Integrating VR and AI Chatbots in Language Education: A Classroom-Based Case Study

Author: Antônio R.M. Simões, Professor
Affiliation: University of Kansas, Lawrence
Contact: asimoes@ku.edu

Abstract

This case study explores the integration of Virtual Reality (VR) and AI chatbots (ChatGPT) in two advanced undergraduate Spanish courses at a U.S. university. Motivated by the rapid advancement of generative technologies, the research examines how educators can adapt to these innovations rather than be marginalized by them. Using an evolutionary lens, it investigates how VR and AI complement existing instructional practices and how students perceive their linguistic and cultural value.

A mixed-methods analysis of student reflections reveals that VR enhances spatialized cultural immersion, while AI offers accessible conversational practice. However, students also reported challenges with interface complexity and the need to critically evaluate AI responses. The study proposes a pedagogical framework outlining the complementary roles of VR and AI across different stages of engagement.

The core argument is that AI represents a structural shift in language practice, feedback, and learner agency, emphasizing task authenticity, process-oriented assessment, and the distinction between learning support and task outsourcing. The findings aim to inform ethical and practical decision-making in modern language education. While primarily addressing foreign language educators and students, the insights also benefit engineers and developers designing VR and AI tools for language learning.

Keywords: Artificial Intelligence (AI), Virtual Reality (VR), ChatGPT, language pedagogy, educational technology

1. Introduction

Recent developments in Virtual Reality (VR) platforms and large language model–based artificial intelligence have expanded the range of digital tools available for educational use. In foreign language teaching, these technologies are frequently associated with the possibility of simulated cultural exposure and increased opportunities for language practice beyond the classroom. At the same time, prior research has repeatedly shown that technological innovation alone does not determine instructional outcomes; instead, learning effects depend on how tools are embedded within pedagogical practices and curricular goals (Pegrum, 2025; Stockwell, 2024).

This study adopts an evolutionary perspective on technology integration in language education. Rather than treating VR and AI chatbots as instructional disruptions, it examines how these tools were incorporated into two advanced Spanish courses and how students experienced their use in practice. The focus is not on technological capability in the abstract, but on classroom-level implementation, learner perception, and instructional constraints.

The study addresses three guiding questions:

1. How do advanced learners describe the linguistic and cultural value of VR-based exploration and AI-mediated conversation?
2. What limitations and challenges do students identify when using these tools?
3. How can student feedback inform pedagogically grounded and ethically responsible classroom use?

To address these questions, the study analyzes student reflections written in Spanish following structured VR and ChatGPT activities. The analysis is informed by constructivist and sociocultural perspectives (Piaget, 1950; Vygotsky, 1978), which provide a framework for understanding experiential learning and mediated interaction. The goal is to contribute empirical evidence that can support instructional decision-making without presupposing the desirability of any particular technology.

Beyond language educators and learners, the findings in this study may inform developers and designers involved in educational VR and conversational AI tools.

1.1. Use of Artificial Intelligence in This Study

This study was conducted with the support of AI, used similarly to human graduate research assistants (GRAs), colleagues, or consultants. AI offered cost-effective, 24/7 assistance, but its use also raised concerns, such as reduced opportunities for GRAs and consultants, and the loss of collaborative academic work. Addressing these challenges requires awareness and inclusive solutions from all stakeholders.

The AI assistant contributed at multiple research stages:

Project Planning: Facilitated initial discussions and streamlined project setup.
Literature Review: Identified key bibliographical sources efficiently.
Idea Development: Brainstormed and refined central research ideas.
Data Analysis: Conducted statistical analyses and aided in result interpretation.
Text Summarization: Generated concise summaries of student narratives.
Language Refinement: Improved manuscript clarity, especially valuable given the author's multilingual communication.

However, limitations were evident. The AI occasionally misaligned with established scholarly perspectives—such as initially misrepresenting the innatist perspective on language acquisition and the chronological relationship between Piaget's and Vygotsky's work—but corrected these after in-depth discussion. Supervision and critical evaluation remained essential.

Beyond AI, the author engaged human scholars for classroom discussions, bibliographical updates, and research support, including planning, technology use, and critical evaluation of AI's role. These contributions are acknowledged at the study's conclusion.

While AI enhanced efficiency and provided valuable insights, its role must be carefully managed. The author applies this experience when guiding students in the responsible use of AI, emphasizing the need for human oversight to identify biases or inaccuracies in AI-generated content.

2. Literature Review and Theoretical Framework

The effective integration of any educational technology must be rooted in a clear understanding of its theoretical underpinnings and its place within existing scholarly discourse. This section establishes that foundation by examining key learning theories and reviewing contemporary research on VR and AI in language education.

2.1. Theoretical Frameworks: Constructivism and Sociocultural Theory

This study draws primarily on two complementary theoretical perspectives that illuminate different facets of the learning process facilitated by VR and AI.

Piagetian Constructivism: Jean Piaget's (1950) theory posits that learners actively construct knowledge through interaction with their environment, undergoing processes of assimilation and accommodation to achieve cognitive equilibrium. VR aligns powerfully with this paradigm. By placing learners inside a simulated Spanish plaza or Andean landscape, VR creates an environment for direct, experiential interaction. The learner is not a passive recipient of cultural facts but an active explorer who constructs understanding through navigation and observation. This active, environmental interaction is a core constructivist principle realized through immersive technology.

Vygotskian Sociocultural Theory: Lev Vygotsky's (1978) work emphasizes the social origins of cognition, highlighting the role of mediation and the Zone of Proximal Development (ZPD)—the space between what a learner can do independently and what they can achieve with guidance. AI chatbots function as a novel form of mediational tool. A chatbot can act as a more knowledgeable other, scaffolding conversational interaction, providing vocabulary,

and offering corrective feedback within a learner's ZPD. This facilitates the internalization of linguistic structures through socially patterned, albeit simulated, dialogue.

These theories do not compete but rather provide a dual lens: constructivism explains the cognitive engagement of immersive exploration, while sociocultural theory explains the developmental potential of interactive, language-focused dialogue. Together, they justify the combined use of VR (for environmental interaction) and AI (for dialogic mediation), providing a solid theoretical rationale for the integrated activities explored in this case study.

2.2. Current State of Research: Affordances, Contradictions, and Gaps

Recent scholarship provides a cautious yet optimistic map of the terrain, highlighting both potential and significant challenges, which this study directly addresses.

Research on VR in language education has documented its potential to support cultural observation and reduce language anxiety, while also noting challenges related to access, cognitive load, and pedagogical alignment (Godwin-Jones, 2023; Lin & Lan, 2023). Studies of AI chatbots in language learning report benefits for confidence and practice frequency, alongside concerns about superficial interaction, accuracy, and learner overreliance (Han, 2024; Warschauer et al., 2023).

VR in Language Learning: Research affirms VR's unique capacity for creating a sense of presence and embodied experience, which can enhance cultural learning and reduce language anxiety by providing a safe space for experimentation (Legault et al., 2021; Godwin-Jones, 2023). Lin and Lan (2023), in their systematic review, note positive impacts on vocabulary acquisition and speaking skills but also call for more longitudinal studies and stronger pedagogical grounding. A key challenge remains the frequent dissociation of immersive visual experiences from target-language narration or interaction, a gap students in this study explicitly noted and which the proposed framework seeks to bridge.

AI Chatbots in Language Learning: The advent of large language models (LLMs) like GPT-4 has dramatically expanded the potential of AI as a conversational partner. Fryer and Carpenter (2020) and Kohnke (2023) discuss chatbots' role in boosting learner confidence and providing always-available practice. However, critical voices are essential. Han (2024) warns against premature claims of efficacy, urging theory-driven research that distinguishes between performance and acquisition. Warschauer et al. (2023) identify core contradictions in using AI for L2 writing: while it can model text and build confidence, it risks eroding authorship and reducing essential cognitive struggle. Stockwell (2024) similarly cautions that uncritical adoption may undermine learner autonomy and interactional competence. This study's findings on the artificiality of chatbot conversation directly engage with these critical concerns.

Synthesizing the Critical Gap: A systematic review by Zawacki-Richter et al. (2019) noted a persistent weak connection between AIEd applications and theoretical pedagogical perspectives, alongside a lack of critical reflection on risks. This study directly addresses that gap by anchoring its analysis in constructivist and sociocultural theory and foregrounding student-voiced ethical and critical concerns. Furthermore, while literature exists on each technology in isolation, there is a paucity of practical, classroom-based research exploring their combined use based on learner feedback—precisely the integrative approach students in this case study spontaneously advocated for and which this manuscript seeks to formalize into a pedagogical framework.

This case study emerges from this evolutionary stance. It responds to a timely need for concrete, experience-based guidance as educators worldwide grapple with the practicalities and pitfalls of AI and VR integration (Hockly, 2023; Godwin-Jones, 2023). The central dilemma transcends mere tool usage, touching on deeper questions of pedagogical integrity, ethical responsibility, and the preservation of humanistic educational values in an age of automation. This research is driven by core questions: How can VR and AI be structured to genuinely enhance, rather than undermine, the goals of advanced language acquisition? What are the realistic affordances and constraints experienced by learners? And how can educators develop the necessary digital literacy to guide students in the critical and ethical use of these powerful technologies?

To move beyond theoretical speculation, this study is anchored in empirical data drawn from the author's own advanced Spanish language classrooms. By analyzing detailed student reflections on their experiences with Meta

Pro VR headsets and the voice-mode function of ChatGPT, this research provides a ground-level view of technological integration.

The ultimate aim is twofold. First, to provide a nuanced, evidence-based account of student perceptions, identifying key themes of benefit, challenge, and unexpected insight. Second, and most crucially, to synthesize these findings into a practical, multi-component framework for educators. This framework encompasses pedagogical design, activity sequencing, ethical guidelines, and a curated compilation of institutional policy resources necessary for responsible implementation. In doing so, this manuscript aims to provide a practical resource for language teachers, curriculum designers, and educational administrators navigating the transition into an AI- and VR-augmented pedagogical landscape.

A recurring gap in the literature concerns classroom-based accounts that examine how students experience the combined use of VR and AI tools and how these experiences inform instructional design. The present study addresses this gap by foregrounding learner reflections from a specific instructional context.

3. Methodology

This study employed a pragmatic, mixed-methods case study design to capture the richness of student experience while allowing for systematic analysis of recurring themes. The approach was naturalistic, integrating technological exploration directly into the curriculum of two existing advanced Spanish courses, with a strong emphasis on ethical practice and student safety.

3.1 Context and Participants

The study was conducted in two advanced undergraduate Spanish courses (XXX-XXX and XXX-XXX) during the Fall 2024 semester, in a large university at the United States. Thirteen students participated. Some students were enrolled in both courses; these cases were treated as repeated measures during analysis to avoid overrepresentation. Students' proficiency ranged from Advanced Low to Advanced High on the ACTFL scale (approximately CEFR B1–B2). Table 1 below presents approximate equivalencies in the Common European Framework of Reference for Languages (CEFR) for these proficiency levels.

ACTFL OPI Level	Approximate CEFR Level
Advanced Low (AL)	B1
Advanced Mid (AM)	B2
Advanced High (AH)	B2
Superior (S)	C1

Table 1. Approximate Alignment of ACTFL and CEFR Proficiency Levels for Study Participants. *Source:* https://en.wikipedia.org/wiki/Common_European_Framework_of_Reference_for_Languages

This advanced proficiency was essential, as the activities required the linguistic ability to engage in complex cultural reflection and sustained conversation.

All participants provided informed consent for the use of anonymized coursework in research.

3.2 Instructional Tools

The technological devices were selected for accessibility and pedagogical alignment.

Virtual Reality: Meta Quest Pro headsets were the primary VR equipment. The study utilized a library of applications and professionally produced 360-degree videos allowing virtual visits to culturally and historically significant sites across the Hispanic world, including Machu Picchu (Peru), Tenochtitlán/Mexico City (Mexico), La Sagrada Familia and La Rambla (Barcelona, Spain), La Alhambra (Granada, Spain), Asunción (Paraguay), and various locations in the Yucatán Peninsula.

AI Chatbot: The voice interaction mode of ChatGPT (GPT-4, paid version) was employed. Students used their own mobile devices or classroom computers to engage in spoken conversations with the AI, maximizing accessibility outside of class time.

Alternative Materials: In strict adherence to ethical and accessibility guidelines, students were provided with the option to substitute VR headset use with interactive 360-degree videos available on platforms like YouTube VR, Google Arts & Culture, and National Geographic VR. This ensured full inclusion and accommodated health-related concerns.

Guiding Protocols: Students received structured, Spanish-language prompt guides for ChatGPT conversations (see Appendix B). These guides provided suggested topics spanning history (e.g., Latin American independence movements), culture (e.g., Semana Santa, Día de los Muertos), gastronomy, literature, art, and explicit prompts for exploring regional linguistic variations.

3.3 Procedures and Data Collection

Ethical Approval: This study was approved by the XXXX XXXX Program (XXX-XXX-XXXX, XXX@XXX.XXX)] at the University XXXX Name of University/Institution] (Protocol/Approval No. [Insert Number]). Informed consent was obtained from individual participants included in the study.

Students completed structured VR explorations and AI-mediated conversations aligned with course content. Following each activity, students submitted reflective reports in Spanish describing their experiences, perceived learning outcomes, and challenges.

The integration was structured across several weeks to allow for depth and reflection, following a clear sequence:

1. *Orientation, Ethics, and Safety Briefing*: Prior to any technological use, students were introduced to the tools within a framework of critical digital literacy. This included discussions on:

Ethical Use: The distinction between AI as a learning aid and as an unauthorized substitute for one's own work.
Data Privacy: Warnings about not inputting personal, sensitive, or proprietary information into public AI tools.
Health and Safety: A formal advisory (included in the syllabus) outlined potential risks of VR (motion sickness, photosensitive seizures, etc.) and mandated consultation with a medical professional for students with specific health conditions. Participation in VR was explicitly voluntary, with the 360-degree video alternative always available.

2. Structured Learning Activities:

Virtual Reality Cultural Exploration: In supervised class sessions, students used the Meta Quest Pro headsets (or alternatives) to explore assigned locations. They were guided by a worksheet prompting observations on the relationship between architecture, urban planning, landscape, and cultural or social values (e.g., "How does the construction of Machu Picchu reflect the Inca's relationship with the natural environment?").
ChatGPT Conversational Practice: Students conducted conversations with ChatGPT both during dedicated class time and as a flexible homework assignment. They were required to use the Spanish voice mode and engage with topics from the prompt guide, moving from simple inquiries to more sustained dialogues. The goal was to practice negotiation of meaning, pronunciation, and explore cultural content interactively.

3. Data Generation via Critical Reflection:

The primary qualitative data source was two required reflective reports per student, submitted in Spanish. The first report focused on the VR experience, the second on the ChatGPT interaction. Prompts asked students to describe their experience, analyze what they learned linguistically and culturally, evaluate the effectiveness of the tool, and discuss any challenges or ethical considerations that arose. These reports provided rich, first-person narratives for thematic analysis.

3.4 Ethical Safeguards and Student Guidelines

Ethical safeguards addressed data privacy, physical safety, and equitable access. VR participation was voluntary, and students could complete equivalent activities using interactive 360-degree videos. Health advisories outlining potential risks (e.g., motion sickness, photosensitivity) were included in the syllabus. Students also received Spanish-language guidance for ChatGPT interactions, emphasizing responsible and task-focused use.

3.5. Data Analysis

A hybrid inductive-deductive thematic analysis was conducted on the corpus of reflective reports to ensure findings were both data-driven and conceptually informed.

Familiarization and Initial Coding: All reports were read multiple times to achieve deep familiarity. Initial descriptive codes were generated (e.g., "feeling of presence," "frustration with turn-taking," "technical glitch," "valuable vocabulary in context").
Theme Development: Initial codes were collated and clustered into broader, analytical candidate themes. For example, codes about "seeing daily life," "understanding scale," and "feeling transported" were grouped under a candidate theme tentatively labeled "Immersive Cultural Experience."
Review and Refinement: The candidate themes were reviewed against the entire data set to ensure they accurately represented the narratives. Themes were refined, split, or merged. This iterative process resulted in the final eight major themes presented in the results.
Quantification for Illustrative Support: To illustrate the prevalence and salience of each theme and to provide a mixed-methods dimension, each student report was systematically coded for the presence (1) or absence (0) of each final theme. Simple descriptive statistics (frequencies and percentages) were then calculated. A critical methodological step was treating each of the 13 students as a single case. For students enrolled in both courses, their responses across both reports were aggregated to form a holistic participant profile before coding, ensuring their feedback was not over-represented in the quantitative illustration.

4. Results: Student Experiences and Emergent Themes

Analysis of the student reflections revealed a nuanced portrait of engagement, identifying eight interconnected themes that illuminate the distinct and complementary affordances of VR and AI chatbots.

4.1. Thematic Analysis Findings

Theme 1: VR as a Tool for Spatialized Cultural Immersion. Students consistently described VR not as watching a video, but as "being there." This sense of presence was emotionally resonant ("impressive," "memorable," "felt real") and cognitively transformative. One student noted that "flying over" Machu Picchu fundamentally altered their understanding of its scale and geographic isolation, moving it from a flat image to a visceral, mountainous settlement. Another described how virtually standing in the plaza of Asunción made the societal coexistence of Spanish and Guaraní feel tangible and immediate. VR enabled culture to be understood as lived, navigable space rather than an abstract textual description, directly fulfilling a constructivist pedagogical goal.

Theme 2: Contextualized Vocabulary and Cultural Learning. Both tools were praised for anchoring language in meaningful, multi-sensory context, enhancing retention. In VR, vocabulary for architecture (e.g., arco, bóveda, patio), geography (montaña, valle), and cultural artifacts (mercado, iglesia) was attached to visual and spatial referents. With ChatGPT, students explored semantic fields within explanatory dialogues about complex traditions like Semana Santa or culinary practices. This contextual embedding was repeatedly reported as more effective for memory and deeper conceptual understanding than studying decontextualized vocabulary lists.

Theme 3: The Dual Nature of AI as a Conversational Partner. Feedback on ChatGPT was markedly bifurcated, revealing a gap between its promise and current reality.

Affordances: It was celebrated for its accessibility, infinite patience, and role as a confidence-builder. Students highlighted its value as a always-available practice partner, particularly for those without access to a Spanish-speaking community outside class. One student noted using it "while cooking," integrating practice into daily life.

Limitations: It was heavily critiqued for artificial and stilted conversational flow. Students found its speech too slow, overly clear, and phonetically "neutral," even when requesting specific accents. It managed turn-taking poorly, often interrupting pauses or delivering monologic responses. Crucially, it struggled to sustain a coherent, deepening dialogue that required building on previous exchanges. It was deemed useful for form-focused practice and rehearsing scripts but inadequate for developing the real-world interactional competence, pragmatic nuance, or spontaneous repair strategies needed for advanced proficiency.

Theme 4: Metalinguistic Awareness vs. Authentic Comprehension. Students demonstrated sophisticated linguistic awareness by valuing ChatGPT's ability to simulate and explain dialectal features (e.g., Argentine voseo, sheísmo, and syllable-final /s/ aspiration). However, they distinguished this metalinguistic awareness—knowing about a linguistic feature—from the ability to comprehend fast, connected, and variable native speech in authentic contexts. The tool was acknowledged as a useful philological informant but a poor substitute for immersive listening practice with human speakers, highlighting a key limitation for auditory skill development.

Theme 5: The Mediating Role of Usability and Access. Practical, logistical factors significantly shaped—and sometimes dictated—the learning experience.

VR: High cognitive load from navigating unfamiliar interfaces, limited headset availability leading to short sessions, and incidents of motion sickness frequently detracted from the primary cultural and linguistic objectives. One student succinctly captured this: "I think 90% of the time I spent using the VR was trying to understand the user interface."
AI Chatbot: Its principal strength was low-barrier access and autonomy. Students could use it anywhere, turning dead time into practice time. This democratized access to conversational interaction, a significant equity consideration noted in the data.

Theme 6: Emergent Critical AI and Digital Literacy. A significant and encouraging finding was students' inherent skepticism and critical perspective. They independently questioned the accuracy and potential bias of information generated by ChatGPT, noted the superficiality or generic nature of some responses, and raised ethical concerns regarding data sourcing and environmental impact. This demonstrated an emerging critical digital literacy, positioning them as evaluative, responsible users rather than passive consumers of technological output—a core goal of modern education.

Theme 7: The Expressed Desire for Technological Integration. An unprompted, recurring suggestion was to combine both tools into a seamless experience. Students spontaneously envisioned a VR tour of the Alhambra with a ChatGPT guide explaining its history in Spanish, or using a chatbot to debrief and ask detailed questions after a VR experience. This student-driven insight is perhaps the most valuable for future design, pointing directly toward a more holistic, multi-modal, and intellectually continuous learning model.

Theme 8: Affective and Motivational Impact. Emotional responses were a powerful mediator of engagement. VR consistently evoked wonder, curiosity, and a sense of adventure, directly driving intrinsic motivation to learn more about the places visited. ChatGPT's non-judgmental nature was reported to reduce speaking anxiety for some learners, allowing them to experiment without fear. This affective dimension—the tools' ability to generate interest and lower affective filters—is a crucial, often overlooked, component of their pedagogical value.

4.2. Quantitative Illustration of Theme Prevalence

To complement the qualitative depth, the coding of all unique participant reports yielded the following frequencies, illustrating how commonly each theme appeared in student narratives:

- Cultural Immersion (VR): 92% (12 of 13 participants)
- Contextualized Learning: 85% (11 of 13)
- Conversational Practice (ChatGPT): 100% (13 of 13)
- Usability/Access Issues: 77% (10 of 13)
- Metalinguistic Awareness: 69% (9 of 13)
- Critical Literacy: 62% (8 of 13)
- Desire for Integration: 54% (7 of 13)

• Affective Impact: 85% (11 of 13)

4.3. Comparative Analysis: VR vs. AI Chatbot

A comparative summary (Table 1) clarifies their distinct and complementary roles, providing a foundational rationale for the integrated framework proposed in the discussion.

Table 1: Comparative Affordances of VR and AI Chatbots in Language Learning

Dimension	Virtual Reality (VR)	AI Chatbot (e.g., ChatGPT)
Primary Strength	Cultural immersion & spatial understanding	Conversational practice & accessibility
Linguistic Focus	Receptive skills, contextual vocabulary	Productive & interactive skills
Cultural Learning	Direct, experiential, embodied	Indirect, explanatory, discursive
Accessibility	Low (requires special equipment, setup)	High (mobile, on-demand)
Cognitive Load	High (navigation, sensory input)	Low to Moderate
Optimal Use Case	Simulated field trips, cultural exploration	Homework tutor, speaking drill partner, writing feedback
Dimension	Virtual Reality (VR)	AI Chatbot (e.g., ChatGPT)

5. Discussion: Toward an Integrative Pedagogical Framework

Before discussing the results, it is important to note that students received explicit instruction throughout the semester regarding the intrusive and potentially invasive capacities of emerging AI technologies. The instructor emphasized that, while these systems can be valuable assets across multiple domains of knowledge, they should be approached with caution and with an awareness of their limitations and uncertainties. Students were reminded that even developers, engineers, or executives in charge of these technologies often cannot fully predict how the systems behave as they are shaped by user inquiries and usage patterns.

As part of this preparation, the class addressed basic digital safety and privacy practices, including turning devices off when not in use, covering cameras, and reflecting critically on data exposure during AI-mediated activities. During sessions involving VR and conversational AI, students were encouraged to develop their own “defensive strategies” or personal safeguards, and to share them with peers. These exchanges informed the classroom culture and surfaced repeatedly in students’ written comments, particularly as they gained experience with the tools.

Taken together, these experiences illustrate a shift from initial feelings of novelty, wonder, and enthusiasm to more critical, less naïve, and more reflective attitudes. Over time, students’ language reveals an evolution in both technological literacy and civic responsibility. Importantly, a number of students ultimately opted not to grant consent for the use of their comments in research, citing concerns over the control and governance of AI systems. This hesitation aligns with a growing pattern of skepticism observable in the United States, although such reluctance does not appear to be as widespread in other national contexts.

The results validate neither techno-utopianism nor blanket skepticism. Instead, they chart a middle path for purposeful, critical integration. This section translates the empirical findings into a comprehensive, four-part framework designed to empower educators to implement these tools effectively and ethically.

5.1. Framework Part 1: Defining Complementary Pedagogical Roles

The data supports using VR and AI for different, synergistic purposes within a lesson or module.

• *VR as a Cultural Simulator and Context-Builder*: Its optimal use is for creating situated, experiential contexts that are otherwise inaccessible. Pedagogical goals should center on observation, description, comparison, and cultural analysis—tasks that leverage its immersive strength. It answers the question, "What does this place look/feel like, and what does that suggest about the culture?"

• *AI as a Dialogic Mediator and Skills Coach*: Its optimal use is for scaffolded, repetitive, and reflective practice. It excels at providing grammar explanations, modeling dialogues, offering low-stakes conversation on defined topics, and giving immediate feedback on written form. It should not be tasked with replicating human spontaneity but rather with reinforcing and practicing linguistic forms within a conversational framework. It answers the question, "How can I practice talking or writing about this?"

The student-driven call for integration suggests a natural pedagogical flow: use VR to generate a rich, shared experiential context (e.g., visiting Machu Picchu), and then use the AI chatbot to help students process, discuss, and linguistically mine that experience through conversation and writing.

5.2. Framework Part 2: A Developmental Model of Student-AI Interaction

Observing the progression in student feedback—from basic use to sophisticated critique—we propose a staged model of AI engagement that aligns with increasing learner autonomy, linguistic proficiency, and criticality. This model provides a roadmap for scaffolding activities across a course or curriculum.

• *Stage 1 - Tool as Tutor (Novice/Intermediate)*: The AI is used as a responsive reference and drill master. Activities include vocabulary lookup, simple grammar Q&A, and rehearsed functional dialogues (e.g., "ChatGPT, act as a market vendor. I will ask for prices."). The focus is on accuracy and basic fluency.

• *Stage 2 - Tool as Collaborator (Intermediate/Advanced)*: The AI is engaged as a conversational partner and creative assistant. Activities involve co-creating a story, debating a current event, summarizing a text, or getting structural feedback on a draft essay. The focus shifts to complexity, discourse management, and idea generation.

• *Stage 3 - Tool as Object of Critique (Advanced/Superior)*: Students critically deconstruct the AI's output and role. Activities include evaluating responses for cultural bias or stereotyping, analyzing the accuracy of historical information provided, comparing its simulated accent to authentic speech samples, and debating the ethical implications of AI in academia. The focus is on critical digital literacy and meta-linguistic/cultural analysis.

5.3. Framework Part 3: Practical Guidelines for Implementation & A Sample Lesson Plan

Drawing from the study's procedural successes and student-identified challenges, these guidelines are offered for educators.

• For VR Implementation:

Always Debrief in the Target Language: The core learning is solidified not in the headset, but in the subsequent, structured reflection. Conduct post-VR discussions, presentations, or written analyses in Spanish.
Curate for Linguistic Objectives: Actively seek out experiences with target-language narration or menus. If unavailable, create a companion vocabulary list or observation worksheet in Spanish to focus the exploration.
Proactively Manage Logistics: Implement a clear safety protocol, train "spotter" students, establish a hygiene plan for shared headsets, and always have a meaningful, equitable alternative activity (like curated 360-degree videos) ready.

• For AI Chatbot Integration:

Explicitly Teach Prompt Engineering: Move students beyond simple queries. Model the "Role-Goal-Format" method (e.g., "Role: Act as a skeptical political science student from Chile. Goal:Debate with me about the economic policies of the 1970s. Format: Use the voseo form, keep arguments concise, and ask me a counter-question after each point.").
Mandate Citation and Critical Verification: Establish a transparent course policy that any factual information drawn from an AI must be cross-referenced with a credible academic source (e.g., a peer-reviewed article, a reputable museum website). Teach students to use AI as a starting point, not a definitive source.
Design "AI-Resistant" Assessments: Redesign evaluations to prioritize process, personal reflection, synthesis, and in-class demonstration of skills. Examples include reflective journals (like those used in this study), oral exams, in-class writing based on a unique prompt, or multimedia projects that document the learning journey.

To concretize this integration, a full sample lesson plan is provided in Appendix C.

5.4. Framework Part 4: Ethical Imperatives and Policy Foundations

The ethical integration of these technologies is non-negotiable. Below is a synthesis of key principles distilled from leading institutional policies in the US and abroad, which should form the cornerstone of any course-level implementation plan. These principles directly address the ethical awareness (Theme 6) demonstrated by students. A full, annotated resource list with active links is provided in Appendix A.

Core Principles for Ethical Integration:

1. *Transparency & Disclosure*: Instructors must model and require students to openly disclose when and how AI has been used in the learning process, from brainstorming to editing (Stanford, Cornell).
2. *Pedagogical Purpose Primacy*: Technology use must be justified by a clear, stated learning objective that cannot be met as effectively through other means. It should augment, not automate, learning (Jisc, QAA).
3. *Human Oversight & Critical Evaluation*: AI output is a draft or a resource, not a final product. The responsibility for accuracy, cultural appropriateness, bias-checking, and final judgment rests unequivocally with the human user (EU Commission, Council of Europe).
4. *Equity of Access*: Course design must proactively ensure that students without personal access to premium AI tools, high-speed internet, or VR hardware are provided with equitable alternatives and support, preventing a digital divide in the classroom.
5. *Data Privacy & Security*: Students must be explicitly warned against inputting personal identifiers, confidential class materials, proprietary research data, or sensitive information into public AI platforms (Harvard IT, EU Guidelines).
6. *Academic Integrity Redefined*: Syllabus policies must move beyond simple "banning" and instead clearly define the boundaries between permitted assistance (e.g., using AI to generate practice quiz questions or to revise the grammar of a final draft) and prohibited outsourcing (e.g., submitting AI-generated content as one's own original creation). This requires nuanced, task-specific guidance (Stanford, UCL).

6. Implications for Engineers and Developers: Toward Innovative VR and Chatbot Integration

While the primary objective of this study is to support instructors and learners in the foreign language classroom, the findings also highlight how these technologies can be refined for a broader audience. The observed experiences suggest significant opportunities for engineers and developers to introduce innovative features tailored to pedagogical needs.

Perhaps most notably, there is a clear demand for the development of "virtual native speakers"—AI-driven avatars capable of simulating regional accents and dialects based on the virtual location the student is visiting. These agents could engage students in real-time dialogue and offer personalized linguistic feedback. Surprisingly, this potential remains largely untapped. Integrating such features would represent a transformative shift in the field of Second Language Acquisition (SLA). Furthermore, developers must address the technical and psychological barriers identified by participants in this study, including interface constraints, usability issues, student tolerance thresholds, and critical concerns regarding privacy and trust.

7. Conclusion: From Integration to Transformation

This case study demonstrates that VR and generative AI are not self-executing magic bullets, but they are transformative tools when placed within a pedagogical, critical, and ethical framework. The student voices at its center reveal a complex reality: these technologies can foster deep cultural connection, democratize language practice, and stimulate critical thinking, yet they simultaneously introduce new challenges in usability, authentic communication, and intellectual integrity.

The primary contribution of this work is the synthesis of these lived experiences into a practical, actionable framework for educators. This framework provides a roadmap, moving from defining the distinct yet complementary roles of VR and AI, through a developmental model for staging student engagement, to concrete classroom guidelines, a sample lesson plan, and an ethical foundation built upon the collective wisdom of leading

international institutions. It promotes an evolutionary approach where technology is deliberately adapted to serve sound, humanistic pedagogy, not the reverse.

Implications for stakeholders are clear and actionable:

• *For Educators*: The call is to develop your own AI and digital literacy. Begin with small, tightly focused integrated activities. Prioritize teaching critical evaluation alongside tool use, and leverage the growing body of policy resources (Appendix A) for institutional support and syllabus development.
• *For Institutions & Administrators*: Support must move beyond procurement to sustained faculty development in digital pedagogy. Establishing clear, pedagogically informed institutional guidelines that empower instructors with choice and nuance is more valuable than top-down restrictive mandates.
• *For Researchers*: Future work should pursue longitudinal studies to measure the impact of integrated models on skill retention and intercultural competence. There is a need to investigate differential impacts across learner profiles and to develop robust assessment methodologies for the unique competencies fostered by immersive, interactive technologies.

The path forward is not defined by being for or against AI and VR, but by being thoughtful, skilled, and critical in determining their place in our educational missions. By grounding integration in learning theory, authentic student experience, and ethical clarity, we can harness these powerful tools to create language learning experiences that are more immersive, personalized, equitable, and critically engaged than ever before—truly evolving our practice for a complex digital age.

8. Acknowledgments

I extend my sincere gratitude to the students in [Course Number]: [Course Name] and [Course Number]: [Course Name] for the invaluable insights and perspectives they shared via Canvas; their critical and original contributions significantly enriched this research. I am especially grateful to [Name], whose introduction to Meta headsets and the pedagogical potential of chatbots served as the catalyst for this project.

My thanks also go to [Name] for his expertise in developing the webpage that facilitated student participation in the Qualtrics survey, and to [Name] for his exceptional technical support in implementing and managing the Meta Pro headsets throughout the study.

Finally, I acknowledge the use of artificial intelligence in supporting the writing and organization of this manuscript, as detailed in the 1.1. section of this article.

Funding: This research received no external funding.

Conflict of Interest: The author declares no conflicts of interest.

9. References

1. Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the age of ChatGPT. Innovations in Education and Teaching International, 61(2), 228–240.
2. Dörnyei, Z., & Ryan, S. (2015). The psychology of the language learner revisited. Routledge.
3. Fryer, L. K., & Carpenter, R. (2020). Chatbots in the classroom: Are we ready? Journal of Educational Technology Systems, 48(3), 319–343.
4. Godwin-Jones, R. (2023). Virtual reality in language learning: The state of the art. Language Learning & Technology, 27(1), 1–12.
5. Han, Z.-H. (2024). ChatGPT in and for second language acquisition: A call for systematic research. Studies in Second Language Acquisition, 46(1), 1–10.
6. Hockly, N. (2023). Technology enhanced language teaching. Oxford University Press.
7. Kohnke, L. (2023). AI chatbots for language practice: Opportunities and challenges. TESOL Quarterly, 57(2), 456–472.
8. Legault, J., et al. (2021). The impact of virtual reality on language anxiety. Computer Assisted Language Learning, 34(5–6), 678–701.

9. Lin, C. H., & Lan, Y. J. (2023). Immersive language learning with VR: A systematic review. Educational Technology & Society, 26(1), 1–14.
10. Pegrum, M. (2025). From revolution to evolution: What generative AI really means for language learning. Language Teaching (FirstView), 1–17.
11. Piaget, J. (1950). The psychology of intelligence. Routledge & Kegan Paul.
12. Stockwell, G. (2024). ChatGPT in language teaching and learning: Exploring the road we're travelling. Computer Assisted Language Learning. Advance online publication.
13. Thorne, S. L. (2024). Generative artificial intelligence, co-evolution, and language education. The Modern Language Journal, 108(2), 567–572.
14. Vygotsky, L. S. (1978). Mind in society: Development of higher psychological processes (M. Cole, V. Jolm-Steiner, S. Scribner, & E. Souberman, Eds.). Harvard University Press.
15. Warschauer, M., Chen, Y., & Zheng, B. (2023). The affordances and contradictions of AI-generated text for second language writing. Journal of Second Language Writing, 62, 101013.
16. Zawacki-Richter, O., Bond, M., Marin, V. I., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education. International Journal of Educational Technology in Higher Education, 16, 39.

Appendices

Appendix A: Annotated Resource Guide for AI Policy & Ethics in Education

This appendix provides a curated, annotated list of key policy documents and guidance from leading institutions in the United States, United Kingdom, and European Union, with active links where available. It is designed to help educators and administrators develop robust, principled approaches to AI integration.

A.1 United States

• Stanford University, Office of Community Standards: "Generative AI Policy Guidance." A clear, honor-code-based framework. https://communitystandards.stanford.edu/generative-ai-policy-guidance
• Stanford Teaching Commons: "Teaching in the AI Era" & "Course Policies on Generative AI Use." Practical teaching guide and adaptable syllabus language. https://tlhub.stanford.edu/docs/teaching-in-the-ai-era/
• Cornell University, Center for Teaching Innovation: "AI & Academic Integrity." Focus on assessment redesign and student verification. https://teaching.cornell.edu/generative-artificial-intelligence/ai-academic-integrity
• Harvard University IT: "Generative AI Guidelines." Essential for data privacy and security protocols. https://www.huit.harvard.edu/ai/guidelines
• Johns Hopkins University, Teaching @ JHU: "Generative AI Tool Guidance." Balanced guidance for classroom use. https://www.huit.harvard.edu/ai/guidelines
• American Council on the Teaching of Foreign Languages (ACTFL): "AI Resources." Language-teaching-specific hub. https://www.actfl.org/educator-resources/ai-resources
https://www.actfl.org/advocacy/actfl-position-statements/artificial-intelligence
• Johns Hopkins University Press: "Generative AI Policy for Authors." Model for disclosure in scholarly work. https://www.press.jhu.edu/sites/default/files/media/2023/07/Generative%20AI%20for%20Authors_final_0.pdf

A.2 United Kingdom & Europe

• Quality Assurance Agency (QAA), UK: "Advice and resources on Generative AI." Quality and standards lens. https://www.qaa.ac.uk/sector-resources/generative-artificial-intelligence/qaa-advice-and-resources
• Jisc (UK): "Exploring AI and assessment" & "Embracing Generative AI in Assessments—Flowcharts." Conceptual framework and practical flowcharts. https://www.jisc.ac.uk/blog/exploring-ai-and-assessment-avoid-outrun-or-embrace
• University College London (UCL): "Generative AI Hub." Comprehensive institutional hub. https://www.ucl.ac.uk/teaching-learning/generative-ai-hub
• European Commission: "Ethical guidelines on the use of artificial intelligence and data in teaching and learning for educators." Ethics-forward, focusing on rights. https://www.europarl.europa.eu/cmsdata/196377/AI%20HLEG_Ethics%20Guidelines%20for%20Trustworthy%20AI.pdf

• Council of Europe: "Artificial Intelligence and Education." High-level values and governance framework. https://www.coe.int/en/web/education/artificial-intelligence
• European University Association (EUA): Publications on AI and quality assurance. Sector perspective. https://www.eua.eu/publications.html

Appendix B: Student-Facing Guide for ChatGPT Interactions (En Español)

Objetivo de la actividad:

• Usar ChatGPT para mejorar su habilidad en español a través de una conversación continua.
• Asegúrese de hablar en español y de formular preguntas claras y específicas.

Cómo empezar:

Para iniciar la conversación, presione el botón de micrófono en la aplicación y diga en voz alta lo que desea preguntar o discutir. Espere la respuesta y continúe la conversación formulando preguntas adicionales.

Sugerencias de temas de conversación:

Historia de los países hispanohablantes: Pregunte sobre los eventos históricos más importantes de países como México, Argentina, Colombia, o España. Por ejemplo: "¿Puede hablarme de la independencia de México?" o "¿Qué papel jugó Simón Bolívar en América Latina?"
Cultura y tradiciones: Pregunte acerca de festividades, costumbres, o tradiciones populares, como el Día de los Muertos, la Semana Santa en Sevilla, o la celebración del Inti Raymi en Perú.
Literatura hispana: Puede preguntar sobre autores importantes, como Gabriel García Márquez, Isabel Allende, o Federico García Lorca, y sus obras más conocidas.
Arte y música: Explore el arte hispano, desde el muralismo mexicano hasta las obras de Pablo Picasso. También puede preguntar sobre géneros de música como el flamenco, el tango, la salsa, o la cumbia.
Gastronomía: Pregunte sobre los platos típicos de cada país, como la paella en España, el ceviche en Perú o las pupusas en El Salvador.
Lugares de interés: Pida información sobre lugares icónicos como Machu Picchu en Perú, la Alhambra en España, o la Ciudad de México. Pregunte sobre su historia, significado y curiosidades.
Lenguas indígenas y diversidad lingüística: Investigue sobre la influencia de lenguas indígenas como el náhuatl, quechua o guaraní en el español.
Consejos para la conversación: Haga preguntas abiertas y siga el flujo de la conversación. Esto le permitirá obtener respuestas más completas y aprender más sobre la diversidad del mundo hispano.

Appendix C: Sample Integrated Lesson Plan – "Machu Picchu: Architecture and Society"

Level: Advanced (ACTFL Advanced Low/Mid, CEFR B1/B2)

Time Frame: Two 75-minute class periods + independent work.
Tech Tools: Meta Quest Pro headsets (or 360-degree video alternative), ChatGPT (voice mode).

Learning Objectives:

• Cultural: Analyze how the architecture and urban planning of Machu Picchu reflect the societal values, religious beliefs, and relationship with the environment of the Inca civilization.
• Linguistic: Utilize descriptive and analytical vocabulary related to architecture, geography, and society in spoken and written Spanish. Practice forming and asking complex questions.

Procedure:

• Day 1 – Pre-VR Context & Exploration (75 mins):

Introduction (20 mins): Instructor provides a brief historical and cultural overview of the Inca Empire and Machu Picchu in Spanish. Introduces key vocabulary (terrazas, templo, observatorio, piedra, montaña, agricultura, astronomía, sociedad).
VR Exploration (40 mins): Students, in pairs or small groups, use VR headsets (or curated 360-video tour) to explore Machu Picchu. They are guided by a worksheet (in Spanish) with prompts: 1) Describe las terrazas. ¿Qué función crees que tenían? 2) Identifica tres tipos diferentes de estructuras (ej. religiosa, residencial). 3) ¿Cómo se integra la ciudad en el paisaje montañoso?
Initial Debrief (15 mins): Quick whole-class share of first impressions and observations in Spanish.

Independent Work – AI-Powered Inquiry:

ChatGPT Interaction (Homework): Students are tasked with using ChatGPT (Spanish voice mode) to delve deeper. Prompt: "Actúa como un arqueólogo especialista en la civilización inca. Yo acabo de hacer un tour virtual de Machu Picchu. Responde a mis preguntas en español, con un lenguaje claro pero académico." Students must prepare at least 5 questions based on their VR observations (e.g., "¿La orientación de los edificios tenía significado astronómico?"). They save a transcript of the conversation.

• Day 2 – Synthesis, Critique, and Creation (75 mins):

AI Output Analysis (25 mins): In small groups, students compare their ChatGPT transcripts. Guided by the instructor, they evaluate: Were the answers consistent? Did any information seem questionable? What new vocabulary did they learn?

Source Verification & Synthesis (30 mins): Students are given a short, credible academic text (or video) about Machu Picchu. They cross-check one key claim from their ChatGPT conversation against this source, noting discrepancies or confirmations.

Final Creative Output (20 mins): Each student writes a brief "informe de arqueólogo" or creates a short spoken presentation (2 mins) summarizing one key insight about Inca society, supported by evidence from their VR observation, AI conversation, and the verified source.

Assessment: Formative assessment via the VR worksheet and participation in discussions. Summative assessment via the final creative output, evaluated on accuracy of cultural insight, effective use of new vocabulary, and demonstrated synthesis of multiple information sources (VR, AI, text).