ACCSCIENCE PUBLISHING

International Journal of Systematic Innovation

REVIEW

Metacognitive cultural intelligence training in the age of generative artificial intelligence: A bibliometric review

Zheyun Zheng, Jiraporn Chano*, and Chowwalit Chookhampaeng

Department of Curriculum and Instruction, Faculty of Education, Mahasarakham University, Kantharawichai, Maha Sarakham, Thailand

Abstract

Generative artificial intelligence (GenAI) is increasingly used in culturally diverse workplaces and educational settings, yet its outputs may reproduce limited, stereotypical, or biased cultural assumptions. This review maps recent scholarship on metacognitive cultural intelligence (CQ), intercultural training, and GenAI-assisted learning and translates the available evidence into a training framework. A descriptive bibliometric review and critical integrative synthesis were conducted for publications dated from January 1, 2021, to July 27, 2026. The final bibliography comprised 58 bibliographically verified sources: 48 topical publications formed the analytical corpus and 10 methodological or indexing sources supported review reporting, evidence appraisal, deduplication, screening transparency, or source verification. Annual publication output in the retained topical corpus increased from one publication in 2021 to 17 in 2024, and 29 of the 48 publications (60.4%) appeared from 2024 onward. Only four publications (8.3%) directly addressed the focal GenAI–intercultural intersection. Accordingly, claims regarding specific training benefits should be interpreted as evidence-informed propositions derived primarily from adjacent CQ, AI-learning, metacognition, and cultural-bias literature rather than as established causal effects. The review proposes a five-component framework linking learner conditions, GenAI affordances, metacognitive regulation, developmental outcomes, and cross-cutting safeguards. It also specifies the evidential status of the proposed relationships, outlines training-design principles, provides an illustrative training scenario, and identifies priorities for testing culturally responsible human–AI learning systems.

Keywords: Cultural intelligence; Metacognition; Generative artificial intelligence; Intercultural training; Bibliometric review; Systematic innovation

***Corresponding author:**
Jiraporn Chano
(jiraporn.j@msu.ac.th)

Citation: Zheng, Z., Chano, J. & Chookhampaeng, C. (2026). Metacognitive cultural intelligence training in the age of generative artificial intelligence: A bibliometric review. *Int J Systematic Innovation*. https://doi.org/10.6977/IJoSI.2026XX_XX(X).026320265

Received: August 7, 2026

Revised: September 9, 2026

Accepted: October 10, 2026

Published online: October 26, 2026

Copyright: © 2026 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Intercultural communication is now commonplace in intercultural project work, international assignments, education, migration, remote work, and digitally mediated collaboration. In these contexts, language practices, socialization, and institutional norms shape every aspect of interpretation. Intercultural competence, then, is not just acquiring static knowledge of national cultures; it means tolerating ambiguity, interpreting provisionally, considering alternatives, suspending judgment, and revising

one's interpretation in light of new information. In recent cultural intelligence (CQ) studies, intercultural effectiveness is thus seen as a learned skill relevant to team performance, adaptation to intercultural environments, and international business (Davaei & Gunkel, 2024; Mammadov & Wald, 2025; Morin & Talbot, 2023).

This capability is underpinned by metacognitive CQ (self-regulation of cultural thinking before, during, and after the encounter). It involves predicting unfamiliar interactions, checking the efficacy of your predictions, noticing the assumptions you make, and changing your strategy as you receive new information. According to recent studies in the field of education and organization, the process of intercultural development is not a natural result of an encounter, but rather depends on reflection, previous experience, involvement, personality, and the quality of the encounter (Chedru & Ostapchuk, 2023; Iskhakova *et al*., 2022; McKay *et al*., 2022; Morin & Talbot, 2024).

Intercultural training is increasingly dynamic, including not only disseminating cultural information but also active involvement, reflection, feedback, and transfer to new cultural contexts (Goldstein, 2022; Iskhakova & Bradly, 2022; Urgun *et al*., 2025). This study's findings reinforce previous situated-learning studies showing that CQ is learned not only through teaching but also through relationships, routines, tensions, and unexpected experiences (Teixeira & Klein, 2024). From a systematic-innovation perspective, training can be seen as continuous rather than standalone.

Generative artificial intelligence (GenAI) offers a versatile space for practice in this learning cycle. These systems can be used to produce cases, simulate conversations, translate, and offer instant feedback, with varying tones or politeness. These functions relate to personalization, content generation, explanation, formative feedback, and learner support in education, as described by Batista *et al.* (2024), Crompton and Burke (2023), Kasneci *et al.* (2023), and S. Wang *et al.* (2024). Meanwhile, issues of academic integrity, misinformation, dependency, access, governance, and others continue to be a concern (Chan, 2023; Cotton *et al*., 2024; Farrokhnia *et al*., 2024; Yusuf *et al*., 2024).

These can support intercultural training when scenario variety and practice time are limited. A GenAI system can modulate the same intercultural conflict by changing who is speaking, who has power, the language, and the context. Previous research has examined GenAI in serious games (Agarwala *et al*., 2024), professional communication (Dai *et al*., 2025), foreign-language learning (Y.M. Lee, 2025), and intercultural competence (H. Wang & Gao, 2025). However, these four direct publications make up only a small share of the retained publications, so the specific effects of intercultural training with GenAI are not yet established.

The same technology can also be detrimental to cultural learning. Model output can oversimplify culture as a set of rules, fail to account for cultural differences, or offer general recommendations with a false sense of security. Previous studies in cultural alignment and benchmarking indicate that language models do not always follow the same behavior in different cultures and might exhibit problems in understanding the implicit values or reasoning under different contexts (Kim *et al*., 2024; Liu *et al*., 2025; Naous *et al*., 2024; Tao *et al*., 2024). Other broader criticisms include data imbalance, representational harm, and insufficient foundations and justification (Bender *et al*., 2021; Liu, 2024; Nyaaba *et al*., 2026).

Another factor of concern is cognitive substitution. GenAI can support reasoning, but it can also reduce learner effort. Evidence suggests that using AI-generated material may foster less critical engagement, while intentional verification and reflection can encourage critical thinking (H.P. Lee *et al*., 2025; Xiao *et al*., 2025). Research on self-regulated learning thus suggests that planning, monitoring, and self-evaluation remain under the learner's control rather than being taken for granted and transferred to the system (Anders & Dux Speltz, 2025; Xia *et al*., 2026). Not only should the question be whether AI should be used, but whether any of the reasoning steps might be supported by AI, and whether any part of the reasoning is to be left to the learner.

The literature lacks coherence on several related topics. Reviews of team intelligence and expatriate CQ primarily focused on academic studies within their respective fields (Davaei & Gunkel, 2024; Mammadov & Wald, 2025). The educational CQ reviews were more comprehensive, incorporating both academic papers and practice (Morin & Talbot, 2023), while the international business application reviews were more recent and centered on academic papers in the field (Soejanto *et al*., 2024). Reviews on AI in education focused more on its use in education, with implementation, teaching practice, ethics, and policy featured prominently (Lo, 2023; Mustafa *et al*., 2024; Tlili *et al*., 2023). Much remains to be explored in the specific dialogue between GenAI and intercultural learning. Therefore, a focused bibliometric review helps distinguish the small body of direct GenAI–intercultural evidence from adjacent literature on CQ, GenAI–learning, metacognition, and cultural bias, and avoids treating them as the same.

The review answers four research questions. The first

research question (RQ1) concerns the growth of publication activity over time in the retained corpus. RQ2 identifies the main evidence streams in the literature and distinguishes between direct and adjacent evidence. RQ3 synthesizes the training mechanisms, risks, and safeguards shared across these streams. RQ4 focuses on how the evidence in the review can be applied to a systematic, testable framework for GenAI-supported metacognitive CQ training. Table 1 illustrates how each research question is mapped onto an analytical procedure and what is intended by the research question.

This review makes four contributions. First, it describes the field without relying on volatile live citation counts and reconstructed database totals. Second, it separates four evidence-base publications that provide direct evidence on GenAI–intercultural evidence from the broader evidence base on related CQ development and training, GenAI learning and metacognition, and cultural bias and governance. Third, it translates these evidence bases into a five-component training architecture and identifies relationships that are explicit, adjacent, or proposed for testing. Fourth, it sees GenAI as a system to systematically vary, provide feedback on, and reflect on, rather than as a megaculturally powerful tool that tells right from wrong. Fourth, it sees GenAI not as a megaculturally powerful tool that tells right from wrong, but as a system to systematically vary, provide feedback on, and reflect on.

2. Data and methods

2.1. Review design and scope

This study used a focused descriptive bibliometric review combined with a critical integrative synthesis. The unit of analysis was the bibliographic publication rather than individual participant data. This design was selected because research specifically addressing metacognitive CQ training and GenAI remains recent and narrow, while adjacent literature is necessary to interpret the emerging direct evidence. The bibliometric component is descriptive: it reports publication-year distributions and organized evidence-stream counts rather than claiming a complete co-citation, co-word, collaboration, or bibliographic-coupling network analysis.

Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 and PRISMA-S (Page *et al*., 2021; Rethlefsen *et al*., 2021) were used to guide the reporting structure, and the work by Donthu *et al.* (2021) was used to guide the bibliometric decisions. OpenAlex was used as an open discovery index and metadata check for scholarly sources (Priem *et al*., 2022). Other review process control items were based on recent guidance regarding transparent screening, deduplication, evidence synthesis, and evidence review auditability (Aromataris *et al*., 2024; Haddaway *et al*., 2022; McGuinness & Higgins, 2021; McKeown & Mir, 2021; Paul *et al*., 2021; van de Schoot *et al*., 2021). The publication eligibility cutoff was retained at July 27, 2026, and the retained-corpus audit was frozen on August 20, 2026. This latter audit verified the integrity of the retained bibliography with respect to the composition, uniqueness, coding, and evidence appraisal. It was not intended to create historical search-stage denominators that were not preserved.

The publication window was specified to capture the contemporary overlap among CQ, metacognitive learning, and rapidly developing GenAI applications. The 2021 lower boundary provides a recent pre-ChatGPT baseline and covers the subsequent period in which GenAI became

Table 1. Research questions, analytical procedures, and intended contributions

Research question	Analytical procedure	Intended contribution
RQ1. How has recent scholarship developed over time?	Annual publication counts and period comparison	Shows the pace and recency of field development
RQ2. Which primary evidence streams organize the retained topical corpus, and how is direct GenAI–intercultural evidence distinguished from adjacent evidence?	Mutually exclusive primary-stream coding with non-exclusive secondary thematic tags	Clarifies stream boundaries, prevents double counting, and separates direct from adjacent evidence
RQ3. Which mechanisms and safeguards are supported?	Cross-stream critical synthesis	Identifies defensible training functions, risks, and controls
RQ4. How can the evidence inform systematic innovation?	Framework construction and research-agenda development	Translates the review into an implementable and testable training architecture

Abbreviations: CQ: Cultural intelligence; GenAI: Generative artificial intelligence; RQ: Research question.

increasingly relevant to education, professional learning, communication, and intercultural research. This recency-oriented window better reflects current technological, pedagogical, and governance conditions but does not capture the longer historical development of CQ or metacognition.

This recency-oriented window is intentionally selective rather than historically comprehensive. Seminal pre-2021 work on CQ, intercultural learning, and metacognition was therefore excluded from the annual frequencies and evidence-stream proportions reported here. The bibliometric results should consequently be interpreted as a profile of recent GenAI-era scholarship relevant to CQ and metacognition, not as a complete historical account of either field.

2.2. Search architecture and source verification

Discovery was conducted using OpenAlex, Crossref/DOI metadata, ACL Anthology, arXiv, discipline-specific repositories, publisher webpages, and backward and forward citation checking. Where supported by the interface, searches were conducted across titles, abstracts, and keywords. Three conceptual query families guided discovery:

(i) Query family 1: ("cultural intelligence" OR "metacognitive cultural intelligence" OR "cultural metacognition" OR "intercultural competence") AND (training OR learning OR education OR development OR simulation OR reflection).
(ii) Query family 2: ("cultural intelligence" OR "metacognitive cultural intelligence" OR "cultural metacognition" OR "intercultural competence") AND ("generative AI" OR ChatGPT OR "large language model*" OR "AI-assisted learning").
(iii) Query family 3: ("generative AI" OR ChatGPT OR "large language model*" OR "AI-assisted learning") AND (culture OR cultural OR intercultural) AND (bias OR alignment OR competence OR training OR learning).

These query families represent the conceptual search logic rather than a claim that an identical executable syntax was used across all platforms. Because database interfaces differed in field structure, phrase searching, wildcard support, and Boolean syntax, the queries were adapted to the requirements of each source while preserving the same conceptual blocks. Table S1 records the query families, discovery sources, eligibility criteria, and available screening-audit information.

Retained records were checked against a publisher page, DOI record, institutional repository, ACL Anthology record, or arXiv record. Bibliographic validation covered the author, year, title, outlet, volume or article number (where available), and DOI or another stable identifier. Search-result snippets were not treated as evidence. Where multiple versions of the same work were available, duplicate versions were consolidated, and the most complete scholarly version was retained. For recent technical contributions without an identifiable peer-reviewed version at the time of review, the available preprint was retained.

The frozen bibliography comprises 58 bibliographically verified sources: 48 topical publications and 10 methodological or indexing sources. Bibliographic verification establishes source identity and metadata; it does not constitute a methodological quality appraisal. A retained-bibliography deduplication audit based on DOI matching, normalized-title matching, and first-author/year inspection found no duplicate DOI/title records among the 58 frozen records (McKeown & Mir, 2021). This zero-duplicate value reflects quality control for the retained bibliography and should not be interpreted as the number of duplicates removed during the original search. Because the initial iterative discovery log did not consistently preserve retrieval, duplicate-removal, title/abstract-screening, or full-text denominators, those historical values cannot be reconstructed. Figure 1 therefore distinguishes unavailable historical screening stages from the auditable frozen retained-corpus counts.

Table 2 summarizes the search architecture, eligibility criteria, source-verification procedure, and analytical denominator.

2.3. Coding and quality control

Each of the 48 topical publications was categorized into one of the four mutually exclusive primary evidence streams: (i) CQ development and training; (ii) GenAI learning and metacognition; (iii) cultural bias and governance; and (iv) direct GenAI–intercultural applications. For this review, the publication's main substantive contribution is an evidence stream, not each theme it covers. Thus, the four primary-stream counts sum to 48, and no publication was counted more than once in the primary-stream percentages. Content that was cross-cutting was also collected separately by non-exclusive secondary thematic tags for learning setting, AI role, metacognitive mechanism, outcome, and risk. Secondary tags were used only in the integrative synthesis, not in calculating frequency in the main stream. The first three streams were treated as adjacent evidence for developing the framework, while the fourth was treated as direct evidence on the focal GenAI–intercultural intersection. Primary-stream assignment was made on the basis of a three-tiered hierarchy that was previously determined: (i) the main RQ

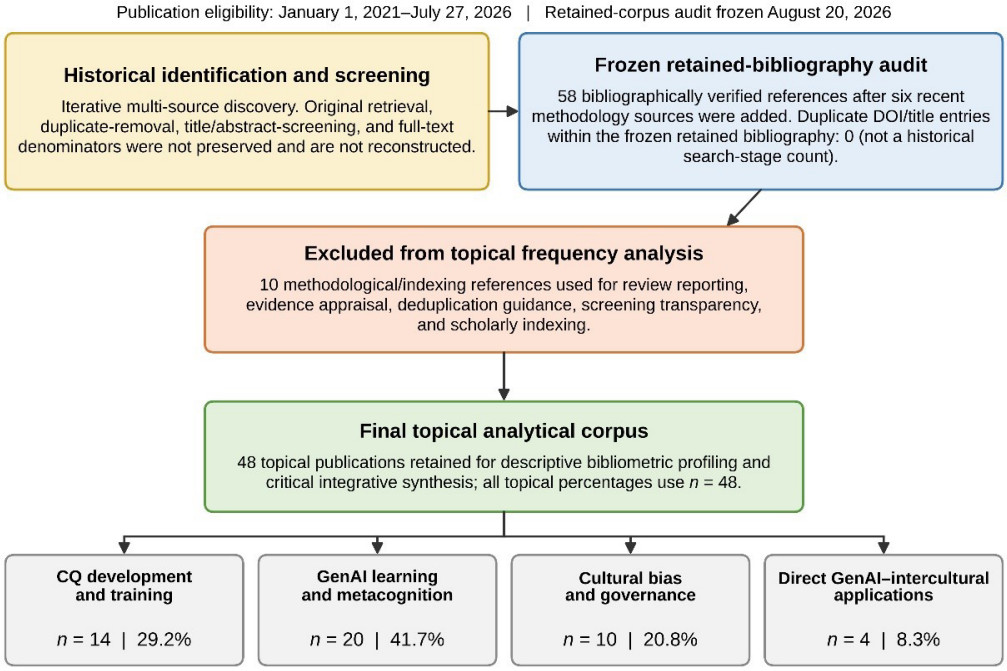


Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses-informed retained-corpus audit. Historical retrieval, duplicate removal, title/abstract screening, and full-text denominators were not preserved and cannot be reconstructed. The frozen retained bibliography contains 58 unique bibliographic records, of which 10 methodological/indexing references were excluded from topical frequencies and 48 topical publications form the analytical corpus. The zero-duplicate value refers only to the frozen retained bibliography, not to the original search-stage duplicate-removal count.
Abbreviations: CQ: Cultural intelligence; GenAI: Generative artificial intelligence.

Table 2. Search, eligibility, and verification protocol

Component	Operational decision	Reason for the decision
Review date and period	Search finalized on July 27, 2026; publication window January 1, 2021–July 27, 2026	Defines the exact review interval: approximately five years and seven months spanning six publication years
Cultural-capability block	Cultural intelligence, metacognitive CQ, cultural metacognition, intercultural competence	Captures the focal capability and closely related training outcomes
Development block	Training, learning, education, development, simulation, reflection	Identifies studies with a developmental mechanism
Technology block	GenAI, ChatGPT, large language models, AI-assisted learning	Captures direct and adjacent technological evidence
Eligible outputs	Peer-reviewed articles and reviews; selected conference papers, guidance, and recent preprints	Balances scholarly quality with coverage of a rapidly changing field
Exclusions	Unrelated uses of culture or intelligence, non-scholarly commentary, unverifiable metadata, duplicate versions	Protects topical relevance and traceability
Verification	Publisher/DOI page, ACL Anthology, institutional repository, arXiv, or other traceable scholarly record	Prevents reliance on unverified search snippets
Analytical denominator	48 topical records; 10 methodological/indexing sources excluded from topical frequencies	Keeps methods, search-reporting, deduplication, evidence-appraisal, and indexing references from distorting topical results

Abbreviations: CQ: Cultural intelligence; GenAI: Generative artificial intelligence.

or objective of the publication, (ii) the main outcome or object of analysis of the publication, and (iii) the principal substantive contribution of the publication.

Only publications that fulfilled two requirements were included in the direct GenAI–intercultural applications stream: (i) GenAI was a central component of the intervention, explanatory framework, or system studied; and (ii) the primary focus or outcome of the intervention was intercultural communication, cultural competence, CQ, or a closely related intercultural construct. For example, the direct stream was used when AI was the focus and intercultural communication competence was the focal outcome. In contrast, Tao *et al*. (2024) was categorized under cultural bias and governance because the study's main contribution focused on the cultural alignment of language models rather than intercultural learner development.

A second substantive coding pass was conducted on the entire text, using the same hierarchy for the coding, and focusing on borderline records. An automated sensitivity analysis based on rules applied to the metadata available was then used to generate 47 of the 48 mainstream assignments, or 97.9%. One discordant case was resolved using the pre-specified hierarchy. This check was not an independent second human coding check, so it is viewed as a classification-stability check, and no Cohen's κ statistic is reported. The coding of the records and the comparison process is documented in the frozen audit workbook.

The process of bibliographic verification and evidence appraisal was separated. The validity of the sources and metadata was checked by a DOI, a publisher page or a repository record, rather than methodological quality. Due to the variation of study designs within the topical corpus, a design-aware, more nuanced, evidential-weight appraisal was used instead of one numerical quality score. Records were classified into higher, moderate, and contextual/emerging evidential weight based on peer-review status, study design, relevance to the review question, and claim supported. Of the 48 topical records, 13 (27.1%) were classified as higher, 15 (31.3%) as moderate, and 20 (41.7%) as contextual/emerging. These categories were used to define the interpretation and are not intended to be interpreted as a common numerical quality scale. Judgements and limitations for records are provided in Table S2.

2.4. Descriptive analysis and synthesis

Annual publication counts were calculated directly from the 48 topical records, and 48 is the denominator for all reported topical percentages. These frequencies describe the retained corpus and should not be interpreted as prevalence estimates for the entire scholarly literature. To strengthen the descriptive bibliometric component without claiming a complete citation-network analysis, a retained-title term co-occurrence analysis was also conducted. All 48 titles were processed using a prespecified concept dictionary; normalized concepts appearing in at least two titles were retained, and an edge was defined when two concepts occurred in the same title. This procedure provides a reproducible check of the conceptual structure of the frozen corpus. It is explicitly not an author-keyword co-word analysis, co-citation analysis, collaboration analysis, or bibliographic-coupling analysis. The frozen dictionary, frequencies, and edges are provided in the audit workbook and Table S4; the corresponding map is shown in Figure S1. The thematic synthesis compared claims within and across evidence streams, with particular attention to convergence, tension, and evidence gaps. Importantly, the review distinguished an AI affordance from a demonstrated learning effect: the capacity of a system to generate dialogue or feedback does not, by itself, demonstrate improvement in metacognitive CQ.

An evidence-to-design process was used to develop the framework. Recurring learner conditions, AI functions, metacognitive processes, developmental outcomes, and safeguards were organized into five components. Three evidential levels were used when interpreting the framework: direct evidence from the four focal GenAI–intercultural publications; adjacent evidence from CQ, AI-learning, metacognition, and cultural-bias studies; and proposed relationships that remain to be tested. Accordingly, arrows in the framework represent theoretically plausible directional relationships rather than demonstrated causal effects. No inferential test of publication growth or causal training effect was conducted.

3. Results

3.1. Publication development and corpus profile

The topical corpus contained 48 publications: one from 2021, four from 2022, 14 from 2023, 17 from 2024, 10 from 2025, and two from the partial 2026 period. Overall, 29 publications (60.4%) appeared from 2024 onward, while 43 (89.6%) appeared from 2023 onward. Publication counts increased descriptively from four in 2022 to 14 in 2023 and 17 in 2024 (Figure 2). Because no inferential temporal-trend test was conducted, this pattern is described as an increase in the retained corpus rather than as a statistically significant trend. The 2026 count is partial and includes publications available through July 27.

The annual pattern largely reflects the recent expansion of GenAI-related scholarship rather than a decline in traditional CQ research. The 2021 count is low by design

because the review uses a recent publication window. Records from 2023 and 2024 span study-abroad and intercultural-development research, CQ mapping, cultural-bias benchmarks, direct applications, and GenAI studies in education; the 2025 and partial 2026 records place greater emphasis on metacognition, professional communication, and AI-supported assessment of cultural competence.

The corpus includes journal articles, reviews, conference papers, selected technical preprints, and policy or guidance documents. Peer-reviewed sources support most training and education claims, whereas preprints are concentrated in rapidly changing technical areas. Preprints are therefore used mainly to identify emerging directions and feasibility, not to establish training effectiveness.

Table 3 presents the descriptive profile of the retained topical corpus.

3.2. Evidence-stream structure

The primary evidence stream with the most publications was GenAI learning and metacognition (20/48; 41.7%), followed by CQ development and training (14/48; 29.2%), cultural bias and governance (10/48; 20.8%), and, finally, direct GenAI–intercultural applications (4/48; 8.3%). The four main categories are mutually exclusive and sum to the full analytical body of 48 publications. These percentages are all based on $n = 48$. Secondary thematic tags were not exclusive; they were applied only to show cross-stream relevance and thus do not appear in the primary-stream counts in Figure 3. The first three streams provide evidence neighboring the proposed framework, while the fourth offers publications directly related to the focal intersection of GenAI and intercultural.

No distribution was found, suggesting an emerging intersection rather than a stand-alone field. While direct evidence is still limited, the four streams highlighted complement each other: CQ research sets the developmental goal; GenAI learning research and metacognition research help identify the learning functions and cognitive risks; cultural bias and governance research focus on the reliability and cultural validity of the model-mediated environment; and the direct stream focuses on how these elements are beginning to converge. The key findings and questions are summarized in each of the four streams in Table 4.

There is limited consistency in outcome evidence within streams. Attitudes, system performance, and perceived usefulness were common themes in GenAI studies, while competence, adaptation, learning, and performance were common themes in intercultural studies. There were relatively few publications that used CQ-related measures that were validated and observed behavior, interaction traces, or delayed transfer. It is therefore more plausible

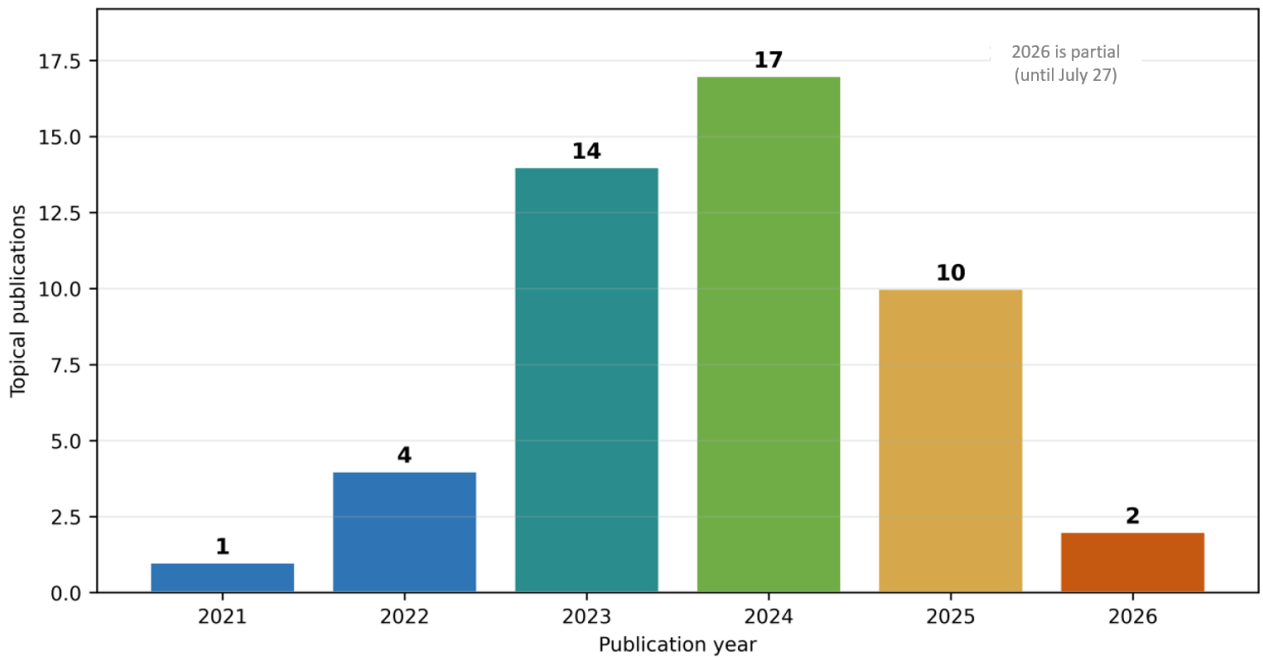


Figure 2. Annual publication distribution of the 48-record topical corpus. The 2026 value includes publications available through July 27.

Table 3. Descriptive profile of the topical corpus.

Indicator	Verified result	Interpretation
Topical publications	48	All frequency results use this denominator
Method/indexing references	10	Used for reporting, bibliometric design, search transparency, deduplication, evidence appraisal, and indexing; not included in topical counts
Publication period	2021–July 27, 2026	Review interval from January 1, 2021 to July 27, 2026
Publications from 2024 onward	29 (60.4%)	The corpus is strongly weighted toward current work
Publications from 2023 onward	43 (89.6%)	Generative artificial intelligence scholarship drives recent expansion
Largest annual count	17 in 2024	Peak year for combined educational and cultural-bias research
Direct generative artificial intelligence–intercultural studies	4 (8.3%)	The exact intersection remains small
Use of live citation counts	None	Avoids age bias and unstable metrics

that the proposed mechanisms explain the findings than durable training effects.

This is corroborated by the design-aware evidence appraisal. Of the 48 topical publications, 13 received higher evidential weight, 15 received moderate weight, and 20 received contextual/emerging weight. Conceptual articles, policy or guidance sources, prototypes, and preprints were more prevalent in the contextual/emerging category, while systematic syntheses were more prevalent in the higher category. These categories were used only to standardize interpretation across designs that vary in the number of categories (Table S2).

The concept map retained is another descriptive bibliometric check. Normalized concepts were most prevalent for education (29 titles), GenAI (23), and CQ (15) across all 48 titles. The highest number of same-title co-occurrences were in the education–GenAI (15), CQ–education (6), education–higher education (5), CQ–study abroad (4), education–governance/ethics (4), and cultural bias–GenAI (4) categories. The relationships are presented in Figure S1, and the frequencies of the strongest co-occurrences are provided in Table S4. The map shows the conceptual structure at the title level and should not be used to interpret a citation or collaboration network.

3.3. Cultural intelligence development as the learning target

The CQ development literature provides strong evidence on which behaviors to develop and how to develop them. Research on study abroad (SAB) shows that successful SAB experiences depend on program design, prior international preparation, cultural distance, reflective practice, and learner engagement (Goldstein, 2022; Iskhakova & Bradly, 2022; Iskhakova *et al*., 2022). Literature and longitudinal evidence also indicate that international experience can be used as a support for the development of CQ, but that the trajectory of change is not consistent between learners and contexts (Chedru & Ostapchuk, 2023; McKay *et al*., 2022; Morin & Talbot, 2024).

Complementary research approaches to intercultural development, such as situated-learning research, can be used alongside classroom-based approaches. CQ can emerge in everyday social interaction, group activities, unexpected events, and the identification of tensions (Teixeira & Klein, 2024). Rather than presenting training with one response as the "culturally correct" response, one should try to avoid doing so. A better design encourages students to recognize knowledge gaps, make tentative interpretations, identify contextual evidence, and revise interpretations as new evidence emerges.

The recent literature that has been related to CQ has included the aspects of team processes and the adaptation of expatriates (Davaei & Gunkel, 2024), the international activities of firms (Mammadov & Wald, 2025), the well-being of professionals working with migrants (Puzzo *et al*., 2024), and cultural competence in nursing education (Ahmadi Forg *et al*., 2025). These applications are not limited to international mobility; they also support a broader understanding of training outcomes and the principles of strategic adaptation, cultural humility, respect, and power awareness.

Mapping and training reviews conducted in educational

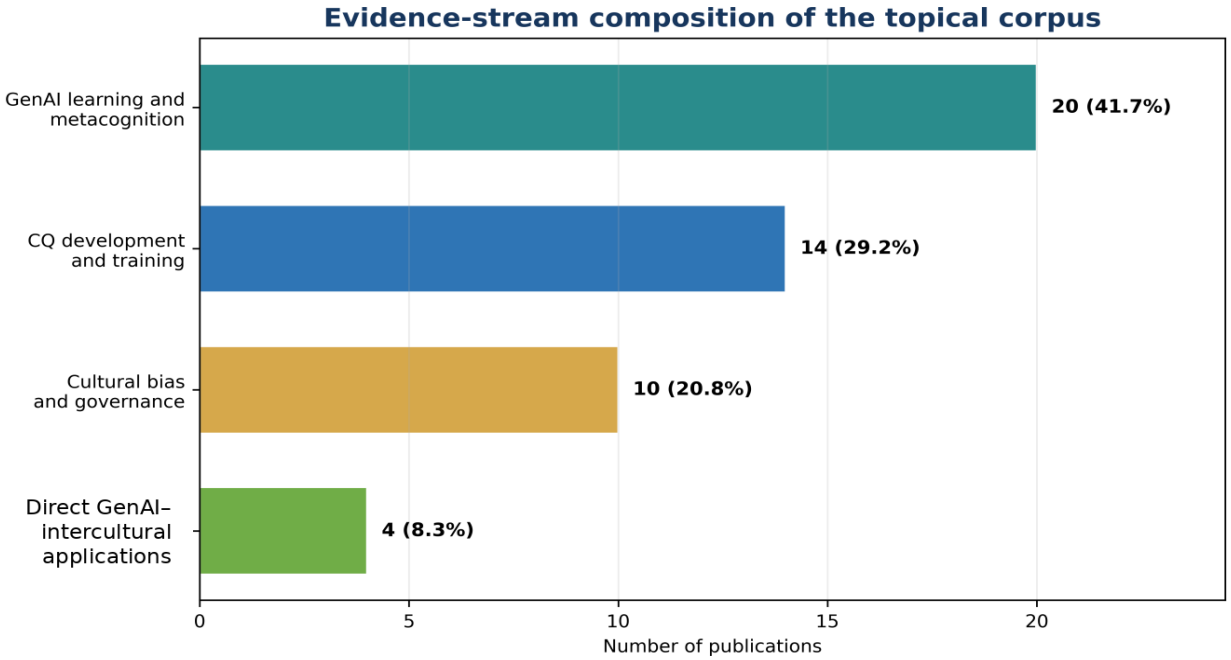


Figure 3. Evidence-stream composition of the 48-record topical corpus. Each publication was assigned to one primary stream for descriptive analysis. Abbreviations: CQ: Cultural intelligence; GenAI: Generative artificial intelligence.

Table 4. Evidence streams, central findings, and unresolved questions

Evidence stream	Recent evidence	Unresolved question
CQ development and training	Experience is more developmental when it includes reflection, feedback, engagement, and transfer	Which training activities produce durable change in metacognitive CQ rather than short-term self-report gains?
GenAI learning and metacognition	AI can personalize dialogue and feedback, but may reduce independent effort or create overtrust	What timing and level of assistance preserve self-regulation and critical thinking?
Cultural bias and governance	Model outputs can privilege dominant cultural values and miss context-sensitive or implicit norms	How should cultural validity, uncertainty, and within-culture variation be audited?
Direct GenAI–intercultural applications	Early systems support scenarios, role-play, feedback, and intercultural communication practice	Do these systems improve behavioral transfer, cultural humility, and ethical judgment?

Note. Each topical publication contributes to one primary evidence stream only. Secondary thematic tags may occur across streams and are not included in stream-frequency counts. The first three streams are treated as adjacent evidence; only direct GenAI–intercultural applications are treated as focal direct evidence.
Abbreviations: AI: Artificial intelligence; CQ: Cultural intelligence; GenAI: Generative artificial intelligence.

settings vary widely in approach, population, and setting (Soejanto *et al.*, 2024; Urgun *et al.*, 2025). This diversity expands the evidence base, but makes measurement more difficult. Observed behaviors, interaction quality, transfers to novel cases, and maintenance over time are less frequently measured. Process traces that could be generated within GenAI environments are more comprehensive; however, they are meant to be analyzed for reasoning and revision and not linguistic fluency.

3.4. Generative artificial intelligence as a learning environment

Adjacent GenAI education research identifies functions relevant to CQ training, including personalized

explanation, rapid feedback, content generation, dialogue, and learner support (Batista *et al.*, 2024; Kasneci *et al.*, 2023; Lo, 2023; Mustafa *et al.*, 2024; S. Wang *et al.*, 2024). Learner-perception studies also report concerns about accuracy, dependency, equity, and assessment (Chan & Hu, 2023; Yusuf *et al.*, 2024). These findings identify candidate affordances and risks; they do not directly demonstrate that GenAI improves metacognitive CQ.

Drawing on adjacent AI-learning and self-regulation research, the functions most relevant to metacognitive CQ are likely non-prescriptive. A system can ask the learner to predict an interlocutor's intended meaning, explain the evidence supporting that prediction, generate alternative explanations, and identify what additional information would reduce uncertainty. It can then reveal further context and require the learner to revise the interpretation. This sequence uses generation to structure inquiry rather than to provide a supposedly correct cultural answer.

This design is consistent with self-regulated learning research. Anders and Dux Speltz (2025) conceptualize GenAI literacy in relation to self-perception and self-regulation. Xia *et al.* (2026) similarly show how GenAI can support forethought, performance, and reflection while poorly designed scaffolds may displace these processes. Xiao *et al.* (2025) illustrate how learners can become more aware of metacognitive processes related to AI use. Together, these adjacent studies support including AI literacy and reflective monitoring in CQ training design, but they do not establish a causal effect on metacognitive CQ.

An important risk is also cognitive substitution. In their study, H.P. Lee *et al.* (2025) reported that higher trust in GenAI-generated answers is related to lower self-reported critical effort, highlighting the importance of maintaining learners' reasoning. Educational opportunities and disruptions due to GenAI have been discussed by Farrokhnia *et al.* (2024), Grassini (2023), Lim *et al.* (2023), and Rahman and Watanobe (2023); and research into academic integrity indicates that AI-generated output may be easy, but not necessarily learning (Cotton *et al.*, 2024). Therefore, the ability to respond fluently cannot be interpreted as a sign of a learner's awareness of cultural ambiguity or a change in an underlying assumption.

Governance scholarship also refutes unchecked adoption and blanket rejection. Dwivedi *et al.* (2023) view GenAI governance as a multi-sectoral challenge, and Tlili *et al.* (2023) see the technology as both an ally and a threat. Designers should therefore be able to articulate which aspects of human capability they want AI to support, which aspects of reasoning they want the learner to own, and how they will be evaluated without the model's help.

3.5. Direct generative artificial intelligence and intercultural applications

Only four publications met the criteria for the direct GenAI–intercultural applications stream. Agarwala *et al.* (2024) describe CultureVo, a serious-game environment for cultural learning and interaction using GenAI. Dai *et al.* (2025) examine GenAI-assisted professional communication training in intercultural contexts. Y.M. Lee (2025) discusses GenAI and intercultural competence in foreign language education. H. Wang and Gao (2025) report associations among AI utilization, enjoyment, trust, instructor competence, and intercultural communication competence among Chinese students.

These publications provide the most directly relevant evidence, but the inferences they support depend on study design. Design and prototype studies can establish technical feasibility; conceptual or review papers can identify pedagogical conditions; and cross-sectional models can estimate associations. None of these designs alone establishes sustained improvement in metacognitive CQ. Stronger tests would include pretest-posttest designs, observed interaction, process traces, delayed transfer, and comparisons among AI-only, human-facilitated AI, and conventional-training conditions.

The proposed framework should therefore be interpreted as an evidence-informed theoretical extension rather than as a training model whose effectiveness has already been established. The framework includes scenario diversity, source grounding, uncertainty signaling, and within-culture variation as design requirements because they address risks documented in adjacent evidence; their effects on learning outcomes still require direct testing.

3.6. Cultural bias, model evaluation, and governance

Generative AI should not be treated as a neutral source of cultural knowledge. Tao *et al.* (2024) identify differences between language-model outputs and cultural value profiles. Naous *et al.* (2024) show how culturally specific prompts can expose model bias, while Kim *et al.* (2024) demonstrate the importance of language- and culture-specific benchmarking. Liu (2024) synthesizes sources of cultural bias and mitigation strategies. Collectively, these studies indicate that factual recall or English-language fluency alone is insufficient evidence of a model's cultural reliability.

Cultural intelligence–bench extends assessment to implicit cultural values and contextual interpretation (Liu *et al.*, 2025), and Dev *et al.* (2026) propose a broader framework for evaluating the CQ of AI systems. These studies are relevant to training because an AI-assisted

learning environment contains two objects of evaluation: the human learner and the model-mediated environment. Claims about learner progress are difficult to interpret when the system generating scenarios and feedback is itself poorly calibrated to cultural context.

Governance sources add further requirements for responsible use. UNESCO (2023) emphasizes human-centered design, inclusion, privacy, and institutional accountability. Nguyen *et al.* (2023) review ethical principles for AI in education, including transparency, privacy, accountability, and human oversight. Nyaaba *et al.* (2026) warn that dominant languages and knowledge systems can be reproduced as universal, reinforcing forms of digital neocolonialism. Bender *et al.* (2021) similarly highlight how scale can amplify data and representation problems.

These findings support a cautious design approach to intercultural training. Systems should communicate uncertainty, avoid presenting generated personas as representatives of an entire culture, document the provenance of scenarios, incorporate local review, and represent variation within cultural settings. Learners should also be encouraged to challenge model outputs. When disagreement with the system requires evidence gathering, comparison, and explicit revision, even model error can become an opportunity for metacognitive learning.

4. Discussion

4.1. From cultural information to regulated cultural inquiry

The synthesis shows that more available cultural information does not necessarily improve intercultural learning with GenAI. More defensible, perhaps, is its ability to manipulate contexts and to encourage multiple questioning. The same problematic issue can arise across different power dynamics, social roles, language choices, time pressures, and/or prior experiences. These differences can reveal what may have been taken for granted, yet they are not necessarily part of cultural reality.

Metacognitive regulation is thus the central process of the framework. Learners should plan and predict, monitor interpretations, compare evidence and revise judgments. Some of these can be facilitated by using GenAI to pose questions, generate alternatives, or uncover additional information, but cannot be entirely completed by GenAI. This interpretation is partially supported by CQ research on reflective experiences and by self-regulated learning research on learner autonomy, and a few studies involving the sequence of improving metacognitive CQ are directly supportive.

This position also rejects cultural essentialism. A fixed cultural profile treats group membership as sufficient to explain behavior, whereas a metacognitive approach treats cultural knowledge as a hypothesis to be tested against the person, relationship, and context. GenAI can support this approach by generating within-culture and non-cultural variation, but it can also reinforce stereotypes when it produces confident generalizations. Its pedagogical value therefore depends heavily on the safeguards built around its use.

4.2. A systematic innovation framework for training design

Based on the reviewed literature, Figure 4 presents a five-component framework for GenAI-assisted metacognitive CQ training. The framework is evidence-informed, not a confirmed causal model. Four publications provide direct evidence at the focal GenAI–intercultural intersection, while adjacent CQ, GenAI-learning, metacognition, and cultural-bias research informs the remaining design relationships. Layers 1–4 describe the proposed learning pathway, and safeguards and governance operate across the full pathway rather than as a final step.

Layer 1 represents learner conditions. Baseline CQ, motivation, language repertoire, AI literacy, bias awareness, and prior intercultural experience may shape training needs and susceptibility to error. Evidence for prior experience and learner characteristics comes mainly from CQ-development research, whereas AI literacy and overtrust are supported by adjacent GenAI research.

Layer 2 includes GenAI affordances such as adaptive cases, role-play, multilingual rehearsal, perspective variation, feedback, and process analytics. These affordances are technically feasible and are supported by adjacent AI-learning and direct design studies, but their effects on sustained metacognitive CQ have not been established.

Layer 3 represents metacognitive regulation: planning, prediction, monitoring, assumption checking, evidence comparison, strategy revision, and transfer. Adjacent self-regulated learning and AI-metacognition research supports these processes as plausible learning mechanisms. The proposed causal link between AI-supported regulation and improved metacognitive CQ remains to be tested directly.

Layer 4 comprises developmental outcomes. The primary outcome of interest is metacognitive CQ, alongside cultural humility, adaptive communication, calibrated trust, and ethical judgment. CQ research

supports adaptation and competence as outcomes, but sustained change after GenAI-supported CQ training remains unestablished.

Layer 5 is cross-cutting rather than sequential. Human facilitation, source grounding, within-culture variation, privacy protection, uncertainty signaling, and auditability should operate throughout scenario generation, interaction, feedback, and assessment. Cultural-bias and governance research supports the need for these safeguards, whereas the effects of particular safeguards on learning outcomes remain empirical questions.

Table 5 maps each framework component to the reviewed evidence and distinguishes directly supported findings from adjacent evidence and proposed relationships.

4.3. Worked example: managing ambiguous feedback in a global virtual team

Imagine that a learner is attending a virtual project meeting that is being conducted in a different country. A colleague responds, "Let's reconsider this." The learner writes an initial interpretation, a confidence level, and the evidence supporting the judgment before reviewing AI-generated suggestions. This step makes the learner's confidence and starting assumptions explicit before AI support is introduced. GenAI then serves as a scripted scenario generator, not a cultural arbiter. The system invites the learner to develop two or more interpretations, at least one of which is non-cultural. Then the context changes, for instance by altering the relative rank, the project's time frame, previous conflicts, or the method of communication. The functions organize inquiry, without suggesting that the model has a "right" answer that is culturally correct.

Layer 3 is stated explicitly by metacognitive prompts: "What did I originally think?", "What evidence would support that interpretation?", "What alternative interpretation seems plausible in the new circumstances?", and "What additional information would I need before responding?" Learner revises interpretation and proposed

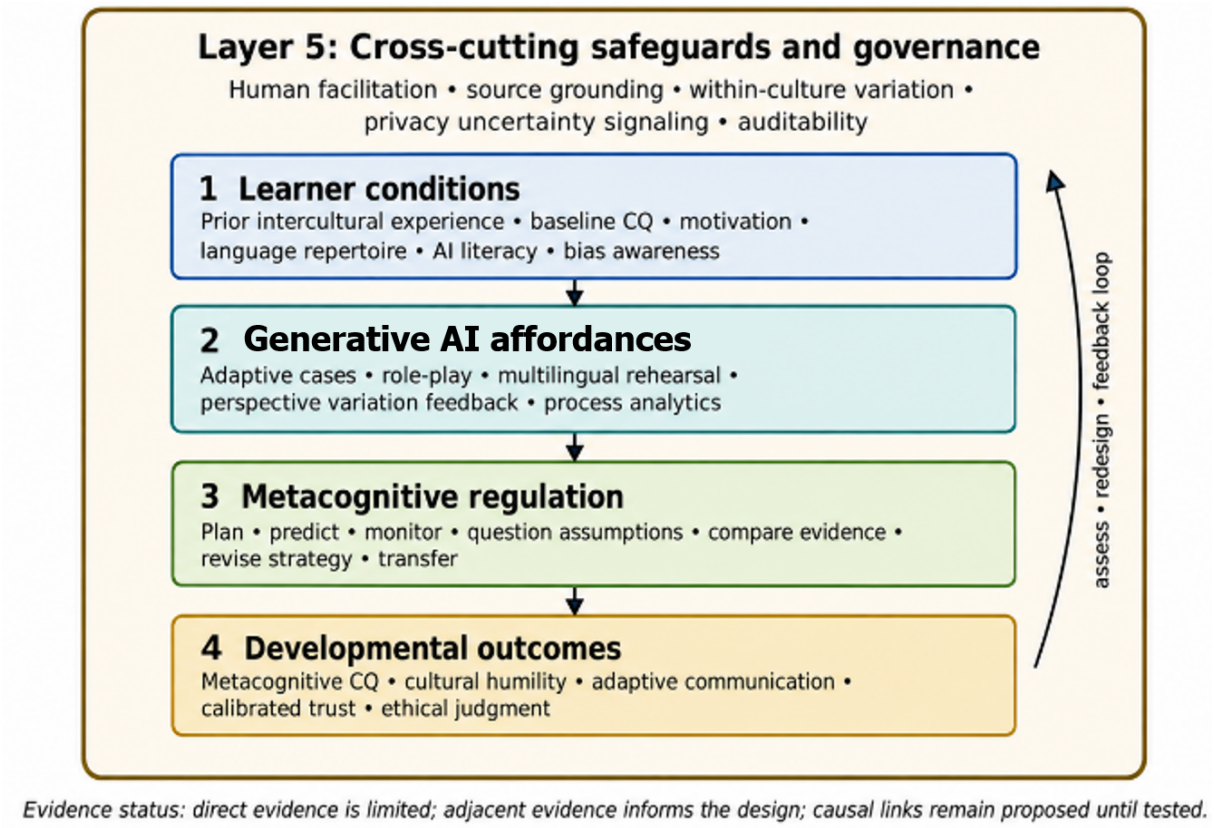


Figure 4. Generative artificial intelligence-enabled metacognitive cultural intelligence training framework. Layers 1–4 describe the proposed learning pathway, while Layer 5 operates as a cross-cutting governance envelope. Direct evidence is limited; adjacent evidence informs the design; causal relationships remain proposed until tested.
Abbreviations: AI: Artificial intelligence; CQ: Cultural intelligence.

Table 5. Evidence base and evidential status of the five-component framework

Framework component	Main supporting literature	What the literature directly supports	Evidential status in the framework
Layer 1: Learner conditions	Chedru & Ostapchuk (2023); Iskhakova *et al.* (2022); McKay *et al.* (2022); Morin & Talbot (2024)	Prior experience, learner characteristics, cultural distance, and engagement are associated with variation in CQ development	Partly empirical; extension to AI literacy and bias awareness is based on adjacent evidence
Layer 2: GenAI affordances	Batista *et al.* (2024); Kasneci *et al.* (2023); Lo (2023); S. Wang *et al.* (2024); Agarwala *et al.* (2024); Dai *et al.* (2025); Lee (2025)	GenAI can provide dialogue, content generation, feedback, adaptation, and multilingual support	Technical affordances are demonstrated; effects on sustained metacognitive CQ are largely untested
Layer 3: Metacognitive regulation	Anders & Dux Speltz (2025); Xiao *et al.* (2025); Xia *et al.* (2026)	AI use can involve planning, monitoring, reflection, and awareness of the learner's own AI-use process	Mechanisms are supported by adjacent evidence; a causal effect on metacognitive CQ remains proposed
Layer 4: Developmental outcomes	H. Wang & Gao (2025); Chedru & Ostapchuk (2023); McKay *et al.* (2022)	Some studies link intercultural competence and CQ-related outcomes to AI use or intercultural experience	Direct testing is needed to assess sustained metacognitive CQ improvement, cultural humility, and durable transfer
Layer 5: Cross-cutting safeguards and governance	Tao *et al.* (2024); Naous *et al.* (2024); Liu (2024); Nguyen *et al.* (2023); UNESCO (2023); Nyaaba *et al.* (2026)	Cultural bias, representational limitations, privacy, accountability, and governance risks are documented	The need for safeguards is supported; the effects of specific safeguards on learning outcomes remain proposed

Abbreviations: AI: Artificial intelligence; CQ: Cultural intelligence.

communication response; explains basis of revisions.

Layer 5 applies to the entire exercise. The system is not a constructed image representing a nation's culture. Factual cultural claims are source-grounded and presented with the appropriate level of uncertainty, and a facilitator can intervene when identity or power issues and/or potentially harmful stereotyping become relevant. These safeguards are design features based on governance and cultural-bias evidence from the adjacent setting; their impact on learning outcomes has yet to be tested.

Finally, the learner performs a similar but new intercultural ambiguity task without AI support. Assessment emphasizes observation and inference, alternatives, contextually relevant evidence, revision of judgements as needed, and an adaptive response. This unaided transfer task operationalizes layer 4 and assesses the ability to implement the regulatory process without AI support.

4.4. Training design principles

Six design principles follow from the synthesis, and each is presented as an evidence-informed recommendation rather than a proven training effect. First, capture the learner's judgment before AI assistance. Second, require the learner to explain the evidence behind that judgment. Third, require multiple plausible interpretations, including at least one non-cultural explanation. Fourth, vary contextual features such as relationships, roles, generations, communication channels, and power. Fifth, ground feedback in evidence by distinguishing fact, inference, and hypothesis, signaling uncertainty, and linking factual cultural claims to sources. Sixth, assess a novel case through an unaided transfer task after feedback. Successful performance with AI assistance alone is not sufficient evidence that regulatory competence has been acquired.

These principles are summarized in an operational design matrix in Table 6, which assigns a failure mode to each AI function and a related control. The controls are precautionary design requirements based on the reviewed evidence; their comparative effectiveness has yet to be tested.

4.5. Theoretical implications

The framework is based on three principles from previous CQ research. First, it creates a virtual space to develop and systematically vary intercultural scenarios. Second,

Table 6. Design matrix for generative artificial intelligence-enabled metacognitive cultural intelligence training

Training function	Metacognitive mechanism	Main failure mode	Required control
Adaptive critical incidents	Prediction, monitoring, strategy revision	Cultural details become stereotyped or fabricated	Use reviewed scenario libraries, source notes, and variation within groups
Conversational role-play	Perspective taking and behavioral rehearsal	The model is treated as an authentic representative of a culture	Disclose simulation status and include multiple personas and contexts
Alternative explanation generation	Assumption checking and cognitive flexibility	False equivalence or unlimited speculation	Require evidence and rank explanations by support and uncertainty
Immediate feedback	Rapid reflection and error correction	Confident but inaccurate guidance	Use rubrics, citations, uncertainty labels, and facilitator review
Multilingual rehearsal	Pragmatic awareness and language adaptation	Translation bias or loss of social meaning	Validate high-stakes cases with proficient local speakers
Process analytics	Detection of repeated assumptions and revision patterns	Privacy loss, surveillance, or simplistic scoring	Minimize data, obtain consent, restrict access, and audit interpretations
Unaided transfer task	Independent strategy use and retention	Learner remains dependent on AI support	Delay assistance and assess a novel case without the model

Abbreviations: AI: Artificial intelligence; CQ: Cultural intelligence.

it specifies that metacognitive CQ is a set of regulatory processes that, in principle, can be studied by analyzing process traces rather than relying only on self-report. Third, it distinguishes cultural adaptation from uncritical conformity: it means purposeful, revisable judgment rather than conforming to an assumed cultural norm.

The framework conceives metacognition as a possible learning goal in GenAI education research and cultural validity as a quality aspect of education. Learning environments can be misrepresentative or untrustworthy if the intercultural scenarios or feedback present narrow cultural assumptions. Cultural validation should therefore be treated as an operational design and evaluation requirement, not just an ethical principle.

The framework also relates model evaluation to human learning. Responsible learning training can be assessed using cultural assessment tools and CQ-related measures. Poor model calibration to a learner's cultural context can undermine the learning context, and good model performance does not imply human learning. Therefore, studying learners' validation, questioning, or refutation of the interpretation(s) produced by the model is worth further research.

4.6. Practical and ethical implications

It should start with a small use case that has been reviewed by trainers. For instance, a pilot could tackle an incorrect assumption in a global virtual team. For designers, the target population, scenario boundaries, reasoning steps, the facilitator's role, the rules governing data retention, and the scenario's outcome measures should be specified beforehand. This makes decisions more auditable and reduces uncontrolled variation in learning and governance.

In high-stakes situations, such as medical treatment, movement, conflict, or disciplinary matters, human facilitation may be essential. Facilitators can elicit emotional issues, power dynamics, and contextual factors in the community that are not represented by the system. While AI can increase practice opportunities, it should not be the ultimate authority on cultural appropriateness.

Institutions should also clearly distinguish between learning support and monitoring. Interaction traces can show how the learner's interpretation has evolved, but they can also be highly sensitive regarding identity, conflict, or performance. Thus, data minimization, informed consent, access limitations, storage time limits, and clearly defined, non-punitive aims are important. Confidentiality and cross-border data transfer should be considered if a commercial system is employed.

Plural review of scenarios, labels, and assumptions is also important for cultural responsibility. Any single reviewer should be regarded not as an all-seeing eye, but

rather as a single voice. A single-bias review needs to be replaced with regular reviews, local or community-based reviews and transparency on how to challenge negative results.

4.7. Research agenda

Stronger causal evidence is needed. Where feasible, randomized or quasi-experimental studies should compare AI-only, human-facilitated AI, conventional training, and appropriate control conditions. Outcomes should include validated CQ measures, process indicators, interaction traces, and delayed transfer. Preregistration can reduce selective reporting, and multisite designs can test whether findings extend across languages and settings.

Research on processes is also required. Learners' prompt logs and think-alouds could reveal whether learners question a claim, check evidence, refine an assumption, or simply restate an AI-generated interpretation at the surface level. The cultural validity of a model can be examined through review by culturally diverse experts and within-culture comparisons. Language coverage, disability access, cost, infrastructure, and local control over training content should also be considered when assessing equity.

Table 7 summarizes the research agenda. The propositions are not conclusions of the current review but testable pathways or directions, based on evidence synthesis, that link innovative design, learning theory, and governance.

4.8. Limitations

Unlike a preserved export of the original, multi-source discovery, the review focuses on a topical corpus collected from a focused topic and verified using sources. The strengths of this retained-corpus audit are that the final bibliography is traceable, the frozen corpus contains no duplicate DOI/title records, and evidence is coded and appraised at the record level. However, the denominators for retrieval, duplicate elimination, title/abstract screening, and full-text screening were not maintained. PRISMA 2020 and PRISMA-S do not indicate that this review is a traditional systematic review with a fully reproducible screening history.

The publication period (January 1, 2021 to July 27, 2026) deliberately focuses on more recent scholarship that is relevant to modern CQ development and the introduction of GenAI-based learning. This option makes the review more relevant to the present and less comprehensive, as pre-2021 seminal works of CQ and metacognition were

Table 7. Future research agenda and testable propositions

Research problem	Example study	Testable proposition
Reflective prompting	Randomized pretest-posttest study with delayed transfer	P1: Reflective prompting will improve metacognitive CQ more than access to unstructured AI dialogue
Assistance timing	Micro-randomized trial varying when hints are released	P2: Delayed AI support will produce stronger unaided transfer than immediate answer provision
Human facilitation	Compare AI-only, facilitator-only, and blended conditions	P3: Human-facilitated AI training will improve ethical judgment and error detection more than AI-only training
Cultural validity	Cross-cultural expert and community review of generated cases	P4: Within-culture variation and source grounding will reduce stereotyping without reducing learner engagement
AI literacy	Measure overtrust, verification behavior, and learning outcomes	P5: AI literacy will moderate training effects by improving calibration and reducing cognitive offloading
Behavioral transfer	Longitudinal field study using observed intercultural interactions	P6: Gains in metacognitive CQ will predict adaptive behavior only when learners demonstrate independent strategy revision
Equity and access	Multi-site mixed-methods study across languages and resource settings	P7: Benefits will vary with language coverage, access quality, and local control over training content
Governance	Evaluate privacy, audit, and user-challenge mechanisms	P8: Transparent uncertainty and contestability will increase trust calibration and reduce harmful reliance

Abbreviations: AI: Artificial intelligence; CQ: Cultural intelligence.

not included in the bibliometric numbers. Therefore, the reported annual distributions and evidence-stream frequencies reflect not the complete literature on CQ or metacognition, but the retained corpus. The bibliometric component is descriptive, and the retained-title term map is a conceptual-structure check that is reproducible; no co-citation, citation-influence, collaboration, or bibliographic-coupling network is claimed due to the lack of the original cited-reference export.

The direct GenAI-intercultural evidence base is very small: only four publications met the direct-stream criterion, and conclusions about training effectiveness are provisional. Judgment also matters in primary-stream coding. The prespecified hierarchy, secondary tags, record-level transparency, and deterministic sensitivity agreement (97.9%) minimize, but do not eliminate, subjectivity in classification. No second human coder was used, so no Cohen's kappa statistic is reported. Finally, technical preprints can be revised prior to formal publication and must be rechecked at submission.

5. Conclusion

Direct evidence on GenAI-assisted intercultural learning remains limited. Only four of the 48 topical publications met the criteria for the direct GenAI–intercultural applications stream. These studies provide early evidence of feasibility for culturally relevant scenarios, communication practice, role-play, feedback, and intercultural learning activities. However, the direct evidence is too limited and heterogeneous to establish that GenAI-supported training produces sustained improvements in metacognitive CQ, cultural humility, behavioral transfer, or ethical judgment.

Adjacent literature provides stronger support for specific design mechanisms than for the overall framework's effectiveness. Recent work across CQ development and training, GenAI learning and metacognition, and cultural bias and governance collectively underscores the importance of reflection, engagement, prior experience, context, learner agency, and cultural validity. These findings inform the design principles proposed in this review, but they should not be interpreted as evidence that the complete GenAI-enabled metacognitive CQ framework is effective.

The five-component framework should therefore be treated as an evidence-informed, testable design framework rather than a validated causal model. Relationships among learner conditions, GenAI affordances, metacognitive regulation, developmental outcomes, and cross-cutting safeguards remain propositions where direct empirical evidence is absent. Future preregistered, experimental or quasi-experimental, longitudinal, and cross-cultural studies should test these relationships using validated CQ measures, process indicators, observed interaction, and unaided transfer. The value of culturally responsible human–AI training will depend on whether it strengthens independent intercultural judgment while preserving learner agency, cultural humility, and human accountability.

Acknowledgments

None.

Funding

None.

Conflict of interest

The authors declare that they have no competing interests.

Author contributions

Conceptualization: Zheyun Zheng, Jiraporn Chano
Formal analysis: Zheyun Zheng
Methodology: All authors
Writing–original draft: Zheyun Zheng
Writing–review & editing: All authors

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

The search logic, eligibility rules, analytical denominator, coding categories, and complete reference corpus are reported in this article.

Further disclosure

During the preparation of this manuscript, the authors used AI-assisted tools for language editing and manuscript preparation. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

References

Agarwala, A., Purwar, A., & Rao, V. (2024). *CultureVo: The serious game of utilizing Gen AI for enhancing cultural intelligence* (Version 2). arXiv.

https://doi.org/10.48550/ARXIV.2407.20685

Ahmadi Forg, F., Doosti, P., Hasani, H., Heydari, F., Roshanfekr, M., Hassangholizadeh, L., Ramezani, F., & Vahedi, S. (2025). Investigating the relationship between cultural intelligence and cultural competence among students in nursing

faculties in Iran: An analytical cross-sectional study. *BMC Nursing, 24*(1), 890.

https://doi.org/10.1186/s12912-025-03555-2

Anders, A. D., & Dux Speltz, E. (2025). Developing generative AI literacies through self-regulated learning: A human-centered approach. *Computers and Education: Artificial Intelligence, 9*, 100482.

https://doi.org/10.1016/j.caeai.2025.100482

Aromataris, E., Lockwood, C., Porritt, K., Pilla, B., & Jordan, Z. (Eds.). (2024). *JBI manual for evidence synthesis*. JBI, Adelaide, Australia.

https://doi.org/10.46658/jbimes-24-01

Batista, J., Mesquita, A., & Carnaz, G. (2024). Generative AI and higher education: Trends, challenges, and future directions from a systematic literature review. *Information, 15*(11), 676.

https://doi.org/10.3390/info15110676

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). Association for Computing Machinery, New York, NY.

https://doi.org/10.1145/3442188.3445922

Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education, 20*(1), 38.

https://doi.org/10.1186/s41239-023-00408-3

Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education, 20*(1), 43.

https://doi.org/10.1186/s41239-023-00411-8

Chedru, M., & Ostapchuk, M. (2023). The effect of study abroad and personality on cultural intelligence: A deeper understanding using the expanded model of cultural intelligence. *International Journal of Intercultural Relations, 92*, 101737.

https://doi.org/10.1016/j.ijintrel.2022.11.001

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International, 61*(2), 228–239.

https://doi.org/10.1080/14703297.2023.2190148

Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education, 20*(1), 22.

https://doi.org/10.1186/s41239-023-00392-8

Dai, D. W., Suzuki, S., & Chen, G. (2025). Generative AI for professional communication training in intercultural contexts: Where are we now and where are we heading? *Applied Linguistics Review, 16*(2), 763–774.

https://doi.org/10.1515/applirev-2024-0184

Davaei, M., & Gunkel, M. (2024). The role of intelligences in teams: A systematic literature review. *Review of Managerial Science, 18*(1), 259–297.

https://doi.org/10.1007/s11846-023-00672-7

Dev, S., Prabhakaran, V., Feman, R. C., Davani, A., Denton, R., Kalia, C., Lertvittayakumjorn, P., Maji, M., Qadri, R., Rostamzadeh, N., Shelby, R., Stella, R., Stepanyan, H., van Liemt, E., Verma, A., Wahltinez, O., Wornyo, E., Zaldivar, A., & Mojsilović, S. (2026). *A unified framework to quantify cultural intelligence of AI* (Version 2). arXiv.

https://doi.org/10.48550/ARXIV.2603.01211

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research, 133*, 285–296.

https://doi.org/10.1016/j.jbusres.2021.04.070

Dwivedi, Y. K., Kshetri, N., Hughes, L., *et al.* (2023). So what if ChatGPT wrote it? Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI. *International Journal of Information Management, 71*, 102642.

https://doi.org/10.1016/j.ijinfomgt.2023.102642

Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2024). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International, 61*(3), 460–474.

https://doi.org/10.1080/14703297.2023.2195846

Goldstein, S. B. (2022). A systematic review of short-term study abroad research methodology and intercultural competence outcomes. *International Journal of Intercultural Relations, 87*, 26–36.

https://doi.org/10.1016/j.ijintrel.2022.01.001

Grassini, S. (2023). Shaping the future of education: Exploring the potential and consequences of AI and ChatGPT in educational settings. *Education Sciences, 13*(7), 692.

https://doi.org/10.3390/educsci13070692

Haddaway, N. R., Page, M. J., Pritchard, C. C., & McGuinness, L. A. (2022). PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. *Campbell Systematic Reviews, 18*(2), e1230.

https://doi.org/10.1002/cl2.1230

Iskhakova, M., & Bradly, A. (2022). Short-term study abroad research: A systematic review 2000–2019. *Journal of Management Education, 46*(2), 383–427.

https://doi.org/10.1177/10525629211015706

Iskhakova, M., Bradly, A., Whiting, B., & Lu, V. N. (2022). Cultural intelligence development during short-term study abroad programmes: The role of cultural distance and prior international experience. *Studies in Higher Education, 47*(8), 1694–1711.

https://doi.org/10.1080/03075079.2021.1957811

Kasneci, E., Sessler, K., Kuchemann, S., *et al.* (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274.

https://doi.org/10.1016/j.lindif.2023.102274

Kim, E., Suk, J., Oh, P., Yoo, H., Thorne, J., & Oh, A. (2024). CLIcK: A benchmark dataset of cultural and linguistic intelligence in Korean. In *Proceedings of the Language Resources and Evaluation Conference* (pp. 3335–3346). European Language Resources Association.

https://doi.org/10.63317/3quu4vzdgf6k

Lee, H.-P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (pp. 1–22). Association for Computing Machinery, New York, NY.

https://doi.org/10.1145/3706598.3713778

Lee, Y. M. (2025). Navigating intercultural competence with ChatGPT: Implications and recommendations for foreign language education. *Journal of Language Teaching, 5*(2), 1–8.

https://doi.org/10.54475/jlt.2025.005

Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., & Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? *The International Journal of Management Education, 21*(2), 100790.

https://doi.org/10.1016/j.ijme.2023.100790

Liu, Z. (2024). Cultural bias in large language models: A comprehensive analysis and mitigation strategies. *Journal of Transcultural Communication, 3*(2), 224–244.

https://doi.org/10.1515/jtc-2023-0019

Liu, Z., Dey, P., Huang, J., Zhao, Z., Jiang, B., Gupta, R., Liu, Y., Du, Y., & Zhao, J. (2025). *Can LLMs grasp implicit cultural values? Benchmarking LLMs' cultural intelligence with CQ-Bench* (Version 2). arXiv.

https://doi.org/10.48550/ARXIV.2504.01127

Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences, 13*(4), 410.

https://doi.org/10.3390/educsci13040410

Mammadov, A., & Wald, A. (2025). The role of cultural intelligence in a firm's international activities: Review, synthesis, and future research. *The International Journal of Human Resource Management, 36*(18), 3251–3315.

https://doi.org/10.1080/09585192.2025.2592773

McGuinness, L. A., & Higgins, J. P. T. (2021). Risk-of-bias VISualization (robvis): An R package and Shiny web app for visualizing risk-of-bias assessments. *Research Synthesis Methods, 12*(1), 55–61.

https://doi.org/10.1002/jrsm.1411

McKay, S., Lannegrand, L., Skues, J., & Wise, L. (2022). International experience and cultural intelligence development: A longitudinal assessment of Australian and French exchange students. *International Journal of Intercultural Relations, 91*, 56–69.

https://doi.org/10.1016/j.ijintrel.2022.08.009

McKeown, S., & Mir, Z. M. (2021). Considerations for conducting systematic reviews: Evaluating the performance of different methods for de-duplicating references. *Systematic Reviews, 10*(1), 38.

https://doi.org/10.1186/s13643-021-01583-y

Morin, G., & Talbot, D. (2023). Cultural intelligence of expatriate workers: A systematic review. *Management Review Quarterly, 73*(1), 413–454.

https://doi.org/10.1007/s11301-021-00245-6

Morin, G., & Talbot, D. (2024). Exploring the effects of international experiences on cultural intelligence. *Journal of International Development, 36*(8), 2959–2985.

https://doi.org/10.1002/jid.3947

Mustafa, M. Y., Tlili, A., Lampropoulos, G., *et al.* (2024). A systematic review of literature reviews on artificial intelligence in education: A roadmap to a future research agenda. *Smart Learning Environments, 11*(1), 59.

https://doi.org/10.1186/s40561-024-00350-5

Naous, T., Ryan, M. J., Ritter, A., & Xu, W. (2024). Having beer after prayer? Measuring cultural bias in large language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics* (pp. 16366–16393). Association for Computational Linguistics.

https://doi.org/10.18653/v1/2024.acl-long.862

Nguyen, A., Ngo, H. N., Hong, Y., Dang, B., & Nguyen, B.-P. T. (2023). Ethical principles for artificial intelligence in education. *Education and Information Technologies, 28*(4), 4221–4241.

https://doi.org/10.1007/s10639-022-11316-w

Nyaaba, M., Wright, A. L., & Choi, G. L. (2026). Generative AI and digital neocolonialism in global education: Towards an equitable framework. *Journal of Digital Educational Technology, 6*(1), ep2608.

https://doi.org/10.30935/jdet/17862

Page, M. J., McKenzie, J. E., Bossuyt, P. M., *et al.* (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71.

https://doi.org/10.1136/bmj.n71

Paul, J., Lim, W. M., O'Cass, A., Hao, A. W., & Bresciani, S. (2021). Scientific procedures and rationales for systematic literature reviews (SPAR-4-SLR). *International Journal of Consumer Studies, 45*(4), O1–O16.

https://doi.org/10.1111/ijcs.12695

Priem, J., Piwowar, H., & Orr, R. (2022). *OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts* (Version 2). arXiv.

https://doi.org/10.48550/ARXIV.2205.01833

Puzzo, G., Sbaa, M. Y., Zappalà, S., & Pietrantoni, L. (2024). The impact of cultural intelligence on burnout among practitioners working with migrants: An examination of age, gender, training, and language proficiency. *Current Psychology, 43*(5), 4443–4457.

https://doi.org/10.1007/s12144-023-04641-x

Rahman, M. M., & Watanobe, Y. (2023). ChatGPT for education and research: Opportunities, threats, and strategies. *Applied Sciences, 13*(9), 5783.

https://doi.org/10.3390/app13095783

Rethlefsen, M. L., Kirtley, S., Waffenschmidt, S., *et al.* (2021). PRISMA-S: An extension to the PRISMA statement for reporting literature searches in systematic reviews. *Systematic Reviews, 10*(1), 39.

https://doi.org/10.1186/s13643-020-01542-z

Soejanto, L. T., Bariyyah, K., Ifdil, I., Andrianie, S., Setyorini, S., Ariyanto, R. D., & Bakar, A. Y. A. (2024). Exploring cultural intelligence in educational studies: A bibliometric perspective on global literature. *COUNS-EDU: The International Journal of Counseling and Education, 9*(1), 37–46.

https://doi.org/10.23916/0020240948410

Tao, Y., Viberg, O., Baker, R. S., & Kizilcec, R. F. (2024). Cultural bias and cultural alignment of large language models. *PNAS Nexus, 3*(9).

https://doi.org/10.1093/pnasnexus/pgae346

Teixeira, K. A. da S., & Klein, A. Z. (2024). The development of cultural intelligence in situated learning. *SAGE Open, 14*(4).

https://doi.org/10.1177/21582440241288056

Tlili, A., Shehata, B., Adarkwah, M. A., *et al.* (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments, 10*(1), 15.

https://doi.org/10.1186/s40561-023-00237-x

UNESCO. (2023). *Guidance for generative AI in education and research*. UNESCO, Paris, France.

https://doi.org/10.54675/EWZM9535

Urgun, D., Seidel, J., Vangeli, E., Borges, M., & de Oliveira, R. F. (2025). Exploring the impact of cross-cultural training on cultural competence and cultural intelligence: A narrative systematic literature review. *Frontiers in Psychology, 16*, 1511788.

https://doi.org/10.3389/fpsyg.2025.1511788

van de Schoot, R., de Bruin, J., Schram, R., *et al.* (2021). An open source machine learning framework for efficient and transparent systematic reviews. *Nature Machine Intelligence, 3*(2), 125–133.

https://doi.org/10.1038/s42256-020-00287-7

Wang, H., & Gao, C. (2025). AI utilization, enjoyment, trust and instructor competence as predictors of intercultural communication competence among Chinese students. *Scientific Reports, 15*(1), 45117.

https://doi.org/10.1038/s41598-025-33052-0

Wang, S., Wang, F., Zhu, Z., Wang, J., Tran, T., & Du, Z. (2024). Artificial intelligence in education: A systematic literature review. *Expert Systems with Applications, 252*, 124167.

https://doi.org/10.1016/j.eswa.2024.124167

Xia, Q., Liu, Q., Tlili, A., & Chiu, T. K. F. (2026). A systematic literature review on designing self-regulated learning using generative artificial intelligence and its future research directions. *Computers & Education, 240*, 105465.

https://doi.org/10.1016/j.compedu.2025.105465

Xiao, Y., Liu, X., & Yao, Y. (2025). Students' development of AI metacognitive awareness in an EAP course: A qualitative inspection through experiential learning theory. *System, 133*, 103790.

https://doi.org/10.1016/j.system.2025.103790

Yusuf, A., Pervin, N., & Roman-Gonzalez, M. (2024). Generative AI and the future of higher education: A threat to academic integrity or reformation? Evidence from multicultural perspectives. *International Journal of Educational Technology in Higher Education, 21*(1).

https://doi.org/10.1186/s41239-024-00453-6