

Research Article:

Navigating Artificial Intelligence (AI) Integration in English Language Teaching: Challenges, Opportunities and Future Directions

Daniel Ginting^{1*}, Rudi Hartono² and Iskandar Iskandar³

¹English Letters Study Program, Faculty of Languages, Universitas Ma Chung, Jl. Vila Puncak Tidar N-1, Malang, East Java, 65151 Indonesia

²Language Education Sciences, Postgraduate Program, Universitas Negeri Semarang, Jl. Raya Sekaran, Gunungpati, Semarang 50229, Central Java, Indonesia

³English Language Education, Postgraduate Program, Universitas Negeri Makassar, Jl. Andi Pangeran Pettarani, Makassar 90221, South Sulawesi, Indonesia

*Corresponding author: daniel.ginting@machung.ac.id

ABSTRACT

This article overviews a study on integrating generative AI tools into education. English language education is critical for AI integration due to its reliance on text-based learning, communication skills and adaptive instructional approaches. Employing a systematic literature review to analyse existing research on AI tools in English language education, it focuses on their benefits and challenges. Articles from Scopus-indexed journals and other scholarly sources published between 2021 and 2024 were reviewed. The study discusses four significant findings: the classification of generative AI, key research topics on AI, the role of AI in cognitive offloading and academic dishonesty and the phenomenon of AI-generated hallucinations. The findings highlight the need for educators to stay updated with AI advancements and adapt their teaching practices accordingly. While AI offers opportunities for personalised learning, it also raises concerns about academic integrity and the reliability of AI-generated content. The review is limited to studies published between 2021 and 2024 and may not encompass all relevant research. Additionally, the focus is primarily on English language education, which may not fully represent the impact of AI tools across other educational contexts. Future research could explore the broader implications of AI integration across different subject areas and examine its long-term effects on pedagogy and student learning outcomes.

Keywords: Artificial intelligence, AI, English language teaching (ELT) critical thinking

Published: 30 September 2025

To cite this article: Ginting, D., Hartono, R., & Iskandar, I. (2025). Navigating artificial intelligence (AI) integration in English language teaching: Challenges, opportunities and future directions. *Asia Pacific Journal of Educators and Education*, 40(2), 279–301. <https://doi.org/10.21315/apjee2025.40.2.13>

INTRODUCTION

Just recently, a colleague shared a screenshot from the website titled “Many academic articles in the journals seem to be written by ChatGPT” (Rossa, 2024). The title is quite provocative, raising academic integrity issues due to the help of Artificial Intelligence (AI). Technology development does not always go hand in hand with human readiness. One group claims that AI should be integrated into education to accelerate educational quality, emphasising its role in fostering efficiency, personalised learning and accessibility (Ng, 2021; Kaswan, 2024). AI can assist students in refining their ideas, improving their writing skills and generating insights more effectively (Jaboob, 2024). The other group, however, argues that AI poses a significant threat to impoverishing students’ cognitive abilities, leading to cognitive dependency, reducing students’ ability to think critically and independently and diminishing originality and intellectual rigour in academic work (Zhai et al., 2024). This debate highlights the need for ethical academic practices on AI use in education, ensuring that technological advancements align with academic integrity and cognitive development. This similar issue has reminded us of the story of Socrates, who once expressed his disagreement with the practice of note-taking as a new learning method used by his students to answer his questions. For Socrates, the note-taking method makes people less critically dependent on unresponsive texts, hindering critical thinking and productive dialogue (Gregorcic & Pendrill, 2023). Similarly, concerns about cognitive offloading have been raised in mathematics education, where some educators discourage early reliance on calculators (LaCour et al., 2019). The fear is that excessive dependence on calculators may weaken students’ mental math skills, problem-solving abilities and conceptual understanding of mathematics.

Emerging technologies, such as virtual reality (VR), augmented reality (AR) and AI, are gaining significant attention. One of the most discussed advancements is generative AI, which focuses on creating new and unique content by learning from large datasets. Transformer-based models, such as GPT-2 and GPT-3, pioneered AI-driven language generation (Aydın & Karaarslan, 2023). More recently, GPT-4 has demonstrated improved reasoning, accuracy and multimodal capabilities (OpenAI, 2023). The latest development, GPT-4 Turbo, offers enhanced efficiency and cost-effectiveness while maintaining high performance (OpenAI, n.d.). These advancements continue to shape AI applications in various fields, including education, research, and content creation. GPT-4 surpasses GPT-3.5-Turbo in its ability to interpret complex instructions and respond with greater accuracy, reducing the chances of generating misinformation (Hofmann et al., 2023). It also demonstrates a stronger ability to reject inappropriate or harmful requests. GPT-4 exhibits enhanced reasoning and contextual awareness.

The latest advancement, GPT-4 Turbo, is a more cost-efficient and faster version of GPT-4, making it a scalable solution for education, research and industry. The continuous evolution of generative AI highlights its expanding influence in medical training, content generation and customer engagement. Microsoft has leveraged GPT-4 to power Bing AI, enabling it to retrieve and incorporate live internet data into its responses—an advantage over ChatGPT, which relies solely on pre-trained knowledge (Hofmann et al., 2023). The swift

progress in generative AI highlights its growing influence across various fields, including education, research and industry. These models demonstrate their expanding real-world applications, reinforcing the transformative potential of AI-driven technologies.

The rapid advancement of generative AI presents both opportunities and challenges in education, particularly in language learning. While AI tools enhance teaching and assessment, their long-term impact on learner autonomy remains uncertain. Concerns over academic dishonesty and AI-generated misinformation further complicate integration. This review is crucial in understanding the evolving role of AI in education, balancing its transformative potential with the risks it poses to academic integrity and cognitive development. Moreover, this paper is also intended to provide a critical analysis of how AI-driven tools can enhance personalised learning, efficiency and accessibility while also addressing concerns such as over-reliance, diminished critical thinking and ethical misuse. By offering insights into best practices, policy considerations and responsible AI integration strategies, this paper serves as a valuable resource for educators, institutions, and policymakers seeking to navigate the complex intersection of AI and education.

ARTIFICIAL INTELLIGENCE

According to UNESCO (2021), AI is “a system which can process data and information in a way that resembles intelligent behaviour, and typically includes aspects of reasoning, learning, perception, prediction, planning or control” (p. 10). Aren’t these abilities critical competencies taught in schools/higher education? These cognitive abilities align closely with essential competencies in schools and higher education, such as critical thinking, problem-solving and decision-making. As AI continues to evolve, its integration into educational settings raises essential questions about its role in enhancing or potentially replacing human cognitive processes. One of the most advanced AI applications in education today is ChatGPT, a generative AI model designed to process and generate human-like text with remarkable accuracy and efficiency. AI, particularly generative models like ChatGPT, mirrors key cognitive skills taught in education. Its growing presence raises important questions about its role in enhancing learning while ensuring it complements rather than replaces human intelligence.

More than 13 million people used AI every day in January 2023, as reported by Hu et al. (2023). Moreover, a researcher named Ward (2023) released the results of his survey, which showed that in January 2023, around 90% of college students in the US used ChatGPT for homework, and more than 50% used it for essay writing. ChatGPT’s popularity is closely tied to its use in written evaluation products. It is not surprising that the popularity of AI models has been skyrocketing, especially among school or campus communities. The widespread adoption of AI, particularly ChatGPT, in academic settings highlights its growing influence on student learning and assessment. As its use continues to rise, it is crucial to examine both its benefits and challenges in education.

While these tools have great potential to enhance learning, there is limited research on whether educators have the necessary training and support to use them in ways that align with educational goals. Studies are needed to explore teachers' readiness, professional development and the challenges they face in adopting AI tools, ensuring that these technologies genuinely benefit students' learning experiences.

ChatGPT's Capabilities and Advancements in AI

ChatGPT is a powerful generative AI model that processes and generates human-like text. Its capabilities extend across various domains, including language understanding, problem-solving and knowledge retrieval. For example, Gao et al. (2023) asked their respondents, comprising senior educators at a university, to distinguish between abstracts created by humans and machines to prove the language quality of the text generated by ChatGPT. The results showed that the respondents had difficulty differentiating between the two abstracts. Moreover, when tested for intelligence, ChatGPT-3 and ChatGPT-4 have an IQ of 150 and a verbal-linguistic IQ of 147 (Ray, 2023). The average IQ for normal humans is around 100. Scores above 130 are considered highly intelligent. ChatGPT even outperforms similar tools like Bing Chat, Davinci, etc., showing excellent performance in various fields such as medical exams, politics, AI, science and general knowledge (Ray, 2023). As such, ChatGPT's advanced language processing and high intelligence scores demonstrate its potential to rival human cognition in various fields. Its ability to generate high-quality text and outperform similar AI models underscores its growing role in education, research and professional domains, raising important discussions about its applications and limitations.

ChatGPT-4, launched on 14 March 2023, introduced significant improvements over ChatGPT-3.5, enhancing reliability, creativity and interactivity (OpenAI, 2023). With an expanded context length, it processes longer passages in a single request, making it more efficient for complex tasks. A key advancement is its ability to handle text and visual prompts, although image-processing capabilities are not publicly available (OpenAI, 2023). Additionally, ChatGPT-4 delivers over 85% accuracy across 25 languages, including Mandarin, Polish and Swahili, and supports coding in all major programming languages, further expanding its versatility (OpenAI, 2023). Moreover, Microsoft integrated ChatGPT-4 into its Bing AI chatbot, demonstrating how generative AI enhances search functionalities, user interactions and content creation (OpenAI, 2023). Looking ahead, OpenAI plans to release ChatGPT-4.5 as an intermediate upgrade before the launch of ChatGPT-5, which is expected in early 2025 (Ism, 2025). ChatGPT-5 could bring ChatGPT to the artificial general intelligence (AGI) level, potentially making its responses indistinguishable from human communication (LazyProgrammer, 2025; Howarth, 2025). ChatGPT-4 has set new standards in AI-driven text processing, multilingual support and coding proficiency. Its integration into major platforms like Bing AI highlights its expanding influence, while future developments, such as ChatGPT-5, signal the potential emergence of AGI, further blurring the line between human and machine communication.

Table 1 compares ChatGPT's performance across various domains against other AI tools, such as Bing Chat and Davinci. It highlights specific achievements of ChatGPT in tasks like medical and legal examinations, academic and general knowledge tests and reasoning abilities (Ray, 2023). Moreover, we also notice that, when comparing its performance to human averages and other AI tools, ChatGPT indicates whether it performs better than humans in specific domains, such as the SAT exam, general knowledge tests and IQ assessments. The comparison in Table 1 highlights ChatGPT's exceptional performance in multiple domains, outperforming other AI tools and even human averages in certain standardised assessments. Its success in medical, legal and academic evaluations demonstrates its advanced reasoning and problem-solving abilities. These findings emphasise ChatGPT's growing influence in education and professional fields, raising important discussions about its potential applications, limitations and ethical considerations in AI-assisted learning and decision-making.

Table 1. Performance comparison of ChatGPT with other AIs

Field	Achievement	Better than human average?	Tool	Testing year
Japan: National Medical Licensure Examination	Bing Chat would achieve 78% [above cut-off grade of 70%], ChatGPT would achieve 38%.	Yes	Bing Chat	2023
Spanish medical examination (MIR)	Bing Chat would achieve 93%, ChatGPT would achieve 70%, both above cut-off grade.	Yes	Bing Chat	2023
Cover of TIME magazine	ChatGPT made the 27/Feb/2023 cover of TIME magazine.	Yes	ChatGPT	2023
Jurisprudence/legal rulings	ChatGPT helps a judge with a verdict (Colombia).	NA	ChatGPT	2023
Politics	ChatGPT writes several Bills (USA).	NA	ChatGPT	2023
MBA	ChatGPT would pass an MBA degree exam at Wharton (UPenn).	Yes	ChatGPT	2023
Accounting	GPT-3.5 would pass the US CPA exam.	Yes	text-davinci-003	2023
Legal	GPT-3.5 would pass the bar in the US.	Yes	text-davinci-003	2022
Medical	ChatGPT would pass the United States Medical Licensing Exam (USMLE).	Yes	ChatGPT	2022

(Continued on next page)

Table 1. (Continued)

Field	Achievement	Better than human average?	Tool	Testing year
IQ (fluid/apptitude)	ChatGPT outperforms college students on the Raven's Progressive Matrices aptitude test.	Yes	text-davinci-003	2022
AWS certificate	ChatGPT would pass the AWS Certified Cloud Practitioner exam.	Yes	ChatGPT	2022
IQ (verbal only)	ChatGPT scores IQ = 147, 99.9th %ile.	Yes	ChatGPT	2022
SAT exam	ChatGPT scores 1020/1600 on SAT exam.	Yes	ChatGPT	2022
General knowledge	GPT-3 would beat IBM Watson on Jeopardy! questions.	Yes	davinci	2021
IQ (Binet-Simon Scale, verbal only)	GPT-3 scores in 99.9th %ile (estimate only).	Yes	davinci	2021
General knowledge	GPT-3 outperforms average humans on trivia.	Yes	davinci	2021
Reasoning	GPT-3 would pass the SAT Analogies subsection.	Yes	davinci	2020

METHODOLOGY

In this review, AI tools in education are defined as technologies utilising machine learning and generative models to support teaching, learning and assessment (Holmes et al., 2019). This study adopts a systematic literature review methodology to examine the application of AI in English language education, focusing on both its benefits and challenges.

LITERATURE SEARCH AND PROCEDURE

The researchers conducted a literature search of peer-reviewed journal articles and reports from recognised educational and technological organisations. The search was limited to publications between 2021 and 2024 to capture recent developments in generative AI and its applications in English language education. Articles were retrieved from Scopus, ProQuest and EBSCOhost, selected for their credibility and scholarly standards (Knopf, 2006). Other databases were excluded.

To identify relevant studies, different combinations of keywords were used, including “AI in education”, “generative AI”, “AI and academic integrity”, “ChatGPT in language teaching” and “AI in assessment”. The initial search yielded 52 publications. These were screened by title, abstract and keywords, reducing the selection to 32 articles. Further

refinement followed the classification method of Ali and Petersen (2014), categorising articles as relevant, uncertain, or irrelevant. Studies unrelated to education, non-academic AI applications, or published in languages other than English were excluded. The reviewers reached 92% agreement. Discrepancies were resolved through discussion, finalising 28 articles for review. Thematic analysis of the selected studies revealed four key themes. First, the classification of AI tools and applications identified various technologies, including text-to-text, text-to-image and AI-assisted assessment tools. Second, the role of AI in teaching methodologies highlighted its impact on personalised learning and instructional strategies. Third, AI-based assessment and academic integrity emerged as a critical theme, focusing on AI-generated content in student work and associated ethical considerations. Lastly, the credibility of AI-generated content is examined in relation to the risks of misinformation and AI hallucinations in educational settings.

The step-by-step article selection process and inclusion rates are summarised in Table 2.

Table 2. Article selection process and inclusion rates in the systematic review

Stage	Description	Total articles	Articles excluded (%)	Articles included (%)
Initial Search	Articles retrieved using keyword combinations.	52	- (0%)	52 (100%)
Title, Abstract, and Keyword Screening	Articles excluded due to irrelevance.	32	20 (38.5%)	32 (61.5%)
Refinement Process	Classification based on the Ali and Petersen (2014) method.	N/A	4 (7.7%)	28 (53.8%)
Final Selection	Articles finalised for review after discussion.	28	- (0%)	28 (53.8%)
Reviewer Agreement	Percentage of agreement among reviewers.	92%	-	-

To ensure alignment with the inclusion criteria, two independent reviewers conducted a thorough analysis of the full texts using a predefined evaluation framework. This framework provided structured guidelines for assessing the relevance of each study based on three key criteria: the integration of AI tools in English language teaching and learning, AI-based assessment methods and their impact on academic integrity and challenges such as AI hallucinations and content credibility. The reviewers employed a standardised coding scheme, which consisted of a structured rubric with predefined categories and criteria to guide their evaluation of the selected studies. The coding scheme assessed each article's relevance, methodological rigour and thematic focus. Key categories included AI tool type, application in English language education, AI-based assessment methods and ethical concerns like academic integrity and misinformation. A three-tier classification system—relevant, uncertain and irrelevant—adapted from Ali and Petersen (2014) ensured

consistency. Discrepancies were resolved through discussion to maintain inter-rater reliability and reduce bias.

DATA ANALYSIS

The researchers employed a directed qualitative content analysis approach, following the method outlined by Assarroudi et al. (2018). This approach begins with predefined categories derived from existing literature, which are refined as the analysis progresses. The primary aim was to systematically examine AI applications in English language education and categorise findings according to key themes. The data analysis process followed a structured approach using a predefined coding framework with three key components: thematic categories based on prior research, criteria for assessing methodological rigour, and an open coding process for identifying emerging themes. The analysis was conducted in several stages.

First, in the initial coding stage, all selected articles were carefully reviewed, and relevant excerpts were assigned to predefined categories, including AI tools, instructional methodologies, assessment practices and ethical concerns. Next, during the refinement and expansion stage, these predefined categories were continuously reassessed, and new subcategories were added whenever novel insights emerged from the data. In the thematic synthesis stage, the coded data were synthesised into broader themes, enabling meaningful comparisons across different studies. Finally, a reliability check was conducted by having a second reviewer independently analyse a subset of the articles. Any discrepancies in coding were discussed and resolved collaboratively to ensure consistency and accuracy in the findings.

RESULTS

Four major points have been deduced about the domain of generative AI and its uses. First, an attempt is made to explain the kinds of generative AI tools being developed and what this may mean for teaching and personalised learning. We then present new developments in NLP and how they will likely affect language education. It then addresses the increasingly important issue of academic dishonesty perpetuated by AI and the strategies being developed as a counteractive measure. Finally, it considers the so-called AI hallucinations, where AI would come up with information that is credible but wrong and all the implications thereof concerning the credibility of AI-generated content.

Generative AI Classifications

Generative AI has revolutionised content creation by automating and enhancing various creative processes. As AI technologies advance, they have been categorised based on their ability to generate different types of content, catering to diverse needs in education, media, and technology. These tools not only improve efficiency but also expand accessibility,

enabling users to produce high-quality outputs with minimal effort. These generative AI tools have been categorised based on their functionality, specifically the type of content they generate: Text-to-Image (T2I), Text-to-Video (T2V), Text-to-Audio (T2A) and Text-to-Text (T2T). T2I models enable rapid image generation, reducing reliance on manual design processes. T2V tools facilitate video content creation without extensive production resources. T2A enhances accessibility through automated speech synthesis, while T2T streamlines writing and translation tasks. AI continues to evolve with capabilities such as Text-to-Motion (T2M), Text-to-Code (T2C) and Brain-to-Text (B2T), expanding its applications across creative, educational and technical domains. The following classification (Table 3) helps us in understanding the diverse applications of generative AI across creative, educational and technical domains.

Table 3. Classification of the AI tools

Classification	Tool name and web page address	Operation mode
Text-to-Image (T2I)	DALL-E 2 (https://openai.com/product/dall-e-2)	Generation
	Stable Diffusion (https://stability.ai/)	Generation
	Craiyon (https://www.craiyon.com/)	Generation
	Jasper (https://www.jasper.ai)	Generation
	Imagen (https://imagen.research.google/)	Generation
	MidJourney (https://www.midjourney.com)	Generation
	NightCafe (https://nightcafe.studio/)	Generation
	GauGAN2 (https://gpt3demo.com/apps/gaugan2-by-nvidia)	Generation
	Wombo (https://www.w.ai/)	Generation
	Wonder (https://www.wonder-ai.com/)	Generation
	Pixray-test2image (https://pixray.gob.io/)	Generation
Text-to-Video (T2V)	Neural.love (https://neural.love/)	Classification, Generation, Transformation
	Runway (https://runwayml.com/)	Generation
	Fliki (https://fliki.ai/)	Generation
	Synthesia (https://www.synthesia.io/)	Generation
	MetaAI (https://ai.facebook.com/)	Generation
	Google AI (https://ai.google/)	Generation
Text-to-Audio (T2A)	Phenaki (https://phenaki.video/)	Generation
	Play.ht (https://play.ht/)	Generation
	Murf.AI (https://murf.ai/)	Generation
	Resemble.AI (https://resemble.ai)	Generation
	WellSaid (https://wellsaidlabs.com/)	Generation
	Descript (https://www.descript.com/)	Classification, Generation, Transformation

(Continued on next page)

Table 3. (Continued)

Classification	Tool name and web page address	Operation mode
Text-to-Text (T2T)	Simplified (https://simplified.com/)	Generation
	Jasper (https://www.jasper.ai)	Generation
	Frase (https://www.frase.io/)	Transformation
	EleutherAI (https://www.eleuther.ai/)	Generation
	Requestor (https://requestory.com/)	Generation
	Grammarly (https://www.grammarly.com)	Transformation
	Copy.ai (https://www.copy.ai/)	Generation
	MarketMuse (https://www.marketmuse.com/)	Generation
	AO21labs (https://www.ai21.com/)	Transformation
	HubSpot (https://www.hubspot.com/)	Classification
	InterKit (https://app.inferkit.com)	Generation
	GooseAI (https://goose.ai/)	Generation
	ResearchAI (https://research-ai.io/)	Generation
	Writesonic (https://writesonic.com/)	Generation
	Co: here (https://cohere.ai/)	Classification
	CHIBI (https://chibi.ai/)	Generation
	Ideas AI (https://ideasai.com/)	Generation
	Copysmith (https://copysmith.ai/)	Generation
	Flowrite (https://www.flowrite.com/)	Generation
	NICHES\$\$ (https://nichess.com/)	Generation
	Sudowrite (https://www.sudowrite.com/)	Generation
	Ideasbyai (http://ideasbyai.com/)	Generation
	Text.cortex (https://textcortex.com/)	Classification, Transformation
	OpenAI GPT3 (https://openai.com/blog/chatgpt)	Generation
	AISEO Blog Idea Generator (https://aiseo.ai/templates/blog-idea-generator.html)	Generation
	Rytr AI https://rytr.me/	Generation
	PepperType AI https://www.peppertype.ai/	Generation
	Kafkai AI https://kafkai.com/en/	Generation
	Texta AI https://texta.ai/dark	Generation
	Anyword AI https://anyword.com/	Generation
	DeepL Write AI https://www.deepl.com/write	Generation
	Perplexity https://www.perplexity.ai/	Generation
	Elicit https://elicit.org/	Generation
Text-to-Motion (T2M)	TREEInd. (https://www.tree.industries/)	Generation
	MDM: Human Motion Diffusion Model (https://guytevet.github.io/mdm-page/)	Generation
Text-to-Code (T2C)	Replit Generate code(https://docs.replit.com/power-ups/ghostwriter/generate-code)	Generation
	Github Copilot (https://github.com/features/copilot)	Generation

(Continued on next page)

Table 3. (Continued)

Classification	Tool name and web page address	Operation mode
Text-to-NFT (T2N)	LensAI (https://lens-ai.com/)	Generation
Text-to-3D (T2D)	DreamFusion (https://dreamfusion3d.github.io/)	Generation
	Clip-Mesh (https://www.nasir.lol/clipmesh)	Generation
	GET3D (https://nv-tlabs.github.io/GET3D/)	Generation
Audio-to-Text (A2T)	Descript (https://www.descript.com/tools/audio-text)	Classification, Generation, Transformation
	AssemblyAI (https://www.assemblyai.com/)	Transformation
	Whisper (OpenAI) (https://github.com/openai/whisper)	Transformation
	Elsa Speaks https://elsaspeak.com/en/	Classification, Generation
Audio-to-Audio (A2A)	AudioLM (https://google-research.github.io/seanet/audiolm/examples/)	Transformation
	VOICEMOD (https://www.voicemod.net/)	Transformation
Brain-to-Text (B2T)	Speech from the brain (Meta AI) (https://ai.facebook.com/blog/ai-speech-brain-activity/)	Generation
	Non-invasive brain recordings (https://hal.science/hal-03808317/document)	Generation
Image-to-Text (I2T)	Neural.love (https://neural.love/)	Classification, Generation, Transformation
	GPT-2 x Image Captions (https://huggingface.co/nlpconnect/vit-gpt2-image-captioning)	Transformation

The classification of generative AI tools significantly impacts English language teaching by enhancing instructional methods and accessibility. T2I and T2V tools provide visual aids and multimedia content, making lessons more engaging. T2A improves listening and pronunciation skills through synthesised speech, while T2T tools assist with writing and grammar correction, fostering independent learning. However, incorporating these technologies requires teachers to develop new skills in using AI in educational contexts (Aydın & Karaarslan, 2023). Without adaptation, they risk falling behind in educational innovation, making their teaching methods less effective and engaging for students who are increasingly exposed to AI-driven learning tools. Furthermore, teachers may struggle to personalise instruction, automate repetitive tasks and provide diverse learning resources, potentially widening the gap between traditional and modern teaching approaches. Additionally, students might rely on AI independently without proper guidance, leading to issues such as academic dishonesty, over-reliance on AI-generated content, and lack of critical thinking skills. In conclusion, teachers must learn how to align their instructions with educational goals. They need to modify their instructional strategies, teaching methods, and classroom activities to incorporate generative AI tools into English language teaching effectively.

State-Of-The-Art Performance: The Most Advanced Technology in Natural Language Processing

The analysis of research trends from 2021 to 2024 reveals a strong emphasis on the integration of AI tools in English language education. Among the 28 reviewed articles, the majority focus on three key areas: chatbots for language help (35.7%), AI-assisted teaching methods (32.1%) and AI-based assessment tools (21.4%). This distribution is illustrated in Figure 1, which shows the relative proportion of research topics on AI in education. This dominance highlights the growing academic interest in AI-driven educational solutions that enhance teaching effectiveness, student engagement and assessment practices. Beyond these core topics, the remaining research articles explore various other aspects of AI in education, each contributing 3.6% to the total. These include AI and emerging technologies, perceptions and attitudes towards AI, AI for writing and speaking skills, ethical and responsible AI use, research trends and literature reviews, technology integration and awareness, and the broader application of AI in language learning and teaching. While these topics receive less attention compared to the primary themes, they play a crucial role in shaping discussions around AI’s ethical, technological, and pedagogical implications.

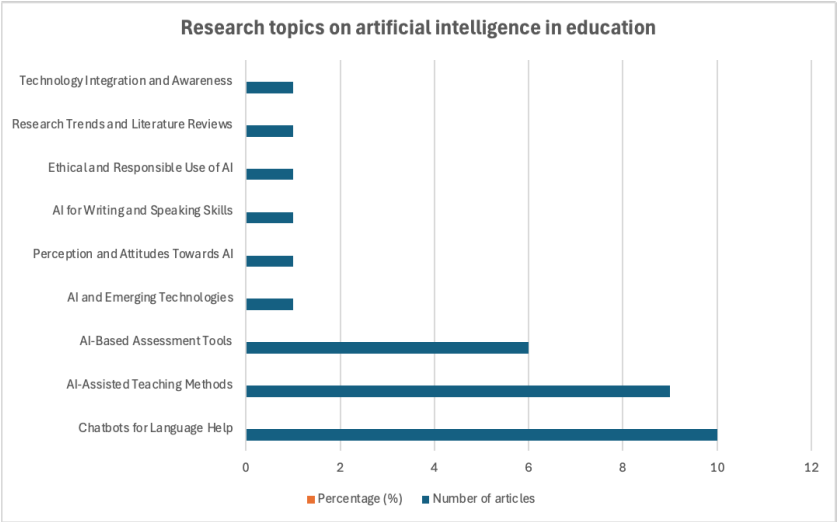


Figure 1. Research topics on AI in education

The research trend from 2021 to 2024 highlights AI’s growing influence on English language teaching, with implications for instructional practices, assessment methods and learner engagement. The emphasis on chatbots, AI-assisted teaching and AI-based assessments suggests a shift towards more technology-integrated pedagogy (Cooper, 2021). This trend encourages personalised learning experiences, real-time feedback and adaptive instruction, making education more accessible and efficient (Bowman, 2023). However, the rapid adoption of AI tools also raises concerns about their effectiveness, ethical use and impact on critical thinking skills. Future research must focus on evaluating AI’s pedagogical

value, addressing biases in AI-generated content and ensuring that AI complements rather than replaces human instruction. Additionally, the long-term effects of AI on student learning outcomes require further empirical investigation to determine best practices for its integration into language education.

AI and Academic Dishonesty

A crucial consideration is the issue of academic integrity that ensures students complete their work honestly, without plagiarism, cheating, or misrepresentation. The rise of generative AI tools presents challenges to academic integrity, as students may use AI-generated content in assignments, essays, or exams without proper attribution or critical engagement. For example, students might submit an AI-generated essay without making significant edits, presenting it as their original work. Similarly, they could use AI-powered paraphrasing tools to alter existing content without truly understanding or engaging with the material. These practices undermine learning and make it difficult for teachers to assess students' actual knowledge and skills.

Several researchers (Ahsan et al., 2022) have identified reasons why students engage in academic dishonesty with AI: they ultimately assign AI to thoroughly do their assignments without giving sufficient effort to prepare their work (Sullivan et al., 2023; Cotton et al., 2023). Short speaking is a complete cognitive offloading.

1. The stakes of exams/tests because they determine graduation or rewards (high stakes),
2. Careless seating arrangements during exams,
3. Lack of supervision of assignments from teachers/professors,
4. Painful past experiences of failure,
5. Previous instances of cheating,
6. Low regard for honest efforts in achieving academic excellence,
7. Orientation towards acquiring grades rather than quality work through personal effort.

Beasley (2014) asked some students what could make them stop cheating. They answered that we need more time, resources and skills to achieve the desired result, better time management and less impact of mistakes on grades. Noam Chomsky referred to the practice of predominantly using ChatGPT to complete academic tasks as “high-tech plagiarism” (Stewart, 2023a, para. 2). Weissman (2023) stated that allowing these cheating practices is the end of the teaching profession.

Efforts have been made to counter this academic dishonesty, such as developing “AI detectors” (Fowler, 2023) embedded in the Turnitin application. A professor earlier this year thwarted his entire class with allegations that his student's work had been produced by AI based on Turnitin information (Agomuo, 2023). A few months later, around June, Turnitin declared its AI detector unreliable (Fowler, 2023) or inaccurate information (Bowman, 2023). Along with the rush to find anti-cheating technology detection methods, some developers have also developed anti-detection AI models. Furthermore, students share tactics to avoid these AI detectors through social media platforms like TikTok (Haensch et al., 2023).

Some researchers propose several strategies for anti-cheating technology:

1. Applying watermarking to content generated by ChatGPT (Eysenbach, 2023).
2. Using ChatGPT to identify ChatGPT material (Khalil & Er, 2023) to verify the content's origin and check for similarities.
3. Measuring perplexity (the model's prediction rate for the next word in a word sequence) and Burstiness (diversity patterns in sentence structure). The higher the perplexity, the more likely a human wrote it. Similarly, the higher the burstiness, the more human-like it is (Bowman, 2023; Gillard & Rorabaugh, 2023).

Traditional assessment methods may be ineffective in fostering genuine learning, as students are often more focused on grades and external rewards than personal growth. The emphasis on high-stakes testing and rigid grading structures may unintentionally push students toward dishonest practices (Ginting & Saukah, 2016). Moreover, the role of educators and institutions is questioned (Kabilan & Veratharaju, 2013). Weak supervision, inadequate guidance and poorly structured assessments create an environment where cheating becomes an easy option. Students are not inherently dishonest but are often driven to cheat due to systemic pressures, lack of resources and fear of failure. This implies that better time management, skills development and reduced grade penalties could deter cheating, indicating that educational reform, rather than just stricter enforcement, is necessary. Finally, the warnings from Chomsky (as cited in Stewart, 2023b) and Weissman (2023) imply that unchecked AI misuse could erode the value of education and even threaten the teaching profession itself. This raises concerns about the long-term implications of technology in academia and the need for ethical AI integration in learning.

Artificial Intelligence Hallucinations

AI has made remarkable advancements in natural language processing, enabling large language models (LLMs) to generate human-like text. However, despite their impressive capabilities, these models are prone to a phenomenon known as hallucination. Hallucination happens when AI produces outputs that, while seemingly plausible, deviate from user input (Adlakha et al., 2024), previously generated context (Liu et al., 2022), or factual knowledge (Min et al., 2023; Muhlgay et al., 2023; Li et al., 2023a), which significantly undermines the reliability of LLMs in real-world scenarios (Kaddour et al., 2023).

Zhang et al. (2023) categorise hallucination within the context of LLMs as follows:

1. Input-conflicting hallucination, where LLMs generate content that deviates from the source input provided by users.
2. Context-conflicting hallucination, where LLMs generate content that conflicts with previously generated information by itself.
3. Fact-conflicting hallucination, where LLMs generate content that is not faithful to established world knowledge.

AI large language models do not possess “knowledge” in the conventional sense; they do not store or retrieve information or search the web like a search engine. Instead, they excel

at predicting the next word in a sequence based on prior learning (Cooper, 2021). As a result, ChatGPT often exhibits “inconsistent factual accuracy” (Vincent, 2022), and it does not ensure that its generated content is accurate, reliable, or valid (Hutson, 2021). In machine learning terms, this can lead to “hallucinations”, where the model produces plausible-sounding but incorrect information (Welborn, 2023).

The finding highlights the fundamental limitations of large language models (LLMs) in producing accurate and reliable information. It emphasises that AI-generated outputs can deviate from user input, contradict prior context, or present misinformation as fact. A key implicit message is that LLMs, despite their advanced capabilities, do not “know” information in the traditional sense. They generate text based on probabilistic predictions rather than factual retrieval, making them prone to inconsistencies and errors. This finding suggests that while AI can enhance content creation and automate tasks, it requires human oversight to verify accuracy.

The real-world implications of AI hallucination are worth noticing, particularly in fields that rely on factual precision, such as education, journalism, and scientific research. It raises ethical concerns about misinformation and the need for critical evaluation of AI-generated content. AI should be seen as a tool to assist human intelligence rather than replace it, reinforcing the importance of responsible AI use and continuous model refinement.

DISCUSSION

Integrating generative AI tools into English language education offers both opportunities and challenges, necessitating that educators adapt their instructional strategies to effectively incorporate AI-generated content. Utilising T2T tools such as ChatGPT or Grammarly can shift traditional grammar instruction towards AI-assisted writing feedback, enabling students to refine their work based on AI suggestions complemented by teacher guidance. AI-generated feedback can enhance students’ writing performance, providing immediate, individualised suggestions that help improve linguistic skills (Gayed et al., 2022). Similarly, T2I tools like DALL-E 2 can enrich storytelling exercises by generating visual prompts, requiring educators to guide students in analysing these AI-generated visuals and aligning them with lesson objectives (Oppenlaender, 2022). This approach not only fosters creativity but also aids in comprehension and engagement. Moreover, these AI tools facilitate personalised learning experiences by offering tailored instructional materials that cater to individual student needs, thereby enhancing the overall learning process (Alharbi, 2024).

However, AI-powered tools like ChatGPT have sparked debates over their ethical implications, raising concerns among educators regarding academic integrity and responsible AI use (García-Peñalvo, 2023). Students may misuse AI for contract cheating, plagiarism and unauthorised assistance (Ahsan et al., 2022). Noam Chomsky has referred to the use of AI like ChatGPT for academic tasks as “high-tech plagiarism” (Stewart, 2023a), emphasising the potential threat it poses to the integrity of education. Weissman (2023) warns that allowing such practices could spell the end of the teaching profession. Ahsan

et al. (2022) have identified key factors that drive students to misuse AI, such as the high stakes of exams, lack of supervision and previous experiences of failure. This phenomenon, often described as cognitive offloading, where students delegate their academic tasks to AI, undermines the essence of personal effort and intellectual growth (Sullivan et al., 2023; Cotton et al., 2023).

Various countermeasures have been proposed to combat the integrity problems: applying watermarking to AI-generated content, using AI to identify AI material and measuring perplexity and burstiness to distinguish human-written from AI-generated text (Eysenbach, 2023; Khalil & Er, 2023; Bowman, 2023). Despite these efforts, students have found ways to evade detection, often sharing tactics on social media platforms like TikTok (Haensch et al., 2023). Unfortunately, the reliability of these detectors has been questioned, as highlighted by incidents where entire classes were falsely accused of AI-generated submissions (Agomuoh, 2023; Fowler, 2023).

The emergence of large language models (LLMs) continuously improves (Min et al., 2023). Banning the use of AI clearly does not solve the problem. Addressing academic dishonesty issues in the age of AI requires a multifaceted approach, including updating academic policies, enhancing supervision and fostering a culture of integrity and critical thinking among students. Educators, policymakers, and technologists must collaborate in developing effective solutions that uphold the principles of academic honesty (Cotton et al. 2023). Implementing AI literacy programs could also help students critically assess AI-generated content and use these tools constructively rather than relying passively (Lodge, 2023; Jaboob, 2024).

Another primary concern is the potential for AI-generated hallucinations, where AI produces seemingly plausible but incorrect or misleading information, resulting in academic misinformation. AI's inability to possess conventional "knowledge" and its reliance on predicting the next word in a sequence (Cooper, 2021) contribute to inconsistent factual accuracy (Vincent, 2022). This issue can lead to situations where AI produces outputs that appear coherent but are fundamentally flawed (Hutson, 2021; Welborn, 2023). AI hallucination issues are categorised into input-conflicting, context-conflicting and fact-conflicting hallucinations, each presenting unique challenges to the integrity and trustworthiness of AI outputs (Zhang et al., 2023). This issue is especially critical in scientific writing and academic research, where accuracy, reliability and source credibility are paramount. AI-generated misinformation can lead to the spread of false claims, the distortion of established knowledge and the misinterpretation of scientific findings (Min et al., 2023). Moreover, researchers who unknowingly rely on hallucinated data may introduce errors into the academic discourse, compromising the validity of their work and undermining public trust in scholarly research (Kaddour et al., 2023; Alkaissi & McFarlane, 2023). Researchers must critically evaluate AI outputs by cross-checking information against credible academic sources, such as peer-reviewed journals and reputable databases (Min et al., 2023). Since AI lacks true understanding and fact-verification capabilities (Cooper, 2021), it should be used as a supplementary tool rather than a primary source of knowledge, requiring scholars to interpret and validate generated content before integration

(Vincent, 2022). Transparency is also essential, as researchers should disclose AI usage in their work to ensure awareness of potential limitations and the extent of human oversight (Hutson, 2021). Furthermore, universities and academic institutions should establish clear guidelines on responsible AI usage, incorporating training programs to educate scholars on identifying and mitigating AI hallucinations (Kaddour et al., 2023).

The exploration of AI in language teaching from 2021 to 2024 has led to significant discussions on its potential impact on various educational stakeholders. Researchers have particularly focused on chatbots for language learning assistance, AI-assisted teaching methods and AI-based assessment tools, reflecting the belief that AI can enhance educational practices (Okonkwo & Ade-Ibijola, 2021). However, while these technologies offer promising advantages, researchers must continue investigating their effectiveness across different educational settings and their long-term impact on learning outcomes (Eysenbach, 2023). A systematic review of AI applications in language teaching highlights the growing academic interest in AI's role in education, emphasising the need for further empirical studies to ensure AI's meaningful integration into pedagogical frameworks (Zawacki-Richter et al., 2019).

Presently, research primarily examines AI's role in improving learning efficiency, engagement, and personalised instruction. Studies have shown that AI-powered chatbots can provide personalised learning experiences, assisting students with homework and study tasks, thereby enhancing engagement and understanding (Okonkwo & Ade-Ibijola, 2021). However, there is a need for empirical studies assessing AI's long-term impact on student outcomes and its effectiveness across diverse educational settings (Zawacki-Richter et al., 2019). Future research should address ethical concerns, including academic integrity, data privacy, and the biases embedded in AI models (Eysenbach, 2023). The role of AI in teacher-student interaction also requires further exploration to ensure that technology complements rather than replaces human instruction. Additionally, the accessibility of AI tools is a critical issue, as disparities in technological infrastructure may limit equitable AI integration in education (Zawacki-Richter et al., 2019). As AI continues to evolve, research must investigate adaptive AI-driven curricula, real-time feedback mechanisms and the psychological effects of AI reliance on students' cognitive and critical thinking skills (Okonkwo & Ade-Ibijola, 2021). Addressing these concerns will be essential for the responsible and effective implementation of AI in language education.

CONCLUSION

The rapid advancement of generative AI technologies from 2021 to 2024 has significantly transformed language teaching and learning. The classification of AI tools into categories such as T2I, T2T, T2A, and others underscores their diverse capabilities, enabling educators to enrich instructional methods through engaging multimedia content and personalised learning experiences. However, while these classifications highlight AI's potential, the successful integration of AI into education depends on how various stakeholders—researchers, educators and students—navigate its opportunities and challenges.

AI-powered tools offer educators new opportunities to enhance language learning through multimodal content delivery, adaptive learning pathways and personalised feedback. However, effective implementation requires professional development and clear guidelines to ensure responsible AI use. Institutions should provide training to help educators integrate AI into their teaching while addressing issues related to academic integrity. The study also contributes to ongoing discussions in AI-assisted language learning by categorising AI tools and examining their impact. It highlights the need for a refined theoretical framework that addresses AI's role in constructivist and communicative approaches to language learning. The findings suggest that AI's effectiveness varies based on the alignment between technological capabilities and pedagogical objectives. Further empirical studies are needed to evaluate AI's long-term impact on student engagement, critical thinking, and learning outcomes. Researchers should examine the ethical dimensions of AI use in education, such as issues related to bias, academic dishonesty, and over-reliance on technology. Future studies should also explore the effectiveness of AI-powered assessments and their role in formative evaluation.

Each AI category presents unique opportunities and challenges. Text-to-image AI enhances visual learning by providing customised illustrations for teaching materials. This feature is particularly useful for abstract concepts and vocabulary acquisition, and educators can integrate these tools into classroom presentations and digital storytelling to improve student engagement. However, content accuracy and potential bias in AI-generated images must be addressed. Text-to-video AI enables the creation of instructional videos tailored to specific learning needs, which can enhance comprehension for auditory and visual learners. Teachers can use these tools to create interactive lessons, but they must ensure that AI-generated content aligns with pedagogical goals and does not misrepresent information. Text-to-audio AI facilitates listening comprehension exercises and language practice by exposing learners to diverse accents and speech patterns. While it promotes linguistic diversity, challenges such as unnatural intonation and incorrect stress patterns require careful selection of AI-generated content. Other AI applications, including chatbots and grammar checkers, assist in language learning by providing instant feedback and fostering independent learning. However, students should be trained to critically evaluate AI-generated responses to avoid over-reliance on technology.

Despite its contributions, this study has several limitations. The analysis is based on a limited selection of AI tools, which may not fully capture the rapidly evolving landscape of AI in education. Additionally, the study does not include extensive empirical data on AI's effectiveness across different learner profiles, making generalisation difficult. Ethical concerns such as data privacy and academic dishonesty also require further exploration. Future research should address these gaps by conducting large-scale experimental studies. In addition, future research should also investigate the ethical and cognitive implications of AI-assisted language learning, particularly regarding academic integrity and critical thinking. Exploring AI's role in fostering student autonomy and metacognition in language learning will provide deeper insights into how these technologies influence learning behaviour.

In conclusion, while AI presents transformative opportunities for language education, its integration must be approached thoughtfully. By addressing pedagogical, theoretical and ethical considerations, educators and researchers can maximise AI's benefits while mitigating its risks, ensuring that AI remains a valuable tool rather than a substitute for human instruction.

ACKNOWLEDGEMENTS

The authors would like to thank colleagues and professors from Universitas Negeri Semarang and Universitas Negeri Makassar for their valuable collaboration and academic support during the completion of this study. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCES

- Adlakha, V., BehnamGhader, P., Lu, X. H., Meade, N., & Reddy, S. (2024). Evaluating correctness and faithfulness of instruction-following models for question answering. *Transactions of the Association for Computational Linguistics*, 12, 681–699. https://doi.org/10.1162/tacl_a_00667
- Agomuoh, F. (2023, March 14). The best ChatGPT alternatives (according to ChatGPT). *DigitalTrends*. <https://www.digitaltrends.com/computing/the-best-chatgpt-alternatives-according-to-chatgpt/>
- Ahsan, K., Akbar, S., & Kam, B. (2022). Contract cheating in higher education: A systematic literature review and future research agenda. *Assessment & Evaluation in Higher Education*, 47(4), 523–539. <https://doi.org/10.1080/02602938.2021.1931660>
- Alharbi, M. (2024). The role of artificial intelligence in advancing English as a foreign language teaching at Saudi universities. *World Journal on Educational Technology: Current Issues*, 16(3), 181–200. <https://doi.org/10.18844/wjet.v16i3.9311>
- Ali, N. B., & Petersen, K. (2014). Evaluating strategies for study selection in systematic literature studies. In *Proceedings of the 8th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM '14)*. (Article 45, pp. 1–4). New York: Association for Computing Machinery. <https://doi.org/10.1145/2652524.2652557>
- Alkaissi, H., & McFarlane, S. I. (2023). Artificial hallucinations in ChatGPT: Implications in scientific writing. *Cureus Journal of Medical Science* 15(2): e35179. <https://doi.org/10.7759/cureus.35179>
- Assarroudi, A., Heshmati Nabavi, F., Armat, M. R., Ebadi, A., & Vaismoradi, M. (2018). Directed qualitative content analysis: The description and elaboration of its underpinning methods and data analysis process. *Journal of Research in Nursing*, 23(1), 42–55. <https://doi.org/10.1177/1744987117741667>
- Aydın, Ö., & Karaarslan, E. (2023). Is ChatGPT leading generative AI? What is beyond expectations? *Academic Platform Journal of Engineering and Smart Systems*, 11(3), 118–134. <https://doi.org/10.21541/apjess.1293702>

- Beasley, E. M. (2014). Students reported for cheating explain what they think would have stopped them. *Ethics & Behavior*, 24(3), 229–252. <https://doi.org/10.1080/10508422.2013.845533>
- Bowman, E. (2023, January 9). A college student created an app to tell whether AI wrote an essay. *NPR*. <https://www.npr.org/2023/01/09/1147549845/gptzero-ai-chatgpt-edward-tian-plagiarism>
- Cooper, K. (2021, November 1). *OpenAI GPT-3: Everything you need to know*. Springboard. <https://www.springboard.com/blog/data-science/machine-learning-gpt-3-open-ai/>
- Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2023, March 13). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *EdArXiv*. <https://doi.org/10.35542/osf.io/mrz8h>
- Eysenbach, G. (2023). The role of ChatGPT, generative language models, and artificial intelligence in medical education. *JMIR Medical Education*, 9, e46885. <https://doi.org/10.2196/46885>
- Fowler, G. A. (2023, August 7). AI is acting ‘pro-anorexia’, and tech companies are not stopping it. *Washington Post*. Retrieved from <https://link.gale.com/apps/doc/A759935629/AONE?u=anon~6a1f13d5&sid=googleScholar&xid=fb70a03f>
- Gao, C. A., Howard, F. M., Markov, N. S., Dyer, E. C., Ramesh, S., Luo, Y., Pearson, A. T., & Knight, K. R. (2023). Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. *npj Digital Medicine*, 6, 75. <https://doi.org/10.1038/s41746-023-00819-6>
- García-Peñalvo, F.J. (2023). The perception of artificial intelligence in educational contexts after the launch of ChatGPT: Disruption or panic? *Education in the Knowledge Society*, 24, e31279. <https://doi.org/10.14201/eks.31279>
- Gayed, J. M., Carlon, M. K. J., Oriola, A. M., & Cross, J. S. (2022). Exploring an AI-based writing assistant’s impact on English language learners. *Computers & Education: Artificial Intelligence*, 3, 100055. <https://doi.org/10.1016/j.caeai.2022.100055>
- Gillard, C., & Rorabaugh, P. (2023, February 3). *You are not going to like how colleges respond to ChatGPT*. Slate. <https://slate.com/technology/2023/02/chat-gpt-cheating-college-ai-detection.html>
- Ginting, D., & Saukah, A. (2016). Tests of writing in the school examination in upper secondary schools. *SAGE Open*, 6(4), 1–7. <https://doi.org/10.1177/2158244016673130>
- Gregorcic, B., & Pendrill, A.-M. (2023). ChatGPT and the frustrated Socrates. *Physics Education*, 58(3), 035021. <https://doi.org/10.1088/1361-6552/acc299>
- Haensch, A. C., Ball, S., Herklotz, M., & Kreuter, F. (2023). Seeing ChatGPT through students’ eyes: An analysis of TikTok data. In *2023 Big Data Meets Survey Science (BigSurv)* (pp. 1–8). IEEE. <https://doi.org/10.1109/BigSurv59479.2023.10486710>
- Hofmann, R., Wimmer, C., Schlecht, J., Eskofier, B. M., & Stanzel, F. (2023). *ChatGPT, Bing, and medical education: Evaluating the performance of AI models on the German Medical State Examination of 2022*. *JMIR Medical Education*, 9, e46482. <https://doi.org/10.2196/46482>
- Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign. <https://unesdoc.unesco.org>

- Howarth, J. (2025, August 11). ChatGPT-5 has arrived: What OpenAI's most powerful AI model can actually do. Exploding Topics. <https://explodingtopics.com/blog/new-chatgpt-release-date>
- Hu, J., Liu, F., Chu, C., & Chang, Y. (2023). Rapid survey study: Health care trainees' and professionals' perceptions of ChatGPT in improving medical knowledge training. *Journal of Medical Internet Research*, 25, e49385. <https://doi.org/10.2196/49385>
- Hutson, M. (2021). Robo-writers: The rise and risks of language-generating AI. *Nature*, 591(7848), 22–25. <https://doi.org/10.1038/d41586-021-00530-0>
- Ism, I. (2025, February 13). GPT-5: Everything we know about OpenAI's new model. *Chatbase*. <https://www.chatbase.co/blog/gpt-5>
- Jaboob, M. (2024). Integration of generative AI techniques and applications in student behavior and cognitive achievement in Arab higher education. *International Journal of Human-Computer Interaction*, 41(1), 353–366. <https://doi.org/10.1080/10447318.2023.2300016>
- Kabilan, M. K., & Veratharaju, K. (2013). Professional development needs of primary school English-language teachers in Malaysia. *Professional Development in Education*, 39(3), 330–351. <https://doi.org/10.1080/19415257.2012.762418>
- Kaddour, J., Harris, J., Mozes, M., Bradley, H., Raileanu, R., & McHardy, R. (2023). Challenges and applications of large language models. *ArXiv*. <https://arxiv.org/abs/2307.10169>
- Kaswan, K. S. (2024). AI in personalised learning. *Advances in Technological Innovations in Higher Education: Theory and Practices*, 103–117. <https://doi.org/10.1201/9781003376699-9>
- Khalil, M., & Er, E. (2023). Will ChatGPT get you caught? Rethinking plagiarism detection. *arXiv preprint*. Retrieved on April 9, 2023, from <https://arxiv.org/abs/2302.04335>
- Knopf, J. W. (2006). Doing a literature review. *PS: Political Science & Politics*, 39(1), 127–132.
- LaCour, M., Cantú, N. G., & Davis, T. (2019). When calculators lie: A demonstration of uncritical calculator usage among college students and factors that improve performance. *PLOS ONE*, 14(10), e0223736. <https://doi.org/10.1371/journal.pone.0223736>
- LazyProgrammer. (2025). *GPT-5: What we know and why it is delayed*. Retrieved from <https://lazyprogrammer.me/gpt-5/>
- Li, K., Patel, O., Viégas, F., Pfister, H., & Wattenberg, M. (2023). Inference-time intervention: Eliciting truthful answers from a language model. *ArXiv*. <https://arxiv.org/abs/2306.03341>
- Liu, J., Liu, A., Lu, X., Welleck, S., West, P., Le Bras, R., Choi, Y., & Hajishirzi, H. (2022). Generated knowledge prompting for commonsense reasoning [Paper presentation]. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers), 3154–3169, Dublin, Ireland: Association for Computational Linguistics.
- Lodge, J. M. (2023, May 21). Cheating with generative AI: Shifting focus from means and opportunity to motive [LinkedIn page]. *LinkedIn*. <https://www.linkedin.com/pulse/cheating-generative-ai-shifting-focus-from-means-motive-lodge/>

- Min, B., Ross, H., Sulem, E., Veysch, A. P. B., Nguyen, T. H., Sainz, O., ... & Roth, D. (2023). Recent advances in natural language processing via large pre-trained language models: A survey. *ACM Computing Surveys*, 56(2), 1–40.
- Muhlgay, D., Ram, O., Magar, I., Levine, Y., Ratner, N., Belinkov, Y., Abend, O., Leyton-Brown, K., Shashua, A., & Shoham, Y. (2023). Generating benchmarks for factuality evaluation of language models. *ArXiv*. <https://arxiv.org/abs/2307.06908>
- Ng, D. T. K. (2021). Conceptualising AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2. <https://doi.org/10.1016/j.caeai.2021.100041>
- Okonkwo, C. W., & Ade-Ibijola, A. (2021). Chatbots applications in education: A systematic review. *Computers and Education: Artificial Intelligence*, 2, 100033. <https://doi.org/10.1016/j.caeai.2021.100033>
- OpenAI. (2023). *GPT-4 technical report*. Retrieved from <https://openai.com/research/gpt-4>
- OpenAI. (n.d.). *GPT-4 Turbo in the OpenAI API*. OpenAI Help Center. <https://help.openai.com/en/articles/8555510-gpt-4-turbo-in-the-openai-api>
- Oppenlaender, J. (2022). The creativity of text-to-image generation. *Proceedings of the 25th International Academic Mindtrek conference (Academic Mindtrek '22)*. ACM, 11. <https://doi.org/10.1145/3569219.3569352>
- Ray, P. P. (2023). ChatGPT: A comprehensive review of background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*, 3, 121–154. Retrieved from www.keaipublishing.com/en/journals/internet-of-things-and-cyber-physical-systems
- Rossa, N.. (2024, March 22). Waduh! Peneliti temukan banyak jurnal akademis yang ternyata ditulis ChatGPT. *DetikEdu*. Retrieved from <https://www.detik.com/edu/edutainment/d-7256390/waduh-peneliti-temukan-banyak-jurnal-akademis-yang-ternyata-ditulis-chatgpt>
- Stewart, I. (2023b, March 8). Noam Chomsky: The false promise of ChatGPT. *The New York Times*. <https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt.html>
- Stewart, J. (2023a, February 17). Noam Chomsky says ChatGPT is a form of “high-tech plagiarism”. *My Modern Met*. <https://mymodernmet.com/noam-chomsky-chat-gpt/>
- Sullivan, M., Kelly, A., & McLaughlin, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning & Teaching*, 6(1), 1–10. <https://doi.org/10.37074/jalt.2023.6.1.17>
- UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000381137>
- Vincent, J. (2022, December 5). AI-generated answers are temporarily banned on coding Q&A sites. *The Verge*. <https://www.theverge.com/2022/12/5/23493932/chatgpt-ai-generated-answers-temporarily-banned-stack-overflow-llms-dangers>
- Ward, P. (2023). Choice, uncertainty, and decision superiority: Is less AI-enabled decision support more? *IEEE Transactions on Human-Machine Systems*, 53(4), 781–791. <https://doi.org/10.1109/THMS.2023.3279036>
- Weissman, J. (2023). “ChatGPT is a Plague Upon Education,” *Inside Higher Ed*. Retrieved April 26, 2023. Available at <https://www.insidehighered.com/opinion/views/2023/02/08/chatgpt-plague-upon-education-opinion>.

- Welborn, A. (2023, March 9). ChatGPT and fake citations. *Duke University Libraries*. <https://blogs.library.duke.edu/blog/2023/03/09/chatgpt-and-fake-citations/>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, 16, 39. <https://doi.org/10.1186/s41239-019-0171-0>
- Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review. *Smart Learning Environments*, 11(1), 28.
- Zhang, S., Pan, L., Zhao, J., & Wang, W. Y. (2023). The knowledge alignment problem: Bridging human and external knowledge for large language models. *ArXiv*. <https://arxiv.org/abs/2305.13669>