Annual Review of Applied Linguistics (2025), **45**, 1–9
doi:10.1017/S0267190525100172

CAMBRIDGE UNIVERSITY PRESS

INTRODUCTION

Artificial intelligence in applied linguistics: Applications, promises, and challenges

Andrea Révész[1], Shungo Suzuki[2] and Yeonwoo Jung[3]

[1]University College London, London, UK; [2]Nagoya University, Nagoya, Japan and [3]Sogang University, Seoul, South Korea
Corresponding author: Andrea Révész; Email: a.revesz@ucl.ac.uk

Recent advancements in artificial intelligence (AI) technologies such as generative AI (GenAI) have sparked a surge of interest within the field of applied linguistics. More and more researchers across various subfields are exploring its potential applications, carefully evaluating both the opportunities it offers and the challenges it presents. Given this heightened interest in AI among applied linguists, it appeared timely to devote this year's issue to this dynamically developing area. To give a snapshot of various perspectives on AI within applied linguistics, we invited experts from a range of subfields, including computational sociolinguistics, corpus linguistics, digital literacies, global Englishes, intercultural communication, language teaching, language assessment, natural language processing (NLP), and second language acquisition, to explore the role of AI in their work and its intersections with their research. The contributors adopted varied formats, such as position papers, theoretical analysis, empirical research, and narrative and systematic reviews, to consider the ethical use and/or potentially negative consequences of AI technologies within their subfield, with a view to identifying constructive and effective pathways forward for applying AI in applied linguistics.

The contributions center around three main themes. The first five papers emphasize the need to critically engage with AI technologies, highlighting key challenges arising from the implicit biases embedded in GenAI tools. The following four articles explore the role of AI in second language (L2) learning and teaching, with the final one also addressing its transformative impact on assessment. This paper serves as a bridge to the next five contributions, which examine potential applications of AI in automated language assessment practices, one of the fields most actively exploring its integration. The issue concludes with a piece critically appraising the benefits and pitfalls of integrating AI into applied linguistics research as a whole. Below, we provide a brief summary of each article.

Implicit bias in GenAI: A need for critical engagement

Taking the perspective of agentive literacy practices, Darvin's position paper provides a critical appraisal of GenAI in the context of human–AI interaction. Given GenAI's

© The Author(s), 2025. Published by Cambridge University Press.

https://doi.org/10.1017/S0267190525100172 Published online by Cambridge University Press

functionality to create texts in response to user input, thereby potentially influencing learners' identity construction, Darvin discusses the "relational, emergent, and distributed" nature of agency in human–AI interaction and underscores the need for L2 learners to negotiate their agency in these interactions. To unpack the affordances and constraints of GenAI, Darvin adopts Darvin and Norton's (2015) model of investment and Fenwick's (2015) concept of sociomateriality to integrate issues of power, agency, and inequality in human–AI interaction. Darvin argues that learners must not only recognize and utilize the potential of GenAI in AI-mediated L2 learning but also develop critical digital literacies to actively resist the implicit biases and power structures present in GenAI tools.

Moving into the domain of intercultural professional communication, Dai et al. investigated the extent to which human–AI interaction may influence the formation of human cultural stereotypes. Using Sequential-Categorical Analysis, a combination of Conversation Analysis and Membership Categorisation Analysis, the authors analyzed interactions between a Chinese-L1 human physiotherapist "Lisa" and three ChatGPT-generated patients. Situated in multicultural communities in Australia, the chatbot was asked to play the role of patients from three different cultural backgrounds: Chinese, a culture Lisa identified with; Australian, which she perceived as mainstream; and Indian, with which she was unfamiliar. The authors observed that human–AI interaction can strengthen and reproduce cultural biases inherent in the datasets on which large language models (LLMs) are based. In light of this, Dai et al. argue that GenAI users need to develop what they term Critical Interactional Competence, with a view to raising humans' awareness of the cultural stereotypes likely to arise in human–AI interactions and fostering the ability to critically reflect on and revise these cultural predispositions.

Alvero et al.'s position paper builds on previous research in AI, text analysis, and sociolinguistics to examine the implications of GenAI use in writing. The authors introduce and explore two theoretical constructs: *digital accents* and *homogeneity-by-design*. They use the term *digital accents* to describe the phenomenon that a writer's social identity is reflected in the linguistic features of their written language. Using this conceptual framework, the authors highlight LLMs' systematic failure to portray the range of linguistic variation found in human writing, a problem they refer to as *homogeneity-by-design*. The paper argues that this built-in homogenization in LLMs potentially reinforces hierarchies of language use, overrepresenting the linguistic practices of those in positions of social power while marginalizing diversity of linguistic expression, a trend reinforced by financial incentives to ensure AI models are predictable and resemble those in power. Given the potentially harmful biases and consequences associated with LLMs, Alvero et al. call on social scientists in general and applied linguists in particular to engage in critical analysis of these models. They emphasize the importance of achieving an in-depth understanding of LLMs and thereby enable the provision of informed critique and expertise to computer scientists, ultimately facilitating the ethical development of AI technologies.

Drawing also on findings from sociolinguistics and AI, Kang and Hirschi shift the focus from writing to speech, examining how social biases, especially those related to accented speech, are reproduced and, in some cases, amplified in AI systems. Building

https://doi.org/10.1017/S0267190525100172 Published online by Cambridge University Press

on the theoretical frameworks of linguistic stereotyping and reverse linguistic stereotyping, the authors synthesize extensive evidence showing that listeners' ethno-racial expectations about speakers can distort speech perception and lead to discriminatory evaluation and decision-making in various settings such as education, employment, and health care. The paper then connects these well-documented human biases to current AI technologies, including automatic speech recognition systems and LLMs, arguing that biases embedded in training data, model architecture, and validation processes contribute to unequal performance across accent and language varieties. For example, the accuracy of AI speech recognition tools can dramatically vary, depending on the discrepancy between the training data of the model and user's accents and gender. The authors conclude with a call for applied linguists to critically engage with AI tools and collaborate with AI researchers and practitioners for more socially responsible and ethical AI technologies and their use.

In addition to confirming the linguistic biases inherent in LLMs, Lee et al. explored techniques to address this issue, adopting a Global Englishes Language Teaching perspective. Specifically, the researchers investigated ChatGPT's ability to create English-language teaching activities and materials that reflect the diversity of how English is used across global communities. The study trialed a combination of prompt engineering and the integration of external English as Lingua Franca (ELF) corpus data to mitigate ChatGPT's inherent tendency to produce outputs corresponding to native speaker norms, which often fail to capture the linguistic diversity critical for global language learning. After repeated refinements of prompts and continued use of the ELF dataset, ChatGPT generated considerably more diverse and culturally acceptable outputs. Nevertheless, the process did not fully eliminate biases favoring standard language use, highlighting the importance of involving human professional expertise and additional resources to make the most of the affordances of GenAI for Global Englishes-informed language teaching.

AI applications in L2 learning and teaching

Also using corpus data, Verratti-Souto et al.'s article takes us to the domain of AI applications in L2 learning research. The aim of their study was to reevaluate, the findings of the English Grammar Profile (EGP) project, a large-scale investigation into what grammatical forms L2 English learners can competently use at different proficiency levels, as defined by the Common European Framework of Reference for Languages (CEFR). Employing NLP, the team designed a system to automatically extract grammatical forms which had previously been mapped to CEFR levels in the EGP project. Analyzing a learner subcorpus, including written texts by English as a Foreign Language (EFL), learners from different proficiency levels enrolled in an online English school, the researchers determined the CEFR levels at which learners began using each grammatical structure significantly more as compared to the preceding CEFR level. The overarching goal was to assess the overlap between the CEFR levels assigned to grammatical structures in the EGP and the levels emerging from the automatic NLP analysis. While the authors found limited agreement between their CEFR-level assignment and that of the EGP, they called for further research using large-scale learner data to verify EGP's level assignment.

https://doi.org/10.1017/S0267190525100172 Published online by Cambridge University Press

Turning to research on the role of AI in classroom learning, Huang and Mizumoto report on an empirical study that examined the relationships between motivation, engagement, and AI utilization in the Japanese EFL context. The study involved 174 participants, who attended two writing workshops during the semester, each lasting 2 weeks. In these workshops, groups of students worked on essays collaboratively and received feedback on their writing using ChatGPT. At the end of the semester, students were asked to complete a questionnaire assessing their motivation and engagement during the workshops. Results from structural equation modeling confirmed that motivation had a positive influence on students' AI usage, which, in turn, impacted all three aspects of students' engagement (affective, behavioral, and cognitive). Huang and Mizumoto's findings suggest that the use of GenAI may serve as a mediating factor between students' motivation and different aspects of their engagement. Based on these results, the researchers concluded that integrating GenAI as a supporting tool is feasible in EFL writing classes, potentially resulting in increased student engagement.

In the next contribution, Fang and Han provide a narrative synthesis of the initial stages of the fast-evolving research on the applications of ChatGPT in foreign language teaching and learning. The researchers' aim was to reflect on current research findings, identify conceptual and methodological issues in research, and outline directions for future inquiry. Based on a review of 71 publications including empirical studies, reviews, position papers, and commentaries, Fang and Han observed several key trends: (a) the majority of studies examined students' and teachers' perceptions of ChatGPT's potential to assist learning and teaching; (b) most investigations took the form of qualitative research, with a few quasi-experimental studies beginning to explore the value of integrating ChatGPT into teaching; (c) descriptive, exploratory, and context-specific research were most widespread; and (d) among various L2 skills, the application of ChatGPT in teaching writing has garnered the most interest. The researchers end the piece by calling on the research community to join forces and engage in systematic investigations into the capabilities of ChatGPT to advance foreign language learning and teaching.

Zooming in on L2 listening and speaking, Goh and Aryadoust's position paper examines the transformative potential of GenAI in teaching, learning, and assessment. The authors offer a comprehensive review of GenAI technologies, such as spoken dialog systems, intelligent personal assistants, and LLMs, highlighting their applications in these domains. Drawing on theoretical frameworks of listening and speaking, the paper evaluates how AI systems align with the complex cognitive, social, and affective processes underpinning oral language competence. While GenAI offers opportunities for interaction, personalized feedback, and task repetition, the authors argue that it also raises significant concerns around authenticity, cognitive validity, data privacy, and pedagogical alignment. Emphasizing the importance of teacher expertise and AI literacy, they propose co-intelligence – collaborative use of AI with human oversight – as a guiding principle for ethical and effective integration. The paper concludes with a call for future research to address not only technological and cognitive aspects but also sociocultural equity and theoretical development in AI-mediated L2 learning and assessment.

https://doi.org/10.1017/S0267190525100172 Published online by Cambridge University Press

AI in language assessment

As the first in a series of papers with a narrower focus on language assessment, Xi's position paper considers the notion of communicative competence in light of recent advances in AI. Xi argues that the conventional definition of communicative competence is fundamentally challenged in today's digital age, where real-world communication is facilitated and enhanced by AI technologies in various domains. To address this shift, Xi reviews four conceptual approaches to defining language ability (task-based, trait-based, interactionalist, and psycholinguistic) and advocates an AI-mediated interactionalist approach, which integrates AI literacy skills (i.e., knowledge of how to utilize AI tools) with broadened linguistic and cognitive skills (i.e., the ability to critically evaluate and modify AI outputs). Xi discusses how the redefinition of communicative competence can better align with recent technological advancements, allowing for a more thorough assessment of communicative competence. She also presents a nuanced perspective when applying this expanded construct, stressing the importance of considering test stakes and target proficiency levels when deciding whether test takers should be given access to editing or generative tools. In addition, the paper highlights several challenges that language testers need to address, particularly in high-stakes, large-scale testing, such as disparities in digital access and AI literacy among learners. Finally, Xi calls for collaborative efforts between language testers and practitioners to develop practical guidelines and drive transformation in language testing, ensuring that assessments reflect the realities of AI-mediated communication.

The next three empirical papers explore the practical applications of AI in a variety of assessment contexts. Drackert et al. investigated the potential of GenAI in developing high-quality input texts for reading comprehension assessments in German as a foreign language. They systematically compared texts generated by ChatGPT (versions 3.5 and 4) with traditional human-written benchmark texts employed in TestDaF, a standardized German test used for university admission purposes in Germany. The texts were analyzed for a range of linguistic features, including measures of readability, lexical diversity, and morphosyntactic complexity, and were assessed by experts. Compared to human-created texts, AI-generated texts exhibited higher lexical density and greater syntactic complexity. However, they also exhibited overuse of nominalizations and nonidiomatic expressions derived from direct English translations. While the authors acknowledge the efficiency benefits of AI, they emphasize the necessity of expert manual revisions, especially for languages other than English, given the issues arising from LLM datasets primarily trained on English-language data. Drackert et al. suggest that future research should refine prompt engineering techniques and conduct longitudinal studies to assess advancements in AI-generated educational content.

Saricaoglu and Bilki's empirical study provides insights into the value of using GenAI in writing assessment. The authors examined the assessment capacity of ChatGPT-4 by evaluating the accuracy and specificity of its feedback in identifying learners' strengths and weaknesses, as well as its alignment with the predetermined assessment criteria of IETLS Writing Task 2 across four dimensions of L2 writing (*task response, coherence and cohesion, lexical resource*, and *grammatical range and*

https://doi.org/10.1017/S0267190525100172 Published online by Cambridge University Press

accuracy). Through qualitative and quantitative analyses of 1,795 pieces of feedback on 35 argumentative essays written by upper-intermediate L2 learners, the study demonstrated that ChatGPT-4 was highly accurate in detecting issues across all dimensions and delivered especially relevant feedback for the criteria in the task response and coherence and cohesion dimensions. However, ChatGPT-4 showed variability in feedback specificity. It generated detailed feedback for the categories of grammatical range and accuracy and lexical resource, while its comments on task response and coherence and cohesion remained general. The authors highlight ChatGPT-4's potential for automating precise linguistic feedback but emphasize persistent challenges, including insufficient specificity in discourse-level feedback and occasional misclassification between grammatical and lexical errors. Saricaoglu and Bilki call for future research to investigate recall accuracy and assess the generalizability of findings across different learner populations and assessment criteria.

Continuing with the topic of developing assessments to evaluate L2 literacy skills, Zhang et al. report on a project exploring how GenAI can support the implementation of a diagnostic English-language assessment, aimed at identifying university students in need of additional academic language support in tertiary education in the context of New Zealand. The first study reported in the paper investigated the suitability of AI for generating reading texts and related test items to assess academic reading ability. The second study examined the comparability of scores generated by automated writing evaluation versus human raters. The project yielded promising findings for AI use in terms of content generation for reading assessments and evaluation of overall writing quality. However, the research also revealed a need for human involvement to fine-tune reading test items and provide feedback that is sufficiently specific to be actionable by L2 writers. The team concluded that, while AI has the potential to promote the adaptivity and personalization of assessments and enhance the efficiency of the test development and implementation process, it is essential to integrate human expertise with AI to maintain a fair, ethical, and responsible approach to assessment.

As reflected in Zhang et al.'s conclusions, assessment experts show a keen interest in ethical integration of AI in language assessment. Addressing this issue, Pastorino and Galaczi propose an ethical framework for the use of AI in language assessment, motivated by both its potential benefits and the negative consequences associated with its use. To this end, the authors conducted iterative inductive and deductive thematic analyses of existing documents, such as international policy and professional standards from general, educational, and language assessment domains. Based on the resulting codes, they identified themes and subthemes then formulated 10 principles and related considerations, covering both domain-general principles (e.g., government and stewardship, human control of technologies) and language assessment-specific ones (e.g., assessment standards). The paper also critically evaluated these principles and considerations across three common scenarios of AI use in language assessment: test item and task creation, automated scoring, and accessible assessment for different populations of test-takers. The authors conclude by highlighting major challenges in applying their principles to practical implementation and recommend that test designers adopt an "ethical-by-design" approach to protect the interests of stakeholders.

https://doi.org/10.1017/S0267190525100172 Published online by Cambridge University Press

Taking stock: Opportunities and challenges of AI and GenAI in applied linguistics

In the final article of the issue, Curry et al. offer a critical appraisal of AI for applied linguistics research, discussing both the affordances and the challenges inherent in the integration of AI and, in particular, GenAI with applied linguistics research. They highlight potential misalignments between applied linguistics and GenAI, specifically, in terms of epistemology, ontology, and ethical traditions, all of which govern researchers' decision-making practices and actions. For instance, contemporary applied linguistics research is oriented toward descriptive, socialized, and contextualized views of language and appreciates the creativity, reflexivity, and criticality of human researchers. By contrast, GenAI, by default, may generate normative language and worldviews represented in its training data, potentially decreasing cultural and linguistic diversity. The authors also evaluate this misalignment in the context of language learning and teaching, emphasizing the need for stakeholders such as educators, researchers, and policymakers to develop a critical awareness of the epistemological, ontological, and ethical concerns associated with GenAI applications. Finally, they propose that applied linguists should focus on enhancing alignments between their research and use of GenAI rather than aiming for the automatization of applied linguistics research procedures and replacement of experts. They make this argument on the grounds that applied linguistics research is expected to critically evaluate language, texts, and their social implications. Curry et al.'s conclusions closely align with our own and reflect the broader perspectives expressed in this issue.

Concluding remarks

As a result of the rapid integration of AI into real-world communication, education, and research, we, as applied linguists, are inevitably confronted with the question of how to engage with AI technologies. To address this question, we must reflect on how we conceptualize AI as a tool, a collaborator, an agent, or, especially in the case of GenAI, a mirror reflecting the human values and biases embedded in its training data.

As several contributions in this issue illustrate, AI technologies offer unprecedented opportunities for applied linguistics research and its applications. In the fields of language education and assessment, AI can support the development of adaptive and personalized materials and feedback, potentially promoting greater accessibility and inclusivity. In addition, AI enables automated evaluation of L2 performance and provision of feedback, enhancing the scalability of teaching materials and assessments. Beyond language teaching and assessment, AI also has the capacity to reshape research practices. For example, applied linguists increasingly employ AI for time-intensive tasks such as transcription, coding, and facets of corpus analysis.

However, AI technologies also come with risks. The potential benefits of AI are not equitably distributed; the effective use of AI tools depends on a certain level of digital literacy and access to technological infrastructure, deepening the existing digital divide. If AI systems are designed and implemented without considering accessibility and support, they may reinforce structural inequities. Also, LLMs are largely trained on English-language datasets that favor standard language use, often failing to account for the linguistic and cultural diversity inherent in real-world language use. Thus, GenAI

https://doi.org/10.1017/S0267190525100172 Published online by Cambridge University Press

tools may entrench the implicit biases and power structures present in the datasets on which they are trained. Additionally, GenAI technologies remain prone to generating errors and hallucinations, lacking the capacity to fully replicate the nuances of human judgment, knowledge, and communication.

As several contributors to this issue have noted – views we also share – it is essential to mitigate these risks through interdisciplinary collaborations with AI experts, by integrating human expertise with AI, and by fostering critical engagement with AI tools in research as well as practice. We believe that applied linguists are well-positioned to lead these efforts through bridging technical innovation and humanistic inquiry through the lens of language, ensuring that we achieve fair, ethical, and responsible AI use. Rather than viewing AI as a threat to human expertise, we see it as a catalyst that prompts us to reflect on our disciplinary identity and contributions.

Acknowledgements and notes from the editorial team

This issue is Andrea Révész's first volume as editor of the Annual Review of Applied Linguistics (ARAL), succeeding Professor Alison Mackey after nearly a decade of inspiring and innovative leadership. Shungo Suzuki served as guest associate editor, bringing expertise on this year's theme in a newly created role to support the editor. Going forward, this position will be held by another early- or mid-career scholar with expertise on the theme of the subsequent issue. Yeonwoo Jung took on the role of editorial assistant. As in previous years, the editorial board played a vital role in determining and shaping the strategic direction of the journal. Members of the editorial board advise about the theme and content of each volume and, from this issue onward, will also make recommendations for the guest associate editor. Once the theme for the annual issue is established, editorial board members are invited to offer suggestions about potential authors and reviewers and may serve as reviewers themselves. We are grateful to the editorial board for their outstanding contributions to this issue, especially to Anne Charity-Hudley, Aek Pakhiti, Luke Plonsky, Bryan Smith, Hansun Waring, and Nicole Ziegler for generous advice and recommendations.

Although the articles in this issue were invited, each was evaluated by at least two blind reviewers. We would like to thank Vahid Aryadoust, Aaron Olaf Batty, Tineke Brunfaut, Alessia Cogo, Alvin Grissom II, Ari Huhta, Lianjiang George Jiang, Rodney Jones, Detmar Meurers, Akira Murakami, Michael Pace-Sigge, Ana Pelicer-Sánchez, Aek Phakiti, Vijay Ramjattan, John Rogers, Ali Fuad Selvi, Huiyang Shen, Lynda Taylor, Aki Tsunemoto, Hansun Waring, Csilla Weninger, Nicole Ziegler, and 10 anonymous reviewers, whose thorough and insightful comments led to refined versions of all articles.

We also appreciate the invaluable support of Amy Laurent, the journal's managing editor at Cambridge University Press. In addition, we are grateful to the American Association for Applied Linguistics (AAAL) Executive Committee for their ongoing support, especially to Peter de Costa who, as AAAL president, played a key role in renewing the contract between ARAL and AAAL, ensuring a continued close relationship between the journal and the association. We also thank Mari Haneda as well as Peter de Costa and Stephanie Link for inviting us to organize a colloquium at the

https://doi.org/10.1017/S0267190525100172 Published online by Cambridge University Press

face-to-face and virtual AAAL 2026 conferences, respectively, based on the contributions of this volume. Above all, we are immensely grateful to Alison Mackey and Erin Fell, the previous editor and editorial assistant, who helpfully introduced us to the workings of ARAL and generously shared all their materials.

References

Darvin, R., & Norton, B. (2015). Identity and a model of investment in applied linguistics. *Annual Review of Applied Linguistics*, *35*, 36–56. https://doi.org/10.1017/S0267190514000191

Fenwick, T. (2015). Sociomateriality and learning: A critical approach. In D. Scott & E. Hargreaves (Eds.), *The SAGE handbook of learning* (pp. 83–93). SAGE. https://doi.org/10.4135/9781473915213.n8

Cite this article: Révész, A., Suzuki, S., & Jung, Y. (2025). Artificial intelligence in applied linguistics: Applications, promises, and challenges. *Annual Review of Applied Linguistics*, *45*, 1–9. https://doi.org/10.1017/S0267190525100172

https://doi.org/10.1017/S0267190525100172 Published online by Cambridge University Press