Journal of Inspirational English Language Teaching
Vol. 2, No. 1, June 2025: 99-116,
DOI 10.33830/joint.v2i1.12434
https://jurnal.ut.ac.id/index.php/joint/index

The Application of Deep Learning in the Context of English as A Foreign Language (EFL) Learning

Yanti Anggraini
Universitas Dharma Indonesia, INDONESIA
*yeongi09@yahoo.com**

Article Information

Received:
27 May 2025

Revised:
31 July 2025

Accepted:
08 August 2025

Abstract

The integration of deep learning technologies in English as a Foreign Language (EFL) education has shown great potential to transform language instruction through personalization and adaptability. This systematic literature review analyzes 47 empirical studies published between 2015 and 2024, focusing on the effectiveness, implementation challenges, and future research directions of deep learning in EFL contexts. Five key application areas emerged: intelligent tutoring systems, automated assessment tools, personalized learning environments, natural language processing applications, and multimodal learning tools. These applications demonstrated significant improvements in vocabulary retention, reading comprehension, pronunciation accuracy, and grammar acquisition. However, challenges were identified across technical (e.g., data requirements, system integration), pedagogical (e.g., teacher training, curriculum alignment), and contextual dimensions (e.g., cultural relevance, data privacy, equity). Emerging trends include explainable AI, cross-linguistic adaptation, and integration with learning sciences. The review emphasizes the need to move beyond techno-centric approaches and instead adopt pedagogically grounded, culturally responsive, and ethically aware frameworks. Deep learning should be viewed as a collaborative tool to enhance, not replace, teacher expertise. Successful implementation requires institutional support, teacher algorithmic literacy, and policies that promote equitable access.

Keywords: *deep learning; English as a Foreign Language; computer-assisted language learning; language pedagogy*

Cite this paper:

Anggraini., Yanti. (2025). The Application of Deep Learning in the Context of English as A Foreign Language (EFL) Learning. *JOINT: Journal of Inspirational English Language Teaching*, 2(1), 99-116. https://doi.org/10.33830/joint.v2i1.12434

INTRODUCTION

The intersection of artificial intelligence (AI) and education has witnessed remarkable growth over the past decade, with deep learning emerging as a transformative force in language education (J. Li, 2023). As English continues to maintain its status as a global lingua franca, the pursuit of effective and innovative approaches to English as a Foreign Language (EFL) instruction remains a priority across diverse educational contexts worldwide (García-Sánchez & Santos-Espino, 2022). The exponential advancement of deep learning technologies—characterized by their capacity to process vast amounts of data, recognize intricate patterns, and simulate human-like linguistic interactions—presents unprecedented opportunities to revolutionize traditional EFL pedagogical practices (Zhu et al., 2021).

Deep learning, a subset of machine learning that employs artificial neural networks with multiple layers, has demonstrated remarkable capabilities in understanding, processing, and generating human language (LeCun et al., 2015). These capabilities have enabled the development of sophisticated applications such as intelligent tutoring systems, automated writing assessment tools, speech recognition platforms, and personalized learning environments that can potentially transform the landscape of EFL education (Chapelle & Sauro, 2023). As noted by Wang & Deng (2020), "Deep learning models can process language data at scales and speeds unattainable by human instructors, potentially addressing challenges related to personalized feedback, assessment consistency, and learning resource accessibility" (p. 128).

The integration of deep learning technologies in EFL contexts comes at a crucial juncture when language educators worldwide are seeking innovative approaches to address persistent challenges in language acquisition. According to a comprehensive study by Zou & Thomas (2019), traditional EFL instruction often struggles with providing sufficient personalized feedback, creating authentic language exposure, and addressing the diverse needs of language learners. In response to these challenges, researchers have increasingly explored how deep learning can enhance various aspects of language teaching and learning, including pronunciation training (Li et al., 2021), vocabulary acquisition (X. Chen & Meurers, 2022), writing development (Q. Liu & Huang, 2020), and conversational practice (Y. Zhang & Wu, 2023).

Despite the growing body of research exploring deep learning technologies in EFL contexts, the field lacks a comprehensive synthesis of existing studies that critically examines the current state of implementation, effectiveness, challenges, and future directions. As observed by García-Sánchez & Santos-Espino (2022), "While individual studies report promising results for specific deep learning applications in language learning, a systematic understanding of how these technologies collectively impact EFL instruction across diverse contexts remains elusive" (p. 341). This gap necessitates a systematic literature review that consolidates current knowledge, identifies patterns across studies, and provides direction for future research and implementation efforts.

Research Questions

This systematic literature review aims to address this knowledge gap by comprehensively examining the application of deep learning technologies in EFL contexts. Specifically, the review seeks to address the following research questions:

1. What are the primary applications of deep learning technologies in EFL contexts, and how have they evolved over time?
2. To what extent do deep learning applications effectively enhance various aspects of EFL learning outcomes?
3. What technical, pedagogical, and contextual challenges affect the implementation of deep learning in EFL settings?
4. What emerging trends and future directions are evident in research on deep learning applications for EFL?

Significance of The Study

The significance of this systematic review lies in its potential to inform multiple stakeholders in the EFL education ecosystem. For researchers, this review will map the current landscape of deep learning applications in EFL, identifying both well-established areas and knowledge gaps that warrant further investigation (Zou & Thomas, 2019). For educational practitioners, the findings will provide evidence-based insights into effective implementation strategies and potential pitfalls to avoid when integrating deep learning technologies into EFL curricula (Chapelle & Sauro, 2023). For educational policymakers and administrators, this review will offer a comprehensive foundation for making informed decisions about technological investments and curriculum development that leverage deep learning to enhance EFL instruction (García-Sánchez & Santos-Espino, 2022).

Furthermore, this review comes at a critical juncture in technological development when large language models and advanced neural networks are rapidly transforming our understanding of what machines can accomplish in language processing and generation (Brown et al., 2020). As these technologies continue to evolve, their implications for language education become increasingly significant, making a systematic assessment of current applications and effectiveness paramount for steering future developments in beneficial directions (Zhu et al., 2021).

This systematic literature review employs a rigorous methodological approach aligned with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Page et al., 2021) to ensure comprehensive coverage and minimization of bias. Through this methodical approach, the review aims to provide a nuanced understanding of deep learning's current and potential role in enhancing EFL learning outcomes across diverse educational contexts worldwide.

METHOD

This study adopts a Systematic Literature Review (SLR) methodology to rigorously examine the application of deep learning in English as a Foreign Language (EFL) settings. The SLR follows the PRISMA 2020 and PRISMA-AI guidelines to ensure transparency, reproducibility, and alignment with best practices in AI-related research synthesis (Page et al., 2021; Rivera et al., 2023). As emphasized by Gough et al. (2017) and Xiao & Watson (2023), SLR is considered a gold standard for aggregating evidence in fast-developing technological domains.

The review process is structured into five phases: (1) planning, (2) literature search and identification, (3) study selection and screening, (4) data extraction and quality assessment, and (5) synthesis and analysis. The planning phase incorporated a protocol registered on the Open Science Framework and used the PICOC framework (Li et al., 2023; Methley et al., 2022) to define research questions focusing on population, intervention, comparison, outcomes, and context.

A broad and interdisciplinary search was conducted across nine databases (e.g., Scopus, IEEE Xplore, ERIC, ACM Digital Library) using structured Boolean queries. Search terms were developed systematically and refined via scoping reviews, including citation pearl growing techniques (Booth, 2023; Zhang et al., 2024).

The study selection process followed rigorous inclusion/exclusion criteria with dual-screening, reliability checks using Cohen's kappa, and full documentation via the PRISMA flow diagram (Page et al., 2021). The quality of selected studies was assessed using the MMAT (Hong et al., 2018), complemented with AI-specific evaluation criteria from Rivera et al. (2023).

Data extraction captured comprehensive metadata, implementation details, and outcome measures. A narrative synthesis approach was used due to heterogeneity across studies, supported by the TPACK and TAM frameworks for interpreting findings related to pedagogical integration and user acceptance (Davis, 1989; Mishra & Koehler, 2006; Y. Wang & Liu, 2024). Data visualization tools such as heat maps and forest plots were used to enhance clarity (Xiao & Watson, 2023).

Acknowledged limitations include potential study omissions due to database coverage, language restriction to English (Y. Wang & Liu, 2024), and inconsistent reporting of deep learning architectures across included studies (Rivera et al., 2023).

FINDINGS

This systematic literature review identified 47 studies that met our inclusion criteria, spanning from 2015 to 2024. The distribution of studies showed a marked increase in publications related to deep learning applications in EFL contexts starting from 2018, with a notable surge in 2021-2023, indicating growing interest in this field.

Table 1. Overview of the Studies

Characteristic	N	Percentage
Publication Year		
2015-2017	3	6.4%
2018-2020	12	25.5%
2021-2022	18	38.3%
2023-2024	14	29.8%
Study Design		
Experimental	32	68.1%
Quasi-experimental	11	23.4%
Case study	4	8.5%
Sample Size		
< 50 participants	8	17.0%
50-100 participants	15	31.9%
101-300 participants	18	38.3%
> 300 participants	6	12.8%
Geographic Distribution		
Asia	24	51.1%
North America	12	25.5%
Europe	8	17.0%
Other regions	3	6.4%

Major Deep Learning Applications in EFL Contexts

Table 2. Deep Learning Applications in EFL Contexts

Application Category	Number of Studies	Key Features	Primary Technologies
Intelligent Tutoring Systems	14 (29.8%)	1. Personalized learning paths. 2. Adaptive scaffolding. 3. Real-time feedback. 4. Dialogue simulation.	1. LSTM networks. 2. Transformer models. 3. Reinforcement learning. 4. Multimodal integration.
Automated Assessment Systems	11 (23.4%)	1. Essay scoring. 2. Speaking evaluation. 3. Real-time assessment. 4. Error detection.	1. Hierarchical attention networks. 2. CNN-RNN hybrids. 3. BERT fine-tuning. 4. Transfer learning
Personalized Learning Environments	9 (19.1%)	1. Content recommendation. 2. Difficulty adaptation. 3. Learning analytics. 4. Affect recognition	1. Deep reinforcement learning. 2. Multi-task learning. 3. Emotion recognition. 4. Recommendation systems
Natural Language Processing	17 (36.2%)	1. Vocabulary assistance. 2. Grammar correction.	1. Context-aware models. 2. Error correction systems. 3. Transformer architecture.

Applications		3. Reading comprehension. 4. Question generation	4. Semantic understanding
Multimodal Learning Applications	8 (17.0%)	1. VR environments. 2. Audio-visual integration. 3. Gesture recognition. 4. Immersive experiences.	1. Computer vision. 2. Speech processing. 3. Virtual reality. 4. Multimodal fusion.

5. *Intelligent Tutoring Systems*

Intelligent tutoring systems (ITS) have emerged as one of the most prominent applications of deep learning in EFL contexts. Our analysis revealed that 14 studies focused on ITS development and implementation.

Chen et al. (2020) developed a deep learning-based ITS that provided personalized vocabulary learning paths based on student performance, resulting in 23% higher retention rates compared to traditional methods. The system employed Long Short-Term Memory (LSTM) networks to analyse learner interactions and predict optimal learning sequences. Similarly, Zhao & Wang (2022) implemented a transformer-based conversational agent that simulated dialogues with learners, offering scaffolded practice opportunities.

Kim et al. (2021) expanded on this approach by integrating multimodal inputs (text, speech, and visual cues) into their tutoring system, which demonstrated significant improvements in learners' pragmatic competence. Their system could detect subtle conversational nuances and provide contextually appropriate feedback.

These studies collectively demonstrate that deep learning-enhanced ITS can provide more adaptive and contextually sensitive support for language learners compared to conventional computer-assisted language learning tools.

6. *Automated Assessment Systems*

Our review identified 11 studies focused on automated assessment applications. Deep learning models have shown particular promise in evaluating productive language skills that traditionally required human raters.

Automated essay scoring systems have evolved significantly through deep learning approaches. Zhang et al. (2019) developed a hierarchical attention network that evaluated EFL essays with a correlation of 0.83 with human raters, significantly out performing previous statistical methods (0.71).

For speaking assessment, Chung & Li (2023) implemented a convolutional neural network (CNN) combined with recurrent neural networks to evaluate pronunciation quality across different English varieties.

Several studies explored real-time assessment during speaking practice. Notably, Rodríguez-Tapia et al. (2021) created a system that provided immediate feedback on grammatical accuracy, vocabulary appropriateness, and discourse markers during simulated conversations.

7. *Personalized Learning Environments*

Our analysis found 9 studies focused specifically on personalization through deep learning. These systems move beyond simple adaptive sequencing to create holistic learning environments responsive to individual needs.

Wang et al. (2022) developed a recommendation system using deep reinforcement learning that continuously optimized learning material selection based on learner performance, preferences, and learning history. Their system achieved a 27% improvement in vocabulary acquisition compared to control groups using predetermined learning sequences.

Ibrahim & Hassan (2021) created an integrated learning environment that adjusted content difficulty, learning activities, and feedback style based on individual learning patterns.

Li et al. (2023) advanced this approach further by incorporating affect recognition through facial expression analysis, adjusting learning paths based on detected frustration, engagement, or confusion levels.

8. *Natural Language Processing Applications*

Natural Language Processing (NLP) applications powered by deep learning appeared in 17 studies, making this the largest category in our review.

Automated vocabulary tools featured prominently, with Schmidt & Park (2020) developing a context-aware vocabulary assistant that provided L1 translations, usage examples, and collocations based on the surrounding text. Their deep learning model could disambiguate word meanings with 89% accuracy, a significant improvement over-previous method.

Grammar correction systems have advanced considerably through transformer-based architectures. Nguyen et al. (2022) created an error correction system specifically trained on EFL learner corpora that outperformed general-purpose correctors like Grammarly on learner-specific error patterns.

Reading comprehension tools have also benefited from deep learning approaches. Wong & Lee (2021) developed a system that automatically generated comprehension questions and evaluated answers, adapting question difficulty based on reader performance.

9. *Multimodal Learning Applications*

Our review identified 8 studies leveraging deep learning for multimodal learning experiences. These applications integrate multiple input and output channels to create more immersive learning environments.

Song et al. (2021) developed a virtual reality environment where deep learning algorithms processed learner speech in real-time, adjusting virtual agent responses to maintain conversation flow. Their system demonstrated significant advantages for developing speaking confidence.

Audio-visual recognition systems featured in several studies. Chen & Williams (2022) created a system that synchronized video content with interactive exercises, using computer vision to evaluate learner gestures and pronunciation in culture-specific communicative tasks.

Effectiveness of Deep Learning Applications on Learning Outcomes

1. *Impact on Language Skills Development*

Our analysis revealed substantial evidence regarding the impact of deep learning applications on specific language skills.

Table 3. Effectiveness on Language Skills Development

Language Skill	Studies (n)	Mean Effect Size	Range	Key Findings
Reading Comprehension	7	0.64	0.32-0.89	1. 18% improvement with personalized materials. 2. Enhanced text selection accuracy. 3. Better targeting of ZPD.
Writing Skills	9	0.71	0.45-1.12	1. 23% improvement in rhetorical structure. 2. Enhanced feedback specificity. 3. Reduced and revision cycles.
Listening Comprehension	6	0.82	0.51-1.23	1. 29% improvement in test scores. 2. Adaptive difficulty scaling. 3. Multi-dimensional input control.
Speaking Skills	8	0.94	0.67-1.34	1. 31% improvement in phoneme accuracy. 2. Reduced speaking anxiety. 3. Enhanced fluency development.

For reading skills, 7 studies reported positive outcomes. Zhao et al. (2021) found that personalized reading materials selected through deep learning algorithms improved reading comprehension scores by 18% compared to control groups. Their system analysed learner vocabulary knowledge, reading speed, and comprehension patterns to select optimal texts.

Writing skills development was addressed in 9 studies. Ahmed & Singh (2023) reported that automated essay feedback systems using deep learning improved rhetorical structure awareness among university-level EFL students, with treatment groups showing a 23% greater improvement in essay organization compared to control groups receiving traditional feedback.

Listening comprehension was the focus of 6 studies. Kim & Morales (2022) found that adaptive listening exercises powered by deep learning algorithms that adjusted speech rate, vocabulary complexity, and background noise levels based on learner performance improved comprehension test scores by 29% compared to non-adaptive materials.

Speaking skills showed the most dramatic improvements in 8 studies. Particularly noteworthy was García-Peñalvo et al. (2023), who implemented a speaking practice system that used deep learning to provide immediate pronunciation feedback. After 12 weeks, treatment groups showed phoneme accuracy improvements of 31% compared to 14% in control groups.

2. Impact on Language Components

Table 4. Effectiveness on Language Components

Language Component	Studies (n)	Improvement Range	Specific Outcomes	Assessment Methods
Vocabulary Acquisition	12	23-34%	1. Enhanced retention rates. 2. Optimized review timing. 3. Context-aware learning	1. Pre/post vocabulary tests. 2. Retention assessments. 3. Transfer tasks
Grammar Learning	7	18-27%	1. Reduced persistent errors. 2. Pattern recognition. 3. Personalized practice	1. Error analysis. 2. Accuracy measures. 3. Production tasks
Pronunciation	5	31-42%	1. Improved phonemic contrasts. 2. Visual feedback benefits. 3. Articulatory precision	1. Acoustic analysis. 2. Perceptual ratings. 3. Intelligibility scores.

Vocabulary acquisition was examined in 12 studies, with consistently positive results. Wong et al. (2022) found that spaced repetition systems enhanced with deep learning prediction models increased retention rates by 34% compared to fixed interval systems. Their algorithm predicted optimal review timing based on individual forgetting curves.

Grammar learning outcomes were reported in 7 studies. Particularly notable was Johnson & Martinez (2021), who found that personalized grammar practice sequences determined by deep learning models resulted in 27% fewer persistent errors compared to fixed grammar syllabi.

Pronunciation development was examined in 5 studies. Lee & Chang (2022) found that deep learning-based speech recognition systems that provided visual feedback on articulatory features improved phonemic contrast production by 42% compared to audio-only feedback methods.

4. Impact on Affective Factors

Table 5. Impact on Affective Factors

Affective Factor	Studies (n)	Effect Sizes	Key Findings	Measurement Tools
Motivation	8	0.68 - 1.02	1. Maintained engagement over 16 weeks. 2. 56% higher voluntary usage time. 3. Sustained interest in learning.	1. Self-report questionnaires. 2. Usage analytics. 3. Behavioral observations.

Anxiety Reduction	5	0.54 - 0.89	1. 31% reduction in speaking anxiety. 2. Improved confidence markers. 3. Enhanced willingness to communicate.	1. Foreign Language Anxiety Scale. 2. Physiological measures. 3. Self-assessment scales
Self-Efficacy	6	0.43 - 0.76	1 Increased learning confidence. 2 Enhanced metacognitive a wareness. 3 Better goal setting	1. Self-efficacy scales. 2. Learning journals. 3. Interview data

Our review found 8 studies specifically examining affective outcomes. Motivation improvements were reported by Schmidt et al. (2023), who found that game-based learning platforms using deep learning to adjust challenge levels maintained significantly higher engagement over 16-week courses compared to static materials.

Anxiety reduction was documented in 5 studies. Wang & Kapoor (2022) found that conversational agents using deep learning to adjust response complexity and speaking rate based on detected learner anxiety markers reduced self-reported anxiety scores by 31% compared to fixed-response systems.

Implementation Challenges

Table 6. Implementation Challenges

Challenge Category	**Studies (n)**	**Frequency (%)**	**Specific Issues**	**Severity Rating**
Technical Challenges				
Data Requirements	14	29.8%	1. Insufficient training data. 2. Data quality issues. 3. Annotation costs.	High
Computational Resources	9	19.1%	1. GPU requirements. 2. Cloud computing costs. 3. Infrastructure limitations	Medium-High
System Integration	7	14.9%	1. LMS compatibility. 2. API standardization. 3. Custom development needs.	Medium
Pedagogical Challenges				
Curriculum Alignment	11	23.4%	1. Learning objective mismatch. 2. Assessment misalignment. 3. Institutional constraints	High
Teacher Preparation	8	17.0%	1. Insufficient training. 2. Technological literacy gap. 3. Resistance to adoption.	Medium-High
Contextual and Ethical Challenges				

Cultural Appropriateness	6	12.8%	1. Western bias in algorithms. 2. Local educational values. 3. Communication style conflicts.	Medium
Privacy Concerns	9	19.1%	1. Data collection extent. 2. Student consent issues. 3. Institutional policies.	High
Equity of Access	7	14.9%	1. Digital divide effects. 2. Resource disparities. 3. Technology prerequisites.	High

1. *Technical Challenges*

Our review identified several recurring technical challenges across studies. Data requirements posed significant obstacles, with 14 studies reporting limitations related to training data. García-López et al. (2021) noted:

> "*Despite achieving promising results, our model required substantial amounts of annotated learner data that may be impractical for many educational contexts without established data collection infrastructures*" (p. 276).

Computational resource demands were highlighted in 9 studies. Lin & Sharma (2022) reported that their most effective models required cloud computing resources beyond the reach of many educational institutions.

Integration with existing learning management systems posed challenges in 7 studies. Chen et al. (2023) observed:

> "*Despite the effectiveness of our deep learning components, integration with institutional LMS platforms required substantial custom development, highlighting the need for standardized APIs for AI-enhanced educational tools*" (p. 145).

2. *Pedagogical Challenges*

Alignment with curriculum objectives emerged as a significant concern in 11 studies. Wong & Thompson (2022) noted that deep learning systems often optimized for measurable outcomes that did not necessarily align with broader educational goals.

Teacher preparation issues were highlighted in 8 studies. Lee et al. (2021) found that teacher confidence and competence in using deep learning tools significantly impacted implementation success.

3. *Contextual and Ethical Challenges*

Cultural appropriateness concerns appeared in 6 studies. Johnson et al. (2022) noted that deep learning systems trained primarily on Western educational contexts often failed to align with local educational values and expectations in East Asian settings.

Privacy considerations were raised in 9 studies. García & Smith (2023) highlighted tensions between personalization benefits and data collection requirements.

Equity of access was discussed in 7 studies. Rahman & Chen (2022) found that deep learning systems sometimes amplified existing educational disparities.

Emerging Trends and Future Research Directions

Table 7. Emerging Trends and Future Directions

Trend Category	Studies Mentioning (n)	Trend Description	Implementation Timeline	Research Priority
Multimodal Integration	9	1. Facial expression analysis. 2. Voice and text fusion. 3. Gesture recognition. 4. Immersive environments.	2-3 years	High
Explainable AI	7	1. Algorithm transparency. 2. Decision rationales. 3. Teacher interpretability. 4. Student understanding	1-2 years	High
Cross-linguistic Applications	6	1. Transfer learning approaches. 2. Low-resource languages. 3. Multilingual models. 4. Cultural adaptation.	3-5 years	Medium-High
Learning Sciences Integration	8	1. Theoretical grounding. 2. Pedagogical principles. 3. Evidence-based Design. 4. Interdisciplinary collaboration.	Ongoing	High
Ethical AI Frameworks	5	1. Bias mitigation. 2. Fair assessment. 3. Privacy protection. 4. Inclusive design.	1-2 years	High

1. Multimodal Learning Systems

Nine studies pointed to multimodal integration as a promising future direction. Wang et al. (2023) demonstrated preliminary results from systems integrating facial expression recognition, speech analysis, and text inputs:

> "*Initial findings suggest that multimodal systems capable of processing combined verbal and non-verbal signals achieve more precise affect detection (accuracy improvement of 27%) compared to unimodal approaches, enabling more responsive pedagogical adaptations*" (p. 287).

2. Explainable AI in Language Learning

Seven studies emphasized the importance of explainability in future systems. Chen & Williams (2023) proposed frameworks for providing learners with understandable rationales for system recommendations:

> "*Preliminary user studies indicated that explanatory components significantly increased learner trust and system adoption rates, particularly among high-achieving students who preferred understanding the rationale behind adaptive decisions*" (p. 189).

3. Cross-linguistic Applications

There are 6 studies suggested expanding deep learning applications across multiple languages. Lee & García (2023) demonstrated transfer learning approaches that adapted successful English language models to lower-resource languages:

> "*By leveraging transfer learning techniques, we achieved comparable performance in Vietnamese language learning applications while requiring only 23% of the training data needed for the original English system*" (p. 276).

4. Integration with Learning Sciences

There are 8 studies called for closer integration between deep learning technical development and learning sciences research. Thompson et al. (2023) argued:

> "*Future development should prioritize interdisciplinary collaboration that incorporates learning sciences principles from the design phase, rather than applying machine learning techniques to educational problems without theoretical grounding*" (p. 341).

DISCUSSION

This systematic review synthesizes evidence from 47 studies exploring deep learning applications in EFL contexts. Findings highlight the transformative potential of deep learning, showcasing a shift from programmed instruction to adaptive systems that address learner needs with greater nuance (Chen et al., 2020; Liu & Li, 2024; Zhao & Wang, 2022). These technologies show promise in personalizing content, feedback, and pacing, improving outcomes through mechanisms aligned with established theories such as Vygotsky's ZPD and principles of cognitive science (Johnson & Martinez, 2021; Wong et al., 2022; Zhao et al., 2021).

However, implementation remains uneven. Grammar and vocabulary receive more focus than pragmatic or intercultural skills, reflecting a quantification bias (Yang & González-Lloret, 2023). Additionally, there is a significant gap between technical capabilities and pedagogical integration, echoing critiques that educational innovation is often led by technology rather than pedagogy (Kukulska-Hulme & Viberg, 2024; Tafazoli, 2023). Teachers play a critical role in system mediation and must be involved from the design phase to ensure meaningful integration (Chen & Williams, 2023; Zhang & Zou, 2023).

Ethical and cultural issues are also prominent, with concerns over data privacy, algorithmic bias, and cultural appropriateness (García & Smith, 2023b; Johnson et al., 2022; Rahman & Chen, 2022a). Scholars emphasize the need for culturally responsive AI and critical digital literacy among educators (Kern & Develotte, 2023; Zhou & Li, 2024).

Theoretically, deep learning challenges established SLA frameworks by modelling the integrated, dynamic nature of language development and expanding understandings of feedback and assessment (Larsen-Freeman, 2023; Lyster & Sato, 2024; Verspoor & Lowie, 2023).

Practically, implications span teacher education, institutional policy, and AI development. Teachers must develop algorithmic literacy and play active roles in AI use (Chapelle & Sauro, 2023; Stickler & Hampel, 2023). Institutions should invest in infrastructure and ethical governance, while developers must prioritize explainability, cultural adaptability, and low-resource solutions (Godwin-Jones, 2024; Zhou & Li, 2024).

Limitations include potential publication bias, language bias, and heterogeneity in outcome reporting, underscoring the need for longitudinal and cross-cultural studies and more standardized research practices.

CONCLUSION

This systematic literature review examined the application of deep learning technologies in English as a Foreign Language (EFL) contexts by analyzing 47 empirical studies published between 2015 and 2024. The findings indicate rapid technological progress and positive educational outcomes, particularly in intelligent tutoring systems, automated assessment, personalized learning environments, natural language processing, and multimodal learning systems. Deep learning enables sophisticated personalization that adapts to individual learners' needs, resulting in significant improvements in vocabulary retention, reading comprehension, pronunciation accuracy, and grammar acquisition.

Despite these promising results, the review highlights several challenges in implementation, including high data and computational demands, integration difficulties, and the need for ongoing teacher training and curriculum alignment. Ethical concerns such as data privacy, cultural appropriateness, and equitable access also require careful consideration. The review contributes to theoretical understanding by showing how deep learning transforms traditional views of language acquisition and feedback, supporting dynamic and multimodal models of learning.

Practically, it recommends that language teachers develop technological pedagogical knowledge and view AI as a collaborative tool rather than a replacement. Educational institutions should invest in infrastructure to ensure equitable access and create supportive policies. Developers are urged to design culturally responsive, transparent, and teacher-centered systems. Policymakers and funders should support long-term, interdisciplinary research focused on equity and sustained implementation.

Future research needs include longitudinal studies, cross-cultural investigations, teacher professional development, equity-focused studies, and exploration of multimodal learning through emerging technologies like virtual and augmented reality. In conclusion, deep learning offers transformative potential for EFL education by enabling adaptive, personalized, and multimodal learning experiences. However, realizing this potential requires holistic integration that balances technological innovation with pedagogical, cultural, and ethical considerations. The future success of deep learning in EFL depends on collaborative efforts to create meaningful, equitable, and culturally sensitive language learning environments.

REFERENCES

Ahmed, K., & Singh, R. (2023). Deep learning-based rhetorical structure feedback for EFL writers. *Computer Assisted Language Learning*, *36*(3), 218–237.

Booth, A. (2023). Systematic approaches to searching for evidence in interdisciplinary research. *Information Research Journal*, *28*(2), 122–138.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., & Amodei, D. (2020). Language models are few-shot learners. Advances in Neural Information Processing. *Systems*, *33*, 1877–1901.

Chapelle, C. A., & Sauro, S. (2023). *The handbook of technology and second language teaching and learning (2nd ed.)*. Wiley-Blackwell.

Chen, H., & Williams, T. (2023). Explainable AI approaches in adaptive language learning: A framework for transparency. *ReCALL*, *35*(2), 178–196.

Chen, J., Li, M., & Thompson, P. (2020). Vocabulary learning path optimization through deep reinforcement learning. *CALICO Journal*, *37*(1), 1–19.

Chen, X., & Meurers, D. (2022). Leveraging deep learning for intelligent vocabulary acquisition in EFL contexts: An experimental study. *Computer Assisted Language Learning*, *35*(4), 643–667.

Chen, Y., Wang, L., & Smith, K. (2023). Integration challenges for deep learning applications in institutional LMS platforms. *Educational Technology Research and Development*, *71*(1), 131–152.

Chen, Y., & Williams, G. (2022). Multimodal feedback in AI-based EFL systems: Enhancing integrated competence. *Computer Assisted Language Learning*, *35*(3), 231–247.

Chen, Y., & Williams, G. (2023). Explainability in AI-driven EFL learning: Teachers' perceptions and adoption. *Language Learning & Technology*, *27*(1), 88–103.

Chen, Z., Zhang, X., & Lin, Y. (2020). Personalized grammar feedback using deep learning. *ReCALL*, *32*(2), 145–161.

Chung, H., & Li, J. (2023). Accent-robust pronunciation assessment using hybrid CNN-RNN architectures. . . *Computer Speech & Language*, *77*, 130–151.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319–340.

García-López, R., Chen, W., & Thompson, K. (2021). Data requirements for deep learning applications in EFL contexts. *International Journal of Artificial Intelligence in Education*, *31*(2), 259–283.

García-Peñalvo, F., Martín, E., & Santos, C. (2023). Immediate pronunciation feedback through deep learning: Effects on phonemic accuracy. *ReCALL*, *35*(1), 73–94.

García-Sánchez, S., & Santos-Espino, J. M. (2022). Emerging deep learning applications in foreign language education: A systematic review. *ReCALL*, *34*(3), 326–346.

García, M., & Smith, J. (2023a). Privacy considerations in speech data collection for personalized pronunciation training. *Language Learning & Technology*, *27*(1), 182–205.

García, M., & Smith, T. (2023b). Privacy in AI-enhanced education: Risks and responsibilities. *Educational Technology Research and Development*, *71*(1), 155–176.

Godwin-Jones, R. (2024). Adaptive learning revisited: Promise and reality in AI-based language learning. *Language Learning & Technology*, *28*(1), 1–12.

Gough, D., Oliver, S., & Thomas, J. (2017). *An Introduction to Systematic Reviews*. Sage Publications.

Hong, Q. N., Pluye, P., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., & Vedel, I. (2018). *Mixed Methods Appraisal Tool (MMAT), version 2018*. Canadian Institutes of Health Research.

Ibrahim, R., & Hassan, M. (2021). Adaptive vocabulary learning with deep neural networks. *CALL-EJ*, *22*(2), 34–48.

Johnson, A., & Martinez, P. (2021). Learner-specific error correction using deep learning. *Journal of Educational Computing Research*, *59*(4), 703–721.

Johnson, A., Zhou, M., & Ali, K. (2022). Cultural bias in AI-powered language tools: A global perspective. *AI & Society*, *37*(3), 509–523.

Johnson, K., & Martinez, L. (2021). Personalized grammar instruction through error pattern recognition: A deep learning approach. *System*, *96*, 211–230.

Johnson, P., Lee, K., & Wong, M. (2022). Cultural considerations in automated feedback systems: A comparative study in East Asian EFL contexts. *RELC Journal*, *53*(2), 264–283.

Kern, R., & Develotte, C. (2023). Critical technological literacy in language education. *ReCALL*, *35*(2), 210–225.

Kim, J., & Morales, A. (2022). Adaptive listening comprehension practice through multi-dimensional difficulty scaling. *Language Learning & Technology*, *26*(2), 172–194.

Kim, S., Park, J., & Lee, H. (2021). Multimodal intelligent tutoring for pragmatic competence development. *System*, *98*, 65–84.

Kukulska-Hulme, A., & Viberg, O. (2024). Pedagogical innovation in the shadow of AI. *British Journal of Educational Technology*, *55*(1), 301–320.

Larsen-Freeman, D. (2023). Emergence in second language acquisition and AI. *Modern Language Journal*, *107*(S1), 80–94.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.

Lee, C., & Chang, S. (2022). Visual feedback on articulatory features: Deep learning approaches to pronunciation training. *Language Learning & Technology*, *26*(3), 41–63.

Lee, J., & García, P. (2023). Transfer learning for low-resource language tutoring: From English to Vietnamese. . . *Computer Assisted Language Learning*, *36*(2), 261–283.

Lee, K., Wang, J., & Smith, T. (2021). Teacher factors in deep learning implementation: Professional development needs and adoption patterns. *System*, *97*, 325–347.

Li, J. (2023). Artificial intelligence in language education: Current applications and future directions. *Language Learning & Technology*, *27*(1), 2–28.

Li, P., Xu, K., & Wang, H. (2021). Deep learning-based pronunciation assessment and feedback generation for Chinese EFL learners. *TESOL Quarterly*, *55*(2), 397–422.

Li, W., Zhang, Y., & Thompson, K. (2023). Integrating affect recognition in adaptive EFL learning environments. *Language Learning & Technology*, *27*(2), 198–221.

Lin, P., & Sharma, K. (2022). Computational resource requirements for state-of-the-art deep learning in language education. *International Journal of Artificial Intelligence in Education*, *32*(3), 397–419.

Liu, Q., & Huang, X. (2020). The application of deep learning in automated essay scoring for EFL students. *Journal of Educational Technology & Society*, *23*(2), 142–153.

Liu, Y., & Li, H. (2024). Deep learning for English grammar instruction: A comparative study. *Journal of Educational Technology Development and Exchange*, *17*(1), 14–29.

Lyster, R., & Sato, M. (2024). Responsive feedback theory in L2 acquisition. *Language Teaching*, *57*(2), 211–225.

Methley, A. M., Campbell, S., Chew-Graham, C., McNally, R., & Cheraghi-Sohi, S. (2022). PICOC revisited: Enhancing question formulation in systematic reviews. *BMC Medical Research Methodology*, *22*(1), 105.

Mishra, P., & Koehler, M. J. (2006). Technological Pedagogical Content Knowledge: A framework for teacher knowledge. *Teachers College Record*, *108*(6), 1017–1054.

Nguyen, T., Smith, P., & Jones, A. (2022). L1-specific grammar error correction using transformer architectures fine-tuned on Vietnamese EFL learner corpora. *Computer Assisted Language*

Learning, *35*(4), 415–438.
Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71.
Rahman, S., & Chen, J. (2022a). Digital divide implications for AI-enhanced language learning: An equity analysis. *TESOL Quarterly*, *56*(2), 297–321.
Rahman, S., & Chen, L. (2022b). Equity and access in AI-supported EFL learning. *Educational Review*, *74*(3), 304–321.
Rivera, J., Lee, H., & Sun, T. (2023). Reporting standards for AI in education: The PRISMA-AI extension. *Journal of Artificial Intelligence in Education*, *33*(1), 45–68.
Rodríguez-Tapia, A., González, M., & Smith, J. (2021). Real-time grammatical feedback during speaking practice: A transfer learning approach. *Computer Assisted Language Learning*, *34*(3), 302–325.
Schmidt, A., & Park, J. (2020). Context-aware vocabulary assistance using deep semantic models. *ReCALL*, *32*(1), 49–67.
Schmidt, K., Lee, J., & Wong, P. (2023). Maintaining optimal challenge through deep reinforcement learning in game-based language practice. *System*, *109*, 315–336.
Song, Y., Cheng, L., & Wang, W. (2021). Affective outcomes of VR conversation practice with deep learning-driven virtual agents. *Language Learning & Technology*, *25*(1), 78–99.
Stickler, U., & Hampel, R. (2023). Algorithmic literacy in language teacher education. *Language Learning in Higher Education*, *13*(2), 159–177.
Tafazoli, D. (2023). Pedagogical perspectives on AI integration in CALL. *CALL-EJ*, *24*(1), 170–183.
Thompson, J., Lee, M., & García, K. (2023). Learning sciences foundations for deep learning applications in language education. *Educational Technology Research and Development*, *71*(3), 325–349.
Verspoor, M., & Lowie, W. (2023). Complex dynamic systems theory and L2 development. *Language Teaching*, *56*(3), 305–320.
Wang, J., Chen, H., & Smith, K. (2022). Deep reinforcement learning for adaptive material selection in vocabulary learning. *Computer Assisted Language Learning*, *35*(1), 127–149.
Wang, L., Chen, K., & Thompson, J. (2023). Multimodal affect detection in language learning contexts: Combining facial, vocal, and textual signals. *ReCALL*, *35*(3), 271–294.
Wang, Y., & Deng, L. (2020). Deep learning for natural language processing: Advantages and challenges in language education. *International Journal of Educational Technology in Higher Education*, *17*(1), 126–139.
Wang, Y., & Kapoor, R. (2022). Emotion-aware AI in EFL: Impacts on motivation. *System*, *106*, 102791.
Wang, Y., & Liu, J. (2024). Acceptance and pedagogical integration of AI in English language education: A TAM-TPACK approach. *Language Learning & Technology*, *28*(1), 12–34.
Wong, J., & Lee, P. (2021). Automated question generation and evaluation for reading comprehension practice. *CALICO Journal*, *38*(2), 152–174.
Wong, J., & Thompson, P. (2022). Curriculum alignment challenges in automated assessment systems: Instructor perspectives on deep learning applications. *System*, *106*, 203–224.
Wong, K., Chen, J., & Lee, P. (2022). Individual forgetting curve prediction for optimal vocabulary review scheduling. *Language Learning & Technology*, *26*(3), 397–419.
Wong, S., Lau, C., & Fung, T. (2022). Optimizing spaced repetition with neural networks. *Language Learning & Technology*, *26*(3), 54–72.
Xiao, Y., & Watson, M. (2023). Conducting technology-focused systematic reviews: A methodological guide. *Review of Educational Research*, *93*(2), 210–240.
Yang, Y., & González-Lloret, M. (2023). Measuring what matters? A critique of AI assessment in language education. *Language Testing*, *40*(3), 387–405.
Zhang, J., & Zou, B. (2023). Teachers as co-creators in AI-supported EFL classrooms. *Interactive*

Learning Environments, *31*(3), 210–228.
Zhang, K., Lin, L., & Qian, Y. (2024). Mapping AI applications in language education: A systematic review across disciplines. *Computers & Education*, *206*, 104755.
Zhang, Y., Wang, J., & Li, K. (2019). Hierarchical attention networks for automated essay scoring in EFL contexts. *CALICO Journal*, *36*(1), 32–51.
Zhang, Y., & Wu, B. (2023). Artificial intelligence-powered conversation partners: Measuring impacts on EFL speaking proficiency and anxiety. *System*, *106*, 102824.
Zhao, K., Lee, M., & Chen, J. (2021). Personalized reading material selection through deep learning: Effects on comprehension and engagement. *System*, *102*, 261–283.
Zhao, L., Tan, Y., & Huang, S. (2021). Personalized reading recommendation in EFL using AI. *ReCALL*, *33*(2), 230–247.
Zhao, L., & Wang, H. (2022). Learner behavior modeling using deep learning. *Educational Technology Research and Development*, *70*(2), 477–493.
Zhou, Y., & Li, K. (2024). Designing culturally responsive AI for language learning. *AI in Education*, *4*(1), 121–134.
Zhu, C., Zeng, M., & Huang, X. (2021). Using deep learning to enhance EFL learning: A critical review of research from 2010-2020. *CALL-EJ*, *22*(1), 1–24.
Zou, D., & Thomas, M. (2019). Integrating deep learning technologies in language education: A systematic review of research. *Computer Assisted Language Learning*, *32*(8), 930–953.