Latin American Business and Sustainability Review

Academic Journal

LABSREVIEW V1 N2 2024
ISSN: 2995-6013
https://doi.10.70469/labsreview.v3i1

An official publication of
ALBUS
Academy of Latin American Business and Sustainability Studies

Latin American Business and Sustainability Review
LABSREVIEW | V2| N2| 2024

ISSN: 2995-6013
https://doi.10.70469/labsreview.v3i1

Table of Contents

Latin American Business and Sustainability Review
Academic Journal

Editorial

Human-Centered Intelligence: Technology, Inclusion, and Sustainability in a Transforming Society

Technological innovation is reshaping how societies learn, communicate, make decisions, and respond to complex environmental challenges. Artificial intelligence, machine learning, assistive technologies, and automated language systems now influence activities ranging from classroom instruction and urban environmental management to bilingual literacy and international destination marketing. Yet the significance of these technologies cannot be assessed solely through their technical performance. Their academic and social value ultimately depends on whether they enhance human capabilities, expand accessibility, respect cultural meaning, and contribute to sustainable development.

The four articles in this issue of *Latin American Business and Sustainability Review* (LABSREVIEW) examine this broader transformation from distinct but complementary perspectives. Together, they suggest that innovation should not be understood as the simple adoption of increasingly sophisticated tools. Instead, meaningful innovation requires the deliberate integration of technology with pedagogical principles, environmental responsibility, linguistic sensitivity, and human judgment. This issue is therefore organized around the theme of human-centered intelligence: the design and application of intelligent systems in ways that remain responsive to people, communities, cultures, and the environments in which they operate.

This theme is particularly relevant for Latin America and other culturally and linguistically diverse regions. Although emerging technologies can help address persistent educational, environmental, and economic challenges, their implementation also raises important questions. Who benefits from technological innovation? Whose knowledge, language, and cultural meanings are represented? How can accessibility be incorporated from the beginning rather than treated as an afterthought? When should automated outputs be accepted, and when is expert human intervention indispensable? The contributions in this issue engage with these questions across education, environmental science, bilingual literacy, and intercultural communication.

The issue begins with Melissa Fiori's article, "Social Constructivism, UDL, and Instructional Design in the AI-Enhanced Spanish L2 Classroom: Operationalizing Theoretical and Practical Implementation." The article addresses a central concern in contemporary higher education: how artificial intelligence can be incorporated into teaching without displacing either learner agency or the educator's pedagogical role. Fiori integrates social constructivism, Universal Design for Learning (UDL), and instructional design into a conceptual model for AI-enhanced second-language education. Social constructivism provides the theoretical foundation for understanding learning as a socially mediated process, while UDL offers a practical framework for designing flexible and accessible learning experiences.

Rather than treating artificial intelligence as an autonomous instructor, the article positions it as a pedagogical resource whose value depends on thoughtful instructional mediation. The educator remains responsible for establishing learning objectives, designing meaningful interactions, evaluating outputs, and guiding students' critical engagement with AI-generated content. This contribution is especially important at a time when educational discussions often oscillate between uncritical enthusiasm for artificial intelligence and calls for its restriction. Fiori offers a more productive position: AI can support accessible, collaborative, and engaging learning, but only when its use is grounded in sound learning theory and deliberate instructional design.

The second article moves from educational innovation to urban environmental sustainability. Daniel Leonardo Gonzalez Torres, María Isabel David Otalvaro, and Ricardo Andres Santa Quintero present "Multi-Output LSTM-Based One-Step-Ahead Hourly Forecasting of $PM_{2.5}$ and PM_{10} at an Urban Station in Bogotá: A Two-Year Performance Analysis." Their study develops a deep-learning model to forecast hourly concentrations of two major atmospheric pollutants at the Fontibón monitoring station in Bogotá, Colombia.

Using a multi-output Long Short-Term Memory network and a seven-day sliding observation window, the authors demonstrate the capacity of machine learning to identify nonlinear temporal patterns in urban air-quality data. The model explains approximately 70% of the variance during the validation period, providing a reproducible proof of concept for the development of urban early-warning systems. The contribution extends

beyond predictive accuracy. Reliable short-term forecasts can support public-health advisories, environmental policy, institutional planning, and more responsive urban management. In this context, intelligent systems become instruments of environmental governance. The study demonstrates how data-driven innovation can contribute to healthier and more sustainable cities, particularly in high-altitude urban environments where atmospheric and meteorological conditions create distinctive pollution challenges.

The third article, "Accessibility Meets the Science of Reading in Spanish English Dual Language Classrooms," by Irene Oramas Sosa and Carlos Zarzalejo, returns to education but addresses it through the intersecting perspectives of bilingual literacy, accessibility, and educational equity. The authors introduce the concept of inclusive biliteracy, defined as the coordinated development of literacy in Spanish and English within instructional systems designed to be equitable from the outset.

Their framework brings together four dimensions: structured literacy, cross-linguistic development, UDL-based accessibility, and technology-supported instruction. This integration is significant because multilingual learners, and particularly multilingual learners with disabilities, are often served through fragmented educational approaches. Literacy development, bilingual education, accessibility, and educational technology may be treated as separate areas, even though they converge in the lived experiences of students. By connecting these domains, the article presents accessibility as an infrastructural condition of effective education rather than a supplementary accommodation. The authors also connect inclusive biliteracy with equity and human-capital development. Accessible bilingual instruction expands opportunities for learners to participate fully in education and society while preserving the cognitive, social, and cultural value of multilingualism. Importantly, the framework generates testable propositions that can guide future empirical research. It therefore provides both a conceptual foundation and a research agenda for examining how inclusive biliteracy can be implemented and evaluated across educational settings.

The issue concludes with Pamela Doran and Carlos Zarzalejo's "Critical MT Praxis: Machine Translation and Tourism Slogans in Latin America." This article examines a less visible but increasingly consequential dimension of automated intelligence: its mediation of cultural meaning. Tourism slogans are not merely informational statements. They are condensed expressions of identity, emotion, and destination positioning intended to create an affective relationship with prospective travelers. When these slogans are processed through automated translation systems, lexical accuracy alone cannot guarantee communicative effectiveness.

The authors analyze 40 Latin American tourism slogans using Google Translate, DeepL, and Microsoft Translator and compare their outputs with human translations. Although 52.5% of the machine-generated translations achieved high fidelity, the analysis identifies recurring problems involving literalism, tonal flattening, lexical imprecision, and culturally misleading cues. DeepL generally produced the most natural results, but none of the systems consistently preserved the full persuasive and emotional force of the original slogans.

Through their proposed Critical MT Praxis framework, the authors demonstrate that machine translation must be assessed not only in terms of semantic correspondence but also through its cultural and affective consequences. This finding has direct implications for tourism marketing, destination branding, and intercultural business communication. Automated systems can facilitate multilingual communication, but human linguistic and cultural expertise remains essential when communication depends on emotion, identity, and persuasion.

Collectively, the four articles show that technological effectiveness and human-centered design are not competing priorities. They are mutually dependent. Artificial intelligence in the classroom requires instructional mediation; environmental forecasting requires connection to public decision-making; bilingual education requires accessibility by design; and machine translation requires cultural and affective evaluation. Across these contexts, technology becomes valuable when it supports human agency, expands participation, strengthens informed decisions, and responds to the particularities of language, place, and community.

This issue invites scholars and practitioners to move beyond asking whether intelligent technologies should be adopted. The more consequential questions concern how they should be designed, governed, interpreted, and integrated into social systems. By addressing these questions across disciplinary boundaries, the articles reinforce LABSREVIEW's commitment to research that connects innovation with sustainability, inclusion, and the realities of Latin America and its global relationships.

The future of intelligent systems will not be determined by computational capacity alone. It will depend equally on the educational, ethical, cultural, and institutional choices that guide their use. Human-centered intelligence offers a framework for making those choices responsibly—and for ensuring that technological progress contributes to more accessible, sustainable, and culturally responsive societies.

Dr. Luis José Camacho
Editor-in-Chief
SUNY Empire State University
USA

Latin American Business and Sustainability Review
Academic Journal

Theoretical

Social constructivism, UDL, and instructional design in the AI-Enhanced Spanish L2 classroom: Operationalizing theoretical and practical implementation

Melissa Fiori

Empire State University, USA, melissa.fiori@sunyempire.edu

Citation: Fiori, M. (2025). Social Constructivism, UDL, and Instructional Design in the AI-Enhanced Spanish L2 Classroom: Operationalizing Theoretical and Practical Implementation. *LABSREVIEW,* 3(1), 1-11. https://doi.org/10.70469/labsreview.v3i1.34.

Academic Editor: Jorge Martín Diez

Received: 11/2/2025
Revised: 1/15/2026
Accepted: 02/18/2026
Published: 30/3/2026

Abstract: Despite growing interest in AI-enhanced (language) learning, there remains a lack of conceptual models that demonstrate how AI can be pedagogically operationalized within established learning frameworks in ways that preserve learner agency and reaffirm the educator's role. Existing literature often emphasizes technological affordances or learner outcomes without articulation about how instructional design serves as a mediator. This paper posits that social constructivism offers the theoretical framework and UDL offers the practical framework that anchor the educator's role as critical to the AI-enhanced learning environment, when mediated by instructional design, and does so on the backdrop of interpersonal communication at the elementary level, Spanish as a second language (L2-Spanish), and higher education courses. It is a conceptual contribution that integrates theory through a design-based, illustrative example.

Keywords: Social constructivism, UDL, Universal Design for Learning, AI, ChatGPT, technology in higher education, instructional design

1. Introduction

How do Social Constructivism, Universal Design for Learning, and Artificial Intelligence work together to make learning more helpful, accessible, engaging, and sustainable? Social Constructivism (SC) is a learning theory that posits that learners learn best through social interactions mediated by experts (Vygotsky, 1979). Universal Design for Learning (UDL) principles center on designing lessons accessible to all learners (CAST, 2025). AI is a tool that can align with both frameworks to support educators and learners alike. By using ChatGPT in a UDL-centered classroom, educators set the stage for the collaborative learning that Social Constructivism encourages while remaining anchored in the process. While the alignment among Social Constructivism, UDL, and AI is readily imagined, the literature has yet to provide clear models for how this alignment can be pedagogically operationalized in instructional practice. Despite the rapid growth of scholarship on artificial intelligence in (language) education, much of the existing literature emphasizes technological affordances, learner outcomes, or ethical concerns without sufficiently articulating how AI can be pedagogically operationalized within established learning frameworks. In particular, there is a lack of theoretically grounded, design-oriented models that demonstrate how AI can be integrated into instructional practice in ways that preserve learner agency and reaffirm the educator's role. This paper is a conceptual piece that addresses that gap by advancing Social Constructivism as the theoretical framework and Universal Design for Learning as the practical framework through which AI-enhanced instruction can be intentionally designed and mediated. Using an interpersonal communication task in the second language Spanish classroom as an illustrative example, this paper demonstrates

how these frameworks can be operationalized through instructional design, while positioning the educator as an essential facilitator of learning. It further discusses the need for institutional investment in professional development for educators and the support personnel offices (Centers for Teaching and Learning, Instructional Design, Instructional Technology, Accessibility Officers) that are the cornerstones of 21st-century education.

2. Literature Review

Artificial Intelligence refers to computer programs designed to mimic human intelligence. Generative AI is a specialized subset of AI that uses large datasets to generate new content. Large Language Models (LLMs) are a subset of generative AI systems that produce human-like output based on the datasets they are trained on. Educational applications of AI have enabled gamified learning tools such as Kahoot!, Quizlet, and Duolingo; auto-generated practice, scoring, and adaptive learning options that textbook publishers now offer; planning assistant apps and rubric generators; and grammar-assist tools, among others. ChatGPT, My Conversation Trainer, and CallAnnie offer conversational practice, while Google Arts & Culture invites virtual field trips, city and museum tours, and free cultural exploration.

Studies on AI and Generative AI's application in language learning are abundant and build on decades of technology-mediated language learning research (Bax, 2003; Fiori, 2005; Hubbard, 2008; Warschauer, 1996, 2000;). While AI offers immediate, personalized feedback (on writing) (Alharbi, 2023; Obaidoon & Wei, 2024), which can result in learning autonomy and self-regulation (Ng et al., 2024; Wu et al., 2025), there are concerns about overreliance (Ducar & Schocket, 2018) and the quality of the output learners were exposed to (Niño, 2009; Sharma et al.,2025). AI can scaffold reading tasks to improve comprehension, thereby reducing task anxiety (Yuan, 2025) and promoting emotional well-being (Rezai et al., 2024), but there are concerns about the distinction between shallow and deep learning (Kos'Myna, 2025). AI can expand learning opportunities beyond the classroom (Crompton et al., 2024; Hockly, 2023; Ji et al., 2022), thereby increasing equity and inclusion, a focal point of technology access for social justice (Dooly & Comas-Quinn, 2025). However, not all learners have access to internet resources, which can contribute to unequal access to learning (Hockly, 2023). AI has the potential to help to improve academic performance (Deng et al., 2025; Zhou, 2023), and L2-speaking skills (Hao & Min, 2025), but with increased use are increased concerns about transparency and ethics surrounding that use (Crompton et al., 2024; Dwivedi et al., 2023; Hockly, 2023; Vinall et al., 2023; Zhu & Wang, 2025), concerns about content accuracy, quality, and reliability (Lee, 2025; Dwivedi, 2023), and a stated need to integrate AI literacy (Lee et al., 2025) into learning contexts.

Benefits for instructors include, but are not limited to, support (Ji et al., 2022), educational efficiency (Delello et al., 2025), innovation in task design (Pérez-Núñez, 2024; Amin, 2023), and improved feedback efficiency and consistency (Amin, 2023). However, there are fears about the role of the educator (student-teacher relationships) in AI-enhanced learning environments (Delello et al., 2025), the long-term pedagogical impact of introducing such tools (Crompton et al., 2024; Lee, 2025; Zhu & Wang, 2025), and a need for training (Kohnke et al, 2025).

Concerns about educational technology are nothing new, however. From resistance to the use of overhead projectors (Allen, 1963) to microcomputers (Hannafin et al., 1993) to graphing calculators (Fisher, 1995), educators have long feared the impact of technological advancements on teacher autonomy and professional identity. Nonetheless, roles have shifted from the educator as the knowledge transmitter to the educator as the facilitator and mediator. As early as the 1980s and 1990s, student-centered approaches were taking shape (Barr & Tagg, 1995; Chickering & Gamson, 1989; King, 1993), with empirical studies in the early 2000s reinforcing the value of student-centric teaching methodologies (Bosworth, 2013; Prince, 2004; Weimer, 2012).

In the era of the AI-infused (language) classroom, instructors are increasingly framed as interpreters of technological affordance (Crompton et al., 2024). While AI can support classroom operations such as assessment (Amin, 2023), writing (Alharbi, 2023; Ducar & Schocket, 2018; Niño, 2009; Obaidoon & Wei, 2024), reading comprehension (Yuan, 2025) and well-being (Rezai et. al, 2024; Yuan, 2025), learner engagement (Wu et al., 2025) and self-regulated learning (Ng et al., 2024), in addition to speaking practice (Zou et al., 2025; Leis, 2025), instructors retain critical judgment and remain central to the educational experience. Researchers consistently report on the need for educators to be adaptable in the face of technological advancements (Allen, 1963; Ducar & Schocket, 2018; Warschauer, 2000), in order to shape technological integration (Bax, 2003), scaffold experiential learning (Ji et al., 2022), and align with the principles of critical pedagogy.

3. Frameworks: Social Constructivism & UDL

Social Constructivism (SC) is a theoretical framework that holds that knowledge is constructed through social interactions in specific contexts. It frames cognitive development as a socially mediated process where interaction with a more capable, knowledgeable peer facilitates movement through the zone of proximal development (ZPD), the space where a knowledge gap is identified and is ripe for amelioration and favorable to intervention, thereby helping the learner move from other-regulated (controlled by the task) to self-regulated.

With scaffolded, strategic support, learners can bridge the gap between what they can achieve with guidance and what they can accomplish independently.

Universal Design for learning (UDL) is an educational framework centered around "learner agency that is purposeful & reflective, resourceful & authentic, strategic & action-oriented" (CAST, 2024). UDL guides learner development through three main principles: Multiple Means of Representation, the "what" of learning, which focuses on how information is processed and perceived consequently emphasizes offering learners different avenues to acquiring information; Multiple Means of Engagement, the "why" of learning, which focuses on motivation, interest, and persistence, and consequently emphasizes challenging learners and motivating them to learn; and Multiple Means of Action & Expression, the "how" of learning, which focuses how learners demonstrate what they know, and consequently emphasizes offering choice with respect to how learners demonstrate acquired knowledge. Consult the CAST UDL Guidelines on the CAST website for a deep dive into their supporting principles (https://udlguidelines.cast.org).

Both frameworks emphasize active, student-centered, collaborative learning, rather than receptivity. Social constructivism holds that knowledge is created (constructed) through social engagement, and UDL encourages collaboration in pairs, small groups, and as a class. Both frameworks emphasize the importance of scaffolding learning activities. Social Constructivism emphasizes providing strategic support within learners' ZPD as central to facilitating eventual learner autonomy (whereby skill gaps are addressed through guided, scaffolded interventions and feedback), whereas UDL emphasizes task sequencing, presenting information in a variety of ways, and providing timely feedback. Both frameworks center on learner agency and encourage learners to take ownership of their learning, recognizing that motivation to learn is linked to personal interests and goals, and that progress is connected to ongoing, meaningful assessment, in which learners have the liberty to choose how to demonstrate their acquired abilities. Naturally, a safe learning environment is a critical component for learning to blossom, and both frameworks recognize the importance of creating the supportive and welcoming space needed for social and emotional well-being and, subsequently, for learning to take root. Both frameworks embrace technological integration as instruments for enhancing social interaction, personalization, collaboration, and expression to personalize and enhance learning. Finally, both frameworks view the role of the educator not as a lecturer or knowledge transmitter, but rather as an expert facilitator and mediator. How do these aforementioned frameworks come to life in an AI-enhanced classroom?

4. Operationalizing SC and UDL in the AI-Enhanced L2 Classroom through Instructional Design

Operationalizing the theoretical and practical frameworks invites a robust, learner-centric experience, which involves the following: identifying measurable learning objectives, designing assessments, and creating learning activities that a) reflect the social context in which they will take place, b) account for learner variability, and c) leverage the tools that facilitate the process. Let us explore this through the lens of interpersonal communication (information exchange between two or more interlocutors) in the second language classroom on the topic of planning a vacation.

Step 1: Identify the desired results

First, identify the desired results. If the goal is to build interpersonal communication skills, then providing opportunities to a) negotiate meaning, b) adopt and adapt conversational strategies, and c) refine their ability to engage with an interlocutor is essential. If the objective is for the learner to plan a vacation, the central question concerns designing the context that facilitates this by considering what learners will be able to do and how they can apply it to contexts beyond the classroom.

Step 1 Example: Plan a vacation in L2 Spanish.

Step 2: Evidence of Learning

Then, establish what constitutes evidence of learning. In the context of planning a vacation, an authentic task, evidence might consist of the learner's ability to formulate and respond to questions involved in planning, and their ability to negotiate meaning with their interlocutor to reach mutual comprehension, manage trip planning, and synthesize elements of the exchange that help them reach their communicative objective.

Step 2 Example

Evidence of Learning: I can discuss and plan a vacation with a more capable peer (AI-interlocutor, the course professor, or an L2 speaker (friend, family, coworker, instructor).

Step 3: Activities and Assessments

Subsequently, determine what learning activities and assessments (formative and summative) best facilitate learning, measure progress, and measure achievement. Focus on application and construction of meaning that require learners to apply their knowledge to new challenges and true-to-life scenarios, rather than merely recalling facts.

Step 3 Example

Prior Knowledge: Students build vacation-focused lexical and grammatical knowledge through a series of activities (interpretive communication (reading, listening) and presentational communication (writing, speaking) that begin with comprehension checks, progress into short-answer responses, and advance into full production of short sentences. They engage with audio and video content that models the topic to build both linguistic and metalinguistic awareness, beginning with comprehension checks and progressing to the examination of how interlocutors negotiate meaning to reach mutual understanding. They will apply the rubric for interpersonal communication to the samples and identify conversational strategies used throughout the process. Infused into creating this background knowledge **is** social constructivism through observation and direct participation, and the UDL principles of Representation (strategic task scaffolding utilizing a variety of resources, examples of successful vacation planning, metacognitive analysis of negotiation of meaning tactics, and discussing strategies for successful communication) and Engagement (activating and building knowledge through meaningful activities).

Step 4: Rubrics

Rubrics that articulate the success criteria are valuable here, as they clarify for all parties what success (evidence of learning) entails and help anchor alignment between task design and assessment from the outset. In the context of vacation planning in the L2, the GPT can be prompted to provide scaffolded feedback based on a set of criteria, and the conversation log can serve as an informative and rich data set that reveals how the learner is negotiating meaning, adopting and adapting conversational strategies, and refining their ability to engage in interpersonal communication. This, in turn, invites opportunities for deeper, individualized, instructor-based feedback that paves the way for success.

Step 4 Example Rubric

Interpersonal Communication Rubric:

- Strong: Participates in and advances the exchange.
  - Responses are appropriate (simple and compound sentences or phrases).
  - Responses are a mix of practiced content and original productions
  - Negotiates meaning through rephrasing for clarification or asking questions (practiced and original)
  - Maintains appropriate register (formal, informal) for the context
- Good: Participates fully in the exchange.
  - Responses are appropriate (simple words and simple sentences).
  - Responses reflect practiced content
  - Negotiates meaning through word substitution for clarification or asking practiced questions
  - Partially maintains appropriate register (formal, informal) for the context
- Developing: Participates partially in the exchange.
  - Some appropriate responses (simple words and formulaic/chunked phrases).
  - Responses reflect memorization
  - Attempts to negotiate meaning through word substitution for clarification or simple questions (¿qué? ¿cómo? ¿repite(s)?
  - Minimally maintains appropriate register (formal, informal) for the context
- Emerging: Participates minimally in the exchange.
  - Some appropriate responses (simple words or lists of words).
  - Responses are either single-word responses or gestures
  - Attempts to negotiate meaning through English
  - Does not maintain appropriate register (formal, informal) for the context

Step 5: Build your Scaffolds

Next, configure the experience. Activate prior knowledge to connect with what the learner brings to the table through prior instruction and lived experience, to provide context for learning, and to increase motivation (UDL

engagement). Chunk and scaffold information (UDL-representation) around social interaction and collaboration (SC), leverage the technology that reinforces the process (SC, UDL), and guide learners (rather than lecture).

Step 5 Example

The Interpersonal Communication Experience:

Students build interpersonal communication (information exchange between two or more interlocutors) skills by conversing with AI (Learner-AI), peers (Learner-Learner), and native/heritage/proficient speakers of the language studied (Learner- L2Speaker), where they have graduated levels of support enabling them to practice communicating in a way that reflects their interest in the topic (UDL-Action & Expression).

Learner-AI

First is a learning activity between the learner and the AI, in which the learner can respond to the AI-interlocutor's feedback and incorporate it into subsequent attempts.

GTP Student Instructions: You are meeting with your travel agent to plan a trip. To begin, type (or say), "Quiero planear un viaje" (I want to plan a trip). The AI-travel agent will ask you questions. Answer in Spanish as best you can. Do not worry about mistakes because the agent will give you feedback and help you improve. You can repeat this activity as many times as you would like. Each time, the questions may be slightly different, providing additional practice with the topic and with incorporating feedback into the exchange. Click on the link to get started: Agente de Viajes.

This activity can be followed up in various ways. The instructor can review the transcript to provide individualized feedback; peers can review the transcripts and recommend ways to respond to the feedback; peers can observe their peer when they engage with the AI-travel agent and coach the conversation from the sideline; peers can examine how meaning was negotiated and identify what learning strategies were useful for advancing the conversation; peers can review the exchanges against the grading criteria and brainstorm ways to make improvements and/or identify or expand on their personal goals for the activity type. At this point, learners may return to the GPT for additional practice, or they might dive into peer conversations.

Learner-Learner

Next is the opportunity to measure progress through peer-to-peer conversations. Role-plays in which learners take turns as travel agents and travelers provide formative assessment opportunities with feedback from both the instructor and peers. Learners practice with assigned partners until they're ready to work with a randomly assigned partner, seeking feedback along the way.

Learner-L2 Speaker

Communication culminates with a conversation (summative assessment) and a reflection. Learners select the interlocutor with whom they believe they'll have the best chance to demonstrate their ability to plan a vacation in the L2 (UDL-Action & Expression). They may choose from among the AI-interlocutor, the course professor, or an L2 speaker (friend, family, coworker, instructor). With or directly following their submission, they submit a reflection (UDL-Engagement) on the learning process and assess themselves against the grading criteria.

Step 6: Reflection

Finally, encourage learners to engage in metacognitive reflection beyond what they completed during the AI-Learner and Learner-Learner activities, through self-assessment (see Appendix C for ideas), as well as through task reflection and refinement for future variations of the activity by educators. Educators can create their own reflection forms or consider using preexisting tools such as Waterloo's Reflection Framework and the DEAL model, purchasing access to Penn State's CCEDIR, or adopting education-based AI tools, such as MirrorTalk or Perplexity.

5. Discussion: Theoretical, Pedagogical, and Institutional Implications

5.1 Theoretical Implications: Social Constructivism and UDL as Anchors for AI Integration

At the intersection of SC and UDL is the shared understanding that learning is socially mediated, scaffolded, and supported through intentional design. Both frameworks advocate learner agency, focus not on how a task is completed but on supporting the journey to realization, and encourage the intentional inclusion of tools that account for variability while supporting meaningful engagement. When applied to AI-enhanced instructional contexts, these frameworks position artificial intelligence not as an autonomous instructional agent but rather as a mediating tool embedded within educator-designed learning experiences. Educators are central to education;

learning is neither passive nor automated, and the human aspects of teaching are not easily replicated by artificial intelligence. From a Social Constructivist perspective, learning occurs within the zone of proximal development through interaction with more capable peers or experts, with scaffolding gradually withdrawn as learners move toward self-regulation. UDL complements this process by emphasizing flexibility in representation, engagement, and action and expression, ensuring that learners have multiple pathways to access content and demonstrate understanding. Together, these frameworks provide a coherent theoretical foundation for integrating AI in ways that support, rather than supplant, human mediation, thereby addressing persistent concerns about the erosion of instructional authority in AI-enhanced classrooms. Recent publications emphasize that AI achieves its greatest pedagogical value when implemented within instructor-mediated, scaffolded designs that center on intentional instructional strategies (Ji et all, 2022), and where educators and learners demonstrate an adaptive mindset (Zhu & Wang, 2025), ensuring that AI complements rather than replaces human facilitation.

5.2 Pedagogical Implications: The Educator's Role in AI-Enhanced Instruction

Arguably, concerns about educators' roles in the face of artificial intelligence have more to do with the shift from knowledge transmitters to facilitators and mentors. Many educators enter higher education as subject-matter experts (SMEs) who have not received formal training in the science of teaching. Often, promotion and tenure focus on research and publication, and on institutional service, rather than on the ability to foster dynamic, engaging andragogical environments. Assuming that the subject matter expert is inherently a skilled educator is flawed and has implications for student success and retention. Framing AI within Social Constructivist and UDL principles reframes concerns about the educator's role in technology-infused classrooms. While artificial intelligence excels at pattern recognition, scalability, and efficiency, it lacks the capacity for sociocultural interpretation and professional judgement that educators bring to instructional contexts. Although AI-generated feedback can approximate human feedback in some domains (Kaliisa, et al., 2025) learner engagement and self-regulated learning are strengthened when AI use is embedded within intentional pedagogical design (Ng et. al., 2024; Wu et al., 2025), underscoring the centrality of the educator and educator judgement. These human dimensions of teaching are central to scaffolding learning, interpreting learner needs, and responding to the affective elements that shape educational experiences. The instructional example presented in this paper highlights the pedagogical labor involved in AI-enhanced instructional design. Behind seemingly simple learner-facing activities lies an iterative process of development, testing, refinement, and alignment with learning objectives and assessment criteria. This design work reflects the educator's role as facilitator, mediator, and designer, roles that are amplified rather than diminished by AI integration. For educators unfamiliar with Social Constructivist or UDL frameworks, engaging in this design process may present challenges, underscoring the importance of pedagogical preparation alongside technological fluency. This aligns with calls for AI literacy development among educators (Kohnke et al., 2025) and longstanding CALL scholarship emphasizing pedagogy over tool adoption (Bax, 2003; Hubbard, 2008).

5.3 Institutional Implications: Supporting Design-Based, AI-Enhanced Pedagogy

The design-intensive nature of today's classrooms, especially those with AI-enhanced instruction, carries important implications for institutions of higher learning. Effective integration of AI requires more than access to tools; it demands sustained professional development and institutional support structures that enable educators to design, implement, and refine pedagogically sound learning experiences. Professional development and institutional support are critical to effective AI integration (Crompton et al., 2024; Zhu & Wang, 2025), particularly given documented educator uncertainty surrounding adoption (Parviz & Arthur, 2025). Investment in professional development should be recognized and rewarded in ways comparable to research and service, acknowledging the expertise required to facilitate inclusive, learner-centered instruction and to integrate AI tools. Consider differentiating between research and teaching lines: research lines focus on grant acquisition and publication, whereas subject-matter experts with classroom assignments are required to actively engage in and document teaching-focused CEUs (continuing education units) through reputable organizations such as the Online Learning Consortium or Quality Matters. Support continuing education through funding and credit reassignment; 150 hours of documented continuing education through approved programs can earn a 3-credit release, with the expectation that the education is completed within a time frame, or conversely, every specified period of time SMEs would be granted a 3-credit release for focused training, with the expectation of x-hours of ongoing continuing ed per semester or academic year; or even require professional development hours be scheduled in the same manner that office hours are held.

In addition to educator-centric professional development, institutions must invest in support offices, such as the Center for Teaching and Learning (CTL), Instructional Design, Instructional Technology, and Accessibility Specialists. These offices play a critical role in supporting educators as they navigate pedagogical frameworks, accessibility standards, and emerging technologies, consistent with research on UDL and accessible online

learning that emphasizes collaboration among faculty, instructional designers, and accessibility specialists (Behling & Tobin, 2018; Lowenthal & Lomellini, 2023). Institutional commitment to these support offices signals recognition that effective AI integration is a pedagogical endeavor rooted in human expertise, collaboration, and intentional design.

6. Conclusion

As artificial intelligence becomes increasingly embedded in (language) education, educators face a persistent challenge: how to integrate these tools in ways that enhance learning without undermining learner agency or diminishing the educator's pedagogical role. From the SC perspective, educators are responsible for scaffolding (the instructor creates the conceptual framework that AI plugs into), facilitating learning in the ZPD (AI can provide practice, feedback, and adaptive learning), curating tools (such as AI), and fostering metacognition (AI strengthens internalization). Through the lens of UDL, the educator's role is to recruit interest, help learners persist, and facilitate learner agency (Engagement); curate resources and create meaningful contexts for learning (Representation); and encourage creative, student-driven demonstrations of achievement (Action & Expression). AI offers numerous options that align with UDL principles. At the foundation of both frameworks, however, is the educator. When the educator is at the helm as a facilitator, concerns about AI overreliance, diminished pedagogical readiness, shallow learning, cognitive load, transparent and ethical use, and related issues are rendered negligible in the classroom. The educator interprets and designs learning with human values, empathy, and context. AI scales and adapts support, provides multimodal access, and reduces administrative burden, while UDL ensures flexibility and inclusion in the shared learning space where human mediation and AI scaffolding converge (ZPD). This paper contributes a theoretically grounded, design-oriented model that integrates Social Constructivism and UDL as complementary frameworks for AI-enhanced instruction. By illustrating how these frameworks can be operationalized in an L2 Spanish interpersonal communication task, the paper demonstrates that AI can serve as a scaffolded resource in educator-mediated learning environments. Although the example presented is situated within an L2-Spanish classroom, the framework integration and design principles advanced here are transferable across disciplines and instructional contexts where educators seek to align emerging technologies with inclusive, learner-centered pedagogy. Ultimately, Social Constructivism and UDL not only anchor the educator's role in AI-enhanced environments but also reveal it as more critical than ever. As AI continues to evolve, pedagogical frameworks that foreground human mediation, intentional design, and learner variability will be essential to ensuring that technological innovation serves educational purposes.

Author Contributions: Melissa Fiori

Funding: No funding was received to conduct this research.

Institutional Review Board Statement: not applicable.

Informed Consent Statement: not applicable

Acknowledgments: not applicable

Conflicts of Interest: There are no known conflicts of interest.

Appendices

A. Instructor Preparation – Custom GPT Generation

1. Register for a paid plan with ChatGPT.
    a. Log into your account, click on your profile icon to upgrade your plan.
2. Fill in the basics: activity name; description, and category.
3. Enter your prompt and test it. Note what is working well and not yet working and refine it by modifying and testing.
4. Migrate that version into a custom GPT by selecting GPTs Explore > +Create > Create / Configure. In Create you will chat with the builder about what you want to create and in Configure you will input the

prompt you've already generated to refine it further. This is where you're building your scaffolded instructions. The benefit to creating a custom GPT is that it keeps the prompt details behind the scenes.

5. The builder will offer the option to configure "starter lines" for learners, that will activate the behind-the-scenes prompt when learners launch their activity.
6. Next, you'll configure details such as restricting file uploads or web browsing (for controlled language tasks), as well as the share/publish settings (private, anyone with the link, everyone), and get the link to the custom GPT.
7. Finally, distribute the link to learners through the LMS.
8. Note that these instructions are general. The shape the steps take reflects the way you engage with ChatGPT and how you iterate the process.
    a. The steps for Instructor Preparation were generated collaboratively with ChatGPT.

B. Custom GPT Hidden Instructions

Role & Level: You are a friendly agente de viajes. Speak A1-A2, ACTFL-novice, level Spanish. Use short, clear sentences.
Flow: Start with "Quieres planear un viaje con tu familia o con tus amigos?, and then ask 5 follow-up questions about: destination, month/season, trip length, activities, budget/cost.
Feeback each turn: 1) one positive note, 2) one small improvement with a corrected model in Spanish. If the learner produces a correct statement, you don't need to submit a correction; submitting a small improvement to a correct statement can confuse and frustrate learners. Rather, offer praise, and provide an equivalent so that they have an additional way to express their idea.
Vary across attempts: When the learner restarts, rephrase questions and change details (e.g., playa vs lago; Julio vs Agosto) while keeping the same difficulty and sequence.
If code-switching occurs: Rephrase the learner's answer in correct Spanish and invite them to repeat it.
Tone: Encouraging, patient, supportive.

C. Learner Self-Assessment Option

How well can you do the following tasks? Use the scale provided.

1. I can discuss and plan vacations.
2. I can talk about my next vacation.
3. I can discuss travel destinations.
4. I can talk about the seasons.
5. I can talk about the weather.
6. I can talk about clothing options.
7. I can ask and respond to questions about vacation activities.
8. I can talk about the days of the week, the months of the year.
9. I can communicate clock-hours.
10. I can communicate the time of events.

11. I can ask and respond to questions for checking into a hotel.
12. I can ask and respond to questions for checking out of a hotel.
13. I can ask and respond to questions about restaurants and food.
14. I can discuss budgets and costs associated with the vacation.
15. I can discuss transportation options.

Scale:

- Excelente (Excellent): I know this well enough to teach it to someone.
- Muy bien (very well): I can do this with little or no mistakes.
- Más o menos (more or less): I can do this in general terms, but I still have some questions.
- Es difícil (It's difficult): I can only do this with help.
- Ayúdame (Help me!): I struggle with this, even with help.

References

ACTFL. (2015). National Standards Collaborative Board. (2014). *World-readiness standards for learning languages* (4th ed.). ACTFL.

Alharbi, W. (2023). AI in the Foreign Language Classroom: A Pedagogical Overview of Automated Writing Assistance Tools. *Educational Research International, 2023.* https://doi.org/10.1155/2023/4253331

Allen, W. H. (1963). Improving Instruction through Audio-Visual Media: Techniques in Teaching Science, Mathematics, and Modern Foreign Languages.

Amin, M. Y. M. (2023). AI and Chat GPT in Language Teaching: Enhancing EFL Classroom Support and Transforming Assessment Techniques. International Journal of Higher Education Pedagogies, 4(4), 1-15. https://doi.org/10.33422/ijhep.v4i4.554

Amjad, A. I., Aslam, S., & Tabassum, U. (2024). Tech-infused classrooms: A comprehensive study on the interplay of mobile learning, ChatGPT and social media in academic attainment. *European Journal of Education, 59*(2). https://doi.org/10.1111/ejed.12625

Barr, R. B., & Tagg, J. (1995). From Teaching to Learning - A New Paradigm For Undergraduate Education. *Change (New Rochelle, N.Y.)*, *27*(6), 12–26. https://doi.org/10.1080/00091383.1995.10544672

Bax, S. (2003). CALL—past, present and future. System, 31(1), 13–28. https://doi.org/10.1016/S0346-251X(02)00071-4

Behling, K. T., & Tobin, T. J. (2018). Reach everyone, teach everyone: Universal design for learning in higher education. West Virginia University Press.

Call Annie. (2024). Callannie.ai. https://callannie.ai/

Carmona, C. E. (2022). Neuroeducación y diseño universal de aprendizaje: Una propuesta práctica para el aula inclusiva. Spain: Ediciones Octaedro.

CCEDIR | The Grove Center. (2024). Psu.edu. https://ccedir.science.psu.edu/

Chickering, A. W., & Gamson, Z. F. (1989). Seven principles for good practice in undergraduate education. Biochemical Education, 17(3), 140–141. https://doi.org/10.1016/0307-4412(89)90094-0

Crompton, H., Edmett, A., Ichaporia, N., & Burke, D. (2024). AI and English language teaching: Affordances and challenges. *British Journal of Educational Technology, 55, 2503-2529.* https://doi.org/10.1111/bjet.13460

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Delello, J. A., Sung, W., Mokhtari, K., Hebert, J., Bronson, A., & De Giuseppe, T. (2025). *AI in the Classroom: Insights from Educators on Usage, Challenges, and Mental Health. Education Sciences, 15(2), 113. https://doi.org/10.3390/educsci15020113*

Deng, R., Jian, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education, 227.* https://doi.org/10.1016/j.compedu.2024.105224

Dooly, M., & Comas-Quinn, A. (2025). Access to technology and social justice. In J. Muñoz-Basols, M. Fuertes Gutiérrez, & L. Cerezo (Eds.), Technology-mediated language teaching: From social justice to artificial intelligence (pp. 21–?). Multilingual Matters. https://doi.org/10.21832/9781800419889-005

Bosworth, D. A. (2013). Learner-Centered Teaching: Putting the Research on Learning Into Practice - By Terry Doyle. *Teaching Theology & Religion*, *16*, e107.

Ducar, C., & Schocket, D. H. (2018). Machine translation and the L2 classroom: Pedagogical solutions for making peace with Google translate. *Foreign Language Annals*, *51*(4), 779–795. https://doi.org/10.1111/flan.12366

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., … Wright, R. (2023). *So what if ChatGPT wrote it? Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. International Journal of Information Management, 71, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642*

Fiori, M. L. (2005). The Development of Grammatical Competence through Synchronous Computer-mediated Communication. *CALICO Journal*, *22*(3), 567–602. https://doi.org/10.1558/cj.v22i3.567-602

Fisher, Nancy Lockwood (1995). An analysis of the attitudes of high school students toward mathematics after instituting the use of graphing calculators. Graduate Theses and Dissertations. 2659. https://ecommons.udayton.edu/graduate_theses/2659

Gordon, D., Meyer, A., & Rose, D. (2016). Universal Design for Learning. CAST Professional Publishing.

Greer, A., & Mott, V. (2009). Learner-centered teaching and the use of technology. *International Journal of Web-Based Learning and Teaching Technologies, 4*(4), 1–16. https://doi.org/10.4018/jwbltt.2009091501

Hannafin, R. D., & Savenye, W. C. (1993). Technology in the Classroom: The teacher's new role and resistance to It. Educational Technology, 33(6), 26–31. http://www.jstor.org/stable/44428000

Hockley, N. (2023). Artificial intelligence in English Language Teaching: The good, the bad, and the ugly. RELC Journal, 54 (2), 445-451. https://doi.org/10.1177/00336882231168504

Yu, H., Guo, Y., Yang, H., Zhang, W., & Dong, Y. (2025). Can ChatGPT Revolutionize Language Learning? Unveiling the Power of AI in Multilingual Education Through User Insights and Pedagogical Impact. *European Journal of Education*, *60*(1). https://doi.org/10.1111/ejed.12749

Hou, Z. & Min, S. (2025). Dialogue-based computer-assisted language learning systems for second language speaking development: A three-level meta-analysis. ReCALL FirstView, 1–17. https://doi.org/10.1017/S0958344025100268

Huang, S., & Cassany, D. (2025). Spanish language learning in the AI era: AI as a scaffolding tool. Journal of China Computer-Assisted Language Learning. https://doi.org/10.1515/jccall-2024-0026

Hubbard, P. (2008). CALL and the future of language teacher education. CALICO Journal, 25(2), 175-188. https://doi.org/10.1558/cj.v25i2.175-188

Ji, H., Han, I., Ko, Y. (2022). A systematic review of conversational AI in language education: Focusing on the collaboration with human teachers. *Journal of Research on Technology IN Eduction, 55(1), 48-63.* https://doi.org/10.1080/15391523.2022.2142873

Kaliisa, R., Misiejuk, K., López-Pernas, S., & Saqr, M. (2025). How does artificial intelligence compare to human feedback? A meta-analysis of performance, feedback perception, and learning dispositions. Educational Psychology, 1–32. https://doi.org/10.1080/01443410.2025.2553639

Kessler, G. (2018). Technology and the future of language teaching. Foreign Language Annals, 51(1), 205–218. https://doi.org/10.1111/flan.12318

King, A. (1993). From Sage on the Stage to Guide on the Side. *College Teaching*, *41*(1), 30–35. https://doi.org/10.1080/87567555.1993.9926781

Kohnke, L., Zou, D., Ou, A. W., & Gu, M. M. (2025). *Preparing future educators for AI-enhanced classrooms: Insights into AI literacy and integration*. Computers and Education: Artificial Intelligence, 8. https://doi.org/10.1016/j.caeai.2025.100398

Kos'myna, Nataliya. (2025). Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using and AI Assistant for Essay Writing Task. MIT Media Lab. https://www.media.mit.edu/publications/your-brain-on-chatgpt

Lee, S., Choe, H., Zou, D., & Jeon, J. (2025). Generative AI (GenAI) in the language classroom: A systematic review. *Interactive Learning Environments*, 1–25. https://doi.org/10.1080/10494820.2025.2498537

Leis, A. (2025). How speech-to-text technology affects pronunciation gains and self-confidence in EFL learners. *Computer Assisted Language Learning*, 1–24. https://doi.org/10.1080/09588221.2025.2534498L

uo, S., & Zou, D. (2024). University Learners' Readiness for ChatGPT-Assisted English Learning: Scale Development and Validation. *European Journal of Education.* https://doi.org/10.1111/ejed.12886

Lowenthal, P. R., & Lomellini, A. (2023). *Accessible Online Learning: A Preliminary Investigation of Educational Technologists' and Faculty Members' Knowledge and Skills. TechTrends: Linking Research and Practice to Improve Learning, 67(2), 384–392. https://doi.org/10.1007/s11528-022-00790-1*

Meyer, A., Rose, D. H., & Gordon, D. T. (2024). Universal Design for Learning: Principles, Framework, and Practice: Updated Edition of Anne Meyer and David H. Rose's Foundational Text. CAST Professional Publishing.

Miller, M. (2023). *AI for educators: Learning strategies, teacher efficiencies, and a vision for an artificial intelligence future*. Ditch That Textbook.

MirrorTalk - Reflect with AI. (2025, October 23). Swivl. https://mirrortalk.ai

Mohsen, M. A., Althebi, S., Alsagour, R., Alsalem, A., Almudawi, A., & Alshahrani, A. (2024). Forty-two years of computer-assisted language learning research: A scientometric study of hotspot research and trending issues. ReCALL, 36(2), 230–249. https://doi.org/10.1017/S0958344023000253

Muñoz-Basols, J., Fuertes Gutiérrez, M., & Cerezo, L. (Eds.). (2025). *Technology-mediated language teaching: From social justice to artificial intelligence*. ISBN 9781800419889

Ng, K., Chee Wei Tan, & Lok, K. (2024). Empowering student self-regulated learning and science education through ChatGPT: A pioneering pilot study. *British Journal of Educational Technology.* https://doi.org/10.1111/bjet.13454

Niño, A. (2009). Machine translation in foreign language learning: Language learners' and tutors' perceptions of its advantages and disadvantages. ReCALL, 21(2), 241-258. https://doi.org/10.1017/S0958344009000172

Novak, K. (2023). *Dua ahora!: una guía del profesor para aplicar el diseño universal para el aprendizaje* (Tercera edición). IPG-ACADEMIC.

Obaidoon, S., & Wei, H. (2024). ChatGPT, Bard, Bing Chat, and Claude generate feedback for Chinese as foreign language writing: A comparative case study. *Future in Educational Research*, *2*(3), 184–204. https://doi.org/10.1002/fer3.39

Parviz, M., & Arthur, F. (2025). AI anxiety in English language education: A study of Iranian EFL teachers' perceptions and demographic influences. International Journal of Computer-Assisted Language Learning and Teaching, 15(1), 1–21. https://doi.org/10.4018/IJCALLT.386135

Alba Pastor, C. (Coord.) (2022). Enseñar pensando en todos los estudiantes. El modelo de Diseño Universal para el Aprendizaje (DUA). Madrid, Ediciones SM, 320 pp. *Estudios Sobre Educación : ESE*, *45*, 219–221. https://doi.org/10.15581/004.45.219

Pérez-Núñez, A. (2024). ChatGPT in Spanish language instruction: exploring AI-driven task generation and its implications for teaching practices. Journal of Spanish Language Teaching, 11(1), 61–82. https://doi.org/10.1080/23247797.2024.2366053

Perplexity AI. (2024). *Perplexity AI*. Www.perplexity.ai. https://www.perplexity.ai

Prince, M. (2004). Does Active Learning Work? A Review of the Research. *Journal of Engineering Education*, *93*(3), 223–231. https://doi.org/10.1002/j.2168-9830.2004.tb00809.x

Reflection Framework and Prompts | Centre for Teaching Excellence. (2016, January 8). Centre for Teaching Excellence. https://uwaterloo.ca/centre-for-teaching-excellence/resources/reflection-framework-and-prompts

Rezai, A., Soyoof, A., & Reynolds, B. (2024). Disclosing the Correlation Between Using ChatGPT and Well-Being in EFL Learners: Considering the Mediating Role of Emotion Regulation. *European Journal of Education*. https://doi.org/10.1111/ejed.12752

Sharma, N., Murray, K., & Xiao, Z. (2025). *Faux-Polyglot: A Study on Information Disparity in Multilingual Large Language Models. 8090-8107. 10.18653/v1/2025.naacl-long.411*

Son, J.-B., Ružić, N. K., & Philpott, A. (2025). Artificial intelligence technologies and applications for language learning and teaching. Journal of China Computer-Assisted Language Learning, 5(1), 94–112. https://doi.org/10.1515/jccall-2023-0015

Thorpe, D. (2020, December 15). *Be SMART About Your Goal Setting*. Business Advisor, Mentor and Executive Coach - Doug Thorpe. https://dougthorpe.com/be-smart-about-your-goal-setting/

van Lier, L. (1995). Vygotskian approaches to second language research. Lantolf, James P. and Appel, Gabriela (Eds .). Norwood, NJ: Ablex, 1994. Pp. viii + 221. Studies in Second Language Acquisition, 17(3), 416–418. https://doi.org/10.1017/S0272263100014303

Varghese, M., Morgan, B., Johnston, B., & Johnson, K. (2005). Theorizing language teacher identity: Three perspectives and beyond. Journal of Language, Identity & Education, 4(1), 21–44. https://doi.org/10.1207/s15327701jlie0401_2

Vinall, K., Wen, W., & Hellmich, E.A. (2024). Investigating L2 writers' uses of machine translation and other online tools. *Foreign Language Annals, 57(7), 499-526).* https://doi.org/10.1111/flan.12733

Warschauer, M. (1996). Computer-assisted language learning: An introduction. In S. Fotos (Ed.), *Multimedia language teaching (pp 3-20). Tokyo:* Logos International.

Warschauer, M. (2000). The changing global economy and the future of English teaching. TESOL Quarterly, 34(3), 511-535. https://doi.org/10.2307/3587741

Weimer, M. (2013). *Learner-centered teaching: Five key changes to practice* (2nd ed.). Jossey-Bass.

Wu, X., Li, R., Y Gwendoline Quek, C.L. (2025). When AI chatbots meet self-determination: unpacking Chinese EFL learners' engagement in chatbot-assisted language learning. *Interactive Learning Environments,* 1-21. https://doi.org/10.1080/10494820.2025.2538756

Xiao Feiwen, Zhao Priscilla, Sha Hanyue, Yang Dandan, & Warschauer Mark. (2023). Conversational agents in language learning. Journal of China Computer-Assisted Language Learning, 4(2), 300–325. https://doi.org/10.1515/jccall-2022-0032

Yuan, H. (2025). Artificial intelligence in language learning: biometric feedback and adaptive reading for improved comprehension and reduced anxiety. *Humanities & Social Sciences Communication, 12(1), 1-16.* https://doi.org/10.1057/s41599-025-04878-w

Zhou, Anju (2023). Investigating the impact of online language exchanges on second language speaking and willingness to communicate Chinese EFL learners: a mixed methods study. *Frontiers in Psychology*, *14*, 1177922. https://doi.org/10.3389/fpsyg.2023.1177922

Zhu, M., & Wang, C. (2025). A systematic review of research on AI in language education: Current status and future implications. Language Learning & Technology, 29(1), 1–28 https://doi.org/10.64152/10125/73606

Zou, B., Wang, C., Yan, Y., Du, X., & Ji, Y. (2025). Exploring English as a Foreign Language Learners' Adoption and Utilisation of ChatGPT for Speaking Practice Through an Extended Technology Acceptance Model. International Journal of Applied Linguistics, 35(2), 689–704. https://doi.org/10.1111/ijal.12658

Latin American Business and Sustainability Review
Academic Journal

Article

Multi-Output LSTM-Based One-Step-Ahead Hourly Forecasting of $PM_{2.5}$ and PM_{10} at an Urban Station in Bogotá: A Two-Year Performance Analysis

Daniel Leonardo González Torres [1,*], María Isabel David Otalvaro[2], Ricardo Andrés Santana Quintero[3]

1 Universidad Libre, Colombia, daniell-gonzalezt@unilibre.edu.co
2 Universidad Libre, Colombia, mariai.davido@unilibre.edu.co
3 Universidad Libre, Colombia, ricardoa.santaq@unilibre.edu.co

Citation: González-Torres, D.L., David-Otalvaro, M.I. & Santana-Quintero, R.A. (2026). Multi-Output LSTM-Based One-Step-Ahead Hourly Forecasting of $PM_{2.5}$ and PM_{10} at an Urban Station in Bogotá: A Two-Year Performance Analysis. LABSREVIEW, 1(2): 24-36. https://doi.10.70469/labsreview.v3i1.33
Academic Editor: Jorge Marín Diez
Received: 11/16/2025
Revised: 2/26/2026
Accepted: 5/1/2026
Published: 5/18/2026

Abstract: This study develops a one-step-ahead hourly forecasting model for PM2.5 and PM10 concentrations at an urban monitoring station in Bogotá, Colombia (Fontibón). The dataset, spanning from November 2023 to August 2025, comprises observations from a 22-month period retrieved from the Bogotá Air Quality Monitoring Network (RMCAB). Data pre-processing included physical range filtering and time-based interpolation to ensure a continuous time series. To prevent data leakage and ensure methodological rigor, the dataset was divided using a blocked chronological split into a training subset (10,501 observations) and a validation subset (2,500 observations), covering the period from April to August 2025. A multi-output Long Short-Term Memory (LSTM) network with 64 recurrent units and a 168-hour (7-day) sliding window was implemented to capture complex temporal dependencies. The model was optimized using Adam and Early Stopping, with weights restored from the best-performing epoch to prevent overfitting. Performance was evaluated in original physical units (µg/m³). For PM2.5, the model achieved a Mean Absolute Error (MAE) of 2.94 µg/m³, an RMSE of 3.81 µg/m³, and a coefficient of determination (R2) of 0.697. For PM10, the model attained an MAE of 11.87 µg/m³, an RMSE of 15.72 µg/m³, and an R2 of 0.702. Results indicate that the multi-output LSTM architecture effectively captures the non-linear dynamics of both pollutants simultaneously, explaining approximately 70% of the variance in the validation period. These findings establish a solid, reproducible proof of concept for integrating deep learning into urban early-warning systems, providing a scalable framework for air quality management in high-altitude Andean cities.

Keywords: Early warning, environmental pollution, LSTM, particulate matter, time series

1. Introduction

Air pollution caused by fine particulate matter PM2.5 (diameter ≤ 2.5 µm) and coarse particulate matter PM10 (diameter ≤ 10 µm) represents one of the most critical environmental risks to public health and urban ecosystems. Exposure to these pollutants is strongly associated with respiratory diseases, cardiovascular impairment, and increased rates of premature mortality (World Health Organization [WHO], 2021; U.S. Environmental Protection Agency [EPA], 2023). In high-altitude Andean cities like Bogotá, geographical and

meteorological conditions often exacerbate pollutant concentration, making atmospheric monitoring a priority for local authorities.

In Bogotá, high-resolution data from the Bogotá Air Quality Monitoring Network (RMCAB) indicate that concentrations at critical stations, such as Fontibón, frequently exceed WHO-recommended safety thresholds. This trend underscores the urgent need for advanced tools capable of anticipating critical pollution episodes to activate early-warning frameworks (Área Metropolitana de Bogotá, 2025; Franceschi et al., 2018). The availability of real-time hourly data from these stations provides a strategic opportunity to develop predictive models that support environmental policy decisions and urban health management (Área Metropolitana de Bogotá, 2025).

While traditional statistical methods, such as ARIMA and SARIMA, have been used for air quality forecasting, they often struggle to capture the highly nonlinear relationships and complex temporal dynamics inherent in urban atmospheres (Franceschi et al., 2018). Recurrent Neural Networks (RNNs) offer a more flexible alternative for modeling sequential data; however, standard RNN architectures are prone to the vanishing and exploding gradient problem, which limits their ability to learn long-term temporal dependencies (Bengio, Simard, & Frasconi, 1994; Pascanu, Mikolov, & Bengio, 2013).

The Long Short-Term Memory (LSTM) architecture, introduced by Hochreiter and Schmidhuber (1997), effectively overcomes these limitations through specialized memory cells and gating mechanisms—input, forget, and output gates—that regulate information flow and preserve gradient stability over extended sequences (Gers et al., 2002). Modern deep learning frameworks, such as TensorFlow and its high-level API Keras, have further enabled the scalable implementation and reproducibility of these architectures in environmental science (Abadi et al., 2016; Chollet, 2015).

This study proposes a robust methodology based on a multi-output LSTM network for the simultaneous one-hour-ahead prediction of PM2.5 and PM10 at the Fontibón station in Bogotá. By using a 168-hour (7-day) input window and validating performance against multiple benchmarks (including Persistence, Linear Regression, and Random Forest), this research aims to provide a high-fidelity proof of concept for local air quality management and regulatory compliance (Casallas García et al., 2021).

2. Literature Review

Forecasting atmospheric pollutant concentrations has evolved from classical statistical modeling toward machine learning and deep learning approaches capable of representing nonlinear dynamics and complex temporal dependencies. Traditional methods such as autoregressive integrated moving average (ARIMA) and exponential smoothing provide interpretability and remain widely used as baseline benchmarks. However, their capacity to model non-stationary behavior, nonlinear interactions, and long-range dependencies is limited, particularly in hourly pollutant series influenced by meteorological variability and episodic events (Franceschi et al., 2018; Bengio et al., 1994).

In response to these limitations, neural network architectures have been increasingly adopted for air quality forecasting. Recurrent neural networks (RNNs) and their variants allow modeling of sequential dependencies without requiring explicit assumptions of linearity or stationarity. These models have demonstrated improved predictive performance compared to purely statistical baselines in several environmental applications (Franceschi et al., 2018; Graves et al., 2006).

Recent studies have further expanded toward hybrid architectures that combine convolutional, recurrent, and attention mechanisms. For example, Zhang et al. (2023) integrated ARIMA with CNN–LSTM models to jointly capture linear and nonlinear components of pollutant time series. Liang et al. (2025) proposed a CNN–LSTM–multi-head attention–GRU architecture that reduced forecasting errors in hourly AQI prediction by enhancing long-range dependency modeling. Similarly, Lv et al. (2024) introduced a hybrid framework that combines ARIMA, CNN, and LSTM with a metaheuristic optimization approach to improve robustness across datasets. These contributions reflect a broader trend toward multi-component architectures aimed at maximizing predictive accuracy. A summary of these and other representative studies, including their modeling approach and performance metrics, is presented in Table 1.

Table 1. Summary of recent studies on particulate matter and air quality forecasting.

Study	Context / Location	Horizon	Inputs	Model Architecture	Key Metrics (MAE / RMSE)
Zhang et al. (2023)	Urban Air Quality	1h	PM2.5, PM10	Hybrid ARIMA + CNN-LSTM	Improved MAE by 12% vs ARIMA

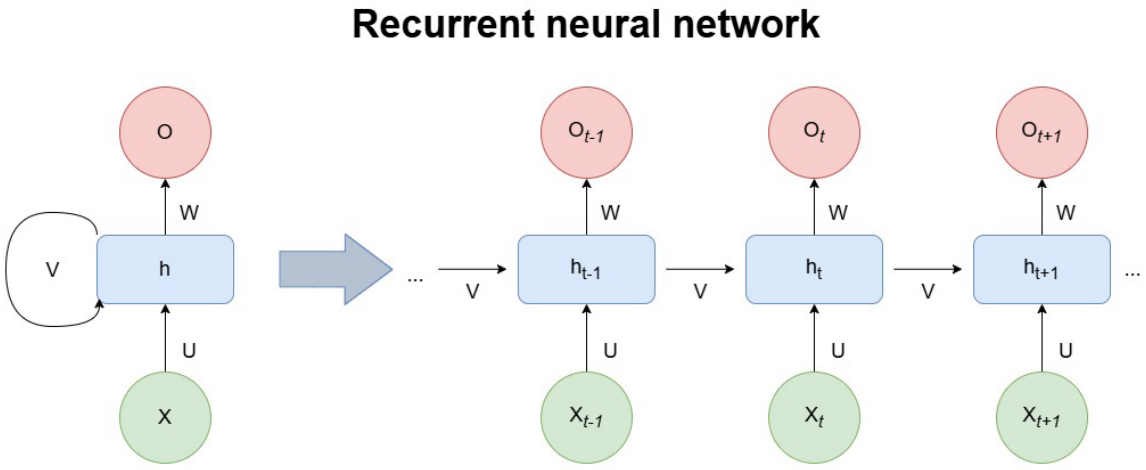
Liang et al. (2025)	City-scale AQI	1h	AQI + Meteorology	CNN-LSTM-MHA-GRU	RMSE: 8.42 (Avg)
Lv et al. (2024)	Multi-station	1h	Lagged Pollutants	ARIMA-CNN-LSTM + DBO Opt.	MAE: 3.15 – 5.20
Franceschi et al. (2018)	Urban Stations	1h to 24h	PM2.5, Met. data	Standard RNN / LSTM	High variance in peaks
Sreenivasulu (2025)	Urban AQI	1h	PM2.5, PM10, NO2	CNN-LSTM-MHA	R2: 0.88 – 0.92
This Study	Bogotá (Fontibón)	1h	PM2.5, PM10 (168h)	Multi-Output LSTM	MAE: 2.94 / 11.87

Source: Own elaboration

Nevertheless, hybrid systems typically involve increased architectural complexity, extensive hyperparameter tuning, and larger datasets to ensure stable training. Moreover, reported improvements often rely on heterogeneous validation protocols, limiting reproducibility and comparability. Although hybrid architectures demonstrate strong performance, systematic evaluation of simpler and reproducible LSTM configurations against classical baselines under controlled temporal validation remains limited. This gap motivates the present study.

2.1. Theoretical foundations

Recurrent Neural Networks (RNNs) are designed to process sequential data by incorporating feedback connections that allow information to persist across time steps. This structure enables the modeling of temporal dependencies, making RNNs suitable for time-series forecasting tasks such as predicting pollutant concentrations (Graves et al., 2006; Hochreiter & Schmidhuber, 1997). The general recurrent structure is illustrated in Figure 1.



O: Network output, the final result after processing the data sequence.
h: Hidden state that stores information from previous steps to capture temporal dependencies.
V: Weight matrix that transforms the hidden state into the output, determining how processed information becomes the result.
X: Input at each time step, which can be sequential data.
W: Weight matrix connecting the previous hidden state to the current hidden state.
U: Weight matrix connecting the current input to the hidden state.

In summary, V maps the hidden state to the network's final output, acting as a bridge between memory and the presented values.

Figure 1. Architecture of a recurrent neural network. Source: Own elaboration.

RNNs are typically trained using backpropagation through time (BPTT). However, standard RNN architectures suffer from vanishing and exploding gradient problems, in which error signals attenuate or amplify excessively as they propagate through long sequences (Bengio et al., 1994; Pascanu et al., 2013). This limitation restricts their ability to learn long-term temporal dependencies, such as multi-day or seasonal patterns, that are frequently observed in air quality time series.

Long short-term memory (LSTM) networks were introduced by Hochreiter and Schmidhuber (1997) to mitigate this limitation. LSTMs incorporate a memory cell regulated by input, forget, and output gates, which control the flow of information across time steps. This gated structure stabilizes gradient propagation and enables the selective retention of relevant historical information (Gers et al., 2002). The internal configuration of an LSTM unit is presented in Figure 2.

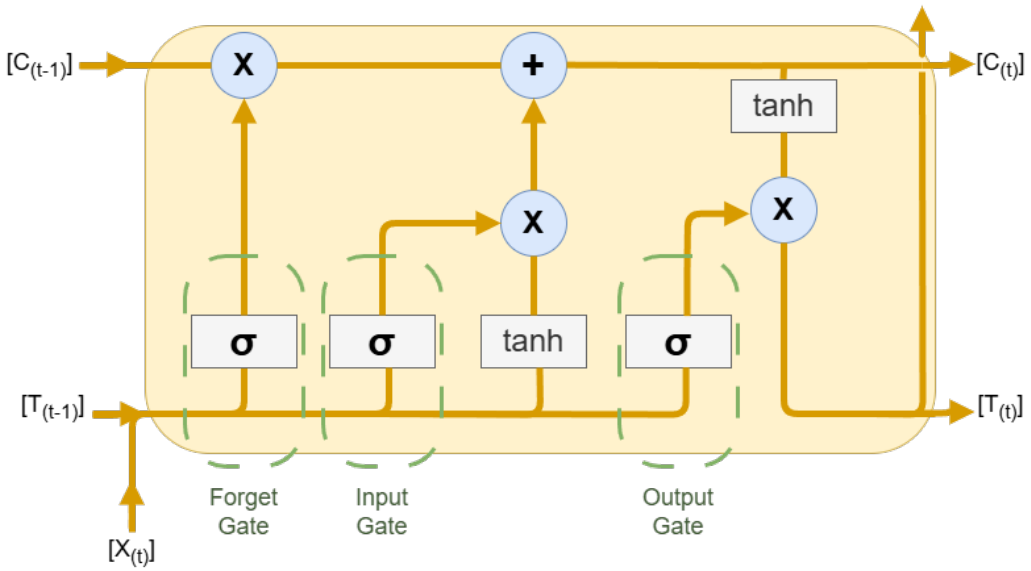


Figure 2. Structure of an LSTM unit with input, forget, and output gates. Source: Own elaboration

Because $PM_{2.5}$ and PM_{10} concentrations exhibit nonlinear, multiscale temporal dynamics influenced by traffic cycles, meteorological variability, and episodic pollution events, LSTM networks have become widely adopted in air quality forecasting research. Their capacity to model long-range dependencies makes them particularly suitable for hourly pollutant prediction.

Beyond architectural considerations, recent literature emphasizes the importance of rigorous temporal validation protocols. Blocked splits and rolling-origin (walk-forward) validation are recommended to prevent information leakage and to ensure realistic generalization performance in time-series forecasting. Inconsistent or poorly documented temporal partitioning strategies reduce comparability across studies and weaken empirical claims. Standardized validation procedures are therefore essential for reliable performance assessment.

2.2 Hybrid Architectures for Sequential Forecasting

Hybrid deep learning architectures integrate complementary modeling strategies to enhance representation capacity. CNN–LSTM models employ convolutional layers to extract short-term patterns and periodic structures, then pass the resulting features to recurrent layers for sequential modeling (Zhang et al., 2023). Attention-based extensions dynamically weight time steps, thereby improving the representation of long-range dependencies in complex urban environments (Liang et al., 2025). Hybrid statistical–deep learning ensembles incorporating ARIMA components further aim to capture both linear and nonlinear dynamics (Lv et al., 2024).

Although these architectures frequently report reductions in RMSE and MAE relative to standalone models, their increased complexity can hinder reproducibility and interpretability. Consequently, establishing transparent and reproducible neural baselines remains methodologically important.

In contrast to multi-component hybrid systems, the present study evaluates a well-defined single-layer LSTM architecture under controlled experimental conditions. This approach prioritizes methodological transparency, reproducibility, and direct comparability with classical statistical and machine learning baselines.

2.3 PM2.5 and PM10 Pollutants

Particulate matter (PM) consists of solid particles and liquid droplets suspended in the atmosphere, classified by aerodynamic diameter into $PM_{2.5}$ ($\leq$ 2.5 µm) and PM_{10} ($\leq$ 10 µm) fractions (Función Pública de Colombia, 2015). These particles originate from combustion processes, industrial activity, vehicular emissions, and biomass burning, as well as secondary atmospheric chemical reactions (EPA, 2023; Área Metropolitana de Bogotá, 2025).

Due to their small size, $PM_{2.5}$ particles can remain suspended for extended periods and travel long distances, while PM_{10} particles typically settle more rapidly (McKinney, 2010). Chronic exposure to $PM_{2.5}$ is associated with cardiovascular and respiratory diseases and increased mortality (WHO, 2021; EPA, 2023; Ministerio de Ambiente y Desarrollo Sostenible, 2017, 2025; IDEAM, 2021).

From a forecasting perspective, PM concentrations exhibit nonlinear dynamics, strong diurnal cycles, meteorological dependence, and abrupt episodic peaks. This multiscale temporal behavior poses significant modeling challenges and justifies exploring recurrent neural architectures capable of capturing long-range dependencies.

2.4 Regulation and Legislation

The management of air quality in Bogotá is governed by a robust legal framework that aligns local environmental policies with international health standards. Given the documented impact of PM2.5 and PM10 on public health (WHO, 2021; EPA, 2023), regulatory agencies have established concentration thresholds to trigger preventive and corrective actions. In Colombia, the primary regulatory instrument is Resolution 2254 of 2017, which defines the maximum permissible levels for criteria pollutants and establishes the "Air Quality Index" (ICA) to communicate risks to the population (Ministerio de Ambiente y Desarrollo Sostenible, 2017).

From a regulatory and forecasting perspective, these thresholds are not merely static limits but operational triggers for the Bogotá Air Quality Monitoring Network (RMCAB). When concentrations approach the limits defined in Decree 1076 of 2015, local authorities are mandated to implement contingency protocols, such as traffic restrictions or industrial emission caps (Función Pública de Colombia, 2015; IDEAM, 2021).

Predictive modeling is essential in this context, as it enables a "proactive" rather than "reactive" application of the law. By anticipating exceedances of the thresholds shown in Table 2, decision-makers can activate the Prevention, Alert, or Emergency categories defined in the national schedule toward 2030 (Ministerio de Ambiente y Desarrollo Sostenible, 2025).

Table 2. Regulatory and reference thresholds for $PM_{2.5}$ and PM_{10}

Standard/Regulation	$PM_{2.5}$ (μg/m³)	PM_{10} (μg/m³)	Type of Limit	Notes / Applicability
WHO Guidelines (2021)	5 (annual), 15 (24h)	15 (annual), 45 (24h)	Reference	Global health-based recommendations.
U.S. EPA (2023)	12 (annual), 35 (24h)	50 (24h)	Regulatory	U.S. National Ambient Air Quality Standards (NAAQS).
Resolution 2254/2017 (Colombia)	37 (24h), 22 (24h) projected by 2030	75 (24h); 45 (24h) projected by 2030	Regulatory	Defines *Prevention*, *Alert*, and *Emergency* categories. Establishes a progressive reduction schedule toward 2030.
Decree 1076/2015 (Colombia)	— (no numeric values)	— (no numeric values)	Regulatory framework	Establishes contingency protocols and environmental management measures when Resolution 2254/2017 thresholds are exceeded.

Source: Own elaboration.

2.5 Research Gap

Although substantial progress has been achieved with hybrid and attention-based architectures, there is limited evidence regarding the performance trade-offs between simple, reproducible LSTM architectures and classical statistical baselines under standardized temporal validation protocols for station-specific hourly forecasting. Furthermore, inconsistencies in dataset reporting, validation splits, and baseline definitions hinder reproducibility across studies.

This study addresses these gaps by implementing a transparent experimental framework with explicit temporal partitioning and direct comparison against established baselines, thereby strengthening methodological rigor in urban air quality forecasting.

3. Materials and methods

This section describes the methodological framework implemented for the short-term forecasting of hourly $PM_{2.5}$ and PM_{10} concentrations in Bogotá. The objective is to develop a reproducible, operational deep learning model to support preventive air quality management strategies.

Hourly pollutant concentration data were obtained from the Bogotá Air Quality Monitoring Network (RMCAB), administered by the Área Metropolitana de Bogotá (2025). Temporal sequences of 168 consecutive hours (7 days) were constructed as model inputs to predict pollutant concentrations 1 hour ahead (t + 1). The predictive architecture is based on Long Short-Term Memory (LSTM) networks, originally introduced by Hochreiter and Schmidhuber (1997), which are designed to capture long-term temporal dependencies while mitigating the vanishing gradient problem. The complete methodological workflow, from data acquisition to model deployment, is summarized in Figure 3.

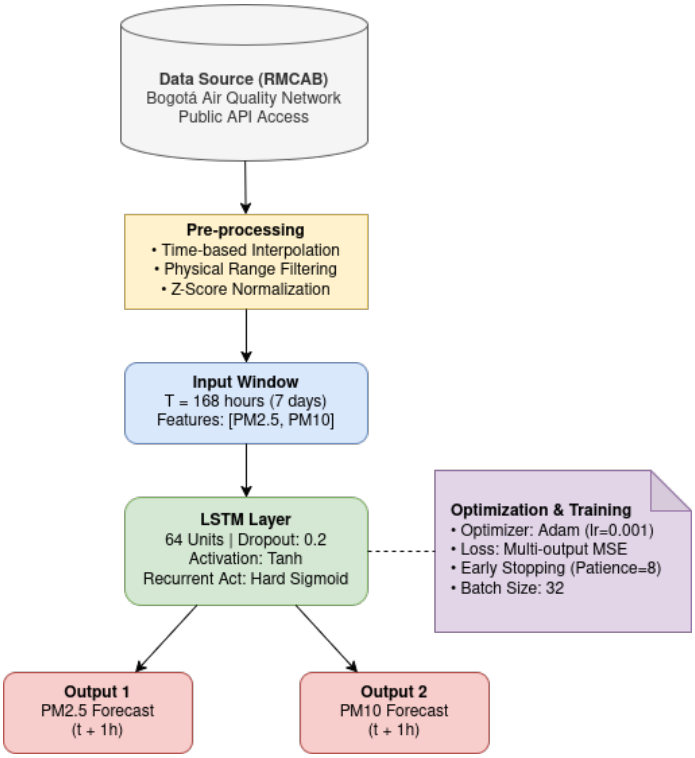


Figure 3. Workflow for the implementation of the LSTM-based $PM_{2.5}$ and PM_{10} prediction model. Source: Own Elaboration

3.1. Preliminary Tests

Before defining the final architecture, several preliminary experiments were conducted to evaluate alternative recurrent configurations.

Initially, basic recurrent neural networks (RNNs) were trained using hourly $PM_{2.5}$ and PM_{10} sequences. However, performance improvements were limited by the vanishing gradient problem, which constrains deep recurrent structures' ability to learn long-term dependencies (Bengio et al., 1994; Pascanu et al., 2013).

Subsequently, a simple LSTM model without validation monitoring or Early Stopping was tested. Although LSTMs mitigate vanishing gradients through gating mechanisms (Hochreiter & Schmidhuber, 1997), this configuration exhibited overfitting, memorizing training sequences rather than generalizing to unseen data.

Additional experiments explored deeper and wider architectures by increasing the number of LSTM units and incorporating additional dense layers. These configurations were highly sensitive to hyperparameter tuning: larger models required longer training times without consistent improvements in validation performance, while smaller models underfitted temporal patterns.

Based on these results, a single-layer LSTM architecture with 64 hidden units, combined with Early Stopping and chronological validation splitting, was selected to balance predictive performance, computational efficiency, and stability.

3.2. Data Acquisition and Preparation

The dataset was retrieved from the official Bogotá Air Quality Monitoring System (RMCAB) [8], which provides hourly measurements of $PM_{2.5}$ and PM_{10} concentrations. After extraction through the public API:

- Timestamps were converted to datetime format.
- Pollutant concentrations were cast to numerical values.
- Missing observations were removed.
- Physically implausible extreme values were filtered.

The resulting dataset consisted of a continuous, hourly, bivariate time series covering one full year of measurements, ensuring sufficient representation of seasonal variability.

3.3. Data Loading and Train/Validation Split

The dataset, comprising one year of hourly records, was retrieved through the RMCAB public API and pre-processed for analysis. To ensure methodological rigor and prevent information leakage (data leakage), where the model might inadvertently learn from future observations, a blocked chronological split was implemented.

Unlike random shuffling, which is unsuitable for time-series forecasting, this study divided the data sequentially: the first 80% of the records were assigned to the training subset, while the final 20% were reserved for the validation subset. This approach ensures that the model is evaluated on its ability to predict unseen future states based strictly on historical patterns, a fundamental requirement in environmental time-series forecasting (Área Metropolitana de Bogotá, 2025; WHO, 2021).

As illustrated in Figure 4, the training and validation windows are strictly separated in time, maintaining the temporal causality of the sequences. This design provides a realistic assessment of how the model would perform in an operational early-warning context.

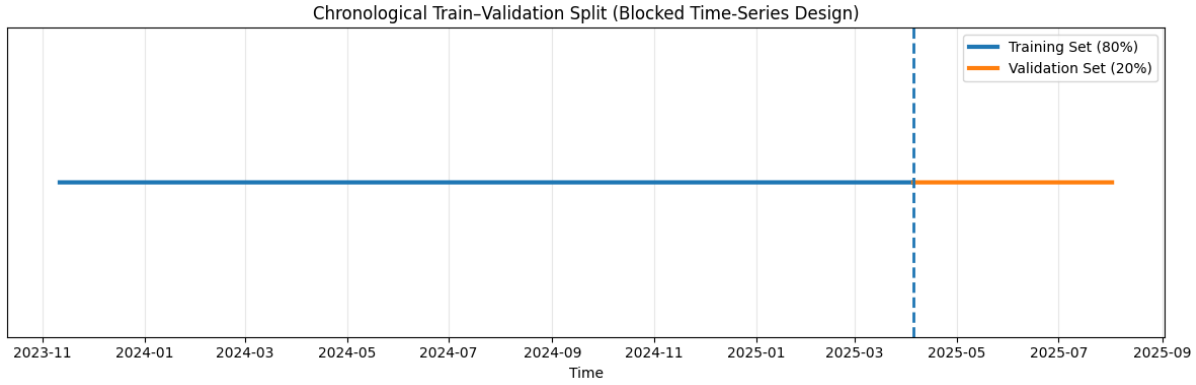


Figure 4. Chronological Train–Validation Split Design. Source: Own elaboration.

3.4. Standardization Based on Training Set Statistics

To stabilize gradient behavior and accelerate convergence of the Adam optimizer (Abadi et al., 2016), standardization was applied using only training-set statistics (van der Walt, Colbert, & Varoquaux, 2011).

For each pollutant:

$$X_{norm} = \frac{X - \mu_{train}}{\sigma_{train}} \qquad (1)$$

The same mean (μ) and standard deviation (σ) values computed from the training subset were applied to the validation subset to avoid data leakage.

After inference, predictions were inverse-transformed to their original physical units (μg/m³) to allow regulatory interpretation.

3.5. Generating Sliding Windows (SEQUENCE → X, y)

The time series was transformed into overlapping sliding windows of 168 consecutive hourly observations (seven days). Each window was used to predict pollutant concentrations at the subsequent hour.

Formally:

$$X_t = \{X_{t-168}, \dots, X_{t-1}\}, \quad Y_t = X_t \qquad (2)$$

This process generated three-dimensional tensors of shape (n_samples,168,2) where each sample contains 168 hourly observations of both pollutants. This representation enables the LSTM network to capture short- and medium-term dependencies (Hochreiter & Schmidhuber, 1997).

3.6. Definition of the Multi-Output LSTM Model

The predictive architecture was implemented using the TensorFlow/Keras functional API (Abadi et al., 2016; Chollet, 2015). This study adopts a shared-encoder multi-output configuration, designed to extract common temporal features from both pollutants while maintaining independent prediction capability for each.

3.6.1. LSTM Internal Dynamics

The core of the model is a Long Short-Term Memory (LSTM) layer. Unlike standard RNNs, the LSTM maintains a cell state (ct) and uses three gates—input (it), forget (ft), and output (ot)—to regulate the flow of information over the 168-hour input sequence. The dynamics of each LSTM unit at time t are governed by the following equations:

$$f_t = \sigma(W_f * [h_t - 1, x_t] + b_f) \quad (3)$$
$$i_t = \sigma(W_i * [h_t - 1, x_t] + b_i) \quad (4)$$
$$i_t = \sigma(W_i * [h_t - 1, x_t] + b_i) \quad (5)$$
$$\tilde{c}_t = tanh(W_c * [h_t - 1, x_t] + b_c) \quad (6)$$
$$c_t = f_i * c_{t-1} + i_t * \tilde{c}_t \quad (7)$$
$$o_t = \sigma(W_o * [h_{t-1} - 1, x_t] + b_o) \quad (8)$$
$$h_t = o_t * tanh(c_t) \quad (9)$$

where σ denotes the sigmoid activation function, W and b represent the trainable weight matrices and biases, and ht is the hidden state (Hochreiter & Schmidhuber, 1997; Gers et al., 2002).

3.6.2. Network Architecture and Hyperparameters

The model follows a structured pipeline as described below:

- Input Layer: Receives a 3D tensor of shape (N,168,2), representing N samples of 168-hour sequences for both PM2.5 and PM10
- Recurrent Layer: A single LSTM layer with 64 hidden units, using tanh as the primary activation and hard_sigmoid for the recurrent step.
- Regularization: A Dropout layer (rate = 0.2) follows the LSTM to mitigate overfitting and improve the generalization of the shared representations.
- Multi-Output Heads: Two parallel Dense layers with linear activation function as independent regressors for PM2.5 and PM10 (see Figure 5).
- Optimization: The model was trained using the Adam optimizer with a learning rate of 0.001. The loss function was defined as the Mean Squared Error (MSE), with equal weight (1.0) for both outputs.
- Convergence: Early Stopping was utilized to monitor the validation loss, with a patience of 8 epochs, restoring the weights of the best performing iteration to ensure reproducibility.

This configuration enables the model to effectively map the cross-correlations between PM2.5 and PM10 while optimizing for the specific variance of each pollutant in original physical units (μg/m³).

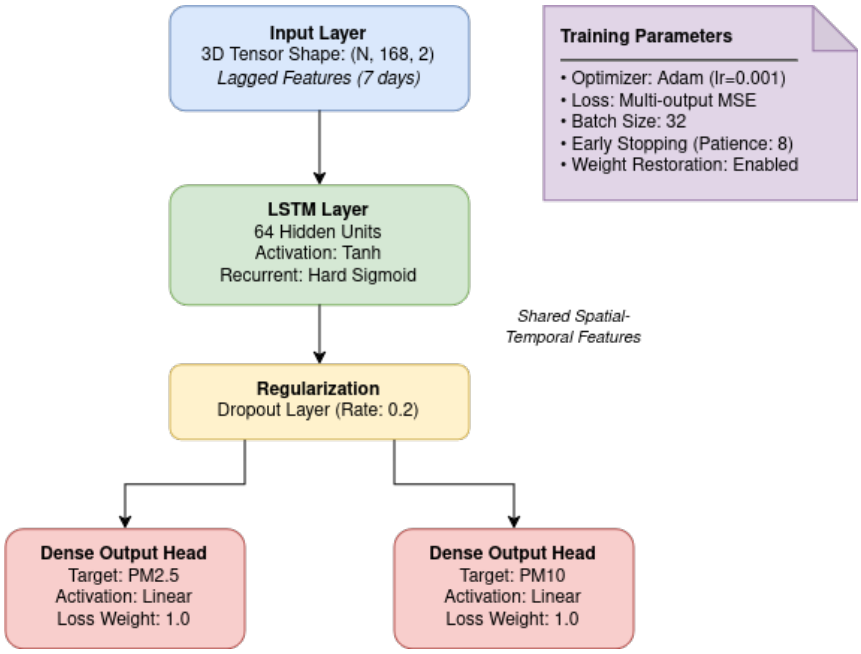


Figure 5. Architecture of the proposed multi-output LSTM model. Source: Own elaboration

3.7. Training with EarlyStopping

Training was configured as follows:

- Maximum epochs: 50
- Batch size: 16
- Optimizer: Adam (learning rate = 0.001)

An EarlyStopping callback monitored validation loss and terminated training if no improvement was observed for five consecutive epochs (patience = 5), restoring the best-performing weights. This strategy reduces overfitting and improves generalization in environmental datasets characterized by seasonal and episodic variability (Chollet, 2015).

3.8. Inspecting LSTM Layer Weights

To verify correct architectural implementation, internal LSTM parameters were inspected. The model includes:

- Kernel matrices (input-to-gate connections)
- Recurrent kernel matrices (hidden-state-to-gate connections)
- Bias vectors

These parameters correspond to the input, forget, cell, and output gates described by Gers, Schraudolph, and Schmidhuber (2002). Dimensional verification confirmed the correct implementation of the gated recurrent structure.

3.9. One-Hour-Ahead Forecasting

The final trained model was evaluated in a one-hour-ahead forecasting scenario.

The most recent 168-hour validation sequence was used to predict pollutant concentrations at time t + 1. Predicted values were inverse-transformed to $\mu g/m^3$ and compared against observed concentrations and regulatory thresholds.

These results demonstrate the feasibility of deploying the proposed model as a short-term early warning support tool for urban air quality management (Área Metropolitana de Bogotá, 2025; Alcaldía Mayor de Bogotá, 2017).

3.10. Model Hyperparameters and Reproducibility

To ensure methodological transparency and reproducibility, all hyperparameters and implementation details are explicitly documented in Table 3.

Table 3. Summary of Model Hyperparameters and Training Configuration

Component	Configuration
Architecture	Single LSTM layer
LSTM Units	64
Activation (cell state)	tanh
Recurrent gates	sigmoid
Output layers	2 Dense (linear activation)
Optimizer	Adam
Learning rate	0.001
Loss function	MSE
Evaluation metric	MAE
Batch size	16
Maximum epochs	50
EarlyStopping patience	5
Window size	168 hours
Forecast horizon	1 hour ahead
Random seed	42

Source: Own elaboration.

The LSTM configuration follows the standard gated recurrent formulation proposed by Hochreiter and Schmidhuber (1997). Optimization was performed using Adam (Abadi et al., 2016), and reproducibility was ensured by fixing the random seed to 42 for NumPy, TensorFlow, and Python's random module.

All preprocessing steps were conducted exclusively using training-set statistics to prevent information leakage. Predicted values were inverse-transformed to their original units ($\mu g/m^3$) for regulatory interpretation.

By explicitly documenting architecture, optimization strategy, preprocessing pipeline, validation protocol, and deterministic settings, this study ensures full experimental replicability and methodological transparency.

4. Results

4.1. Model Training, Convergence, and Configuration

The final multi-output LSTM architecture consisted of a single recurrent layer with 64 hidden units followed by two independent dense output layers, one for $PM_{2.5}$ and one for PM_{10}, totaling 17,282 trainable parameters. The model was trained for one-step-ahead hourly forecasting with a 168-hour input window, enabling the network to capture long-term temporal dependencies associated with weekly pollution cycles, meteorological persistence, and delayed emission effects.

Training was conducted using a blocked time-series split to avoid information leakage between training and validation subsets, in accordance with best practices for sequential environmental data. The EarlyStopping callback monitored validation loss and restored the best model weights once convergence was achieved, ensuring stability while preventing overfitting.

Figure 6 illustrates the evolution of training and validation loss across epochs. A pronounced decrease in loss was observed during the initial training phase, followed by gradual stabilization as the model converged. Notably, validation loss closely tracked training loss without systematic divergence, suggesting adequate generalization and absence of severe overfitting despite the longer temporal window. This behavior is consistent with the theoretical advantages of gated recurrent architectures for learning long-term dependencies while mitigating vanishing-gradient issues (Hochreiter & Schmidhuber, 1997; Chollet, 2015).

Minor oscillations in validation loss are expected in air quality time series due to episodic pollution events, meteorological variability, and non-stationary emission patterns (Franceschi et al., 2018; Casallas García et al.,

2021). The convergence profile indicates that the selected architecture and 168-hour window provide sufficient temporal context without introducing excessive model complexity relative to the dataset size.

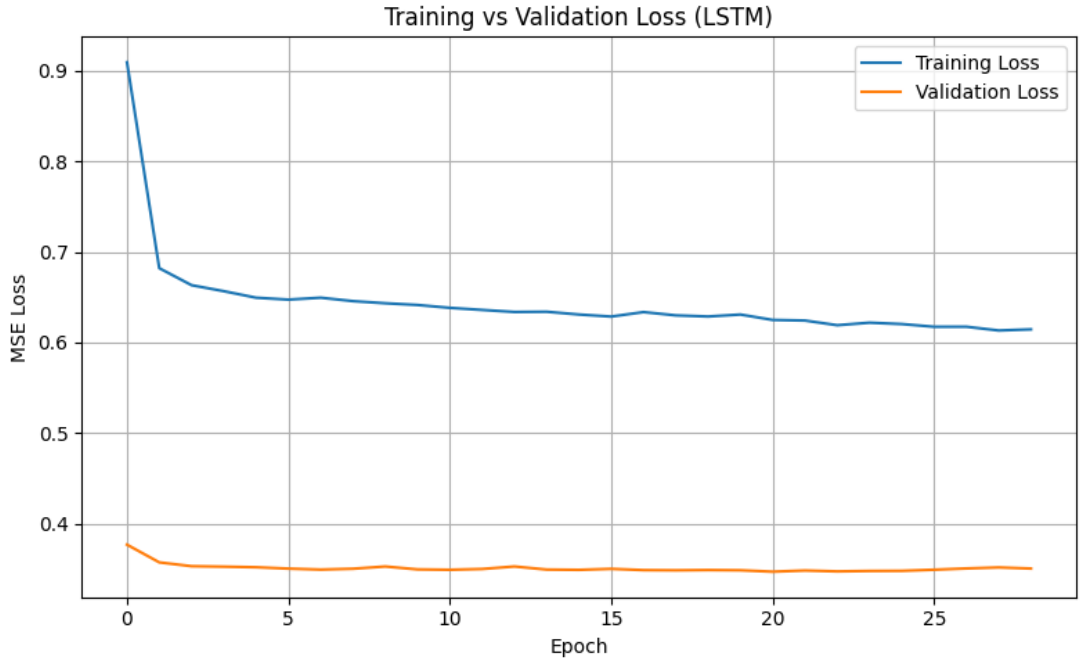


Figure 6. Training and validation loss curves during model convergence (one-step-ahead forecasting). Source: Own elaboration.

4.2. One-Step-Ahead Forecast Performance (Validation Set)

Model performance was evaluated on the validation subset using standard regression metrics computed in physical units (µg/m³) after inverse scaling. The evaluation protocol followed a blocked validation scheme to ensure that the one-hour-ahead forecasting objective was assessed on entirely unseen future data.

For the 168-hour input window, the multi-output LSTM achieved the following performance metrics:

- PM2.5: MAE = 2.943 µg/m³; RMSE = 3.812 µg/m³; R2 = 0.697
- PM10: MAE = 11.873 µg/m³; RMSE = 15.715 µg/m³; R2 = 0.702

These results indicate a strong short-term predictive capability, explaining approximately 70% of the variance for both pollutants. The extended 168-hour contextual window allows the architecture to encode weekly temporal patterns and delayed pollution dynamics, which are critical in urban environments governed by cyclical traffic and meteorological regimes.

The lower prediction error for PM2.5 may be attributed to its smoother temporal dynamics and stronger short-term autocorrelation structure. In contrast, PM${10}$ concentrations are influenced by heterogeneous and episodic sources such as resuspended dust, construction activity, and localized emissions, which increase hourly variability and reduce predictability (WHO, 2021; EPA, 2023).

To assess predictive alignment, Figure 7 displays the scatter plots of predicted versus observed concentrations for the validation set. The PM2.5 scatter plot shows wider dispersion, reflecting the reduced predictability at the hourly scale, a phenomenon widely reported in urban air quality studies due to the stochastic nature of coarse particles (Franceschi et al., 2018). Despite this dispersion, the high R2 values confirm that the model maintains a strong correlation with ground-truth observations across the entire validation period.

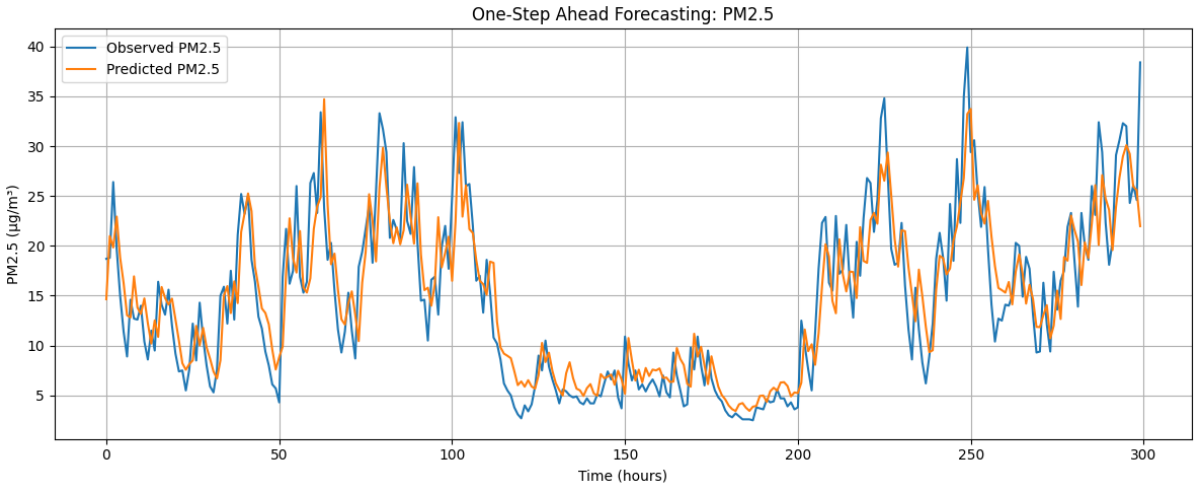


Figure 7. Predicted vs. observed concentrations (scatter plots). Source: Own elaboration

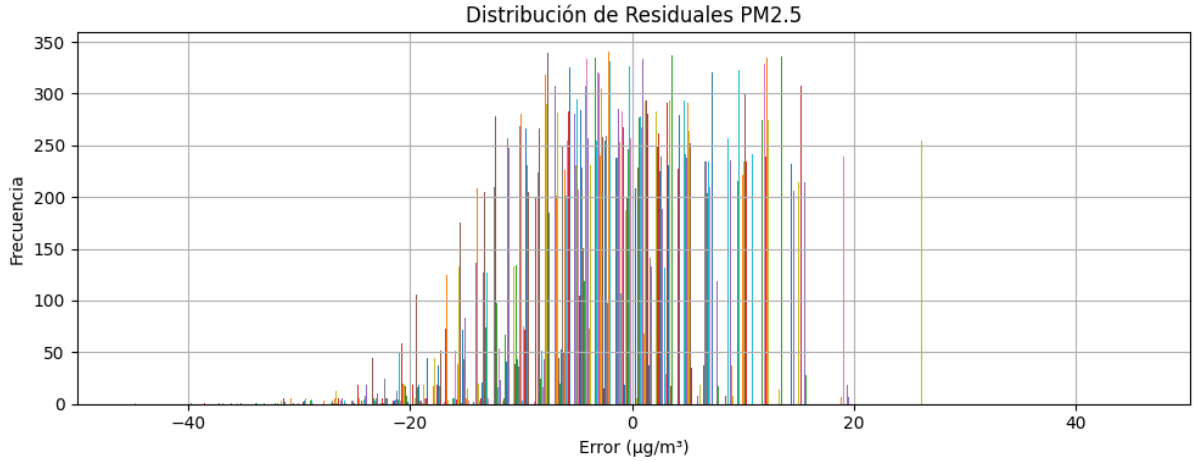
4.3 Error Distribution and Residual Analysis

The distribution of prediction errors was rigorously examined using error histograms and temporal residual plots (ypred − ytrue) for both pollutants. This analysis is critical for verifying the model's reliability and identifying potential systematic biases.

As shown in Figure 8, the error histograms for PM2.5 exhibit a near-zero, centered distribution with moderate spread (MAE = 2.94 µg/m³), indicating low systematic bias and stable short-term forecasting performance. In contrast, the PM10 error distribution exhibits greater variance and heavier tails. This suggests greater sensitivity to sudden spikes in concentration and exogenous disturbances, which is characteristic of the behavior of coarser particulate matter in urban corridors.

The residual plots over time reveal that most prediction errors remain within a narrow range under stable atmospheric conditions. However, larger residuals—indicating higher uncertainty—tend to occur during abrupt pollution episodes and rapid concentration transitions. This behavior is expected in environmental time series characterized by non-linearity, episodic emission events, and complex meteorological perturbations (Franceschi et al., 2018; Casallas García et al., 2021).

Importantly, the residual sequence does not exhibit a clear temporal autocorrelation pattern, suggesting that the 168-hour (7-day) input window successfully captures a substantial portion of the temporal dependency structure in the Bogotá Air Quality Monitoring Network (RMCAB) data. Nonetheless, the occasional underestimation of peak PM10 events points toward a limitation in modeling extreme short-term variability when only historical pollutant concentrations are utilized as predictors. This finding reinforces the need to integrate meteorological covariates in future work to further stabilize the error variance.



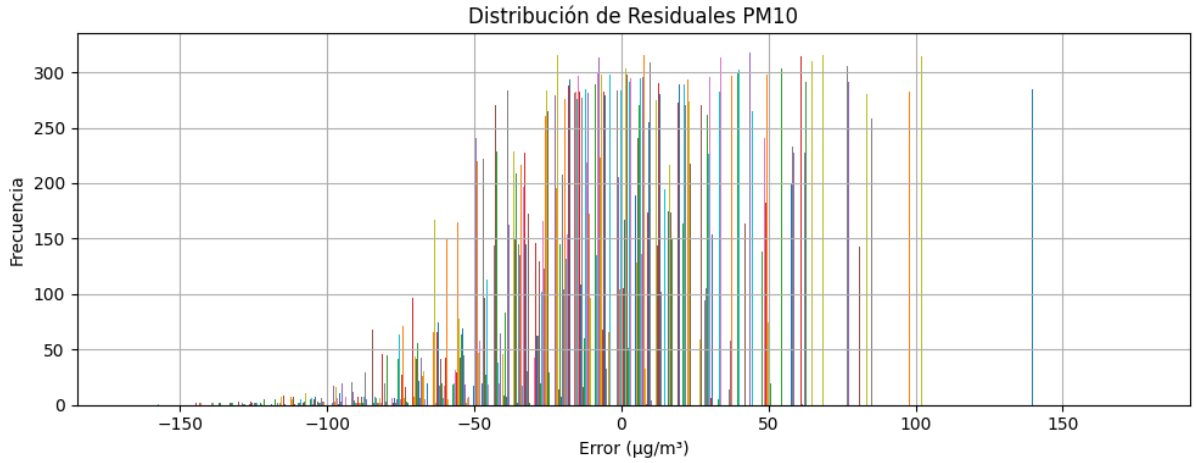


Figure 8. Error distribution histograms for PM2.5 and PM10 on the validation set. Source: Own elaboration

4.4. Baseline Considerations and Model Justification

To evaluate the robustness of the proposed multi-output LSTM, its performance was benchmarked against three standard forecasting approaches: (1) a Persistence model (naïve baseline), (2) a Linear Regression model with 24-hour lagged features, and (3) a Random Forest Regressor.

The persistence approach assumes that the concentration at t+1 equals the observation at t. While competitive in one-hour-ahead forecasting due to high temporal autocorrelation, it inherently fails to capture non-linear transitions and atmospheric regime shifts. Traditional RNNs were also considered but dismissed due to vanishing gradient constraints that limit their effectiveness for the 168-hour temporal window used here (Bengio et al., 1994; Pascanu et al., 2013). In contrast, the LSTM's gating mechanisms allow selective retention of information over extended horizons (Hochreiter & Schmidhuber, 1997; Gers et al., 2002). As summarized in Table 4, the LSTM model consistently outperformed all benchmarks across both pollutants.

Table 4. Performance comparison between the proposed LSTM and baseline models on the validation set.

Pollutant	Model	MAE (µg/m³)	RMSE (µg/m³)	R^2
PM2.5	LSTM (Proposed)	2.94	3.81	0.697
	Linear Regression	2.98	3.87	0.686
	Random Forest	3.08	3.98	0.669
	Persistence	3.10	4.18	0.636
PM10	LSTM (Proposed)	11.87	15.72	0.702
	Linear Regression	11.90	15.69	0.700
	Random Forest	12.08	15.98	0.689
	Persistence	12.21	16.49	0.672

Source: Own elaboration.

The results indicate that the LSTM architecture achieved the highest explained variance ($R2 \approx 0.70$), successfully mapping complex dependencies that classical lagged models struggle to represent. While Linear Regression showed competitive RMSE for PM10, the LSTM provided superior predictive coherence and lower Mean Absolute Error (MAE) for both particles.

Consistent with previous research (Casallas García et al., 2021), this performance gap confirms that the 168-hour input window, combined with the LSTM's memory cells, enhances temporal context encoding. This enables the model to learn weekly cycles and delayed accumulation patterns—capabilities that simpler models like

Persistence or Random Forest lack in high-granularity urban datasets. This validation strengthens the case for using deep learning architectures in developing Bogotá's early-warning systems.

4.5. Implications for Air Quality Management

From regulatory and urban management perspectives, accurate one-hour-ahead forecasting is a relevant step toward data-driven early warning systems in metropolitan areas (Alcaldía Mayor de Bogotá, 2017; MinAmbiente, 2017). The results demonstrate the technical feasibility of applying recurrent neural networks with extended temporal windows for short-term $PM_{2.5}$ and PM_{10} prediction in Bogotá's urban context.

However, the present implementation should be interpreted as a proof of concept rather than an operational forecasting system. The current model relies exclusively on historical pollutant concentrations and does not account for key exogenous drivers, including meteorological variables, traffic intensity, boundary-layer dynamics, or regional pollutant transport. Additionally, validation was performed using a single blocked temporal split and a single forecasting horizon.

Operational deployment would require multi-station datasets, walk-forward validation across multiple temporal blocks, and multi-horizon forecasting evaluation (e.g., 6 h, 12 h, and 24 h ahead). Hybrid architectures incorporating atmospheric and meteorological covariates could further improve robustness, peak prediction, and real-world applicability (Casallas García et al., 2021). Accordingly, conclusions regarding the applicability of early warning and regulatory decision support should be considered conditional on these methodological extensions and broader validation frameworks.

5. Limitations

Although the results demonstrate the potential of LSTM networks for air quality forecasting in Bogotá, several limitations restrict the generalizability and operational use of this proof of concept. First, the input space was limited to pollutant concentrations alone. The exclusion of meteorological variables such as temperature, relative humidity, wind speed, and atmospheric pressure reduces the model's ability to account for external drivers of particulate matter dynamics. These variables are known to strongly influence dispersion, accumulation, and chemical transformation processes, and their absence partially explains the weaker performance observed for PM_{10}.

Second, the dataset size constrained the training process. Although the Bogotá Air Quality Monitoring Network provides continuous hourly measurements, the relatively short historical window available for this study limited the diversity of pollution scenarios included in the training and validation sets. This scarcity increases the risk of overfitting and reduces the model's robustness across seasons and atypical conditions.

Third, the model's capacity to generalize to real-world scenarios remains limited. Situations such as wildfire smoke intrusions, holiday traffic peaks, or industrial events often cause abrupt changes in pollutant concentrations that the current model struggles to capture. Without extensive testing on such extreme events, the predictive accuracy demonstrated here should not be interpreted as readiness for operational deployment.

In summary, the model should be regarded as a preliminary step that demonstrates feasibility rather than a reliable tool for decision support. Future work must address these limitations by incorporating meteorological covariates, expanding the dataset through longer monitoring periods, and validating the approach under diverse real-world conditions.

6. Future development

Building on the limitations identified in this proof-of-concept, the next stage of this research involves fully implementing the intelligent air quality monitoring and advisory system illustrated in Figure 9. This proposed framework aims to evolve the current LSTM-based architecture into a comprehensive Decision Support System (DSS) that integrates high-performance data architectures, interactive interfaces, and advanced Artificial Intelligence components.

First, the system will transition from a standalone model to a robust database infrastructure capable of managing high-velocity streams of air-quality records. This will ensure efficient storage, retrieval, and automated pre-processing of RMCAB data. A specialized user interface will provide real-time interaction with the forecasting engine. Simultaneously, the integration of Large Language Models (LLMs) coupled with Retrieval-Augmented Generation (RAG) modules will enable the system to provide intelligent, context-aware responses grounded in both empirical evidence and local legal frameworks (Hochreiter & Schmidhuber, 1997; Pascanu, Mikolov, & Bengio, 2013; Chollet, 2015; Gers et al., 2002). These elements will transform the predictive model into an intelligent monitoring assistant capable of not only forecasting pollutant levels but also offering technical explanations and actionable recommendations for both the general public and environmental authorities.

Second, to enhance predictive robustness and reduce variance in PM10 forecasts, future iterations will incorporate meteorological covariates, including ambient temperature, relative humidity, and atmospheric pressure. These factors are critical drivers of particulate matter dispersion and concentration; their inclusion will allow the model to capture complex atmospheric dynamics more effectively (Pascanu et al., 2013; World Health Organization, 2021; Casallas García et al., 2021). Furthermore, expanding the training horizon to include multi-year historical datasets stored in optimized environments, such as SQLite or NoSQL databases, will improve the model's capacity to identify seasonal cycles and long-term trends (Abadi et al., 2016; Franceschi et al., 2018; Zaharia et al., 2020).

Third, the development of a natural language interface powered by LLMs will bridge the gap between complex numerical outputs and public understanding. Users will be able to query the platform through a graphical interface, receiving plain-language interpretations of forecasted pollution levels, automated health recommendations based on scientific literature, and context-specific references to regulatory mandates like Resolution 2254 of 2017 (Alcaldía Mayor de Bogotá, 2017; Ministerio de Ambiente y Desarrollo Sostenible, 2017). The RAG engine will ensure that these responses are dynamically enriched with information retrieved from verified legal databases and peer-reviewed articles, enhancing the transparency and reliability of the advisory system (Chollet, 2015; Docker Inc., 2020; U.S. EPA, 2023).

In summary, the proposed developments aim to evolve this experimental prototype into a holistic early-warning and advisory platform aligned with international sustainability agendas. By synergizing LSTMs, meteorological data, and generative AI, the platform will contribute to Sustainable Development Goal 13 (Climate Action) and support Bogotá's strategic efforts to reduce PM2.5 and PM10 exceedance days by 2030 (Ministerio de Ambiente y Desarrollo Sostenible, 2025).

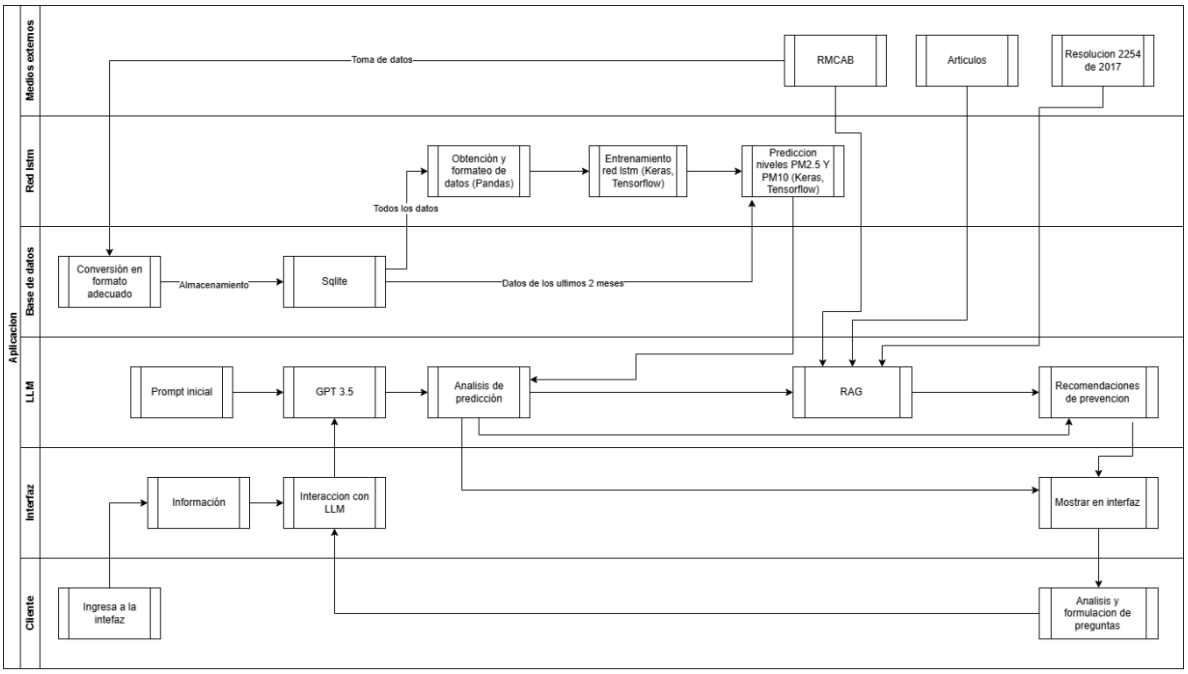


Figure 9. Process diagram for a prediction and recommendations system for $PM_{2.5}$ and PM_{10}. Source: Own elaboration

6. Conclusions

This study developed and tested a predictive system for hourly $PM_{2.5}$ and PM_{10} concentrations in Bogotá, based on 168-hour (7-day) temporal windows and a multi-output Long Short-Term Memory (LSTM) architecture with 64 recurrent units. Implemented in Keras/TensorFlow and trained with one year of data from the Bogotá Air Quality Monitoring Network (RMCAB), the model outperformed robust benchmarks, including persistence, Linear Regression with 24-hour lags, and Random Forest Regressors, particularly for $PM_{2.5}$. While the LSTM showed superior ability to capture non-linear temporal dependencies, performance for PM_{10} remains an area for further refinement, confirming that this research serves as a validated proof-of-concept for urban forecasting.

From a regulatory perspective, the system demonstrates potential as a foundation for early-warning frameworks aligned with Resolution 2254 of 2017. Nonetheless, its current accuracy and generalizability are

constrained by several limitations: the exclusion of meteorological covariates and the difficulty of capturing episodic pollution events such as wildfires or traffic peaks. Addressing these gaps will be essential before any deployment in real-world decision-making contexts.

The study highlights the importance of reproducibility and scalability in environmental modeling. The integration of collaborative platforms such as Google Colab, containerization with Docker, and monitoring frameworks such as MLflow and TensorBoard provides a transferable workflow for future projects (Zaharia et al., 2020).

Future work should expand the input space to include meteorological variables (temperature, humidity, atmospheric pressure), test hybrid architectures (e.g., CNN–LSTM, attention mechanisms), and evaluate forecasts over longer horizons (24–72 hours). Recent advances show that hybrid designs integrating CNNs, LSTMs, and attention can substantially reduce forecasting errors and improve robustness compared to standalone models (Zhang et al., 2023; Liang et al., 2025; Lv et al., 2024). Incorporating these strategies will be essential to move from experimental prototypes to operational systems for air quality management.

In summary, this research contributes a significant step toward data-driven air quality forecasting in Bogotá. It offers both methodological lessons and a pathway for developing more reliable, interpretable, and impactful predictive systems that can ultimately support public health strategies, regulatory compliance, and broader sustainability goals, including Sustainable Development Goal 13 (Climate Action).

6.1. Managerial Implications

For policymakers and environmental authorities, this research illustrates how machine learning can eventually support proactive air quality management in Bogotá. Even at a proof-of-concept stage, the framework suggests the potential to use LSTM-based models to anticipate pollution episodes and align responses with Resolution 2254 of 2017. If refined and validated, such tools could guide contingency measures, improve communication through Air Quality Index platforms, and foster greater public awareness. The study, therefore, highlights a pathway for integrating predictive analytics into regulatory practice while recognizing that further development is necessary before operational adoption.

6.2. Theoretical Implications

From an academic standpoint, this work contributes to the literature on deep learning for environmental forecasting by adapting a multi-output LSTM to simultaneously predict $PM_{2.5}$ and PM_{10}. This approach underscores the value of shared representations in time series modeling and provides an example of how recurrent architectures can be tailored to complex urban datasets. Beyond the model itself, the use of reproducible tools such as TensorFlow, Docker, and MLflow emphasizes the importance of transparent, scalable workflows in applied machine learning research. These insights open avenues for future studies exploring hybrid neural architectures and integrating meteorological and regulatory data, advancing both machine learning methodology and environmental science.

References:

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... Kudlur, M. (2016). TensorFlow: A system for large-scale machine learning. In *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI'16)* (pp.265–283). USENIX Association. https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi

Alcaldía Mayor de Bogotá. (2017). Resolución 2254 de 2017 Ministerio del Medio Ambiente. SISJUR. Retrieved from https://www.alcaldiabogota.gov.co/sisjur/normas/Norma1.jsp?i=82634

Área Metropolitana de Bogotá. (2025). Sistema de Monitoreo de la Calidad del Aire de Bogotá (RMCAB). Retrieved from http://rmcab.ambientebogota.gov.co/dynamicTabulars/index

Casallas García, A., Ferro, C., & Celis Mayorga, N. (2021). Long short-term memory artificial neural network approach to forecast meteorology and $PM_{2.5}$ local variables in Bogotá, Colombia. Modeling Earth Systems and Environment, 8(3), 2951–2964. https://doi.org/10.1007/s40808-021-01274-6

Chollet, F. (2015). Keras [Computer software]. GitHub. https://github.com/fchollet/keras

Docker Inc. (2020). What is Docker? Retrieved from https://docs.docker.com/get-started/docker-overview/

Franceschi, F., Cobo, M., & Figueredo, M. (2018). Discovering relationships and forecasting PM_{10} and $PM_{2.5}$ concentrations in Bogotá, Colombia, using artificial neural networks, principal component analysis, and k-means clustering. Atmospheric Pollution Research, 9(5), 912–922. https://doi.org/10.1016/j.apr.2018.02.006

Función Pública de Colombia. (2015). Decreto 1076 de 2015: Sector Ambiente y Desarrollo Sostenible. Gestor Normativo. Retrieved from https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=78153

Gers, F. A., Schraudolph, N. N., & Schmidhuber, J. (2002). Learning precise timing with LSTM recurrent networks. *Journal of Machine Learning Research, 3*, 115–143. https://www.jmlr.org/papers/volume3/gers02a/gers02a.pdf

Google. (2025). Welcome to Colab. Colaboratory. Retrieved from https://colab.research.google.com/

Google Research. (2025). Colaboratory FAQ. Retrieved from https://research.google.com/colaboratory/faq.html

Graves, A., Fernández, S., Gómez, F., & Schmidhuber, J. (2006). Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks. In Proceedings of the 23rd International Conference on Machine Learning (ICML '06) (pp. 369–376). ACM. https://doi.org/10.1145/1143844.1143891

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM). (2021). *Proporción de datos del Índice de Calidad del Aire (ICA), Bogotá* [Informe técnico]. http://archivo.ideam.gov.co/documents/11769/641368/2.01+HM+%C3%ADndice+calidad+aire.pdf/5130ffb3-a1bf-4d23-a663-b4c51327cc05

McKinney, W. (2010). Data structures for statistical computing in Python. In Proceedings of the 9th Python in Science Conference (pp. 51–56). SciPy. https://doi.org/10.25080/Majora-92bf1922-00a

Ministerio de Ambiente y Desarrollo Sostenible. (2017). Resolución 2254 de 2017. Retrieved from https://www.minambiente.gov.co/wp-content/uploads/2021/10/Resolucion-2254-de-2017.pdf

Ministerio de Ambiente y Desarrollo Sostenible. (2025). Contaminación atmosférica. Retrieved from https://www.minambiente.gov.co/asuntos-ambientales-sectorial-y-urbana/contaminacion-atmosferica/

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... Chintala, S. (2019). PyTorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 32). Curran Associates, Inc. https://papers.neurips.cc/paper_files/paper/2019/file/bdbca288fee7f92f2bfa9f7012727740-Paper.pdf

Pascanu, R., Mikolov, T., & Bengio, Y. (2013). On the difficulty of training recurrent neural networks. In S. Dasgupta & D. McAllester (Eds.), *Proceedings of the 30th International Conference on Machine Learning* (Vol. 28, pp. 1310–1318). PMLR. https://proceedings.mlr.press/v28/pascanu13.html

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830. https://jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf

U.S. Environmental Protection Agency. (2023). Particulate matter (PM) basics. U.S. Environmental Protection Agency. Retrieved September 16, 2025, from https://www.epa.gov/pm-pollution/particulate-matter-pm-basics

van der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). The NumPy array: A structure for efficient numerical computation. Computing in Science & Engineering, 13(2), 22–30. https://doi.org/10.1109/MCSE.2011.37

World Health Organization. (2021). WHO global air quality guidelines: Particulate matter ($PM_{2.5}$ and PM_{10}), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. Geneva: WHO. Retrieved from https://apps.who.int/iris/handle/10665/345329

Zaharia, M., Chen, A., Davidson, A., Ghodsi, A., Hong, M., Konwinski, A., & Xin, R. (2020). Accelerating the machine learning lifecycle with MLflow. IEEE Data Engineering Bulletin, 43(3), 34–45. Retrieved from https://people.eecs.berkeley.edu/~matei/papers/2018/ieee_mlflow.pdf

Duan, J., Gong, Y., Luo, J., & Zhao, Z. (2023). Air-quality prediction based on the ARIMA–CNN–LSTM combination model optimized by dung beetle optimizer. Scientific Reports, 13(1), 12127. https://doi.org/10.1038/s41598-023-36620-4

Sreenivasulu, T., & Mokesh Rayalu, G. (2025). Accurate hourly AQI prediction using temporal CNN–LSTM–MHA+GRU: A case study of seasonal variations and pollution extremes in Visakhapatnam, India. Results in Engineering, 27, 106303. https://doi.org/10.1016/j.rineng.2025.106303

Nguyen, A. T., Pham, D. H., Oo, B. L., Ahn, Y., & Lim, B. T. H. (2024). Predicting air quality index using attention hybrid deep learning and quantum-inspired particle swarm optimization. Journal of Big Data, 11, Article 71. https://doi.org/10.1186/s40537-024-00926-5

Latin American Business and Sustainability Review
Academic Journal

Article

Accessibility meets the Science of Reading in Spanish English Dual language classrooms

Irene Oramas Sosa [1], Carlos Zarzalejo [2]

Citation: Oramas-Sosa, I. & Zarzalejo, C. (2026). Accessibility meets the Science of Reading in Spanish-English dual language classrooms. LABSREVIEW, 1(2): 24-36. https://doi.10.70469/labsreview.v3i1.39
Academic Editor: Jorge Martín Diez.
Received: 12/12/2025
Revised: 3/24/2026
Accepted: 5/19/2026
Published: 7/14/2026

[1] Empire State University; irene.oramassosa@sunyempire.edu
[2] La Era Digita; azarzalejo@gmail.com
* Correspondence: irene.oramassosa@sunyempire.edu

Abstract: This article advances a conceptual framework defining inclusive biliteracy—the coordinated development of literacy across Spanish and English within instructional systems designed to be equitable by default—and argues that effective biliteracy instruction requires coordinated design across four dimensions: structured literacy, cross-linguistic development, UDL-based accessibility, and technology-supported instruction. Drawing on an integrative review across cognitive reading science, bilingual education, accessibility research, and educational technology, the article establishes that foundational literacy components remain essential in bilingual contexts when coordinated across languages and supported by oral language development and cross-linguistic transfer; that multilingual learners with disabilities achieve strong outcomes under systematic, scaffolded, accessible instruction; and that assistive and AI-mediated technology extends access when aligned with instructional goals. The framework positions inclusive biliteracy as both an educational construct and an infrastructural condition for equity and human capital development, and generates five testable propositions operationalizable through multilevel modeling, structural equation modeling, or cluster-randomized designs. The contribution is conceptual: a theoretically grounded, multi-domain model specifying the conditions under which biliteracy instruction becomes inclusive for all learners.

Keywords: accessibility; bilingual education; Science of Reading; Spanish English dual language; AI-mediated learning.

1. Introduction

U.S. public schools educate a large and growing population of students who develop literacy across more than one language. English learners represented 10.6% of all public-school students in 2021, or approximately 5.3 million learners, with higher concentrations in the early grades. In kindergarten, this proportion reaches 14.7% before declining as some students are reclassified as English proficient (National Center for Education Statistics, 2024). Multilingualism is not peripheral to elementary education. It is a defining condition that must shape the design and delivery of literacy.

Spanish–English dual language immersion (DLI) programs respond directly to this reality. These programs aim to develop bilingualism, biliteracy, and academic achievement through sustained instruction in two languages. National reports document their continued expansion across states and districts. The What Works Clearinghouse defines dual language programs as models that provide at least 50% of instruction in a partner language and extend across the elementary grades. Evidence from RAND and related studies shows positive long-term academic outcomes, particularly for English learners (Institute of Education Sciences, What Works Clearinghouse, 2022; Steele et al., 2017).

Early literacy remains a critical threshold within this context. In grades K–3, approximately 15% of students are multilingual learners. By third grade, students are expected to transition from learning to read to reading to

learn, a shift that continues to shape instructional priorities and intervention systems. Students who do not develop strong foundational literacy skills during this period face increased risk of long-term academic difficulty (Amplify Education, 2023; Annie E. Casey Foundation, 2010). Literacy design in the early grades carries lasting consequences for multilingual learners.

The Science of Reading (SoR) has become a dominant force in shaping literacy policy and instructional practice. The term refers to converging evidence from cognitive psychology, linguistics, neuroscience, and education on how reading develops and how it can be taught effectively. In practice, SoR is associated with explicit and systematic instruction in phonological awareness, phonics, decoding, word recognition, fluency, vocabulary, oral language, and comprehension (Castles et al., 2018; Snowling et al., 2022). These principles influence legislation, curriculum adoption, and teacher preparation across the United States.

Current SoR implementation frameworks, however, contain three specific structural limitations that this article directly addresses. First, most presuppose phonemic inventories and orthographic features native to English, creating a phonological mismatch when applied to Spanish-dominant learners without modification. Second, existing SoR frameworks lack a principled model for cross-language instructional bridging – for determining when and how to make the relationships between two linguistic systems explicit. Third, the monolingual assessment benchmarks embedded in prevailing SoR-aligned policy frameworks (e.g., state literacy laws) increase the risk of misidentification for multilingual learners and fail to account for the bilingual development trajectories documented in the research literature. These are not peripheral concerns; they are structural gaps that undermine the effectiveness of SoR implementation in dual language contexts (Kittle et al., 2024; National Academies of Sciences, Engineering and Medicine, 2017).

Spanish–English bilingual learners illustrate these dynamics clearly. Spanish and English share alphabetic features but differ in orthographic transparency, phonology, and morphology. Spanish supports early decoding through consistent grapheme–phoneme relationships, while English introduces greater variability. Skills such as phonological awareness, vocabulary, and metalinguistic knowledge operate across languages when instruction makes these relationships explicit. Cross-linguistic transfer supports reading development when instruction is coordinated across both languages (Míguez-Álvarez et al., 2022; Sun et al., 2021; van der Velde et al., 2019).

Literacy instruction in dual language classrooms must address linguistic complexity alongside learner variability. Students differ in language proficiency, disability status, prior educational experience, and access to resources. Instruction that assumes uniform development creates barriers to participation. Accessibility is not an additional layer. It is a core condition of effective literacy instruction in bilingual settings.

Universal Design for Learning (UDL) provides a framework for designing instruction that anticipates variability from the outset. UDL emphasizes multiple means of engagement, representation, and action and expression (Meyer et al., 2014). In bilingual literacy contexts, this includes multimodal supports, scaffolded language opportunities, and a clear separation between learning goals and the means used to access them. UDL supports rigorous instruction by ensuring that all learners can engage with it.

Bilingual learners with disabilities require explicit inclusion within these systems. Research shows that these students achieve biliteracy and academic success in dual language programs when instruction is intentional, coordinated, and adequately supported (Genesee et al., 2022). Exclusion reflects institutional design, not learner capacity. Bilingualism serves as an asset for cognitive and academic development when supported by equitable instructional practices (Cummins, 1983; Genesee et al., 2022).

Educational technology further shapes literacy instruction. Digital tools, assistive technologies, and AI-mediated systems support access, practice, and feedback. In Spanish–English contexts, these tools provide multimodal input, pronunciation support, text-to-speech, and structured language scaffolds. For students with disabilities, they reduce barriers related to decoding, written expression, and language production. Their effectiveness depends on alignment with instructional goals and pedagogical design (Hurwitz et al., 2022; Wood et al., 2018).

This article defines literacy instruction in Spanish–English dual language settings as an integrated system. It advances a conceptual framework that aligns structured literacy, bilingual development, accessibility, and technology-supported instruction. The analysis examines how SoR-informed instruction operates across languages, how UDL supports access, and how technology extends biliteracy development. This framework positions inclusive biliteracy as a condition for equitable and sustainable educational systems.

2. Literature Review

This review draws on five intersecting bodies of scholarship: cognitive reading science, bilingual education, sociocultural theory, accessibility-oriented instructional design, and educational technology. Together, these domains explain how literacy develops in bilingual learners, how instruction operates across languages, and how access is structured for diverse learners. The review focuses on points of convergence that support a coherent model of biliteracy instruction.

2.1 Science of Reading and Foundational Literacy

The Science of Reading (SoR) literature establishes reading as an integrated process involving decoding and linguistic comprehension. The Simple View of Reading defines comprehension as the product of these two components (Gough & Tunmer, 1986; Hoover & Gough, 1990). Scarborough's Reading Rope extends this model by identifying the interacting strands that contribute to reading development, including phonological awareness, decoding, vocabulary, language structures, and background knowledge (Scarborough, 2001).

The National Reading Panel identified phonemic awareness, phonics, fluency, vocabulary, and comprehension as central components of effective instruction (National Reading Panel, 2000). Subsequent work situates these components within broader linguistic systems, including morphology, syntax, and semantics (Snowling et al., 2022; Seidenberg, 2017).

This body of research is widely applied in policy and practice. However, its original scope did not center on multilingual learners, and the current SoR-aligned policy framework has largely extended monolingual models without systematic adaptation. Critics have noted that applying phonics-first SoR models without modification can pathologize reading development that follows a bilingual rather than a monolingual trajectory (cf. Paris, 2005). This framework directly addresses that concern: the problem is not with structured literacy as such, but with its decontextualized application to populations whose linguistic development differs structurally from the populations on which the original research was conducted. Current scholarship extends SoR principles to bilingual contexts by examining how foundational skills operate across languages (Castles et al., 2018; Kittle et al., 2024).

2.2 Bilingual Literacy and Cross-Linguistic Development

Literacy development in bilingual learners is shaped by cross-linguistic interaction. August and Shanahan (2006) demonstrate that literacy-related skills transfer across languages and that instruction must address oral language, linguistic relationships, and prior experiences with print. The National Academies (2017) define home languages as developmental resources that support cognition, identity, and academic learning. Cummins's theory of linguistic interdependence explains how proficiency in one language supports literacy development in another through shared underlying processes (Cummins, 1979).

Research on cross-linguistic transfer specifies these relationships. Transfer depends on orthographic structure, typological similarity, oral language proficiency, and metalinguistic awareness. In Spanish–English contexts, phonological awareness, decoding, and metalinguistic skills transfer across languages when instruction makes these relationships explicit. Vocabulary development requires direct instruction in both languages (Míguez-Álvarez et al., 2022; Sun et al., 2021; van der Velde et al., 2019). Biliteracy development, therefore, requires coordinated instruction across languages: foundational skills remain central, but their acquisition is shaped by linguistic interaction rather than isolated instruction.

2.3 Dual Language Education and Program Design

Research on dual language education defines the institutional and instructional conditions that support biliteracy. Lindholm-Leary (2001) shows that sustained implementation and instructional quality are central to long-term outcomes. Howard et al. (2018) define dual language education as a comprehensive model that integrates curriculum, assessment, leadership, and community engagement. Biliteracy development depends on coordinated program design rather than language allocation alone. Large-scale studies confirm that dual language programs support academic achievement while promoting bilingualism and biliteracy (Institute of Education Sciences, What Works Clearinghouse, 2022; Steele et al., 2017). These findings establish that structured literacy and bilingual instruction operate within the same instructional system, not as parallel but unconnected tracks.

2.4 Sociocultural Perspectives on Literacy

Literacy development in bilingual contexts is shaped by social interaction and cultural meaning-making. Vygotsky's theory positions learning as socially mediated, with language functioning as a central tool in cognitive development. The concept of funds of knowledge demonstrates that learners draw on cultural and linguistic resources in meaning-making processes (Moll et al., 1992).

The multiliteracies framework (New London Group, 1996) and translanguaging theory (García & Wei, 2014) both expand what counts as literacy beyond print and monolingual norms. Translanguaging theory defines bilingual language use as an integrated system in which learners mobilize their full linguistic repertoire dynamically rather than switching between two separate codes. This framework draws productively on this insight while noting an important boundary: translanguaging frameworks center language use but do not operationalize instruction in SoR-aligned foundational skills. Multiliteracies theory expands the definition of text but does not

address disability access systematically. The integrated model proposed here extends both frameworks by adding explicit instructional design requirements (structured literacy) and proactive accessibility planning (UDL) that neither framework addresses in depth. This is the specific contribution of the inclusive biliteracy framework over existing approaches.

2.5 Accessibility and Universal Design for Learning

Universal Design for Learning (UDL) provides a framework for designing instruction that anticipates learner variability rather than retrofitting accommodations after barriers emerge. UDL defines three core principles: multiple means of engagement, representation, action and expression (Meyer et al., 2014). In literacy instruction, UDL supports access to rigorous content by varying how learners engage with material and demonstrate understanding. Foundational skills instruction remains explicit and systematic, while access pathways are flexible.

Accessibility in bilingual contexts includes both disability-related supports and linguistic access. Instruction must remove barriers related to decoding, language comprehension, and expression. UDL operationalizes access by aligning instructional goals with multiple entry points, ensuring that the construct being assessed is not conflated with the mode of access. This distinction between what is taught and how it is accessed is critical for maintaining both rigor and equity.

2.6 Bilingual Learners with Disabilities

Research on bilingual learners with disabilities establishes inclusion as a core condition of equitable literacy instruction. Cummins (1983) challenged deficit-based assumptions that restricted access to bilingual education for students with disabilities. Recent research confirms that bilingual learners with disabilities achieve strong academic and linguistic outcomes in dual language programs when instruction is coordinated and supported (Genesee et al., 2022). Exclusion reflects institutional barriers rather than incompatibility between bilingualism and disability.

This literature also identifies persistent issues of misidentification and disproportionality. Assessment systems that do not account for bilingual development increase the risk of inappropriate classification. Instructional systems must distinguish language development from disability and provide equitable access to biliteracy pathways.

2.7 Educational Technology

Educational and assistive technology expands access to, practice of, and feedback in literacy instruction. Assistive tools such as text-to-speech, read-aloud systems, and speech recognition support comprehension and expression for students with reading and language-based disabilities (Wood et al., 2018). Digital literacy platforms support structured skill development when integrated with explicit instructional goals (Hurwitz et al., 2022). In bilingual contexts, translation scaffolds, bilingual glossaries, and pronunciation tools support vocabulary and comprehension when used as part of coordinated instruction.

AI-mediated tools represent an emerging layer with potential for adaptive feedback and personalized practice (Paglialunga & Melogno, 2025). The evidence base remains preliminary. Critics have raised a substantive concern: that AI-mediated and algorithmic literacy tools risk reducing instruction to decontextualized skill rehearsal, displacing teacher judgment and culturally sustaining practice (cf. Selwyn, 2019, on the datafication of education). This framework addresses this concern directly by establishing that technology must be pedagogically aligned and teacher-mediated; it cannot function as a standalone solution. This constraint is a feature, not a limitation, of the model.

2.8 Synthesis and Theoretical Positioning

Literature establishes a coherent model of literacy development. Reading science defines foundational processes. Bilingual education research shows that these processes operate across languages through transfer and coordinated instruction. Sociocultural theory positions literacy as socially mediated and culturally grounded. UDL defines access as a core feature of instructional design. Research on bilingual learners with disabilities establishes inclusion as essential to equitable systems. Technology literature identifies conditions under which digital tools support rather than displace learning.

These domains converge on a single conclusion: biliteracy instruction requires coordinated design across linguistic, cognitive, social, and accessibility dimensions. The framework proposed in this article is distinguished from existing models in three specific ways. First, unlike translanguaging frameworks, it operationalizes foundational skill instruction explicitly across languages rather than treating language use as inherently flexible and student-directed. Second, unlike multiliteracies frameworks, it systematically addresses the needs of learners

with disabilities through UDL. Third, unlike SoR implementation frameworks as currently applied in policy, it builds cross-linguistic transfer and bilingual developmental trajectories into the architecture of literacy instruction rather than treating them as supplementary adaptations.

3. Conceptual Framework for Inclusive Biliteracy

3.1 Framework Overview

This article advances a conceptual framework for inclusive biliteracy built on four coordinated dimensions: (1) structured literacy aligned with the Science of Reading, (2) bilingual and cross-linguistic development, (3) accessibility through Universal Design for Learning (UDL), and (4) technology-supported instruction.

Inclusive biliteracy, as a construct, is defined along four corresponding dimensions. The linguistic dimension refers to proficiency and literacy development in both Spanish and English, including the ability to mobilize cross-linguistic knowledge to support comprehension and production. The accessibility dimension refers to the proactive removal of barriers for learners with disabilities and linguistic differences, designed into instruction from the outset rather than added as accommodation after the fact. The instructional design dimension refers to coordinated, cross-linguistic, structured instruction aligned with SoR, and to explicit, cumulative, and systematically adapted instruction across both languages. The participation dimension refers to equitable engagement in biliteracy pathways regardless of disability status or language background. This four-part definition distinguishes inclusive biliteracy from general biliteracy frameworks (which do not systematically address disability access) and from UDL-only approaches (which do not address cross-linguistic transfer). The term "inclusive biliteracy framework" is used throughout to refer to the instructional design model; "inclusive biliteracy" refers to the educational outcome construct.

The framework defines literacy instruction as a system rather than a sequence of independent components. Structured literacy establishes the developmental architecture of reading. Bilingual development determines how those skills are acquired and applied across languages. Accessibility ensures that all learners can engage with instruction. Technology extends access and supports differentiation. When a single dimension is isolated or underdeveloped, instructional coherence breaks down, and access is reduced.

3.2 Core Constructs

3.2.1 Structured Literacy (Science of Reading)

Structured literacy defines the essential components of reading development: phonological awareness, phonics, fluency, vocabulary, and comprehension. Instruction is explicit, systematic, and cumulative. In bilingual contexts, these components are not duplicated across languages. They are coordinated to reflect differences in orthography, phonology, and morphology. Spanish supports early decoding through consistent sound-symbol correspondence; English requires greater flexibility in mapping print to sound. Instruction must account for both shared and language-specific features, making structural differences visible through explicit metalinguistic comparison.

3.2.2 Bilingual and Cross-Linguistic Development

Biliteracy develops through interaction between two linguistic systems. Skills such as phonological awareness, decoding, and metalinguistic knowledge transfer across languages when instruction makes these relationships explicit. Spanish and English differ in transparency, phonotactics, and morphological structure. Instruction must address these differences directly while supporting cross-linguistic connections, including attention to cognates, morphological patterns, syllable structures, and phoneme–grapheme correspondences.

3.2.3 Accessibility and Universal Design for Learning (UDL)

Accessibility is a design condition, not an accommodation. UDL operationalizes this principle through multiple means of engagement, representation, and action and expression. In bilingual literacy contexts, accessibility includes both disability-related supports and linguistic access. The critical distinction is between the construct being taught and the mode through which it is accessed. This distinction allows rigorous content to remain stable while access pathways vary. Inclusion of multilingual learners with disabilities is not a specialized pathway; it is the baseline expectation of equitable instructional design.

3.2.4 Technology-Supported Instruction

Technology serves as instructional support, extending access to practice and feedback. It serves two distinct functions: a compensatory function (reducing barriers to access for students with disabilities) and a developmental function (supporting practice and skill acquisition). These functions are not interchangeable. Instruction must distinguish between tools that provide access and those that build capacity and must deploy each accordingly. Technology is effective when it is aligned with instructional goals and integrated within the broader literacy system. When used in isolation, it does not produce sustained learning outcomes.

3.3 Relationships Among Constructs

The four constructs operate as an integrated system with reciprocal relationships:

- Structured literacy defines what is taught.
- Bilingual development defines how literacy skills transfer and develop across languages.
- Accessibility defines who can participate and under what conditions.
- Technology defines how instruction is extended and supported.

Instructional effectiveness depends on alignment across all four dimensions. Structured literacy that is not coordinated across languages limits transfer. Bilingual instruction that does not incorporate accessibility restricts participation. Technology that is not aligned with pedagogy functions as an add-on rather than as an instructional support. Alignment produces coherence. Misalignment produces fragmentation.

3.4 Outcome: Inclusive Biliteracy Development

The immediate outcome of this framework is inclusive biliteracy development, defined as:

- Proficiency in reading across Spanish and English
- Development of cross-linguistic awareness and transfer
- Access to grade-level literacy instruction for learners with diverse profiles
- Sustained participation through accessible and responsive design

At the broader level, inclusive biliteracy functions as part of the educational infrastructure. It determines whether learners can access, engage with, and benefit from schooling, and downstream, whether they can participate in multilingual economies, civic institutions, and professional pathways. The long-term human capital and sustainability outcomes described in Section 5 are contingent on the four instructional dimensions being operative simultaneously. This connection integrates the sustainability argument into the core of the framework rather than appending it as a separate consideration.

3.5 Propositions for Future Research

This framework generates testable propositions that can guide empirical research:

Proposition 1: Coordinated structured literacy instruction across Spanish and English produces stronger biliteracy outcomes than monolingual implementation models. Measurable outcomes: Oral Reading Fluency (ORF) and Developmental Reading Assessment (DRA) scores in both languages. Feasible design: quasi-experimental comparison using multilevel modeling (students nested within classrooms and schools).

Proposition 2: Explicit cross-linguistic instructional design strengthens transfer of phonological, morphological, and metalinguistic skills. Measurable outcomes: phonological awareness task performance across languages, morpheme identification scores. Feasible design: pre-post experimental study with cross-linguistic bridging as the instructional condition.

Proposition 3: Accessibility-informed literacy instruction increases participation and reading outcomes for multilingual learners, including students with disabilities. Measurable outcomes: participation rates in biliteracy pathways, longitudinal literacy trajectory data by disability status. Feasible design: longitudinal comparison of UDL-implementing vs. non-implementing DLI programs.

Proposition 4: Technology enhances differentiation and access when it is aligned with structured literacy and UDL principles. Measurable outcomes: comprehension scores, task completion rates for students with and without assistive technology. Feasible design: randomized assignment to technology-integrated vs. technology-as-supplement conditions within classrooms.

Proposition 5: Alignment among structured literacy, bilingual development, accessibility, and technology yields more equitable and sustained literacy outcomes than the isolated implementation of any single component. Feasible designs: (a) structural equation modeling estimating path coefficients among the four dimensions toward a latent biliteracy outcome; (b) cluster-randomized trial comparing whole-framework vs. partial implementation conditions.

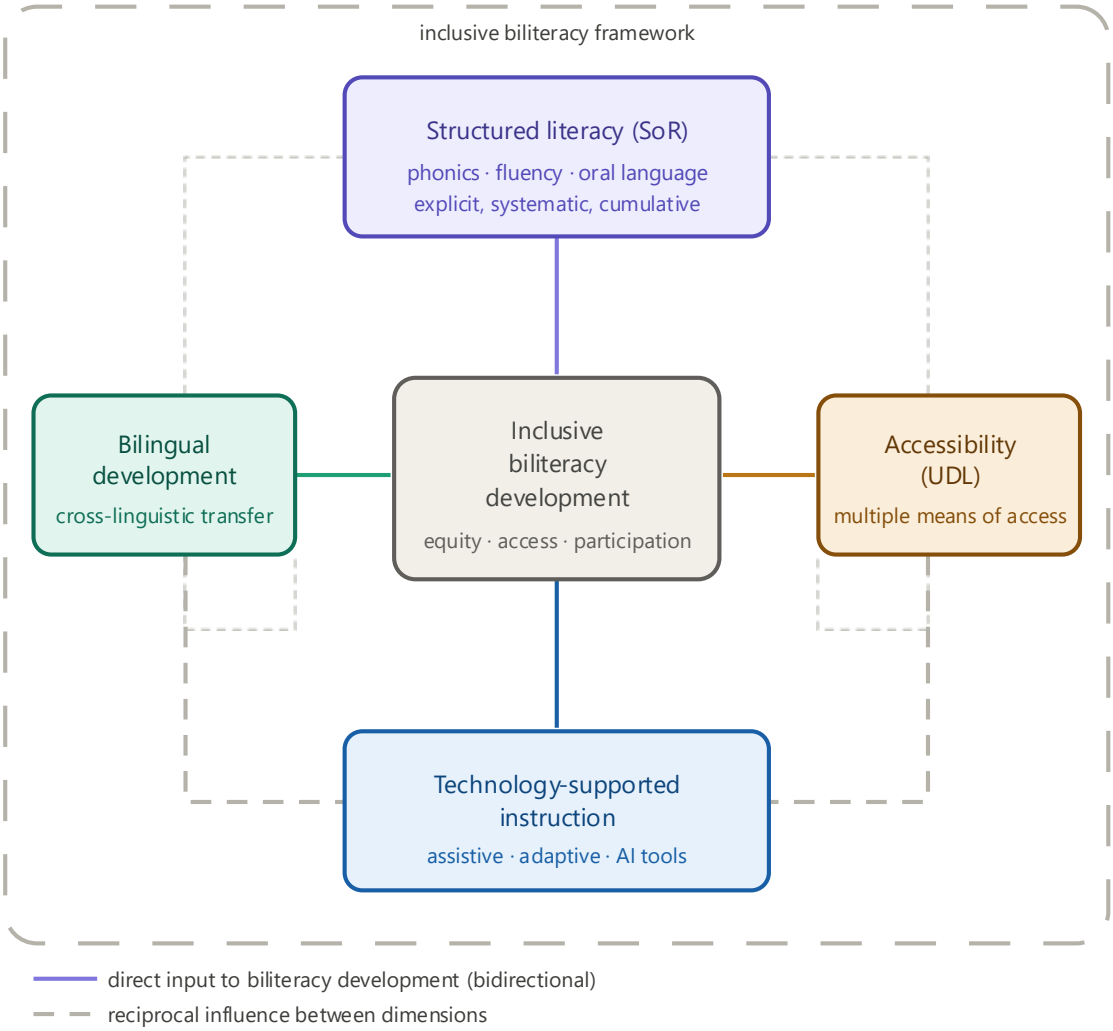


Figure 1. Conceptual framework for inclusive biliteracy integrating structured literacy (SoR), bilingual development, accessibility (UDL), and technology-supported instruction.

Figure 1 presents the framework. Each of the four dimensions provides direct, bidirectional input to inclusive biliteracy development (solid arrows) and exerts reciprocal influence on the other dimensions (dashed lines), within an integrated instructional system.

4. Materials and Methods

4.1 Research Design

This study uses an integrative theoretical literature review to synthesize scholarship across multiple domains and to construct a conceptual framework for literacy instruction in Spanish–English dual-language contexts. Integrative reviews are suited to research problems that span disciplines, incorporate diverse forms of evidence, and require conceptual development rather than narrow intervention analysis (Torraco, 2005, 2016; Whittemore & Knafl, 2005; Snyder, 2019). The research problem requires engagement with cognitive reading science, bilingual education, sociocultural theory, accessibility and disability studies, Universal Design for Learning, and educational technology. No single domain captures the full structure of biliteracy instruction. This design supports the identification of key constructs, the analysis of relationships among them, and the development of a unified explanatory model (Jabareen, 2009). This study is theory-building. Systematic review principles inform it, but it does not function as a meta-analysis or a bounded systematic review of intervention effects (Grant & Booth, 2009).

4.2 Review Purpose and Guiding Questions

The purpose of the review is to synthesize interdisciplinary scholarship to answer the following conceptual question: How can Science of Reading–informed literacy instruction be designed for Spanish–English dual language classrooms through coordinated attention to bilingual development, accessibility, and technology?

Four guiding questions structure the analysis:

- What scholarship defines the core components of Science of Reading–aligned literacy instruction?
- What does the literature on bilingualism and biliteracy indicate about cross-linguistic transfer and language-specific adaptation?
- How do UDL, accessibility research, and disability studies inform equitable participation in literacy instruction?
- How does educational and assistive technology extend access and support biliteracy development?

4.3 Search Strategy and Source Identification

The review used a structured, iterative search process across five domains: reading science, bilingual education, sociocultural and inclusive theory, accessibility and UDL, and educational technology. Search methods included keyword searches across academic databases (ERIC, PsycINFO, Google Scholar), backward and forward citation tracking from anchor studies, and targeted identification of foundational and highly cited works. Search terms included combinations of: "Science of Reading," "structured literacy," "dual language education," "Spanish–English bilingual," "biliteracy," "cross-linguistic transfer," "multilingual learners," "Universal Design for Learning," "accessibility," "students with disabilities," "assistive technology," and "AI in literacy."

The final corpus comprised approximately 35 sources spanning the period 1979–2025, distributed across five domains: cognitive reading science, bilingual and dual language education, sociocultural and inclusive theory, accessibility and disability studies, and educational technology, with three policy and institutional reports included for contextual grounding. Foundational works established theoretical grounding, while recent studies captured current developments in multilingual literacy and technology.

4.4 Inclusion and Exclusion Criteria

Sources were included if they: established theoretical or conceptual foundations in reading science, bilingual education, UDL, sociocultural theory, or disability inclusion; provided empirical or synthesized evidence relevant to literacy, biliteracy, or multilingual learners; addressed Spanish–English contexts, cross-linguistic transfer, accessibility, or instructional design; or focused on multilingual learners, students with disabilities, or their intersection. Sources were excluded if they: focused exclusively on monolingual populations without transferable relevance, lacked methodological rigor or scholarly grounding, presented technical descriptions without instructional or conceptual implications, or duplicated stronger or more comprehensive sources. Policy and practitioner documents were included selectively to contextualize implementation.

4.5 Source Appraisal and Analytic Procedure

Sources were evaluated based on their contribution to the conceptual framework rather than treated as equivalent in evidentiary weight. Empirical studies were assessed for methodological transparency and alignment between evidence and claims. Foundational and theoretical works were evaluated for conceptual clarity and sustained scholarly influence.

The analytic process was qualitative and iterative. Sources were organized into a domain-based matrix: cognitive reading science, bilingualism and biliteracy, sociocultural and inclusive theory, accessibility and disability, and educational technology. Within each domain, texts were coded for: core constructs (e.g., decoding, transfer, accessibility), instructional principles, theoretical claims, and points of convergence across domains. Codes were grouped into higher-order themes, including structured literacy, cross-linguistic transfer, oral language development, learner variability, accessibility through UDL, inclusive participation, and technology as instructional support. Constant comparison across sources ensured that claims reflected patterns across the literature rather than individual studies (Braun & Clarke, 2006; Saldaña, 2016).

Where sources produced conflicting findings, patterns were evaluated based on the preponderance of the evidence rather than on individual studies. Conflicts arising from differences in sample characteristics (e.g., monolingual vs. bilingual populations, age range, disability status) or instructional context were noted and incorporated as boundary conditions within the framework's propositions rather than treated as disconfirming evidence. This approach is consistent with established integrative review methodology (Whittemore & Knafl, 2005).

4.6 Framework Construction, Reflexivity, and Limitations

The final stage translated thematic patterns into a conceptual model. The analysis adopts a defined interpretive stance: it focuses on Spanish–English dual language contexts in U.S. elementary education, treats bilingualism and disability as dimensions of learner variability rather than deficits, and distinguishes reading science as a body of scholarship from its policy interpretation.

Several limitations apply. The review synthesizes heterogeneous sources and does not provide pooled effect sizes or formal evidence grading. The inclusion of both empirical and conceptual work requires interpretive judgment. Technology and AI literature are evolving, and some findings remain preliminary. The contribution of this study is conceptual: it provides a theoretically grounded framework that organizes existing scholarship into a coherent model for inclusive biliteracy instruction.

5. Results

The literature establishes that the core components of the Science of Reading, including phonological awareness, phonics, word recognition, fluency, vocabulary, oral language, and comprehension, remain essential in Spanish–English dual language programs. These components do not change in bilingual contexts. Their implementation does.

Instruction must account for the interaction between two linguistic systems. Foundational skills develop through coordinated exposure to Spanish and English rather than through isolated, monolingual sequences. A recent synthesis of research on multilingual learners confirms that SoR-aligned instruction supports reading development when it incorporates oral language, linguistic assets, and the relationship between decoding and language proficiency (Kittle et al., 2024). Cross-linguistic transfer functions as a central mechanism in this process. Skills developed in one language support reading development in another, particularly in domains such as phonological awareness, decoding, and metalinguistic knowledge. A meta-analysis of bilingual learners identifies consistent relationships between phonological awareness and reading outcomes across languages (Míguez-Álvarez et al., 2022). Longitudinal research on Spanish–English learners shows that early literacy development in Spanish supports later word reading in English (Sun et al., 2021).

Transfer is not automatic; it depends on the explicit instructional design described in Section 3.2. Empirically, when learners engage in structured cross-linguistic comparison, they develop metalinguistic awareness that strengthens reading development in both systems (Míguez-Álvarez et al., 2022; Sun et al., 2021; van der Velde et al., 2019).

Oral language operates as a structural component of reading development. Vocabulary, syntax, and discourse knowledge support comprehension across languages. Multilingual learners often develop these domains unevenly across Spanish and English. Instruction must integrate oral language development with foundational skills rather than treat it as a separate domain (Kittle et al., 2024; National Academies of Sciences, Engineering, and Medicine, 2017). The evidence supports a unified instructional model. Structured literacy and bilingual pedagogy operate within the same system. When foundational skills instruction is explicit and coordinated across languages, dual language programs support both reading development and bilingual proficiency (Howard et al., 2018; Lindholm-Leary, 2001; Steele et al., 2017).

Accessibility defines participation in biliteracy instruction. Multilingual learners with disabilities have historically been excluded from bilingual programs through referral practices, program design, and deficit-based assumptions about capacity. The literature rejects these assumptions: multilingual learners with disabilities develop biliteracy and academic competence when instruction is coordinated, scaffolded, and sustained (Genesee et al., 2022). Outcomes are not constrained by bilingualism. They are shaped by access to instruction.

Applying the UDL principles defined in Section 3.2.3, accessible biliteracy instruction provides multimodal input, structured language supports, visual representation, and flexible options for demonstrating understanding—keeping learning goals stable while access pathways vary, so that rigor and access are maintained together.

Inclusive biliteracy instruction requires systematic scaffolding: oral language modeling, guided practice, visual supports, and structured opportunities for cross-language interaction. These supports benefit multilingual learners broadly and are essential for students with identified disabilities. Assessment practices must also reflect bilingual development; monolingual assessment systems increase the risk of misidentification, and evaluations must account for language proficiency, cross-linguistic transfer, and variation in language exposure.

Consistent with the compensatory and developmental functions distinguished in Section 3.2.4, the evidence shows technology amplifies rather than replaces foundational teaching. Assistive technologies—text-to-speech, read-aloud systems, speech recognition—provide the strongest evidence of impact, improving access to grade-level content for students with reading and language-based disabilities while instruction continues to develop foundational skills (Wood et al., 2018). Digital literacy tools support structured skill engagement when integrated with explicit pedagogical goals; without that alignment, their impact is limited (Hurwitz et al., 2022).

In bilingual contexts, technology also supports linguistic access through translation scaffolds, bilingual glossaries, and pronunciation tools. AI-mediated tools offer potential for adaptive feedback and personalized

practice, though the current evidence base is preliminary (Paglialunga & Melogno, 2025). The compensatory and developmental functions of technology are distinct and must be deployed accordingly: tools that provide access to content are not substitutes for tools that build literacy capacity, and neither replaces direct instruction.

Technology is effective when aligned with structured literacy, bilingual development, and accessibility principles. When integrated within this system, it extends instructional reach and supports participation. When treated as an independent solution—or when AI tools are allowed to substitute for teacher-mediated, culturally sustaining instruction—it does not produce sustained literacy outcomes and may reinforce the decontextualized skill-rehearsal patterns that critics of educational technology have documented (Selwyn, 2019).

5.1 Relevance to Sustainability, Human Capital, and Public Policy

Inclusive biliteracy instruction carries implications beyond classroom practice that connect directly to the four framework dimensions established in Section 3. The long-term outcomes described in this section- expanded human capital, workforce participation, and intergenerational mobility- are downstream consequences of whether structured literacy is coordinated across languages, whether bilingual development is supported through explicit cross-linguistic design, whether accessibility is embedded through UDL, and whether technology is integrated with rather than substituted for direct instruction. Sustainable educational systems require all four conditions to be operative simultaneously.

International frameworks identify equitable access to education as central to sustainable development. Sustainable Development Goal 4 and the Education 2030 agenda define literacy, inclusion, and participation as core components of educational systems (United Nations, 2015; UNESCO, 2016). These frameworks emphasize access to quality education across linguistic and social contexts, which requires precisely the kind of inclusive instructional design this framework operationalizes.

Biliteracy development contributes directly to human capital formation. Proficiency across languages expands access to academic content, professional pathways, and civic participation. In Spanish–English contexts, biliteracy supports engagement in multilingual economies and institutions (World Bank, 2018). In Latin American and Latinx contexts, language, access to education, and structural inequality shape patterns of opportunity; instructional models that integrate bilingual development, accessibility, and participation expand pathways into education and employment across generations.

The relationship between literacy and economic participation is well established: individuals with strong literacy skills demonstrate higher rates of employment, income stability, and civic engagement. When literacy systems exclude multilingual learners or students with disabilities by failing to address any one of the four framework dimensions, these outcomes are unevenly distributed. Educational design decisions produce long-term consequences for participation at scale.

6. Discussion

The analysis establishes that the Science of Reading, bilingual education theory, and accessibility-oriented design converge on a shared instructional model. This convergence is not incidental. It reflects a consistent pattern across the literature: literacy development depends on the interaction of decoding, language, access, and participation.

Two countervailing positions warrant acknowledgment before elaborating on the implications. First, some scholars caution that rigid phonics-first SoR models can pathologize the reading development of multilingual learners whose phonemic systems differ from English, effectively treating bilingual developmental patterns as deficits (Paris, 2005). This framework addresses this concern by requiring language-specific adaptation of structured literacy rather than wholesale transfer of monolingual SoR models; the problem is in the application, not in structured literacy as a principle. Second, critics of AI-mediated and digital literacy tools argue that algorithmic feedback risks reducing literacy to decontextualized skill rehearsal, displacing teacher judgment and culturally responsive practice (Selwyn, 2019). The framework's condition that technology must align with structured literacy and UDL goals and cannot function as a standalone solution directly addresses this risk. Naming these tensions sharpens the framework's scope conditions.

Structured literacy and bilingual pedagogy operate within the same system. Explicit and systematic instruction in foundational skills supports multilingual learners when it is coordinated across languages. Cross-linguistic transfer, oral language development, and language-specific adaptation shape how these skills develop. Instruction that isolates languages or applies monolingual models limits this process and misses the central mechanism through which biliteracy develops.

Accessibility is embedded in the architecture of effective instruction, not appended to it. Multilingual learners with disabilities participate fully in biliteracy development when supports are proactive and systematically designed. Access to bilingualism is not a specialized pathway; it is part of equitable educational

design. When accessibility is not embedded from the outset, participation is restricted regardless of instructional quality elsewhere in the system.

Technology operates as an instructional mediator within this system. Digital and assistive tools extend opportunities for practice, feedback, and access to content, but only when their compensatory and developmental functions are clearly distinguished and aligned with literacy goals. Technology does not strengthen instruction on its own; it extends instruction that is already coherent.

The current evidence base presents clear limitations. Multilingual learners and students with disabilities remain underrepresented in research that informs large-scale literacy policy. Implementation claims often extend beyond the populations included in foundational studies. Future research must examine cross-linguistic sequencing of foundational skills, the role of oral language in bilingual comprehension, the impact of accessibility on literacy trajectories, and the conditions under which technology produces durable outcomes in authentic classroom contexts. The five propositions in Section 3.5 provide specific, operationalized directions for this research.

Culturally and linguistically sustaining pedagogy remains central to this model. Learners' linguistic resources, cultural knowledge, and identities shape engagement and meaning-making. Instruction that incorporates these dimensions supports comprehension and participation across languages and integrates naturally with the framework's sociocultural foundation.

7. Conclusions

This article establishes that literacy instruction in Spanish–English dual language classrooms operates as an integrated system that aligns structured literacy, bilingual development, accessibility, and technology. The Science of Reading, defined as a multidisciplinary body of evidence on decoding, word recognition, language comprehension, fluency, vocabulary, and meaning-making, supports biliteracy development when implemented within a multilingual and accessibility-oriented framework (August & Shanahan, 2006; Kittle et al., 2024; National Academies of Sciences, Engineering, and Medicine, 2017).

The inclusive biliteracy framework proposed here is distinguished from existing models in three ways: it operationalizes foundational skill instruction explicitly across languages (unlike translanguaging frameworks), it systematically addresses the needs of learners with disabilities through UDL (unlike multiliteracies frameworks), and it builds cross-linguistic transfer and bilingual developmental trajectories into the architecture of instruction (unlike SoR frameworks as currently applied in policy). The four-dimensional construct definition—linguistic, accessibility, instructional design, and participation—provides a basis for operationalization, measurement, and empirical testing.

Structured literacy remains essential. Its implementation requires coordination across Spanish and English to reflect orthographic, phonological, and morphological differences. Foundational skills develop through explicit instruction in both languages, supported by cross-linguistic awareness and sustained oral language development. Biliteracy does not emerge from parallel monolingual instruction; it develops through coordinated design.

Accessibility defines whether learners can participate in biliteracy instruction. Multilingual learners with disabilities develop literacy and academic competence when instruction is scaffolded, linguistically responsive, and systematically designed from the start. Access to bilingualism and biliteracy is a matter of educational equity, not program eligibility.

Technology supports biliteracy development by extending access, practice, and feedback—but only when its compensatory and developmental functions are clearly distinguished and aligned with instructional goals. Emerging AI-mediated tools show potential for adaptive support; the same alignment requirements as other technologies must govern their integration.

Future research should examine cross-linguistic sequencing of foundational skills, the role of oral language in bilingual comprehension, the impact of accessibility on literacy trajectories, and the function of technology in authentic classroom contexts. The five propositions in Section 3.5 provide operationalized directions for this work, with specific outcome variables and research designs.

Inclusive biliteracy functions as part of the educational infrastructure. It shapes access to learning, participation in schooling, and long-term social and economic outcomes. Instruction that integrates structured literacy, bilingual development, accessibility, and technology—all four dimensions, in coordination—supports equitable literacy development and expands educational opportunity in multilingual societies.

References

Annie E. Casey Foundation. (2010). *Early warning! Why reading by the end of third grade matters*.https://ed.psu.edu/sites/default/files/inline-files/Anne%20E%20Casey%202010%20Early%20Warning%20Special_Report_Executive_Summary.pdf

August, D., & Shanahan, T. (Eds.). (2006). *Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth. Lawrence Erlbaum.* https://doi.org/10.1080/1086296X.2010.503745

Braun, V., & Clarke, V. (2006). *Using thematic analysis in psychology.* Qualitative Research in Psychology*, 3(2), 77–101*. https://doi.org/10.1191/1478088706qp063oa

Castles, A., Rastle, K., & Nation, K. (2018*). Ending the reading wars: Reading acquisition from novice to expert.* Psychological Science in the Public Interest*,* 19(1), 5–51. https://doi.org/10.1177/1529100618772271

Cummins, J. (1979). *Linguistic interdependence and the educational development of bilingual children*. Review of Educational Research, 49(2), 222–251. https://doi.org/10.2307/1169960

Cummins, J. (1983). *Bilingualism and special education: Program and pedagogical issues*. Learning Disability Quarterly, 6(4), 373–386. https://doi.org/10.2307/1510525

García, O., & Wei, L. (2014). *Translanguaging: Language, bilingualism and education*. Palgrave Macmillan. https://doi.org/10.1057/9781137385765

Genesee, F., Hilliard, J., Sánchez-López, C., & Young, T. (2022). *Welcoming bilingual learners with disabilities into dual language programs: Research-informed implications for inclusive programs and practice*. Center for Applied Linguistics. https://www.cal.org/wp-content/uploads/2022/09/Welcoming-Bilingual-Learners-with-Disabilities-into-Dual-Language-Programs-3.pdf

Gough, P. B., & Tunmer, W. E. (1986). *Decoding, reading, and reading disability*. Remedial and Special Education, 7(1), 6–10. https://doi.org/10.1177/074193258600700104

Grant, M. J., & Booth, A. (2009). *A typology of reviews: An analysis of 14 review types and associated methodologies*. Health Information and Libraries Journal, 26(2), 91–108. https://doi.org/10.1111/j.1471-1842.2009.00848.x

Hammer, C. S., Hoff, E., Uchikoshi, Y., Gillanders, C., Castro, D. C., & Sandilos, L. E. (2014). *The language and literacy development of young dual language learners: A critical review*. Early Childhood Research Quarterly, 29(4), 715–733. https://doi.org/10.1016/j.ecresq.2014.05.008

Hoover, W. A., & Gough, P. B. (1990). *The simple view of reading*. Reading and Writing, 2, 127–160. https://doi.org/10.1007/BF00401799

Howard, E. R., Lindholm-Leary, K. J., Rogers, D., Olague, N., Medina, J., Kennedy, B., Sugarman, J., & Christian, D. (2018). *Guiding principles for dual language education (3rd ed.*). Center for Applied Linguistics. https://dlenm.org/wp-content/uploads/2024/11/Guiding-Principles-for-Dual-Language-Education-3rd-edition.pdf

Hurwitz, L. B., Solari, E. J., Parsley, J., & Puranik, C. (2022). *Educational technology in support of elementary students with disabilities*. AERA Open, 8. https://doi.org/10.1177/00222194221141093

Institute of Education Sciences, What Works Clearinghouse. (2022). *Dual language programs intervention report. U.S. Department of Education*. https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/WWC_DLP_IR-Report.pdf

Jabareen, Y. (2009). *Building a conceptual framework: Philosophy, definitions, and procedure*. International Journal of Qualitative Methods, 8(4), 49–62. https://doi.org/10.1177/160940690900800406

Kittle, J. M., Amendum, S. J., & Budde, C. M. (2024). *What does research say about the science of reading for K–5 multilingual learners? A systematic review of systematic reviews*. Educational Psychology Review. Advance online publication. https://doi.org/10.1007/s10648-024-09942-6

Lindholm-Leary, K. J. (2001). *Dual language education*. Multilingual Matters. https://www.multilingual-matters.com/page/detail/?k=9781853595318

Meyer, A., Rose, D. H., & Gordon, D. (2014). *Universal Design for Learning: Theory and practice*. CAST. https://udltheorypractice.cast.org

Míguez-Álvarez, C., Miller-Cotto, D., Cuadro, A., & Auxiliadora Moreira, M. (2022). *Relationships between phonological awareness and reading in bilingual children: A meta-analysis*. Language Learning, 72(S2), 667–706. https://doi.org/10.1111/lang.12493

Moll, L. C., Amanti, C., Neff, D., & González, N. (1992). *Funds of knowledge for teaching: Using a qualitative approach to connect homes and classrooms*. Theory Into Practice, 31(2), 132–141. https://doi.org/10.1080/00405849209543534

National Academies of Sciences, Engineering, and Medicine. (2017). *Promoting the educational success of children and youth learning English: Promising futures*. National Academies Press. https://doi.org/10.17226/24677

National Center for Education Statistics. (2024). *English learners in public schools.* The Condition of Education 2024. U.S. Department of Education. https://nces.ed.gov/programs/coe/indicator/cgf/english-learners-in-public-schools

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. National Institute of Child Health and Human Development. https://www.nichd.nih.gov/sites/default/files/publications/pubs/nrp/Documents/report.pdf

New London Group. (1996). *A pedagogy of multiliteracies: Designing social futures*. Harvard Educational Review, 66(1), 60–92. https://doi.org/10.17763/haer.66.1.17370n67v22j160u

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). *The PRISMA 2020 statement: An updated guideline for reporting systematic reviews*. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71

Paglialunga, A., & Melogno, S. (2025). *The effectiveness of artificial intelligence-based interventions for students with learning disabilities: A systematic review*. Brain Sciences, 15(8), 806. https://doi.org/10.3390/brainsci15080806

Paris, S. G. (2005). *Reinterpreting the development of reading skills*. Reading Research Quarterly, 40(2), 184–202. https://doi.org/10.1598/RRQ.40.2.3

Saldaña, J. (2016). *The coding manual for qualitative researchers (3rd ed.).* SAGE. https://study.sagepub.com/saldanacoding3e

Scarborough, H. S. (2001). *Connecting early language and literacy to later reading (dis)abilities: Evidence, theory, and practice.* In S. B. Neuman & D. K. Dickinson (Eds.), Handbook of early literacy research (pp. 97–110). Guilford Press.

Seidenberg, M. S. (2017). *Language at the speed of sight: How we read, why so many can't, and what can be done about it*. Basic Books. https://www.hachettebookgroup.com/titles/mark-seidenberg/language-at-the-speed-of-sight/9780465080656/

Selwyn, N. (2019). *Should robots replace teachers? AI and the future of education*. Polity Press. https://politybooks.com/bookdetail/?isbn=9781509528967

Snyder, H. (2019). *Literature review as a research methodology: An overview and guidelines*. Journal of Business Research, 104, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039

Snowling, M. J., Hulme, C., & Nation, K. (Eds.). (2022). *The science of reading: A handbook (2nd ed.)*. Wiley. https://doi.org/10.1002/9781119705116

Steele, J. L., Slater, R., Zamarro, G., Miller, T., Li, J. J., Burkhauser, S., & Bacon, M. (2017). *Dual-language immersion programs raise student achievement in English. RAND Corporation*. https://www.rand.org/pubs/research_briefs/RB9903.html

Sun, X., Marks, R. A., Pratt, A. S., Slusser, E. B., Bedore, L. M., & Kovelman, I. (2021). *What's in a word? Cross-linguistic influences on Spanish–English bilingual children's word reading development*. Journal of Child Language, 48(5), 983–1006. https://doi.org/10.1017/S0305000920000665

Torraco, R. J. (2005). *Writing integrative literature reviews: Guidelines and examples*. Human Resource Development Review, 4(3), 356–367. https://doi.org/10.1177/1534484305278283

Torraco, R. J. (2016). *Writing integrative literature reviews: Using the past and present to explore the future*. Human Resource Development Review, 15(4), 404–428. https://doi.org/10.1177/1534484316671606

UNESCO. (2016). *Education 2030: Incheon Declaration and Framework for Action for the implementation of Sustainable Development Goal 4*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000245656

United Nations. (2015). *Transforming our world: The 2030 Agenda for Sustainable Development*. https://sdgs.un.org/2030agenda

van der Velde, L., Kremin, L. V., & de Bree, E. H. (2019). *The effects of Spanish heritage language literacy on English reading for Spanish–English bilingual children in the United States*. International Journal of Bilingual Education and Bilingualism. Advance online publication. https://doi.org/10.1080/13670050.2016.1239692

Whittemore, R., & Knafl, K. (2005). *The integrative review: Updated methodology*. Journal of Advanced Nursing, 52(5), 546–553. https://doi.org/10.1111/j.1365-2648.2005.03621.x

Wood, S. G., Moxley, J. H., Tighe, E. L., & Wagner, R. K. (2018*). Does use of text-to-speech and related read-aloud tools improve reading comprehension for students with reading disabilities? A meta-analysis*. Journal of Learning Disabilities, 51(1), 73–84. https://doi.org/10.1177/0022219416688170

World Bank. (2018). *World development report 2018: Learning to realize education's promise*. World Bank. https://www.worldbank.org/en/publication/wdr2018

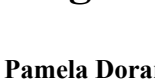
Latin American Business and Sustainability Review
Academic Journal

Article

Critical Machine Translation Praxis in Tourism Discourse: Translating Latin American Destination Slogans Across AI-Based MT Engines

Pamela Doran [1], and Carlos Zarzalejo [2]

Citation: Doran, P. & Zarzalejo, C. (2026). Critical Machine Translation Praxis in Tourism Discourse: Translating Latin American Destination Slogans Across AI-Based MT Engines. LABSREVIEW, 1(2): 24-36. https://doi.10.70469/labsreview.v3i1.47.
Academic Editor: Jorge Martín Diez
Received: 10/5/2025
Revised: 1/21/2026
Accepted: 16/5/2026
Published: 7/14/2026



[1] SUNY Empire State University; pamela.doran@sunyempire.edu
[2] La Era Digita; azarzalejo@gmail.com
* Correspondence: pamela.doran@sunyempire.edu

Abstract: Tourism slogans function as cultural invitations and are often among the first messages encountered by international travelers. Increasingly, these messages are mediated through free machine translation (MT) platforms such as Google Translate, DeepL, and Microsoft Translator. This pilot study examines whether these systems transmit not only semantic meaning but also the persuasive tone embedded in tourism discourse. To address this question, the study introduces Critical Machine Translation Praxis, an analytical approach that combines Baker's taxonomy of translation shifts with an affective semiotic perspective. Forty Latin American tourism slogans were translated across three MT engines and compared with human benchmark translations. Each output was evaluated for fidelity, translation-shift patterns, and persuasive effect. The findings reveal a notable divide: although 52.5% of the translations maintained high semantic fidelity, many of the outputs exhibited tonal flattening, excessive literalism, or culturally misleading cues. DeepL produced the most natural-sounding translations overall, while Google Translate and Microsoft Translator showed greater variability, particularly in the handling of idioms, affective nuance, and culturally embedded expressions. These results suggest that tonal drift and lexical imprecision can weaken the persuasive force of tourism slogans. The study argues that translation quality in tourism communication should not be reduced to accuracy alone; rather, emotional resonance, cultural nuance, and destination appeal remain central to how places are represented, interpreted, and received by global audiences.

Keywords: Affective blindness, Artificial Intelligence, intercultural communication, machine translation, tourism slogans.

1. Introduction

Tourism is a cornerstone of Latin America's economy, contributing billions of dollars annually and sustaining millions of jobs. In this highly competitive sector, slogans serve as strategic assets, condensing national identity and brand promise into a few memorable words. For many destinations, these short phrases are the first, and sometimes only, encounter international audiences have with a country's image. Increasingly, that encounter is mediated by free machine translation (MT) systems such as Google Translate, DeepL, and Microsoft Translator.

As tourists browse websites, social media posts, or promotional campaigns, these tools often provide the gateway through which slogans are understood. This reliance reaches beyond convenience. For travelers with limited language proficiency, or for those relying on screen readers, machine translation can determine how

accessible and inclusive a destination appears. When a translation flattens tone or distorts cultural meaning, the effect extends beyond weakened persuasion. It risks creating barriers to access and shaping perceptions in ways that exclude rather than welcome. At stake is not only linguistic accuracy, but also the preservation of persuasive force, emotional resonance, and accessibility that global tourism communication requires.

Machine translation systems have made substantial progress in grammatical and semantic accuracy, yet they continue to display what can be described as "affective blindness" or the inability to perceive and reproduce emotional undertones. In tourism marketing, persuasion relies not only on linguistic precision but also on tone, rhythm, and cultural resonance. Even small distortions can carry significant consequences. A slogan crafted to radiate warmth may appear flat, while an idiom intended to spark excitement may surface as awkward or confusing. For example, the Spanish expression "se me sale el corazón del pecho", which conveys overwhelming joy, frequently emerges in MT as "my heart comes out of my chest," a literal rendering that risks unsettling visitors rather than welcoming them.

This pilot study introduces Critical MT Praxis, an applied framework that brings together Baker's (1992) taxonomy of translation shifts with an affective semiotic lens. The framework evaluates not only structural distortions, level, category, structure, unit, and intra-system, but also the degree to which a slogan's emotional and cultural force is preserved in translation. By linking technical analysis with intercultural communication theory, Critical MT Praxis foregrounds both meaning and affect as core criteria for evaluation.

Our exploratory analysis of 40 Latin American tourism slogans highlights both the promise and the limits of contemporary MT Just over half (52.5%) of slogans retained clarity and persuasive tone. Nearly half (47.5%), however, suffered meaningful losses through tone flattening, misleading literalism, or diluted cultural cues. These distortions matter: mistranslations can weaken brand identity, erode trust, and, in some cases, create accessibility risks for multilingual users or those relying on assistive technologies.

The central research question guiding this pilot is:

To what extent do free MT tools preserve the persuasive and affective qualities of Latin American tourism slogans, and what risks emerge when they do not?

Answering this question has both academic and practical value. For scholars, it provides an empirical case of how affective blindness manifests in short-form marketing discourse. For practitioners, it offers a straightforward pre-launch screening method: test slogans in free MT engines, identify distortions, and revise as needed.

By integrating semantic and affective dimensions, Critical MT Praxis makes significant contributions to translation studies, intercultural communication, and marketing ethics. It underscores that in tourism, language is never just informational; it is persuasive, affective, and inclusive, shaping how destinations are experienced and trusted by global audiences.

2. Literature Review

Tourism descriptions and slogans have long been recognized as a domain where language functions not only to describe destinations but also to persuade and shape identity (Dann, 1996). Slogans carry particular force, condensing cultural resonance and emotion into short, memorable lines. However, when these slogans circulate through machine translation (MT), much of their persuasive impact can be weakened or lost.

2.1 Baker's Shift Taxonomy

Baker's (1992) taxonomy of translation shifts, level, category, structure, unit, and intra-system. It offers a systematic way to document semantic distortions. While MT systems are increasingly accurate at the structural level, they struggle with affective nuance. This limitation can be described as "affective blindness", or the incapacity of neural systems to reproduce tone, warmth, or cultural resonance. Barthes's (1980) distinction between studium (the shared, literal meaning) and punctum (the affective spark that moves us) provides a valuable lens. MT often retains studium but erases punctum, flattening language into information devoid of affect. For example, "Colombia es pasión" becomes "Colombia is passion", a translation that is lexically correct but emotionally flat. Similarly, idioms such as "se me sale el corazón del pecho" appear as "my heart comes out of my chest," producing grotesque or confusing imagery rather than heartfelt excitement. These distortions are not trivial or small. They illustrate how affective meaning, central to persuasion, is lost through machine-rendered Literalism.

2.2 Intercultural Communicative Competence (ICC)

Byram (1997) defines intercultural communicative competence (ICC) as the ability to navigate across cultures with curiosity, empathy, and sensitivity. Human translators draw on ICC when choosing between “usted” and “tú," or when reshaping metaphors so they resonate in another cultural context. MT systems, by contrast, predict word sequences without awareness of context. As House (2015) observes, this often yields surface-level accuracy but socially off-key phrasing. A typical case is Curaçao, where "siente el calor" is rendered as "feel the warmth" by a human translator, but MT typically outputs "feel the heat." This can give the impression of discomfort instead of hospitality. ICC highlights why humans still outperform machines in contexts where persuasion depends on affect and cultural nuance.

2.3 Research Gap and Contribution

Most MT research focuses on extended texts such as literature, speeches, or idioms (Bowker & Buitrago Ciro, 2015; Klimova, 2025). Short-form marketing discourse, despite its commercial and cultural weight, has received less systematic study. Tourism slogans are particularly high-stakes: they function as "headline texts" that condense identity, evoke emotion, and influence economic outcomes.

Recent studies confirm this gap. Chen (2025) shows that ChatGPT, when guided by culturally tailored prompts, can outperform Google Translate and DeepL in translating tourism text. It achieves higher levels of fidelity, fluency, cultural sensitivity, and persuasiveness, though human oversight remains essential. Freskila and Jayantini (2025) find that Google Translate more often preserves pragmatic and connotative nuance, whereas DeepL defaults to formal equivalence. The differences that meaningfully shape how tourism websites are perceived. Complementing these findings, Naveen (2024) surveys the broader limitations of machine translation, noting idiomatic and affectively dense texts as persistent weaknesses. Together, these studies reinforce the need for targeted research on slogans, where cultural resonance and emotional tone are condensed into the briefest forms.

Foundational work in translation studies reinforces this gap. Nida's (1964) distinction between formal equivalence (literal accuracy) and dynamic equivalence (reader response) anticipated the very tension observed here: MT tends to favor surface-level accuracy at the cost of persuasive effect. Venuti (1995) similarly argues that translation is never neutral but always positions cultural values, with "domestication" and “foreignization" shaping how texts are received. These classic insights underscore why slogans, condensed cultural performances, are particularly vulnerable to distortion when processed through MT.

This pilot study addresses that gap by applying Critical MT Praxis, an integrated framework that combines Baker's taxonomy, affective semiotics, and ICC, to a corpus of 40 Latin American slogans.

The framework rests on three insights:

- Semantic distortions (Baker's shifts) can be systematically measured.
- Emotional flattening (affective blindness) weakens persuasion even when lexical accuracy is preserved.
- Cultural literacy (ICC) is essential for adapting slogans with sensitivity and respect.

2.4 Applied AI, Accessibility, and Custom GPTs

Recent advances in applied AI suggest possible responses. Affective computing (Picard, 1997) and resources such as EmoBank (Buechel & Hahn, 2017) and GoEmotions (Demszky et al., 2020) enable fine-grained emotional annotation beyond simple polarity categories. Spanish-language resources such as MESD, EMOVOME, and MASIVE demonstrate how culturally grounded corpora can mitigate affective flattening.

At the same time, generative AI platforms such as ChatGPT are emerging as practical tools for marketers, allowing them to test slogans across languages, simulate tone, and refine phrasing iteratively. Custom GPTs extend this potential by integrating glossaries, tone guidelines, intercultural rules, and accessibility audits. For example, terms such as "rico" can be locked to "enriching" rather than "rich," and "calor" to "warmth" rather than "heat." Beyond accuracy, such systems can flag ambiguous terms, suggest alternative phrasing, and align with ADA Title II and WCAG 2.2 accessibility standards, which are critical for travelers who use assistive technologies or multilingual platforms.

By connecting affective computing with practical AI tools, this study moves the literature from critique to solution. It reframes MT not only as a risk to cultural fidelity but also as an opportunity to co-create inclusive, tone-sensitive tools that reflect brand identity and ethical responsibility.

3. Materials and Methods

3.1 Corpus Construction

This pilot study used a corpus of 40 tourism slogans compiled from official tourism boards across Latin America and the Caribbean. Selection criteria prioritized brevity (fewer than 120 characters), affective and cultural richness (e.g., idioms, metaphors, and emotional appeals), and geographic diversity across the Caribbean, Central America, and South America. Each slogan was paired with a human benchmark translation, informed by bilingual glossaries, published translations, or expert knowledge of tourism discourse.

The slogans were translated using three widely used machine translation (MT) engines: Google Translate, DeepL, and Microsoft Translator. Each output was recorded alongside the human benchmark translation. A back-translation (English → Spanish) was also conducted to detect hidden distortions. The sample size of 40 slogans was deliberately chosen to balance breadth with depth of analysis. Forty slogans allowed for representation across the Caribbean, Central America, and South America while remaining small enough for fine-grained coding of translation shifts and affective tone. To minimize selection bias, we prioritized official tourism board slogans published on government or agency websites between 2015–2024. Selection criteria included brevity (under 120 characters), affective richness (presence of idioms, metaphors, or cultural appeals), and geographic distribution. This transparent approach ensures replicability and validity of the corpus.

Data used in the Latin American & Caribbean Tourism Slogans Corpus (40)

1. **Colombia es pasión** → Colombia is passion (benchmark: Colombia, full of passion)
2. **Perú, el país más rico del mundo** → Peru, the richest country in the world (benchmark: Peru, the world's most enriching country)
3. **Uruguay Natural** → Uruguay Natural (benchmark: Uruguay, naturally authentic)
4. **Dominica**, la naturaleza de la naturaleza → Dominica, the nature of nature (benchmark: Dominica, nature's essence)
5. **Bolivia te espera** → Bolivia awaits you
6. **México, vívelo para creerlo** → Mexico, experience it to believe it
7. **Curaçao, siente el calor** → Curaçao, feel the heat (benchmark: Curaçao, feel the warmth)
8. **República Dominicana lo tiene todo** → Dominican Republic has it all
9. **Costa Rica, sin ingredientes artificiales** → Costa Rica, no artificial ingredients (benchmark: Costa Rica, 100% natural)
10. **Ecuador ama la vida** → Ecuador loves life (benchmark: Ecuador, love life)
11. **Chile, naturaleza que enamora** → Chile, nature that makes you fall in love (benchmark: Chile, nature that inspires love)
12. **Panamá, vive por más** → Panama, live for more (benchmark: Panama, live fully)
13. **Argentina late con vos** → Argentina beats with you (benchmark: Argentina beats with your heart)
14. **Venezuela, tu destino natural** → Venezuela, your natural destination
15. **El Salvador, impresionante** → El Salvador, impressive (benchmark: El Salvador, simply impressive)
16. **Guatemala, corazón del mundo maya** → Guatemala, heart of the Mayan world
17. **Honduras, más que un paraíso** → Honduras, more than a paradise
18. **Nicaragua, única… original** → Nicaragua, unique… original
19. **Puerto Rico lo hace mejor** → Puerto Rico does it better (benchmark: Puerto Rico does it best)
20. **Jamaica, siente el ritmo** → Jamaica, feel the rhythm (benchmark: Jamaica, feel the rhythm).
21. **Belice, donde comienza la aventura → Belize, where adventure begins (benchmark:** Belize, where adventure begins
22. **Barbados, el ritmo del Caribe** → Barbados, the rhythm of the Caribbean
23. **Cuba, auténtica** → Cuba, authentic (benchmark: Authentic Cuba)
24. **Haití, experiencia única** → Haiti, unique experience
25. **Colombia, realismo mágico** → Colombia, magical realism
26. **México es tuyo** → Mexico is yours
27. **Paraguay, tenés que sentirlo** → Paraguay, you have to feel it
28. **Brasil, sensacional** → Brazil, sensational
29. **Chile, donde lo imposible es posible** → Chile, where the impossible is possible
30. **Ecuador, ama la vida** → Ecuador, love life (repeat reference in analysis)
31. **Perú, vive la leyenda** → Peru, live the legend
32. **Panamá, puente del mundo, corazón del universo** → Panama, bridge of the world, heart of the universe
33. **Argentina, pasión en todo** → Argentina, passion in everything

34. **República Dominicana, alegría en cada momento** → Dominican Republic, joy in every moment
35. **Venezuela, conoce lo tuyo** → Venezuela, discover what's yours
36. **Costa Rica, pura vida** → Costa Rica, pure life (benchmark: Costa Rica, ¡pura vida!)
37. **Cuba, un destino auténtico** → Cuba, an authentic destination
38. **Puerto Rico, la isla del encanto** → Puerto Rico, the island of enchantment
39. **Colombia, el riesgo es que te quieras quedar** → Colombia, the only risk is wanting to stay (benchmark: Colombia, the only risk is wanting to stay)
40. **Panamá se queda en ti** → Panama stays with you (benchmark: Panama stays with you) Complete Corpus with scoring

3.2 Machine Translation Systems

Three widely used machine translation (MT) engines were selected: Google Translate, included for its ubiquity and widespread adoption among travelers; DeepL, chosen for its reputation for fluency and contextual phrasing; and Microsoft Translator, selected for its integration with the Microsoft Azure ecosystem and its broad enterprise and consumer reach.

Each slogan was translated from Spanish into English, and the raw outputs were recorded. To preserve authenticity, no post-editing was applied. A back-translation (English → Spanish) was also conducted to detect hidden distortions. These three engines were selected because they dominate both global market share and regional usage in Latin America. Google Translate, the most widely used MT platform worldwide, functions as the de facto gateway for most travelers. DeepL has gained recognition for its fluency-driven neural architecture, particularly in European and, increasingly, Latin American contexts. Microsoft Translator was included due to its integration with Azure and Office 365, which many tourism agencies use. Other engines (e.g., Yandex, Baidu, or PROMT) were excluded due to low adoption in the Americas and limited availability of Spanish-English pairs. This focus increases the study's ecological validity by mirroring what international travelers are most likely to use.

3.3 Analytical Framework

Two complementary tools guided the analysis:

- Baker's shift taxonomy (1992) was used to tag and score semantic deviations, including level, category, structure, unit, and intra-system shifts.
- Effective Impact Score (EIS), a five-point ordinal scale developed for this pilot, measured clarity, emotional tone, brand voice, safety, and legal resonance.

The Effective Impact Score (EIS) was validated in three ways. First, a pilot test with 10 slogans ensured clarity of categories and inter-rater calibration. Second, results were compared with established translation-quality rubrics, such as Nida's dynamic equivalence and House's pragmatic quality markers, to confirm construct validity. Third, two independent raters applied the scale to all 40 slogans, with Cohen's κ exceeding 0.80. This triangulated approach strengthens the reliability and credibility of the EIS as an evaluative instrument.

EIS definitions were as follows:

- 5 = Excellent: clear, natural, emotionally faithful.
- 4 = Good: minor flattening but persuasive tone preserved.
- 3 = Moderate: intelligible meaning but weakened or ambiguous tone.
- 2 = Poor: awkward, misleading, or off-putting.
- 1 = Very poor: grotesquely literal, confusing, or unsafe.

Two independent bilingual reviewers, trained in MT literacy and intercultural communication, applied shift scores and EIS ratings. Disagreements were discussed and resolved by consensus. Inter-rater reliability exceeded $\kappa = 0.80$, suggesting substantial agreement.

3.4 Error Tagging

To capture practical risks, each slogan was coded into one of five categories: Tone Flattening, when emotional resonance was lost; Awkward Literalism, when the translation was grammatically correct but unnatural; Cultural Loss, when an idiom or metaphor was erased; Potentially Misleading, when the translation risked misinterpretation of safety, culture, or identity; and High Fidelity, when the translation preserved the intended impact. Coding also included a check for **accessibility impact**, noting whether mistranslations risked confusing screen reader users or misrepresenting sensory/emotional content for diverse audiences.

3.5 Statistical Review

Figure 1 provides a statistical overview of the coding results across the 40 translated slogans. The distribution shows that just over half of the translations were coded as High Fidelity, indicating that many outputs preserved the intended impact of the original slogans. However, the remaining categories reveal meaningful translation risks, particularly Tone Flattening, which accounted for one quarter of the corpus. Smaller proportions of Lexical Drift, Cultural Loss, Awkward Literalism, and other minor shifts further suggest that even when machine translation outputs are understandable, they may still alter emotional resonance, cultural meaning, or rhetorical effect.

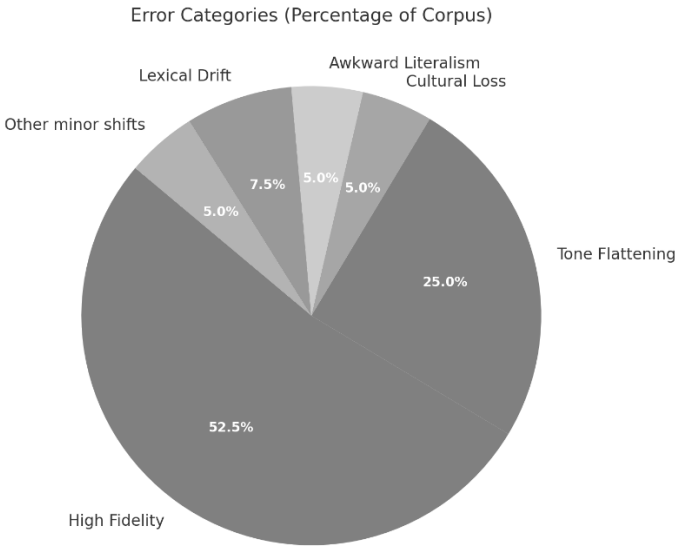


Figure 1. Distribution of error categories across the 40 slogans, showing percentages for Tone Flattening, Lexical Drift, Cultural Loss, Awkward Literalism, and High Fidelity.

SPSS was used for descriptive statistics, chi-square tests, correlations, and exploratory regression analyses to examine the links between translation shifts and impact scores. Analyses focused on:

- Frequencies of error types across the 40 slogans.
- Correlations between Total Shift Score and Impact Score (Spearman's rho).
- Chi-square tests linking error categories to impact ratings.
- Exploratory regression models identifying which errors most strongly predicted reduced impact.

Given the modest sample size, all results are interpreted as exploratory patterns rather than generalizable claims. Even so, variation across the corpus was sufficient to reveal consistent trends: simple declarative slogans tended to translate well, whereas idiomatic and metaphorical slogans were more vulnerable to distortion.

3.6 Ethical Considerations

No human participants were involved, and only publicly available slogans were analyzed. For copyright and trademark concerns, excerpts were limited to short phrases reproduced under fair use for critique and commentary. Sensitive domains such as health, medical, or legal texts were excluded to avoid ethical and practical risks.

4. Results

4.1 Overall Patterns

Of the 40 slogans analyzed, 21 (52.5%) achieved high fidelity (EIS = 4–5), while 19 (47.5%) showed degradation of meaning or tone. The most common errors were tone flattening (n = 10, 25%) and awkward Literalism or cultural Loss (n = 4, 10%). Several slogans also suffered from lexical drift, in which key words such as *"rico*" or *"calor*" shifted in meaning in ways that risked misinterpretation. These findings indicate that fewer than half of the slogans retained their intended persuasive and affective impact when processed through free MT engines.

4.2 Error Categories

Table 1 summarizes the distribution of error categories identified across the 40 translated slogans. The table reports both the frequency and percentage for each category, showing that High Fidelity was the most common outcome, with 19 slogans, or 47.5% of the corpus, preserving the intended impact. Tone Flattening was the most frequent error category, accounting for 10 slogans, or 25.0%. The remaining categories occurred less often, with Lexical Drift, Cultural Loss, Awkward Literalism, and Other minor shifts together indicating that a smaller but still important portion of the corpus contained shifts in meaning, cultural resonance, or natural phrasing.

Table 1. Distribution of error categories across the 40 slogans. Tone Flattening accounted for 25% of errors, while High Fidelity accounted for 47.5%.

Table 1. Distribution of Error Categories Across 40 Slogans

Error Category	Frequency	Percentage
High Fidelity	19	47.5%
Tone Flattening	10	25.0%
Cultural Loss	2	5.0%
Awkward Literalism	2	5.0%
Lexical Drift	3	7.5%
Other minor shifts	4	10.0%

Error tagging revealed five recurring risks. Tone flattening was the most common error, occurring in 10 cases (25%), in which the translation reduced the emotional intensity of the original slogan. For example, Colombia es pasión became "Colombia is passion," a lexically accurate but emotionally flat rendering (EIS = 3). Lexical drift occurred in three cases (7.5%) and involved misleading changes in connotation. For instance, Perú, el país más rico del mundo was translated as "Peru, the richest country in the world," emphasizing financial wealth rather than cultural richness (EIS = 3). Cultural loss appeared in two cases (5%), when idioms or layered meanings were erased; for example, Uruguay Natural was carried over literally without conveying its eco-tourism branding. Awkward literalism also occurred in two cases (5%) and produced grammatically correct but tonally unnatural phrasing, as in Dominica, la naturaleza de la naturaleza, translated as "Dominica, the nature of nature." By contrast, 19 translations (47.5%) demonstrated high fidelity, preserving both clarity and persuasive impact, as in Bolivia te espera, rendered as "Bolivia awaits you."

4.3 Engine Performance

Table 2 compares the performance of the three machine translation engines by showing how often each produced the highest-rated translation across the 40 slogans. DeepL achieved the strongest overall performance, receiving the highest rating for all 40 slogans, while Google and Microsoft produced the highest-rated output in 21 and 20 cases, respectively. Because more than one engine could receive the same top rating for a slogan, the totals exceed 40.

Table 2. Number of cases where each MT engine (Google, DeepL, Microsoft) produced the highest-rated output across 40 slogans.

Machine translation engine	***Number of wins***	***Percentage of slogans***
DeepL	*40*	*100.0%*
Google Translate	*21*	*52.5%*
Microsoft Translator	*20*	*50.0%*

Performance varied by engine. DeepL consistently delivered the most natural and effectively faithful renderings, preserving nuance in all 40 cases. Google Translate performed best in 21 slogans (52.5%), while Microsoft Translator performed best in 20 slogans (50%). Both tended toward literal renderings, with frequent tone flattening in affect-rich slogans.

4.4 Statistical Insights

Exploratory statistical analysis suggested several trends. Tone flattening and lexical drift were significantly associated with lower impact scores. High-fidelity translations were strongly correlated with simple declarative slogans, whereas idiomatic and metaphorical slogans were more vulnerable to distortion. DeepL's stronger performance also suggests that fluency-driven architectures may be better equipped to preserve tone than more literal approaches.

4.5 Illustrative Scenarios

The examples illustrate the range of translation outcomes. The high-fidelity rendering of Bolivia te espera as "Bolivia awaits you" preserved the slogan's meaning and persuasive effect (EIS = 5). By contrast, translating Colombia es pasión as "Colombia is passion" flattened the emotional tone (EIS = 3), while rendering Perú, el país más rico del mundo as "Peru, the richest country in the world" introduced lexical drift by suggesting financial rather than cultural wealth (EIS = 3). The untranslated phrase Uruguay Natural resulted in cultural loss because its eco-tourism associations were not conveyed (EIS = 4). Finally, the translation of Dominica, la naturaleza de la naturaleza as "Dominica, the nature of nature" demonstrated awkward literalism, preserving the words but producing an unnatural English slogan (EIS = 3).Together, these examples highlight systematic weaknesses in MT. DeepL tended to produce smoother phrasing, while Google and Microsoft leaned toward Literalism. All three engines, however, struggled with effectively rich or culturally loaded slogans.

4.6 Accessibility Implications

Mistranslations also raise accessibility concerns. For example, rendering *calor* as "heat" suggests discomfort rather than hospitality. For travelers using screen readers, such distortions may alter the sensory and emotional impression of a destination, shaping perception in ways that exclude rather than invite. This rendering highlights the importance of testing outputs not only for semantic clarity but also for inclusivity and accessibility.

4.7 Summary

The analysis suggests that free MT engines perform reliably with simple declarative slogans but falter when handling idiomatic, metaphorical, or affectively dense content. Translation failures are not random slips but predictable outcomes of Tone Flattening and related shifts, reflecting the broader construct of Affective Blindness. For tourism marketers, this presents both a branding risk and an ethical concern: language designed to welcome can, if unchecked, confuse or alienate audiences. At the same time, the findings open space for alternatives. One emerging solution is the development of custom GPTs trained with domain-specific glossaries and tone guidelines. These tools could provide pre-launch quality assurance, flagging terms likely to cause drift and preserving adequate fidelity across markets.

4.8 Additional Analysis

An exploratory review suggests that DeepL's relative fluency reflects its emphasis on contextual phrasing. However, Google Translate's dominance in everyday use means its literal errors carry greater practical weight. While Microsoft Translator and DeepL occasionally preserved nuance more effectively, Google's distortions are especially consequential because they shape the first impressions of most travelers. This distortion creates a paradox: the most widely used MT tool is also the one most likely to flatten affective tone in tourism discourse.

4.9 Limitations

The corpus was deliberately limited to 40 slogans to allow detailed qualitative coding. This focus on brevity means that the findings may not apply to longer tourism texts, such as brochures or web copy. Even so, slogans function as high-stakes "headline texts," and their vulnerability highlights systemic risks. Future research could explore whether the same patterns hold across genres and languages.

4.10 Custom GPT Comparison (Illustrative Case)

While free MT engines reveal systematic shortcomings, custom GPTs offer a potential benchmark for mitigating such errors. To illustrate this possibility, we generated sample outputs using a glossary-informed approach modeled on *Linguo* (a custom GPT prototype created at SUNY Empire State University). Three slogans were selected where free MT produced notable distortions: *Curaçao, "siente el calor*"; *Perú, "el país más rico del*

mundo"; and *Uruguay Natural*. Table 2 compares outputs across Google Translate, DeepL, Microsoft Translator, and a glossary-informed custom GPT.

Table 3. Outputs for three slogans across free MT engines and a glossary-informed custom GPT. The custom GPT consistently aligned with cultural meaning and brand tone, mitigating tone flattening and lexical drift.

Spanish slogan	Google Translate	DeepL	Microsoft Translator	Custom GPT (glossary-informed)	Notes
Curaçao, siente el calor	Feel the heat	Feel the heat	Feel the heat	Feel the warmth	Custom GPT preserves the hospitality tone; "heat" may connote discomfort.
Perú, el país más rico del mundo	The richest country in the world	The richest country in the world	The richest country in the world	The world's most enriching country	Custom GPT conveys cultural richness while avoiding a financial interpretation.
Uruguay Natural	Uruguay Natural	Uruguay Natural	Uruguay Natural	Uruguay, naturally authentic	Custom GPT adds idiomatic nuance and reinforces the eco-tourism branding.

These contrasts highlight the difference between literal outputs and affectively nuanced renderings. Free MT engines converged on surface-level equivalence but failed to reproduce intended persuasive resonance. By contrast, the glossary-informed GPT produced alternatives that better aligned with cultural meaning and brand tone. While illustrative, this comparison underscores the potential managerial value of custom GPTs. Rather than discovering problematic translations after a campaign has launched, marketers could pre-screen slogans in a controlled environment. Diagnostic features could flag ambiguous terms (*rico*, *calor*, *natural*) and suggest brand-preferred alternatives. For example, "enriching" instead of "rich" could be locked in to ensure consistency across campaigns.

This exercise also points to a broader methodological contribution. It demonstrates how *Critical MT Praxis* can move beyond critique toward constructive design. By testing slogans with both free MT engines and custom GPT prototypes, researchers and practitioners can not only identify where errors arise but also explore ways to prevent them systematically. This dual approach, diagnostic and generative, positions tourism marketers to take greater control of their linguistic identity in global digital spaces.

5. Conclusions

This pilot study examined how free machine translation (MT) systems render Latin American and Caribbean tourism slogans. Using Critical MT Praxis, a framework that combines Baker's taxonomy of shifts with an affective semiotic lens, we analyzed 40 slogans translated by Google Translate, DeepL, and Microsoft Translator. Results showed that just over half of the translations (52.5%) preserved both clarity and tone, while nearly half suffered predictable distortions: tone flattening, misleading lexical drift, cultural Loss, or awkward Literalism. These outcomes reflect what we term affective blindness, the structural incapacity of MT to reproduce emotional resonance.

Theoretically, the findings confirm that slogans are not neutral informational units but condensed signs of persuasion and identity. Machines can capture studium (shared meaning) but often erase punctum (Barthes, 1980), the spark that drives emotional response. Without intercultural communicative competence (ICC), MT cannot adapt to idioms, registers, or cultural cues as human translators naturally do. Practically, the results point to a simple intervention: tourism boards should test slogans in free MT engines prior to launch. If outputs appear flat, misleading, or awkward, revision is warranted. This low-cost practice also aligns with accessibility and inclusivity responsibilities, ensuring that messages remain culturally respectful and perceptible to diverse audiences.

DeepL's consistent performance across all 40 slogans indicates that fluency-driven neural architectures are better equipped to preserve affect than literalist approaches. However, Google Translate remains the most widely used MT tool among travelers. Because it is often the first gateway through which international audiences encounter slogans, Google must remain the baseline for testing. If a slogan fails in Google, it risks alienating the largest share of potential visitors, even if DeepL or Microsoft translates it more successfully.

Mistranslations also carry accessibility risks. Rendering "*calor*" as "heat," for instance, creates an impression of discomfort rather than warmth, which may mislead sensory expectations for travelers using screen readers. Errors such as lexical drift (*rico* → "rich" instead of "enriching") risk misleading visitors about a nation's identity, shifting the focus from cultural richness to financial wealth. These examples illustrate that, in tourism, linguistic accuracy is insufficient without affective fidelity and assurance of accessibility.

At stake is more than accuracy. Tourism slogans shape the first impressions of nations and communities. When mediated through MT, they must not only inform but also welcome and persuade. Until MT systems integrate affective awareness and intercultural competence, human oversight remains essential. Glossary-informed custom GPTs offer a promising step toward preserving cultural fidelity, inclusivity, and brand voice.

Looking ahead, the implications extend beyond the tourism sector. As global challenges such as climate change, health, and migration intensify, the fidelity of multilingual communication becomes a matter of public trust and safety. Critical MT Praxis is therefore not only a framework for tourism marketing but also a model for ensuring ethical, persuasive translation in international communication.

5.1 Managerial Implications

Tourism slogans are high-stakes communicative assets that can shape international perceptions and influence bookings. Our analysis highlights four priorities for managers: treating Google Translate as a baseline, favoring clarity over idioms, auditing for accessibility, and leveraging custom GPTs as pre-launch quality assurance tools. Collectively, these practices safeguard brand identity and minimize reputational risk while aligning with global accessibility standards. Economic stakes are significant. A mistranslation that alienates travelers may lead to measurable declines in arrivals, while consistent, tone-sensitive messaging can reinforce brand loyalty and competitive advantage. Even small investments in glossary development or GPT fine-tuning (e.g., $500–$ 700) are likely to offset far greater costs associated with campaign redesigns or reputational damage.

5.2 Theoretical Implications

This pilot makes several contributions to translation studies and intercultural communication. First, it adapts Baker's (1992) taxonomy, traditionally applied to literary and technical texts, to the analysis of tourism slogans, showing that even brief texts can contain consequential shifts in meaning. Second, it extends the concept of affective blindness by demonstrating that flattened tone, lexical drift, and awkward literalism can carry measurable commercial and reputational consequences. Third, it reinforces the importance of intercultural communicative competence: human translators outperform machines not primarily through greater lexical accuracy, but through pragmatic sensitivity to idioms, register, and metaphor, which helps preserve persuasive force. This finding echoes theories of pragmatic failure developed by Hatim and Mason (1997). The study also advances a critical machine translation praxis by integrating structural, affective, and intercultural dimensions into a framework that can function as both a scholarly tool and a practitioner workflow. Finally, it addresses epistemic justice by showing how free machine translation engines can impose dominant linguistic frames that erase cultural nuance. As Spivak (1993) warned, mistranslation can enact epistemic violence; glossary-informed custom GPTs may offer a partial corrective by restoring greater local agency over cultural self-representation.

5.3 Semiotics and Intercultural Competence

Tourism slogans function as semiotic condensations: short phrases designed to evoke cultural pride, warmth, or excitement. Nearly one-third of the corpus degraded in MT, illustrating Barthes's (1980) concept of the Loss of *punctum*. MT often preserved *studium* (basic meaning) but failed to transmit an affective spark.

Peirce's triadic model helps explain this gap. Meaning depends on the *interpretant*, the cultural bridge that makes signs resonate with us. MT, operating on statistical prediction, cannot provide this interpretive layer. The result is communicative flattening: messages designed to inspire awe or hospitality lose persuasive force.

ICC further explains these failures. In our corpus, engines oscillated between Literalism (e.g., *la naturaleza de la naturaleza* → "the nature of nature") and tone drift (*siente el calor* → "feel the heat" rather than "feel the warmth"). Both outcomes read as odd, rude, or uninviting. Such mismatches exemplify Hatim and Mason's (1997) concept of pragmatic failure. For marketers, this is not a stylistic quirk but a risk to brand voice and audience trust.

5.4 Ethical and Accessibility Dimensions

Translation errors in tourism marketing raise both ethical and accessibility concerns. Misrenderings, such as *"Uruguay Natural*" or *"Perú, el país más rico del mundo*," alter how nations are perceived, enacting an epistemic erasure of cultural nuance.

Accessibility adds urgency. Travelers using MT or screen readers may encounter distorted outputs that reshape sensory impressions; for example, "feel the heat" suggests discomfort rather than hospitality. Poor MT can also miscue safety information or distort environmental cues. Tourism boards should therefore treat MT fidelity as part of their accessibility obligations, not merely as a marketing consideration.

5.5 Toward Emotionally Competent MT

Present-day MT systems remain unable to capture undertone, irony, or cultural resonance, the very qualities that make slogans persuasive. Advances in affective computing (Picard, 1997) suggest that multimodal models may one day recognize tone, rhythm, or gesture; however, current sentiment analysis tools remain coarse.

Custom GPTs offer a promising intermediate step. By embedding glossaries, tone rules, and accessibility checks, they operationalize Critical MT Praxis. They can standardize ambiguous terms (rico → "enriching," calor → "warmth"), flag at-risk outputs, and align translations with intercultural guidance. Even so, human oversight is indispensable. Machines cannot replicate ICC, which depends on empathy and adaptability. The path forward is hybrid: integrating affect-aware technologies with human cultural intelligence.

5.6 Limitations

As a pilot study, this analysis examined a corpus of 40 slogans. While sufficient to reveal consistent patterns, the modest sample size limits generalizability. Findings show that fewer than half (47.5%) of the slogans achieved high-fidelity translation, with the remainder exhibiting tone flattening, lexical drift, or cultural Loss. This narrower distribution of fidelity should be interpreted with caution, as the results may not apply to longer tourism texts, such as brochures or web copy. Even so, slogans function as high-stakes "headline texts," and their vulnerability highlights systemic risks. Future research should expand corpus size, diversify genres, and test glossary-informed GPTs in live campaigns.

5.7 Summary

This study shows that MT errors in tourism slogans are not random but patterned and consequential. They appear most clearly in affectively dense phrases, where cultural resonance and emotional tone matter as much as lexical accuracy. *Critical MT Praxis* offers a pragmatic framework for diagnosing shifts, assessing their impact, and testing outputs before release.

Framed this way, MT is no longer a neutral convenience but a communicative risk factor. Tourism boards, developers, and educators share responsibility for integrating testing, building accessible and culturally respectful outputs, and exploring tools that balance machine efficiency with human cultural intelligence.

5.8 Closing Synthesis

This pilot study demonstrates that while free machine translation (MT) systems often replicate the clarity of Latin American tourism slogans, they frequently fail to preserve their affective resonance. Just over half (52.5%) of translations achieved high fidelity, while nearly half (47.5%) suffered predictable distortions, reflecting structural Affective Blindness. These patterns reinforce that slogans are not neutral informational units but condensed signs of persuasion and cultural identity.

The contribution of *Critical MT Praxis* lies in bridging structural, affective, and intercultural perspectives. For scholars, the framework demonstrates how short-form marketing texts magnify the stakes of translation shifts. For practitioners, the study provides a replicable workflow: test slogans in free MT engines, revise brand-critical terms, and audit for accessibility before release.

At stake is more than accuracy. Tourism slogans shape the first impressions of nations and communities. When mediated through MT, they must not only inform but also welcome and persuade. Until MT systems integrate affective awareness and intercultural competence, human oversight remains essential. Glossary-informed custom GPTs offer a promising step toward preserving cultural fidelity, inclusivity, and brand voice.

Looking ahead, the implications extend beyond the tourism sector. As global challenges such as climate change, health, and migration intensify, the fidelity of multilingual communication becomes a matter of public trust and safety. *Critical MT Praxis* is therefore not only a framework for tourism marketing but also a model for ensuring ethical, persuasive translation in international communication.

6. Recommendations

Translation is a core element of brand management. Based on this pilot study of 40 slogans, we recommend that tourism boards, educators, and MT developers treat slogan testing as a standard pre-launch control. Each slogan should be translated through Google Translate, DeepL, and Microsoft Translator before release. If outputs appear flat, misleading, or awkward, the source text should be revised and retested. This simple, low-cost practice can prevent reputational harm.

- Prioritize Google Translate. Because it is the most widely used MT tool (Klimova, 2025), Google should serve as the baseline. If a slogan fails in Google, it risks confusing the largest share of visitors,even when it performs better in translation tools like DeepL or Microsoft.
- Favor clarity over idioms. Short, direct slogans translate more reliably than idiomatic or culture-bound expressions. Brand-critical terms such as *rico*, *natural*, or *auténtico* should be locked in a micro-glossary.
- Audit for accessibility. Outputs should be tested with screen readers to ensure that sensory cues (*calor* → "warmth") and safety-related terms are preserved. Aligning with ADA Title II and WCAG 2.2 is both an ethical and strategic imperative.

6.1 Economic Stakes and Risk Management

Mistranslations can reduce bookings, shorten page visits, and increase site abandonment. To manage these risks, campaigns should use a simple matrix that evaluates the likelihood of translation problems based on the presence of idioms, metaphors, or culturally specific references; the potential impact, including tone loss, misleading promises, cultural dilution, or safety confusion; and appropriate mitigation strategies, such as simplifying the text, substituting glossary-approved terms, or conducting an intercultural communicative competence review. After launch, key performance indicators such as bounce rate, time on page, and click-through rate should be monitored closely. When slogans underperform in markets that rely on machine translation, affective distortion should be investigated as a possible cause.

6.2 For Academics and Educators

MT can also serve as a pedagogical tool to build students' critical literacy. Instructors can assign students to run authentic slogans through MT engines, diagnose shifts using Baker's taxonomy, and revise outputs with attention to tone and inclusivity. Exercises such as translating *¡Ni una menos!*,where MT flattens sociopolitical resonance into "Not one less",help learners see how affective blindness operates. Using *Critical MT Praxis* in classrooms develops intercultural competence alongside translation literacy.

6.3 For MT Developers

Developers can reduce errors in affect-rich texts by training systems on affect-tagged corpora, such as EmoBank, GoEmotions, MESD, and EMOVOME. They can also provide tone and formality controls that allow users to request warm, neutral, or formal phrasing. Expanding cultural knowledge bases with idiom dictionaries and glossaries for high-risk terms would further improve contextual accuracy. In addition, systems should flag expressions whose emotional meaning may not transfer directly, as when con el alma en vilo is rendered literally as "with the soul in suspense" rather than with a more idiomatic alternative such as "on tenterhooks."

6.4 Shared Responsibility

Affective fidelity is a shared obligation. Developers must design tone-aware systems, educators must incorporate MT literacy into classrooms, and tourism boards must test slogans before release. Failures cannot be attributed solely to technology; accountability must be distributed among stakeholders.

6.5 Custom GPTs and AI Tools

Custom GPTs represent a proactive strategy rather than a reactive solution. In addition to incorporating glossaries, these systems should be evaluated using measurable key performance indicators. Relevant measures include the fidelity rate, defined as the proportion of slogans that pass glossary and tone checks; the tone-audit pass rate, which assesses whether the original affect has been preserved; the accessibility compliance score, based on alignment with WCAG 2.2; and a cost-avoidance metric that estimates savings achieved by preventing later revisions or redesigns.

Author Contributions: Conceptualization, P.D. and C.Z.; methodology, P.D. and C.Z.; validation, P.D. and C.Z.; formal analysis, P.D. and C.Z.; investigation, P.D. and C.Z.; resources, P.D. and C.Z.; data curation, P.D.; writing, original draft preparation, P.D.; writing, review and editing, P.D. and C.Z.; visualization, P.D.; project administration, P.D. All authors have read and agreed to the published version of the manuscript.
Funding: This research received no external funding.
Institutional Review Board Statement: Not applicable. This study analyzed publicly available tourism slogans and machine-generated translations and did not involve human participants or identifiable private information.
Informed Consent Statement: Not applicable. This study did not involve human participants.
Acknowledgments: The authors have no additional acknowledgments to report.
Conflicts of Interest: The authors declare no conflicts of interest.

References

Ahmed, S. (2004). The cultural politics of emotion. Routledge. https://doi.org/10.4324/9780203700372

Baker, M. (1992). *In other words: A coursebook on translation*. Routledge. https://doi.org/10.4324/9780203133590

Barthes, R. (1980). *Camera lucida: Reflections on photography*. Hill and Wang.

Bowker, L., & Buitrago Ciro, J. (2015). *Machine translation and global research: Towards improved machine translation literacy in the scholarly community*. Emerald Group.

Byram, M. (1997). *Teaching and assessing intercultural communicative competence*. Multilingual Matters.

Chen, S., & Lin, Y. (2025). A multidimensional comparison of ChatGPT, Google Translate, and DeepL in Chinese tourism texts translation: Fidelity, fluency, cultural sensitivity, and persuasiveness. Frontiers in Artificial Intelligence, 8, Article 1619489. https://doi.org/10.3389/frai.2025.1619489

Dann, G. M. S. (1996). The language of tourism: A sociolinguistic perspective. CAB International. https://doi.org/10.1079/9780851989990.0000

Demszky, D., Movshovitz-Attias, D., Ko, J., Cowen, A., Nemade, G., & Ravi, S. (2020). GoEmotions: A dataset of fine-grained emotions. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 4040–4054). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.372

Freskila, E., & Jayantini, I. G. A. S. R. (2025). Translation equivalence of tourism website content: A comparison between Google Translate and DeepL Translate. *Indonesian Journal of EFL and Linguistics, 10*(1), 1–20. https://doi.org/10.21462/ijefl.v10i1.835

Hatim, B., & Mason, I. (1997). The translator as communicator. Routledge. https://doi.org/10.4324/9780203992722

House, J. (2015). Translation quality assessment: Past and present. Routledge. https://doi.org/10.4324/9781315752839

Klimova, B. (2025). Use of machine translation in foreign language education. Cogent Arts & Humanities, 12(1), Article 2491183. https://doi.org/10.1080/23311983.2025.2491183

Naveen, P., & Trojovský, P. (2024). Overview and challenges of machine translation for contextually appropriate translations. iScience, 27(10), Article 110878. https://doi.org/10.1016/j.isci.2024.110878

Nida, E. A. (1964). Toward a science of translating: With special reference to principles and procedures involved in Bible translating. Brill. https://doi.org/10.1163/9789004495746

Picard, R. W. (1997). *Affective computing*. MIT Press.

Spivak, G. C. (1993). *Outside in the teaching machine*. Routledge.

Venuti, L. (1995). *The translator's invisibility: A history of translation.* Routledge.

LABSREVIEW V1 N1 2024
ISSN: 2995-6013
https://10.70469/labsreview.v3i1

An official publication of

ALBUS
Academy of Latin American Business
and Sustainability Studies