Fantastic AI text generations and where to trust them: It’s not magic, it’s science!

Lisa Straub
Julius-Maximilians-Universität Würzburg, lisa.straub@uni-wuerzburg.de

Myriam Schaschek
Julius-Maximilians-Universität Würzburg, myriam.schaschek@uni-wuerzburg.de

Christoph Tomitza
Julius-Maximilians-Universität Würzburg, christoph.tomitza@uni-wuerzburg.de

Axel Winkelmann
Julius-Maximilians-Universität Würzburg, axel.winkelmann@uni-wuerzburg.de

Fantastic AI Text Generations and Where to Trust Them: It's not Magic, it's Science!

Research in Progress Paper

Lisa Straub, Myriam Schaschek, Christoph Tomitza, and Axel Winkelmann

Julius-Maximilians-Universität Würzburg, Würzburg, Germany
{lisa.straub, myriam.schaschek, christoph.tomitza, axel.winkelmann}@uni-wuerzburg.de

Abstract. Enter the web application, type in a question, and get a human-like answer in no time. Especially with the advent of ChatGPT, text-generating artificial intelligence permeates daily life. As a result, end-users are trying out new applications bearing risks, such as overconfidence. This research-in-progress paper investigates the main factors affecting end-user perception regarding human-like AI-generated output and corresponding trust. With the overarching goal of appropriate protection by creating a standardized information structure for integration into websites as our artifact, we conduct a structured literature review in the first step to determine what causes overconfidence and the issues that need to be addressed by an appropriate solution. Therefore, we contribute to the broader aim of preventing end users from misinterpreting AI output. Our findings highlight AI literacy, difficulties in detecting misinformation, and a lack of transparency and explainability as critical factors to consider during solution development.

Keywords: overtrust, generative artificial intelligence, misinformation, transparency, AI

1 Introduction

AI sometimes feels like magic for inexperienced end-users, but it is not. Regarding magic, enthusiasm for the results causes people to disregard the obvious and sometimes fail to scrutinize it. Does the same apply to AI? Outputs generated with genAI are both fascinating and frightening for inexperienced users of AI applications (Bankins et al. 2023). Just a few weeks after ChatGPT was made available to the public, millions of interested end-users not only tested out the tool but also experimented with it and used the application for a wide variety of use cases (Peres et al. 2023, Teubner et al. 2023). This was due to the development of text-generating artificial intelligence (text-genAI), labelled as a disruptive technological advance with a pioneering large language model (LLM), currently the subject of intense debates (Peres et al. 2023, Ray 2023).

Only sometimes is the AI output received by end users trustworthy or correct, and it can also happen that answers are given that do not comply with legal requirements (Bankins et al. 2023, Tomitza et al. 2023, Schaschek & Engel 2023). End users must know that the chatbot's response is based on probability calculations and misjudges the confidence of the answer given (Ray 2023). A fact that occasionally leads to false confidence in AI outputs, which can have serious consequences (Lockey et al. 2021, Ray 2023, Banh & Strobel 2023, Schaschek & Engel 2023). Blind trust is therefore inappropriate (Dwivedi et al. 2023), and the question arises how this can be avoided.

The topic of trust in AI has been represented in the literature for many years and in various disciplines (Glikson & Woolley 2020). However, there are still apparent research gaps, particularly in the area of genAI and how end-users perceive and interpret the generated results. The adjacent topic of overtrust and how to prevent it is also under-researched (Tomitza et al. 2023). In academia and practice, especially with text-genAI, researchers and practitioners must explore new ways to ensure that end-users do not perceive simply false or unverifiable text phrases as trustworthy, ethical, and accurate sources (Bankins et al. 2023, Tomitza et al. 2023). This is also confirmed by the current AI regulations being discussed and considered by governments (Banh & Strobel 2023).

To provide short-term assistance for end users to correctly assess content and question the magic of applications such as ChatGPT, we develop a standardized information structure for integration into websites (inspired by an imprint structure or trust label) to ensure that the most important information for end users is comprehensible and available at all times. We value an easily understandable presentation for all user groups and an equally simple option for website operators with text-genAI integration. We do not see the solution we aim for as a final solution to protect users from false content permanently but as an early approach that can be implemented quickly. It is an initial security implementation whose structure and elements can also provide and establish essential building blocks in the field of explainable AI (Schaschek et al. 2023).

Therefore, in this research-in-progress paper, we answer the following questions with a structured literature review before expanding our research:

RQ: *What are the key factors leading to excessive confidence, overtrust, and inappropriate reliance on the genAI outcomes from the end-user?*

The paper is composed as follows: We describe the results of our structured literature search in section 2, followed by an outlook on our future research project 3. Finally, we discuss research opportunities and the limitations in section 4.

2 Elaborated Key Factors for Inappropriate Confidentiality

In order to work out the decisive factors that influence users' trust in AI and to be able to take these into account in a future solution, we carried out a structured literature research according to vom Brocke et al. (2009). Our search string encompassed the keywords: *(trust OR ethical OR explainable OR responsible OR human-centered) AND ("generative AI" OR "generative artificial intelligence").* We selected AIS eLibrary, IEEE, Science Direct, and Web of Science as the relevant databases. Due to the subject's topicality and the manageability of the search results, we defined no further inclusion and exclusion criteria. From the *n=264* articles identified, *n=17* relevant articles were identified through duplicate elimination and a comprehensive search process, including title *(n=63)*, abstract *(n=26)* and full text *(n=17)* analysis. A forward and backward search ultimately expands the relevant contributions to *n=47*. All authors read the literature to eliminate bias and performed independent coding related to the research questions.

We organized influencing factors listed in the literature regarding genAI overtrust into three main categories: (1) lack of AI expertise, (2) difficulty identifying misinformation, and (3) lack of transparency and explainability.

(1) Lack of AI Literacy: AI literacy refers to the ability not only to use AI but also to understand it, to question the output and technology critically, and to use it in a goal-oriented manner in private and business life (Druga & Ko 2021, Long & Magerko 2020, Ng et al. 2021). Various literature pay particular attention to children's AI literacy (Ali et al. 2021, Lee et al. 2021, Su & Ng 2023). Furthermore, the literature emphasizes the significance that everyone builds up AI skills, whether children or adults, to keep up with technological progress (Ng et al. 2021, Yang 2022). In this context, AI literacy is even compared to classical writing and calculating ability that everyone should have (Ng et al. 2021). As Laupichler et al. (2022) aptly describes: "Since artificial intelligence (AI) is finding its way into more and more areas of everyday life, improving the AI skills of non-experts is important and will become even more relevant in the future." In this regard, numerous articles show how AI literacy can be increased in the population, especially concerning children (Ali et al. 2021, Lee et al. 2021, Su & Ng 2023).

However, experts also highlight that the rapid advancement of AI makes it challenging to provide comprehensive education on the topic within schools, including both students and teachers. AI education is progressing too slowly compared to technological progress, and it will not be possible to create an equally high level of knowledge in all segments of the population (e.g., due to the origin, intellectual capacity, educational opportunities) (Lee et al. 2021, Su & Ng 2023). Additionally, there needs to be more basic knowledge about AI within the general population (Lockey et al. 2021, Su & Ng 2023). Although the prerequisite for the responsible use of AI technology is understanding how the technology works (Druga & Ko 2021), there needs to be a greater understanding of how AI operates (Gaube et al. 2021, Lee et al. 2021). Even more pernicious is that it is difficult to detect possible biases or mistakes (Kong et al. 2021, Ray 2023).

Since the potential applications go far beyond research and science, and applications such as ChatGPT have made them more accessible and more commonplace, the topic must not be only addressed in higher education (Bankins et al. 2023, Su & Ng 2023). Therefore, it is also necessary to analyze protection options that illuminate the problem of transparency from another side, for example, by preventing users from being shown sources that are not true or made up (Tomitza et al. 2023). Consequently, a lack of AI literacy might be an antecedent of overconfidence in generative AI outputs (Nourani et al. 2022, Sieck & Arkes 2005), which could cause misinformation (Gaube et al. 2021).

(2) Difficulty to detect misinformation: Today, news information spreads swiftly on social media, and psychological biases, alongside misleading information, rapidly induce misinformation. Particularly with digital sources, and thus also with text-genAI (is even itself a creator of digital context), verification of the information for the end-user is very laborious or not possible at all (Ali et al. 2021, Fielding 2019). Previous approaches to checking web page content are often based on visual elements and superficial features (Griesbaum 2022). Mainly about text-genAI, these common checking elements are not helpful for people to the same extent, and well-known methods, such as the CRAAP test, for analyzing suitable sources can only be used to a limited extent to protect the end-user from incorrect information (Ali et al. 2021, Fielding 2019, Griesbaum 2022).

Enriching media with LLM-generated text further amplifies those biases (Borji 2023, Datta et al. 2021, Deng & Lin 2022, Van Dis et al. 2023). Concerns regarding the future are potential harm induced by the misinformation of end-users generated by text-genAI. This is particularly important as the role of text-genAI shifts from mere problem solver to problem finder, and thus, it acts as a publicly available content creator (Ali et al. 2021, Seeber et al. 2020). Something striking about artificially created content is that it can be based on AI hallucinations and might be grammatically correct but contain misleading content (Alkaissi & McFarlane 2023, Ziwei et al. 2023, Banh & Strobel 2023). In light of AI hallucinations, the issue of misinformation is gaining momentum, prompting researchers and the media to warn about the output generated by text-genAI (Dwivedi et al. 2023, Gao et al. 2022, Qu et al. 2020, Ray 2023). Accordingly, numerous articles point out the limitations of LLM-generated texts in correspondence with the potential harm of misinformation (e.g. Deng & Lin (2022), Teubner et al. (2023)). Limitations include the quality of input data to train the LLM (Harrer 2023, Lim et al. 2023).

In the face of potential harm, literature investigates (text)-genAI risk factors to derive mitigation pathways (Harrer 2023, Weidinger et al. 2022) and design principles (Weisz et al. 2023). For example, it is suggested that models provide an empty answer if they could cause harm by giving the wrong answer (e.g. in the area of questions on types of suicide) (Weidinger et al. 2022)., potential approaches that should be taken into account in prevention solutions. Further, literature demands interdisciplinary innovations in the realms of technology, policy, and practice to prevent intentional misinformation by text-genAI and provides first solution approaches, such as content provenance methods (Horvitz 2022) or watermarking tools (Mackenzie 2023). It is not easy to accurately distinguish between text generated by genAI technology and text written by humans (Gao et al. 2022, Rikab 2023, Banh & Strobel 2023). To effectively combat misinformation, it is essential to use interdisciplinary and socio-technical approaches.

(3) Paucity of transparency and explainability: Transparency and explainability are practical tools for users with limited AI literacy to understand output better and identify misinformation. The issue when using text-genAI is that the factors mentioned earlier are not accessible to end-users (Dwivedi et al. 2023, Herm et al. 2023). The nature of AI applications has been criticized due to their "black box" characteristics (Herm et al. 2023, Shin 2021, Banh & Strobel 2023, Schaschek et al. 2023). Accordingly, it is difficult for AI novices (Mohseni et al. 2021) to comprehend the mechanisms and decision-making or interpret its outputs. Despite several studies showing that the understanding and evaluation of AI systems play an essential role in AI system acceptance (Wolf 2019), text-genAI neglects to provide information to the end-user. This also involves dependencies on open and closed source models (Banh & Strobel 2023). Intuitively, one might assume that transparency leads to the highest level of trust in a positive sense. A study by Schmidt, Biessmann, and Teubner (2020) suggests that maximum transparency might negatively influence trust perception and lead to overtrust. In this context, we emphasize the need for caution when relying on transparency to build trust because this might further amplify the overconfidence in AI-generated outputs. Thus, developers must consider that transparency and explainability can positively and negatively impact trust when developing appropriate solutions.

3 Further Research Plans

To guarantee the integration of relevant literature and related work, we started our research journey with a structured literature review, following the process according to vom Brocke et al. (2009). The derivated key factors, presented in chapter 2, form the starting point for further research. Overall, our work follows a Design Science Research process according to Kuechler & Vaishnavi (2012) outlined in Figure1.

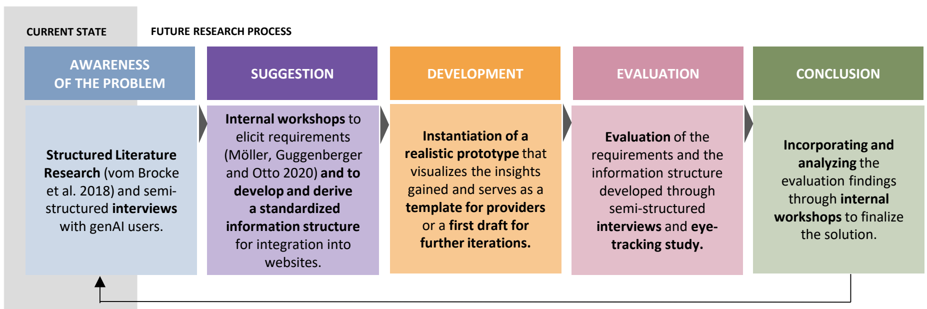

Figure 1. Design Science Research based on Kuechler & Vaishnavi (2012)

(1) Awareness of the problem: During the next steps, we will use the data from the structured literature review to develop an interview structure for a targeted exchange with end users. This will expand our knowledge base, identify further decisive influencing factors, and generate ideas for an adequate design.

(2) Suggestion: To achieve a user-centered and purposeful result, we derive requirements before establishing the solution, following the process outlined by Möller et al. (2020), which explicitly targets requirement development. The key factors derived from section 2, such as different AI Literacy levels, need to be considered, as well as related mitigation pathways (Harrer 2023, Weidinger et al. 2022), as their coverage and consideration are essential for a successful prototype. We then dedicate the standardized information structures as our artifact and prepare the associated core elements to ensure their helpfulness for end-users. The artefact is intended to provide a basis for helping users interpret the results when interacting with text-genAI, and to provide users with key information to help them assess whether to trust the service. Depending on the requirements, it can be a static or dynamic artefact. In considering points of reference for a possible solution, we analyze trust seals and information subpages, such as imprints and data protection notices on websites, based on scientific findings (Tomitza et al. 2023).

(3) Development: We implement a prototype that embeds the desired requirements in a suitable context to test if the solution meets users' expectations and needs. In addition, we use the development of the prototype to strive for a visualization that is as language-neutral, barrier-free, and targeted as possible to achieve maximum dissemination and acceptance of the overall results of our research work. Elements are derived that take into account the key points addressed in the literature research and in the interviews, such as the consideration of different AI literacy levels among end users, in order to

enable all user groups to understand and quickly grasp the information provided (e.g., in the form of a traffic light system, trust labels).

(4) Evaluation: We conduct semi-structured interviews with genAI end-users and an eye-tracking study to demonstrate that the developed solution supports users. We explicitly consider the AI literacy of the end users, for example, by addressing appropriate measurement methods (Wang et al. 2023, Weber et al. 2023). During the evaluation phase, we assess the requirements, the proposed solution, the prototype implementation, and the associated visualization elements.

(5) Conclusion: Subsequently, we analyze the data collected during the evaluation and incorporate improvements into the solution. If we make far-reaching changes, we begin a new iteration process.

4 Conclusion and Next Steps

Due to the advantages that arise with the emergence of AI-generated texts, solutions should be prioritized that do not slow down technical progress and acceptance in the population but rather help AI end-users to assess content, e.g., by showing them the technical background of the solution or providing them with warning messages and comprehensible information (Tomitza et al. 2023). Our research-in-progress paper examines which factors lead to excessive confidence in the output. We identify AI literacy, the difficulty in detecting misinformation in the output, and the paucity of transparency and explainability as crucial factors. As with any research, our findings should be weighed against possible limitations related to the nature of a representative literature review. Thus, there is a possibility that contributions that did not appear in our review exist. We are confident in the validity of our findings since we followed the structured approach by vom Brocke et al. (2009) and cross-checked the analysis results with multiple authors.

We have found that there is currently a lack of scientific approaches that help people with different levels of knowledge to interpret genAI output through structured information processing. This opens up numerous approaches for future research that can make a decisive contribution to science. In future research, we aim to provide an initial supporting approach that addresses a large user community and highlights essential information elements by structuring the information requirements. We address this in developing a standardized information structure as our artifact. Accordingly, we contribute to the protection of end-users and promote the use of AI output and its acceptance by society. Our goal is a solution that is ideally also language-neutral or can be embedded in multilingual websites to create an overarching protection concept with a chance of widespread dissemination.

5 Acknowledgements

This work has been developed in the project Hyko and is partly funded by the Bavarian Ministry of Economic Affairs, Regional Development and Energy enabled by the European Regional Development Fund (ERDF) under grant number 1.2-StMWK-F.4-UFR-002. The authors are responsible for the content of this publication.

References

Ali, S., DiPaola, D., Lee, I., Sindato, V., Kim, G., Blumofe, R. & Breazeal, C. (2021), 'Children as creators, thinkers and citizens in an AI-driven future', *Computers and Education: Artificial Intelligence* **2**, 100040.

Alkaissi, H. & McFarlane, S. I. (2023), 'Artificial Hallucinations in ChatGPT: Implications in Scientific Writing.', *Cureus* **15**(2).

Banh, L. & Strobel, G. (2023), 'Generative artificial intelligence', *Electronic Markets* **33**(1), 63.

Bankins, S., Ocampo, A. C., Marrone, M., Restubog, S. L. D. & Woo, S. E. (2023), 'A multilevel review of artificial intelligence in organizations: Implications for organizational behavior research and practice', *Journal of Organizational Behavior* .

Borji, A. (2023), 'A Categorical Archive of ChatGPT Failures'.

Datta, P., Whitmore, M. & Nwankpa, J. K. (2021), 'A Perfect Storm', *Digital Threats: Research and Practice* **2**(2).

Deng, J. & Lin, Y. (2022), 'Frontiers in Computing and Intelligent Systems The Benefits and Challenges of ChatGPT: An Overview', *Frontiers in Computing and intelligent Systems* **2**(2), 81–83.

Druga, S. & Ko, A. J. (2021), 'How do children's perceptions of machine intelligence change when training and coding smart programs?', *Proceedings of Interaction Design and Children 2021* pp. 49–61.

Dwivedi, Y. K., Kshetri, N. & et al., H. (2023), '"So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy', *International Journal of Information Management* **71**.

Fielding, J. A. (2019), 'Rethinking CRAAP: Getting students thinking like fact-checkers in evaluating web sources', *College & Research Libraries News* **80**(11), 620.

Gao, C. A., Howard, F. M. & Markov, N. S. e. a. (2022), 'Comparing scientific abstracts generated by ChatGPT to original abstracts using an artificial intelligence output detector, plagiarism detector, and blinded human reviewers', *bioRxiv* **12**, 1–18.

Gaube, S., Suresh, H., Raue, M., Merritt, A., Berkowitz, S. J., Lermer, E., Coughlin, J. F., Guttag, J. V., Colak, E. & Ghassemi, M. (2021), 'Do as AI say: susceptibility in deployment of clinical decision-aids', *npj Digital Medicine 2021 4:1* **4**(1), 1–8.

Glikson, E. & Woolley, A. W. (2020), 'Human Trust in Artificial Intelligence: Review of Empirical Research', **14**(2), 627–660.

Griesbaum, J. (2022), *Informationskompetenz*, Springer Berlin Heidelberg, pp. 67–98.

Harrer, S. (2023), Attention is not all you need: the complicated case of ethically using large language models in healthcare and medicine Personal View, Technical report.

Herm, L.-V., Heinrich, K., Wanner, J. & Janiesch, C. (2023), 'Stop ordering machine learning algorithms by their explainability! A user-centered investigation of performance and explainability', *International Journal of Information Management* **69**, 102538.

Horvitz, E. (2022), 'On the Horizon: Interactive and Compositional Deepfakes', *ACM International Conference Proceeding Series* **22**, 653–661.

Kong, S. C., Man-Yin Cheung, W. & Zhang, G. (2021), 'Evaluation of an artificial intelligence literacy course for university students with diverse study backgrounds', *Computers and Education: Artificial Intelligence* **2**, 100026.

Kuechler, W. & Vaishnavi, V. (2012), 'A framework for theory development in design science research: multiple perspectives', *Journal of the Association for Information systems* **13**(6), 3.

Laupichler, M. C., Aster, A., Schirch, J. & Raupach, T. (2022), 'Artificial intelligence literacy in higher and adult education: A scoping literature review', *Computers and Education: Artificial Intelligence* **3**, 100101.

Lee, I., Ali, S., Zhang, H., Dipaola, D. & Breazeal, C. (2021), 'Developing Middle School Students' AI Literacy', *SIGCSE 2021 - Proceedings of the 52nd ACM Technical Symposium on Computer Science Education* **7**, 191–197.

Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I. & Pechenkina, E. (2023), 'Generative AI and the future of education: Ragnarök or reformation?', *International Journal of Management Education* **21**(2).

Lockey, S., Gillespie, N., Holm, D. & Someh, I. A. (2021), A review of trust in artificial intelligence: Challenges, vulnerabilities and future directions, *in* 'Proceedings of the Annual Hawaii International Conference on System Sciences', Vol. 2020-Janua, pp. 5463–5472.

Long, D. & Magerko, B. (2020), What is AI Literacy? Competencies and Design Considerations, *in* 'Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems', CHI '20, Association for Computing Machinery, New York, NY, USA, pp. 1–16.

Mackenzie, D. (2023), 'Surprising Advances in Generative Artificial Intelligence Prompt Amazement—and Worries', *Engineering* .

Mohseni, S., Zarei, N. & Ragan, E. D. (2021), 'A Multidisciplinary Survey and Framework for Design and Evaluation of Explainable AI Systems', *ACM Transactions on Interactive Intelligent Systems (TiiS)* **11**(3-4).

Möller, F., Guggenberger, T. M. & Otto, B. (2020), Towards a method for design principle development in information systems, *in* 'Designing for Digital Transformation. Co-Creating Services with Citizens and Industry: 15th International Conference on Design Science Research in Information Systems and Technology, DESRIST 2020, Kristiansand, Norway, December 2–4, 2020, Proceedings 15', Springer, pp. 208–220.

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W. & Qiao, M. S. (2021), 'Conceptualizing AI literacy: An exploratory review', *Computers and Education: Artificial Intelligence* **2**, 100041.

Nourani, M., Roy, C. & Block, J. E. e. a. (2022), 'On the Importance of User Backgrounds and Impressions: Lessons Learned from Interactive AI Applications', *ACM Transactions on Interactive Intelligent Systems* **12**(4), 1–29.

Peres, R., Schreier, M., Schweidel, D. & Sorescu, A. (2023), 'On ChatGPT and beyond: How generative artificial intelligence may affect research, teaching, and practice', *International Journal of Research in Marketing* .

Qu, Y., Liu, P., Song, W., Liu, L. & Cheng, M. (2020), A Text Generation and Prediction System: Pre-training on New Corpora Using BERT and GPT-2, *in* '2020 IEEE 10th

International COnference on Electronics Information and Emergency Communication'.

Ray, P. P. (2023), 'ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope'.

Rikab, W. (2023), 'Detectors of AI-Generated Text Often Fail. Here Is What to Do.'.

Schaschek, M. & Engel, S. (2023), Measuring trustworthiness of ai systems: A holistic maturity model, *in* 'Rising like a Phoenix: Emerging from the Pandemic and Reshaping Human Endeavors with Digital Technologies ICIS 2023', pp. 1–9.

Schaschek, M., Gwinner, F., Hein, B. & Winkelmann, A. (2023), From black box to glass box: Evaluating faithfulness of process predictions with gcnns, *in* 'XKDD Workshop Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases', pp. 1–17.

Schmidt, P., Biessmann, F. & Teubner, T. (2020), 'Transparency and trust in artificial intelligence systems', *Journal of Decision Systems* **29**(4), 260–278.

Seeber, I., Bittner, E. & Briggs, R. O. e. a. (2020), 'Machines as teammates: A research agenda on AI in team collaboration', *Information & Management* **57**(2), 103174.

Shin, D. (2021), 'The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI', *International Journal of Human-Computer Studies* **146**, 102551.

Sieck, W. R. & Arkes, H. R. (2005), 'The recalcitrance of overconfidence and its contribution to decision aid neglect', *Journal of Behavioral Decision Making* **18**(1), 29–53.

Su, J. & Ng, D. T. K. (2023), 'Artificial intelligence (AI) literacy in early childhood education: The challenges and opportunities', *Computers and Education: Artificial Intelligence* p. 100124.

Teubner, T., Flath, C. M., Weinhardt, C., van der Aalst, W. & Hinz, O. (2023), 'Welcome to the Era of ChatGPT et al.', *Business and Information Systems Engineering* .

Tomitza, C., Schaschek, M., Straub, L. & Winkelmann, A. (2023), What is the minimum to trust ai?—a requirement analysis for (generative) ai-based texts, *in* 'Wirtschaftsinformatik 2023 Proceedings'.

Van Dis, E. A. M., Bollen, J., Van Rooij, R., Zuidema, W. & Bockting, C. L. (2023), ChatGPT: five priorities for research, Technical report.

vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R. & Cleven, A. (2009), Reconstructing the giant: On the importance of rigour in documenting the literature search process, *in* '17th European Conference on Information Systems (ECIS)', p. 161.

Wang, B., Rau, P.-L. P. & Yuan, T. (2023), 'Measuring user competence in using artificial intelligence: validity and reliability of artificial intelligence literacy scale', *Behaviour & information technology* **42**(9), 1324–1337.

Weber, P., Baum, L. & Pinski, M. (2023), 'Messung von ai literacy–empirische evidenz und implikationen'.

Weidinger, L., Uesato, J. & Rauh, M. e. a. (2022), 'Taxonomy of Risks posed by Language Models', *ACM International Conference Proceeding Series* **22**, 214–229.

Weisz, J. D., Muller, M., He, J. & Houde, S. (2023), 'Toward General Design Principles for Generative AI Applications'.

Wolf, C. T. (2019), 'Explainability scenarios: Towards scenario-based XAI design', *International Conference on Intelligent User Interfaces, Proceedings IUI* **Part F1476**, 252–257.

Yang, W. (2022), 'Artificial Intelligence education for young children: Why, what, and how in curriculum design and implementation', *Computers and Education: Artificial Intelligence* **3**, 100061.

Ziwei, J., Nayeon, L., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Jin, B., Madotto, A. & Fung, P. (2023), 'Survey of Hallucination in Natural Language Generation', *ACM Computing Survey2* **55**(12), 1–38.