Preprints.org

Article

ChatGPT-5 in Education: New Capabilities and Opportunities for Teaching and Learning

Wan Chong Choi [*] and Chi In Chang

Article

ChatGPT-5 in Education: New Capabilities and Opportunities for Teaching and Learning

Wan Chong Choi [1,2,*] and Chi In Chang [3]

[1] School of Informatics, Computing and Cyber Systems, Northern Arizona University, U.S.

[2] Department of Computer Science, Illinois Institute of Technology, U.S.

[3] Department of Psychology, Golden Gate University, U.S.

* Correspondence: wc388@nau.edu

Abstract

This study examined the technological advancements of OpenAI's GPT-5 and their implications for teaching and learning. Building on previous GPT iterations, GPT-5 introduced a unified model architecture featuring real-time routing, enhanced reasoning, and substantial improvements in factual accuracy, multimodal capabilities, and instruction following. Benchmark evaluations demonstrated notable performance gains across mathematics, coding, visual reasoning, and professional knowledge tasks, surpassing prior models in both depth and reliability. Through a narrative synthesis, this paper addressed two research questions: (1) What were GPT-5's architectural and performance improvements compared to earlier models? and (2) How did these advancements translate into educational opportunities? Findings indicated that GPT-5 supported diverse instructional scenarios, including intelligent tutoring, assessment design, content generation, and cross-linguistic learning. A key innovation, Study Mode, enhanced educational alignment by walking learners through problems step-by-step, prompting reasoning, and offering scaffolded hints—mirroring evidence-based teaching strategies such as formative feedback and the Socratic method. These features enabled more personalized, interactive, and pedagogically sound engagement, particularly in self-directed and differentiated learning contexts. Although large-scale classroom studies remain limited, GPT-5 represented a major step forward in the use of AI for education, offering improved reliability, flexibility, and alignment with instructional goals. Responsible integration and ongoing evaluation were identified as essential for maximizing its educational impact.

Keywords: GPT-5; ChatGPT; OpenAI; large language models (LLMs); artificial intelligence in education; intelligent tutoring systems; study mode; educational technology; AI-enhanced teaching and learning; responsible AI integration; education

1. Introduction

Large language models (LLMs) have rapidly evolved in their capabilities and applications in education over the past few years [1]. From the initial release of GPT-3, which offered a glimpse of AI's potential to assist learners, to the widespread adoption of ChatGPT (based on GPT-3.5) in 2023, and the subsequent improvements in GPT-4, each generation has expanded what is possible in educational settings. These models have been used for tasks ranging from answering students' questions and providing writing feedback, to serving as conversational partners for language practice [1] [2].

The impact has been significant [3]– for instance, within a year of its launch, ChatGPT reached hundreds of millions of users, fueling both excitement and concern in academia [4] [5]. Enthusiasts have envisioned "expert tutors available on demand through every smartphone" [4], and even world leaders acknowledged AI's potential to transform education by supporting personalized tutoring at scale [4]. At the same time, educators noted limitations of earlier models: while they could answer questions, they were not specifically designed to follow pedagogical best practices and sometimes

exhibited uncanny confidence when giving an incorrect answer [6]. GPT-4 marked a substantial improvement in intelligence [7] [8] – yet it still occasionally produced factual errors or inconsistent responses that posed challenges for classroom use.

In this context, the debut of GPT-5 in 2025 is regarded as a potential watershed moment for AI in education. OpenAI's GPT-5 is described as "our best AI system yet", a model that represents a significant leap in intelligence over its predecessors [9]. GPT-5 combines state-of-the-art performance across domains like math, coding, writing, health, and even visual understanding, with architectural innovations aimed at making the model more versatile and reliable [9].

One notable new feature is Study Mode [10], which enables GPT-5 to walk learners through problems step-by-step, prompting their reasoning and offering targeted hints rather than simply giving away solutions. This feature is specifically aligned with instructional goals such as scaffolding and formative feedback, and represents a more deliberate pedagogical orientation than seen in earlier models.

Moreover, GPT-5 introduces a unified architecture that can dynamically adjust its reasoning depth and significant advances in factual accuracy and safety, addressing many prior concerns about LLMs in educational use [9]. As such, GPT-5 arrives at a pivotal time, offering new possibilities for teaching and learning while raising important questions about integration and ethics in educational environments.

This study examined GPT-5's technological advancements and their educational implications through two research questions:

RQ1: What are GPT-5's architectural and performance advancements compared to previous versions?

RQ2: How do these improvements create new opportunities for teaching and learning?

We begin by situating GPT-5 in the context of prior GPT models' use in education, then describe our review methodology. We next detail GPT-5's architecture and capabilities (RQ1), including its unified model design, new features, and benchmark performance. We then explore potential educational applications enabled by these advances (RQ2), such as intelligent tutoring, multilingual and multimodal learning support, content generation, assessment, and professional training.

By addressing these questions, we aim to provide a comprehensive picture of how GPT-5 may transform education and what steps are necessary to harness its benefits responsibly.

2. Literature Review

2.1. Early Promise of GPT Models in Education

Early experiments with GPT-series education models demonstrated significant promise and clear limitations. Personalized tutoring has been a long-envisioned application of AI in education, often framed by Benjamin Bloom's finding that one-on-one tutoring can dramatically improve learning outcomes [4]. LLMs like ChatGPT began to approximate this vision by offering on-demand, interactive help to students.

Studies reported that ChatGPT could provide instant explanations of complex concepts, customized feedback, and 24/7 support, thereby improving accessibility to learning resources [1] [2] [5]. For example, a recent randomized controlled trial by Kestin et al. [4] implemented a carefully designed AI physics tutor (using GPT-4) and found that students learn significantly more in less time when using the AI tutor, compared with in-class active learning, with students also feeling more engaged and motivated.

This indicates that, under guided conditions, current generative AI can mimic some benefits of expert human tutors. Major educational organizations have also explored GPT-powered tutors; Khan Academy's Khanmigo, built on GPT-4, is one notable example of an AI teaching assistant designed to guide learners through problems rather than just giving answers [11].

2.2. Limitations and Risks of Earlier GPT Models

However, previous GPT models also had notable limitations that complicated their use in education [1] [6]. One well-documented issue is the tendency to hallucinate — that is, to produce incorrect information with a confident tone. ChatGPT (GPT-3.5) and even GPT-4 sometimes provided answers that were factually wrong or flawed reasoning, yet expressed in a manner that could mislead learners [5].

For instance, educators observed that ChatGPT might assert an incorrect math solution or mark a correct answer as incorrect, all with the same smooth fluency that it exhibits for reliable answers [4]. This "confident in error" behavior is dangerous in an educational context, as novice learners might uncritically trust the AI's responses.

Another challenge is the lack of pedagogical strategy: these models were trained to be generally helpful conversational agents, not specifically to foster learning. As a result, without careful prompting or system design, they might simply give away answers or perform tasks for the student, short-circuiting the learning process. Researchers have pointed out that unguided use of ChatGPT can enable students to complete assignments with minimal thinking, undermining the development of problem-solving skills [1] [5] [6].

The models do not inherently apply teaching techniques like scaffolding or eliciting student reasoning, since they were not explicitly trained as intelligent tutoring systems.

2.3. Integrity and Privacy Concerns

Previous work also highlights academic integrity concerns. By late 2022 and 2023, as ChatGPT's popularity surged, educators grew worried about students using it to cheat on essays or problem sets [12]. ChatGPT could generate entire essays or solutions that evade plagiarism detectors (because the text is newly generated), raising what one article called "academic integrity hanging in the balance."

Some school districts and universities reacted with outright bans on AI-generated content or the tool itself [5] [12]. In contrast, others adopted a more accommodating stance, issuing guidelines for responsible AI use in coursework. There is an emerging consensus that simply prohibiting these tools is not sustainable; education must adapt through new honor codes, assessment designs less vulnerable to AI, and detection tools (though AI-generated text detection remains an unsolved problem).

Alongside cheating, privacy is another concern: AI models like ChatGPT require large amounts of data and typically send user queries to centralized servers [6]. This raises issues if students input personal information or school data – it may conflict with student data protection laws if not appropriately handled [13]. Given past incidents, caution has been urged about what data teachers and students share with such platforms.

2.4. Constructive Applications in Practice

Despite these challenges, numerous positive applications of GPT models in education have been explored [14] [15]. In the domain of language learning, for example, learners have used ChatGPT for conversational practice in foreign languages, receiving instant feedback on grammar and vocabulary. The model's patience and non-judgmental nature can encourage students to practice more freely [2] [16] [17].

Early analyses suggest ChatGPT can indeed help improve writing and language skills by offering corrections and alternative phrasing [5]. For translation tasks, GPT-4 reached near-professional quality on many language pairs, suggesting that GPT-5 could further assist in multilingual education, both by translating educational materials and by allowing students to ask questions in their native language and receive answers in the target language.

Another emergent use is question generation and assessment: instructors experimented with using GPT-3.5/4 to generate quiz questions, practice problems, or even entire exam papers. This can save considerable time for teachers and enable the creation of large question banks for test prep or

formative assessment [1] [16]. Preliminary studies indicate AI-generated questions can be of acceptable quality, though often requiring human review for accuracy and alignment to curriculum [18].

In subjects like medicine and law, GPT-4's strong performance on professional exams (e.g., USMLE for medical licensing, the Bar exam for law) hinted that it could serve as a valuable training aid for students in those fields – for instance, by simulating clinical case questions or mock trial scenarios.

2.5. Summary and Research Gap

In summary, prior to GPT-5, the use of GPT models in education was marked by a dual character: transformative potential (personalized tutoring at scale, improved access and support, automated content and feedback generation) on one hand, and substantial concerns (hallucinations, misguided usage, cheating, privacy issues) on the other.

Researchers and educators have called for more studies to better understand these tools' impact and for the development of frameworks to integrate AI into education ethically.

Our study is the first to examine the use of GPT-5 in education, addressing this clear gap in the literature and providing updated insights on how newer models may overcome previous limitations.

3. Methodology

This article presents a narrative synthesis [19] of sources relevant to GPT-5 and its potential role in education. The study draws on three bodies of information. First, we consulted official material released by OpenAI about GPT-5. Second, we searched academic and industry publications that discuss how earlier GPT models have been used in teaching and learning. Third, we examined recent media coverage and expert commentary that describe the features of GPT-5 and early impressions from the field.

3.1. Search Strategy and Selection Criteria

We conducted structured searches in Scopus, Web of Science, arXiv, leading conference proceedings, and education-technology reports. Keywords combined model names (GPT-3, GPT-4, GPT-5) with terms such as "education", "tutoring", and "learning". Items were screened by title and abstract, then by full text, keeping studies that reported empirical findings or detailed case descriptions. For GPT-5, because empirical classroom studies are not yet available, we relied on OpenAI's published evaluation metrics and independent benchmarks released since August 2025.

3.2. Data Extraction and Synthesis

We recorded aims, context, data type, reported outcomes, and noted strengths and limitations for each included source. We then used a narrative approach to organise findings under two research questions. RQ1 asks what technical advances GPT-5 offers relative to GPT-4 and earlier models, with attention to reasoning, accuracy, context length, and safety controls. RQ2 asks what educational opportunities these advances may introduce. In addressing RQ2, we mapped documented experiences with GPT-4 and similar systems onto the new capabilities of GPT-5 and considered how changes in performance might influence usefulness, reliability, and alignment with learning goals. Where evidence was insufficient, we highlighted the need for empirical work and marked statements as projections.

3.3. Limitations

The recency of GPT-5 means that peer-reviewed classroom studies are lacking. Therefore, this study combines established findings on prior models with early technical reports on ChatGPT. Conclusions should be read as provisional until more data emerge.

4. Results for RQ1: GPT-5 Architecture and Capability Improvements

4.1. Unified System and Realtime Router

One of the most significant architectural innovations in GPT-5 is its design as a unified system with an intelligent routing mechanism. In previous iterations like GPT-4, users had to choose between different model versions—for example, the default fast model versus a "GPT-4 with advanced reasoning" that was slower and had higher limits.

GPT-5 eliminates the need for users to manually select a model for different tasks [9]. Instead, GPT-5 automatically incorporates multiple sub-models specialized for different levels of complexity, along with a real-time router that automatically directs queries to the appropriate component on the fly [9], as shown in Figure 1. As OpenAI describes, GPT-5 consists of:

(i) a "smart, efficient model" that can answer most straightforward questions rapidly,

(ii) a deeper "GPT-5 thinking" model that engages in extended reasoning for challenging or complex problems, and

(iii) a router which automatically decides which model (or how much reasoning) is needed, based on factors such as the question's complexity, the conversation context, tool usage, and even explicit user instructions to "think hard" [9].
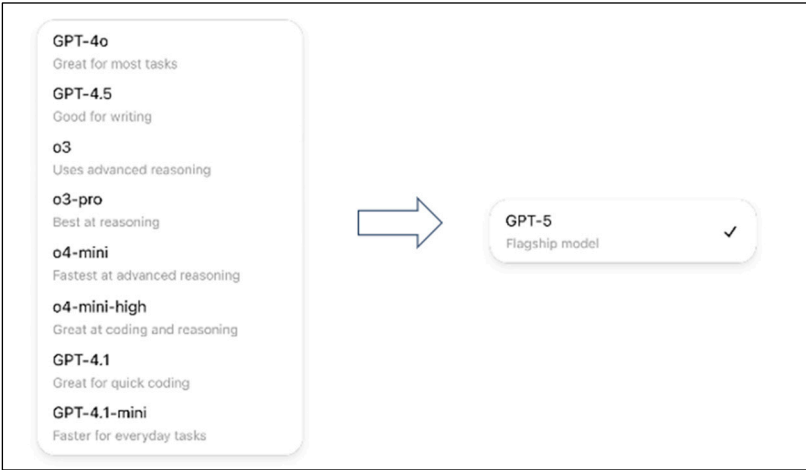

Figure 1. GPT-5's Unified Model and Real-Time Routing System.

This routing mechanism reflects a dynamic form of what researchers call meta-reasoning, where the AI allocates computational effort based on the nature of the task.

The unified system significantly simplifies the user experience: you constantly interact with one GPT-5, and under the hood, it "knows when to respond quickly and when to think longer" [9]. For instance, a simple factual question or a request for a summary might trigger the lightweight fast response mode, returning an answer almost immediately. In contrast, a multi-step math word problem or a request to critique an essay may invoke the deeper reasoning mode, which takes more time to generate a well-considered response.

The real-time router is continuously trained and refined using feedback signals from actual usage [9]. For example, if users frequently press "Try again", switch to a different model after receiving a response, or indicate dissatisfaction with an answer, the router adapts its decisions accordingly. Over time, this should make GPT-5's reasoning more efficient and contextually appropriate.

This architecture addresses a practical challenge seen with GPT-4: some tasks did not require the advanced model's full reasoning power (and latency), whereas others benefited from it. GPT-5 can now handle both types of tasks seamlessly within one system.

OpenAI's eventual plan is to integrate the fast and deep models into a single model in the future [9]. For now, however, the router-based architecture offers a pragmatic solution. An added benefit is

graceful handling of usage limits: GPT-5 also includes scaled-down versions (internally referred to as "mini" and "nano" models) that the system can fall back to if a user exceeds their allocation of the main model's usage [9].

In practice, if a student on the free tier reaches their usage limit, GPT-5 may automatically switch to a lightweight model for remaining queries, rather than cutting the user off entirely. While this may result in reduced response quality, it ensures continuity of service. For educators and learners, the unified GPT-5 means less friction—no need to guess which mode to use—and it promises the "best of both worlds" in terms of speed and depth of reasoning within a single assistant.

4.2. Software Generation and Multitasking

GPT-5 extends what an AI assistant can do by providing information, performing tasks, and generating software. OpenAI's CEO, Sam Altman, described "software on demand" as a defining capability of the GPT-5 era. GPT-5 can take high-level requests and produce working programs or applications in one prompt. Compared to GPT-4, GPT-5 shows specific improvements in complex front-end generation and debugging larger codebases [9].

In educational settings, GPT-5 can serve as a programming assistant, explaining code and generating substantial portions of it. For example, a student might ask for help building a simple physics simulation, and GPT-5 can generate a working codebase with minimal further prompting.

GPT-5 has also been optimized for agentic behavior, meaning it can invoke tools or APIs and handle multi-step tasks autonomously. According to OpenAI, GPT-5 shows significant gains in benchmarks that test instruction following and multi-step reasoning, including coordination across tools such as browsers and code execution environments [9].

This allows GPT-5 to handle complex educational tasks end-to-end. For example, when asked a real-world data question, GPT-5 could retrieve the data, process it using a code interpreter, and return a graph and written analysis. These kinds of orchestrated tasks are now more robust than GPT-4, which often require careful prompting.

GPT-5's capabilities also make it useful for students and teachers in creative and planning contexts. It can help solve a math problem, then switch to generating a list of supplies for a science project, or even draft an experimental procedure.

The ability to generate working software and complete multi-step tasks leads to new pedagogical possibilities. Students can focus on a project's design and conceptual aspects while delegating parts of the implementation to the AI. However, this raises the challenge of balancing automation and learning. GPT-5's multitasking improvements are a key technical advancement supporting this shift [9].

Figure 2 shows GPT-5's response to a prompt asking for a complete HTML-based jumping game. The model generates a visually styled, fully functional, single-page application with user interaction, sound effects, and gameplay logic from a high-level natural language description. This example illustrates GPT-5's ability to perform complex multi-step software generation tasks relevant for education, particularly in computer science and STEM teaching contexts [9].
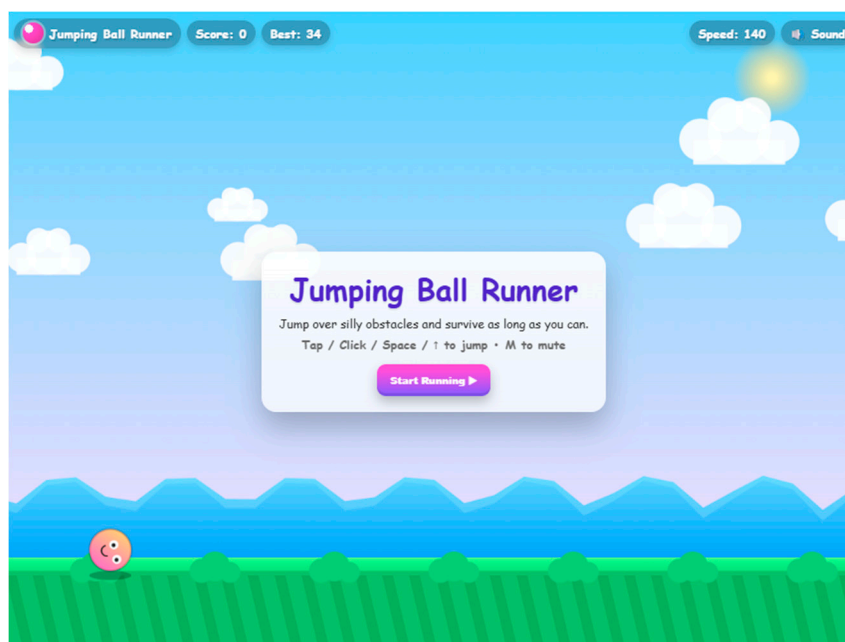
Figure 2. GPT-5's Capability to Autonomously Generate Educational Software from Natural Language Prompts [9].

4.3. Reliability and Safety Enhancements

OpenAI invested heavily in making GPT-5 more reliable, factual, and safe than its predecessors. A headline claim is that GPT-5 is "by far our most reliable, factual model ever," with substantially reduced rates of hallucination. In internal evaluations, GPT-5's responses were about 45% less likely to contain factual errors than GPT-4 (GPT-4o) on real user prompts with web browsing enabled. In its full "thinking" mode, GPT-5 achieved around an 80% reduction in factual errors compared to the earlier o3 model [9].

GPT-5 also exhibits improvements in minimizing sycophancy. For example, when asked to describe images that were not shown, o3 fabricated responses 86.7% of the time, while GPT-5 did so only 9% of the time. GPT-5 also showed about half the deception rate of o3 in tasks designed to test bluffing [9].

GPT-5 introduces a more flexible safety strategy called safe completions. Instead of blunt refusals, it offers partial and helpful responses when possible. When refusal is necessary, GPT-5 explains why and suggests alternative queries [9].

In biomedical contexts, GPT-5's advanced reasoning mode is treated as a high-capability model. It was red-teamed for over 5,000 hours in collaboration with groups like CAISI to evaluate its safety in bioscience scenarios [9]. These safeguards help ensure GPT-5's responses are more appropriate for health and education-related use.

In summary, GPT-5 significantly improves factuality, self-awareness of limitations, and safety handling, making it more suitable as an educational tool.

4.4. Expanded Compute Modes and Personalization Options

GPT-5 introduces new usage modes and personalization options that enhance its applicability in educational contexts. All users can access GPT-5, while Plus subscribers receive higher usage limits, and Pro subscribers gain access to GPT-5 Pro—a variant with extended reasoning capabilities that enables more detailed and accurate responses to complex queries [9].

GPT-5 is structured as a unified system comprising a standard model for fast responses, a deeper reasoning model (GPT-5 Thinking), and a real-time router that dynamically selects between them based on prompt complexity, user intent, and tool usage. When usage limits are reached, a mini version of the model handles subsequent interactions [9].

Personalization features in ChatGPT with GPT-5 include the ability to choose interface theme colors and conversational personalities. Users can select between different interaction styles, such as more concise, more supportive, or more professional responses. These settings allow the AI to adapt its tone to better fit individual or classroom preferences [9].

Voice interaction has also improved: users can now engage GPT-5 using natural speech with adjustable speed and speaking style. This benefits language learning and accessibility use cases.

On the API side, developers can access GPT-5 Pro for tasks that require higher reasoning depth. GPT-5 shows improved instruction following, reliability, and reduced hallucination rates across domains, including coding, math, health, and writing [9].

These enhancements make GPT-5 a more adaptable and capable platform for educational applications. Its configurability allows instructors to shape the model's behavior to meet specific pedagogical needs, and its improved reasoning supports complex, multi-step educational interactions.

4.5. Safer Behavior and Reduced Sycophancy

GPT-5 includes improvements that make its responses more accurate, responsible, and appropriate in broader contexts. Compared to previous versions, the model demonstrates reduced hallucinations, better instruction following, and a significant decrease in sycophantic behavior, where the model might have previously agreed with incorrect or misleading user statements. These changes are especially relevant in educational settings, where factual integrity and ethical consistency are critical [9].

OpenAI shows GPT-5 is less likely to echo false assertions even when users phrase them confidently. Instead, the model maintains factual correctness and adheres to established guidelines. This is important in classrooms, where students may present misconceptions or controversial opinions. In such cases, GPT-5 is more likely to correct misinformation gently but clearly, rather than reinforcing misunderstandings [9].

In addition to improvements in accuracy and alignment, GPT-5 demonstrates enhanced stylistic flexibility. It can adjust its responses' formality, complexity, and tone based on user input and contextual cues. This allows the model to respond more appropriately across various educational levels—from simplifying concepts for younger learners to using precise terminology with advanced students. A notable example shared in OpenAI's release compared GPT-5 and GPT-4 on a poetic task, where GPT-5's response was described as having more explicit imagery and more refined emotional structure, showing better adherence to the requested style [9], as shown in Figure 3.

Together, these advancements make GPT-5 a more suitable tool for educational use. Its improved judgment, reduced tendency to reinforce errors, and adaptable communication style contribute to safer, more trustworthy interactions with learners. These are traits commonly expected from human tutors, and GPT-5's training has made measurable progress toward achieving them.
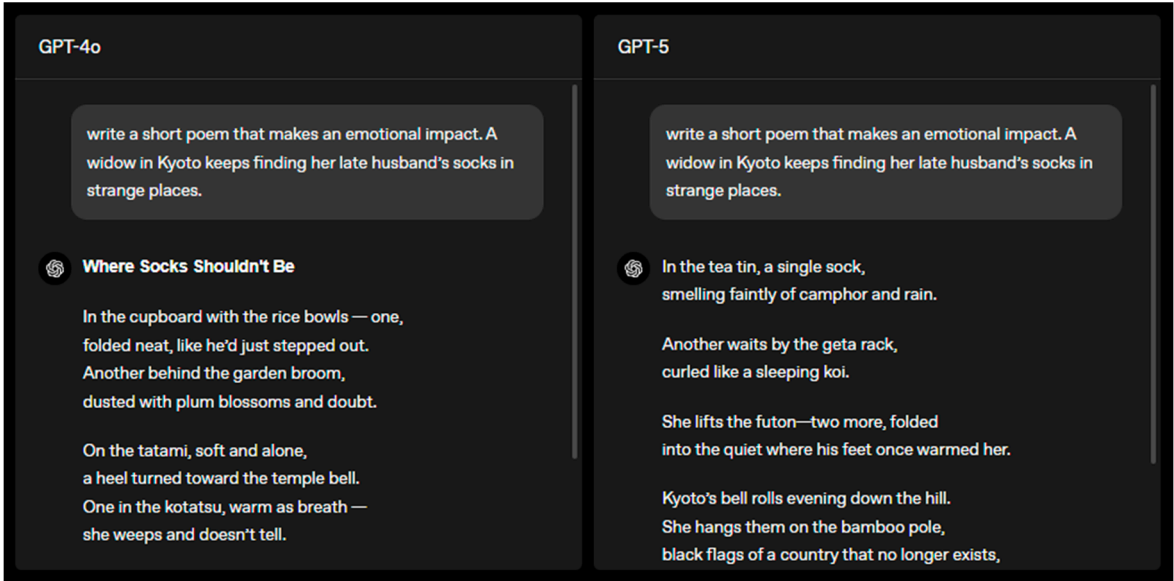
Figure 3. Comparative Output on a Poetic Task Between GPT-4o and GPT-5 [9].

4.6. Performance on Key Benchmarks

OpenAI reports that GPT-5 achieves state-of-the-art results on a wide range of academic and professional benchmarks, reflecting improvements in reasoning and real-world utility [9].

In mathematics, GPT-5 achieved 94.6% accuracy on AIME 2025 without tool use, and GPT-5 Pro achieved 100% accuracy with tool-assisted reasoning. This places GPT-5 well above prior models regarding advanced mathematical problem solving. GPT-5 scored 74.9% in software engineering on SWE-bench Verified and 88.0% on Aider Polyglot, benchmarks that evaluate real-world and multi-language programming capabilities [9].

GPT-5 also demonstrates strong multimodal reasoning abilities. It achieved 84.2% on MMMU (college-level visual tasks) and 78.4% on MMMU Pro (graduate-level visual tasks). Additional multimodal evaluations, such as VideoMMMU and CharXiv Reasoning, show similarly high accuracy, suggesting GPT-5's ability to interpret visual and scientific inputs with greater precision than previous models [9].

In the health domain, GPT-5 scored 46.2% on HealthBench Hard, significantly outperforming prior models. It adapts responses based on the user's knowledge level and location, and provides proactive and context-sensitive guidance. While it is not a substitute for medical professionals, it is designed to support users in understanding medical information more effectively [9].

GPT-5 also outperforms previous models on comprehensive general knowledge assessments. On the Humanity's Last Exam, which includes expert-level questions across various fields, GPT-5 Pro scored 42.0%, nearly double GPT-4o's performance [9]. This benchmark highlights GPT-5's general reasoning ability across disciplines such as history, law, science, and literature.

Finally, in an internal benchmark evaluating performance across more than 40 professional domains—including law, logistics, and engineering—GPT-5 was comparable to or better than human experts in roughly half of the tasks when reasoning was enabled. These results suggest that GPT-5 can assist in many forms of advanced knowledge work and training contexts.

5. Results for RQ2: Educational Application Scenarios for GPT-5

5.1. Virtual Tutoring and Personalized Learning with GPT-5's Study Mode

GPT-5's enhanced reasoning, accuracy, and adaptation make it suitable as a virtual tutor across education levels. Compared to earlier models, GPT-5 is better at adjusting explanations based on user needs, thanks partly to features like Study Mode [10] and personality customization [20].

In Study Mode, as shown in Figure 4, the model breaks problems into smaller steps, prompts student thinking with questions, and offers hints when needed, closely mimicking human tutoring strategies such as the Socratic method.

Personalization is supported on several levels. GPT-5 can dynamically adjust the depth and complexity of its responses, depending on the learner's apparent understanding. If a student starts with a fundamental question and then asks a deeper follow-up, the model can respond with more technical detail and examples. Its large context window (up to 400k tokens) allows it to retain earlier parts of a tutoring session, enabling context-aware feedback and continuity over long interactions [20]. Teachers can also set the model's tone and persona—for example, "Kind Coach" or "Socratic Professor"—to match student preferences and improve engagement.

Previous research has already demonstrated educational benefits from AI tutors [2]. In a randomized controlled trial, GPT-4-based tutoring led to significant learning gains when paired with effective pedagogy [4]. With GPT-5's improvements, these gains may increase, as the model can now apply such methods more fluently, without relying solely on prompt design.

Beyond supporting struggling students, GPT-5 can assist advanced learners. A student interested in topics beyond their curriculum—such as quantum mechanics or general relativity—can use GPT-5 to explore these subjects through progressively deeper explanations. This enables self-guided study even when qualified human mentors are unavailable.

While GPT-5 cannot replace the emotional support provided by human teachers, it can complement them by handling routine instruction and providing personalized practice. Teachers may use GPT-5 for homework support, differentiated instruction, or as a resource during supervised tutoring sessions. Early feedback may suggest that students find AI tutors non-judgmental and approachable, making them more likely to ask questions or request repeated explanations [4].
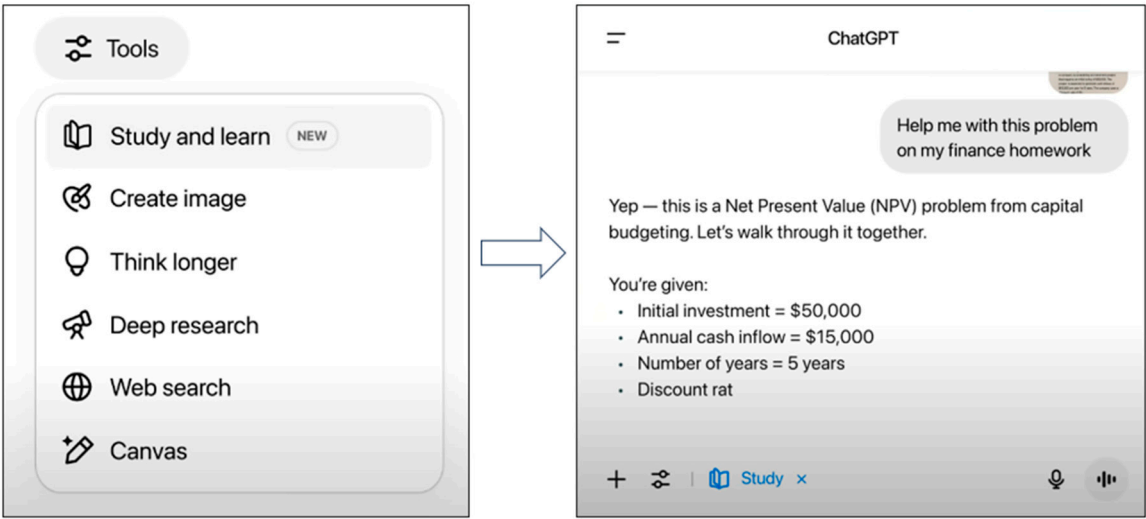

Figure 4. ChatGPT-5's Study Mode Enabling Step-by-Step Interactive Tutoring [10].

5.2. Cross-Linguistic and Multimodal Support

GPT-5's enhanced multilingual and multimodal capabilities may help address language and format barriers in education. Compared to GPT-4, GPT-5 performs better in cross-linguistic tasks, including multi-language code editing benchmarks, suggesting more robust polyglot abilities [9]. In practice, this may allow students to engage in conversations across languages, receive corrections, and translate educational content. For example, a Spanish-speaking student attending an English class may use GPT-5 to understand course material in their native language or interact with the AI using bilingual input and output [5].

In multilingual classrooms, GPT-5 may assist teachers by generating translated instructions or adapting reading materials to students' proficiency levels. Unlike static translation tools, GPT-5 can

modify its translations according to the learner's reading ability. Its voice input and output also allow for pronunciation practice and listening comprehension support [20].

GPT-5's multimodal support enables users to engage through both text and images. Students may upload visual content—such as diagrams, graphs, or historical artifacts—and ask questions about them. GPT-5 can analyze and interpret these visuals, offering explanations or identifying errors [20]. In STEM education, it may assist in solving geometry problems, analyzing lab data, or interpreting circuit diagrams.

These capabilities also support inclusive learning. For visually impaired students, GPT-5 offers voice responses; for hearing-impaired learners, it delivers complete text-based interactions [20]. Visual learners may benefit from GPT-5's descriptive output, including its ability to convey imagery or explain visual data using language.

While limitations remain—such as image generation being external to ChatGPT itself—GPT-5's multilingual and multimodal features may allow it to act as a flexible educational assistant. It could be particularly valuable for learners in under-resourced or linguistically diverse environments, helping to translate, interpret, and explain content in ways that promote accessibility and equity.

5.3. Creative Writing and Instructional Content Generation

GPT-5's improvements in language generation may support both creative writing and educational content development. OpenAI describes it as their "most capable writing collaborator yet," with improved stylistic control and instruction adherence [9].

For students, GPT-5 may serve as a writing assistant by helping brainstorm ideas, suggesting revisions, or demonstrating literary forms. For instance, it can provide examples of sonnets or help restructure argumentative essays while preserving core ideas. It's reduced hallucination and more precise instruction-following mean that feedback is more likely to stay relevant to the student's original intent. Students learning poetry or narrative writing may benefit from concrete demonstrations of form and tone, such as GPT-5 maintaining poetic meter or replicating a particular author's voice [9].

GPT-5 may also aid in revision by offering context-aware feedback. It can explain grammatical issues, assess coherence across long essays, or evaluate tone consistency, taking advantage of its extended 400k-token context window. Unlike grammar tools that flag issues without explanation, GPT-5 can provide reasoning behind suggested changes.

For teachers, GPT-5 may function as a content-generation tool, producing practice problems, simplified texts, quizzes, and example explanations. It can adapt output to student reading levels or preferred formats (e.g., summaries, dialogues), which may help in differentiating instruction. Teachers might also use GPT-5 to generate reading comprehension passages, conceptual dialogues, or scaffolded writing assignments.

In language learning contexts, students may draft essays in their native language and use GPT-5 to produce versions in the target language. Because GPT-5 captures both meaning and style, it may help learners avoid common translation pitfalls and improve fluency.

Although concerns about misuse remain, structured assignments can incorporate GPT-5 transparently. For example, students might be asked to use GPT-5 to generate an outline, write their essay, and revise using the model's feedback. This process fosters critical engagement with AI-generated content while supporting writing development.

In summary, GPT-5 may enhance writing education by providing flexible assistance in drafting, revising, and modeling. It can reduce teacher workload in content creation and offer students personalized, on-demand feedback. Research suggests student engagement increases when AI is used as a writing partner, and GPT-5's advancements may strengthen this effect. Used thoughtfully, GPT-5 may support students in developing stronger writing skills and deeper metacognitive awareness of their own work.

5.4. Assessment Item Generation and Scenario Simulation

GPT-5 may assist educators in generating high-quality assessment items and simulating realistic learning scenarios. One persistent challenge in education is crafting questions that assess true understanding rather than rote memorization. GPT-5 can generate varied question types on demand, such as conceptual and calculation problems related to specific topics (e.g., Newton's laws), complete with answers and explanations. Early use of GPT models suggests that AI-generated questions can approximate teacher-written items, though review remains necessary [18].

A key strength of GPT-5 is its ability to create scenario-based and open-ended items. For example, it may simulate patient cases for medical students, business case studies, or legal dilemmas. Such scenarios can include realistic distractions or ambiguity, helping students practice applied reasoning. In interactive settings, GPT-5 can adopt roles—such as a patient, client, or judge—responding dynamically during student-led interviews or arguments. This could provide experiential learning opportunities typically requiring human role-play or scripted software.

Language learners may also benefit from conversational simulations. GPT-5 can adopt consistent personas (e.g., travel agent, historical figure), helping students practice dialogue with immediate feedback. Improved persona consistency contributes to a more authentic interaction.

GPT-5 may also support assessment design by helping educators identify challenging items for AI. By generating a pool of candidate questions and flagging those it finds difficult, GPT-5 may help highlight items requiring deeper human understanding, potentially helpful in countering AI-assisted cheating and designing assessments that emphasize reasoning.

In feedback and grading, GPT-5 may pre-screen student responses based on a rubric. For instance, it can highlight missing elements or summarize common misconceptions across a class set of essays. While not replacing human grading, this function may reduce workload and inform targeted instruction.

In professional education, the use of simulation may expand. Law students might practice courtroom exchanges with GPT-5 playing opposing counsel; business students could pitch ideas to a simulated investor posing critical questions. These role-based simulations allow repeated practice in complex settings, offering feedback and increasing student confidence.

In summary, GPT-5 may support both formative and summative assessment by enabling diverse item generation, scenario-based learning, and structured feedback. While human oversight remains essential to ensure alignment and fairness, GPT-5 could reduce the burden of content creation and make experiential learning more accessible at scale.

5.5. Professional Education and Training (Medicine, Law, and Beyond)

GPT-5's wide-ranging capabilities may support professional and vocational education, particularly in fields requiring complex knowledge and applied reasoning.

In medical education, GPT-5 may assist students preparing for exams like the USMLE by explaining complex topics such as pharmacology or pathology. Its improved factual accuracy over GPT-4 may reduce errors in medical guidance. GPT-5 also tends to prompt users with relevant follow-up questions, for example, asking if a differential diagnosis has considered certain symptoms, which mimics clinical reasoning by a supervisor [9].

It may also simulate patient interactions, helping trainees practice history-taking and communication. When playing a patient persona, GPT-5 can model realistic emotions or confusion. Afterward, it can provide feedback or annotate dialogues for instructional review. Its long context capacity also allows tracking ongoing clinical scenarios over multiple interactions.

In legal education, GPT-5 may help students analyze hypothetical fact patterns or explore case law. It can identify legal principles, list potential arguments, and suggest relevant precedents. It may also assist in writing tasks like drafting contract clauses or reviewing legal briefs.

Beyond medicine and law, GPT-5 may contribute to engineering, business, and finance education. In engineering, students may use it to troubleshoot designs or understand practical constraints. Its coding capabilities allow for debugging, code review, or collaborative programming.

Civil engineering or architecture may help interpret building codes or materials data. In business education, GPT-5 can analyze case studies, role-play stakeholders, or critique business strategies. For communication training, it can simulate difficult HR conversations and suggest better phrasing.

GPT-5 may also support preparation for professional licensing exams (e.g., CPA, PE), offering explanations, adaptive drills, and targeted remediation. If a learner repeatedly struggles on a topic, GPT-5 may shift its focus, offering review before more practice questions – a function similar to adaptive tutoring systems.

One broader implication is accessibility. For learners without access to expert instructors, such as those in remote or under-resourced areas, GPT-5 may offer support otherwise unavailable. While it cannot replace expert judgment in high-stakes decisions, it may help learners rehearse, explore, and test their knowledge more deeply.

For example, in a recent study on AI tutors in education, aligned AI feedback improved student outcomes [4]. While similar studies in medicine, law, or business are still emerging, GPT-5's simulated expertise may eventually bring comparable benefits in domains where traditional one-on-one training is limited.

5.6. Additional Emerging Applications

Beyond the main instructional uses, GPT-5 may support several emerging educational applications.

One is metacognitive skill development. GPT-5 may help students reflect on how they learn by prompting them to examine their study habits, offering strategies like self-quizzing over re-reading, or guiding weekly reflections on learning progress. Given its exposure to educational psychology literature, it may provide evidence-based advice on studying and goal setting.

Another area is accessibility. GPT-5 could support students with disabilities by simplifying instructions, breaking content into smaller steps, or rephrasing complex text for those with reading challenges like dyslexia or ADHD. It may also offer literal, consistent communication for students on the autism spectrum and provide practice for interpreting social cues in a low-pressure setting.

In the arts and music, GPT-5 might generate prompts or discuss compositions. For example, it may explain music theory when given chord progressions or suggest how to develop a melody. In creative writing or drama, it could help students script scenes or build dialogue for rehearsal.

For collaborative learning, GPT-5 may act as a neutral assistant, tracking group decisions, translating conversations in multilingual teams, or summarizing discussions. This could help students improve teamwork, planning, and communication skills.

In educational research and administration, GPT-5 might analyze documents, extract insights from student feedback, or draft reports. For example, it could help review hundreds of curriculum comments and surface key issues for faculty consideration.

GPT-5 may also support lifelong learning. Adults exploring new topics or changing careers could use it as an on-demand tutor—for example, someone learning coding later in life or studying a second language without formal instruction.

Another possible use is for parental support. GPT-5 may help parents understand modern curricula, such as explaining math methods or reading development strategies. It could also assist in community education, helping residents learn about health information, digital skills, or civic processes in accessible language.

All these scenarios rely on GPT-5's general strengths—language processing, reasoning, memory, and adaptability. While many of these applications remain experimental, they illustrate how GPT-5 might integrate into diverse educational settings. The key will be responsible use, with human educators shaping how GPT-5 is applied to meet learners' needs.

6. Conclusions

This study examined the technological advancements of GPT-5 (RQ1) and their implications for teaching and learning (RQ2).

In response to RQ1, GPT-5 introduced a unified model architecture with real-time routing, allowing it to adjust reasoning depth based on task complexity dynamically. This design eliminated the need for users to manually select different versions for different purposes, a limitation seen in GPT-4. GPT-5 demonstrated improved performance across multiple domains—including mathematics, software engineering, visual reasoning, and health—supported by benchmark results that showed substantial gains over earlier models. Enhancements in factual accuracy, reduced hallucination rates, improved sycophancy handling, and refined safe completion strategies made GPT-5 more reliable and appropriate for educational use. New features such as Study Mode, software generation capabilities, expanded personalization, and longer context windows further increased its practical utility in various educational contexts.

For RQ2, these advancements created new opportunities for learners and educators. GPT-5 supported interactive, personalized tutoring through its Study Mode, guiding students step-by-step and adapting explanations to their level of understanding. Its multilingual and multimodal capabilities helped overcome language and format barriers, enabling more inclusive educational support. GPT-5 also assisted in creative writing, instructional content generation, and assessment design, reducing teacher workload and enabling differentiated instruction. In professional education, GPT-5 showed potential in simulating complex scenarios in medicine, law, and engineering, offering learners structured practice that previously required human experts or dedicated systems. Additional applications emerged in accessibility support, metacognitive development, and lifelong learning, highlighting GPT-5's broad educational relevance.

In conclusion, GPT-5 represented a significant advancement in developing language models for education. It addressed key limitations of prior GPT models—particularly hallucinations, lack of pedagogical alignment, and the need to manage different model versions—while opening new possibilities across subject areas and learner profiles. Although empirical classroom studies remain limited due to the model's recency, the available evidence indicates that GPT-5 was more aligned with instructional goals and more adaptable to diverse educational needs than its predecessors. Its practical and responsible use will depend on thoughtful integration into curricula, clear pedagogical guidance, and ongoing evaluation of its educational impact.

References

1. W. C. Choi, I. C. Choi, and C. I. Chang, 'The Impact of Artificial Intelligence on Education: The Applications, Advantages, Challenges and Researchers' Perspective', *Preprints*, Jan. 2025, doi: 10.20944/preprints202501.1420.v1.

2. W. C. Choi and C. I. Chang, 'A Survey of Techniques, Design, Applications, Challenges, and Student Perspective of Chatbot-Based Learning Tutoring System Supporting Students to Learn in Education', 2025, *Preprints.org*. doi: 10.20944/preprints202503.1134.v1.

3. W. C. Choi, C. I. Chang, I. C. Choi, and L. C. Lam, 'A Review of Large Language Models (LLMs) Development: A Cross-Country Comparison of the US, China, Europe, UK, India, Japan, South Korea, and Canada', *Preprints*, Apr. 2025, doi: 10.20944/preprints202504.2136.v1.

4. G. Kestin, K. Miller, A. Klales, T. Milbourne, and G. Ponti, 'AI tutoring outperforms in-class active learning: an RCT introducing a novel research-based design in an authentic educational setting', *Sci. Rep.*, vol. 15, no. 1, p. 17458, 2025.

5. A. M. Hasanein and A. E. E. Sobaih, 'Drivers and consequences of ChatGPT use in higher education: Key stakeholder perspectives', *Eur. J. Investig. Health Psychol. Educ.*, vol. 13, no. 11, pp. 2599–2614, 2023.

6. C. I. Chang, W. C. Choi, and I. C. Choi, 'Challenges and Limitations of Using Artificial Intelligence Generated Content (AIGC) with ChatGPT in Programming Curriculum: A Systematic Literature Review', in *Proceedings of the 2024 7th Artificial Intelligence and Cloud Computing Conference*, 2024.

7. OpenAI, 'GPT-4'. [Online]. Available: https://openai.com/index/gpt-4-research/

8. W. C. Choi, I. C. Choi, C. I. Chang, and L. C. Lam, 'Comparison of Claude (Sonnet and Opus) and ChatGPT (GPT-4, GPT-4o, GPT-o1) in Analyzing Educational Image-based Questions from Block-Based

Programming Assessments', in *2025 14th International Conference on Information and Education Technology (ICIET)*, IEEE, 2025.

9. OpenAI, 'Introducing GPT-5'. [Online]. Available: https://openai.com/index/introducing-gpt-5/

10. OpenAI, 'Introducing Study Mode in ChatGPT'. 2025. [Online]. Available: https://www.youtube.com/watch?v=XDYilxy1dn8

11. S. Wang, F. Wang, Z. Zhu, J. Wang, T. Tran, and Z. Du, 'Artificial intelligence in education: A systematic literature review', *Expert Syst. Appl.*, vol. 252, p. 124167, 2024.

12. A. Johnson, 'ChatGPT In Schools: Here's Where It's Banned—And How It Could Potentially Help Students'. [Online]. Available: https://www.forbes.com/sites/ariannajohnson/2023/01/18/chatgpt-in-schools-heres-where-its-banned-and-how-it-could-potentially-help-students/

13. A. L. Turner, 'ChatGPT's Impact on Education and Student Data Privacy'. [Online]. Available: https://www.spilmanlaw.com/resource-article/chatgpts-impact-on-education-and-student-data-privacy/

14. I. Adeshola and A. P. Adepoju, 'The opportunities and challenges of ChatGPT in education', *Interact. Learn. Environ.*, vol. 32, no. 10, pp. 6159–6172, 2024.

15. W. C. Choi, J. Peng, I. C. Choi, H. Lei, L. C. Lam, and C. I. Chang, 'Improving Young Learners with Copilot: The Influence of Large Language Models (LLMs) on Cognitive Load and Self-Efficacy in K-12 Programming Education', in *Proceedings of the 2025 International Conference on Artificial Intelligence and Education (ICAIE)*, Suzhou, China, 2025.

16. C. I. Chang, W. C. Choi, and I. C. Choi, 'A Systematic Literature Review of the Opportunities and Advantages for AIGC (OpenAI ChatGPT, Copilot, Codex) in Programming Course', in *Proceedings of the 2024 7th International Conference on Big Data and Education*, 2024.

17. W. C. Choi and C. I. Chang, 'A Survey of Techniques, Key Components, Strategies, Challenges, and Student Perspectives on Prompt Engineering for Large Language Models (LLMs) in Education', 2025, *Preprints.org*.

18. S. T. Vu, H. T. Truong, O. T. Do, T. A. Le, and T. T. Mai, 'A ChatGPT-based approach for questions generation in higher education', in *Proceedings of the 1st ACM Workshop on AI-Powered Q&A Systems for Multimedia*, 2024, pp. 13–18.

19. J. Popay *et al.*, 'Guidance on the Conduct of Narrative Synthesis in Systematic Reviews: A Product from the ESRC Methods Programme', Lancaster University, Lancaster, UK, 2006.

20. OpenAI, 'GPT-5 is here'. [Online]. Available: https://openai.com/gpt-5/