

Potential and hallucinations of AI in ELT: Balancing innovation and scepticism with critical insight

LUISA LUPOLI

Università degli Studi di Napoli Federico II
luisa.lupoli@unina.it

ABSTRACT

As generative artificial intelligence (GenAI) increasingly permeates educational domains, English Language Teaching (ELT) is undergoing a profound transformation. Drawing on theoretical frameworks, reflective pedagogy, and recent classroom applications, this paper explores the dual nature of AI integration in ELT, presenting both its innovative potential and the challenges posed by AI-generated inaccuracies, or ‘hallucinations’. By examining contemporary AI-driven platforms, such as ChatGPT, Twee.ai, Natural Reader, Gemini, as well as image-generation tools like DeepAI, the article demonstrates their applicability to real-world classroom contexts, including writing support for Cambridge exam preparation, adaptive reading tasks, and multimodal resources for learners with special needs, with a primary focus on how AI can enhance engagement, personalisation, and accessibility in ELT. Simultaneously, it addresses the limitations of AI-generated content, such as inaccuracies, biases, and ethical concerns, emphasising the importance of pedagogical mediation and critical digital literacy, with a call for responsible AI integration and continuous professional development. Ultimately, it argues

that AI should serve as a complement – not a substitute – for human judgment, creativity, and contextual awareness in language education, where cultivating critical thinking among learners and educators is essential to balance innovation and enthusiasm with ethical responsibility.

KEYWORDS

Generative Artificial Intelligence (GenAI), English Language Teaching (ELT), Inclusivity, Learner autonomy, Critical digital literacy

1. INTRODUCTION

The rapid evolution of artificial intelligence (AI) and its integration into educational contexts worldwide have sparked a paradigm shift in how language is taught and learned, responding to learner input in dynamic, personalised ways and offering new possibilities for scaffolding, feedback, and learner agency (see Kundu & Bej 2025). Particularly in the context of ELT, where multilingualism and digital competence are strategic priorities (European Commission 2020), AI applications are reshaping pedagogical strategies, assessment methods, and learner engagement. Among the most prominent technologies leading this transformation is Generative Artificial Intelligence (GenAI), a suite of tools capable of *producing* human-like text, speech, and images, thus presenting promising opportunities to enhance learner engagement, personalise instruction, and automate feedback.

As AI technologies become more embedded in everyday educational practice, the discourse surrounding their pedagogical value has grown more complex. On one hand, the unprecedented access they offer to personalised learning, real-time feedback, and interactive learning experiences has introduced substantial opportunities. Tools such as OpenAI ChatGPT, Google Gemini, or Twee.ai demonstrate remarkable capabilities in supporting both receptive and productive language skills, allowing teachers to automate certain instructional tasks, create engaging content, and cater to diverse learner needs.

On the other hand, however, while the pedagogical possibilities are exciting, these tools are not without limitations, as this rapid evolution also introduces considerable risks and complex challenges. Issues such as misinformation, algorithmic bias, and the phenomenon commonly referred to as ‘hallucination’, where AI generates false or misleading content, pose significant concerns and raise critical questions about the reliability, accountability,

and educational integrity of AI-driven instruction. Moreover, the ease of access to these tools may inadvertently encourage passive learning, undermine critical engagement, and lead to a diminished role for human judgement, contextual nuance, and cultural sensitivity.

This article seeks to provide a balanced analysis of the role of AI in ELT, emphasising the need to temper innovation with scepticism. Through a synthesis of research, classroom observation, and reflective analysis, it explores both the pedagogical potential and the limitations of AI tools by highlighting best practices while acknowledging ethical and practical risks. Central to the argument is the assertion that critical thinking must remain at the heart of any AI-assisted educational approach, meaningfully supporting language learning without compromising its human-centred foundations.

2. AI IN CONTEMPORARY ELT: A LITERATURE REVIEW

The integration of artificial intelligence into English Language Teaching (ELT) has evolved in tandem with broader developments in educational technology. Understanding its role today requires a look at its historical foundations, the current pedagogical landscape, recent GenAI advances, and the theoretical frameworks that inform its educational use.

2.1 HISTORICAL BACKGROUND OF AI IN EDUCATION

The application of AI in education is not entirely new. It dates back to the 1960s with early work on intelligent tutoring systems (ITS), such as SCHOLAR and PLATO, which sought to simulate human tutors by adapting feedback to student input (Carbonell 1970). These systems, though relatively basic in comparison to today's technologies, laid the groundwork for individualised learning through computational models.

During the 1980s and 1990s, AI remained largely confined to experimental environments due to limitations in processing power and natural language understanding, until the advent of machine learning and more sophisticated algorithms in the early 2000s paved the way for more robust educational tools, especially in adaptive learning platforms and automated grading systems (Luckin et al. 2016).

However, in language education specifically, AI applications were traditionally limited to grammar-checking and speech recognition. It was the rise of natural language processing (NLP)-based tools in the 2010s – see, for ex-

ample, Duolingo's adaptive exercises or Grammarly's context-aware suggestions, that marked a significant shift towards more 'intelligent', responsive systems that could support learners in dynamic ways.

2. CURRENT STATE OF ELT AND EDUCATIONAL TECHNOLOGY

The realm of ELT has increasingly embraced technology to meet the needs of diverse classrooms and cross-border education initiatives. Digital learning environments, virtual exchange programmes, and mobile learning apps are now commonplace in many educational systems.

In the European context, for instance, according to the Eurydice Report (European Commission 2022) over 70% of language teachers regularly use some form of digital tool to support instruction, whether for grammar drills, listening tasks, or formative assessment purposes.

Despite this integration, the gap between pedagogical promise and actual implementation remains wide, as many ELT practitioners face challenges in integrating technology meaningfully due to limited training, infrastructural disparities, or a lack of alignment with curricular goals (Kessler & Hubbard 2017). The push towards digital transformation – accelerated by the COVID-19 pandemic – has made it more urgent than ever to evaluate not only the availability of technology, but also its pedagogical value and critical implications.

2.3 RECENT DEVELOPMENTS IN GENERATIVE AI

While AI-driven language learning tools have existed since the early 2000s (see Han 2020), recent years have seen a proliferation of GenAI tools designed to support learners' acquisition of English as a second or foreign language.

Large language models (LLMs) such as OpenAI ChatGPT and Google Gemini represent a significant leap in NLP capabilities as they can produce human-like text outputs, correct grammar, suggest vocabulary alternatives, provide feedback on writing quality, and offer conversational interfaces that simulate real dialogue, allowing learners to practise English in realistic contexts. Studies have shown that these tools enhance speaking confidence (Du & Daniel 2024) and writing fluency (Marzuki 2023), while also transforming how teachers approach content creation, learner feedback, and individualised instruction, with a notable shift in the ability to provide immediate, context-aware feedback at scale – a task that once required considerable time and effort.

In this context, educators are experimenting with AI-enhanced lessons that adapt to student needs in real time, with these tools serving as virtual tutors and enabling learners to engage in extended written and spoken interactions aimed at improving fluency and accuracy (Godwin-Jones 2022). Similarly, platforms like Twee.ai use NLP to generate context-sensitive quizzes, lesson plans, and learning materials, thus reducing the teacher's administrative burden and allowing for more differentiated instruction, not to mention the support they offer to differentiated learning (Chen et al. 2022).

Beyond text generation, AI-powered multimodal tools are also being implemented into ELT. Platforms like Natural Reader offer sophisticated text-to-speech (TTS) capabilities, enabling learners to hear written texts read aloud in natural, human-like voices by providing audio outputs with varied accents and adjustable playback speeds, which can be particularly beneficial for receptive language development, e.g. pronunciation and listening skills, as observed by Bione et al. (2017) in a study carried out on the evaluation of TTS as a pedagogical tool. In addition, image generators such as DALL-E or DeepAI are designed to create visual stimuli for storytelling, creative writing, and most importantly vocabulary acquisition (Wang et al. 2024), thus aiding learners with strong visual learning preferences.

In inclusive and diverse classroom settings, these technologies are also beneficial to support learners with disabilities or specific learning needs. Text-to-speech, real-time captioning, and adaptive learning systems have shown promise in improving accessibility (Holmes et al. 2019), also aligning with the broader goals of the European Commission Digital Education Action Plan (2020), which encourages the use of emerging technologies to foster equity and inclusion in education.

These developments have begun to alter the traditional teacher-student dynamic, as AI tools increasingly perform instructional and evaluative roles, while the teacher's function is shifting towards that of a facilitator or orchestrator of learning experiences (Popenici & Kerr 2017), which is particularly salient given the constant need to balance standardised language instruction with individualised support – a gap that AI can help bridge through careful pedagogical framing and equitable and effective use.

However, despite their many potential benefits, AI tools are to be accompanied by a healthy dose of criticism (for a broader investigation on the subject, see Wach et al. 2023). One of the primary concerns raised in the literature is the issue of AI-generated misinformation – often referred to as 'hallucinations', which can range from minor factual inaccuracies to more serious misrepresentation of cultural or linguistic norms. Such outputs can

be confusing or misleading for language learners, especially if they are not trained to critically evaluate digital content.

Another recurring concern is bias. Studies have shown that LLMs can inadvertently reproduce cultural, racial, or gender biases present in their training data (Blodgett et al. 2020) which, particularly in the context of a language classroom, raises ethical questions about the kind of content learners are exposed to and the values implicitly conveyed through AI-generated materials.

Additionally, the lack of transparency in how many AI tools operate has led scholars to call for increased digital literacy among both teachers and students. As Selwyn (2019) argues, educators must be equipped not only to use AI but also to interrogate its assumptions, limitations, and socio-political implications, especially in an era where digital literacy is increasingly recognised as a core educational competence.

2.4 THEORETICAL FRAMEWORKS INFORMING AI INTEGRATION IN ELT

A critical understanding of AI's role in ELT benefits from established educational theories and contemporary insights into digital pedagogy. Constructivist learning theory, rooted in the seminal work of Piaget (1952) and Vygotsky (1978), emphasises active, learner-centred knowledge construction, a framework with which AI tools that foster interaction, feedback, and discovery align well – particularly when technology encourages learners to engage with language through exploration rather than rote memorisation (Jonassen 1991).

Vygotsky's concept of the Zone of Proximal Development (ZPD) is also particularly relevant in the context of AI tutors and support systems. AI can serve as a scaffold, offering assistance that allows learners to complete tasks slightly beyond their independent ability – a scaffolding that is temporary and strategic as its ultimate goal is to foster learner autonomy.

Another key model is the Technological Pedagogical Content Knowledge (TPACK) framework (Mishra & Koehler 2006), which outlines the interplay between technology, pedagogy, and content knowledge. According to this framework, effective integration of AI in ELT requires teachers to understand not only how the tools function, but also how they intersect with pedagogical strategies and language content. For instance, using ChatGPT to facilitate a collaborative writing task demands more than technical know-how, as it also involves understanding how to structure the activity, guide learner interaction, and critically evaluate AI-generated output.

Complementing these foundational theories, Pegrum et al. (2022) highlight the importance of teacher mediation in digital learning environments,

arguing that the effective integration of technologies requires educators to balance innovative possibilities with ethical considerations and pedagogical intent. Their work underscores the vital role of teachers in guiding learners through the complexities of machine-generated content and fostering critical digital literacies.

Together, these theoretical lenses help ensure that AI in ELT is not employed merely for novelty or convenience, but rather as a deliberate, evidence-based pedagogical approach. By grounding AI applications in constructivist principles, educators can prioritise learner engagement, critical thinking, and personalised support, fostering deeper language acquisition and avoiding the pitfalls of passive consumption or overreliance on technology, thus ensuring that AI serves to complement – not replace – the essential human elements of teaching and learning.

3. METHODOLOGY AND ANALYTICAL APPROACH: A PRACTITIONER'S PERSPECTIVE IN ITALIAN HIGHER EDUCATION

This section presents a practitioner-oriented account of how artificial intelligence has been integrated into my English as a second language teaching practice within the context of an Italian university. The analysis draws on first-hand classroom experience, particularly in preparing students for the Cambridge C1 Advanced examination as well as supporting those undertaking English curricular exams at either intermediate or upper-intermediate levels (CEFR B1-B2). Rather than presenting empirical data, the approach is reflective as I offer structured insights into the pedagogical use of generative AI tools, informed by critical pedagogy and practical experimentation with AI tools. The focus is on practical integration, learner response, and ethical awareness, particularly around individualised feedback and inclusive pedagogy. This practitioner-oriented methodology seeks to offer grounded, real-world illustrations of the role of AI in contemporary ELT, with the aim of highlighting how these tools can enhance, rather than replace, principled language teaching when guided by theory, reflection, and critical evaluation.

3.1 MODELLING ACADEMIC WRITING WITH CHATGPT

In preparing my advanced learners (CEFR C1) for an academic writing module, I used ChatGPT to co-construct a model argumentative essay in response to a sample prompt on the following topic: “Should governments regulate

social media to protect public mental health?”. I prompted the tool to write a 250-word academic essay adhering closely to the four standardised assessment criteria: content, communicative achievement, organisation, and language.

Having reviewed and lightly edited the generated output, I used it as a comparative model for a peer analysis activity: after completing a gist task on the model text – submitted with blanks in place of linking devices – students were asked to reflect on the types of cohesive devices they could use. They then compared their ideas with the linking words/phrases and discourse markers found in the original model text, and evaluated them using the Cambridge rubric we had previously studied, focusing on strengths in coherence and cohesion. This proved highly effective as it helped learners internalise formal writing features through clear modelling, and it prompted deeper engagement with the criteria, with many students expressing surprise at the fluency and accuracy of the text. Nevertheless, in spite of the initial enthusiasm, this also became a springboard for discussing both the strengths of AI-generated content and its potential shortcomings, such as overgeneralisation and lack of nuanced argumentation. The fluent but slightly mechanical tone of the model text helped them distinguish more clearly what constitutes ‘natural’ vs. ‘artificial’, and the activity raised critical awareness of how such tools could be useful for modelling – but not replacing – authentic, original student writing.

3.2 LEARNER-AUTONOMY AND FEEDBACK WITH CHATGPT

Beyond modelling, I encouraged C1 students to use ChatGPT to request feedback on their own essays. One student, preparing for the Writing Part 1 exam task, submitted their own writing piece to the chatbot and asked for feedback in line with the Cambridge assessment criteria, using the following prompt that I had suggested: “You are an assessor for an exam at advanced level. Please provide detailed feedback on this writing task, focusing on content, communicative achievement, organisation, and language, as per Cambridge assessment criteria”. As a result, the AI provided structured comments on language accuracy and complexity, content and relevance, organisation, tone, and register, which we then analysed together in class, evaluating both the correctness and relevance of the AI-generated feedback.

This prompted a reflective discussion: to what extent was the feedback valid? Where did it overreach or misunderstand the task? I found this experience particularly valuable not only for encouraging learner autonomy but also for building critical digital literacy. It shifted the conversation from “Is this tool accurate?” to “How do we assess and refine the feedback we receive?”, which

I value as a fundamental skill in today's AI-influenced learning environment. The most pedagogically significant outcome was the way students developed *feedback literacy*, that is, the ability to interpret and critique formative input, whether human or machine-generated. As an added benefit, for learners with processing difficulties, such as dyslexia, this AI support provided immediate, digestible guidance that complemented my own feedback and allowed for paced, private reflection outside the classroom.

3.3 FLEXIBLE LESSON DESIGN WITH TWEE.AI

Twee.ai has become a cornerstone of my lesson planning, particularly for learners preparing for English curricular exams at intermediate and upper-intermediate levels. I use the platform weekly to generate targeted vocabulary quizzes, cloze texts, reading comprehension tasks on a wide range of topics, discussion questions, or even creating writing prompts. For instance, during a unit on environmental issues, I used Twee.ai to generate a list of open-ended questions for a fishbowl discussion, a cloze exercise using target vocabulary, and a role-play scenario between a policymaker and a climate activist. The platform allowed me to select CEFR level, task type, and even include distractors for more advanced students. Needless to say, review and adaption of the output are always needed before using it in class, particularly to ensure cultural appropriateness and alignment with learning outcomes. An example of this is the approach adopted in most of the AI-generated reading tasks, which I have observed frequently tend to focus primarily on reading for gist subskills (e.g. understanding the main message or overall tone), while almost entirely neglecting other essential competencies, such as detailed comprehension or making inferences.

Apart from inevitable critical considerations and consequent adaptation, the undeniable strength of the platform lies in its wide range of formats, as well as the flexibility that Twee.ai offers, significantly reducing my preparation time while allowing me to easily *differentiate* tasks more meaningfully than static worksheets from textbooks. As a teacher, this enables me to address the needs of students who require either extra challenge or simplified support – something particularly valuable in university curricular classes that often combine diverse levels of language proficiency.

3.4 ENHANCING PRONUNCIATION AND LISTENING SKILLS WITH NATURAL READER

To support pronunciation and listening comprehension in both my intermediate and advanced classes, I have incorporated Natural Reader, an advanced TTS tool that enables the conversion of written texts into natural-sounding audio, with a choice of voices varying in accent, gender, and tone.

This feature offers several pedagogical advantages. Firstly, it provides students with ready-made listening materials that are free from copyright restrictions, a significant concern in university settings, where sourcing authentic audio can be legally and logistically complex. Instead of relying solely on pre-recorded, often limited or culturally specific materials, Natural Reader allows me to generate customised audio that matches course content, students' proficiency levels, and thematic units.

Secondly, the diversity of voices helps address learner preferences and exposure to varied accents and intonations, which helps students develop flexible listening skills that are critical for real-world comprehension.

Moreover, the ability to control playback speed and segment audio into smaller chunks enhances accessibility, especially for students with special needs such as auditory processing difficulties or dyslexia, thus supporting differentiated learning by allowing learners to repeatedly listen at their own pace, as well as strengthening phonological awareness and pronunciation accuracy.

In my teaching practice, for example, I assign texts related to course topics, such as excerpts from academic essays or adapted news articles, and provide the corresponding Natural Reader audio files for out-of-class listening or pronunciation drills. The high-quality, clear pronunciation of the tool acts as an accessible model, and students can record themselves repeating phrases to compare intonation and rhythm. Feedback sessions then integrate this self-monitoring with teacher guidance.

Natural Reader thus exemplifies how AI can supplement traditional listening materials, offering flexibility, legal safety, and inclusivity. Its integration aligns with the objective of empowering learners to engage autonomously with spoken English, enhancing both receptive and productive language skills.

3.5 HUMAN-MACHINE DIALOGUES WITH GOOGLE GEMINI

In my English language classroom, I have incorporated Gemini as a tool to foster spoken interaction, critical thinking, and lexical development, acknowledging its potential to simulate meaningful dialogue around complex

topics – which makes it particularly effective for speaking practice in academic contexts.

One example involved a discussion task as part of a B2-level unit on technology and society – an engaging topic that allowed learners to explore abstract ideas while practising useful language for expressing opinions as well as discourse markers. To encourage spoken fluency, which remains a subsidiary aim even in lessons not explicitly focused on speaking, I invited learners to use Google Gemini and simulate a real-time conversation about the role of humans in a world increasingly shaped by robotics and automation. Students were prompted to ask the AI: “What is the future of human employment in an age of robots and intelligent machines?” and to record or summarise the responses provided by Gemini in order to prepare their own counterpoints, which they used in a structured in-class debate. Moreover, Gemini’s multimodal capabilities allowed students to engage not only with text but also, when relevant, with related images or links to background sources, further enriching their critical engagement.

During follow-up class discussions, students reflected on the stylistic tone and argumentative strategies of Gemini. Some found it useful as a starting point for engaging with a contemporary topic, expressing their own ideas, and developing speaking skills, especially when rehearsing presentations or debates, while others noted its limitations, for instance its repetitiveness, excessive optimism, or tendency to avoid controversy and use vague generalisations.

From my perspective as a teacher, the platform proved particularly effective in promoting autonomous exploration, as students were able to explore ideas, clarify vocabulary, and refine their arguments in preparation for in-class presentations and debates, thus supporting both language development and the cultivation of soft skills such as digital communication, reflection, and audience awareness. Additionally, without displacing student agency or peer interaction, even the criticism became in itself a teachable moment on the importance of critical thinking when engaging with generative tools.

3.6 VISUAL SCAFFOLDING AND CREATIVE TASKS WITH DEEPAI

To support vocabulary development and stimulate visual thinking in language learning, I have also integrated DeepAI, an image-generation tool that creates visuals from text prompts. I found this especially effective for learners in mixed-ability groups, where visual input can serve as a powerful scaffold for those who may struggle with abstract or decontextualised vocabulary.

In one unit focused on describing places and environments, I asked students to input descriptive prompts such as “a foggy medieval village in the mountains” or “a vibrant urban square during a summer festival” into DeepAI. The generated images were then used in class for speaking activities, including paired dialogues and descriptive writing tasks, which not only sparked creativity but also encouraged learners to produce more specific and varied lexical choices, as they had a clear visual anchor.

Beyond vocabulary enrichment, DeepAI has also shown potential in supporting inclusive practice. Abstract concepts can be particularly challenging for some learners, especially those with different cognitive processing profiles. For example, although learning difficulties or special educational needs are not always disclosed in university contexts, in one instance a student voluntarily revealed visual-spatial processing difficulties, sharing that AI-generated visuals helped them retain new lexical items more effectively than text-only materials. Furthermore, image prompts helped reduce cognitive load and anxiety, offering lower-pressure entry points into speaking tasks, especially for those who were less confident in spontaneous language production. The visual stimuli provided structure and direction, allowing learners to anchor their responses in concrete imagery and activate relevant vocabulary in a more supported, imaginative way.

Another application was in a creative writing module, where students used DeepAI to visualise narrative settings or characters they were planning to write about. This decision was grounded in the pedagogical goal of fostering integrated skills development, encouraging students to engage in ideation and lexical exploration before committing to extended written production. One student, for example, used the prompt “a futuristic city at night with glowing towers and flying cars” as the inspiration for a science fiction short story, using the AI tool to support the use of descriptive language, genre conventions, and cohesive narrative structures. These tasks helped bridge the gap between visual literacy and language production, reinforcing both learner agency and the creative use of language in context.

The key pedagogical value of DeepAI lies not only in its capacity to generate engaging imagery but also in the fact that the images are original and free from traditional copyright concerns – which is particularly important in a university context, where the legal use of visuals in course materials must be carefully managed. Used critically and creatively, DeepAI has proven to be a meaningful addition to my toolkit, supporting both receptive and productive skills while enabling differentiation and learner-centred engagement.

To address the range of receptive skills (i.e. reading and listening) in my curricular English classes, I experimented with Diffit to create differentiated lesson materials from a Ted Talk video on how great leaders inspire action¹. I input the video link into Diffit, which then generated a summary text at three CEFR-aligned levels (B1, B2, and C1), along with vocabulary glossaries, comprehension questions, and short-answer prompts. This allowed me to assign different versions of the same content to learners based on their individual proficiency and confidence in either reading or listening, without sacrificing topic cohesion.

The feature I found particularly useful was the ability to instantly produce discussion questions and vocabulary exercises embedded in the content, allowing for an inclusive, learner-centred approach in a mixed-ability classroom – which is often a pedagogical challenge in language education. Students working at different proficiency levels could engage with the same topic using texts tailored to their needs, enabling whole-class discussions without excluding less confident readers or listeners, while the built-in glossary feature proved particularly valuable in addressing the needs of lower-achieving students, who often benefit from embedded vocabulary scaffolding and multi-modal input. The efficiency of this tool facilitated more targeted use of lesson time, creating space for more focused interaction and reflection rather than pre-teaching or clarification.

4. A PRACTITIONER'S PERSPECTIVE: CHALLENGES AND ETHICAL CONSIDERATIONS

While the pedagogical benefits of AI tools are increasingly recognised, integrating them into English language teaching also entails inevitable complications, as observed by Godwin-Jones (2024) who, through the balanced perspective he offers on the 'hype and hope' surrounding AI in language education, cautions against overreliance on AI-generated content due to risks such as inaccuracies, hallucinations, and ethical issues, as well as highlighting the importance of critical teacher mediation to navigate these challenges and uphold educational integrity when integrating AI tools.

Through classroom-based examples, this section outlines the challenges that I have encountered as a university-based language practitioner, with the aim of reinforcing the importance of reflective, responsible use of AI in ELT.

1 https://www.ted.com/talks/simon_sinek_how_great_leaders_inspire_action?language=en

4.1 INACCURACIES IN AI-GENERATED CONTENT

One of the most immediate and recurring issues with generative AI tools is factual or linguistic inaccuracy. In my experience, tools like Twee.ai and ChatGPT can occasionally generate error-prone multiple-choice items, especially when based on cloze texts.

In one instance, I used Twee.ai to generate a vocabulary gap-fill exam task for my B2 students, which required choosing one correct word out of four options to complete each sentence. Upon close review, I noticed that in two of the four items, more than one option could reasonably fit the sentence, depending on context or slight shifts in meaning. Below is an example of one of the AI-generated items, in which more than one choice were contextually appropriate, though the AI marked only *decline* as correct:

He decided to _____ the invitation after thinking it over carefully.

- a) decline
- b) refuse
- c) ignore
- d) cancel

Similarly, another gap-fill exam task based on a cloze text about urban planning – both generated by ChatGPT – resulted in the following ambiguity, which shows how both options a) and c) below were correct, with c) also being to a certain extent possible:

In cities around the world, urban, territorial, landscape, and environmental planning are becoming increasingly crucial (1) _____ populations rise and green spaces shrink.

- a) since
- b) as
- c) while
- d) due

Such inaccuracies, if unchecked, risk undermining learner trust and assessment validity, which highlights the necessity of teacher oversight when deploying AI-generated materials, using them as a base rather than a finished product. Additionally, they can serve as a starting point for integrating a

metacognitive element into activities by asking students to critique the tasks themselves, thereby promoting critical thinking alongside language skills.

4.2 BIAS AND CULTURAL LIMITATIONS

Another challenge concerns bias and cultural framing in AI-generated content. When generating reading prompts or discussion questions with tools like ChatGPT or Diffit, I have noticed a strong tendency towards Western-dominant narratives and examples – particularly American educational or sociocultural contexts.

In one case, while preparing a reading text about educational systems using Diffit, the platform generated an article focusing exclusively on U.S. school structures and lexis, which may have caused confusion for my B1 students who – as native Italians – lacked the background knowledge to relate to terms such as ‘sophomore’ or ‘GPA’. Although such texts can be pedagogically useful as comparative material, their use requires adaptation and contextualisation, for example by shifting the focus towards deducing meaning skills to elicit specific vocabulary learning.

In another example, while demonstrating to the whole-class how to use ChatGPT to generate arguments for a debate on gender roles, the examples provided by the AI skewed towards stereotypical framings, presenting only traditional binary perspectives and failing to reflect broader, intersectional situations. This lack of inclusivity led to a productive class conversation about representation in algorithmic outputs, thus serving as a gateway to discussing digital literacy and AI limitations in light of an inclusive and open-minded approach, encouraging learners to interrogate the cultural assumptions behind what may appear to be ‘neutral’ sources.

4.3 ETHICAL AMBIGUITIES AND LEARNER DEPENDENCE

AI also raises ethical concerns around authorship, learner dependence, and data use. One ethical grey area I encountered involved students submitting AI-generated work without disclosure. In one of the writing modules for my students preparing for the C1 Advanced exam, a learner turned in an exceptionally fluent review. When prompted with simple questions, the student admitted to having used ChatGPT for the first draft, which initiated an open discussion about academic integrity and compelled me to revise my course policy to explicitly address AI-assisted writing.

Instead of banning its use outright, I chose to incorporate AI use transparently. For example, students may use ChatGPT for brainstorming or drafting, provided they clearly document how they used the tool, adopting an approach that fosters responsible engagement and helps learners understand the boundary between support and substitution – a distinction that I believe is vital in higher education.

Moreover, the temptation to offload critical thinking to AI models is quite tangible. In one lesson on preparing arguments for speaking exams, several students returned nearly identical ideas they had sourced from Google Gemini. While the content was linguistically appropriate, it lacked depth, personal engagement, and originality. I responded by adding reflective prompts such as “Do you agree with this point? Why/ Why not?”, to re-centre the learner’s voice in the task.

Finally, there are pressing concerns about data security and privacy. Although platforms like ChatGPT now offer data controls, many free tools do not provide adequate transparency regarding how student inputs are stored or used. In classroom practice, I advise students to avoid entering personal data or identifiable details and ensure that all AI use complies with institutional digital policies and GDPR regulations (European Data Protection Board 2021).

5. CONCLUSION AND FUTURE DIRECTIONS

This article has explored the evolving role of generative artificial intelligence in English Language Teaching through the lens of a practitioner perspective rooted in the context of an Italian university. Drawing on classroom-based applications of tools such as ChatGPT, Twee.ai, Diffit, Natural Reader, DeepAI, and Google Gemini, it has offered both practical strategies and critical reflections on how AI can be used to support a range of pedagogical aims, from scaffolding writing and reading to enhancing listening skills and promoting inclusivity. Such technologies, when used critically and with pedagogical intent, have been shown to enrich learner engagement and autonomy, personalise feedback, and enable more differentiated and accessible learning environments.

However, the integration of AI into ELT also introduces new risks, including inaccuracies in generated content, underlying cultural or linguistic biases, ethical concerns related to authorship and transparency, and the danger of learner over-reliance. The examples provided have highlighted the importance of careful mediation, teacher oversight, and critical digital literacy – not only among educators, but also among students. As AI becomes more embed-

ded in educational technologies, it is crucial that language teachers position these tools not as authoritative sources, but as support systems that require human judgment, contextual knowledge, and ethical awareness.

Looking to the future, several key directions emerge for ELT practitioners and researchers. First of all, there appears to be an urgent need for language curricula to incorporate AI literacy, not merely as an external supplement but as a core component of digital and communicative competence. Language learners, especially in higher education, should be equipped with the ability to use AI tools responsibly and reflectively.

In addition, it is advisable that teacher professional development evolves to include a deeper understanding of AI tools, in terms of both technical knowledge and ethical frameworks, raising awareness of limitations and adopting strategies to address them and to adapt AI-generated content for specific learner groups. Training should also foreground the cultural specificity and potential biases of AI outputs, especially when teaching in contexts that encompass linguistic and sociocultural diversity. Moreover, while AI tools hold promise for more inclusive pedagogy – particularly through customisable, multimodal resources that may benefit students with special educational needs or diverse cognitive styles – what remains essential is further exploration into how such tools can be effectively tailored to support individual learner needs. The focus should now shift from potential to practical implementation, ensuring that AI addresses not only accessibility but also meaningful, personalised engagement.

Furthermore, the ethical and institutional implications of AI in education cannot be overlooked. As students increasingly use AI tools to support coursework, academic settings should develop clear, transparent policies regarding authorship, originality, fairness, and permissible use. Educators are increasingly expected to strike a balance between ensuring academic integrity and embracing the pedagogical opportunities that AI provides.

Finally, more practitioner-based and collaborative research will be essential in monitoring the long-term impact of AI tools in real-world classrooms. As technologies evolve rapidly, it is only through sustained reflection, sharing of practices, and context-sensitive enquiry that educators can refine the use of AI over time to align with learning goals and ethical standards.

Ultimately, AI presents both challenges and opportunities for English language education, and its role is neither neutral nor predetermined. The task before educators is not to passively adopt or reject these tools, but to engage with them critically, leveraging their benefits while remaining attentive to their limitations. When grounded in ethical awareness, pedagogical soundness, and contextual relevance, AI can serve not as a substitute for teach-

ing, but as an enhancement to it, supporting learners' development in ways that are inclusive, adaptive, and intellectually engaging, always in service of meaningful life-long learning.

BIBLIOGRAPHY

Bione T., Grimshaw J. & Cardoso W. (2017) "An evaluation of TTS as a pedagogical tool for pronunciation instruction: The 'foreign' language context", *Research-publishing.net*.

Blodgett S. L., Barocas S., Daumé III H. & Wallach H. (2020) "Language (technology) is power: A critical survey of 'bias' in NLP", *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 5454–5476.

Carbonell J. R. (1970) "AI in CAI: An artificial-intelligence approach to computer-assisted instruction", *IEEE Transactions on Man-Machine Systems*, 11(4), pp. 190–202.

Chen X., Zou D., Xie H., Cheng G. & Liu C. (2022) "Two decades of Artificial Intelligence in education: Contributors, collaborations, research topics, challenges, and future directions", *Educational Technology & Society*, 25(1), pp. 28–47.

Du J. & Daniel B. K. (2024) "Transforming language education: A systematic review of AI-powered chatbots for English as a foreign language speaking practice", *Computers and education. Artificial intelligence*, 6, 100230.

European Commission (2020) *Digital Education Action Plan 2021–2027: Resetting education and training for the digital age*. <https://education.ec.europa.eu/focus-topics/digital-education/action-plan>. Accessed 21/05/2025.

European Commission (2022) *Eurydice report: Digital education at school in Europe*. <https://eurydice.eacea.ec.europa.eu/publications/digital-education-school-europe>, Accessed 27/05/2025.

European Data Protection Board (2021) *Guidelines 05/2020 on consent under Regulation 2016/679*. https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-052020-consent-under-regulation-2016679_en. Accessed 30/04/2025.

Godwin-Jones R. (2022) “Chatbots in language learning: AI systems on the rise, *Intelligent CALL, granular systems, and learner data: short papers from EUROCALL 2022*, pp. 124-128.

Godwin-Jones R. (2024) “Distributed agency in second language learning and teaching through generative AI”, *Language Learning & Technology*, 28(2), pp. 4–31.

Han Y. (2020) “Connecting the past to the future of computer-assisted language learning: Theory, practice, and research”, *Issues and trends in learning technologies*, 8(1).

Holmes W., Bialik M. & Fadel C. (2019) *Artificial intelligence in education: Promises and implications for teaching and learning*, Boston, MA, Center for Curriculum Redesign.

Jonassen D. H. (1991) “Objectivism versus constructivism: Do we need a new philosophical paradigm?”, *Educational Technology Research and Development*, 39(3), pp. 5–14.

Kessler G. & Hubbard P. (2017) “Language teacher education and technology”, *The handbook of technology and second language teaching and learning*, pp. 278-292.

Kundu A. & Bej, T. (2025) “Transforming EFL teaching with AI: A systematic review of empirical studies”, *International Journal of Artificial Intelligence in Education*, pp. 1-34.

Luckin R., Holmes W., Griffiths M. & Forcier, L. B. (2016) *Intelligence unleashed: An argument for AI in education*. London, UCL Knowledge Lab.

Marzuki, Widiati U., Rusdin D., Darwin & Indrawati I. (2023) “The impact of AI writing tools on the content and organization of students’ writing: EFL teachers’ perspective”, *Cogent Education*, 10(2), 2236469.

Mishra P. & Koehler M. J. (2006) “Technological pedagogical content knowledge: A framework for teacher knowledge”, *Teachers College Record*, 108(6), pp. 1017–1054.

Pegrum M., Hockly N. & Dudeney G. (2022) *Digital literacies*, 2nd edition, London, Routledge.

Piaget J. (1952) *The origins of intelligence in children*, New York, International Universities Press.

Popenici S. A. D. & Kerr S. (2017) “Exploring the impact of artificial intelligence on teaching and learning in higher education”, *Research and Practice in Technology Enhanced Learning*, 12(1), pp. 1–13.

Selwyn N. (2019) *Should robots replace teachers? AI and the future of education*, Cambridge, Polity Press.

Sinek, S. (2009) ‘How great leaders inspire action’, *TED*, September. <https://tinyurl.com/3s5eftzk>. Accessed 11/11/2024.

Vygotsky L. S. (1978) *Mind in society: The development of higher psychological processes*, Massachusetts, Harvard University Press.

Wach K., Duong C. D., Ejdys J., Kazlauskaitė R., Korzynski P., Mazurek G., Paliszkiewicz J. & Ziemba E. (2023) “The dark side of generative artificial intelligence: A critical analysis of controversies and risks of ChatGPT”, *Entrepreneurial Business and Economics Review*, 11(2), pp. 7-30.

Wang Y., Wu J., Chen F., Wang Z., Li J. & Wang L. (2024) “Empirical assessment of AI-powered tools for vocabulary acquisition in EFL instruction”, *IEEE Access*, 12.