



Investigating an Assessment Design that Prevents Students from Using ChatGPT as the Sole Basis to Pass Assessment at the Tertiary Level

Michael Stack ¹ 

¹ Department of Social Sciences and Commerce Education, University of the Free State, Bloemfontein, South Africa.

ABSTRACT

This paper investigated how to design undergraduate assessments that could not be adequately answered by ChatGPT. The paper was embedded in the revised Bloom's Taxonomy as a theoretical model. ChatGPT has presented significant challenges to lecturers when setting assessments at the tertiary level. There is enormous potential for students to attempt to use ChatGPT to write and pass assessments. Design Based Research formed the basis of this paper's research design. Qualitative research methods were used to interpret recordings of interaction with ChatGPT during Hermeneutic research cycles. The paper found that it was possible to design an assessment that could not be satisfactorily answered solely by ChatGPT. Interactions with ChatGPT were found to be an essential aspect of the research process. In light of the findings, efforts should be made to revise undergraduate assessments to encourage students to engage with them rather than copy and paste from ChatGPT. The paper recommended ChatGPT as a useful tool or form of interactive Google that can support writing assessment but is unable to replace the student effectively. The paper suggests that students should receive training on the limitations of ChatGPT and how they can use it more effectively in their studies.

Correspondence

Michael Stack

Email:

StackMDL@ufs.ac.za

Publication History

Received 25th August, 2023

Accepted 6th October, 2023

Published online:

30th November, 2023

Keywords: *Assessment Design, ChatGPT, Higher Education, Bloom's Taxonomy*

INTRODUCTION

There are concerns among academics that the recent proliferation of Chat Generative Pre-Trained Transformers (ChatGPT) is harming assessment. These concerns form part of an ongoing debate regarding the utility of ChatGPT in tertiary education. ChatGPT is an interactive language model that can imitate human writing styles.¹ The AI tool was released in November 2022 and further iterations have since been launched, namely, Chat GPT 3.5 and ChatGPT 4 on March 14, 2023. ChatGPT has presented higher education with both challenges and opportunities. For example, there is scope to use ChatGPT for more personalised learning. Assessment can be submitted to ChatGPT for more immediate feedback.² ChatGPT can also assist students and support the writing of assignments and research tasks.³ However, using ChatGPT is not without challenges. The free version of ChatGPT has a corpus of knowledge that is limited to September 2021. The lack of current data can affect the quality of responses. The main challenge at the tertiary level is that students may rely too heavily on ChatGPT. This can negatively impact critical thinking skills and reduce the quality of assessment.⁴ The ready availability

¹ Robert Dale, "GPT-3: What's It Good for?," *Natural Language Engineering* 27, no. 1 (2021): 113.

² Kevin Fuchs, "Exploring the Opportunities and Challenges of NLP Models in Higher Education: Is Chat GPT a Blessing or a Curse?," in *Frontiers in Education*, vol. 8 (Frontiers, 2023), 2.

³ Hesham Allam et al., "Artificial Intelligence in Education: An Argument of Chat-GPT Use in Education," in *2023 9th International Conference on Information Technology Trends (ITT)* (IEEE, 2023), 154.

⁴ Fuchs, "Exploring the Opportunities and Challenges of NLP Models in Higher Education: Is Chat GPT a Blessing or a Curse?" 2.

of a free version of ChatGPT has resulted in calls to either ban the AI tool or integrate it into teaching and learning in higher education.

Assessment design approaches before ChatGPT were susceptible to satisfactory imitations by ChatGPT. This has led either to lecturers integrating ChatGPT with assessment or to designing assessment that is ChatGPT 'proof'. ChatGPT has therefore created an urgent need to redesign assessment, particularly at the undergraduate level. A major concern is that ChatGPT is not picked up by plagiarism software. It, therefore, falls to the marker to scrutinise papers for AI assistance and provide for the burden of proof should a student be implicated. Tools have been developed to detect AI assistance; however, they are not foolproof and are not mainstream enough to be integrated into the university assessment infrastructure.

This study employed Bloom's taxonomy to design assessment questions that can be tested with ChatGPT in a sandbox with no live participants. The aim was to assess ChatGPT's capabilities and develop a question design that cannot be fully answered by ChatGPT. Questions that can be easily answered by ChatGPT facilitate extensive use of ChatGPT by students that do not promote teaching and learning whereas a more challenging question has the potential to make ChatGPT's role as a supporting tool more apparent. Without a ChatGPT policy in place and a lack of accurate AI detection tools, assessment design remains a vital avenue to engage students in viable assessment tasks. The research focused on the two research questions below.

- To what extent can assessment be rendered ChatGPT proof through the use of Bloom's taxonomy?
- How can context be used in assessment to challenge the use of ChatGPT by students?

LITERATURE REVIEW

Several studies have considered the issue of assessment in higher education in light of ChatGPT. Strzelecki explored students' use and acceptance of ChatGPT.⁵ He found that students readily adopted new forms of technology. His findings imply that there is social pressure to adopt ChatGPT and the development of university policies was considered. Chaudhry et al. conducted a quasi-experimental study that tested Chat GPT against a variety of courses with a student control group.⁶ Turnitin was found to be unable to detect ChatGPT and generally scored top or high marks for assessment.⁷ ChatGPT was not able to write assessments longer than 650 words.⁸ GPTZero and Copyleaks were found to be not fully effective in detecting AI-generated answers.⁹

Castillo et al, have further explored the impact of ChatGPT on digital education systems among Peruvian students.¹⁰ They found that ChatGPT encourages a high degree of laziness as it has a faster and higher word output than a student.¹¹ They called for a renewal of teaching and learning approaches. They argued that certain conditions would have to be met to use Chat GPT.¹² Crawford et al. posit that leadership is required for the ethical use of ChatGPT.¹³ Significant numbers of students are cheating and using ChatGPT. They argue that ChatGPT can be used to create and support written assessments.¹⁴ They call for ChatGPT to be embedded in subjects and courses. However, they contest that subjects and degrees should be future-proofed. AI cannot substitute learning through it can be used to facilitate it.¹⁵

An interdisciplinary and quasi-experimental study by Zhu et al explored the use of ChatGPT by undergraduates.¹⁶ They found that non-STEM students were more engaged with learning without using ChatGPT whereas STEM students have no major differences in learning with or without ChatGPT. Students demonstrated some awareness of the limitations of using ChatGPT.¹⁷ It was concluded that ChatGPT could revolutionise education; however, there were concerns regarding the impact of ChatGPT on students' self-

⁵ Artur Strzelecki, "To Use or Not to Use ChatGPT in Higher Education? A Study of Students' Acceptance and Use of Technology," *Interactive Learning Environments*, May 8, 2023, 10-11, <https://doi.org/10.1080/10494820.2023.2209881>.

⁶ Iffat Sabir Chaudhry et al., "Time to Revisit Existing Student's Performance Evaluation Approach in Higher Education Sector in a New Era of ChatGPT—A Case Study," *Cogent Education* 10, no. 1 (2023):7.

⁷ Chaudhry et al., "Time to Revisit Existing Student's Performance Evaluation Approach in Higher Education Sector ...," 20.

⁸ Chaudhry et al., "Time to Revisit Existing Student's Performance Evaluation Approach in Higher Education Sector ...," 23.

⁹ Chaudhry et al., "Time to Revisit Existing Student's Performance Evaluation Approach in Higher Education Sector ...," 24.

¹⁰ Alejandro Guadalupe Rincón Castillo et al., "Effect of Chat GPT on the Digitized Learning Process of University Students," *Journal of Namibian Studies: History Politics Culture* 33 (2023): 1–15.

¹¹ Castillo et al., "Effect of Chat GPT on the Digitized Learning Process of University Students," 11.

¹² Castillo et al., "Effect of Chat GPT on the Digitized Learning Process of University Students," 13.

¹³ Joseph Crawford, Michael Cowling, and Kelly-Ann Allen, "Leadership Is Needed for Ethical ChatGPT: Character, Assessment, and Learning Using Artificial Intelligence (AI)," *Journal of University Teaching & Learning Practice* 20, no. 3 (2023): 6.

¹⁴ Crawford, Cowling, and Allen, "Leadership Is Needed for Ethical ChatGPT...," 8.

¹⁵ Crawford, Cowling, and Allen, "Leadership Is Needed for Ethical ChatGPT...," 11.

¹⁶ Gaoxia Zhu et al., "Embrace Opportunities and Face Challenges: Using ChatGPT in Undergraduate Students' Collaborative Interdisciplinary Learning," *ArXiv Preprint ArXiv:2305.18616*, 2023.

¹⁷ Zhu et al., "Embrace Opportunities and Face Challenges...," 25.

discipline and critical thinking.¹⁸ Rudolph et al also conducted a desktop analysis into the effect of ChatGPT on traditional assessment in higher education.¹⁹ Face-to-face oral exams were considered a solution as well as handwritten tests. Although others argued against policing ChatGPT. The main recommendation was the design of an assessment that ChatGPT would not be able to complete successfully.

This paper takes up this recommendation and explores the designing and testing of written assessments at the tertiary level with a dialogic engagement with ChatGPT. This study is unique in that the assessments were designed using current literature and input from ChatGPT to test ChatGPT's capabilities in answering senior undergraduate assessments. It forms part of an ongoing project of designing an assessment that is ChatGPT 'proof' and encourages students to critically apply skills in a way that ChatGPT is unable to assist beyond a limited starting point. Bloom's taxonomy forms a crucial part of the design process and provides a framework for assessing ChatGPT's capabilities in answering higher and lower-order questions.

THEORETICAL FRAMEWORK

The theoretical framework of this study is Bloom's taxonomy. The theory was developed by Dr. Benjamin Bloom and his colleagues in 1956. The theory was revised in 2001 by Education and Cognitive psychologists.²⁰ The *raison d'être* for Bloom's taxonomy was a document that could be used to improve communication between Educators. In Bloom's taxonomy, learning is divided into cognitive, psychomotor and affective domains. The revision in 2001 into a tabular format made use of the knowledge dimension and cognitive process dimension.²¹ Bloom's taxonomy's main assumption is that teaching and assessment should progress from lower-order domains to higher-order domains. This theory's relevance to two-fold. Firstly, Bloom's taxonomy can be applied to assessment in order to foster higher levels of thinking and critical thinking skills required as outcomes for higher education modules. Secondly, is a framework that can assist in the design of questions that present a significant challenge to the overreliance on ChatGPT.

This paper's focus is on the design and the testing of assessment questions. This theory is an ideal vehicle for the assessment design process and will also allow the study to better situate feedback from ChatGPT. Therefore, the theory will assist in the development of questions during the research process. This theory is of paramount importance to the study in terms of grading the level of questions designed to test ChatGPT.

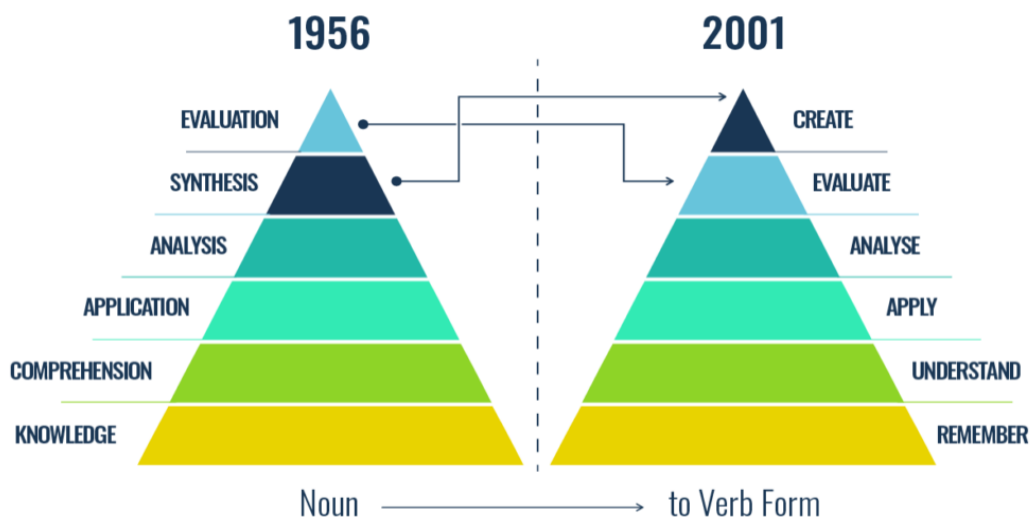


Figure 1. Transition from the 1956 Bloom's Taxonomy to the 2001 revised Taxonomy

¹⁸ Zhu et al., "Embrace Opportunities and Face Challenges...", 28.

¹⁹ Jürgen Rudolph, Samson Tan, and Shannon Tan, "ChatGPT: Bullshit Spewer or the End of Traditional Assessments in Higher Education?," *Journal of Applied Learning and Teaching* 6, no. 1 (2023).

²⁰ Rani Gul, Shazia Kanwal, and Sadia Suleman Khan, "Preferences of the Teachers in Employing Revised Blooms Taxonomy in Their Instructions," *Sjesr* 3, no. 2 (2020): 258.

²¹ Sónia Rolland Sobral, "Bloom's Taxonomy to Improve Teaching-Learning in Introduction to Programming," *International Journal of Information and Education Technology* 11, no. 3 (2021): 148.

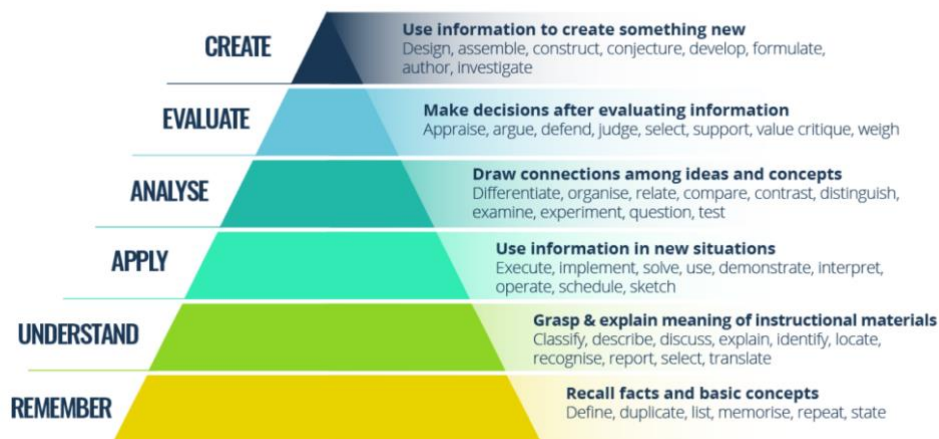


Figure 2. Definition of Levels for Bloom's Taxonomy 2001

METHODOLOGY

This paper was situated within a constructivist paradigm whereby theory is substantiated by the data drawn from research. The research derives meaning from the data through processes from working with the data.²² The interactive process is primarily between the researcher and ChatGPT. The research design for this study is Design Based Research (DBR). DBR involves identifying a problem, designing a solution, testing it and then evaluating the test. The research goes through a number of hermeneutic cycles before the study is completed.²³ This paper makes use of Qualitative research methods where the researcher is considered the main instrument. The research method relies primarily on text and image data.²⁴ Purposeful non-probability sampling was employed to select a single research participant, ChatGPT 3.5. This technique allows the researcher's preference to guide sampling selection. Purposeful sampling involves the selection of research samples that are deemed by the researcher to fulfil the research objectives.²⁵ Chat GPT3.5 was specifically selected as it is a form of ChatGPT that is freely available on Google. This is considered a version that is easy and free to access by undergraduates. More advanced forms of ChatGPT were not considered due to their varying paywalls.

ChatGPT was explored by the researcher for the design of the assessment questions that could challenge ChatGPT. Assessment questions were entered into ChatGPT 3.5's chat box. The answers were saved on the ChatGPT platform and PDF copies were saved as a form of data backup. Additional prompts were entered into ChatGPT and were saved similarly. Even though there were no live participants in the study, an ethical issue that was considered was whether Open AI has made use of unethical labour practices. This was considered to remove toxic material from the body of knowledge. OpenAI has not been fully transparent regarding the human labour required to maintain ChatGPT's image as intelligent.²⁶ The assessments will be replaced with Assessments A, B, and C rather than indicating their full module details. The school context in Assessment C will be referred to as Primary School A.

Inductive thematic analysis was used to code the data. Themes were allowed to come from the data rather than using preexisting themes from the research literature. To promote validity this research used a rich description in order to communicate the findings. It aims to communicate negative information that may contrast

²² Charles Kivunja and Ahmed Bawa Kuyini, "Understanding and Applying Research Paradigms in Educational Contexts.," *International Journal of Higher Education* 6, no. 5 (2017): 33.

²³ Emily E. Scott, Mary Pat Wenderoth, and Jennifer H. Doherty, "Design-Based Research: A Methodology to Extend and Enrich Biology Education Research," *CBE—Life Sciences Education* 19, no. 3 (September 2020): es11, <https://doi.org/10.1187/cbe.19-11-0245>.

²⁴ John W Creswell and J David Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (Sage publications, 2017), 295.

²⁵ Ezezi Isaac Obilor, "Convenience and Purposive Sampling Techniques: Are They the Same," *International Journal of Innovative Social & Science Education Research* 11, no. 1 (2023): 3-4.

²⁶ Konstantinos Konstantis et al., "Ethical Considerations in Working with ChatGPT on a Questionnaire about the Future of Work with ChatGPT," *AI and Ethics*, June 20, 2023,3. <https://doi.org/10.1007/s43681-023-00312-6>.

with the themes of the research.²⁷ To promote reliability, the PDF soft copies were compared with the online saved copies to avoid errors. An effort was made to prevent the definition of codes from drifting.²⁸

FINDINGS AND DISCUSSION

Assessments were designed using a variety of levels from the 2001 revised Bloom's taxonomy. Questions ranged from Remembering to Evaluate to assess ChatGPT's capabilities across all six cognitive levels. A number of the questions were designed in conjunction with context. ChatGPT was able to produce a series of guidelines for designing assessments. Below are the guidelines most relevant to written assessment prompted from the Chat GPT 3.5.

Prompt: How do I make it challenging for university students to use ChatGPT to write an assessment?

1. **Complex Questions:** Frame the assessment questions in a way that requires students to analyse, synthesise, and apply knowledge from multiple sources. These questions should go beyond simple factual recall and encourage critical thinking.
2. **Application-Oriented Questions:** Craft questions that ask students to apply concepts to real-world scenarios or case studies. This will force them to demonstrate their ability to use the knowledge they have learned in practical situations.
3. **Originality and Creativity:** Ask students to provide original responses or creative solutions that go beyond what the AI model could generate. For example, require them to propose innovative ideas or approaches to a problem.
4. **Open-Ended Questions:** Frame questions that have multiple valid answers, promoting different perspectives and interpretations. This will challenge students to articulate their own ideas rather than relying on the AI model for a "correct" response.²⁹

This was followed up by the prompt: How can I design an essay question that cannot be answered by ChatGPT? Rather than giving a list of guidelines, ChatGPT provided an example of a complex question and the reasons for the design of the question.

Designing an essay question that cannot be answered solely by ChatGPT requires combining elements that challenge the AI model's limitations and demand higher-order thinking skills from the students. Here's an example of such a question:

Essay Question:

Topic: The Impacts of Ethical Dilemmas on Autonomous Vehicles Adoption

Prompt: Autonomous vehicles are emerging as a transformative technology with the potential to revolutionize transportation systems worldwide. However, the widespread adoption of autonomous vehicles comes with ethical challenges that must be carefully considered. In an essay, discuss the various ethical dilemmas associated with the deployment of autonomous vehicles and their potential impacts on society, economy, and the environment. Consider the roles of stakeholders, such as manufacturers, governments, consumers, and pedestrians, in addressing these dilemmas and fostering responsible autonomous vehicle adoption. Provide evidence from scholarly articles, real-world case studies, and ethical frameworks to support arguments.

Reasoning Behind the Challenge:

1. **Real-World Context:** The essay question requires students to delve into the current ethical challenges related to autonomous vehicles, which may involve recent developments and debates that ChatGPT might not be up-to-date with.
2. **Integration of Multiple Sources:** Students are expected to use a wide range of academic sources, real-world case studies and ethical frameworks to support their arguments. This demands a deeper understanding of the topic than ChatGPT can provide, as it lacks access to the internet and up-to-date scholarly databases.

²⁷ Creswell and Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 314-315.

²⁸ Creswell and Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 316.

²⁹ OpenAI, ChatGPT to Michael Stack, Output, July 2023.

3. **Critical Analysis and Synthesis:** The question necessitates critical analysis and synthesis of information from various sources, which involves forming coherent and well-structured arguments that build upon existing research. ChatGPT's responses might lack the depth and complexity required to address these aspects effectively.
4. **Subject Expertise:** To answer the question comprehensively, students need a deep understanding of autonomous vehicles, ethical dilemmas, and the complexities of societal implications. While ChatGPT can provide general information, it may lack the expertise and contextual understanding necessary to address this multidimensional topic thoroughly.
5. **Originality and Creativity:** Students are encouraged to present original insights and ideas beyond what ChatGPT can offer. This includes considering unique perspectives, proposing novel solutions, and discussing implications not readily available in the model's responses.³⁰

The point of most interest to this research was the use of real-world contexts. ChatGPT was prompted successively to produce a series of similar answers. A further prompt was included: "What are the limits of your corpus of knowledge? This line of questioning was aimed at working out what information had been input into the corpus of knowledge.

1. **Knowledge Cutoff:** My training data only includes information up until September 2021. This means I do not have access to any events, developments, or research that occurred after that date.
2. **Dynamic and Evolving World:** The world is constantly changing, and new information is continually being generated. I do not have access to real-time data or updates, so I might not be aware of the latest news, discoveries, or advancements.
3. **Lack of Internet Access:** I do not have direct access to the Internet, databases, or any external sources. This means I cannot browse the web or refer to information beyond what I was trained on.
4. **Subjective and Biased Information:** My training data includes text written by various individuals with different perspectives and biases. As a result, my responses might inadvertently reflect some of those biases.
5. **Inability to Form New Information:** While I can generate responses based on patterns in my training data, I cannot create new information or perform independent reasoning. My responses are limited to what I have learned from the data.
6. **Specialized Knowledge:** While I have knowledge across a broad range of topics, my understanding might not be as deep or specialized as that of domain experts in specific fields.
7. **Sensitive Information:** I am designed to follow ethical guidelines and avoid providing harmful or sensitive information. Therefore, I may not be able to answer certain questions that involve confidential, harmful, or inappropriate content.
8. **Ambiguity and Context:** Like any language model, I am not immune to misunderstandings or misinterpretations. My responses are based on patterns in the data and may not always grasp the full context or nuances of a question.
9. **Language Limitations:** My primary mode of communication is text, and I may not have access to visual or auditory information, which can sometimes limit my ability to respond appropriately to certain types of questions.
10. **Legal and Ethical Guidelines:** I am designed to comply with OpenAI's guidelines, which means I may refuse certain requests or avoid providing responses that could be harmful or violate ethical principles.³¹

It was apparent that the major design characteristic of ChatGPT 3.5 is that its corpus of knowledge only went up to September and could only access data that it was programmed with.³² This indicated that contexts stored online, may not necessarily be part of the corpus of knowledge. Therefore, through testing, the limits of the

³⁰ OpenAI, ChatGPT to Michael Stack, Output, July 2023.

³¹ OpenAI, ChatGPT to Michael Stack, Output, July 2023.

³² David Baidoo-Anu and Leticia Owusu Ansah, "Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning," *Journal of AI* 7, no. 1 (2023): 52–62.

stored knowledge could also be tested. It was decided to incorporate context in several different levels and compare the efficacy of the design process rather than focus context only on the higher levels of Bloom's taxonomy. Jagdishbhai and Thakkar indicate that ChatGPT struggles to understand and engage accurately with the context of a question.³³

Test Discussion

Three assessments each of which had a component that relied on personal experience were tested on ChatGPT. The questions below focused mainly on Remember and Comprehension according to Bloom's Taxonomy. Some examples from Assessment A are as follows:

Question 1

Describe how you were taught history in high school.

Question 2

How were you taught history at university before you started the history education module for 3rd year?

Question 3

Compare the methods and approaches of your school history educators and your university history lecturers.

Question 4

How well do think that your school history education and two years of academic history training before 3rd year have prepared you to be a history teacher? (Assessment A)

When these questions were used as direct prompts, ChatGPT was only able to give broad and general information in the third person. The questions require students to recall their personal experiences as learners and as students. This would require the first person and a more familiar tone. This personal context had been suggested by Rudolph et al. as a method for preventing ChatGPT use in written assessments.³⁴

Prompt: Describe how you were taught history in high school.

I don't have personal experiences or consciousness, but I can provide you with a general idea of how history is typically taught in high schools.

In high schools, history is usually taught through a combination of lectures, textbooks, discussions, projects, and multimedia resources. The exact approach can vary depending on the school, the teacher's style, and the curriculum in place. Here's a typical overview of how history might be taught in high school:

Chronological Framework: History is often presented in a chronological framework, starting from ancient times and progressing through various historical periods to the present day. This helps students understand the development of civilizations, cultures, and societies over time...³⁵

This answer does not contain references to personal experiences and would not be sufficient for a passing grade for the question. However, ChatGPT is capable of writing in the first person and imitating experience if requested. The change in style can be observed from the next prompt.

Ah, high school history classes, what a journey those were! Looking back, I can't help but feel a mix of nostalgia and appreciation for the way those classes shaped my understanding of the world. While some days were a bit of a slog, overall, I think my history teachers did a commendable job of making the past come alive.

I remember vividly the excitement of walking into the classroom, wondering what era we would be diving into that day. From ancient civilizations like Egypt and Greece to more recent events like World War II,

³³ K Thakkar and N Jagdishbhai, "Exploring the Capabilities and Limitations of GPT and Chat GPT in Natural Language Processing," *J. Manag. Res. Anal.*, 2023, 19.

³⁴ Rudolph, Tan, and Tan, "ChatGPT: Bullshit Spewer or the End of Traditional Assessments in Higher Education?" 355.

³⁵ OpenAI, ChatGPT to Michael Stack, Output, August 2023.

each unit felt like stepping into a time machine that transported us to a different world. The textbooks might have been a bit heavy, but they were like treasure troves of stories waiting to be uncovered...³⁶

While this style is more convincing, it is only a plausible answer on a superficial level. AI tool usage could be suspected as Greece is not part of the CAPS history curriculum and Egypt would fall under the intermediate phase at the primary level rather than high school. (DOE, 2011: 17) Cross-referencing the CAPS curriculum and knowledge of the South African education system would demonstrate the superficial nature of the response. Direct prompts clearly do not elicit answers that fully engage with the question while personal experience is elicited with less direct prompting. Therefore, lower-order questions linked to context can be answered in what appears to be a convincing manner.

Question 5

You will complete your BEd at the end of 2024. How will you teach history in your classroom?
(Assessment A)

Question 5 Assessment A required a discussion of teaching approaches that included less formal material. ChatGPT provided a variety of plausible teaching methods that would have required some editing to present as a paragraph answer. This question was designed as a Create-level question where a new topic was to be constructed briefly for inclusion in the South African CAPS History Curriculum.

Question 6

If you could choose any topic to include in the history curriculum at the school level, which topic would that be and why would this topic be important to teach in a South African history classroom?
(Assessment A)

With a direct prompt, ChatGPT generally selected topics that were already present in the CAPS curriculum document for history. The question aimed to elicit a topic that did not feature any phase of CAPS history. With a more indirect prompt ChatGPT was able to consider non-South African topics. However, the Silk Road features in several topics in CAPS such as: “Explores from Europe find Southern Africa” in Grade 6,³⁷ “The Kingdom of Mali and the City of Timbuktu” in Grade 7,³⁸ and “The Ming Dynasty” in Grade 10.³⁹

Prompt: Choose a history topic not already in the CAPS history document to be included.

ChatGPT

Topic: The Silk Road

The Silk Road was a network of interconnected trade routes that spanned thousands of miles across Asia, connecting the East to the West. It played a significant role in the exchange of goods, ideas, cultures, and technologies between the regions it linked. The Silk Road was not a single road but rather a complex web of routes that facilitated trade.

This indicates that ChatGPT 3.5 has not been specifically trained with Curriculum Statements from the Department of Education for any of the 2 phases of Social Sciences or the FET phase for history.

Question 7

Make a balanced argument for compulsory history education for FET history grades 10 to 12 in South Africa that considers the benefits and challenges of implementing this reform. (Assessment A)

Question 7 is structured as a longer essay question and was designed as an evaluation question. The question is linked to a South African context and debate. ChatGPT was able to compile a balanced argument with specific reference to benefits and challenges. Clearly, ChatGPT can compile an Evaluation level answer with contextual information that it has been trained.

³⁶ OpenAI, ChatGPT to Michael Stack, Output, August 2023.

³⁷ Department of Basic Education, *CAPS Social Sciences Grade 4-6* (South Africa: Department of Education, 2011), 43.

³⁸ Department of Basic Education, *CAPS Social Sciences Grade 7-9* (South Africa: Department of Education, 2011), 33.

³⁹ Department of Basic Education, *CAPS History Grade 10 - 12* (South Africa: Department of Education, 2011).

Assessment A comprised the first test. Assessment B was adjusted to include a final more complex higher-level question.

Question 1:

Comparatively, discuss the teaching approaches you experienced in Social Sciences at school while you were a learner with the teaching approaches you have experienced in your 2nd-year content modules for the Social Science Intermediate Phase. (Assessment B)

This question was designed as a comprehension question linked to the lived experience of the student. ChatGPT was only able to give general information that was clearly unrelated to personal experiences. With a more indirect prompt requesting ChatGPT to pretend to be a student, it was able to give a more believable and personal response. It did, however, refer to learners as students. This is not the correct terminology in the South African Education system.

Question 2:

Comment on how you have observed the use of PowerPoint and board presentations at the school and university level and whether this was effective. Consider how digital resources can be used in university lecture theatres and the adjustments that would have to be made for intermediate-phase learners in using those resources. For example: YouTube. (Assessment B)

For question 2, ChatGPT provided a well-balanced response with a little caveat in that it referred to middle school and junior high school. These distinctions are not a feature of South African state schools.

Question 3:

Evaluate how the aims and objectives of Social Sciences in conjunction with the conceptualisation of the discipline can assist a Social Science educator in understanding their own context in relation to the context of their learners in the classroom. Refer to the diverse contexts that can be found in 21st-century South African Classrooms. Comment on how your understanding of Social Sciences as cultural transmission, value analysis, personal development, social criticism and action training will impact your approach to understanding context. Take a position with respect to the value of context to a Social Science Education in the classroom. (Assessment B)

ChatGPT answered this complex question with relative ease. It was interesting to note that the answer included references to learners rather than students. Framing the South African context clearly in the question removed a common error in previous answers. ChatGPT is unable to deduce context that is not input into prompts. Up to this point ChatGPT had dealt with contexts that it could either imitate or were evidently included in its corpus of knowledge. The final assessment, assessment C consisted of a question that drew on a personal context while the second question involved a specially selected context.

Question 1

Comparatively, discuss the teaching approaches you experienced as a learner during Social Science Lessons with those you experienced in your 2nd Year Social Science content modules and 3rd Year Social Science Didactic module. Comment on the success of these approaches in terms of assessment as evidence of learning. Consider examples of factors that may impact the practical implementation of teaching approaches that are theoretically sound. (Assessment C)

Again, direct prompts lacked the first person and personal experiences. Prompting ChatGPT to pretend to be a student resulted in a more reasonable answer. However, the response was of limited length.

Question 2:

Critically evaluate how you would implement the Social Sciences Intermediate CAPS document in your classroom in the context of the school listed below. Your paper should address how you would adapt planning, assessment and teaching approaches in order to ensure that learning could take place for all learners in your classroom. Your classroom is taken to be a grade 5 class. Please follow the links below in order to explore the context of the school.

In terms of how you adapt planning, assessment and teaching approaches. This is not a theoretical exercise. You will need to refer to the context of the school and support how you will adapt according to your interpretation of the context.

Primary School A (Assessment C)

Question 2 was designed as an evaluation-level question linked to a context that had a minimal internet footprint. The context was Primary School A in the Western Cape which was unlikely to be part of students' lived experiences. ChatGPT could be prompted to confirm that it had no information on the school.

I'm sorry, but I don't have specific information about "Primary School A" in my current database. It's possible that the school is a smaller or less well-known institution, or that it has gained prominence or relevance after my last update in September 2021. If you could provide more context or details about what you're looking for, I'd be happy to try to assist you further.⁴⁰

The aim of this question was whether ChatGPT would be able to compile answers for higher-order questions in conjunction with a context that was not part of the corpus of knowledge. The first part of the response is as follows.

I'm sorry, but I don't have access to external content such as specific school websites or documents. Therefore, I am unable to directly access the Social Sciences Intermediate CAPS document or the context of Primary School A that you mentioned. However, I can provide you with a general framework on how to adapt planning, assessment, and teaching approaches for a grade 5 Social Sciences class, taking into consideration the context of the school.⁴¹

What follows is a breakdown of tools that ChatGPT is unable to apply. There is no evidence of any application in the context of Primary School A. Various teaching approaches were recommended and ChatGPT referred to students rather than learners. The answer was general and did not show direct engagement with the question. A human, however, would be able to answer the question by employing research skills in order to collate the small quantity of publicly available data on the internet. It was deemed sufficient information to attempt to answer the question. ChatGPT while not able to answer this question, could perhaps provide a reasonable starting point. However, the first prompt may not have sufficient material for a long essay question. Alternative prompts that broke the question into segments could elicit more material.

ChatGPT indicated that specific formatting and references could be used to challenge ChatGPT use.⁴² References are an important aspect of undergraduate assessment. A more direct prompt was used and Chat GPT responded below:

I'm sorry for any confusion, but I must clarify that I don't have direct access to external sources or the ability to provide citations as a machine learning model. My responses are generated based on the information available up until my last update in September 2021, and I don't have real-time access to the internet or databases to fetch specific references.⁴³

This result concurs with Rudolph et al's assertion that ChatGPT is incapable of writing academic references.⁴⁴ ChatGPT was then prompted to rewrite its answer as an academic paper. It started with its limitations.

I understand your request, and I'll provide an example of how the information could be structured in an academic paper format with Harvard referencing. Please note that the references provided are placeholders and should be replaced with actual sources you've consulted.⁴⁵

The list of references below seemed convincing and well-formatted but are all spurious inventions on the part of ChatGPT.

Brown, A., & Green, B. (2017). Community engagement projects in South African classrooms. *Journal of Social Science Education*, 14(2), 45-61.

⁴⁰ OpenAI, ChatGPT to Michael Stack, Output, August 2023.

⁴¹ OpenAI, ChatGPT to Michael Stack, Output, August 2023.

⁴² OpenAI, ChatGPT to Michael Stack, Output, July 2023.

⁴³ OpenAI, ChatGPT to Michael Stack, Output, August 2023.

⁴⁴ Rudolph, Tan, and Tan, "ChatGPT: Bullshit Spewer or the End of Traditional Assessments in Higher Education?" 355.

⁴⁵ OpenAI, ChatGPT to Michael Stack, Output, August 2023.

Johnson, L. M. (2015). Value analysis and identity discourse in South African education. *International Journal of Educational Research*, 72, 58-67.

Jones, R. K. (2013). Social criticism in the South African classroom: Addressing inequalities through education. *Comparative Education*, 49(4), 512-530.

Smith, J. W. (2010). Cultural transmission in diverse classrooms: Strategies for South African educators. *South African Journal of Education*, 30(2), 221-236.

Williams, S. A. (2018). Personal development and citizenship education in South Africa. *Journal of Curriculum Studies*, 50(3), 325-342.

Williams, S. A., & Lee, M. J. (2019). Embracing context in South African Social Science Education. *Teaching and Teacher Education*, 80, 40-49.

There is a serious limitation of ChatGPT in its inability to provide authentic academic sources which is in line with Day's assessment of this issue.⁴⁶ However, the references appear credible and need to be confirmed one by one. The risk is that students may try to employ ChatGPT references based on their appearance rather than their substance. That said, it is considered another method, if not time consuming, to investigate cases of ChatGPT usage. According to Day, this could be challenging to pick up as ChatGPT often uses names of real authors and journals in order to compile a fake reference.⁴⁷

DISCUSSION

It is clear that Bloom's taxonomy is not sufficient for the task of rendering assessment proof from extensive ChatGPT usage. ChatGPT can answer a range of higher and lower-order questions with minor errors with respect to the assessment that was tested. This finding is in agreement with Susnak.⁴⁸ However, these errors would be sufficient cause to investigate and should be picked up by a careful marker. Higher-order questions do typically present a greater challenge than lower order and would require a degree of editing to resemble a paper. This study first considered the use of personal context in setting assessment questions. ChatGPT was found to give unconvincing third-person answers. Less direct prompts resulted in narratives that were believable but could miss important characteristics such as the CAPS curriculum and terminology unique to South Africa. This may indicate the redundancy of personal context given the imitation ability of ChatGPT. Thus, adding further evidence to the assertion that ChatGPT can imitate a variety of writing styles.⁴⁹

This indicates that employing Bloom's taxonomy and contexts that ChatGPT can access and imitate would not constitute best practices for challenging students in their use of AI tools. The final question considered in this study involved a context that ChatGPT was not trained on and could not be accessed in its Corpus of Knowledge. This reduced ChatGPT to a brainstorming role where it could only provide tools that the student would still have applied to the appropriate context. Therefore, the choice of the real-world context is key. It would be essential that the selected context is tested on ChatGPT during the assessment design process, rather than operating under the assumption that it is an inaccessible context. This provides more clarity concerning the efficacy of ChatGPT concerning real-world contexts.⁵⁰ ChatGPT's inability to answer Question 2 (Assessment C), speaks to its limitations with respect to the data it was trained on. While ChatGPT gives a convincing impression of being knowledgeable, it was not trained on all available data. This is compounded by ChatGPT's limited ability to comprehend context; this can also affect the quality of answers.⁵¹

⁴⁶ Terence Day, "A Preliminary Investigation of Fake Peer-Reviewed Citations and References Generated by ChatGPT," *The Professional Geographer*, 2023, 3.

⁴⁷ Day, "A Preliminary Investigation of Fake Peer-Reviewed Citations and References Generated by ChatGPT," 3.

⁴⁸ Teo Susnjak, "ChatGPT: The End of Online Exam Integrity?," *ArXiv Preprint ArXiv:2212.09292*, 2022, 2.
<https://doi.org/http://dx.doi.org/10.48550/arXiv.2212.09292>.

⁴⁹ Lea Bishop, "A Computer Wrote This Paper: What Chatgpt Means for Education, Research, and Writing," *Research, and Writing (January 26, 2023)*, 2023, 9.

⁵⁰ Marco Cascella et al., "Evaluating the Feasibility of ChatGPT in Healthcare: An Analysis of Multiple Clinical and Research Scenarios," *Journal of Medical Systems* 47, no. 1 (2023): 33, 1.

⁵¹ Walid Hariri, "Unlocking the Potential of ChatGPT: A Comprehensive Exploration of Its Applications, Advantages, Limitations, and Future Directions in Natural Language Processing," *ArXiv Preprint ArXiv:2304.02017*, 2023, 14.
<https://doi.org/http://dx.doi.org/10.13140/RG.2.2.23433.11360/3>.

RECOMMENDATIONS

It is clear from the findings that ChatGPT is capable of writing apparently convincing answers for both lower and higher-order questions. The only assessment that is designed using contexts that ChatGPT cannot access in conjunction with higher-order questions presents a challenge to the AI tool. This paper recommends that training in assessment design and moderation be considered at the tertiary level for academic staff. This will provide the tools to render ChatGPT use more challenging for students. Part of moderating assessment post-November 2022 should include ChatGPT testing. Alternatives to written assessments such as face-to-face handwritten tests should be considered. Further examples may include presentations and oral exams. If ChatGPT is to be better integrated at the tertiary level, students should be educated regarding the limitations of ChatGPT and how to use it as a tool more productively. Given the convincing style of ChatGPT and the difficulty of plagiarism detection, alternative tools such as ChatGPTzero and Copyleaks should be integrated with assessment management platforms. Until such a time, experience could be pooled to present workshops on how to manually detect extensive ChatGPT usage during the marking process. ChatGPT lacked detailed knowledge of South Africa's specific curricular and education system context. The discrepancies in the ChatGPT answers could be used to detect cases where ChatGPT was used to write an assessment. This paper recommends that undergraduate assessment should undergo an extensive review that considers redesigning assessment so that it scaffolds higher-order thinking in light of the ongoing developments of ChatGPT. Written assessment in particular, due to their susceptibility to ChatGPT should incorporate real-world contexts that ChatGPT cannot access.

CONCLUSION

The value of writing and research skills is at risk of debasement from assessment design flaws that can be capitalised by AI tools. ChatGPT is able to mimic and imitate academic writing styles in a manner that appears convincing. While studies have indicated that there is an awareness of the limitations of ChatGPT among the student body, there are valid concerns regarding the use of AI tools to pass the assessment.⁵² This could also impact the quality of conferred degrees. This study used ChatGPT and research literature to investigate guidelines for designing assessments. Assessment questions were then designed and tested on the OpenAI ChatGPT 3.5 platform. The results indicate a need for a broader assessment renewal at the undergraduate level as well as consideration of the structure of assessment within a module. Further policy guidelines on ChatGPT would support lecturers in their design and management of assessment. While the debate to integrate or ban ChatGPT is ongoing, AI tools are readily available on the internet and students will make use of them. The question should be how far or to what extent should these tools be integrated? This study has made modest use of ChatGPT, in interrogating the platform for design guidelines that are also drawn from the literature. It has been demonstrated that Bloom's taxonomy needs to be combined with real-world contexts that have a low internet footprint to render ChatGPT's limitations as a tool more apparent. This research was limited in that the questions were only tested on ChatGPT. This calls for further research where written assessments are designed and tested on student participants in conjunction with ChatGPT. Nonetheless, assessment design practices will need to be changed given the advent of ChatGPT and accessibility for students.

BIBLIOGRAPHY

- Allam, Hesham, Juan Dempere, Vishwesh Akre, Divya Parakash, Noman Mazher, and Jinesh Ahamed. "Artificial Intelligence in Education: An Argument of Chat-GPT Use in Education." In *2023 9th International Conference on Information Technology Trends (ITT)*, 151–56. IEEE, 2023.
- Baidoo-Anu, David, and Leticia Owusu Ansah. "Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning." *Journal of AI* 7, no. 1 (2023): 52–62.
- Bishop, Lea. "A Computer Wrote This Paper: What Chatgpt Means for Education, Research, and Writing." *Research, and Writing (January 26, 2023)*, 2023.
- Cascella, Marco, Jonathan Montomoli, Valentina Bellini, and Elena Bignami. "Evaluating the Feasibility of ChatGPT in Healthcare: An Analysis of Multiple Clinical and Research Scenarios." *Journal of Medical Systems* 47, no. 1 (2023): 33.
- Castillo, Alejandro Guadalupe Rincón, Giovanna Jackeline Serna Silva, Javier Pedro Flores Arocutipa, Haydeé Quispe Berrios, Marco Antonio Marcos Rodriguez, Guillermo Yanowsky Reyes, Hugo Ricardo Prado Lopez, Rosa Marina Vera Teves, Herbert Victor Huaranga Rivera, and José Luis Arias-González. "Effect of Chat GPT on the Digitized Learning Process of University Students." *Journal of Namibian*

⁵² Zhu et al., "Embrace Opportunities and Face Challenges...", 25.

Studies: History Politics Culture 33 (2023): 1–15.

- Chaudhry, Iffat Sabir, Sayed Ahmad M Sarwary, Ghaleb A El Refae, and Habib Chabchoub. “Time to Revisit Existing Student’s Performance Evaluation Approach in Higher Education Sector in a New Era of ChatGPT—A Case Study.” *Cogent Education* 10, no. 1 (2023): 2210461.
- Crawford, Joseph, Michael Cowling, and Kelly-Ann Allen. “Leadership Is Needed for Ethical ChatGPT: Character, Assessment, and Learning Using Artificial Intelligence (AI).” *Journal of University Teaching & Learning Practice* 20, no. 3 (2023): 02.
- Creswell, John W, and J David Creswell. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Sage publications, 2017.
- Dale, Robert. “GPT-3: What’s It Good for?” *Natural Language Engineering* 27, no. 1 (2021): 113–18.
- Day, Terence. “A Preliminary Investigation of Fake Peer-Reviewed Citations and References Generated by ChatGPT.” *The Professional Geographer*, 2023, 1–4.
- Department of Basic Education. *CAPS History Grade 10 - 12*. South Africa: Department of Education, 2011.
- . *CAPS Social Sciences Grade 4-6*. South Africa: Department of Education, 2011.
- . *CAPS Social Sciences Grade 7-9*. South Africa: Department of Education, 2011.
- Fuchs, Kevin. “Exploring the Opportunities and Challenges of NLP Models in Higher Education: Is Chat GPT a Blessing or a Curse?” In *Frontiers in Education*, 8:1–4. Frontiers, 2023.
- Gul, Rani, Shazia Kanwal, and Sadia Suleman Khan. “Preferences of the Teachers in Employing Revised Blooms Taxonomy in Their Instructions.” *Sjesr* 3, no. 2 (2020): 258–66.
- Hariri, Walid. “Unlocking the Potential of ChatGPT: A Comprehensive Exploration of Its Applications, Advantages, Limitations, and Future Directions in Natural Language Processing.” *ArXiv Preprint ArXiv:2304.02017*, 2023. <https://doi.org/http://dx.doi.org/10.13140/RG.2.2.23433.11360/3>.
- Kivunja, Charles, and Ahmed Bawa Kuyini. “Understanding and Applying Research Paradigms in Educational Contexts.” *International Journal of Higher Education* 6, no. 5 (2017): 26–41.
- Konstantis, Konstantinos, Antonios Georgas, Antonis Faras, Konstantinos Georgas, and Aristotle Tympas. “Ethical Considerations in Working with ChatGPT on a Questionnaire about the Future of Work with ChatGPT.” *AI and Ethics*, June 20, 2023. <https://doi.org/10.1007/s43681-023-00312-6>.
- Obilor, Ezezi Isaac. “Convenience and Purposive Sampling Techniques: Are They the Same.” *International Journal of Innovative Social & Science Education Research* 11, no. 1 (2023): 1–7.
- OpenAI, ChatGPT to Michael Stack, Output, August 2023. <https://chat.openai.com/c/bd67e062-b307-4e6f-ba2b-3939ad1d5c24>
- OpenAI, ChatGPT to Michael Stack, Output, August 2023. <https://chat.openai.com/c/7502475c-f759-4483-99ce-65bf07c13ccf>
- OpenAI, ChatGPT to Michael Stack, Output, August 2023. <https://chat.openai.com/c/d78e778d-c9cf-458d-934f-0cb2dad5d51f>
- OpenAI, ChatGPT to Michael Stack, Output, August 2023. <https://chat.openai.com/c/a2ae4a94-7710-4630-995a-3f5485347a03>
- OpenAI, ChatGPT to Michael Stack, Output, August 2023. <https://chat.openai.com/c/168d7d11-81a9-47d6-acdf-74273752ffa4>
- OpenAI, ChatGPT to Michael Stack, Output, August 2023. <https://chat.openai.com/c/20a3a02f-62d4-456c-a158-a21b16558fab>
- OpenAI, ChatGPT to Michael Stack, Output, August 2023. <https://chat.openai.com/c/750ece33-7c36-488a-8bdd-43db76f9aa04>
- OpenAI, ChatGPT to Michael Stack, Output, August 2023. <https://chat.openai.com/c/a551faa2-1795-44e0-91de-617069b00f47>
- OpenAI, ChatGPT to Michael Stack, Output, July 2023. <https://chat.openai.com/c/54596ce8-13fe-43b0-9e95-6374a4cfa67e>
- Rudolph, Jürgen, Samson Tan, and Shannon Tan. “ChatGPT: Bullshit Spewer or the End of Traditional Assessments in Higher Education?” *Journal of Applied Learning and Teaching* 6, no. 1 (2023).
- Scott, Emily E., Mary Pat Wenderoth, and Jennifer H. Doherty. “Design-Based Research: A Methodology to Extend and Enrich Biology Education Research.” *CBE—Life Sciences Education* 19, no. 3 (September 2020): es11. <https://doi.org/10.1187/cbe.19-11-0245>.
- Sobral, Sónia Rolland. “Bloom’s Taxonomy to Improve Teaching-Learning in Introduction to Programming.” *International Journal of Information and Education Technology* 11, no. 3 (2021): 148–53.
- Strzelecki, Artur. “To Use or Not to Use ChatGPT in Higher Education? A Study of Students’ Acceptance and Use of Technology.” *Interactive Learning Environments*, May 8, 2023, 1–14. <https://doi.org/10.1080/10494820.2023.2209881>.

Susnjak, Teo. "ChatGPT: The End of Online Exam Integrity?" *ArXiv Preprint ArXiv:2212.09292*, 2022.
<https://doi.org/http://dx.doi.org/10.48550/arXiv.2212.09292>.

Thakkar, K, and N Jagdishbhai. "Exploring the Capabilities and Limitations of GPT and Chat GPT in Natural Language Processing." *J. Manag. Res. Anal.*, 2023.

Zhu, Gaoxia, Xiuyi Fan, Chenyu Hou, Tianlong Zhong, Peter Seow, Annabel Chen Shen-Hsing, Preman Rajalingam, Low Kin Yew, and Tan Lay Poh. "Embrace Opportunities and Face Challenges: Using ChatGPT in Undergraduate Students' Collaborative Interdisciplinary Learning." *ArXiv Preprint ArXiv:2305.18616*, 2023.

ABOUT AUTHOR

Michael Stack is a Lecturer in the Department of Social Sciences and Commerce Education at the University of the Free State in South Africa. He received a Bachelor of Arts Honours in Historical Studies from the University of Cape and two master's degrees from Stellenbosch University, in Curriculum Studies and Historical Studies. He is currently pursuing his PhD studies at the University of the Free State. His research interests include Game Based Learning, Gamification and ICT integration in Education.