own." (P67)*
	fixation	GLM ideas narrowed thinking or governed the ideation process. *"It definitely felt like my thought process was being governed by the AI" (P48)*
Effort (in-)dispensability		Participant reflects on the (non-)necessity, reduction, or redirection of their own effort.
	effort reduced	The participant relied on the GLM, putting in less effort themselves. *"i became lazy" (P61)*
	effort dispensable	Participant perceives that their own effort or own contribution is (somewhat) unnecessary (compared to the GLM's contribution) for (parts of) the task. *"i am not that creative" (P60)*
	effort indispensable	The participant perceives their effort/ contribution as still essential. *"The AI ideas seemed uncreative and boring" (P51)*
	effort shift	The participant switched focus, e.g., from generating to curating or extending ideas. *"All I had to do was read the ideas put forth and decide which ones make the cut." (P48)*
Responsibility (non-)diffusion		The participant provides their rationale for their perceived share of responsibility for the result.
	contribution-based	Rationale based on the origin of ideas, or share of ideas used or contributed (incl. copy/editing). *"I selected 3 of my answers and 1 AI answer" (P54)*
	curation responsibility	Rationale based on selection or evaluation effort for the ideas. *"AI can give ideas, but I felt like the selection of the 4 ideas is the key point in concluding such [a] task." (P52)*
Cognitive load		Participant describes how mentally demanding or effortless the task felt.
	load reduction	The task felt easier or more comfortable. *"The burden fell off my shoulders and instead of thinking hard, I let the AI write ideas." (P46)*
	mental disengagement	The participant mentally disengaged or stopped thinking. *"I feel like using the AI made my brain shut off a little bit" (P47)*
	load increase	Working with GLM required additional mental effort. *"I had to reject the ai suggestions and pay attention not to be influenced too much by it or to be bogged down in the ai's way of thinking." (P75)*

References

Althuizen N, Reichel A (2016) The Effects of IT-Enabled Cognitive Stimulation Tools on Creative Problem Solving: A Dual Pathway to Creativity. Journal of Management Information Systems 33(1):11–44. doi:10.1080/07421222.2016.1172439.

Dell'Acqua F, McFowland E, Mollick ER, Lifshitz-Assaf H, Kellogg K, Rajendran S, Krayer L, Candelon F, Lakhani KR (2023) Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality. Harvard Business School Working Paper.

Haase J, Hanel PHP (2023) Artificial muses: Generative Artificial Intelligence Chatbots Have Risen to Human-Level Creativity. Journal of Creativity 33(3). doi:10.48550/arXiv.2303.12003.

Hart SG, Staveland LE (1988) Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In: Human Mental Workload. Elsevier, pp 139–183. doi:10.1016/S0166-4115(08)62386-9.

Memmert L, Cvetkovic I, Bittner E (2024) The More Is Not the Merrier: Effects of Prompt Engineering on the Quality of Ideas Generated By GPT-3. In: Proceedings of the 57th Hawaii International Conference on System Sciences, Honolulu, HI, USA, pp 7520–7529. https://hdl.handle.net/10125/107289.

Memmert L, Tavanapour N (2023) Towards Human-AI Collaboration in Brainstorming: Empirical Insights into the Perception of Working with a Generative AI. In: 31st European Conference on Information Systems. https://aisel.aisnet.org/ecis2023_rp/429.

Mishra S, Khashabi D, Baral C, Choi Y, Hajishirzi H (2022) Reframing Instructional Prompts to GPTk's Language. In: Findings of the Association for Computational Linguistics: ACL 2022, pp 589–612.

Nijstad BA, Dreu CKW de, Rietzschel EF, Baas M (2010) The dual pathway to creativity model: Creative ideation as a function of flexibility and persistence. European Review of Social Psychology 21(1):34–77. doi:10.1080/10463281003765323.

Organisciak P, Acar S, Dumas D, Berthiaume K (2023) Beyond semantic distance: Automated scoring of divergent thinking greatly improves with large language models. Thinking Skills and Creativity 49. doi:10.1016/j.tsc.2023.101356.

Paulus P (2000) Groups, Teams, and Creativity: The Creative Potential of Idea-generating Groups. Applied Psychology 49(2):237–262. doi:10.1111/1464-0597.00013.

Reinig BA, Briggs RO, Nunamaker JF (2007) On the Measurement of Ideation Quality. Journal of Management Information Systems 23(4):143–161. doi:10.2753/MIS0742-1222230407.

Schneider J, Meske C, Kuss P (2024) Foundation Models. Business & Information Systems Engineering 66(2):221–231. doi:10.1007/s12599-024-00851-0.

Siangliulue P, Arnold KC, Gajos KZ, Dow SP (2015a) Toward Collaborative Ideation at Scale. In: CSCW '15 Proceedings. ACM, NY, USA, pp 937–945. doi:10.1145/2675133.2675239.

Siangliulue P, Chan J, Gajos KZ, Dow SP (2015b) Providing Timely Examples Improves the Quantity and Quality of Generated Ideas. In: C&C '15: Creativity and Cognition. ACM, NY, USA, pp 83–92. doi:10.1145/2757226.2757230.

Specht AR, Buck EB (2019) Crowdsourcing Change: An Analysis of Twitter Discourse on Food Waste and Reduction Strategies. Journal of Applied Communications 103(2). doi:10.4148/1051-0834.2240.

Zhu Y, Ritter SM, Dijksterhuis A (2021) The effect of rank-ordering strategy on creative idea selection performance. Eur. J. Soc. Psych. 51(2):360–376. doi:10.1002/ejsp.2743.