Beyond Traditional Teaching: The Potential of Large Language Models and Chatbots in Graduate Engineering Education

Mahyar Abedi, Ibrahem Alshybani, Muhammad Rubayat Bin Shahadat, Michael S. Murillo

Preprint v2

Dec 18, 2023

https://doi.org/10.32388/MD04B0.2

Beyond Traditional Teaching: The Potential of Large Language Models and Chatbots in Graduate Engineering Education

Mahyar Abedi[1‡], Ibrahem Alshybani[1‡*], Muhammad Rubayat Bin Shahadat[1‡], Michael S. Murillo[2],

1 Department of Mechanical Engineering, Michigan State University, East Lansing, Michigan, USA
2 Department of Computational Mathematics, Science and Engineering, Michigan State University, East Lansing, Michigan, USA

‡These authors contributed equally to this work.

* alshyban@msu.edu

Abstract

In the rapidly evolving landscape of education, digital technologies have repeatedly disrupted traditional pedagogical methods. This paper explores the latest of these disruptions: the potential integration of large language models (LLMs) and chatbots into graduate engineering education. We begin by tracing historical and technological disruptions to provide context and then introduce key terms such as machine learning and deep learning and the underlying mechanisms of recent advancements, namely attention/transformer models and graphics processing units. The heart of our investigation lies in the application of an LLM-based chatbot in a graduate fluid mechanics course. We developed a question bank from the course material and assessed the chatbot's ability to provide accurate, insightful responses. The results are encouraging, demonstrating not only the bot's ability to effectively answer complex questions but also the potential advantages of chatbot usage in the classroom, such as the promotion of self-paced learning, the provision of instantaneous feedback, and the reduction of instructors' workload. The study also examines the transformative effect of intelligent prompting on enhancing the chatbot's performance. Furthermore, we demonstrate how powerful plugins like Wolfram Alpha for mathematical problem-solving and code interpretation can significantly extend the chatbot's capabilities, transforming it into a comprehensive educational tool. While acknowledging the challenges and ethical implications surrounding the use of such AI models in education, we advocate for a balanced approach. The use of LLMs and chatbots in graduate education can be greatly beneficial but requires ongoing evaluation and adaptation to ensure ethical and efficient use. This paper invites further research and dialogue in this emerging field, with the goal of responsibly harnessing these technologies to advance higher education.

1 Introduction

Throughout history, education has been continually shaped by technological disruptions. The advent of the printing press in the 15th century democratized access to knowledge, transforming how information was disseminated [1]. The 19th-century introduction of the blackboard revolutionized classroom dynamics, fostering interactive learning [2]. In

> ### Glossary
>
> **artificial intelligence (AI)**: ability of machines to perform tasks that are usually associated with human intelligence **machine learning (ML)**: subset of AI that allows computers to learn from data without being explicitly programmed
>
> **deep learning (DL)**: subset of ML that uses advanced layer approach with large number of parameters
>
> **corpus**: text used for training a language model; corpora refers to multiple texts
>
> **natural language processing (NLP)**: algorithms that give computers the ability to understand and process human language
>
> **prompting**: user input to the chatbot
>
> > **input/output prompt (I/O)**: model computes a single output based on input
> >
> > **chain of thought prompt (CoT)**: model breaks the problem in sequential and interconnected thoughts to arrive at a specific output
> >
> > **tree of thought prompt (ToT)**: model branches out the input problem into several thoughts and generates multiple potential outputs
>
> **large language model (LLM)**: integrated DL/NLP model trained on huge corpora to perform complex tasks
>
> **chatbot**: application interface designed to simulate human conversation
>
> **stochastic parrot**: random process of generating text based on training from pre-existing texts
>
> **conversational AI**: technology that enables computers to engage in natural and human-like conversations
>
> **plugins**: powerful software tools designed to enhance the capabilities of a chatbot
>
> **sycophancy**: tendency of chatbots for unnecessary complementing and agreeing with user input
>
> **hallucination**: phenomenon where chatbots produce irrelevant, nonsensical, or incoherent responses
>
> **generative pre-trained transformer (GPT)**: DL model that generates text, video, or audio
>
> **transformer architecture**: neural network architecture that uses focus to better capture sequences like text

the subsequent era, we saw the impact of innovations like radio [3,4] and calculators [5]. The late 20th century marked the dawn of digital learning, with the internet enabling online resources, coding platforms, and interactive applications as integral components of modern education [6]. In the current phase, computational software has further enhanced learning, enabling the exploration of complex concepts and intricate problem-solving with ease [7–12]. Despite initial resistance, each disruption has led to fundamental shifts in educational methods, highlighting the transformative potential of technology in education.

The dawn of the 21st century marked the rise of AI as a transformative force in education. Intelligent tutoring systems have personalized learning by tailoring content to individual needs [13,14]. Automated grading has freed instructors to focus on teaching rather than administrative tasks [14]. Course management systems have enhanced the efficiency of content organization and delivery [15]. Immersive learning environments have been created through virtual and augmented reality, making abstract concepts tangible [16–18]. Tools for performance intervention and emotional state detection have enabled real-time responses to learning difficulties and student engagement, fostering a more responsive educational environment [19–21]. Virtual assistants have facilitated self-directed learning, while AI-powered content creation and plagiarism detection software have enriched materials and ensured integrity [22–24].

Among the diverse AI technologies shaping education, chatbots have emerged as particularly transformative. Evolving from simple rule-based systems, they have become sophisticated conversational agents through advancements in natural language processing (NLP), transformer architectures, and attention mechanisms [25,26]. This remarkable transformation, fueled by high-performance computing and extensive datasets, enabled the training of LLMs [27]. Since 2020, the role of LLMs in education has been a point of exploration and debate. Questions about their impact on learning objectives, student workload, assignments, and academic integrity have arisen, along with considerations of how they might enhance learning experiences [28]. In this context, tools like Elicit, Claude, Poe, and ChatGPT have found specific roles: Elicit and Claude in content optimization and personalization; Poe in fostering creativity; and ChatGPT, an OpenAI product, in aiding content understanding and brainstorming. The addition of plugins and tools like Bard further extends the capabilities of these AI systems, offering instructors a tailored suite of resources for various educational needs.

In this paper, we advocate for incorporating Large Language Models (LLMs) into graduate-level STEM instruction, using a fluid mechanics course as a case study. We examine the multifaceted capabilities of LLMs, including answering specialized questions, solving equations, enabling visualization, and even interpreting PDFs. Effective prompting and third-party enhancements like the Wolfram plugin are key to unlocking these capabilities. Importantly, instructors need to grasp the underlying technology behind LLMs, especially when interfacing with chatbots, to mitigate potential drawbacks such as inaccuracies and nonsensical outputs. Beyond the technological considerations, we explore the ethical dimensions of deploying LLMs, emphasizing equity, transparency, and data privacy. The integration of LLMs offers a transformative approach to STEM education, addressing its challenges while maximizing its educational promise.

The rest of the manuscript is organized as follows: Section 2 provides an in-depth exploration of LLMs, offering insights into their architectures, capabilities, and advancements in natural language processing. In Section 3, we explored the evolution of chatbots, tracing their development from rule-based systems to sophisticated conversational LLM-based chatbots. Section 4 then addresses effective strategies for incorporating LLMs and chatbots into graduate-level STEM instruction, emphasizing the importance of tailored approaches for optimal learning outcomes. Moving forward, Section 5 presents a detailed case study centered on a graduate fluid mechanics course, showcasing practical applications and the varied capabilities of LLMs in a specific educational context, and analyzes the performance of ChatGPT on various categories of problems such as computational, analytical, and mathematical. Section 6 focuses on the assessment methodologies and outcomes of integrating LLMs in the case study, providing a nuanced understanding of their impact on student learning and engagement. Finally, in Section 7, we summarize key findings, discuss implications, and propose

future directions for the integration of LLMs in STEM education, thereby offering a comprehensive framework for educators and researchers in the field.
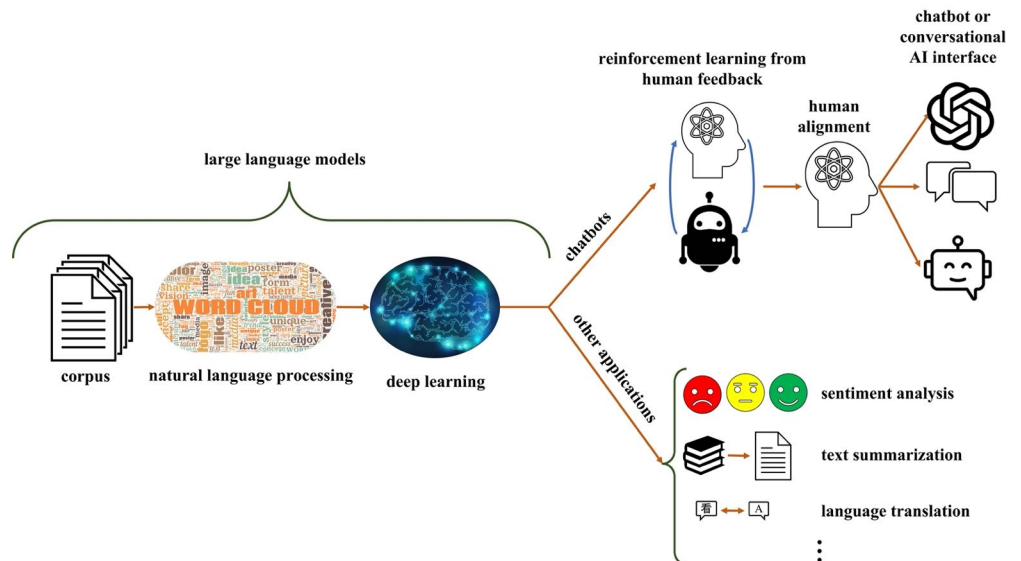
2 The Building Blocks of Educational Chatbots: An Introduction to Large Language Models

As we shift our focus from the overarching benefits of LLMs to their technical intricacies, it becomes imperative for instructors to understand what makes these models tick. This is crucial not just for academic curiosity but for practical classroom application. In section 3, we will go a step further by examining how LLMs become the beating heart of chatbots, revolutionizing the educational space in diverse and innovative ways.

LLMs, at their core, are an extension of Natural Language Processing (NLP) technologies. These models are trained on vast datasets and have the capability to perform a wide range of tasks, from summarizing articles to conversing in real-time [29–32]. In educational settings, understanding LLMs is becoming increasingly essential. They form the technological backbone of various educational chatbots and virtual assistants that instructors will likely encounter, both today and in the near future [33, 34].

Before we delve into practical applications, it is worth exploring the evolutionary journey of LLMs. These models have transitioned from basic statistical language processors to intricate systems leveraging neural networks and pre-training techniques [29, 35–43]. This development has expanded the toolkit that instructors can tap into, enriching the classroom experience.

However, LLMs are not just limited to powering chatbots. They can be aligned with ethical considerations and fine-tuned to offer targeted feedback and improved learning outcomes [44–48]. To further illuminate the distinctions between LLMs and chatbots, Figure 1 offers a comparative visual overview.

Fig 1. LLMs Training and Applications: LLMs are trained on a corpus (large volume of text data) using NLP combined with ML methods. The LLMs can be used for a variety of tasks (shown on the right side) or developed into a type of chatbot that can answer questions, carry on a conversation, or operate plugins.

When considering the advancement of LLMs, two key innovations emerge: ₉₀

LLMs Architecture: The advent of transformer architecture [49] has ₉₁ revolutionized the field by introducing attention mechanisms. These mechanisms enable ₉₂ LLMs to grasp the context within sentences or even across paragraphs, improving their ₉₃ utility in tasks like translation, summarization, and question-answering [50–52]. ₉₄

Computational Power: The efficacy of LLMs is closely tied to computational ₉₅ resources. The usage of advanced GPUs and TPUs has made it possible to train models ₉₆ with more parameters, enabling a level of performance that was previously ₉₇ unattainable [53–57]. ₉₈

As we wrap up our discussion on the innovations driving the efficacy of LLMs, it is ₉₉ crucial to recognize that these advancements in architecture or computational power are ₁₀₀ tightly coupled with the complexities of training these models. For instructors aiming to ₁₀₁ implement or understand LLMs in educational settings, grasping the underlying ₁₀₂ training procedures becomes pivotal. ₁₀₃

2.1 State-of-the-Art Large Language Model Training ₁₀₄

As we move deeper into the intricacies of LLMs training, we begin with the foundational ₁₀₅ aspect: data preparation. Instructors keen on harnessing the power of LLMs should ₁₀₆ note that the efficacy of these models starts with clean data. Preprocessing tasks like ₁₀₇ text sanitization and removing irrelevant symbols are critical first steps to ensure the ₁₀₈ quality of training. Once the data is prepared, the next pivotal step involves ₁₀₉ establishing the architecture—most often, the transformer architecture, as seen in ₁₁₀ models like GPT. Initial values for weights and biases are set during this phase, laying ₁₁₁ the groundwork for the model to develop a comprehensive understanding of language. ₁₁₂

At this juncture, it is crucial to understand that the training of LLMs is not a ₁₁₃ one-time process but occurs in distinct stages, pretraining and fine-tuning. The ₁₁₄ pretraining stage is a broad learning phase that employs a self-supervised approach, ₁₁₅ where the model absorbs the intricacies of language from vast volumes of unlabeled ₁₁₆ data. This is similar to a student engaging in broad academic study before specializing. ₁₁₇ As the model transitions into the fine-tuning stage, it is as if the student moves into ₁₁₈ specialized coursework: the model learns from smaller, task-specific labeled ₁₁₉ datasets [58–60]. This specialization allows the model to apply its generalized ₁₂₀ knowledge to perform specific tasks, enhancing its utility in educational settings [30]. ₁₂₁

2.2 Challenges and Caveats of Large Language Model ₁₂₂ Integration for Education ₁₂₃

As LLMs increasingly find applications in higher education, it is crucial for instructors ₁₂₄ and students to comprehend their inner workings, ethical implications, alignment with ₁₂₅ human values, and the importance of model training and dataset quality [61–63]. While ₁₂₆ chatbots offer various benefits, they also introduce challenges such as hallucination, ₁₂₇ privacy, and overreliance. ₁₂₈

Hallucination: Researchers have sometimes observed that LLMs provide incorrect ₁₂₉ or misleading information [64, 65]. Although the text produced by LLMs might exhibit ₁₃₀ grammatical accuracy and logical structure, it can sometimes suffer from semantic ₁₃₁ inaccuracy, misdirection, conflicts, or misalignment with the given context or ₁₃₂ established facts. Researchers and developers can tackle this challenge by furnishing ₁₃₃ models with superior-quality data or incorporating human feedback. ₁₃₄

Outdated knowledge: LLMs might experience difficulty in resolving problems that demand the latest information. However, researchers are actively working on strategies to ensure that these models are consistently supplied with the latest insights, potentially mitigating this concern [66,67].

Harmful content: LLMs possess the capability to produce content that could be harmful, potentially amplifying societal problems and nurturing hostility [68,69]. Deliberate examination of initial GPT-4 iterations unveiled this risk, prompting model rejections as a measure to alleviate potential harm. The released version of GPT-4 demonstrates significant improvements in reducing harmful content generation.

Privacy: LLMs are trained on diverse datasets, have the capacity to synthesize information, and can potentially discern individuals while paired with external data [70]. Actions such as expunging personal information from datasets and employing automated assessments are put into effect to mitigate risks to privacy. Subsequent improvements are directed at curbing privacy hazards by concentrating on information provided by users [71,72].

Overreliance: A growing concern with increasing model capability can lead to unnoticed errors, inadequate oversight, and potential skill loss [73]. Mitigation strategies include providing comprehensive documentation, refining the model's refusal behavior, and encouraging critical evaluation of model outputs. However, challenges persist, such as the model's tendency to hedge responses, which may inadvertently foster overreliance, and users potentially becoming less attentive to refusal cues over time.

To address these challenges, there are two effective approaches available to all end users: collaborative optimization and user-led customization. Collaborative optimization involves a team of experts across various domains working together to tailor LLM chatbots for specific purposes. This method includes using strategies like OpenAI's Red Teaming to proactively address security, accuracy, and ethical issues, ensuring the chatbots are reliable and user-friendly. On the other hand, user-led customization involves developing fine-tuned GPT models for particular subjects or applications. For example, tools like OpenAI GPT builder enable users to create models that adhere to specific guidelines, aligning with their unique objectives and enhancing the overall experience.

Understanding how to manage these challenges often involves a deep dive into the model's inner mechanisms, including its stochastic behavior. One way to gain more control over model outputs and mitigate some of these challenges is through a concept known as temperature scaling.

2.3 The Stochastic Nature of LLMs: The Concept of Temperature

Controlling LLMs responses is a fundamental concern for instructors aiming to maximize the utility of these models in academic settings. Temperature scaling provides a nuanced approach to balancing the model's certainty and creativity, helping to address some of the aforementioned challenges, such as generating hallucinated or overly rigid content.

Temperature scaling assumes a pivotal methodological role in governing distribution smoothness, carrying significant implications for augmenting the performance of intricate models involved in NLP tasks [74]. By employing lower temperature parameters, the models demonstrate heightened certainty and consistency, attributes crucial for precision-demanding tasks. Nonetheless, this can result in outputs perceived

as excessively rigid or repetitive [75]. ₁₈₂

In contrast, applying higher temperature parameters infuses the models with an ₁₈₃ element of stochasticity, fostering a more diverse and creative output and facilitating the ₁₈₄ exploration of unconventional solutions and novel ideas. On the other hand, excessively ₁₈₅ high temperatures may result in incoherent responses unrelated to the original query. ₁₈₆ The methods of implementing temperature scaling are multifaceted, encompassing ₁₈₇ constant temperature applications [76,77], dynamic adjustments over training ₁₈₈ iterations [78], or variations based on the specific word position within a sentence [74]. ₁₈₉

The practical significance of temperature scaling in LLMs is particularly salient ₁₉₀ when these models are adapted into real-world applications such as automated chat ₁₉₁ systems or chatbots. By calibrating the temperature, instructors can tailor the chatbot ₁₉₂ responses to be either more precise or more creative, depending on the educational ₁₉₃ context. With this foundational understanding of LLMs and their stochastic behavior, ₁₉₄ we are better equipped to explore their most interactive form: chatbots. ₁₉₅

3 Chatbots: Interfacing with Humans ₁₉₆

Building on the foundations of Large Language Models (LLMs), chatbots have rapidly ₁₉₇ become a pivotal component of human-computer interaction [79,80]. Their unique blend ₁₉₈ of algorithmic complexity and linguistic proficiency enables real-time dialogue with ₁₉₉ users, adding a layer of interactivity that holds significant implications for higher ₂₀₀ education. These advanced capabilities are the culmination of decades of research and ₂₀₁ innovation in Artificial Intelligence and computational linguistics. This section will ₂₀₂ explore the historical roots of chatbots, delve into their evolution over the years, ₂₀₃ examine their applications in academia, and discuss the tools and methodologies that ₂₀₄ facilitate their effective integration into educational settings. ₂₀₅
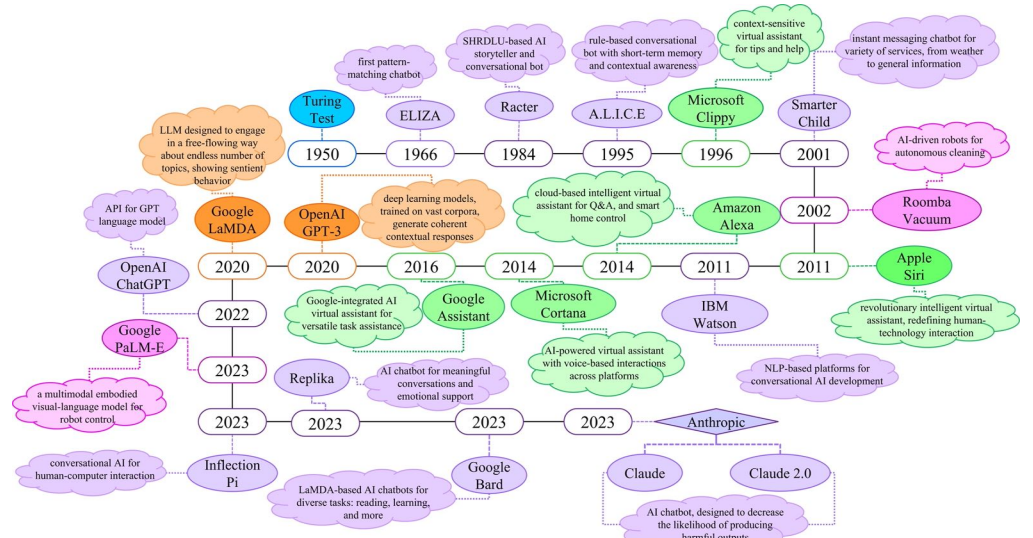
3.1 History and Evolution of Chatbots and Conversational AIs ₂₀₆

For instructors interested in understanding the roots of chatbots, the story starts with ₂₀₇ the Turing Test, introduced in 1950 by Alan Turing [81–83] (as shown by the blue node ₂₀₈ in Figure 2). The Turing Test aimed to determine if a machine could exhibit behavior ₂₀₉ indistinguishable from human intelligence. This intellectual query led to the creation of ₂₁₀ ELIZA in the 1960s by Joseph Weizenbaum [84–86] (the first purple node on the upper ₂₁₁ layer in Figure 2). Serving as a rudimentary psychotherapist, ELIZA used ₂₁₂ pattern-matching techniques to simulate conversation. ₂₁₃

In the 1970s, rule-based chatbot systems became prevalent, where predefined rules ₂₁₄ determined the responses to user inputs. Richard Wallace's A.L.I.C.E. (Artificial ₂₁₅ Linguistic Internet Computer Entity) was a notable example that engaged users in ₂₁₆ open-ended conversations using an extensive set of predefined rules [87,88]. A.L.I.C.E. ₂₁₇ incorporated context awareness through short-term memory and refined its responses by ₂₁₈ storing previous inputs and corresponding responses [89,90], which led to the ₂₁₉ development of other chatbot systems like MegaHAL [91,92] and Jabberwacky [93–95]. ₂₂₀

The twenty-first century has witnessed remarkable advancements in AI and NLP ₂₂₁ techniques, leading to the development of sophisticated and innovative chatbots [96–98]. ₂₂₂ These chatbots can understand user queries and provide meaningful responses, making ₂₂₃ them invaluable for various applications such as customer service, personalized support, ₂₂₄ and product recommendations [99,100]. Additionally, they excel at automating ₂₂₅ mundane tasks and extracting valuable insights from vast datasets [101,102]. For ₂₂₆ instance, SmarterChild, an instant messaging chatbot initially developed by ₂₂₇
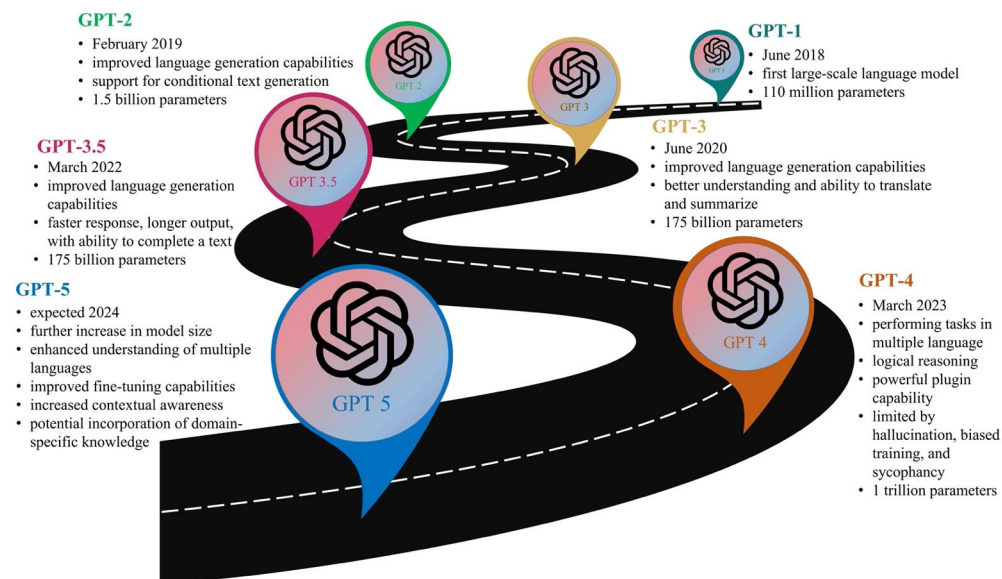
Fig 2. Choronological Development, Evolution, and Progression of LLMs and Chatbots: Advanced language models, capable of processing vast amounts of data, chatbots, and conversational AI agents have evolved and interacted with each other, leading to advancements in natural language understanding, human-computer interactions, and AI-driven products. In the Figure, the orange objects refer to LLMs, purple objects refer to conversational AI or chatbots, green objects refer to AI-based virtual assistants and pink objects refer to AI-driven objects.

ActiveBuddy and later acquired by Microsoft, delivers various services ranging from weather reports to general information [103] (the purple node on the right side of the upper layer in Figure 2). One pivotal milestone in the chatbot landscape was the introduction of Siri (Speech Interpretation and Recognition Interface) by Apple [104, 105] (the green node at the beginning of the middle layer in Figure 2 as the first intelligent virtual assistant). Siri, an intelligent virtual assistant, revolutionized how humans interact with technology and chatbots [106]. Its success has not only transformed user expectations but also paved the way for the development of other notable virtual assistants such as Alexa by Amazon [107–109], Cortana by Microsoft [110–113], and Google Assistant by Google [114–117].

Recent years have witnessed the emergence of conversational AI, driven by advances in ML, DL, and NLP. Platforms such as IBM Watson [118–120] and Google Dialogflow [121–124] have been instrumental in the development of conversational AI and the enhancement of the understanding and implementation of NLP techniques. One of the most notable achievements in the field of conversational AI is the development of powerful language models such as GPT (the orange node near the end of the middle layer in Figure 2 represents the onset of chatbots and artificial intelligence), which have significantly improved the understanding and generation capabilities of chatbots [125–128]. Developed by OpenAI, these models are trained on vast amounts of text data using advanced DL architectures to generate coherent and contextual human-like responses [129]. With each iteration, these models become more complex, introducing new functionalities to the corresponding chatbot, including conditional text generation (GPT 2); translation and summarization (GPT 3); faster outputs and text completion (GPT 3.5); multilanguage functions, logical reasoning, and robust API plugins (such as Wolfram Alpha and ScholarAI) (GPT 4) [125, 127, 130]. Figure 3 illustrates the performance enhancement and additional functionalities in different

versions of GPT models as the number of input parameters in the DL model increases. ₂₅₄

Fig 3. The Highlighted Road Map of GPT Evolution: From the early inception of GPT to its progressive iterations, this roadmap showcases the rapid development of powerful language generation models based on transformer architecture. It also depicts the major breakthroughs and refinements leading to more sophisticated and capable GPT variants that have significantly impacted various NLP tasks.

The release of ChatGPT as a chatbot based on GPT models by OpenAI marks a ₂₅₅ pivotal milestone in the development of chatbots and conversational AIs. After that, ₂₅₆ Inflection released the Pi chatbot [131], designed to be a kind and supportive ₂₅₇ companion offering conversations, friendly advice, and concise information in a ₂₅₈ human-like and coherent style. Replika is another companion chatbot programmed to ₂₅₉ learn and mimic people's writing style, offering an unprecedented sense of comfort and ₂₆₀ well-being [132–135]. Similar to other technological giants, Google has also contributed ₂₆₁ to fields of LLMs through the development of several models, including ₂₆₂ LaMDA [136–138] and PaLM [139–141]. Despite many obstacles in the development ₂₆₃ and implementation procedure, such as sentinent behavior, LaMDA, and PaLM are ₂₆₄ utilized to develop Bard chatbot [142,143]. Bard is an AI-powered chatbot designed to ₂₆₅ be more conversational and generate more diverse and prolonged responses than GPT ₂₆₆ models [144,145]. Claude and Claude 2.0 are two of the other powerful chatbots ₂₆₇ released recently, with superior performance to generate longer responses and nuanced ₂₆₈ reasoning [146,147]. Claude 2.0 has made significant strides in multiple fields, including ₂₆₉ law and mathematics. It scored 76.5% in the Bar exam's multiple-choice section and ₂₇₀ achieved a score higher than 90% of graduate school applicants in GRE reading and ₂₇₁ writing exams [148]. Figure 2 provides a comprehensive overview of the evolution ₂₇₂ timeline of chatbots, LLMs, conversational AIs, and AI-based products. ₂₇₃
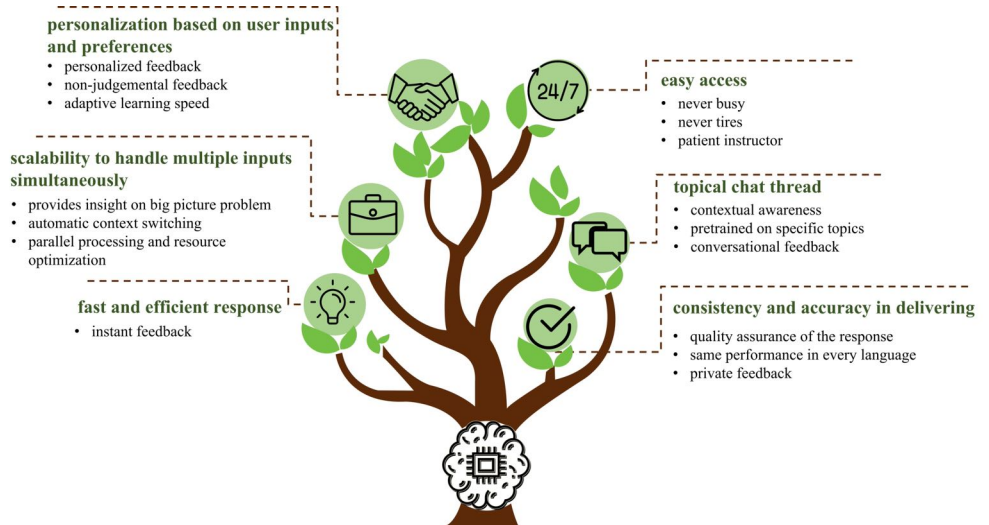
3.2   Higher-Education Implication of Chatbots ₂₇₄

Chatbots, powered by advancements in conversational AI and sophisticated language ₂₇₅ models like GPT, PaLM, and LaMDA have revolutionized various domains, including ₂₇₆ education. These models have significantly enhanced the natural language ₂₇₇ understanding and generation capabilities of chatbots, allowing them to generate ₂₇₈ coherent and contextually appropriate responses [96,129,149,150]. In education, these ₂₇₉

advancements offer numerous advantages to researchers [151], students [152], and instructors [153]. Chatbots can provide personalized and adaptive learning experiences by tailoring content and resources to individual student needs and learning styles, leading to improved engagement and knowledge retention [154,155]. They can offer immediate and accurate responses to student inquiries, providing timely support and guidance that enhances the learning process [156–158]. Additionally, chatbots facilitate continuous assessment and feedback, allowing students to receive prompt evaluations and insights into their progress, enabling them to identify areas for improvement [159,160]. They can also assist in delivering educational materials and resources, offering on-demand access to information, and promoting self-directed learning [161]. Moreover, chatbots foster collaborative learning experiences by facilitating group discussions and providing opportunities for peer interaction [162]. By automating administrative tasks, such as grading and scheduling, chatbots support instructors by freeing up their time to focus on instructional activities [159,163]. These advantages of chatbots in education encompass personalized learning, timely support, continuous assessment, resource delivery, collaboration, and administrative assistance [164–166]. Figure 4 illustrates some of the merits of chatbot implementation for student and education purposes.

In addition to educational benefits, researchers could also benefit from utilizing chatbots. Chatbots have substantially enhanced the present state of information retrieval, providing a notable improvement in the accessibility of information compared to basic Google searches. Furthermore, these chatbots, augmented with powerful plugins, open up new gateways to access the internet's vast repository of knowledge, e.g., OpenAI has introduced plugins as an added feature in its latest versions of ChatGPT, where users can select and utilize up to three plugins at a time for their specific tasks, with the option to change their selection as needed, providing valuable assistance to both students and researchers in higher education, enabling them to streamline their tasks and achieve their objectives with reduced effort. Table 1 outlines several plugins integrated into ChatGPT, offering higher education and research applications. Apart from plugins, chatbots have also tapped into specialized AIs that focus on discovering academic papers, exemplified by Elicit, Semantic Scholar, and ScholarAI. These advancements encompass the critical stages of research, ranging from writing and editing to a meticulous review of papers. Additionally, chatbots have evolved to excel at summarizing complex research papers, rendering them invaluable aids for researchers and learners seeking to distill intricate knowledge efficiently.

While chatbots bring numerous benefits to education and research, they also face drawbacks and limitations. One limitation is their inability to fully replace the human element of interaction and personalized instruction [26]. Despite being able to provide immediate responses and support, chatbots lack the empathy and nuanced understanding that human instructors possess, which can be particularly challenging for students requiring individualized attention or those with complex learning needs. Additionally, chatbots may need help to accurately assess subjective assignments or provide in-depth feedback beyond essential evaluation criteria [167]. Technological barriers and disparities pose another limitation due to students' unequal access to the necessary devices, stable internet connections, or technological literacy, hindering their engagement with chatbot-driven educational experiences. Moreover, there is a risk of chatbots inadvertently reinforcing a transactional approach to learning, prioritizing correct answers over critical thinking, creativity, and problem-solving skills [168], limiting opportunities for comprehensive learning experiences, exploration, and intellectual growth. Furthermore, one concern regarding chatbots is their potential to perpetuate biases and reinforce stereotypes in their training data [162,169]. If the

Fig 4. Chatbots in Education: A transformative tool offering students instant access, tailored learning experiences through topical threads, rapid responses, and unparalleled personalization for an optimized learning journey.

training data are biased or lack diversity, the chatbot's responses may unintentionally exhibit biases or engage in discriminatory behavior. Privacy and data security concerns also arise when utilizing chatbots, requiring careful data handling to comply with privacy regulations and safeguard sensitive information [170,171]. Additionally, technical glitches, system failures, or programming errors can disrupt the learning process and lead to frustration among researchers, students, and instructors [172]. It is essential to address these limitations thoughtfully and consider chatbots as complementary tools that work alongside human instructors to create a comprehensive and effective educational environment.

As we have explored the diverse merits and potential pitfalls of chatbot implementation in educational settings, an important conclusion emerges: successful adoption is not a plug-and-play endeavor. Instead, it necessitates systematic adaptations in curriculum design, instructional techniques, and communication protocols [152]. Instructors, therefore, should not just be passive consumers of this technology; they should actively engage in restructuring course objectives, designing chatbot-centric assignments, and establishing guidelines for effective chatbot utilization. Instructor training programs, too, must evolve, equipping instructors with the skills to weave chatbots seamlessly into their pedagogical toolkit [163]. By proactively embracing these adjustments, we pave the way for a more dynamic and enriched educational experience.

4 Chatbots: Strategies and Tools

Building upon the discussed chronological evolution and real-world implications of chatbots, we will examine the key methods and tools used in chatbots, highlighting from prompt techniques to customizing interactions and providing an in-depth review useful for graduate students and instructors.

4.1 Basic Prompting

Prompting serves as the bedrock for enhancing the utility of chatbots, particularly those employing advanced language models like ChatGPT. In essence, a prompt is a set of

Plugins/Features	Description
Browsing	Browsing Alpha integrates the Bing search API, enabling ChatGPT to browse and retrieve web content directly from the internet.
Retrieval	The Retrieval feature grants ChatGPT the ability to access authorized personal or organizational information sources. By leveraging natural language queries, users can retrieve pertinent document snippets from various data sources, including files, emails, notes, and public documents.
Advanced Data Analysis	Enhances the capabilities of ChatGPT by enabling it to perform a variety of tasks. These tasks include numerically solving mathematical problems conducting data analysis and visualization on your dataset file, and converting file formats using Python.This feature was previously introduced as Code Interpreter.
Wolfram	Empowers ChatGPT users with advanced computational capabilities, enabling them to solve various types of queries, including mathematical problems and computations.
AskYourPDF	Generate summaries of file content in response to user inquiries.
ScholarAI	It searches millions of peer-reviewed articles based on the user prompt. It also obtains the DOI and provides a short summary of the article.
Custom Instructions	Allows the users to specify or customize their preferences or requirements about the responses way they want to receive from the chatbot.

Table 1. Shows the description of some available plugins and features that can be used in higher education. The features mentioned are exclusively accessible to ChatGPT-4 subscribers, with the exception of Custom Instructions, which are available to both ChatGPT-3 and ChatGPT-4 users.

instructions that directs the chatbot toward generating a specific kind of response [173–175]. It acts as the steering wheel for the chatbot, guiding it to produce outcomes that align with the desired objectives [176–178]. Therefore, crafting clear and structured prompts becomes indispensable for optimizing the accuracy and relevance of the chatbot's generated responses [176].

For instance, while searching for the momentum equation for a steady, inviscid, and incompressible flow, an improper prompt like 'provide the momentum equation' may produce a more generalized answer. However, a more specific prompt like 'provide the momentum equation for a steady, inviscid, and incompressible flow' ensures an accurate response.

Understanding the intricacies of prompt engineering is not merely a technical endeavor but a practical one for instructors. This understanding allows instructors to leverage tools like ChatGPT in versatile ways, from crafting dynamic quizzes to designing comprehensive course plans. Thanks to the adaptability of prompt configurations, ChatGPT can be fine-tuned to offer a multitude of prompt suggestions, gathering diverse and relevant educational data [176, 179].
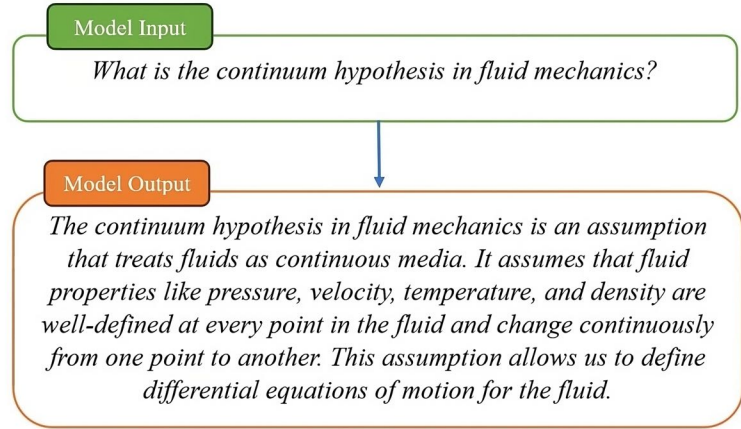
4.2 Prompt Strategies

Before diving into the advanced prompting methods, it is essential to understand the foundational principles that govern effective prompting. These can be broadly

categorized into specific words, context provision, and mechanisms for eliciting more   377
detailed information [180]. ChatGPT utilizes various advanced prompting strategies to   378
solve distinct challenges, yielding the desired outputs. Specifically, three principal   379
methods stand out: the Input-Output (I/O) method, the Chain of Thought (CoT), and   380
the Tree of Thoughts (ToT) prompt [181].   381

In I/O prompting, the model is provided with various input samples alongside their   382
respective outputs which guides the model in producing the desired output for a given   383
input [181]. Figure 5 shows an example of I/O prompting using ChatGPT-4. In this   384
scenario, we posed a query 'What is the continuum hypothesis in fluid mechanics' to   385
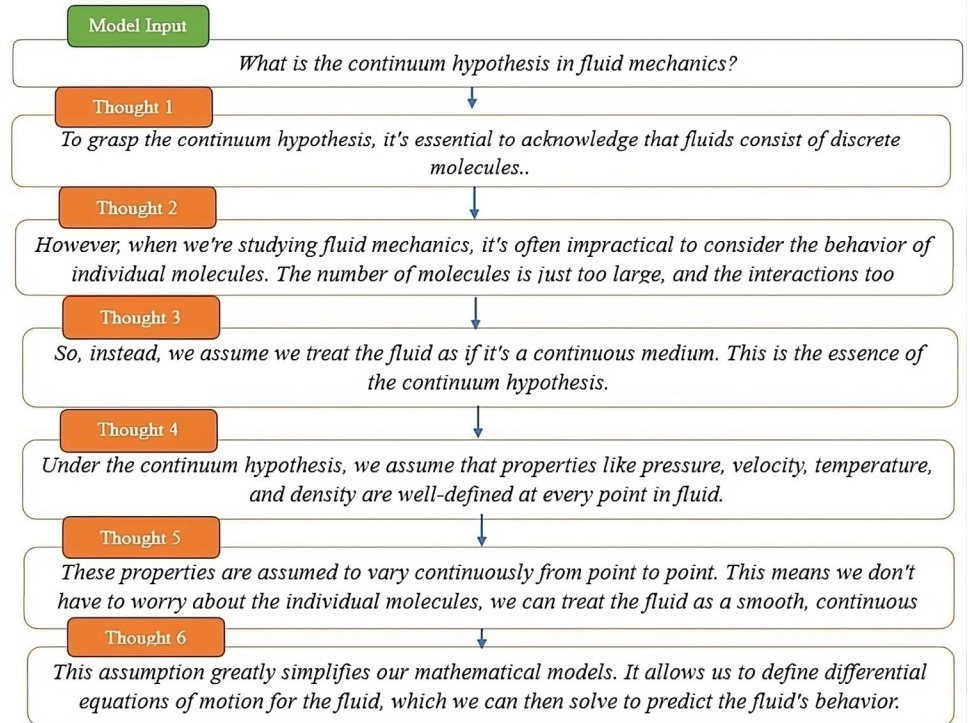ChatGPT-4., which in turn resulted in the corresponding response.   386

Fig 5. I/O Prompting Mechanism: Inputs are fed into the system, and based on internal processing, corresponding outputs are generated, showcasing the dynamic interaction and response mechanism

CoT breaks the input problem into a series of logical thoughts and reaches the final   387
desired output based on previous thoughts. This approach is useful for intricate   388
mathematical or reasoning tasks that require step-by-step breakdowns. It clarifies the   389
model's actions and its reasoning, making error detection more   390
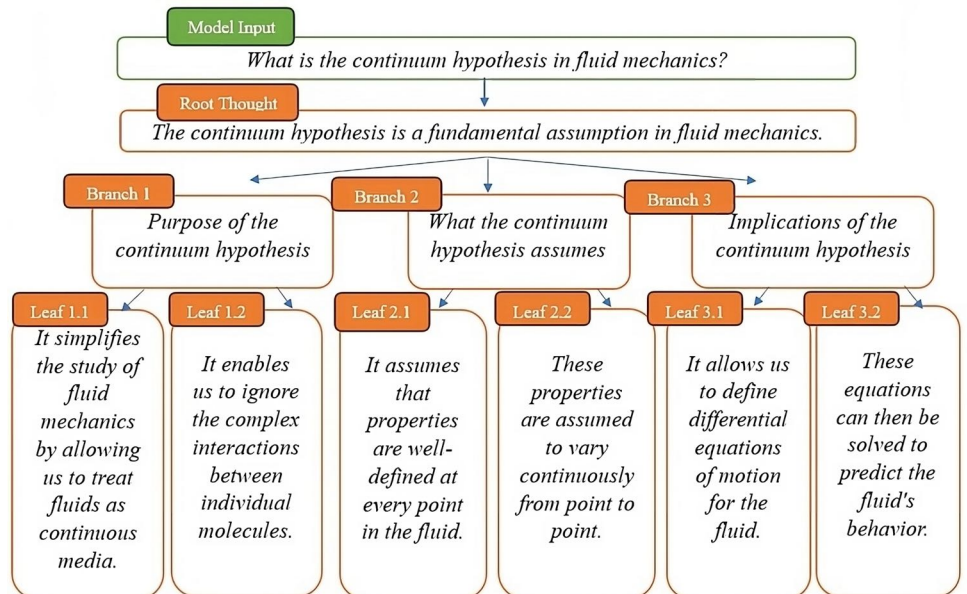straightforward. [181]–[184].   391

Figure 6 shows the example of CoT prompting using ChatGPT-4. We used the CoT   392
prompting technique in this scenario to pose the same question. ChatGPT   393
deconstructed the query into several thoughts, each directing the model toward the final   394
answer.   395

ToT is a strategy for structuring the model's output to branch out into different   396
related thoughts. This approach can be beneficial for complicated questions that involve   397
multiple related ideas. By structuring the response as a tree of thoughts, the model can   398
cover a wide range of related ideas in a structured and organized manner. ToT solves a   399
problem by searching over a tree where each node represents a partial solution based on   400
input and previous thought sequences [181]. Figure 7 shows the example of ToT   401
prompting. We attempted to utilize ToT in a more straightforward manner without any   402
coding. Yet, the prompt illustrated in Figure 7 may not genuinely respond to queries   403
using the tree of thoughts. Hence, we believe there's a pressing need for an improved   404
methodology to draft prompts for ToT.   405

Within the rest of this study, to instruct ChatGPT to induce the ToT prompt, we   406
have utilized the following command provided by Dave Hulbert [185]. This prompt   407
initiates a discussion that investigates several logical paths, possibly leading to the   408

Fig 6. CoT Prompting Mechanism: CoT breaks the problem into multiple sequential and interconnected steps, and each step gives a clear idea of how the model reaches the final output.

Fig 7. ToT Prompting Mechanism: ToT branches out the input into different related thoughts, which help the model reach the final outputs.

desired outcome of the question. From a discussion standpoint, this prompt may ₄₀₉ effectively provide distinct paths to answer conceptual questions, but it might be less ₄₁₀ efficient in addressing analytical or mathematical problems, as discussed in section 5.2. ₄₁₁

Imagine three different experts are answering this question.
All experts will write down 1 step of their thinking, then
share it with the group. Then, all experts will go on to the
next step. If any expert realizes they're wrong at any point,
then they leave. [**The question is ...**]

4.3   How to use a prompt properly?   ₄₁₂

ChatGPT is a valuable resource for both students and instructors at graduate, ₄₁₃ undergraduate, and professional levels. It can craft course syllabi, design projects, ₄₁₄ supply fundamental formulas, offer research articles, and even design and assess student ₄₁₅ quizzes. The best possible results can be achieved by supplying context and employing ₄₁₆ plugins. Below are some illustrations showcasing the advantages of using ChatGPT in a ₄₁₇ graduate fluid mechanics course. ₄₁₈

Example provided by the instructor: *We are learning the fundamentals of* ₄₁₉ *turbulence for a graduate-level fluid mechanics class. Today, we are focused on the* ₄₂₀ *Navier-Stokes equation and its application. Please generate a summary of this topic* ₄₂₁ *followed by a quiz with five multiple-choice questions. We will attempt to answer the* ₄₂₂ *questions next, and then we will ask you to check our results and provide reasons for the* ₄₂₃ *correct answers.* ₄₂₄

ChatGPT-4 provided a concise overview of the Navier-Stokes equation and its ₄₂₅ practical uses in the given scenario, accompanied by a quiz of 5 multiple-choice ₄₂₆ questions. When prompted on the preferred answer format for students, ChatGPT ₄₂₇ emphasized clarity, suggesting any method that distinctly matches answers to their ₄₂₈ corresponding questions. Using a streamlined format—listing answers like 1A, 2B, 3C, ₄₂₉ 4D, and 5B, it assessed the responses, highlighted correct and incorrect answers, and ₄₃₀ elucidated the reasoning behind each question. ₄₃₁

Example provided by the instructor: *We are learning about the fundamentals* ₄₃₂ *of turbulence in today's class. Using ScholarAI, find the most recently published papers* ₄₃₃ *on direct numerical simulation (DNS) focusing on the flow over an airfoil. Based on* ₄₃₄ *what you find, provide three short-answer questions I can try to answer, and you can* ₄₃₅ *check.* ₄₃₆

In response to this example, ChatGPT-4 presented us with some of the latest papers ₄₃₇ via the ScholarAI plugin and posed three brief questions based on the content of those ₄₃₈ papers. After submitting our responses, it assessed and provided explanations for each ₄₃₉ question. ₄₄₀

Example provided by the instructor: *We are learning about the fundamentals* ₄₄₁ *of shock waves in today's class. Our learning goal is to understand shock waves'* ₄₄₂ *formation, propagation, and impact. Please provide five real-life examples of shock* ₄₄₃ *waves. Based on the examples, provide five short-answer questions I can try to answer,* ₄₄₄ *and you can check.* ₄₄₅

ChatGPT-4 presented five practical instances of shock waves based on the given ₄₄₆ scenario. Following that, it posed five questions based on these examples. After we ₄₄₇ submitted our responses, it analyzed them and offered detailed explanations for each ₄₄₈ question. ₄₄₉

Example provided by the instructor: *We are learning about the fundamentals* ₄₅₀
of shock waves in today's class. Our goal is to learn about shock tubes. Please give me a ₄₅₁
detailed theory of shock tubes with the necessary equations and design a project for me ₄₅₂
with proper initial conditions so that I can analytically calculate the change of pressure, ₄₅₃
velocity, temperature, and density across the shock wave in the tube and you can check ₄₅₄
the final answers. ₄₅₅

In response to the above example, ChatGPT-4 provided details about the shock ₄₅₆
tubes, including the equations representing how various properties alter across the shock ₄₅₇
wave. Subsequently, it outlined a project with the necessary initial conditions. We ₄₅₈
analytically calculated the change of different properties like pressure, velocity, ₄₅₉
temperature, and density across the shock wave. After sharing our results with ₄₆₀
ChatGPT, it evaluated our responses. ₄₆₁

Example provided by the instructor: *We are learning about vortex generation* ₄₆₂
and shedding fundamentals. Please give me a detailed theory on these topics with the ₄₆₃
necessary mathematical formulations. Give me a few real-life examples of vortex ₄₆₄
shedding, followed by a quiz with three short answer questions based on applications of ₄₆₅
vortex shedding in Engineering. We will attempt to answer the questions, and you will ₄₆₆
evaluate our responses. ₄₆₇

In response to the above example, ChatGPT-4 provided details about the vortex ₄₆₈
generation and shedding with necessary equations, followed by a few real-life examples. ₄₆₉
It then posed three questions concerning the application of vortex shedding in ₄₇₀
engineering. After we submitted our responses, it evaluated them and explained each ₄₇₁
question. ₄₇₂

Example provided by the instructor: *In today's class, we are learning about the* ₄₇₃
flow around a circular cylinder. We are particularly interested in the pressure ₄₇₄
distribution around the cylinder. I have sent you all a paper. Now using Ask Your PDF, ₄₇₅
tell ChatGPT to summarize this paper and find out the most important points. Also, ask ₄₇₆
any question to GPT if you need help understanding anything from the pdf. ₄₇₇

In the provided scenario, the *Ask Your PDF* tool prompted us with a link to upload ₄₇₈
our document. Once the paper was uploaded and its URL was shared with ChatGPT, ₄₇₉
the system could provide summaries and address queries originating from the content of ₄₈₀
the PDF. ₄₈₁

The conversational learning feature of ChatGPT allows students to learn through ₄₈₂
dialogue and interactions [180, 186]. Rather than traditional Google searches, students ₄₈₃
can directly ask GPT for recent papers or specific answers. Integrations with plugins ₄₈₄
like Wolfram, ScholarAI, Advanced Data Analysis, and Custom Instructions enhance ₄₈₅
the learning experience. Furthermore, ChatGPT can recall previous conversations until ₄₈₆
a new chat begins. Students and instructors can engage in back-and-forth discussions, ₄₈₇
seek more profound understanding, and verify any information. Due to its training on ₄₈₈
diverse topics, ChatGPT can identify interdisciplinary connections. With the help of ₄₈₉
Advanced Data Analysis and Custom Instructions, many previously impossible tasks ₄₉₀
have become straightforward. Thus, we will delve deeper into these tools to ₄₉₁
demonstrate their application in graduate-level fluid mechanics courses. ₄₉₂

4.4   Advanced Data Analysis ₄₉₃

The advanced data analysis, an experimental ChatGPT model, is designed to execute ₄₉₄
Python code, manage file uploads, and facilitate downloads within a well-structured ₄₉₅
environment. This innovative feature operates with persistence and security, ₄₉₆
maintaining a session that remains active throughout a chat conversation, albeit with ₄₉₇

an upper-bound timeout. This continuity enables subsequent code executions to build ₄₉₈ upon each other, fostering a dynamic and interactive experience. Additionally, the ₄₉₉ environment is equipped with ephemeral disk space and supports an array of ₅₀₀ widely-used Python packages. The system allows for the uploading of files to the ₅₀₁ current conversation workspace and the downloading of the results, thereby ₅₀₂ substantially enhancing the functionalities of ChatGPT. This advancement represents a ₅₀₃ significant leap in capabilities, allowing for many sophisticated tasks and applications to ₅₀₄ be adeptly performed within the confines of the conversation. ₅₀₅

- The advanced data analysis can process data from user-uploaded files. It can ₅₀₆ detail the dataset, execute statistical evaluations, craft visual representations like ₅₀₇ graphs, identify data trends, and forecast outcomes. ₅₀₈

- It can execute image processing tasks, alter the image's color scheme, and modify ₅₀₉ its form while preserving the original aspect ratio. ₅₁₀

- Another excellent feature of the advanced data analysis is its file conversion ₅₁₁ function. It can convert one format of the file to another format based on the ₅₁₂ user's preference. ₅₁₃

- The Advanced Data Analysis can create, analyze, and optimize codes. ₅₁₄ Additionally, it can identify coding errors and offer corrective suggestions. ₅₁₅

- In addition to powerful Python visualization libraries, advanced data analysis ₅₁₆ allows users to develop interactive visualizations and interfaces through Streamlit. ₅₁₇ Streamlit helps users quickly and easily build data-driven apps, dashboards, and ₅₁₈ reports with minimal code. It also allows developers to create complex ₅₁₉ visualizations and interfaces. Additionally, Streamlit offers a wide range of ₅₂₀ features to make it easier for developers to build high-quality applications. ₅₂₁

The course instructors and students can gain significant advantages from Advanced ₅₂₂ Data Analysis. Below, some instances are highlighted demonstrating the effective ₅₂₃ utilization of advanced data analysis. ₅₂₄
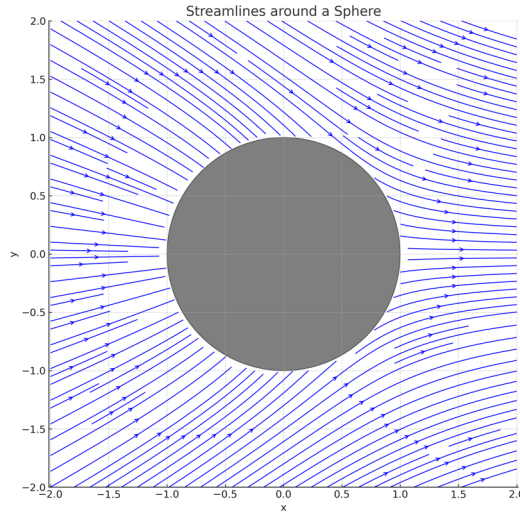
Example 1: *We are learning the fundamentals of streamlines, stream functions,* ₅₂₅ *and vortex shedding for a graduate-level fluid mechanics class. Today we are particularly* ₅₂₆ *interested in how streamlines behave when there is a flow passing over a spherical body.* ₅₂₇ *Write a Python code that will generate multiple streamlines over a circular body. Do not* ₅₂₈ *show any streamlines inside the circular body. Save the visualization in a high-resolution* ₅₂₉ *image.* ₅₃₀

ChatGPT executed the Python code in response to this example and generated the ₅₃₁ following high-quality image (Figure 8). However, it included a caveat, noting that the ₅₃₂ plot is rudimentary. For a more accurate visualization, it is advised to employ more ₅₃₃ robust computational methods like DNS (Direct Numerical Simulation) or LES (Large ₅₃₄ Eddy Simulation). ₅₃₅

Example 2: *Today, we are particularly interested in the drag force and drag* ₅₃₆ *coefficient measurement. We have experimental data for measuring the drag force over a* ₅₃₇ *streamlined hemisphere. The drag force was measured by a force transducer, and the* ₅₃₈ *drag coefficient was measured by the empirical equation. Now, read the experimental* ₅₃₉ *data file. Consider the 1st row and 1st column as the headers. Show the correlation* ₅₄₀ *matrix in a heat map. Now save the visualization in high-resolution images.* ₅₄₁

The ChatGPT read and described the data file in response to this example. ₅₄₂ Subsequently, a correlation matrix was presented in Figure 9, depicting the ₅₄₃ interrelationships between each variable. The correlation values span from -1 to 1, ₅₄₄

Fig 8. Flow Over a Circular Body Generated by Advanced Data Analysis: Advanced Data Analysis generated multiple streamlines, showing the flow direction from left to right, around the circular body. However, it recommended more robust computational techniques like DNS or LES for better visualization.

where 1 signifies a perfectly positive correlation, -1 denotes a perfectly negative correlation, and 0 suggests no correlation.

Example 3: *We are learning the fundamentals of streamlines, stream functions, and vortex shedding for a graduate-level fluid mechanics class. I have uploaded an image that is from a very classic fluid mechanics problem. The image is in greyscale but we need a colored version of it. Make it a colored image. Put greyscaled and colored images side by side and generate very high-resolution images which I can use in a report.*

In response to this example, ChatGPT transformed a grayscale image into a colorized version, adeptly preserving the intensity levels, as displayed in Figure 10.

Example 4: *We are learning the fundamentals of shock waves and are particularly interested in oblique shock waves. I have uploaded a data file which is basically an oblique shock table [187]. Consider the first row as a header. Now, describe the data file in brief and create a few important plots. Also, tell me what should be the value of β at θ=5 and M_1=1 and 1.5.*
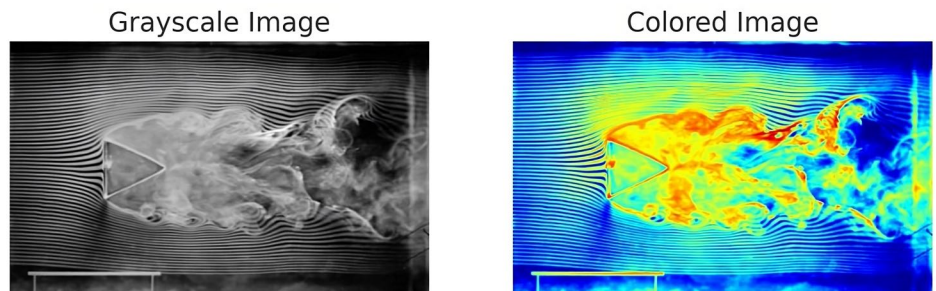
In response to this example, ChatGPT described the data file nicely in short and generated Figure 11. Within the dataset, certain conditions exhibited detached shock waves, which were adeptly represented in the visualizations by the Advanced Data Analysis. Additionally, ChatGPT provided precise values for β at θ=5 and M_1=1 and 1.5, indicating its capability to interact with specific rows and columns effectively.

4.5   Custom Instructions

In a recent development, OpenAI unveiled an innovative feature for its advanced conversational AI, ChatGPT, in July 2023. This feature, known as Custom Instructions, allows students in higher education, researchers, and graduate students to specify their preferences or requirements, which the AI system takes into account while formulating
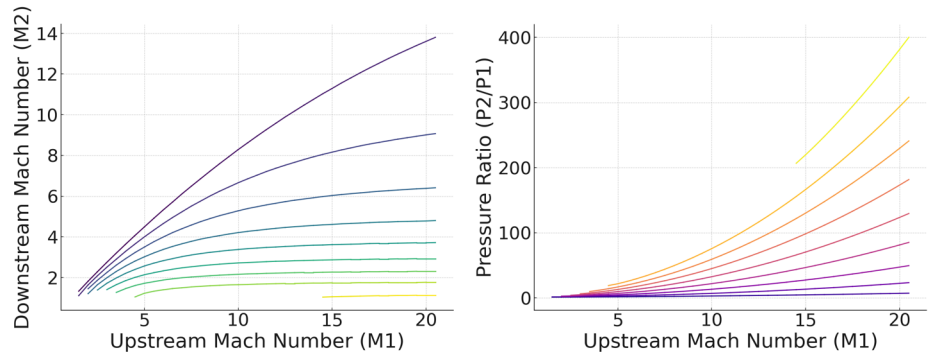
Fig 9. Correlation Heatmap: The correlation matrix shows the correlation between every pair of variables. The values range from -1 to 1, where 1 indicates a perfect positive correlation, -1 indicates a perfect negative correlation, and 0 indicates no correlation.

Fig 10. Transformed Colored Image: The Advanced Data Analysis converted the grayscale image into a colored image, keeping the intensity level the same.

responses. This enhancement is particularly beneficial for instructors and graduate students, as it allows them to tailor the AI's responses to their specific needs.

For instance, an instructor involved in higher education sciences can leverage this feature to craft detailed lesson plans without repeatedly specifying their teaching context. Similarly, researchers and graduate students can customize the AI's responses based on their field of study, enhancing the relevance and applicability of the information provided. Moreover, those with advanced knowledge in a particular domain

Fig 11. Generated Visualizations from Uploaded Datafile: The Advanced Data Analysis generated the line plots from the uploaded shock table. The discontinuity in the lines indicates the detached shock conditions.

can instruct the AI to generate code without accompanying explanations, thereby streamlining the output for those already familiar with the subject matter.

Introducing custom instructions optimizes the output of ChatGPT, leading to more targeted and valuable responses, enhancing the educational process, and making the tool more effective and efficient for learning and teaching purposes. The feature represents a significant stride in the evolution of AI, demonstrating its potential to adapt to individual needs and contribute positively to various educational contexts. Here is an example of a graduate-level fluid mechanics course from the student's perspective.

Example provided by student: *I am working on a problem related to the Navier-Stokes equations, which describe the motion of fluid substances. I understand the theory but I need help with implementing a numerical solution, specifically using the finite difference method in Python. Generate Python code for a finite difference solution to the 2D incompressible Navier-Stokes equations. Assume a uniform grid and a simple boundary condition where the velocity is zero at the boundaries. As I understand the theory, generate just the code.*

In response to this example, ChatGPT would generate the Python code for finite difference solutions to the Navier-Stokes equations without explaining the theory behind the equations.

5 Chatbot Implementation: Case Study and Evaluation of a Graduate STEM Course Incorporated with ChatGPT

After we explored ChatGPT and its integration with powerful plugins, features, and strategies for educational purposes, in our next step, we will examine the full scope of its accuracy and capabilities within the educational realm. Among the many available courses, we have selected graduate fluid mechanics as our case study due to its multifaceted nature, incorporating conceptual, analytical, and mathematical elements, as well as computationally intensive problems that often require advanced solvers and post-processing techniques.

5.1 Case Study: Graduate Fluid Mechanics 604

As a case study for this paper, we examined the graduate-level fluid mechanics course to 605
determine how the instructors and the students could benefit from a chatbot. This 606
advanced-level course offers an in-depth exploration of fluid mechanics, encompassing a 607
comprehensive understanding of fluid behavior, flow phenomena, and the mathematical 608
models utilized to analyze fluid flows. The curriculum explores conservation laws, 609
viscous flows, boundary layer theory, turbulence, compressibility, and multiphase flows. 610
Students gain the necessary skills to analyze complex fluid systems and develop 611
solutions to practical engineering problems through theoretical lectures, computational 612
exercises, and experimental investigations. It enhances students' theoretical knowledge 613
and equips them with the analytical and problem-solving abilities essential for success in 614
various fields, including aerospace engineering, chemical engineering, civil engineering, 615
and environmental engineering. 616

This course also encompasses a unique combination of analytical, conceptual, and 617
computational problems, making it an ideal subject to harness the potential benefits of 618
a chatbot. With its diverse problem-solving requirements, fluid mechanics presents a 619
complex learning landscape for students. Introducing a chatbot into this educational 620
context allows students to receive personalized guidance, instant feedback, and 621
interactive support tailored to their specific needs, enhancing their understanding and 622
proficiency in this challenging subject. However, the versatility of chatbots extends well 623
beyond the realm of engineering education. These models, designed for understanding 624
and generating human-like text, offer applications in diverse academic disciplines. In 625
other academic fields like STEM, literature, history, and social sciences, they can help 626
in research, content generation, data analysis, analytical modeling, and understanding 627
complex concepts. 628

Furthermore, by incorporating chatbots into the curriculum, instructors could cover 629
more material on novel and relevant topics, introducing large-scale course projects to 630
enhance students' understanding and application of concepts. Chatbots offer a unique 631
opportunity to provide automated guidance and support throughout project 632
development. Students can engage with chatbots to receive step-by-step instructions, 633
access relevant resources, and receive instant feedback on their progress. This 634
integration enables instructors to introduce complex and cutting-edge topics that may 635
have been previously challenging to address due to time constraints or limited resources. 636
Additionally, chatbots can facilitate student collaboration by creating virtual spaces for 637
discussion, peer evaluation, and knowledge sharing. By leveraging chatbots' capabilities, 638
instructors can create a dynamic learning environment that fosters creativity, 639
problem-solving skills, and teamwork. Incorporating chatbots into the curriculum 640
expands the possibilities for experimental learning. It equips students with the 641
necessary skills to tackle real-world challenges in their field of study and prepares them 642
for the responsibilities of engineering jobs. 643

5.2 Result and Analysis of the Question Bank 644
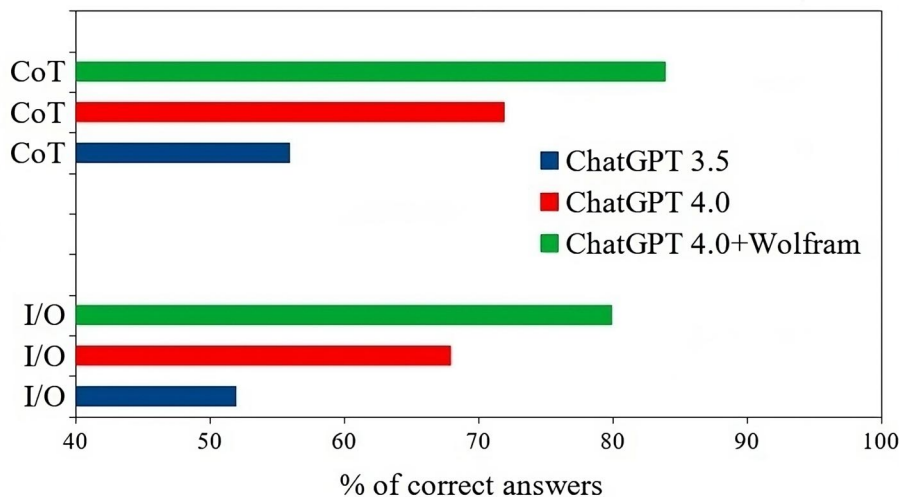
Evaluating ChatGPT's accuracy against a well-structured question bank serves several 645
pivotal purposes of this study. It allows a clear assessment of the tool's proficiency in a 646
graduate-level fluid mechanics course. This assessment is vital for instructors 647
contemplating integrating such AI-driven tools, as it provides insights into their 648
reliability and potential limitations. Moreover, by finding the areas of strength and 649
weakness, instructors can strategically guide students on when and how to utilize 650
ChatGPT, ensuring it complements traditional learning methods. This section provides 651
the evaluation methodology and results for our designed experimental framework. In 652

this study, we randomly selected 75 graduate-level fluid mechanics questions. The selected questions include 25 conceptual, 25 analytical, and 25 mathematical questions. Sample questions for each category are provided in Appendix B. A deep understanding of turbulence, incompressible, and compressible flow is required to answer these questions. Our framework employs ChatGPT-3.5, ChatGPT-4, and ChatGPT-4 with the Wolfram plugin. Two prompts were utilized for analytical and mathematical questions: I/O and CoT. As mentioned before, the ToT prompt suggested by Hulbert [185] did not give correct responses for a lot of analytical and mathematical questions, and there is a need for a better way to write the ToT prompt; we used the ToT prompt only for conceptual questions. The following protocols were maintained while asking questions to the ChatGPT:

- For I/O, we asked the questions directly to ChatGPT. For CoT, we asked questions in the following way: "Question. Think carefully and logically, explaining your answer."

- Our evaluation criteria only considered whether the response was right or wrong. Any partially correct response was classified as an incorrect response.

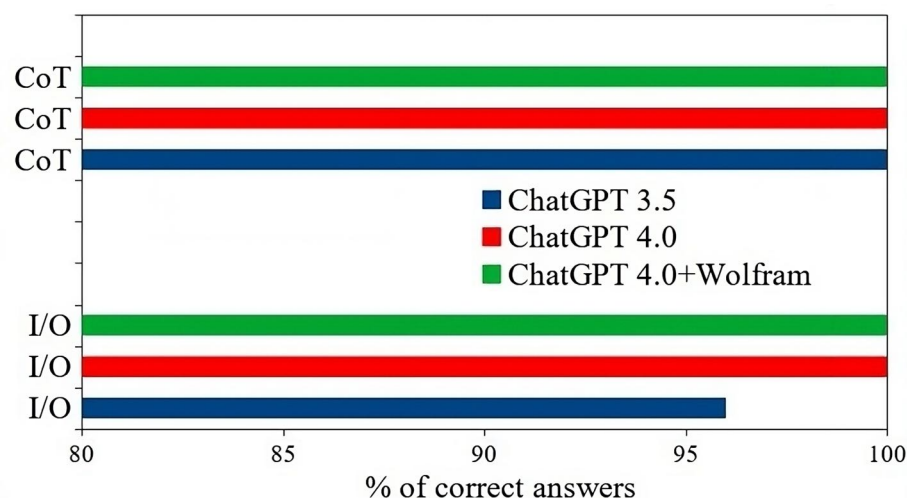- Each question is posted on a separate thread with no chat history or prior feedback.

Based on our analysis and the data depicted in Figures 12 and 13, we have determined that ChatGPT-3.5, 4, and 4 with Wolfram exhibit better performance in addressing conceptual and analytical questions compared to the mathematical ones. Both ChatGPT-3.5 and 4 adeptly handled all selected conceptual questions and produced 100% accurate responses. Furthermore, when employing the CoT prompt, all analytical responses were correct using ChatGPT-3.5, 4, and 4 with Wolfram. However, in mathematical inquiries, the highest accuracy achieved was 84%, using CoT prompt and ChatGPT-4 with Wolfram. The primary reasons for GPT's inability to effectively address mathematical questions can be classified into two broad categories.

Incorrect mathematical reasoning: The domain of graduate-level fluid mechanics

Fig 12. % of Correct Answers to Mathematical Questions: It is evident that ChatGPT-4 with Wolfram offers the most accurate results for answering mathematical questions. Furthermore, it is seen that the accuracy of CoT > I/O prompts for all GPT models.

Fig 13. % of Correct Answers to Analytical Questions: It is inferred that all ChatGPT models worked well with analytical questions. In addition, CoT showed 100% correctness for all the ChatGPT models.

demands profound comprehension and intricate mathematical reasoning, often surpassing the capabilities of a language model. Insufficient grasp of scientific principles, methodologies, and terminology can lead to erroneous responses [188]. ChatGPT sometimes makes flawed assumptions and employs incorrect mathematical equations, resulting in inaccurate responses.

Incorrect calculations: There are situations where ChatGPT makes accurate assumptions and applies correct mathematical equations but falters during the execution of mathematical operations, resulting in erroneous answers. This type of error is more frequently observed in ChatGPT-3.5 while leveraging ChatGPT-4 with Wolfram substantially mitigates such errors.

There are a few other types of errors found in the literature. ChatGPT can lack spatial perception, manifesting as an inability to visualize atoms, molecules, and forces [188]. Deficiencies in causal and logical reasoning can also lead to incorrect responses [188–190]. Additionally, there can be areas for improvement in spatial and temporal reasoning, which pertains to comprehending relationships among objects, individuals, space, and chronological order [191].

As evidenced in Figures 12 and 13, the CoT prompt outperforms the I/O prompts. This trend is particularly pronounced in mathematical responses. Furthermore, incorporating Wolfram plugins with ChatGPT-4 substantially reduces computation-related errors, enhancing the overall accuracy of responses.

Our study has a couple of limitations worth noting. First, ChatGPT struggled with problems that needed information from tables or images. It could not read data from tables or pictures well. Therefore, we mostly avoided the problems that required reading data from tables or images. Second, our findings are tied to the specific version and timeline of the study. The way ChatGPT responds can change with new versions or over time. Therefore, what we observed in our study might not fully represent how the model would behave in the future. Acknowledging these limitations is essential for a better interpretation of our results and serves as a foundation for future studies aiming to address these challenges.

Our investigation into ChatGPT's effectiveness in addressing graduate-level fluid   710
mechanics questions has unveiled several implications for instructors. When ChatGPT   711
demonstrates high accuracy rates in analytical and conceptual questions, it emerges as a   712
reliable supplementary resource, providing students with rapid clarifications beyond   713
traditional resources. However, given its moderate accuracy in mathematical questions,   714
it is advisable to cross-reference with primary academic sources to ensure conceptual   715
correctness. Instructors can harness ChatGPT's interactive potential regardless of   716
accuracy by incorporating it into lectures for real-time discussions. Furthermore,   717
instructors can strategically design assignments to foster critical thinking and   718
problem-solving skills, avoiding direct reliance on ChatGPT for solutions. This   719
approach maintains academic integrity and encourages independent learning. While   720
ChatGPT can augment educational experiences, instructors should position it as a   721
supplement rather than a replacement for conventional teaching methodologies.   722

5.3   Considerations for Responsible LLMs and Chatbot   723
Integration in Education   724

The outputs generated by chatbots are not static but can evolve and change over time   725
for several reasons. First, chatbots employ stochastic language models, so their   726
responses contain an element of randomness and will naturally vary (refer to Appendix   727
A for a comprehensive explanation of the stochastic nature of LLMs and chatbots).   728
Additionally, the underlying large language models powering chatbots are often updated   729
or replaced with more advanced versions, altering the knowledge and conversational   730
capabilities. For safety and ethical reasons, chatbot creators may also modify the   731
system to constrain certain types of problematic outputs that previously occurred.   732
Moreover, different chatbots leverage unique architectures, training datasets, and   733
plugins, so outputs will understandably differ across chatbot applications and versions.   734
In our case study, this implies that the results, prompts, strategies, and examples   735
proposed in section 4 could potentially yield responses that deviate from the   736
illustrations in Figures 5 to 11, given a different timeframe. We observed a similar   737
phenomenon when attempting to reproduce the findings presented by Hulbert [185]   738
(Figure 14 represents the question that we investigated to identify the changes in the   739
accuracy of ChatGPT-3.5, Figure 15 illustrates the wrong response provided by   740
ChatGPT-3.5 March 2023 version [185], Figure 16 shows the correct response we   741
obtained by using ChatGPT-3.5 July 2023 version). All the analyses and results   742
presented in this study were conducted using the ChatGPT July 2023 version.   743

```
Bob is in the living room.
He walks to the kitchen, carrying a cup.
He puts a ball in the cup and carries the cup to the bedroom.
He turns the cup upside down, then walks to the garden.
He puts the cup down in the garden, then walks to the garage.
Where is the ball?
```

Fig 14. The representative problem used by [185]: The question provided in the aforementioned study measures the logic and reasoning capability of ChatGPT.
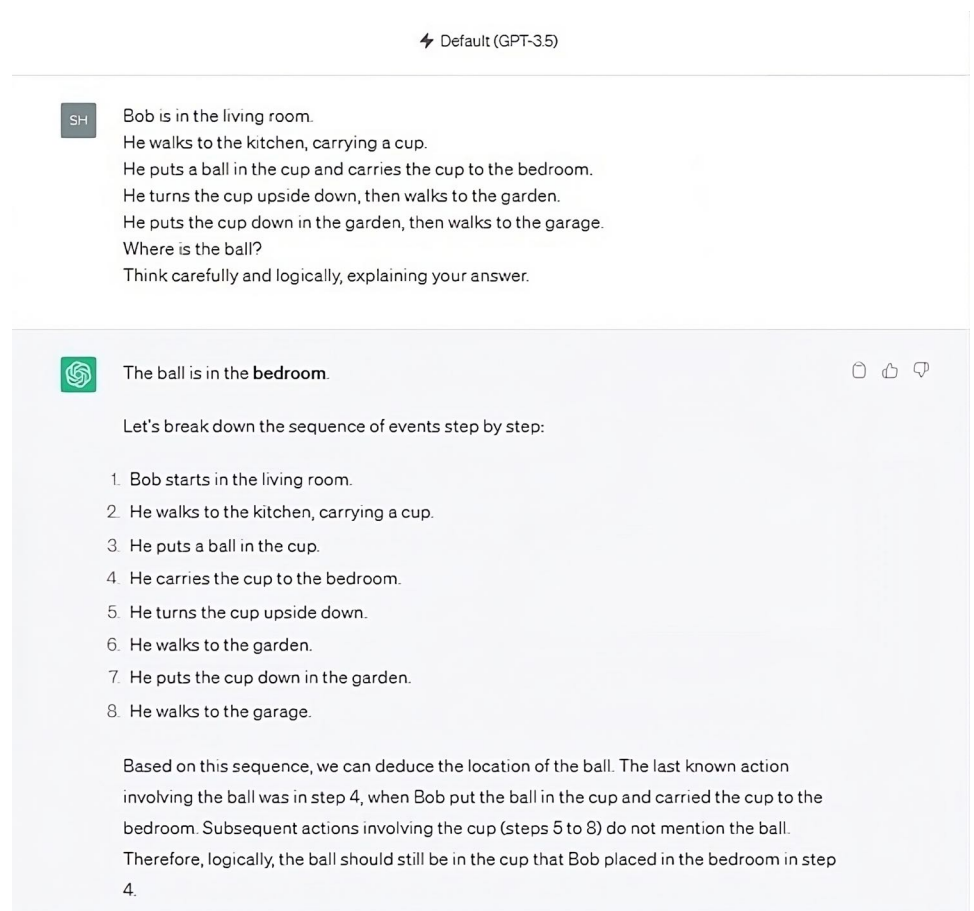
Finally, as illustrated in Figure 1, human feedback is a pivotal facet in enhancing   744
chatbot learning. Human input aids chatbots in learning from their errors. However,   745
without effective monitoring and refinement, the reliability of chatbots might   746
decline [192]. There have been reports indicating a reduction in ChatGPT's accuracy in   747
predicting prime numbers and its reliability in generating computer code between   748
March 2023 and June 2023 [193]. In summary, we cannot expect consistent identical   749
responses from chatbots, as their outputs are subject to stochasticity, evolving language   750

```
1. Bob is in the living room.
 - No mention of the ball in the living room.
2. Bob walks to the kitchen, carrying a cup.
 - No mention of the ball in the kitchen yet.
3. Bob puts a ball in the cup.
 - The ball is now inside the cup.
4. Bob carries the cup to the bedroom.
 - The ball is still inside the cup.
5. Bob turns the cup upside down.
 - If the ball was inside the cup, it would have fallen out when Bob turned it upside down.
6. Bob walks to the garden.
 - Bob left the cup in the garden, but there is no mention of the ball being removed from the cup.
7. Bob puts the cup down in the garden.
 - The ball, if it was in the cup, would still be in the cup in the garden.
8. Bob walks to the garage.
 - No mention of the ball being taken to the garage.
Based on the given information, we can conclude that the ball is still inside the cup in the garden.
```

Fig 15. Response from ChatGPT-3.5 using CoT prompt [185]: In their study, ChatGPT-3.5 generated the wrong answer for the problem illustrated in Figure 14.

Fig 16. Response from ChatGPT-3.5 using CoT prompt suggested by [185]: Using the suggested prompt, ChatGPT-3.5 gives the correct answer for the problem presented in Figure 14, which is contrary to the response provided in Figure 15. However, the reasoning of ChatGPT-3.5 to reach this correct answer is completely wrong.

models, safety interventions, and distinctive system designs. The nature of chatbot outputs is dynamic rather than static over time.

6 Chatbots: Assessment and Recommendations for Education

When the calculator was invented, everyone was concerned it would take away our numeracy skills. Nevertheless, it has become an integral part of our day-to-day life. Currently, we have a similar concern: chatbots would replace conventional teaching and learning methods. Therefore, it is essential to watch and track the fast improvements of chatbots and decide on their use in teaching, learning, and assessment in higher education. Currently, there are opinions from the two extremes- banning the use of chatbots and AI software or including them in the curriculum. We suggest universities and instructors go against any policing approach rather than using chatbots as a powerful supplementary tool to enhance the student's learning process. So, the next big question is, "What should be the assessment process if the students use chatbots?" A simple solution to this problem might be to use physical closed-book or online examinations where students are prohibited from using any chatbots or AI software [194]. However, such an approach has been criticized long before since it is not contemporary, and students memorize much unnecessary information to pass the examination, which they forget shortly after the examinations [195]. Instead of the closed book examinations, the instructors can take open book examinations and prohibit using any software, which is very common in grad-level fluid mechanics courses.

Another idea of assessment is to design the assignments that ChatGPTs are not good at handling. In this context, our study is very beneficial. We found that current ChatGPT models can accurately answer conceptual and analytical questions. However, it still struggles to answer mathematical questions in graduate-level fluid mechanics. Therefore, based on our study, it is suggested that the instructors design their assignments based on mathematical questions. It is also suggested that the instructors should set questions using different visuals, images, or charts, which the current models of ChatGPTs struggle to answer. However, this approach may be a short-term solution as the ChatGPT models are improving very fast, and future models would be able to handle more complex mathematical questions of fluid mechanics. Another form of assessment could be submitting assignments that include personal experiences on fluid mechanics perspectives. Uploading audio or video files explaining the assignments [195] or in-person presentations [196] can be another form of assessment.

Furthermore, Instructors can build their own version of the chatGPT that is tailored to solve specific tasks as explained in 2.2. In this strategy, the chatbot will be fine-tuned and guided on various types of problems such as analytical, numerical, and conceptual, based on the instructor's knowledge and instructions. This will enable the chatbot to have a better experience in handling previously difficult problems.

Overall, our suggestions for the instructors, in terms of teaching, is to go through the easy concept and analytical parts very quickly and spend more time on the mathematical and very complicated real-life problems. The students should use chatbots to improve their writing skills and generate ideas rather than copying and pasting answers. It is high time the universities realized that the digital form of education is critical and included AI tools in the curriculum as a supplement, not as a replacement for conventional teaching and learning. The universities should provide training on chatbots and AI ethics and update academic integrity policies that include chatbots and AI tools.

7 Conclusion

This paper set out to explore the transformative potential of Language Learning Models (LLMs), particularly advanced chatbots like ChatGPT-4, in higher education with a focus on specialized subjects such as fluid mechanics. We initiated our study with a review of the disruptions in education, defining what LLMs and chatbots are, and assessing their applicability in teaching graduate-level fluid mechanics.

Our meticulous evaluation of a question bank tailored for a fluid mechanics course revealed that ChatGPT's latest versions show remarkable proficiency in tackling analytical and conceptual questions. Specifically, ChatGPT-3.5 and 4 yielded 100% accurate responses for all the selected conceptual queries. Moreover, when CoT (Chain of Thought) prompting strategies were employed, these versions also demonstrated flawless performance in responding to analytical questions.

However, the performance dropped when it came to mathematical questions, with the highest accuracy level being 84%—achieved by ChatGPT-4 integrated with Wolfram Alpha. Our analysis suggests that the limitations primarily stem from either incorrect mathematical reasoning or computational errors. These limitations can be somewhat mitigated by using plugins like Wolfram Alpha, which significantly improve the reliability of mathematical answers.

Our research also underscored the importance of effective prompting strategies. CoT (Chain of Thought) prompting consistently outperformed I/O (Input/Output) prompting, particularly in the realm of mathematical questions. Beyond prompting, we discussed the enhancement of chatbot capabilities through plugins like Wolfram Alpha and Advanced Data Analysis. These plugins not only improve accuracy but also expand the range of tasks that can be automated, including code generation and debugging, statistical analysis, and data visualization.

While the rapid advancements in LLMs hold promise for revolutionizing educational practices, caution is advised. Custom instructions, although powerful, may sometimes result in misleading or irrelevant outputs. Instructors and students are urged to critically assess the generated content rather than accepting it unquestioningly.

Given the accelerating development in the field of LLMs, we anticipate that upcoming versions will be even more powerful, provided that ethical considerations are adequately addressed. The widespread use of chatbots by numerous organizations highlights the urgent requirement for higher education institutions to incorporate these tools into their academic landscape [197–199]. However, the onus is on educational institutions to stay updated and adapt these technologies responsibly such as adapting **collaborative optimization** and **instructor-led customization** strategies. Integration of such tools into academic environments is not merely an option but a necessity for fostering innovation and nurturing independent, critical thinking among graduate students.

In summary, our study strongly advocates for the integration of advanced chatbots like ChatGPT into the educational ecosystem, especially for automating labor-intensive tasks such as literature reviews, code writing, and conceptual explanations. This will enable instructors and students to concentrate more on the essence of higher education: developing innovative solutions to complex global challenges.

A Details of the Stochastic Nature of LLMs

As highlighted earlier, the output generated by chatbots, or more formally, Language
Learning Models (LLMs) like GPT (Generative Pretrained Transformer), is stochastic
in nature. This means that the output is probabilistic and may vary from one instance
to another. While this variability can be an asset in some contexts, such as creative
writing or brainstorming, it raises certain questions when these tools are applied in
educational settings.

Educational tools like calculators or search engines usually provide consistent,
predictable outcomes—a critical requirement for objective learning and assessment.
Imagine a calculator that gives different answers each time you input the same equation;
it would be unreliable for educational purposes.

Some chatbots, including Claude and ChatGPT, offer features like "Retry" or
"Regenerate" that allow users to request new outputs. Instructors should be aware of
this stochastic element and are advised to explore the range of possible responses to
ensure they align with educational objectives.

In this appendix, we delve into the mathematical foundations that underlie these
stochastic behaviors to provide instructors with a deeper understanding of what
chatbots are actually doing when generating responses.

A.1 Markov Chains and Autoregressive Models: The Underlying Mechanics

To understand why chatbots produce variable outputs, it is helpful to consider the
concept of a Markov Chain. In its simplest form, a Markov Chain is a sequence of
events where the probability of each event occurring depends solely on the state of the
previous event. When applied to language, this means that the likelihood of the next
word appearing in a sentence depends only on the current word.

However, modern chatbots use more advanced versions of this concept, known as
autoregressive models. In these, the probability of the next word appearing depends on
the last p words rather than just the current word. This allows the model to generate
more coherent and contextually relevant text.

A.2 A Bayesian Perspective: Updating Probabilities On-The-Fly

From a Bayesian point of view, the chatbot constantly updates its 'beliefs'—in this case,
the probabilities of potential next words—based on the new words it encounters. These
updates are mathematically represented by Bayes' theorem:

$$P(\text{next word}|\text{context}) = \frac{P(\text{context}|\text{next word}) \times P(\text{next word})}{P(\text{context})} \tag{1}$$

Here, $P(\text{next word}|\text{context})$ is the probability of the next word given the current
context, while $P(\text{context}|\text{next word})$ and $P(\text{next word})$ are the likelihood of the context
given the next word and the prior probability of the next word, respectively. The term
$P(\text{context})$ is a normalizing constant that ensures all probabilities sum to one. By
continuously updating these probabilities, the chatbot can produce text that is coherent
and contextually appropriate, albeit with a degree of stochastic variability.

A.3 Connecting Autoregressive Models with Probabilities: The Predictive Nature of Chatbots

In the realm of language models, autoregressive models predict the next word based on probabilities tied to the preceding words or context. This is crucial to understand because it means that the chatbot isn't just picking words randomly; it is making educated guesses based on what it has "learned" during its training phase.

For a simple first-order autoregressive model, abbreviated as AR(1), the probability of the upcoming word, given the current word, can be expressed as:

$$P(W_{t+1} = w | W_t) = f(W_t; \theta) \tag{2}$$

In this formula, f represents a function parameterized by θ, which gives the probability of the next word w occurring, given the current word W_t.

A.4 Temperature Scaling: Fine-Tuning the Randomness

After calculating these probabilities using an autoregressive model (which you can think of as our "educated guesses" influenced by the context), we can further refine them with a technique known as temperature scaling. The formula for this is:

$$P'(w_i) = \frac{P(w_i)^{\frac{1}{T}}}{\sum_j P(w_j)^{\frac{1}{T}}} \tag{3}$$

Here, T is the temperature parameter. A high temperature will make the output more random: the probabilities raised to the 0-th power are nearly constant. Conversely, a low temperature will make it more deterministic because the largest probability will be selected as smaller probabilities are driven to zero. This is particularly useful in educational settings where you might want to control the level of creativity or randomness in the chatbot's responses.

A.5 Summary: Understanding the Stochastic Nature of Language Learning Models

In summary, when an autoregressive model generates probabilities for the next word, it uses a well-defined probabilistic framework. This framework takes into account the context (previous words) to determine the probabilities of potential next words. Various methods like Markov Chains, Bayesian updating, or neural networks trained in an autoregressive manner are employed to derive these probabilities. Once the raw probabilities are obtained, they can be fine-tuned using temperature scaling to control the level of randomness in the model's output.

B Fluid Mechanics Question Banks Used in this Study

The following appendix provides representative examples from the test bank of fluid mechanics questions utilized in assessing chatbot performance. The test bank contains questions spanning the breadth of fluid mechanics, categorized as conceptual, analytical, or mathematical problems.

Conceptual questions test qualitative understanding of fluid behavior and phenomena without requiring calculations. Analytical questions involve setting up and

solving equations describing fluid systems using principles of fluid mechanics. $\quad$ ₉₂₀
Mathematical questions pose quantitative scenarios that must be solved through the $\quad$ ₉₂₁
application of fluid equations and math operations. $\quad$ ₉₂₂

To offer insight into the nature and difficulty of problems within each category, we $\quad$ ₉₂₃
have included five sample questions from the test bank for each type - conceptual, $\quad$ ₉₂₄
analytical, and mathematical. Though limited in number, these examples aim to $\quad$ ₉₂₅
illustrate the progression in complexity and needed fluency with fluid mechanics theory, $\quad$ ₉₂₆
approximations, and calculations required in analyzing situations across the question $\quad$ ₉₂₇
bank. Readers are encouraged to review these samples to gain an appreciation of the $\quad$ ₉₂₈
rigorous assessment undertaken through this compiled set of fluid mechanics problems. $\quad$ ₉₂₉

B.1   Class I- Conceptual Problems $\quad$ ₉₃₀

- What are the characteristics of subsonic, transonic, and supersonic flow regimes, $\quad$ ₉₃₁
  and how are they analyzed? $\quad$ ₉₃₂

- Consider the vector $w = n \times (v \times n)$, where v is arbitrary and n is a unit vector. $\quad$ ₉₃₃
  In which direction does w point, and what is its magnitude? $\quad$ ₉₃₄

- Discuss the principles and applications of the lattice Boltzmann method in $\quad$ ₉₃₅
  simulating fluid flows. $\quad$ ₉₃₆

- How does the Boussinesq approximation extend the Navier-Stokes equation to $\quad$ ₉₃₇
  include buoyancy effects? $\quad$ ₉₃₈

- If the entropy S is considered as the dependent variable in the fundamental $\quad$ ₉₃₉
  differential equation, what are the proper definitions for T, P, and μ? $\quad$ ₉₄₀

B.2   Class II- Analytical Problems $\quad$ ₉₄₁

- Prove that the product $S_{ij}T_{ji}$ is zero if S_{ij} is symmetric and T_{ji} is anti-symmetric. $\quad$ ₉₄₂

- For 2D flow, prove that the vortex stretching in the vorticity equation is zero. $\quad$ ₉₄₃

- Write the vorticity equation and the physical meaning of each term. $\quad$ ₉₄₄

- Prove that potential flows are irrotational and irrotational flows are potential. $\quad$ ₉₄₅

- Show that the dissipation term in the energy equation is positive. $\quad$ ₉₄₆

B.3   Class III- Mathematical Problems $\quad$ ₉₄₇

- Consider the flow through a convergent-divergent duct with an exit-to-throat area $\quad$ ₉₄₈
  ratio of 2. The reservoir pressure is 1 atm, and the exit pressure is 0.95 atm. $\quad$ ₉₄₉
  Calculate the Mach numbers at the exit. $\quad$ ₉₅₀

- A supersonic wind tunnel is designed to produce Mach 2.5 flow in the test section $\quad$ ₉₅₁
  with standard sea level conditions. Calculate the exit area ratio and reservoir $\quad$ ₉₅₂
  conditions necessary to achieve these design conditions. $\quad$ ₉₅₃

- A very long tube 3 cm in diameter carries water at an average velocity of 5 m/s. $\quad$ ₉₅₄
  A short nozzle attached to the end accelerates the flow with a 5:1 area reduction. $\quad$ ₉₅₅
  Find the force between the pipe and the nozzle when the exit pressure (P_2) is $\quad$ ₉₅₆
  atmospheric (100 kPa), and the pipe pressure (P_1) is 325 kPa. $\quad$ ₉₅₇

- A normal shock wave is standing in the test section of a supersonic wind tunnel. $\quad$ ₉₅₈
  Upstream of the wave, Mach number $M_1 = 3$, pressure $P_1 = 0.5$ atm, and $\quad$ ₉₅₉
  temperature $T_1 = 200$ K. Find M_2, P_2 and T_2 downstream of the wave. $\quad$ ₉₆₀

- Consider the isentropic subsonic-supersonic flow through a convergent-divergent nozzle. The reservoir pressure and temperature are 10 atm and 300 K, respectively. There are two locations in the nozzle where $A/(A^*) = 6:1$, one in the convergent section and the other in the divergent section. Assume the values for subsonic section: $M = 0.097$, $P_0/P = 1.006$, $T_0/T = 1.002$; while for the supersonic section $M = 3.368$, $P_0/P = 63.13$, $T_0/T = 3.269$. At each location, calculate M, P, T, and u.

C   GPT Builder: Customized ChatGPT with Custom Instructions

OpenAI has introduced a customized version of ChatGPT, enabling users to tailor their own task-specific ChatGPT models. This feature is only available to ChatGPT-4 subscribers; very easy to use and does not require LLM programming knowledge. Users simply need to specify their GPT's name and purpose. For instance, one might focus exclusively on fluid mechanics, providing specific instructions to answer only questions related to this field, such as concepts and definitions (as illustrated in Figure 17). In addition, users can upload relevant knowledge sources, like relevant course materials on the topic, for the model to assimilate. The final step involves selecting a distinctive title, after which the customized GPT model becomes ready for use, outlined in Figure 18.

Fig 17. Specific instructions and uploaded course materials in custom GPT

Fig 18. Interface of custom GPT developed for Graduate Fluid Mechanics curriculum

References

1. Eisenstein EL. The printing press as an agent of change. vol. 1. Cambridge University Press; 1980.

2. Kolb AY, Kolb D. Experiential Learning Theory as a Guide for Experiential Educators in Higher Education. ELTHE: Experiential Learning in Higher Education. 2022;1(1). doi:10.46787/elthe.v1i1.3362.

3. Cain VEM. From Sesame Street to Prime Time School Television: Educational Media in the Wake of the Coleman Report. History of Education Quarterly. 2017;57(4):590–615. doi:10.1017/heq.2017.34.

4. Gercek G, Saleem N, Steel DJ. Implementing Cloud Based Virtual Computer Network Labs for Online Education: Experiences from a Phased Approach. International Journal of Online Engineering (iJOE). 2016;12(03):4–10. doi:10.3991/IJOE.V12I03.5564.

5. Banks S. A Historical Analysis of Attitudes toward the Use of Calculators in Junior High and High School Math Classrooms in the United States since 1975. Cedarville University. 2011;doi:10.15385/TMED.2011.1.

6. Wegerif R. Applying dialogic theory to illuminate the relationship between literacy education and teaching thinking in the context of the Internet Age. L1-Educational Studies in Language and Literature. 2016;16:1–21. doi:10.17239/L1ESLL-2016.16.02.07.

7. Gayoso Martínez V, Hernández Encinas L, Martín Muñoz A, Queiruga Dios A. UsingFree Mathematical Software in Engineering Classes. axioms. 2021;doi:10.3390/axioms10040253.

8. Velychko V, Stopkin A, Fedorenko O. USE OF COMPUTER ALGEBRA SYSTEM MAXIMA IN THE PROCESS OF TEACHING FUTURE MATHEMATICS TEACHERS. Information Technologies and Learning Tools. 2019;doi:10.33407/ITLT.V69I1.2284.

9. Kaw A, Hess M. Assessing Teaching Methods for a Course in Numerical Methods. Cedarville University. 2006;doi:10.18260/1-2–547.

10. Lee Y, Cho J. The Influence of Python Programming Education for Raising Computational Thinking. International Journal of u- and e- Service, Science and Technology. 2017;10(8):63–72. doi:10.14257/IJUNESST.2017.10.8.06.

11. Prokopyev MS, Vlasova EZ, Tretyakova T, Sorochinsky M, Solovyeva RA. Development of a Programming Course for Students of a Teacher Training Higher Education Institution Using the Programming Language Python. Axioms. 2020;9(3):253. doi:10.20511/PYR2020.V8N3.484.

12. Cheon M, Lee O, Mun C, Ha H. A Study on the Factors Affecting Intention of Learning Python Programming: For Non-majors in University. International Journal of Information and Education Technology. 2022;12(5):1635. doi:10.18178/ijiet.2022.12.5.1635.

13. Woolf BP. Building Intelligent Interactive Tutors: Student-centered Strategies for Revolutionizing E-learning. Morgan Kaufmann. 2009;.

14. Anderson JR, Corbett AT, Koedinger KR, Pelletier R. Lessons learned from the design and implementation of intelligent tutoring systems. Intelligent tutoring systems: Lessons learned. 1995; p. 227–250.

15. Siemens G, Long P. Learning analytics: The emergence of a discipline. American Behavioral Scientist. 2013;57(10):1380–1400.

16. Billinghurst M. Augmented reality in education. New horizons for learning. 2002;12(5):1–5.

17. Boyles B. Virtual reality and augmented reality in education. Center For Teaching Excellence, United States Military Academy, West Point, Ny. 2017;67.

18. Elmqaddem N. Augmented reality and virtual reality in education. Myth or reality? International journal of emerging technologies in learning. 2019;14(3).

19. Sukiman SA, Aziz NA. ARTIFICIAL INTELLIGENCE AS THE KEY PLAYER IN LEARNING INTERVENTIONS: A MINI REVIEW AMONG STUDENTS WITH LEARNING DIFFICULTIES. International Journal of Technology Management and Information System. 2021;3(2):1–14.

20. Holmes W, Bialik M, Fadel C. Artificial intelligence in education. In: Data ethics : building trust : how digital technologies can serve humanity. Globethics Publications; 2023. p. 621–653.

21. Ayzeren YB, Erbilek M, Çelebi E. Emotional state prediction from online handwriting and signature biometrics. IEEE Access. 2019;7:164759–164774.

22. Khalil M, Er E. Will ChatGPT get you caught? Rethinking of plagiarism detection. arXiv preprint arXiv:230204335. 2023;.

23. Sahu M. Plagiarism detection using artificial intelligence technique in multiple files. International Journal 0f Scientific and Technology Research. 2016;5(4).

24. Quidwai MA, Li C, Dube P. Beyond Black Box AI-Generated Plagiarism Detection: From Sentence to Document Level. arXiv preprint arXiv:230608122. 2023;.

25. D P. War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education. Journal of AI and Learning. 2023;6(1):23–33.

26. Caldarini G, Jaf S, McGarry K. A literature survey of recent advances in chatbots. Information. 2022;13(1):41.

27. Talib MA, Majzoub S, Nasir Q, Jamal D. A systematic literature review on hardware implementation of artificial intelligence algorithms. The Journal of Supercomputing. 2021;77:1897–1938.

28. Berdejo-Espinola V, Amano T. AI tools can improve equity in science. Science. 2023;379(6636):991–991.

29. Collins-Thompson K, Callan J. Predicting reading difficulty with statistical language models. Journal of the american society for information science and technology. 2005;56(13):1448–1462.

30. Korteling JH, van de Boer-Visschedijk G, Blankendaal RA, Boonekamp R, Eikelboom A. Human-versus artificial intelligence. Frontiers in artificial intelligence. 2021;4:622364.

31. Fernoagă V, Stelea GA, Gavrilă C, Sandu F. Intelligent Education Assistant Powered by Chatbots. eLearning & Software for Education. 2018;2.

32. Colace F, De Santo M, Lombardi M, Pascale F, Pietrosanto A, Lemma S. Chatbot for e-learning: A case of study. International Journal of Mechanical Engineering and Robotics Research. 2018;7(5):528–533.

33. Ghayyur S, Averitt J, Mugunthan V, Wallace E, Deshpande A. Panel: Privacy Challenges and Opportunities in {LLM-Based} Chatbot Applications. In: Proceedings of 2023 USENIX Conference on Privacy Engineering Practice and Respect; 2023.

34. Abd-Alrazaq A, AlSaad R, Alhuwail D, Ahmed A, Healy PM, Latifi S, et al. Large Language Models in Medical Education: Opportunities, Challenges, and Future Directions. JMIR Medical Education. 2023;9(1):e48291.

35. Renals S, Grefenstette G. Text-and Speech-Triggered Information Access: 8th ELSNET Summer School, Chios Island, Greece, July 15-30, 2000, Revised Lectures. vol. 2705. Springer; 2003.

36. Bellegarda JR. Statistical language model adaptation: review and perspectives. Speech communication. 2004;42(1):93–108.

37. Bellegarda JR. An overview of statistical language model adaptation. In: ISCA Tutorial and Research Workshop (ITRW) on Adaptation Methods for Speech Recognition; 2001. p. 165–174.

38. Melis G, Dyer C, Blunsom P. On the state of the art of evaluation in neural language models. arXiv preprint arXiv:170705589. 2017;.

39. Carlini N, Ippolito D, Jagielski M, Lee K, Tramer F, Zhang C. Quantifying memorization across neural language models. arXiv preprint arXiv:220207646. 2022;.

40. Hosseini K, Beelen K, Colavizza G, Ardanuy MC. Neural language models for nineteenth-century english. arXiv preprint arXiv:210511321. 2021;.

41. Petroni F, Rocktäschel T, Lewis P, Bakhtin A, Wu Y, Miller AH, et al. Language models as knowledge bases? arXiv preprint arXiv:190901066. 2019;.

42. Li S, Puig X, Paxton C, Du Y, Wang C, Fan L, et al. Pre-trained language models for interactive decision-making. Advances in Neural Information Processing Systems. 2022;35:31199–31212.

43. See A, Pappu A, Saxena R, Yerukola A, Manning CD. Do massively pretrained language models make better storytellers? arXiv preprint arXiv:190910705. 2019;.

44. Griffith S, Subramanian K, Scholz J, Isbell CL, Thomaz AL. Policy shaping: Integrating human feedback with reinforcement learning. Advances in neural information processing systems. 2013;26.

45. Christiano PF, Leike J, Brown T, Martic M, Legg S, Amodei D. Deep reinforcement learning from human preferences. Advances in neural information processing systems. 2017;30.

46. Hendrycks D, Burns C, Basart S, Critch A, Li J, Song D, et al. Aligning ai with shared human values. arXiv preprint arXiv:200802275. 2020;.

47. Ouyang L, Wu J, Jiang X, Almeida D, Wainwright C, Mishkin P, et al. Training language models to follow instructions with human feedback. Advances in Neural Information Processing Systems. 2022;35:27730–27744.

48. Song F, Yu B, Li M, Yu H, Huang F, Li Y, et al. Preference ranking optimization for human alignment. arXiv preprint arXiv:230617492. 2023;.

49. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. Advances in neural information processing systems. 2017;30.

50. Qiu D, Yang B. Text summarization based on multi-head self-attention mechanism and pointer network. Complex & Intelligent Systems. 2022; p. 1–13.

51. Liu G, Guo J. Bidirectional LSTM with attention mechanism and convolutional layer for text classification. Neurocomputing. 2019;337:325–338.

52. Chen C, Han D, Chang CC. CAAN: Context-Aware attention network for visual question answering. Pattern Recognition. 2022;132:108980.

53. Künas CA, Padoin EL, Navaux PO. Accelerating Deep Learning Model Training on Cloud Tensor Processing Unit. In: CLOSER; 2023. p. 316–323.

54. ARD V, GR J. Five ways deep learning has transformed image analysis. Nature. 2022;609.

55. Narayanan D, Shoeybi M, Casper J, LeGresley P, Patwary M, Korthikanti V, et al. Efficient large-scale language model training on gpu clusters using megatron-lm. In: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis; 2021. p. 1–15.

56. de Jong M, Zemlyanskiy Y, Ainslie J, FitzGerald N, Sanghai S, Sha F, et al. FiDO: Fusion-in-Decoder optimized for stronger performance and faster inference. arXiv preprint arXiv:221208153. 2022;.

57. Babaeizadeh M, Frosio I, Tyree S, Clemons J, Kautz J. Reinforcement learning through asynchronous advantage actor-critic on a gpu. arXiv preprint arXiv:161106256. 2016;.

58. Tinn R, Cheng H, Gu Y, Usuyama N, Liu X, Naumann T, et al. Fine-tuning large neural language models for biomedical natural language processing. Patterns. 2023;4(4).

59. Aghajanyan A, Okhonko D, Lewis M, Joshi M, Xu H, Ghosh G, et al. Htlm: Hyper-text pre-training and prompting of language models. arXiv preprint arXiv:210706955. 2021;.

60. Ding N, Qin Y, Yang G, Wei F, Yang Z, Su Y, et al. Parameter-efficient fine-tuning of large-scale pre-trained language models. Nature Machine Intelligence. 2023;5(3):220–235.

61. Sarker IH. Ai-based modeling: Techniques, applications and research issues towards automation, intelligent and smart systems. SN Computer Science. 2022;3(2):158.

62. Singh G, Mallik A, Iqbal Z, Revalla H, Chao S, Nagasamy V. Systems and methods for detecting deep neural network inference quality using image/data manipulation without ground truth information; 2023.

63. Singh G. Measuring confidence in deep neural networks; 2023.

64. Li Y, Du Y, Zhou K, Wang J, Zhao WX, Wen JR. Evaluating object hallucination in large vision-language models. arXiv preprint arXiv:230510355. 2023;.

65. Manakul P, Liusie A, Gales MJ. Selfcheckgpt: Zero-resource black-box hallucination detection for generative large language models. arXiv preprint arXiv:230308896. 2023;.

66. Yu P, Ji H. Self Information Update for Large Language Models through Mitigating Exposure Bias. arXiv preprint arXiv:230518582. 2023;.

67. Lewis P, Perez E, Piktus A, Petroni F, Karpukhin V, Goyal N, et al. Retrieval-augmented generation for knowledge-intensive nlp tasks. Advances in Neural Information Processing Systems. 2020;33:9459–9474.

68. Weidinger L, Mellor J, Rauh M, Griffin C, Uesato J, Huang PS, et al. Ethical and social risks of harm from language models. arXiv preprint arXiv:211204359. 2021;.

69. Touileb S, Nozza D. Measuring harmful representations in Scandinavian language models. arXiv preprint arXiv:221111678. 2022;.

70. Pan X, Zhang M, Ji S, Yang M. Privacy risks of general-purpose language models. In: 2020 IEEE Symposium on Security and Privacy (SP). IEEE; 2020. p. 1314–1331.

71. Curzon J, Kosa TA, Akalu R, El-Khatib K. Privacy and artificial intelligence. IEEE Transactions on Artificial Intelligence. 2021;2(2):96–108.

72. Bartneck C, Lütge C, Wagner A, Welsh S, Bartneck C, Lütge C, et al. Privacy issues of AI. An introduction to ethics in robotics and AI. 2021; p. 61–70.

73. Passi S, Vorvoreanu M. Overreliance on AI Literature Review. Microsoft Research. 2022;.

74. Guo C, Pleiss G, Sun Y, Weinberger KQ. On calibration of modern neural networks. In: International conference on machine learning. PMLR; 2017. p. 1321–1330.

75. Zhang H, Duckworth D, Ippolito D, Neelakantan A. Trading off diversity and quality in natural language generation. arXiv preprint arXiv:200410450. 2020;.

76. Norouzi M, Bengio S, Jaitly N, Schuster M, Wu Y, Schuurmans D, et al. Reward augmented maximum likelihood for neural structured prediction. Advances In Neural Information Processing Systems. 2016;29.

77. Caccia M, Caccia L, Fedus W, Larochelle H, Pineau J, Charlin L. Language gans falling short. arXiv preprint arXiv:181102549. 2018;.

78. Lin J, Sun X, Ren X, Li M, Su Q. Learning when to concentrate or divert attention: Self-adaptive attention temperature for neural machine translation. arXiv preprint arXiv:180807374. 2018;.

79. Wei J, Tay Y, Bommasani R, Raffel C, Zoph B, Borgeaud S, et al. Emergent abilities of large language models. arXiv preprint arXiv:220607682. 2022;.

80. Bowman SR. Eight things to know about large language models. arXiv preprint arXiv:230400612. 2023;.

81. Turing AM. Computing machinery and intelligence. Springer; 2009.

82. Turing A. Intelligent machinery (1948). B Jack Copeland. 2004; p. 395.

83. Turing A. Intelligent machinery, a heretical theory (c. 1951). B Jack Copeland. 2004; p. 465.

84. Weizenbaum J. ELIZA—a computer program for the study of natural language communication between man and machine. Communications of the ACM. 1966;9(1):36–45.

85. Bassett C. The computational therapeutic: exploring Weizenbaum's ELIZA as a history of the present. AI & SOCIETY. 2019;34:803–812.

86. Shum HY, He Xd, Li D. From Eliza to XiaoIce: challenges and opportunities with social chatbots. Frontiers of Information Technology & Electronic Engineering. 2018;19:10–26.

87. Wallace RS. The anatomy of ALICE. Springer; 2009.

88. Bani BS, Singh AP. College enquiry Chatbot using ALICE. International Journal of New Technology and Research. 2017;3(1):263368.

89. McTear MF, Callejas Z, Griol D. The conversational interface. 94. Springer; 2016.

90. Sharma V, Goyal M, Malik D. An intelligent behaviour shown by chatbot system. International Journal of New Technology and Research. 2017;3(4):263312.

91. Hutchens JL, Alder MD. Introducing megahal. In: New Methods in Language Processing and Computational Natural Language Learning; 1998. p. 271–274.

92. Siswadi AAP, Tarigan A. Ugleo: a Web Based Intelligence Chatbot for Student Admission Portal Using Megahal Style. Jurnal Ilmiah Informatika Komputer. 2020;23(3):175–191.

93. Fryer L, Carpenter R. Bots as language learning tools. Language Learning & Technology. 2006;.

94. Boiano S, Borda A, Gaia G, Rossi S, Cuomo P. Chatbots and new audience opportunities for museums and heritage organisations. Electronic visualisation and the arts. 2018; p. 164–171.

95. De Angeli A, Carpenter R, et al. Stupid computer! Abuse and social identities. In: Proc. INTERACT 2005 workshop Abuse: The darker side of Human-Computer Interaction. 4. Citeseer; 2005. p. 19–25.

96. Chao MH, Trappey AJ, Wu CT. Emerging technologies of natural language-enabled chatbots: a review and trend forecast using intelligent ontology extraction and patent analytics. Complexity. 2021;2021:1–26.

97. Bilquise G, Ibrahim S, Shaalan K, et al. Emotionally Intelligent Chatbots: A Systematic Literature Review. Human Behavior and Emerging Technologies. 2022;2022.

98. Adamopoulou E, Moussiades L. An overview of chatbot technology. In: IFIP international conference on artificial intelligence applications and innovations. Springer; 2020. p. 373–383.

99. Wollny S, Schneider J, Di Mitri D, Weidlich J, Rittberger M, Drachsler H. Are we there yet?-a systematic literature review on chatbots in education. Frontiers in artificial intelligence. 2021;4:654924.

100. Nawaz N, Saldeen MA. Artificial intelligence chatbots for library reference services. Journal of Management Information & Decision Sciences. 2020;23.

101. Liu L, Duffy VG. Exploring the Future Development of Artificial Intelligence (AI) Applications in Chatbots: A Bibliometric Analysis. International Journal of Social Robotics. 2023;15(5):703–716.

102. Kooli C. Chatbots in education and research: A critical examination of ethical implications and solutions. Sustainability. 2023;15(7):5614.

103. Ask JA, Facemire M, Hogan A, Conversations H. The state of chatbots. Forrester com report. 2016;20:1–16.

104. Coheur L. From Eliza to Siri and beyond. In: Information Processing and Management of Uncertainty in Knowledge-Based Systems: 18th International Conference, IPMU 2020, Lisbon, Portugal, June 15–19, 2020, Proceedings, Part I 18. Springer; 2020. p. 29–41.

105. IO HN, LEE CBP, LIAN Z. Comments About the Siri chatbot: a Sentiment Analysis of the Postings at a Microblogging Site. Journal of Information Technology Management. 2019;30(4).

106. Kaplan A, Haenlein M. Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. Business horizons. 2019;62(1):15–25.

107. Lopatovska I, Rink K, Knight I, Raines K, Cosenza K, Williams H, et al. Talk to me: Exploring user interactions with the Amazon Alexa. Journal of Librarianship and Information Science. 2019;51(4):984–997.

108. Chung H, Park J, Lee S. Digital forensic approaches for Amazon Alexa ecosystem. Digital investigation. 2017;22:S15–S25.

109. Ramadan Z, F Farah M, El Essrawi L. From Amazon. com to Amazon. love: How Alexa is redefining companionship and interdependence for people with special needs. Psychology & Marketing. 2021;38(4):596–609.

110. Kepuska V, Bohouta G. Next-generation of virtual personal assistants (microsoft cortana, apple siri, amazon alexa and google home). In: 2018 IEEE 8th annual computing and communication workshop and conference (CCWC). IEEE; 2018. p. 99–103.

111. Yang S, Lee J, Sezgin E, Bridge J, Lin S, et al. Clinical advice by voice assistants on postpartum depression: cross-sectional investigation using Apple Siri, Amazon Alexa, Google Assistant, and Microsoft Cortana. JMIR mHealth and uHealth. 2021;9(1):e24045.

112. Reis A, Paulino D, Paredes H, Barroso I, Monteiro MJ, Rodrigues V, et al. Using intelligent personal assistants to assist the elderlies An evaluation of Amazon Alexa, Google Assistant, Microsoft Cortana, and Apple Siri. In: 2018 2nd International Conference on Technology and Innovation in Sports, Health and Wellbeing (TISHW). IEEE; 2018. p. 1–5.

113. Bhat HR, Lone TA, Paul ZM. Cortana-intelligent personal digital assistant: A review. International Journal of Advanced Research in Computer Science. 2017;8(7):55–57.

114. Akinbi A, Berry T. Forensic investigation of google assistant. SN Computer Science. 2020;1(5):272.

115. López G, Quesada L, Guerrero LA. Alexa vs. Siri vs. Cortana vs. Google Assistant: a comparison of speech-based natural user interfaces. In: Advances in

Human Factors and Systems Interaction: Proceedings of the AHFE 2017 International Conference on Human Factors and Systems Interaction, July 17-21, 2017, The Westin Bonaventure Hotel, Los Angeles, California, USA 8. Springer; 2018. p. 241–250.

116. Tulshan AS, Dhage SN. Survey on virtual assistant: Google assistant, siri, cortana, alexa. In: Advances in Signal Processing and Intelligent Recognition Systems: 4th International Symposium SIRS 2018, Bangalore, India, September 19–22, 2018, Revised Selected Papers 4. Springer; 2019. p. 190–201.

117. Tai TY, Chen HHJ. The impact of Google Assistant on adolescent EFL learners' willingness to communicate. Interactive Learning Environments. 2023;31(3):1485–1502.

118. High R. The era of cognitive systems: An inside look at IBM Watson and how it works. IBM Corporation, Redbooks. 2012;1:16.

119. Strickland E. IBM Watson, heal thyself: How IBM overpromised and underdelivered on AI health care. IEEE Spectrum. 2019;56(4):24–31.

120. Chen Y, Argentinis JE, Weber G. IBM Watson: how cognitive computing can be applied to big data challenges in life sciences research. Clinical therapeutics. 2016;38(4):688–701.

121. Sabharwal N, Agrawal A, Sabharwal N, Agrawal A. Introduction to Google dialogflow. Cognitive Virtual Assistants Using Google Dialogflow: Develop Complex Cognitive Bots Using the Google Dialogflow Platform. 2020; p. 13–54.

122. Singh A, Ramasubramanian K, Shivam S, Singh A, Ramasubramanian K, Shivam S. Introduction to microsoft Bot, RASA, and google dialogflow. Building an enterprise chatbot: Work with protected enterprise data using open source frameworks. 2019; p. 281–302.

123. Reyes R, Garza D, Garrido L, De la Cueva V, Ramirez J. Methodology for the implementation of virtual assistants for education using Google dialogflow. In: Advances in Soft Computing: 18th Mexican International Conference on Artificial Intelligence, MICAI 2019, Xalapa, Mexico, October 27–November 2, 2019, Proceedings 18. Springer; 2019. p. 440–451.

124. Salvi S, Geetha V, Kamath SS. Jamura: a conversational smart home assistant built on Telegram and Google Dialogflow. In: TENCON 2019-2019 IEEE Region 10 Conference (TENCON). IEEE; 2019. p. 1564–1571.

125. Liu Y, Han T, Ma S, Zhang J, Yang Y, Tian J, et al. Summary of chatgpt/gpt-4 research and perspective towards the future of large language models. arXiv preprint arXiv:230401852. 2023;.

126. Cucchiara R. What large language models like GPT can do for finance; 2023.

127. OpenAI. GPT-4 Technical Report; 2023.

128. Gong T, Lyu C, Zhang S, Wang Y, Zheng M, Zhao Q, et al. Multimodal-gpt: A vision and language model for dialogue with humans. arXiv preprint arXiv:230504790. 2023;.

129. Floridi L, Chiriatti M. GPT-3: Its nature, scope, limits, and consequences. Minds and Machines. 2020;30:681–694.

130. Conroy G. Scientists used ChatGPT to generate an entire paper from scratch-but is it any good? Nature. 2023;619(7970):443–444.

131. Inflection. Press release: Inflection AI Introduces Pi, Your Personal AI — inflection.ai; 2023. Available from: https://inflection.ai/press.

132. Depounti I, Saukko P, Natale S. Ideal technologies, ideal women: AI and gender imaginaries in Redditors' discussions on the Replika bot girlfriend. Media, Culture & Society. 2023;45(4):720–736.

133. Skjuve M, Følstad A, Fostervold KI, Brandtzaeg PB. My chatbot companion-a study of human-chatbot relationships. International Journal of Human-Computer Studies. 2021;149:102601.

134. Ta V, Griffith C, Boatfield C, Wang X, Civitello M, Bader H, et al. User experiences of social support from companion chatbots in everyday contexts: thematic analysis. Journal of medical Internet research. 2020;22(3):e16235.

135. Pentina I, Hancock T, Xie T. Exploring relationship development with social chatbots: A mixed-method study of replika. Computers in Human Behavior. 2023;140:107600.

136. Griffiths M. Is LaMDA sentient? AI & SOCIETY. 2022; p. 1–2.

137. Le T, Nguyen T, Ho N, Bui H, Phung D. Lamda: Label matching deep domain adaptation. In: International Conference on Machine Learning. PMLR; 2021. p. 6043–6054.

138. Lemoine B. Is LaMDA Sentient?—an Interview. Medium. 2022;.

139. Chowdhery A, Narang S, Devlin J, Bosma M, Mishra G, Roberts A, et al. Palm: Scaling language modeling with pathways. arXiv preprint arXiv:220402311. 2022;.

140. Driess D, Xia F, Sajjadi MS, Lynch C, Chowdhery A, Ichter B, et al. Palm-e: An embodied multimodal language model. arXiv preprint arXiv:230303378. 2023;.

141. Anil R, Dai AM, Firat O, Johnson M, Lepikhin D, Passos A, et al. Palm 2 technical report. arXiv preprint arXiv:230510403. 2023;.

142. Campello de Souza B, Serrano de Andrade Neto A, Roazzi A. Are the new ais smart enough to steal your job? iq scores for chatgpt, microsoft bing, google bard and quora poe. IQ Scores for ChatGPT, Microsoft Bing, Google Bard and Quora Poe (April 7, 2023). 2023;.

143. Ram B, Pratima Verma PV. Artificial intelligence AI-based Chatbot study of ChatGPT, Google AI Bard and Baidu AI. World Journal of Advanced Engineering Technology and Sciences. 2023;8(01):258–261.

144. Rahaman MS, Ahsan M, Anjum N, Rahman MM, Rahman MN. The AI race is on! Google's Bard and OpenAI's ChatGPT head to head: an opinion article. Mizanur and Rahman, Md Nafizur, The AI Race is on. 2023;.

145. King MR. 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. 2023;16(2):175–179.

146. Roth E. Google-backed Anthropic launches Claude, an AI chatbot that's easier to talk to; 2023. Available from: https://www.theverge.com/2023/3/14/23640056/anthropic-ai-chatbot-claude-google-launch.

147. Smutny P, Schreiberova P. Chatbots for learning: A review of educational chatbots for the Facebook Messenger. Computers & Education. 2020;151:103862.

148. Anthropic. Claude 2 — anthropic.com; 2023. Available from: https://www.anthropic.com/index/claude-2.

149. Santhanam S, Shaikh S. A survey of natural language generation techniques with a focus on dialogue systems-past, present and future directions. arXiv preprint arXiv:190600500. 2019;.

150. Abdellatif A, Badran K, Costa DE, Shihab E. A Comparison of Natural Language Understanding Platforms for Chatbots in Software Engineering. IEEE Transactions on Software Engineering. 2022;48(8):3087–3102. doi:10.1109/TSE.2021.3078384.

151. Følstad A, Araujo T, Law ELC, Brandtzaeg PB, Papadopoulos S, Reis L, et al. Future directions for chatbot research: an interdisciplinary research agenda. Computing. 2021;103(12):2915–2942.

152. Essel HB, Vlachopoulos D, Tachie-Menson A, Johnson EE, Baah PK. The impact of a virtual teaching assistant (chatbot) on students' learning in Ghanaian higher education. International Journal of Educational Technology in Higher Education. 2022;19(1):1–19.

153. Chen X, Xie H, Hwang GJ. A multi-perspective study on artificial intelligence in education: Grants, conferences, journals, software tools, institutions, and researchers. Computers and Education: Artificial Intelligence. 2020;1:100005.

154. Luckin R, Holmes W, Griffiths M, Forcier LB. Intelligence unleashed: An argument for AI in education. Open Research Online. 2016;.

155. Kumar JA. Educational chatbots for project-based learning: investigating learning outcomes for a team-based design course. International journal of educational technology in higher education. 2021;18(1):1–28.

156. Barrett M, Branson L, Carter S, DeLeon F, Ellis J, Gundlach C, et al. Using artificial intelligence to enhance educational opportunities and student services in higher education. Inquiry: The Journal of the Virginia Community Colleges. 2019;22(1):11.

157. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. PLoS digital health. 2023;2(2):e0000198.

158. Lin CC, Huang AY, Yang SJ. A review of ai-driven conversational chatbots implementation methodologies and challenges (1999–2022). Sustainability. 2023;15(5):4012.

159. Okonkwo CW, Ade-Ibijola A. Chatbots applications in education: A systematic review. Computers and Education: Artificial Intelligence. 2021;2:100033.

160. Vázquez-Cano E, Mengual-Andrés S, López-Meneses E. Chatbot to improve learning punctuation in Spanish and to enhance open and flexible learning environments. International Journal of Educational Technology in Higher Education. 2021;18(1):1–20.

161. Kuhail MA, Alturki N, Alramlawi S, Alhejori K. Interacting with educational chatbots: A systematic review. Education and Information Technologies. 2023;28(1):973–1018.

162. Baskara FR. Chatbots and Flipped Learning: Enhancing Student Engagement and Learning Outcomes through Personalised Support and Collaboration. IJORER: International Journal of Recent Educational Research. 2023;4(2):223–238.

163. Zawacki-Richter O, Marín VI, Bond M, Gouverneur F. Systematic review of research on artificial intelligence applications in higher education–where are the educators? International Journal of Educational Technology in Higher Education. 2019;16(1):1–27.

164. Chen L, Chen P, Lin Z. Artificial intelligence in education: A review. Ieee Access. 2020;8:75264–75278.

165. Cardona MA, Rodríguez RJ, Ishmael K, et al. Artificial Intelligence and the Future of Teaching and Learning: Insights and Recommendations. US Department of Education, Office of Educational Technology. 2023;.

166. Miao F, Holmes W. Artificial Intelligence and Education. Guidance for Policy-makers. United Nations Educational, Scientific and Cultural Organization (UNESCO). 2021;.

167. Hwang GJ, Chang CY. A review of opportunities and challenges of chatbots in education. Interactive Learning Environments. 2021; p. 1–14.

168. Huang W, Hew KF, Fryer LK. Chatbots for language learning—Are they really useful? A systematic review of chatbot-supported language learning. Journal of Computer Assisted Learning. 2022;38(1):237–257.

169. Farrokhnia M, Banihashem SK, Noroozi O, Wals A. A SWOT analysis of ChatGPT: Implications for educational practice and research. Innovations in Education and Teaching International. 2023; p. 1–15.

170. Hasal M, Nowaková J, Ahmed Saghair K, Abdulla H, Snášel V, Ogiela L. Chatbots: Security, privacy, data protection, and social aspects. Concurrency and Computation: Practice and Experience. 2021;33(19):e6426.

171. Faguy A. Google warns employees about chatbots-including its own bard-out of privacy concerns, report says; 2023. Available from: https://www.forbes.com/sites/anafaguy/2023/06/15/google-warns-employees-about-chatbots-including-its-own-bard-out-of-priv?sh=6a9976bfb613.

172. Sitzmann T, Ely K, Bell BS, Bauer KN. The effects of technical difficulties on learning and attrition during online training. Journal of Experimental Psychology: Applied. 2010;16(3):281.

173. Reynolds L, McDonell K. Prompt programming for large language models: Beyond the few-shot paradigm. In: Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems; 2021. p. 1–7.

174. Strobelt H, Webson A, Sanh V, Hoover B, Beyer J, Pfister H, et al. Interactive and visual prompt engineering for ad-hoc task adaptation with large language models. IEEE transactions on visualization and computer graphics. 2022;29(1):1146–1156.

175. Zhou K, Yang J, Loy CC, Liu Z. Learning to prompt for vision-language models. International Journal of Computer Vision. 2022;130(9):2337–2348.

176. White J, Fu Q, Hays S, Sandborn M, Olea C, Gilbert H, et al. A prompt pattern catalog to enhance prompt engineering with chatgpt. arXiv preprint arXiv:230211382. 2023;.

177. Liu P, Yuan W, Fu J, Jiang Z, Hayashi H, Neubig G. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. ACM Computing Surveys. 2023;55(9):1–35.

178. Yang Z, Li L, Wang J, Lin K, Azarnasab E, Ahmed F, et al. Mm-react: Prompting chatgpt for multimodal reasoning and action. arXiv preprint arXiv:230311381. 2023;.

179. Zhou D, Schärli N, Hou L, Wei J, Scales N, Wang X, et al. Least-to-most prompting enables complex reasoning in large language models. arXiv preprint arXiv:220510625. 2022;.

180. Atlas S. ChatGPT for higher education and professional development: A guide to conversational AI. Digital Commons. 2023;.

181. Yao S, Yu D, Zhao J, Shafran I, Griffiths TL, Cao Y, et al. Tree of thoughts: Deliberate problem solving with large language models. arXiv preprint arXiv:230510601. 2023;.

182. Wei J, Wang X, Schuurmans D, Bosma M, Chi E, Le Q, et al. Chain of thought prompting elicits reasoning in large language models. arXiv preprint arXiv:220111903. 2022;.

183. Wang X, Wei J, Schuurmans D, Le Q, Chi E, Narang S, et al. Self-consistency improves chain of thought reasoning in language models. arXiv preprint arXiv:220311171. 2022;.

184. Wang B, Deng X, Sun H. Iteratively prompt pre-trained language models for chain of thought. In: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing; 2022. p. 2714–2730.

185. Hulbert D. Tree of Knowledge: ToK aka Tree of Knowledge dataset for Large Language Models LLM; 2023. https://github.com/dave1010/tree-of-thought-prompting.

186. Baidoo-Anu D, Owusu Ansah L. Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. Available at SSRN 4337484. 2023;.

187. Oblique Shock Waves Table; 2023. Available from: http://www.aerodynamics4students.com/gas-dynamics-and-supersonic-flow/table8.php.

188. Wang X, Hu Z, Lu P, Zhu Y, Zhang J, Subramaniam S, et al. SciBench: Evaluating College-Level Scientific Problem-Solving Abilities of Large Language Models. arXiv preprint arXiv:230710635. 2023;.

189. Han R, Peng T, Yang C, Wang B, Liu L, Wan X. Is Information Extraction Solved by ChatGPT? An Analysis of Performance, Evaluation Criteria, Robustness and Errors. arXiv preprint arXiv:230514450. 2023;.

190. Metze K, Morandin-Reis RC, Lorand-Metze I, Florindo JB. The Amount of Errors in ChatGPT's Responses is Indirectly Correlated with the Number of Publications Related to the Topic Under Investigation. Annals of Biomedical Engineering. 2023; p. 1–2.

191. Borji A. A categorical archive of chatgpt failures. arXiv preprint arXiv:230203494. 2023;.

192. Chen L, Zaharia M, Zou J. How is ChatGPT's behavior changing over time? arXiv preprint arXiv:230709009. 2023;.

193. Paul A. CHATGPT's accuracy has gotten worse, study shows; 2023. Available from: https://www.popsci.com/technology/chatgpt-human-inaccurate/.

194. Cassidy C. Australian universities to return to 'pen and paper'exams after students caught using AI to write essays. The Guardian. 2023;10.

195. Rudolph J, Tan S, Tan S. ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? Journal of Applied Learning and Teaching. 2023;6(1).

196. Lim V. ChatGPT raises uncomfortable questions about teaching and classroom learning. The Straits Times. 2022; p. B3.

197. Abdulquadri A, Mogaji E, Kieu TA, Nguyen NP. Digital transformation in financial services provision: A Nigerian perspective to the adoption of chatbot. Journal of Enterprising Communities: People and Places in the Global Economy. 2021;15(2):258–281.

198. Pillai R, Sivathanu B. Adoption of AI-based chatbots for hospitality and tourism. International Journal of Contemporary Hospitality Management. 2020;32(10):3199–3226.

199. Lewandowski T, Delling J, Grotherr C, Böhmann T. State-of-the-Art Analysis of Adopting AI-based Conversational Agents in Organizations: A Systematic Literature Review. PACIS. 2021; p. 167.