GENERATIVE AI AND CHATGPT USING GAN’S AND CHATBOT

NAME: DADI. SIVA NAGA GANESH

BRANCH: COMPUTER SCIENCE AND ENGNIEERING WITH ARTIFICIAL INTELLIGENCE AND MACHINE LEARING

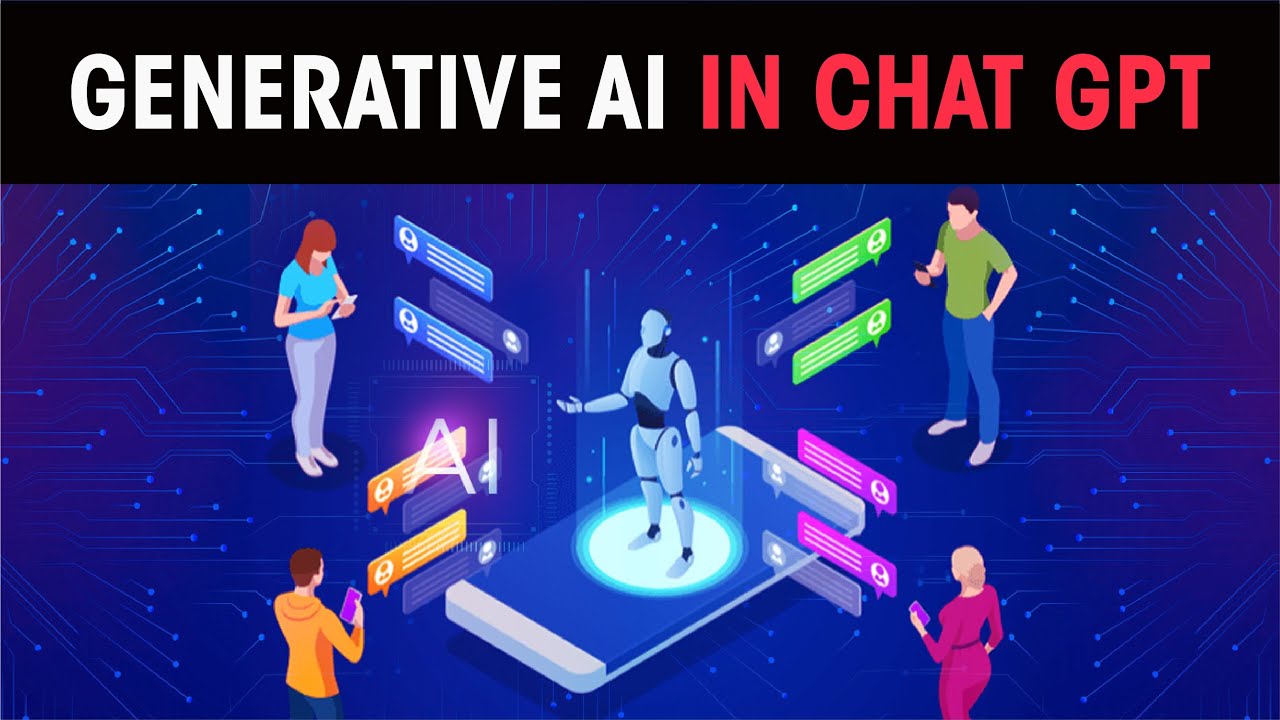
DOMIAN: GENERATIVE AI AND CHATGPT

PROJECT TITTLE: CHATBOT DEVELOPMENT BY USING GPT-3

COMPANY :BLACKBUCKS SOLUTIONS

DURATION:MAY 27TH -JULY 20TH





ChatGPT and Generative AI

USING GAN’S

# ABSTRACT

*Generative artificial intelligence (GenAI) is transforming the education landscape, with ChatGPT emerging as a pervasive force in student engagement and pedagogical methods. A systematic rapid review of 43 studies dissects ChatGPT’s nuanced role in education, highlighting its capacity to transform passive learning into an interactive and engaging process. The analysis ventures into the dark side of ChatGPT from the perspectives of educators and students, highlighting concerns about academic misconduct, the reliability of the information provided by ChatGPT, digital safety, bias, data privacy, academic integrity, threats to traditional and online assessments, and the potential erosion of critical thinking skills. To responsibly integrate AI into educational frameworks, the chapter delineates strategic recommendations that ensure technological advancements align with safeguarding educational standards and fostering a secure learning environment. This chapter offers a guide for educators and academic institutions to harness AI’s educational potential while mitigating its risks ethically.*

# INTRODUCTION

In the digital dawn of the 21st century, the realm of education continues to undergo a seismic shift driven by the persistent advancement of technology. Generative Artificial Intelligence (GenAI) is among the most groundbreaking technological vanguards, with systems like ChatGPT at The forefront, revolutionizing the educational landscape. These sophisticated AI platforms are collaborative partners in learning, transforming student engagement, and reshaping pedagogy (Schiff, 2023). Over 70% of educators believe that generative AI has the potential to transform education, according to a recent survey by Capgemini Research Institute (2023). This statistic highlights the positive perceptions of generative AI among educators and their belief in its potential to improve student learning.

Integrating AI into education offers opportunities and challenges, such as personalized learning and instantaneous feedback (Lam et al., 2018), as well as ethical concerns (Holmes et al., 2021), data privacy risks, and academic integrity threats (Cotton et al., 2023). Educators must prioritize digital safety to ensure both secure and stimulating learning environments. The extensive and diverse literature on using ChatGPT within educational environments documents its potential to positively and negatively influence teaching and learning practices (Lo, 2023; Grassini, 2023; Adeshola & Adepoju, 2023). For instance, ChatGPT can enhance educational experiences by providing instant, tailored feedback to learners, but it may also inadvertently promote reliance on automated responses at the expense of critical thinking skills. However, this body of research lacks common agreement on how ChatGPT contributes to or detracts from student engagement, a cornerstone for transformative education. The notion of transformative education can be described as “education that (i) is transformative (i.e., the “idea”) in its (ii) process (i.e., the “how”) and (iii) product (i.e., the “what”) in response to (iv) grand challenges (i.e., the “so what”)” (Lim et al., 2023, p. 2).

As we embark on this exploration, it is crucial to recognize the multi-faceted nature of AI’s influence on education. Through a Systematic Literature Review (SLR), specifically a Rapid Review, this chapter aims to dissect the nuanced layers of generative AI’s integration into learning environments. It seeks to uncover how generative AI, particularly ChatGPT, is shaping student engagement—transforming it from a passive to an active process, and how educators can leverage this technology to foster a more dynamic form of learning while simultaneously navigating the challenges of digital safety. By synthesizing the current research, we will illuminate the pathways through which AI can enrich education and the safeguards needed to anchor its promise in a foundation of security and trust. The high-level research questions guiding this research are: “How are ChatGPT and generative AI currently utilized in educational ecosystems, and how do they impact teaching and learning

processes? What ethical challenges and digital safety concerns are associated with using ChatGPT and generative AI in educational settings, and what strategies can be employed to address them?”

The chapter is organized as follows: We begin by providing background information on Generative AI and its application in education through ChatGPT. This is followed by a description of our methodology, detailing the steps of our rapid review process. We then present our findings, which include the educational applications of ChatGPT, the ethical considerations it raises, and issues of digital safety. The subsequent discussion delves into these findings in greater depth and explores the policy implications. The paper concludes with a summary of the principal outcomes, their educational implications, and a comprehensive list of references.

# BACKGROUND

**Generative AI: Definition and Capabilities**

Generative AI can be defined as “a technology that (i) leverages deep learning models to (ii) generate human-like content (e.g., images, words) in response to

(iii) complex and varied prompts (e.g., languages, instructions, questions)” (Lim et al., 2023, p. 2). Chat Generative Pre-trained Transformer (hereafter referred to as ChatGPT) is a cutting-edge natural language processing (NLP) model developed by OpenAI (Mhlanga, 2023) through the lens of human-centric functional modeling (Williams, 2022). This highly advanced chatbot has the ability to handle an array of text-based tasks, from responding to straightforward inquiries to performing complex assignments like crafting appreciation notes and facilitating nuanced conversations on challenging topics like productivity concerns (Liu et al., 2021). The foundation of ChatGPT is GPT-3, the third model developed for NLP projects (Entrepreneur, 2023; Thorp, 2023), which was upgraded in April 2023 to GPT-4. In the realm of GenAI, DALL-E is another notable OpenAI project; however, unlike ChatGPT, which generates text, DALL-E is designed to create digital images. Table 1 concisely summarizes the primary concepts related to GenAI.

*Table 1. Overview of key ChatGPT concepts*

|  |  |
| --- | --- |
| **Key Concepts** | **Definition** |
| Chatbot | A computer program designed to mimic human-user communication, particularly via the Internet (Adamopoulou & Moussiades, p. 373) |
| Generative Model | A Generative Model is a sophisticated method for learning any type of data distribution using unsupervised learning that has shown significant success in recent years (Pandey, 2018). |
| Generative Pre- Trained Transformer | Generative Pre-Trained Transformer (GPT) is a machine learning model that uses unsupervised and supervised learning techniques to understand and generate human- like language (Radford et al., 2018) |
| Natural Language Processing | NLP is a field of computer science and AI focused on the interaction between computers and human languages, involving the study and computational modeling of language’s myriad aspects to develop diverse systems (Reshamwala et al., 2013). |

Source: Author

# The Evolution of AI in Education: ChatGPT’s Role

The trajectory of AI in Education (AIED) has followed a path of progression that mirror technological advancements. In the mid-20th century, Skinner (1958), regarded as the progenitor of behaviorism, introduced his ‘teaching machine,’ which diverged from multiple-choice by prompting students to formulate responses actively. Skinner envisioned his device as akin to a personal tutor’s capacity to offer tailored learning experiences, presaging today’s Intelligent Tutoring Systems (ITS) capabilities. Parallel to Skinner’s work, Pask developed the Self-Adaptive Keyboard Instructor (SAKI) in the same era. This adaptive device advanced training for punch-card operators by adjusting tasks based on the learner’s prior performance, a precursor to adaptive learning technologies (Pask, 1982). However, it was Jaime Carbonell who, in his 1970 Ph.D. thesis, first applied mainstream AI techniques to computer-aided instruction with his system called SCHOLAR (Carbonell, 1970). SCHOLAR was groundbreaking, capable of generating tailored responses through a network of interlinked semantic concepts, establishing a foundational model for the conversational AI that would follow. The advent of AIED has not only facilitated the emergence of ‘smart’ educational technology but has also emerged as a potent instrument for delving into the ‘black box of learning’ (Luckin & Holmes, 2016). It allows for more detailed insights into the learning process, considering factors such as socio-economic backgrounds, environmental contexts, and technological influences (Blanchard, 2015).

The transition from traditional AI to the cutting-edge realm of GenAI holds the

promise of significantly diminishing complexity in numerous applications, including education. At the vanguard of this AIEd revolution is ChatGPT, which represents a

paradigm shift in how educational content is delivered, personalized, and assessed. Built on the robust GPT architecture, ChatGPT has been integrated into various educational frameworks, fundamentally altering the teaching and learning landscape. ChatGPT, powered by the sophisticated GPT architecture, has become a key component in educational settings, revolutionizing pedagogical methods and learning experiences. It functions as an interactive dialogue system that supports students by answering queries, elucidating concepts, and guiding research, effectively replicating the beneficial aspects of personal tutoring (Javaid et al., 2023). In language learning, ChatGPT’s proficiency in multiple languages provides a unique avenue for students to enhance their skills. Natural and engaging conversational exchanges offer a more dynamic mode of language practice, complementing traditional methodologies with

instant feedback and grammatical support.

ChatGPT has proven to be a versatile asset for educators, capable of crafting customized teaching materials, quizzes, and diverse content to suit various educational needs and complexities (Hashem et al., 2023; Javaid et al., 2023). It supports a personalized approach to instruction, catering to individual student’s learning trajectories. Educators have found ChatGPT a dynamic ally for creating customized teaching materials. It can generate examples, quizzes, and explanatory paragraphs across various subjects and complexities (Hashem et al., 2023; Javaid et al., 2023). This not only aids teachers in preparing course content but also provides students with additional resources tailored to their learning pace and style.

The capacity of ChatGPT to comprehend, produce, and translate text across various languages presents significant advantages for those learning new languages. Students can improve their language skills more interactively and engagingly through conversational practice and grammatical assistance than traditional study methods. The IDEE (Identify, Determine, Ensure, Evaluate) framework, developed by Su and Yang (2023), offers a structured approach to deploying these technologies in

the classroom consisting of:

* Identify the Desired Outcomes: Educators must begin by defining clear objectives for incorporating ChatGPT into their curricula
* Determine the Appropriate Level of Automation: Deciding on the level of automation involves balancing AI-driven instruction with human interaction.
* Ensure Ethical Considerations: Upholding ethical standards when employing ChatGPT in educational settings is critical. Key concerns involve safeguarding personal information, reducing bias in AI systems, and preventing the unintentional reinforcement of social disparities.
* Evaluate the Effectiveness: Finally, the integration of ChatGPT should be subject to continuous evaluation to determine its impact on learning outcomes.

# GenAI and Educational Theories

Exploring the intersection of GenAI and established educational theories, such as behaviorism, cognitivism, constructivism, and social learning theory, offers intriguing insights into how these advanced technologies can be harmoniously integrated to enhance and revolutionize learning experiences.

As an educational theory, behaviorism strongly emphasizes the observable behaviors of learners and how these behaviors are shaped and modified by various external stimuli (Strain et al., 1992). In this context, GenAI can be used as a tool for reinforcement learning. For example, AI-driven learning platforms can provide immediate feedback to students, reinforcing correct responses and guiding them away from incorrect ones, much like a traditional conditioning process (Vashishth et al., 2024). Cognitivism emphasizes the role of mental processes in learning, such as memory, attention, and problem-solving (Clark, 2018). GenAI can aid in these cognitive processes by organizing and presenting information in ways that are easier for the brain to process and remember. AI can also assist in problem- solving by offering hints or breaking down complex problems into manageable steps. Constructivism posits that learners construct knowledge through experiences and interactions with their environment (Miller-First & Ballard, 2017). GenAI can enhance this by creating personalized learning experiences. For instance, GenAI can simulate real-world scenarios where students apply concepts, thus constructing their understanding through interactive and engaging environments (Salinas-Navarro et al., 2024). Social learning theory posits that people learn from one another, through observation, imitation, and modeling (Laland, 2004). GenAI can facilitate social learning by connecting learners from different backgrounds in virtual environments.

# Methodology

Snyder (2019) posits that “if the purpose of the review is to investigate and synthesize evidence of the effect of a specific factor, a systematic review approach should be used” (p. 336). In this study, the “specific factor” pertains to the implications raised by ChatGPT. Traditional SLRs often encompass a broader timeframe, typically examining articles published over a decade (Korovessis & Repantis, 2019; Kucharski et al., 2009; Urschl et al., 2005). Given ChatGPT’s recent emergence, a complete SLR that spans such an extensive period is not feasible. Therefore, this study will adopt the rapid review methodology, recognized for its efficiency in synthesizing research within a condensed time frame. This approach has demonstrated effectiveness across various fields (e.g., Meis-Harris et al., 2021; Ponce et al., 2019) and has been particularly beneficial in education-related research (Lo, 2023), making it well-suited to the aims of this study.

A rapid review is a streamlined form of knowledge synthesis that accelerates the systematic review process by simplifying or omitting specific steps to produce timely evidence (Khangura et al., 2012). Although rapid reviews aim to maintain the same rigorous standards as systematic reviews, they prioritize expediency and are particularly well-suited to emerging fields where immediacy is paramount. The dual objectives of this review are to delineate the challenges associated with the nascent technology of ChatGPT and to explore the ethical considerations, impacts, and potential mitigation strategies within the realm of GenAI implementation.

While rapid reviews offer advantages such as timeliness, accuracy, comprehensiveness, cost-effectiveness, and feasibility, they also carry an inherent risk of bias (Tricco et al., 2015). Our review will be anchored in a structured, transparent, and reproducible methodology to mitigate this risk. We will adhere to Tranfield et al.’s (2003) established three-stage procedure of planning, execution, and reporting, which has been successfully applied in various disciplines and contexts (e.g., Calabrò et al., 2019; Crossan & Apaydin, 2010), adapting it to the unique requirements of a rapid review.

# Planning the Review

The formulation of our research problems is anchored in the transformative role that ChatGPT and GenAI play within educational ecosystems. As these technologies expand to become more in academic settings, it is critical to understand their significance and impact on pedagogical practices and the breadth of ethical and safety considerations they raise. The research questions have been designed to highlight the dual necessity of harnessing AI technologies’ educational capabilities, such as ChatGPT, and proactively managing the complex issues surrounding their use. They stem from a recognized need to synthesize the current state of knowledge and aid educators, policymakers, and technologists in making informed decisions that maximize benefits and mitigate risks.

The inclusion and exclusion criteria for this rapid review, outlined in Table 1, were systematically developed to ensure a focused and relevant body of literature is examined.

*Table 2. Inclusion and exclusion criteria for the review process*

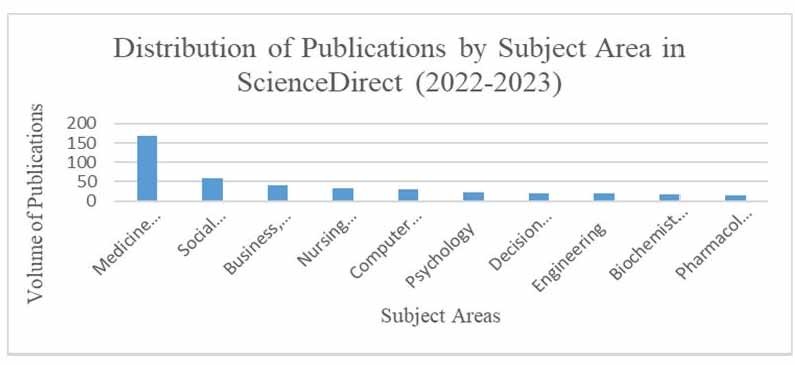
|  |  |  |
| --- | --- | --- |
| **Criterion** | **Inclusion Criteria** | **Exclusion Criteria** |
| Article Topic | Articles that specifically discuss the use and impact of ChatGPT, or generative AI, within educational settings. | Articles that do not specifically address ChatGPT or its application in the educational domain that serve the review objectives. |
| Article Type | Peer-reviewed academic journal articles and conference proceedings. | Non-academic articles, editorials, opinion pieces, and social media blogs or articles were excluded from the study to maintain research rigor and academic focus. |
| Timeframe | Articles published from 1 January 2022, capturing the most recent developments following ChatGPT’s release. | Articles published outside this date range, given the focus on recent developments in the field. |
| Language | Articles published in English to ensure the review can be conducted within the resource constraints and timeline. | Articles in languages other than English, due to translation resources and time limitations. |

Source: Author

# Execution of the Review

In this process, three channels can be considered: electronic database search, backward searching, and forward searching (Xiao & Watson, 2019). To investigate the research question, we used “electronic database search” as the only tool in our case. We reviewed articles published in peer-reviewed journals for the period between January 2022 and October 2023 from the following three databases: ScienceDirect, Google Scholar, and Business Source Complete. ScienceDirect, in particular, provided academic articles relevant to our research question across various subject areas, as shown in Figure 1. For instance, the “Social Sciences” category within ScienceDirect includes publications pertinent to education, thereby ensuring that our search captured interdisciplinary studies that address educational aspects of GenAI technologies like ChatGPT.

*Figure 1. ScienceDirect: Overview of publications in generative AI and education in different areas*



As for keywords, they can be derived from the research question and applied with both Boolean operators “AND” and “OR.” The search string was (“ChatGPT” OR “Generative AI” OR “Generative Artificial Intelligence”) AND (“education” OR “educational ecosystems”) AND (“safety” OR “digital safety”). This rapid review adhered to the guidelines set forth by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) for selecting pertinent articles, as Stevens et al. (2018) recommended. Figure 2 illustrates the review research flowchart design, similar to Xiao & Watson (2019), showing four main steps: identification, screening, eligibility, and inclusion.

# Reporting the Review

Following the PRISMA framework, this rapid review transparently navigates article selection, applying necessary adjustments to expedite the process. The modifications, particularly the adoption of an electronic database search as the sole channel for literature identification, were tailored to address the objective of this study.

Quality was assured using a checklist that rigorously evaluated articles against our inclusion criteria, refining the full-text selection to a substantive collection of 43 relevant studies. This streamlined yet rigorous process enabled a targeted synthesis of current material, providing implications for educational practice and finding fruitful grounds for future research.

While the accelerated nature of this review facilitated a prompt assessment of emerging literature, it is acknowledged that the scope was narrower than that of a traditional SLR. This limitation is inherent to rapid reviews prioritizing timeliness

over exhaustive coverage. Nonetheless, the selected articles provide a pivotal snapshot of the current state of ChatGPT’s application in educational settings, offering valuable insights while paving the way for more comprehensive, in-depth analyses in the future.

# Findings of the Rapid Review

The rapid review revealed several recurrent themes in the abstracts regarding the application of GenAI like ChatGPT in educational settings. One prominent theme is the use of ChatGPT for personalized learning, with numerous studies highlighting its capability to tailor educational content to individual student needs (Author A et al., 2022; Author B et al., 2023). Another theme centers around the challenges of integrating GenAI into existing curriculums, with concerns about digital safety and the potential for algorithmic bias emerging as critical issues (Author C et al., 2022). These themes underscore the transformative potential of AI in education, along with the critical need for robust ethical frameworks to safeguard digital interactions (Author D et al., 2023).

# Fields of ChatGPT in Education

In the realm of mathematics, Shihab et al. (2023) suggest that ChatGPT serves as a competent guide, assisting students with complex problems and encouraging critical thinking. However, Frieder et al. (2023) put ChatGPT’s capabilities into perspective using the GHOSTS dataset – a comprehensive set of advanced mathematical problems and solutions designed specifically to benchmark AI capabilities against the expertise of graduate-level students – and revealing that ChatGPT’s mathematical proficiency, particularly in advanced topics, tends to fall short of graduate student levels. Yet, while ChatGPT showed some promise for the same authors, particularly in elementary set theory and logic, its overall performance was below that of typical math graduate students, often understanding but not correctly solving the problems. Ardos et al. (2023) conducted a study using the Open Adaptive Tutoring system (OATutor) to evaluate the effectiveness of ChatGPT-generated prompts in teaching algebra, involving 77 Mechanical Turk participants. They observed that while 70% of ChatGPT’s prompts met quality standards and facilitated learning, human-crafted prompts resulted in more significant learning gains, with scores notably exceeding those achieved through ChatGPT’s assistance.

Physics education research reflects a diverse assessment of ChatGPT’s capabilities. Cameron (2023) touched on its ability to tackle complex concepts such as quantum gravity, suggesting a depth of understanding in specialized topics. Other studies examined ChatGPT and physics from an examination’s perspective. Kortemeyer

(2023) critically evaluates the performance of ChatGPT in an introductory calculus- based physics course, highlighting the model’s human-like response generation that often blurs the line between AI and human output. Despite its ability to pass the course, ChatGPT displayed common misconceptions and errors akin to those of a novice learner. West (2023) further evaluated ChatGPT using the Force Concept Inventory and found its performance on physics concept problems was on par with, or slightly better than, the average college student after one semester, scoring between 50% and 65%. Çoban and Çoban (2023) highlighted ChatGPT’s versatility in physics education for tasks like lecturing and designing experiments. However, they also noted its challenges with mathematical precision and the potential to propagate common student misconceptions, emphasizing careful usage for optimal educational outcomes. Within the same context, Liang et al. (2023) explored the pedagogical advantages of using ChatGPT in physics education, detailing the effectiveness of prompting techniques in guiding the AI to solve calculation problems, clarify solutions, and create new exercises with proficiency comparable to human performance.

The application of ChatGPT in chemistry education has been explored by Nguyen et al. (2023b) in the context of Vietnamese high schools, assessing its strengths and weaknesses for enhancing the educational process. The paper emphasized ChatGPT’s capacity for personalized learning, fostering critical thinking, and aiding in the conceptual understanding of chemistry, indicating its promising role in educational advancement. Humphry and Fuller (2023) discussed similar findings. Yet, the authors noted that ChatGPT is prone to conceptual errors in many of its chemistry answers and explanations and should not be relied upon for chemistry instruction. A significant body of work has explored ChatGPT’s impact on language learning, particularly in the context of foreign languages (e.g., Kohnke et al., 2023; Baskara et al., 2023) and English (Kostka & Toncelli, 2023; Koraishi, 2023). The consensus among these studies is that ChatGPT offers substantial opportunities for both teaching and research, especially for personalized learning experiences. Hong (2023) mainly focused on the concerns of language educators grappling with the implications of ChatGPT for writing assessments, revealing both the promise and the complexities

of its integration into language pedagogy.

In medical education, scholars like Abd-Alrazaq et al. (2023), Hisan and Amri (2023), Khan et al. (2023), and Lee (2023) have highlighted the dual-edged nature of implementing LLMs like the GPT series. On the one hand, these AI models are revolutionizing medical curriculum development, teaching methodologies, creating a more interactive, and engaging educational environment. On the other hand, they present challenges such as potential biases and the risk of fostering overreliance, which could undermine essential clinical reasoning skills and introduce misinformation. The scholars emphasize a prudent approach to leveraging LLMs, ensuring synergy between their innovative potential and the critical needs of medical education. One

study has examined the intersection of AI with clinical and educational dimensions of dentistry, suggesting the need for curriculum updates in dental education to match the pace of AI advancements, particularly with tools like ChatGPT reshaping diagnostics and patient communication (Thurzo et al., 2023).

Across disciplines, the studies reviewed suggest that ChatGPT can be a valuable educational tool, albeit with limitations. ChatGPT serves multiple functions in science learning, aiding educators and learners across the five domains of science education and underscores the need for technology that aligns with the sophistication required for effective science pedagogy (Purnama & Wijanarko, 2023). These articles raise essential considerations for the future of teaching and testing in physics education, suggesting a need for more in-depth research and curriculum development to address the implications of AI in academic settings.

# Ethical and Practical Challenges of Generative AI in Education

## The Dark Side of ChatGPT

ChatGPT experienced meteoric growth, amassing one million users within five days of its launch and reaching a remarkable 100 million users within two months, achieving a new record for the fastest-growing consumer application (Hu, 2023). Despite the meteoric rise in popularity, ChatGPT raised concerns among academic institutions (Benuyenah, 2023).

Recognizing these concerns from an educator perspective, Dalalah and Dalalah (2023) pinpointed challenges in discerning AI-generated text from human writing, citing risks of misclassification and the overlapping characteristics of genuine and AI-produced content. Within this context, educators, particularly those in research- driven institutions, face the daunting task of distinguishing between genuine knowledge and AI-generated content (Lim et al., 2023; Simon, 2023; Cooper, 2023; Elkhatat et al., 2023). This challenge underscores the crucial need to uphold academic integrity and foster original thinking among students. The diffusion of GenAI threatens the fairness and validity of online exams and assignments (Rahman & Watanobe, 2023) and might render traditional assessments like essays obsolete, as these tools can produce detailed, human-like responses (Zhai, 2022; Rudolph et al., 2023) and elaborating a debate about the effectiveness of multiple-choice questions (Gonsalves, 2023).

From the student’s perspective, the advent of ChatGPT and similar systems in the dynamic realm of digital education presents a series of potential challenges. First, excessive dependence on this advanced AI tool for academic support may inadvertently hinder the cultivation of critical thinking and research abilities, which are fundamental pillars for educational advancement (Ningrum, 2023; Ciampa et

al., 2023; Limna et al., 2023). According to Hong (2023), students’ overreliance on ChatGPT in assignments inhibits the development of their language skills and critical thinking. The threat to academic integrity further exacerbates this reliance, as the ease with which ChatGPT generates sophisticated content could entice students to engage in plagiarism, jeopardizing their learning trajectory. Second, the issue of accuracy looms large, with studies by Limna et al. (2023), Mhlanga (2023), and Fiialka et al. (2023) exposing ChatGPT’s occasional propagation of misinformation, which could lead students to internalize errors. Kung et al. (2023) quantified ChatGPT’s accuracy at approximately 60%, indicating a significant margin for error. Furthermore, the requirement for data input to generate personalized responses raises considerable privacy concerns, as the potential for personal information disclosure exists. Lastly, the availability of instant AI assistance could potentially undermine student motivation to engage deeply with their coursework, nurturing a preference for superficial solutions over a thorough exploration of academic materials (Wang, 2023). Michel-Villarreal et al. (2023) encapsulated these concerns into four critical challenges: quality control, expertise and authority, personalized learning, and communication and collaboration, each representing a crucial dimension in the evolving educational narrative shaped by AI technologies.

## Trust in the Age of Generative AI

In the emerging era of GenAI, trust in technology is pivotal and has been defined by Mayer et al. (1995) as “the extent to which a user is willing to depend on a technology and its outcomes” (p. 718). In the context of ChatGPT, *trust* refers to “users’ confidence in the accuracy and reliability of ChatGPT’s responses” (Salah et al., 2023, p. 4). Choudhury and Shamszare (2023) identified a direct correlation between trust and user engagement with ChatGPT, highlighting that an increase in trust correlates with a heightened intention to use the technology. They further elaborate on how trust not only directly propels the actual use of ChatGPT but also has an indirect impact by shaping users’ behavioral intentions, serving as a partial mediator between trust and usage.

Amid the utility of ChatGPT and other large language models (LLMs) in higher education—where they assist with tasks such as idea generation and learning enhancement—Rasul et al. (2023) cautioned against potential pitfalls, including academic dishonesty and the dissemination of misinformation. This underscores the need for a judicious and ethical approach to applying these AI tools in educational contexts. The academic community shares such reservations in Krabi, Thailand, where interviews conducted by Lim et al. (2023) reveal educators’ and students’ concerns over trust issues with ChatGPT. The fear of inaccurate information, data privacy breaches, and the diminishment of crucial personal interactions in the

learning environment that foster rapport and in-depth understanding are prominent among these concerns.

Echoing the need for caution, Firaina and Sulisworo (2023) recognized ChatGPT’s assistance in various academic activities like information gathering, idea production, text translation, and crafting questions. They emphasized, however, the critical need for users to verify the AI’s output against other sources and maintain an active, critical engagement with the information provided. Tlili et al. (2023) delved deeper into these issues through a three-stage instrumental case study examining the early adoption of ChatGPT in the educational sphere via social media analytics and interviews. Their study accentuated the requirement for structured guidelines that ensure AI’s safe and productive utilization, suggesting future research directions for integrating chatbots in education.

To conclude, as posited by Dalalah and Dalalah (2023), the thoughtful yet vigilant deployment of ChatGPT and akin language models holds transformative potential for academic research and pedagogy. When applied with consideration and caution, these tools can catalyze innovation, provide context, and reinforce analytical rigor, marking a significant shift in how academic endeavors are approached.

# Digital Safety Considerations

## AI and Ethical Tendencies

In exploring AI’s ethical framework, Krügel et al. (2023) investigated ChatGPT’s moral compass by providing numerous iterations of the trolley problem. They found that ChatGPT’s responses lacked a consistent moral orientation, which may have an unintended impact on users’ ethical reasoning. Borji (2023) also found differences in ChatGPT’s logical thinking, mathematical calculations, and coding jobs and recognized prejudices, indicating the possibility of factual and ethical errors. These tendencies highlight the importance of users being skeptical and cautious when accessing ChatGPT for moral or factual assistance.

Expanding on these concerns, Zhuo et al. (2023) conducted a comprehensive evaluation of ChatGPT, assessing its moral hazard, bias, reliability, robustness, and susceptibility to producing toxic content. While ChatGPT demonstrated a marginal improvement over other state-of-the-art language models, it exhibited notable deficiencies across these ethical dimensions. The study emphasized the pressing ethical challenges accompanying the development of sophisticated language models, advocating for a strategic focus on creating AI that upholds ethical standards.

As we integrate such tools into our digital lives, both students and educators must stay vigilant about these ethical and safety implications. Software developers, in particular, face the critical task of enhancing the ethical design of future language

models (Rane et al., 2023; Sohail et al., 2023). This challenge requires a concerted effort to balance technological innovation with moral responsibility. For users, especially in educational and professional settings, awareness and a discerning approach to AI interactions are crucial for maintaining integrity and safety in the digital age.

## Principles for Ethical AI in Education

The rapid integration of AI into education underscores the urgent need for comprehensive ethical guidelines that resonate with all stakeholders—educators, students, parents, AI developers, and policymakers (Berendt et al., 2020). A UNICEF policy reviewed 20 national AI strategies and revealed a concerning oversight: children’s unique needs are often overlooked, with few strategies addressing them specifically (Penagos et al., 2020). Masters (2023) underlined this in his examination of AI in healthcare education, identifying a set of ethical principles such as data gathering, privacy, anonymity, consent, data ownership, security, bias, transparency, autonomy, responsibility, and beneficence that parallel concerns in the educational sphere. Similarly, Adams et al. (2023) explored ethical frameworks guiding AI policies for children in K–12 education, uncovering principles of: transparency, justice and fairness, non-maleficence, responsibility, privacy, beneficence, freedom and autonomy, pedagogical appropriateness, children’s rights, AI literacy, and teacher well-being. Yet, Schiff (2021) pointed out a misalignment in national AI policies, which emphasize preparing individuals for AI rather than leveraging AI to enhance education. This disparity suggests a need to reorient policy discussions to prioritize AI’s role in supporting and improving educational outcomes.

The ethical deployment of AI in education is not without controversy, especially regarding data privacy. Jones et al. (2014) differentiated between data created by students, such as assignments, and data generated by institutions, like tracking information and grades, each with its own set of ethical considerations. Balancing the protection of student data with the educational benefits of AI personalization remains a contentious issue, highlighting the delicate interplay between advancing educational technologies and safeguarding student rights.

The Institute for Ethical AI in Education (2021) has outlined an Ethical Framework for AI in Education, advocating for AI to enhance educational institutions’ capacities while maintaining the integrity of human relationships. This framework promotes the equitable use of AI, ensuring support across all learner demographics without bias, and empowers learners by granting them greater agency in their educational endeavors. It emphasizes the need for a balanced approach to individual privacy and data utilization for defined educational benefits. Central to the framework is the importance of transparency and accountability, urging that all stakeholders,

including learners and educators, achieve a foundational understanding of AI’s capabilities and limitations. The framework also calls for ethical AI design, insisting that educational tools comply with moral standards and contribute constructively to the academic environment.

However, while AI has the potential to uphold educational values such as child- centered learning and personalized education, Carr (2020) cautioned about the unintended consequences on mental health and cognitive autonomy. Complementing these insights, Nguyen et al. (2023a) identified additional ethical principles necessary for AIED, such as robust governance, inclusiveness, transparency, accountability, sustainability, proportionality, and a focus on human-centered approaches. These principles collectively form a cornerstone for the ethical deployment of AI, ensuring that its integration into education prioritizes enhancing human potential and safeguards against any adverse effects.

## Specific Ethical Concern: Mental health

The intersection of AI interaction and mental health is becoming increasingly scrutinized as research uncovers the nuanced psychological effects of technology on individuals. Studies by Xie & Pentina (2022) and Pentina et al. (2023) have highlighted that emotional bonds can form between users and chatbots, influencing emotions, attitudes, and behaviors—a phenomenon corroborated by Elyoseph et al. (2023), who observed ChatGPT’s significant emotional awareness. This finding opens up potential applications for ChatGPT in areas such as cognitive training, which could aid individuals with emotional awareness deficits.

Designing responsible AI for educational purposes, as Kuhail (2023) and Tlili et al. (2023) argue, requires a user-centered approach that considers both academic and social and emotional factors. This perspective is underscored by Rad and Rad (2023), whose research with ten psychology students identified a range of emotional responses to chatbots, encompassing both positive reactions to their utility and negative responses when expectations are not met, as well as concerns about data privacy and the erosion of human interaction. The implications of these emotional attachments extend to learner motivation, as Ali et al. (2023) noted, with chatbots like ChatGPT providing structured responses that can engage but potentially limit psychological growth. Uludag (2023) critiques this limitation, observing that ChatGPT’s repetitiveness in offering solutions might not fully address the complex needs of individuals facing psychological issues.

The concept of Action Identification Theory (Fennell, 2023) introduces another layer to this discussion, referring to the traits and mental processes behind how people think about and describe actions in more concrete s—a concept that ChatGPT’s responses might mirror, as Kosinski (2023) examination of the Theory of Mind

suggests emergent but incomplete signs of such cognitive processing. In a related study by Szabo (2023), ChatGPT’s understanding of exercise addiction was tested, revealing a blend of accuracy and error, which indicates a gap in the AI’s capability to handle complex psychological topics.

Finally, Aditama et al. (2023) recognize the potential of AI as an immediate support tool but reiterate that AI cannot substitute professional mental health care. They advocate for the need to develop new engagement strategies that leverage technology while maintaining the central role of human practitioners, especially for the youth who may turn to these accessible platforms for support.

# DISCUSSION

Our rapid review has unveiled the expanding role of ChatGPT across educational disciplines, demonstrating its involvement in teaching subjects such as chemistry, physics, mathematics, and foreign languages (Frieder et al., 2023; Cameron, 2023; Nguyen et al., 2023b; Liang et al., 2023; Kostka & Toncelli, 2023). The findings indicate that while ChatGPT holds the potential for augmenting educational practices, its integration is not without challenges.

From an educator’s standpoint, issues such as distinguishing AI-generated content from student work, maintaining academic integrity, and ensuring effective use of ChatGPT without compromising learning quality have been underscored (Lim et al., 2023; Simon, 2023; Cooper, 2023; Elkhatat et al., 2023). From students’ perspective, ChatGPT raises concerns over developing critical thinking skills, the risk of academic misconduct, and the reliability of the information provided by ChatGPT (Mhlanga, 2023; Fiialka et al., 2023). The study also delves into the crucial role of ‘trust’ in GenAI, affecting both the uptake and effective use of ChatGPT in educational contexts (Sulisworo, 2023; Tlili et al., 2023; Lim et al., 2023). Further, it underscores the need for robust digital safety protocols and ethical frameworks to navigate AI’s potential risks to mental health and data privacy. As a case study illustrating the use of ChatBot technology in universities, Georgia State University provides an exemplary model. Faced with a significant increase in their summer melt rates, which escalated from 12% to almost 19%, the university sought a technological solution. They deployed a ‘campus coach’ chatbot, a tool that not only encompassed the institution’s collective knowledge but also managed to capture the spirit of the school’s community. This strategic move highlights the potential of AI-driven chatbots in enhancing student engagement and retention in higher education settings1. Also, Quibeldey-Cirkel (2023) elaborated a web plateform named ChatGPT Codex allows students to ask questions anonymously and receive answers from ChatGPT in real-time, providing them with the support they need to succeed in their studies.

Building upon the findings of the rapid review conducted by Lo (2023), which highlighted issues of accuracy and reliability in the application of ChatGPT within educational settings, our review delves deeper into the landscape of emerging challenges these advanced AI tools address. Significant among these is the misclassification of AI-generated content as human writing, a concern that strikes at the core of academic evaluation and the integrity of students’ scholarly output. This misclassification bears profound implications, undermining the trustworthiness of the academic work and threatening the fairness and validity of both online and traditional assessment forms. Further, the reliance on AI for academic support opens up avenues for academic misconduct and plagiarism, as students may submit AI-crafted responses as their own, bypassing the intellectual rigor expected in higher education. Moreover, the capabilities of ChatGPT, though impressive, are not without fault, the main reason for raising substantial privacy concerns and the specter of misuse, in addition to sharing personal data. Lastly, the availability of instantaneous AI-generated assistance risks cultivating a preference among students for superficial engagements with their learning material. Thus, it undermines the student’s motivation for deep learning due to easy access to AI assistance.

One area that has prompted action is authorship in academic publishing. Prominent journal editors and professional societies are clarifying authorship guidelines in response to the capabilities of large language models like ChatGPT. As per Flanagin et al. (2023) and Zielinski et al. (2023), such AI tools, despite their advanced abilities to generate text, do not meet the criteria for authorship because they cannot take responsibility for the content they generate. This stance by academic gatekeepers not only defines the scope of AI’s role in research dissemination but also upholds the integrity of scholarly work, emphasizing the human element that is accountable for interpreting and applying knowledge.

There’s a discernible gap in research concerning students’ digital safety awareness when using ChatGPT in educational settings. Despite the growing presence of AI, education on its safe use isn’t keeping pace. Addressing this lack of awareness is critical, as it involves protecting personal data and equips students to evaluate AI- generated content critically. Future research must enhance students’ awareness to ensure their digital well-being as they navigate AI in learning.

After addressing the crucial issue of students’ digital safety awareness when using ChatGPT, it is equally important to delve into the technical requirements and practical considerations for effectively integrating GenAI in various educational contexts. This includes understanding the infrastructural needs, software compatibility, and training requirements for educators and students in different settings, such as K-12 schools, higher education institutions, and online learning platform. In K-12 schools, the integration of ChatGPT may require a focus on creating age-appropriate interfaces and ensuring compatibility with existing educational software. . For higher education

institutions, the technical considerations extend to robustness and scalability, given the complexity of subjects and the larger student base. In the realm of online learning platforms, the discussion revolves around ensuring seamless integration of ChatGPT for a diverse, global audience, focusing on accessibility, user experience, and the ability to handle high volumes of concurrent interactions.

# Policy Implication

The burgeoning significance of AI in enhancing educational practices and the importance of nurturing AI literacy within educational curricula are rapidly emerging as focal points in policy discussions (Miao & Holmes, 2021). Continuing this discourse, the present study enriches the policy debate by delineating three critical policy implications that are integral for the strategic integration of GenAI in educational environments.

The adoption of AI in educational environments necessitates stringent risk management policies. Institutions should consider integrating established frameworks (e.g., NIST Framework for Improving Critical Infrastructure Cybersecurity, ISO 31000 Risk Management, FAIR (Factor Analysis of Information Risk), and the Montreal Declaration for ethical AI into their risk management strategies to address the unique challenges AI systems pose. Furthermore, developing an AI Risk Impact Framework tailored to educational settings could guide stakeholders (academic institutions, educators, and students) in evaluating AI’s ethical implications, ensuring data privacy, and aligning AI use with educational objectives. Policymakers and educational leaders must collaborate to implement these frameworks, thereby institutionalizing best practices for AI risk management in education.

A critical policy consideration underscored by Lim et al. (2023) is the transition of ChatGPT to a paid subscription model, which poses a significant risk to the democratization of knowledge. Such a move could create or widen existing socio- economic divides, as it may constrain access to this educational tool for lower- income students and communities. This development directly influences Sustainable Development Goal 4 (SDG4), which is centered on ensuring inclusive and equitable quality education for all.

To fully leverage the potential of GenAI technologies such as ChatGPT within the educational sphere, professional development programs provide instructors with the required skills to integrate these technologies into the classroom efficiently. Such training should emphasize the practical application of AI and a thorough awareness of its limitations, ethical implications, and critical evaluation of AI-generated content. This may guarantee that instructors are equipped to guide their students’ AI usage and improve the learning experience without relying too heavily on technology.

# CONCLUSION

“If Big Data is the new oil, then AI is the new electricity, as it enables many developments” (Market Insights, 2023, p. 3). This review has comprehensively explored the integration of ChatGPT and GenAI within educational ecosystems, marking a significant juncture in how students engage with learning materials and how educators facilitate teaching processes. In response to the research questions, our findings have established that ChatGPT serves various educational purposes, from personalized tutoring to promoting language acquisition, showcasing its versatility across multiple disciplines, including mathematics, physics, chemistry, and foreign languages. However, with these advancements come nuanced ethical challenges and concerns for digital safety. Our review has brought to light the ‘dark side’ of ChatGPT, revealing the potential for these technologies to perpetuate biases, undermine academic integrity, and even affect mental health. Trust has been identified as a central factor influencing GenAI adoption and proper utilization in educational settings. The principles for ethical GenAI in education—centered around transparency, accountability, inclusivity, pedagogical appropriateness, student’s rights, AI literacy, and teacher well-being—provide a framework for navigating these issues, yet their implementation remains a critical challenge.

As our study concludes, three pivotal policy implications have crystallized. First, the necessity for educational institutions to adopt a robust risk management framework is clear, mitigating potential negative impacts of AI technologies like ChatGPT. Secondly, there is a need to address the ways in which such technologies could inadvertently hinder the achievement of SDG 4, which aims for inclusive and equitable quality education. Third, developing sophisticated educational tools is required to provide educators with the skills needed to effectively and seamlessly integrate these technologies into the classroom.

Future research endeavors must close the divide between ‘AI for Education’—the use of AI to support and enhance learning outcomes—and ‘Education for AI’— preparing students to thrive in a future shaped by AI. This necessitates a concerted effort to formulate educational policies that foresee a future interwoven with AI and actively employ AI to foster educational advancement. Such an initiative demands collaborative synergy among AI developers, academic researchers, and policymakers to create AI tools that embody ethical integrity, are grounded in effective pedagogy, and resonate with the foundational values of education.

Generative AI, including ChatGPT, is a fascinating field at the intersection of artificial intelligence and natural language processing. Here are some key points about generative AI and specifically ChatGPT:

1. \*\*Definition of Generative AI\*\*: Generative AI refers to models and algorithms that can generate new content such as text, images, or even music, often mimicking human creativity and intelligence.

2. \*\*Neural Networks\*\*: Many generative AI models, including ChatGPT, are based on deep neural networks. These networks are trained on vast amounts of data to understand and generate human-like text.

3. \*\*Language Understanding\*\*: ChatGPT, for example, understands and generates text based on the context provided by users. It uses a variant of the Transformer architecture to process and respond to text inputs.

4. \*\*Training Data\*\*: Generative models like ChatGPT require large-scale datasets to learn patterns of human language. OpenAI used diverse internet text to train earlier versions of ChatGPT.

5. \*\*Applications\*\*: Generative AI has numerous applications beyond chatbots, including content generation, creative writing support, virtual assistants, and more. It's also used in fields like healthcare for generating medical reports or in gaming for creating interactive narratives.

6. \*\*Ethical Considerations\*\*: The capabilities of generative AI raise ethical concerns, such as misuse for generating fake news, impersonation, or biased outputs based on the training data.

7. \*\*Improvements Over Time\*\*: Generative AI models continue to improve with advancements in model architectures, training techniques, and access to larger datasets. Each new version aims to produce more coherent and contextually relevant responses.

8. \*\*Human-AI Interaction\*\*: ChatGPT and similar models are designed to interact seamlessly with humans, understanding natural language queries and providing appropriate responses, which enhances user experience in applications like customer service or education.

9. \*\*Limitations\*\*: Despite their capabilities, generative AI models like ChatGPT can still produce inaccurate or nonsensical responses, especially when faced with ambiguous or novel inputs.

10. \*\*Future Directions\*\*: The future of generative AI involves improving model understanding, enhancing contextual awareness, and addressing ethical concerns while expanding applications into new domains.

Understanding these aspects provides insight into how generative AI, exemplified by ChatGPT, is evolving and shaping various aspects of technology and human interaction.