June 2024

Worked Examples to Facilitate the Development of Prompt Engineering Skills

Antonia Tolzin
University of Kassel, antonia.tolzin@uni-kassel.de

Nils Knoth
University of Kassel, nils.knoth@uni-kassel.de

Andreas Janson
University of St. Gallen, andreas.janson@unisg.ch

WORKED EXAMPLES TO FACILITATE THE DEVELOPMENT OF PROMPT ENGINEERING SKILLS

Short Paper

Antonia Tolzin, University of Kassel, Kassel, Germany, antonia.tolzin@uni-kassel.de

Nils Knoth, University of Kassel, Kassel, Germany, nils.knoth@uni-kassel.de

Andreas Janson, University of St. Gallen, St. Gallen, Switzerland, andreas.janson@unisg.ch

Abstract

This paper explores the evolving field of prompt engineering in Artificial Intelligence (AI), with a focus on Large Language Models (LLMs). As LLMs exhibit remarkable potential in various educational domains, their effective use requires adept prompt engineering skills. We introduce a skill-based approach to prompt engineering and explicitly investigate the impact of using worked examples to facilitate prompt engineering skills among students interacting with LLMs. We propose hypotheses linking prompt engineering, worked examples, and perceived anthropomorphism to the quality of LLM output. Our initial findings support the critical relationship between proficient prompt engineering and the resulting output quality of LLMs. Subsequent phases will further explore the role of worked examples in prompt engineering, aiming to provide practical recommendations for educational improvement and industry application. Additionally, this research aims to shed light on the responsible utilization of LLMs in education and contribute insights to educational practice, research, and organizational development.

Keywords: Large Language Model, Prompt Engineering, Worked Examples, AI Interaction, Education.

1 Introduction

Over the years, Artificial Intelligence (AI) has undergone rapid development across multiple disciplines, demonstrating advances in image identification, speech recognition, language modeling, and more (Berg *et al.*, 2023). Large Language Models (LLMs) use iterative predictions of subsequent words based on previous sequences, facilitating the construction of human-like language (Bommasani *et al.*, 2021; McCoy *et al.*, 2023). These LLMs and conversational interfaces are also increasingly being used in human communication, providing smooth multi-turn dialogues without the need for extensive data or programming knowledge, thereby enhancing the conversational user experience (Dwivedi *et al.*, 2023). The exceptional ability of LLMs to produce high quality output has generated excitement in higher education. This transformational potential extends to essay and assignment writing, surpassing human performance in areas such as law, and revolutionizing the bespoke learning experience (Dwivedi *et al.*, 2023; Choi *et al.*, 2023; Cao *et al.*, 2023).

ChatGPT with more than 100 million weekly active users (Shewale, 2023) signals the widespread accessibility of generative AI. Recent research highlights the potential of ChatGPT to enhance learning through adaptive, personalized, and self-directed methods (Rahman and Watanobe, 2023; Rasul *et al.*, 2023; Zhu *et al.*, 2023). This is where prompt engineering plays an important role: Prompt engineering, also known as prompt design, prompt programming, or prompting, involves the creation of input instructions for generative AI models (Oppenlaender et al., 2023). This emerging field focuses on the creation, refinement, and implementation of instructions that guide the output of LLMs in various tasks. Due to the industry demand for competent users of AI tools, it is imperative to train students in prompt engineering. Dell'Acqua et al. (2023) highlighted the positive impact on consultants' work outcomes, signaling wider industry adoption of this technology, necessitating the preparation of students in LLM prompting. As users continue to face challenges in managing LLM-generated content, educators need to be adept at teaching effective LLM interactions. However, Dang et al. (2022) point out that prompt engineering is primarily an iterative process, often characterized by trial and error, which makes it difficult to create successful prompts and promote robust interactions. Nevertheless, the ability to engineer effective prompts is increasingly critical for communicating with LLMs (White *et al.*, 2023). Despite the interest in LLMs, there is still limited understanding of how individuals without formal AI training (non-experts), engineer prompts and their ability to do so (Zamfirescu-Pereira *et al.*, 2023).

This research in progress aims to investigate non-experts' prompt engineering for LLMs and its impact on LLM performance in higher education. Improving prompt engineering skills, similar to any other essential skill, requires a better understanding of basic technology principles, practical exposure to technology-integrated systems, and continuous skill refinement through feedback iterations (Meskó, 2023). One potential strategy borrowed from educational psychology to enhance understanding and practical experience is to provide students with worked examples that illustrate prompt engineering (Wittwer and Renkl, 2010; Atkinson *et al.*, 2000; Sweller, 1988). These examples provide comprehensive, beginner-centered approaches to solutions, allowing for direct application or later use based on memorized schemata (Wittwer and Renkl, 2010). Research consistently demonstrates the superior effectiveness of worked example learning, particularly in the early stages of cognitive skill acquisition, outperforming results obtained using problem solving methods alone (Atkinson *et al.*, 2000). Furthermore, worked examples could enhance the perceived anthropomorphism of the LLM as an assistant with agentic qualities. This enhancement fosters increased communicative interaction and a sense of trustworthiness among students in their engagement with prompt engineering tasks, as suggested by Fink (2012). Perceiving the LLM as more anthropomorphic might trigger LLM agency, leading to task delegation, prompting purposeful strategies, and could also result in persona assignment to the LLM (Baird and Maruping, 2021; Puranam and Vanneste, 2021).

Currently, the predominant focus of prompt engineering research has been on a technological perspective (Ding *et al.*, 2021; Liu *et al.*, 2023). In this study, we aim to introduce a skills-oriented methodology to prompt engineering, recognizing its crucial role in empowering students for effective management of LLMs. The guiding research question (RQ) is as follows: *How do worked examples contribute to the acquisition of prompt engineering skills?*

2 Theoretical Background and Hypotheses Development

2.1 Prompt engineering

To prompt an LLM to generate or modify text output, specific input text or instructions must be formulated (White *et al.*, 2023). The quality of a prompt significantly influences interactions with an LLM, as it establishes clear guidelines and rules for the LLM's dialogue, using predetermined norms to inform the structure, importance of information, and intended output form and content (White *et al.*, 2023). However, LLMs also have significant challenges (Bommasani *et al.*, 2021) because they require expertise for using this technology (Dwivedi *et al.*, 2023; Zamfirescu-Pereira *et al.*, 2023).

Despite the importance of developing efficient prompts for LLMs like ChatGPT, empirical quantitative findings are limited concerning prompt engineering. However, Oppenlaender et al. (2023) explored the potential of prompt engineering in generating art with generative AI, revealing that proficiency requires practice, and excellent prompt writing demands knowledge of relevant terminologies. Zamfirescu-Pereira et al. (2023) examined prompt engineering by non-experts using an LLM-based chatbot design tool. Non-experts could generate prompt ideas but struggled to advance systematically due to limited awareness of LLM capabilities, tending to create prompts resembling human-to-human instructions. Dang et al. (2022) identified challenges in prompt creation through an human-computer-interaction (HCI) focus group, including a lack of clear direction, poor depiction of activities, concerns about computing costs, and ethical implications. Participants reported difficulties in formulating efficient prompts optimized for specific LLM tasks.

In contrast, our study focuses on how non-expert users write and utilize prompts with an LLM, offering insights into user practices beyond technology-centric approaches. Thus, we hypothesize the following:

H1: Students with higher prompt engineering skills will demonstrate a better LLM output quality.

2.2 Worked examples

Worked examples, also called example-based learning, constitute an extensively studied instructional method in educational psychology (Wittwer and Renkl, 2010). This pedagogical approach acts as a scaffold for learning by providing students with detailed task instructions, effectively reducing the cognitive load associated with ineffective problem-solving strategies (Vogel *et al.*, 2022; Janson *et al.*, 2020). Worked examples present a comprehensive and correct solution procedure in written form, typically designed to reflect the approach a novice student would take, allowing for direct utilization or later application based on a memorized schema (Wittwer and Renkl, 2010). Extensive research has consistently demonstrated the superior effectiveness of worked examples, particularly in the early phases of cognitive skill acquisition, surpassing the outcomes achieved through problem-solving alone—a phenomenon known as the worked example effect (Wittwer and Renkl, 2010; Paas and van Merriënboer, 1994; Atkinson *et al.*, 2000; Sweller, 1988).

The efficacy of studying worked examples for learning is explicable through Cognitive Load Theory (CLT) (Sweller, 1988). In unfamiliar knowledge domains, learners often employ domain-independent solution strategies, imposing high demands on working memory and hindering the construction of problem-solving schemata. Worked examples, by preventing learners from engaging in irrelevant search processes, enable focused attention on the presented problem, reducing cognitive load and facilitating the construction of problem-solving schemata with reference to underlying domain principles. Therefore, studying worked examples is instrumental in acquiring knowledge that can be flexibly applied to new problems, addressing cognitive overload and promoting effective learning (Atkinson *et al.*, 2000; Wittwer and Renkl, 2010; Sweller, 2020). Thus, we hypothesize the following:

H2: Students who are provided with a worked example of how to prompt will produce better LLM output quality than students who are not provided with a worked example.

H3: Students who are provided with a worked example of how to prompt will engage in more sophisticated prompt engineering behavior by using more purposeful prompting strategies than students who are not provided with a worked example.

2.3 Perceived anthropomorphism

According to Fink (2012), design elements that resemble human traits and social characteristics can increase the familiarity and acceptance of robots. Anthropomorphism, the attribution of human characteristics to non-human entities (Fink, 2012), helps to rationalize their behavior. Epley et al. (2007) present two main reasons for the human tendency to anthropomorphize objects: the need to understand the environment in order to minimize uncertainty, and the inherent tendency to establish social relationships with other entities. Fong et al. (2003) suggest that people want to engage with machines in a manner similar to human-to-human interactions. Additionally, cues such as facial expressions, gestures, and conversational tones substantially contribute to users perceiving machines as more human-like in their interactions (Feine *et al.*, 2019). Moreover, social features, joint action and embodiment are also mechanisms to enhance human-agent interaction (Tolzin and Janson, 2023; Tolzin *et al.*, 2023).

Worked examples serve as mental models for prompt engineering, which promotes analogical reasoning. This reasoning leads students to anthropomorphize the LLM, perceiving it as knowledgeable and agentic, consistent with the theoretical framework of delegation to agentic information systems artifacts (Baird and Maruping, 2021). Clarity in the worked examples facilitated understanding of prompt engineering, encouraging students to attribute human-like characteristics to the LLM, increasing its perceived communicativeness. Furthermore, well-designed examples aligned with students' cognitive processes increase perceptions of the LLM's agency (Puranam and Vanneste, 2021), while relatability and familiarity in these examples encourage students to delegate tasks, viewing the LLM as a trusted and skilled guide (Baird and Maruping, 2021). Therefore, we assume:

H4: Students who are provided with a worked example of how to prompt will perceive the LLM as more anthropomorphic than students who are not provided with a worked example.

Perceiving the LLM as more anthropomorphic promotes greater engagement and triggers a sense of agency in students (Baird and Maruping, 2021). This sense leads to delegation of tasks to the system, encouraging reliance on the LLM for sophisticated strategies. Moreover, anthropomorphism evokes cognitive empathy (Janson, 2023), prompting students to adopt purposeful prompting strategies. Empathy for the anthropomorphic qualities of the LLM leads to the attribution of higher agency to the system (Puranam and Vanneste, 2021), which encourages purposeful prompting strategies. In addition, students who perceive the LLM as more anthropomorphic are more likely to assign a persona to the LLM, which is a desirable prompt strategy. Perceiving the LLM as intelligent due to anthropomorphism leads students to engage in more sophisticated prompt engineering, expecting an adept response from the system based on its perceived intelligence (Baird and Maruping, 2021). Thus, we assume:

H5: The more students perceive the LLM as anthropomorphic, the more they will engage in sophisticated prompt engineering behavior by using more purposeful prompting strategies.

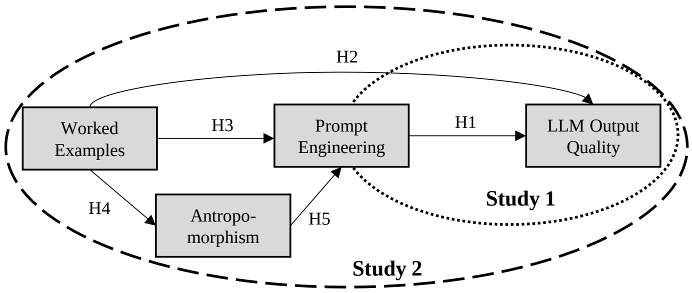

Figure 1. Conceptual model of worked examples for prompt engineering and LLM output quality.

3 Study Design

The conceptual model will be evaluated in two consecutive studies. The first study focuses the influence of prompt engineering on LLM output quality, while the second study examines the role of worked examples for prompt engineering and the quality of the LLM Output (Figure 1).

3.1 Study 1

To investigate how non-experts interact with LLMs and engage in prompt engineering, and to validate Hypothesis 1, we conducted an experiment in May 2023. The sample size included N = 45 university students, aged between 19 to 35 years, thereof n = 15 women, n = 28 men, and n = 2 non-specified. They studied different subjects (Mechanical Engineering: n = 15, Psychology: n = 6, Business and Economics: n = 21, and n = 3 non-specified). These participants were tasked with completing two tasks using a GDPR-compliant platform, using Open AI's application programming interface (API) along with the gpt-3.5-turbo model for conversational interactions.

To assess students' prompt engineering skills, two tasks were designed, each requiring a solution via a LLM: creating a comprehensive travel plan for Andorra (Task 1) and planning a scientific project on automated essay scoring (Task 2). Task 1 represented a generic prompt engineering scenario for leisure, while Task 2 contextualized the requirements of higher education. We collected behavioral indicators through written protocols focusing on student generated prompts and LLM outputs for each task. LLM output quality was assessed using an integrative complexity score (Suedfeld *et al.*, 1992), evaluating cognitive traits of differentiation and integration. Two coders independently scored the outputs on a 10-point scale, achieving high inter-rater reliability (IRR; Task 1: $r = .96$; $n = 42$; $p < .001$; Task 2: $r = .96$;

n = 42; p < .001). In addition, prompt quality was assessed on the basis of the six prompt components proposed by Eager and Brunton (2023), with each prompt component receiving one point, demonstrating substantial inter-rater reliability between the two coders (IRR, Task 1: r = .83; n = 42; p < .001;Task 2: r = .80; n = 42; p < .00). The components include (1) verb, (2) focus, (3) context, (4) focus and condition, (5) alignment, and (6) constraints and limitations. Additionally, students completed reflection protocols assessing perceived ease of writing prompts, task complexity, and overall user experience with the generative AI. To control for other variables potentially influencing prompt engineering behavior, assessments were conducted for trust in generative AI (Schmitt *et al.*, 2022) and personal innovativeness (Agarwal and Prasad, 1998).

3.2 Study 2

To test the effects of worked examples, we evaluate worked examples in an AI-based learning environment embedded in a between-subject experiment. This allowed us to investigate how worked examples influence prompt engineering behavior and perceived anthropomorphism of the LLM. This second experiment with 251 undergraduate business students was conducted in December 2023. The analysis is ongoing. Students were randomly assigned to one of two groups: the experimental group received a worked example of prompts. This worked example consisted of seven recommendations with a description and a good and a bad example of a prompt based on those recommendations. We created the worked examples based on existing typologies of prompting strategies such as for example Svendsen and Garvey (2023); Liu et al. (2023); and Eager and Brunton (2023). The seven recommendations are: (1) assign a role to the AI, (2) prime the AI and set the context, (3) set structural specifications, (4) limit the length of the AI's output, (5) give precise descriptions of the AI's procedure and result, (6) segment a task and generate sequences of prompts from it, (7) avoid ambiguous fillers and adjectives.

The control group was given a text about "How to make major events sustainable", which was the same length as the prompting guide. Both groups were given five minutes to read the text. Afterwards, both groups were asked to complete a task using the same platform as in study 1. We used tasks from Dell'Acqua et al. (2023) to assess the impact of AI on performance in realistic, complex, knowledge-intensive tasks designed with expert input to ensure ecological validity. Students were tasked with generating new beverage product ideas, assessing creativity, analytical, persuasive and writing skills. Behavioral indicators, including prompting, perceived anthropomorphism and LLM outputs, were collected via post-task surveys. Prompt quality was evaluated based on Eager and Brunton's (2023) six components, each scored one point, while perceived anthropomorphism was measured using Moussawi et al.'s (2023) six-item scale. LLM output quality was assessed using the integrative complexity score from Study 1. Additionally, students completed a survey on prompt writing ease, task complexity (Gupta and Bostrom, 2013), and user experience with the AI. Cognitive load (Krieglstein *et al.*, 2023), prior prompt engineering knowledge, and self-perception (Miron *et al.*, 2004) were recorded for statistical control.

4 Preliminary Findings

In study 1, students reported a predominantly positive interaction experience with the GPT-based platform, showcasing satisfaction in meeting their expectations, evaluating output quality, and overall user experience. Additionally, they expressed a willingness to utilize generative AI again for similar tasks and perceived a low level of difficulty in crafting prompts. Moreover, students exhibited a notable interest in using generative AI, reflecting a more favorable attitude toward its adoption, revealing potential implications for future AI educational endeavors, as interest and attitudes are influential factors for learning success (Eccles and Wigfield, 2002). No anomalies were detected regarding trust in generative AI and personal innovativeness, as possible outliers were examined using box plots. As no specific hypotheses were formulated concerning these constructs, further calculations were not undertaken. To validate hypothesis 1, linear regression analyses were conducted, using the rated quality of prompt engineering as the predictor for the rated quality of the generated output (criterion). For the travel plan task (Task 1), the model revealed a significant beta coefficient for the quality of the prompt engineering towards the quality of the travel plan output (β = 1.49, $t(40) = 6.78$, $p < .001$), accounting for 53% of the variance in output quality ($R^2 = .535$, $F(1, 40) = 46.01$, $p < .001$) (Figure 2). Similarly, for the scientific project planning task (Task 2), a significant beta coefficient was observed (β = 1.376, $t(37) = 11.502$, $p < .001$), explaining approximately 78% of the variance in the output quality ($R^2 = .782$, $F(1, 37) = 132.3$, $p < .001$). The consistent effect found across tasks confirms that higher-quality prompt engineering behavior correlates with improved LLM output, indicating a significant association between prompt engineering skills and LLM output quality. Therefore, H1 is supported.

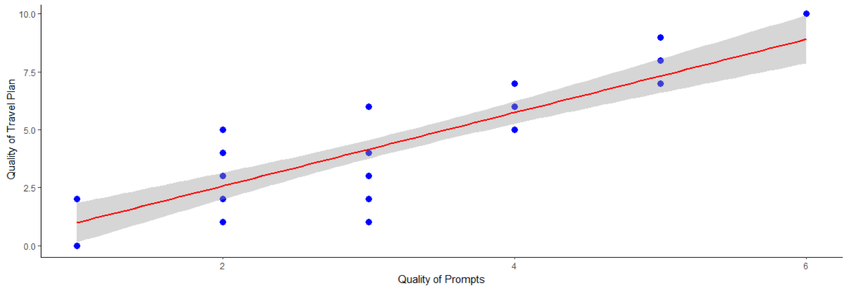

Figure 2. Impact of quality of prompt engineering on LLM output (solution of task 1, travel plan).

5 Concluding Implications

This research in progress paper investigates the emerging skill of prompt engineering and provides a first glimpse into the role of worked examples for the development of prompt engineering skills. First, in our initial study, we found evidence that proficient prompt engineering reliably predicts the quality of LLM output. Our empirical findings indicate that effective prompt engineering significantly improves the generation of better LLM output, allowing users with advanced prompt skills to fully exploit the enormous potential of this technology. In our second study, we will explore the role of worked examples for prompt engineering and the quality of LLM output in more detail and provide practical recommendations for students to improve their interactions with LLMs (hypotheses 2-5). We predict that worked examples will improve prompt engineering skills, perceived anthropomorphism, and LLM output quality. In doing so, we contribute to educational practice by demonstrating how students' prompt engineering skills can be trained using worked examples. This also points to future research investigating the trainability of this particular skill. The results will help educational institutions, such as universities, to move effectively towards skills-based education. There is also a need to explore the implications for the responsible use of LLM in education, particularly in the creation of effective instructional prompts. In addition, with the completed research, we will contribute to research by (1) conceptualizing prompt engineering as an emerging skill from a human-centered perspective, and (2) providing empirical evidence on how worked examples affect students' ability to engage in prompt engineering behaviors, and (3) the role of anthropomorphism in this context. Finally, we will contribute to organizational practice as these findings will help companies to develop learning tools to upskill employees for interdisciplinary roles. In doing so, we will provide a deeper understanding of prompt engineering and open avenues for future research. Future research should extend this study by developing a more comprehensive prompt taxonomy, utilizing larger sample sizes, and including tasks that require more prompt engineering. This approach will provide more rigorous and nuanced insights into the effects of worked examples and anthropomorphism on prompt engineering skills education.

Acknowledgements

The results presented were partially developed in the research projects: Komp-HI funded by the German Federal Ministry of Education and Research (BMBF, grant 16DHBKI073) and Managing the Algorithm: Prompt Engineering for AI-based Systems as an Emerging Business Skill by the Swiss National Science Foundation (SNSF, grant number: 221281). We thank the BMBF and SNSF for supporting our research.

References

Agarwal, R. and Prasad, J. (1998). "A Conceptual and Operational Definition of Personal Innovativeness in the Domain of Information Technology", *ISR*, 9 (2), pp. 204–215.

Atkinson, R.K., Derry, S.J., Renkl, A. and Wortham, D. (2000). "Learning from Examples: Instructional Principles from the Worked Examples Research", Rev. Educ. Res. , 70 (2), pp. 181–214.

Baird, A. and Maruping, L.M. (2021). "The Next Generation of Research on IS Use: A Theoretical Framework of Delegation to and from Agentic IS Artifacts", *MIS Quarterly*, 45 (1), pp. 315–341.

Berg, J., Raj, M. and Seamans, R. (2023). "Capturing Value from Artificial Intelligence", *AMD* .

Bommasani, R., Hudson, D.A., Adeli, E., Altman, R., Arora, S., Arx, S. et al. (2021), *On the Opportunities and Risks of Foundation Models*.

Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., Yu, P.S. and Sun, L. (2023), *A Comprehensive Survey of AI-Generated Content (AIGC): A History of Generative AI from GAN to ChatGPT*.

Choi, J.H., Hickman, K.E., Monahan, A. and Schwarcz, D.B. (2023). "ChatGPT Goes to Law School", *SSRN Electronic Journal* .

Dang, H., Benharrak, K., Lehmann, F. and Buschek, D. (2022). "Beyond Text Generation: Supporting Writers with Continuous Automatic Text Summaries", in Agrawala, M., Wobbrock, J.O., Adar, E. and Setlur, V. (Eds.), *ACM Symposium* , ACM, NY, USA.

Dell'Acqua, F., McFowland, E., Mollick, E.R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F. and Lakhani, K.R. (2023). "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality", *SSRN Electronic Journal* .

Ding, N., Hu, S., Zhao, W., Chen, Y., Liu, Z., Zheng, H.-T. and Sun, M. (2021), *OpenPrompt: An Open-source Framework for Prompt-learning*.

Dwivedi, Y.K., Kshetri, N., Hughes, L., Slade, E.L., Jeyaraj, A., Kar, A.K., et al. (2023). "So what if ChatGPT wrote it? Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy", *IJIM*, 71 .

Eager, B. and Brunton, R. (2023). "Prompting Higher Education Towards AI-Augmented Teaching and Learning Practice", *Journal of University Teaching and Learning Practice*, 20 (5).

Eccles, J.S. and Wigfield, A. (2002). "Motivational beliefs, values, and goals", Annu. Rev. Psychol., 53 pp. 109–132.

Epley, N., Waytz, A. and Cacioppo, J.T. (2007). "On seeing human: a three-factor theory of anthropomorphism", *Psychological review*, 114 (4), pp. 864–886.

Feine, J., Gnewuch, U., Morana, S. and Maedche, A. (2019). "A Taxonomy of Social Cues for Conversational Agents", *International Journal of Human-Computer Studies*, 132 pp. 138–161.

Fink, J. (2012). "Anthropomorphism and Human Likeness in the Design of Robots and Human-Robot Interaction", in Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J.M., Mattern, F., Mitchell, J.C., et al. (Eds.), *Social Robotics: 4th International Conference, ICSR 2012*, Vol. 7621, pp. 199–208.

Fong, T., Nourbakhsh, I. and Dautenhahn, K. (2003). "A survey of socially interactive robots", *Robotics and Autonomous Systems*, 42 (3-4), pp. 143–166.

Gupta, S. and Bostrom, R. (2013). "Research Note —An Investigation of the Appropriation of Technology-Mediated Training Methods Incorporating Enactive and Collaborative Learning", *Information Systems Research*, 24 (2), pp. 454–469.

Janson, A. (2023). "How to leverage anthropomorphism for chatbot service interfaces: The interplay of communication style and personification", *Computers in Human Behavior*, 149 p. 107954.

Janson, A., Söllner, M. and Leimeister, J.M. (2020). "Ladders for Learning: Is Scaffolding the Key to Teaching Problem-Solving in Technology-Mediated Learning Contexts?", *Academy of Management Learning & Education*, 19 (4), pp. 439–468.

Krieglstein, F., Beege, M., Rey, G.D., Sanchez-Stockhammer, C. and Schneider, S. (2023). "Development and Validation of a Theory-Based Questionnaire to Measure Different Types of Cognitive Load", *Educational Psychology Review*, 35 (1).

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H. and Neubig, G. (2023). "Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing", *ACM.*, 55 (9), 1–35.

McCoy, R.T., Yao, S., Friedman, D., Hardy, M. and Griffiths, T.L. (2023), *Embers of Autoregression: Understanding Large Language Models Through the Problem They are Trained to Solve*.

Meskó, B. (2023). "Prompt Engineering as an Important Emerging Skill for Medical Professionals: Tutorial", *Journal of medical Internet research*, 25 e50638.

Miron, E., Erez, M. and Naveh, E. (2004). "Do personal characteristics and cultural values that promote innovation, quality, and efficiency compete or complement each other?", *Journal of Organizational Behavior*, 25 (2), pp. 175–199.

Moussawi, S., Koufaris, M. and Benbunan-Fich, R. (2023). "The role of user perceptions of intelligence, anthropomorphism, and self-extension on continuance of use of personal intelligent agents", *European Journal of Information Systems*, 32 (3), pp. 601–622.

Oppenlaender, J., Linder, R. and Silvennoinen, J. (2023), *Prompting AI Art: An Investigation into the Creative Skill of Prompt Engineering*.

Paas, F.G.W.C. and van Merriënboer, J.J.G. (1994). "Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach", *Journal of Educational Psychology*, 86 (1), pp. 122–133.

Puranam, P. and Vanneste, B. (2021). "Artificial Intelligence, Trust, and Perceptions of Agency", *SSRN Electronic Journal* .

Rahman, M.M. and Watanobe, Y. (2023). "ChatGPT for Education and Research: Opportunities, Threats, and Strategies", *Applied Sciences*, 13 (9), p. 5783.

Rasul, T., Nair, S., Kalendra, D., Robin, M., Oliveira Santini, F. de, Ladeira, W.J., Sun, M., Day, I., Rather, R.A. and Heathcote, L. (2023). "The role of ChatGPT in higher education: Benefits, challenges, and future research directions", *Journal of Applied Learning and Teaching (*6(1)).

Schmitt, A., Wambsganss, T. and Janson, A. (2022). "Designing for Conversational System Trustworthiness: The Impact of Model Transparency on Trust and Task Performance", *ECIS 2022*.

Shewale, R. (2023). "ChatGPT Statistics: Detailed Insights On Users (2023)", available at: https://www.demandsage.com/chatgpt-statistics/.

Suedfeld, P., Tetlock, P.E. and Streufert, S. (1992). "Conceptual/integrative complexity", in Smith, C.P. (Ed.), *Motivation and personality: Handbook of thematic content*, Cambridge University Press, NewYork, pp. 393–400.

Svendsen, A. and Garvey, B. (2023). "Prompt-engineering testing ChatGPT4 and Bard for assessing Generative-AI efficacy to support decision-making", *SSRN Electronic Journal* .

Sweller, J. (1988). "Cognitive Load During Problem Solving: Effects on Learning", *Cognitive Science*, 12 (2), pp. 257–285.

Sweller, J. (2020). "Cognitive load theory and educational technology", *ETR& D*, 68 (1), pp. 1–16.

Tolzin, A. and Janson, A. (2023). "Mechanisms of Common Ground in Human-Agent Interaction: A Systematic Review of Conversational Agent Research", *HICSS 2023*.

Tolzin, A., Körner, A., Dickhaut, E., Janson, A., Rummer, R. and Leimeister, J.M. (2023). "Designing Pedagogical Conversational Agents for Achieving Common Ground", in Gerber, A. and Baskerville, R. (Eds.), *Design Science Research for a New Society: Society 5.0, Lecture Notes in Computer Science*, Vol. 13873, Springer Nature Switzerland, Cham, pp. 345–359.

Vogel, F., Kollar, I., Fischer, F., Reiss, K. and Ufer, S. (2022). "Adaptable scaffolding of mathematical argumentation skills: The role of self-regulation when scaffolded with CSCL scripts and heuristic worked examples", *ijCSCL*, 17 (1), pp. 39–64.

White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J. and Schmidt, D.C. (2023), *A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT*.

Wittwer, J. and Renkl, A. (2010). "How Effective are Instructional Explanations in Example-Based Learning? A Meta-Analytic Review", *Educational Psychology Review*, 22 (4), pp. 393–409.

Zamfirescu-Pereira, J.D., Wong, R.Y., Hartmann, B. and Yang, Q. (2023). "Why Johnny Can't Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts", in *Proceedings CHI 2023*, pp. 1–21.

Zhu, C., Sun, M., Luo, J., Li, T. and Wang, M. (2023). "How to harness the potential of ChatGPT in education?", *Knowledge Management & E-Learning: An International Journal (*15(2)), pp. 133–152.