Journal of Language Teaching, *5*(2), 1-8, 2025
https://doi.org/10.54475/jlt.2025.005

Review OPEN ACCESS

Navigating intercultural competence with ChatGPT: Implications and recommendations for foreign language education

Yeon Mi Lee

Department of World Languages and Cultures, Samford University, Homewood, Alabama, United States

Received: May 10, 2025 / Accepted: June 2, 2025 / Published Online: June 4, 2025

Abstract

Despite the increasing corpus of research concerning the application of ChatGPT in adaptive learning, grammar, and lexical acquisition in foreign language education (FLE), its effectiveness on the cultivation of intercultural competence (IC) as an essential learning outcome in FLE remains comparatively underexplored. This review article introduces the theoretical background and definitions of intercultural competence in second language acquisition (SLA) before analyzing empirical studies that examine ChatGPT's cross-cultural performance across multiple languages, cultures, and methodologies. It then discusses the implications of utilizing ChatGPT for intercultural engagement in FLE. The article concludes by underscoring the imperative of critical artificial intelligence literacy (CAIL) among language learners and offers recommendations for best practices in FLE.

Keywords intercultural competence, foreign language education, second language acquisition, ChatGPT, critical artificial intelligence literacy

1. Introduction

> *Language is not merely a reproducing instrument for voicing ideas, but is itself a shaper of ideas. …We see and hear and otherwise experience very largely as we do because language habits of our community predispose certain choices of interpretation.*
>
> –Benjamin Lee Whorf, Language, Thought, and Reality (1959, p. 212)

While the integration of the intercultural learning outcomes into the mainstream L2 curricula and courses did not begin until the 1990s, driven by the new demands of globalization and international affairs, its theoretical underpinnings in SLA can be traced back to the 1920s. Benjamin Lee Whorf's ideas on the intrinsic link between language and culture, as illustrated in the preceding epigraph, derived from the rising notions in the field of cultural anthropology and linguistics of the 1920s and 1930s, primarily from the work of his mentor and linguist Edward Sapir (Subbiondo, 2005). Sapir and Whorf's ideas on the cultural determinants of human language laid the groundwork for the sociocultural theoretical framework in the following decades (Thorne & Tasker, 2011). In psychology, Sapir's ideas met the theoretical construct of Vygotsky's Sociocultural Theory (SCT), whose emphasis on the significance of contextual knowledge in learning later brought a great impact to the field of second language acquisition; for example, notions such as "interaction and negotiation" in speech analysis and development was introduced by Michael Long (1980), building awareness about the constructivist process wherein speech is co-constructed, navigated, and negotiated between speakers. Thus, sociocultural awareness shifted the traditional paradigms of SLA by taking language acquisition beyond the confines of lexical and grammatical structures and integrating cultural competence as an essential learning outcome. Wiseman et al. (1989) assert that "cultural knowledge is an important determinant of one's ability to minimize misunderstandings with someone from another culture. Cultural knowledge has a positive effect on other [cross-cultural competence] attributes and maximizes intercultural competency" (p. 351).

In recent years, the swift adoption of OpenAI's ChatGPT as a digital learning tool in FLE has driven significant scholarly interest, presenting its impact on adaptive learning, personalized feedback, grammar, and lexical acquisition (Anjum et al., 2024). While the advantages of integrating ChatGPT into L2 pedagogy continue to be widely explored, discussion concerning its impact and implications on the development of intercultural competence among L2 learners has garnered less attention. The current article introduces the theoretical frameworks and definitions that have shaped the discourse around intercultural competence in FLE. It then provides an analysis of the empirical studies that focus primarily on assessing ChatGPT's performance in engaging cultural knowledge and interaction across multiple languages and cultures. Drawing upon this review, the current article offers insights on potential implications and recommendations for FLE. Finally, the article

concludes by highlighting the significance of cultivating critical artificial intelligence literacy (CAIL) to educate L2 learners and instructors about the potential risks and benefits of utilizing ChatGPT in cross-cultural contexts.

2. Intercultural competence in FLE: Theoretical background and definitions

The expansion of globalization and international trade during the late 20th century created a high demand for professionals with optimal skills capable of navigating the new multicultural landscape (Garrett, 2025). Intercultural competence emerged as an essential outcome for L2 or foreign language programs, as language associations and institutions provided official guidelines on its implementation; in the United States, the World-Readiness Standards was established in 1996 by the American Council on the Teaching of Foreign Languages (ACTFL), offering specific guidelines for achieving cultural learning outcomes, known as the 5 C's framework– Communications, Cultures, Connections, Comparisons, Communities. In Europe, the Common European Framework for Languages began to provide similar guidelines (Garrett, 2025).

Since the 1990s, intercultural competence in FLE has garnered significant attention from scholars calling for this outcome to "be examined and interpreted as a multifaceted process" (Stier 2006, p. 5). A range of pedagogical frameworks aimed at cultivating this skill were proposed via traditional classroom instruction, cultural immersion via study-abroad or experiential learning, and other comparable methods (Deardorff 2006; Stier 2006; Leask 2015). Along with increased interest in methodology came varied attempts to define this concept: Alvino Fantini defines intercultural competence as "the complex of abilities needed to perform effectively and appropriately when interacting with others who are linguistically and culturally different from oneself" (2005, p.1). On the other hand, Hammer's definition underlines the dynamic, nuanced process of interaction where the speakers acquire "the capability to shift one's cultural perspective and appropriately adapt behavior to cultural differences and commonalities" (2015, p. 483). Likewise, Spitzberg and Changnon define intercultural proficiency as "the appropriate and effective management of interaction between people who, to some degree or another, represent different or divergent affective, cognitive, and behavioral orientations to the world" (2009, p. 7).

Theoretical considerations on the concept of intercultural competence have followed a long trajectory of epistemic genealogy since Edward Sapir and Benjamin Lee Whorf introduced their foundational notion on the interrelationship between language and culture (Subbiondo, 2005). The development of Lev Vygotsky's Sociocultural Theory in psychology, for whom the individual cognitive competence is dialectically linked to their community's culture, is a prime example of the overlapping impact of the emerging ideas that began to circulate at the time (Vygotsky, 1978). Vygotsky's ideas later influenced key concepts in applied linguistics, including the notion of "comprehensible input" by Stephen Krashen (1982) and "comprehensible output" by Merrill Swain (1985), both of which underscore the significance of cultural and contextual knowledge in achieving effective L2 comprehension and production.

Acceptance and practice of intercultural competence were further validated and recognized thanks to its incorporation into the discipline of Intercultural Communication Studies (ICS). Until the early 1990s, the domain of cultural competence in FLE had been strongly influenced by comparativist and positivist models of cross-cultural psychology, where culture was primarily understood in terms of nationality and compared through generalized constructs (Hofstede, 1991; Triandis, 1990). However, the rise of ICS in the 1990s reinforced the understanding of human communication as a dynamic, interpersonal process for effective meaning making and negotiation (Liddicoat & Scarino, 2013). Michael Byram's intercultural framework was developed based on this notion, outlining practical and ethical objectives in achieving comprehensive linguistic and cross-cultural proficiency in five competencies, succinctly summarized by Schenker (2012) as "knowledge of self and other, attitudes of openness and curiosity, skills or interpreting and relating, skills of diversity and interaction, and critical cultural awareness" (p. 450). The new intercultural frameworks shifted the old paradigms of FLE towards a more holistic, integrated framework prioritizing meaningful intercultural engagement in language education, as well as its broad acknowledgement across L2 language programs and curricula. As Byram et al. (2002) suggest, the goal of L2 intercultural competence is "to cultivate learners' cross-cultural competence alongside linguistic proficiency; to prepare them for engagement with individuals from diverse cultures; to facilitate their understanding and acceptance of others as unique entities possessing different perspectives, values, and behaviors; and to assist them in recognizing that such interactions serve as enriching experiences" (2002, p. 10).

3. Empirical research on ChatGPT's intercultural performance and its impact on FLE

The use of ChatGPT for a designated purpose in FLE has incited substantial interest within the domain of applied linguistics (see, e.g., Huang et al., 2022; Xiao et al, 2023; Zou et al., 2023), garnering both excitement as well as concern for its adoption in educational settings. One area of concern is L2 learners' lack of awareness and uncritical engagement with the AI chatbots, utilizing them as a search engine or ultimate authority for the knowledge they are seeking. This new digital environment has redefined the L2 learners' educational experience and critical decision-making processes for language programs (Darvin, 2025). Kirschenbaum and Raley (2024) highlight the epistemological concerns deriving from the development of new GPT models; GPT-4 features approximately 1.8 trillion parameters compared to the 175

billion parameters of the previous model GPT-3. What this indicates "is not yet another novel computational application or approach but rather a general condition of language and life" (2024, p. 509). This condition raises ethical and epistemological concerns as the chatbots' competence in language mimicry may not inherently translate into output that is free from cultural bias or misinterpretation. For the purposes of FLE, the application of ChatGPT in L2 intercultural engagement carries profound implications for culturally sensitive and inclusive pedagogy, particularly concerning the representation of cultural perspectives and worldviews pertaining to ethnic minority cultures and languages. This section provides a selective review of empirical studies on ChatGPT's intercultural performance, guided by three criteria: 1) the studies experimented with multiple languages and/or (sub)cultures to measure the IC performance of ChatGPT; 2) the studies represented methodological diversity employing computational benchmarking against human survey data as well as qualitative analysis from direct AI-human interactive engagement; 3) the studies addressed the impact of ChatGPT's IC performance on educational settings.

3.1. Empirical research from computational benchmarking against pre-existing human survey data

Georgiou (2025) investigated the cultural performance of ChatGPT by evaluating the chatbot's response to a prompt eliciting a general description of a selection of developed and developing countries based on the Human Development Index (HDI). The study employed ChatGPT-3.5 to prompt cultural descriptions of 20 countries–10 developed or high HDI countries, such as Singapore, Sweden, Denmark, etc., and 10 developing or low-HDI nations, such as Sierra Leone, Niger, South Sudan, and others. Drawing on the chatbot's responses, Georgiou (2025) conducted a sentiment analysis using the quantitative discourse analysis (QDAP) package within the R programming environment and implemented the Bayesian regression model to analyze the data. The study found that although ChatGPT produced generally positive descriptions across all countries, Bayesian statistical analysis showed that ChatGPT used language that encompassed more positive sentiments for countries with elevated HDI scores–predominantly European nations–compared to their low-HDI counterparts, largely from Africa. ChatGPT's description of the different countries and their cultures indicated a higher sentiment score associated with high-HDI nations, while the language used for low-HDI countries generated a lower sentiment score. These findings have important implications for the cultivation of intercultural competence in FLE, as ChatGPT users' lack of critical awareness and holistic understanding of cultural diversity within developing nations "may perpetuate perceptions of superiority or inferiority based on national economic status" (Georgiou, 2025, p. 5), posing the risk of reinforcing cultural stereotypes. While Georgiou's research (2025) is somewhat limited due to its concentration on a single language (English), it is nonetheless a compelling study that highlights the potential effects of utilizing ChatGPT in cross-cultural educational contexts.

Cao et al. (2023) examined the cultural alignment of ChatGPT by evaluating its responses against human-generated data from pre-existing responses to the Hofstede Culture Survey. Their experiment included three standardized prompts with questions across five distinct languages and cultures (English, Chinese, German, Spanish, Japanese) and employed Hofstede's framework across six dimensions (e.g., Power Distance, Individualism, Masculinity, Uncertainty Avoidance, etc.). A multi-turn interaction strategy with three categories of knowledge injection was employed to evaluate the model's adaptability and consistency with culture-specific questions. Findings revealed that the type of interaction strategy used in multi-turn conversations can impact the responses from ChatGPT, uncovering the dependency of the model's performance on prompts and conversation structures. Moreover, the study found that ChatGPT's cultural alignment scores varied significantly across the different cultures. ChatGPT performed in higher alignment with contexts familiar to American cultural values, which increased even more when prompted in English. Consistency rates for English prompts exceeded 70% and tended to favor American norms. ChatGPT responses showed better alignment when prompted using the native language of the target culture rather than in English, suggesting the prompt language has a great influence on ChatGPT-generated responses. The study by Cao et al. (2023) also found that ChatGPT's cultural alignment decreased when prompted in non-English languages, but Chinese and German tended towards a stronger alignment. These findings are particularly telling for L2 education other than English, given that ChatGPT's training data is predominantly English (96%, according to Ouyang et al.), resulting in its tendency to default to American cultural norms and paradigms. ChatGPT's language dominance poses challenges for non-English L2 learners and their intercultural outcomes, underscoring concerns about culturally misaligned information regarding language minorities of the Global South.

Similar to the study by Cao et al. (2023), the research undertaken by Wang et al. (2023) assessed cultural alignment of LLM's outputs in relation with pre-existing human-generated data from the World Values Survey. They utilized two models of LLMs, OpenAI's ChatGPT-4 and text-davinci-003. The study developed a benchmark dataset, incorporating tangible cultural artifacts (e.g., holidays, films, literature, songs, etc.) as well as intangible cultural values (e.g., opinions and beliefs), and categorized the analysis into two cultural dimensions –survival versus self-expression values, and traditional versus secular-rational values. Wang et al. compared the different cultural responses in six distinct languages: English, Chinese, Russian, Indonesian, Hindi, and Arabic. Similar to the findings uncovered by Cao et al. (2023), this study found that ChatGPT's responses were consistently more congruent with the English World Values Survey data, even when prompted to respond within specific cultural frameworks, such as Chinese or Arabic. The study found that chatbot responses increased in cultural alignment with English-speaking cultures and nations, corroborating Cao et al.'s findings about the linguistic and cultural

dominance of the English language found in AI-assisted environments. For Wang et al. (2023), the predominant influence of English in data training of LLMs reinforces a systemic issue dominated by the English cultural paradigms. As for L2 education, the study suggests caution and awareness when interacting with ChatGPT for cultural learning outcomes in non-English language pedagogy, as the chatbots' cultural misrepresentations may hinder the acquisition and engagement of culturally sensitive skills. As a result, the study suggests the need for robust, human-assisted intercultural education in FLE and calls for a critical approach to AI-assisted technologies within controlled settings.

Resembling in methodology and findings with Wang et al.'s study (2023), the experiment conducted by Tuna et al. (2024) deployed two LLM models, GPT-3.5-turbo and GPT-4, and compared their responses against human subject responses derived from the World Values Survey. Similar to previous studies, this study compared chatbots' output against pre-existing data on cultural values across cultures and nations, rather than engaging in human responses in the conventional form of interviews or qualitative data. The study probed ChatGPT's cultural performance in five distinct languages: German, French, English, Spanish, and Portuguese. Moreover, the experiment included studies within ten subcultural variations: English-speaking Great Britain and the USA, German-speaking Germany and Austria, Spanish-speaking Spain and Mexico, French-speaking Canada and France, and Portuguese-speaking Brazil and Portugal. Tuna et al.'s assessment (2024) employed inquiries on topics such as Trust, Faith, and Happiness, and measured the proximity of the GPT's response to existing human cultural norms, using the Euclidean distance. Similar to the findings from previous studies, Tuna et al. (2024) found that the chatbots showed closer proximity to Euclidean distance when prompted in English within Western-centric cultures, and that a downgraded version of GPT-3.5-turbo performed better than GPT-4 in cultural alignment, especially within the German language setting. However, both models showed lower cultural alignment when engaging subcultures and languages, particularly in Mexican Spanish and Brazilian Portuguese. For instance, in Mexican Spanish, the GPT-3.5-turbo's average distance from human values was significantly higher, indicating a clear lack of cultural nuance essential for language interaction in this subculture. The chatbots' high degree of distance from human cultural values in a specific subculture suggests that learners using these models may be exposed to culturally inappropriate or oversimplified content that can hinder the development of intercultural communicative abilities. Without accurate cultural representations, language learners can adopt and internalize expressions or views that are misaligned with the sociocultural expectations of native cultures and speakers. Likewise, these findings underscore the imperative of teaching intercultural competencies so that students are equipped to engage with the cultural mindset of the native cultures and speakers.

Ahmad et al. (2024) further investigated the efficacy of ChatGPT in capturing the cultural competence and nuances inherent in languages other than English. They examined the chatbots' performance in Hausa, a low-resource language utilized primarily in the West African region. The research assessed ChatGPT's outputs and compared them to those provided by 18 native Hausa speakers located in Nigeria. It employed 37 culturally relevant prompts that elicited responses on cultural norms and emotional expectations. The methodology encompassed two distinct phases, initially with participants generating their own open-ended responses to the questions, and then having them evaluate the cultural and emotional authenticity of ChatGPT's output utilizing a Likert scale. This two-phase process allowed the researchers to evaluate both the semantic similarity and emotional resonance, which are essential elements for navigating intercultural language exchange. Ahmad et al. (2024) found significant cultural and emotional discrepancies when compared to human responses. While ChatGPT displayed semantic similarity, the authors found that the chatbots lacked emotional depth, defaulting frequently to neutral tones or perspectives. For example, when prompted on the question "How would you feel if your student calls you by your first name?", ChatGPT responded with a polite acceptance, while the majority of Hausa speakers regard it to be culturally inappropriate and disrespectful. An average of only 8.2 participants said GPT's responses were likely articulated by a native Hausa speaker, while 5.2 expressed otherwise. Ahmad et al.'s study (2024) demonstrates that overreliance on LLM models could inadvertently promote generic or culturally inappropriate notions, particularly within underrepresented linguistic communities.

3.2. Empirical research from direct human interaction with ChatGPT

The studies examined thus far provided insights utilizing standardized prompts and pre-existing human responses to static survey data that measured ChatGPT's cultural proficiency. However, within their limitations is the absence of direct user engagement with the chatbots in the form of iterative conversation, interaction, direct feedback, and provision of additional contexts that more closely resemble real-world exchange in a dynamic environment. Therefore, in what follows, we will review empirical studies that have been conducted by means of interactive user feedback and direct engagement with the chatbots to assess their impact in a more fluid, collaborative setting.

Masato Tahara's study (2024) brings significant insights regarding the application of ChatGPT in the domain of L2 translation and cultural competence. Tahara investigated ChatGPT-4's cultural performance in the context of collaborative translation and dialogic reading, instructing five Japanese and two Malaysian Chinese students to engage with AI-translated adaptations of the Japanese novel *Jimmy* by Aoumi in English and Chinese. The author instructed the students to analyze cultural interpretation and intricate complex meanings from AI-generated translation. This case study integrated a dual-phase translation framework that incorporated GPT's translation with human collaboration. The method followed AI-translated text (Translation 1), followed by a

refinement process with human-addressed critical reading questions. The chatbots' responses were uploaded as analytical texts to inform a subsequent translation (Translation 2). The research showed how the diversity of cultural backgrounds from participants influenced their interpretations. For example, the Japanese participants resonated with societal pressures represented in the narrative, while Malaysian Chinese participants gave multiple perspectives on multicultural norms. Their results found that Translation 2 was more effective in conveying cultural nuances and characters' psychology, suggesting that the collaborative framework built upon human input and interaction with ChatGPT can enhance cross-cultural understanding. Tahara's study (2024) on AI-human interaction shows that new meanings can arise from the confluence of varied cultural inputs generated by users, lending to the chatbots' potential for developing interpretive flexibility and adjustment based on direct human feedback.

In engaging ChatGPT's direct interaction with human subjects, Darvin (2025) examined the interactions of six secondary school students (Grades 8-12) in British Columbia who spoke languages other than English in their home, and explored the implications of machine-human interaction for cross-cultural competence. The research utilized a case study methodology and data consisting of semi-structured interviews and multimodal discourse analysis of GenAI interactions with ChatGPT, Copilot, and CharacterAI. This methodology enabled the author to observe the ways the students engaged with these tools, the digital literacies enacted during the interaction, and the cultural knowledge reflected in the generated outputs. The results revealed that, while the students' interaction with ChatGPT in areas specific to lexicality, grammatical correction, and rewriting was promising, a majority of students displayed a lack of critical awareness about the mechanisms by which ChatGPT generated its responses. Many students, lacking critical digital literacy, approached the tool as a neutral source of information, failing to think critically about its cultural assumptions inherent in those outputs. One example is that ChatGPT frequently aligned with prevailing discourses or cultural assumptions about academic writing, discouraging the use of first-person pronouns, which represents a specific cultural bias. These behaviors reflect potential hazards of promoting language skills without concurrent development of cultural competence. Another critique that the author makes illustrates how particular AI features like avatar customization or premium access influence learners' engagement levels that can open or restrict their participation in more complex writing tasks or cross-referencing contexts, depending on the type of electronic device in use. This is an important finding because it affects the students' learning experience, what they learn, and what they are exposed to. Moreover, the study brings a critical perspective to the AI tools' lack of ability to accommodate low-resource languages, resulting in misinterpretations of languages such as Vietnamese or Malayalam, which exacerbates the cultural bias against non-dominant cultural identities. As a result, while Darvin (2025) does validate the utility of ChatGPT for L2 learning as a practical tool for areas specific to lexical or grammar acquisition, the author critiques its potential risks for cross-cultural engagement. Among the limitations of Darvin's study (2025) is its relatively small pool of participants, which may not reflect adequately the diversity of L2 learners on a larger scale. Despite this, Darvin's study (2025) is compelling research that supports the argument that effective L2 learning must involve a culturally critical perspective beyond mere technical proficiency.

In assessing ChatGPT's impact on the development of culturally-proficient L2 curriculum and instructional design, Kim et al.'s study (2023) employed a virtual learner persona – a Korean undergraduate student learning business English writing competencies for employment. The simulated learner had intermediate English proficiency and five years of language experience. While the study is limited due to the lack of live interaction with real-human learners, the experiment nevertheless enabled the authors to assess and gain insight into the quality of GPT's response in the area of instructional design and cultural competence in educational scenarios. The study deployed a two-phase methodological framework: first, GPT was instructed to design a business English curriculum based on H. Douglas Brown's model for L2 course development. Then, the chatbot was prompted to engage in teaching this course using a Task-Based Language Teaching (TBLT) approach. The research measured GPT's performance in building lesson plans, task assignment, and feedback capability. Kim et al. (2023) found that the chatbot was successful in generating structure, topic-relevant curriculum, offering effective examples for writing assignments, and linguistic scaffolding. However, in the area of cross-cultural language proficiency, the authors found that ChatGPT lacked the ability to replicate real-world business scenarios that engage in nuanced, distinct cultural settings. While the chatbots performed the prompted tasks, the level of engagement was simplistic and superficial, displaying limited connection with sociocultural variation and context-specific considerations in real-life communication. For instance, the study found that ChatGPT feedback was highly confined to grammatical and lexical correction, undermining deeper communicative and cultural dimensions like politeness strategies, business cultural norms, and regional differences in business etiquette. Further, Kim et al. (2023) found that the chatbots are overly focused on grammatical accuracy rather than the ability to engage and perform interculturally and pragmatically. These shortcomings exposed the learner to missed opportunities for essential intercultural engagement and effective communication in multicultural business settings. While the GPT is a promising tool for self-directed learning, its cultural performance is limited when addressing sociolinguistic depth for authentic intercultural communication. Given the drawbacks, Kim et al. (2023) suggest that L2 language learners should use the tool under the supervision and guidance of educators who can deliver expertise on culturally-appropriate knowledge and critical engagement. Kim et al.'s research (2023) underscores the importance of human-led instructional oversight and guidance for adequate cultural skills to ensure L2 learners are equipped with the tools needed to

navigate diverse communicative contexts with cultural sensitivity and confidence in multicultural business settings.

4. Implications and recommendations for foreign language education

Empirical studies on the integration of ChatGPT into the acquisition of intercultural skills in SLA demonstrate limitations and opportunities for L2 learners and educators alike. While the chatbots' performance in supporting FLE in areas like linguistic scaffolding, grammar, and lexical acquisition are promising, challenges remain when adopting the tool to enhance L2 intercultural outcomes: among the persisting challenges are the prevalence of cultural and linguistic dominance of algorithmic contexts that resource the AI-pre-training datasets, which results in increased chatbot's alignment with dominant cultural paradigms (Wang et al., 2023; Cao et al., 2023). This has important implications for L2 intercultural pedagogy, for the teaching of cultural norms of subcultures and linguistic minority groups can be impacted by the overwhelming prevalence of linguistic dominance encountered in AI-enhanced environments. This, in turn, can lead to diminished cultural perspectives and visibility of minority languages and subcultures in FLE settings, course materials, curriculum, instructional design, and more (Broadhead, 2024). To tackle this problem, the current article calls for the imperative of cultivating critical artificial intelligence literacy (CAIL) among students and instructors to inform best practices in FLE. The purpose of CAIL is to raise awareness around the epistemic impact of LLMs and the algorithmic processes whereby knowledge is generated and disseminated in digital environments. As the empirical studies have demonstrated, algorithmic processes can often perpetuate prevailing cultural viewpoints while marginalizing alternative perspectives. Therefore, the current article offers four actionable insights or recommendations for cultivating CAIL in FL intercultural contexts:

1) Assist FL learners with specific guidelines and pedagogy to foster critical evaluation of cross-cultural information sourced by algorithmic systems. Supervise the use of ChatGPT for intercultural outcomes, facilitating access to human-led cultural expertise and cross-referencing practices.
2) Provide holistic guidance to facilitate AI-human collaborative learning and critical thinking practices through iterative, dialogic processes of human-generated input and feedback (Tahara, 2024).
3) Offer opportunities for reflective practices in the form of cultural response, interpretation, or reflective journal to address the discrepancies between AI-generated content and humans' lived experiences and cultivate metacognitive awareness about the process of knowledge construction and representation (Darvin, 2025).
4) Practice critical evaluation and cross-cultural comparison by contextualizing ChatGPT's culture-specific prompts and by examining its response within non-dominant cultural frameworks (Wang et al., 2023; Tuna et al., 2024).

While the chatbots may perform well syntactically, their output often fails to capture in-depth cross-cultural representations that reflect nuanced sensitivity to underrepresented linguistic communities. Educators and learners alike should be aware of these practical recommendations that can assist in achieving a more culturally inclusive and pedagogically relevant experience in FLE.

5. Conclusion

The seemingly boundless use of ChatGPT in educational contexts has generated a large corpus of research in FLE, noting its benefits in enriching lexical and grammar acquisition, L2 writing, and self-directed and adaptive learning. However, its impact on fostering intercultural competence continues to be developed and should be approached with caution and acute awareness about its potential risks. Critical thinking skills and methods should be promoted among L2 learners to equip them with the tools to evaluate knowledge production and cultural output generated by AI, particularly in assessing culturally sensitive knowledge of language minorities and subcultures. Algorithmic cultural (mis)representations can inadvertently entrench systemic inequities through their classification processes, and hence diminish the potential for sustainability and preservation of heritage and less-commonly taught languages. As Tahara (2024) suggests, GPT should be leveraged to advance language acquisition, but users should be critically aware of its best practices and risks involving L2 intercultural engagement.

Professor Yeon Mi Lee received her PhD degree from the University of California, Los Angeles. Dr. Lee is a faculty member in the Department of World Languages and Cultures at Samford University in the United States.
Email: ymlee1331@gmail.com

References

Ahmad, I. S., Dudy, S., Ramachandranpillai, R., & Church, K. (2024). Are generative language models multicultural? A study on Hausa culture and emotions using ChatGPT. *arXiv:2406.19504*. https://doi.org/10.48550/arXiv.2406.19504

Anjum, F., Raheem, B. R., & Ghafar, Z. N. (2024). The impact of ChatGPT on enhancing students' motivation and learning engagement in second language acquisition: Insights from students. *Journal of e-learning Research*, *3*(2), 1-11. https://doi.org/10.33422/jelr.v3i2.679

Broadhead, L. (2024). Insidious chatter versus critical thinking: Resisting the Eurocentric siren song of AI in the classroom. *Journal of Applied Learning and Teaching*, *7*(2). 28-37. https://doi.org/10.37074/jalt.2024.7.2.9

Byram, M., Gribkova, B, & Starkey, H. (2002). *Developing the intercultural dimension in language teaching. A practical introduction for teachers.*

Council of Europe Publishing, Language Policy Division.
Cao, Y., Zhou, L., Lee, S., Cabello, L., Chen, M., & Hershcovich, D. (2023). Assessing cross-cultural alignment between ChatGPT and human societies: An empirical study. *arXiv:2303.17466*. https://doi.org/10.48550/arXiv.2303.17466
Crawford, Kate. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.
Darvin, R. (2025). The need for critical digital literacies in generative AI-mediated L2 writing. *Journal of Second Language Writing, 67*, 101186. https://doi.org/10.1016/j.jslw.2025.101186
Deardorff, D. K. (2006). Identification and assessment of intercultural competence as a student outcome of internationalization. *Journal of Studies in International Education, 10*(3), 241–266. https://doi.org/10.1177/10283153062870
Fantini, A. (2005). *About intercultural communicative competence: A construct*. School for International Training.
Garrett-Rucks, P. (2025). Assessing intercultural competence. In C. Facke, X. Gao & P. Garrett-Rucks (Eds.), *The Handbook of plurilingual and intercultural language learning* (pp. 247-260). John Wiley & Sons.
Georgiou, G. P. (2025). ChatGPT exhibits bias toward developed countries over developing ones, as indicated by a sentiment analysis approach. *Journal of Language and Social Psychology*, *44*(1), 132-141. https://doi.org/10.31234/osf.io/49gzc_v1
Hammer, M. R. (2015). Intercultural competence development. In J. M. Bennett (Ed.), *The SAGE encyclopedia of intercultural competence* (pp. 483-486). Sage Publications.
Hofstede, G. (1991). *Cultures and organisations: Software of the mind*. McGraw Hill.
Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning—are they really useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning, 38*(1), 237–257. https://doi.org/10.1111/jcal.1261
Kim, S.-Y., Shim, J., & Shim, J. (2023). A study on the utilization of OpenAI ChatGPT as a second language learning tool. *Journal of Multimedia Information System*, *10*(1), 79–88. https://doi.org/10.33851/jmis.2023.10.1.79
Kirschenbaum, M. & Raley, R. (2024). AI and the University as a service. *Pmla/Publications of the Modern Language Association of America, 139*(3), 504-515. https://doi.org/10.1632/S003081292400052X
Krashen, S. (1982). *Principles and practice in second language acquisition*. Oxford University Press.
Leask, B. (2015). *Internationalizing the curriculum*. New York: Routledge.
Liddicoat, A. & Scarino, A. (2013). *Intercultural language teaching and learning*. Wiley-Blackwell.
Long, M. H. (1980). *Input, interaction, and second language acquisition*. University of California, Los Angeles.
Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Siemens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). Training language models to follow instructions with human feedback. *arXiv:2203.02155*. https://doi.org/10.48550/arXiv.2203.02155
Schenker, T. (2012). Intercultural competence and cultural learning through telecollaboration. *CALICO Journal, 29*(3), 449–470. https://doi.org/10.11139/cj.29.3.449-470
Spitzberg, B., & Changnon, G. (2009). Conceptualizing intercultural competence. In D. K. Deardorff (Ed.), *The Sage handbook of intercultural competence* (pp. 2–52). SAGE Publications. https://doi.org/10.4135/9781071872987.n1
Stier, J. (2006). Internationalisation, intercultural communication and intercultural competence. *Journal of Intercultural Communication, 11*(1), pp. 1–12. https://doi.org/10.36923/jicc.v6i1.422
Subbiondo, J.L. (2005). Benjamin Lee Whorf's theory of language, culture, and consciousness: A critique of Western science. *Language and Communication*, *25*(2), 149-159. https://doi.org/10.1016/J.LANGCOM.2005.02.001
Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Newbury House.
Tahara, M. (2024). How does Generative AI function as an active cross-cultural communication agent? *INContext*, *4*(2), 63–83. https://doi.org/10.54754/incontext.v4i2.103
Thorne, S. L., & Tasker, T. (2011). Sociocultural and cultural-historical theories of language development. In *The Routledge handbook of applied linguistics* (pp. 487-500). Routledge. https://doi.org/10.4324/9780203835654.ch34
Triandis, H.C. (1990). Cross-cultural studies of individualism and collectivism. In J.J. Berman (ed.), *Cross-cultural perspective* (pp. 41–33). University of Nebraska Press.
Tuna, M., Schaaff, K., & Schlippe, T. (2024, November). Effects of language-and culture-specific prompting on ChatGPT. In *2024 2nd International Conference on Foundation and Large Language Models (FLLM)* (pp. 73-81). https://doi.org/10.1109/fllm63129.2024.10852463
Vygotsky, L. (1978). *Mind in society. The development of higher psychological processes*. (M. Cole, V. John-Steiner, S. Scribner & E. Souberman (Eds.). Harvard University Press. https://doi.org/10.2307/j.ctvjf9vz4
Wang, W., Jiao, W., Huang, J., Dai, R., Huang, J., Tu, Z., & Lyu, M. R. (2023). *Not all countries celebrate Thanksgiving: On the cultural dominance in Large Language Models. arXiv:2310.12481*. https://doi.org/10.48550/arXiv.2310.12481
Whorf, B. L. (1959). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf* (J. B. Carroll, Ed.). MIT Press.

Wiseman, R.L., Hammer, M.R. and Nishida, H. (1989). Predictors of intercultural competence, *International Journal of Intercultural Relations*, *13*(3), 349–370. https://doi.org/10.1016/0147-1767(89)90017-5

Xiao, F., Zhao, P., Sha, H., Yang, D., & Warshauer, M. (2023). Conversational agents in language learning. *Journal of China Computer-Assisted Language Learning*. https://doi.org/10.1515/jccall-2022-0032

Zou, B., Reinders, H., Thomas, M., & Barr, D. (2023). Editorial: Using artificial intelligence technology for language learning. *Frontiers in Psychology, 14*, 1287667. https://doi.org/10.3389/fpsyg.2023.1287667

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.