ESJ Humanities Peer-reviewed

Rethinking Language Education After Crisis: AI-Driven Learning Analytics as a Site of Social Science Inquiry

Nurdan Kavakli Ulutas, PhD
Izmir Demokrasi University, Turkiye
Vasil Kikutadze, PhD
Grigol Robakidze University, Georgia

Doi:10.19044/esj.2026.v22n20p98

Submitted: 10 April 2026
Accepted: 02 July 2026
Published: 31 July 2026

Cite As:
Ulutas, N.K., & Kikutadze, V. (2026). *Rethinking Language Education After Crisis: AI-Driven Learning Analytics as a Site of Social Science Inquiry*. European Scientific Journal, ESJ, 22 (20), 98. https://doi.org/10.19044/esj.2026.v22n20p98

Abstract

Promising to give personalised feedback and monitor learner behaviour, artificial intelligence-driven learning analytics (AI-LA) gained prominence in language education during the COVID-19 pandemic. However, its adoption has grown faster than the social science research needed to make sense. Conceptually, a structured integrative synthesis of recent peer-reviewed literature is conducted following Jaakkola's (2020) theory-synthesis approach, prioritizing work published since 2020, and the regulatory shift marked by the European Union's 2024 Artificial Intelligence Act. Bringing together critical data studies (Boyd & Crawford, 2012; Williamson, Komljenovic & Gulson, 2024), recent systematic reviews of generative AI and personalisation in language learning (Jeon, 2025; Teng, 2025), scholarship on algorithmic bias and linguistic legitimacy (Baker & Hawn, 2022; Koenecke et al., 2020), and post-crisis education research (Charitonos et al., 2025; Menashy & Zakharia, 2022), it argues that AI-LA should be understood as a sociomaterial arrangement rather than a technical pipeline. Its contribution is a conceptual framework, with testable propositions, showing where the pedagogical promise of personalisation breaks down epistemically, why bias in language-recognition systems is political as much as technical and bound up with whose language counts as legitimate, and how the displaced learner exposes the limits of current AI-LA. Implications for researchers, language teachers, curriculum developers, learners, and policy makers are listed thereafter.

ISSN: 1857-7881 (Print)
e-ISSN 1857-7431 (Online)

Keywords: AI-driven learning analytics, language education, algorithmic bias, sociomaterial perspectives, post-crisis pedagogy, conceptual framework

Introduction

Generally speaking, the pandemic did not introduce AI-LA into language education but rather removed the last barriers to an adoption that was already in progress. Within weeks of the initial lockdowns, language teachers were obliged to adapt not only to novel modes of communication but also to new systems of oversight, among them dashboards reporting who was engaged, automated feedback evaluating learner writing almost instantly, and voice-recognition applications assessing pronunciation against models built from native speakers the learners had never encountered. Notably, within these settings, the teachers themselves were frequently the last to be consulted.

Today, those infrastructures still exist and have even increased in number. In this respect, the European Union's Artificial Intelligence Act (the EU AI Act) has been in force since August 2024, categorizing a wide range of educational AI as high-risk and imposing obligations on its creators and users (European Parliament and Council, 2024). However, the field still lacks a definitive understanding of the impact these systems have on language education and on whom. To highlight this, the relevant literature mainly presents two major perspectives. The first is celebratory, in which AI-LA serves as a means for the personalisation of language education, customising content and feedback that no teacher could match, but has heightened enthusiasm with the recent developments in AI (Lee & Yeung, 2024; Teng, 2025). On the other hand, the second is technical-reparative, where the celebratory approach views bias and privacy as leftovers to be fixed. Thus, this perspective regards them as engineering problems to be addressed with improvements in data, more equitable algorithms, and a more robust governance (Baker & Hawn, 2022; Barocas et al., 2019). Both perspectives have their value, and we do not aim to disregard either; however, the primary focus lies on another question that both perspectives often neglect. *What type of object is AI-LA in language education when it is functioning effectively? Whose teaching authority does it replace, and for what purpose? What does it not acknowledge, and what is the price of that oversight?*

Taking these questions as the starting point, we will argue, as Selwyn (2019, 2022) has argued, whether it is more accurate to view educational AI not as a pure technical tool, albeit as a sociomaterial arrangement engaged in political labour, since it reallocates authority, legitimises accounts of meaningful and natural learning, and limits alternatives subtly (see also Perrotta et al., 2022). This argument has significantly impacted educational research in a broader context, yet it has been less advanced specifically in

ISSN: 1857-7881 (Print)
e-ISSN 1857-7431 (Online)

language education, where these inquiries have sharpened since language serves as a sphere of social distinction. More typically, whose accent will be modelled, whose syntax will be classified as erroneous, whose pragmatics will be read, are not atypical instances; instead, they constitute the essence of what AI-LA performs when it works with language. What is more, they are also well-aligned with the divisions of class, race, and any other regional differences that standardisation efforts in language education have traditionally sought to control (Pennycook, 2001).

Working through that line of thinking and developing it, we aim to proceed with a structured synthesis of recent literature in learning analytics, applied linguistics, and social sciences, and read through theoretical resources from critical data studies in the related fields. Although there is a myriad of studies conducted as proof-of-concept, short-term, and self-reported evaluations (Jeon, 2025; Teng, 2025) in terms of technical demonstrations of AI-LA, there is a lack of a framework that can bring the technical, pedagogical, and political dimensions together to see how they interact with one another. To build this framework, we aim to seek answers to the below-listed research questions:

1. What does AI-LA achieve pedagogically in language learning, and what are the epistemic limits?
2. In what ways does the algorithmic mediation of language assessment reduce the inequities of linguistic legitimacy, and how does the recent regulatory shift alter that scenario?
3. What can a sociomaterial perspective on AI-LA offer that the prevalent technical-reparative view overlooks, especially in post-crisis situations where learners are underrepresented?

Literature review

Without first revisiting the story the field tells about itself, it becomes difficult to discuss what technology means in language education. In this account, LA emerged in the early 2010s to fulfil the promise that the digital footprints learners create can be transformed into practical assessments of progress and, ultimately, into personalised instruction (Baker & Inventado, 2014; Siemens, 2013). Because language learning unfolds through interaction, it produces traces rich in both form and meaning, and these lend themselves readily to the creation of a record (Romero & Ventura, 2013). Predictive modelling has been followed by adaptive content sequencing and automated feedback, and, together with the emergence of generative AI, the current literature presents AI-LA as deliverable at scale. There is, however, a disparity between the capabilities of these systems and the evidence for their actual performance, and that disparity is the natural place to begin.

ISSN: 1857-7881 (Print)
e-ISSN 1857-7431 (Online)

To start with, Teng (2025), in a systematic review of forty-nine studies on generative AI in the language classroom published between 2023 and 2024, finds that these tools are now utilised across speaking, writing, and feedback activities; however, the evidence base is predominantly reliant on short-term, self-reported data mostly from higher education and English, with minimal focus on other languages, younger learners, or low-resource environments. Additionally, Jeon's (2025) complementary review of recent reviews arrives at a unified conclusion that generative AI consistently aids in surface-level precision and the rapidity of feedback, but learners and teachers still prefer human feedback for more profound and contextual revisions that the models cannot manage. Similarly, Godwin-Jones (2022) reports a significant difference in how AI writing-feedback tools perform across proficiency levels, whereas Huang et al. (2023) showcase varying effects of adaptive systems across learners' motivational profiles, pinpointing that personalisation reinforces inequalities rather than helping. A recent systematic review by Wah (2025) on AI, personalization, and learner engagement in language learning reaches a similar conclusion, documenting genuine gains in vocabulary acquisition and engagement. In the meantime, it is also concluded that the field still lacks longitudinal, interdisciplinary, and culturally adaptive frameworks for assessing the sustainable impact on language acquisition. When considered cumulatively, these studies suggest that personalisation is more of an unresolved issue than an area where technical progress has surpassed the development of teaching methods and supporting evidence.

What is striking is the variability, where a system underperforms for some learners, concluding that the algorithm is not good enough yet. However, the assumption is that learner behaviour, noted as data, is a dependable proxy for learning, which has long been pursued quietly in the field of applied linguistics. The sociocultural perspective, since Vygotsky (1978), asserts that higher mental functions are social in nature before they are individual, and the ethnography of communication as a competence is inherently tied to context, albeit not a standalone trait (Hymes, 1972). Accordingly, research on situated learning has purported that the same performance can convey different meanings based on the practices in which it takes place (Gee, 2004). Because learning is context-specific, dependent, and interactionally generated via collaborative efforts. The signs are real, but what is questionable is whether they are adequate to explain learning.

This is where it becomes necessary to stop treating AI-LA as a technical instrument and to start treating it as something else. A growing body of work, drawing on critical data studies, insists that data is never simply found but always produced, framed at every stage by what is selected for measurement and by the assumptions about what is worth knowing that the selection encodes (Boyd & Crawford, 2012; Kitchin, 2014). Brought into

ISSN: 1857-7881 (Print)
e-ISSN 1857-7431 (Online)

education, this orientation has produced an account of datafication as a transformation of the field itself, which is of its governance, its institutional cultures, and its political economy, as the apparatus of monitoring settles into the everyday and decisions come to count as legitimate only when they are also data-mediated (Williamson, Komljenovic, & Gulson, 2024). Selwyn (2019, 2022) has pressed the argument furthest, and most usefully for present purposes, by refusing the comforting view that the displacement of pedagogical judgement by algorithmic recommendation is an incidental by-product of automation to be corrected with better design. It is, he argues, closer to the operating principle of the technology, and the rhetoric of technical solutionism that surrounds educational AI works precisely to keep that principle from view. Perrotta, Selwyn, and Ewin (2022) carry the point into the generative era, showing how a language model performs what they call the affective labor of understanding as an imitation of interpretation that does none of the relational work interpretation requires.

Seen this way, the problems the technical literature treats as defects to be debugged look rather different. Consider bias, which the corrective framing approaches as an engineering fault as a matter of unrepresentative training data, to be remedied by better sampling and fairer algorithms. The empirical record of the harm is not in dispute. Koenecke et al. (2020) documented word-error rates in commercial speech recognition running roughly twice as high for African American speakers as for white speakers; Baker and Hawn (2022) review consistent underperformance of educational AI for learners from racial, ethnic, and linguistic minorities; and Gallegos et al. (2024), surveying bias in large language models, show how the very determination of which varieties count as standard reinforces existing social hierarchies. What the engineering framing cannot easily accommodate is that these are not isolated malfunctions, but the predictable outputs of systems built on data that records a long history of whose language has been listened to and whose has been counted as correct. The standardising work that formal language education has always done, devaluing non-dominant varieties in the name of a norm (Pennycook, 2001), is now being performed at scale and at speed by training corpora and the objectives that optimise against them.

The focus here is not on accuracy, but rather on who is the authority. When a system where AI and LA take over tasks, it does not just substitute the teacher.

It embodies a notion of language competence, in which the criteria for evidence of learning and the perspectives that are valued are established. Crawford (2021) states that this system also helps form a perspective of the world. And this perspective normalises certain experiences while obscuring others. This notion can be applied to language education. In this domain, an analytics framework not only assesses a person’s language proficiency but

ISSN: 1857-7881 (Print)
e-ISSN 1857-7431 (Online)

also assists in clarifying the definition of competence. Basically, it decides whose language it can understand and whose it will not. The concept of "datafication" (Selwyn, 2019) is useful in this context. It involves turning processes into simple variables that a computer can work with. This process leaves out the details that come from studying language classrooms in depth (Gee, 2004; Hymes, 1972). The loss of these details is not a technical problem, yet it is a part of what the system does.

If these concerns are general, they sharpen to a point in the settings this article takes as its analytical limit. Crises do not merely interrupt education; rather, they expose and frequently deepen the inequalities already running through it. Menashy and Zakharia (2022), in a vertical case study of Syrian refugee education in Lebanon, show how the pandemic operated as a crisis layered upon an earlier humanitarian one, and how the asymmetric power relations structuring education governance in such contexts were reproduced rather than disrupted by the emergency response. The rapid scaling of digital provision during the pandemic, surveyed by Reich et al. (2020), reached the most marginalised learners least well, and Dryden-Peterson (2017) supplies the longer frame, theorising refugee education as a domain shaped by displacement and interrupted schooling, conditions that produce precisely the linguistic repertoires least likely to be represented in any standard training corpus. Generative AI has entered this picture as both promise and threat. Creely and Henderson (2025) describe a digital dis/empowerment for migrant and refugee learners, in which genuine openings, on-demand translation, conversational practice, accessible explanation of unfamiliar bureaucratic language, arrive bound to a documented Western bias that risks cultural homogenisation; and Charitonos et al. (2025), writing on English-language teaching in refugee settings, show teachers left to adapt, through their own professional judgement, tools never designed for the learners in front of them. The displaced learner is not a marginal case for AI-LA to accommodate in due course. She is the case against which any equity claim must be tested, the learner least likely to be well represented in the data and most dependent on the language the system purports to teach.

It is tempting to suppose that regulation has begun to answer these concerns, and in one respect it has. The European Union's Artificial Intelligence Act (European Parliament and Council, 2024), in force from August 2024, classifies AI systems that evaluate learning outcomes or determine access to education as high-risk, attaching binding obligations on data governance, bias auditing, technical documentation, and human oversight, and so converting what had been the voluntary transparency principles of bodies such as UNESCO and the OECD into enforceable requirements falling on identifiable parties. For example, providers, who are the creators that develop and launch an AI-LA system into the market, need to

ISSN: 1857-7881 (Print)
e-ISSN 1857-7431 (Online)

implement a risk-management framework and document the validation data. On the other hand, deployers, which are the institutions that put the system to use, need to guarantee rigorous human supervision and track operational performance that learners may be affected by. This distinction is vital for language education since a university that authorises an automated writing-evaluation tool acts as a deployer. However, it must also focus on the responsibilities of monitoring and transparency, even when it lacks access to the provider's training data. Yet the gap between adhering to these regulations and achieving fairness is precisely where the current argument lies. A system might fulfil all procedural demands yet still put its users at a disadvantage as it continues to be structured around the language of its creators. Conformity assessment typically asks if appropriate measures were followed, rather than if the system can understand the language of individuals whose dialects deviate from the dominant norm. Regulation can require that bias auditing be performed; it cannot, by itself, ensure that the audit asks the right question. That question, of whose language counts and at what cost, is not a technical one, and it will not be answered by better engineering or by tighter documentation alone. It is a question of the kind of sociomaterial and sociolinguistic analysis that the remainder of this article sets out to develop.

Theoretical and conceptual framework

The framework that follows draws on two intersecting bodies of work, in continuous conversation with a third. Critical data studies, as developed by Boyd and Crawford (2012) and Kitchin (2014) and extended in educational settings by Williamson and colleagues (Williamson, Komljenovic & Gulson, 2024), supply the orientation. Data is not a record but a production, shaped at every stage by what institutions choose to measure, by the historical patterns of inclusion and exclusion their measurement instruments encode, and by the assumptions about what counts as worth knowing that the choice of variables makes operative. Bringing that orientation to bear on AI-LA in language education means treating the systems not as instruments that report on learning, but as participants in the production of what learning is taken to be.

Language learning is not something that happens in our heads. It is something that we do with people. The people we talk to, the groups we belong to, and the words we use all play a role in how we learn to communicate. The way we learn to talk and write is shaped by the people around us. Our language is influenced by our friends, family, and community. It is also influenced by the schools we go to and the jobs we have. The kind of language we use can affect how people see and treat us. At that point, the ethnography of communication (Hymes, 1972), critical applied linguistics (Pennycook, 2001), the sociocultural tradition (Vygotsky, 1978), and the situated-learning scholarship that follows Gee (2004) have long been in pursuit of scrutinising

ISSN: 1857-7881 (Print)
e-ISSN 1857-7431 (Online)

how people learn to communicate. They have found that language learning involves how to talk and write in a way that makes sense to the people around us. But when we use computers to assess language, we must ask some questions. Whose language are we using as a standard? Who decides what so-called good and bad language is? Who benefits when the computer gets it right, and who gets in trouble when it gets it wrong? Language learning systems, like AI-LA, cannot answer these questions on their own. We need to think about how we use these systems and how they affect the people who use them.

In this highlight, post-crisis education scholarship provides the temporal frame within which both bodies of work matter acutely. A crisis can cause problems for the way things are normally done, and this can be an opportunity to make changes. It can also lead to the use of technology that is not thought through; however, this can make things unfair for some people even after the crisis is over (Williamson et al., 2020). Specifically, language education is affected over time. This is due to the fact that people's ability to learn and use language is directly impacted by what happens after a crisis. Language education is crucial since it determines whether individuals and communities of practice can deal with the aftermath of a crisis. Recent work on AI and displaced learners makes this point with unusual clarity. Generative AI tools offer real openings for refugee and migrant learners while simultaneously embedding cultural biases and linguistic norms that risk further marginalising the very populations they are claimed to serve (Charitonos et al., 2025; Creely & Henderson, 2025). The displaced learner, in this argument, is not a special case. They are the analytical limit against which any claim to equitable AI-LA needs to be tested.

The aforementioned bodies of work discipline one another in ways that make the integration productive. Critical data studies prevent sociolinguistic analysis from sliding into pure description by insisting on the material and institutional stakes of how data is produced. Sociolinguistic and critical applied-linguistic work prevents critical data studies from staying at a level of abstraction that loses contact with what people do with language. Post-crisis education scholarship reminds both bodies of work that the conditions in which AI-LA gets deployed are never neutral conditions. They, in fact, carry histories of disruption, of unequal recovery, and of institutional response that determines who bears the cost of getting deployment wrong. The argument the rest of this paper develops is, accordingly, that the work of building AI-LA worth wanting, in language education, is not work that can be left to system designers and platform vendors; it is, rather, work that requires the sustained involvement of researchers and practitioners trained to take the social production of language and learning seriously.

ISSN: 1857-7881 (Print)
e-ISSN 1857-7431 (Online)

Method

This article is a conceptual study, and more precisely a theory-synthesis article in the sense developed by Jaakkola (2020). It aims to integrate concepts that have developed within separate literatures into a single framework capable of addressing a problem that none of them resolves alone. Thus, it proceeds by the purposive selection and critical re-reading of scholarship rather than by empirical testing, and its claims are, in essence, to be judged on conceptual coherence, on explanatory reach, and on the capacity of the resulting framework to generate questions that subsequent empirical research can pursue. These criteria, rather than statistical validity or sample representativeness, constitute the standard against which the argument is asked to be assessed.

The literature that underpins the synthesis was identified through structured searching in Scopus, Web of Science, and the Education Resources Information Centre (ERIC), supplemented by targeted Google Scholar searches to capture very recent work and the grey literature on regulation. Searches were conducted between September 2025 and January 2026 and were restricted to English-language sources. Search strings combined the technological object ("learning analytics," "generative AI," "automated writing evaluation," "intelligent tutoring," "ASR") with the educational domain ("language education," "language learning," "EFL," "second language") and with the critical-contextual dimensions of interest ("algorithmic bias," "datafication," "equity," "refugee education," "displacement"). Within each database, these were joined with the Boolean operators AND and OR, so that a representative string took the form ("learning analytics" OR "generative AI") AND ("language education" OR "EFL") AND ("algorithmic bias" OR "datafication"). Peer-reviewed work published between 2020 and 2025 was prioritized, with foundational older sources retained where they remain the standard reference for a concept (Hymes, 1972, on communicative competence; Vygotsky, 1978, on the social formation of mind; Pennycook, 2001, on critical applied linguistics).

Since the aim was conceptual synthesis rather than a comprehensive systematic review, the screening process was structured to produce an analytically adequate rather than an exhaustive corpus. The initial searches yielded around 480 records. Following the elimination of duplicates and items whose titles or abstracts did not align with the intersection of educational technology, language education, and the critical-contextual themes, approximately 110 sources were fully reviewed. Inclusion at this stage relied on three criteria. The first is to possess peer-reviewed or authoritative recognition, including major scholarly monographs and the official text of the European Union regulation. The second one is directly relevant to at least one of the study's three guiding questions related to the pedagogical, ethical, and

ISSN: 1857-7881 (Print)
e-ISSN 1857-7431 (Online)

governance dimensions of AI-LA. The third one is the capacity to enhance the conceptual argument rather than simply supporting it. Sources were eliminated if they only mentioned AI in education briefly, if they repeated a point made more authoritatively elsewhere, or if they could not be verified. As a result, the collection of studies was summarised and reported thoroughly as the references. The synthesis was thematic in nature. Each source was analysed according to the three guiding questions and the framework elements outlined in Table 1. Lastly, the classifications were compared with each other to highlight the areas of shared support and conflict that formed the basis of the framework.

Sources were selected for their capacity to advance the conceptual argument rather than for comprehensive coverage of any one subfield. This is a deliberate feature of theory-synthesis work, and a limitation that the conclusion returns to. Post-crisis settings are treated throughout not as a passing illustration but as a structured analytical case, partly because such settings sharpen every problem the framework identifies, and partly because the populations most likely to be misrepresented in training data are very often the displaced and crisis-affected learners whose access to language is most consequential for their lives (Creely & Henderson, 2025; Menashy & Zakharia, 2022). For the analytical procedure, passages from the literature were read against one another, points of mutual reinforcement and mutual disturbance were identified, and the resulting framework was tested for internal consistency by working it back through the research questions.

Table 1: The sociomaterial framework for AI-LA in language education

Framework component	Guiding question it raises for AI-LA	Grounding literature
Datafication	What is rendered visible, and what is lost, when learning is reduced to the traces an analytics system can capture?	Boyd & Crawford (2012); Kitchin (2014); Selwyn (2019); Williamson, Komljenovic & Gulson (2024);
Pedagogical personalisation	What can adaptive personalisation accomplish, and where do its epistemic limits lie?	Godwin-Jones (2022); Huang et al. (2023); Jeon (2025); Teng (2025); Wah (2025)
Algorithmic bias	How do the training distributions of language-recognition systems disadvantage speakers of non-dominant varieties?	Baker & Hawn (2022); Gallegos et al. (2024); Koenecke et al. (2020)
Linguistic legitimacy	Whose language is treated as the norm, and how does standard-language ideology operate through the system?	Hymes (1972); Gee (2004); Pennycook (2001)
Governance	What do high-risk regulation and institutional data governance require, and where does compliance fall short of equity?	European Parliament and Council (2024); Crawford (2021); Prinsloo & Slade (2017)

ISSN: 1857-7881 (Print)
e-ISSN 1857-7431 (Online)

Framework component	Guiding question it raises for AI-LA	Grounding literature
Post-crisis displacement	How does the displaced learner function as the limit case against which any claim to equitable AI-LA is tested?	Charitonos et al. (2025); Creely & Henderson (2025); Dryden-Peterson (2017); Menashy & Zakharia (2022)

Note. The six components are analytically distinct but mutually conditioning; the conceptual analysis that follows works through them in turn, and the testable propositions in the conclusion are keyed to them.

Conceptual analysis

What personalisation actually does

Consider what an adaptive system is doing, in practice, when it personalises a learner's sequence. A model receives a stream of inputs and uses those inputs to update its estimate of the learner's current state. From that estimate, it selects the next task. The pedagogical work, on this account, is the calibration, which is the matching of difficulty to estimated ability. And it is precisely this account, in its smooth simplicity, that decades of research on language learning as a social practice have been complicating.

Take a learner who responds slowly to a vocabulary prompt. The model has limited interpretive options. It can treat the latency as evidence of difficulty and serve up easier items, or it can treat the same latency as evidence of deliberation and continue. What the model cannot do is what a sensitive teacher does as a matter of course by noticing that this particular learner is processing the word in a second language whose orthography differs from her first, or that the prompt arrived at a moment when her infant was crying in the next room, or that the cultural reference embedded in the example sentence is one she does not share. Ethnographic and sociocultural research on language classrooms has made precisely this kind of situated interpretive work its central object (Gee, 2004; Vygotsky, 1978). What that research shows, repeatedly, is that learning is not a property the learner has but a process produced in social activity, and that the work of producing it depends on resources, shared history, mutual knowledge, and the reading of context, which lie outside anything an interaction trace can pursue.

The implication is not that AI-LA is useless for personalisation. The implication is that the personalisation it performs is operating on a thin slice of what learning actually involves. The slice can be useful (it is, for instance, plausibly useful for vocabulary drilling), and it can also be quietly misleading when systems trained to optimise for the thin slice are presented as optimising for learning. Recent reviews of generative AI in language education return repeatedly to this concern. Students and teachers value the speed and breadth of automated feedback, but they reserve for human feedback the deeper, contextual revision work that, on closer inspection, turns out to be most of

ISSN: 1857-7881 (Print)
e-ISSN 1857-7431 (Online)

what writing pedagogy is for (Jeon, 2025; Teng, 2025). The challenge for AI-LA design is not to do without the thin slice. It is to stop pretending that the thin slice is the whole.

What helps here is a different kind of integration than the one usually proposed. The standard move is to layer qualitative data onto quantitative outputs after the fact. Focus groups interpret what the dashboards show, and ethnographic notes contextualize what the models predict. Although it is useful, it leaves the underlying instrument as it is. A more rigorous integration would entail creating analytics systems where their variables, thresholds, and intervention strategies are guided from the outset by descriptions of learning as a contextual social activity. This would involve the utilisation of the descriptive findings from ethnographic and sociolinguistic research on language classrooms, not to validate but to establish design limitations, such as recognising the signal, deliberately ignoring the signal, and specifying where human judgment is needed instead of being enhanced (Gee, 2004; Hymes, 1972; Pennycook, 2001). Such integration is less common than the term "mixed methods" implies.

Bias, voice, and the question of whose language counts

While personalisation has been the primary educational promise of AI-LA, automated evaluation has been its most pronounced ethical challenge. The systems that evaluate writing, that assess pronunciation, and that identify potentially plagiarised content do not generate unbiased results. They generate results influenced by the data they were trained on, and that data, nearly always, is skewed towards the linguistic variations of individuals from specific social backgrounds. The consequences are concrete in the language classroom. An automated writing evaluation (AWE) tool that flags transfer features from a learner's first language, an article omission characteristic of a Slavic-language background, say, as grammatical error rather than developmental interlanguage, penalises the learner for precisely the marker a human teacher would read diagnostically. An automatic speech recognition (ASR) system underpinning a pronunciation-feedback application returns higher error rates for a learner whose accent departs from the inner-circle norms on which it was trained, and the learner is told to repeat an utterance that was, communicatively, entirely adequate. A teacher-facing dashboard, aggregating these judgments, then reports the learner as struggling. Koenecke et al. (2020) recorded this dynamic for commercial automatic speech recognition, showing that word error rates for African American speakers were approximately double those for white speakers across five key systems. The trend reappears in broader natural language processing applications (Barocas et al., 2019), and Baker and Hawn (2022) provide consistent evidence that learners from racial,

ISSN: 1857-7881 (Print)
e-ISSN 1857-7431 (Online)

ethnic, and linguistic minorities underperform across a range of educational AI applications.

What an applied linguist reads in these findings differs from what the engineering literature tends to see. The performance discrepancies are not merely isolated technical malfunctions; they are a contemporary manifestation of a much longer history. Formal language education has been intertwined with the enforcement of standard forms and the devaluation of all others (Pennycook, 2001). The work of standardisation was once carried out by teachers and examination boards, but it is now being accomplished by training corpora at scale and speed. The mechanism is new, but the selectivity is old. Speakers of non-dominant English varieties, multilingual learners whose writing exhibits traits from a first language, and refugees whose spoken language reflects disrupted education and displacement (Charitonos et al., 2025; Creely & Henderson, 2025) are not merely statistical anomalies for AI-LA to eventually adapt to. They represent the groups for which the system's implicit standards are established.

Crawford (2021) regards this as an element of a broader process she refers to as atlas-making. This involved the selective assembly of worldwide visuals that legitimise specific experiences while obscuring others. Herein, the language learning resources are helpful to acquire the language, and an AL language system not only assesses someone's proficiency in a language but also determines what it signifies to be proficient in that language. It also chooses which language to understand and which language not to. This implies that certain individuals will be deliberately misunderstood by the system. To confirm, studies reveal more insights into why the system makes certain choices (e.g., Hymes, 1972; Pennycook, 2001). Thus, taking early action is crucial during the design phase of the system, rather than waiting until it is in use. If we happen to postpone, modifying the language learning material will become more challenging. Additionally, the AI system must be evaluated to ensure it is equitable for all users. The system, fundamentally, must assist individuals in learning, albeit no abandon them.

Certain recent developments make this position more urgent than ever. The initial aspect is the shift in the regulatory landscape. The European Union's Artificial Intelligence Act categorises various educational AI systems as high-risk, including those that determine access to education, assess learning outcomes, or monitor learners during assessment (European Parliament and Council, 2024). High-risk categorisation entails significant responsibilities, including technical documentation, human supervision, and the mitigation of bias in training and validation datasets. The effects on language education are seen as immediate. The second development is the rapid integration of generative AI into language learning and teaching (Lee & Yeung, 2024; Teng, 2025), which raises the question of which forms of

ISSN: 1857-7881 (Print)
e-ISSN 1857-7431 (Online)

fluency the model recognises and which it disregards. A system may pass a conformity assessment but still encode sociolinguistic biases that disadvantage speakers of rarer varieties. The divide between formal compliance and genuine equity is precisely the area where sociolinguistic research can engage most effectively.

Post-crisis contexts and the displaced learner

Each technology brings its own inquiries in the aftermath of a crisis. The pandemic's rapid introduction of digital learning platforms revealed disparities in connectivity, device availability, and platform usability among different socioeconomic groups, with language learners in low-income countries being particularly impacted (Reich et al., 2020). Menashy and Zakharia (2022), based on a vertical case study of Syrian refugee education in Lebanon, highlight an aspect frequently overlooked in pandemic literature, that is, the true crisis was not merely the pandemic itself, but rather the pandemic added to an existing humanitarian crisis, with the unequal global power dynamics. This also shaped education governance in these contexts, being reinforced rather than altered by the emergency response. Refugee learners were navigating Arabic-to-host-language transitions under conditions of displacement long before COVID arrived, and the digital language tools designed for standard-dialect speakers worked poorly for populations whose linguistic repertoires reflect non-dominant varieties and interrupted schooling (Dryden-Peterson, 2017).

Generative AI has, since then, sharpened the dynamic rather than resolved it. Creely and Henderson (2025) describe a digital dis/empowerment for migrant and refugee learners through real openings (e.g., on-demand translation, conversational practice) accompanied by a documented Western bias in training data that risks cultural homogenisation and the erosion of linguistic diversity. Charitonos and colleagues (2025), writing on English language teaching in refugee settings, show how teachers in these contexts are left to negotiate technological tools never designed for displacement, exercising professional judgement to adapt systems that assume linguistic profiles their learners do not have. The displaced learner is, in effect, the limit case for the framework this paper has been developing since the person is least likely to be well represented in the training data, most likely to be misread by the model, and most dependent on the language the system is supposed to help them acquire. An AI-LA framework that cannot account for this learner is not, in any useful sense, an equitable one.

Governing AI-LA in such settings is a matter of arrangements at several institutional levels (Prinsloo & Slade, 2017). System designers need accountability frameworks that require equity-impact assessment before deployment, and that take linguistic diversity in training corpora as a

ISSN: 1857-7881 (Print)
e-ISSN 1857-7431 (Online)

substantive design requirement rather than a documentation tick-box. Educational institutions need data-governance policies that address the full lifecycle of learner data and that align it with learner interests, not platform interests (Crawford, 2021). Educators need access to interpretable analytics outputs and the professional agency to exercise judgment when an algorithmic recommendation looks inconsistent with what they know about their learners (Selwyn, 2019, 2022). Policy makers, post-2024, now operate in a legal environment in which high-risk obligations attach to a significant share of the AI-LA being deployed in education (European Parliament and Council, 2024), and the work of translating those obligations into substantively equitable practice is the work the field has not yet done.

Social science research has something to contribute to every one of these levels. It can produce the empirical evidence base that equity-impact assessment requires, it can illuminate the institutional conditions that enable or constrain responsible AI governance, it can document, through careful qualitative inquiry, the professional-practice implications of analytics integration in language classrooms (Selwyn, 2022), and it can inform policy through systematic comparative analysis of how different educational and national contexts have approached governance (Kitchin, 2014; Williamson, Komljenovic, & Gulson, 2024). The argument is that the work of social science should be positioned, institutionally, as a co-productive partner in it, with the standing to shape design rather than only the standing to comment on systems already deployed.

Implications for research, practice, and policy

The implications follow from the argument rather than from a separate list. For researchers, the most consequential one is methodological. Studying AI-LA in language education productively requires designs in which computational and pedagogical work is in sustained dialogue with social science perspectives, rather than applied to it in sequence. Mixed-methods designs that combine large-scale interaction data with close-grained classroom ethnography and discourse analysis, with longitudinal designs capable of tracking effects over time, and with the political-economy attention of critical data studies (Williamson, Komljenovic, & Gulson, 2024) can produce interpretive depth that any one of these methods produces only partially. The kind of single-snapshot, self-reported study that has dominated the recent generative AI literature (Teng, 2025) is unlikely to deliver the framework-testing the field now needs.

For language educators, course designers, and institutional technology coordinators, the implication is professional. Critical literacy with respect to AI-LA tools has become a part of the job rather than an optional specialisation. Educators who can read the normative assumptions encoded in an automated

ISSN: 1857-7881 (Print)
e-ISSN 1857-7431 (Online)

feedback system, who can recognise when an adaptive recommendation is operating on a thin slice of what their learners are doing, and who are positioned to advocate for learners whose linguistic profiles fall outside the model's training distribution, are simply better placed to use the tools well. Initial teacher education and in-service professional development have begun to address this, but unevenly, and the institutions that treat critical AI literacy as central to language teacher preparation remain the exception rather than the rule.

For policymakers and institutional leaders, the implications now run through a binding regulatory environment. Adoption of AI-LA at scale in language education should be preceded by equity-impact assessment, should involve consultation with the learner communities most likely to be affected by misclassification, should include mandatory bias-auditing protocols that interrogate the linguistic norms of the training data, and should be accompanied by data-governance frameworks that protect learner rights (Barocas et al., 2019; European Parliament and Council, 2024; Prinsloo & Slade, 2017). International organisations, such as UNESCO, UNHCR, and the World Bank, have a significant role in setting shared standards and supporting capacity-building in contexts where national institutional capacity is mostly limited.

Ultimately, the implication is fundamental for learners. In AI-LA contexts, learner agency is often treated in the design literature as a dependent variable. Henceforth, the risks are greatest for learners whose language profiles differ from norms and whose educational access has already been most disrupted. Within this context, meaningful agency is not sufficiently represented by the ability to opt out. It necessitates active involvement in system design, clear access to the rationale behind algorithmic decisions, and institutional processes through which learners can challenge evaluations they consider unfair (Prinsloo & Slade, 2017). The displaced learner represents the extreme case in this situation, as the framework unable to support her does not align with the framework the field requires.

Conclusions

Beyond question, AI-LA matters in language education. The abilities are genuine, and the institutional forces pushing for adoption will not lessen. Even more, the regulatory framework implemented in 2024 has only heightened the importance of successful deployment. Rejecting the technology would be just as analytically unproductive as embracing it without critique; thus, this article aims to position itself between these two choices.

Accordingly, the central claim is that AI-LA in language education must be understood as a sociomaterial arrangement rather than a technical conduit. What this reframing adds to existing critical data studies and

ISSN: 1857-7881 (Print)
e-ISSN 1857-7431 (Online)

educational-AI scholarship is its specificity to language. In that, it shows that the epistemic limits of personalisation, the political character of algorithmic bias, and the workings of standard-language ideology are not three separate concerns, but a single problem reflected through the domain in which linguistic difference carries the most social weight. The displaced learner is where that problem becomes visible in its sharpest form, and the framework summarised in Table 1 is offered as a means of holding these dimensions together rather than treating them in isolation.

Besides, the contribution of this article is conceptual rather than empirical. It offers a framework, drawing on critical data studies, on sociolinguistic and critical applied-linguistic scholarship, and on post-crisis education scholarship, that has not been assembled in this form before, and that aims to hold the technical, pedagogical, and political dimensions of AI-LA together for long enough to see them interact. The limitations follow from the genre, as a theory-synthesis paper can only do so much. What it can produce is a set of testable propositions for the empirical work that has to follow, and these are stated explicitly below, each keyed to a component of the framework set out in Table 1.

- **P1 (datafication).** Interaction traces capture a diminishing share of the variance in language-learning outcomes as the construct assessed moves from discrete form towards situated communicative competence, so that analytics-driven inferences lose validity precisely where the pedagogical stakes are highest.
- **P2 (personalisation).** Adaptive personalisation improves measurable gains for narrowly specified, form-focused tasks but produces differential effects across learners' motivational and linguistic profiles, such that uniform deployment widens rather than narrows existing gaps.
- **P3 (algorithmic bias).** Automated assessment of speech and writing systematically scores speakers of non-dominant varieties and multilingual learners more harshly than trained human raters, and the size of the gap tracks the under-representation of those varieties in the training distribution rather than any difference in communicative adequacy.
- **P4 (linguistic legitimacy).** Participatory, sociolinguistically informed design that treats non-dominant varieties as legitimate at the level of the training data measurably reduces performance gaps relative to systems optimised against a single standard norm.
- **P5 (governance).** Compliance with high-risk regulatory requirements is necessary but not sufficient for equity, since a system can satisfy a conformity assessment while leaving the sociolinguistic assumptions embedded in its training data unexamined.

ISSN: 1857-7881 (Print)
e-ISSN 1857-7431 (Online)

P6 (post-crisis displacement). The performance of an AI-LA system for displaced and crisis-affected learners is a more sensitive indicator of its overall equity than its aggregate accuracy, because these learners concentrate the linguistic and contextual features least represented in standard training data.

Pursued through genuine collaboration between researchers in the learning and social sciences, applied linguists, and technology developers, the empirical examination of these propositions looks like the most promising route toward AI-LA in language education that is technically capable, socially responsible, and educationally equitable in the same breath.

Acknowledgments

This article is based on work previously presented at the 14th Eurasian Multidisciplinary Forum (EMF 2025), held on 20-21 November 2025 at Grigol Robakidze University, Tbilisi, Georgia. The conference presentation version has since been substantially revised, expanded, and refined for publication in its current form. The authors are grateful to the conference participants for their constructive comments on the earlier version of this work. The authors also acknowledge the use of an AI-based language tool for proofreading and linguistic refinement of the final version of the text.

Author contributions

Conceptualization, literature review, writing the original draft (NKU); Review and editing, and final approval of the version to be published (VK).

Conflict of Interest: The authors reported no conflict of interest.

Data Availability: All data are included in the content of the paper.

Funding Statement: The authors did not obtain any funding for this research.

References:

1. Baker, R. S., & Hawn, A. (2022). Algorithmic bias in education. *International Journal of Artificial Intelligence in Education*, *32*(4), 1052-1092. https://doi.org/10.1007/s40593-021-00285-9
2. Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In J. A. Larusson & B. White (Eds.), *Learning analytics: From research to practice* (pp. 61-75). Springer. https://doi.org/10.1007/978-1-4614-3305-7_4
3. Barocas, S., Hardt, M., & Narayanan, A. (2019). *Fairness and machine learning: Limitations and opportunities*. MIT Press. https://fairmlbook.org

ISSN: 1857-7881 (Print)
e-ISSN 1857-7431 (Online)

4. Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, *15*(5), 662-679. https://doi.org/10.1080/1369118X.2012.678878
5. Charitonos, K., Khalil, B., Ross, T. W., Bonfini-Hotlosz, C., Aristorenas, M., & Webster, B. (2025). Teacher autonomy in refugee settings: English language teachers' negotiation of professional agency in displacement contexts. *ReCALL*, *37*(1), 24-41. https://doi.org/10.1017/S0958344024000156
6. Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.
7. Creely, E., & Henderson, M. (2025). Artificial intelligence and the digital dis/empowerment of migrant and refugee learners. In E. Tour, E. Creely, P. Waterhouse, & M. Henderson (Eds.), *Digital empowerment for refugee and migrant learners: Applying strengths-based practice to adult education* (pp. 47-63). Routledge. https://doi.org/10.4324/9781003422891-5
8. Dryden-Peterson, S. (2017). Refugee education: The crossroads of globalization. *Educational Researcher*, *46*(9), 493-506. https://doi.org/10.3102/0013189X17723028
9. European Parliament and Council. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonized rules on artificial intelligence (Artificial Intelligence Act)*. Official Journal of the European Union, L series. https://eur-lex.europa.eu/eli/reg/2024/1689/oj
10. Gallegos, I. O., Rossi, R. A., Barrow, J., Tanjim, M. M., Kim, S., Dernoncourt, F., Yu, T., Zhang, R., & Ahmed, N. K. (2024). Bias and fairness in large language models: A survey. *Computational Linguistics, 50*(3), 1097-1179. https://doi.org/10.1162/coli_a_00524
11. Gee, J. P. (2004). *Situated language and learning: A critique of traditional schooling*. Routledge.
12. Godwin-Jones, R. (2022). Partnering with AI: Intelligent writing assistance and instructed language learning. *Language Learning & Technology*, *26*(2), 5-24. https://doi.org/10125/73474
13. Huang, X., Zou, D., Cheng, G., Chen, X., & Xie, H. (2023). Trends, research issues and applications of artificial intelligence in language education. *Educational Technology & Society*, *26*(1), 112-131. https://doi.org/10.30191/ETS.202301_26(1).0009
14. Hymes, D. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics: Selected readings* (pp. 269-293). Penguin.

ISSN: 1857-7881 (Print)
e-ISSN 1857-7431 (Online)

15. Jaakkola, E. (2020). Designing conceptual articles: Four approaches. *AMS Review*, *10*(1), 18-26. https://doi.org/10.1007/s13162-020-00161-0
16. Jeon, J. (2025). Artificial intelligence in ESL/EFL education: Evidence from recent reviews (2024-2025). *International Journal of Learning, Teaching and Educational Research*, *24*(9), 1-22. https://doi.org/10.26803/ijlter.24.9.1
17. Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures & their consequences*. SAGE.
18. Koenecke, A., Nam, A., Lake, E., Nudell, J., Quartey, M., Mengesha, Z., Toups, C., Rickford, J. R., Jurafsky, D., & Goel, S. (2020). Racial disparities in automated speech recognition. *Proceedings of the National Academy of Sciences*, *117*(14), 7684-7689. https://doi.org/10.1073/pnas.1915768117
19. Lee, I., & Yeung, M. W. (2024). Generative AI and English language teaching and learning: Insights from expert interviews. *Computer Assisted Language Learning*. Advance online publication. https://doi.org/10.1080/09588221.2024.2347234
20. Menashy, F., & Zakharia, Z. (2022). Crisis upon crisis: Refugee education responses amid COVID-19. *Peabody Journal of Education*, *97*(3), 309-325. https://doi.org/10.1080/0161956X.2022.2079895
21. Pennycook, A. (2001). *Critical applied linguistics: A critical introduction*. Lawrence Erlbaum.
22. Perrotta, C., Selwyn, N., & Ewin, C. (2022). Artificial intelligence and the affective labour of understanding: The intimate moderation of a language model. *New Media & Society*. Advance online publication. https://doi.org/10.1177/14614448221075296
23. Prinsloo, P., & Slade, S. (2017). An elephant in the learning analytics room: The obligation to act. In A. Wise, P. Winne, & G. Lynch (Eds.), *Proceedings of the 7th International Learning Analytics & Knowledge Conference* (pp. 46-55). ACM. https://doi.org/10.1145/3027385.3027406
24. Reich, J., Buttimer, C. J., Fang, A., Hillaire, G., Hirsch, K., Larke, L. R., Littenberg-Tobias, J., Moussapour, R. M., Napier, A., Thompson, M., & Slama, R. (2020). *Remote learning guidance from state education agencies during the COVID-19 pandemic: A first look*. EdArXiv. https://doi.org/10.35542/osf.io/437e2
25. Romero, C., & Ventura, S. (2013). Data mining in education. *WIREs Data Mining and Knowledge Discovery*, *3*(1), 12-27. https://doi.org/10.1002/widm.1075
26. Selwyn, N. (2019). *Should robots replace teachers? AI and the future of education*. Polity Press.

ISSN: 1857-7881 (Print)
e-ISSN 1857-7431 (Online)

27. Selwyn, N. (2022). The future of AI and education: Some cautionary notes. *European Journal of Education*, *57*(4), 620-631. https://doi.org/10.1111/ejed.12532
28. Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioural Scientist*, *57*(10), 1380-1400. https://doi.org/10.1177/0002764213498851
29. Teng, M. F. (2025). Generative AI (GenAI) in the language classroom: A systematic review. *Interactive Learning Environments*. Advance online publication. https://doi.org/10.1080/10494820.2025.2498537
30. Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
31. Wah, J. N. K. (2025). Artificial intelligence in language learning: A systematic review of personalization and learner engagement. *Forum for Linguistic Studies, 7*(9), 327-341. https://doi.org/10.30564/fls.v7i9.10336
32. Williamson, B., Eynon, R., & Potter, J. (2020). Pandemic politics, pedagogies and practices: Digital technologies and distance education during the coronavirus emergency. *Learning, Media and Technology*, *45*(2), 107-114. https://doi.org/10.1080/17439884.2020.1761641
33. Williamson, B., Komljenovic, J., & Gulson, K. N. (Eds.). (2024). *World yearbook of education 2024: Digitalization of education in the era of algorithms, automation and artificial intelligence*. Routledge. https://doi.org/10.4324/9781003359722