CHATGPT AND BARD IN EDUCATION: A COMPARATIVE REVIEW

Gustavo Simas da Silva and Vânia Ribas Ulbricht
Federal University of Santa Catarina (UFSC), Brazil
R. Eng. Agronômico Andrei Cristian Ferreira, s/n - Trindade, Florianópolis - SC, Brazil

ABSTRACT

ChatGPT and Bard, two chatbots powered by Large Language Models (LLMs), are propelling the educational sector towards a new era of instructional innovation. Within this educational paradigm, the present investigation conducts a comparative analysis of these groundbreaking chatbots, scrutinizing their distinct operational characteristics and applications as depicted in current scholarly discourse. ChatGPT emerges as an exemplary tool in task-oriented textual interactions, while Bard brandishes unique features such as Text-To-Speech (TTS) functionality, which enhances accessibility and inclusive education, as well as integration with Google Workspace applications. This research critically examines their utilization in various spheres such as pedagogy, academic research, Massive Open Online Courses (MOOCs), mathematics, and software programming. Findings accentuate ChatGPT's superior efficacy in content drafting, code generation, language translation, and providing clinically precise responses, notwithstanding Bard's significant potential encapsulated in its exclusive features. Furthermore, the study traverses' crucial ethical aspects, including privacy concerns and inherent bias, underscoring the profound implications of these Artificial Intelligence (AI) technologies on literature and advocating against the indiscriminate reliance on such models.

KEYWORDS

ChatGPT, Bard, Education, Large Language Models, Chatbots, Artificial Intelligence

1. INTRODUCTION

The domain of Artificial Intelligence (AI) has experienced considerable progression, particularly within the subfield of conversational AI. This progression has been marked by the advent of sophisticated Large Language Models (LLMs) demonstrating the capacity for generating text that strikingly mirrors human discourse and has the potential to transform various aspects of our lives, including education (Tlili et al., 2023).

Two eminent examples of such advanced models are OpenAI's ChatGPT and Google's Bard. Chatbots like these are being analyzed in the literature concerning its potential and risks in healthcare (Sallam, 2023), consumer studies (Paul et al., 2023), educational settings (Lund & Wang, 2023) and a wide variety of areas. Despite the recent enthusiasm, there are also voices that approach its use cautiously. Investigations of user experiences revealed various issues, including cheating, honesty and truthfulness of ChatGPT, privacy misleading, and manipulation, which highlight the need for research directions that ensure a safe and responsible adoption of chatbots, especially in learning and education (Tlili et al., 2023).

Besides that, it is observed that recent discussions need to tackle advantages of emerging technologies and cutting-edge chatbots or AI assistants like ChatGPT and Bard, but also highlight the ethical and practical challenges associated with their use in education. There is emphasis on the importance of responsible and ethical use of AI in education while promoting its successful incorporation into the educational setting to benefit teachers and students (Adiguzel et al., 2023). Dis and colleagues (2023) point out the challenges associated with generative AI and lay out an agenda for future research. This agenda underscores several priority areas including the necessity for human verification, the formulation of accountability measures, the investment in genuinely open LLMs, the embrace of AI benefits, and the expansion of the discourse on these technologies.

Therefore, in accordance with the aforementioned scope, the fundamental research question this article aims to answer is: "In what ways can ChatGPT and Bard be utilized comparatively within the educational sector?" This study embarks on a comparative analysis, delving into the functionalities and features of Google's Bard and ChatGPT in the context of education, while leveraging pertinent current literature. Viewing these tools through the educational prism enables an exploration of their utility and the quality of responses provided to users. Concurrently, it illuminates the ethical considerations integral to their deployment. By delving into these aspects, the study hopes to provide a comprehensive understanding of the role these language models can play in educational settings, their strengths and limitations, and the ethical implications of their usage.

2. CHATGPT AND BARD

ChatGPT, Bard and other conversational systems are interfaces that permit Human-AI Interaction (HAI) powered by Large Language Models. LLMs embody the quintessence of Deep Learning (DL) models, expertly engineered to decipher and generate cogent responses using a multilayer neural network configuration (Rahsepar et al., 2023). These models are trained utilizing colossal volumes of textual data, employing unsupervised learning techniques, and are capable of discerning relationships between words within the text. Moreover, they boast the ability to predict the succeeding word in a string of words, based on their antecedents. LLMs, unique in their construction, are tailored specifically for natural language processing tasks, including but not limited to, language translation, text summarization, and question-answering. This differentiates them from other DL model variants such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), which are predominantly used for tasks related to image and speech recognition (LeCun et al., 2015).

The interaction between Artificial Intelligence and education is not a novel concept, as AI has been progressively integrated into Learning Management Systems (LMS) to augment the overall pedagogical experience. However, leveraging the capabilities of Natural Language Processing (NLP) and cognitive reasoning within LMS has fostered a more interactive and personalized learning environment (Ali et al., 2023). The amalgamation of AI and NLP in an educational context holds the potential to enhance both communication and interaction between students and their virtual pedagogical counterparts or peers (Kasneci et al., 2023). Yet, with the advent of Large Language Models, the landscape of AI interactions has experienced a paradigm shift. Individuals are now able to engage with AI systems using natural language, a revolution that has broadened the horizon for Human-AI interaction, decision-making processes, and natural language comprehension across diverse domains (Clavié, 2023).

As of July, 2023, ChatGPT is based on Generative Pre-trainer Transformer (GPT), version 3.5 and 4 (OpenAI, 2023), while Bard is based on Pathways Language Model (PaLM), version 2 (Google, 2023b). Both are equipped with Application Programming Interfaces (APIs) to facilitate their integration with external applications. Unique features of OpenAI's ChatGPT system encompass the availability of plugin features within the premium version (ChatGPT Plus). In contrast, Google's Bard offers distinctive capabilities including real-time internet information retrieval, image utilization in prompts and a Text-to-Speech (TTS) feature to audibly relay responses.

3. METHODOLOGY

This research unfolds a scoping review through distinct phases. Firstly, a review protocol predicated on the PCC framework (Population, Concept, Context), an established tool utilized across diverse research domains to direct the formation of research questions and guide the process of scoping reviews (Pollock et al., 2023). By providing a structured approach to these elements, the PCC framework enables researchers to generate more focused and contextually relevant research questions, thereby enhancing the potential significance and applicability of the findings in the wider scholarly discourse. This study has:

- Population: students, teachers and self-directed learners in any education level;
- Concept: the use of AI powered chatbots, specifically ChatGPT and Bard;
- Context: learning and educational settings.

Secondly, with the research question formulated, the selection criteria are derived, encapsulating both inclusion and exclusion parameters, which subsequently informs data extraction and data synthesis processes.

Given the emergent and novel nature of the topic, inclusion and exclusion criteria for this study have been crafted such that the articles embody keywords such as 'ChatGPT', 'Bard', 'chatbots', 'Large Language Models' and 'Education', and are indexed in databases such as Scopus, Web of Science, IEEEXplore, along with a deep-dive into gray literature, like technical reports and companies websites, to compensate for the relative dearth of formal literature on the topic.

The search horizon was set to publications issued in 2022 and 2023, written in English, with open access, in line with the developmental timeline of ChatGPT and Google's Bard. The search process is conducted manually by traversing specific journal papers in the identified databases. At this juncture, a total of 230 articles were discovered. Consequently, a process of removing duplicates and excluding papers that did not directly address the use of chatbots, specifically ChatGPT or Bard, in educational contexts was conducted based on the publication's titles and abstracts. Lastly, after a quality assessment, 22 articles were finalized for this research study.

4. RESULTS AND DISCUSSION

ChatGPT, specifically calibrated for task-oriented textual interactions, serves manifold purposes encompassing language translation, composition of product descriptions, and summarization of transcripts. Given its expansive training dataset compared to LaMDA (the previous LLM behind Bard), ChatGPT displays superior performance in more intensive tasks such as drafting content, generating code, and executing translations (Ahmed et al., 2023). Conversely, Bard is primarily designed to retrieve information through succinct responses, akin to digital assistants like Alexa and Siri. Bard's distinctive focus on the generation of creative language imbues it with utility across a wide array of applications, spanning writing, publishing, marketing, and advertising.

Lund and Wang (2023) discussed the benefits of ChatGPT, such as improving search and discovery, reference and information services, cataloging and metadata generation, and content creation. However, the study also emphasized the ethical considerations that need to be taken into account, such as privacy and bias.

In the learning field, educators can use ChatGPT to create role-playing exercises or simulate the writing style of famous authors, for example. This can be used to attract students who are not interested in the mainstream teaching style but find, for instance, contemporary music more relatable. By adapting a generated or existing text to the style of different genres, educators manage to retain the scientific integrity of their educational content while increasing its relevance (Panagopoulou et al., 2023). Also, ChatGPT can be used to generate human-like responses to student queries, engage in critical thinking, and assist with idea generation (Halaweh, 2023).

Another option is that of generating pros and cons with respect to a specific issue. ChatGPT has the potential to "humanize" web search, i.e., help users locate and retrieve information in the same manner as asking a fellow or colleague (Kumar, 2023). A set of pros and cons can be used either as part of a more general research project or as part of a debate exercise, where students are asked to support or find weaknesses to a specific argument.

Regarding online learning, specifically in Massive Open Online Courses (MOOCs), some implications were explored in a study (Alseddiqi et al., 2023). The work discussed the benefits of ChatGPT in addressing challenges associated with MOOCs, such as lack of personalized support and feedback, approaching the capabilities of ChatGPT, such as language understanding, response generation, personalization, and continuous learning. The study conducted experiments to measure the accuracy and effectiveness of ChatGPT in answering questions from various subject areas and difficulty levels. The potential impact of this technology on students' learning skills was discussed, concluding that it is highly effective in answering multiple-choice and true or false questions from various MOOCs platforms. The findings suggest that ChatGPT can actually enhance students' critical thinking skills by providing them with more opportunities to apply their knowledge in real-world contexts.

Meanwhile, Bard, in its current iteration, demonstrates limitations in generating original content following the process of paraphrasing, with similar struggles observed in crafting responses to inquiries (Aydin, 2023). Upon comparison with ChatGPT, a congruent outcome is discerned in the context of paraphrased text generation. However, ChatGPT exhibits lower matching rates with respect to replication or plagiarism in the responses it generates, indicating a greater degree of originality. The implications of these findings suggest that the utilization of these and related AI applications could potentially lead to a divergence in scholarly literature.

Regarding the field of mathematics, Friedes and colleagues (2023) evaluated ChatGPT for its mathematical capabilities. It has been tested on publicly available datasets and hand-crafted ones to measure its performance in tasks such as question answering and theorem searching. ChatGPT's performance has been compared to other models trained on mathematical corpora and the development of datasets that cover advanced mathematics has been proposed to further evaluate ChatGPT's mathematical comprehension.

With respect to academic research, ChatGPT has been used to complete various tasks, including essay writing, speech writing, summarizing literature, and generating ideas. It was discussed that it has the potential to assist researchers in idea generation and scientific writing (Rahman et al., 2023). However, challenges have been identified in areas such as literature synthesis, citations, problem statements, research gaps, and data analysis. From that, researchers need to be cautious when using ChatGPT in academic research, check the veracity of references and establish necessary guidelines.

Not only between themselves, but also comparisons between these systems and traditional search mechanisms are being researched. Sezgin et al. (2023) conducted a study evaluating the response quality of ChatGPT, Google Bard and Google Search, with respect to recurrently queried information about Postpartum Depression (PPD). The evaluation criterion was based on clinical accuracy and the results indicated a relatively superior performance by ChatGPT in generating clinically accurate responses, as compared to Google Bard and Google Search. This finding posits that, particularly in the context of addressing mental health disorders like PPD, ChatGPT might possess a competitive edge in terms of precision.

Also, the most recent version, GPT-4, advances ChatGPT's proficiency, enabling it to supersede Bard in the domain of imaging-related tasks, while Bard exhibits a heightened propensity towards generating hallucinations (Ahmed et al., 2023). Notably, Bard responded incorrectly to between 50% and 75% of elementary SAT queries. Despite these shortcomings, Bard has seen substantial improvements in its latest update, displaying newfound capabilities in code generation, debugging, and providing explanations for code spanning over 20 programming languages, inclusive of C++, Java, JavaScript, and Python.

Although there are some investigations about programming and software development, the most researched area is, currently, concerning the health sector and medical themes. Rahsepar et al. (2023) embarked on a comparative analysis of the performance metrics of ChatGPT-3.5 and Google Bard, in addressing medically-themed questions. A discernible disparity was observed, with ChatGPT-3.5 demonstrating a superior accuracy and consistency in the delivery of correct responses in contrast to Bard. However, it is crucial to acknowledge that neither of the models achieved a 100% success rate in providing correct and consistent answers. The study unveiled several restricting factors that could potentially influence the results as the scope of the sample size was limited, and the chosen lung cancer queries utilized in the research might not encompass the entire diversity of potential medical questions. Such limitations suggest cautious interpretation and more experiment replications to validate the findings.

About reliability analysis against human performance, the inter-reliability of the two chatbots was assessed against human evaluators in terms of recognizing and assessing the complexity of writing prompts (Khademi, 2023). It was discovered that the agreement between ChatGPT and human evaluators in assessing the perceived complexity of writing prompts was more pronounced ($r = .22$) compared to the agreement between Bard and human evaluators ($r = .05$). In general, the outcomes revealed moderate to substantial concordance between the AI tools and human evaluators concerning the perceived complexity of the writing prompts. Nonetheless, the AI models demonstrated lower agreement with human evaluators when it came to assessing the quality of writing prompts. These findings indicate that while AI tools can potentially aid human evaluators in recognizing the complexity of writing prompts, their reliability in substituting human evaluators in assessing the quality of these prompts might be questionable.

Table 1. Comparison between ChatGPT and Bard capabilities, limitations and outcomes regarding education

System	Capabilities	Limitations	Outcomes
ChatGPT	LLMs usual capabilities (NLP, code interpretation, etc.) Computer Vision (OpenAI, 2023) Superior accuracy and consistency (Alseddiqi et al, 2023; Sezgin et al., 2023; Rahsepar et al., 2023) Plugins (ChatGPT Plus) GPT API	LLMs usual limitations (hallucinations, ethical considerations such as privacy and bias, etc.) (Sallam, 2023; Irfan, 2023; Koos, 2023; Panagopoulou, 2023; Rahman & Watanobe, 2023) Lack of different idioms in the interface	Improved teaching experience (Tlili et al., 2023; Siegle, 2023; Kasneci et al., 2023; Rahman & Watanobe, 2023) Improved learning experience (Kasneci et al., 2023; Alseddiqi et al., 2023; Rahman & Watanobe, 2023) Enhanced critical thinking (Panagopoulou et al., 2023; Sharma et al., 2023; Siegle, 2023) New teaching philosophy (Tlili et al., 2023) Upskilling (Tlili et al., 2023) Increased accessibility (Rahman et al., 2023; Siegle, 2023; Kasneci et al., 2023; Rahman & Watanobe, 2023)
Bard	LLMs usual capabilities (NLP, code interpretation, etc.) Computer Vision (Google, 2023c) Current real time internet information retrieval (Google, 2023a) Integration with Gmail and productivity applications from Google Workspace (Google, 2023c) Files and images in prompts and responses (Google, 2023c) Accessibility features (TTS) (Google, 2023c) PaLM API	LLMs usual limitations (hallucinations, ethical considerations such as privacy and bias, etc.) (Sallam, 2023; Irfan, 2023; Koos, 2023; Panagopoulou, 2023; Rahman & Watanobe, 2023) Limitations in generating original content (Aydin, 2023)	

With respect to Citation Diversity Statement[1], an experimental study by King (2023) investigated the potential of Bard to assist researchers in analyzing the diversity of their scientific citation practices. The premise was to leverage the chatbot's capabilities to gauge the representation of diverse groups in referenced

[1] The goal of this statement is to prompt authors to consciously evaluate the proportion of their cited references that have been authored by women or people of color - demographics that have been historically underrepresented in scientific literature. The intent is to encourage researchers to enhance the diversity of the authors they cite in their scholarly work.

scientific literature. However, it was observed that Bard, in its current iteration, fell short of delivering a reliable performance for this task. Despite demonstrating a slight improvement in the accuracy of reference demographics, it was evident that the chatbot's live search functionalities were not fully optimized for this application. In response to this analytical discrepancy, Bard quickly agreed to the proposed explanation, affirming that it had indeed based its demographic analysis on prior average data. Nevertheless, the author remains hopeful about the potential for leveraging such technology for diversity analysis in scientific references as Bard continues to evolve and improve.

The comparisons between OpenAI's and Google's chatbots even join the field of personalities and anthropomorphism. In a comparative examination, distinct personality types were found to be consistently exhibited. ChatGPT, regardless of contextual variations or instructions, invariably maintains a Myers-Briggs Type Indicator (MBTI) ENFJ (Extraversion, Intuition, Feeling, Judging) personality type. Conversely, Bard corresponds to an ISTJ (Introversion, Sensing, Thinking, Judging) type. A key facet of the study was the examination of chatbots' personality consistency across seven additional languages: Chinese, Korean, Spanish, French, German, Italian, and Arabic. Using 60 questions from the 16Personalities tool, translated into the aforementioned languages, the consistency of the tools exhibited personality was scrutinized. A focal point of this investigation was to ascertain whether LLMs could manifest changes in their exhibited personalities in response to instructional or contextual cues. Despite the variations of "happy" and "sad" contexts, the results underscored that ChatGPT unwaveringly maintains its original personality type, defined as ENFJ, unaffected by contextual alterations or provided instructions. This illustrates the consistent nature of LLM personalities, and could serve as a reference point for future research on LLM behavior and potential manipulation.

Having this, Table 1 shows a compiled comparison between ChatGPT and Bard capabilities, limitations and outcomes regarding education. Indeed, it is critical to highlight that this synthesis of information is current as of July 2023. Given the swift progression of advancements in the Artificial Intelligence field, the capabilities and limitations of these tools may rapidly be updated over time. Nonetheless, the core implications and outcomes delineated in this comparison should persist, illustrating the potential that these sophisticated AI models, ChatGPT and Bard, possess in reshaping our engagement with technology.

5. CONCLUSION

This article reviewed the potentials and risks regarding ChatGPT and Bard, comparatively, in the educational field. In terms of opportunities, these technologies can be employed as effective tools for personalized learning support, as it can analyze student performance data, identify areas where students are facing challenges, and generate customized resources and learning activities to boost their skill sets. Furthermore, ChatGPT's and Bard's proficiency in understanding and answering learners' questions can enhance the learning experience by providing comprehensive answers and additional pertinent resources, such as articles, videos, or tutorials. The application of such chatbots also extends to academic evaluation and assessment, where it can swiftly assess and grade assignments and quizzes, including the generation of variable-difficulty questions. Lastly, OpenAI's and Google's systems can support research activities by summarizing published work, suggesting novel research ideas, and even generating code.

However, the usage of those LLMs powered tools also poses certain threats: for instance, its capacity to generate human-like text can potentially compromise the integrity of online exams and assignments due to the risk of cheating. To counter this, it is necessary to educate students about the repercussions of academic dishonesty and employ plagiarism-detection tools to identify AI-generated content. Another issue is the risk of over-reliance on AI, which could undermine problem-solving skills. To mitigate this, it is important to guide students on how to effectively utilize ChatGPT while also promoting a critical approach to the information it generates. Finally, the increasingly blurred line between AI-generated content and human-produced text presents an escalating challenge for educators and researchers. Overcoming this challenge requires the development of innovative technologies that can distinguish between human and AI-generated content and label them, providing clear disclaimers.

Also, many ethical questions emerge within this scope. A body of literature has deliberated on the ethical considerations pertinent to the employment of ChatGPT in educational settings (Sallam, 2023; Irfan, 2023; Koos, 2023; Panagopoulou, 2023). These studies underscore the imperative of confronting ethical repercussions, fostering responsible utilization, and enlightening both educators and learners about the

prospective perils and constraints associated with LLMs. Additionally, these studies have proffered strategies and recommendations aimed at warranting the ethical incorporation of chatbots or AI assistants into pedagogical practices. AI systems' prospective contribution to the refinement of pedagogical methodologies, the fortification of health education and scientific exploration, the facilitation of mathematical problem-solving, and the advancement of academic inquiry holds considerable potential. Teachers must play a key role in mediating AI integration, ensuring a balance between technology usage and development of independent learning skills (Siegle, 2023).

A crucial concern warranting attention pertains to the potential implications of LLMs on diversity and existing disparities within research fields. While promising to democratize access to information by potentially eliminating language barriers and enhancing the quality of written output, these models might inadvertently exacerbate existing inequalities (Dis et al., 2023). Similar to past technological innovations, it is plausible that resource-rich countries and elite researchers could leverage LLMs to bolster their own research productivity, thereby widening the existing disparities. Consequently, it is imperative that discussions surrounding the implementation and governance of these "stochastic parrots" involve individuals from underrepresented groups in research and those from communities directly impacted by the research outcomes (Bender et al., 2021). This inclusivity is essential to ensure that lived experiences serve as valuable resources in shaping the trajectory and ethical use of these advanced AI models (Bolukbasi, 2016).

Also, there is a pressing need for further research and investigation to acquire a holistic understanding of future OpenSource technologies competencies and features vis-à-vis ChatGPT and Bard. Comparative research dedicated expressly to new LLMs, encompassing performance across various domains, would yield insightful revelations concerning their assets and constraints. Furthermore, it is of paramount importance to address ethical considerations and potential biases in HAI, in order to ensure ethical progression and deployment in diverse applications.

Addressing the limitations of general LLMs within specialized fields, domain-specific models are being trained on text corpora pertinent to specific areas of interest. These specialized LLMs, such as Google's Med-PaLM for the medical domain and Sec-PaLM for cybersecurity, will tend to exhibit improved performance within their respective disciplines, demonstrating an enhanced understanding and language generation in their specialized contexts (Google, 2023b). These models symbolize the convergence of advanced AI capabilities and field-specific requirements, promising enhanced efficacy and precision within these targeted domains.

As they persistently evolve, both ChatGPT and Bard find themselves in a competitive equilibrium, each progressively broadening their respective scopes of application (Ahmed et al., 2023). The acceleration of knowledge gathering and representation through AI seems imminent, simultaneously decreasing reliance on human-centered methods for these operations (Aydin, 2023). The emergence of these technologies promises an exciting future where more intelligent systems could become an integral part of our knowledge ecosystems.

REFERENCES

Adiguzel, T. et al. (2023). Revolutionizing education with AI: Exploring the transformative potential of ChatGPT. *Contemporary Educational Technology*, 15(3), ep429.

Ahmed, I. et al. (2023). ChatGPT vs. Bard: A Comparative Study. *In UMBC Student Collection.*

Ali, N. M. et al. (2023). Artificial Intelligence In Learning Management System: a Case Study Of The Students Of Mass Communication. *Voyage Journal of Educational Sciences*, 2(3), 92-114.

Alseddiqi, M. et al. (2023). Revolutionizing Online Learning: The Potential of ChatGPT in Massive Open Online Courses. *European Journal of Education and Pedagogy*, 4(4), 1-5.

Aydin, Ö. (2023). Google Bard generated literature review: metaverse. *Journal of AI*, 7(1), 1-14.

Bender, E. M. et al. (2021, March). On the dangers of stochastic parrots: Can language models be too big? 🦜 In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, pp. 610-623.

Bolukbasi, T. et al. (2016). Man is to computer programmer as woman is to homemaker? debiasing word embeddings. *Advances in neural information processing systems*, 29.

Clavié, B., et al. (2023). Towards Simple Hybrid Language Model Reasoning Through Human Explanations Enhanced Prompts. *In HHAI 2023: Augmenting Human Intellect*, pp. 379-381.

Dis, E. A. M. v. et al. (2023). Chatgpt: Five Priorities For Research. *Nature*, 7947(614), 224-226.

Frieder, S. et al. (2023). Mathematical capabilities of chatgpt. *arXiv preprint* arXiv:2301.13867.

Google (2023a), An important next step on our AI journey, https://blog.google/technology/ai/bard-google-ai-search-updates/, last accessed 2023/07/16.

Google (2023b), AI Across Google: PaLM 2, https://ai.google/discover/palm2/, last accessed 2023/07/17.

Google (2023c), Bard Updates, https://bard.google.com/updates, last accessed 2023/07/17.

Halaweh, M. (2023). Chatgpt In Education: Strategies For Responsible Implementation. *Contemporary Educational Technology*, 2(15), ep421.

Huang, J. T. et al. (2023). ChatGPT an ENFJ, Bard an ISTJ: Empirical Study on Personalities of Large Language Models. *arXiv preprint* arXiv:2305.19926.

Kasneci, E. et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and individual differences*, 103, 102274.

Khademi, A. (2023). Can ChatGPT and bard generate aligned assessment items? A reliability analysis against human performance. *arXiv preprint* arXiv:2304.05372.

King, M. R. (2023). Can Bard, Google's Experimental Chatbot Based on the LaMDA Large Language Model, Help to Analyze the Gender and Racial Diversity of Authors in Your Cited Scientific References? *Cellular and Molecular Bioengineering*, 16(2), 175-179.

Kumar, P. (2023). Large Language Models Humanize Technology. *arXiv preprint* arXiv:2305.05576.

LeCun, Y. et al. (2015). Deep learning. *Nature*, 521(7553), 436-444.

Lund, B. D., & Wang, T. (2023). Chatting about ChatGPT: how may AI and GPT impact academia and libraries? *Library Hi Tech News*, 40(3), 26-29.

OpenAI (2023). GPT-4 Technical Report. ArXiv, abs/2303.08774.

Panagopoulou, F. et al. (2023). Legal and ethical considerations regarding the use of ChatGPT in education. *arXiv preprint* arXiv:2306.10037.

Paul, J. et al. (2023). chatgpt and Consumers: Benefits, Pitfalls And Future Research Agenda. *Int J Consumer Studies*, 4(47), 1213-1225.

Pollock, D. et al. (2023). Recommendations for the extraction, analysis, and presentation of results in scoping reviews. *JBI Evidence Synthesis*, 21(3), 520–532.

Rahman, M. M. et al. (2023). ChatGPT and academic research: a review and recommendations based on practical examples. *Journal of Education, Management and Development Studies*, 3(1), 1-12.

Rahman, Md. M., & Watanobe, Y. (2023). ChatGPT for Education and Research: Opportunities, Threats, and Strategies. *Applied Sciences*, 13(9), 5783.

Rahsepar, A. A. et al. (2023). How AI Responds to Common Lung Cancer Questions: ChatGPT vs Google Bard. *Radiology*, 307(5), e230922.

Sallam, M. (2023). ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. *In Healthcare*, Vol. 11, No. 6, p. 887.

Sezgin, E. et al. (2023). Clinical Accuracy Of Large Language Models and Google Search Responses To Postpartum Depression Questions: A Cross-sectional Study. JMIR Preprints. 22/05/2023:49240.

Siegle, D. (2023). A Role for ChatGPT and AI in Gifted Education. *Gifted Child Today*, 46(3), 211-219.

Tlili, A. et al. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, 10(1), 15.