Language Teaching Research Quarterly **EUROKD**

2026 (56), 66–83
https://doi.org/10.32038/ltrq.2026.56.04

Reframing Generative AI Bias as a Pedagogical Artifact for Critical AI Literacy in TESOL: A Conceptual Framework and Research Agenda

Jessie S. Barrot*
National University, Philippines
Hassan Mohebbi
European Knowledge Development Institute, Turkey

Correspondence
Email: jessiebarrot@yahoo.com

Abstract

This article reconceptualizes bias in generative artificial intelligence (GenAI) as a sociotechnical and pedagogical phenomenon rather than merely a technical flaw to be mitigated. Drawing on critical language awareness, critical applied linguistics, Global Englishes scholarship, and emerging research on AI-mediated language learning, it examines how bias is shaped by training data, language ideologies, institutional priorities, user practices, and broader social inequalities. In TESOL contexts, GenAI may privilege standardized English norms, marginalize linguistic diversity, reproduce culturally narrow representations, and weaken learner agency when its outputs are accepted uncritically. However, we argue that biased AI outputs can also serve as pedagogical artifacts for examining language, culture, identity, and power. To explain this potential, we propose the Pedagogical Transformation Framework of AI Bias, which conceptualizes learners' movement from invisible bias to critical AI literacy through five stages: invisible bias, bias recognition, critical interrogation, transformative engagement, and critical AI literacy. From this perspective, we discuss its applications in writing, communication, intercultural learning, and assessment, while highlighting the central role of teacher agency and ethical safeguards. It concludes that the educational significance of AI bias depends not only on its presence but on how teachers and learners critically engage with and transform the assumptions embedded in AI-generated discourse.

ARTICLE HISTORY
Received: 28 April 2026
Revised: 18 August 2026
Accepted: 25 August 2026

KEYWORDS
Generative Artificial Intelligence, Bias, Computer-Assisted Language Learning, Writing Assistant, Critical AI Literacy

How to cite this article (APA 7th Edition):
Barrot, J. S., & Mohebbi, H. (2026). Reframing Generative AI bias as a pedagogical artifact for critical AI literacy in TESOL: A conceptual framework and research agenda. *Language Teaching Research Quarterly*, *56*, 66–83. https://doi.org/10.32038/ltrq.2026.56.04

Introduction

The rapid diffusion of generative artificial intelligence (GenAI) tools such as ChatGPT, Gemini, and Claude has introduced significant changes to English language teaching and learning. These technologies have expanded access to linguistic support, provided immediate feedback on language production, facilitated writing development, and created opportunities for interaction across diverse communicative contexts (Al-Khresheh, 2024; Barrot, 2024; Dong, 2026; Kostka & McMartin-Miller, 2026; Law, 2024; Pack & Maloney, 2023; Pitychoutis & Spathopoulou, 2026; Seo, 2024). Their growing presence in TESOL has also sparked substantial debate over authorship, academic integrity, teacher roles, and the ethical implications of AI-mediated learning (Barrot, 2023; Bittle & El-Gayar, 2025; Yeo, 2023). Among these concerns, bias has emerged as one of the most persistent and complex issues.

Current discussions often conceptualize bias as a technical defect that requires detection, mitigation, or elimination (Afreen et al., 2025). Such a view assumes that bias exists primarily within algorithms and training data and that educational concerns can be resolved through improvements in model design. This perspective, however, offers only a partial account of how bias operates in language education. Research in applied linguistics and AI studies indicates that GenAI systems do not merely generate language; they also reproduce linguistic ideologies, cultural assumptions, and social hierarchies embedded within the data, institutions, and communities that shape their development (Jenks, 2025; Lee et al., 2025). A sociotechnical phenomenon refers to a process, system, or outcome that emerges from the interaction between technological structures and social forces (Boy, 2026). Its effects extend beyond algorithmic outputs to influence whose language practices receive legitimacy, whose cultural knowledge gains visibility, and whose communicative norms acquire authority.

This issue carries particular significance in TESOL. English language education has long grappled with questions of linguistic legitimacy, native-speakerism, cultural representation, and unequal distributions of symbolic power (Canagarajah, 1999; Galloway & Rose, 2015). GenAI systems frequently privilege standardized varieties of English, especially American and British norms, while providing limited representation of World Englishes, multilingual practices, and localized forms of communication (Lee et al., 2025; Rivero & Yin, 2025). Such tendencies risk reinforcing existing hierarchies that TESOL scholars have challenged for decades. As Dixon-Román et al. (2020) observed, AI algorithms can function as a "racializing assemblage" that imposes standards of writing aligned with white, middle-class literacy norms. Thus, uncritical acceptance of AI-generated outputs may perpetuate linguistic exclusion, cultural misrepresentation, and reduced learner agency.

At the same time, an exclusive focus on harm risks overlooking an important pedagogical possibility. Critical traditions in language education have long treated texts, discourses,

and communicative practices as sites through which learners can examine relations of power, ideology, and identity (Fairclough, 1992; Pennycook, 2004). AI-generated texts warrant similar scrutiny. Biased outputs can function as pedagogical artifacts that make underlying assumptions visible and open to critique. As Kohnke and Zou (2025) argued, teachers can use their pedagogical expertise to contextualize AI feedback and adapt it to diverse learners' linguistic needs. This fosters linguistic inclusivity rather than merely accepting AI-generated corrections as authoritative. Examination of whose voices are represented, whose perspectives are marginalized, and whose norms are privileged can support the development of critical language awareness, intercultural understanding, and ethical engagement with AI-mediated communication.

This article argues that the educational significance of AI bias depends not solely on its presence but on how teachers and learners engage with it. Bias can reproduce inequities when it remains invisible or unquestioned. It can also serve as a catalyst for critical inquiry when pedagogically mediated through reflective analysis, dialogue, and prompt design. The article reconceptualizes AI bias as a sociotechnical and pedagogical phenomenon and examines how TESOL educators can transform encounters with biased AI outputs into opportunities for critical AI literacy, critical language awareness, and learner agency.

Reconceptualizing AI Bias as a Sociotechnical Phenomenon

Discussions of bias in GenAI frequently portray it as a technical problem that originates from flawed datasets, imperfect algorithms, or inadequate model training procedures (Afreen et al., 2025). This perspective has shaped much of the contemporary discourse on AI ethics, where efforts often focus on bias detection and mitigation. Although such efforts remain important, a purely technical interpretation offers an incomplete understanding of how bias operates within educational contexts. Language does not exist outside sociocultural and political structures. Consequently, AI systems trained on human-produced language inevitably reflect the values and assumptions embedded within the societies that generate their training data.

Scholarship in critical applied linguistics has long demonstrated that language functions as a site of ideological struggle rather than a neutral medium of communication (Fairclough, 1992; Pennycook, 2004). Questions about which language varieties receive legitimacy, whose voices acquire authority, and which communicative norms become institutionalized remain deeply connected to broader relations of power. GenAI systems inherit these dynamics because their outputs emerge from large-scale collections of texts that reflect dominant sociocultural and linguistic practices.

Evidently, recent research on AI and language education supports this broader perspective. Lee et al. (2025) rightly argued that GenAI systems frequently privilege standardized varieties of English while offering limited representation of linguistic

diversity. Such tendencies reinforce ideologies that position American and British English as normative models of language use. Rivero and Yin (2025) likewise contend that AI-generated language often reproduces assumptions associated with monolingual and native-speaker-oriented paradigms. These tendencies do not emerge solely from technical design. They reflect historical inequalities in linguistic representation, knowledge production, and technological development. Zhang et al. (2026) similarly demonstrated that large language models (LLMs) do not merely reflect culturally entrenched ideologies but actively participate in shaping them within educational discourse, with different models manifesting bias in culturally specific ways.

A sociotechnical perspective situates AI bias within a network of interacting influences. Training data constitute one important source of bias, but data alone cannot explain how particular outputs acquire authority. Institutional priorities, commercial interests, regulatory frameworks, user interactions, and dominant language ideologies also shape the development and deployment of AI systems (Jenks, 2025). Outputs that privilege standardized English, for example, reflect not only statistical patterns within training corpora but also longstanding educational traditions that equate linguistic legitimacy with conformity to native-speaker norms.

This issue is particularly significant in TESOL. The field has increasingly challenged deficit views of multilingualism and has advocated recognition of linguistic diversity through perspectives such as World Englishes, English as a Lingua Franca, and Global Englishes Language Teaching (Canagarajah, 1999; Galloway & Rose, 2015). GenAI systems often operate according to assumptions that conflict with these developments. Corrections of localized lexical choices, grammatical constructions, or discourse conventions frequently align with standardized norms even when alternative forms remain contextually appropriate. Such responses risk marginalizing multilingual identities and reducing the legitimacy of diverse English varieties. Learners may consequently internalize narrow conceptions of communicative competence that privilege conformity over linguistic flexibility.

Cultural representation constitutes another important dimension of sociotechnical bias. GenAI systems often draw upon knowledge structures that overrepresent perspectives from economically and technologically dominant regions. Responses related to communication, professionalism, education, or social interaction might therefore reflect assumptions rooted in Western contexts while overlooking alternative cultural frameworks (Jenks, 2025). Such patterns influence how learners understand language use across cultural settings and may contribute to the normalization of particular worldviews. Zhang et al. (2026) found that both Western and non-Western LLMs exhibited an inherent positivity bias that framed culturally and linguistically responsive teaching as either depoliticized multiculturalism or superficial harmony, revealing how bias operates differently across cultural contexts.

Teacher and learner interactions with AI systems further shape the manifestation of bias. Users do not encounter AI outputs as passive recipients. Interpretations, prompt choices, classroom practices, and institutional expectations influence how AI-generated texts are understood and utilized. Wang et al. (2025) note that teacher beliefs and pedagogical decisions significantly affect the educational consequences of GenAI integration. This claim is corroborated by empirical human-AI collaboration research outside TESOL. Beck et al. (2026) reported that reviewers' pre-existing attitudes toward AI predict their ability to catch erroneous AI suggestions more strongly than any demographic factor, and even small increases in the effort required to correct AI output measurably reduce correction rates. Similarly, Romeo and Conti's (2026) review of automation bias found that verification effort and AI literacy, not the AI system's technical quality, are the strongest predictors of whether biased AI output is accepted or challenged. The significance of a biased output depends not only on what the system produces but also on how teachers and learners respond to it. Educational contexts can either reinforce problematic assumptions or create opportunities for critical examination. Todd (2025) emphasized that preparing for GenAI-influenced language education necessitates training in AI literacy, prompt engineering, and critical evaluation, while cautioning that teacher training in GenAI lags far behind its applications.

A sociotechnical understanding of AI bias carries important implications for TESOL. Bias should not be viewed exclusively as an error to be eliminated. Such a perspective risks obscuring the social and ideological processes that produce inequity in the first place. Critical examination of biased outputs can reveal underlying assumptions about language, culture, identity, and power. This perspective aligns with traditions of critical language awareness that seek to make ideological dimensions of discourse visible and open to scrutiny (Fairclough, 1992).

This reconceptualization shifts the focus from whether bias exists to how bias functions within AI-mediated language education. Such a shift provides a foundation for understanding how encounters with AI bias can support the development of critical AI literacy, critical language awareness, and learner agency. The educational challenge does not concern the complete eradication of bias. The challenge concerns cultivating pedagogical approaches that enable teachers and learners to interrogate and transform the assumptions embedded in AI-generated discourse.

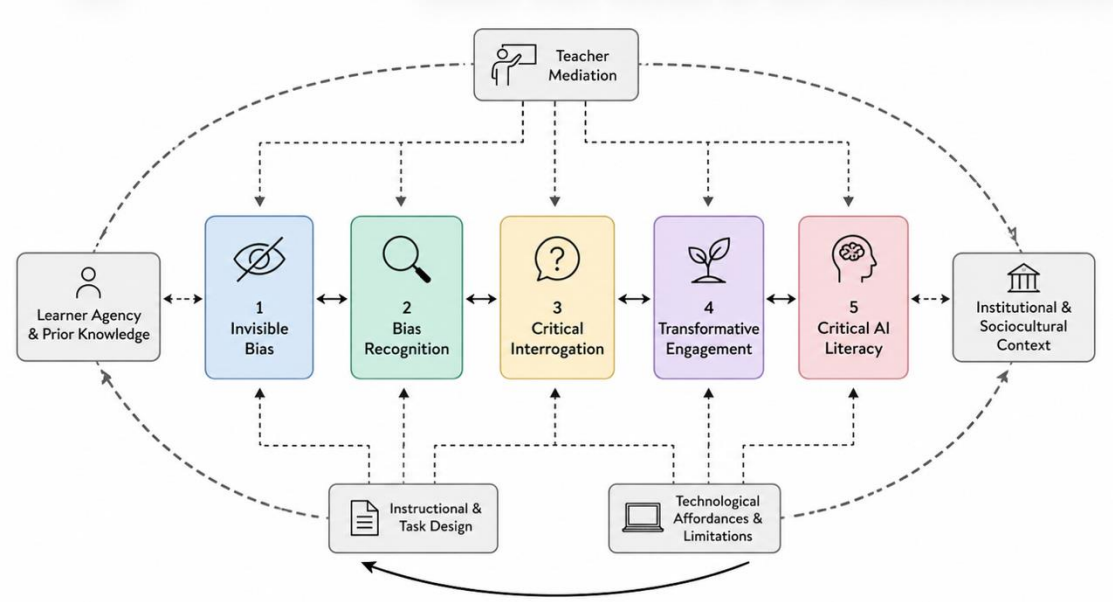
A Conceptual Framework of Pedagogical Transformation

Educational value emerges from the ways teachers and learners recognize and respond to biased outputs. A biased response generated by an AI system can reinforce linguistic and cultural inequalities when users accept it as authoritative. The same response can also function as a catalyst for critical inquiry when learners examine the assumptions that shape it.

To explain this process, this article proposes the Pedagogical Transformation Framework of AI Bias (See Figure 1). The framework conceptualizes AI bias as a pedagogical artifact that moves through five interconnected stages: invisible bias, bias recognition, critical interrogation, transformative engagement, and critical AI literacy. The framework draws on critical language awareness (Fairclough, 1992), critical applied linguistics (Pennycook, 2004), critical approaches to English language teaching (Canagarajah, 1999), Global Englishes scholarship (Galloway & Rose, 2015), and emerging research on AI-mediated language learning (Lee et al., 2025; Wang et al., 2025). It explains how encounters with biased AI outputs can facilitate the development of linguistic, cultural, ethical, and technological awareness in TESOL contexts.

Figure 1

Pedagogical Transformation Framework of AI Bias



Stage 1: Invisible Bias

The process begins with invisible bias. At this stage, learners interact with AI-generated outputs without questioning the assumptions embedded within them. Responses generated by GenAI often appear authoritative because they exhibit fluency and linguistic sophistication. Such characteristics may encourage users to equate fluency with accuracy, neutrality, or objectivity. Learners may accept AI-generated recommendations, corrections, and explanations without examining their ideological foundations.

Invisible bias often manifests through the normalization of dominant linguistic and cultural norms. Standardized varieties of English may be preferred over localized forms, while Western communicative practices may be seen as universal models of effective communication (Lee et al., 2025). Learners who encounter such outputs may remain unaware of the sociocultural assumptions that shape them. Educational risks emerge when AI-generated discourse acquires unquestioned authority and displaces opportunities for critical reflection.

Stage 2: Bias Recognition

Educational transformation requires recognizing bias. This stage involves identifying linguistic, cultural, ideological, or representational patterns embedded in AI-generated outputs. Teachers play a central role in directing learners' attention toward features that might otherwise remain invisible.

Recognition may occur through comparative analysis of AI responses, examination of variations in prompts, or evaluation of alternative linguistic and cultural perspectives. For example, students may compare an AI-generated description of an "ideal English speaker" with descriptions that reflect multilingual or Global Englishes perspectives. Such comparisons expose assumptions regarding linguistic legitimacy and communicative competence. Bias recognition is an important threshold because critical examination cannot occur unless learners first identify underlying assumptions.

Stage 3: Critical Interrogation

Recognition alone does not guarantee critical understanding. Learners must also interrogate the social and ideological foundations of biased outputs. This stage draws heavily on critical language awareness, which emphasizes examination of the relationship between discourse and power (Fairclough, 1992).

Critical interrogation encourages learners to ask a series of analytical questions. Whose voices appear in the response? Which perspectives receive legitimacy? What linguistic norms receive preference? Which identities or experiences remain absent? Such questions shift attention from the content of AI-generated outputs to the assumptions that underpin them.

This stage also foregrounds the sociotechnical nature of AI bias. Learners examine how training data, institutional priorities, technological design, and historical power relations shape AI-generated discourse (Jenks, 2025). This means AI bias becomes evidence of broader social processes rather than a simple computational error.

Stage 4: Transformative Engagement

Critical interrogation creates opportunities for transformative engagement. At this stage, learners move beyond critique and actively reshape their interactions with AI systems. Educational activities emphasize agency and knowledge construction.

Prompt revision provides one example. Students may reformulate prompts that produce culturally narrow or linguistically restrictive outputs and then evaluate how responses change. Alternative perspectives can also emerge through tasks that require learners to challenge standardized assumptions and generate more inclusive representations of language use. Such activities position learners as active participants rather than passive consumers of AI-generated knowledge. Kohnke and Zou (2025) illustrated this through

their application of the TPACK and SAMR frameworks, showing how teachers can move from substitution (relying on AI for feedback) to modification or redefinition.

Stage 5: Critical AI Literacy

The final stage involves developing critical AI literacy. This outcome extends beyond technical competence in using AI tools. Critical AI literacy encompasses the ability to evaluate AI-generated information, recognize embedded assumptions, understand sociotechnical influences, and make informed ethical decisions regarding AI use.

Learners who reach this stage understand that AI systems are neither neutral nor infallible. They recognize that outputs reflect particular forms of knowledge, representation, and power. Such learners demonstrate greater capacity to engage with AI critically while maintaining awareness of linguistic diversity, cultural complexity, and ethical responsibility.

Critical AI literacy also strengthens learner agency. Rather than relying on AI as an unquestioned authority, learners develop the capacity to evaluate, adapt, and challenge AI-generated discourse in light of context, purpose, and audience. This outcome aligns with broader educational goals that emphasize critical thinking, innovation, intercultural competence, and informed participation in technologically mediated societies. As Lu (2025) argues, GenAI bias can itself become a resource for critical thinking when learners engage in activities such as comparative text analysis, discourse reconstruction, critical debate, and prompt engineering.

Dynamic Relationships across Stages

The framework does not assume a linear progression. Learners may move between stages as they encounter different AI systems and tasks. Teacher mediation remains central throughout the process. Pedagogical decisions determine whether bias remains invisible or serves as a stimulus for deeper critical engagement. Educational outcomes depend on the interaction among technological affordances, learner agency, and instructional design.

The Pedagogical Transformation Framework of AI Bias offers a conceptual lens through which TESOL scholars and practitioners can understand the educational potential of AI bias. The framework shifts attention away from the question of whether bias should be eliminated and toward how encounters with bias can support critical learning. Such a perspective positions AI-generated texts as sociocultural artifacts that reveal underlying assumptions about language, identity, culture, and power. Critical engagement with these artifacts creates opportunities to develop critical AI literacy and more equitable forms of AI-mediated language education.

Applications in TESOL

The pedagogical transformation of AI bias requires instructional practices that move beyond the instrumental use of GenAI. Educational applications should not focus exclusively on efficiency, accuracy, or language production. Such goals remain important, but they do not address the sociotechnical assumptions embedded within AI-generated discourse. A critical approach positions AI outputs as objects of inquiry through which learners examine language ideologies, cultural representations, and power relations. This section illustrates how the Pedagogical Transformation Framework of AI Bias can inform TESOL practices across writing instruction, speaking and communication, intercultural learning, and assessment.

Writing Instruction

Writing represents one of the most common applications of GenAI in language education. Learners frequently use AI systems to generate ideas, revise drafts, improve grammatical accuracy, and receive feedback on written texts (Law, 2024; Pack & Maloney, 2023). Nevertheless, AI-generated feedback often reflects assumptions about what constitutes appropriate academic or professional writing.

Research suggests that GenAI systems frequently privilege standardized forms of English and may discourage linguistic variation (Lee et al., 2025). Feedback may favor conventions associated with dominant English-speaking contexts while overlooking rhetorical traditions that emerge from multilingual and multicultural settings. Such tendencies create opportunities for critical analysis.

Teachers can encourage learners to compare AI-generated revisions with their original texts and examine the assumptions underlying suggested changes. Questions such as "What linguistic choices did the AI modify?" and "What norms appear to guide these revisions?" can facilitate critical reflection. Learners can also evaluate whether AI recommendations enhance clarity or simply promote conformity to dominant language standards. Such activities transform writing feedback from a corrective mechanism into a site of critical inquiry.

Prompt revision offers another pedagogical opportunity. Students may ask AI systems to evaluate the same text from different perspectives, such as a native-speaker-oriented perspective, a Global Englishes perspective, or an English as a Lingua Franca perspective. Comparative analysis of responses can reveal how language ideologies influence evaluations of writing quality.

Speaking and Communication

GenAI systems increasingly support speaking instruction through conversational simulations, dialogue generation, and communicative practice. Such tools provide learners with opportunities to engage in interaction beyond traditional classroom

settings. However, AI-generated dialogues often reflect culturally specific assumptions regarding professionalism and interpersonal communication.

Communicative norms vary across sociocultural contexts. Recommendations regarding disagreement, persuasion, requests, or workplace communication may privilege interactional conventions associated with dominant English-speaking societies. Learners who encounter these responses may incorrectly assume that such conventions possess universal applicability.

Critical engagement with AI-generated dialogues can help learners examine these assumptions. Teachers can present AI-generated conversations alongside authentic examples from different linguistic and cultural contexts. Comparative analysis allows students to identify variations in discourse practices and communicative expectations. Discussion can then focus on questions of appropriateness, audience, and context rather than on simplistic notions of correctness.

Intercultural Learning and Global Englishes

Intercultural competence constitutes a central objective in contemporary language education. GenAI systems frequently generate explanations of cultural practices, social norms, and communicative conventions. Such outputs may appear informative and comprehensive. Nevertheless, cultural representations often reflect dominant perspectives and may oversimplify complex social realities (Jenks, 2025).

A critical examination of AI-generated cultural descriptions can foster intercultural awareness. Teachers may ask learners to evaluate AI responses against local knowledge, personal experience, or alternative sources of information. Discrepancies between AI-generated content and lived realities can stimulate discussion regarding representation and cultural authority.

Global Englishes pedagogy offers a particularly useful framework for such work (Galloway & Rose, 2015). Students can analyze how AI systems represent English speakers from different linguistic and cultural backgrounds. Questions concerning whose voices receive visibility and whose perspectives remain marginalized can encourage reflection on linguistic diversity and global inequalities.

Research by Rivero and Yin (2025) suggests that critical engagement with AI-generated representations can support multilingual learners' capacity to challenge dominant language ideologies and develop alternative linguistic possibilities. Educational activities that foreground representation and diversity, therefore, contribute to broader goals of equity and inclusion within TESOL.

Assessment and Feedback Practices

The integration of GenAI into language assessment presents significant challenges and opportunities. Traditional assessment models often assume clear distinctions between human-authored and technology-assisted work. GenAI complicates these distinctions because learners increasingly compose texts through interaction with AI systems.

A critical perspective suggests that assessment should evaluate not only final products but also learners' decision-making processes. Reflective components can encourage students to explain how they used AI tools, why they accepted or rejected particular suggestions, and how they evaluated the quality of AI-generated feedback. Such practices shift attention from detection and surveillance toward critical engagement and responsible use.

Bias analysis can also become an assessment activity. Students may examine AI-generated responses and identify linguistic, cultural, or ideological assumptions embedded within them. Evaluation criteria can focus on critical reasoning, evidence-based argumentation, and awareness of sociotechnical influences. These tasks assess learners' capacity to engage critically with AI-generated discourse rather than their ability to reproduce AI-generated content.

This approach aligns with calls for assessment reform in AI-mediated educational environments (Toncelli & Kostka, 2024). Educational objectives increasingly require learners to evaluate, interpret, and critique information produced through AI systems. Assessment practices should therefore reflect these emerging competencies.

Implications for Teacher Practice

The applications discussed above highlight the central role of teacher agency in the pedagogical transformation of AI bias. Teachers function not as gatekeepers who simply approve or reject AI use but as mediators who create opportunities for critical engagement. Educational outcomes depend on how teachers frame AI-generated outputs, design learning activities, and facilitate classroom dialogue.

Teacher expertise remains essential because AI systems cannot independently cultivate critical awareness. GenAI can generate language, but it cannot determine which values should guide interpretation or which assumptions deserve scrutiny. Wang et al. (2025) and Barrot (2026) emphasize that teacher agency remains a crucial factor in shaping meaningful educational uses of AI. Therefore, effective integration requires teachers to possess not only technical knowledge but also a critical understanding of the sociotechnical dimensions of AI.

Risks and Ethical Boundaries

The pedagogical use of AI bias requires careful ethical delimitation. Although biased outputs can serve as resources for critical inquiry, this use does not imply that bias should be tolerated, normalized, or reproduced without constraint. Bias can cause real educational harm, especially for learners whose language varieties, cultural identities, or social positions already face marginalization. A critical approach must therefore distinguish between bias as an object of analysis and bias as a condition that learners must endure.

A first ethical risk concerns the possible normalization of inequity. If teachers present biased AI outputs as useful classroom materials without adequate critical framing, learners may interpret those outputs as acceptable representations of language and culture. This risk has particular relevance in TESOL because GenAI systems may privilege standardized English norms and weaken recognition of World Englishes, multilingual repertoires, and local communicative practices (Galloway & Rose, 2015; Lee et al., 2025). These outputs may reinforce native-speakerist assumptions and narrow conceptions of linguistic competence. Critical pedagogy must therefore make clear that the educational value of bias resides in critique, not in the biased content itself.

A second risk concerns learner harm. AI-generated stereotypes, deficit portrayals, or culturally narrow responses may affect learners who identify with the groups or language practices represented in problematic ways. Classroom use of biased outputs must therefore follow principles of care, consent, and contextual sensitivity. Teachers should avoid examples that expose learners to humiliation, identity threat, or stigmatization. Critical analysis should focus on discourse, ideology, and system behavior rather than on learners as representatives of particular groups. This distinction protects classroom dialogue from becoming personally burdensome for minoritized students.

A third ethical boundary concerns epistemic authority. GenAI systems often produce fluent and coherent responses that appear credible even when they reflect partial, biased, or inaccurate assumptions. This fluency can obscure the ideological nature of AI-generated discourse. Learners may assign unwarranted authority to AI outputs and treat them as neutral evidence. This risk aligns with broader concerns about trust and bias in LLMs (Jenks, 2025). Teachers must position AI-generated text as provisional, contestable, and context-dependent. Classroom tasks should require verification, comparison with other sources, and reflection on the limits of AI-generated knowledge.

A fourth risk concerns teacher agency and professional responsibility. Critical use of AI bias demands more than access to digital tools. It requires teachers who can mediate ethical discussion, guide discourse analysis, and connect AI outputs to broader issues of language, power, and culture. Without such mediation, AI use may deepen dependency on automated feedback and reduce the role of teacher judgment. Wang et al. (2025) and

Barrot (2026) emphasize that teacher agency remains central in educational responses to GenAI. Teachers must retain authority over instructional design, assessment criteria, and ethical safeguards.

A fifth boundary concerns assessment. Bias analysis can support critical AI literacy, but assessment tasks must not reward uncritical reliance on AI-generated content. Evaluation should focus on learners' capacity to identify assumptions, justify interpretations, evaluate sources, and make context-sensitive decisions. These criteria align with critical language awareness, which treats language as a social practice connected to power and ideology (Fairclough, 1992). Assessment should also require transparency regarding AI use so that learners can account for their decisions and demonstrate judgment and ethical responsibility.

These risks suggest that AI bias should not be used as a pedagogical resource under all conditions. Its educational value depends on explicit safeguards. Teachers should select examples carefully, prepare learners for critical analysis, establish norms for respectful dialogue, and provide alternative representations that affirm linguistic and cultural diversity. Institutions should also support teachers through professional development, policy guidance, and assessment reform. Thus, the ethical goal is not to celebrate bias but to transform encounters with bias into opportunities for critique and agency. Bias becomes pedagogically defensible only when it remains visible as a problem, subject to analysis, and linked to broader efforts toward linguistic equity. Without these conditions, the classroom use of biased AI outputs risks reproducing the very inequalities that critical TESOL seeks to challenge.

Implications for Research, Practice, and Policy

The reconceptualization of AI bias as a sociotechnical and pedagogical phenomenon has important implications for research, classroom practice, and policy in TESOL. It shifts attention from a narrow focus on bias detection to a broader concern with how learners, teachers, institutions, and AI systems interact within unequal linguistic and cultural conditions. This perspective requires TESOL scholarship to examine not only whether GenAI produces biased outputs but also how such outputs affect language ideologies, learner agency, teacher authority, and educational equity.

Realizing this broader agenda begins with research. Future research should move beyond descriptive accounts of AI use and examine the pedagogical consequences of biased AI outputs in specific TESOL contexts. Empirical studies need to investigate how learners interpret AI-generated feedback and cultural explanations across diverse linguistic backgrounds. These kinds of studies can reveal whether learners accept AI outputs as authoritative or develop critical awareness of the assumptions embedded within them. Research should also examine how GenAI systems represent multilingual repertoires and local communicative norms. Current scholarship suggests that GenAI may privilege

standardized English and dominant cultural perspectives (Lee et al., 2025; Rivero & Yin, 2025). However, more context-specific studies are necessary to understand how these tendencies appear across regions, proficiency levels, task types, and classroom settings. Another research priority concerns teacher mediation. Teacher beliefs, professional knowledge, and institutional conditions shape how GenAI enters classroom practice (Toncelli & Kostka, 2024; Wang et al., 2025). Future studies should examine how teachers design tasks that transform biased AI outputs into opportunities for critical language awareness. Longitudinal research can also investigate how sustained exposure to AI bias analysis affects learner agency, intercultural awareness, and ethical AI use.

These research priorities carry direct implications for classroom practice. Classroom practice should treat AI-generated outputs as texts for critical examination rather than as neutral sources of linguistic authority. Teachers can ask learners to compare AI responses across prompts, evaluate whose language norms receive preference, and examine how cultural assumptions shape AI-generated advice. These practices align with critical language awareness, which emphasizes the relationship between discourse, ideology, and power (Fairclough, 1992). Teachers also need to integrate prompt literacy into language pedagogy. Prompt design should not be reduced to a technical skill. It should function as a rhetorical and ethical practice that requires attention to audience, purpose, context, and representation. Learners should understand that different prompts can produce different assumptions about language, culture, and identity. Assessment practices should also change. Instead of relying primarily on AI detection or punishment, educators can require reflective documentation of AI use. Learners may explain which AI suggestions they accepted, rejected, or revised and why. This kind of reflection allows assessment to value judgment, authorship, source evaluation, and ethical responsibility. It also supports a more realistic view of literacy in AI-mediated environments.

Sustaining these research and classroom priorities, in turn, depends on institutional and policy support. TESOL institutions need policies that support critical and equitable AI integration. Restrictive policies that prohibit GenAI use may fail to prepare learners for AI-mediated communication. Unrestricted adoption may also reinforce linguistic bias, academic integrity concerns, and dependency on automated feedback. A balanced policy approach should define acceptable AI use and promote critical AI literacy. Teacher education programs should include systematic preparation for AI-related ethical and pedagogical issues. Such preparation should address bias, data ethics, authorship, linguistic diversity, and assessment reform. Teachers require more than operational knowledge of AI tools. They need conceptual and ethical frameworks that help them evaluate how AI systems shape language education. Policy makers should also recognize the importance of linguistic equity in AI adoption. GenAI tools used in TESOL should be evaluated for their treatment of multilingual practices and culturally diverse communicative norms. Institutions should avoid policies that position standardized

English as the only legitimate target of AI-assisted language development. Policy should instead affirm linguistic diversity and require critical examination of AI-generated norms. Research, practice, and policy must converge around a central principle: AI bias should neither be ignored nor accepted as inevitable. It should become an object of inquiry, a focus of teacher education, and a catalyst for more equitable language pedagogy. This approach positions TESOL not as a passive recipient of technological change but as a field capable of shaping AI use through critical, ethical, and context-sensitive educational practice.

Directions for Future Research

The reconceptualization of AI bias advanced in this article opens several avenues for future inquiry. The following questions are offered to guide subsequent research on bias, critical AI literacy, assessment, and ethical practice in AI-mediated TESOL:

1. How do multilingual English learners interpret and evaluate AI-generated feedback and cultural explanations that privilege standardized or Western-centric norms?
2. How do different GenAI systems represent World Englishes, English as a Lingua Franca, and localized communicative practices, and how consistent are these representations across platforms?
3. What pedagogical strategies most effectively support learners' movement through the five stages of the Pedagogical Transformation Framework, from invisible bias to critical AI literacy?
4. How do teacher beliefs, professional knowledge, and institutional conditions mediate learners' critical engagement with biased AI outputs?
5. What is the longitudinal impact of sustained instruction in AI bias analysis on learner agency, intercultural competence, and ethical AI use?
6. How can prompt literacy be systematically integrated into TESOL curricula, and what measurable effect does it have on learners' critical AI literacy?
7. What assessment models can validly evaluate learners' critical engagement with AI-generated discourse, as distinct from their ability to reproduce or comply with AI-generated content?
8. What ethical codes or institutional safeguards are needed to ensure that pedagogical engagement with AI bias does not cause harm to learners from marginalized linguistic or cultural backgrounds?
9. How does the manifestation of AI bias vary across language proficiency levels, task types, and regional or institutional contexts?
10. What institutional and national policy frameworks most effectively balance the promotion of critical AI literacy with equitable, inclusive AI integration in English language education?

Future research addressing these questions can help establish empirical grounding for the conceptual claims advanced in this article and support the development of evidence-based pedagogical and policy responses to AI bias in TESOL.

Conclusion

Bias in generative artificial intelligence presents one of the most significant challenges facing contemporary TESOL. Dominant discussions often frame bias as a technical flaw that requires mitigation or elimination. This article has argued that such a perspective remains incomplete because AI bias operates as a sociotechnical phenomenon shaped by language ideologies, cultural representations, institutional priorities, and unequal distributions of power. Although unexamined bias can reinforce linguistic hierarchies, marginalize diverse Englishes, and constrain learner agency, critical pedagogical engagement can transform biased outputs into valuable sites of inquiry. The Pedagogical Transformation Framework proposed in this article explains how learners can move from passive acceptance of AI-generated discourse toward critical AI literacy through processes of recognition, interrogation, and transformative engagement. This perspective does not seek to legitimize bias or diminish its potential harms. Rather, it positions bias as a pedagogical artifact that can reveal the assumptions embedded within AI-mediated communication and stimulate reflection on language, culture, identity, and power. As GenAI continues to reshape language education, the central challenge for TESOL does not concern whether AI should be adopted or rejected. The challenge concerns how educators, researchers, and policymakers can cultivate critical, ethical, and context-sensitive approaches that enable learners to engage with AI thoughtfully and responsibly. This orientation offers a pathway toward a more equitable vision of language education in which learners become not only users of AI technologies but also critical interpreters of the sociotechnical systems that increasingly shape communication in the twenty-first century.

ORCID

https://orcid.org/0000-0001-8517-4058

https://orcid.org/0000-0003-3661-1690

Publisher's Note

The claims, arguments, and counter-arguments made in this article are exclusively those of the contributing authors. Hence, they do not necessarily represent the viewpoints of the authors' affiliated institutions, or EUROKD as the publisher, the editors and the reviewers of the article.

Acknowledgements

Not applicable

Funding
This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

CRediT Authorship Contribution Statement
Jessie S. Barrot: Conceptualization, Writing - Original Draft
Hassan Mohebbi: Writing - Review & Editing

Generative AI Use Disclosure Statement
The authors used ChatGPT to improve readability, figure, and language accuracy. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Ethics Declarations
World Medical Association (WMA) Declaration of Helsinki–Ethical Principles for Medical Research Involving Human Participants
Not applicable to this paper which is conceptual in nature.

Competing Interests
The authors have no competing interest.

Data Availability
No data is avilable as this is a conceptual paper.

References

Afreen, J., Mohaghegh, M., & Doborjeh, M. (2025). Systematic literature review on bias mitigation in generative AI. *AI and Ethics*, *5*(5), 4789–4841. https://doi.org/10.1007/s43681-025-00721-9

Al-Khresheh, M. H. (2024). The future of artificial intelligence in English language teaching: Pros and cons of ChatGPT implementation through a systematic review. *Language Teaching Research Quarterly*, *43*, 54–80. https://doi.org/10.32038/ltrq.2024.43.04

Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, *57*, 100745. https://doi.org/10.1016/j.asw.2023.100745

Barrot, J. S. (2024). Leveraging ChatGPT in the writing classrooms: Theoretical and practical insights. *Language Teaching Research Quarterly*, *43*, 43–53. https://doi.org/10.32038/ltrq.2024.43.03

Barrot, J. S. (2026). Generative artificial intelligence for automated essay scoring: Exploring teacher agency through an ecological perspective. *Assessing Writing*, *67*, 100990. https://doi.org/10.1016/j.asw.2025.100990

Beck, J., Eckman, S., Kern, C., & Kreuter, F. (2026). Bias in the loop: How humans evaluate AI-generated suggestions. *Harvard Data Science Review*, *8*(2). https://doi.org/10.1162/99608f92.0e98898d

Bittle, K., & El-Gayar, O. (2025). Generative AI and academic integrity in higher education: A systematic review and research agenda. *Information*, *16*(4), 296. https://doi.org/10.3390/info16040296

Boy, G. A. (Ed.). (2026). *Handbook of sociotechnical systems: A human systems integration approach*. CRC Press.

Canagarajah, A. (1999). *Resisting linguistic imperialism in English teaching*. Oxford University Press.

Dixon-Román, E., Nichols, T. P., & Nyame-Mensah, A. (2020). The racializing forces of/in AI educational technologies. *Learning, Media and Technology*, *45*(3), 236–250. https://doi.org/10.1080/17439884.2020.

Dong, L. (2026). When machines teach writers: Reflecting on GenAI's role in giving corrective feedback on linguistic CAF facets. *Feedback Research in Second Language*, *4*, 28–39. https://doi.org/10.32038/frsl.2026.04.04

Fairclough, N. (1992). *Discourse and social change*. Polity Press.

Galloway, N., & Rose, H. (2015). *Introducing global Englishes*. Routledge. https://doi.org/10.4324/9781315734347

Jenks, C. (2025). Communicating the cultural other: Trust and bias in generative AI and large language models. *Applied Linguistics Review*, *16*(2), 787–795. https://doi.org/10.1515/applirev-2024-0196

Kohnke, L., & Zou, D. (2025). Artificial intelligence integration in TESOL teacher education: Promoting a critical lens guided by TPACK and SAMR. *TESOL Quarterly*, *59*, S267-S278. https://doi.org/10.1002/tesq.3396

Kostka, I., & McMartin-Miller, C. (2026). Reflections on using GenAI for multilingual writing feedback: Implications for practice. *Feedback Research in Second Language*, *4*, 16–27. https://doi.org/10.32038/frsl.2026.04.03

Law, L. (2024). Application of generative artificial intelligence (GenAI) in language teaching and learning: A scoping literature review. *Computers and Education Open*, *6*, 100174. https://doi.org/10.1016/j.caeo.2024.100174

Lee, S., Jeon, J., McKinley, J., & Rose, H. (2025). Generative AI and English language teaching: A global Englishes perspective. *Annual Review of Applied Linguistics*, *45*, 85–108. https://doi.org/10.1017/s0267190525100184

Lu, H. (2025). Cultivating critical thinking in foreign language education: Pedagogical strategies for leveraging generative AI biases. In Q. Ding, J. Luo, H. Li, & Y. Wang (Eds.), *Proceedings of the 9th International Seminar on Education, Management, and Social Sciences* (pp. 26–32). Atlantis Press. https://doi.org/10.2991/978-2-38476-462-4_4

Pack, A., & Maloney, J. (2023). Using generative artificial intelligence for language education research: Insights from using OpenAI's ChatGPT. *TESOL Quarterly*, *57*(4), 1571–1582. https://doi.org/10.1002/tesq.3253

Pitychoutis, K. M., & Spathopoulou, F. (2026). AI-mediated written corrective feedback in L2 writing: What changes, what doesn't, and what we still don't know. *Feedback Research in Second Language*, *4*, 6–15. https://doi.org/10.32038/frsl.2026.04.02

Pennycook, A. (2004). Critical applied linguistics. In A. Davies & C. Elder (Eds.), *The Handbook of Applied Linguistics* (pp. 784–807). Wiley. https://doi.org/10.1002/9780470757000.ch32

Rivero, E., & Yin, P. (2025). Navigating the contradictions of AI: Critiquing AI's standardized English and developing creative bilingual possibilities. *Frontiers in Education*, *10*, Article 1616458. https://doi.org/10.3389/feduc.2025.1616458

Romeo, G., & Conti, D. (2026). Exploring automation bias in human-AI collaboration: A review and implications for explainable AI. *AI & Society*, *41*(1), 259–278. https://doi.org/10.1007/s00146-025-02422-7

Seo, J. Y. (2024). Exploring the educational potential of ChatGPT: AI-assisted narrative writing for EFL college students. *Language Teaching Research Quarterly*, *43*, 1–21. https://doi.org/10.32038/ltrq.2024.43.01

Todd, R. W. (2025). Generative AI as a disrupter of language education. *International Journal of TESOL Studies*, *250127*, 1–9. https://doi.org/10.58304/ijts.250127

Toncelli, R., & Kostka, I. (2024). A love-hate relationship: Exploring faculty attitudes towards GenAI and its integration into teaching. *International Journal of TESOL Studies*, *6*(3), 77–94. https://doi.org/10.58304/ijts.20240306

Wang, J., Meng, W., & Zhang, H. (2025). EFL teaching in the age of generative AI: Challenges, opportunities, and teacher agency. *Testing, Psychometrics, Methodology in Applied Psychology*, 32(3), 862–866. https://tpmap.org/submission/index.php/tpm/article/view/2275

Yeo, M. (2023). Academic integrity in the age of Artificial Intelligence (AI) authoring apps. *TESOL Journal, 14*(3), e716. https://doi.org/10.1002/tesj.716

Zhang, Y., Tan, L., & O'Halloran, K. L. (2026). Bias in GenAI: Exploring cultural variations between large language models. *Linguistics and Education*, *92*, 101515. https://doi.org/10.1016/j.linged.2026.101515