Memoire: Harnessing Generative AI to Bridge the Metacognitive Gap in Reflective Writing

Matea Tashkovska[1,*], Seyed Parsa Neshaei[1,*,†], Paola Mejia-Domenzain[1] and Tanja Käser[1]

[1]*EPFL, Switzerland*

Abstract

Reflective writing is considered an important metacognitive skill, especially in vocational education where students must bridge theoretical knowledge and practical experiences. However, meaningful reflection often requires supervision and guidance, as students struggle to make connections between classroom concepts and workplace experiences. While generative large language models (LLMs) have shown promise in such educational applications including personalized learning and writing support, their effectiveness is hindered by inherent issues such as hallucination and lack of personalization. To address this, retrieval-augmented generation (RAG) has emerged as a solution to enable models to integrate external information in text generation. While RAG has demonstrated success in various domains, its potential for enhancing reflective writing remains untapped. In this work-in-progress, we introduce `Memoire`, a writing assistant designed to utilize the capabilities of RAG to support students in reflective writing. By leveraging external knowledge and memory from prior reflections of each student, `Memoire` helps them write reflections that are both insightful and grounded in accurate prior information. We also conduct a pre-study to evaluate and compare three modalities of providing writing support in the domain of reflective writing from prior works. Finally, we introduce our study design plan for an in-classroom evaluation of `Memoire`[1].

Keywords

Reflective Writing, Writing Assistants, Generative AI, Large Language Models, Intelligent Learning Support, Retrieval-Augmented Generation

1. Introduction and Related Work

Reflective writing (i.e., journaling) is the process of writing about one's thoughts, experiences, and insights on certain scenarios or events [1]. The process of writing reflectively is considered a method to improve the metacognitive skills of students [2]. It is also known to help learners discover deeper insights into their actions and lead to improvement in their tasks and learning [3]. In particular, students in vocational schools frequently participate in writing-to-learn and journaling activities, as reflective writing has been shown to be effective in developing and acquiring the necessary knowledge for such students [4].

With that said, prior works have highlighted challenges for novice learners to come up with well-structured reflective writings, specifically regarding thinking back on their thoughts and emotions during the event or using specific reflective models [5, 6]. This has led to a research stream of works trying to design and evaluate tutoring systems to help learners write reflectively [7]. The difficulty in reflective writing is specifically pronounced among vocational school students, as in dual vocational systems, the students need to connect their theoretical knowledge obtained from the theory sessions in the classroom to the situations in the workplace they experience during their practical studies [8]. Constructivism, a learning theory emphasizing the importance of building knowledge through personal

[1]The code, the details of the prompts, the explanation of the personas, and the study questions can be found on: https://github.com/epfl-ml4ed/memoire.

Second International Workshop on Generative AI for Learning Analytics, 2025

*These authors contributed equally.
†Corresponding author.
matea.tashkovska@epfl.ch (M. Tashkovska); seyed.neshaei@epfl.ch (S. P. Neshaei); paola.mejia@epfl.ch (P. Mejia-Domenzain); tanja.kaeser@epfl.ch (T. Käser)
0009-0005-3036-8652 (M. Tashkovska); 0000-0002-4794-395X (S. P. Neshaei); 0000-0003-1242-3134 (P. Mejia-Domenzain); 0000-0003-0672-0415 (T. Käser)

experience and active reflection, highlights the need for students to not just absorb knowledge but to actively engage with it, and with that, link new concepts to their prior experiences [9]. In this context, reflective writing serves as a powerful tool that enables students to integrate theoretical knowledge with practical experiences, enhancing their understanding and preparing them for professional competence [10]. However, reflection, like most other writing tasks, does not always occur spontaneously, and learners often need support to overcome writer's block and make meaningful connections in their writings [11, 12].

The recent advances in natural language processing (NLP), specifically large language models (LLMs), have shown to be promising, effective, and useful for learning and writing assistance in various domains [13, 11, 14, 15]. However, they still face notable challenges. Notably, systems designed around LLMs can suffer from hallucination, when the models generate information that can be classified as unreliable or inaccurate [16, 17]. This issue is particularly concerning in pedagogical writing support tools, as it can negatively impact the learning or experience of the students using the tools. Additionally, intelligent tools built around LLMs are not personalized by default, that is, tailoring to each learner's own learning path and building upon the existing information from each learner to provide intelligent, relevant, and useful insights [18].

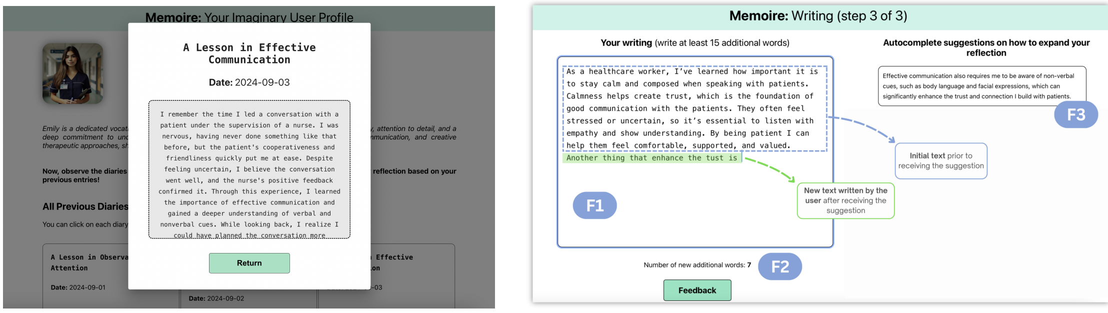
To mitigate the issue of hallucination in LLMs and to enable more personalized output based on prior text entries, previous works have proposed Retrieval-Augmented Generation (RAG) [19]. RAG enhances the capabilities of LLMs by incorporating external data, including previous information of each user related to their query, during the generation process. This helps the model produce more accurate and reliable outputs, as it relies on factual data retrieved from additional sources, thereby reducing the risk of generating misleading or incorrect information [20]. Although RAG-based methods have shown promise across a range of writing assistants and pedagogical tools [21, 22], their usage in reflective writing, and specifically their potential to link entries to prior journals from theory or practice sessions of learners, remains underexplored.

To address the research gap mentioned above, we suggested integrating a RAG-based pipeline into a writing assistant for reflective writing. By doing this, we specifically aimed to answer the following research questions (RQs): **RQ1)** How can we best provide AI-assisted support using a RAG-based method to help users in writing reflective texts? **RQ2)** How do users perceive the usefulness, benefits, or drawbacks of a RAG-based writing assistant for reflective writing?

To answer our RQs, we designed and developed `Memoire`, our writing assistant for reflective writing with intelligent suggestions grounded in previous reflection entries written by each learner. We implemented a RAG pipeline to be able to provide personalized suggestions connecting to the previous reflections of each user. We extracted three types of suggestions from prior work in learning sciences and writing assistants: A) critical questions [23, 24, 25], B) autocomplete suggestions [11, 26, 14], and C) summarizing feedback [27, 28]. We embedded the three approaches in `Memoire` and evaluated them in an online pre-study simulation conducted on Prolific with 17 users. Our results show a higher overall preference for autocomplete suggestions and critical questions compared to summarizing feedback, accompanied by a qualitative analysis of the open-answer comments provided by the participants in our pre-study. We finally provide our study plan for a larger-scale real-world classroom study, evaluating `Memoire` with vocational school students and their reflective entries written over a school semester. This enables us to test how our tool facilitates reflecting on and connecting the writing of the learners to their prior knowledge in real-life scenarios. Our work sheds light on the applicability of RAG-based methods for reflective writing in terms of mitigating the risks of hallucination and limited personalization, with the goal of helping learners reduce writer's block when reflecting, and assisting them in forming their reflective texts.

2. Design and Implementation of `Memoire`

To answer our research questions, we designed `Memoire`, our intelligent writing assistant for reflective writing support. `Memoire` is developed as a React-based web application with the MUI library for the



(a) (b)

Figure 1: The interface of `Memoire`, showing the past entries (a) and the writing interface (b).

front-end and the Python-based Flask framework for the back-end. The main interface of `Memoire`, designed to have a similar look and feel to many existing writing assistants [11], can be seen in Figure 1.

The learners first see their user profile as well as a grid list of their past reflective entries (Figure 1-a). We show up to 15 prior reflections to each user on this page. Then, the learners enter the writing phase (Figure 1-b), in which they can write their reflection in the designated text area (F1), up to a certain number of words (F2). After writing their reflection, they receive suggestions (F3) from one of the three suggestion types (see Section 2.1), relevant to their written text and the context of their most similar prior reflections.

2.1. Suggestion Types

To find the types of suggestions to show to the learners in `Memoire`, we searched the literature on intelligent and interactive writing assistants [11] to extract suggestion types used and validated in prior works. In particular, we picked three possible types of suggestions to use in `Memoire`:

- **A) Critical Questions:** Based on the prior research on the usefulness of asking questions to prompt metacognition and affect students' learning [24, 23, 25], we added this suggestion type to `Memoire`. We implemented `Memoire` such that it formulates three contextual questions for the learners to reflect on and help them maintain a thoughtful and reflective tone in their writing. `Memoire` shows the questions in a numerated list to the learners.
- **B) Autocomplete Suggestions:** Prior works in the domain of writing assistants have used autocomplete interfaces as a means to provide users with ideas on how to write and mitigate writer's block [11, 26, 14]. We also included this type of suggestions in `Memoire`. We implemented `Memoire` to assist the learner by generating the next sentence in their writing, focusing on overall ideas, and providing a continuation that is coherent and thoughtful. We ensured that `Memoire` only returns a single sentence at a time, naturally continuing the current writing of the user, to avoid excessively lengthy replies.
- **C) Summarizing Feedback:** Finally, similar to several prior works in the domain of writing assistants [27, 28], we also included a module in `Memoire` to provide a brief summary of the most relevant prior reflection, focusing on insights, and lessons, and main ideas, to inspire writing the current reflection and help the learners recall insights from their previous writings.

2.2. Retrieval Pipeline

We used RAG to implement the retrieval pipeline of `Memoire`. As explained in Section 1, RAG enhances language model outputs by referencing an external knowledge base, ensuring responses are more accurate and reliable.

In general, RAG consists of two primary components: a retriever and a generator. The *retriever* selects relevant documents from a knowledge base based on the query. In our system, this query corresponds to the reflection the learner is currently writing. The *generator* then uses the retrieved documents and the original query to generate a response. The retriever component in our RAG system is designed to efficiently select past reflections that are most relevant to the learner's current writing. To achieve this, we used OpenAI's "text-embedding-3-small" model. To implement the retrieval, we first embedded each prior reflection, creating a base of representations that serves as our retrieval dataset. We then embedded the student's current reflection (i.e., the query) using the same model. We used cosine similarity to measure the relevance between the query and each past reflection.

Finally, for the generator, we used the GPT-4o model provided by OpenAI, an efficient model released in May 2024, claimed by OpenAI as being their "most advanced" GPT model to date and used by researchers to inform the design and implementation of intelligent assistants [29]. The prompts used for the text generation phase were designed, refined, and finally approved, collaboratively by three learning sciences researchers in a workshop.

3. Experimental Evaluation

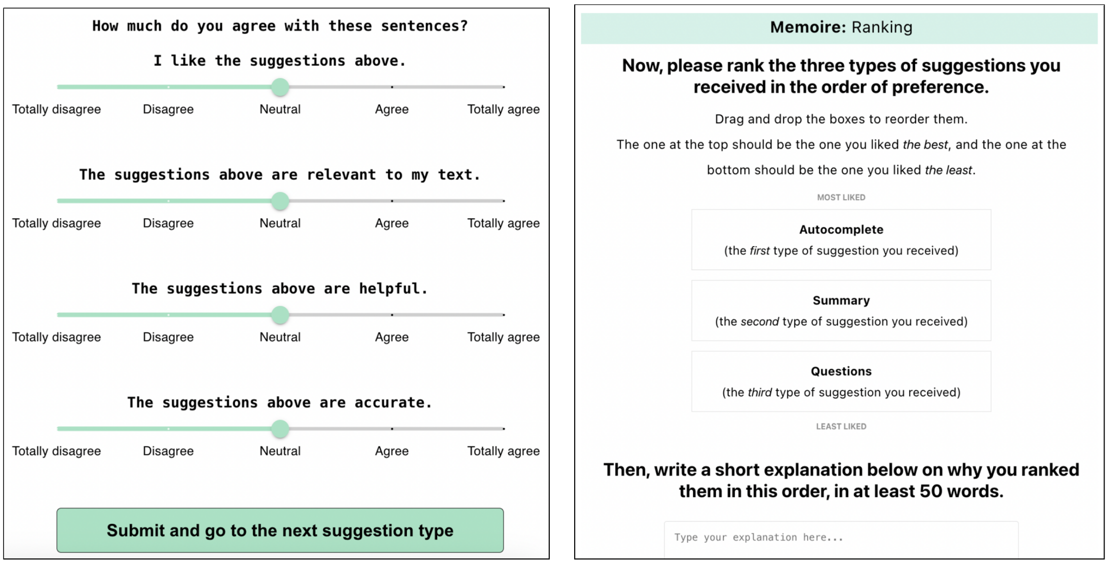
To find the answer to our RQs (as mentioned in Section 1), we A) conducted an experimental evaluation over Prolific as a pilot pre-study, with early results provided in this paper, and B) planned a classroom study with vocational school students, who will have already participated in reflective writing throughout a semester, to conduct as our main study in a follow-up.

For our pilot study, we conducted a crowdsourced pre-study over Prolific to observe early indications of the usefulness and benefits of `Memoire` among users, as well as to compare the effects and perceptions of users towards the three different suggestion types we used (as discussed in Section 2.1). In total, 17 people (13 identified as female and 4 identified as male; average age = 25.53, SD = 2.30) participated in our study. As we did not have access to the prior reflections of users in Prolific, nor were they necessarily confirmed to have practiced in reflective writing sessions before using a tool, we changed `Memoire` to provide a "simulation" environment for users to be able to conduct our pilot study. We defined three different personas of vocational school students and distributed them randomly to the online participants. Each persona came with a definition and a set of prior reflections[1]. The participants had to read at least five of the prior reflections before getting to start the writing session; we used this as a means of trying to simulate a real-world scenario in which learners have written their previous entries themselves, with the aim of helping each user relate to their assigned persona. The pre-study included the following components:

Pre-test: After logging in with their Prolific IDs, users first entered a pre-test, during which demographic information was collected. Additionally, we asked the users a set of two behavioral constructs to ensure correct randomization, picked from prior works on writing assistants [26, 30]: A) feedback-seeking behavior of participants [31], and B) information technology usage model [32], both measured in a 7-point Likert scale. We also asked them if they had participated in reflective writing studies before. The findings did not show any specific differences among users in our small sample, ensuring a valid randomization.

Writing task: After the pre-test, users entered the writing interface (see Figure 1), where they were instructed to act as their assigned persona and write a new reflection. To allow us to compare the three suggestion types (see Section 2.1), we made the writing interface such that it shows suggestions from the three types on three consecutive pages in a randomized order per user. When users entered their first text before receiving any suggestion, `Memoire` saved their text to use as a starting point for the subsequent pages. This ensured that the quality or relevance of the suggestions did not depend only on the specific content of the user's writing at each step, but was based on the consistency of the starting point across all suggestions. To be able to find the perception of the users towards each suggestion type, we added four questions (see Figure 2-a) asking the users to grade from 1 to 5 before going to the next page, whether A) they liked the suggestions, B) they considered them as relevant to their text, C) they found them helpful, and D) they perceived them as accurate. After the end of the writing tasks,

[1]Two researchers in the domain of learning sciences ensured and confirmed the validity and relevance of the reflective texts we used for the personas.



(a) (b)

Figure 2: The questions added to the interface of `Memoire` for our experimental study.

they were asked to 1) rank the three suggestion types based on their preference by a drag-and-drop interface, and 2) write a brief explanation behind their decision (see Figure 2-b).

Post-test: After the writing task, the users answered our post-test. The post-test consisted of an open question asking for feedback on the tool, including what they liked and did not like about `Memoire` and their suggestions for improvement, as well as a set of perception constructs picked from prior works on writing assistants [26, 30] and measured in 1-7 Likert scale: A) excitement after interaction, B) perceived usefulness, C) perceived ease of use, D) technology acceptance, E) correctness of the suggestions, F) perceived improvement in writing, and G) perceived improvement in writing in the long run.

Classroom study: We additionally plan to evaluate `Memoire` in an in-person vocational school classroom in a Western European country, among a set of learners who have been writing journal entries during the semester as a part of their curriculum. The in-class version will not have the "simulation" aspect of our Prolific pre-study, as we will use the prior real reflections of each user to inform the generations. In this work-in-progress workshop paper, only the results of the pilot pre-study are presented and discussed.

4. Results

We analyzed the logged data from the interactions of the online Prolific participants to provide early pre-study indications of the answers to our two RQs (see Section 1).

4.1. RQ1: Best Method of Providing Support

We investigated the scores the participants provided to each of the suggestion types, in terms of personal preference, relevancy, helpfulness, and accuracy of the suggestions. Table 1 shows the scores given by the participants. The data shows a higher mean for the autocomplete suggestion type compared to the rest in all four scoring questions, suggesting a higher preference towards this method of providing suggestions. Moreover, on the ranking page, 7 participants ranked autocomplete suggestions and 7 ranked critical questions as the best suggestion type, leaving summarizing feedback at a distant third with only 3 participants ranking it as the best suggestion type.

In addition, we analyzed the answers the participants (P1 to P17) provided to the open question asking them for the reasons behind their provided ranking, to find indications of justifications behind the effectiveness of each of the three suggestion types:

Table 1
The scores given by the participants to each of the suggestion types (Mean ± SD).

Suggestion Type	Personal Preference	Relevancy	Helpfulness	Accuracy
Critical Questions	3.41 ± 0.60	3.29 ± 0.67	3.24 ± 0.64	3.06 ± 0.73
Autocomplete Suggestions	**3.53** ± 0.70	**3.71** ± 0.57	**3.59** ± 0.60	**3.35** ± 0.68
Summarizing Feedback	3.35 ± 0.76	3.24 ± 0.73	3.24 ± 0.64	3.18 ± 0.86

Critical Questions: Five participants (P1, P3, P5, P6, and P10) mentioned that the critical questions made them reflect, think more, and challenge them positively. Four participants (P4, P8, P15, and P17) believed that the questions gave clear and straightforward instructions on what they should write next. None of the users mentioned any negative points about this suggestion type in the open-ended responses.

Autocomplete Suggestions: Three participants (P5, P12, and P14) believed that this suggestion type provided a perfect way to wrap up their thoughts and allowed them to develop ideas further. Specifically, five participants (P4, P8, P13, P15, and P17) considered this suggestion type as "helpful" or mentioned that the autocomplete suggestions provided their ideas and concepts in "better ways." On the other hand, four participants (P3, P6, P10, and P11) mentioned that the autocomplete suggestions seemed too general, were a repetition of what was already discussed, and did not cause them to think or reflect on their experiences.

Summarizing Feedback: Five participants (P6 and P11-P14) found it helpful to see the past reflections summarized and were able to leverage it to see "connections." However, six participants (P1-P3, P5, P8, and P17) provided negative opinions towards this type of suggestion and believed it was too long, not related to the current reflection, lacking clear suggestions to change the writing, or explaining something the users already knew.

4.2. RQ2: User Perception

To find early indications for answering RQ2, we both A) investigated the provided answers to the post-test perception constructs and B) explored the responses to the open question asking for feedback.

Regarding the perception constructs, we received values higher than an average of 4.0 for the Likert scale for each of the post-test constructs, i.e., excitement after interaction (M = 5.18, SD = 0.64), perceived usefulness (M = 5.35, SD = 0.61), perceived ease of use (M = 5.25, SD = 0.54), technology acceptance (M = 5.18, SD = 0.92), correctness of the suggestions (M = 4.88, SD = 0.78), perceived improvement in writing (M = 5.12, SD = 0.76), and perceived improvement in writing in the long run (M = 5.12, SD = 0.58), suggesting indications of positive perception towards `Memoire` in our pre-study.

Moreover, the responses to open questions show a highly positive perception towards `Memoire`, with users indicating that: they enjoyed using `Memoire` (2 participants), it was easy to use and understand (3 participants), the suggestions and the tool were relevant, helpful, and accurate (7 participants), and the tool provided "personalized" suggestions (2 participants). Users also provided suggestions for improvement, which we plan to consider in future studies. The main suggestions included being able to see a precise indication of the area within a text that the suggestion refers to (1 participant), faster response time (1 participant), and updates to the font and the visual design being in use in the interface (1 participant).

5. Discussion and Conclusion

Reflective writing is often a challenging process for students, as they struggle to meaningfully connect their thoughts to past experiences and knowledge. In this paper, we designed and developed `Memoire`, an intelligent writing assistant for providing suggestions on the reflective writings of learners within the context of prior entries using a RAG pipeline. We then conducted a pilot pre-study Prolific experimental evaluation on 17 participants to gather early insights on the usefulness of our tool, and finally, provided a plan for our future main classroom study. The literature on NLP has identified challenges with systems using generative AI LLMs, including hallucination and lack of personalization, which are avenues in which RAG can offer a viable alternative. In this work, we provided the first indications of the applicability of RAG to the domain of intelligent reflective writing assistants. This approach allows users to receive suggestions not only with regards to their current writing, but also with the personalized context of their own past interactions with the tool, and thus enables them to link what they write to prior entries from their theory and practice sessions. The early results of our study, as well as the analysis of the open answers, show an early success of `Memoire` in mitigating writer's block in the domain of reflection and helping learners to write reflective texts.

With that said, our study comes with a set of limitations. We only conducted our pre-study on a small sample of 17 people, necessitating more testing. Future works can also expand upon the types of suggestions we provided, the LLMs we used as agents in the generator module in the RAG pipeline, and the interaction data collected from the participants.

In conclusion, `Memoire` uses the concept of RAG in the domain of reflective writing to help learners connect their current writings with their prior journal entries. Our study sheds light on the applicability of RAG for reflective writing support, sets the stage for a larger-scale classroom study of `Memoire`, and provides early insights for researchers in learning sciences and educational technologies on the perception of RAG-enabled writing assistants.

Acknowledgments

This project was substantially financed by the Swiss State Secretariat for Education, Research, and Innovation (SERI).

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT and Apple Intelligence for grammar and spelling checks, as well as to improve writing clarity. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] K. Williams, M. Woolliams, J. Spiro, Reflective writing, Bloomsbury Publishing, 2020.
[2] V. D. O'Loughlin, L. M. Griffith, Developing student metacognition through reflective writing in an upper level undergraduate anatomy course, Anatomical Sciences Education 13 (2020) 680–693. Publisher: Wiley Online Library.
[3] J. Colomer, T. Serra, D. Cañabate, R. Bubnys, Reflective learning in higher education: Active methodologies for transformative practices, 2020. Issue: 9 Pages: 3827 Volume: 12.
[4] A. Cattaneo, E. Boldrini, Individual and collaborative writing-to-learn activities in vocational education: An overview of different instructional strategies, Writing for professional development (2016) 188–208. Publisher: Brill.
[5] I. S. Adeani, R. B. Febriani, S. Syafryadin, Using GIBBS'reflective cycle in making reflections of literary analysis, Indonesian EFL Journal 6 (2020) 139–148.
[6] J. Prior, S. Ferguson, J. Leaney, Reflection is hard: teaching and learning reflective practice in a software studio, in: Proceedings of the Australasian Computer Science Week Multiconference, 2016, pp. 1–8.
[7] H. Alrashidi, N. Almujally, M. Kadhum, T. Daniel Ullmann, M. Joy, Evaluating an automated analysis using machine learning and natural language processing approaches to classify computer science students' reflective writing, in: Pervasive computing and social networking: Proceedings of ICPCSN 2022, Springer, 2022, pp. 463–477.

[8] B. A. Schwendimann, A. A. Cattaneo, J. Dehler Zufferey, J.-L. Gurtner, M. Bétrancourt, P. Dillenbourg, The 'Erfahrraum': A pedagogical model for designing educational technologies in dual vocational systems, Journal of Vocational Education & Training 67 (2015) 367–396. Publisher: Taylor & Francis.

[9] S. O. Bada, S. Olusegun, Constructivism learning theory: A paradigm for teaching and learning, Journal of Research & Method in Education 5 (2015) 66–70.

[10] L. E. McGuire, K. A. Lay, Reflective pedagogy for social work education: Integrating classroom and field for competency-based education, Journal of Social Work Education 56 (2020) 519–532. Publisher: Taylor & Francis.

[11] M. Lee, K. I. Gero, J. J. Y. Chung, S. B. Shum, V. Raheja, H. Shen, S. Venugopalan, T. Wambsganss, D. Zhou, E. A. Alghamdi, others, A design space for intelligent and interactive writing assistants, in: Proceedings of the CHI conference on human factors in computing systems, 2024, pp. 1–35.

[12] M. Rose, M. A. Rose, Writer's block: The cognitive dimension, SIU Press, 2009.

[13] A. Göldi, T. Wambsganss, S. P. Neshaei, R. Rietsche, Intelligent support engages writers through relevant cognitive processes, in: Proceedings of the CHI conference on human factors in computing systems, 2024, pp. 1–12.

[14] X. Su, T. Wambsganss, R. Rietsche, S. P. Neshaei, T. Käser, Reviewriter: AI-Generated instructions for peer review writing, in: Proceedings of the 18th workshop on innovative use of NLP for building educational applications (BEA 2023), 2023, pp. 57–71.

[15] F. Weber, T. Wambsganss, S. P. Neshaei, M. Soellner, LegalWriter: An intelligent writing support system for structured and persuasive legal case writing for novice law students, in: Proceedings of the CHI conference on human factors in computing systems, 2024, pp. 1–23.

[16] T. Liu, Y. Zhang, C. Brockett, Y. Mao, Z. Sui, W. Chen, B. Dolan, A token-level reference-free hallucination detection benchmark for free-form text generation, in: S. Muresan, P. Nakov, A. Villavicencio (Eds.), Proceedings of the 60th annual meeting of the association for computational linguistics (volume 1: Long papers), Association for Computational Linguistics, Dublin, Ireland, 2022, pp. 6723–6737. URL: https://aclanthology.org/2022.acl-long.464/. doi:10.18653/v1/2022.acl-long.464.

[17] D. Muhlgay, O. Ram, I. Magar, Y. Levine, N. Ratner, Y. Belinkov, O. Abend, K. Leyton-Brown, A. Shashua, Y. Shoham, Generating benchmarks for factuality evaluation of language models, in: Y. Graham, M. Purver (Eds.), Proceedings of the 18th conference of the european chapter of the association for computational linguistics (volume 1: Long papers), Association for Computational Linguistics, St. Julian's, Malta, 2024, pp. 49–66. URL: https://aclanthology.org/2024.eacl-long.4/.

[18] A. Alam, The secret sauce of student success: Cracking the code by navigating the path to personalized learning with educational data mining, in: 2023 2nd international conference on smart technologies and systems for next generation computing (ICSTSN), IEEE, 2023, pp. 1–8.

[19] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, others, Retrieval-augmented generation for knowledge-intensive NLP tasks, Advances in Neural Information Processing Systems 33 (2020) 9459–9474.

[20] W. Fan, Y. Ding, L. Ning, S. Wang, H. Li, D. Yin, T.-S. Chua, Q. Li, A survey on RAG meeting LLMs: Towards retrieval-augmented large language models, in: Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining, 2024, pp. 6491–6501.

[21] M. Galarnyk, R. Routu, K. Bheda, P. Mehta, A. Shah, S. Chava, ACL ready: RAG based assistant for the ACL checklist, arXiv preprint arXiv:2408.04675 (2024).

[22] D. Thüs, S. Malone, R. Brünken, Exploring generative AI in higher education: a RAG system to enhance student engagement with scientific literature, Frontiers in Psychology 15 (2024) 1474892. Publisher: Frontiers Media SA.

[23] K. Shridhar, J. Macina, M. El-Assady, T. Sinha, M. Kapur, M. Sachan, Automatic generation of socratic subquestions for teaching math word problems, arXiv preprint arXiv:2211.12835 (2022).

[24] K. Martha, Improving critical thinking in written assignments: Human vs. ChatGPT tutor in socratic questioning intervention (2023).

[25] L. Favero, J. A. Pérez-Ortiz, T. Käser, N. Oliver, Enhancing critical thinking in education by means

of a socratic chatbot, arXiv preprint arXiv:2409.05511 (2024).
[26] S. P. Neshaei, R. Rietsche, X. Su, T. Wambsganss, Enhancing peer review with AI-powered suggestion generation assistance: Investigating the design dynamics, in: Proceedings of the 29th international conference on intelligent user interfaces, 2024, pp. 88–102.
[27] H. Dang, K. Benharrak, F. Lehmann, D. Buschek, Beyond text generation: Supporting writers with continuous automatic text summaries, in: Proceedings of the 35th annual ACM symposium on user interface software and technology, 2022, pp. 1–13.
[28] A. Wischgoll, Improving undergraduates' and postgraduates' academic writing skills with strategy training and feedback, in: Frontiers in education, volume 2, Frontiers Media SA, 2017, p. 33.
[29] S. Thapa, S. Adhikari, GPT-4o and multimodal large language models as companions for mental wellbeing, Asian Journal of Psychiatry 99 (2024) 104157. Publisher: Elsevier.
[30] T. Wambsganss, A. Janson, J. M. Leimeister, Enhancing argumentative writing with automated feedback and social comparison nudging, Computers & Education 191 (2022) 104644. Publisher: Elsevier.
[31] S. J. Ashford, Feedback-seeking in individual adaptation: A resource perspective, Academy of Management journal 29 (1986) 465–487. Publisher: Academy of Management Briarcliff Manor, NY 10510.
[32] R. Agarwal, E. Karahanna, Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage, MIS quarterly (2000) 665–694. Publisher: JSTOR.