

Enhancing AI Literacy and Critical Thinking through Comparative Analysis of LLMs: An Action Research in a Bilingual Senior High School Classroom in Taiwan

粘靜容 Ching-Jung Nien
National Yilan Senior High School

Research Overview

This action research examines how structured comparative analysis of large language models (LLMs) fosters critical AI literacy and intercultural awareness among 27 twelfth-grade students in a bilingual experimental class in Taiwan. Facilitated by a designated bilingual program English teacher over five weeks, the study integrates language education, social science inquiry, and AI literacy within a single curriculum framework.

Purpose of Study

As AI tools become deeply embedded in students' academic routines, concerns have grown not merely about frequency of use, but about whether students possess the critical capacity to interrogate AI-generated content. Situated within Taiwan's Bilingual Nation 2030 initiative, this study pursues two primary objectives: (1) to cultivate students' critical AI literacy by developing their ability to evaluate, compare, and question the outputs of multiple LLM platforms; and (2) to foster intercultural awareness by engaging students in structured cross-cultural analysis of how AI tools developed in different linguistic and cultural contexts encode divergent values, assumptions, and rhetorical strategies. A secondary aim is to provide a replicable curriculum framework for bilingual educators seeking to integrate AI literacy into non-STEM secondary instruction.

Research Methods

This study adopts an action research design, drawing on pre- and post-intervention questionnaires, structured evaluation worksheets, and classroom discussions to enable triangulation across quantitative and qualitative data sources. The five-week intervention comprised ten class sessions organized as a progressive five-stage learning sequence. An initial survey mapped students' existing AI literacy, including functional and operational use, cross-cultural analysis and bias detection, human-AI collaboration, and practical application, establishing a baseline for post-intervention comparison. Students then received foundational AI literacy instruction using materials from the Taiwan AI Academy. They were required to reflect on their own usage habit and evaluate the risks of AI tools. In the third stage, a linguistic analysis activity required students to identify framing differences and connotational nuances

between English and Chinese versions of a questionnaire translated by AI from Chinese to English—an exercise carrying particular pedagogical significance in the bilingual classroom context, where navigating between languages is central to students' learning experience.

In the fourth stage, students independently formulated moral dilemma questions connected to their intended college majors—such as "As a lawyer, should I defend a client who has objectively committed a serious breach of ethics?" and "Should an accountant hide company debts to help secure a bank loan?"—and submitted them in English to two Western-based LLMs, such as Claude, Gemini, or ChatGPT. Responses were evaluated across eight criteria: organization, tone and style, language difficulty, visual design, usefulness, evidence and explanation, bias or perspective, and missing information, each was accompanied by suitable descriptors for selection and students were required to provide a qualitative written explanation. Data collection is ongoing. In the final stage, currently underway, students are reformulating the same questions in Chinese and submitting them to Chinese-based LLMs, primarily DeepSeek. Responses will be evaluated using four targeted criteria—clarity, depth, balance, and domain relevance and credibility—before students identify the most significant contrasts between platforms, select the tool they found most useful, and reflect on how they might purposefully integrate AI tools into their future academic learning. Post-intervention questionnaire data will be collected upon completion of this stage.

Results and Discussion

Preliminary findings, drawn from initial surveys, evaluation worksheets, and classroom discussions, reveal several significant patterns. Students identified concrete instances of Western-centric bias in English-based LLMs: one student observed that both Gemini and ChatGPT implicitly assumed the user resided in a democratic country, while another noted that ChatGPT recommended reporting concerns to a U.S. government agency—positioning the user within an American institutional framework irrelevant to their Taiwanese context. Multiple students raised similar observations independently, suggesting a systemic pattern rather than an isolated incident.

Students also articulated meaningful differences in reasoning styles across Western platforms, noting ChatGPT's tendency to guide users toward a conclusion, whereas Gemini tends to present theoretical perspectives using language that is too advanced for EFL high school students rather than providing a direct answer. At a deeper level, students demonstrated emergent epistemic awareness: one reflection captured the insight that "AI guesses the user's viewpoint and preferences," signaling a shift from passive consumption to active interrogation

of AI output. Post-intervention data, including cross-cultural comparison findings from the final stage, will be incorporated into the full paper to present a complete before-and-after account of students' developing critical AI literacy. Some students found AI responses overly detailed or linguistically demanding, pointing to the need for scaffolded language support in future bilingual implementations.

This study contributes a replicable, cross-disciplinary framework for integrating critical AI literacy into bilingual secondary classrooms. By positioning the bilingual classroom—where students routinely navigate between English and Chinese—as a uniquely productive site for cross-cultural LLM analysis, the study advances both AI literacy pedagogy and bilingual education practice. It further provides preliminary qualitative evidence that structured comparative analysis of culturally divergent LLMs can shift students from passive AI consumers to reflective, culturally aware critical thinkers capable of asking: whose assumptions does this AI reflect, and for whom was it designed?

Keywords

critical AI literacy, bilingual education, large language models, cross-cultural comparison, algorithmic bias