MDPI

Commentary

Exploring the Ethical Dimensions of Using ChatGPT in Language Learning and Beyond

Silvia Vaccino-Salvadore

Department of English, American University of Sharjah, Sharjah P.O. Box 1797, United Arab Emirates;
ssalvadore@aus.edu

Abstract: The emergence of ChatGPT in the field of education has opened up new opportunities for language learning, but it has also brought about significant ethical considerations that must be carefully considered and addressed to ensure that this technology is used responsibly. With the field of artificial intelligence (AI) advancing at an unprecedented rate, it is imperative for educators and administrators to remain vigilant in monitoring the ethical implications of integrating ChatGPT into language education and beyond. This paper will explore several ethical dimensions concerning the use of ChatGPT, a sophisticated language model developed by OpenAI, in language education. It will discuss privacy, bias, reliability, accessibility, authenticity, and academic integrity as significant ethical implications to consider while integrating ChatGPT into the language classroom. By gaining an initial understanding of the ethical implications involved in utilizing ChatGPT in language education, students, teachers, and administrators will be able to make informed decisions about the appropriate use of the technology, ensuring that it is employed in an ethical and responsible manner.

Keywords: ChatGPT; ethical considerations; language learning

1. Introduction

The rise of artificial intelligence (AI) has introduced new possibilities for language learning, including the use of AI chatbots such as ChatGPT. ChatGPT, a variant of GPT-3 (Brown et al. 2020), is a large language model (LLM) developed by OpenAI (San Francisco, CA, USA) that has been trained on vast amounts of text data (Hughes 2023) and can generate human-like responses to text inputs. Simply put, ChatGPT is designed to mimic human communication in a conversational manner. With almost no computer skills, "users simply have to write their request in natural language in the 'prompt'"; ChatGPT then responds by attempting to match the answer with the question (Pistilli 2022). Its simplicity makes ChatGPT an attractive tool for language learners seeking to improve their skills through interactive conversations with an AI-powered tutor.

Over the past few years, LLMs have made remarkable strides in the field of natural[1] language processing (NLP). LLMs have been employed in a range of real-world contexts, such as utility companies (Government of Dubai 2023), language translation (Vilar et al. 2022), and search engines (Stokel-Walker and Van Noorden 2023), just to name a few. The release of ChatGPT to the public in November 2022 caused quite a stir. From teachers who enthusiastically embraced and incorporated this new technology into their lesson plans (Ortiz 2023), to academic institutions (Castillo 2023; Yang 2023) and even entire countries (Martindale 2023; McCallum 2023) that imposed restrictions on or outright banned the service, ChatGPT is widely recognized as, in the words of academic integrity expert Sarah Elaine Eaton, "the most creative, disruptive technology in a generation" (Nicoll 2023).

Studies have shown that LLMs such as ChatGPT have the capacity to be utilized for promoting social good in different real-world scenarios (Jin et al. 2021); however, whenever a new technological innovation emerges, ethical concerns inevitably come into play. Like

any technology that becomes a part of our daily lives, the development and use of ChatGPT give rise to important ethical questions (Ruane et al. 2019). This paper will explore several ethical dimensions surrounding the use of ChatGPT in language education, such as privacy, bias, accessibility, reliability, authenticity, and academic integrity. These are significant ethical implications to consider before integrating ChatGPT into the language classroom.

2. Ethical Considerations

2.1. Data Privacy and Security

One of the primary ethical concerns associated with using ChatGPT in language learning is privacy. As mentioned above, ChatGPT, like other AI models, requires an enormous amount of data to train and improve its language generation capabilities. These data may include information shared by users during their interactions with the model, which can raise concerns about data privacy and security. Learners may express personal thoughts, emotions, and experiences while using ChatGPT for language learning purposes, and there may be risks associated with the collection, storage, and usage of such sensitive data.

To explain further, the data collected by ChatGPT during language learning interactions may be used "for purposes other than education" (Kasneci et al. 2023, p. 6), including, but not limited to, improving the model's performance and conducting research. The potential uses of learner data raise questions about consent, transparency, and control. Are learners fully informed about the data collection practices of ChatGPT? Do they have control over how their data are used and shared? Are there robust safeguards in place to protect learner privacy and prevent the misuse of their data? Recently, an Italian governmental agency charged with data protection implemented a country-wide ban on ChatGPT, accusing the service of engaging in the unauthorized data collection of users and neglecting to safeguard against minors accessing inappropriate content (McCallum 2023). OpenAI made the necessary modifications to comply with these Italian regulatory requirements, resulting in its service being restored (Robertson 2023). This example shows the importance of being vigilant and ensuring that the use of ChatGPT in language learning respects and protects learners' (and other users') privacy rights, especially in the context of K-12 education, where underage users are involved (Akgun and Greenhow 2021).

2.2. Bias and Lack of Diversity

Bias is another significant ethical concern associated with the use of ChatGPT. Biases have been found in the large datasets used to train LLMs. Consequently, using such datasets can result in biased language generation and the perpetuation of stereotypes, discrimination, and prejudice in language learning interactions. Furthermore, by perpetuating existing societal biases and unfairness, LLMs have the potential to detrimentally affect the teaching and learning processes, as well as their outcomes (Kasneci et al. 2023).

Language often reflects, whether directly or indirectly, human values that vary across culture, time, and place (Occhionero 2000; Hofstede 2001). When we communicate through written text, such as when we write on the internet, the resulting content typically reflects deeply rooted socio-cultural values, identities, and perspectives (Johnson et al. 2022). As a result, by utilizing internet text to train LLMs, which make probabilistic choices guided by these training datasets, we often witness this reflection in the resulting outputs (Zhuo et al. 2023). By analyzing these value-laden outputs, studies have demonstrated the ways in which LLMs show persistent biases against gender, race, and religion. Gender bias, for example, manifests itself when students in language learning courses use AI to translate between languages with different levels of gender specificity (Akgun and Greenhow 2021; Ullman 2022). Studies have also shown a persistent anti-Muslim bias in LLM outputs. By examining the occurrence of certain words alongside religious terms, researchers have found that words such as "violent", "terrorism", and "terrorist" were more commonly associated with Islam compared to other religions (Abid et al. 2021; Garrido-Muñoz et al. 2021). The danger of using biased datasets for training is twofold: at a macro level, the outputs generated by AI have the potential to perpetuate existing biases,

while on a micro level, they can exert influence by subtly and unconsciously shaping one's opinions, which can then impact one's behavior in the real world (Weidinger et al. 2021; Rayne 2023).

Biased language generation by ChatGPT can impact learners' perceptions, attitudes, and understanding of different cultures, languages, and communities. It can reinforce dominant cultural norms, marginalize minority groups, or perpetuate linguistic inequalities. Johnson et al. (2022), who explored the constitution of the training data for GPT-3, found that 93% of the training data were in English (Brown et al. 2020) and that the "values" embedded in the data seemed to manifest an American socio-cultural worldview. This lack of diversity—understood not only as a representation of gender, ethnicity, or religion, but also as an actual language—in the training data contributes to the erosion of linguistic and cultural diversity, propagating a monoculture that is counter to "the pluralism of values in which we live in our diverse societies" (Pistilli 2022). Furthermore, the biases in ChatGPT's language generation may not always be obvious to learners—wherein lies the ethical concern—as the model's outputs are generated based on learned patterns and are not necessarily reflective of accurate, balanced, or unbiased information.

2.3. Accessibility and Reliability

In previous ethical discussions, it has been advocated that AI systems should play a role in promoting global justice and strive for equitable access among all individuals (European Commission 2018; Nguyen et al. 2022). As a result, language learning should be accessible to all learners, regardless of their socioeconomic background, abilities, or limitations. In reality, however, the use of ChatGPT in language learning raises concerns about accessibility. AI-powered language learning tools may require specific hardware, software, or internet connectivity, which may not be available or affordable to all learners, leading some students to gain an advantage over others. For example, the use of ChatGPT by a student to generate more sophisticated written assignments in their additional language can create an unfair advantage over their peers who lack access to the model. As a result, this discrepancy has the potential to introduce inequalities to the assessment process (Cotton et al. 2023).

Access to the internet is one factor to keep in mind. According to Petrosyan (2023), there were 5 billion active internet users, or two thirds of the global population, as of January 2023. Despite this digital population spread, internet access and availability are unevenly distributed. Specifically, the world average internet penetration rate is approximately 64.6%: this percentage jumps to 97.3% in Northern Europe, but in some African countries, the internet access rate drops to 23%. This inequitable internet access can create disparities and exclude certain learners from accessing ChatGPT and other AI-powered tools. This inequity also speaks to the biased nature of the training datasets mentioned above: if we were to "include the entire Internet in all languages, large sections of humanity would still not be represented in the resulting training dataset" (Johnson et al. 2022, p. 3).

Even when accessibility is not a concern, there is still the potential issue of reliability, as it pertains to ChatGPT and its capability "to provide precise and dependable information" (Zhuo et al. 2023). ChatGPT's ability to generate text that resembles human language could pose a challenge for language learners in discerning between verified knowledge and unverified information. As a result, learners may unknowingly accept false, misleading, or outdated[2] content as true without critically evaluating its validity (Kasneci et al. 2023). This difficulty in distinguishing between the two may result, once again, from the model being trained on outdated, inaccurate, or biased[3] data.

2.4. Authenticity

Authenticity is an important aspect of language learning. It involves the use of real-life situations, contextually relevant content, and meaningful interactions with native or near-native speakers (Gilmore 2007). However, the use of ChatGPT in language learning raises questions about the authenticity of the learning experience. While ChatGPT can

generate language-based responses, it lacks the depth, richness, and authenticity of human interactions. AI-generated content may not capture the subtleties of cultural norms, gestures, or nuances of language use, for example, when asked to translate from one language to another (Ducar and Schocket 2018).

In addition, the use of ChatGPT may not be able to provide learners with the emotional, social, and cultural context that is integral to language learning. For example, Kushmar et al. (2022) surveyed over 400 undergraduate English language learners at Ukrainian universities about their perceptions of AI use in the language classroom. The results showed that, by using AI, the students feared losing an authentic learning environment with speakers, and 92% of the respondents were concerned about being assessed by an AI "because of their pronunciation, accent, way of speaking and emotions" (p. 270). In other words, they were worried that the AI model would not be able to understand them.

Can AI-generated content truly replicate the nuances and authenticity of human language, culture, and communication? Given what we have just read about bias in the training data, it raises salient questions about the potential limitations and distortions present in the outputs produced by these language models. Furthermore, an over-reliance on ChatGPT may result in learners acquiring language skills that are disconnected from real-life communication, leading to a superficial understanding of language and culture.

2.5. Academic Dishonesty

Perhaps one of the more salient ethical concerns within the educational community in recent months is the concept of academic integrity and the potential for ChatGPT to undermine it. Language learners could use ChatGPT for assistance in completing language assignments or assessments, which raises concerns about plagiarism, cheating, and the authenticity of learners' work (Currie 2023). They could use it to generate written essays in their additional language, given the right parameters or prompts, and submit these essays as their own work (e.g., Dehouche 2021). Additionally, they can use it in real time to cheat on exams (Susnjak 2022), thereby compromising the fairness of these exams and potentially resulting in "inaccurate assessments of students' knowledge and skills" (Currie 2023, p. 5).

Furthermore, with sophisticated inputs (prompts) come ever-more sophisticated outputs, so much so that the ability to differentiate between text generated by machines and text generated by humans is increasingly posing a significant challenge for teachers and educators[4] (Elkins and Chun 2020; Susnjak 2022; Cotton et al. 2023). AI-generated detection tools have been developed to analyze the language utilized in written text and detect patterns or irregularities that might suggest the work was generated by a machine as opposed to a human (Cotton et al. 2023). The text generated by ChatGPT, however, is unique, which makes it undetectable by anti-plagiarism software.

Apart from the challenges associated with assessment, some academics believe that another key concern of using ChatGPT is its possible negative impact on learners' "critical thinking, learning experience, research training, and imagination" (Shiri 2023). When learners choose to use ChatGPT or other AI-powered tools, significant learning shortcuts may occur, particularly in the development of transferable skills. By relying too heavily on ChatGPT's responses without seeking additional input from teachers or other authoritative sources, learners could become dependent on its use, one outcome of which could be the inability to develop critical thinking skills. As a result, this form of academic dishonesty compromises the fundamental intent of education—to challenge and educate students (Cotton et al. 2023).

3. Concluding Remarks

Ultimately, the integration of ChatGPT, and other AI-powered tools, into language education and beyond requires a thoughtful approach that prioritizes privacy protection, bias mitigation, accessibility, authenticity, and academic integrity. By addressing these ethical considerations, educators and developers can maximize the benefits of AI-powered language learning tools while minimizing their potential risks, thus fostering a more inclu-

sive, fair, and effective learning environment. All told, AI-powered tools such as ChatGPT are here to stay and will increasingly become an integral part of learners' educational experiences. As a result, it may be "better to teach students what [AI tools] are—with all of their flaws, possibilities, and ethical challenges—than to ignore them" (McMurtrie 2023).

Notes

[1] The term "natural" language pertains specifically to human communication (for ex., Italian, Arabic, Mandarin), in contrast to the "artificial" language utilized by machines, often referred to as code languages.

[2] While ChatGPT itself lacks Internet access, other LLMs such as Google Bard claims to leverage Internet search capabilities to generate responses, thus bypassing the constraints of the dataset used by the OpenAI model.

[3] As of May 12, the following disclaimer was added to ChatGPT "Free Research Preview. ChatGPT may produce inaccurate information about people, places, or facts. ChatGPT May 12 Version".

[4] Out of curiosity, the author prompted ChatGPT to provide a list of ethical concerns of using the model in education. Most of the ideas were similar to the ones initially considered by the author.

References

Abid, Abubakar, Maheen Farooqi, and James Zou. 2021. Persistent Anti-Muslim Bias in Large Language Models. Paper presented at the 2021 AAAI/ACM Conference on AI, Ethics, and Society, Virtual, May 19–21.

Akgun, Selin, and Christine Greenhow. 2021. Artificial Intelligence in Education: Addressing Ethical Challenges in K-12 Settings. *AI and Ethics* 2: 431–40. [CrossRef] [PubMed]

Brown, Tom, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, and et al. 2020. Language Models Are Few-Shot Learners. Paper presented at the 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Virtual, December 6–12. Available online: https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf (accessed on 24 May 2023).

Castillo, Evan. 2023. These Schools Have Banned ChatGPT and Similar AI Tools. *BestColleges*. March 27. Available online: https://www.bestcolleges.com/news/schools-colleges-banned-chat-gpt-similar-ai-tools (accessed on 26 May 2023).

Cotton, Debby R. E., Peter A. Cotton, and J. Reuben Shipway. 2023. Chatting and Cheating: Ensuring Academic Integrity in the Era of ChatGPT. *Innovations in Education and Teaching International*, 1–12. [CrossRef]

Currie, Geoffrey M. 2023. Academic Integrity and Artificial Intelligence: Is ChatGPT Hype, Hero or Heresy? *Seminars in Nuclear Medicine* 53: 719–30. [CrossRef]

Dehouche, Nassim. 2021. Plagiarism in the Age of Massive Generative Pre-Trained Transformers (GPT-3). *Ethics in Science and Environmental Politics* 21: 17–23. [CrossRef]

Ducar, Cynthia, and Deborah Houk Schocket. 2018. Machine Translation and the L2 Classroom: Pedagogical Solutions for Making Peace with Google Translate. *Foreign Language Annals* 51: 779–95. [CrossRef]

Elkins, Katherine, and Jon Chun. 2020. Can GPT-3 Pass a Writer's Turing Test? *Journal of Cultural Analytics* 5: 1–16. [CrossRef]

European Commission. 2018. Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems. Available online: https://op.europa.eu/en/publication-detail/-/publication/6b1bc507-af70-11e8-99ee-01aa75ed71a1/language-en/format-PDF (accessed on 25 May 2023).

Garrido-Muñoz, Ismael, Arturo Montejo-Ráez, Fernando Martínez-Santiago, and L. Alfonso Ureña-López. 2021. A Survey on Bias in Deep NLP. *Applied Sciences* 11: 3184. [CrossRef]

Gilmore, Alex. 2007. Authentic Materials and Authenticity in Foreign Language Learning. *Language Teaching* 40: 97–118. [CrossRef]

Government of Dubai. 2023. DEWA Is the First Utility in the World to Enrich Its Services with ChatGPT Technology. *Dubai Electricity & Water Authority (DEWA)*, February 8. Available online: https://www.dewa.gov.ae/en/about-us/media-publications/latest-news/2023/02/chatgpt-technology (accessed on 23 March 2023).

Hofstede, Geert. 2001. *Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations across Nations*, 2nd ed. Thousand Oaks: Sage.

Hughes, Alex. 2023. ChatGPT: Everything You Need to Know about OpenAI's GPT-3 Tool. *BBC Science Focus Magazine*, January 4. Available online: https://www.sciencefocus.com/future-technology/gpt-3/ (accessed on 25 May 2023).

Jin, Zhijing, Geeticka Chauhan, Brian Tse, Mrinmaya Sachan, and Rada Mihalcea. 2021. How Good Is NLP? A Sober Look at NLP Tasks through the Lens of Social Impact, Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. Paper presented at the Association for Computational Linguistics Conference, Bangkok, Thailand, August 1–6; pp. 3099–113. [CrossRef]

Johnson, Rebecca L., Giada Pistilli, Natalia Menédez-González, Leslye Denisse Dias Duran, Enrico Panai, Julija Kalpokiene, and Donald Jay Bertulfo. 2022. The Ghost in the Machine Has an American Accent: Value Conflict in GPT-3. *arXiv*. [CrossRef]

Kasneci, Enkelejda, Kathrin Sessler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günnemann, Eyke Hüllermeie, and et al. 2023. ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education. *Learning and Individual Differences* 103: 102274. [CrossRef]

Kushmar, Lesia Viktorivna, Andrii Oleksandrovych Vornachev, Iryna Oleksandrivna Korobova, and Nadia Oleksandrivna Kaida. 2022. Artificial Intelligence in Language Learning: What Are We Afraid of. *Arab World English Journal Special Issue on CALL* 8: 262–73. [CrossRef]

Martindale, Jon. 2023. These Are the Countries Where ChatGPT Is Currently Banned. *Digital Trends*, April 12. Available online: https://www.digitaltrends.com/computing/these-countries-chatgpt-banned/ (accessed on 26 May 2023).

McCallum, Shiona. 2023. ChatGPT Banned in Italy over Privacy Concerns. *BBC News*, March 31 sec. Technology. Available online: https://www.bbc.com/news/technology-65139406 (accessed on 26 May 2023).

McMurtrie, Beth. 2023. ChatGPT Is Already Upending Campus Practices. Colleges Are Rushing to Respond. *The Chronicle of Higher Education*, March 6. Available online: https://www.chronicle.com/article/chatgpt-is-already-upending-campus-practices-colleges-are-rushing-to-respond (accessed on 27 March 2023).

Nguyen, Andy, Ha Ngan Ngo, Yvonne Hong, Belle Dang, and Bich-Phuong Thi Nguyen. 2022. Ethical Principles for Artificial Intelligence in Education. *Education and Information Technologies* 28: 4221–41. [CrossRef] [PubMed]

Nicoll, Doreen. 2023. ChatGPT and Artificial Intelligence in the Classroom. *Rabble.ca*, May 16. Available online: https://rabble.ca/education/chatgpt-and-artificial-intelligence-in-the-classroom/ (accessed on 27 May 2023).

Occhionero, Marisa Ferrari. 2000. Generations and Value Change across Time. *International Review of Sociology* 10: 223–33. [CrossRef]

Ortiz, Sabrina. 2023. This Professor Asked His Students to Use ChatGPT. The Results Were Surprising. *ZDNET*, February 23. Available online: https://www.zdnet.com/article/this-professor-asked-his-students-to-use-chatgpt-the-results-were-surprising/ (accessed on 27 May 2023).

Petrosyan, Ani. 2023. Internet Usage Worldwide—Statistics & Facts. *Statista*, April 26. Available online: https://www-statista-com/topics/1145/internet-usage-worldwide/ (accessed on 25 May 2023).

Pistilli, Giada. 2022. What Lies behind AGI: Ethical Concerns Related to LLMs. *Revue Ethique et Numérique*, March. Available online: https://hal.science/hal-03607808 (accessed on 5 May 2023).

Rayne, Elizabeth. 2023. AI Writing Assistants Can Cause Biased Thinking in Their Users. *Ars Technica*, May 26. Available online: https://arstechnica.com/science/2023/05/ai-writing-assistants-can-cause-biased-thinking-in-their-users/ (accessed on 27 May 2023).

Robertson, Adi. 2023. ChatGPT Returns to Italy after Ban. *The Verge*, April 28. Available online: https://www.theverge.com/2023/4/28/23702883/chatgpt-italy-ban-lifted-gpdp-data-protection-age-verification (accessed on 30 April 2023).

Ruane, Elayne, Abeba Birhane, and Anthony Ventresque. 2019. Conversational AI: Social and Ethical Considerations. Paper presented at CEUR Workshop Proceedings, Galway, Ireland, December 5–6, vol. 2563, pp. 104–15. Available online: https://ceur-ws.org/Vol-2563/aics_12.pdf (accessed on 24 May 2023).

Shiri, Ali. 2023. ChatGPT and Academic Integrity. *Information Matters*, February 2. Available online: https://informationmatters.org/2023/02/chatgpt-and-academic-integrity/ (accessed on 28 March 2023).

Stokel-Walker, Chris, and Richard Van Noorden. 2023. What ChatGPT and Generative AI Mean for Science. *Nature* 614: 214–16. [CrossRef] [PubMed]

Susnjak, Teo. 2022. ChatGPT: The End of Online Exam Integrity? *arXiv*. [CrossRef]

Ullman, Stefanie. 2022. Gender Bias in Machine Translation Systems. In *Artificial Intelligence and Its Discontents: Critiques from the Social Sciences and Humanities*. Edited by Ariane Hanemaayer. Cham: Palgrave Macmillan, pp. 123–44.

Vilar, David, Markus Freitag, Colin Cherry, Jiaming Luo, Viresh Ratnakar, and George Foster. 2022. Prompting PaLM for Translation: Assessing Strategies and Performance. *arXiv*. [CrossRef]

Weidinger, Laura, John Mellor, Maribeth Rauh, Conor Griffin, Jonathan Uesato, Po-Sen Huang, Myra Cheng, Mia Glaese, Borja Balle, Atoosa Kasirzadeh, and et al. 2021. Ethical and Social Risks of Harm from Language Models. Available online: https://arxiv.org/pdf/2112.04359.pdf (accessed on 25 May 2023).

Yang, Maya. 2023. New York City Schools Ban AI Chatbot That Writes Essays and Answers Prompts. *The Guardian*, January 6 sec. US news. Available online: https://www.theguardian.com/us-news/2023/jan/06/new-york-city-schools-ban-ai-chatbot-chatgpt (accessed on 8 March 2023).

Zhuo, Terry Yue, Yujin Huang, Chunyang Chen, and Zhenchang Xing. 2023. Exploring AI Ethics of ChatGPT: A Diagnostic Analysis. *arXiv*. [CrossRef]