Expositions

Generative Artificial Intelligence and Applied Linguistics

Sowmya Vajjala
National Research Council, Canada

Since the advent of ChatGPT in November 2022, there has been a growing interest and widespread speculation on how Artificial Intelligence (AI), more specifically generative AI, has the potential to revolutionize research and applications across disciplines, with real-world implications. Applied linguistics researchers and practitioners have long adapted to the use of technology in language learning and teaching, where AI already plays a role in the form of Natural Language Processing (NLP), Machine Learning (ML), and other related technologies. This *Expositions* article introduces generative AI, explains how it works and what distinguishes it from other AI technologies, and discusses its growing influence in the applications relevant to applied linguists. The article concludes with some guidance on how to navigate the generative AI space as an applied linguist while acknowledging the current limitations, including how to use generative AI in research and practice.

2022年11月のChatGPTの登場以来、人工知能（AI）、より具体的には生成AIが、現実の世界にどのような影響を及ぼし、分野横断的な研究や応用に革命をもたらす可能性があるのか、関心が高まり、さまざまな憶測が現在広がっている。応用言語学の研究者や実践者は、自然言語処理（NLP）、機械学習（ML）、その他の関連技術の形でAIがすでに役割を果たしている言語学習や教育における技術の使用に、長い間適応してきた。このExpositionsの論文では、生成AIを紹介し、その仕組みと他のAI技術との違いを説明し、応用言語学にどのように影響をもたらすのかについて論じている。最後には、応用言語学者として、現在の限界を認識しながら、生成AIをどのように研究・実践に活用するかなど、生成AIの空間をどのようにうまく使っていけるかを紹介している。

https://doi.org/10.37546/JALTJJ46.1-3

JALT Journal, Vol. 46, No. 1, May 2024

Keywords: educational technologies; generative AI; natural language processing; speech recognition

Artificial Intelligence technologies in the form of Natural Language Processing (NLP) and Machine Learning (ML) have been used in Intelligent Computer Assisted Language Learning (ICALL), particularly in the development of support tools for reading, writing, speaking, or listening, for intelligent tutoring systems and automated assessment (Heift, 2012). Software applications targeting teachers as well as students that rely on such technologies have been developed and researched for almost two decades now. The potential of AI technologies in building support tools for teachers to select course materials (Brown & Eskenazi, 2004; Sheehan et al., 2014), automated grading of written and spoken language (Burstein et al., 2013; Chen et al., 2018), and automated creation of questions and assessment items (Chinkina & Meurers, 2017) have all been well explored in the past. These technologies have also been employed to build language learner support tools for writing (Madnani et al., 2018) and speaking (Kheir et al., 2023). The availability of language learning mobile apps such as Duolingo (duolingo.com), general purpose writing assistants such as Grammarly (grammarly.com), pronunciation and speaking apps such as elsaspeak (elsaspeak.com) are examples of how this strand of research evolved into practical everyday tools. The use of AI in language learning and teaching can thus be considered an active and established area of research and practice. Vajjala (2018) and Meurers (2021) give an overview of some of the research on the role of machine learning and natural language processing respectively in language learning and teaching.

ChatGPT (https://chat.openai.com/) was released as an open-access web tool in November 2022 and has since played an important role in the discussion around the applications of artificial intelligence, more specifically, generative artificial intelligence in various areas. Within the realm of education, it has been utilized across a range of subject areas such as medical education (Kung et al., 2023; Tsang, 2023), computing education (Denny et al., 2023), and science education (Cooper, 2023). Although there was an initial wave of skepticism and a call to ban the use of such tools in education contexts, there are now calls to think about ways to incorporate them into education policies. The New York City public schools AI policy lab is an example of such an initiative (Klein, 2023). Khan Academy (https://www.khanacademy.org/), an online education provider, announced Khanmigo, an AI-powered teaching assistant, earlier this year, leveraging GPT-4, ChatGPT's successor

(Khan Academy, 2023). In the language learning space, applications such as Duolingo (Duolingo, 2023) and Grammarly (Grammarly, 2023) quickly moved towards incorporating generative AI into their existing software applications. The rapid adaptation of this new technology into these various areas of educational technologies indicates its importance for educational technologies in general, and language learning technologies in particular.

Considering that AI-based technologies have already been used in language learning and teaching for some time now, what new things does generative AI bring into the picture? Does it just do existing things better, or does it enable new possibilities? This *Expositions* article explores the role of generative AI in language learning and teaching technologies and addresses the following questions:

1. What is generative AI and how does it work?
2. How does it impact the technologies related to language learning and teaching?
3. How should one work with generative AI, as an applied linguist?
4. What are some limitations of generative AI, and caveats to working with it?

The target audience is expected to be primarily applied linguists familiar with the use of language technologies and artificial intelligence in the context of language learning and teaching, and interested in knowing more about how the recent developments in generative AI are useful for research and practice in this area. The next four sections address the four questions listed above, respectively.

Here is a quick note on the terminology before diving in: While discussing what AI systems can and cannot do, it is common to use words such as "learning", "understanding", "reasoning" etc. Such words are only used in a metaphorical sense, for easier comprehension, and there are no parallels with human learning/understanding/reasoning abilities. Readers are advised to not conflate machine processes with human processes as they explore this article further.

Generative AI—An Overview

The ultimate goal of any AI system is to achieve a semblance of human-like intelligence in the tasks it is expected to perform. There are many ways of achieving this goal, from using hard-coded rule-based reasoning to learning to perform different tasks in a data-driven manner, from a large volume of ex-

amples, without explicit specification of rules. *Generative AI* refers to the form of AI that is capable of processing and generating new content for a range of input/output forms (e.g., text, image, audio, video, a combination of these et cetera). Generative AI models today are responsible for creating human-like texts, realistic images and videos, and natural-sounding audio. *Deep learning*, a form of data-driven learning based on artificial neural networks, is the force behind all the recent developments in generative AI. There are many different forms of generative models for processing different forms of data, and some of these models can also learn multimodally, that is, working with different forms of input or output at the same time. In this article, we will focus on one type of generative AI that is more relevant to our context—*Large Language Models* (LLMs).

Language models learn to assign probabilities to a sequence of words (Jurafsky & Martin, 2023; Chapter 3). They learn the probabilities of word sequences by using the frequency information from large amounts of textual data. Readers familiar with the use of word concordance models in corpus linguistics may be familiar with this approach. Language models go further and use that knowledge to predict probabilities for future sequences, which can then be used to perform a range of language processing tasks, from text classification to machine translation. *Neural language models*, based on artificial neural networks, use massive amounts of textual data to learn these probabilities. Such massive data is available in many languages in the form of web texts, Wikipedia dumps, and other such sources. This process of learning the probabilities is known as "pre-training". Performing pre-training on increasing amounts of textual data from various sources resulted in more and more powerful language models over the past five years, since the arrival of the BERT language model a few years ago (Devlin et al., 2018). A pre-trained language model that passed through this process with massive amounts of generic text data can be further "fine-tuned" with smaller amounts of task-specific data to perform specific tasks (e.g., question answering, machine translation et cetera), by a process known as "transfer learning" (Howard & Ruder, 2018).

Autoregressive language models are a form of neural language models that undergo pre-training by repeatedly predicting the next token given a sequence of tokens. A token can be understood as a machine equivalent of a word. Note that what humans understand as one word is considered to be composed of multiple tokens by a neural language model. For example, consider this sentence - "Sara vociferously denied to comment". A traditional NLP system may split it linguistically and identify six tokens ["sara," "vo-

ciferously," "denied," "to," "comment"] in this sentence, like a human would perhaps do. However, GPT-4 splits it into 9 tokens instead, as ["S", "ara", " voc", "ifer", "ously", " denied", " to", " comment", "."]. The tokens are not necessarily morphologically meaningful, and this tokenization is machine-learned by processing word patterns in the data, to create a finite vocabulary for the language model.

The task of next token prediction may seem like a simple task from a layperson's perspective. Yet, it forms the foundation for all the modern-day LLMs, as many NLP tasks can be framed as text completion tasks, laying the foundations for a generative language model. For example, if one gives an input "What is the capital of Canada?", a pre-trained LLM can respond with "Ottawa" as an answer. As the amount of pre-training data increased, the models became capable of learning to perform a task based on a description, with very few or no examples, without requiring any explicit further fine-tuning. GPT-3 (Brown et al., 2020), an LLM developed by OpenAI and trained on half a trillion tokens, is an example of such a general-purpose LLM. Today's LLMs (such as ChatGPT) follow this autoregressive approach to text generation and show some ability to process human input and generate an appropriate output for a given input from a human user, in a human language. The current generation of LLMs can also generate natural-sounding text following human instructions. Development of new techniques to improve over what a language model "learns" during pre-training resulted in the latest generative large language models we see today, such as ChatGPT, GPT-4 (OpenAI, 2023), Gemini (Gemini Team, 2023) and Claude (Anthropic, 2023). In addition to such commercial LLMs, a wide range of non-commercial, open-source alternatives, such as Zephyr (Tunstall et al., 2023), Falcon (Almazourei et al., 2023), and LLaMa2 (Touvron et al., 2023), to name a few, are other alternatives. There is also a growing body of work on developing small, focused language models (e.g., Li et al., 2023; Zhang et al., 2024) that are good at reasoning from data and performing tasks that require some form of natural language understanding. The generative LLMs mentioned here are only a few examples, and the readers are suggested to refer to Zhao et al. (2023) for a detailed listing of LLMs.

Two key ideas that made large language models go from models such as BERT to systems like ChatGPT are *Supervised Fine-Tuning (SFT)* and *Reinforcement Learning with Human Feedback (RLHF)*, both of which involve a large number of human annotators. In SFT, the LLM is taught to follow instructions for different use cases (e.g., machine translation, text classification, chat, writing a short story, et cetera), by providing task descriptions

along with example items and soliciting responses from humans for a large data sample. This data is then used to fine-tune and optimize the original pre-trained LLM to perform diverse tasks. For a given prompt, many outcomes are possible from a language model, considering that the output generation process is probabilistic. Which is the most preferred by human users? If a human user ranks a set of responses by an LLM for a given prompt in terms of how good they are, can a model learn to generate "good" responses? RLHF is the technique that addresses this question by learning a "reward model" and optimizing an LLM to generate responses that align with human preferences. The data to learn such a reward model is again collected on a large scale by setting up an annotation task where humans choose a preferred output from the given machine responses. InstructGPT (Ouyang et al., 2022), a generative language model from OpenAI which is a predecessor of ChatGPT, and GPT4, was among the first to describe this approach, which soon became a standard procedure for building large generative AI models.

Any computer system built for a specific purpose can be evaluated on how it performs on specific tasks that achieve that purpose, and machine-learned systems are no exception. However, how should we evaluate Generative AI systems, more specifically, LLMs such as ChatGPT? This is an ongoing and active area of research, and the current practices include evaluating LLMs on popular benchmarks that cover multiple tasks and languages as well as other aspects such as toxicity and harmfulness. Note that there are several *LLM evaluation benchmarks*, and there is no single LLM that performs the best on all the benchmarks. A *public leaderboard* offers a quick lookup of how different LLMs compare against each other on various benchmarks (Huggingface.co, 2023). Liu et al. (2023) and Guo et al. (2023) present comprehensive surveys on the evaluation of large language models. Note that the performance on such standard evaluation benchmarks should not be equated to real-world performance in a given application scenario and it is possible for an LLM to do well on such benchmarks but not be useful for a given real-world task.

There is much more to LLMs and generative AI than what was presented so far, and this only aimed to provide a short overview of what generative AI is, how it differs from other forms of AI, and how generative LLMs such as ChatGPT are built, trained, and evaluated. For a more comprehensive discussion about the topic, refer to Jurafsky and Martin (2023). For a contemporary introduction to the artificial neural network models that power modern generative AI, refer to Prince (2023). With this introduction to what generative AI is, let us now turn to how it is impacting the language learning

and technology space.

Impact of Generative AI on Language Learning Technology

The past year witnessed the impact of generative AI in a range of disciplines that were not already adapted to AI in general. Hence, it is natural that educational technologies, that have already adapted AI across many applications, were impacted by generative AI. Some applications such as providing reading/writing/speaking support for learners or teaching support in the form of grading and creating assessment items have improved, and others that were previously considered too specific, such as providing personalized, explicit feedback, are now enabled by these new advances. There is also a huge potential for previously under-explored use cases for AI such as helping teachers with lesson planning or for multimodal content generation. Recent research on the use of generative AI, more specifically large language models, in language learning technologies can perhaps be grouped into three categories: content and test generation, assessment, and assistive tool development. Let us take a closer look at each of them below:

Test item generation: Generation of diverse, high-quality questions from a given content, adhering to a given criteria, can reduce the teachers' workload while increasing content quality. It is also useful in the development of intelligent tutoring systems. NLP techniques have been used for various forms of automated question generation in the past, ranging from fill-in-the-blank and multiple-choice questions to generating open-ended questions. Recent research discussed the utility of large language models for question item generation for English and Swedish texts (Elkins et al., 2023; Goran & Abed Bariche, 2023). Other research also showed how ChatGPT can be useful in generating questions for assessing English reading comprehension (Lee et al., 2023; Shin & Lee, 2023). Human validation studies were conducted in all these studies to verify the usefulness of machine-generated questions. Going a step further, Xiao et al. (2023) demonstrate the usage of ChatGPT for both reading text generation as well as exercise generation for English reading comprehension. They also report an evaluation study with Chinese middle school teachers who concluded the generated texts and exercises to be appropriate for their students.

Assessment: Assessment is another area in which the important application of Natural Language Processing and AI for language learning and technology has been investigated. Automated scoring of essays for language proficiency or short answers for content accuracy has been well-studied in the literature. Over the past year, some work in the NLP community has explored the usefulness of generative AI models for this purpose. Naismith

et al. (2023) show the use of GPT4 in automatic writing evaluation for discourse coherence. Their research showed that GPT4's ratings correlate well with human evaluations, and GPT4 performance is better than a linguistic feature-based model baseline for the dataset under consideration. Further, the GPT4 response can be accompanied by rationales for the evaluations, if necessary. Note that the "rationales" are generated by the model, and need not necessarily align with a human evaluator's rationales.

In contrast to Naismith et al. (2023), another recent work evaluating the ability of GPT3.5 and GPT4's ability to rate short essays on the CEFR scale (Yancey et al., 2023) showed that although GPT4 performs on par with existing approaches when calibration examples are provided in the prompt, agreement with human ratings vary depending on the test taker's first language. Another recent work by Mizumoto and Eguchi (2023) shows that a GPT-based LLM model combined with linguistic feature information performs better than just using an LLM by itself. One major concern with using some inherently opaque large and complex models is the lack of interpretability and explainability of their predictions. Fiacco et al. (2023) developed a method to extract and understand the implicit rubrics of such neural network models when used as essay scorers. Even though this discussion is not exhaustive, it clearly shows the adaptation of generative AI and LLMs into automated language assessment research, and we could expect more practical utilities in the coming years.

Support tools for language learners: Davis et al. (2024) present a comprehensive evaluation of both open-source and proprietary LLMs for (English) Grammatical Error Correction tasks and show that they do not always outperform custom-built machine learning models for the task when used as-is. However, the quick adaptation to generative AI by language learning and writing support software such as Duolingo and Grammarly, which was discussed earlier, clearly points to the value these technologies bring to language learners when customized to the task. Beyond a language learner context, Speakerly (Kumar et al., 2023), a new language learning platform by Grammarly, shows how large language models and speech recognition can be integrated to build a voice-based writing assistant. Raheja et al. (2023) explored instruction tuning, which was described in Section 2, to build a text editing system for writing assistance. Expanding the horizons beyond the commonly seen applications of NLP in the development of such support tools, emerging research has begun to investigate using generative large language models for grammatical error correction beyond English (Kwon et al., 2023). Duolingo (2023) discusses the use of LLMs for generating person-

alized feedback for learners. Kew et al.'s (2023) recent work on benchmarking large language models for automatic text simplification shows that such generative AI models can assist in making texts easier to read for learners, by producing rephrased versions of the input text with simpler vocabulary and syntactic structure.

The use of AI in most of the above-mentioned areas is an existing practice, which underwent considerable improvement with the new generative AI methods. Language technologies such as machine translation and chatbots too have been studied in the context of language learning and teaching in the past for quite some time (Freyer et al., 2020; Hellmich et al., 2023). However, the limitations of the technologies themselves resulted in their use being limited to research studies. Recent advances in neural network techniques improved the generative capability of NLP systems. Hence, we may see more research into the usefulness of such technologies in language learning research in the future (Huang et al., 2022; Tyen et al., 2022; Zhou et al., 2023).

New developments in generative AI can potentially enable new use cases too. Several recent studies (Kasneci et al., 2023; Yan et al., 2023; Yu & Guo, 2023) provide a broader overview of the potential applications and challenges of using generative AI technologies in various aspects of education (not specifically language education). Caines et al. (2023) take the specific case of language teaching and assessment technologies and discuss how generative AI technologies such as large language models can be used in novel ways for content generation, providing feedback, open-ended chatting at the level of a learner, providing document level assessment and feedback, and supporting "plurilingual" learning. Aryadoust et al. (2024) studied the use of LLMs for developing listening assessments targeting test takers at different proficiency levels and concluded that LLMs can be adopted at different stages of listening test development and validation.

Considering pronunciation training in particular, Kheir et al. (2023) predict that the advances in conversational capabilities of generative AI models, coupled with other developments in low-resource and end-to-end speech processing, may lead to the development of more sophisticated and personalized virtual tutors, and support multilingual applications for spoken language learning resources such as pronunciation tutors, which have been primarily English-focused so far (e.g., Ding et al., 2019; Thompson, 2012; Yonesaka, 2017). Asthana et al. (2023) describe an initiative to incorporate generative AI into a higher education course and study how automated generation of course metadata could support broader instructional goals. Matelsky et al. (2023) explore how large language models can be used to

provide rapid personalized feedback to students for open-ended questions. The discussion in this article revolved around written or spoken texts, but we have to remember that language learning involves interaction between learners and a range of semiotic modes beyond printed or spoken texts. Future developments may lead to the maturing of multimodal learning environments with text, images, audio, and other media integrated into the learning process using generative AI technologies.

Most of the developments discussed in this section so far show a high degree of interest in utilizing generative AI in the language learning and technology space. This interest and the push towards adopting generative AI into applied linguistics research and practice necessitates a discussion around the ethics of using generative AI in this context, particularly on how to use the technology appropriately and responsibly. How do applied linguists working in the language teaching and assessment context see the rise of these technologies so far?

There has been some discourse in this regard, particularly in language testing research. Summarizing the debate on allowing the use of assistive technologies including generative AI by test takers for language assessment, Voss et al. (2023) suggest that language teachers must have sufficient expertise to understand and integrate such technologies into their language instruction and assessment practice and recommend collaboration between test creators and AI developers for ensuring appropriate usage of assistive technologies. Taking a holistic perspective on the role of AI methods in the language testing and assessment process, Bolender et al. (2023) also recommend a collaboration between AI scientists, psychometricians, and subject matter experts to address issues around reliability, validity, and fairness in language test development. Another recent article by Xi (2023) echoes this strand of thought, emphasizing developing best practices for the ethical and responsible use of generative AI technologies specifically in the context of language testing. It is not surprising that the discussion around the responsible use of generative AI in this area started with language testing, as that can be considered as a high-stakes application scenario for generative AI compared to others such as the development of teaching and learning support tools.

Working with Generative AI

We've seen how recent advances in generative AI, especially with large language models, have improved upon existing use cases within the realm of language learning and technology, and how they opened pathways for potential new use cases that were not possible before. Kohnke et al. (2023)

in a recent study on generative AI preparedness among university language instructors pointed to the need for tailored support for teachers to develop AI-related competencies. Some research recommends training both the faculty and the students about the effective use of these new technologies (Fuchs, 2023; Huallpa et al., 2023). With widespread speculation around how ubiquitous generative AI would be in our personal and professional lives, how should applied linguists learn to work with generative AI? There are two ways:

Prompting: The most common means of interacting with such systems is through *prompting*. A prompt is similar to a "query" given to a search engine and can be understood as the input (including any instructions) to the AI describing the expected outcome. While having a natural language interface to generative AI systems is tempting to get started right away, creating proper prompts is more of an art than a science, and it would be useful to know some basics to get started. Saravia (2022) provides a comprehensive, constantly updated, collection of resources on prompting large language models. Understanding efficient and effective prompting methodologies could lead to applied linguists exploring the use of generative AI to pursue some of the prospective directions mentioned earlier, as well as add another tool to their research methods basket. Vee et al. (2023) compiled exercises to incorporate generative AI into the practice of teaching writing, which could serve as a useful resource for applied linguistics interested in pursuing this direction.

AI Coding Assistant: Another interesting possibility to work with generative AI as an applied linguist is by using it as a software coding assistant. Applied linguists, especially those who work on topics such as corpus linguistics or CALL have been learning to write software programs across universities. However, available teaching and learning material is not often geared towards students coming from a language teaching background, making learning challenging. The advent of generative AI-based assistants to write code over the past few years has shown promising results in its use in introductory programming classrooms (Porter & Zingaro, 2024; Puryear, 2022).

As for how generative AI is useful in applied linguistics research, the applications discussed in the previous section hopefully provide useful pointers in that direction. Most such research has been traditionally conducted on English language resources, considering the amount of available datasets and software support. The advent of large language models that have some form of knowledge about various languages provides an opportunity to explore them for other languages (e.g., in the Japanese as a Second Language

context). The same applications (content generation, question generation, content assessment, learner support tools, etc.) can all be explored and the capabilities and limitations of current generative AI methods in a broader language learning and teaching technology context can be evaluated for other languages as well.

Let us turn to the question: What can applied linguistics contribute to the discourse around generative AI itself? With the widespread increase in both interest and adoption of generative AI technologies in various application domains, there is also a lot of emerging discourse around the responsible usage of the technologies to ensure reliability and integrity. Note that this discussion is field-specific. For example, a discussion around the ethics of AI system development typically focuses on issues such as fairness and bias in the models, privacy concerns, explainability, and accountability. But when it comes to actually using such AI systems in, say, education, there are other (or additional) concerns such as the question of what is appropriate usage for a student who is learning a topic, or taking part in an assessment to evaluate their understanding. This is where the applied linguistics community can contribute to the general discourse around the ethics of generative AI usage.

A guideline on the ethical usage of generative AI in language teaching, learning, and testing (and more broadly, encompassing other areas of applied linguistics) is needed considering the growing interest in the community on the topic. Yan et al. (2023) discuss ethical concerns around the use of AI broadly in the context of education, and Mohammad (2022) suggests an "ethics sheets" approach for different AI applications, listing the specific questions that need to be addressed, which can have different answers depending on the task at hand. Both these references are useful in thinking about developing guidelines for applied linguistics. The call for developing best practices in using generative AI for language testing (Xi, 2023) can be considered a starting point in this direction.

The annual state of AI reports published by the Montreal AI Ethics Institute (Gupta et al., 2023) are useful to give a broader perspective on various topics around AI ethics, for readers interested in exploring this aspect further. The EU AI Act (European Union, 2024) which proposes to regulate the development, deployment, and use of AI in the European Union region is another example of a broader discussion around addressing the ethical issues around AI and ensuring responsible development of technology.

Limitations and Caveats

Generative AI and its implications and applications are speculated upon

and adopted widely, but this is not immune to challenges. Over the past year, researchers have widely discussed the technological as well as behavioral limitations of generative AI systems (see Kaddour et al., 2023 for a comprehensive discussion). Here are some limitations one needs to be aware of while using generative AI systems:

- Brittleness of the prompt-based querying process: Small changes in the prompts given to generative AI systems can sometimes result in drastic changes in output, which pose problems in terms of reliability and reproducibility of the process.
- Hallucinations: Generative AI systems such as large language models can produce potentially inaccurate, and at times, completely false information, which may be hard to detect, as the text itself is highly fluent.

Although the above-mentioned limitations arise from the working of the systems themselves, the abilities of these systems, along with their ubiquity now, pose two other problems:

- Distinguishing between machine and human-generated output is sometimes difficult owing to the fluency and human-like text patterns. However, there is some ongoing research into watermarking AI-generated output, which can potentially help address such issues in the future.
- Access to such AI systems could potentially compromise the integrity of computer-based testing scenarios, as some recent research showed (de Winter, 2023). Research into alternative formats of assessment may help overcome the challenges that arise out of this issue.

With existing limitations and the potential problems that may arise from the use of these technologies, should we prohibit their use until some solutions have been found? Current discussion in the research community instead suggests acknowledging the ubiquity of generative AI today, and adapting the teaching and evaluation approaches accordingly (Yu, 2023). Finally, it has to be noted that these limitations and caveats reflect the current state-of-the-art, and the mitigation of such issues is currently an active area of research. Thus, we could expect future research to develop new systems that can overcome such challenges, as far as the technology itself is concerned. However, responsible and ethical use of any technology needs to

be separately addressed for a given application scenario, irrespective of how good or advanced the technology is. The guidelines on the responsible use of generative AI should be field-specific, and application-specific, and developing more specific guidance on the use of generative AI covering different topics in applied linguistics would be a worthwhile direction to pursue, as the adoption of these technologies increases.

Summary

In this *Expositions* article, I aimed to give a broader overview of generative AI and its implications for applied linguistics researchers and practitioners. In doing so, I attempted to summarize recent research on generative AI in areas related to applied linguistics from the Natural Language Processing community, as well as the perspectives from applied linguistics research and practice. Some guidelines were provided for applied linguists who are interested in getting started with generative AI technologies, and potentially new research directions were identified. Generative AI was described as an active research area with a blurring divide between research and practice today. Hence, it is important to be aware of its current limitations and potential issues that may arise, and I have provided some guidance in that direction. The capabilities of current generative AI methods open up new avenues for applied linguists, and I hope this article serves as a starting point for a deeper exploration of these technologies and their relevance to the field.

Sowmya Vajjala works as a Natural Language Processing (NLP) researcher at National Research Council, Canada's largest federal research and development organization. Her research interests lie in information extraction from text, multilingual modeling, and studying the relevance of NLP in other disciplines. She co-authored a book: "Practical Natural Language Processing: A Comprehensive Guide to Building Real World NLP Systems", published by O'Reilly Media, which was also translated into Japanese as "実践 自然言語処理 ―実世界NLPアプリケーション開発のベストプラクティス"in 2023.

References

Almazrouei, E., Alobeidli, H., Alshamsi, A., Cappelli, A., Cojocaru, R., Hesslow, D., Launay, J., Malartic, Q., Mazzotta, D., Noune, B., Pannier, B., & Penedo, G. (2023). *The Falcon series of open language models* [Preprint]. arXiv. https://doi.org/10.48550/arxiv.2311.16867

Anthropic. (2023, March 14). Introducing Claude. *Anthropic.* https://www.anthropic.com/index/introducing-claude

Aryadoust, V., Zakaria, A., & Yichen, J. (2024). Investigating the affordances of OpenAI's large language model in developing listening assessments. *Computers and Education: Artificial Intelligence, 6*, Article 100204. https://doi.org/10.1016/j.caeai.2024.100204

Asthana, S., Arif, T., & Thompson, K. C. (2023, December 15). *Field experiences and reflections on using LLMs to generate comprehensive lecture metadata* [Workshop]. Generative AI for Education (GAIED): Advances, Opportunities, and Challenges, San Diego, CA, United States. https://gaied.org/neurips2023/files/31/31_paper.pdf

Bolender, B., Foster, C., & Vispoel, S. (2023). The criticality of implementing principled design when using AI technologies in test development. *Language Assessment Quarterly, 20*(4–5), 512–519. https://doi.org/10.1080/15434303.2023.2288266

Brown, J., & Eskenazi, M. (2004). Retrieval of authentic documents for reader-specific lexical practice. In R. Delmonte, P. Delcloque, & S. Tonelli (Eds.), *Proceedings of InSTIL/ICALL2004 – NLP and speech technologies in advanced language learning systems.* Unipress.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., ... Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems, 33,* 1877–1901.

Burstein, J., Tetreault, J., & Madnani, N. (2013). The e-rater automated essay scoring system. In M. D. Shermis & J. Burstein (Eds.), *Handbook of automated essay evaluation: Current applications and new directions* (pp. 55–67). Routledge/Taylor & Francis.

Caines, A., Benedetto, L., Taslimipoor, S., Davis, C., Gao, Y., Andersen, Ø., Yuan, Z., Eilliott, M., Moore, R., Bryant, C., Rei, M., Yannakoudakis, H., Mullooly, A., Nicholls, D., & Buttery, P. (2023). *On the application of large language models for language teaching and assessment technology* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2307.08393

Chen, L., Zechner, K., Yoon, S. Y., Evanini, K., Wang, X., Loukina, A., Tao, J., Davis, L., Lee, C. M., Ma, M., Mundkowsky, R., Lu, C., Leon, C. W., & Gyawali, B. (2018). Automated scoring of nonnative speech using the speechrater sm v. 5.0 engine. *ETS Research Report Series, 2018*(1), 1-31. https://doi.org/10.1002/ets2.12198

Chinkina, M., & Meurers, D. (2017). Question generation for language learning: From ensuring texts are read to supporting learning. In J. Tetreault, J. Burstein, C. Leacock, & H. Yannakoudakis (Eds.), *Proceedings of the 12th workshop on innovative use of NLP for building educational applications* (pp. 334–344). Association for Computational Linguistics. https://doi.org/10.18653/v1/W17-5038

Cooper, G. (2023). Examining science education in ChatGPT: An exploratory study of generative artificial intelligence. *Journal of Science Education and Technology, 32*(3), 444–452. https://doi.org/10.1007/s10956-023-10039-y

Davis, C., Caines, A., Andersen, Ø., Taslimipoor, S., Yannakoudakis, H., Yuan, Z., Byrant, C., Rei, M., & Buttery, P. (2024). *Prompting open-source and commercial language models for grammatical error correction of English learner text* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2401.07702

de Winter, J. C. F. (2023). Can ChatGPT pass high school exams on English language comprehension? *International Journal of Artificial Intelligence in Education.* https://doi.org/10.1007/s40593-023-00372-z

Denny, P., Prather, J., Becker, B. A., Finnie-Ansley, J., Hellas, A., Leinonen, J., Luxton-Reilly, A., Reeves, B. N., Santos, E. A., & Sarsa, S. (2023). *Computing education in the era of generative AI* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2306.02608

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In J. Burstein, C. Doran, & T. Solorio (Eds.), *Proceedings of the 2019 conference of the North American chapter of the Association for Computational Linguistics: Human language technologies, volume 1 (long and short papers)* (pp. 4171–4186). Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1423

Ding, S., Liberatore, C., Sonsaat, S., Lučić, I., Silpachai, A., Zhao, G., Chukharev-Hudilainen, E., Levis, J., & Gutierrez-Osuna, R. (2019). Golden speaker builder–An interactive tool for pronunciation training. *Speech Communication, 115,* 51–66. https://doi.org/10.1016/j.specom.2019.10.005

Duolingo. (2023, March 14). Introducing Duolingo Max, a learning experience powered by GPT-4. *Duolingo Blog*. https://blog.duolingo.com/duolingo-max/

Elkins, S., Kochmar, E., Serban, I., & Cheung, J. C. K. (2023). How useful are educational questions generated by large language models? In N. Wang, G. Rebolledo-Mendez, V. Dimitrova, N. Matsuda, & O. C. Santos (Eds.), *Home artificial intelligence in education. Posters and late breaking results, workshops and tutorials, industry and innovation tracks, practitioners, doctoral consortium and blue sky* (pp. 536–542). Springer. https://doi.org/10.1007/978-3-031-36336-8_83

European Union. (2024). *Artificial intelligence act.* https://artificialintelligenceact.com/

Fawzi, F., Amini, S., & Bulathwelaa, S. (2023). *Small generative language models for educational question generation* [Workshop]. Generative AI for Education (GAIED): Advances, Opportunities, and Challenges, San Diego, CA, USA. https://gaied.org/neurips2023/files/18/18_paper.pdf

Fiacco, J., Adamson, D., & Ros, C. (2023). Towards extracting and understanding the implicit rubrics of transformer based automatic essay scoring models. In E. Kochmar, J. Burstein, A. Horbach, R. Laarmann-Quante, N. Madnani, A. Tack, V. Yaneva, Z. Yuan, & T. Zesch (Eds.), *Proceedings of the 18th workshop on innovative use of NLP for building educational applications (BEA 2023)* (pp. 232–241). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.bea-1.20

Fryer, L. K., Coniam, D., Carpenter, R., & Lăpușneanu, D. (2020). Bots for language learning now: Current and future directions. *Language Learning & Technology, 24*(2), 8–22. https://doi.org/10125/44719

Fuchs, K. (2023). Exploring the opportunities and challenges of NLP models in higher education: Is Chat GPT a blessing or a curse? *Frontiers in Education, 8*, Article 1166682. https://doi.org/10.3389/feduc.2023.1166682

Gemini Team. (2023). *Gemini: A family of highly capable multimodal models*. DeepMind. https://storage.googleapis.com/deepmind-media/gemini/gemini_1_report.pdf

Godwin-Jones, R. (2023). Emerging spaces for language learning: AI bots, ambient intelligence, and the metaverse. *Language Learning & Technology, 27*(2), 6–27. https://doi.org//10125/73501

Goran, R., & Abed Bariche, D. (2023). *Leveraging GPT-3 as a question generator in Swedish for high school teachers* [Bachelor's thesis, KTH Royal Institute of Technology]. Digital Scientific Archive. https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1789082&dswid=6855

Grammarly. (2023, April 25). Grammarly brings personalized generative AI to your writing process*. Grammarly Blog*. https://www.grammarly.com/blog/grammarlygo-personalized-ai-writing/

Guo, Z., Jin, R., Liu, C., Huang, Y., Shi, D., Supryadi, Yu, L., Liu, Y., Li, J., Xiong, B., & Xiong, D. (2023). *Evaluating large language models: A comprehensive survey* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2310.19736

Gupta, A., Wright, C., Bergamaschi Ganapini, M., Sweidan, M., & Butalid, R. (2022). *State of AI ethics report (Volume 6, February 2022)* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2202.07435

Heift, T. (2012). Intelligent computer-assisted language learning. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics.* Wiley-Blackwell. https://doi.org/10.1002/9781405198431.wbeal0548

Hellmich, E. A., & Vinall, K. (2023). Student use and instructor beliefs: Machine translation in language education. *Language Learning & Technology, 27*(1), 1–27. https://doi.org/10125/73525

Howard, J., & Ruder, S. (2018). Universal language model fine-tuning for text classification. In I. Gurevych & Y. Miyao (Eds.), *Proceedings of the 56th annual meeting of the Association for Computational Linguistics (volume 1: long papers)* (pp. 328–339). Association for Computational Linguistics. https://doi.org/10.18653/v1/P18-1031

Huallpa, J. J., Arocutipa, J. P. F., Panduro, W. D., Huete, L. C., Limo, F. A. F., Herrera, E. E., Callacna, R. A. A., Flores, V. A. A., Romero, M. Á. M., Quispe, I. M., & Hernández Hernández, F. A. (2023). Exploring the ethical considerations of using Chat GPT in university education. *Periodicals of Engineering and Natural Sciences, 11*(4), 105–115.

Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning—Are they really useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning, 38*(1), 237–257. https://doi.org/10.1111/jcal.12610

Huggingface.co. (2023). *Open LLM leaderboard*. https://huggingface.co/spaces/HuggingFaceH4/open_llm_leaderboard

Jurafsky, D., & Martin, J. H. (2023). *Speech and language processing* (3rd ed). https://web.stanford.edu/~jurafsky/slp3/

Kaddour, J., Harris, J., Mozes, M., Bradley, H., Raileanu, R., & McHardy, R. (2023). *Challenges and applications of large language models* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2307.10169

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103,* Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kew, T., Chi, A., Vásquez-Rodríguez, L., Agrawal, S., Aumiller, D., Alva-Manchego, F., & Shardlow, M. (2023). BLESS: Benchmarking large language models on sentence simplification. In H. Bouamor, J. Pino, & K. Bali (Eds.), *Proceedings of the 2023 conference on empirical methods in natural language processing* (pp. 13291–13309). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.emnlp-main.821

Khan Academy. (2023, March 14). Harnessing GPT-4 so that all students benefit. A nonprofit approach for equal access. *Khan Academy Blog.* https://blog.khanacademy.org/harnessing-ai-so-that-all-students-benefit-a-nonprofit-approach-for-equal-access/

Kheir, Y., Ali, A., & Chowdhury, S. (2023). Automatic pronunciation assessment: A review. In H. Bouamor, J. Pino, & K. Bali (Eds.), *Findings of the Association for Computational Linguistics: EMNLP 2023* (pp. 8304–8324). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.findings-emnlp.557

Klein, A. (2023, October 5). 180 degree turn: NYC district goes from banning ChatGPT to exploring AI's potential. *EducationWeek*. https://www.edweek.org/technology/180-degree-turn-nyc-schools-goes-from-banning-chatgpt-to-exploring-ais-potential/2023/10

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). Exploring generative artificial intelligence preparedness among university language instructors: A case study. *Computers and Education: Artificial Intelligence, 5*, Article 100156. https://doi.org/10.1016/j.caeai.2023.100156

Kumar, D., Raheja, V., Kaiser-Schatzlein, A., Perry, R., Joshi, A., Hugues-Nuger, J., Lou, S., & Chowdhury, N. (2023). Speakerly: A voice-based writing assistant for text composition. In M. Wang & I. Zitouni (Eds.), *Proceedings of the 2023 conference on empirical methods in natural language processing: Industry track* (pp. 396–407). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.emnlp-industry.38

Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., Madriaga, M., Aggabao, R., Diaz-Candido, G., Maningo, J., & Tseng, V. (2023). Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLoS Digital Health, 2*(2), Article e0000198. https://doi.org/10.1371/journal.pdig.0000198

Kwon, S., Bhatia, G., & Abdul-Mageed, M. (2023). Beyond English: Evaluating LLMs for Arabic grammatical error correction. In H. Sawaf, S. El-Beltagy, W. Zaghouani, W. Magdy, A. Abdelali, N. Tomeh, I. A. Farha, N. Habash, S. Khalifa, A. Keleg, H. Haddad, I. Zitouni, K. Mrini, & R. Almatham (Eds.), *Proceedings of ArabicNLP 2023* (pp. 101–119). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.arabicnlp-1.9

Lee, U., Jung, H., Jeon, Y., Sohn, Y., Hwang, W., Moon, J., & Kim, H. (2023). Few-shot is enough: Exploring ChatGPT prompt engineering method for automatic question generation in English education. *Education and Information Technologies*, 1–33. https://doi.org/10.1007/s10639-023-12249-8

Li, Y., Bubeck, S., Eldan, R., Del Giorno, A., Gunasekar, S., & Lee, Y. T. (2023). *Textbooks are all you need II: Phi-1.5 technical report* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2309.05463

Liu, Y., Yao, Y., Ton, J. F., Zhang, X., Cheng, R. G. H., Klochkov, Y., Taufiq, M. H., & Li, H. (2023). *Trustworthy LLMs: A survey and guideline for evaluating large language models' alignment* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2308.05374

Madnani, N., Burstein, J., Elliot, N., Klebanov, B. B., Napolitano, D., Andreyev, S., & Schwartz, M. (2018). Writing Mentor: Self-regulated writing feedback for struggling writers. In D. Zhao (Ed.), *Proceedings of the 27th international conference on computational linguistics: System demonstrations* (pp. 113–117). Association for Computational Linguistics. https://aclanthology.org/C18-2025

Matelsky, J. K., Parodi, F., Liu, T., Lange, R. D., & Kording, K. P. (2023). *A large language model-assisted education tool to provide feedback on open-ended responses* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2308.02439

Meurers, D. (2021). Natural language processing and language learning. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics.* Wiley-Blackwell. https://doi.org/10.1002/9781405198431.wbeal0858.pub2

Min, B., Ross, H., Sulem, E., Veyseh, A. P. B., Nguyen, T. H., Sainz, O., Agirre, E., Heinz, I., & Roth, D. (2023). Recent advances in natural language processing via large pre-trained language models: A survey. *ACM Computing Surveys, 56*(2), 1–40. https://doi.org/10.1145/3605943

Mizumoto, A., & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics, 2*(2), Article 100050. https://doi.org/10.1016/j.rmal.2023.100050

Mohammad, S. (2022). Ethics sheets for AI tasks. In S. Muresan, P. Nakov, & A. Villavicencio (Eds.), *Proceedings of the 60th annual meeting of the Association for Computational Linguistics (volume 1: Long papers)* (pp. 8368–8379). Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.acl-long.573

Naismith, B., Mulcaire, P., & Burstein, J. (2023). Automated evaluation of written discourse coherence using GPT-4. In E. Kochmar, J. Burstein, A. Horbach, R. Laarmann-Quante, N. Madnani, A. Tack, V. Yaneva, Z. Yuan, & T. Zesch (Eds.), *Proceedings of the 18th workshop on innovative use of NLP for building educational applications (BEA 2023)* (pp. 394–403). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.bea-1.32

OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Batescu, P., Bao, H., Bavarian, M., Belgum, J., … Zoph, B. (2023). *GPT-4 technical report* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2303.08774

OpenAI. (2023, December 20). *Prompt engineering.* Openai.com. https://platform.openai.com/docs/guides/prompt-engineering

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, Z., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Siemens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems, 35*, 27730–27744.

Porter, L., & Zingaro, D. (2024). *Learn AI-assisted Python programming: With Github Copilot and ChatGPT*. Manning Publications.

Prince, S. J. (2023). *Understanding deep learning*. MIT Press.

Puryear, B., & Sprint, G. (2022). Github Copilot in the classroom: Learning to code with AI assistance. *Journal of Computing Sciences in Colleges, 38*(1), 37–47.

Raheja, V., Kumar, D., Koo, R., & Kang, D. (2023). COEDIT: Text editing by task-specific instruction tuning. In H. Bouamor, J. Pino, & K. Bali (Eds.), *Findings of the Association for Computational Linguistics: EMNLP 2023* (pp. 5274–5291). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.findings-emnlp.350

Saravia, E. (2022). *Prompt engineering guide*. https://github.com/dair-ai/prompt-engineering-guide

Sheehan, K. M., Kostin, I., Napolitano, D., & Flor, M. (2014). The TextEvaluator tool: Helping teachers and test developers select texts for use in instruction and assessment. *The Elementary School Journal, 115*(2), 184–209. https://doi.org/10.1086/678294

Shin, D., & Lee, J. H. (2023). Can ChatGPT make reading comprehension testing items on par with human experts? *Language Learning & Technology, 27*(3), 27–40. https://doi.org/10125/73530

Thomson, R. I. (2012). *English Accent Coach* (Version 2) [Computer software]. https://www.englishaccentcoach.com/

Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., Bashlylov, N., Batra, S., Bhargava, P., Bhosale, S., Bikel, D., Bleacher, L., Canton Ferrer, C., Chen, M., Cucurull, G., Esiobu, D., Fernandes, J., Fu, J., Fu, W., … Scialom, T. (2023). *Llama 2: Open foundation and fine-tuned chat models* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2307.09288

Tsang, R. (2023). Practical applications of ChatGPT in undergraduate medical education. *Journal of Medical Education and Curricular Development, 10*. https://doi.org/10.1177/2382120523117844

Tunstall, L., Beeching, E., Lambert, N., Rajani, N., Rasul, K., Belkada, Y., H, Shengyi., von Werra, L., Fourrier, C., Habib, N., Sarrazin, N., Sanseviero, O., Rush, A. M., & Wolf, T. (2023). *Zephyr: Direct distillation of LM alignment* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2310.16944

Tyen, G., Brenchley, M., Caines, A., & Buttery, P. (2022). Towards an open-domain chatbot for language practice. In E. Kochmar, J. Burstein, A. Horbach, R. Laarmann-Quante, N. Madnani, A. Tack, V. Yaneva, Z. Yuan, & T. Zesch (Eds.), *Proceedings of the 17th workshop on innovative use of NLP for building educational applications (BEA 2022)* (pp. 234–249). Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.bea-1.28

Vajjala, S. (2018). Machine learning in applied linguistics. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics.* Wiley-Blackwell. https://doi.org/10.1002/9781405198431.wbeal1486

Vee, A., Laquintano, T., & Schnitzler, C. (Eds.) (2023). *TextGenEd: Teaching with text generation technologies.* The WAC Clearinghouse. https://doi.org/10.37514/TWR-J.2023.1.1.02

Voss, E., Cushing, S. T., Ockey, G. J., & Yan, X. (2023). The use of assistive technologies including generative AI by test takers in language assessment: A debate of theory and practice. *Language Assessment Quarterly, 20*(4–5), 520–532. https://doi.org/10.1080/15434303.2023.2288256

Xiao, C., Xu, S. X., Zhang, K., Wang, Y., & Xia, L. (2023). Evaluating reading comprehension exercises generated by LLMs: A showcase of ChatGPT in education applications. In E. Kochmar, J. Burstein, A. Horbach, R. Laarmann-Quante, N. Madnani, A. Tack, V. Yaneva, Z. Yuan, & T. Zesch (Eds.), *Proceedings of the 18th workshop on innovative use of NLP for building educational applications (BEA 2023)* (pp. 610–625). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.bea-1.52

Xi, X. (2023). Advancing language assessment with AI and ML–Leaning into AI is inevitable, but can theory keep up? *Language Assessment Quarterly, 20*(4–5), 357–376. https://doi.org/10.1080/15434303.2023.2291488

Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., & Gašević, D. (2023). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology, 55*(1), 90–112. https://doi.org/10.1111/bjet.13370

Yancey, K. P., Laflair, G., Verardi, A., & Burstein, J. (2023). Rating short L2 essays on the CEFR scale with GPT-4. In E. Kochmar, J. Burstein, A. Horbach, R. Laarmann-Quante, N. Madnani, A. Tack, V. Yaneva, Z. Yuan, & T. Zesch (Eds.), *Proceedings of the 18th workshop on innovative use of NLP for building educational applications (BEA 2023)* (pp. 576–584). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.bea-1.49

Yonesaka, S. M. (2017). Learner perceptions of online peer pronunciation feedback through P-Check. *JALT CALL Journal, 13*(1), 29–51. https://doi.org/10.29140/jaltcall.v13n1.210

Yu, H. (2023). Reflection on whether Chat GPT should be banned by academia from the perspective of education and teaching. *Frontiers in Psychology, 14,* Article 1181712. https://doi.org/10.3389/fpsyg.2023.1181712

Yu, H., & Guo, Y. (2023). Generative artificial intelligence empowers educational reform: Current status, issues, and prospects. *Frontiers in Education, 8*. https://doi.org/10.3389/feduc.2023.1183162

Zhang, P., Zeng, G., Wang, T., & Lu, W. (2024). *TinyLlama: An open-source small language model* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2401.02385

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., … Wen, J. R. (2023). *A survey of large language models* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2303.18223

Zhou, W. (2023). *Chat GPT integrated with voice assistant as learning oral chat-based constructive communication to improve communicative competence for EFL learners* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2311.00718