Review
Education and Learning

Generative AI in Korean Language Learning: A Scoping Review

Weijian Kong, **Yundong Wu**

[The author informations are in the declarations section. This article is published by ETFLIN in Education and Learning, Volume 2, Issue 2, 2026, Page 56-66. DOI: 10.58920/edu0202767]

Received
30 July 2026

Revised
08 September 2026

Accepted
23 October 2026

Published
01 November 2026

Editor
Kamaludin Yusra

Keywords: generative artificial intelligence (AI), Korean language learning, scoping review, pragmatic competence, generative AI-assisted language learning, responsible classroom use.

CC BY 4.0

AUTHOR (2026)

Abstract: Generative artificial intelligence (GenAI) is increasingly being integrated into language education, yet evidence concerning its use in Korean language learning remains fragmented. This scoping review examines the emerging research on GenAI and related artificial intelligence applications in Korean language learning, with particular attention to pedagogical affordances, language skills, learner agency, and Korean-specific linguistic and pragmatic characteristics. Following a scoping review framework, relevant literature was identified and screened across major international and Korean academic databases. The final corpus comprised 56 sources, including direct studies of GenAI in Korean language learning, earlier Korean AI and chatbot research, broader research on GenAI in second and foreign language education, and conceptual, ethical, and policy-oriented studies. The findings indicate that existing Korean-language research is concentrated primarily on writing, prompt use, learner perceptions, and university-level instruction, while evidence concerning speaking, listening, reading, and sustained classroom implementation remains comparatively limited. The review further identifies Korean-specific challenges involving honorifics, speech-level selection, contextual appropriateness, and discourse management, which may not be adequately addressed by general-purpose GenAI systems. Across the literature, effective use of GenAI appears to depend on instructional design, learner agency, human mediation, and responsible assessment rather than technology alone. The review concludes by identifying research priorities for empirically validating GenAI-supported Korean language pedagogy and developing context-sensitive, ethically responsible approaches to its integration.

Introduction

Artificial intelligence (AI) has moved from a specialist topic in educational technology to a practical concern in teaching, assessment, and learner support. Reviews of AI in education show a broad range of applications, while recent work on generative systems has emphasized open-ended interaction, content generation, and conversational support (1). A systematic review of AI in higher education also showed that pedagogical roles and educator involvement require explicit attention rather than being treated as secondary implementation details (2). In this article, "generative AI" is used consistently for systems that generate new text, dialogue, explanations, feedback, or learning materials from user prompts; "AI" is retained only for the broader field.

Generative AI differs functionally from earlier computer-assisted language-learning tools. Rule-based parsers, retrieval systems, fixed-response exercises, automated scoring tools, and narrowly scripted chatbots generally operate within predefined response spaces. Contemporary large language models (LLMs), by contrast, generate token sequences conditioned on prompts and conversational context, allowing reformulation, follow-up questioning, changes in difficulty, and open-ended text production (3). The distinction is not absolute because earlier chatbots could also generate responses; the relevant difference is the breadth of learned language patterns, the openness of generation, and the flexibility of interaction. This architectural distinction matters because a responsive conversational interface does not guarantee factual, linguistic, or pragmatic reliability.

Korean is a useful case because linguistic choices frequently encode social relationships. Learners must coordinate particles, verb endings, speech levels, honorific marking, word order, sound change, and lexical choice with status, age, familiarity, institutional role, and communicative purpose (4). Korean requests can combine indirectness with language-specific honorific resources, and the appropriateness of a form depends on the interactional context (5). Consequently, grammatical well-formedness is necessary but not sufficient for pragmatic appropriateness.

This distinction exposes a limitation in broad AI-language reviews. Existing reviews commonly organize

evidence around writing support, feedback, learner perceptions, technology acceptance, or general chatbot use (1). Those categories are useful but do not reveal whether a generated Korean utterance remains appropriate when social variables alter speech level, honorific marking, or directness. A Korean-focused synthesis therefore requires a different unit of analysis: not only what a system can generate, but whether the generated form aligns with the social conditions that license that form.

The present review addresses that gap by mapping four evidence groups separately: (**1**) direct Korean-language GenAI studies; (**2**) earlier Korean AI or chatbot studies; (**3**) general L2/foreign-language AI studies used only for comparative interpretation; and (**4**) conceptual, policy, ethics, and methodological sources. This classification prevents evidence of a proposed affordance from being presented as evidence of a demonstrated Korean-learning effect.

Despite the rapid expansion of research on generative AI in language education, the evidence base remains uneven across languages and learning contexts. Existing reviews have primarily examined English as a second or foreign language or language education more broadly, leaving the Korean language learning context comparatively under-synthesized. This gap is important because Korean presents linguistic and pragmatic characteristics—including honorifics, politeness, register, and context-sensitive expression—that may require forms of learner support and human verification different from those emphasized in general L2 research (4, 5). Accordingly, the present scoping review maps the available evidence on generative AI in Korean language learning while also considering relevant evidence from broader L2/FL contexts where direct Korean-language evidence remains limited.

The review addresses three questions: (**1**) What types of generative-AI applications have been reported or proposed for Korean language learning? (**2**) What learning activities, skills, and pedagogical functions are associated with these applications? and (**3**) What limitations, risks, and requirements for teacher or learner verification are identified across the evidence base?

Literature Review

AI-Supported Language Learning: From Technological Affordances to Generative AI

Technology-supported language learning has developed from computer-assisted language learning (CALL) toward mobile, interactive, and increasingly adaptive forms of learner support. Across this literature, the educational value of technology depends less on access to a tool than on how the tool is integrated with exposure, practice, feedback, interaction, and clearly defined learning objectives (6, 7). Research on intercultural competence similarly emphasizes the importance of culturally and contextually situated communication when technology is used for language learning (8). This principle is important for interpreting GenAI because the ability to generate language does not, by itself, establish language acquisition.

Earlier CALL systems commonly relied on predefined exercises, automated scoring, rule-based feedback, or constrained interaction, while educational chatbot research expanded the possibilities for guided interaction but continued to show variation in instructional design and limited evidence about durable learning outcomes (9–12). Mobile-assisted language learning expanded access to practice and collaboration, but research likewise shows that implementation conditions, task design, and learner activity must be considered alongside outcomes (7, 13, 14). Generative AI extends this trajectory by allowing learners to request explanations, reformulations, examples, and alternative versions within a single interaction. The technological shift is therefore substantial, but the pedagogical question remains how these affordances are embedded in meaningful learning tasks.

Recent language-education research further illustrates the transition from fixed digital resources toward open-ended conversational support. ChatGPT and related systems have been used for content generation, feedback, learner interaction, and teaching support, with writing and learner perceptions among prominent areas of investigation (15). Research on intercultural competence also emphasizes that ChatGPT-supported foreign-language education must attend to culturally and contextually situated communication (8). Research on L2 writing identifies potential support for language form, style, coherence, cohesion, and revision while emphasizing problems of accuracy and intelligibility (16). These findings establish the relevance of GenAI to language education but do not demonstrate equivalent learning effects across languages or contexts.

The distinction between technological affordance and instructional evidence is therefore central. A system may generate multiple dialogues, explanations, or revisions without those outputs necessarily being accurate, pedagogically appropriate, or conducive to durable learning (17). The literature consequently supports examining GenAI as a mediated instructional resource whose value depends on task design, learner activity, feedback, and teacher judgment rather than as an autonomous language authority (2).

Generative AI in L2 and Foreign-Language Education

Recent evidence syntheses indicate that the GenAI literature in L2/FL education is expanding rapidly but remains heterogeneous. A review of 44 ChatGPT studies identified learner perceptions and writing as prominent research foci, while a systematic review of empirical ESL/EFL studies similarly documented a predominantly English-focused evidence base (15, 18). A broader scoping review concluded that research on GenAI in language teaching and learning remains diverse in methods and applications and called for stronger empirical evidence concerning effectiveness, language skills, ethics, and responsible integration (19).

This concentration on English-language settings creates a transfer problem for Korean language education. Findings from ESL/EFL research can illuminate mechanisms such as feedback, drafting, interaction, and learner engagement, but they cannot automatically establish equivalent effects in Korean (18). The transfer problem is especially important when the learning target includes forms whose interpretation depends on social

relationships and discourse context rather than grammatical form alone.

Writing provides the most developed comparison point in the current L2 literature. Studies of ChatGPT-supported teacher feedback describe blended workflows in which generated feedback is treated as input to teacher judgment rather than as a replacement for instructional expertise (20). Comparative research also indicates differences between AI- and teacher-generated feedback in the amount and distribution of attention across content, organization, and language (21). Mixed-method research further shows that learner engagement with AI-generated corrective feedback includes behavioral, cognitive, and affective dimensions that shape how suggestions are interpreted and used during revision (22).

Other empirical findings reinforce the need for contextual interpretation. Research comparing AI-generated and tutor-generated writing feedback does not identify a universally superior feedback source, while intervention research shows that ChatGPT can be incorporated into foreign-language instruction within particular learner populations and instructional settings (23, 24). The broader implication is not that GenAI has a uniform effect on L2 learning, but that its contribution depends on the task, learner, instructional design, and conditions under which AI output is evaluated.

Korean Language Education and the Emerging Local Evidence Base

Korean-focused research provides the necessary local context for interpreting this broader literature. A 2024 trend analysis of ChatGPT research in English and Korean language education found that domestic studies were concentrated largely on university and adult learners, with writing and speaking receiving greater attention than listening and grammar (25). This pattern indicates an emerging research area in which text-based and interaction-oriented applications have developed more quickly than evidence for other language skills.

Direct research with KFL learners also shows that effective GenAI use cannot be assumed from access to the tool. An analysis of writing-task interactions identified recurrent patterns of prompt uniformity, discreteness, and passivity, suggesting that learners may require explicit support for prompt literacy and reflective use (26). The finding is important because it shifts attention from the availability of GenAI to learner agency and the instructional conditions that shape how the technology is used.

Korean writing research has begun to examine ChatGPT across different stages of the composing process, while a university writing-course case study has considered how GenAI can be adapted to course objectives and academic-writing practices (27, 28). Earlier Korean chatbot research provides a developmental reference point for conversational learning, but it used chatbot-builder technology rather than contemporary LLM-based GenAI and should therefore not be treated as direct evidence of current GenAI effects (29). Recent work also addresses the strategic use of generative AI in Korean writing education (30).

Taken together, the Korean literature suggests a developing but still uneven evidence base. Direct GenAI evidence is concentrated in writing, university settings, learner use, and perceptions, while evidence for listening, speaking, reading, assessment, and sustained proficiency development remains comparatively limited (25–30). This distribution supports treating Korean evidence as a distinct evidence tier rather than assuming that the much larger English-language literature directly represents Korean language learning.

Korean Linguistic and Pragmatic Demands

Korean is particularly important for examining the limits of generic GenAI claims because linguistic form is closely connected to social relationships. Descriptions of Korean emphasize speech-level distinctions, honorific marking, sentence-final choices, and lexical resources that interact with context and interlocutor relationships (4). Research on Korean requests further shows that linguistic indirectness and honorific resources contribute to politeness and that appropriate expression depends on interactional conditions (5).

These characteristics mean that Korean GenAI tasks cannot be evaluated solely through grammatical accuracy. A generated request may be syntactically well formed while remaining inappropriate because the selected speech level, honorific marking, degree of indirectness, or lexical choice does not match the relationship between interlocutors (4, 5). Korean therefore provides a useful test case for whether AI-supported language learning can address pragmatic appropriateness as well as linguistic form.

Theoretical Foundations for AI-Supported Korean Learning

Interaction-oriented accounts of second-language acquisition emphasize interaction and meaning negotiation, while Swain's output hypothesis treats learner production as an opportunity to notice gaps and modify output (31, 32). These perspectives provide a theoretical rationale for structured learner–LLM interaction, such as reformulation, comparison, and revision. They do not, however, establish that interaction with an LLM reproduces the social and cognitive conditions of interaction with human interlocutors.

Sociocultural theory conceptualizes second-language development as mediated and socially situated activity (33). From this perspective, GenAI can function as a mediational resource, but its learning value depends on how learners and teachers interpret and use the resource. The noticing literature provides a complementary rationale for comparison tasks in which learners attend explicitly to differences among linguistic forms (34).

Research syntheses focusing specifically on instructed second-language acquisition frequently caution researchers, language educators, and curriculum designers against prematurely equating mere classroom exposure or initial practice opportunities with durable, long-term acquisition (35). Broader, multifaceted accounts of second-language development likewise consistently emphasize multiple interacting cognitive, social, and psychological processes rather than a simple, linear accumulation of input (36). Vocabulary-learning research consistently supports rich, contextualized encounters and sustained, repeated engagement, which can effectively inform

AI-generated lexical activities while still necessitating rigorous accuracy and contextual-fit checks (37).

Learner Agency, Motivation, and Self-Regulated Use

Learner agency is an additional dimension of GenAI-supported language learning. Self-determination theory identifies autonomy, competence, and relatedness as dimensions of motivation, while self-efficacy theory emphasizes learners' beliefs about their capabilities (38, 39). These perspectives suggest that research should examine not only observable language performance but also whether AI-supported activities encourage active decision-making or passive dependence on generated answers.

The relevance of learner agency connects directly to the governance literature. AI in education raises questions concerning pedagogy, assessment, responsibility, access, and institutional oversight (40). UNESCO's guidance on GenAI emphasizes human-centred use, including privacy, transparency, equity, and responsible implementation (41). These principles are particularly relevant to language education because generated explanations, corrections, and reformulations can directly shape learners' linguistic choices.

Ethics, Governance, and the Need for Human Mediation

Research on self-regulated use of machine translation and GenAI in academic writing further indicates that learners' cognitive and sociocultural engagement with AI tools is central to understanding how the tools are actually used (42). For Korean instruction, this supports tasks that require learners to justify prompt choices, compare alternatives, verify explanations, and revise outputs rather than simply accepting the first response.

Across the literature, the resulting pedagogical position is one of mediated use. GenAI can expand access to language input, interaction, and feedback, but learners and teachers remain responsible for evaluating accuracy, contextual fit, pragmatic appropriateness, and educational purpose. This principle also provides the conceptual basis for distinguishing demonstrated findings from proposed classroom applications in the present review.

Rationale for a Scoping Review

A scoping review is appropriate because the literature is heterogeneous in study design, learner population, technology, language skill, and outcome. The purpose is to map concepts and evidence rather than calculate a pooled effect size (43). Methodological guidance emphasizes transparent eligibility criteria, systematic data charting, and explicit reporting of the evidence landscape (44), while PRISMA-ScR provides reporting principles for transparent scoping-review selection and synthesis (45).

Methodology

Research Design

This study employed a scoping review design. The objective was to map the available evidence concerning generative AI in Korean language learning, identify Korean-specific linguistic and pragmatic constraints, and synthesize instructional and governance implications. Because the included sources differed in design and outcome, no meta-analysis or pooled effectiveness estimate was attempted.

Data Sources and Search Strategy

Searches were conducted in Scopus, Web of Science, ERIC, Google Scholar, RISS, and KCI. The documented search window was 10 May 2026 to 15 June 2026, with 15 June 2026 serving as the final search cut-off. This date is a methodological boundary of the evidence map and must be distinguished from manuscript preparation, submission, and evaluation dates; the revised manuscript is interpreted as a version prepared after the search cut-off.

Database-Level Retrieval

The database-level retrieval counts were Scopus (412), Web of Science (286), ERIC (198), Google Scholar (176), RISS (143), and KCI (121), yielding 1,336 records. After 286 duplicate records were removed, 1,050 unique records remained for title and abstract screening. These counts are reported separately to make the audit trail reproducible.

Search Strategy and Search Terms

The English search strategy combined terms for generative AI, Korean language learning, language-learning skills, pragmatics, assessment, and ethics. The principal Boolean string was:

("generative artificial intelligence" OR "ChatGPT" OR "large language model" OR "AI chatbot" OR "generative AI") AND ("Korean language learning" OR "Korean language education" OR "Korean learning" OR "second language learning" OR "foreign language learning") AND (honorific OR pragmatics OR "speech level" OR writing OR speaking OR listening OR reading OR assessment OR ethics)*

For RISS and KCI, the documented Korean-language search was:

("생성형 인공지능" OR "생성형 AI" OR "대규모 언어 모델" OR "대형 언어 모델" OR ChatGPT OR "AI 챗봇") AND ("한국어 학습" OR "한국어 교육" OR "한국어 학습자" OR "제2언어 학습" OR "외국어 학습") AND (존댓말 OR 높임말 OR 경어 OR 화용론 OR 말하기 OR 듣기 OR 읽기 OR 쓰기 OR 평가 OR 윤리).

Platform-specific operators were adapted to the syntax supported by each database while preserving these conceptual term groups.

The search was broad enough to retrieve direct Korean GenAI studies, earlier Korean chatbot work, and broader second- or foreign-language studies. During screening, these evidence types were classified separately and were not treated as equivalent evidence.

Screening and Selection

Two screening stages were used. Titles and abstracts were first assessed against the eligibility criteria. Potentially relevant records were then assessed through full text or, when full text was unavailable, through detailed bibliographic information sufficient to establish identity, scholarly venue, relevance, and eligibility.

Of the 1,050 unique records, 823 were excluded at title and abstract screening: 341 were unrelated to language education, 219 concerned general AI technology without a language-learning connection, 187 focused on non-Korean language learning without sufficient relevance

to the review questions, and 76 lacked sufficient scholarly information. This left 227 records for second-stage assessment.

At the second stage, 171 records were excluded: 62 lacked a relevant Korean-language learning or pragmatic connection; 21 were duplicate or substantially overlapping publication records; 34 lacked sufficient methodological or bibliographic information; 29 were commentary or opinion sources without a substantive contribution to the review questions; and 25 could not be confirmed because full-text information was unavailable. The final corpus therefore contained 56 sources.

For operational clarity, "insufficient methodological or bibliographic information" meant that the available record did not provide enough information to identify the author or title, establish the scholarly venue or publication year, or determine relevance to the eligibility criteria. "Commentary or opinion source" referred to a source that did not provide substantive empirical evidence, review synthesis, policy guidance, or a clearly documented conceptual contribution relevant to the research questions. The resulting screening pathway is summarized in **Figure 1**.

Reviewer Roles and Disagreement Procedures

Two reviewers independently assessed records using the same eligibility criteria. Screening decisions were recorded separately before comparison. Disagreements were resolved through discussion using the predefined criteria; when agreement could not be reached, a third reviewer adjudicated the decision. The same procedure was applied when classifying included sources by evidence group.

Data Charting and Analysis

For each included source, the charting form recorded publication year, educational context, learner group, language-learning focus, AI type, directness of Korean relevance, source type, learning skill, pragmatic or cultural

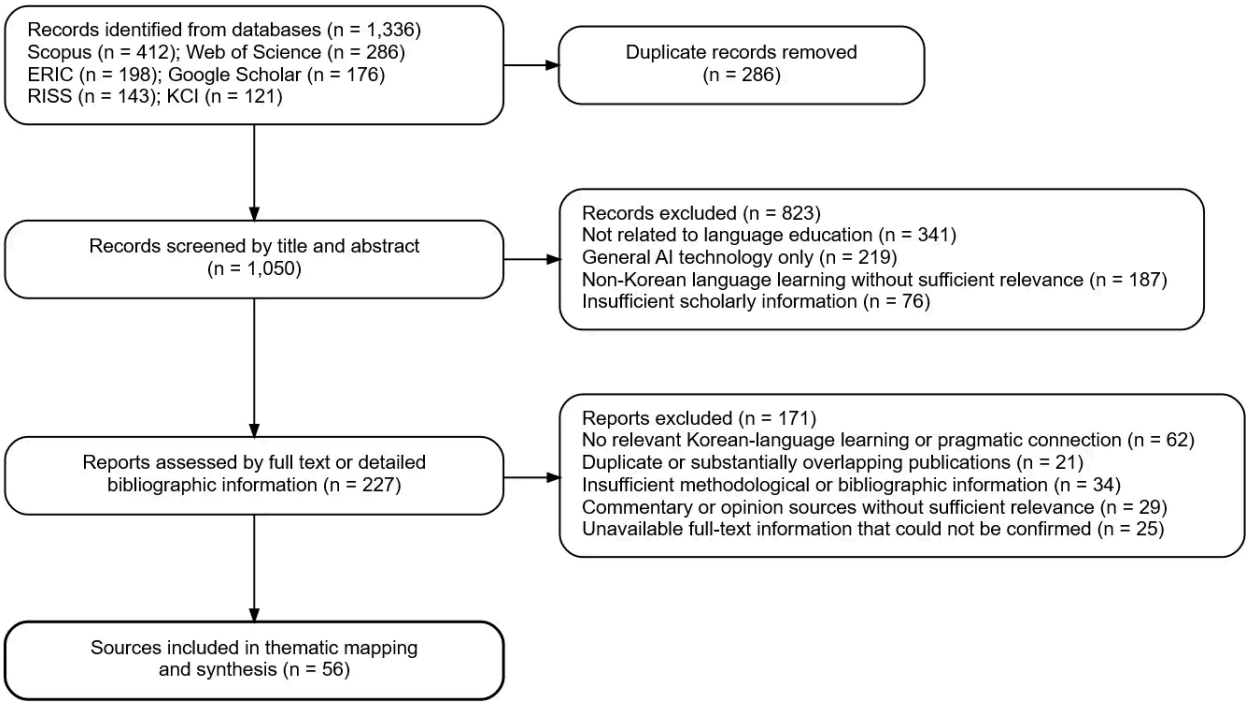


Figure 1. PRISMA-ScR flow diagram of the literature search and study selection process.

Table 1. Inclusion and exclusion criteria for the scoping review.

Inclusion criteria	Exclusion criteria
Peer-reviewed studies, reviews, books, and policy reports relevant to AI, language education, Korean learning, or ethics.	Sources unrelated to education or language learning.
Sources published mainly from 2019 to June 2026, with older theoretical or methodological works retained when directly relevant.	Sources discussing AI only as a general technology without a learning connection.
Sources addressing Korean pragmatics, honorifics, speech levels, learner needs, or comparative L2/foreign-language evidence relevant to the review questions.	Records for which available metadata could not establish author/title, scholarly venue, publication year, or relevance to the eligibility criteria.
Sources with sufficient information to classify evidence type and directness.	Commentary or opinion sources without substantive empirical evidence, review synthesis, policy guidance, or a clearly documented conceptual contribution relevant to the review questions.

Table 2. Composition of the 56-source review corpus by evidence category.

Evidence category	Number	Main focus	Role in synthesis
Direct Korean-language GenAI studies	8	Korean learning, writing, and GenAI/LLM use	Primary Korean-specific evidence
Earlier Korean AI/chatbot studies	7	Chatbots and AI-assisted Korean learning	Developmental context before current GenAI
General L2/foreign-language AI studies	18	Writing, interaction, feedback, learner use	Comparative evidence, not direct Korean evidence
Conceptual, policy, ethics, and review sources	23	AI literacy, assessment, privacy, pedagogy, methods	Framework and governance interpretation

feature, reported benefit or affordance, reported limitation or risk, and teaching or assessment implications. The synthesis used thematic comparison rather than quantitative effect estimation.

Operational Definitions and Eligibility Criteria

Generative AI was defined as a digital system capable of generating new text, dialogue, feedback, explanation, or learning material in response to user prompts, with current LLM-based systems as the principal object of interest. Earlier rule-based, retrieval-based, automated-scoring, and narrowly scripted chatbot systems were excluded from the GenAI category unless they served as historical or comparative evidence. Korean language learning included formal classes, self-directed learning, TOPIK preparation, study-abroad preparation, workplace Korean, and culturally motivated learning.

Sources were included when they addressed AI in language learning, Korean language education, Korean pragmatics, review methodology, ethics, or relevant learning theory. Sources were excluded when they were unrelated to education or language learning; discussed AI only as a general technology without a learning connection; duplicated another record; or lacked sufficient bibliographic information to establish scholarly provenance and relevance.

Screening and Data Preparation

The screening process was operationalized using the criteria in **Table 1**. The criteria were designed to retain sources that could contribute either direct Korean-language evidence or clearly bounded comparative, theoretical, methodological, or governance evidence.

Data Synthesis

The synthesis proceeded in three stages. First, sources were grouped by evidence type and research focus. Second, recurring findings were compared across listening, speaking, reading, writing, pragmatics, assessment, equity, privacy, and governance. Third, the findings were aligned with RQ1–RQ3. The four evidence groups were retained throughout the synthesis so that comparative evidence was not presented as direct Korean evidence.

Characteristics of the Review Corpus

The final corpus was organized into four evidence categories so that direct Korean-language evidence could be distinguished from developmental, comparative, and conceptual sources. **Table 2** summarizes the composition and role of each category in the synthesis.

Table 2 indicates that the review corpus is intentionally evidence-tiered: the 8 direct Korean GenAI studies provide the closest answer to the review questions, whereas the other categories are used to establish developmental context, comparative L2 evidence, or conceptual and governance boundaries. This distinction is maintained throughout the Results and Discussion so that broader evidence is not presented as direct evidence of Korean-learning effects.

Results

Composition and Scope of the Evidence Base

The final corpus contained 56 sources distributed across four evidence categories: 8 direct Korean-language GenAI studies, 7 earlier Korean AI/chatbot studies, 18 general L2/foreign-language AI studies, and 23 conceptual, policy, ethics, or review sources. Korean GenAI studies represent the primary evidence for the review questions, while broader L2/FL studies serve a comparative role and conceptual or governance sources provide interpretive and methodological context.

The distribution of evidence is important for interpreting the review. The corpus is not a homogeneous set of intervention studies and therefore cannot support a single estimate of GenAI effectiveness. Instead, it maps a developing evidence landscape in which direct Korean evidence is smaller than the broader L2/FL literature and is concentrated in writing, university contexts, learner use, and perceptions (25–30).

Generative AI Applications Across Language Skills

Across the 56-source corpus, GenAI was associated with potential or documented applications in listening, speaking, reading, and writing, but the strength and directness of evidence differed by skill. The review therefore distinguishes documented uses from pedagogical task designs and does not interpret the existence of an available AI function as evidence of a learning gain.

For speaking, the mapped literature supports conversational practice, role-play, and chatbot-mediated interaction as plausible opportunities for repeated output. However, direct Korean GenAI evidence remains limited, and earlier Korean chatbot research represents developmental evidence rather than evidence equivalent to current LLMs (29). Role-play activities involving requests, service encounters, or institutional interactions should therefore be understood as evidence-informed designs requiring empirical evaluation.

For listening, the evidence is comparatively indirect. AI-supported activities can be designed around transcript comparison, controlled repetition, and variation in speaking rate, but the present corpus does not establish a Korean-specific longitudinal effect on listening proficiency. Such activities should therefore be evaluated through explicit listening outcomes rather than inferred to be effective from the existence of the technology.

For reading, GenAI can support summarization, vocabulary explanation, and alternative representations of a text. The principal risk is that generated summaries may omit stance, cultural cues, or interactionally important information. A more defensible instructional use is therefore comparison between an AI-generated summary and the source text, with learners identifying omissions and evaluating whether the generated interpretation preserves the original meaning.

Writing has the most developed evidence base. The broader L2 literature documents uses of GenAI for drafting, revision, feedback, language form, style, coherence, and cohesion, while also identifying accuracy and intelligibility concerns (16, 20–24). Korean-specific studies similarly examine writing processes, prompt literacy, and strategic GenAI use (25–28, 30). These findings support process-oriented writing activities but do not establish that GenAI independently improves Korean writing proficiency.

Table 3 translates these patterns into skill-specific instructional possibilities and corresponding verification requirements. The proposed activities are derived from the mapped evidence and are not presented as universally validated Korean classroom interventions.

Korean-Specific Constraints and Risks

The most consistent Korean-specific boundary concerns pragmatic fit. Korean honorifics, speech levels, indirectness, and address forms operate through interactional variables, so the appropriateness of an utterance cannot be determined from grammatical accuracy alone (4, 5). The corpus supports treating pragmatic mismatch as an important risk mechanism, but it does not provide enough controlled Korean LLM testing to calculate a population-level error rate.

Two particularly important mechanisms for future testing are subject-oriented honorification and hearer-oriented speech-level selection. A generated form may be grammatically possible while failing to match whether a referent should be honorified, or a sentence-final form may remain unchanged when the interlocutor's status, familiarity, or institutional role changes (4, 5). These are analytically useful failure conditions, not quantified error rates.

A second risk concerns multi-turn discourse-state stability. In extended interactions, an LLM may receive several social variables and later produce an utterance inconsistent with an earlier role assignment. The relevant research question is therefore whether the model maintains a stable relationship between social variables and linguistic choices across turns. The present review identifies this as an empirical priority rather than an established prevalence finding.

Broader GenAI literature also identifies inaccurate or fabricated output as a recurring concern (17–19). In Korean learning, the same problem can affect explanations of register, lexical nuance, cultural conventions, and pragmatic appropriateness. Assessment validity is another concern when AI assistance changes the relationship between a learner's observable process and final product; process records, revision explanations, oral follow-up, or timed writing can make learner decision-making more visible.

Equity and privacy are governance concerns rather than Korean-specific learning effects. Access to paid models, reliable devices, stable internet, and digital literacy varies across learners, while institutional use raises questions about data protection and disclosure (40, 41). These concerns support transparent policies and alternative learning pathways rather than assumptions of universal access.

Human Mediation, Instructional Design, and Under-Researched Areas

The evidence supports a mediated instructional model in which AI-generated feedback or language is treated as an input for learner and teacher evaluation. Research on teacher feedback, learner engagement, and AI-generated writing feedback indicates differences in feedback distribution, learner uptake, and preferences across AI and human sources (20–23). These findings support blended workflows but do not establish that teacher mediation itself causes improved Korean proficiency.

For Korean instruction, teacher mediation is particularly relevant when the learning objective includes pragmatic appropriateness. The teacher may act as task designer, cultural interpreter, output checker, assessment designer, and ethics guide. Learners should likewise compare alternatives, question explanations, verify claims, and justify revisions rather than accepting generated

Table 3. Skill-specific GenAI applications and required human verification.

Skill	Korean focus	AI-assisted task	Required human check
Listening	Sound change, final consonants, speech rate	Compare AI-generated dialogue/transcript variants with a teacher-selected recording.	Teacher verifies natural pronunciation and transcript accuracy.
Speaking	Honorifics, requests, address terms	Role-play a request to a professor and annotate the interaction log.	Teacher checks speech level, honorific marking, tone, and relationship fit.
Reading	Formal style, stance, cultural cues	Compare an AI summary with the original text and mark omissions.	Teacher checks omitted meaning and culturally important cues.
Writing	Particles, cohesion, genre, email etiquette	Submit first draft, AI feedback, revised draft, and reflection.	Teacher checks whether revisions are justified and independently understood.

language as authoritative (40 – 42).

The review identifies four priorities for empirical research. First, Korean GenAI tasks should be tested longitudinally with explicit proficiency and pragmatic outcomes. Second, studies should manipulate social variables such as subject honorification, hearer-oriented speech level, address terms, and indirectness. Third, assessment research should compare process-sensitive and final-product measures under clearly documented AI-use conditions. Fourth, teacher-readiness research should examine prompt design, output verification, privacy, workload, and assessment practices in actual Korean classrooms.

These priorities reflect the composition of the evidence base: the literature is sufficiently developed to identify recurring applications and risks, but not sufficiently developed to support broad causal claims about GenAI effectiveness in Korean language learning.

Discussion

Synthesis of the Evidence Base

The review indicates that GenAI has expanded the range of language-learning activities that can be designed around repeated interaction, reformulation, drafting, and feedback. This finding is consistent with the broader development of CALL and MALL, in which educational value depends on task design, learner participation, and implementation conditions rather than access to technology alone (6, 7, 11–14). The distinctive contribution of GenAI is therefore the breadth and flexibility of generated language, not evidence of an automatic improvement in learning outcomes.

The 56-source corpus also shows a clear imbalance in the evidence base. Direct Korean GenAI research is smaller than the broader L2/FL literature and is concentrated in writing, university contexts, and learner use (25–30). Broader L2 studies provide useful comparative evidence about feedback, writing, interaction, and learner engagement (15–24), but their findings cannot be assumed to transfer directly to Korean.

The most defensible interpretation is consequently that GenAI represents a potentially useful supplementary resource whose educational contribution depends on the conditions of use. The review does not establish a pooled or causal effect on Korean proficiency, and the heterogeneity of the evidence is itself an important finding.

Korean Pragmatic Competence as a Central Boundary

Korean-specific evidence changes the unit of analysis from language generation alone to contextually appropriate language use. Korean speech levels, honorific marking, indirectness, and address forms interact with social relationships and communicative purposes (4, 5). Consequently, a generated sentence can be grammatically acceptable while remaining pragmatically inappropriate.

The theoretical literature provides a useful explanation for why structured learner–AI interaction may have educational value. Interaction and output perspectives emphasize meaning negotiation and opportunities to notice and modify language (31, 32), while sociocultural theory frames learning as mediated and socially situated (33). Noticing and instructed-SLA research further caution that exposure and practice should not be equated with durable acquisition (34–36). These theories support structured comparison, revision, and reflection but do not constitute empirical evidence that LLM interaction itself produces acquisition.

For Korean, an appropriate evaluation protocol should manipulate social variables rather than testing isolated sentences only. At minimum, studies should examine subject-oriented honorification, hearer-oriented speech-level shifts, relative status, familiarity, institutional role, indirectness, and address terms. This would allow researchers to test whether model output changes appropriately when the social conditions governing Korean expression change.

The review therefore identifies pragmatic competence as a central boundary condition for Korean GenAI use. Future studies should report model version, prompt conditions, learner proficiency, scenario characteristics, and evaluation criteria so that linguistic accuracy and pragmatic appropriateness can be distinguished.

Learner Agency, Assessment, Equity, and Responsibility

The evidence also suggests that learner agency should remain visible in AI-supported learning. Self-determination theory and self-efficacy theory provide complementary perspectives on autonomy, competence, and learners' beliefs about their capabilities (38, 39). These concepts are relevant because unrestricted AI assistance may either support active decision-making or encourage passive acceptance of generated answers. The present review identifies this as an important pedagogical issue rather than as a demonstrated causal effect.

Assessment requires similar caution. When learners submit only polished AI-assisted products, the final text may no longer provide a transparent representation of the learner's independent performance. Process evidence such as drafts, prompt logs, revision explanations, oral follow-up, and timed writing can make learner decision-making more visible. Such measures should be aligned with institutional assessment policies rather than treated as universally required.

Equity, privacy, and institutional responsibility are broader governance concerns. AI-in-education scholarship and UNESCO guidance emphasize human-centred implementation, transparency, privacy, equity, and responsible use (40, 41). Research on self-regulated use of GenAI further indicates that learners' cognitive and sociocultural engagement with AI shapes how the technology is actually used (42).

Responsibility should therefore be distributed across institutions, teachers, and learners. Institutions need clear policies on privacy, disclosure, and acceptable tools; teachers need to define learning objectives, design tasks, verify outputs, and determine acceptable assistance; and learners need to evaluate generated language and remain responsible for final choices.

Pedagogical Implications

The synthesis supports a process-based framework for GenAI-assisted Korean instruction: define the learning objective; specify the social context; generate or compare

language; verify linguistic and pragmatic fit; revise; and document the learner's reasoning. Role-play, contrastive examples, summary checking, writing revision, and reflective portfolios can fit this framework, but they should be presented as evidence-informed task designs rather than proven superior methods.

Teacher preparation should include Korean-specific output verification. Professional development can address prompt specification, factual checking, honorific and speech-level analysis, pragmatic appropriateness, privacy, disclosure, and process-sensitive assessment. Learners should be positioned as critical agents who compare alternatives, ask why a form was selected, test it against the stated social situation, and justify revisions.

Limitations

This review has four principal limitations. First, it searched selected international and Korean databases and used a defined set of English and Korean terms, so relevant studies indexed elsewhere may have been missed. Second, the corpus contains heterogeneous source types, including empirical studies, reviews, conceptual papers, policy documents, and methodological sources. Third, direct empirical evidence on GenAI in Korean language classrooms remains smaller than the broader L2 literature. Fourth, GenAI systems change rapidly, so findings tied to a specific model version may have limited temporal stability. The review should therefore be interpreted as a map of evidence available up to 15 June 2026, not as a quantitative estimate of GenAI effectiveness.

Conclusion

This scoping review maps a heterogeneous 56-source evidence base and yields three bounded conclusions. First, GenAI offers plausible and in some contexts documented support for listening, speaking, reading, and writing, but the evidentiary basis is stronger in broader L2/foreign-language research than in direct Korean classroom studies. Second, Korean introduces a distinctive evaluation requirement because linguistic form must be interpreted together with speech level, honorific marking, relationship, communicative purpose, and genre. Third, responsible implementation requires explicit task design, output verification, critical AI literacy, transparent disclosure, and process-sensitive assessment; the review does not establish a causal effect of teacher mediation on Korean learning outcomes.

The pedagogical boundary introduced by GenAI in Korean language acquisition can be stated as a five-part alignment requirement: linguistic form, social relationship, speech level, communicative purpose, and genre must remain mutually compatible. This boundary is more specific than a general recommendation to "check AI output," because it identifies the dimensions that teachers and learners should evaluate in Korean.

Future research should test these dimensions in real classrooms using transparent AI-use records and clearly defined proficiency and pragmatic outcomes. Priority designs should manipulate subject- and hearer-honorific conditions, speech-level shifts, indirectness, and institutional roles, and should report model version, prompt conditions, learner proficiency, and evaluation criteria. Repeated database searches should preserve database-level logs so that future updates can distinguish changes in the literature from changes in model architecture or training data.

Declaration

Author Information

Weijian Kong *(*Corresponding author)*, Higher Education, Psychology, and Policy Analysis Laboratory, Naju, 58245, South Korea. **Contribution:** Writing – Review & Editing, Writing – Original Draft, Supervision, Project Administration, Formal Analysis, Conceptualization.

Yundong Wu *(*Corresponding author)*, Higher Education, Psychology, and Policy Analysis Laboratory, Naju, 58245, South Korea. **Contribution:** Writing – Original Draft, Writing – Review & Editing, Conceptualization, Data Curation.

Acknowledgment

During the preparation of this manuscript, the authors used Google Gemini to assist with grammar refinement, syntax editing, and language optimization. Following the use of this tool, the author(s) thoroughly reviewed and edited the text as necessary and take full responsibility for the content and integrity of the final publication.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability

Not applicable.

Ethics Statement

Not applicable.

Funding Information

This work received no external funding.

References

1. Crompton H, Burke D. Artificial intelligence in higher education: the state of the field. Int J Educ Technol High Educ. 2023;20(1):22. doi: https://doi.org/10.1186/s41239-023-00392-8

2. Zawacki-Richter O, Marín VI, Bond M, Gouverneur F. Systematic review of research on artificial intelligence applications in higher education – where are the educators?. Int J Educ Technol High Educ. 2019;16(1):39. doi: https://doi.org/10.1186/s41239-019-0171-0

3. Godwin-Jones R. Emerging spaces for language learning: AI bots, ambient intelligence, and the metaverse. Llt. 2023;27(2):6-27. doi: https://doi.org/10.64152/10125/73501

4. Cho H, Kim Y. Learning Korean honorifics through individual and collaborative writing tasks and written corrective feedback. Appl Linguist Rev. 2022;13(1):19-47. doi:10.1515/applirev-2018-0075

5. Byon AS. The role of linguistic indirectness and honorifics in achieving linguistic politeness in Korean requests. Journal of Politeness Research. Language,

Behaviour, Culture. 2006;2(2):247-276. doi: https://doi.org/10.1515/pr.2006.013

6. Yang SC, Chen YJ. Technology-enhanced language learning: A case study. Computers in Human Behavior. 2007;23(1):860-879. doi: https://doi.org/10.1016/j.chb.2006.02.015

7. Burston J. Twenty years of MALL project implementation: A meta-analysis of learning outcomes. ReCALL. 2014;27(1):4-20. doi: https://doi.org/10.1017/s0958344014000159

8. Lee YM. Navigating intercultural competence with ChatGPT: Implications and recommendations for foreign language education. Jlt. 2025;5(2):1-8. doi: https://doi.org/10.54475/jlt.2025.005

9. Chapelle CA. Computer Applications in Second Language Acquisition. Cambridge, UK: Cambridge University Press; 2001. doi: https://doi.org/10.1017/cbo9781139524681

10. Hwang GJ, Chang CY. A review of opportunities and challenges of chatbots in education. Interactive Learning Environments. 2021;31(7):4099-4112. doi: https://doi.org/10.1080/10494820.2021.1952615

11. Warschauer M, Healey D. Computers and language learning: an overview. Lang. Teach. 1998;31(2):57-71. doi: https://doi.org/10.1017/s0261444800012970

12. Hubbard P. Guest Editorial. Computer Assisted Language Learning. 2004;17(1):1-6. doi: https://doi.org/10.1076/call.17.1.1.29708

13. Stockwell G. Using mobile phones for vocabulary activities: Examining the effect of platform. Llt. 2010;14(2):95-110. doi: https://doi.org/10.64152/10125/44216

14. Kukulska-Hulme A, Shield L. An overview of mobile assisted language learning: From content delivery to supported collaboration and interaction. ReCALL. 2008;20(3):271-289. doi: https://doi.org/10.1017/s0958344008000335

15. Yang L, Li R. ChatGPT for L2 learning: Current status and implications. System. 2024;124:103351. doi: https://doi.org/10.1016/j.system.2024.103351

16. Barrot JS. Using ChatGPT for second language writing: Pitfalls and potentials. Assessing Writing. 2023;57:100745. doi: https://doi.org/10.1016/j.asw.2023.100745

17. Kasneci E, Sessler K, Küchemann S, Bannert M, Dementieva D, Fischer F, *et al*. ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences. 2023;103:102274. doi: https://doi.org/10.1016/j.lindif.2023.102274

18. Lo CK, Yu PLH, Xu S, Ng DTK, Jong MSY. Exploring the application of ChatGPT in ESL/EFL education and related research issues: a systematic review of empirical studies. Smart Learn. Environ. 2024;11(1):50. doi: https://doi.org/10.1186/s40561-024-00342-5

19. Law L. Application of generative artificial intelligence (GenAI) in language teaching and learning: A scoping literature review. Computers and Education Open. 2024;6:100174. doi: https://doi.org/10.1016/j.caeo.2024.100174

20. Han J, Li M. Exploring ChatGPT-supported teacher feedback in the EFL context. System. 2024;126:103502. doi: https://doi.org/10.1016/j.system.2024.103502

21. Guo K, Wang D. To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. Educ Inf Technol. 2023;29(7):8435-8463. doi: https://doi.org/10.1007/s10639-023-12146-0

22. Yan D, Zhang S. L2 writer engagement with automated written corrective feedback provided by ChatGPT: A mixed-method multiple case study. Humanit Soc Sci Commun. 2024;11(1):1086. doi: https://doi.org/10.1057/s41599-024-03543-y

23. Escalante J, Pack A, Barrett A. AI-generated feedback on writing: insights into efficacy and ENL student preference. Int J Educ Technol High Educ. 2023;20(1):57. doi: https://doi.org/10.1186/s41239-023-00425-2

24. Karataş F, Abedi FY, Ozek Gunyel F, Karadeniz D, Kuzgun Y. Incorporating AI in foreign language education: An investigation into ChatGPT's effect on foreign language learners. Educ Inf Technol. 2024;29(15):19343-19366. doi: https://doi.org/10.1007/s10639-024-12574-6

25. Kohnke L, Moorhouse BL, Zou D. ChatGPT for language teaching and learning. RELC J. 2023;54(2):537-550. doi:10.1177/00336882231162868

26. Su Y, Yang M. Unlocking the power of ChatGPT: a framework for applying generative AI in education. ECNU Rev Educ. 2023;6(3):355-366. doi:10.1177/20965311231168423

27. Kim M. A Study on Teaching and Learning Methods for Korean Writing using ChatGPT. The Journal of Korean Literary Creative Writng. 2023;22(2):55-86. doi: https://doi.org/10.47057/jklcw.2023.58.03

28. Lee MN. Researching Academic Writing Practices and Pedagogy with ChatGPT : A Case Study of H University's "Logical Thinking and Writing" Course. Korean Lang Lit. 2024;128:265-302. doi: https://doi.org/10.21793/koreall.2024.128.265

29. Jung JK, Kim HS. A study on the development and validation of Korean language education chatbots using chatbot builder: Focusing on learners' perceptions. Foreign Languages Education. 2024;31(3):101-131. doi: https://doi.org/10.15334/fle.2024.31.3.101

30. Su Y, Lin Y, Lai C. Collaborating with ChatGPT in argumentative writing classrooms. Assess Writing. 2023;57:100752. doi:10.1016/j.asw.2023.100752

31. Long MH. The Role of the Linguistic Environment in Second Language Acquisition. Amsterdam, Netherlands: Elsevier; 1996. doi: https://doi.org/10.1016/b978-012589042-7/50015-3

32. Swain M, Lapkin S. Problems in output and the cognitive processes they generate: a step towards second language learning. Appl Linguist. 1995;16(3):371-391. doi:10.1093/applin/16.3.371

33. Byrnes H. Sociocultural Theory and the Genesis of Second Language Development. James P. Lantolf and Steven L. Thorne. Oxford: Oxford University Press, 2006. Pp. xi + 398. $29.50 paper. Stud Second Lang Acquis. 2008;30(3):394-396. doi: https://doi.org/10.1017/s0272263108080546

34. Schmidt RW. The Role of Consciousness in Second Language Learning1. Applied Linguistics. 1990;11(2):129-158. doi: https://doi.org/10.1093/applin/11.2.129

35. Norris JM, Ortega L. Effectiveness of L2 Instruction: A Research Synthesis and Quantitative Meta-analysis. Language Learning. 2000;50(3):417-528. doi: https://doi.org/10.1111/0023-8333.00136

36. Spada N. Form-focussed instruction and second language acquisition: a review of classroom and laboratory research. Lang Teach. 1997;30(2):73-87. doi:10.1017/S0261444800012799

37. Nation ISP. Learning Vocabulary in Another Language.Cambridge, UK: Cambridge University Press; 2001. doi: https://doi.org/10.1017/cbo9781139524759

38. Deci EL, Ryan RM. The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior. Psychological Inquiry. 2000;11(4):227-268. doi: https://doi.org/10.1207/s15327965pli1104_01

39. Bandura A. Social cognitive theory: an agentic perspective. Annu Rev Psychol. 2001;52(1):1-26. doi:10.1146/annurev.psych.52.1.1

40. Holmes W, Bialik M, Fadel C. Artificial intelligence in education. Boston: Center for Curriculum Redesign: Globethics Publications; 2023. doi: https://doi.org/10.58863/20.500.12424/4276068

41. Lodge JM, Thompson K, Corrin L. Mapping out a research agenda for generative artificial intelligence in tertiary education. Australas J Educ Technol. 2023;39(2):1-8. doi:10.14742/ajet.8695

42. Wang Y. Cognitive and sociocultural dynamics of self-regulated use of machine translation and generative AI tools in academic EFL writing. System. 2024;126:103505. doi: https://doi.org/10.1016/j.system.2024.103505

43. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology. 2005;8(1):19-32. doi: https://doi.org/10.1080/1364557032000119616

44. Peters MD, Marnie C, Tricco AC, Pollock D, Munn Z, Alexander L, *et al*. Updated methodological guidance for the conduct of scoping reviews. JBI Evidence Synthesis. 2020;18(10):2119-2126. doi: https://doi.org/10.11124/jbies-20-00167

45. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, *et al*. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med. 2018;169(7):467-473. doi: https://doi.org/10.7326/m18-0850

Additional Information

How to Cite

APA 7th Edition: Kong, W. & Wu, Y. (2026). Generative AI in Korean Language Learning: A Scoping Review. Education and Learning, 2(2), 56-66. https://doi.org/10.58920/edu0202767

Vancouver: Kong W, Wu Y. Generative AI in Korean Language Learning: A Scoping Review. Education and Learning. 2026;2(2):56-66. https://doi.org/10.58920/edu0202767

Harvard: Kong, W. & Wu, Y. (2026) 'Generative AI in Korean Language Learning: A Scoping Review', Education and Learning, 2(2), pp. 56-66. doi: 10.58920/edu0202767

Publisher Note

All claims expressed in this article are solely those of the authors and do not necessarily reflect the views of the publisher, the editors, or the reviewers. Any product that may be evaluated in this article, or claim made by its manufacturer, is not guaranteed or endorsed by the publisher. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access

This article is licensed under a Creative Commons Attribution 4.0 International License. You may share and adapt the material with proper credit to the original author(s) and source, include a link to the license, and indicate if changes were made.