Intelligent Human Systems Integration (IHSI 2025), Vol. 160, 2025, 378–388
https://doi.org/10.54941/ahfe1005843

Evaluating AI-Generated Research Plans: Expert Insights From a Blind Authorship Study

Dam Thi Thien Nhi and Leonhard Glomann

LINC Interactionarchitects GmbH, Munich, Germany

ABSTRACT

Artificial Intelligence (AI) is increasingly being adopted in academia to enhance various research activities and has proven to be a valuable research accelerator in recent studies. This study examines the capabilities of large language models (LLMs), specifically ChatGPT and Gemini, in generating comprehensive research plans during the early stages of research. We tasked both models with developing research plans on the topic of "Gen Z's willingness to sacrifice convenience for environmental benefits" including interview guidelines and survey questions. Eight expert researchers evaluated these research plans without knowing they were AI-generated. Our findings provide in-depth insights into the perceptions of expert researchers regarding the quality of AI-generated research plans, identifying missing elements and pitfalls of utilizing AI in the planning activity of research. The necessity for researchers to oversee and intervene in AI outputs is emphasised in our research to fully leverage the advantages offered by this technology.

Keywords: Artificial intelligence, Large language model, ChatGPT, Gemini, Academic research

INTRODUCTION

Artificial Intelligence (AI) has recently emerged as a transformative force in academia, streamlining various stages of the research process. Previous studies have explored its applications in literature reviews (Wagner et al., 2021, Bolanos et al., 2024), research material preparation (Chubb, 2023), data analysis (Hamilton et al., 2023), and scientific writing (Macdonald et al., 2023; AlZaabi et al., 2023; Khalifa et al., 2024). AI is widely recognized for its ability to enhance research efficiency (Macdonald et al., 2023; Alshater, 2022) and minimize human error (Burger et al., 2023). However, it also faces significant limitations, including a lack of common sense (Bian et al., 2023), limited contextual understanding (Farrokhnia et al., 2023), susceptibility to inaccuracies, lack of transparency (AlZaabi et al., 2023), and response biases (Ekundayo et al., 2024; Kasneci et al., 2023). These drawbacks fuel researchers' mistrust, restricting AI's role to tasks like paraphrasing and reference searches (Abdelhafiz et al., 2024).

Building on previous research exploring AI applications in academia, our study specifically examines the capacity of Large Language Models (LLMs) to assist in the research planning stage. We tasked two leading LLMs, ChatGPT

(OpenAI) and Gemini (Google), with generating research plans on the topic of "*Gen Z's willingness to sacrifice convenience when using goods for environmental benefits*." To ensure objective evaluation, expert researchers assessed these plans without knowing their origin, mitigating potential bias against AI-generated work (Bellaiche et al., 2023). Through subsequent interviews, we gathered researcher perceptions and insights regarding the quality and utility of the AI-generated plans. This study aims to provide a comprehensive understanding of the potential benefits, limitations, and best practices for integrating LLMs into research workflows. By contributing to the ongoing discourse surrounding AI in academia, we offer valuable insights for researchers navigating the evolving landscape of research and seeking to effectively leverage AI in their work.

STUDY DESIGN

To assess the capability of LLMs in generating research plans, we utilized ChatGPT (GPT-4 Plus) and Google's Gemini (1.5 Pro), providing both models with identical prompts to develop plans on the topic: "*Investigate Gen Z users' insights into their willingness to sacrifice convenience for environmental benefits when using goods*." This topic was chosen for its broad relevance and accessibility to researchers across various disciplines with expertise in qualitative and quantitative methodologies.

Both LLMs proposed mixed-methods approaches, incorporating surveys for quantitative data and interviews for qualitative insights. We further prompted both models to generate corresponding interview guidelines and survey questions. The research plans generated by ChatGPT (Plan A) and Gemini (Plan B) share a similar structure with varied section names and content arrangement. Overall, both plans include sections for Research Objectives, Research Questions, Methodology, Data Collection Tools, Data Analysis, Ethical Considerations, Timeline, and Expected Outcomes. Plan A uniquely includes a list of hypotheses and a budget section, while Plan B features an "Additional Considerations" section with tips for enhancing engagement, recruiting participants, and pilot testing. Full materials of AI-generated outputs can be found in the Appendix.

The outputs were compiled into a single document, labelled as Research Plan A and Research Plan B, and shared with researchers blinded to the plans' origins. The document was prefaced with an introduction: "*We are initiating a research project to explore the perspectives of Gen-Z users on their willingness to sacrifice convenience for environmental benefits when using goods. To ensure the success of this study, we have developed two distinct research plans and would appreciate your insights in evaluating them.*"

Eight researchers from the United States, Canada, and Australia were recruited for this study. Participants were chosen for their extensive research experience, with six having over 10 years of experience and two having between 6 and 10 years. Research is a core part of their professional roles, and five out of the eight participants hold a PhD. All participants are proficient in both qualitative and quantitative research methods with their expertise

spanning various domains, detailed participant information is depicted in Table 1. Feedback on the research plans was gathered through individual online interviews.

Table 1. Participants information.

ID	Research Experience	Research Field	Country
R1	6–10 years of experience, 8+ publications, PhD	Healthcare/Medicine	Canada
R2	10+ years of experience, 4–7 publications, PhD	Social Sciences	US
R3	10+ years of experience, 4–7 publications	Social Sciences	Australia
R4	10+ years of experience, 8+ publications, PhD	Environmental Science	Canada
R5	10+ years of experience, 8+ publications, PhD	Healthcare and Engineering applied to healthcare	Canada
R6	10+ years of experience, 8+ publications, PhD	Engineering/Technical	US
R7	10+ years of experience, 8+ publications	Business/Marketing	US
R8	6–10 years of experience, 8+ publications	Social Sciences	US

FINDINGS

AI-Generated Research Plan is a Good Draft to Start With

Researchers generally agreed that both research plans provide a solid foundational structure outlining essential components for a research plan. While the research questions and research objectives in both plans require further refinement, they were considered as a promising start. Regarding methodology, the mixed-methods approach proposed in both plans were supported by all researchers. In terms of the interview guidelines and survey questions, most researchers rated the quality between 4 and 5 out of 6 for both plans, despite acknowledging room for improvement. Many questions received positive feedback, with researchers highlighting and praising their ability to effectively explore and capture key aspects of the research topic.

In Research Plan B, the "Additional Considerations" section, which includes four practical tips for conducting the research, received notable praise from researchers. For example, participant R4 highlighted the first point: "*Use engaging visuals and interactive elements in the survey to enhance participation rates*", and remarked "*whoever did this is I think comes from an understanding that it's hard to get participants sometimes.*" With experience on a research committee evaluating proposals, R5 shared that these considerations are a "green flag" for him in a research proposal.

Need for Justification and Specificity in Research Plans

Researchers emphasized the need for greater focus and justification in the research plans, citing omissions and vague terms that could undermine the study's validity. They frequently suggested that the proposed research topic was overly broad and vague for a focused study. To address this, they recommended narrowing the scope by specifying a particular geographical context, which would significantly influence various aspects of the plan, such as the interview and survey questions. In addition to geographic factors, researchers also recommended refining the study to target a specific product sector, noting that consumer behavior and environmental attitudes can vary widely between categories like food and fashion products (R6).

Both plans propose a mixed-methods approach but fail to explain this choice or specify whether the research will be conducted online or in person, impacting the study's design. Plan A includes sample sizes of "*300–500 participants for quantitative analysis; 20–30 participants for in-depth qualitative interviews*" without rationale, while Plan B omits them entirely.

In terms of Ethical Considerations, the provided details were generally deemed as "*really basic*" (R5). R3 raised concerns about confidentiality, especially if data is collected online, and R5 questioned whether an ethical committee would be involved, asking, "*Will the research project be sent for evaluation to an ethical board? If so, which one? Will it be university based?*" Additionally, protocols for handling data from participants who withdraw were not addressed. R5, from the viewpoints of a research committee, criticised *"if I would have received such a proposal for evaluation, I would have rejected it with this ethical consideration"*. Discussing the Timeline, researchers pointed out that the number of people working on the project was not mentioned while it is a decisive factor. R1 and R3 also inquired about the project's stakeholders, asking whether it is intended for a business, an institute, or a government body.

The use of vague terms, such as "convenience," "environmental awareness," and "eco-friendly," was criticized for potentially leading to inconsistent interpretations. For example, Plan A's objective - "*to understand the extent to which Gen-Z is willing to compromise convenience for environmentally friendly practices*" - left "*convenience*" undefined, prompting questions from researchers that what it refers to, *"Is it affordability, mobility, or the time it takes to get the service?"* (R1). Researchers recommended providing clear definitions and examples to ensure consistent understanding among participants.

Lack of Coherence and Empathy Factors

Researchers frequently highlighted disconnections within both research plans. R7 noted that the hypotheses in Plan A were not aligned with the research questions, while R6 recommended rearranging interview questions for a more coherent narrative. Additionally, Plan A proposed a two-week timeline for the literature review as the first step, but this task was absent from the rest of the plan. Both plans included interview and survey questions

poorly aligned with the research objectives. R8 remarked, *"They ask this question in the interview, but it didn't relate to the research questions."* In Plan B, R7 pointed out a disconnect between the survey and interview guidelines, as the survey included a question on social media's influence on user choices, but the interview only focused on family and friends.

Researchers also discussed the lack of narrative and empathising factors from constructed interview guidelines. R4 indicated the important aspect of conducting interview is "*to build trust with people, making them comfortable enough to share with you how they actually feel about something"*, therefore, she did not agree with the arrangement of interview questions in plan A when it immediately asks "*How would you describe your daily habits in terms of environmental consciousness?*" right after the question about demographic background. R6 shared a similar viewpoint and recommended rearranging the interview questions to create a more coherent narrative for smooth conversations with participants.

Unrealistic Factors and Mistakes

Researchers highlighted unrealistic elements and critical mistakes in the two research plans. They questioned the feasibility of Plan A's timeline, noting that four weeks for data collection was insufficient to recruit 300–500 survey respondents and conduct 20–30 interviews. R1 observed that achieving this within the proposed timeframe would require an exceptional recruitment strategy and high response rates. Other phases of the research were also seen as overly brief, leading researchers to recommend significant extensions to the timeline. Plan B's timeline was viewed as more realistic, with R4 suggesting it reflected the input of someone with more experience.

Besides, researchers identified critical mistakes in the research plan B. R5 and R7 remarked that the age categories in the demographic question (13–17; 18–21; 22–25) were unsuitable for the intended Gen Z audience. The inclusion of the 13–17 age group is particularly problematic, as it does not accurately reflect the Gen Z cohort. Furthermore, involving participants under 18 leads to serious ethical concerns and requires compliance with applicable laws and regulations.

Comparison Between Two Research Plans

When comparing the two research plans, three researchers (R3, R6, R7) chose Plan A, while four (R1, R4, R5, R8) favoured Plan B. One researcher (R2) remained neutral, noting that Plan A was more academically structured, while Plan B offered actionable insights for companies.

Researchers R3 and R6 were drawn to Plan A for its comprehensive detail. R6, after examining the Output section of Plan B, found it overly simplistic, describing it as *"coarse and not comprehensive, almost as if copied from somewhere."* He concluded, *"every part of Plan A is better than Plan B"*, which was agreed by R7. Conversely, four researchers supporting plan B agreed that the specific focus made it more compelling. R5 commented *"If two students came to me with Plan A and Plan B, I'd recommend going with*

Plan B and refining it." R8 particularly appreciate its consistency throughout the proposal.

Overall, researchers all suggested combining strengths from both plans, such as integrating Plan A's insights on social impact and budget details into Plan B, and each plan has its own strong interview and survey questions that could complement each other.

Researchers' Experiences and Insights Drive Plan Refinement

While evaluating the plans, researchers contributed diverse perspectives shaped by their individual experiences and perceptions of the topic, helping to refine the research plans. For participant recruitment, R4, R7, and R8 proposed conducting the survey first, then strategically selecting interviewees from respondents to gain deeper insights and further build on the survey data. R1 and R7 emphasized the need for flexibility in survey timelines to account for factors that might influence response rates.

Regarding interview questions, R3 highlighted that eco-friendly products are often more expensive and recommended including questions about participants' incomes to explore how financial factors might affect ecological choices. R2 noted that eco-awareness is frequently discussed online, suggesting that frequent internet users may exhibit greater eco-awareness and articulate opinions on the topic more effectively than those with less online engagement.

Researchers also brought contrasting viewpoints on certain aspects of the plans. For instance, while R2, R4, and R7 raised concerns that lengthy questions in both plans might lead to participant disengagement, others found the question lists acceptable. Similarly, Plan B's proposal to collaborate with environmental groups sparked mixed reactions: R3 commended this approach as an effective way for recruiting participants, but R1 cautioned that it might introduce pro-environmental biases, given that individuals in such organizations are likely to be more environmentally aware.

Unveiling the Author: Expert Reactions to AI Authorship

After completing their evaluations, the researchers were informed that both research plans had been generated by AI, surprising six of the eight participants. Some admitted they hadn't considered AI authorship at all. Notably, even before the reveal, R5 suspected the plans might be AI-generated. When asked about his preference between the two, he responded, "*First question I have to ask: were these really written by a real person?*" Being a ChatGPT user himself, R5 elaborated, "*I felt that some questions reminded me of how ChatGPT typically formulates things.*" Especially the "Additional Considerations" section of Plan B as exhibiting "a behaviour you expect from ChatGPT." Ironically, this plan was generated by Gemini, not ChatGPT. Similarly, R6, an AI researcher himself, was unsurprised to learn about the AI authorship, despite not showing any scepticism during the evaluation process.

The remaining six researchers were surprised by the revelation and generally praised AI's capabilities. They noted that AI did an effective job of

covering the basics, particularly in crafting interview and survey questions, which R1 described as "quite exhaustive" and having AI's help on this part is valuable.

Following this reveal, researchers began to attribute certain shortcomings, such as inconsistencies and lack of coherence, to AI's limitations. As R8 observed, *"I think it makes sense then why there is a lack of continuity between them. AI is really bad at maintaining context."* The perceived disconnection and quality inconsistencies led her to conclude, *"It seemed like one junior researcher came up with this part, and another junior researcher came up with that part, and then they put it all together."*

Shifting the discussion toward the broader use of AI, the researchers generally agreed that AI has proven valuable for repetitive, simple, and automated tasks such as refining writing, programming, summarising papers, and looking up information. Still, they expressed reluctance to fully trust AI in more complex academic endeavours. R4, for instance, stated, *"for my type of research, I'm a bit reluctant to rely on AI, because it might not pick up everything I need, so I'd rather read it and gain that understanding myself."* Similarly, R2 asserted that her research area of Alzheimer's disease is not something AI can handle, *"at least at this point"*, she emphasised.

Reflecting on AI's role in drafting research plans, researchers also raised concerns about potential risks to academic integrity. Five out of eight researchers noted that this could be particularly problematic for inexperienced researchers who may lack the expertise to validate AI's output, potentially leading to poor quality outcomes or even misinformation. R4 expressed her worries for future researchers, noting that creating research plans is a skill she teaches her students and it's fundamental for any researcher. If students rely on AI, they may not grasp the underlying principles and fail to develop essential research skills, then she raised her concern *"what is the quality of research going forward?"*

DISCUSSION

Building on recent studies of AI in academia, this research explores the potential of ChatGPT and Gemini to assist in early-stage research planning, with expert researchers evaluating AI-generated plans. Our findings reaffirm existing insights on AI's academic applications and provide fresh perspectives on its use in research.

Our study affirms the advantage of LLM models in research planning, particularly in their capacity to efficiently establish foundational structures and core elements of a research plan. Expert researchers in our study also found AI-generated interview guidelines and survey questions to be especially helpful, as formulating these questions is often a challenging part of study preparation. Specifically, Gemini impressed experts by some characteristics comparable to those of an experienced researcher.

Despite its advantages, our study highlights certain limitations of AI in crafting research plans. Consistent with Michel-Villarreal's analysis (Michel-Villarreal et al., 2023), and as ChatGPT itself disclaims, AI lacks genuine knowledge and domain-specific expertise. Our findings underscore this

shortfall, as the AI-generated research plans frequently missed critical details and failed to provide justifications for key decisions, such as methodologies and sample sizes. It's worth noting that the AI outputs in our study were based on single-prompt inputs, whereas research indicates that advanced prompt engineering techniques can yield more comprehensive outputs for complex tasks (Mostafapour et al., 2024; Koyuturk et al., 2023). Indeed, LLMs like ChatGPT have demonstrated their impressive ability to reason with provided contextual information (Joublin et al., 2023). The need for greater specificity in the research topic, such as clearly defining geographic and product-sector contexts, as highlighted in our findings, could be addressed by applying refined prompt development.

Additionally, our study reinforces earlier findings about AI's limitations in contextual comprehension (Mostafapour et al., 2024; Bano et al., 2023; Alshater, 2022; Farrokhnia et al., 2023; Bian et al., 2023), illustrating that AI often fails to unpack the nuanced and implicit meanings of a conversation. As a result, its responses may include irrelevant details or tend to be overly generic. In our study, both ChatGPT and Gemini generated vague terms lacking specific context, potentially leading to misinterpretations and negatively impacting data collection. Furthermore, aligning with Hamilton's findings on AI's lack of emotional intelligence (Hamilton et al., 2023), expert researchers in our study noted the absence of empathy in the generated interview guidelines. Our findings also point to AI's difficulty in maintaining narrative coherence across research plans, interview guides, and survey questions, which is likely due to memory limitations, which can impede consistency over extended content generation (Kim et al., 2024). Additionally, expert reviews spotted errors and unrealistic factors in the research plans that demonstrated the inaccuracies and misinformation in AI outputs as discussed in previous studies (Mostafapour et al., 2024; Michel-Villarreal et al., 2023; Alshater, 2022).

When discussing the potential risks of using AI in research, Kasneci et al. (2023) and Qasem (2023) highlighted the drawbacks of over-reliance, particularly in educational settings, where critical skills like problem-solving and critical thinking could be undermined. Similarly, experts in our study expressed concerns that novice researchers might struggle to grasp the rationale behind or validate AI-generated outputs, potentially compromising the quality of research in the long term.

Supporting the survey findings by Abdelhafiz et al. (2024) that researchers remain wary of AI and limit its use to simple tasks, the expert researchers in our study reported minimal reliance on AI in their research endeavours. Instead, they primarily used AI for information retrieval and expressed continued mistrust of its application in complex academic work.

Given AI's limitations, the role of human oversight over AI-generated content remains essential, which has been consistently reinforced by previous research (Bano et al., 2023; Burger et al., 2023; Wagner et al., 2021) and strongly underscored by our findings. Expert feedback identified substantial room for improvement, suggesting that while AI is useful for initial idea generation and foundational structuring, it is crucial to approach its outputs with a critical eye, recognizing the potential for missing information,

inconsistencies, and errors. In the research context, we found that individual perspectives and experience significantly shape how studies are planned and designed, aligning with the discussion by Bano et al. (2023) comparing human reasoning versus LLM-driven data analysis, where human analysts heavily lean on their individual experiences to analyse data. Therefore, researchers must actively engage with AI-generated content, leveraging their expertise to refine and contextualise the output. Interestingly, our study revealed no clear preference among expert researchers for either ChatGPT or Gemini in research planning tasks. This suggests that a holistic approach, combining the strengths of both tools under careful human supervision and insights, can optimise efficiency and quality of final outcomes.

CONCLUSION

The use of AI in research continues to demonstrate its promise, particularly in the early stages of study design. By employing large language models to generate research plans for specific topics and having expert researchers evaluate the outcomes, our findings provide valuable insights into the perceptions of these experts regarding AI-generated plans. Generally, researchers view AI-generated outcomes as a solid starting point; however, significant limitations persist, including the omission of critical information, a lack of justification for certain choices, generic elements that lack precision, and an absence of coherence and empathy in the interview guidelines. The importance of researchers' oversight is once again emphasised as essential and irreplaceable. Their individual experiences may influence the direction of the plan and contribute to the diversity of research. While there has been considerable exploration of AI's applications at various stages of the research process, its ability to craft complete research plans remains limited. Our study is expected to serve as a foundation for future research in this area. Subsequent studies could complement our study by exploring the capabilities of AI in generating research plans on more specific topics, employing advanced prompt engineering techniques as guided in prior studies to uncover its full potential and limitations.

APPENDIX

Research plans including interview guideline and survey questions generated by ChatGPT and Gemini can be found at https://doi.org/10.5281/zenodo.14213763.

REFERENCES

Abdelhafiz, A. S., Ali, A., Maaly, A. M., Ziady, H. H., Sultan, E. A., & Mahgoub, M. A. (2024). Knowledge, Perceptions and Attitude of Researchers Towards Using ChatGPT in Research. Journal of Medical Systems, 48.

Alshater, M. M. (2022). Exploring the Role of Artificial Intelligence in Enhancing Academic Performance: A Case Study of ChatGPT. SSRN Electronic Journal.

AlZaabi, A. A., ALAmri, A., Albalushi, H., Aljabri, R., & AAlAbdulsalam, A. K. (2023). ChatGPT applications in Academic Research: A Review of Benefits, Concerns, and Recommendations. bioRxiv.

Bano, M., Zowghi, D., & Whittle, J. (2023). Exploring Qualitative Research Using LLMs. ArXiv, abs/2306.13298.

Bellaiche, L., Shahi, R., Turpin, M. H., Ragnhildstveit, A., Sprockett, S., Barr, N., Christensen, A., & Seli, P. (2023). Humans versus AI: Whether and why we prefer human-created compared to AI-created artwork. Cognitive Research: Principles and Implications, 8.

Bian, N., Han, X., Sun, L., Lin, H., Lu, Y., & He, B. (2023). ChatGPT Is a Knowledgeable but Inexperienced Solver: An Investigation of Commonsense Problem in Large Language Models. ArXiv, abs/2303.16421.

Bolanos, F., Salatino, A., Osborne, F., & Motta, E. (2024). Artificial Intelligence for Literature Reviews: Opportunities and Challenges. ArXiv, abs/2402.08565.

Burger, B., Kanbach, D. K., Kraus, S., Breier, M., & Corvello, V. (2023). On the use of AI-based tools like ChatGPT to support management research. European Journal of Innovation Management.

Chubb, L. A. (2023). Me and the Machines: Possibilities and Pitfalls of Using Artificial Intelligence for Qualitative Data Analysis. International Journal of Qualitative Methods, 22.

Ekundayo, T., Khan, Z., & Nuzhat, S. (2024). Evaluating the Influence of Artificial Intelligence on Scholarly Research: A Study Focused on Academics. Human Behavior and Emerging Technologies.

Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. E. (2023). A SWOT analysis of ChatGPT: Implications for educational practice and research. Innovations in Education and Teaching International, 61, 460–474.

Hamilton, L., Elliott, D., Quick, A., Smith, S., & Choplin, V. (2023). Exploring the Use of AI in Qualitative Analysis: A Comparative Study of Guaranteed Income Data. International Journal of Qualitative Methods, 22.

Joublin, F., Ceravola, A., Deigmoeller, J., Gienger, M., Franzius, M., & Eggert, J. (2023). A Glimpse in ChatGPT Capabilities and its impact for AI research. ArXiv, abs/2305.06087.

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G. L., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., Weller, J., Kuhn, J., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences.

Khalifa, Mohamed & Albadawy, Mona. (2024). Using Artificial Intelligence in Academic Writing and Research: An Essential Productivity Tool. Computer Methods and Programs in Biomedicine Update. 5. 100145. 10.1016/j.cmpbup.2024.100145.

Kim, M., Shim, K., Choi, J., & Chang, S. (2024). InfiniPot: Infinite Context Processing on Memory-Constrained LLMs. *ArXiv, abs/2410.01518*.

Koyuturk, C., Yavari, M., Theophilou, E., Bursic, S., Donabauer, G., Telari, A., Testa, A., Boiano, R., Gabbiadini, A., Leo, D. H., Ruskov, M., & Ognibene, D. (2023). Developing Effective Educational Chatbots with ChatGPT prompts: Insights from Preliminary Tests in a Case Study on Social Media Literacy. ArXiv, abs/2306.10645.

Macdonald, C., Adeloye, D., Sheikh, A., & Rudan, I. (2023). Can ChatGPT draft a research article? An example of population-level vaccine effectiveness analysis. Journal of Global Health, 13.

Michel-Villarreal, R., Vilalta-Perdomo, E. L., Salinas-Navarro, D. E., Thierry-Aguilera, R., & Gerardou, F. S. (2023). Challenges and Opportunities of Generative AI for Higher Education as Explained by ChatGPT. Education Sciences.

Mostafapour, M., Fortier, J. H., Pacheco, K., Murray, H., & Garber, G. E. (2024). Evaluating Literature Reviews Conducted by Humans Versus ChatGPT: Comparative Study. JMIR AI, 3.

Qasem, F. (2023), "ChatGPT in scientific and academic research: future fears and reassurances", Library Hi Tech News, Vol. 40, No. 3, pp. 30–32. https://doi.org/10.1108/LHTN-03-2023-0043

Wagner, G., Lukyanenko, R., & Paré, G. (2021). Artificial intelligence and the conduct of literature reviews. Journal of Information Technology, 37, 209–226.