INTERNATIONAL JOURNAL OF SOCIAL SCIENCES BULLETIN
ISSN: (E) 3007-1917 (P) 3007-1909
Volume 4, Issue 5, 2026

ALGORITHMIC LANGUAGE EXPOSURE AND LINGUISTIC IDENTITY FORMATION: EXAMINING THE IMPACT OF AI-DRIVEN EDUCATIONAL PLATFORMS ON MULTILINGUAL LEARNERS IN PAKISTAN

Muniba Tabassum[*1], Priya Inam Shaikh[2], Syeda Mahnoor Saeed[3], Muhammad Farhan Amjad[4]

[*1]*Student, English, Mirpur University of Science and Technology*
[2]*Junior School Teacher, Education Department*
[3]*Lecturer, Research Assistant, Department of English, Abasyn University, Peshawar*
[4]*Manager Performance at a Power Plant*

[*1]muniba914@gmail.com, [2]priyashaikh1111@gmail.com, [4]farhan.sheikh@ymail.com

Corresponding Author: *
Muniba Tabassum
DOI: https://doi.org/10.5281/zenodo.20022615

Received	Accepted	Published
11 March 2026	21 April 2026	04 May 2026

ABSTRACT

The increasing integration of artificial intelligence (AI)-driven educational platforms in language learning has introduced new dynamics of algorithmic language exposure and its influence on linguistic identity formation. This study examined the impact of AI-mediated language learning environments on multilingual learners in Pakistan, with a particular focus on how continuous algorithmic feedback and standardized linguistic input shape language preferences and identity construction. Grounded in Sociocultural Theory, the study employed a mixed-methods research design involving 320 multilingual learners selected through stratified random sampling, along with in-depth interviews from 20 participants. Quantitative findings revealed a significant relationship between algorithmic language exposure and linguistic identity transformation, with AI-driven platforms strongly influencing English language preference and contributing to a gradual decline in local language usage. Regression analysis indicated that algorithmic exposure, linguistic standardization pressure, and English preference shift collectively explained 62% of the variance in linguistic identity transformation. Qualitative findings further supported these results, highlighting themes of algorithmic English dominance, reduced linguistic flexibility, and increased dependence on AI systems for linguistic validation. The study concludes that while AI-driven educational tools enhance language learning efficiency and accessibility, they also contribute to linguistic standardization and identity shifts in multilingual contexts. The findings underscore the need for culturally responsive AI design and balanced pedagogical integration to preserve linguistic diversity.

Keywords: *Artificial Intelligence; Algorithmic Language Exposure; Linguistic Identity; Multilingual Learners; AI-Driven Education; Sociocultural Theory; Language Learning; Pakistan*

INTRODUCTION

The rapid integration of Artificial Intelligence (AI)–driven educational platforms into language learning environments has fundamentally transformed the ways in which multilingual learners engage with linguistic input, process meaning, and construct linguistic identity. These

algorithmic systems, including adaptive learning applications, generative AI chatbots, and automated feedback tools, are increasingly shaping educational experiences by providing personalized, data-driven, and interactive learning pathways. In multilingual contexts such as Pakistan, where linguistic diversity intersects with complex sociocultural hierarchies, the emergence of AI-mediated language learning raises critical questions regarding algorithmic language exposure and its implications for linguistic identity formation.

Recent research indicates that AI-based language technologies are not neutral tools but are embedded with sociotechnical and ideological assumptions that may privilege dominant linguistic norms, particularly standardized English varieties, while marginalizing local linguistic repertoires (Abbasi et al., 2026; Memon et al., 2026). In Pakistan, English occupies a powerful yet contested position within education, governance, and socioeconomic mobility, and AI-driven platforms are increasingly reinforcing this dominance through algorithmic feedback systems that prioritize native-speaker norms and globalized linguistic standards (Abbasi et al., 2026). Consequently, learners are frequently exposed to algorithmically curated language models that may inadvertently shape their perceptions of linguistic correctness, prestige, and identity.

From a sociolinguistic perspective, linguistic identity is understood as a dynamic construct shaped by interaction, ideology, and language use within specific sociocultural contexts. In multilingual educational environments, learners continuously negotiate identities across multiple languages such as Urdu, English, and regional languages including Punjabi, Pashto, Sindhi, and Balochi. However, AI-mediated educational platforms may reconfigure these identity negotiations by privileging English-centric linguistic input and limiting the visibility of local linguistic practices within digital learning ecosystems (Khalid et al., 2025). This phenomenon raises concerns regarding digital linguistic inequality and algorithmic reinforcement of linguistic hierarchies.

Furthermore, studies on AI adoption in Pakistani higher education highlight that learners and educators increasingly rely on generative AI tools for writing, translation, and language practice, which significantly influences learner autonomy, engagement, and communicative competence (Abbasi et al., 2025; Khan et al., 2025). While such technologies enhance accessibility and learning efficiency, they also risk standardizing linguistic production and reducing exposure to culturally embedded language use. This aligns with broader critiques that AI-mediated language learning systems may function as instruments of digital colonialism by embedding Western linguistic ideologies into educational technologies (Memon et al., 2026).

In addition, algorithmic opacity in educational AI systems limits learners' and educators' understanding of how language feedback and evaluation are generated, thereby shaping language learning behaviors in largely invisible ways (Raza & Aslam, 2024). This lack of transparency can reinforce dependency on algorithmic judgments of correctness, potentially weakening critical linguistic awareness and reducing opportunities for localized pedagogical adaptation in multilingual classrooms.

Given these dynamics, it becomes essential to critically examine how algorithmic language exposure through AI-driven educational platforms influences linguistic identity formation among multilingual learners in Pakistan. Understanding this relationship is crucial for developing equitable, context-sensitive, and ethically grounded AI applications in language education that support linguistic diversity rather than suppress it.

Problem Statement

The rapid integration of artificial intelligence (AI)-driven educational platforms into language learning environments has fundamentally altered the nature of linguistic exposure, interaction, and identity construction among multilingual learners. In Pakistan, where linguistic diversity encompasses Urdu, English, and multiple regional languages, AI-mediated language learning tools are increasingly shaping how learners acquire, process,

and reproduce language. These platforms—ranging from intelligent tutoring systems to generative AI chatbots—provide continuous algorithmic feedback and standardized linguistic input, which may significantly influence learners' linguistic choices and communicative identities.

While AI-based language learning technologies are widely acknowledged for enhancing accessibility, personalization, and learner engagement, their socio-linguistic implications remain insufficiently explored, particularly in developing multilingual contexts. Existing research has largely focused on pedagogical effectiveness and cognitive learning outcomes, overlooking how algorithmic language exposure may implicitly reinforce dominant language ideologies, especially privileging standardized English forms over local linguistic variations. This raises critical concerns regarding the potential marginalization of indigenous linguistic identities and the subtle reshaping of learners' self-perception as language users within algorithmically governed environments.

Furthermore, there is a limited empirical understanding of how continuous interaction with AI-driven platforms contributes to linguistic identity formation among multilingual learners in Pakistan. The absence of localized research addressing the intersection of AI, language exposure, and identity construction creates a significant theoretical and practical gap. This study therefore addresses the need to critically examine how algorithmic mediation of language learning influences linguistic identity development, agency, and multilingual expression in Pakistani educational contexts.

Research Questions

1. How do AI-driven educational platforms influence algorithmic language exposure among multilingual learners in Pakistan?
2. In what ways does AI-mediated language learning affect learners' linguistic identity formation?
3. How do multilingual learners perceive the role of AI systems in shaping their language preferences and linguistic confidence?
4. To what extent do AI-driven platforms reinforce or challenge existing linguistic hierarchies in Pakistani educational contexts?
5. What are the socio-cultural implications of sustained interaction with AI-based language learning tools on multilingual identity negotiation?

Research Objectives

1. To examine the nature and extent of algorithmic language exposure through AI-driven educational platforms among multilingual learners in Pakistan.
2. To analyze the impact of AI-mediated language learning on linguistic identity formation and self-representation.
3. To explore learners' perceptions of AI systems in influencing their language use, preferences, and communicative behavior.
4. To investigate the role of AI-driven platforms in reinforcing or disrupting linguistic hierarchies within multilingual educational settings.
5. To assess the broader socio-cultural implications of AI-mediated language learning on multilingual identity development in Pakistan.

Significance of the Study

This study is significant in advancing both theoretical understanding and practical insights into the emerging intersection of artificial intelligence (AI), language education, and multilingual identity formation. As AI-driven educational platforms increasingly mediate language learning processes, particularly through adaptive algorithms and generative language tools, there is a growing need to critically evaluate their broader linguistic and socio-cultural implications.

From a theoretical perspective, the study contributes to applied linguistics, sociolinguistics, and digital language education by conceptualizing *algorithmic language exposure* as a distinct phenomenon shaping linguistic behavior and identity construction. It extends existing literature on computer-assisted language learning (CALL) and AI-assisted language learning (AIALL) by integrating a critical lens that interrogates how algorithmic systems influence language

hierarchies, learner agency, and identity negotiation in multilingual contexts.

Practically, the findings of this study are expected to inform educators, curriculum designers, and policymakers in Pakistan regarding the implications of integrating AI-based language learning technologies into educational systems. By highlighting how these platforms may reinforce or challenge linguistic hierarchies—particularly the dominance of standardized English over local languages—the study provides evidence-based insights for developing more inclusive and culturally responsive digital learning environments.

Furthermore, the study holds relevance for AI developers and educational technology designers by emphasizing the need for culturally sensitive algorithmic models that accommodate linguistic diversity rather than homogenizing language use. This is particularly important in multilingual societies such as Pakistan, where linguistic identity is closely tied to cultural heritage and social belonging.

At the learner level, the study offers an understanding of how continuous interaction with AI systems shapes learners' linguistic self-perception, communicative confidence, and identity formation. Such insights can support the development of pedagogical strategies that balance technological innovation with linguistic and cultural preservation.

Overall, this research contributes to bridging the gap between technological advancement and socio-linguistic equity, ensuring that AI-driven education supports rather than constrains multilingual identity development.

Literature Review

Artificial Intelligence in Language Learning

The integration of artificial intelligence (AI) into language education has transformed traditional pedagogical approaches by enabling adaptive, data-driven, and learner-centered environments. AI-driven language learning systems, including intelligent tutoring systems, chatbots, and generative language models, have been widely recognized for their ability to provide personalized feedback, immediate correction, and interactive linguistic input (Zhang & Dong, 2024). These technologies are part of a broader shift from conventional Computer-Assisted Language Learning (CALL) to AI-Assisted Language Learning (AIALL), where machine learning algorithms actively mediate language acquisition processes rather than simply supporting instruction (Bahari et al., 2025).

Recent studies emphasize that AI-based platforms enhance learner autonomy by allowing continuous engagement with target language input beyond classroom constraints. For instance, generative AI tools such as conversational agents simulate authentic communication, thereby increasing exposure to contextualized language use (Zou et al., 2025). However, scholars also caution that such systems often rely on standardized corpora dominated by native-speaker norms, which may inadvertently reinforce linguistic homogeneity and marginalize local language varieties (Abbasi et al., 2026).

Algorithmic Language Exposure

The concept of algorithmic language exposure refers to the continuous and structured linguistic input generated and filtered through AI systems. Unlike traditional exposure, which is socially and contextually grounded, algorithmic exposure is shaped by predictive models, training datasets, and optimization algorithms that determine which linguistic forms are considered "correct" or "preferred" (Khan et al., 2025). This form of exposure significantly influences learners' vocabulary acquisition, syntactic preferences, and discourse patterns.

Studies indicate that algorithmic systems tend to prioritize standardized English forms due to the predominance of high-resource datasets used in training AI models (Zou et al., 2025). As a result, learners interacting with AI-driven platforms may gradually align their linguistic output with algorithmically reinforced norms. This raises concerns about linguistic narrowing, particularly in multilingual societies where linguistic diversity is central to cultural identity.

Furthermore, algorithmic exposure is not neutral; it reflects embedded ideological assumptions about language correctness and proficiency.

Abbasi et al. (2026) argue that AI-driven language tools may perpetuate digital linguistic colonialism by privileging global English over indigenous languages, thereby influencing learners' linguistic behavior in subtle but powerful ways.

Linguistic Identity Formation in Multilingual Contexts

Linguistic identity formation is a dynamic process shaped by social interaction, cultural context, and language use practices. In multilingual societies, individuals often negotiate multiple linguistic identities depending on context, audience, and communicative purpose. Traditionally, this process has been examined through sociocultural and poststructuralist frameworks, emphasizing identity as fluid and socially constructed.

Recent literature suggests that digital environments are increasingly central to identity formation processes. AI-mediated language learning platforms introduce new dimensions to this process by embedding learners in algorithmically curated linguistic spaces (Abbas et al., 2026). These environments influence not only what language is learned but also how learners perceive their linguistic competence and legitimacy.

In Pakistan, where English functions as a gateway to academic and professional advancement while Urdu and regional languages represent cultural identity, linguistic identity formation is particularly complex. Exposure to AI-driven platforms may intensify this complexity by privileging English proficiency as the dominant marker of academic success, potentially reshaping learners' linguistic self-concept (Khan et al., 2025).

AI-Driven Educational Platforms and Pedagogical Transformation

The adoption of AI-driven educational platforms in language learning has significantly altered pedagogical practices. These systems enable adaptive learning pathways, automated assessment, and real-time corrective feedback, thereby enhancing instructional efficiency and learner engagement (Bahari et al., 2025). Moreover, they support blended and hybrid learning models, which have become increasingly prevalent in higher education institutions in Pakistan.

However, while these technologies improve accessibility and scalability, they also raise critical pedagogical concerns. One major issue is the reduction of teacher mediation in language learning, which traditionally plays a crucial role in contextualizing linguistic input and supporting identity negotiation. AI systems, in contrast, rely on algorithmic standardization that may lack sensitivity to cultural and linguistic diversity (Zhang & Dong, 2024).

Additionally, the reliance on AI-generated feedback may lead to over-dependence on automated correction systems, potentially diminishing learners' critical awareness of language variation and pragmatic flexibility.

Multilingual Learners in Pakistan and Digital Language Learning

Pakistan presents a highly multilingual environment where Urdu serves as the national language, English functions as an official and academic language, and numerous regional languages contribute to linguistic diversity. In such a context, language learning is not merely a cognitive process but also a socio-political and identity-based phenomenon.

The introduction of AI-driven language learning platforms into this environment has created new opportunities for exposure to English and other global languages. However, research suggests that these platforms often prioritize English-centric datasets, thereby limiting exposure to indigenous linguistic structures and expressions (Zou et al., 2025). This imbalance may contribute to linguistic stratification, where English is increasingly associated with technological competence and socio-economic mobility, while local languages are relegated to informal domains.

Despite growing adoption, empirical studies examining the socio-linguistic implications of AI in Pakistan remain limited. Most existing research focuses on technological adoption and learning outcomes, with insufficient attention to identity formation and linguistic equity (Abbasi et al., 2026).

Although substantial literature exists on AI in language learning and CALL systems, there is a clear gap in understanding how algorithmic mediation influences linguistic identity formation in multilingual contexts. Specifically, limited research has examined the socio-cultural consequences of continuous AI-mediated language exposure in developing countries such as Pakistan.

Furthermore, existing studies rarely integrate critical linguistic perspectives with AI-based educational research, resulting in a fragmented understanding of how technology shapes both language learning and identity construction. This study addresses this gap by examining the intersection of algorithmic language exposure, AI-driven educational platforms, and linguistic identity formation among multilingual learners in Pakistan.

Underpinning Theory: Sociocultural Theory of Language Learning (Vygotskian Perspective)

This study is primarily grounded in the Sociocultural Theory (SCT) of language learning, originally proposed by Lev Vygotsky, which conceptualizes learning as a socially mediated process rather than an individual cognitive activity. According to this theory, language development occurs through interaction with more knowledgeable others and is shaped by cultural, historical, and technological contexts. Central to SCT is the idea that cognitive and linguistic development are mediated through tools and signs, including language itself, which act as instruments for meaning-making and knowledge construction.

In the context of AI-driven educational platforms, SCT provides a strong theoretical foundation for understanding how algorithmic mediation functions as a new form of learning tool**.** AI systems such as intelligent tutoring systems, chatbots, and generative language models act as mediational artifacts that shape learners' exposure to linguistic input, feedback patterns, and communicative structures. These tools essentially replace or supplement human "more knowledgeable others" by providing continuous, automated scaffolding within the learner's Zone of Proximal Development (ZPD)**.**

Within this framework, algorithmic language exposure can be understood as a technologically mediated extension of scaffolding, where AI systems guide learners toward linguistically acceptable forms based on pre-trained models and statistical language norms. However, unlike human mediation, AI-based scaffolding is largely shaped by datasets that often privilege dominant linguistic varieties, particularly standardized English. This raises important theoretical concerns regarding the cultural neutrality of mediation tools and their influence on identity formation.

From an identity perspective, SCT also emphasizes that learning is inherently tied to participation in socially and culturally situated practices. In multilingual contexts such as Pakistan, learners negotiate multiple linguistic identities through interaction with both local languages and global languages like English. AI-driven platforms, by continuously exposing learners to algorithmically filtered linguistic input, may influence how learners position themselves within these linguistic hierarchies, thereby shaping their evolving linguistic identities.

Thus, Sociocultural Theory provides a comprehensive lens to analyze how AI-mediated learning environments function not merely as instructional tools but as powerful mediational systems that shape both language acquisition and identity formation.

Methodology

Research Design

The study adopted a mixed-methods research design to comprehensively examine the relationship between algorithmic language exposure and linguistic identity formation among multilingual learners. This design was selected to integrate both quantitative measurement of learner perceptions and qualitative exploration of lived linguistic experiences within AI-mediated educational environments. The combination of approaches enabled a more holistic understanding of how AI-driven platforms influenced language learning and identity construction.

Population of the Study

The population of the study consisted of multilingual learners enrolled in higher education institutions in Pakistan, who were actively engaged in AI-driven language learning platforms. This included students from disciplines such as English language studies, education, computer science, and social sciences, where digital learning tools were frequently integrated into academic instruction. The population represented learners exposed to Urdu, English, and regional languages, reflecting the multilingual educational context of Pakistan.

Sample and Sampling Technique

A total sample of 320 participants was selected from the target population. The sample was drawn using a stratified random sampling technique to ensure proportional representation across gender, academic discipline, and proficiency levels in English. This approach enhanced the representativeness of the sample and minimized selection bias.

In addition, for the qualitative component, 20 participants were purposively selected from the main sample. These participants were chosen based on their extensive use of AI-driven language learning platforms and their willingness to provide in-depth insights into their linguistic experiences.

Data Collection Instruments

For the quantitative phase, a structured questionnaire was developed, consisting of Likert-scale items measuring algorithmic language exposure, perceived linguistic influence, and identity-related outcomes. The instrument was adapted from existing validated scales in digital language learning research and modified to suit the Pakistani multilingual context.

For the qualitative phase, semi-structured interviews were conducted to explore learners' experiences with AI-based educational platforms, focusing on language preference shifts, identity negotiation, and perceived linguistic changes.

Data Collection Procedure

Data were collected after obtaining formal permission from the selected institutions. The questionnaire was distributed both physically and electronically to ensure wider accessibility. Interviews were conducted in person and via online platforms, depending on participant availability. Informed consent was obtained from all participants prior to data collection, and confidentiality was strictly maintained throughout the research process.

Data Analysis Analysis

Quantitative data were analyzed using descriptive statistics (mean, standard deviation, and frequency distribution) and inferential statistical techniques, including correlation and regression analysis, to examine relationships between variables. Qualitative data were analyzed using thematic analysis, where interview transcripts were coded and categorized into emerging themes related to linguistic identity formation and AI-mediated language exposure.

Data Analysis

Quantitative Analysis

The quantitative data collected from 320 multilingual learners were analyzed using descriptive and inferential statistics. The analysis focused on examining algorithmic language exposure, linguistic identity formation, and the relationship between AI-driven platform usage and identity shifts.

Table 1: Descriptive Statistics of Key Variables (n = 320)

Variable	Mean	Std. Deviation	Interpretation
Algorithmic Language Exposure	3.89	0.74	High exposure
Linguistic Standardization Pressure	3.76	0.81	Moderately high
English Language Preference Shift	4.02	0.69	High
Local Language Usage Decline	3.58	0.88	Moderate
Linguistic Identity Transformation	3.91	0.77	High

The results indicated a high level of algorithmic language exposure (M = 3.89) among respondents, suggesting frequent interaction with AI-driven language learning platforms. Similarly, linguistic identity transformation also showed a high mean score (M = 3.91), indicating that learners perceived noticeable changes in their linguistic self-concept due to AI-mediated exposure.

A comparatively higher English language preference shift (M = 4.02) suggests that AI systems significantly influenced learners toward greater reliance on English, reinforcing its dominance as the preferred academic and communicative language. Conversely, the moderate decline in local language usage (M = 3.58) reflects a gradual but noticeable reduction in indigenous language engagement within digital learning environments.

Table 2: Correlation Analysis

Variables	ALE	LST	ELPS	LLUD
Algorithmic Language Exposure (ALE)	1			
Linguistic Standardization Pressure (LST)	0.71**	1		
English Language Preference Shift (ELPS)	0.68**	0.74**	1	
Local Language Usage Decline (LLUD)	0.59**	0.62**	0.66**	1

Note: $p < 0.01$

The correlation results revealed a strong positive relationship between algorithmic language exposure and linguistic standardization pressure ($r = 0.71$, $p < 0.01$), indicating that increased exposure to AI systems was associated with stronger perceptions of standardized language norms.

Similarly, a significant positive correlation was observed between algorithmic language exposure and English language preference shift ($r = 0.68$, $p < 0.01$), suggesting that AI-driven platforms actively encouraged English dominance in learners' linguistic behavior.

Furthermore, the positive relationship between algorithmic exposure and local language usage decline ($r = 0.59$, $p < 0.01$) indicated that increased reliance on AI systems may contribute to reduced engagement with indigenous languages.

Table 3: Regression Analysis (Predicting Linguistic Identity Transformation)

Predictor Variable	Beta (β)	t-value	Significance
Algorithmic Language Exposure	0.41	7.82	0.000***
Linguistic Standardization Pressure	0.36	6.54	0.000***
English Language Preference Shift	0.33	5.91	0.000***

$R^2 = 0.62$, $F = 88.45$, $p < 0.001$

The regression model explained 62% of the variance in linguistic identity transformation ($R^2 = 0.62$), indicating a strong predictive capacity of the independent variables.

Algorithmic language exposure emerged as the strongest predictor ($\beta = 0.41$), confirming that increased interaction with AI-driven platforms significantly influenced learners' linguistic identity formation. Linguistic standardization pressure ($\beta = 0.36$) and English language preference shift ($\beta = 0.33$) also showed significant predictive effects, highlighting the combined role of AI systems in reshaping linguistic identity through normalization of standardized English usage.

Qualitative Analysis

Thematic analysis of 20 semi-structured interviews revealed four major themes:

Theme 1: Algorithmic English Dominance

Participants reported that AI tools consistently corrected or prioritized English structures, leading

to gradual internalization of English-centric norms.

"Whenever I use AI tools, it always suggests formal English. Even my Urdu-English mixing gets corrected." (Participant 7)

Theme 2: Decline of Linguistic Flexibility

Learners expressed reduced confidence in using Urdu or regional language structures in academic contexts due to AI correction patterns.

Theme 3: Identity Shift Toward Global Linguistic Identity

Many respondents reported feeling more "global" but less connected to local linguistic identity.

"I feel more confident in English now, but sometimes I feel I am losing my natural way of speaking Urdu." (Participant 12)

Theme 4: Dependence on AI for Linguistic Validation

Participants increasingly relied on AI systems for grammatical correctness and vocabulary selection, reducing autonomous language decision-making.

The integration of findings demonstrated that AI-driven educational platforms significantly influenced both linguistic behavior and identity formation among multilingual learners in Pakistan. Quantitative results confirmed strong statistical relationships between algorithmic exposure and identity transformation, while qualitative insights revealed the lived experience of linguistic standardization and identity negotiation.

Overall, the findings suggest that AI systems function not only as pedagogical tools but also as **ideological mediators of language**, shaping learners toward standardized English norms while gradually weakening engagement with indigenous linguistic identities.

Discussion

The findings of this study demonstrate that AI-driven educational platforms play a significant role in shaping both linguistic practices and identity formation among multilingual learners in Pakistan. The quantitative results revealed a strong positive relationship between algorithmic language exposure and linguistic identity transformation, suggesting that continuous interaction with AI systems meaningfully influences how learners perceive and construct their linguistic selves. This aligns with Sociocultural Theory, which posits that cognitive and linguistic development is mediated through tools; however, in this context, mediation is increasingly performed by algorithmic systems rather than human interlocutors.

The study further indicates that AI platforms tend to reinforce standardized English norms through automated correction and predictive language modeling. This finding is consistent with prior research highlighting the dominance of high-resource languages in AI training datasets, which often marginalize linguistic diversity. As learners increasingly depend on AI-generated feedback, their linguistic output gradually aligns with algorithmically preferred structures, leading to a subtle but significant shift in language preference and usage patterns. In multilingual contexts like Pakistan, this creates a tension between global linguistic integration and local linguistic preservation.

Qualitative insights reinforced these patterns by revealing learners' subjective experiences of linguistic adjustment and identity negotiation. Participants reported increased confidence in English communication but simultaneously expressed concerns about diminishing fluency and emotional connection with Urdu and regional languages. This duality reflects a broader identity dilemma where AI-mediated language learning simultaneously empowers learners linguistically while reshaping their cultural-linguistic affiliations.

Conclusion

This study concludes that AI-driven educational platforms significantly influence algorithmic language exposure and contribute to the reshaping of linguistic identity among multilingual learners in Pakistan. The findings confirm that continuous engagement with AI systems leads to increased alignment with standardized English norms, while simultaneously reducing the active use and

perceived relevance of indigenous languages in academic contexts.

Although AI technologies enhance accessibility, efficiency, and personalization in language learning, they also introduce unintended socio-linguistic consequences. These include linguistic standardization, reduced linguistic diversity in learner output, and evolving identity shifts toward a more globalized but less locally grounded linguistic self-concept. Therefore, AI in language education must be understood not only as a pedagogical innovation but also as a powerful socio-cultural force.

Implications of the Study

Theoretically, this study extends Sociocultural Theory by introducing the concept of algorithmic mediation as a new form of learning scaffolding. It highlights that mediation is no longer exclusively human-driven but increasingly governed by data-trained systems that shape linguistic input and feedback. This contributes to emerging discussions in applied linguistics regarding digital mediation and identity construction.

Practically, the findings have important implications for educators, policymakers, and educational technology developers. For educators, there is a need to critically integrate AI tools in ways that encourage balanced multilingual development rather than exclusive reliance on standardized English. For policymakers, the study underscores the importance of developing digital education policies that protect linguistic diversity in AI-integrated learning environments. For developers, it highlights the necessity of designing culturally responsive AI systems that recognize and incorporate local linguistic variations.

Future Directions

Future research should explore longitudinal effects of AI-mediated language learning on linguistic identity to determine whether observed shifts are temporary adaptation effects or long-term transformations. Comparative studies across different multilingual societies could also provide broader insights into how cultural and linguistic contexts mediate AI influence. Additionally, future studies should investigate the role of specific AI tools, such as large language models and conversational agents, in shaping pragmatic competence and discourse identity.

There is also a need for interdisciplinary research combining computational linguistics, education, and sociolinguistics to better understand algorithmic biases in language learning systems. Exploring learner agency in resisting or negotiating algorithmic language norms would further enrich the field.

Recommendations

It is recommended that educational institutions adopt a balanced approach to AI integration in language teaching by combining AI-based instruction with human-led pedagogical practices. Teachers should be trained to critically evaluate AI-generated feedback and guide students in understanding its limitations.

Furthermore, AI developers should be encouraged to include multilingual datasets that reflect regional linguistic diversity, particularly languages such as Urdu and regional Pakistani languages. Curriculum designers should also incorporate translanguaging practices that allow learners to fluidly move between languages rather than restricting them to standardized forms.

Finally, awareness programs should be introduced to help learners critically engage with AI tools, enabling them to use technology as a supportive resource rather than an authoritative linguistic gatekeeper.

Limitations of the Study

This study is limited by its cross-sectional design, which restricts the ability to establish long-term causal relationships between AI exposure and linguistic identity formation. The reliance on self-reported data may also introduce response bias, as participants' perceptions may not fully reflect actual linguistic behavior.

Additionally, the study focused primarily on higher education learners in Pakistan, which may limit the generalizability of findings to other educational levels or cultural contexts. The rapidly evolving nature of AI technologies also means that findings may need continuous updating as new platforms and models emerge.

Despite these limitations, the study provides valuable foundational insights into the socio-linguistic impact of AI-driven language learning in multilingual environments.

References

Abbas, H., Muneer, K., & Meeraj, A. (2026). AI-mediated translanguaging in multilingual EFL classrooms: Emerging practices in South Asia. *Contemporary Journal of Social Science Review, 4*(1), 310–328.

Abbasi, K. J., Tunio, W. A., Buriro, Z. A., & Memon, M. Y. (2026). Algorithmic bias and digital colonialism in AI-driven English language learning systems. *Journal of Applied Linguistics and TESOL, 9*(1), 1–20.

Ahn, S., & Lee, J. (2024). Generative AI in second language acquisition: Pedagogical opportunities and challenges. *Language Learning & Technology, 28*(2), 45–63.

Bahari, A., Han, F., & Strzelecki, A. (2025). Integrating CALL and AIALL: A hybrid model for intelligent language learning systems. *Education and Information Technologies, 30*, 14305–14333.

Chapelle, C. A. (2023). Technology-mediated language learning and sociocultural perspectives. *Modern Language Journal, 107*(1), 15–30.

Dai, Z., & Wu, H. (2024). Artificial intelligence and personalized language learning: A systematic review. *Computers & Education: Artificial Intelligence, 5*, 100178.

Godwin-Jones, R. (2023). Emerging technologies: Generative AI and language learning. *Language Learning & Technology, 27*(1), 1–12.

Gong, Y., & Zhang, D. (2025). AI-mediated language learning and learner autonomy: A critical review. *System, 117*, 103–118.

Huang, X., & Hew, K. F. (2024). Chatbots in language education: Enhancing interaction and feedback. *Educational Technology Research and Development, 72*(3), 811–830.

Khan, G. Z., Khan, A. S., & Sher, S. F. (2025). Artificial intelligence in second language acquisition: Evidence from Pakistan. *Journal of Applied Linguistics and TESOL, 12*(2), 55–72.

Li, C., & Wang, Y. (2024). Algorithmic mediation in digital language learning environments. *ReCALL, 36*(3), 289–305.

Lin, T., & Mubarok, H. (2025). Linguistic identity in AI-driven educational ecosystems. *Language, Culture and Curriculum, 38*(1), 21–39.

Long, M. (2023). Sociocultural theory and second language acquisition in digital contexts. *Applied Linguistics Review, 14*(2), 201–219.

Nakamura, Y., & Zhou, M. (2024). AI bias in language learning models and implications for multilingual education. *Computers and Composition, 68*, 102725.

Reinders, H., & Pegrum, M. (2023). Digital language learning in the age of AI. *Language Teaching Research Quarterly, 35*, 1–18.

Sun, L., & Zhang, Y. (2025). Generative AI in education: Transforming language learning environments. *Educational Research Review, 42*, 100556.

Thorne, S. L. (2024). Digital mediation and identity in language learning. *The Modern Language Journal, 108*(2), 345–362.

Warschauer, M. (2023). Technology and equity in language learning. *Annual Review of Applied Linguistics, 43*, 1–19.

Zhang, Y., & Dong, C. (2024). Generative AI and foreign language education: A digital transformation perspective. *Systems, 12*(11), 462.

Zou, M. M., Reinders, H., & Thomas, M. (2025). GenAI-mediated informal digital English learning in global contexts. *ReCALL, 38*(1), 1–18.