EPISTEMOLOGICAL BIAS IN DIGITAL FOREIGN LANGUAGE LEARNING: ALGORITHMIC SYSTEMS VERSUS SOCIAL REALITY OF LANGUAGE

Muhammad Rusydi[1]
Hastang[2]
Abd. Basir[3]
Rezqiatul Fitriah Binti Suharyono[4]

[1]Universitas Islam Negeri Alauddin Makassar, Indonesia
[2]Institut Agama Islam Negeri Bone, Indonesia
[3]Sekolah Tinggi Agama Islam Syekh Manshur Pandeglang, Indonesia
[4]Kolej Universiti Perguruan Ugama Seri Begawan, Brunei Darussalam

E-mail: muh.rusydi@uin-alauddin.ac.id

ABSTRACT

This study examines epistemological bias in digital foreign language learning by positioning algorithmic systems and social reality of language as two interrelated yet unequal dimensions. As a library-based study, it finds that epistemological bias emerges from the dominance of algorithmic structures built upon global linguistic datasets that fail to fully capture the socio-cultural complexity of language. Consequently, this leads to semantic reduction, oversimplification of linguistic structures, and the marginalization of dialects and local contexts in digital learning materials. Such tendencies align with broader findings that language technologies often privilege dominant language patterns while neglecting contextual variation. Moreover, AI-based systems may generate inequitable assessments and feedback for learners from diverse linguistic backgrounds. The tension between standardized algorithmic logic and dynamic nature of language underscores the need for a more contextual and inclusive epistemological reconstruction.

Keywords: Epistemological Bias, Algorithmic Systems, Social Reality of Language.

1) INTRODUCTION

The development of digital technology, especially Artificial Intelligence (AI) as an algorithmic system, has brought significant changes in foreign language learning. AI-based learning platforms, automatic translators, and language recognition systems have now become an integral part of global language education practices (Min He, Babar Nawaz Abbasi, & Jinhua He, 2025). This technology is often seen as an innovative solution

capable of enhancing the efficiency, personalization, and accessibility of learning. However, behind these advancements, fundamental epistemological issues arise concerning how language is represented, processed, and understood in digital systems. Epistemologically, language is not merely a static system of symbols but a dynamic, contextual social phenomenon rich with cultural meaning (Desi Purnama Sari, et. al., 2025) Language develops through social interaction, historical experiences, and continuously changing communication practices. However, the algorithmic systems in language learning technology tend to reduce this complexity into statistical patterns that can be calculated and predicted. In this process, language is represented as standardized data, thus neglecting the social and cultural dimensions inherent in language practices (Sarthak Sharma, 2023)

This reductionist tendency is increasingly evident in the design of artificial intelligence systems that rely on large data corpora. These systems are trained to recognize common patterns in language, but often neglect more subtle and contextual linguistic variations (Agung Susilo & Masagus Firdaus, 2025). Research shows that AI-based language technology only optimally represents a small fraction of the world's languages, particularly those with political and economic dominance, while other languages are underrepresented. This indicates the presence of an epistemological bias rooted in the processes of data selection and representation. Furthermore, algorithmic bias in foreign language learning is not only technical but also ideological. Digital foreign language learning systems tend to internalize certain language norms—for example, native speaker standards or dominant variations—as a benchmark of linguistic correctness. In the context of English language learning, for instance, digital technology often reproduces Western standards and ignores local variations or multilingual practices (Jaeho Jeon, Seongyong Lee, & Serafín M. Coronel-Molina, 2025). As a result, the socially diverse reality of language is simplified into a single standard considered universal.

In addition, algorithmic systems operate based on computational logic that is static and deterministic, whereas language in social practice is dynamic, negotiable, and contextual. This mismatch creates what can be referred to as an epistemological tension between 'language as

data' and 'language as social practice.' (Desi Purnama Sari, et. al., 2025) In many cases, algorithms are unable to capture implicit meaning, cultural nuances, or pragmatic context that are an important part of language communication. This phenomenon is also related to the concept of epistemic injustice, where technological systems indirectly ignore or distort certain forms of knowledge. Bias in language technology can lead to the loss of meaning diversity, simplification of cultural concepts, as well as marginalization of certain language communities (Agung Susilo & Masagus Firdaus, 2025) Indeed, in some cases, language technology produces "linguistic impoverishment" due to the loss of lexical and morphological variation in the language modeling process. This indicates that epistemological bias not only affects language representation but also the quality of the language itself. Therefore, it is important to view foreign language learning technology not as a neutral tool, but as a socio-technological construct that carries certain assumptions, values, and interests. A critical perspective on algorithmic systems becomes essential to understand how knowledge about language is produced, selected, and disseminated in the digital ecosystem (Khalil Jibran Abbasi, Wazir Ali Tunio, Zaheer Abbas Buriro, & Muhammad Yaseen Memon, 2026). This approach opens space to reconsider the relationship between technology, language, and power in an increasingly digitalized global context.

Based on this background, this study aims to critically examine epistemological biases in digital-based foreign language learning, particularly in the struggle between algorithmic systems and the social reality of language. Scientifically, this study contributes to enriching the study of critical applied linguistics and educational technology studies by providing an epistemological perspective on the use of AI in language learning. Practically, this study is expected to serve as a basis for developing language learning designs that are more sensitive to social, cultural, and linguistic diversity contexts, so that technology no longer reduces the reality of language but instead represents it more fairly and comprehensively.

2) METHODS

This article is the result of a literature study aimed at critically examining the construction of knowledge regarding epistemological bias in digital foreign language learning by comparing

algorithmic systems and the social reality of language through the analysis of various relevant scientific literature sources. The data collection process was carried out systematically by tracing journal articles, academic books, and scientific publications related to the researched theme. This study applies a structured search strategy by setting inclusion and exclusion criteria so that the literature used has relevance, validity, and strong theoretical contribution to the focus of the study (Abdurrahman, 2024) The data collected were then analyzed through the stages of reduction, categorization, and thematic synthesis to identify patterns, tendencies, and conceptual relationships among previous research findings. This approach allows researchers to build a comprehensive analytical framework. In addition, this method also emphasizes transparency and systematicity in the literature review process in order to produce interpretations that can be academically accountable (Qomaruddin & Sa'diyah, 2024).

3) RESULTS

Dominance of Global Linguistic Algorithms and Reduction of Sociocultural Complexity of Language

The dominance of global linguistic algorithms in AI-based language learning systems shows that the construction of linguistic knowledge today is heavily determined by big data-driven computational logic. Modern language models are trained using massive global corpora, which inherently emphasize frequency, probability, and pattern regularity. This approach causes language to be represented as a statistical entity, rather than as a living social practice (Desi Purnama Sari, et. al., 2025). In this context, language loses its complex historical, cultural, and interactional dimensions. Such data-based representations tend to abstract language from the social reality that gives rise to it As a result, language is treated as a neutral system, whereas in reality it is full of values and ideology. Dependence on big data also causes the dominance of global languages that have abundant data sources. This creates an imbalance in linguistic representation in digital learning systems (Khalil Jibran Abbasi, Wazir Ali Tunio, Zaheer Abbas Buriro, & Muhammad Yaseen Memon, 2026). Thus, the dominance of algorithms is not only technical, but also epistemological.

This phenomenon becomes even stronger when we see that algorithms do not merely represent language, but also shape the way language is understood. In digital learning systems, language is processed through mathematical models that prioritize consistency and predictability. This leads to the standardization of language forms considered "correct" based on their frequency of occurrence in corpora. Linguistic variations that are local, contextual, or minority are often marginalized. In fact, from a sociolinguistic perspective, these variations are actually the core of language dynamics. When algorithms ignore these variations, a reduction in language complexity occurs. This reduction is not only linguistic but also social and cultural. In other words, algorithms create a simplified language reality. This reality is then reproduced in digital language learning practices.

Furthermore, epistemological bias in this system arises from the assumption that data is an objective representation of reality. In fact, the linguistic data used in model training contains historical and social biases. Research shows that language models tend to reproduce the biases contained in their training data. These biases can be in the form of gender, racial stereotypes, or a preference for dominant languages. When these biases are not critically examined, they become part of the knowledge structures produced by AI. In the context of language learning, this affects how users understand language and culture. Language is no longer understood as a dynamic social practice, but as a fixed and universal system (Pingkan Verginia Monding, 2025). This indicates that bias in AI is not just a technical issue, but also an epistemological one. Therefore, it is important to view data as a social construction, not merely as a fact.

The dominance of global corpora also leads to the marginalization of local languages and minority dialects. Digital foreign language learning systems such as AI tend to be more accurate in processing languages that have large data representations, such as English. In contrast, languages with limited resources often experience misrepresentation. This creates an imbalance in access to language technology. Language models show performance disparities against non-standard dialects. As a result, speakers of minority languages have to adjust to the language standards accommodated by the systems. This phenomenon can be referred to as a form of 'digital colonization' in the linguistic realm. Local languages lose their authority in the

digital space. Thus, the dominance of algorithms contributes to the homogenization of global languages (Sarthak Sharma, 2023).

The reduction of language complexity is also evident in the loss of pragmatic dimensions in digital foreign language learning systems such as AI. Language in real life is not only composed of grammatical structures, but also meanings that depend on context. However, statistical-based language models tend to struggle to capture complex social contexts. They focus more on the relationships between words rather than the meanings generated in interactions. This leads to limitations in understanding irony, humor, or implicature. In language learning, pragmatic aspects are very important for communicative competence. When these aspects are ignored, learning becomes non-contextual. Language is learned as a formal system, not as a social communication tool. This indicates that algorithms are not yet capable of capturing the full reality of language (Eric Zhao, 2026)

In addition, the algorithmic structure in AI also reinforces bias through the system design process itself. Bias does not only come from data, but also from the way the model is developed and evaluated. Bias can arise from various stages, including data selection, annotation, and model design. This shows that bias is systemic, not incidental. In the context of language learning, this bias can affect the materials, evaluations, and feedback given to users. Digital foreign language learning systems like AI can unconsciously promote certain forms of language as the standard (Esa Peggy Nerida Manurung, Jesika Lolo Karina Manik, Gracesela Septiana Sibuea, & M.Surip, 2025). As a result, users internalize these standards without realizing there are alternatives. This reinforces the dominance of global languages in learning practices. Thus, algorithmic bias becomes part of the digital education structure.

Furthermore, the reduction of language to statistical patterns also impacts the loss of linguistic creativity. Human language is productive and innovative, allowing the creation of new meanings in various contexts. However, AI models tend to reproduce patterns that already exist in the data. This leads to limitations in generating authentic language variation. In language learning, creativity is an important aspect of language mastery. When a system only offers standardized patterns, the space for language exploration becomes limited. Users tend

to follow the patterns considered "correct" by the system. This creates homogenization of linguistic expression. Thus, algorithms not only represent language but also restrict its potential (Khalil Jibran Abbasi, Wazir Ali Tunio, Zaheer Abbas Buriro, & Muhammad Yaseen Memon, 2026)

This phenomenon is also related to what is called 'epistemic injustice' in language technology. When digital foreign language learning systems like AI are unable to represent certain linguistic experiences, speakers of that language experience epistemic injustice. They do not have space to express their linguistic identity authentically. Studies show that biases in language technology can result in inaccurate representations of certain groups. In the context of education, this impacts the exclusion of local knowledge. Language as a medium of knowledge becomes reduced in a narrow form. This shows that technology is not neutral, but carries social and political implications. Therefore, it is important to critique the epistemological structures in digital foreign language learning, such as those using AI.

The dominance of algorithms also creates an illusion of objectivity in digital language learning. Because it is based on data and mathematics, digital foreign language learning systems like AI are often considered neutral and value-free. However, as has been explained, data and algorithms contain unseen biases (Alexandre Puttick, et. al., 2025). This illusion is dangerous because it leads users to accept system results without critique. In the context of education, this can hinder the development of critical thinking. Users tend to rely on the system as the sole source of truth. In reality, language is a complex and contextual phenomenon (Pingkan Verginia Monding, 2025). When this complexity is ignored, learning becomes superficial. Therefore, it is important to integrate a critical approach in the use of AI.

The dominance of global linguistic algorithms has broad implications for language understanding. Language is no longer seen as a dynamic social practice, but as a standardized system. This reduction impacts the loss of cultural, pragmatic, and identity dimensions in language. The epistemological bias that arises is not only due to data but also to the underlying algorithmic structures. Therefore, an alternative approach that is more contextual and inclusive in language technology development is needed. The integration of sociolinguistic perspectives

becomes important to address these limitations. In addition, there needs to be critical awareness in the use of AI in language learning. Thus, technology can be used more fairly and reflectively (Nuryani, Siti Isnaniah, & Ixsir Eliya, 2021).

The dominance of global linguistic algorithms not only changes how language is processed, but also how language is understood. The reduction of language complexity into statistical patterns creates significant epistemological biases. These biases affect language representation, learning practices, and user experience. Therefore, it is important to develop digital foreign language learning systems like AI that are more sensitive to social and cultural contexts (Eric Zhao, 2026). This requires collaboration between technology, linguistics, and education experts. With a multidisciplinary approach, it is hoped that language technology can represent linguistic diversity more fairly. In addition, strict regulation and evaluation of biases in AI are necessary. Thus, the future of digital language learning can be more inclusive and equitable.

Simplification of Linguistic Structure and Marginalization of Local Variation in the Digital Ecosystem

The simplification of linguistic structures in digital language learning ecosystems is a direct consequence of algorithmic design oriented towards computational efficiency. Digital foreign language learning systems such as AI are designed to process language in the form of mathematical representations that can be processed quickly and consistently (Alvina Fridella Anggraini Br Sitepu, et. al., 2025). In this process, linguistic complexity is often reduced to patterns that most commonly appear in the data. As a result, rich and layered language structures are simplified into a more predictable form. This approach ignores the fact that language is not only composed of formal structures, but also meanings that depend on social context. This reduction results in the loss of important semantic nuances in communication. Language is treated as a code system, rather than as a living social practice. This indicates a paradigm shift in understanding language. Thus, linguistic simplification becomes a key feature in the digital ecosystem of AI-based foreign language learning.

In practice, this simplification is seen in how language models prioritize standard forms that are considered most representative. Languages that have high frequency in global corpora are easier for the system to learn. Conversely, linguistic variations that rarely appear tend to be ignored or misunderstood. This leads to the dominance of standard language forms in AI output. Studies show that language models often "default" to standard language forms such as Standard English. This phenomenon indicates that algorithms are not neutral, but rather selective towards certain types of data. This selectivity affects how language is represented in the system. Standard language becomes the norm, while other variations are considered deviations. Thus, a process of linguistic normalization occurs in the digital space.

The marginalization of local variations is a logical consequence of this normalization process. Non-standard dialects and language varieties often do not have adequate representation in training data. As a result, digital foreign language learning systems such as AI experience difficulties in accurately recognizing and processing these variations. There is a disparity in model performance when dealing with non-standard dialects. This indicates that AI capabilities are uneven in handling language diversity. Users who use local dialects often receive less accurate responses (Eric Zhao, 2026). In some cases, the system even fails to understand the conveyed meaning. This indicates the presence of structural bias in language technology. Thus, local variations experience marginalization in the digital ecosystem.

Furthermore, linguistic simplification impacts the reduction of meaning in digital communication. Language not only conveys information but also emotions, identity, and social relationships (Ridha Hasnul Ulya, 2024). However, AI models tend to process language at the surface level, namely the relationships between words. Implicit nuances of meaning are often not captured by the system. This results in communication becoming less rich and contextual. In language learning, this aspect is very important for building communicative competence. When nuances of meaning are ignored, learning becomes mechanistic. Language is learned as a structure, not as a social practice. This highlights the limitations of machine learning-based approaches in fully understanding language. The dominance of standard languages is also related to the history and politics of global language. English, for example,

holds a dominant position in global knowledge production. This is reflected in the data used to train AI models. Most of the linguistic corpora come from English-language sources. As a result, language models are more optimal at processing that language. This condition reinforces the position of English as a digital lingua franca. This dominance also impacts discrimination against non-native speakers. Thus, language technology reproduces global linguistic inequalities. Local languages become less visible in digital spaces.

The limitations of models in recognizing dialects also indicate problems in linguistic representation. Dialects are not merely phonological or lexical variations, but also reflect social identity. However, AI models often fail to capture this dimension. Models have difficulty understanding dialects such as African American English. In fact, model performance can decline significantly when processing these dialects. This indicates that complex linguistic variation is not easily captured by statistical approaches. Models tend to generalize based on dominant patterns. As a result, linguistic diversity is not adequately accommodated. This underscores the epistemological limitations in machine learning (Roni Subhan, Alejandro Diaz, & Karan Singh, 2025) In addition, bias against dialects can also generate stereotypes in AI output. Studies show that responses to non-standard dialects tend to contain stereotypes and negative judgments. This indicates that models not only fail to understand dialects, but also reproduce existing social biases in the data. In the context of education, this is very problematic. Users can receive unfair representations of certain linguistic groups. Language becomes a tool for reproducing social inequality. Thus, language technology is not neutral, but carries ideological implications. Linguistic bias becomes part of the system's structure.

This phenomenon also impacts the learning experience of users within the digital ecosystem. Users who use local variations may feel unrepresented in the system. They have to adapt to the standard language in order to be understood by AI, which serves as a platform for digital foreign language learning (Jaeho Jeon, Seongyong Lee, & Serafín M. Coronel-Molina, 2025). This creates pressure for linguistic standardization. In the long term, this can reduce the use of local dialects. Local languages lose space in digital practices. This indicates a process of linguistic homogenization (Esa Peggy Nerida Manurung, Jesika Lolo Karina Manik,

Gracesela Septiana Sibuea, & M. Surip, 2025). Thus, the digital ecosystem has the potential to erode language diversity (Pingkan Verginia Monding, 2025).

From a pedagogical perspective, linguistic simplification also affects language learning design. Materials produced by AI tend to follow standard structured patterns (Jaeho Jeon, Seongyong Lee, & Serafín M. Coronel-Molina, 2025). This can be helpful for basic learning, but less effective for more complex contexts. Language learning should include variation and context of language use. However, digital foreign language learning systems such as AI often cannot provide this. As a result, learning becomes less contextual and less relevant to social realities (Ulya, 2024). This indicates the need for a more holistic approach in learning design. Integration of sociolinguistic aspects becomes very important. The simplification of linguistic structures and the marginalization of local variations indicate the presence of systemic bias in the digital language ecosystem. Language is reduced to measurable and predictable entities, while its social complexity is ignored. This impacts language representation, comprehension, and learning (Alvina Fridella Anggraini Br Sitepu, et. al., 2025). Therefore, efforts are needed to develop more inclusive technologies. Enriching data with local variations becomes one important solution. In addition, critical awareness in the use of AI is necessary. In this way, language technology can evolve without sacrificing linguistic diversity.

Epistemic Injustice in Language Assessment: The Tension between Algorithmic Standardization and the Dynamic Reality of Language

Epistemic injustice in AI-based language assessment is rooted in the fundamental tension between the logic of algorithmic standardization and the dynamic nature of language. Digital foreign language learning systems such as AI are designed to provide evaluations based on measurable and consistent parameters. However, language as a social practice does not always conform to that logic. It evolves through context, interaction, and the collective experience of its speakers. When algorithms impose certain standards, language diversity becomes reduced. The resulting assessment tends to overlook socially legitimate contextual variations (Khalil Jibran Abbasi, Wazir Ali Tunio, Zaheer Abbas Buriro, & Muhammad Yaseen Memon, 2026). This creates a gap between the reality of language and its representation in

digital systems. In this context, epistemic injustice emerges as a structural consequence. Thus, AI language evaluation is not entirely neutral.

AI-based assessment systems often rely on probabilistic models that evaluate the "correctness" of language based on frequency in data. This approach assumes that the most common form of language is the most correct. In reality, however, social language variations have different legitimacies (Ridha Hasnul Ulya, 2024). When the system ignores these variations, bias in assessment occurs. Learners from certain linguistic backgrounds can be considered "less accurate" simply because they do not conform to dominant standards. This indicates that standards in AI are hegemonic. Assessment becomes a tool for reproducing certain linguistic norms. Thus, algorithms function as mechanisms of language normalization.

Bias in Natural Language Processing (NLP) systems is also closely related to power relations in language. Studies show that language technology cannot be separated from the social structures that produce it. In many cases, dominant languages reflect the social groups that have access to data production. As a result, other groups become underrepresented in the system (Baiq Iling Kiranawati, Umi Latifah, Noffa Kharimatul Khusnah, 2025). When AI is used to evaluate language, this inequality is also reproduced. Evaluations not only reflect linguistic ability but also the social position of the user. This emphasizes that language is an arena of power. Thus, bias in the use of AI in digital foreign language learning is a reflection of broader social inequalities.

Epistemic injustice is also evident in how digital foreign language learning systems such as AI provide feedback to learners. The feedback is often normative and does not take the user's context into account. For example, the use of local dialects can be considered a mistake. In fact, in certain communities, these dialects are legitimate forms of communication. When the system does not recognize this legitimacy, local knowledge is delegitimized. Learners are forced to conform to standards that are not always relevant. This creates unfair linguistic pressure. Thus, AI feedback can reinforce the dominance of the standard language. Furthermore, bias in NLP can reinforce social stereotypes that already exist in society. Training data often contains stereotypes that are then reproduced by the model. These stereotypes can

appear in the form of word associations, classifications, or system responses (Tolga Balukbasi, et. al. 2016) In the context of language learning, this is very problematic. Learners can receive inaccurate representations of certain groups. Language becomes a medium for reproducing social biases. This shows that technology is not ideologically neutral. Therefore, it is important to critically examine data sources in the development of digital foreign language learning using AI.

Tension between standardization and language dynamics is also evident in the limitations of models in understanding context. Human language heavily depends on situation, social relations, and communication goals. However, AI models tend to process language decontextually (Roni Subhan, Alejandro Diaz, & Karan Singh, 2025). This leads to errors in interpreting meaning. In language assessment, context is crucial in determining whether a form is considered correct or not. When context is ignored, the assessment becomes inaccurate. This shows that algorithmic approaches have epistemological limitations. Therefore, a more contextual approach is needed (Ridha Hasnul Ulya, 2024) Epistemic injustice is also related to the concept of "quality-of-service harm" in NLP. Studies show that digital foreign language learning systems such as AI often provide different services to different groups. For example, system accuracy may be lower for certain dialects (Talukder Hasnat Zadid, et. al., 2026). This affects the user experience in language learning. Learners from certain groups may receive less accurate feedback. This creates inequality in access to quality learning. Thus, bias is not only representational but also operational. Its impact is tangible in educational practice.

In addition, bias in language assessment can also affect learners' linguistic identity. Language is an important part of the identity of individuals and communities. When digital foreign language learning systems such as AI evaluate certain language forms as "wrong," that identity is also degraded (Desi Purnama Sari, et. al., 2025). Learners may experience pressure to abandon their language variations. This can lead to the erosion of linguistic identity. In the long term, this impacts cultural diversity. Local languages become less used in digital spaces. Thus, epistemic injustice has significant cultural implications. This phenomenon underscores the need for epistemological reconstruction in the design of digital language learning. An

approach that relies solely on data and algorithms is not sufficient to capture the complexity of language. It requires the integration of sociolinguistic and humanistic perspectives in system development (Ridha Hasnul Ulya, 2024). This includes recognizing language diversity and user contexts. In addition, there needs to be community involvement in the design process. This approach can help reduce bias in systems. Thus, technology can become more inclusive. Epistemological reconstruction becomes an urgent necessity.

The tension between algorithmic standardization and the dynamic reality of language shows that language technology is at a crossroads between efficiency and fairness (Kiranawati, Latifah, and Khusnah, 2025). Digital foreign language learning systems such as AI offer convenience and consistency, but they risk overlooking linguistic diversity. The epistemic injustice that arises is not only a technical issue, but also a social and philosophical one. Therefore, the development of language technology must consider ethical and epistemological dimensions. System evaluation should be based not only on accuracy but also on fairness. With a more reflective approach, digital language learning can become more inclusive. This becomes both a challenge and an opportunity for the development of education in the digital era.

4) DISCUSSION

The dominance of global linguistic algorithms in AI-based digital foreign language learning indicates a fundamental shift in the construction of linguistic knowledge from being social in nature to computational. Modern language models are built on large data corpora that emphasize frequency and probability, so language is reduced to predictable statistical patterns. Within this framework, language is no longer understood as a living social practice, but as a mathematical entity detached from its historical and cultural context (Alvina Fridella Anggraini Br Sitepu, et. al., 2025) These findings are in line with studies showing that NLP models heavily rely on training data as the main source of bias and language representation. Dependence on global data causes dominant languages to obtain greater representation compared to minority languages (Esa Peggy Nerida Manurung, Jesika Lolo Karina Manik, Gracesela Septiana Sibuea, & M. Surip, 2025). As a result, there is an epistemological imbalance in how languages are understood and taught. Languages rich in social context are instead reduced to a uniform form.

Thus, the dominance of algorithms not only has technical impacts but also shapes the way languages themselves are perceived.

This phenomenon becomes even more problematic when algorithms not only represent language but also standardize certain linguistic forms as the norm. Digital foreign language learning systems like AI tend to consider the most frequently occurring language forms in the data as the most correct forms (Alexandre Puttick, et. al., 2025). This leads to the marginalization of local linguistic variations that are actually central to sociolinguistic studies. Dialectal variations, registers, and language usage contexts are often ignored because they do not meet the frequency logic in the corpus. In reality, language in social contexts is always dynamic and contextual. When these variations are eliminated, the complexity of language is also significantly reduced. This reduction impacts not only the linguistic structure but also the social dimension and the identity of speakers (Roni Subhan, Alejandro Diaz, & Karan Singh, 2025) In other words, the algorithm creates a simplified and standardized linguistic reality. This reality is then reproduced in the digital foreign language learning ecosystem without adequate criticism.

Epistemological bias in digital foreign language learning systems such as AI arises from the mistaken assumption that data is an objective representation of reality. In fact, linguistic data contains historical, social, and cultural biases that accumulate in the process of its production (Jaeho Jeon, Seongyong Lee, & Serafín M. Coronel-Molina, 2025). Language models can reproduce social stereotypes and discrimination present in the training data. This bias can take the form of gender, racial stereotypes, or preferences toward certain linguistic groups. When such biases go unrecognized, they become part of the knowledge structure that is considered neutral. In the context of language learning, this affects the way users understand linguistic norms. Language is no longer understood as a diverse social practice, but rather as a homogeneous universal system. This reinforces the dominance of certain perspectives in knowledge production. Thus, bias in AI is epistemological because it influences the way of knowing and understanding language (Desi Purnama Sari, et. al., 2025)

The simplification of linguistic structures in the digital ecosystem also has a direct impact on the marginalization of local variations. AI models are more optimal at processing languages that have large amounts of data, such as English, while other languages face limitations in representation. NLP research is still very focused on English, so global linguistic diversity is less accommodated (Talukder Hasnat Zadid, et. al., 2026). This situation causes digital foreign language learning systems such as AI to be less accurate in understanding dialects and minority languages (Baiq Iling Kiranawati, Umi Latifah, Noffa Kharimatul Khusnah, 2025) As a result, local language speakers often have to adapt to global language standards in order to be understood by systems. This phenomenon creates pressure to carry out linguistic standardization in the digital space. In the long term, this can erode language diversity and local identity. Local languages lose their authority in the digital ecosystem dominated by algorithms. Thus, linguistic simplification contributes to the global homogenization of languages (Esa Peggy Nerida Manurung, Jesika Lolo Karina Manik, Gracesela Septiana Sibuea, & M. Surip, 2025)

Ultimately, epistemic injustice in AI-based foreign language assessment becomes a consequence of the tension between algorithmic standardization and the dynamic reality of language. Digital foreign language learning systems like AI assess language based on measurable parameters, while language in social practice cannot always be reduced to these metrics. This results in assessments that tend to be biased toward certain forms of language and overlook the user's context (Alexandre Puttick, et. al., 2025) Bias in digital foreign language learning can result in service inequities and stereotypes against certain groups. In the educational context, this affects unequal learning experiences. Learners from certain linguistic backgrounds can be disadvantaged because they do not conform to algorithmic standards. Thus, AI language evaluation is not entirely neutral, but rather laden with social and ideological implications. Therefore, a more critical, contextual, and inclusive approach is needed in the development of language learning technology in the digital era.

5) CONCLUSION

The dominance of linguistic algorithms in digital foreign language learning has shifted the understanding of language from a complex social practice to a computational representation based on probability and frequency, thus reducing language to statistical patterns detached from its cultural and historical context. The reliance of NLP models on large-scale training data not only shapes skewed language representations—with the dominance of global languages and the marginalization of local languages—but also reproduces the social and linguistic biases contained in that data. In this process, algorithms do not merely represent language but also standardize certain forms as norms, which has implications for the simplification of linguistic structures, the elimination of dialectal variation, and the reinforcement of global language homogenization. The impact is epistemological, as it affects the way language is understood, taught, and assessed, while also creating inequities in the learning experience for speakers from diverse linguistic backgrounds. Therefore, a critical and contextual approach is needed in the development of language learning technology in order to accommodate the social complexity of language and prevent the reproduction of bias in the digital ecosystem.

REFERENCES

Abbasi, K. J., Tunio, W.A., Buriro, Z. A. & Memon, M. Y., (2026). AI-Mediated Language Learning and Digital Colonialism: A Critical Applied Linguistics (CAL) Analysis of Algorithmic Bias in English Language Teaching Technologies in Pakistan. Journal of Applied Linguistics and TESOL, 9 (1), 232-240. https://doi.org/10.63878/jalt1798

Abdurrahman, (2024). Metode Penelitian Kepustakaan dalam Pendidikan Islam. Adabuma: Jurnal Pendidikan dan Pemikiran, 3 (2), 102-113. https://doi.org/10.38073/adabuna.v3i2.1563

Balukbasi, T., Chang, K., Zon, J., Saligrama, V., & Kalai, A., (2016). Quantifying and Reducing Stereotypes in Word Embeddings. Workshop Article in Workshop on Machine Learning in Social Good Applications. 41-45. https://doi.org/10.48550/arXiv.1606.06121

He, M., Abbasi, B. N., & He, J., (2025). AI-Driven Language Learning in Higher Education: An Empirical Study on Self-Reflection, Creativity, Anxiety, and Emotional Resilience in EFL Learners. Humanities and Social Sciences Communications, 12 (1525), 1-20. https://doi.org/10.1057/s41599-025-05817-5

Jeon, J., Lee, S., & Coronel-Molina, S. M., (2025). Rethinking AI: Bias in Speech Recognition Chatbots for ELT. ELT Journal, 78 (4). 435-445. https://doi.org/10.1093/elt/ccae035

Kiranawati, B.I., Latifah, U., & Khusnah, N. K., (2025). Pemanfaatan Model AI dalam Analisis Sintaksis Bahasa Indonesia: Kajian Linguistik Komputasional. Jurnal Progresif, 3 (1), 7-26. https://journal.univgresik.ac.id/index.php/progresif/article/view/443

Manurung, E.P.N., Manik, J.L.K., Sibuea, G.S. & Surip, M., (2025). Dinamika Bahasa Indonesia di Era Digital: Antara Perkembangan Linguistik dan Tantangan Preservasi. Morfologi: Jurnal Ilmu Pendidikan, Bahasa, Sastra dan Budaya, 3 (2), 248-257. https://doi.org/10.61132/morfologi.v3i2.1551

Monding, P. V., (2025). Pemanfaatan Kecerdasan Buatan (AI) dalam Linguistik Terapan: Antara Efisiensi, Tantangan, dan Pelestarian Bahasa. Kajian Linguistik, 10 (2), 16-19. https://doi.org/10.35796/kl.10.2.2025.62474

Nuryani, Isnaniah, S., & Eliya,, I., (2021). Sosiolinguistik dalam Pengajaran Bahasa Berbasis Multikultural: Teori dan Praktik Penelitian. Bogor: In Media.

Puttick, A., Ikae, C., Rigotti, C., Fosch-Villaronga, E., Kharas, M.W., Søraa, R.A., & Kurpicz-Briki, M. (2025). A Systematic Review of Bias Detection Methods for Non-English Word Embeddings and Language Models. Artificial Intelligence Review, 58 (370), 1-56. https://doi.org/10.1007/s10462-025-11375-8

Qomaruddin & Sa'diyah, H., (2024). Kajian Teoritis tentang Teknik Analisis Data dalam Penelitian Kualitatif: Perspektif Spradley, Miles dan Huberman. Journal of Management, Accounting, and Administration, 1 (2), 77–84. https://doi.org/10.52620/jomaa.v1i2.93

Sari, D.P., Syihabbudin, Zifana, M., Lestaria, E., Iskandar, & Lushinta, I.P., (2025). Dari 'Cogito' ke 'Algorithm': Pergeseran Paradigma Epistemologis Mahasiswa di Era AI. Jurnal Pendidikan Tambusai, 9 (3), 41318-41329. https://doi.org/10.31004/jptam.v9i3.35544

Sharma, S., (2023), Natural Language Processing in Education: Enhancing Learning Through AI. International Machine Learning Journal and Computer Engineering, 6(6). https://mljce.in/index.php/Imljce/article/view/63

Sitepu, A. F. A.B., Sihotang, Y. N.O., Syaharani, R. S., Sinaga, H., Priskila G, Y., & Tansliova, L., (2025). Membangun Literasi Digital Pembelajaran Bahasa Indonesia: Strategi dan Tantangan di Era AI. Jurnal Inovasi Penelitian Ilmu Pendidikan Indonesia, 2 (1), 30-35. https://jipipi.org/index.php/jipipi

Subhan, R., Diaz, A., & Singh, K., (2025). The Ethics of AI in Language Assessment: A Critical Examination of Algorithmic Bias in Automated Speaking and Writing Tests. International Journal of Language and Ubiquitous Learning, 3 (6), 307–317.https://doi.org/10.70177/ijlul.v3i6.2838

Susilo, A. & Firdaus, M., (2025). A Review of Artificial Intelligence in Language Assessment: Trends, Challenges, and Opportunities. International Journal of Research in Education, 5 (2), 437-446. https://doi.org/10.26877/ijre.v5i2.1995

Ulya, R. H., (2024). Transformasi Makrolinguistik Bahasa Indonesia dalam Gamitan Media Digital: Analisis Wacana Kritis pada Platform Media Sosial. Jurnal Ilmiah Langue and Parole, 8 (1), 91-99. https://doi.org/10.36057/jilp.v8i1.717

Zadid, T. H., et. al., (2026). A Review of Fairness Challenges in Natural Language Processing. Array, 30 (100773), 1-18. https://doi.org/10.1016/j.array.2026.100773

Zhao, E., (2026) Linguistic and Dialectal Bias in Large Language Models: A Review of Evidence, Mechanisms, and Implications (2018–2025). Advances in Humanities Research, 13 (1). 73-80. DOI:10.54254/2753-7080/2026.31302