

## Conducting Academic Research with the AI Interface ChatGPT: Challenges and Opportunities

Ismail Dönmez<sup>1</sup> Sahin Idil<sup>2</sup> Salih Gulen<sup>3</sup>

### ABSTRACT

**Editorial**

**Article History Received:**

3 Mar 2023

**Received in revised form:**

11 Mar 2023

**Accepted:** 14 Mar 2023

**Published:** 1 Jun 2023

Artificial intelligence technologies have become popular in recent years. The question of whether chatbots can be used in academic writing or research in artificial intelligence applications attracts the attention of many researchers. Can conducting a scientific research with artificial intelligence help in the research process? Can AI work like a co-writer? In this study, the opportunities and challenges experienced in the scientific research process using the ChatGPT application were examined. Its critiqued and included a previous study in the scientific research process under specific themes. The results show that while artificial intelligence technologies provide opportunities to researchers in terms of approval, creativity, and offering different perspectives, it is seen that there are reliability problems in producing content and it brings ethical and plundering issues. Therefore, artificial intelligence does not write articles for you but provides some tips and support in article writing.

**Keywords:** Artificial Intelligence, ChatGPT, Academic research, Chatbots

Donmez, I., Idil., S. & Gulen, S. (2023). Conducting academic research with the AI interface ChatGPT: Challenges and opportunities. *Journal of STEAM Education*, 6(2), 101-118.  
<https://doi.org/10.55290/steam.1263404>

## INTRODUCTION

Ian Turing's thought on Artificial Intelligence dates back to 1950: "Are machines capable of thinking?" Since then, many technologies have been developed that attempt to pass the Turing Test, such as ELIZA in 1966 and ALICE in 1995, as well as Apple Siri, Amazon Alexa, and Microsoft Cortana in 2021 (Xu et al, 2021). AI chatbots use natural language processing models to interpret human language, which is at the core of artificial intelligence (AI) chatbots (Nagarhalli et al., 2020). ChatGPT is a speech-based artificial intelligence interface (OpenAI, 2023) that interacts realistically and even uses natural language processing (NLP) to "answer follow-up questions, admit mistakes, challenge false premises, and reject inappropriate requests". ChatGPT is a powerful language model based on deep learning, released in 2020, the largest neural network ever produced with 175 billion parameters (Chatterjee & Dethlefs, 2023). It requires further research on how humans can coexist with artificial intelligence and how to minimize the negative impact of technology (Duan et al., 2019). Advances in artificial intelligence may bring new opportunities to experience power. For example, people may feel high power regarding a human-like autonomous digital assistant (Fast & Schroeder, 2020). AI-based digital assistants offer significant opportunities, but can also become a threat (Reddy, 2017; Maedche et al., 2019). In this research, we examine the convenience and difficulties that the use of ChatGPT will bring to researchers in the context of an exemplary scientific research process and published research.

Chatbots are conversational agents that can interact with users through natural languages and can also be defined by the broader term conversational user interfaces (Smestad, 2018). The term "chatbot" is derived from "chatterbot", a term first proposed by Michael Mauldin in 1997 to describe robots with which humans can chat (Derugina, 2010). ChatGPT is a public chatbot developed by OpenAI and based on the GPT language model technology (Kirmani, 2022). The platform reached one million users in its first week alone (Mollman, 2022; Vallance, 2022) and was soon dubbed the "next big disruptor of the industry" (Rudolph et al., 2023) due to the perceived quality of the response outputs from the model. ChatGPT uses machine learning to continuously improve its capabilities and learn from the internet. GPT technology is a powerful tool for NLP (Natural language processing) tasks, but it has limitations. One of the main limitations is that GPT models rely on a statistical approach that can perpetuate existing biases and stereotypes in the data, learning patterns from a large text dataset (Lucy & Bamman, 2021). The ChatGPT was trained on a variety of internet texts, including books, articles, and websites, and covering a wide range of topics such as news, and fiction (Shen et al., 2023). ChatGPT is set up specifically for conversational tasks. Because of its dialog format, ChatGPT can answer follow-up questions, admit mistakes, dispute false premises, and reject inappropriate requests. Question answering, storytelling, logical reasoning, code debugging, machine translation, etc. It integrates various capabilities of natural language processing, including (Jiao et al., 2023). A broad perspective on the emerging role of artificial intelligence in the production of scientific research has been taken by (Grimaldi & Ehrler, 2023). While Alshater (2022) suggested that ChatGPT should be useful for a range of tasks involved in constructing a research study without empirical testing, most applied research focused on the creation of research summaries and literature synthesis. For example, Aydın and Karaarslan (2022) try to create a health literature review suitable for an academic journal and state that although this is possible, there is a

significant amount of 'plagiarism' or insufficient paraphrasing. Chatterjee and Dethlefs (2023) suggest interacting with ChatGPT and sharing their positive and negative experiences with developers and the world so that together they can raise the ethical, integrity, and moral values of future artificial intelligence models like this one. Numerous blog posts and media outlets have reported the educational advantages of ChatGPT (Zhai, 2022); some even provided instructions for its use in classrooms (Lieberman, 2023). Tlili et al. (2023) followed a threestage instrumental case study: a social network analysis of tweets, content analysis of interviews, and research of user experiences to examine concerns about chatbot use in education through the ChatGPT use study. The results revealed that although ChatGPT is a powerful tool in education, it should be used more carefully and more guidelines should be created on how to use it safely in education ( Tlili et al., 2023). Although ChatGPT's responses are still flawed and its limitations are easily circumvented, ChatGPT's capabilities have caused much excitement and hysteria among the public since its launch, as reported by many researchers (Taecharungroj, 2023). Therefore, any part of a paper written by an NLP system should be checked by a domain expert for accuracy, bias, relevance, and reasoning. Second, the use of NLP systems raises issues of accountability. If part of an article written by an NLP system contains errors or biases, co-authors must be held accountable for its accuracy, credibility, and integrity (Hosseini et al., 2023). While research on speech systems and user interfaces has been conducted for decades, chatbot research and design are still in their relative infancy (Følstad et al. 2021). Among the many industries affected by ChatGPT, scientific publishing is one that should soon address the potential implications (Grimaldi & Ehrler, 2023). This research tries to examine some stages used in scientific research with ChatGPT, which uses artificial intelligence technology through a case study. It examines the answers given by ChatGPT. It creates a discussion about the opportunities and risks that it will create in its use. The purpose of this study can be determined by the following research questions:

- What are the challenges encountered in using AI interface ChatGPT to support academic research?
- What are the opportunities to use AI interface ChatGPT to support academic research?

## **METHOD**

This study is a qualitative research study using narrative research as an inquiry strategy. Qualitative research allows the researcher to access detailed data in their natural environment and gives the researcher the opportunity to interpret the data as it is interpretive (Cresswell, 2003). It also focuses on the experiences and ideas of the participants. As an inquiry strategy, narrative research aims to understand “the result of interpretation rather than explanations” by providing an opportunity to collect data from real life and lived experiences (Kramp, 2004). There are various approaches developed and widely used regarding the types of narratives in the literature. “Shared Experience Field Notes” were used in this study. In this approach, field records collected through participant observation in a shared practical setting are one of the primary tools of narrative inquiry work (Connelly & Clandinin, 1990).

## Analysis of Data

Narratives include only a part of the human experience and stories are based on narratives (Kim, 2016). In narrative analysis, researchers consider personal stories within the personal experiences of the participants (occupation, home), cultures, and historical contexts (time and space) (Uğuz Arsu & Tekindal, 2021). Just as there are various ways of understanding narrative research, digital technologies can be explored in different ways. One approach could examine the ways in which users interact with digital technologies in daily life. This type of research seeks to answer the questions of who (what kind of users), what (technology and impact), when (time of day or event), where (location matters), how (infrastructure or technique requirements), and why (motivation and intended effect) questions around people's activities related to technology (Green et al., 2019). In this study, the content based on the responses of ChatGPT was used in the analysis. Then, according to the answers, the researchers commented on what these answers meant. The findings, data, and experiences discussed in the analysis are expressed as they are and then discussed with the literature.

## RESULTS

### *Finding a Research Question*

With the research on ChatGPT, We asked some questions and looked for answers. For this, it is necessary to answer the question of how can I find a research question. Let's start by choosing a topic: We chose "STEM" as our topic and "STEM motivation" in particular (we will have the opportunity to review an article on this subject in the future). We asked my first question: Can you write research questions about STEM Motivation for students?

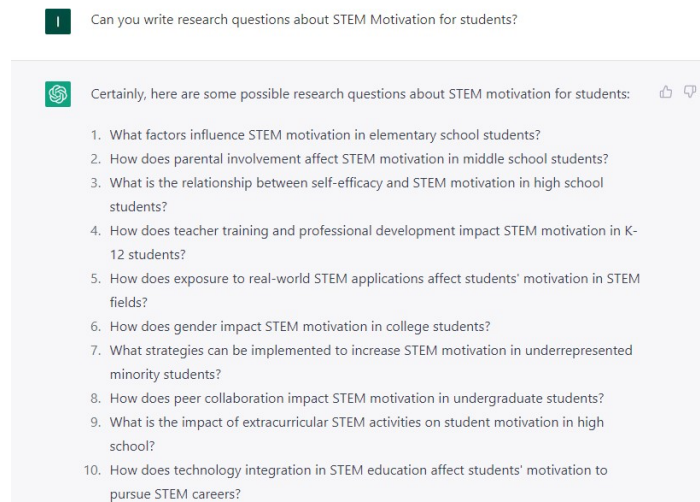


Fig 1. Finding a research question

The questions look interesting and offer different suggestions about motivation. It seems to be consistent within itself from different perspectives. It can be said that the questions are creative. It seems to go out of bounds in propositions 5 and 10. But something that interests me is that in most of the questions, I can only measure indirectly. I want him to iterate on his suggestions in

a more creative way. My question is “Can you write more creative questions about STEM motivation?”

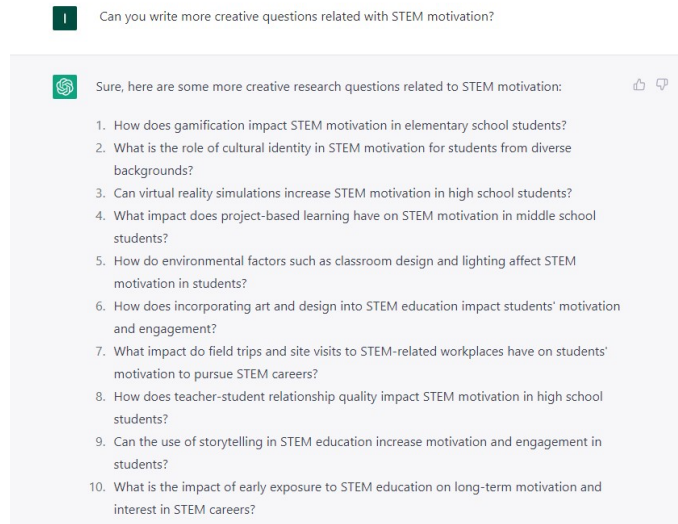


Fig 2. Revision of the research questions

At this stage, it is seen that it raises questions about STEM motivation. The questions are seen as consistent and meaningful in themselves. This time it seems to use the influence of secondary and tertiary tools such as storytelling, environmental factors, and culture. But as a problematic point, how can we choose one of them as a research question? Has any research been conducted on these issues before? Is there such a gap in the literature? As a result, if the question we chose has been asked before, it will not contribute to the literature and will cause unnecessary time and effort. At this stage, we asked ChatGPT if this question has been the subject of research before.

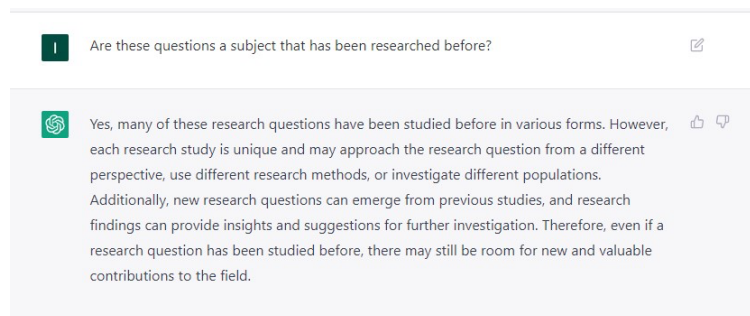


Fig 3. Examination of the research

ChatGPT gives a general answer. It states that these have been revealed from previous studies and may have been handled from different aspects. I know that some of these topics have been researched before. However, for the research process, “How do factors such as gender, ethnicity, and socio-economic status affect students' STEM motivation and how are these factors handled in STEM teaching? we chose the question. Now we need some advice on how to continue the research process. We would like to get an idea about how we should do this

research with ChatGPT. We pose a question like this: “How do factors such as gender, ethnicity, and socioeconomic status affect students' STEM motivation and how are these factors addressed in STEM teaching?” How can we create a research design for the problem?

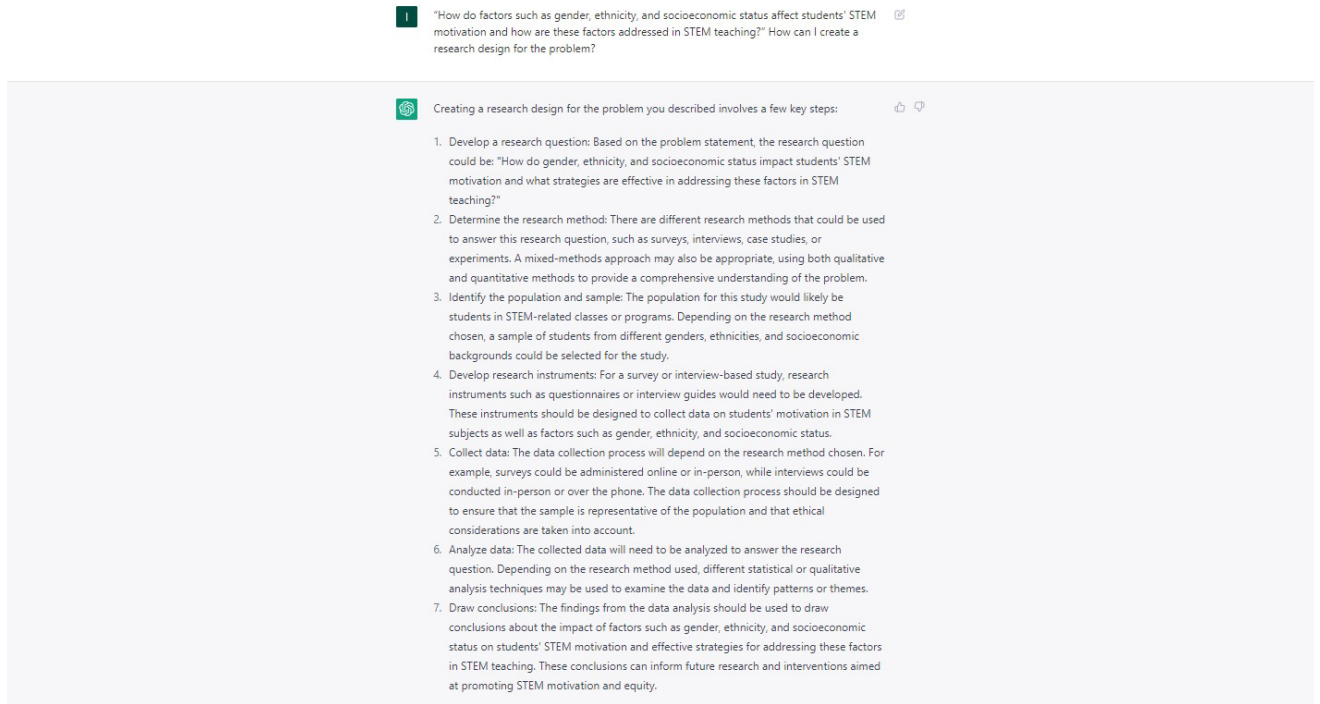


Fig 4. Research design

It lists the stages of scientific research. Asking questions, research method, sample and participant selection, data collection, data analysis, and determining the results are listed as. It also explains what we need to do in these steps. The content can be evaluated logically and systematically. We continue from the participant selection. We will do research for students. However, may we add teachers that it would be appropriate to include them? By posing another question, we are posing the question of “which student and teacher groups should I work with”.

1 With which class students does it make sense to conduct research? And should I collect data from all teachers?

The choice of which class of students to conduct research with depends on the specific research question and the population of interest. For example, if the research question is about the impact of gender on STEM motivation, it may be appropriate to select classes with an equal number of male and female students. Similarly, if the research question is about the impact of socioeconomic status, classes with a diverse range of socioeconomic backgrounds may be selected. It's important to consider the research question and the population of interest when selecting the class of students for the study.

As for collecting data from all teachers, this may not be necessary or feasible, depending on the research question and the scope of the study. If the research question is focused on the impact of teaching strategies on students' STEM motivation, it may be appropriate to collect data from a subset of teachers who are implementing those strategies. Alternatively, if the research question is focused on the impact of student demographics on STEM motivation, it may not be necessary to collect data from all teachers. Instead, data could be collected from a sample of teachers who are representative of the population of interest. The key is to ensure that the sample of teachers and students selected for the study is representative of the population and that the data collected is relevant to the research question.

Fig 5. Research design

At this step, It states that it is meaningless to do research with teachers since we mentioned the students in my research question. Seems totally logical and correct. ChatGPT gives general answers to every question. we need to customize it further with each step. Are there any resources we should refer to when determining the sample size?

1 Are there any resources I should refer to when determining sample size?

Yes, there are various resources that can be helpful when determining sample size for a research study. Here are a few suggestions:

1. Power analysis calculators: Power analysis calculators can help you determine the appropriate sample size for your study based on factors such as the effect size, level of significance, and power. Examples of online power analysis calculators include G\*Power (<https://gpower.software.informer.com/>) and Sample Size Calculator (<https://www.surveymonkey.com/mp/sample-size-calculator/>).
2. Sample size guidelines: Different research designs and methods may have different guidelines for sample size. For example, the American Psychological Association (APA) provides general guidelines for sample size in their publication manual. Consulting resources like this can be helpful in determining a starting point for sample size considerations.
3. Consulting with a statistician: A statistician can help guide you in determining an appropriate sample size based on the specific research question, study design, and statistical analysis plan. It can be helpful to consult with a statistician early on in the research planning process to ensure that sample size considerations are appropriately addressed.

Ultimately, sample size considerations will depend on various factors such as the research question, study design, and statistical analysis plan. By consulting these resources and working with experienced professionals, you can ensure that your sample size is appropriate and supports valid and reliable conclusions.

Fig 6. Sample size references

It makes three suggestions for sample size. It shows two sources for calculating the sample with two sources. Resources are running. In the second suggestion, it may be correct to reach the source through the APA. The third item suggests that I should consult a statistician. Is this

information reliable? Here, our expectation is that it will show an academic reference or not. This information is wrong, what are the sources? And are the sources reliable? I'm asking a new question. We want your sources and references.

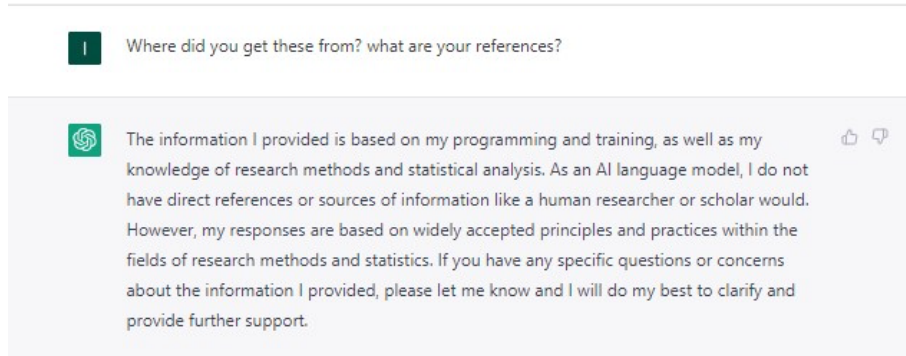


Table 7. Sources and References

It states that the texts it creates cannot provide direct references or sources of information as artificial intelligence would do a human researcher or academic. However, it states that we should trust answers. ChatGPT understands that we have expectations and concerns about this issue. At this point, it becomes clear that the AI application is using a language to relieve the user's concerns. Understanding that I cannot find any other information on this subject, I ask which surveys I should use.

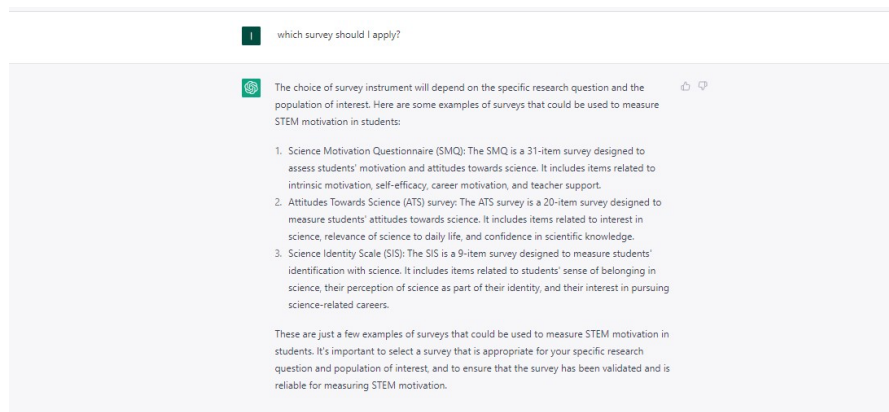


Table 8. Data collection tools

To measure STEM motivation according to the information it gives, STEM Interest Survey, The STEM Career Interest Survey recommends measurement tools such as the STEM Attitude Scale. If we have a chance to measure directly, of course we should prefer it. There are STEM motivation surveys in the literature. It states details such as how many items the questionnaire consists of. But for me the resource is important. We asked where we could find these scales.



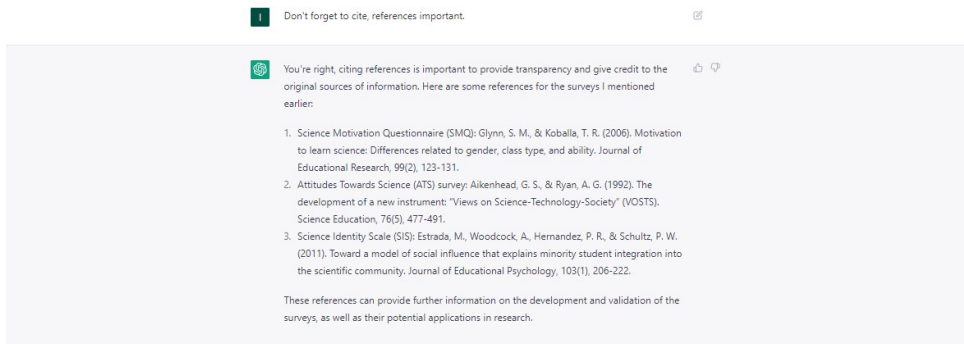


Table 9. Sources of data collection tools

We've been going to the sources and looking. The number one source is science motivation. In fact, the concept of STEM does not pass for the text. As the title suggests, it's an article about success and emotion at an early age. When I go to source number two, I find the article in "Google scholar". However, the source is not about motivation again. I'm going to the third source. However, it is not related to what I want. At this point, it does not meet my expectations. Also, the sources aren't very new, suggesting that either the language's scan dates are out of date or the AI hasn't been swayed. In this regard, it loses its reliability, which is perhaps one of the most important criteria in academic studies. I'm thinking of using different methods to use this program. Therefore, at this stage, ChatGPT should not be trusted in citing and writing literature. Otherwise, fake texts and research content will be supposedly trusted. Is this something that was done on purpose? we don't know yet. The next step is to review a published article before, "Dönmez, I., İdin, S., & Gürbüz, S. (2022). Determining Lower-Secondary Students' STEM Motivation: A Profile from Turkey. *Journal of Baltic Science Education*, 21(1), 38-51." we want to do a review on it.

### ***Revising the Title and Abstract***

I want something more creative in your title for the article. At this point, I expected ChatGPT to offer different examples. I decided to listen to some of his suggestions regarding an earlier published article using AI. Can you find a more creative and attractive title for the article "Determining Lower-Secondary Students' STEM Motivation: A Profile From Turkey"? I direct the question.

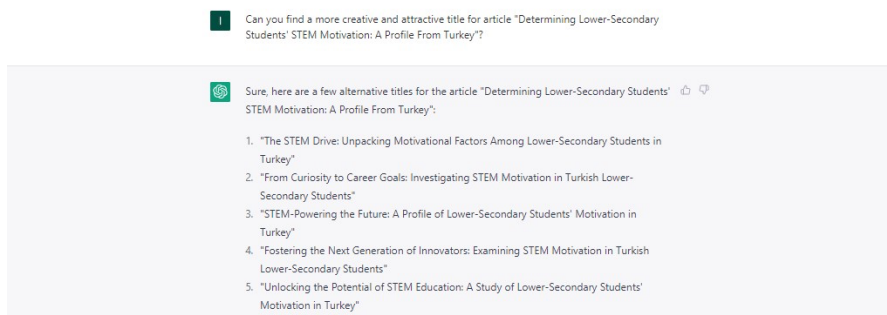


Fig 10. Finding a title

The titles look interesting, but they also bring some methodological errors. It makes a conceptual mistake in items 1. and 2. Other results are more acceptable. It explains the concepts with close concepts and lists the possibilities. It creates partially creative content. We also wonder if it got the titles it prepared from somewhere. We are going through a plagiarism program. The plagiarism program gives 0%. One of the difficulties encountered in academic journal applications is exceeding the word limits. It can be a difficult process to fit your content into certain word limits. We would like him to take the abstract of the same article and shorten it and present it in a more creative way.

1 Could you summarize the summary I shared with 100 words and present it in a more creative way? "The ongoing motivation of students toward the fields of science, technology, engineering, and mathematics (STEM) is a process related to their orientation toward those fields and their active participation in related activities. This study aimed to examine the ongoing STEM motivations of lower-secondary school students in a sample from Turkey according to demographic variables. In this study, in which 1926 students from 12 provinces participated, the students' STEM motivations were compared according to the variables of gender, place of residence, grade level, school type, "mothers' and fathers' education level, and family income level. T-tests and ANOVA testing were used in the analysis process via SPSS package program. As a result, male students' STEM motivations were found to be at higher levels than those of female students. The STEM motivations of gifted students were found to be higher than those of students studying in religiously oriented education institutions. The results also revealed that STEM motivation is directly proportional to the education level of the parents. The results obtained in this study, reflecting the general profile of STEM motivation among students in Turkey, may guide education makers, program developers, teachers, prospective teachers, and researchers on STEM education."

"STEM education is the key to unlocking a world of innovation, but how do we keep students motivated to pursue these fields? A recent study surveyed over 1,900 lower-secondary school students from 12 provinces in Turkey to examine their ongoing STEM motivations, comparing results based on gender, place of residence, grade level, school type, parental education level, and family income level. The study found that male students showed higher levels of STEM motivation than females, gifted students were more motivated than those in religious education institutions, and parental education levels had a direct impact on student motivation. These findings provide insight for educators, program developers, and researchers to improve STEM education for the next generation of innovators."

Fig 11. Revising the summary

When we give ChatGPT the content we had prepared and ask to shorten it, it offers a successful performance. It successfully summarizes the structure consisting of 196 words that we shared first, up to 76 words. It does not appear to have made any difference in meaning. It offers a successful performance in this regard. It is observed that it successfully reduces some points such as the analysis process

### *Analyzing and determining the results*

We have a number of data obtained from my research. Can it summarize what data might mean when we present an obtained table as text in ChatGPT? The same article can be considered as a positive outcome in confirming the accuracy of the content. We are trying to copy the analysis table that we obtained in the same article as the text. We want ChatGPT to interpret the findings.

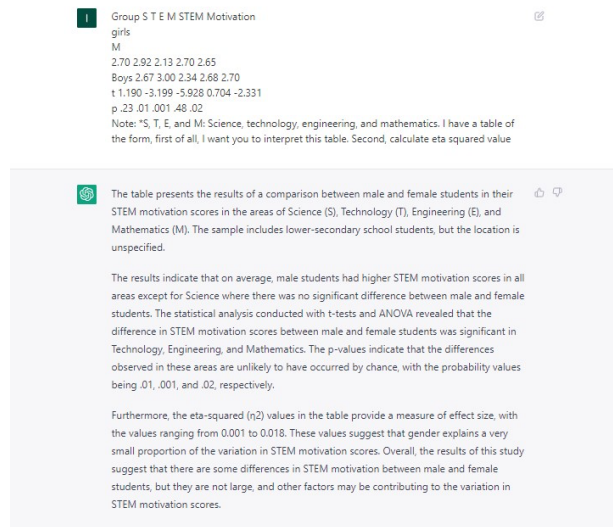


Fig 12. Examining the findings

Fig 12. shows the t-test results for girls and boys related to motivation in the fields of science, technology, engineering, and mathematics. There is a significant difference in favor of boys in the technology dimension ( $t = -3.199$ ,  $p < .05$ ,  $\eta^2 = 0.005$ ) and the engineering dimension ( $t = 5.928$ ,  $p < .05$ ,  $\eta^2 = 0.001$ ). There is no significant difference in the dimensions of science and mathematics. For overall STEM motivation ( $t = -2.331$ ,  $p < .05$ ,  $\eta^2 = 0.003$ ), there was again a significant difference in favor of male students. In Fig 12, S, T, E, and M are given to describe the dimensions of the STEM field (science, technology, engineering, and mathematics). Results appear to be in agreement with the results we found. Our questions are endless and we want to know if you can present this table graphically. However, none of the graphic links it offers are working every time. <https://i.ibb.co/gD5y0V0/STEM-Motivation-by-Gender-andSubcategory.png>

### ***Ethical Issues***

There seems to be a real risk that AI-generated articles will soon blur the distinction between original human-written and AI-written or re-edited content. The increase in such applications also raises the question of who the author really is. It opens the door to further discussions on validity and reliability. Predatory adds to the never-ending debate about publishing. Applications such as ChatZero, GPTGenerator, and GPTKit are thought to offer some solutions for understanding artificial intelligence-generated texts. We used the content ChatGPT writes through these applications and examine whether it is written with AI.

| Chatzero | After Deleted Text with Chatzero                      | AI Text Classifier   | After Deleted Text with AI Text Classifier | GPTKit | After Deleted Text with GPTKit |
|----------|---|--|--|--------|--------------------------------|
| Fig 1.   | Your text is likely to be written entirely by a human | No data  | No Data                                    |        |                                |
| Fig 2.   | Your text is likely to be written entirely by a human | The classifier considers the text to be likely AI-generated. | No Data                                    |        |                                |
| Fig 3.   | Your text may include parts written by AI             | No data  | No Data                                    |        |                                |
| Fig 4.   | Your text is likely to be written entirely by AI      | The classifier considers the text to be likely AI-generated. | No Data                                    |        |                                |
| Fig 5.   | Your text is likely to be written entirely by AI      | The classifier considers the text to be likely AI-generated. | No Data                                    |        |                                |
| Fig 6.   | Your text is likely to be written entirely by AI      | The classifier considers the text to be likely AI-generated. | No Data                                    |        |                                |
| Fig 7.   | Your text is likely to be written entirely by AI      | No data  | No data                                    |        |                                |
| Fig 8.   | Your text is likely to be written entirely by AI      | The classifier considers the text to be likely AI-generated. | No Data                                    |        |                                |
| Fig 9.   | Your text may include parts written by AI             | The classifier considers the text to be likely AI-generated. | No Data                                    |        |                                |
| Fig 10.  | Your text may include parts written by AI             | No data  | No Data                                    |        |                                |
| Fig 11.  | Your text is likely to be written entirely by AI      | No data  | No Data                                    |        |                                |
| Fig 12.  | Your text may include parts written by AI             | The classifier considers the text to be likely AI-generated. | No Data                                    |        |                                |

Fig 13. Scanning text written with Chatzero, AI Text Classifier and GPTKit We've been trying the content it has produced so far with Chatzero, AI Text Classifier, and GPTKit programs. We tried the outputs it gave me in Fig 13 with Chatzero, and we copy and paste the text. Chatzero states that it was written by artificial intelligence, except for figures 1. and 2. Fig 3., 9., 10. 12. states that some chapters are written with AI. It states that there is a good chance that 80% of it was written with AI. AI Text Classifier, on the other hand, states that only figures 6., and 9. are not written by AI. In other tables, it cannot reach any result because it is less than 1000 words. GPTKit, on the other hand, thinks that figures 3., 4., 5., 6., 7., and 8. may have been written by AI. Therefore, the findings show great diversity. I'm deleting the content from ChatGPT to test if ChatGPT is getting the text. ChatGPT changes its decision in three ways. AI Text Classifier differs in 4 ways. GPTKit, on the other hand, gives closer rates. Then we did an itenticate search. Naturally, since these texts are not published anywhere, they cannot get any rate. However, the findings are not sufficient in terms of whether the findings were written by AI.

### Review the bibliography

One of the most difficult issues in academic research is revising the bibliography. We asked ChatGPT to delete some references to the previously written article and convert it to APA 7 format. Can you revise the text below in accordance with APA 7? (I wrote "DELETED" where I omitted and removed those parts. Can't find the first source. The second source is not correct. The fourth source's name is incorrect. The fifth source is incorrect. The date indicated in the fifth source is incorrect. Cannot find the seventh source. The eighth source is incorrect. Therefore, it provides 40% accuracy out of the 10 references given. It's a low rate It's unknown where it got the data from or whether it was knowingly manipulating it, but what is known is that it's not reliable.

1 Can you revise the text below in accordance with APA 7? I couldn't find what it should be where it says DELETED. An, G., Wang, J., DELETED, Y., & Du, X. (2018). A study on the effects to students' STEM academic achievement with Chinese parents' participative styles in school education. *Educational Sciences: Theory & Practice*, 19(1), 41-54. DELETED  
 Biggs, J., & Tang, C. (2011). Teaching for quality learning at university. DELETED.  
 Bybee, R. W. (DELETED). The case for STEM education: Challenges and opportunities. NSTA press.  
 Chachashvili-Bolotin, S., Milner-Bolotin, M., & Lissitsa, S. (2016). DELETED. *International Journal of Science Education*, 38(3), 366-390.  
<https://doi.org/10.1080/09500693.2016.1143137>  
 Chavatzia, T. (DELETED). Cracking the code: Girls' and women's education in science, technology, engineering, and mathematics (STEM). DELETED.  
 Carrico, C., Murzi, H., & Matusovich, H. (2016, October). The roles of socializers in career choice decisions for high school students in rural central Appalachia: "Who's doing what?". In 2016 IEEE Frontiers in Education Conference (FIE) (pp. 1-6). IEEE.  
 Chadwell, B. (2016). The effect of socio-economic status and gender on high school student perceptions about career and technical education (Doctoral dissertation). DELETED.  
 Craig, C., Verma, R., Stokes, D., Evans, P., & Abrol, B. (2018). The influence of parents on undergraduate and graduate students' entering the STEM discipline sand STEM careers. DELETED, 40(6), 621- 643. <https://doi.org/10.1080/09500693.2018.1431853>  
 DELETED, J. (2003). Research design: Qualitative, quantitative and mixed methods approach (2nd ed.). SAGE Publications.  
 Ünü, Z. K., & Dökme, İ. (2020). Multivariate assessment of middle school students' interest in STEM career: A profile from Turkey. *Research in Science Education*, DELETED(3), 1217-1231. <https://doi.org/10.1007/s11165-018-9729-4>

2 Sure! Here's the revised text:

An, G., Wang, J., DELETED, Y., & Du, X. (2018). A study on the effects to students' STEM academic achievement with Chinese parents' participative styles in school education. *Educational Sciences: Theory & Practice*, 19(1), 41-54.

Biggs, J., & Tang, C. (2011). Teaching for quality learning at university. Open University Press.

Bybee, R. W. (2013). The case for STEM education: Challenges and opportunities. NSTA press.

Chachashvili-Bolotin, S., Milner-Bolotin, M., & Lissitsa, S. (2016). Engaging students in STEM education early: A closer look at STEM in early childhood education. *International Journal of Science Education*, 38(3), 366-390. <https://doi.org/10.1080/09500693.2016.1143137>

Chavatzia, T. (2020). Cracking the code: Girls' and women's education in science, technology, engineering, and mathematics (STEM). Routledge.

Carrico, C., Murzi, H., & Matusovich, H. (2016, October). The roles of socializers in career choice decisions for high school students in rural central Appalachia: "Who's doing what?". In 2016 IEEE Frontiers in Education Conference (FIE) (pp. 1-6). IEEE.

Chadwell, B. (2016). The effect of socio-economic status and gender on high school student perceptions about career and technical education (Doctoral dissertation). DELETED.

Craig, C., Verma, R., Stokes, D., Evans, P., & Abrol, B. (2018). The influence of parents on undergraduate and graduate students' entering the STEM disciplines and STEM careers. *Studies in Higher Education*, 43(6), 1036-1051. <https://doi.org/10.1080/03075079.2018.1462789>

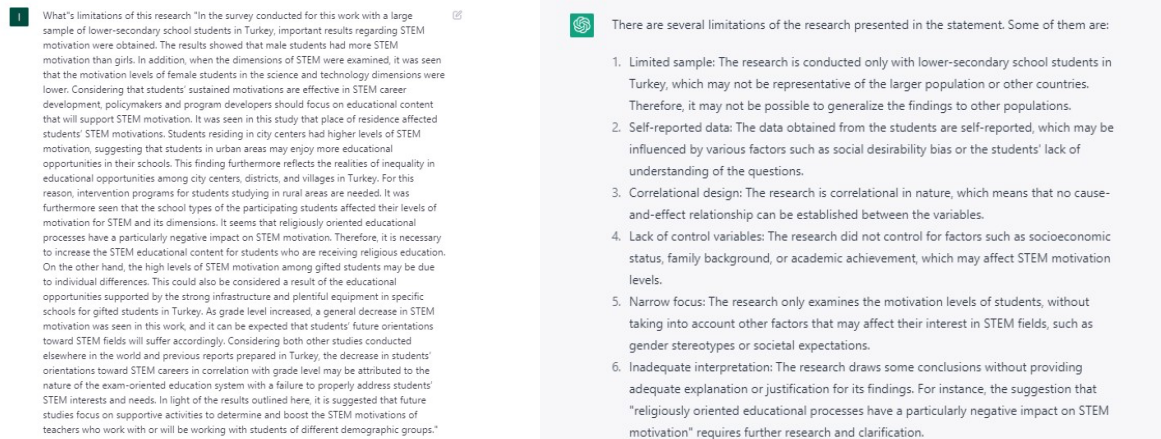
Creswell, J. W. (2003). Research design: Qualitative, quantitative and mixed methods approach (2nd ed.). SAGE Publications.

Ünlü, Z. K., & Dökme, İ. (2020). Multivariate assessment of middle school students' interest in STEM careers: A profile from Turkey. *Research in Science Education*, 50(3), 1217-1231. <https://doi.org/10.1007/s11165-018-9729-4>

Fig 14. Examination of the bibliography

## Writing limitations

I give the findings part of the research to ChatGPT. Then I want the research to identify its limitations.



The screenshot shows two columns of text. The left column contains a paragraph of text starting with "What's limitations of this research". The right column contains a list of six limitations of the research, such as "Limited sample", "Self-reported data", "Correlational design", "Lack of control variables", "Narrow focus", and "Inadequate interpretation".

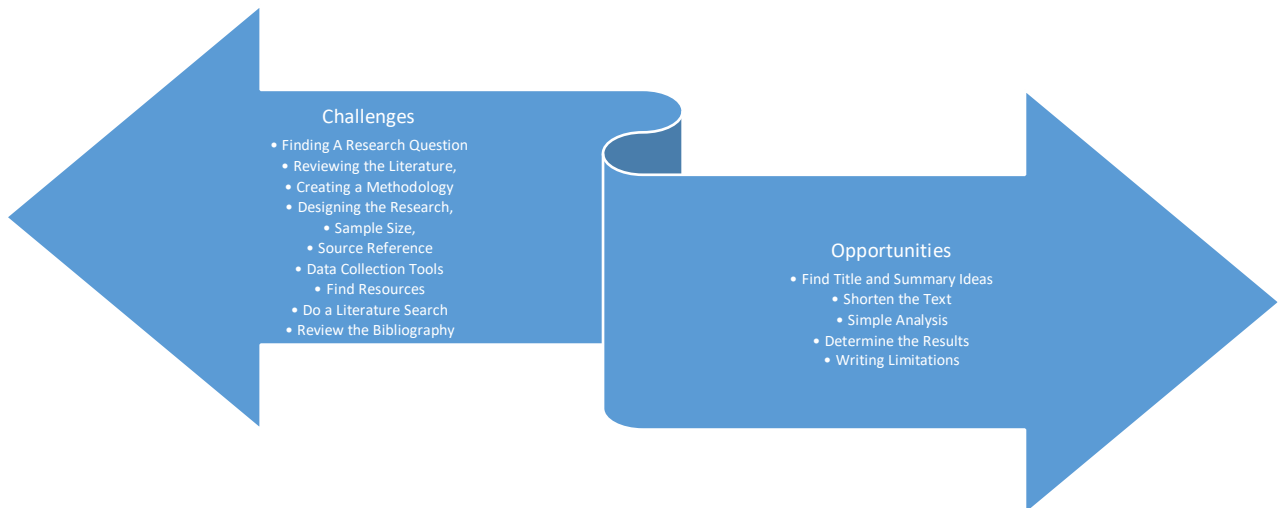
Fig 15. Examination of the limitations

ChatGPT examines the limitations of the research under 4 headings. It makes meaningful comments such as Limited sample, Self-reported data, Correlational design, Lack of control variables, Narrow focus, Inadequate interpretation. The content looks meaningful and creative.

## CONCLUSIONS AND DISCUSSIONS

The presented challenges and opportunities can contribute to future studies on the use and challenges of chatbots in scientific research and its broad implications. In summary, the technology always exists with its pros as well as its cons. Artificial intelligence technologies can provide some benefits for studies, such as finding title and summary ideas, shortening the text, identifying simple analyzes and conclusions, and writing limitations. Challenges; can be listed as finding a research question, examining the literature, finding a methodology, designing the research, sample size, source reference, data collection tools, finding sources, reviewing the literature, and reviewing the bibliography. Therefore, if you have some data, you can make some inferences from them. It can be said that it is successful in creativity. But it does not meet the expectations in terms of originality, the information it offers is mostly estimates obtained from "Google" search. One of the concerns is the accuracy of the predictions and predictions made by AI algorithms. While these algorithms can be very accurate, they are only as good as the data they are trained on. The conclusion is that it's not a good idea to have all or part of the research content written entirely on AI yet. Not to mention the ethical issues that come with it. Chatbots often fail to answer users' questions for three reasons: they have trouble with unknown concepts, out-of-field tasks, and misconduct (Lou et al., 2022). ChatGPT is like a smart friend, but you can't figure out when it's saying it right. It seems that more content is needed to explore how artificial intelligence technologies can contribute to scientific research. Additionally,

ChatGPT tends to follow instructions rather than have any real interaction. For example, when the information provided by users is insufficient, ChatGPT tends to make assumptions about what the user wants to hear rather than asking clarifying questions. As a result, Chatbots can have unintended consequences and become a double-edged sword (Shen et al., 2023). The findings overlap with the study of Aydın and Karaarslan (2022); see 'plagiarism' or inadequate paraphrasing. At this point, plagiarism programs should be reviewed and updated. It will cause consequences such as ethics and plunder in academic publishing. Writing the prepared text by AI will also bring a new problem in academic publishing. The software prepared for detecting the topics written by AI seems insufficient. It is important that human authors using artificial intelligence can be held accountable when the software fails or produces incorrect or incorrect output (Hosseini et al., 2023). Another problem is the reference to falsification of ideas, i.e. plagiarism, where ideas sought by AI-driven software like ChatGPT can benefit from others' ideas but is then used to claim originality. It is possible to reach different conclusions from the different experiences of different researchers. Maybe writing scientific papers is another smart activity where computers can learn to be better than humans. But ChatGPT is incapable of generating an article, was it intentionally created by the generators or coders? or future applications can be developed and trained to give better results. It is seen that different applications such as scite, elicite, bing, and concensus have recently been revised with academic concerns about the shortcomings of ChatGPT. Therefore, more research is needed on the advantages or disadvantages of other AI interfaces and other applications that can be used in scientific research.



## REFERENCES

- Alshater, M. (2022). Exploring the role of artificial intelligence in enhancing academic performance: A case study of ChatGPT. <http://dx.doi.org/10.2139/ssrn.4312358>
- Aydin, O., & Karaarslan, E. (2022). *OpenAI ChatGPT generated literature review: Digital twin in healthcare*. In Ö. Aydın (Ed.), *Emerging Computer Technologies 2* (pp. 22-31). Izmir Academy Association.
- Chatterjee, J., & Dethlefs, N. (2023). This new conversational AI model can be your friend, philosopher, and guide... and even your worst enemy. *Patterns*, 4(1), 1-3. <https://doi.org/10.1016/j.patter.2022.100676>
- Connelly, F. M., & Clandinin, D. J. (1990). Stories of experience and narrative inquiry. *Educational researcher*, 19 (5), 2-14. <https://doi.org/10.3102/0013189X019005002>
- Cresswell, J. W. (2003). *Research design: qualitative, quantitative and mixed methods approaches*. Sage Publications, Thousand Oaks.
- Deryugina, O. V. (2010). Chatterbots. *Scientific and Technical Information Processing*, 37, 143-147. <https://doi.org/10.3103/S0147688210020097>
- Dönmez, I., Idin, S., & Gürbüz, S. (2022). Determining lower-secondary students' STEM motivation: A profile from Turkey. *Journal of Baltic Science Education*, 21(1), 38-51.
- Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data—evolution, challenges and research agenda. *International journal of information management*, 48, 63-71. <https://doi.org/10.1016/j.ijinfomgt.2019.01.021>
- Fast, NJ, & Schroeder, J. (2020). Power and decision-making: New directions for research in the age of artificial intelligence. *Current opinion in psychology*, 33, 172-176. <https://doi.org/10.1016/j.copsyc.2019.07.039>
- Følstad, A., Araujo, T., Law, E. L. C., Brandtzaeg, P. B., Papadopoulos, S., Reis, L., ... & Luger, E. (2021). Future directions for chatbot research: an interdisciplinary research agenda. *Computing*, 103(12), 2915-2942. <https://doi.org/10.1007/s00607-021-01016-7>
- Gordijn, B., Have, HT (2023). ChatGPT: evolution or revolution?. *Med Health Care and Philos.* <https://doi.org/10.1007/s11019-023-10136-0>
- Green, L., Van Royen, K., & Vermeulen, A. (2019). Narrative research methods particularly focused upon digital technology use in everyday life. *Narratives in Research and Interventions on Cyberbullying among Young People*, 47-60.
- Grimaldi, G. & Ehrler, B. (2023). AI et al.: machines are about to change scientific publishing forever. *ACS Energy Lett.* 8, 878–880. <https://doi.org/10.1021/acscenergylett.2c02828>
- Hosseini, M., LM Rasmussen, and DB Resnik. (2023). Using AI to write scholarly publications. *Accountability in Research*. 1–9. <https://doi.org/10.1080/089896211.2023.2168535>



- Jiao, W., Wang, W., Huang, JT, Wang, X., & Tu, Z. (2023). Is ChatGPT a good translator? *A preliminary study*. <https://doi.org/10.48550/arXiv.2301.08745>
- Kim, J. H. (2016). *Understanding narrative inquiry: The crafting and analysis of stories as research*. California: Sage.
- Kirmani, A. R. (2022). Artificial intelligence-enabled science poetry, *ACS Energy Letters*, 8, 574-576. <https://doi.org/10.1021/acseenergylett.2c02758>
- Kramp, M. K. (2004). *Exploring life and experience through narrative inquiry*. In: de Marris, K., Lapan, S. (eds.) *Foundations of Research: Methods of Inquiry in Education and the Social Sciences*, pp. 103–122. Lawrence Erlbaum, NJ.
- Lieberman, M. (2023). What Is ChatGPT and how Is It used in education?. Education week. (Access 15 February 2023: at; <https://www.edweek.org/technology/what-is-chatgptand-how-is-it-used-in-education/2023/01>)
- Lucy, L. & Bamman, D. (2021). Gender and representation bias in GPT-3 generated stories , *Proceedings of the Workshop on Narrative Understanding*, 3, 48-55. <https://doi.org/10.18653/v1/2021.nuse-1.5> \_
- Luo, B., Lau, RY, Li, C., & Si , Y. W. (2022). A critical review of state-of-the-art chatbot designs and applications. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 12 (1), 1-12. <https://doi.org/10.1002/widm.1434>
- Maedche, A., Legner, C., Benlian, A., Berger, B., Gimpel, H., Hess, T., ... & Söllner, M. (2019). AI-based digital assistants: Opportunities, threats, and research perspectives. *Business & Information Systems Engineering*, 61, 535-544.
- Mollman, S. (2022, December 9). ChatGPT has gained 1 million followers in a single week. Here's why the AI chatbot is primed to disrupt search as we know it. *Yahoo Finance* . (Accessed 15 February 2023, at: [https://consent.yahoo.com/v2/collectConsent?sessionId=3\\_cc-session\\_2184255279ae-41b3-849d-bcfc80fc9db6](https://consent.yahoo.com/v2/collectConsent?sessionId=3_cc-session_2184255279ae-41b3-849d-bcfc80fc9db6))
- Nagarhalli, T. P., Vaze, V., Rana, N. (2020). *A review of current trends in the development of chatbot systems*. In *Proceedings of the 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India, 6–7 March 2020; pp. 706–710.
- Open AI, (2023) Openai blog chatgpt. (Accessed 1 February 2023, at: <https://openai.com/blog/chatgpt/>)
- Reddy, T. (2017). How chatbots can help reduce customer service costs by 30%. <https://www.ibm.com/blogs/watson/2017/10/how-chatbots-reduce-customer-servicecosts-by-30-percent>. Accessed 22 March 2019

- Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?. *Journal of Applied Learning and Teaching*, 6(1), 1-22. <https://doi.org/10.37074/jalt.2023.6.1.9>
- Shen, Y., Heacock, L., Elias, J., Hentel, K. D., Reig, B., Shih, G., & Moy, L. (2023). ChatGPT and other large language models are double-edged swords. *Radiology*, 230163. <https://doi.org/10.1148/radiol.230163>
- Smestad, T. L. (2018). *Personality Matters! Improving The User Experience of Chatbot Interfaces-Personality provides a stable pattern to guide the design and behavior of conversational agents* (Master's thesis, NTNU).
- Taecharungroj, V. (2023). “What Can ChatGPT Do?” analyzing early reactions to the innovative AI chatbot on Twitter. *Big Data and Cognitive Computing*, 7(1), 1-10. <https://doi.org/10.3390/bdcc7010035>
- Tlili, A., Shehata, B., Adarkwah, M. A. et al. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *SmartLearn. environ.* 10, 15. <https://doi.org/10.1186/s40561-023-00237-x>
- Uğuz Arsu, S. & Tekindal, M. (2021). Definition, scope and process of narrative research in qualitative research. *Beyond the Horizon Science Journal*, 21 (1), 85-124.
- Vallance, C. (2022). ChatGPT : New AI chatbot has everyone talking to it. (Accessed 23 December 2022; at; <https://www.bbc.com/news/technology-63861322>)
- Xu, L., Sanders, L., Li, K.; Chow, J.C. (2021). Chatbot for health care and oncology applications using artificial intelligence and machine learning: Systematic review. *JMIR Cancer*, 7 (4), 1-18. <https://doi.org/10.2196/27850>
- Zhai, X. (2022). ChatGPT user experience: Implications for education. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4312418>