AI in language learning: Opportunities and threats

Robert Godwin-Jones, Virginia Commonwealth University

Abstract

This short paper discusses the benefits of generative AI such as *ChatGPT* for language learning and teaching. Those include the availability of conversation practice with an always available chatbot who can be instructed to use a specific proficiency level, even particular vocabulary or grammatical structures. Chatbots can take on particular roles or identities, as well as provide corrective feedback. For both learners and teachers, AI systems supply the ability to generate second language (L2) learning materials (stories, exercises, games). AI systems are of substantial potential use in assisting in L2 writing in a variety of ways. But there are as well a good many downsides to AI systems. Those include their cultural inauthenticity and their built-in linguistic and social biases. AI systems use a statistical model of language which limits their understanding of the texts they create as well as of the real world inhabited by humans. That can lead to misinformation as well as to limitations in dealing with pragmatic aspects of language use. Students and teachers will need to use AI with their eyes wide open to both the opportunities and threats. Critical AI literacy is required.

Artificial intelligence itself is not new, but generative AI represents a sharp break from earlier efforts. While AI was initially designed to function as an "expert system" within a narrow range of functionality, *ChatGPT* and other generative AI are general use systems that work within a broad range of contexts and can generate language in a wide variety of genres. Additionally, natural language processing was developed in traditional AI by training systems in the rules of how language works (i.e., syntax, morphology, etc.) as well as programming them with knowledge about the social and natural world in which humans live (something akin to humans' common sense). Early AI systems met with only limited success. It turns out to be too complex and vast a task to program computers with enough rules and facts to deal adequately with realities as complex as human language and human society.

A different approach has yielded incredibly better results, namely that used in generative AI. Generative AI is based on "large language models," built from feeding the system with a huge amount of data (digital texts taken from online sources) and having the system look for patterns and regularities to be able to predict next items in a text string (Godwin-Jones, 2021). That results not in a linguistic model of language, but a mathematical one, as the machine learning process converts text into mathematical symbols (see Wolfram, 2023, for a detailed account of the process).

This is very different from how humans acquire language, through a gradual process of socialization. The statistical model of language in AI is very effective at generating output that seems in its fluency, flow, and idiomaticity to have come from a human source. Yet, because AI has no real understanding of the texts it generates and no lived experience in our world to go by, it can "hallucinate", producing statements that are false.

The way that generative AI produces output is different from rule-based AI, and it is also radically different from how computer software is created in traditional programming. Programmers use symbolic reasoning, writing code that features sequencing, loops, and conditional logic; they understand what each line of code is intended to do and have a clear understanding of how a program's output results from the written code. Code can be edited and compiled to change the behavior of the program.

The machine learning process in large language models does not have the same level of transparency. While AI engineers understand the broad outlines of how AI systems work, that is, their basic architecture, they do not comprehend just how the output comes about. In fact, the creators of *ChatGPT* were surprised by some of its capabilities, which they had not expected, much less "programmed". Some adjustments are possible to guide AI output in particular directions, particularly through human reinforcement training. Fundamentally, however, generative AI remains largely a "black box," whose performance seems mysterious, almost magical. Tweaking AI systems is more a matter of trial and error than it is of software engineering (Heaven, 2024).

Likewise, writing prompts for AI systems is an unpredictable process, with sometimes quite different outputs resulting from the same or very similar prompts. From that perspective, "prompt engineering" is a misleading formulation; the process can be better described as "incantation", with users finding by iterative prompting just the write sequence of words to generate the desired output.

While AI output seems genuine (as if written by a human), and is certainly grammatically correct, studies have shown that the texts it produces tend to be bland and uninspired (Kramsch, 2023). They lack the spark of human creativity, representing as they do a repackaging of their textual training data. AI output is form without substance.

Studies have shown also that, unsurprisingly, AI systems lack social and cultural sensitivity (Cao et al., 2023). That derives in part from their training data, that represents the largely Western, "first world", male-oriented cultural orientation of the Internet, its principal training source (Atari et al., 2023). That training data is also overwhelmingly in English, resulting in inferior performance in other languages (Lai et al., 2023). Because AI systems have no lived social experience, their output is often pragmatically challenged (Barattieri di San Pietro et al., 2023). From the dialogues in their dataset, they can learn appropriate sociolinguistic practices (politeness formulas, for example), but studies have shown that they have difficulty dealing with the social nuances in speech acts and other socially determined language use (Lee & Wang, 2023).

One approach that has been used in analyzing AI's pragmatic abilities is to test the systems for alignment with Grice's cooperative principle (Grice, 1989), a frequently used concept in applied linguistics. The principle asserts that for conversational exchanges in social settings, human interlocutors strive to cooperate to be able to communicate effectively. Grice outlined four maxims used to evaluate the degree of cooperation. Those point to the degree to which speakers are truthful, informative, relevant, and clear. Studies of *ChatGPT* from this perspective have shown shortcomings in adhering to those maxims, showing obvious discrepancies with the maxim of truthfulness (Chen et al., 2024). Particularly troublesome as well is informativeness, as that maxim stipulates the speakers use just enough language to be informative and no more. *ChatGPT* tends to

be overly verbose (Thorne, 2024) as well as repetitive, self-referencing, and apologetic (Harnad, 2024).

If AI systems are linguistically and culturally deficient in authenticity that limits their usefulness as a substitute for humans in practicing a second language. Contrary to studies that have advocated integration of AI as an "authentic" language partner in L2 education (Chiu et al., 2023), actual experimentation with AI (and studies on AI pragmatics) demonstrate the infelicity of that idea. That, however, by no means translates into AI being unusable in language learning. On the contrary, AI can be a highly useful addition or complement to other L2 learning activities and practices – if implemented with an awareness of its limitations (Godwin-Jones, 2024). In that process, the role of the teacher is essential, in guiding learners through the use and discussion of AI's benefits and limitations.

The development of critical AI literacy is needed to gain a "calibrated trust" (Ranalli, 2021, p. 14) in AI output. That includes, as indicated above, an understanding of how the process used in creating AI systems limits substantially their ability to use language in socially and culturally appropriate ways. At the same time, AI systems are evolving rapidly with improved functionality arriving on a regular basis, so that current limitations may change in the future. However efficient AI systems become, it will still be important in instructed language learning that any AI use be complemented by human-to-human interactions, such as participation in virtual exchange, which builds interactional and pragmatic competence, not achievable through AI.

Despite limitations, there have been many ways explored in which AI systems can be integrated into second language (L2) learning and teaching. Indeed, since the release of *ChatGPT* in late 2022 there have been a multitude of blog posts, conference papers, and journal articles on AI and language learning. Many have suggested using AI as a chat partner to practice language use (Godwin-Jones et al., 2024). AI chatbots can serve as a conversation partner for practicing the target language in written or spoken form, an especially useful feature for learners in foreign language settings who may not have good access to L2 conversational opportunities.

In contrast to scripted, rule-based chatbots, AI systems are fully capable of conversing coherently on a wide range of subjects. An AI chatbot can take on different identities, fulfilling for example the role of language tutor, or taking on the persona of the user's favorite singer. It can be instructed to use the targeted language at a particular level of proficiency (novice, intermediate, advanced), matching that of the learner. Before initiating a conversation with *ChatGPT*, the learner can direct the AI to correct errors and provide explanations for particular language constructions.

In contrast to traditional voice assistants like *Siri*, *ChatGPT* remembers conversations and thus builds a personal profile of the user. That allows AI chatbots potentially to be pro-active in introducing topics of interest to the learner based on previous encounters. That means that an AI system could function as a personal learning and living companion, adapting to the student's interests and abilities progressing up the educational and professional ladder (Dans, 2024). The customizability of AI systems can be used as well to target particular areas of language, as demonstrated in a tailored GPT tutor for L2 storytelling (Lan & Chen, 2024).

AI could supply a study plan for learners, providing suggested topics and timetables as well as creating learning materials including texts for reading or listening practice, exercises and explanations for learning grammar, and corrective feedback on writing. Those materials could be valuable for teachers as well as for learners. Teachers have the additional option of using AI systems to assist in evaluating student written work, something that is likely of considerable interest to instructors with large numbers of students, as providing corrective feedback individually is a time-consuming process.

One of the important ways in which AI can benefit language learning is in L2 writing. In fact, AI powered tools have been available for some time (see Godwin-Jones, 2022, for an overview). Those include automated writing evaluation systems, like *Criterion*, and interactive text editors like *Grammarly*. These "narrow" AI tools (Schmidt & Strasser, 2022), some of which can be quite expensive, can potentially be replaced with systems like *ChatGPT*. A recent study found that *ChatGPT* was more effective than *Grammarly* in providing helpful feedback (Mizumoto et al. 2024). Another found that providing graded examples of student writing along with directions in the prompt improved the quality of feedback (Poole & Coss, 2023).

Using AI will involve learning how to write prompts appropriately throughout the writing process. Discussing that dynamic with students can be helpful in moving views on writing from product to process. Jacob et al. (2023) describes how *ChatGPT* was used in multiple phases of L2 writing, from idea generation, to evaluating a first draft, to suggesting improvements in grammar and style. The student writer in the case study critically examined AI output, accepting some, rejecting others, and insisting that the final version reflect her voice and persona.

Many teachers will likely be concerned that students will simply have AI complete writing assignments for them. That certainly can happen, although it is not new that students find means to avoid assigned work through cheating or plagiarism. Part of training for students in the use of AI — a necessary practice in education today — should involve discussion of ethics in AI use. Optimally, guidelines for AI are developed cooperatively with students, so that there is a degree of student buy-in. While some students are likely to use AI inappropriately, it is helpful to look at studies of learner use of machine translation to anticipate how AI might be used. Those have shown that students mostly use machine translation to look up words and phrases and to check on the accuracy of individual sentences (Vinall & Hellmich, 2022). Early studies of AI suggests that pattern may hold for its use as well (Baek et al., 2023).

One of the options for integrating AI into L2 writing that can be beneficial is to assign written tasks to connect AI use to real-world contexts. Poole and Polio (2023) point out that doing that accords with the use of AI for "functional authenticity" (Buendgens-Kosten, 2014) in that it aligns with actual use outside of educational settings. Suggestions from that article include assigning students to write blog posts, online reviews, or professionally oriented emails. The tasks could include comparing AI text versions with those written by students and then analyzing and discussing differences. That can lead to greater awareness of the nature of AI output, contributing to AI literacy. Such tasks emphasize the importance of audience in writing, something often lost sight of in classroom emphasis on academic writing. Tasks that show real-world integration also

demonstrate that AI skills will be useful not only for students' academic careers, but in the workplace as well.

There have been many different reactions to the use of AI in education, ranging from asserting that AI has "no place in education" (Caplan, 2023, April 30), to instructors fully integrating AI use in all assigned work. For language learning, it is too early to tell what the learning benefits will be. However, it is clear that AI will play a significant role in our students' current and future lives. Therefore, it behooves us to not ignore AI but to discuss and model its ethical and effective use, as preparation for life after graduation.

References

Atari, M., Xue, M. J., Park, P. S., Blasi, D., & Henrich, J. (2023). Which humans? *PsyArXiv*. https://psyarxiv.com/5b26t/

Barattieri di San Pietro, C., Frau, F., Mangiaterra, V., & Bambini, V. (2023), The pragmatic profile of *ChatGPT*: Assessing the communicative skills of a conversational agent, *Sistemi Intelligenti*, *3*5(2), 379-399. https://doi.org/10.31234/osf.io/ckghw

Baek, C., Tate, T., & Warschauer, M. (2023). "*ChatGPT* seems too good to be true": College students' use and perceptions of generative AI. *OSFPreprints*. https://doi.org/10.31219/osf.io/6tjpk

Buendgens-Kosten, J. (2014). Authenticity. *ELT journal*, *68*(4), 457-459. https://doi.org/10.1093/elt/ccu034

Cao, Y., Zhou, L., Lee, S., Cabello, L., Chen, M., & Hershcovich, D. (2023). Assessing cross-cultural alignment between *ChatGPT* and human societies: An empirical study. *arXiv preprint arXiv:2303.17466*.

Caplan, N. (2023, April 30). *Why I'm not excited by (or even using) generative AI*. Nigel Caplan. https://nigelcaplan.com/2023/04/30/why-im-not-excited-by- or-even-using-generative-ai/

Chen, X., Li, J., & Ye, Y. (2024). A feasibility study for the application of AI-generated conversations in pragmatic analysis. *Journal of Pragmatics, 223*, 14-30. https://doi.org/10.2139/ssrn.4545327

Chiu, T.K.F., Xia, Q., Zhou, X., Chai, C.S., & Cheng, M. (2023). Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, *4*, 100118. https://doi.org/10.1016/j.caeai.2022.100118

Dans,E. (2024, May 9). Let's use AI to liberate students and create a personalized learning process. *Medium*. https://medium.com/enrique-dans/lets-use-ai-to-liberate-students-and-create-a-personalized-learning-process-1fb1228d9428

Godwin-Jones, R. (2021). Big data and language learning: Opportunities and challenges. *Language Learning & Technology*, *25*(1), 4–19. http://hdl.handle.net/10125/44747

Godwin-Jones, R. (2022). Partnering with AI: Intelligent writing assistance and instructed language learning. *Language Learning & Technology*, *26*(2), 5–24. http://doi.org/10125/73474

Godwin-Jones, R. (2024). Distributed agency in language learning and teaching through generative AI. *Language Learning & Technology*, *28*(2), 5–31. https://hdl.handle.net/10125/73570

Godwin-Jones, R., O'Neill, E., & Ranalli, J. (2024). Integrating AI Tools into Instructed Second Language Acquisition. In Chapelle, C. A., Beckett, G., & Ranalli, J. (Eds.), *Exploring artificial intelligence in applied linguistics*. Iowa State University Digital Press.

Grice, P. (1989). *Studies in the Way of Words*. Harvard University Press.

Harnad, S. (2024). Language Writ Large: LLMs, *ChatGPT*, Grounding, Meaning and Understanding. *arXiv preprint arXiv:2402.02243*. https://doi.org/10.48550/arXiv.2402.02243

Heaven, W. (2024, March 4). Large language models can do jaw-dropping things. But nobody knows exactly why. MIT Technology Review. https://www.technologyreview.com/2024/03/04/1089403/large-language-models-amazing-but-nobody-knows-why/

Jacob, S., Tate, T., & Warschauer, M. (2023). Emergent AI-assisted discourse: Case study of a second language writer authoring with *ChatGPT*. *arXiv*. https://doi.org/10.48550/arXiv.2310.10903

Kramsch, C. (2023). Global English: The Indispensable Bridge in Intercultural Communication?. *TEANGA, the Journal of the Irish Association for Applied Linguistics*, *30*(1), 1-32. https://journal.iraal.ie/index.php/teanga/article/view/6797

Lai, V. D., Ngo, N. T., Veyseh, A. P. B., Man, H., Dernoncourt, F., Bui, T., & Nguyen, T. H. (2023). ChatGPT beyond English: towards a comprehensive evaluation of large language models in multilingual learning. *arXiv preprint arXiv:2304.05613*. https://doi.org/10.48550/arXiv.2304.05613

Lan, Y.-J., & Chen, N.-S. (2024). Teachers' agency in the era of LLM and generative AI: Designing pedagogical AI agents. *Educational Technology & Society*, *27*(1), i-xviii. https://doi.org/10.30191/ETS.202401_27(1).PP01

Lee, S. H., & Wang, S. (2023). Do language models know how to be polite*?. Proceedings of the Society for Computation in Linguistics, 6*(1), 375-378. https://doi.org/10.7275/8621-5w02

Poole, F. J., & Coss, M. (2023, December 19). Can *ChatGPT* Reliably and Accurately Apply a Rubric to L2 Writing Assessments? The Devil is in the Prompt(s). https://doi.org/10.35542/osf.io/3r2zb

Poole, F. J., & Polio, C. (2023). From sci-fi to the classroom: Implications of AI in task-based writing. *TASK*, *3*(2), 243-272. https://doi.org/10.1075/task.00022.poo

Ranalli, J. (2021). L2 student engagement with automated feedback on writing: Potential for learning and issues of trust. *Journal of Second Language Writing*, *52*, Article 100816. https://doi.org/10.1016/j.jslw.2021.100816

Schmidt, T., & Strasser, T. (2022). Artificial intelligence in foreign language learning and teaching: A CALL for intelligent practice. *Anglistik: International Journal of English Studies*, *33*(1), 165– 184. https://doi.org/10.33675/angl/2022/1/14

Thorne, S. L. (2024). Generative artificial intelligence, co-evolution, and language education. *Modern Language Journal*, 1–6. https://doi.org/10.1111/modl.12932

Vinall, K., & Hellmich, E. (2022). Do you speak translate?: Reflections on the nature and role of translation. *L2 Journal*, *14*(1), 4–25. https://doi.org/10.5070/l214156150

Wolfram, S. (2023). *What is ChatGPT doing... and why does it work? Stephen Wolfram Writings*. https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/