Online-Journals.org

iJEP International Journal of Engineering Pedagogy

iJEP | eISSN: 2192-4880 | Vol. 16 No. 3 (2026) | OPEN ACCESS

https://doi.org/10.3991/ijep.v16i3.54465

PAPER

Development of an AI-based Spanish-Japanese Voice Translator for Engineering Education

Jose Antonio Chinchay-Delgado, Nicole Stephania Salazar-Jo, Cristian Castro-Vargas (✉)

Universidad Privada del Norte, Lima, Peru

cristian.castro@upn.pe

ABSTRACT

Linguistic diversity represents a significant challenge in engineering education, particularly in contexts that require international collaboration and access to multilingual technical content. Although commercial machine translation tools are widely available, few solutions are specifically designed to support pedagogical objectives in engineering training environments. This study presents the design and implementation of an artificial intelligence (AI)-based Spanish–Japanese voice translator with a modular architecture tailored for educational use. The system integrates speech recognition, optical character recognition, and machine translation components and was evaluated under controlled conditions using typed text, handwritten input, and voice data. Experimental results show an overall accuracy above 94%, with particularly strong performance in handwritten text recognition (98%) and Spanish audio transcription (96%). Beyond translation functionality, the proposed system supports project-based learning (PBL) by enabling students to interact with and modify AI modules, fostering competencies in natural language processing and intelligent systems design. The findings suggest that AI-driven multilingual tools can enhance digital inclusion and intercultural communication in engineering education contexts.

KEYWORDS

machine translation, Japanese, speech recognition, natural language processing (NLP), engineering education, artificial intelligence (AI), digital inclusion

1 INTRODUCTION

The technological revolution has radically transformed our communication, making language translation a fundamental necessity for an increasingly interconnected society [1]–[4]. This transformation is particularly relevant in engineering education, where interdisciplinary and global collaboration requires tools that eliminate linguistic and cultural barriers. Artificial intelligence (AI)-based technologies, such as speech recognition and automatic language processing, have enabled more fluid interactions between people from diverse linguistic backgrounds [5]–[13].

Chinchay-Delgado, J. A., Salazar-Jo, N. S., Castro-Vargas, C. (2026). Development of an AI-based Spanish-Japanese Voice Translator for Engineering Education. *International Journal of Engineering Pedagogy (iJEP)*, 16(3), pp. 67–85. https://doi.org/10.3991/ijep.v16i3.54465

Article submitted 2025-01-16. Revision uploaded 2026-03-01. Final acceptance 2026-03-01.

In addition to facilitating communication, these tools promote digital inclusion by reducing the technological gap between communities and enabling equitable access to information [14], [15]. Digital inclusion, understood as equal access and participation in technological environments, is strengthened by using multilingual translators in education, facilitating non-native speakers' understanding of content. Furthermore, these technologies allow for preserving identity elements of the cultures of origin, contributing to intercultural exchange [16]–[18]. For example, a Spanish-Japanese translator can serve as a bridge to understand Japanese technical texts in engineering or promote language learning in double-degree programs [19]–[21]. Unlike general-purpose commercial translators, this project is explicitly oriented toward engineering education and incorporates an open and modular architecture that allows pedagogical adaptation and experimental modification. Its implementation is not limited to literal translation but seeks to promote semantic comprehension and interaction with handwritten texts and voice input, even with varied pronunciations [22]–[25].

The choice of the Spanish-Japanese language pair responds to the need to connect communities with strong cultural traditions and growing scientific collaboration. Japan is a global technological leader, and Peru, the project's country of origin, maintains academic and commercial relations with the government. Japan is a global technological leader, and Peru, the project's country of origin, maintains long-standing academic and commercial relations with Japan [26]. Furthermore, the Japanese language has complex linguistic features such as multiple writing systems (hiragana, katakana, and kanji), which represent an ideal challenge for the implementation of AI models [27]. From a Latin American perspective, recent initiatives demonstrate the growing adoption of AI-driven educational technologies, including AI-based applications for Peruvian Sign Language [28], automatic speech recognition advancements for underrepresented languages in the Americas [29], and broader regulatory and policy developments in artificial intelligence implementation [30]. At a global level, AI-driven industrial and communication applications [31] establish relevant benchmarks, demonstrating the transformative potential of intelligent systems across multiple sectors, including education. This project, framed within this context, proposes a solution that not only translates between two languages but also develops students' technical skills, promotes intercultural understanding, and enables experimentation with emerging technologies in education.

In this context, the present study contributes to the field of engineering pedagogy in three main ways. First, it proposes a modular AI-based Spanish–Japanese voice translation architecture specifically designed for engineering learning environments. Second, it evaluates the technical performance and pedagogical applicability of the system in controlled academic scenarios, integrating speech recognition, optical character recognition, and machine translation components. Third, it offers a replicable framework for incorporating AI-driven multilingual tools into engineering education, supporting the development of technical competencies in natural language processing and intelligent systems design. By addressing both technological implementation and educational integration, this work advances the use of artificial intelligence as an active learning instrument in multilingual engineering contexts.

2 LITERARY REVIEW

Several recent studies have explored how emerging technologies, such as AI, natural language processing (NLP), and speech recognition systems, transform

engineering education and language teaching. In the study [32], the impact of humanoid educational robots on the motivation and participation of primary school students was evaluated. Significant increases in intrinsic motivation (+12.07%) and cognitive skills (+21.56%) were observed. Although the context was at a basic level, the findings show the potential of AI to strengthen socio-emotional skills in learning, which is relevant in initial engineering training. For their part, [33] analyzed how NLP influences digital and social interaction through immersive technologies such as the metaverse. The study identified ethical challenges in NLP algorithms and highlighted their impact on modifying digital behaviors. These findings reinforce the need to integrate ethical approaches into the design of educational translation systems. In [34], they developed language tutoring systems with automatic speech recognition optimized by semi-supervised learning, highlighting the system's accuracy and adaptability for real-time linguistic feedback. This is directly related to the functionality of our voice translator, especially in multilingual contexts.

Likewise, [35] emphasized the importance of incorporating emotional intelligence into educational chatbots, arguing that systems must be culturally sensitive to improve the user experience. They also pointed out challenges in semantic interpretation and response personalization, which were addressed in our project. Furthermore, recent research has addressed innovative AI-based pedagogical approaches. [36] presented a framework for integrating generative AI (GenAI) into technical education, highlighting its impact on educational tools and practices and its role in transforming intercultural communication and curriculum reform. The study also underlines the need to address ethical and legal challenges, proposing interdisciplinary models. On the other hand, [37] explores the use of AI in English teaching in Japan, focusing on how this technology can reduce students' anxiety, improve their fluency, and enable the personalization of learning programs. These findings are especially relevant in contexts where multilingualism and communicative fluency, such as engineering training, are key competencies. Other relevant studies include the work of [38], who analyzed the integration of translation systems in augmented reality environments to support language learning in engineering, and the research of [39], which explores how machine translation can improve access to content for non-native learners. In summary, recent literature shows a convergence between AI, NLP, and technical education, identifying key trends: adaptive learning, ethical integration, cultural sensitivity, and multilingual accessibility. However, gaps persist in research on translators designed explicitly for engineering training contexts, reinforcing this study's relevance and originality.

3 JUSTIFICATION OF THE ARTIFICIAL INTELLIGENCE MODELS USED

The AI models used in this project were selected based on a comprehensive analysis of technical, pedagogical, and accessibility criteria. Priority was given to using tools that balanced performance, ease of integration, available documentation, and relevance to the educational environment.

The Google Cloud Speech-to-Text API was chosen as the speech recognition engine due to its support for multiple languages, including Latin American Spanish; its real-time processing capabilities; and its ease of integration using Python libraries.

These features make it a viable option for learning environments that require speech processing under diverse conditions.

GoogleTrans, an unofficial Google Translate interface, was selected for its ease of use, support for bidirectional translation, and adaptability within open-source applications. Although its use is intended for educational or experimental, non-commercial purposes, it allows students to explore linguistic structures and translation workflows without implementation barriers.

4 MATERIALS AND METHODS

4.1 Research design

This study follows an applied research design with a case study approach in an academic engineering context. The objective was to design, implement, and technically validate an AI-based Spanish–Japanese voice translation system intended for potential use in engineering education environments.

The technical validation of the proposed system was conducted through controlled test scenarios simulating academic engineering use cases. The evaluation focused on system functionality, translation accuracy, and multimodal input processing, including typed text, handwritten text through optical character recognition, and voice input through speech recognition modules.

Accuracy rates were calculated by comparing system outputs with verified reference translations in each test scenario. The study did not involve experimental manipulation of independent variables nor the application of structured educational measurement instruments. At this stage, no formal pedagogical experimental study was conducted. The primary objective was to validate the technical performance of the system and explore its potential applicability in engineering education contexts.

4.2 Evaluation procedure

The evaluation procedure consisted of three stages: (1) text-based translation tests, (2) handwritten character recognition tests using OCR, and (3) Spanish voice transcription and translation tasks. System accuracy was calculated by comparing output results with validated reference translations.

4.3 Data analysis

Descriptive statistical analysis was applied to determine system accuracy rates across different input modalities. Accuracy percentages were calculated by dividing correct outputs by total test cases. No inferential statistical tests were performed, as the study aimed at technical validation rather than hypothesis testing.

4.4 Ethical considerations and data privacy

During the development of the Spanish-Japanese translation system, the ethical principles established in the Code of Ethics for Scientific Research of the Universidad

Privada del Norte (UPN, version 2024) were respected. The study was part of a technological research project without clinical intervention or the collection of sensitive data.

Informed consent: Participants participating in the voice tests were previously informed about the project objectives, the educational use of the system, and their right to withdraw without repercussions. Their consent was given voluntarily and documented by the institutional guidelines established by the Research Department.

Anonymizsation and data protection: All collected voice samples were anonymized before processing. Names, biometric data, and personal metadata were not included. Alphanumeric codes were assigned, and the recordings were stored in secure environments, with access restricted to the research team. Data management complied with Law No. 29733—Peru's Personal Data Protection Law—and the confidentiality provisions defined by the UPN.

Ethical Approval: Since the study did not involve risks to the physical or emotional integrity of the participants, nor did it directly intervene with vulnerable populations, and considering that the research focused on technological development without the collection of sensitive information, approval from the UPN's Institutional Research Ethics Committee (CIEI) was not required, as provided in Articles 8 and 20 of the current Code of Ethics (Rector's Resolution No. 028-2024-UPN).

This ethical treatment is aligned with international principles (Helsinki, Belmont, CIOMS) and with the institutional principles of beneficence, truthfulness, confidentiality, and participant respect.

4.5 Introduction to the system

The system's structure is based on a logical architecture that integrates multiple components for speech and text processing (see Figure 1). The workflow begins with the capture of Spanish input, either in audio or text format, and continues through several processing stages, including speech recognition, translation using AI models, and the generation of Japanese output. This approach fosters robust and flexible processing, suitable for addressing the complexities inherent in both languages.

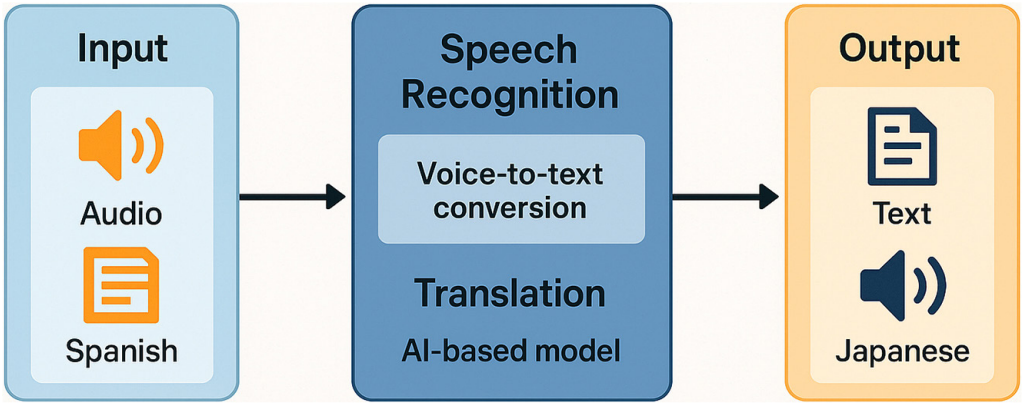

Fig. 1. Logical software architecture

4.6 Logical architecture

The logical architecture of the system is structured into four main functional modules organized in a modular and scalable configuration, enabling future educational and technical adaptations (see Figure 2).

The first module corresponds to the input layer, responsible for capturing user data in Spanish, either as audio signals or typed text. The second module performs voice recognition using an AI-based engine that converts speech into textual representation. The third module implements machine translation through advanced machine learning algorithms that transform Spanish input into Japanese output. Finally, the output generation module delivers the translated content either as Japanese text or synthesized audio.

This modular configuration enables efficient processing, system scalability, and functional adaptability to educational and research environments.

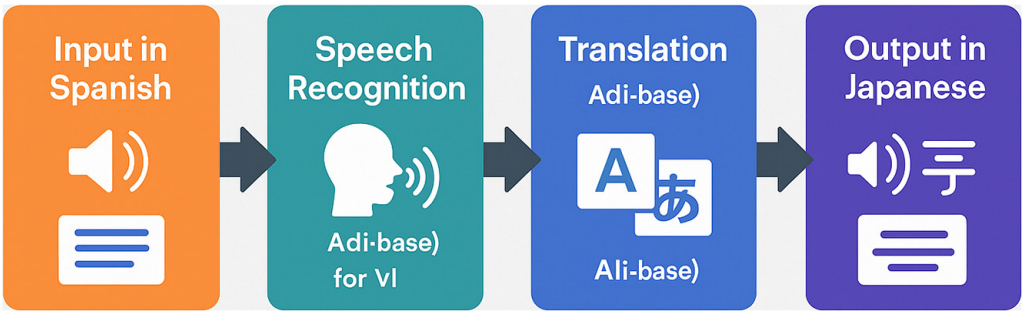

Fig. 2. Logical software architecture

4.7 Physical architecture

The physical implementation of the system is distributed across several interconnected components that support real-time processing and modular scalability (see Figure 3).

The user interaction layer is implemented through a web-based interface developed in Streamlit, enabling intuitive access to text input, handwritten input, and voice interaction functionalities. This interface acts as the entry point for data acquisition and system control.

The processing layer is supported by web and AI servers responsible for speech recognition and machine translation services. These services manage audio preprocessing, text normalization, and communication with external APIs required for translation and speech-to-text operations.

Additionally, specialized modules handle audio encoding, optical character recognition (OCR), and text formatting processes to ensure coherent data flow between system components. A linguistic database supports translation accuracy by providing structured lexical resources and contextual references that enhance semantic interpretation.

This distributed architecture enables functional separation, scalability, and adaptability, ensuring compatibility with diverse educational deployment environments.

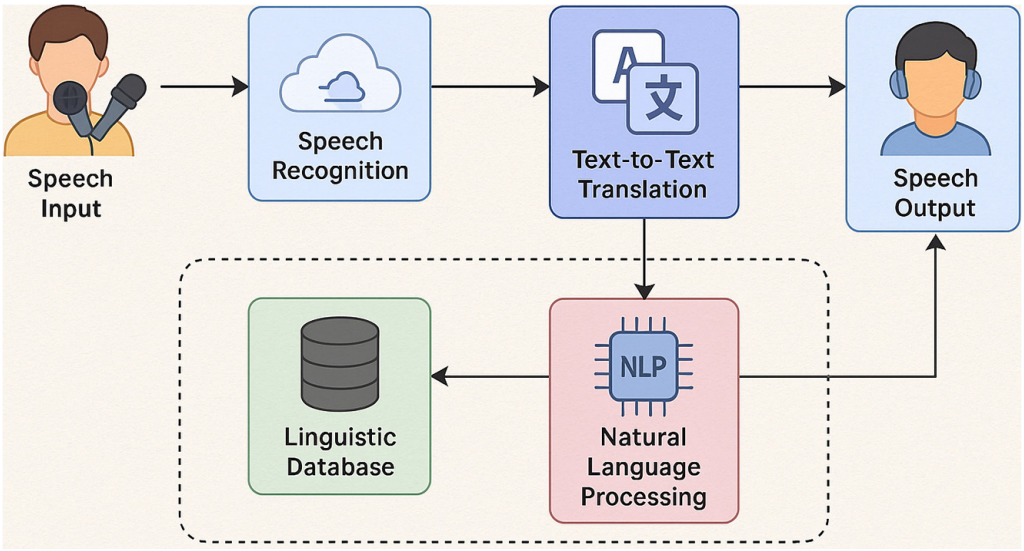

Fig. 3. Physical software architecture

4.8 Technical specifications

Real-time processing of audio signals and running AI-based translation algorithms require specific computing resources. To ensure smooth operation and an optimal experience, the following minimum specifications were established (refer to Table 1), which were sufficient for the smooth execution of all tests developed during this study:

Table 1. Minimum system requirements

Requirements	Minimum	Recommended
Processor (CPU)	8th Generation Intel Core i3 or equivalent	8th Generation Intel Core i5 or higher
Memory (RAM)	8 GB	16 GB
Storage	256 GB SSD	512 GB SSD
Operating System (OS)	Windows 10, macOS, or Linux later than 18.4	Windows 10/11, macOS, or Linux after 20.04
Graphics Card (GPU)	Integrated	NVIDIA GTX 1050 or higher

4.9 Tools and libraries

The software architecture, developed in Python using Visual Studio Code, integrates several specialized components. The libraries used and their functions are presented in Table 2:

Table 2. Libraries implemented in the project

Bookshop	Function
Streamlight	Rapid creation of web applications
Pillow	Image processing
Pytesseract	Tesseract OCR Engine
Googletrans	Unofficial Google Translate Interface
Streamlight drawable canvas	Streamlight extension for web annotations
Pykakasi	Romanization of Japanese
Google Cloud Speech	Google Cloud Speech Recognition API

4.10 Pedagogical impact

The system's modular design facilitates maintenance and future expansion, providing a unique learning environment. This environment enables project-based learning (PBL) implementation, as students can integrate their modules, enhance existing ones, or experiment with various AI and NLP libraries. In this process, future engineers acquire technical skills and develop competencies in system design, computational thinking, and complex problem-solving.

5 RESULTS

5.1 Overall system performance

The Spanish-Japanese translator was evaluated under controlled conditions using three input formats: typed text, handwritten text, and voice. Translations were processed in both directions (Spanish—Japanese) and evaluated using automated metrics (BLEU, METEOR, and WER) complemented by human validation. Factors such as latency, robustness to noise, accuracy in voice commands, and the system's ability to maintain semantic consistency across linguistic variations were also considered.

5.2 Evaluation of textual entries

The system's evaluation of text input yielded favorable results regarding accuracy and translation quality. For typed texts, an accuracy of 97% was achieved, with a BLEU metric of 0.76 and a METEOR metric of 0.68. The most common errors were associated with the omission of grammatical elements typical of formal Japanese, indicating specific areas for improvement in the cultural and syntactical adaptation of the translation. Regarding handwritten input, the system demonstrated remarkable recognition capabilities, achieving 98% accuracy for legible handwriting and 91% for irregular handwriting, thanks to the performance of the integrated OCR module. Translations for these inputs achieved BLEU scores of 0.80, METEOR scores of 0.72, and a word error rate (WER) of 6.8%, making this modality the most robust regarding adaptability and accuracy.

This performance reinforces the system's educational value in contexts where students work with non-digital or handwritten input, facilitating active learning in optical character recognition and natural language processing. Table 3 summarizes the main results obtained, while Figures 4–8 illustrate representative examples of the complete flow of the system: from the capture of handwriting (see Figures 4 and 5), its digital conversion and translation into Japanese (see Figures 6 and 7), to the back-translation into Spanish (see Figure 8).

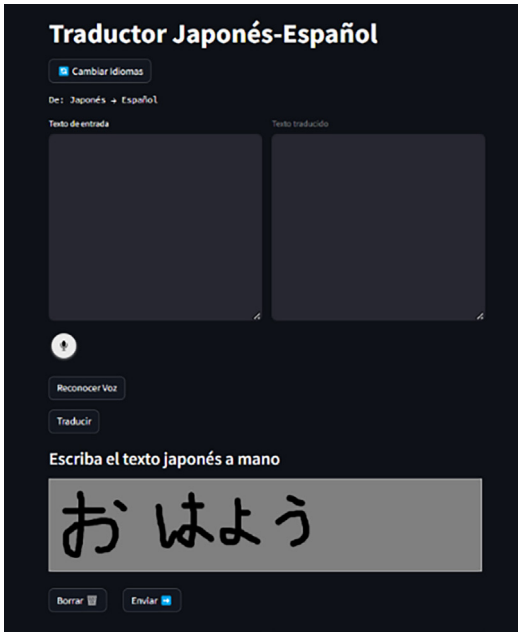

Fig. 4. Handwriting input

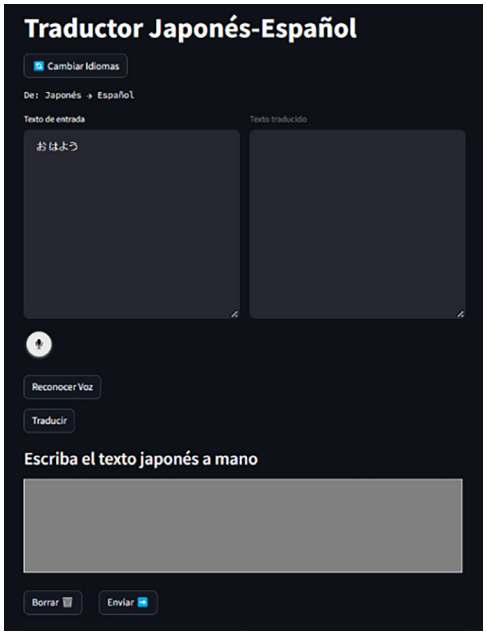

Fig. 5. Converting handwriting to text

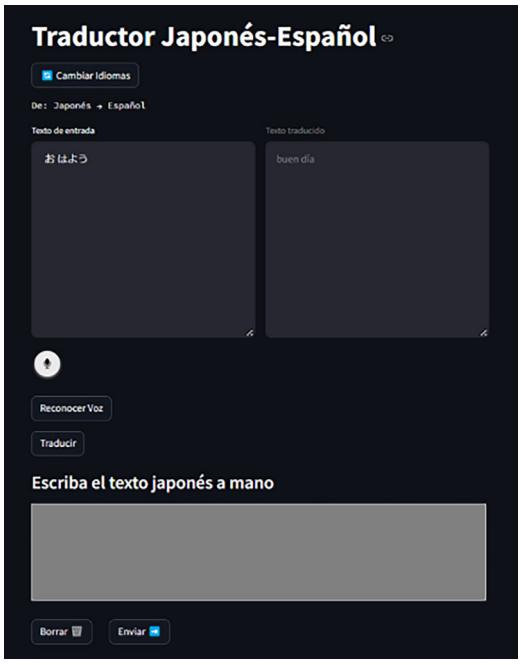

Fig. 6. Translation of the written text

Fig. 7. Translation of a text from Spanish to Japanese

Fig. 8. Translation of a text from Japanese to Spanish

Table 3. Evaluation metrics for textual input

Entry Type	Accuracy (%)	BLUE	METEOR	WATER (%)
Typewritten text	97	0.76	0.68	—
Legible handwriting	98	0.80	0.72	6.8
Irregular handwriting	91	0.80	0.72	6.8

5.3 Voice input evaluation

In Spanish audio testing, transcription accuracy was 96%, with BLEU at 0.74, METEOR at 0.66, and WER at 7.4%. The human evaluation indicated a 93% semantic match with reference translations.

The results for Japanese audio were slightly lower: accuracy of 92%, BLEU of 0.68, METEOR of 0.62, and WER of 9.1%. The main difficulties arose in interpreting cultural nuances, idiomatic expressions, and regional accents.

As for voice commands, activation was recognized with 95% accuracy and an average latency of 1.2 seconds, making it easy to use in accessibility applications, especially for users with mobility disabilities.

Figures 9–14 show the visual results for Spanish and Japanese voice input. These demonstrate the flow from audio capture to recognition activation, transcription, and bidirectional translation.

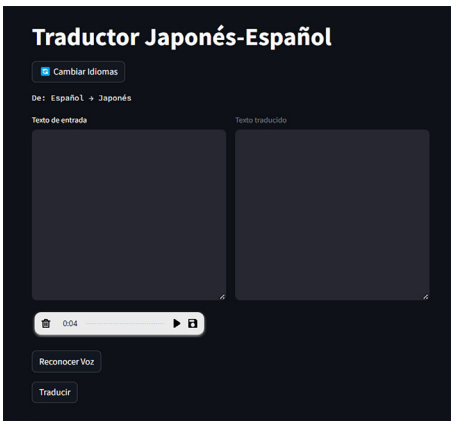

Fig. 9. Spanish voice input

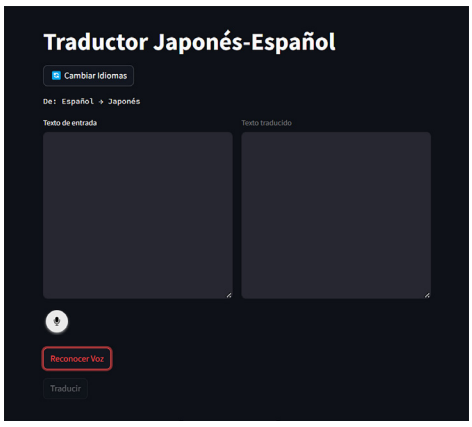

Fig. 10. Enable voice recognition as input

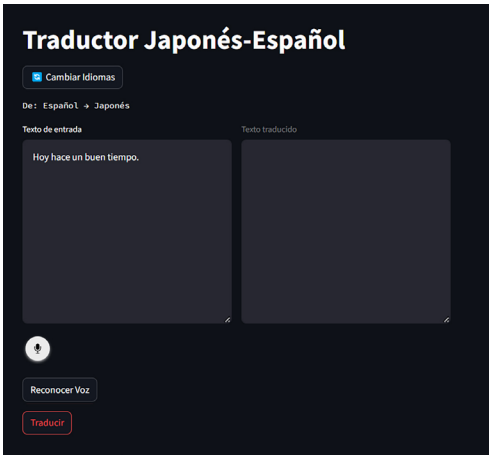

Fig. 11. Translation of the text obtained by audio

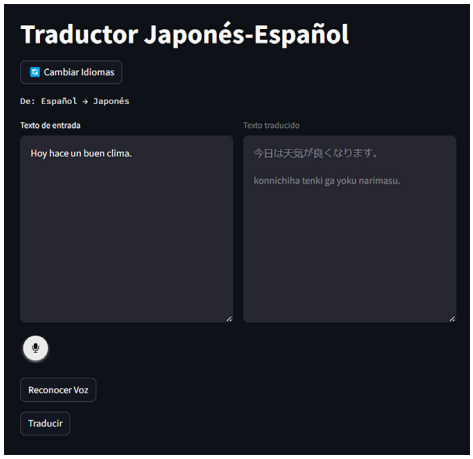

Fig. 12. Text translated from audio

Fig. 13. Japanese audio input

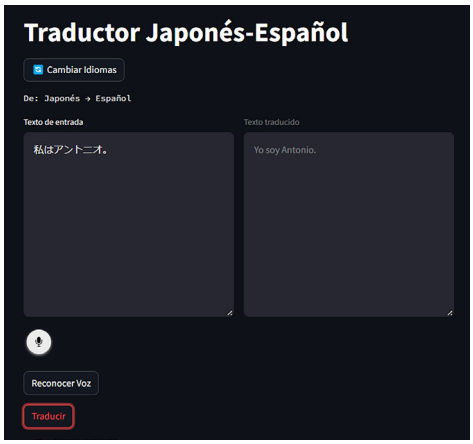

Fig. 14. Translation of the text from Japanese to Spanish

5.4 Error analysis by category

Table 4 presents the types of errors observed during testing, their relative frequency, and representative examples. Grammatical and contextual errors were the most frequent, affecting the structure of complex sentences and polysemous terms. Phonetic errors, common in voice input, were linked to similar pronunciations. The low impact of illegible handwriting confirms the effectiveness of OCR. Meanwhile, acoustic interference increased WER in uncontrolled environments, revealing areas where the system could still be optimized.

Table 4. Error analysis by category

Error Category	Frequency	Example
Grammatical	High	Incorrect use of “wa” particles
Phonetic	Average	Confusion between “ka” and “ga”
Contextual	High	Ambiguous translation of “career”
Illegible handwriting	Low	Requires OCR preprocessing
Noise interference	Average	WER increased by up to 10%

5.5 Integrated results by entry type

Accuracy by Text Input Type: Below are the consolidated accuracy percentages by textual input modality (refer to Table 5 and see Figure 15). Handwriting achieved the highest accuracy, reinforcing the effectiveness of the system's OCR module. Performance on typed text was also high, whereas complex characters (accents and special symbols) presented slightly greater challenges, indicating the need to strengthen semantic and graphical preprocessing.

Table 5. Accuracy by type of text input

Entry Type	Accuracy (%)
Handwriting	98
Typewritten text	97
Complex characters	94

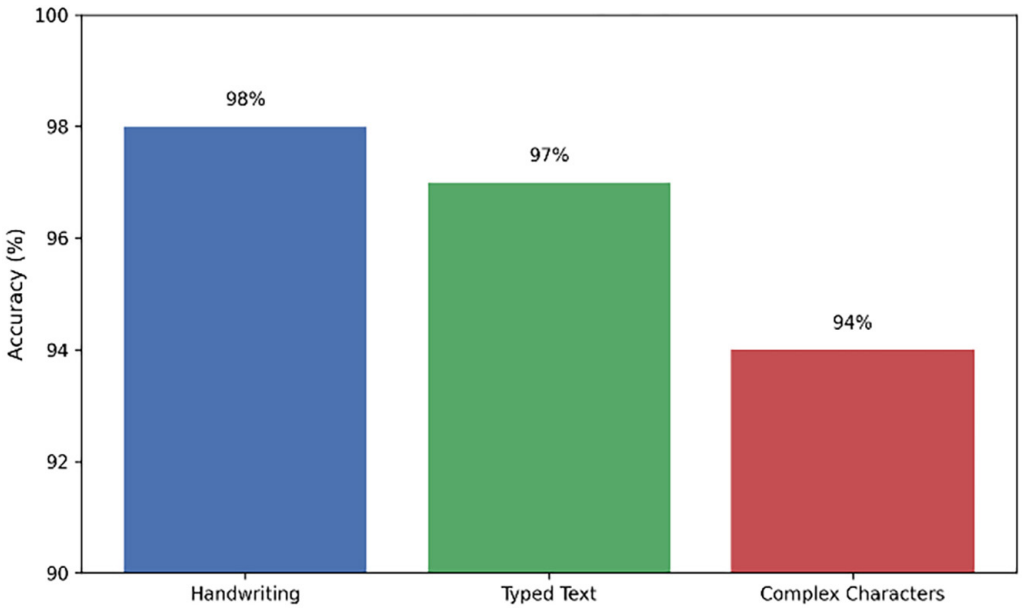

Fig. 15. Text Input Accuracy by Type

The results reveal differentiated system performance across input modalities. Handwriting achieved an accuracy of 98%, demonstrating the robustness of the OCR module under controlled conditions and supporting multimodal interaction in educational environments. Typed text input yielded 97% accuracy, maintaining stable recognition performance with minor discrepancies attributable to formatting variability. In contrast, complex characters resulted in a slightly lower accuracy of 94%, suggesting limitations in semantic and graphical preprocessing stages.

Overall, the system demonstrates strong performance across textual modalities, although further optimization is recommended to enhance processing of linguistically complex inputs.

Accuracy by Audio Input Scenario: Table 6 and Figure 16 present the accuracy results obtained under different audio input conditions. The system demonstrated high effectiveness when processing Spanish audio and structured voice commands.

However, slightly lower performance was observed for Japanese audio input, suggesting the need for improved exposure to dialectal variations, intonation patterns, and contextual linguistic nuances.

Table 6. Accuracy by audio input scenario

Audio Stage	Accuracy (%)
Audio in Spanish	96
Voice commands	95
Japanese audio	92

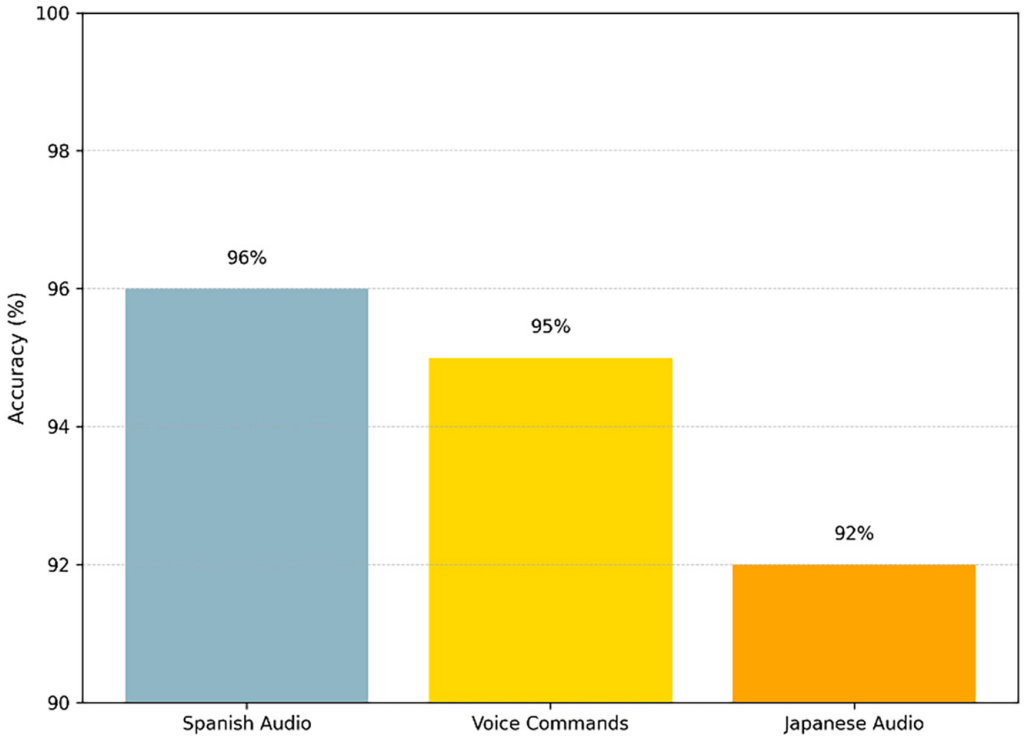

Fig. 16. Audio Input Accuracy by Scenario

Spanish audio input achieved an accuracy of 96%, confirming the system's robustness in transcribing spoken Spanish under controlled testing conditions. Voice commands yielded a comparable accuracy of 95%, highlighting the reliability of speech-triggered functionalities, which are particularly relevant for accessibility and interactive educational environments.

In contrast, Japanese audio input resulted in an accuracy of 92%. This decrease suggests challenges related to phonetic variability, dialectal diversity, and culturally embedded linguistic expressions. These findings indicate the need for expanded training datasets and enhanced contextual modeling to improve recognition performance in Japanese speech scenarios.

Overall, the system demonstrates solid performance across audio modalities; however, targeted optimization for Japanese audio processing would significantly enhance multilingual usability and pedagogical applicability.

6 DISCUSSION

The development and implementation of the proposed AI-based Spanish–Japanese voice translator demonstrated encouraging levels of technical accuracy and functional adaptability. Beyond reporting performance metrics, this discussion interprets these results in relation to existing literature, engineering pedagogy implications, and methodological boundaries.

6.1 Comparison with previous studies

The results obtained are aligned with recent advances in AI-supported language learning systems. Kang et al. [34] demonstrated that automatic speech recognition (ASR) integrated into intelligent tutoring systems enhances adaptive linguistic feedback and pronunciation training. In a comparable manner, the present system achieved over 94% accuracy in controlled multimodal scenarios, confirming the reliability of AI-driven speech processing in educational environments.

However, whereas the ASR-based tutoring architecture described in [34] focuses primarily on structured pronunciation exercises, the proposed system extends this approach by integrating speech recognition, handwritten text recognition, and machine translation within a unified modular framework. This multimodal design shifts the user role from passive recipient to active system-level explorer, which is particularly relevant in engineering education contexts.

Xiao et al. [35] emphasized the importance of emotional awareness and cultural sensitivity in AI-based educational tools. Consistent with this perspective, the contextual and phonetic inaccuracies detected in Japanese audio processing reveal persistent challenges in achieving culturally adaptive translation. These findings underscore the need for more diverse linguistic corpora and culturally representative datasets.

Finally, Chen [36] discussed the transformative potential of generative AI in translation and technical education, highlighting its role in reshaping intercultural communication and curriculum design. While Chen's contribution remains primarily conceptual, the present study operationalizes this vision through the development of a functional and replicable prototype tailored to engineering training, thereby bridging the gap between theoretical frameworks and applied educational technology.

6.2 Pedagogical impact and areas of opportunity

From an engineering pedagogy perspective, the system's primary contribution lies not only in its translation capability but also in its role as a learning artifact. The modular architecture enables students to analyze, modify, and experiment with individual AI components—including speech recognition engines, OCR modules, and translation APIs—thereby fostering PBL and systems-level thinking.

Unlike commercial platforms such as Google Translate or Microsoft Translator, which function as closed, black-box services, the proposed system exposes structural and computational layers for academic exploration. This transparency allows engineering students to develop competencies in NLP, API integration, software

modularization, and human–computer interaction, as well as critical awareness of AI ethics and technological dependency analysis.

Such integration of technical development with reflective pedagogical application strengthens computational thinking and problem-solving skills, which are core competencies in contemporary engineering curricula.

6.3 Comparative evaluation of AI models used

The selection of Google Cloud Speech-to-Text and Google Trans was guided by operational feasibility and pedagogical accessibility rather than exclusively by peak technical performance.

The Google Cloud Speech-to-Text API demonstrated recognition accuracy above 94% in controlled validation scenarios, with low latency suitable for real-time classroom interaction. Such responsiveness is essential in engineering education environments, where synchronous feedback supports active and participatory learning.

Open-source alternatives such as Whisper report comparable recognition accuracy (94–98%) in prior studies; however, their implementation typically requires dedicated GPU resources and advanced configuration. These technical requirements may limit adoption in resource-constrained educational institutions. Similarly, while DeepL offers high-quality text translation, it does not natively support integrated speech input, reducing its applicability in multimodal engineering training contexts.

Therefore, the tools selected represent a pragmatic balance between technical performance, accessibility, integration simplicity, and pedagogical feasibility. This decision reflects realistic deployment conditions in higher education, where reproducibility and ease of integration are as critical as algorithmic precision.

To address potential reproducibility concerns associated with proprietary APIs, the system was designed with a modular architecture that allows future substitution of engines such as Whisper or MarianNMT without altering the overall functional structure.

6.4 Technical and cultural limitations

Despite promising results, several limitations must be acknowledged:

First, Japanese linguistic complexity—including regional pronunciation variability, politeness levels, and context-dependent semantics—affected translation accuracy in idiomatic and polysemous expressions. This limitation highlights the inherent difficulty of handling linguistically distant language pairs in real-time AI systems.

Second, dependence on proprietary APIs introduces potential reproducibility constraints, as internal algorithm updates may alter performance over time. Although mitigated through modular system design, full algorithmic transparency remains limited.

Third, the linguistic databases employed were restricted in size and domain specificity. As a result, highly specialized engineering terminology may not be optimally processed. Future work should incorporate domain-adapted corpora to enhance contextual precision.

7 CONCLUSION

This study presented the design and technical validation of a modular AI-based Spanish–Japanese voice translation system tailored to engineering education contexts. The system successfully integrates speech recognition, optical character recognition, and machine translation components within a unified architecture, achieving accuracy levels above 94% under controlled validation scenarios.

Beyond its technical performance, the primary contribution of this work lies in its pedagogical positioning. Unlike commercial black-box translators, the proposed system functions as an educational artifact that exposes computational processes for academic exploration. Its modular architecture enables engineering students to engage directly with NLP pipelines, API integration, and AI system design, thereby fostering PBL, computational thinking, and interdisciplinary technological competencies.

Nevertheless, several limitations must be acknowledged. First, the technical validation was conducted exclusively under controlled test conditions rather than in longitudinal classroom implementations. Second, dependence on proprietary APIs introduces potential reproducibility constraints and technological dependency. Third, linguistic variability and cultural nuances in Japanese remain challenges for context-sensitive translation, particularly in specialized engineering discourse.

Future research should focus on empirical classroom validation through structured pedagogical interventions and quantitative learning assessments. Additional developments may include the integration of open-source alternatives such as Whisper or MarianNMT, expansion of domain-specific corpora, and extension to other multilingual engineering education contexts.

Overall, this study provides a replicable framework for integrating AI-driven multilingual tools into engineering curricula. By aligning technological implementation with educational objectives, it contributes to the advancement of inclusive, technology-enhanced engineering pedagogy in increasingly globalized academic environments.

8 REFERENCES

[1] D. Wang, X. Wang, and S. Lv, "An overview of end-to-end automatic speech recognition," *Symmetry*, vol. 11, no. 8, p. 1018, 2019. https://doi.org/10.3390/sym11081018

[2] R. Haeb-Umbach, J. Heymann, L. Drude, S. Watanabe, M. Delcroix, and T. Nakatani, "Far-field automatic speech recognition," *Proceedings of the IEEE*, vol. 109, no. 2, pp. 124–148, 2021. https://doi.org/10.1109/JPROC.2020.3018668

[3] J. Kahn, A. Lee, and A. Hannun, "Self-training for end-to-end speech recognition," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing*, 2020, pp. 7084–7088. https://doi.org/10.1109/ICASSP40776.2020.9054295

[4] A. Hannun, "The history of speech recognition to the year 2030," *arXiv preprint arXiv:2108.00084,* 2021. https://doi.org/10.48550/arXiv.2108.00084

[5] S. Zhou and H. Beigi, "A transfer learning method for speech emotion recognition from automatic speech recognition," *arXiv preprint arXiv:2008.02863*, 2020. https://doi.org/10.48550/arXiv.2008.02863

[6] W. Han *et al.*, "ContextNet: Improving convolutional neural networks for automatic speech recognition with global context," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, vol. 2020, 2020, pp. 3610–3614. https://doi.org/10.21437/Interspeech.2020-2059

[7] U. Kamath, J. Liu, and J. Whitaker, *Deep Learning for NLP and Speech Recognition*, 2019. https://doi.org/10.1007/978-3-030-14596-5

[8] S. Feng, O. Kudina, B. M. Halpern, and O. Scharenborg, "Quantifying bias in automatic speech recognition," *arXiv preprint arXiv:2103.15122*, 2021. https://doi.org/10.48550/arXiv.2103.15122

[9] M. Ustaszewski, "Towards a machine learning approach to the analysis of indirect translation," *Translation Studies*, vol. 14, no. 3, pp. 313–331, 2021. https://doi.org/10.1080/14781700.2021.1894226

[10] M. Popel *et al.*, "Transforming machine translation: A deep learning system reaches news translation quality comparable to human professionals," *Nature Communications*, vol. 11, no. 1, pp. 1–15, 2020. https://doi.org/10.1038/s41467-020-18073-9

[11] M. L. Romana García and B. Hernández Pardo, "Teaching innovation and technological competencies in translation: Artificial intelligence and corporate management," 2021, Accessed: May 31, 2024. [Online]. Available: https://repositorio.comillas.edu/xmlui/handle/11531/55599

[12] J. Canavilhas, "Artificial intelligence applied to journalism: Machine translation and content recommendation in the 'A European Perspective' (UER) project," *Latin American Journal of Communication*, no. 80, pp. 1–13, 2022. https://doi.org/10.4185/RLCS-2022-1534

[13] C. Vargas-sierra, "The translator's workstation in the age of artificial intelligence. Towards knowledge-assisted translation," *Pragmalingüística*, no. 28, pp. 166–187, 2020. https://doi.org/10.25267/Pragmalinguistica.2020.i28.09

[14] Z. Zhou *et al.*, "Sign-to-speech translation using machine-learning-assisted stretchable sensor arrays," *Nature Electronics*, vol. 3, no. 9, pp. 571–578, 2020. https://doi.org/10.1038/s41928-020-0428-6

[15] C. Ortiz-Leon, F. Yupanqui-Allcca, and B. Meneses-Claudio, "Using Artificial Intelligence for sign language translation: A systematic literature review," *Health, Science and Technology – Conference Series*, vol. 2, pp. 446–446, 2023. https://doi.org/10.56294/sctconf2023446

[16] S. López Jara, "The Spanish and Japanese information structure: Problems with the topic and the focus in teaching Spanish to Japanese people," in *Profiles, Factors and Contexts in the Teaching and Learning of ELE/EL2*, 2020, pp. 609–624. [Online]. Available: https://dialnet.unirioja.es/servlet/articulo?codigo=8188157

[17] L. Asquerino Egoscozábal, "The training of Japanese-Spanish translators. Current situation and prospects," *TDX (Doctoral Theses in the Network), Autonomous University of Barcelona*, 2021, Accessed: May 31, 2024. [Online]. Available: https://www.tdx.cat/handle/10803/673347

[18] L. A. Egoscozábal and A. H. Albir, "Experimental study on the acquisition of Japanese-Spanish translation competence. Design and results of the pilot study," *Meta (Canada)*, vol. 65, no. 2, pp. 394–419, 2020. https://doi.org/10.7202/1075842ar

[19] M. Ribés Fortanet, "Analysis of the translation of cultural references in the Japanese/Spanish dubbing of spirited away," 2023, Accessed: May 31, 2024. [Online]. Available: https://repositori.uji.es/xmlui/handle/10234/203479

[20] L. A. Egoscozábal, "The training of Japanese-Spanish translators in Spain. A study on the current situation," *Sendebar*, vol. 32, pp. 196–218, 2021. https://doi.org/10.30827/sendebar.v32.15903

[21] L. A. Egoscozábal, "Japanese-Spanish university translator training in Spain: Status of the issue," *CLINA Interdisciplinary Journal of Translation, Interpreting and Intercultural Communication*, vol. 8, no. 2, pp. 9–29, 2022. https://doi.org/10.14201/clina202282929

[22] A. B. Nassif, I. Shahin, I. Attili, M. Azzeh, and K. Shaalan, "Speech recognition using deep neural networks: A systematic review," *IEEE Access*, vol. 7, pp. 19143–19165, 2019. https://doi.org/10.1109/ACCESS.2019.2896880

[23] Y. He *et al.*, "Streaming end-to-end speech recognition for mobile devices," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 2019, 2019, pp. 6381–6385. https://doi.org/10.1109/ICASSP.2019.8682336

[24] M. Ravanelli, T. Parcollet, and Y. Bengio, "The pytorch-kaldi speech recognition toolkit," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 2019, 2019, pp. 6465–6469. https://doi.org/10.1109/ICASSP.2019.8683713

[25] T. Kano, S. Sakti, and S. Nakamura, "Transformer-based direct speech-to-speech translation with transcoder," in *2021 IEEE Spoken Language Technology Workshop, SLT 2021*, 2021, pp. 958–965. https://doi.org/10.1109/SLT48900.2021.9383496

[26] E. Castellanos and J. M. Cortéz, "Las relaciones comerciales Perú-Japón: la necesidad de un cambio," *Apunt. Rev. Ciencias Soc.*, no. 24, pp. 99–118, 1989. https://doi.org/10.21678/apuntes.24.287

[27] Y. Harun and F. N. Biduri, "Historical analysis of Japanese writing systems Hiragana, Katakana, and Kanji," *Int. J. Soc. Serv. Res.*, vol. 4, no. 2, pp. 612–618, 2024. https://doi.org/10.46799/ijssr.v4i02.720

[28] J. Huamani-Malca *et al.*, "Lessons from deploying the first bilingual peruvian sign language—Spanish online dictionary," in *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, Torino, Italia. ELRA and ICCL, 2024, pp. 10316–10323. Accessed: Jan. 15, 2025. [Online]. Available: https://aclanthology.org/2024.lrec-main.901/

[29] M. Romero, S. Gómez-Canaval, and I. G. Torre, "Automatic speech recognition advancements for indigenous languages of the Americas," *Appl. Sci.*, vol. 14, no. 15, p. 6497, 2024. https://doi.org/10.3390/app14156497

[30] Y. Walter, "Managing the race to the moon: Global policy and governance in Artificial Intelligence regulation—A contemporary overview and an analysis of socioeconomic consequences," *Discov. Artif. Intel.* vol. 4, no. 1, pp. 1–24, 2024. https://doi.org/10.1007/s44163-024-00109-4

[31] A. Bin Rashid and M. A. K. Kausik, "AI revolutionizing industries worldwide: A comprehensive overview of its diverse applications," *Hybrid Adv.*, vol. 7, p. 100277, 2024. https://doi.org/10.1016/j.hybadv.2024.100277

[32] K. Y. Fung, L. H. Lee, K. F. Sin, S. Song, and H. Qu, "Humanoid robot-empowered language learning based on self-determination theory," *Educ. Inf. Technol.*, vol. 29, no. 14, pp. 18927–18957, 2024. https://doi.org/10.1007/s10639-024-12570-w

[33] R. I. Sumon, S. M. I. Uddin, S. Akter, M. A. I. Mozumder, M. O. Khan, and H. C. Kim, "Natural language processing influence on digital socialization and linguistic interactions in the integration of the metaverse in regular social life," *Electron. 2024*, vol. 13, no. 7, p. 1331, 2024. https://doi.org/10.3390/electronics13071331

[34] B. O. Kang, H. B. Jeon, and Y. K. Lee, "AI-based language tutoring systems with end-to-end automatic speech recognition and proficiency evaluation," *ETRI J.*, vol. 46, no. 1, pp. 48–58, 2024. https://doi.org/10.4218/etrij.2023-0322

[35] Y. Xiao, T. Zhang, and J. He, "The promises and challenges of AI-based chatbots in language education through the lens of learner emotions," *Heliyon*, vol. 10, no. 18, p. e37238, 2024. https://doi.org/10.1016/j.heliyon.2024.e37238

[36] Y. Chen, "Generative Artificial Intelligence empowering translation: Current status, challenges, and prospects," *Educ. Rev. USA*, vol. 8, no. 12, pp. 1447–1454, 2024. https://doi.org/10.26855/er.2024.12.005

[37] A. Busso and B. Sanchez, "Advancing communicative competence in the digital age: A case for AI tools in Japanese EFL education," *Technol. Lang. Teach. Learn.*, vol. 6, no. 3, pp. 1211–1211, 2024. https://doi.org/10.29140/tltl.v6n3.1211

[38] R. O. Flores-Castañeda, S. Olaya-Cotera, and O. Iparraguirre-Villanueva, "Benefits of metaverse application in education: A systematic review," *Int. J. Eng. Pedagog.*, vol. 14, no. 1, pp. 61–81, 2024. https://doi.org/10.3991/ijep.v14i1.42421

[39] S. Beltozar-Clemente, E. Díaz-Vega, R. Tejeda-Navarrete, and J. Zapata-Paulini, "We can rely on ChatGPT as an educational tutor: A cross-sectional study of its performance, accuracy, and limitations in university admission tests," *Int. J. Eng. Pedagog.*, vol. 14, no. 1, pp. 50–60, 2024. https://doi.org/10.3991/ijep.v14i1.46787

9 AUTHORS

Jose Antonio Chinchay-Delgado is with the Facultad de Ingeniería, Universidad Privada del Norte, Lima, Peru. He holds a degree in computer systems engineering and specializes in cybersecurity. He has obtained multiple industry certifications and has developed advanced technical expertise in threat detection and security operations. He currently works as a Threat Detection Security Engineer on a project for a major financial institution in Peru, where he focuses on security event analysis, correlation, and continuous monitoring. His research interests include robotics, software engineering, artificial intelligence, cybersecurity, and penetration testing (E-mail: N00281648@upn.pe).

Nicole Stephania Salazar-Jo is with the Facultad de Ingeniería, Universidad Privada del Norte, Lima, Peru. She holds a bachelor's degree in computer systems engineering. She contributed to the development and implementation of the Spanish–Japanese translation system presented in this study, including software programming in Python, user interface design, and integration of speech recognition and machine translation tools. She has experience in data analysis and intelligent system development, with interests in applied artificial intelligence, data-driven technologies, and software engineering (E-mail: N00235661@upn.pe).

Cristian Castro-Vargas is with the Facultad de Ingeniería, Universidad Privada del Norte, Lima, Peru. He is an electronic engineer who graduated from the National University of Callao. He holds a doctorate in education and a master's degree in university teaching. He has extensive experience in applied research, electronic system development, automation, telecommunications, and academic innovation. His work integrates active learning methodologies with technological development in engineering education. He is a member of the IEEE Robotics and Automation Society (IEEE-RAS) (E-mail: cristian.castro@upn.pe).