

Can AI replace humans? Comparing the capabilities of AI tools and human performance in a business management education scenario

Dinuka B. Herath¹  | Egena Ode² | Gayanga B. Herath³ 

¹School of Business, Education and Law, The University of Huddersfield, Huddersfield, UK

²Department of Strategy, Enterprise and Sustainability, Faculty of Business and Law, Manchester Metropolitan University, Manchester, UK

³Department of Business Administration, Inland School of Business and Social Sciences, Innlandet, Norway

Correspondence

Dinuka B. Herath, The University of Huddersfield Business School, Queensgate, Huddersfield HD1 3DH, UK.
Email: d.herath@hud.ac.uk

Abstract

This study provides a comparative assessment of the capabilities of leading artificial intelligence (AI) tools and human participants in a business management education context. Specifically, we (a) assess how well current language models perform in providing answers to standardised essay-type assessments in a business and management education context, (b) examine the efficacy of emergent tools in detecting AI-generated texts and (c) evaluate online AI re-writing and paraphrasing tools and their efficacy in evading detection. Using an exploratory qualitative design, this study generated and evaluated 15 standard essays using ChatGPT ($n=5$), Bard ($n=5$) and human ($n=5$). A comparison is provided between the average performance of AI-derived essays and that of ChatGPT-generated essays across all five essays. The results suggest that AI-generated content can achieve reasonably high marks in management and business assessments. According to the findings of the study, AI's performance is highly influenced by the types of prompts used, the user's experience and the degree to which the user can discern between relevant and irrelevant content. According to the findings, Turnitin's AI detection tool is highly effective at detecting content that has been created by AI, but the effectiveness is reduced by rewriters. The Turnitin AI detection tool, however, is significantly more effective at identifying content generated by Bard compared

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *British Educational Research Journal* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

with content generated by ChatGPT. According to the results, ChatGPT produced better results when the user provided a clear context, outlined the topic and expectations, divided the assessment tasks into sections and fed the prompts in a conversational manner to train the model. By utilising AI chatbots effectively, traditional teaching and assessment methods can be supplemented with targeted and engaging learning experiences.

KEYWORDS

artificial intelligence, Bard, ChatGPT, general language models, management education, rewriting tools

Key insights

What is the main issue that the paper addresses?

In this paper we assess how well current AI tools perform in providing answers to standardised essay-type assessments in a business and management education context, along with examining the efficacy of emergent tools in detecting AI-generated texts and evaluating online AI rewriting and paraphrasing tools as to their efficacy in evading detection.

What are the main insights that the paper provides?

AI-generated content can achieve reasonably high marks in management assessments. AI's performance is highly influenced by the types of prompts used and the user's experience. By utilising AI effectively, traditional teaching and assessment methods can be supplemented with targeted and engaging learning experiences.

INTRODUCTION

Recent developments in the field of natural language processing (NLP) have heightened the need to explore developments in generative pre-trained transformers (GPTs) (Gozalo-Brizuela & Garrido-Merchán, 2023; Prajapati et al., 2024; The Economist, 2021; The New York Times, 2023a; Thorp, 2023). These GPTs have gained popularity due to their ability to generate relatively coherent and contextually suitable text in a considerably short amount of time (Deng & Lin, 2023; Thorp, 2023). Lately, this technology has become widely used in various content creation-related use cases (Anantrasirichai & Bull, 2021; Thorp, 2023), such as journalism (Kothari & Cruikshank, 2021; Marconi, 2020; Túnñez-López et al., 2021), visual design (AlShaikh et al., 2024; Kumar et al., 2022), answering exam questions (Ali & OpenAI, 2023; Rudolph et al., 2023a), summarising text (Ying et al., 2024) and translating languages (Green et al., 2015). The most notable example of such a tool would be the ChatGPT platform by OpenAI (Deng & Lin, 2023; Thorp, 2023). In a short time, this platform

has managed to attract an enormous user base due to its human-like ability to produce content that is relatively difficult to detect as computer-generated (Thorp, 2023; Yenduri et al., 2024). This in turn has sparked a plethora of use cases, thus affecting the way work is conducted in both professional and educational contexts (Jensen et al., 2024). As such, the growing dependence on these language models has produced several challenges and concerns (Deng & Lin, 2023; Kolbjørnsrud, 2024). Recent evidence suggests that given the growing popularity of such language models, as well as recent acquisitions of GPT tools by tech giants (Financial Times, 2023; The Economist, 2021) and the introduction of new AI tools to the market, their use among the general public will undoubtedly increase. This could increase the likelihood of misuse and raise ethical concerns, which heightens the need for a deeper understanding of GPTs and their impact on society. Considering that the average student in 2024 belongs to Generation Z (Hasas et al., 2024; Seemiller & Grace, 2017) and since they are more likely to use technology (Cilliers, 2017; Hernandez-de-Menendez et al., 2020), it is fair to assume that AI-based GPTs will have a major influence on the education sector (Hernandez-de-Menendez et al., 2020; Jensen et al., 2024; Maloni et al., 2019; Rahiman & Kodikal, 2024).

Whilst research on the application of AI tools in education is still emerging, questions have been raised about the efficacy, detection capability and effects of prolonged use of GPT in education (Skavronskaya et al., 2023). Such questions in turn have increased concerns, particularly concerning the use of GPTs in essay-type assessments in educational settings (Baidoo-Anu & Ansah, 2023; Deebel & Terlecki, 2023; Wood et al., 2023). For instance, Skavronskaya et al. (2023) and Jensen et al. (2024) emphasised the potential disruptive capacity of generative AI chatbot technology in education. Similarly, Cooper (2023) describes the emergence of generative AI tools as a seismic shift with far-reaching implications. According to Ratten and Jones (2023), generative AI tools provide a systematic computing capability that can produce human-like answers in management studies. Notably, assessments conducted as part of business and management-related educational programmes tend to involve critical thinking-related tasks where students must use a combination of theories, analytical tools and techniques (Bandyopadhyay & Szostek, 2018; Calma & Cotronei-Baird, 2021; McEwen, 1994).

Taking this into account, GPT tools are particularly effective at utilising existing theories, practices and knowledge to produce text that is largely articulate and meaningful. As a result, we contend that these technologies could significantly affect the field of business and management education by, for example, affecting the quality and originality of education. This formulates the purpose of our study, which is to investigate the implications of developing language model tools (particularly GPTs) for business and management education. In so doing, we introduce an additional layer of investigation into the generative AI-driven educational debate, in the form of exploring the utility of word rewriters that allow students to reduce the AI percentage in their generative AI-generated content drastically. The use of such an AI word rewriter to disguise generative AI-written content is a factor that has not yet been captured in the current research debate (Jensen et al., 2024; McGrath et al., 2024).

This study is guided by three objectives: (a) to assess how well current language models perform in providing answers to standardised essay-type assessments in business and management education; (b) following the emerging debates on the danger of AI in promoting plagiarism and academic dishonesty in management education, this study examined the efficacy of emergent tools in detecting AI-generated texts; and (c) the study evaluates the efficiency of online rewriting and paraphrasing tools and assessed their efficacy in evading detection.

In this paper, through a comprehensive analysis of written assessments, the study provides insight into the potential implications of language models for business and management education. First, the study critically reviewed the literature surrounding the use of GPT

models in education and highlighted the shortcomings of current models and discussed the fundamental ethical concerns related to AI material generation. Then, using a three-phased exploratory approach, we designed assessment criteria for an organisational behaviour module and compared student responses with those generated by AI models (e.g., ChatGPT and Bard). The results were then assessed by management experts/lecturers using a blind review process. This process allowed us to not only experiment with the effectiveness of current GPT models in generating written material suitable for business and management assessments but also explore the potential consequences of such use cases. Second, the study assessed the effectiveness of the Turnitin AI detection tool in detecting AI-generated text by ChatGPT and Bard. Third, the research used an online rewriter to rewrite the different essays generated by ChatGPT and Bard to explore if they can evade detection by the Turnitin AI detection tool. Although some research has been carried out on the performance of AI tools and applications in education (Baidoo-Anu & Ansah, 2023; Deebel & Terlecki, 2023; Ratten & Jones, 2023; Skavronskaya et al., 2023; Wood et al., 2023), there have been few empirical comparative analysis investigations into the potency of AI detection and combined role of word rewriters in business and management studies. Thus, this study contributes to a more nuanced understanding of the significance of language models in business and management education, and in turn highlights the ethical criteria for their use in educational settings.

THEORETICAL BACKGROUND

The significance of NLP

NLP is a rapidly growing important subfield within modern AI (Cambria & White, 2014; Kaddari et al., 2020; Pugalenthi et al., 2020; Zhou et al., 2020). It is concerned with training computers on how to comprehend, interpret and articulate natural human language (Green et al., 2015; Hirschberg & Manning, 2015). To accomplish this, NLP combines expertise from disciplines such as linguistics (Bender, 2016), computer science (Bird et al., 2009; Sri, 2020) and mathematics (Bird et al., 2009), thus making it an interdisciplinary field. Collectively, it facilitates the development of powerful algorithms and models that can comprehend the intricacies of natural language. In the current digital age, with an overwhelming amount of unstructured data in the form of text, audio and video, the importance of NLP is highly evident (Kaddari et al., 2020; Zhou et al., 2020). As such, NLP provides us with the tools and techniques to make meaning of this data and extract valuable insights and knowledge. By this means, NLP has become an essential technology in various applications, from sentiment analysis (Mohanty, 2021; Solangi et al., 2018) and language translation (Green et al., 2015) to chatbots (Chong et al., 2021) and virtual assistants (Roshan & Meenu Garg, 2020; Sri, 2020).

The utility of NLP is not just limited to one industry, rather it makes significant contributions to various sectors, such as healthcare (Sadiku et al., 2018; Wen et al., 2019), finance (Fisher et al., 2016; Xing et al., 2017) and education (Ali & OpenAI, 2023; Rudolph et al., 2023b). Accordingly, NLP-powered systems can, for example, automate processes, improve decision-making and provide personalised recommendations. As such, one could credit this level of success to the development of intuitive natural language interfaces, which have managed to reinvent the way we interact with computers and other systems (AlShaikh et al., 2024; Just, 2024). For instance, consider how virtual assistants such as Siri, Alexa and Google Assistant have integrated into our daily lives. These technologies have enabled us to effortlessly converse with computers in a natural and intuitive manner (Hirschberg & Manning, 2015; Pugalenthi et al., 2020). This breakthrough in communication is a clear indication of NLP's effectiveness and utility. Besides, as we continue to generate more and

more data, the need for NLP-powered tools and applications will only increase, making it a field of great importance and relevance in today's world. NLP coupled with AI has the power to revitalise the way we function in our professional and daily personal lives. These technologies, when used methodically, have the potential to not only enhance communication and streamline decision-making but also promote greater social inclusivity. Ultimately, the continued advancements in NLP hold immense potential for transformative applications in various domains of modern society.

GPTs and how they work

In recent times, GPTs have garnered much attention in the domain of NLP (Gozalo-Brizuela & Garrido-Merchán, 2023; The New York Times, 2023; Thorp, 2023). These transformer-based models are trained on a vast corpus of text data to produce responses that are coherent and contextually appropriate for a diverse range of natural language tasks (Deng & Lin, 2023; Thorp, 2023). The distinctive feature of GPTs is that they can learn the underlying patterns and relationships in language through unsupervised pre-training, without any specific task in mind (Radford et al., 2019; Thorp, 2023). Once pre-trained, GPTs can be fine-tuned for specific assignments, such as translating a body of literature or summarising research papers or large reports. One of the most significant advantages of GPTs is their ability to generate responses that mimic human-like language with a nuanced understanding of context (Radford et al., 2019). Recently, OpenAI launched ChatGPT, a new GPT model specifically designed for conversational AI applications (The Economist, 2021; The New York Times, 2023). ChatGPT has demonstrated the ability to generate engaging and contextually relevant responses to various conversational prompts, including greetings, questions and statements (Deng & Lin, 2023; Gozalo-Brizuela & Garrido-Merchán, 2023; Thorp, 2023). Similarly, tech giant Google has also recently introduced its very own GPT named Bard, and it too is claimed to feature a similar capability (Business Insider, 2023; The New York Times, 2023b).

There are numerous industries where GPTs can be applied with great versatility. In journalism, they can be utilised to create news articles, write summaries and even headlines (Kothari & Cruikshank, 2021; Marconi, 2020; Túnñez-López et al., 2021). In visual design, GPTs can assist in generating captions, product descriptions and creating marketing material (AlShaikh et al., 2024; Kumar et al., 2022). In the field of education, GPTs can be employed to answer exam questions and generate study materials (Ali & OpenAI, 2023; Rudolph et al., 2023a). Furthermore, GPTs can also perform text summarization (Ying et al., 2024) and language translation (Green et al., 2015). This shows that GPTs are incredibly powerful tools that have the potential to revolutionise various industries. With recent advancements like ChatGPT, GPTs are becoming increasingly sophisticated and capable of generating human-like responses to a range of natural language tasks. Their broad array of uses makes them a valuable resource for businesses and even individuals seeking to streamline their NLP workflows (e.g., students could use such tools to answer their exam assignments and examination questions).

Implications of language models for business and management education

GPTs are a type of language model that can understand large amounts of text data and produce contextually relevant outputs for various natural language tasks (Deng & Lin, 2023; Thorp, 2023). Thus, due to their remarkable abilities, some predict that GPTs

will revolutionise education and in turn trigger an adverse impact on the education sector (Dehouche, 2021; Rudolph et al., 2023a; Selwyn, 2022). Business and management education may be a prime example of a field that could be greatly impacted due to its fundamental learning and assessment criteria. Considering this, one area where GPTs can be incredibly useful is critical thinking. In business and management education, students must apply different theories, analysis tools and techniques to solve complex problems (Bandyopadhyay & Szostek, 2018; Calma & Cotronei-Baird, 2021; McEwen, 1994). GPTs can be trained to perform tasks like summarization, translation and most notably the production of comprehensive written material (Ali & OpenAI, 2023; Rudolph et al., 2023b). One could argue that this could save students time and allow them to focus on understanding course material (Rudolph et al., 2023b). However, relying too much on GPTs may hinder students' ability to develop critical thinking skills, which are deemed essential in the learning and development process of an education programme (Ali & OpenAI, 2023; Rudolph et al., 2023a). Thus, in recent times this has generated great concern in the educational sector.

Similarly, another potential impact of GPTs on education is the possibility that students may produce work that lacks originality and quality (Ali & OpenAI, 2023; Rudolph et al., 2023b). This is especially dangerous given that students may attempt to bypass the essential learning processes that relate to aspects such as creativity (Kaplan, 2019), planning (Chuvgunova & Kostromina, 2016; Smith, 2007), brainstorming (Navaresta, 2021; Rao, 2007), reviewing (Favero, 2011) and reflective thinking (Chamdani et al., 2022; Odiba & Baba, 2013; Rodgers, 2002). Most of all, while GPTs can produce coherent and relevant text, educators may struggle to tell the difference between work produced by GPTs and work written by students. This raises concerns about plagiarism and uniqueness (Dehouche, 2021), especially given the fact that most institutions have not clearly accounted for the influence of AI in their policies and regulations.

In line with this, there are also ethical concerns associated with using GPTs in education. For instance, GPTs are trained on vast amounts of text data, which may contain biased, inaccurate and harmful language (Packin & Lev-Aretz, 2018; Selwyn, 2022; Srinivasan & Chander, 2021). In such an instance, such trained data could have a negative impact on students and in turn, reinforce stereotypes (Criado & Such, 2019; Selwyn, 2022; Weyerer & Langer, 2020). Therefore, as some have highlighted already (Srinivasan & Chander, 2021), it is crucial that measures be set in place to ensure that GPTs are trained on diverse and inclusive datasets (Srinivasan & Chander, 2021). Lastly, even though GPTs can greatly enhance a plethora of teaching-related processes (Rudolph et al., 2023a), institutions should also keep in mind the cost of using GPTs as this may create a digital divide between institutions that can afford them and those that cannot (Selwyn, 2022).

METHODS

This section outlines the methodology adopted in this study. The research method follows a three-phase exploratory approach, namely, preliminary preparation and assessment, evaluation and expert review. Several previous studies, such as Luiz de Mattos Nascimento et al. (2019) and Ustun and Tracey (2020), have adopted the phased qualitative methodology. Adopting a structured exploratory methodology that synthesises knowledge from multiple sources can enhance rigour, leverage existing perspectives in creative ways, adhere to a process and produce strongly defensible research (Dev Singh, 2015). In this study, output from phase 1 was evaluated in phases 2 and 3, first by the three authors independently and collectively, and then the assessments were evaluated by purposively selected experts through a moderation procedure. According to Robinson et al. (2007), adopting a structured exploratory methodology can improve

predictive accuracy. The three phases were driven by the research aims and the need to adopt a structured methodology in evaluating the role of two language models (ChatGPT and Bard) in business management education. Each of these phases is described below (see Figure 1).

As shown in Figure 1, first, the authors designed learning outcomes and assessment criteria for an undergraduate business management module. This stage generated real-time student responses to the assessment ($n=5$). To select the human participants, a simple random technique was used. A simple random sampling technique ensures that all essays generated in the population had an equal chance of being selected (Dhivyadeepa, 2015; Narayan et al., 2023; Noor et al., 2022). This helped to minimise bias, thus enhancing the representativeness of the human essay samples. Each essay was assigned a unique number, and a random number generator was used to select five essays from the available population. This method was preferred because it provided an unbiased way to select the essays whilst ensuring that no subjective criteria influenced the selection process (Narayan et al., 2023). This was critical in ensuring that the comparison between AI-generated essays and human responses was objective. This method was also implemented because it is straightforward, and guarantees that the selected

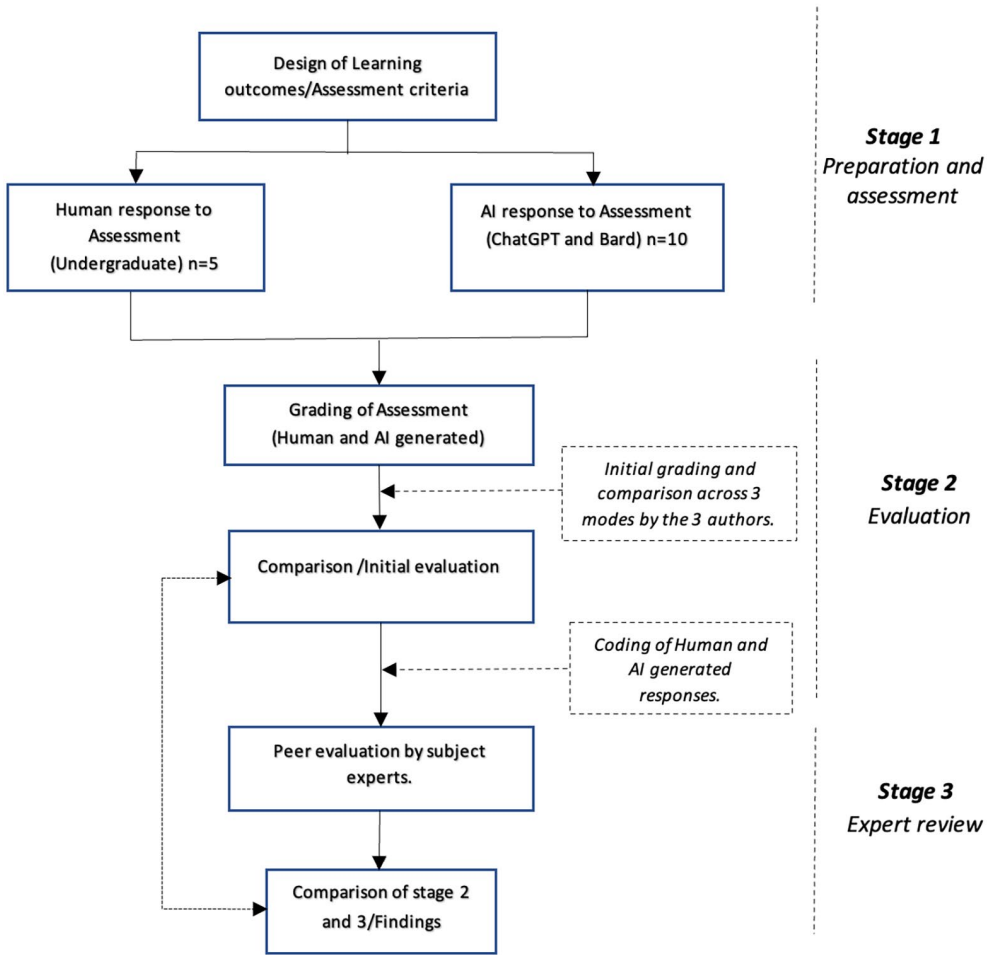


FIGURE 1 Overview of three-phase methodology. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/terms-and-conditions)]

sample is representative of the larger population (Dhivyadeepa, 2015; Noor et al., 2022). This makes the findings more generalisable to a wider academic context. By adopting this approach, the researchers were able to rigorously and fairly assess the relative performance of human and AI-generated essays.

The authors also used a chat-based approach to elicit ten ($n=10$) responses from two AI language models (ChatGPT and Bard). Adopting 5 purposive samples per group and 15 samples in total was based on best practice for exploratory studies (Boddy, 2016; Etikan et al., 2016). The aim is to provide adequate depth and variety for comparing patterns between groups and attaining data saturation where adding more data does not provide additional insights (Gibbert et al., 2008). The purposive sample aims to strike a balance between obtaining representative responses to compare AI and human performance while minimising the overall sample size for feasibility. Comparative studies such as those of Khademi (2023), Rudolph et al. (2023a) and Yıldız (2023) have adopted similar strategies. Other similar studies have adopted a chat-based approach to provide valuable insights into the potential of AI-generated content (Gunawan, 2023; Lund & Wang, 2023; Mogali, 2024). The second phase involved marking and comparing the responses between human-generated content and AI-generated content. In this phase, the authors compared results from two AI language models (ChatGPT and Bard) using inter- and intra-mode comparison. In this approach, the researchers compared responses between two AI language models and between AI and human-generated responses based on the assessment criteria. In grading the assessments, each of the three authors independently graded and moderated the assessments.

In the final phase, an independent external moderator assessed the scores from the initial grading and juxtaposed them with the marking rubric. Previous evidence suggests that relying on the knowledge of an expert—especially those who are responsible for the development, implementation and control of a solution—is important (Döringer, 2021; Moro et al., 2018). Alenezi et al. (2019) note that expert reviews have increasingly become a popular method in social sciences, mostly because of the specialist knowledge they have acquired from their professional activities. Other studies, such as that of Al-Worafi et al. (2023), have used this approach to evaluate the appropriateness and accuracy of answers generated by ChatGPT from three professors. Considering that research on the role of AI in education is emerging, adopting structured methodologies that can be replicated and tested can help to improve the replicability, validity and generalizability of research findings.

The assessment task

To test the performance of AI chatbots in business and management education and compare their performance with humans, the researchers prepared an assessment brief (see Table 1) that was used to compare the results of human participants with AI-generated essays. The assessment explores the evolution of management theory from the classical, neo-classical and modern management perspectives and how they have shaped the development of organisational structure in relation to two quotes by Drucker (2007, p. 3): 'There is or there must be ONE right organisation structure'. 'There is or there must be ONE right way to manage people'. The expected word count for the essay is 2500 words excluding references and appendices. In total, 15 essays were generated (human response $n=5$, ChatGPT $n=5$, Bard $n=5$). The essays went through two stages of marking and rounds of moderation, using the same format adopted by most business schools in evaluating student assessments.

To generate responses from ChatGPT, the researchers used various instruction formats to elicit responses using ChatGPT 3.5. The nature of the input used determined the kind of response generated. In generating the first essay, the assessment instruction was fed into

TABLE 1 Assessment brief.

Since the study of management first began, two assumptions regarding the realities of management have been held by most scholars and practitioners: 'There is or there must be ONE right organisation structure'. 'There is or there must be ONE right way to manage people' (Drucker, 2007, p. 3).
Question How relevant are these two traditional management assumptions for organisations in the twenty-first century? Critically discuss and justify your answer.
Key points to focus on when writing your answer • The evolution of management theory (referring to classical, neo-classical and modern perspectives of management). • How did organisational 'structures' evolve? • How did our approaches to 'management' evolve? • What role does 'technology', 'social trends' and 'the drive for globalisation' play in this evolution?
Task-specific guidance The work needs to be presented in the form of an essay using concepts and theories to support the discussion. It is expected that you will draw upon academic journals to support the work with examples of research carried out in the field of management, work and/or organisational behaviour. Avoid description of the content of material referred to—critical evaluation is required where specified. Read widely from textbooks, journals and authoritative commentaries in forming your views. Pay close attention to the assessment criteria (see Appendix A). These criteria will be used to inform electronic feedback on your marked assignment.
General study guidance Cite all information used in your work which is clearly from a source. Try to ensure that all sources in your reference list are seen as citations in your work, and all names cited in the work appear in your reference list. Reference and cite your work in accordance with the APA 7th system. Avoid plagiarism. Do not exceed the word limit/time/other limit. Word/time limit: 2500 (excluding references & appendices).

ChatGPT as a single instruction. This meant that the researchers inserted the assessment brief and the assessment criteria into ChatGPT. However, this changed with the remaining four assessments. Generally, the authors used a trial-and-error approach to find out the method that produced the most efficient outcome. The researchers found that generating the best response follows a four-step approach. These steps involved (a) establishing the context, (b) explaining the topic, (c) specifying the task and (d) providing follow-up questions and prompts. The findings suggest that ChatGPT produced a better result when the user set a clear context, fed in the details about the topic and expectations, broke down the assessment tasks into sections and fed the prompts using a conversation-style approach to train the model. In this instance the researcher established clear boundaries such as the course (i.e., business and management), level (i.e., first year university-level course), expectations (i.e., assessment should be 2500 words containing an introduction, relevant literature and a conclusion), topics covered (i.e., classical, neo-classical, modern management, organisational structures and the role of technology) and clarified some of the basic concepts (i.e., be specific about the contributions of various proponents of each school of thought) (see Figure 2). Prompts such as 'continue', 'elaborate', 'provide a counter-argument and justification' or 'the entire section is dependent on a single citation' or 'there is no citation to support the section, please use APA 7th format' and 'provide a list of references and supporting documents used' were used to improve the quality of the content generated.

In total, five essays were generated by the researchers using ChatGPT 3.5. A similar format was used to generate five essay responses from Bard (see Figure 3).

In total, 15 essays were evaluated. The evaluation used a standardised method for assessing student essay-type assessments. The reports were independently graded using five broad criteria: (a) detailed knowledge and critical understanding of relevant knowledge in the subject (25%); (b) argument and reasoning (25%); (c) use of referenced evidence and

First instruction:
 "There 'is' or there 'must be' - O/E right organisation structure". "There 'is' or there 'must be' - O/E right way to manage people". (Drucker, 2007, p3). How relevant are these two traditional management assumptions for organisations in the 21st Century? Critically discuss and justify your answer using in-text citations and references from academic journals.

Second instruction:
 Relate the discussion above to the evolution of management theory (Referring to classical, neo-classical, and modern perspectives of management). Use in-text citations from academic journals.
 'Continue'.

Third instruction:
 Relate the discussions above to the question: How did organizational "structures" evolve? Use in-text citations from academic journals.

Fourth instruction:
 With reference to the discussions above, what role does "technology", "social trends" and "the drive for globalisation" play in this evolution? Use in-text citations from recent academic journals.

Fifth instruction:
 Provide a conclusion that wraps the entire discussion from the beginning of the conversation.

Sixth instruction:
 Provide an introduction to the essay.

The entire section was dependent on wood

You are correct, and I apologize for that mistake. Here is a revised version that includes proper APA 7th edition citations and a more diverse range of sources:

The evolution of management theory has had a significant impact on the traditional assumptions regarding organizational structures and management approaches. Classical management theory, which emphasized the importance of efficiency, control, and hierarchy, has given way to more modern management theories, such as human relations theory and contingency theory (Robbins, Coulter, & DeCenzo, 2020). Human relations theory emphasized the importance of employee satisfaction, motivation, and teamwork, while contingency theory recognized the need for flexibility and adaptability in management approaches.

With the rise of the knowledge economy, the importance of human capital has become increasingly recognized. As a result, management approaches have shifted towards a more human-centric approach, with a focus on employee empowerment, autonomy, and creativity. This has led to the development of more participative management styles, such as transformational and servant leadership, which emphasize the importance of collaboration, communication, and trust (Northouse, 2021).

Technological advancements have also had a significant impact on the evolution of management. The proliferation of digital technologies has transformed the way we work and communicate, making it easier for organizations to operate globally and for employees to work remotely. As a result, many organizations are moving towards more flexible and virtual work arrangements, which require new management approaches to ensure that employees

Evolution of Management Theory

Management theory is a body of knowledge that seeks to understand and explain how organizations work. It is concerned with the processes, functions, and activities that are necessary for organizations to achieve their goals. Management theory has evolved over time, reflecting the changing nature of organizations and the environment in which they operate.

Classical Management Theory

Classical management theory emerged in the early 20th century in response to the increasing size and complexity of organizations. The classical theorists, such as Frederick Taylor, Henri Fayol, and Max Weber, sought to develop a scientific approach to management that would increase efficiency and productivity.

- **Taylor's scientific management theory** focused on the division of labor, specialization, and time and motion studies. Taylor believed that the best way to increase productivity was to break down work into its smallest components, analyze each component, and then find the most efficient way to perform each component. He also believed that managers should use time and motion studies to measure the efficiency of workers and to identify areas where productivity could be improved (Taylor, 1911).
- **Fayol's general management theory** identified the five primary functions of management: planning, organizing, commanding, coordinating, and controlling. Fayol believed that managers should develop a set of principles that could be used to guide their decision-making (Fayol, 1949).
- **Weber's bureaucratic theory** proposed a rational, hierarchical structure of authority that would ensure efficiency and order. Weber argued that bureaucracy was characterized by a clear hierarchy of authority, a division of labor, and a set of rules and procedures that governed the behavior of employees (Weber, 1947).

Neo-Classical Management Theory

Neo-classical management theory emerged in the mid-20th century in response to the perceived shortcomings of classical management theory. The neo-classical theorists, such as Chester Barnard, Elton Mayo, and Douglas McGregor, argued that classical management theory was too mechanistic and did not take into account the human element of organizations.

- **Barnard's cooperative system theory** believed that organizations are cooperative systems that are based on the consent of the governed. He argued that managers should focus on creating an environment in which employees are motivated to contribute to the goals of the organization (Barnard, 1938).
- **Mayo's Hawthorne studies** showed that social factors, such as group dynamics and morale, could have a significant impact on productivity. Mayo's work led to a greater focus on the human element of organizations and to the development of new management techniques, such as employee participation and team building (Mayo, 1933).
- **McGregor's Theory X and Theory Y** proposed two different views of human nature: Theory X assumes that people are lazy and need to be controlled. Theory Y assumes that people are self-motivated and can be trusted to take responsibility for their own work. McGregor's work has had a significant impact on the development of management theory and practice (McGregor, 1960).

Neo-classical management theory has helped to humanize the workplace and to make organizations more responsive to the needs of employees. However, it has also been criticized for being too soft and for failing to provide managers with the tools they need to effectively manage organizations.

FIGURE 2 Sample of ChatGPT instructions and response. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/berj.4111)]

sources to support task—APA 7th (20%); (d) structure and style in supporting the development of ideas (criteria relevant for essay-style work) (20%); and (e) formatting of work (font, pagination, labelling) (10%).

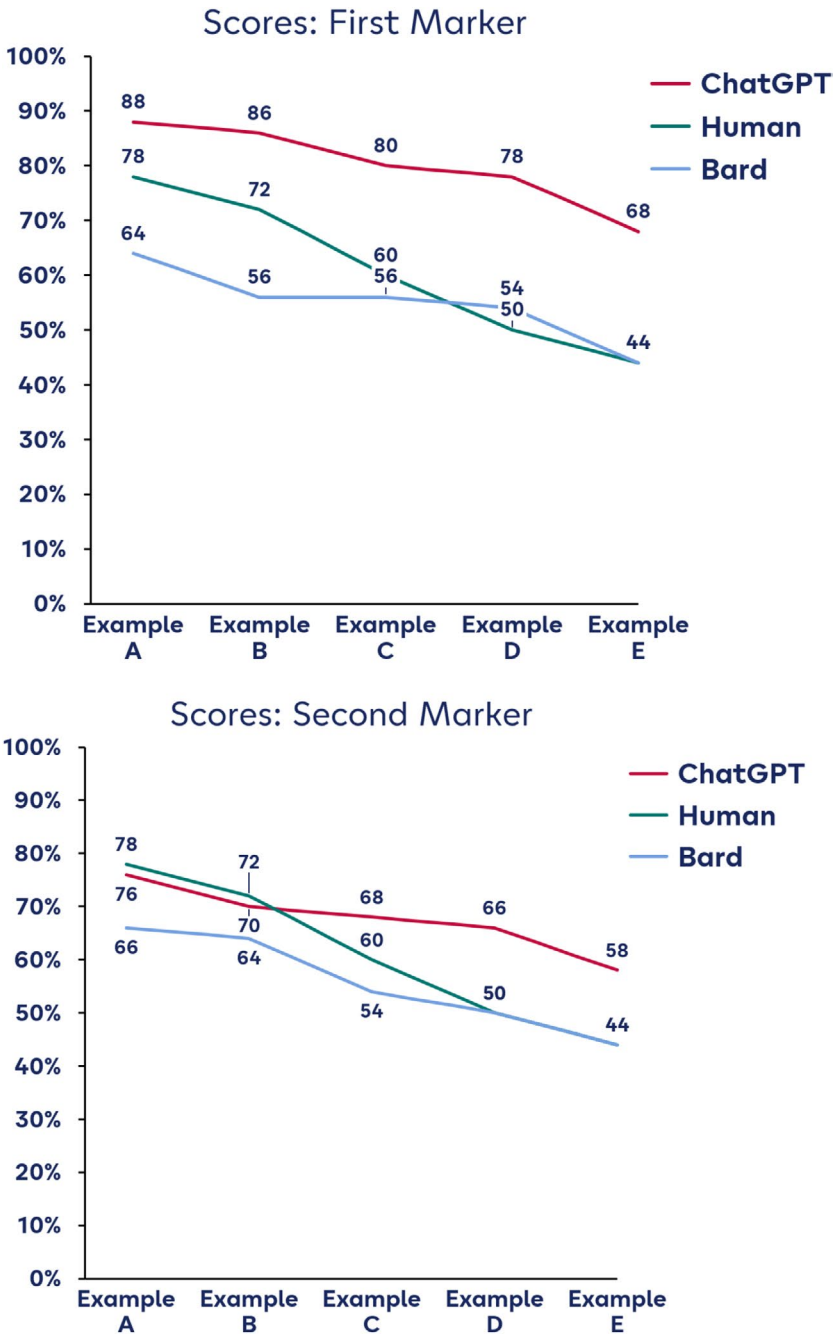


FIGURE 3 Comparative scores for first and second markers. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/berj.4111)]

RESULTS

This study investigated the implications of emergent language model tools (particularly GPTs) and their application to business and management education. To accomplish this goal, in the first instance the study sought to systematically assess how well current language models perform in providing answers to standardised essay-type assessments in

business and management education. To achieve this aim, using a standardised assessment, the study compared the performance of two language models, ChatGPT and Bard, with human responses in the context of management education.

Second, following the emerging debates on the danger of AI in promoting plagiarism and academic dishonesty in management education, this study examined the capability of Turnitin in detecting academic dishonesty and plagiarism from AI-generated text. To accomplish this goal, the study assessed the efficacy of the Turnitin AI detection tool in detecting AI-generated text by ChatGPT and Bard.

Third, this study evaluated the efficiency of online rewriting and paraphrasing tools and assessed their efficacy in evading detection by plagiarism and AI detection tools. To accomplish this, the research used WordTune and Quillbot online rewriters to rewrite the different essays generated by ChatGPT and Bard. Thereafter, the result was assessed and analysed.

In the following subsections we present the findings, followed by a discussion section. We compare the three groups (human, ChatGPT and Bard) and explore the efficacy of AI detection tools, along with the role of emergent tools such as word rewriters and paraphrasing tools. We discuss their implications for business and management education.

Performance: ChatGPT versus Bard versus human responses

The results of the essays across the three observations are reported in [Table 2](#).

As shown in [Figure 4](#), AI-generated essays performed relatively well. Specifically, on average, ChatGPT-generated essays performed better across all five essays. The results confirm that AI can produce relatively high grades in business and management assessments. However, how AI performs is hugely dependent on the nature of prompts used, the user's experience and breadth of knowledge in deciphering between relevant and irrelevant content.

To evaluate the performance of AI tools (ChatGPT and Bard) and human scores, a one-way analysis of variance (ANOVA) was conducted to compare the performance among the different groups. We had the following hypothesis: There is a statistically significant

TABLE 2 Grades and moderation.

First marker score			
Code	Human**	ChatGPT	Bard
Example A	78%	88%	64%
Example B	72%	86%	56%
Example C	60%	80%	56%
Example D	50%	78%	54%
Example E	44%	68%	44%
Second marker score			
	Human**	ChatGPT	Bard
Example A	78%	76%	66%
Example B	72%	70%	64%
Example C	60%	68%	54%
Example D	50%	66%	50%
Example E	44%	58%	50%

Note: **Moderated score.

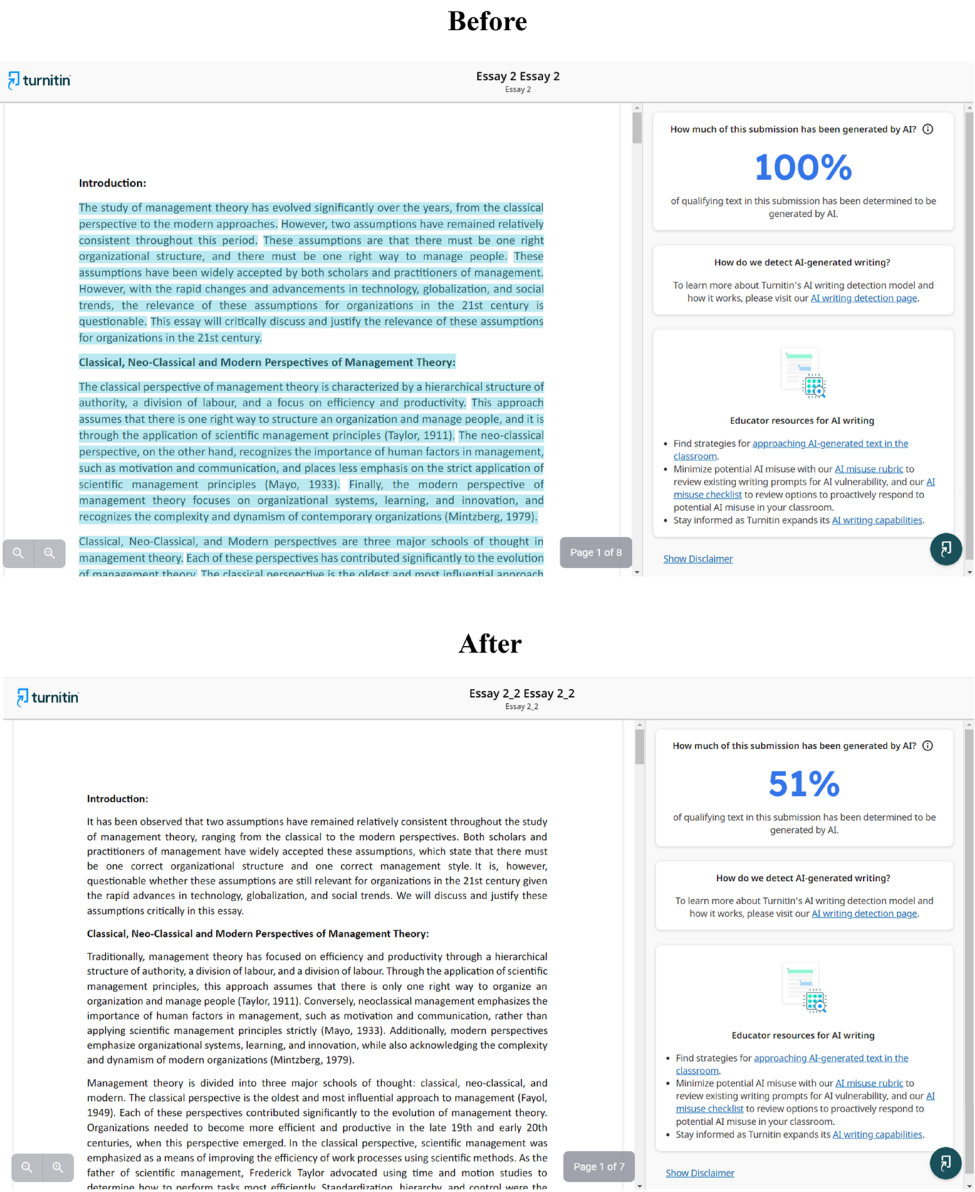


FIGURE 4 Sample of AI report before and after using a rewriter. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/berj.4111)]

difference in performance of essays generated by ChatGPT, Bard and human students. To test this hypothesis, the comparison involved three steps. First, we compared the mean of the first marker. Second, a one-way ANOVA was performed to compare the scores for the moderator. Third, a comparison of the average scores between the first marker and the moderator. In the first step, the results revealed a statistically significant difference between the groups ($F(2, 12)=8.163, p=0.06$) with a large effect size (57.6%). A post-hoc Tukey test demonstrated that ChatGPT significantly outperformed both Bard ($M=25.20, p=0.006$) and human essays ($M=19.20, p=0.030$). However, the results show no statistically significant difference between human essays and Bard ($M=6.00, p=0.638$). The ANOVA results for the second marker (moderator) revealed that there was no statistically significant

TABLE 3 Results of ANOVA and group performance differences.

Measure	First marker	Moderator	Average
Sum of squares (between groups)	1732.800	298.133	863.333
d.f.	2	2	2
Mean square (between groups)	866.400	149.067	431.667
F-Value	8.163	1.456	4.245
p-Value	0.006	0.272	0.040
Effect size (eta squared)	0.576	0.195	0.414
ChatGPT mean score	80.0	67.6	73.8
Bard mean score	54.8	56.8	55.8
Human mean score	60.8	60.8	60.8
Significant difference ($p < 0.05$)	ChatGPT > Bard ($p = 0.006$); ChatGPT > human ($p = 0.030$)	None	ChatGPT > Bard ($p = 0.038$)

difference in essay performance across the three groups ($F(2, 12) = 1.456$, $p = 0.272$) with a smaller effect size (19.5%). Post-hoc Tukey tests also indicate that there is no significant difference between ChatGPT and Bard ($M = 10.80$, $p = 0.250$) or between ChatGPT and human essays ($M = 6.80$, $p = 0.554$). Likewise, the results show no statistical difference between Bard and human essays ($M = 4.00$, $p = 0.809$). To provide a more comprehensive view of the performance across the first marker and the moderator, the scores from the first and second markers were averaged, and a one-way ANOVA was performed. As shown in Table 3, the results revealed a statistically significant difference between the three groups ($F(2, 12) = 4.245$, $p = 0.040$), with a moderate effect size (41.4%) of the variance in scores. To test further, a post-hoc Tukey test was performed. The results showed that ChatGPT essays significantly outperformed Bard essays ($M = 13.00$, $p = 0.038$). However, the results show that there was no statistically significant difference between ChatGPT and human essays ($M = 13.00$, $p = 0.145$), or between human and Bard essays ($M = -5.00$, $p = 0.720$).

As shown in Table 3, the findings of the first marker and the combined average analyses provide support for the hypothesis that there is a significant difference in the performance of ChatGPT, Bard and human essays. The results indicated that ChatGPT consistently outperformed Bard and according to the results of the first marker, it also outperforms human essays. However, when the scores of the first marker and the second marker are averaged, there is no significant difference between ChatGPT and human essays, suggesting that ChatGPT essays can be comparable to human-generated essays.

These findings show that AI tools, particularly ChatGPT, can perform at or above the level of human students in the context of business education. While ChatGPT consistently outperforms Bard, the performance of human essays is more comparable to ChatGPT in the combined analysis. These results suggest that ChatGPT can serve as a valuable tool in academic writing tasks, but educators should also be aware of the discrepancies between different AI tools as demonstrated in this study. Moreover, the lack of significant difference in some of the comparisons demonstrates that AI-generated content can closely mimic human submissions, raising important questions about assessments and detection methods in academic contexts. The emergence of AI chatbots has raised concerns among academics about the possibility that students will rely heavily on the technology to provide academic assessment solutions. It is understandable that students can use AI chatbots as a crutch to complete their assignments since they can answer questions and provide solutions to problems. Increasing reliance on AI technology could lead to students becoming overly dependent on it, which could negatively affect their academic performance. The use of AI chatbots

in academic assessment, however, is subject to ethical considerations as well. In what ways is it fair to allow students to work with AI chatbots when others may not be able to use these technologies? What are the implications for the educational system at large? Because students are increasingly relying on AI chatbots to complete academic assessments, academics must be aware of the dangers and take proactive measures to ensure that students are using this technology responsibly. As part of this, AI chatbots must be monitored, clear guidance should be provided on their use and assessments should be designed to limit their use. To ensure that AI chatbots do not hinder students' academic performance, academics must first ensure that AI chatbots are not detrimental to the learning process. To effectively achieve this, systematic evidence needs to be provided to enhance our overall understanding and performance of the technology in academic assessments.

AI chatbots, online paraphrasing tools and plagiarism detection

One of the objectives of this research is to explore the efficacy of plagiarism detection and AI paraphrasing tools. Specifically, the researchers used an online word rewriting tool to rewrite ten AI-generated essays. Before using the rewriting tool, the essays were assessed using Turnitin. As shown in Table 4, the highest similarity report for AI-generated content is 35% for ChatGPT and 33% for Bard, while the highest similarity report for human-generated essays is 23%.

The findings of this research show that these tools allow users to avoid plagiarism detection by rephrasing and reorganising sentences to create content that is unique to the original source. This is because the AI chatbot mimics human conversation and provides students with automated assistance, which could be original. These bots are programmed to provide responses that are relevant to the user's query. This raises concerns about the future of plagiarism detection. As shown in Table 4, after running the ChatGPT essays one time each through a word rewriter, the AI percentage dropped by over 50%. To test this further, the word rewriter tool was used to rewrite the Bard essays three times for each essay. This drastically reduced both the similarity index (see Table 5) and the AI detection (see Table 6). This shows that continuous use of word rewriting tools can mask text generated by AI language models. The findings show that despite using a rewriter tool, the Turnitin AI detection tool is more effective in detecting Bard-generated content than ChatGPT-generated content. As shown in Table 5, the results also show that using an online word rewriter can effectively evade plagiarism detection by drastically reducing the similarity index.

These findings demonstrate that with the development of AI chatbots and word paraphrasing tools, it is becoming increasingly difficult to detect plagiarism and AI-generated content for academic assessments. Although the findings showcase the effectiveness of the Turnitin AI detection tool, for plagiarism and AI detection to keep pace with the evolving tools, deploying more advanced AI technology may become necessary. This may involve

TABLE 4 Summary of similarity report.

Similarity report (Turnitin)			
Code	Human	ChatGPT	Bard
Example A	23%	25%	14%
Example B	12%	25%	33%
Example C	17%	31%	24%
Example D	15%	28%	19%
Example E	15%	35%	22%

TABLE 5 AI detection and word rewriting tool report.

AI report before and after using a word rewriter					
Code	Human	ChatGPT		Bard	
Example A	—	100%	43% ^a	100%	100% ^a
Example B	—	100%	51% ^b	99%	82% ^b
Example C	—	100%	34% ^c	100%	74% ^c
Example D	—	100%	57% ^d	100%	80% ^d
Example E		100%	41% ^e	100%	82% ^e

Note: ^{a–e}AI report after using a word rewriter.

TABLE 6 Similarity report after using word rewriter.

Code	Human	ChatGPT	Bard
Example A	23%	25% (15%)	14% (7%)
Example B	12%	25% (18%)	33% (15%)
Example C	17%	31% (18%)	24% (10%)
Example D	15%	28% (22%)	19% (4%)
Example E	15%	35% (26%)	22% (3%)

Note: Score in brackets = Turnitin report after using a word rewriter tool.

using machine learning algorithms and NLP to improve the detection of plagiarised and AI-generated content. These findings demonstrate that students' use of AI chatbots and word paraphrasing tools has both positive and negative impacts. While they can help assist with assignments and research, they raise questions regarding academic integrity and plagiarism detection. To maintain academic standards, educators and institutions must stay on top of technological advancements as they continue to evolve.

Whilst many previous findings report that AI tools hallucinate references, the findings in this research demonstrate that in the context of management, ChatGPT and Bard generated valid references. All the references generated were checked on Google Scholar. Although the findings show that some of the references did not exist, the results show that the percentage of valid references generated was more than those that were hallucinated.

DISCUSSION AND IMPLICATIONS

A case for promoting ethical use

In modern pedagogy, the use of AI chatbots such as ChatGPT, Bard and word rewriters/paraphrasing tools for academic purposes has become an interesting discourse (Ali & OpenAI, 2023; Burger et al., 2023; Dwivedi et al., 2023; Korzynski et al., 2023; Rudolph et al., 2023a). This study offers fresh perspectives on the ongoing debate by comparing the performance of two popular AI chatbots with human participants in the context of business and management assessments. This is quite interesting because previous studies have focused on assessing these tools (ChatGPT and Bard) separately, and not in the context of management education (Dehouche, 2021; Rudolph et al., 2023a; Selwyn, 2022). Specifically, very limited studies have explored the role of Bard and rewriters in management education (Jensen et al., 2024; Khademi, 2023; McGrath et al., 2024). By combining and comparing these tools in the context of business and management education, this study

aims to contribute to the emerging debates on the role of AI chatbots in management education. The result of this study provides valuable insights for management education and contributes to emerging debates on the role of AI tools in management (Burger et al., 2023; Dwivedi et al., 2023; Korzynski et al., 2023).

First, the findings confirm that AI chatbots such as ChatGPT and Bard can perform relatively well in business and management essay-type assessments. The study found that the performance of the AI chatbots was comparable to that of human participants. However, there were some areas where the chatbots struggled, particularly when it came to more complex tasks that required a deeper understanding of the subject matter, synthesis and analysis of multiple concepts in business management. Despite this limitation, the study concludes that the use of AI chatbots in management education could be a valuable tool for both students and educators. It is important to note, however, that the performance of AI chatbots improves significantly with the format of prompts and the experience of the user, which suggests that AI bots are not yet ready to replace human students in all academic contexts as they depend significantly on humans to generate relevant content. It is also found that the effectiveness of AI tools depends on the student's capability and his or her ability to differentiate relevant from irrelevant information. Consequently, AI chatbots are not a 'one size fits all' solution for all educational needs. Instead, they are tools that can be used to varying degrees, depending on the user's expertise. It is possible to use AI chatbots in conjunction with human-led instruction to provide the best learning experience to students. Ultimately, AI tools in business and management education should be viewed as an opportunity to enhance the student's learning experience rather than a replacement for traditional teaching methods. A learning experience tailored to each student's needs and abilities can be created by utilising the numerous benefits that AI tools offer. With the right use of AI chatbots, traditional teaching and assessment methods can be supplemented with targeted and engaging learning experiences for students. Students can improve their performance by combining human and AI-led instruction. While providing students with personalised and meaningful learning experiences, AI-led instruction can also help educators leverage AI tools.

Consistent with the findings of Rudolph et al. (2023a), the findings in this study have shown that students can use AI tools in assessments without getting caught. Paraphrasing and other AI tools are part of this tool set. This is potentially dangerous given that it may empower students to bypass the essential learning processes that relate to aspects such as creativity (Kaplan, 2019), planning (Chuvgunova & Kostromina, 2016; Smith, 2007), brainstorming (Navaresta, 2021; Rao, 2007), reviewing (Favero, 2011) and reflective thinking (Chamdani et al., 2022; Odiba & Baba, 2013; Rodgers, 2002). This research argues that rather than trying to prevent students from using AI tools, it is necessary to focus on ethical use. It is important to emphasise ethical use to ensure that students can use these tools responsibly, without compromising their assessments' integrity. These tools can be quite helpful, but they also raise important questions about academic dishonesty and the role of technology in education (Dehouche, 2021; Jensen et al., 2024). Educational systems can continue to benefit from AI tools without compromising their integrity by fostering a culture of ethical use and integrating them into pedagogy development. The findings suggest that AI tools have the potential to revolutionise assessment methods. However, to ensure that AI tools continue to play a positive and productive role in education, we should keep in mind the ethical implications of their use and promote responsible practices.

Using AI chatbots such as ChatGPT, Bard and word rewriters/paraphrasing tools for academic purposes is an interesting discourse in modern pedagogy. This study offers fresh perspectives on the ongoing debate by comparing the performance of AI chatbots with human participants in business and management assessments. In summary, the results of this study provide valuable insights for management education practitioners on how to prevent and detect plagiarism and academic dishonesty in the future. The research makes three

important contributions by comparing different AI tools with human performance. First, the quality of responses generated depends on the breadth of knowledge of the user. Second, the quality of output depends largely on the ability of the user to integrate other resources and technology. Third, it is possible to evade detection by combining other tools to improve the accuracy of AI-generated content; the use of references alone to check the authenticity of generated content can be misleading.

Broader implications

Trying to prevent, or limit, AI use in educational settings is not only becoming increasingly impractical but also effectively deskills those pupils compared to others who have learned these tools. This can exacerbate already existing disparities between student groups in higher education (HE) settings (Lim et al., 2023). In addition, it can create unease and disparities that might be very complicated to reverse in the long term. Our findings indicate that the quality of human content produced using AI is strongly dependent on an individual's expertise and proficiency. Those who are better educated tend to be better at effective use of AI prompts, and at detecting when AI makes an error or a dubious assertion. This shows that traditional learning is not necessarily at odds with AI, but rather it is complementary to its effective use.

Furthermore, the prevalent dynamic between AI usage and the tools used to detect AI usage locks us into a never-ending tussle between AI use and its detection. Instead, rethinking our approach to this dynamic would mean AI integration in our teaching and learning methods. For instance, one approach is to use veracity-based AI teaching methods, where students are encouraged to apply AI and one of the tasks in the assignment could be to establish the veracity of the AI-produced content. This approach ensures students need to plan, be creative, research and evaluate AI-produced content in relation to the assignment task. This ensures the preservation of some of the essential learning processes noted by others such as Kaplan (2019) and Chamdani et al. (2022), while also enabling students to explore the effective use of AI tools.

Moreover, as educators, understanding the skills mix a modern student requires is essential in producing competent and well-functioning graduates. As such, it is increasingly challenging to make a case for producing graduates without effective AI skills in an attempt to conserve traditional learning (McGrath et al., 2024; Rahiman & Kodikal, 2024; Wu & Yu, 2024). For instance, before the advent of the motor car, learning to ride a horse would have been an essential skill any individual needed to master. However, with the advent of the motor car, effective equestrian skills became obsolete. One's time was better spent learning to drive than riding a horse. As such, our approaches to AI integration into student teaching and learning need to take into account the landscape and trajectory modern students find themselves in. At this pivotal initial juncture between AI and HE, particularly given the capability of generative AI as shown in this study, efforts to incorporate AI with HE seem promising (McGrath et al., 2024; Wu & Yu, 2024). Some have already begun to formulate ways in which we could incorporate AI with HE. For instance, Perkins et al. (2024) have developed a framework for ethically integrating generative AI in educational assessments.

Specifically in terms of management, another critical consideration is what management as a discipline and an organisational function entails. It entails working with people. This means that even if we take the argument that AI will replace many (if not all) management functions at face value, it does not follow that the people who used to perform those functions effectively disappear. Instead, the more realistic scenario is that how people relate to work and derive meaning from work would change as a result of AI replacing management functions. As such, a future manager (at least in the medium term) would still need to be effective in people management, albeit with the added variable of AI in the mix (Kolbjørnsrud, 2024). It could also be a catalyst for future managers to be even more people-oriented to cater to the

increasing and varied ways people seek meaning and see their relationship with work (Hasas et al., 2024) in a world where AI has replaced many traditional aspects of work.

CONCLUSION

A threefold contribution has been made to the existing debate on AI's application in education in this study. In particular, the study focused on management education, an area that has been underexplored by most research evaluating AI in education. First, a comparative analysis of the performance of two AI tools was compared with human-generated content in a business and management context. The results demonstrated that AI performs well in essay-type assessments. Second, the findings demonstrated that Turnitin is a powerful tool capable of detecting AI-generated content, although the effectiveness degenerates with the use of word rewriters. Finally, this study demonstrated the role rewriters can play in aiding academic dishonesty. The role of AI in education has rarely been explored comprehensively in previous studies. The approach adopted in this study is critical because it allows us to gain a better understanding of how AI can be used to detect academic dishonesty, as well as the potential implications of its use. It also allows us to consider potential strategies to prevent academic dishonesty, such as the use of word rewriters. In addition, the study highlights the need for further research in this area to better understand the potential of AI in education.

As AI develops rapidly, the delivery and management of education will remain at risk. With the proliferation of a variety of available AI tools, students are increasingly exploring how to use them whilst finding ways to avoid detection. In this sense, it raises concerns that over-reliance on AI technology can lead to passiveness on the part of students. When students spend more time using AI systems, they may have less interaction with their teachers and classmates as a result. Besides student cheating and exploring ways educators can detect and avoid it, there are also other concerns such as bias in AI algorithms, privacy issues, an apparent overreliance on technology, equity issues and a risk of decreased social connections. There is also the potential for student disengagement or a decrease in critical thinking skills that may result from reliance on AI. These issues require further exploration. Since the research in the field is still emerging, and considering that this study is exploratory, future studies can assess other methods of assessment in business and management. Future studies can also explore nuances and the role of education and educational institutions in developing effective systems and strategies to cope with the rising AI technology. Going forward, management educators must rethink how assessments are conducted as students increasingly rely on AI tools to guide their assessments. There are concerns about promoting ethical use and setting boundaries for students. This way, there is an established boundary of dos and don'ts. Despite AI's transformative capability, the use of AI tools is subject to limitations, including the possibility of producing inaccurate or inappropriate content, because they are still being developed.

FUNDING INFORMATION

This research has no funding to report.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to report.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in the [Online Supplementary Material](#).

ETHICS STATEMENT

This paper and the research within it comply with BERA's ethics policy. There are no ethical issues to report.

ORCID

Dinuka B. Herath  <https://orcid.org/0000-0001-9235-7004>

Gayanga B. Herath  <https://orcid.org/0000-0002-4014-8339>

REFERENCES

- Alenezi, A., Atlam, H. F., & Wills, G. B. (2019). Experts' reviews of a cloud forensic readiness framework for organizations. *Journal of Cloud Computing*, 8(1), 1–14. <https://doi.org/10.1186/S13677-019-0133-Z/FIGURES/3>
- Ali, F., & OpenAI. (2023). Let the devil speak for itself: Should ChatGPT be allowed or banned in hospitality and tourism schools? *Journal of Global Hospitality and Tourism*, 2(1), 1–6. <https://doi.org/10.5038/2771-5957.2.1.1016>
- AlShaikh, R., Al-Malki, N., & Almasre, M. (2024). The implementation of the cognitive theory of multimedia learning in the design and evaluation of an AI educational video assistant utilizing large language models. *Heliyon*, 10(3), e25361. <https://doi.org/10.1016/j.heliyon.2024.e25361>
- Al-Worafi, Y. M., Hermansyah, A., Goh, K. W., & Ming, L. C. (2023). Artificial intelligence use in university: Should we ban ChatGPT? <https://doi.org/10.20944/PREPRINTS202302.0400.V1>
- Anantrasirichai, N., & Bull, D. (2021). Artificial intelligence in the creative industries: A review. *Artificial Intelligence Review*, 55(1), 589–656. <https://doi.org/10.1007/s10462-021-10039-7>
- Baidoo-Anu, D., & Ansah, L. O. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI*, 7(1), 52–62. <https://dergiipark.org.tr/en/pub/jai/issue/77844/1337500>
- Bandyopadhyay, S., & Szostek, J. (2018). Thinking critically about critical thinking: Assessing critical thinking of business students using multiple measures. *Journal of Education for Business*, 94(4), 259–270. <https://doi.org/10.1080/08832323.2018.1524355>
- Bender, E. M. (2016). Linguistic typology in natural language processing. *Linguistic Typology*, 20(3), 645–660. <https://doi.org/10.1515/lingty-2016-0035>
- Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: Analysing text with the natural language toolkit*. O'Reilly Media. <http://my.safaribooksonline.com/9780596516499>
- Boddy, C. R. (2016). Sample size for qualitative research. *Qualitative Market Research: An International Journal*, 19(4), 426–432. <https://doi.org/10.1108/QMR-06-2016-0053>
- Burger, B., Kanbach, D. K., Kraus, S., Breier, M., & Corvello, V. (2023). On the use of AI-based tools like ChatGPT to support management research. *European Journal of Innovation Management*, 26(7), 233–241. <https://doi.org/10.1108/EJIM-02-2023-0156>
- Business Insider. (2023). Google Bard is here. Catch up on everything you need to know about the CHATGPT alternative. Business Insider Retrieved 5 April 2023 from <https://www.businessinsider.com/google-bard-ai-chatgpt-alternative-2023-3?r=US&IR=T>
- Calma, A., & Cotronei-Baird, V. (2021). Assessing critical thinking in business education: Key issues and practical solutions. *The International Journal of Management Education*, 19(3), 100531. <https://doi.org/10.1016/j.ijme.2021.100531>
- Cambria, E., & White, B. (2014). Jumping NLP curves: A review of natural language processing research. *IEEE Computational Intelligence Magazine*, 9(2), 48–57. <https://doi.org/10.1109/MCI.2014.2307227>
- Chamdani, M., Ali Yusuf, F., Salimi, M., & Fajari, L. E. (2022). Meta-analysis study: The relationship between reflective thinking and learning achievement. *Journal on Efficiency and Responsibility in Education and Science*, 15(3), 181–188. <https://doi.org/10.7160/eriesj.2022.150305>
- Chong, T., Yu, T., Keeling, D. I., & de Ruyter, K. (2021). AI-Chatbots on the services frontline addressing the challenges and opportunities of agency. *Journal of Retailing and Consumer Services*, 63, 102735. <https://doi.org/10.1016/j.jretconser.2021.102735>
- Chuvgunova, O., & Kostromina, S. (2016). Planning as a learning skill of students. *Procedia—Social and Behavioural Sciences*, 217, 132–138. <https://doi.org/10.1016/j.sbspro.2016.02.045>
- Cilliers, E. J. (2017). The challenge of teaching generation Z. *PEOPLE: International Journal of Social Sciences*, 3(1), 188–198. <https://doi.org/10.20319/pijss.2017.31.188198>
- Cooper, G. (2023). Examining science education in ChatGPT: An exploratory study of generative artificial intelligence. *Journal of Science Education and Technology*, 32(3), 444–452. <https://doi.org/10.1007/s10956-023-10039-y>
- Criado, N., & Such, J. M. (2019). Digital discrimination. *Algorithmic Regulation*, 5, 82–97. <https://nms.kcl.ac.uk/hasp/pubs/Digital-Discrimination-AV.pdf>

- Deebel, N. A., & Terlecki, R. (2023). ChatGPT performance on the American Urological Association (AUA) self-assessment study program and the potential influence of artificial intelligence (AI) in urologic training. *Urology*, 177, 29–33. <https://doi.org/10.1016/j.urology.2023.05.010>
- Dehouche, N. (2021). Plagiarism in the age of massive generative pre-trained transformers (GPT-3). *Ethics in Science and Environmental Politics*, 21, 17–23. <https://doi.org/10.3354/esepe00195>
- Deng, J., & Lin, Y. (2023). The benefits and challenges of ChatGPT: An overview. *Frontiers in Computing and Intelligent Systems*, 2(2), 81–83. <https://doi.org/10.54097/fcis.v2i2.4465>
- Dev Singh, K. (2015). Creating your own qualitative research approach: Selecting, integrating and operationalizing philosophy, methodology and methods. *Vision: The Journal of Business Perspective*, 19(2), 132–146. <https://doi.org/10.1177/0972262915575657>
- Dhivyadeepa, E. (2015). *Sampling techniques in educational research*. Lulu Press.
- Döringer, S. (2021). 'The problem-centred expert interview'. Combining qualitative interviewing approaches for investigating implicit expert knowledge. *International Journal of Social Research Methodology*, 24(3), 265–278. <https://doi.org/10.1080/13645579.2020.1766777>
- Drucker, P. (2007). *Management challenges for the 21st century (Classic Drucker Collection)*. Routledge. <https://www.abebooks.com/9780750685092/Management-Challenges-21st-Century-Peter-0750685093/plp>
- Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., & Wright, R. (2023). 'So what if ChatGPT wrote it?' Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, 102642. <https://doi.org/10.1016/j.ijinfomgt.2023.102642>
- Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics*, 5(1), 1–4. <https://doi.org/10.11648/j.ajtas.20160501.11>
- Favero, T. G. (2011). Active review sessions can advance student learning. *Advances in Physiology Education*, 35(3), 247–248. <https://doi.org/10.1152/advan.00040.2011>
- Financial Times (2023). Google invests \$300mn in artificial intelligence start-up Anthropic. Retrieved 23 March 2023 from <https://www.ft.com/content/583ead66-467c-4bd5-84d0-ed5df7b5bf9c>.
- Fisher, I. E., Garnsey, M. R., & Hughes, M. E. (2016). Natural language processing in accounting, auditing and finance: A synthesis of the literature with a roadmap for future research. *Intelligent Systems in Accounting, Finance and Management*, 23(3), 157–214. <https://doi.org/10.1002/isaf.1386>
- Gibbert, M., Ruigrok, W., & Wicki, B. (2008). What passes as a rigorous case study? *Strategic Management Journal*, 29(13), 1465–1474. <https://doi.org/10.1002/smj.722>
- Gozalo-Brizuela, R., & Garrido-Merchán, E. (2023). ChatGPT is not all you need: A state-of-the-art review of large generative AI models. *ArXiv*. <https://doi.org/10.48550/arXiv.2301.04655>
- Green, S., Heer, J., & Manning, C. D. (2015). Natural language translation at the intersection of AI and HCI. *Queue*, 13(6), 30–42. <https://doi.org/10.1145/2791301.2798086>
- Gunawan, J. (2023). Exploring the future of nursing: Insights from the ChatGPT model. *Belitung Nursing Journal*, 9(1), 1–5. <https://doi.org/10.33546/bnj.2551>
- Hasas, A., Hakimi, M., Shahidzay, A. K., & Fazil, A. W. (2024). AI for social good: Leveraging artificial intelligence for community development. *Journal of Community Service and Society Empowerment*, 2(2), 196–210. <https://doi.org/10.59653/jcsse.v2i02.592>
- Hernandez-de-Menendez, M., Escobar Díaz, C., & Morales-Menendez, R. (2020). Technologies for the future of learning: state of the art. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, 14(2), 683–695.
- Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. *Science*, 349(6245), 261–266. <https://doi.org/10.1126/science.aaa8685>
- Jensen, L. X., Buhl, A., Sharma, A., & Bearman, M. (2024). Generative AI and higher education: A review of claims from the first months of ChatGPT. *Higher Education*, 15, 1–7. <https://doi.org/10.1007/s10734-024-01265-3>
- Just, J. (2024). Natural language processing for innovation search – reviewing an emerging non-human innovation intermediary. *Technovation*, 129, 102883. <https://doi.org/10.1016/j.technovation.2023.102883>
- Kaddari, Z., Mellah, Y., Berrich, J., Belkasm, M. G., & Bouchentouf, T. (2020). Natural language processing: Challenges and future directions. *Artificial Intelligence and Industrial Applications*, 144, 236–246. https://doi.org/10.1007/978-3-030-53970-2_22
- Kaplan, D. E. (2019). Creativity in education: Teaching for creativity development. *Psychology*, 10(2), 140–147. <https://doi.org/10.4236/psych.2019.102012>
- Khademi, A. (2023). Can ChatGPT and Bard generate aligned assessment items? A reliability analysis against human performance. *Journal of Applied Learning & Teaching*, 6(1), 75–80. <https://doi.org/10.48550/arXiv.2304.05372>
- Kolbjørnsrud, V. (2024). Designing the intelligent organization: Six principles for human–AI collaboration. *California Management Review*, 66(2), 44–64. <https://doi.org/10.1177/00081256231211020>

- Korzynski, P., Mazurek, G., Altmann, A., Ejdays, J., Kazlauskaitė, R., Paliszkievicz, J., & Ziemba, E. (2023). Generative artificial intelligence as a new context for management theories: Analysis of ChatGPT. *Central European Management Journal*, 31(1), 3–13. <https://doi.org/10.1108/CEMJ-02-2023-0091>
- Kothari, A., & Cruikshank, S. A. (2021). Artificial intelligence and journalism: An agenda for journalism research in Africa. *African Journalism Studies*, 43(1), 17–33. <https://doi.org/10.1080/23743670.2021.1999840>
- Kumar, S., Ghai, V., Jha, A., & Sharma, S. (2022). Role of artificial intelligence in generating video, paper presented at the 8th International Conference on Advanced Computing and Communication Systems (ICACCS).
- Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., & Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *The International Journal of Management Education*, 21(2), 100790. <https://doi-org.libaccess.hud.ac.uk/10.1016/j.ijme.2023.100790>
- Luiz de Mattos Nascimento, D., Alencastro, V., Luiz Gonçalves Quelhas, O., Caiado, R., Arturo Garza-Reyes, J., Rocha Lona, L., & Luz Tortorella, G. (2019). Exploring Industry 4.0 technologies to enable circular economy practices in a manufacturing context – a business model proposal. *Journal of Manufacturing Technology Management*, 30(3), 607–627. <https://doi.org/10.1108/JMTM-03-2018-0071>
- Lund, B. D., & Wang, T. (2023). Chatting about ChatGPT: How may AI and GPT impact academia and libraries? Library Hi Tech News, ahead-of-print(ahead-of-print). <https://doi.org/10.1108/LHTN-01-2023-0009>
- Maloni, M., Hiatt, M. S., & Campbell, S. (2019). Understanding the work values of gen Z business students. *The International Journal of Management Education*, 17(3), 100320. <https://doi.org/10.1016/j.ijme.2019.100320>
- Marconi, F. (2020). *Newsmakers: Artificial intelligence and the future of journalism*. Columbia University Press. <https://doi.org/10.7312/marc19136>
- McEwen, B. C. (1994). Teaching critical thinking skills in business education. *Journal of Education for Business*, 70(2), 99–103. <https://doi.org/10.1080/08832323.1994.10117733>
- McGrath, C., Farazouli, A., & Cerratto-Pargman, T. (2024). Generative AI chatbots in higher education: A review of an emerging research area. *Higher Education*, 24, 1–7. <https://doi.org/10.1007/s10734-024-01288-w>
- Mogali, S. R. (2024). Initial impressions of ChatGPT for anatomy education. *Anatomical Sciences Education*, 17(2), 444–447. <https://doi.org/10.1002/ASE.2261>
- Mohanty, H. (2021). NLP for sentiment computation. *Studies in Computational Intelligence*, 2021, 1–21. https://doi.org/10.1007/978-981-33-6815-6_1
- Moro, A., Boelman, E., Joanny, G., & Lopez-Garcia, J. (2018). A bibliometric-based technique to identify emerging photovoltaic technologies in a comparative assessment with expert review. *Renewable Energy*, 123, 407–416. <https://doi.org/10.1016/J.RENENE.2018.02.016>
- Narayan, K. G., Sinha, D. K., & Singh, D. K. (2023). Sampling techniques. In *Veterinary public health & epidemiology* (pp. 111–123). Springer Nature. <https://doi.org/10.1007/978-981-19-7800-5>
- Navaresta, D. (2021). The use of brainstorming in learning writing. *Journal of English Language and Education*, 6(1), 1–9. <https://doi.org/10.26486/jele.v6i1.1021>
- Noor, S., Tajik, O., & Golzar, J. (2022). Simple random sampling. *International Journal of Education & Language Studies*, 1(2), 78–82. <https://doi.org/10.22034/ijels.2022.162982>
- Odiba, I. A., & Baba, P. A. (2013). Using reflective thinking skills for education quality improvement in Nigeria. *Journal of Education and Practice*, 4, 196–201.
- Packin, N. G., & Lev-Aretz, Y. (2018). Learning algorithms and discrimination. In *Research handbook on the law of artificial intelligence* (pp. 88–113). Edward Elgar.
- Perkins, M., Furze, L., Roe, J., & MacVaugh, J. (2024). The artificial intelligence assessment scale (AIAS): A framework for ethical integration of generative AI in educational assessment. *Journal of University Teaching and Learning Practice*, 21(6), 1–18. <https://doi.org/10.53761/q3azde36>
- Prajapati, J. B., Kumar, A., Singh, S., Prajapati, B., Thakar, Y., Tambe, P. R., & Ved, A. (2024). Artificial intelligence-assisted generative pretrained transformers for applications of ChatGPT in higher education among graduates. *SN Social Sciences*, 4(2), 19. <https://doi.org/10.1007/s43545-023-00818-0>
- Pugalenth, R., Prabhu Chakkaravarthy, A., Ramya, J., Babu, S., & Rasika Krishnan, R. (2020). Artificial learning companion using machine learning and natural language processing. *International Journal of Speech Technology*, 24(3), 553–560. <https://doi.org/10.1007/s10772-020-09773-0>
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI Blog*, 24(2), 1–8. <https://doi.org/10.1007/s10772-020-09773-0>
- Rahman, H. U., & Kodikal, R. (2024). Revolutionizing education: Artificial intelligence empowered learning in higher education. *Cogent Education*, 11(1), 2293431. <https://doi.org/10.1080/2331186X.2023.2293431>
- Rao, Z. (2007). Training in brainstorming and developing writing skills. *ELT Journal*, 61(2), 100–106. <https://doi.org/10.1093/elt/ccm002>
- Ratten, V., & Jones, P. (2023). Generative artificial intelligence (ChatGPT): Implications for management educators. *The International Journal of Management Education*, 21(3), 1–7. <https://doi.org/10.1016/j.ijme.2023.100857>
- Robinson, M. A., Sparrow, P. R., Clegg, C., & Birdi, K. (2007). Forecasting future competency requirements: A three-phase methodology. *Personnel Review*, 36(1), 65–90. <https://doi.org/10.1108/00483480710716722/FULL/PDF>

- Rodgers, C. (2002). Defining reflection: Another look at John Dewey and reflective thinking. *Teachers College Record: The Voice of Scholarship in Education*, 104(4), 842–866. <https://doi.org/10.1111/1467-9620.00181>
- Roshan, A., & Meenu Garg, D. N. (2020). Intelligent voice assistant for desktop using NLP and AI. *International Journal for Modern Trends in Science and Technology*, 6(12), 328–331. <https://doi.org/10.46501/IJMTST061261>
- Rudolph, J., Tan, S., & Tan, S. (2023a). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching*, 6(1), 342–363. <https://doi.org/10.37074/jalt.2023.6.1.9>
- Rudolph, J., Tan, S., & Tan, S. (2023b). War of the chatbots: Bard, Bing chat, ChatGPT, Ernie and beyond: The new AI gold rush and its impact on higher education. *Journal of Applied Learning and Teaching*, 6(1), 364–389. <https://doi.org/10.37074/jalt.2023.6.1.23>
- Sadiku, M. N. O., Zhou, Y., & Musa, S. M. (2018). Natural language processing in healthcare. *International Journal of Advanced Research in Computer Science and Software Engineering*, 8(5), 1–17.
- Seemiller, C., & Grace, M. (2017). Generation Z: Educating and engaging the next generation of students. *About Campus: Enriching the Student Learning Experience*, 22(3), 21–26. <https://doi.org/10.1002/abc.21293>
- Selwyn, N. (2022). The future of AI and education: Some cautionary notes. *European Journal of Education*, 57(4), 620–631. <https://doi.org/10.1111/ejed.12532>
- Skavronskaya, L., Hadinejad, A., & Cotterell, D. (2023). Reversing the threat of artificial intelligence to opportunity: A discussion of ChatGPT in tourism education. *Journal of Teaching in Travel & Tourism*, 23(2), 253–258. <https://doi.org/10.1080/15313220.2023.2196658>
- Smith, P. A. (2007). Case study: Planning as learning. *Action Learning: Research and Practice*, 4(1), 77–86. <https://doi.org/10.1080/14767330701233897>
- Solangi, Y. A., Solangi, Z. A., Aarain, S., Abro, A., Mallah, G. A., & Shah, A. (2018). Review on natural language processing (NLP) and its toolkits for opinion mining and sentiment analysis. In *Proceedings of the 2018 IEEE 5th International Conference on Engineering Technologies and Applied Sciences (ICETAS)* (pp. 1–4). IEEE. <https://doi.org/10.1109/ICETAS.2018.8629198>
- Sri, M. (2020). *Practical natural language processing with Python: With case studies from industries using text data at scale*. Apress. <https://doi.org/10.1007/978-1-4842-6246-7>
- Srinivasan, R., & Chander, A. (2021). Biases in AI systems. *Communications of the ACM*, 64(8), 44–49. <https://doi.org/10.1145/3464903>
- The Economist. (2021). Investors are going nuts for CHATGPT-ish artificial intelligence. The Economist, Retrieved 22 March 2023 from <https://www.economist.com/business/2023/02/28/investors-are-going-nuts-for-chatpt-ish-artificial-intelligence>
- The New York Times. (2023a). How CHATGPT kicked off an AI arms race. The New York Times. <https://www.nytimes.com/2023/03/21/technology/google-bard-chatbot.html>; <https://www.nytimes.com/2023/02/03/technology/chatgpt-openai-artificial-intelligence.html>
- The New York Times. (2023b). Google releases Bard, its competitor in the race to create AI chatbots. The New York Times. <https://www.nytimes.com/2023/03/21/technology/google-bard-chatbot.html>
- Thorp, H. H. (2023). ChatGPT is fun, but not an author. *Science*, 379(6630), 313. <https://doi.org/10.1126/science.adg7879>
- Túñez-López, J.-M., Fieiras-Ceide, C., & Vaz-Álvarez, M. (2021). Impact of artificial intelligence on journalism: Transformations in the company, products, contents and professional profile. *Communications Society*, 34(1), 177–193. <https://doi.org/10.15581/003.34.1.177-193>
- Ustun, A. B., & Tracey, M. W. (2020). An effective way of designing blended learning: A three phase design-based research approach. *Education and Information Technologies*, 25(3), 1529–1552. <https://doi.org/10.1007/S10639-019-09999-9/TABLES/5>
- Wen, A., Fu, S., Moon, S., El Wazir, M., Rosenbaum, A., Kaggal, V. C., Liu, S., Sohn, S., Liu, H., & Fan, J. (2019). Desiderata for delivering NLP to accelerate healthcare AI advancement and a Mayo Clinic NLP-as-a-service implementation. *npj Digital Medicine*, 2(1), 1–7.
- Weyerer, J. C., & Langer, P. F. (2020). Bias and discrimination in artificial intelligence. *Interdisciplinary Approaches to Digital Transformation and Innovation*, 2020, 256–283. <https://doi.org/10.4018/978-1-7998-1879-3.ch011>
- Wood, D. A., Achhpilia, M. P., Adams, M. T., Aghazadeh, S., Akinyele, K., Akpan, M., Allee, K., Allen, A., Almer, E., Ames, D., Arity, V., Barr-Pulliam, D., Basoglu, K., Belnap, A., Bentley, J., Berg, T., Berglund, N., Berry, E., Bhandari, A., ... Kuruppu, C. (2023). The ChatGPT artificial intelligence chatbot: How well does it answer accounting assessment questions? *Issues in Accounting Education*, 38(4), 81–108. <https://doi.org/10.2308/issues-2023-013>
- Wu, T., & Yu, Z. (2024). Bibliometric and systematic analysis of artificial intelligence chatbots' use for language education. *Journal of University Teaching and Learning Practice*, 21(6), 1–19. <https://doi.org/10.53761/6jmnst54>
- Xing, F. Z., Cambria, E., & Welsch, R. E. (2017). Natural language based financial forecasting: A survey. *Artificial Intelligence Review*, 50(1), 49–73. <http://hdl.handle.net/1721.1/116314>

- Yenduri, G., Ramalingam, M., Selvi, G. C., Supriya, Y., Srivastava, G., Maddikunta, P. K. R., & Gadekallu, T. R. (2024). GPT (generative pre-trained transformer) – a comprehensive review on enabling technologies, potential applications, emerging challenges, and future directions. *IEEE Access*, 12, 54608–54649. <https://doi.org/10.1109/ACCESS.2024.3389497>
- Yildiz, M. S. (2023). Comparing response performances of Chatgpt-3.5, Chatgpt-4 and Bard to health-related questions: Comprehensiveness, accuracy and being up to date. <https://doi.org/10.2139/ssrn.4503443>
- Ying, L., Liu, Z., Fang, H., Kusko, R., Wu, L., Harris, S., & Tong, W. (2024). Text summarization with ChatGPT for drug labeling documents. *Drug Discovery Today*, 29(6), 104018. <https://doi.org/10.1016/j.drudis.2024.104018>
- Zhou, M., Duan, N., Liu, S., & Shum, H.-Y. (2020). Progress in neural NLP: Modelling, learning, and reasoning. *Engineering*, 6(3), 275–290. <https://doi.org/10.1016/j.eng.2019.12.014>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Herath, D. B., Ode, E. & Herath, G. B. (2025). Can AI replace humans? Comparing the capabilities of AI tools and human performance in a business management education scenario. *British Educational Research Journal*, 51, 1073–1096. <https://doi.org/10.1002/berj.4111>