A Personal Translation Assistant Based on Artificial Intelligence and Cloud Computing

Omar Otoniel Flores-Cortez*, Alberto Oviedo†
*Direccion de Investigaciones
Universidad Tecnológica de El Salvador, El Salvador
Email: omar.flores@utec.edu.sv
†Direccion de Investigaciones
Universidad Salvadoreña Alberto Masferrer, El Salvador
Email:alberto.oviedo01@liveusam.edu.sv

***Abstract*—This paper presents the development of a personal language translation assistant, focused on the particularities of Salvadoran Spanish to English, using artificial intelligence (AI) and cloud computing. The technical and economic feasibility is analyzed, and Microsoft Azure is selected as the most appropriate platform for its implementation. The system integrates advanced natural language processing (NLP) and deep learning technologies to improve the accuracy and fluency of translations. The results highlight an innovative solution adapted to specific cultural contexts that positively impacts global communication and access to information.**

***Index Terms*—Automatic translation, artificial intelligence, cloud computing, natural language processing, deep learning.**

I. Introduction

Communication between people speaking different languages has been a persistent challenge in an increasingly globalized world. In a context such as El Salvador, where Salvadoran Spanish presents cultural and linguistic particularities, conventional translation tools often fail to capture the nuances necessary for accurate and contextualized interpretation. This work addresses these limitations by creating a personal translation assistant based on artificial intelligence (AI) and cloud computing, combining the best technologies to overcome linguistic and cultural barriers.

Progress in machine translation has been significant in recent decades, evolving from grammatical rule-based approaches to implementing neural models that use deep networks to improve fluency and accuracy. Although tools like Google Translate have popularized machine translation, these solutions are primarily designed for generalized uses and lack customization, especially for regional dialects and specific idiomatic expressions. On the other hand, cloud computing has revolutionized the access and processing of large volumes of data, enabling robust and scalable systems in critical applications such as translation.

Despite these advances, significant challenges remain in the field of machine translation. These include the difficulty of handling complex contexts, adapting to individual communication styles and translating cultural nuances and colloquial expressions. These challenges are particularly relevant in the case of Salvadoran Spanish, a dialect rich in idioms and syntactic particularities that are not considered by standard translation systems. Resolving these limitations improves the usefulness of these tools in local contexts and contributes to intercultural understanding and linguistic inclusion.

In this context, the presented project seeks to answer the question: How to develop a personal assistant that facilitates the language translation process (Salvadoran Spanish-English) using artificial intelligence and cloud computing? The central hypothesis is that a system based on deep learning, fed by Salvadoran dialect-specific data and supported by the scalability and real-time processing capabilities of cloud computing, can overcome the current challenges and provide a more accurate and contextualized solution.

This research is motivated by the need to democratize access to personalized translation tools, allowing linguistically diverse communities to benefit from artificial intelligence's advantages. This project provides a practical solution by solving machine translation's technical and cultural challenges. It advances the field of applied artificial intelligence, demonstrating how integrating specific data and advanced technology can improve the effectiveness and adaptability of modern systems. This work fills a critical gap in machine translation, offering a model that combines linguistic accuracy, cultural adaptability and technological scalability and can serve as a reference for future developments in this field.

This document is organized as follows: Section II describes the related work. Section III shows ... Section IV presents ... Finally, section V summarizes ...

II. Related Work

Automatic or machine translation has evolved significantly from initial grammar rule-based approaches to advanced deep learning models. This progress has been driven by advances in artificial intelligence, natural language processing (NLP) and cloud computing. The following reviews the most relevant work, highlighting its contributions and limitations. In the early years of this century, statistical machine translation (SMT) systems dominated the landscape [1]. These systems relied on probabilistic models trained on large bilingual corpora to generate more accurate translations than rule-based approaches [2] [3]. However, the quality of the translations was limited by the quantity and quality of the available training

data and by the inability of these models to handle complex linguistic contexts [4] [5].

A significant area of research is automatic post-editing (APE), which aims to enhance machine translation systems' outputs through machine learning-based techniques. Despite initial challenges due to neural systems' limitations, it has recently re-emerged with effective applications in commercial workflows [6] [7] [8]. Furthermore, cloud computing has facilitated the implementation of robust and scalable translation systems. Platforms such as Microsoft Azure and Google Cloud offer advanced NLP services that allow for the integration of translation models with real-time processing and distributed storage [9] [10].

Notwithstanding these advancements, considerable challenges persist. The current systems cannot adequately address cultural and linguistic context nuances, particularly in regional dialects such as Salvadoran Spanish. Moreover, personalization represents a crucial domain for enhancement, as most models are trained on generalized data that does not consistently reflect the specific nuances of end users [11].

Despite notable advancements in automatic translation, significant challenges remain in personalization and cultural understanding. This project aims to address these limitations by developing a personal assistant that combines the capabilities of AI and cloud computing to deliver tailored and accurate translations. This approach contributes to existing research by demonstrating how these technologies can be integrated to address specific challenges in multilingual contexts. This work differs from previous proposals by focusing on a specific context: the translation of Salvadoran Spanish into English, considering this dialect's unique linguistic and cultural particularities. Unlike existing generalist systems, this work integrates Salvadoran dialect-specific data into neural machine translation models and uses cloud computing to offer a scalable and personalized system.

III. Methodology

Methodology This AI—and cloud-based personal translation assistant was developed following a methodology comprising four key stages: requirements analysis, technological platform selection, system design and implementation and validation and testing. Each stage was designed to ensure that the system met the project's specific requirements, including linguistic accuracy, scalability and customization.

A. Requirements Analysis

In this stage, the requirements for translating from Salvadoran Spanish to English were identified, including idiomatic expressions and idioms and distinctive cultural characteristics. A representative linguistic corpus was assembled to train the translation models, with particular attention paid to selecting texts that exhibited vocabulary and grammatical structures that faithfully reflected the Salvadoran context. Furthermore, this analysis investigated the constraints inherent to existing machine translation systems, such as Google Translate, concerning this particular dialect.

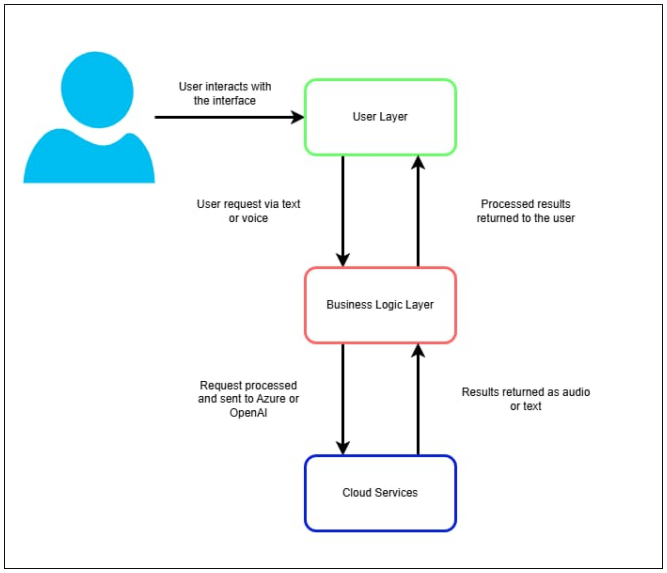

Fig. 1: Text to voice translation workflow of the developed application.

B. Technology Platform Selection

After evaluating several cloud computing options, including Amazon Web Services, Google Cloud and Microsoft Azure, the latter was selected due to its optimal balance between cost, scalability and integrated artificial intelligence tools. Microsoft Azure offered advanced services for developing natural language processing (NLP) models, including the Cognitive Services Translator and Azure Machine Learning, which facilitated the integration and customization of the system.

C. System Design and Implementation

The system architecture includes three main components: User Interface: An intuitive web application that allows users to enter text or voice for translation. NLP Module: This component processes linguistic inputs using neural translation models optimized with data specific to the Salvadoran dialect. Cloud Infrastructure: Manages data storage and processing, ensuring high availability and real-time performance.

The design was based on agile methodologies, allowing for rapid iterations and adaptations in response to user feedback during development.

D. Validation and Testing

The system was evaluated through tests with Salvadoran users, which demonstrated that the translations were accurate and that the application was user-friendly.

IV. Results and Discussion

A. Results

The Salvadoran Spanish-to-English translation application has been successfully installed on a local server, thereby ensuring efficient accessibility and performance. Figure 2 illustrates the user-friendly interface of the application, designed with intuitive navigation in mind. Figure 3 illustrates the

application's diverse speech-to-text and text-to-speech translation capabilities, showcasing its adaptability in addressing various communication requirements. A preliminary testing phase was conducted to evaluate the application's functionality. The results of this phase are visually represented in Figure 3, which shows translations that incorporate authentic Salvadoran Spanish expressions. These results demonstrate the effectiveness of the application in capturing the nuances of the local language.

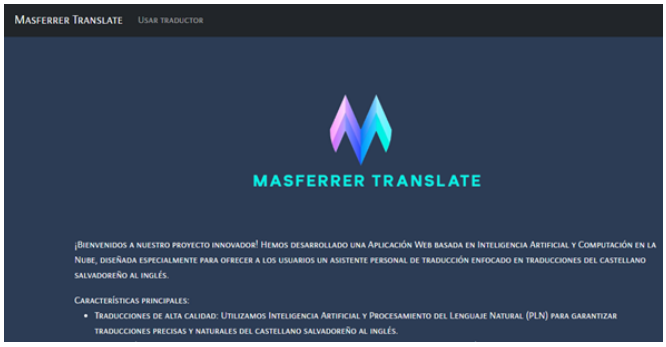

Fig. 2: Main screen of the developed translation application.

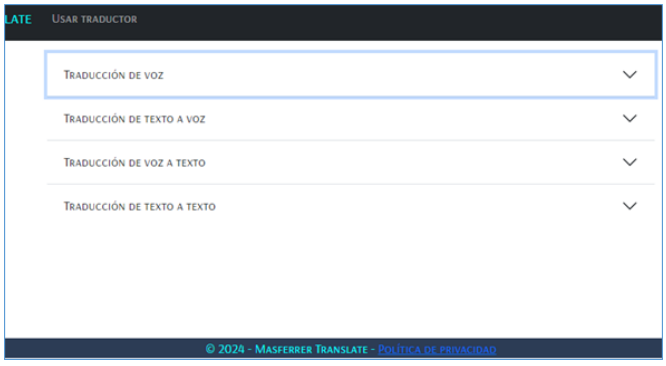

Fig. 3: Translate options of the developed translation application.

B. Discussion

These results highlight the importance of integrating specific data and advanced technologies to overcome the limitations of generalist translation systems. Although the system meets the stated objectives, there are still areas for improvement, such as incorporating an even more extensive corpus and expansion to other dialects of Spanish. In addition, continuous feedback from users could further optimize the system's adaptability and accuracy in future developments.

The potential impact of this system transcends the technical field. In an environment such as El Salvador, where a significant part of the population seeks to improve their job and educational opportunities by learning English, this assistant can become an essential tool to foster linguistic inclusion. This is particularly relevant in sectors such as tourism and export, where cultural and linguistic understanding is key to establishing effective business relationships.

The Salvadoran environment also presents specific technological challenges, such as limited connectivity in rural areas. Although integration with Microsoft Azure allows for robust performance under ideal conditions, the system must consider optimization strategies to operate efficiently under low connectivity conditions. Implementing offline features or lightweight models could be an area of future development to ensure accessibility in communities with limited infrastructure.

V. Conclusions

The development of a personal translation assistant based on artificial intelligence and cloud computing has proven to be an effective solution to address the specific needs of Salvadoran Spanish. By integrating linguistic and cultural data unique to the dialect, this system overcame the limitations of generic translation tools by providing more accurate and contextualized results. Its ability to handle idioms and cultural expressions, as well as its real-time performance through the use of Microsoft Azure, highlights its potential to foster linguistic inclusion and improve intercultural communication.

Furthermore, this project contributes significantly to machine translation by demonstrating how model personalization can overcome linguistic and cultural barriers. The results validate the initial hypothesis and reinforce the importance of considering regional contexts in the design of translation technologies. This approach can be a reference for similar initiatives in other dialects and languages with distinctive cultural characteristics.

Despite the achievements, this project has key areas for improvement and expansion. First, it is recommended that the language corpus be expanded, incorporating a greater diversity of texts and expressions from Salvadoran Spanish further to improve the accuracy and adaptability of the system. It is also suggested that the assistant's functionality be extended to other dialects of Spanish and indigenous languages in the region, broadening its impact on underserved communities.

Another area for improvement is optimizing the system to operate under conditions of limited connectivity, which is critical in rural areas. Developing offline versions of the system through compressed models or hybrid solutions could ensure its accessibility to a broader population.

Integrating advanced functions such as multimodal translation (text, voice and images) and dynamic personalization based on the user's style and preferences would also be a significant advance. These additions would not only improve the user experience but also position the system as a complete and adaptable solution for global and local contexts.

References

[1] H. Mercan, Y. Akgün, and M. C. Odacıoğlu, "The evolution of machine translation: A review study," *International Journal of Language and Translation Studies*, vol. 4, no. 1, pp. 104–116, 2024.

[2] K. Kumar, A. Faiz, K. Shruti, M. Chhabra, and S. Setia, "Translation systems: A synoptic survey of deep learning approaches and techniques," in *2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)*. IEEE, 2024, pp. 1–6.

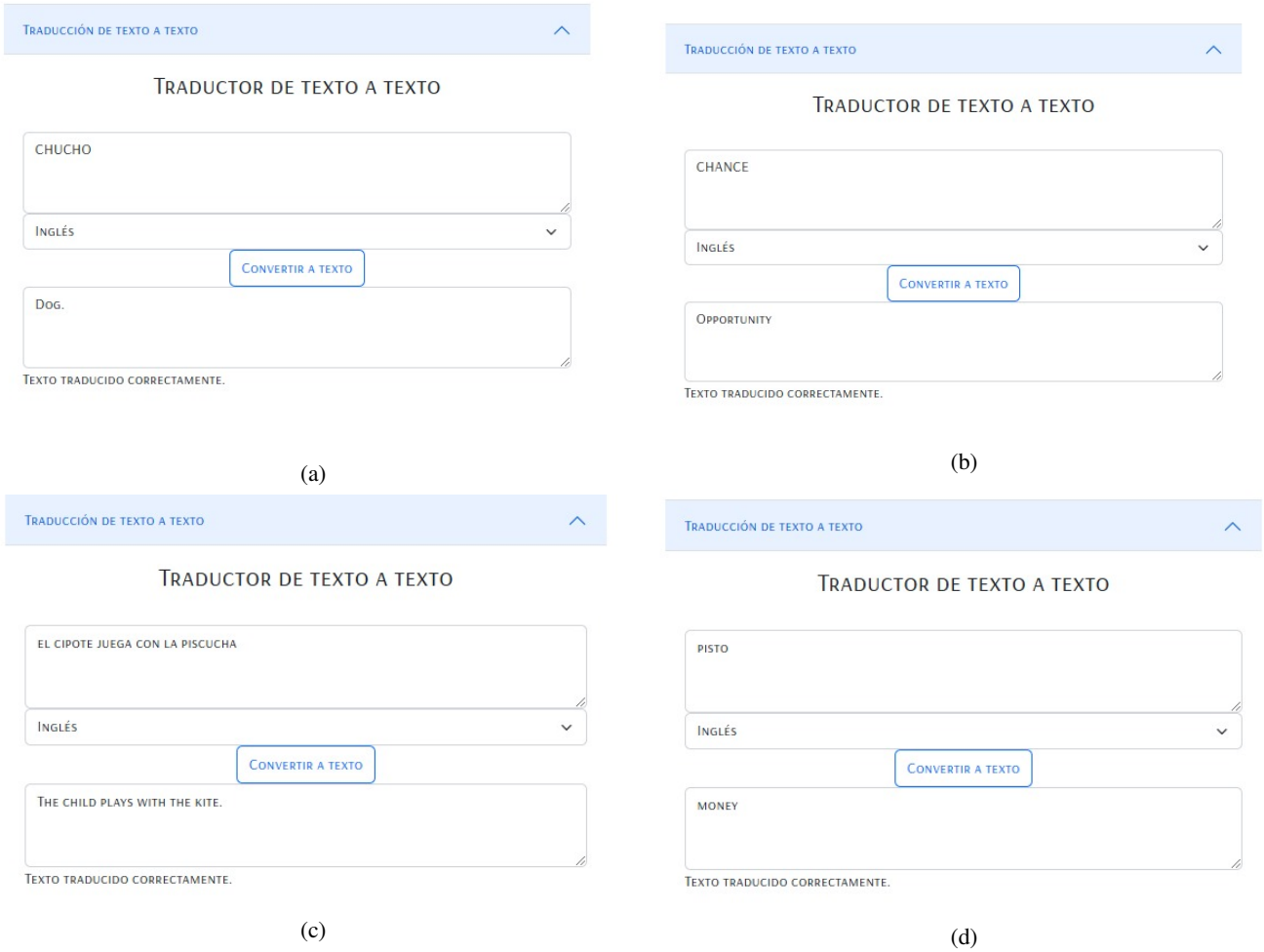

Fig. 4: The figures present a variety of translations of Salvadoran Spanish words and expressions, accompanied by their corresponding English translations.

[3] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "Bleu: a method for automatic evaluation of machine translation," in *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ser. ACL '02. USA: Association for Computational Linguistics, 2002, p. 311–318. [Online]. Available: https://doi.org/10.3115/1073083.1073135

[4] H. Hassan, A. Aue, C. Chen, V. Chowdhary, J. Clark, C. Federmann, X. Huang, M. Junczys-Dowmunt, W. Lewis, M. Li, S. Liu, T.-Y. Liu, R. Luo, A. Menezes, T. Qin, F. Seide, X. Tan, F. Tian, L. Wu, S. Wu, Y. Xia, D. Zhang, Z. Zhang, and M. Zhou, "Achieving human parity on automatic chinese to english news translation," 2018. [Online]. Available: https://arxiv.org/abs/1803.05567

[5] M. Shubham, "Breaking language barriers: Advancements in machine translation for enhanced cross-lingual information retrieval," *J. Electrical Systems*, vol. 20, no. 9s, pp. 2860–2875, 2024.

[6] F. do Carmo, D. Shterionov, J. Moorkens, J. Wagner, M. Hossari, E. Paquin, D. Schmidtke, D. Groves, and A. Way, "A review of the state-of-the-art in automatic post-editing," *Machine Translation*, vol. 35, pp. 101–143, 2021.

[7] L. I. Omar and A. A. Salih, "Systematic review of english/arabic machine translation postediting: Implications for ai application in translation research and pedagogy," in *Informatics*, vol. 11, no. 2. MDPI, 2024, p. 23.

[8] H. Zhao, Y. Liu, S. Tao, W. Meng, Y. Chen, X. Geng, C. Su, M. Zhang, and H. Yang, "From handcrafted features to llms: A brief survey for machine translation quality estimation," *arXiv preprint arXiv:2403.14118*, 2024.

[9] S. Guo, "Intelligent english automatic translation system based on human-computer interaction enhancement algorithm," in *2023 Asia-Europe Conference on Electronics, Data Processing and Informatics (ACEDPI)*. IEEE, 2023, pp. 465–469.

[10] K. K. Naveen, V. Priya, R. G. Sunkad, and N. Pradeep, "An overview of cloud computing for data-driven intelligent systems with ai services," *Data-Driven Systems and Intelligent Applications*, pp. 72–118, 2024.

[11] M. Tsujikawa, K. Okabe, K. Hanazawa, and Y. Kajikawa, "Automatic speech translation system selecting target language by direction-of-arrival information," in *2018 26th European Signal Processing Conference (EUSIPCO)*. IEEE, 2018, pp. 2315–2319.