



Large Language Models in Traditional Korean Medicine: Current applications and linguistic implications*

Suji Lee^{**a,b} · Sejin Kim^{a,b,c} · Jong-In Kim^{a,b} · Seulkee Park^{***a} ·
Jong-Bok Kim^a
(Kyung Hee University^a · Kyung Hee University Medical Center^b · Korea University^c)

Lee, Suji, Sejin Kim, Jong-In Kim, Seulkee Park, and Jong-Bok Kim. 2026. **Large Language Models in Traditional Korean Medicine: Current applications and linguistic implications**. *Linguistic Research* 43(Special Edition): 341-371. This study reviews recent applications of Large Language Models (LLMs) in Traditional Korean Medicine (TKM) and examines their research implications from a linguistic perspective. A PRISMA-ScR, guided scoping review of 23 studies published between 2023 and 2025 shows that current research is heavily concentrated on education and clinical training, with limited direct clinical application. Technical efforts increasingly rely on retrieval-augmented generation and structured dialogue systems to mitigate knowledge gaps and hallucination. Beyond technical issues, TKM presents distinct linguistic challenges. Clinical discourse is multilingual, semantically dense, and highly context-dependent, often relying on metaphorical expressions and non-sentential responses. These features require pragmatic and discourse-level interpretation rather than simple information retrieval. The review suggests that effective LLM use in TKM depends on how clinical language is modeled and evaluated. We suggest that future research integrate linguistic insights, particularly from pragmatics, to enable safe and reliable deployment of LLMs in TKM. (Kyung Hee University · Kyung Hee University Medical Center · Korea University)

Keywords Traditional Korean Medicine, large language models, pragmatics, medical discourse, scoping review

* This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2025S1A5C3A02006302).

** First Author

*** Corresponding author

1. Introduction

Large Language Models (LLMs), including ChatGPT, have been rapidly adopted across a wide range of medical domains, such as clinical decision support, medical education, and the summarization of clinical records. Despite this rapid expansion, the extent to which these technologies have been applied within the field of Traditional Korean Medicine (TKM), as well as the developmental stage of the associated research, has not yet been systematically examined. TKM is characterized by domain-specific features, including distinctive pattern identification logic, extensive classical textual corpora, and clinically grounded narratives that rely heavily on metaphorical expression (S-Y. Park and C-E. Kim 2023; Hou et al. 2025). These characteristics impose clear limitations on the straightforward application of general-purpose LLMs (Kang et al. 2024b). To date, most discussions on LLMs in East Asian medicine have focused on Traditional Chinese Medicine (TCM) (T-H. Kim et al. 2023; B. Kang et al. 2024), leaving the research landscape specific to the medical, educational, and institutional context of TKM largely unmapped. Metaphors observed in TKM discourse should not be understood as mere rhetorical ornamentation or figurative comparison in the everyday sense. Rather, metaphor in clinical TKM interaction involves pragmatic interpretation processes through which speakers' intentions, presuppositions, and implicatures are inferred in context to construct meaning. Accordingly, metaphor in this domain constitutes a context-sensitive meaning-making process rather than a problem of surface-level information retrieval. Addressing such phenomena therefore requires a linguistic, particularly pragmatic—perspective that foregrounds inferential meaning, relevance, and speaker intention (Grice 1975; Gibbs 1994; Sperber and Wilson 2015).

Contemporary TKM education is situated at a paradigmatic crossroads, where the transmission of classical knowledge grounded in long-standing epistemological traditions must be integrated with diagnostic frameworks informed by modern biomedical rationality. Within this context, the ideal future practitioner in Korean medicine has been described as a professional equipped with medical bilingualism: the ability to navigate and mediate between TKM-specific pattern identification logic and biomedical indicators (Hanson and Kumar 2025). The emergence of ChatGPT-4o has further intensified discussions surrounding both the opportunities and limitations of generative artificial intelligence in complementary and alternative medicine, serving

as a catalyst for more systematic consideration of LLMs in TKM education (T-H. Kim et al. 2023). As leading medical schools worldwide increasingly incorporate AI competencies into core curricula, TKM education likewise faces the imperative to critically evaluate the potential of generative AI for learning support and clinical reasoning (S. Kim et al. 2025). From a linguistic perspective, TKM constitutes a prototypical low-resource domain within the LLM ecosystem. The scarcity of digitized corpora, combined with a multilingual structure in which Classical Chinese, Korean, and English coexist, generates structural challenges that extend beyond mere data insufficiency. These challenges manifest as imbalanced distributions across morphological, semantic, and discourse levels, forming a linguistic bottleneck for model training and evaluation (Bender 2011; Joshi et al. 2020).

Moreover, TKM knowledge is characterized by high semantic density, whereby multiple layers of meaning, symptoms, etiological assumptions, and diagnostic patterns, are compressed into relatively short expressions. Such density is not readily captured at the sentence level but instead emerges through discourse-level meaning construction, necessitating linguistic approaches that go beyond propositional analysis (Halliday 2004). Current LLMs, trained predominantly on English-centered corpora, exhibit structural limitations in bridging the gap between the condensed, Classical Chinese-based written style of canonical TKM texts and the colloquial, interactional language used in contemporary clinical settings. Empirical evidence underscores this limitation: a study evaluating ChatGPT's performance on the Korean National Licensing Examination for Korean Medicine Doctors reported higher scores when identical items were translated into English before being input to the model (D. Jang et al. 2023). This improvement does not necessarily reflect enhanced general reasoning ability. Rather, it suggests that the conceptual structures and discourse conventions presupposed by the input language modulate the model's semantic interpretation and pragmatic inference load. In other words, although the task remains constant, different languages impose distinct presuppositions, implicatures, and information-structural requirements, differences that current LLM evaluations tend to overlook by reducing outcomes to aggregate performance metrics. This observation calls for a linguistic re-examination of LLM evaluation practices, one that incorporates semantic appropriateness and pragmatic inference conditions alongside accuracy-based measures (Hovy and Spruit 2016; Bender and Koller 2020). Beyond concerns of performance validity, such linguistic asymmetries raise broader epistemic questions

regarding the potential simplification or distortion of East Asian medical knowledge when filtered through Western-centric data infrastructures (B. Kang et al. 2024; Mehta and Ring 2025).

Against this backdrop, the present study conducts a scoping review of 23 studies that have examined the use of LLMs in TKM education and clinical contexts. The aim is to systematically map the current research landscape of TKM-specific LLM applications, identifying their thematic distribution, methodological orientations, and technical characteristics. By doing so, this review seeks to illuminate research directions that can maximize the technological benefits of LLMs while preserving the epistemological and linguistic features of Korean medicine. Ultimately, the findings demonstrate that the central challenge of LLM utilization in TKM lies less in raw model performance than in how the structure and meaning of clinical language are represented, interpreted, and mediated, an inherently linguistic problem at its core.

2. Methods

2.1 Study design

To systematically examine the scope and characteristics of LLM-related research in the field of Traditional Korean Medicine (TKM), we conducted a scoping review. This review followed the scoping review methodology proposed by the Joanna Briggs Institute (JBI) and was carried out in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for scoping reviews (PRISMA-ScR) guidelines (Tricco et al. 2018). The study protocol was prospectively registered with the Open Science Framework (OSF; DOI: 10.17605/OSF.IO/RD6CS). To clearly define the scope of the review and formulate the core research questions, we adopted the Population-Concept-Context (PCC) framework. The Population comprised all scholarly literature related to TKM healthcare, education, and research. The Concept focused on the use of large language models and generative artificial intelligence technologies. The Context was defined as the clinical, educational, and research environments of TKM in Korea. This PCC framework guided the establishment of inclusion and exclusion criteria, the design of data extraction items, and the overall thematic synthesis process.

2.2 Search strategy

A comprehensive literature search was conducted using both international and Korean databases. For international sources, we searched MEDLINE via PubMed, Embase, and the Cochrane Library. Given the scoping review design, Google Scholar was additionally included to capture relevant gray literature. For domestic sources in Korea, KISS and OASIS, a specialized database for Traditional Korean Medicine (TKM), were searched. The search strategy was constructed by combining terms related to large language models (e.g., “large language model,” “LLM,” “ChatGPT,” “generative artificial intelligence”) with terms related to TKM (e.g., “traditional Korean medicine,” “Korean medicine,” “oriental medicine”). Search queries were adapted to the specific characteristics and indexing systems of each database. The full search strategy is provided in Supplementary File 1.

2.3 Eligibility criteria

The inclusion criteria were defined as follows:

1. studies that used or analyzed large language models (LLMs) or other generative language models as a core research tool, and
2. studies that explicitly addressed clinical, educational, or research contexts of Traditional Korean Medicine (TKM).

The exclusion criteria were:

1. general artificial intelligence studies not involving LLMs,
2. studies focusing exclusively on Traditional Chinese Medicine (TCM) without relevance to the TKM context, and
3. essays, news articles, commentaries, or other publications that had not undergone an academic peer-review process.

2.4 Study selection

All retrieved records were imported into a reference management software (EndNote 20; Clarivate, UK), and duplicate entries were identified and removed using the

software's automated duplicate-checking function. Two reviewers then independently screened titles and abstracts to identify studies that met the eligibility criteria. Subsequently, the full texts of the selected articles were reviewed in detail to determine final inclusion. Any disagreements between the reviewers during the selection process were resolved through discussion, with the involvement of a third reviewer when necessary to reach consensus.

2.5 Data extraction

To systematically extract data from the final set of included studies, a standardized data charting form was developed using Microsoft Excel 365 (Microsoft, USA). Data extraction was conducted independently by members of the research team, and any discrepancies were resolved through discussion. The extracted items included basic bibliographic information (first author, year of publication, and journal title) and key study characteristics, such as study design and type. In addition, technical aspects, including the type of LLM used and the methodological approach, were recorded, along with the role of the model, application domain, and target users. Finally, the main findings and outcome types reported in each study, as well as the limitations identified by the authors and areas requiring further improvement, were comprehensively collected.

2.6 Data synthesis

The extracted data were synthesized using a narrative synthesis approach in accordance with the research questions. To systematically organize the distribution and characteristics of LLM-related studies within the TKM domain, each study was classified along two dimensions: application domain and research focus.

Application domains were defined based on the study objectives and contexts reported in the literature and were categorized into educational pedagogy, clinical training and clinical practice, and technical foundation. Research focus was classified into four stages, conceptual, perception, development, and performance comparison and validation (benchmarking), according to the primary aims and methodological orientations of the studies. This classification scheme is grounded in prior scoping

reviews and conceptual models of medical AI research, which describe a progression from initial exploration and awareness to system development and outcome validation (Gordon et al. 2024).

Using this analytical framework, the present study visualized the distribution of LLM research in TKM through a thematic evidence map structured around application domain and research focus, thereby providing an overview of the current research landscape in TKM education and clinical contexts.

3. Results

3.1 Study selection

A total of 163 records were identified through searches across six international and Korean databases. After removing 40 duplicate records using reference management software, 123 articles were screened based on titles and abstracts. Two reviewers independently conducted the screening, resulting in the identification of 36 articles deemed potentially relevant and thus eligible for full-text review.

Following full-text assessment of these 36 articles, 12 were excluded for not meeting the final inclusion criteria. The primary reasons for exclusion were lack of relevance to the TKM domain or context ($n=6$), inappropriate use or role of LLMs ($n=3$), and misalignment in study design or research focus ($n=3$). During the review process, cases were identified in which substantially identical study content had been reported both as a preprint (D. Jang and C-E. Kim 2023) and as a peer-reviewed journal article (D. Jang et al. 2023); these were treated as a single study. As a result, a total of 23 independent studies were included in the final analysis (Figure 1).

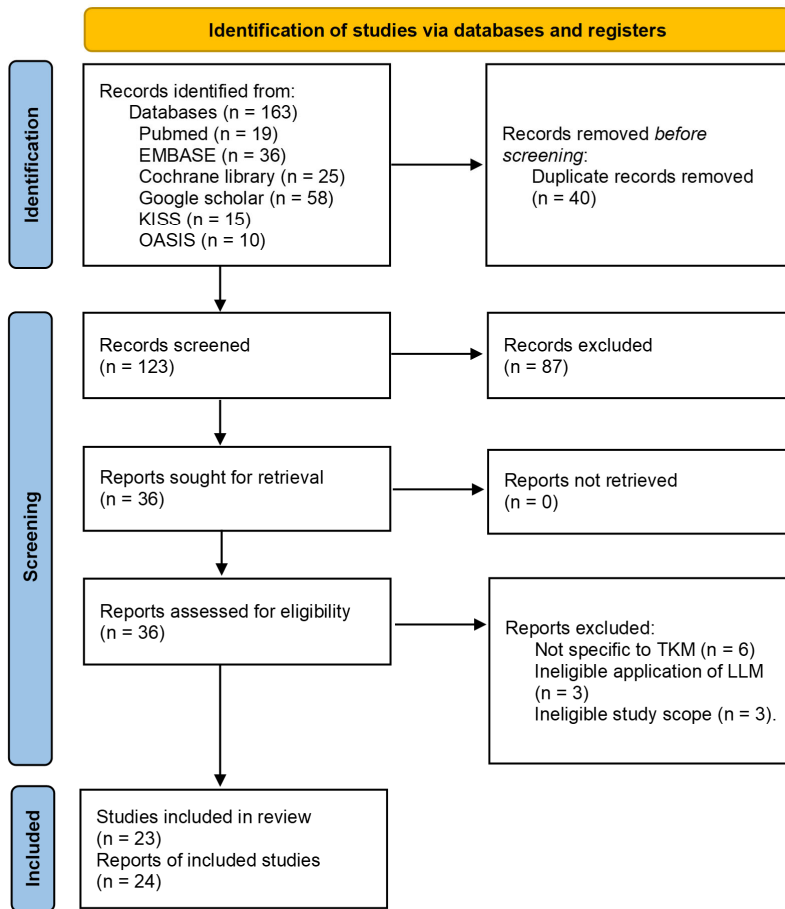


Figure 1. PRISMA flow diagram

3.2 Study characteristics

The characteristics of the 23 studies included in the final analysis are summarized in Table 1. The publication years ranged from 2023 to 2025, with a clear year-by-year increase in the number of studies, reflecting the rapidly growing interest in the application of LLMs within the TKM field.

Examination of the temporal distribution shows that three studies (13.0%) were

published in 2023, the year in which generative AI technologies began to be introduced in earnest into TKM education. Research during this initial phase primarily consisted of exploratory studies aimed at examining the academic feasibility of LLMs and their capacity for basic knowledge retrieval, such as evaluations of LLM performance on the Korean National Licensing Examination for Korean Medicine Doctors (D. Jang et al. 2023) and assessments of foundational competencies related to acupuncture prescription (H. Lee 2023).

In 2024, nine independent studies (39.1%) were published, marking a shift in research focus toward system implementation at the level of specific subjects or functional units. Studies conducted during this period concentrated on applications such as suicide risk assessment training (C-Y. Kwon 2024; J. Ryu et al. 2024), pattern identification practice systems (W-Y. Lee et al. 2024), and the automated classification and normalization of TKM literature (H. Jang 2024a, b). This trend indicates a transition toward structuring generative AI technologies as practical educational and clinical support tools. The patient-practitioner interactions and clinical scenarios addressed in these studies suggest that model performance is closely tied not to descriptions of social context per se, but to the accuracy of pragmatic language use, including the appropriateness of questions, the inferential adequacy of responses, and the handling of presuppositions and implicatures. This observation highlights the need for linguistically informed analyses.

In 2025, nearly half of the included studies ($n=11$, 47.8%) were published, confirming a further acceleration of research activity at the intersection of TKM and AI. During this period, empirical studies aiming to demonstrate objective effects in educational and clinical settings became predominant. These included evaluations of the non-inferiority of AI chatbot simulations in educational contexts (H-Y. Lee et al. 2025), the application of multimodal RAG-based educational systems (S. Seo et al. 2025), and analyses of correlations between AI use and learning outcomes within formal curricula (H. Lee and Y. Kim 2025). However, these effectiveness-oriented studies largely focused on learning achievement or system-level accuracy, while providing relatively limited analysis of how pragmatic factors, such as speaker intention, indirectness, and inferential load in clinical discourse, contribute to or constrain model performance. Overall, whereas early studies emphasized the technical feasibility of generative AI, later research shows a gradual progression toward validating the effects of specific educational scenarios or clinical applications.

From a technical perspective, there was a clear transition in the models used, shifting rapidly from GPT-3.5 to GPT-4o or GPT-5o. In addition, studies reporting the qualitative strengths of models such as Claude-3.5-Sonnet began to emerge (Kim et al. 2024), indicating increasing model diversification. To address hallucination issues arising from the characteristics of TKM knowledge, researchers moved beyond prompt engineering alone and increasingly adopted retrieval-augmented generation (RAG) architectures that integrate external knowledge bases (B. Kang et al. 2024; W-Y. Lee et al. 2025) as well as multimodal approaches combining textual and visual information (S. Seo et al. 2025). These developments point to substantial technical advancement. Despite such progress, the extent to which clinical dialogue success depends on pragmatic language processing, such as contextual interpretation, implicature resolution, and response appropriateness, rather than on social role relations, has not been systematically addressed in the existing literature.

Table 1. Characteristics of included studies

Study ID	First author (year)	Study design	LLM model	Application domain	Main result	Limitation	Linguistic focus
1 [24]	Choi (2025)	Course design and educational implementation study	Generative LLMs (not specified)	Educational pedagogy (Donguibogam curriculum; flipped learning)	LLM-supported flipped learning facilitated self-directed study and in-class discussion in classical TKM education	No objective learning outcome comparison; single course context	Summarization and explanation of classical medical texts; instructional dialogue
2 [25]	Kim (2025)	Conceptual framework and pilot application	Not applicable (conceptual)	Clinical training (acupuncture education)	Proposed AI-assisted acupuncture education framework integrating simulation and feedback	Lack of empirical validation	Procedural explanation language; instructional prompting
3 [26]	Kim (2025)	Survey-based needs assessment	Generative LLMs (not specified)	Educational pedagogy (AI literacy; anatomy education)	Students showed high interest in AI-assisted anatomy learning but limited	Self-reported data; no intervention	Information-seeking and explanatory language

					structured exposure		
4 [27]	Kim (2025)	Development and educational application study	Generative LLM (ChatGPT; version not specified)	Educational pedagogy (TKM-specific chatbot for education)	Demonstrated feasibility of TKM-specialized chatbot for student learning support	No learning outcome evaluation	Domain-adapted explanatory dialogue; terminology translation
5 [20]	Lee (2025)	Comparative educational evaluation	Generative LLM (ChatGPT; version not specified)	Educational pedagogy (TKM history education)	LLM-assisted flipped learning improved engagement and comprehension	Limited sample size; subjective outcomes	Narrative summarization; historical explanation language
6 [18]	Lee (2025)	Pilot randomized controlled trial	Generative LLMs (ChatGPT; GPT-4o, Claude 3.5)	Clinical training (OSCE preparation)	AI chatbot simulation was comparable to peer role-play for OSCE practice	Small sample; short intervention	Simulated patient dialogue; structured interview language
7 [22]	Lee (2025)	System development and validation study	Generative LLM (ChatGPT; GPT-4o)	Clinical training (syndrome differentiation & diagnosis)	Gen-SynDi enabled dual training in syndrome differentiation and disease diagnosis	Limited clinical scenario diversity	Diagnostic reasoning language; structured pattern discourse
8 [28]	Park (2025)	Case-based educational implementation	Generative LLMs (ChatGPT; GPT-4o, Claude)	Clinical training (personalized TKM clinical practice education)	AI-supported simulations enhanced realism and learner engagement	No objective competency metrics	Case-based clinical dialogue; adaptive questioning
9 [19]	Seo (2025)	System development study	Schema-based LLM with RAG	Technical foundation (standardized patient system; CURA)	Developed schema-driven standardized patient platform for TKM	No downstream educational or clinical outcome	Schema-constrained dialogue; structured clinical language
10 [29]	Shin (2025)	Cross-sectional survey	Generative LLMs (not specified)	Educational pedagogy (student perception; AI literacy)	High awareness but low trust for TKM-specific learning tasks	Single institution; self-reported	Meta-linguistic evaluation of trust and accuracy
11 [30]	Yoon (2025)	Comparative performance study	Generative LLM (ChatGPT; GPT-4)	Clinical practice (acupuncture treatment planning)	GPT-4 showed partial overlap with clinicians in acupoint selection	Text-only cases; no outcome validation	Prescriptive clinical reasoning language

12 [16]	Jang (2024)	Text classification and retrieval study	OpenAI text-embedding-ada-002, BERT	Technical foundation (semantic embedding for TKM literature)	LLM embeddings improved disease-prescription document classification	Non-generative; no application validation	Semantic representation and similarity language
13 [17]	Jang (2024)	Named entity normalization study	BERT, Generative LLMs (ChatGPT; GPT-4o, GPT-4o mini)	Technical foundation (prescription name normalization)	LLM outperformed traditional ML in entity equivalence judgment	Black-box decision process	Context-sensitive semantic equivalence language
14 [23]	Kang (2024)	Methodological comparative study	Generative LLMs (ChatGPT; GPT-4, GPT-3.5)	Technical foundation (Prompt-RAG for TKM)	Prompt-RAG outperformed embedding-based RAG in TKM QA	Limited corpus; higher cost	Discourse-level retrieval and reference-constrained generation
15 [21]	Kim (2024)	Development & evaluation study	Generative LLMs (ChatGPT; GPT-4o, Claude-3.5)	Clinical training (CPX practice; automated scoring & feedback)	Two-stage LLM framework enabled accurate CPX scoring and feedback	Single scenario; simulated data	Evaluative rubric-based language; feedback meta-language
16 [31]	Kim (2024)	Pilot development study	Generative LLM (ChatGPT; GPT-4)	Clinical training (CPX simulation chatbot)	Prompt-engineered chatbot feasible for CPX practice	No learner outcome validation	Role-conditioned simulated patient dialogue
17 [13]	Kwon (2024)	Educational intervention study	Generative LLM (Claude-3)	Clinical training (TKM neuropsychiatry; suicide risk assessment)	AI simulation improved learner confidence in high-risk interviews	Subjective outcomes only	Empathy-trained, safety-aware clinical dialogue
18 [32]	Lee (2024)	Web-based expert survey	Generative LLM (ChatGPT; GPT-4)	Clinical practice (facial palsy care)	Experts found limited reliability for direct TKM clinical use	Perception-based; no clinical outcomes	Patient-facing vs clinician-facing explanatory language
19 [15]	Lee (2024)	Feasibility and development study	Generative LLM (ChatGPT; version not specified)	Clinical training (pattern identification education)	Simulated patient enabled structured pattern-differentiation practice	Single pattern; no performance assessment	Pattern-constrained diagnostic dialogue
20 [14]	Ryu (2024)	Development and pilot application study	Generative LLM (ChatGPT; GPT-4o-mini)	Clinical training (CPX-based TKM simulation)	High student satisfaction and perceived skill improvement	Self-reported outcomes; single scenario	High-stakes simulated patient communication

21 [7]	Jang (2023)	Benchmark capability evaluation	Generative LLMs (ChatGPT; GPT-4, GPT-3.5)	Technical foundation (LLM capability in TKM domain)	GPT-4 passed TKM licensing exam benchmark without fine-tuning	Exam-based; no clinical reasoning depth	Cross-lingual and culturally localized medical language
22 [12]	Lee (2023)	Comparative educational study	Generative LLMs (ChatGPT; GPT-4, GPT-3.5)	Educational pedagogy (acupuncture learning tool)	Discrepancies with case reports framed as learning opportunities	No direct learner outcome measurement	Exploratory, non-authorita- tive suggestion language
23 [1]	Park (2023)	Conceptual and narrative review	Not applicable (conceptual)	Educational pedagogy (educational AI strategy)	Proposed strategic directions for AI-enhanced TKM education	No empirical validation	Meta-educatio- nal explanatory discourse

AI, artificial intelligence; BERT, bert-base-uncased; CPX, clinical performance examination; GPT, generative pretrained transformer; TKM, traditional korean medicine; LLM, large language model; ML, machine learning; OSCE, objective structured clinical examination; QA, question answering; RAG, retrieval-augmented generation.

3.3 Data synthesis

Analysis of the selected studies revealed that the impacts of generative AI on TKM education and clinical preparation can be grouped into three core themes: (1) AI-based clinical simulation using virtual patients, (2) support for knowledge acquisition and self-directed learning, and (3) development of TKM-specific knowledge infrastructure. These themes showed distinct distributional patterns depending on application domain and research focus (Figure 2).

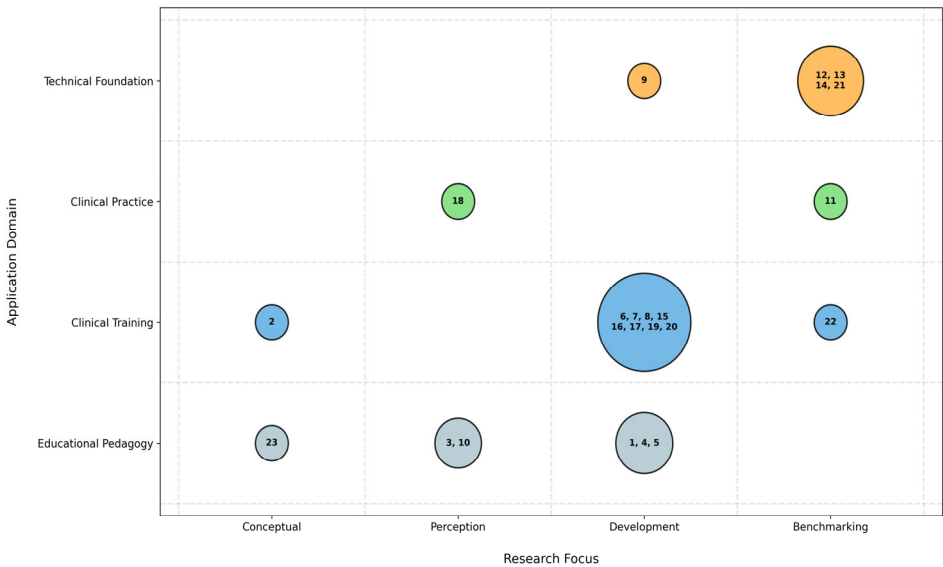


Figure 2. Mapping of large language model research in Traditional Korean Medicine by application domain and research focus

The map illustrates the distribution of the 23 included studies across four application domains and four research focus areas. Bubble size represents the density of studies in each cell, with individual study IDs labeled for reference.

3.3.1 AI-based clinical simulation

The most actively studied area involved training for the Clinical Performance Examination (CPX) and the Objective Structured Clinical Examination (OSCE) using virtual standardized patients. Studies in this category were predominantly situated within the clinical training domain and were largely concentrated at the development stage. Early studies focused primarily on implementing simulated patient personas through prompt engineering (J. Kim et al. 2024a; C-Y. Kwon 2024), whereas more recent work has evolved into integrated educational platforms that provide real-time feedback to learners and automated scoring functions (J. Kim et al. 2024b; J. Ryu et al. 2024).

Randomized controlled trials demonstrated that AI chatbot-based simulations were non-inferior to traditional peer role-play in terms of educational effectiveness (H-Y. Lee et al. 2025). In addition, in high-risk interview training contexts-such as suicide risk assessment-AI-based simulations were positively evaluated for offering emotional safety and opportunities for repeated practice (C-Y. Kwon 2024; J. Ryu et al. 2024). These findings suggest that virtual patient-based simulations constitute a particularly promising application of generative AI in TKM clinical education.

3.3.2 Support for knowledge acquisition and self-directed learning

Studies addressing support for knowledge acquisition and self-directed learning were predominantly situated within the educational pedagogy domain, with research foci mainly corresponding to the perception or development stages. Research on classical text learning, such as instruction based on Donguibogam (J. Choi and Y. Kim 2025), and analyses of learning outcomes in Korean medical history education (H. Lee and Y. Kim 2025) reported positive correlations between the extent of AI utilization and academic performance. In addition, students expressed a strong willingness to use AI systems as personal tutors in foundational medical subjects, including anatomy (S-K. Kim et al. 2025).

Overall, these studies primarily documented patterns of AI use during the learning process, learners' perceptions and attitudes toward AI-assisted instruction, and associations between AI engagement and educational outcome measures, rather than providing detailed analyses of underlying cognitive or linguistic mechanisms.

3.3.3 TKM-specific knowledge engineering and optimization

Studies focusing on the development of TKM-specific knowledge infrastructure were primarily situated within the technical foundation domain and corresponded to the development and benchmarking stages. These studies aimed to adapt the general-purpose capabilities of LLMs to the specific characteristics of the TKM domain through targeted technical strategies. Representative approaches included fine-tuning models using TKM-specific datasets (Y. Kim 2025) and integrating retrieval-augmented generation (RAG) architectures to mitigate hallucination by grounding model outputs in external knowledge bases (S-Y. Park and C-E. Kim 2023; B. Kang et al. 2024).

In particular, research on prescription name normalization (H. Jang 2024b) and automatic classification of herbal formula literature (H. Jang 2024a) contributed to transforming fragmented traditional medical knowledge into machine-readable and learnable structures. Furthermore, multimodal approaches that combined textual data with visual information, such as tongue diagnosis images or thermographic indicators, expanded the potential for digital transformation in TKM diagnostic education (S. Seo et al. 2025).

3.3.4 Linguistic perspective

A linguistic review of the selected studies indicates that the language processed by generative AI in TKM education and clinical preparation differs substantially from the explicit question–answer structures typically observed in Western biomedical discourse. Prior research shows that Western medical discourse is largely organized around information-oriented exchanges aimed at symptom verification and diagnostic decision-making, whereas East Asian traditional medical discourse places greater emphasis on interpretive reasoning grounded in patients’ experiential narratives (Mishler 1984; Heritage and Maynard 2006).

In particular, TKM clinical discourse frequently conceptualizes symptoms and bodily states through metaphorical expressions. Such metaphors function not merely as rhetorical devices but as core inputs to diagnostic reasoning. Expressions such as “qi is blocked” or “heat is accumulating” are grounded in conceptual metaphors that structure bodily experience in abstract terms; their interpretation depends on pragmatic context and shared domain-specific knowledge (Lakoff and Johnson 1980, 2008; Yu 2009; Kövecses 2010). Across multiple studies included in this review, metaphorical symptom descriptions were found to play a central role in interactions within clinical simulations and educational chatbots.

TKM clinical interactions also exhibit frequent use of non-sentential utterances in question-answer pairs. Patient responses are often limited to noun phrases or adverbial phrases, with interpretation relying heavily on the preceding question and the broader discourse context. These utterances are not treated as grammatically deficient but have been analyzed as pragmatically licensed phenomena motivated by discourse economy and inferential recoverability (Ginzburg and Sag 2000; Merchant 2005; Culicover and Jackendoff 2005). For generative AI systems to process TKM

clinical discourse appropriately, such context-dependent mechanisms of meaning construction must therefore be taken into account.

Taken together, the linguistic challenges faced by generative AI in the TKM domain extend beyond analyses of social role relations to encompass pragmatic processes such as metaphor interpretation, discourse-level reconstruction of non-sentential utterances, and context-based inference. This suggests that the application of LLMs in TKM is not simply a matter of sentence generation or information retrieval, but rather requires more refined modeling that reflects the semantic and pragmatic properties of language use in TKM clinical and educational contexts.

4. Discussion

This study systematically mapped the research landscape of LLM applications in TKM education and clinical contexts by analyzing 23 studies published between 2023 and 2025. As shown in Figure 2, AI-related research in TKM has been most heavily concentrated in the education and training domain (69.6%), followed by technical foundations (21.7%) and clinical practice (8.7%). Early studies largely remained exploratory, focusing on whether LLMs could retrieve or reproduce TKM knowledge to a certain degree. More recent research, however, has rapidly evolved toward higher levels of sophistication, including clinical skills training using virtual standardized patients, automated feedback and assessment systems, and the construction of TKM-specific knowledge infrastructures based on retrieval-augmented generation (RAG). This trajectory suggests that LLM research in TKM is moving beyond proof-of-concept demonstrations toward structural integration into educational and clinical processes (M. Gordon et al. 2024).

LLM research in TKM also shows clear contrasts with patterns observed in conventional medicine. In biomedical contexts, LLMs are typically applied to well-structured data such as clinical guidelines and electronic health records, where they excel at rule-based reasoning within explicit diagnostic and treatment algorithms (Rajpurkar et al. 2022; Marinelli et al. 2025). By contrast, TKM clinical reasoning involves a process of semantic synthesis, in which clinicians interpret patients' subjective and unstructured complaints to derive a higher-level diagnostic construct, namely pattern identification. This process can be understood as a fundamentally

linguistic operation that integrates contextual interpretation and meaning construction. In TKM clinical discourse, metaphorical symptom expressions, non-sentential utterances, and inferences based on presupposition and implicature function as key diagnostic cues, phenomena that are difficult to account for without pragmatic interpretation and semantic structuring (Grice 1975; Lakoff and Johnson 1980; Levinson 1983). Consistent with our findings, TKM clinical discourse prioritizes the organic integration of fragmented symptoms rather than the simple accumulation of discrete information units. Consequently, LLMs in the TKM domain are required not merely to retrieve medical knowledge, but to extract and integrate contextual meaning from highly unstructured text, demanding a more advanced level of linguistic inference (Hou et al. 2025). This linguistic complexity further explains why hybrid approaches, such as RAG architectures and knowledge graph-based systems, have emerged as essential technical trends in TKM-specific LLM research. These differences are attributable not only to data structure but also to discourse organization. While Western biomedical discourse typically follows explicit question-answer conventions oriented toward information verification, TKM discourse tends to construct meaning through patients' experiential narratives and implicit expressions (Mishler 1984; Heritage and Maynard 2006). Accordingly, LLM applications in TKM must prioritize pragmatic inference and discourse-sensitive meaning reconstruction over purely rule-based information processing.

In comparison with TKM, research in Traditional Chinese Medicine (TCM) has pursued large-scale, domain-specific foundation models supported by substantial capital and data resources, with several systems integrating knowledge graphs and LLMs advancing to clinical application stages (Chen et al. 2025; Hou et al. 2025). TKM research, by contrast, has emphasized linguistic precision and pedagogical utility. Efforts to activate knowledge through code-switching and to normalize terminology within a medical bilingualism environment, where Korean, Classical Chinese, and English coexist, represent a distinctive characteristic of TKM-oriented approaches.

A major gap identified in this review is the limited engagement with real-world clinical settings. More than half of the analyzed studies focused on educational contexts, while relatively few directly addressed patient–physician interactions or unstructured clinical records from actual practice. This limitation is closely related to the narrative-heavy and weakly standardized nature of TKM clinical data. In addition, concerns regarding the opacity of AI decision-making and associated legal

and ethical responsibilities remain significant constraints in this domain (Marinelli et al. 2025). As a result, most current research has prioritized validation in lower-risk educational environments. These limitations are not merely technical; they are closely tied to the difficulty of ensuring pragmatic safety in clinical discourse. In TKM consultations, incomplete utterances and metaphorical expressions are common, and misinterpretation or excessive explicitization of such language may lead to diagnostic errors or inappropriate recommendations. This highlights the need for generative AI systems to operate under refined constraints on speech act types and implicature interpretation, rather than relying solely on surface-level accuracy metrics (Ginzburg and Sag 2000; Culicover and Jackendoff 2005).

In conclusion, LLMs in the TKM domain have emerged as key drivers of educational innovation and technical infrastructure development, extending beyond their role as simple knowledge delivery tools. Although substantial challenges remain for direct clinical application, advances in semantic representation and medical bilingualism processing are likely to help bridge this gap. Future research should move beyond educational validation toward large-scale, multi-institutional benchmarking datasets and the establishment of technical safeguards and ethical guidelines capable of handling the complexity of real clinical discourse. Achieving these goals will require not only technical infrastructure but also linguistically informed annotation schemes and benchmark resources that reflect the distinctive properties of TKM clinical language. In particular, language resources that systematically encode pragmatic features, such as metaphorical symptom descriptions, non-sentential responses, and speech act types, will be essential foundations for the safe and reliable clinical deployment of TKM-specific LLMs (Joshi et al. 2020; Bender 2011).

5. Linguistic implications

The findings of this scoping review indicate that the effective use of large language models (LLMs) in Traditional Korean Medicine (TKM) depends not only on technical performance but also on how well the linguistic properties of TKM discourse are modeled. In particular, several pragmatic phenomena that are central to TKM clinical interaction remain underexplored in current LLM-based approaches. Accordingly, this section illustrates four such phenomena to highlight key linguistic challenges for LLM

applications in TKM: (i) metaphorical symptom descriptions, (ii) non-sentential utterances in question-answer pairs, (iii) indirectness and circumlocution in symptom reporting, and (iv) presupposition and implicature in diagnostic interaction.

5.1 Metaphorical symptom descriptions

A central linguistic characteristic of TKM clinical discourse is the pervasive use of metaphorical expressions in symptom description. These metaphors are not merely stylistic expressions; they serve as diagnostically meaningful inputs that require pragmatic interpretation rather than literal understanding. For example, in TKM consultations and simulation scenarios, patients frequently describe their condition using metaphorical expressions, as follows:

- (1) a. Kiwun-i mak-hi-n kes kath-a-yo.
 energy-NOM block-PASS-ADN thing seem-DECL-POL
 ‘It feels like my qi is blocked.’
- b. Sok-ey yel-i kyeysoK ssahi-nun
 inside-LOC heat-NOM continuously accumulate-ADN
 nukkim-i-eyyo.
 feeling-COP-POL
 ‘I feel like heat is accumulating inside.’

These utterances instantiate conceptual metaphors that map embodied experiences onto abstract diagnostic dimensions (Lakoff and Johnson 1980, 2008). Their interpretation depends on shared domain knowledge and contextual assumptions regarding bodily processes, rather than on compositional semantics alone. From a pragmatic perspective, such expressions invite the clinician or the AI system to infer clinically relevant implications, such as stagnation or excess patterns, that are not explicitly stated (Gibbs 1994; Yu 2009).

Several studies reviewed in this paper reported that LLM-based systems tended to either literalize these expressions or paraphrase them into biomedical terminology without preserving their diagnostic function, indicating a gap in pragmatic modeling. This suggests that future TKM-specific LLMs must treat metaphor interpretation as

a core inferential task, not as a peripheral natural language understanding problem.

5.2 Non-sentential utterances in question-answer pairs

Another salient feature of TKM clinical interaction is the frequent occurrence of non-sentential utterances (NSUs) in question-answer sequences. Patient responses are often fragmentary yet pragmatically complete, relying on discourse context for interpretation. Such fragmentary responses are facilitated by the clinician's questioning style in TKM, which is oriented toward fine-grained understanding of the patient's subjective bodily experience rather than eliciting fully propositional answers, thereby creating interactional conditions in which non-sentential utterances are pragmatically natural. Consider the following exchange pattern, commonly observed in simulated patient interactions used in CPX and OSCE training:

- (2) Doctor: Encey-pwuthe sok-i taptap-ha-ko
 when-from inside-NOM stuffy-be-CONJ
 kapkap-ha-sy-ess-na-yo?
 tight-be-HON-PST-Q-POL
 'Since when have you been feeling stuffy and constricted in your stomach?'
 Patient: Han tal ccum-yo.
 One month about-POL.
 'About a month.'

In (2), the response *hal dal ccumyo* 'about a month' constitutes a noun phrase rather than a full sentence, yet it is fully interpretable due to its dependence on the preceding question for semantic completion. Such utterances have been extensively analyzed in linguistic theory as pragmatically licensed forms that maximize discourse efficiency (Ginzburg and Sag 2000; Culicover and Jackendoff 2005).

Several reviewed studies involving AI-based clinical simulations implicitly relied on this interactional pattern, but none explicitly addressed how LLMs recover the missing propositional structure or evaluate the pragmatic adequacy of such responses. This omission is non-trivial: failure to properly interpret NSUs may lead to

inappropriate follow-up questions or incorrect diagnostic inferences. Consequently, NSU handling should be recognized as a core discourse competence for LLMs operating in TKM clinical contexts.

5.3 Indirectness and circumlocution in symptom reporting

TKM discourse also exhibits a strong tendency toward indirectness and circumlocution, particularly in the expression of discomfort or uncertainty. Rather than explicitly stating symptoms, patients often employ mitigated or hedged formulations that require pragmatic enrichment. Consider the following conversation between a TKM doctor and an elderly female patient:

- (3) Doctor: Yak mek-ki himtul-kena kule-ci-n
 medicine take-MOD difficult-OR such-COMP-TOP
 anh-u-sy-ess-cyo?
 not-do-HON-PST-QUE
 ‘Wasn’t it difficult for you to take the medicine?’
- Patient: Yak-i com mingmingha-ko mwul kath-a.
 medicine-NOM a.bit clear-COMP water seem-DECL
 Amwu mas-to eps-tay?
 any taste-FOC not.exist-REPORT
 ‘The medicine is kind of bland and tastes almost like water.
 It doesn’t taste like anything.’
- Doctor: Emeni, yak-ul sey-key hay-se ciu-myen
 mother medicine-ACC strong-ADV do-CONN make-if
 sok-i ta mangka-cye-yo.
 stomach-NOM all ruin-become-DECL
 ‘Ma’am, if we make the medicine too strong, your stomach will
 be completely upset.’

In (3), the patient does not state her concern about the medication in a direct or explicit way. Rather than saying that the medicine is ineffective or causing discomfort, she comments on its sensory qualities, such as its smell and lack of taste, and cautiously

suggests that it seems weak. This indirect way of speaking allows the patient to express doubt about the treatment while avoiding direct criticism or confrontation. The concern is conveyed through implication rather than overt assertion. Importantly, the TKM doctor is able to recognize the communicative intention behind this indirect expression. By immediately addressing the possibility that the medicine was made too strong, the doctor shows an understanding that the patient is implicitly questioning whether the herbal ingredients were used in sufficient quantity. This illustrates that meaning in TKM clinical interaction is often constructed through pragmatic inference, where the clinician interprets what the patient is suggesting rather than what is explicitly said. Such expressions represent scalar judgments and epistemic stance rather than discrete symptom categories. From a pragmatic perspective, they involve implicature and pragmatic mitigation that reflect sociocultural norms of medical interaction (Grice 1975; Levinson 1983). LLMs focusing on explicit symptom extraction may misinterpret these utterances as clinically insignificant, thereby underestimating their diagnostic relevance.

The reviewed literature suggests that current AI systems tend to regularize such indirect expressions into categorical labels, potentially failing to preserve clinically meaningful gradience. This indicates the need for pragmatically informed representations that preserve degrees of certainty, intensity, and speaker stance.

5.4 Presupposition and implicature in diagnostic interaction

Many clinically relevant inferences are conveyed indirectly, relying on what is taken for granted in the interaction or what is pragmatically implicated rather than overtly asserted. The following turn sequence illustrates a typical interaction observed during ongoing acupuncture treatment in TKM clinical practice.

- (4) Patient: Nwun-to com chimchim-hay-yo.
 eye-even a.bit dim-be-POL
 ‘My eyes also feel somewhat dim.’
 Doctor: (inserting acupuncture needles)
 Elkwul-ey-to cokum nwa-tuli-lkkey-yo.
 face-LOC-even a.little place-give-FUT-POL

‘I’ll apply acupuncture to your face as well.’

In (4), the patient’s utterance presupposes the existence of an ongoing treatment context and an already established symptom trajectory. The use of *-to* ‘also’ signals that visual discomfort is being added to a set of previously discussed complaints, rather than introduced as a new, isolated symptom. Crucially, the patient does not explicitly request a change in treatment; instead, the utterance pragmatically implicates a relevance relation between the reported symptom and the ongoing acupuncture procedure. The doctor’s response demonstrates sensitivity to this presuppositional and implicature-based cue. By immediately adjusting the treatment site, for example, ‘I’ll apply acupuncture to your face as well’, the clinician treats the patient’s remark as diagnostically and therapeutically actionable, without requiring further clarification. The diagnostic relevance thus lies not in an explicit causal statement, but in the shared pragmatic inference that links symptom reporting to treatment adjustment.

For an LLM, failure to track such presuppositional additions and implied requests may result in missed opportunities for appropriate intervention or overly literal interpretations of symptom statements. From a Gricean perspective, the interaction relies on cooperative principles of relevance and informativeness, where clinically meaningful intentions are conveyed indirectly rather than overtly stated (Grice 1975; Levinson 1983). In TKM diagnostic interaction, these implicated meanings often carry greater practical significance than explicit symptom labels, as they reflect patients’ expectations about ongoing treatment and bodily response.

Several studies reviewed in this paper involved AI-based clinical simulations in which learner performance was evaluated based on the appropriateness of follow-up questions. However, these studies rarely foregrounded whether AI systems were able to systematically track presuppositional commitments or recover implicated meanings as explicit objects of analysis. This may lead to redundant questioning or premature diagnostic closure. These observations suggest the need for linguistically informed evaluation frameworks that explicitly incorporate presupposition tracking and implicature-sensitive inference in TKM-oriented LLM systems.

The observations imply that the primary linguistic challenges for LLMs in TKM lie not at the level of surface linguistic form, but at the level of pragmatic inference and contextual interpretation. Metaphorical expressions, non-sentential utterances, and indirect symptom descriptions all require context-sensitive interpretation grounded

in shared medical and cultural knowledge. From a methodological perspective, this points to several concrete directions for future research:

1. the need for linguistically annotated TKM corpora that explicitly encode pragmatic features such as metaphor type, NSU structure, and speech act function;
2. analytic frameworks examining pragmatic felicity rather than surface-level accuracy alone;
3. hybrid modeling approaches in which LLMs are constrained or guided by discourse-level and domain-specific representations.

By accounting for these linguistic approaches, future work can move toward LLM systems that are not only technically proficient but also pragmatically reliable in TKM educational and clinical settings.

6. Conclusion

This scoping review mapped the current landscape of large language model (LLM) research in Traditional Korean Medicine (TKM) education and clinical training, synthesizing 23 studies published between 2023 and 2025. The findings demonstrate a rapid expansion of LLM applications, particularly in educational and training contexts, with growing technical sophistication through simulation-based learning, automated feedback systems, and retrieval-augmented generation architectures.

At the same time, this review highlights a persistent gap between technical performance and linguistic adequacy in TKM-oriented LLM research. Unlike conventional biomedical discourse, TKM clinical interaction relies heavily on pragmatic inference, including metaphorical symptom conceptualization, non-sentential question–answer pairs, indirect expression, and inference driven by presupposition and implicature. These features are not peripheral stylistic choices but constitute core mechanisms of diagnostic reasoning and clinical meaning construction. Current evaluation frameworks, however, tend to reduce model performance to accuracy or task completion, without adequately accounting for these discourse-level demands.

From a linguistic perspective, the central challenge of LLM deployment in TKM

lies not only in improving model capacity, but in accurately modeling how meaning is constructed, inferred, and negotiated in clinical discourse. Addressing this challenge requires linguistically informed annotation schemes, benchmark datasets that encode pragmatic phenomena, and evaluation criteria sensitive to semantic and pragmatic appropriateness rather than surface correctness alone.

In conclusion, the effective and responsible integration of LLMs into TKM education and clinical practice depends on interdisciplinary collaboration between medical researchers, AI developers, and linguists. By foregrounding the linguistic structure of TKM discourse, future research can move beyond technology-centered optimization toward models that more faithfully mediate between traditional medical knowledge and contemporary clinical reasoning.

References

- Bender, Emily M. 2011. On achieving and evaluating language-independence in NLP. *Linguistic Issues in Language Technology* 6(3): 1-26.
- Bender, Emily M. and Alexander Koller. 2020. Climbing towards NLU: On meaning, form, and understanding in the age of data. In Dan Jurafsky, Joyce Chai, Natalie Schluter, Joel Tetreault (eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5185-5198. Online. Association for Computational Linguistics.
- Chen, Zijia, Wenxi Peng, Dezheng Zhang, Xin Liu, and Zhifei Wang. 2025. Application, challenges, and prospects of Large Language Model in the field of Traditional Chinese Medicine. *Medical Journal of Peking Union Medical College Hospital* 16(1): 83-89. <https://doi.org/10.12290/xhyxzz.2024-0315>.
- Choi, Jeongbin and Yongjin Kim. 2025. Course design and instructional strategies for Donguibogam using flipped learning-based LLM. *Journal of Korean Medical Classics* 38(1): 37-50. <https://doi.org/http://dx.doi.org/10.14369/jkmc.2025.38.1.037>.
- Culicover, Peter W. and Ray Jackendoff. 2005. *Simpler syntax*. Oxford: Oxford University Press.
- Gibbs, Raymond W. 1994. *The poetics of mind: Figurative thought, language, and understanding*. Cambridge: Cambridge University Press.
- Ginzburg, Jonathan and Ivan A. Sag. 2000. *Interrogative investigations*. Stanford, CA: CSLI Publications.
- Gordon, Morris, Michelle Daniel, Aderonke Ajiboye, Hussein Uraiby, Nicole Y. Xu, Rangana Bartlett, Janice Hanson, Mary Haas, Maxwell Spadafore, and Ciaran Grafton-Clarke. 2024. A scoping review of artificial intelligence in medical education: BEME Guide No. 84. *Medical*

- Teacher* 46(4): 446-470.
- Grice, Herbert P. 1975. Logic and conversation. In Peter Cole and Jerry L. Morgan (eds.), *Speech acts*, 41-58. Berlin: Brill.
- Halliday, Michael Alexander Kirkwood. 2004. *Computational and quantitative studies: Volume 6*. New York, NY: Bloomsbury Publishing.
- Hanson, Marta and Victor Kumar. 2025. Medical bilingualism: Reframing the overlap between medical systems in recent medical history and anthropology. *Chinese Medicine and Culture* 8(3): 239-255.
- Heritage, John and Douglas W. Maynard. 2006. *Communication in medical care: Interaction between primary care physicians and patients*. Cambridge: Cambridge University Press.
- Hou, Chengbin, Yanzhuo Gao, Xinyu Lin, Jinchao Wu, Ning Li, Hairong Lv, and William Cheng-Chung Chu. 2025. A review of recent artificial intelligence for traditional medicine. *Journal of Traditional and Complementary Medicine* 15(3): 215-228. <https://doi.org/10.1016/j.jtcme.2025.02.009>.
- Jang, Dongyeop and Chang-Eop Kim. 2023. Exploring the potential of large language models in traditional Korean medicine: A foundation model approach to culturally-adapted healthcare. *arXiv preprint arXiv:2303.17807*.
- Jang, Dongyeop, Taerim Yun, Choong-Yeol Lee, Young-Kyu Kwon, and Chang-Eop Kim. 2023. GPT-4 can pass the Korean National Licensing Examination for Korean Medicine Doctors. *PLOS Digit Health* 2(12): e0000416. <https://doi.org/10.1371/journal.pdig.0000416>.
- Jang, Ho. 2024a. Automatic classification of herbal formula documents related to diseases using LLM-based embedding vectors. *Herbal Formula Science* 32(4): 377-388. <https://doi.org/http://dx.doi.org/10.14374/HFS.2024.32.4.377>.
- Jang, Ho. 2024b. Named entity normalization for traditional herbal formula mentions. *Journal of The Korea Society of Computer and Information* 29(10): 105-111. <https://doi.org/10.9708/jksci.2024.29.10.105>.
- Joshi, Pratik, Sebastin Santy, Amar Budhiraja, Kalika Bali, and Monojit Choudhury. 2020. The state and fate of linguistic diversity in the NLP World. *arXiv preprint arXiv:2004.09095*.
- Kang, Bongsu, Jundong Kim, Tae-Rim Yun, and Chang-Eop Kim. 2024a. Prompt-rag: Pioneering vector embedding-free retrieval-augmented generation in niche domains, exemplified by Korean medicine. *arXiv preprint arXiv:2401.11246*.
- Kang, Bongsu, Sangyeon Lee, Hyojin Bae, and Chang-Eop Kim. 2024b Current status and direction of generative Large Language Model applications in medicine - Focusing on East Asian medicine. *Journal of Physiology & Pathology in Korean Medicine* 38(2): 49-58. <https://doi.org/http://dx.doi.org/10.15188/kjopp.2024.04.38.2.49>.
- Kim, Jundong, Hye-Yoon Lee, Ji-Hwan Kim, and Chang-Eop Kim. 2024a. Development of an LLM-based CPX practicing chatbot for Korean medicine education: Implementation of automated scoring and feedback generation framework. *Journal of Korean Medicine* 45(4): 215-230. <https://doi.org/http://dx.doi.org/10.13048/jkm.24067>.

- Kim, Jundong, Hye-Yoon Lee, Ji-Hwan Kim, and Chang-Eop Kim. 2024b. Pilot development of a 'Clinical Performance Examination (CPX) practicing chatbot' utilizing prompt engineering. *Journal of Korean Medicine* 45(1): 200-212. <https://doi.org/10.13048/jkm.24013>.
- Kim, Kyeong Han, Hyein Jeong, Gyeong Seo Lee, and Seung-Hee Lee. 2025. Exploring the potential of acupuncture practice education using artificial intelligence. *Integrative Medicine Research* 14(1): 101123. <https://doi.org/10.1016/j.imr.2025.101123>.
- Kim, Seon Kyoung, Myung Ho Jin, Seung Bae Choi, and Bum Hoi Kim. 2025. Proposal for an AI-based anatomy education model through analysis of AI utilization among students in colleges of Korean medicine. *Anatomy & Biological Anthropology* 38(3): 181-193. <https://doi.org/10.11637/aba.2025.38.3.181>.
- Kim, Suyoun, Su Hyun Kim, Hansea Kim, and Young-Mee Lee. 2025. Integrating artificial intelligence into medical curricula: Perspectives of faculty and students in South Korea. *Korean Journal of Medical Education* 37(1): 65-70. <https://doi.org/10.3946/kjme.2025.324>.
- Kim, Tae-Hun, Jung Won Kang, and Myeong Soo Lee. 2023. AI Chat bot - ChatGPT-4: A new opportunity and challenges in complementary and alternative medicine (CAM). *Integrative Medicine Research* 12(3): 100977. <https://doi.org/10.1016/j.imr.2023.100977>.
- Kim, Yongjin. 2025. Development and instructional application of a Korean medicine-specific LLM Chatbot in Traditional Korean Medical education. *Journal of Korean Medical Classics* 38(2): 11-26. <https://doi.org/http://dx.doi.org/10.14369/jkmc.2025.38.2.011>.
- Kövecses, Zoltán. 2010. *Metaphor: A practical introduction*. Oxford: Oxford University Press.
- Kwon, Chan-Young. 2024. Utilization of generative artificial intelligence chatbot for training in suicide risk assessment of depressed patients: Focusing on students at a college of Korean medicine. *Journal of Oriental Neuropsychiatry* 35(2): 153-162. <https://doi.org/10.7231/jon.2024.35.2.153>.
- Lakoff, George and Mark Johnson. 1980/2008. *Metaphors we live by*. Chicago, IL: The University of Chicago Press.
- Lee, Haedeum and Yongjin Kim. 2025. Assessing how Large Language Models enhance learning outcomes in flipped Korean medical history education. *The Journal of Korean Medicine History* 38(1): 81-96. <https://doi.org/http://dx.doi.org/10.15521/jkmh.2025.38.1.081>.
- Lee, Hye-Yoon, Jundong Kim, Hyojae Choi, Hyojin Bae, Aram Jeong, Sungyul Choi, Ji-Hwan Kim, and Chang-Eop Kim. 2025. Comparing AI chatbot simulation and peer role-play for OSCE preparation: A pilot randomized controlled trial. *BMC Medical Education* 25(1): 1755. <https://doi.org/10.1186/s12909-025-08308-y>.
- Lee, Hyeonhoon. 2023. Using ChatGPT as a learning tool in acupuncture education: Comparative study. *JMIR Medical Education* 9: e47427. <https://doi.org/10.2196/47427>.
- Lee, Ji-Su, Sung-A Kim, Taegon Kim, Seunghoon Lee, Tae-Hun Kim, and Jung Won Kang. 2024. Exploring Korean medicine professions' perspectives on the applicability of ChatGPT in facial palsy practice: A web-based survey. *European Journal of Integrative Medicine* 72: 102422. <https://doi.org/10.1016/j.eujim.2024.102422>.

- Lee, Won-Yung, Sang-Yun Han, Ji-Hwan Kim, Byung-Wook Lee, Yejin Han, and Seungho Lee. 2025. Gen-SynDi: Leveraging knowledge-guided generative AI for dual education of syndrome differentiation and disease diagnosis. *Applied Sciences* 15(9): 4862.
- Lee, Won-Yung, Sang Yun Han, and Seungho Lee. 2024. Exploring the feasibility of developing an education tool for pattern identification using a large language model: Focusing on the case of a simulated patient with fatigue symptom and dual deficiency of the heart-spleen pattern. *Herbal Formula Science* 32(1): 1-9. <https://doi.org/http://dx.doi.org/10.14374/HFS.2024.32.1.1>.
- Levinson, Stephen C. 1983. *Pragmatics*. Cambridge: Cambridge University Press.
- Marinelli, S., L. De Paola, M. Stark, and G. Montanari Vergallo. 2025. Artificial Intelligence in the service of medicine: Current solutions and future perspectives, opportunities, and challenges. *Clinica Terapeutica* 176(2): 77-82. <https://doi.org/10.7417/CT.2025.5192>.
- Mehta, Darshan H. and Melinda Ring. 2025. Artificial intelligence in integrative medicine: Transforming education for a digital future. *Integrative Medicine Research* 14(4): 101259.
- Merchant, Jason. 2005. Fragments and ellipsis. *Linguistics and Philosophy* 27: 661-738.
- Mishler, Elliot. 1984. *The discourse of medicine: Dialectics of medical interviews*. Norwood, NJ: Ablex.
- Park, Miso, Ju Hyun Jeon, Junyoung Hur, Horyong Yoo, and Ji-Yeon Lee. 2025. Integrating generative AI into clinical practice education: Enhancing personalized medicine delivery skills for Korean medicine students. *Journal of Korean Medicine* 46(1): 87-102. <https://doi.org/10.13048/jkm.25007>.
- Park, Sa-Yoon and Chang-Eop Kim. 2023. Enhancing Korean medicine education with Large Language Models: Focusing on the development of educational artificial intelligence. *Journal of Physiology & Pathology in Korean Medicine* 37(5): 134-138.
- Rajpurkar, Pranav, Emma Chen, Oishi Banerjee, and Eric J. Topol. 2022. AI in health and medicine. *Nature Medicine* 28(1): 31-38.
- Ryu, Jinwoo, Chanyoung Kwon, Jeaseong Park, Seongryong Lim, Hyeli Jeon, Huijun Kim, and Sunghee Kim. 2024. Development and application of a ChatGPT-based simulation training platform for Korean medicine. *Journal of Oriental Neuropsychiatry* 35(4): 413-427. <https://doi.org/http://dx.doi.org/10.7231/jon.2024.35.4.413>.
- Seo, Sumin, Anna Kim, and Yongtaek Oh. 2025. Development of an AI-based standardized patient system (CURA) utilizing diagnostic-pattern identification schema. *The Journal of the Society of Korean Medicine Diagnostics* 29(1): 251-263.
- Shin, Seungwon, Jihyun Sang, and Ji-Sun You. 2025. Use and perception of generative artificial intelligence in traditional Korean medicine education: A cross-sectional survey of undergraduate students in Korea. *Integrative Medicine Research* 14(4): 101216. <https://doi.org/10.1016/j.imr.2025.101216>.
- Sperber, Dan and Deirdre Wilson. 2015. Beyond speaker's meaning. *Croatian Journal of Philosophy* 15(44): 117-149.

- Tricco, Andrea C, Erin Lillie, Wasifa Zarin, Kelly K. O'Brien, Heather Colquhoun, Danielle Levac, David Moher, Micah DJ Peters, Tanya Horsley, and Laura Weeks. 2018. PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine* 169(7): 467-473.
- Yoon, Da-Eun, Cheol-Han Kim, Yeonhee Ryu, Ye-Seul Lee, and Younbyoung Chae. 2025. Performance of GPT-4 for planning acupuncture treatment: Comparison with human clinician performance. *Frontiers in Medicine* 12: 1632303. <https://doi.org/10.3389/fmed.2025.1632303>.
- Yu, Ning. 2009. *From body to meaning in culture: Papers on cognitive semantic studies of Chinese*. Berlin: Mouton de Gruyter.

Suji Lee

Clinical Associate Professor
Department of Acupuncture and Moxibustion
Kyung Hee University / Kyung Hee University Medical Center
26 Kyungheedaero, Dongdaemun-gu,
Seoul, 02447, Korea
E-mail: sjstarry41@naver.com

Sejin Kim

Resident
Department of Acupuncture and Moxibustion
Kyung Hee University / Kyung Hee University Medical Center
26 Kyungheedaero, Dongdaemun-gu,
Seoul, 02447, Korea;
Graduate Student
Department of Psychology
Korea University
145 Anam-ro, Seongbuk-gu,
Seoul, 02841, Korea
E-mail: acu.say.jin@gmail.com

Jong-In Kim

Professor
Department of Acupuncture and Moxibustion
Kyung Hee University / Kyung Hee University Medical Center
26 Kyungheedaero, Dongdaemun-gu,
Seoul, 02447, Korea
E-mail: hann8400@hanmail.net

Seulkee Park

Academic Research Professor
Department of English Linguistics and Literature
Kyung Hee University
26 Kyungheedaero-ro, Dongdaemun-gu,
Seoul, 02447, Korea
E-mail: seulkeepark@khu.ac.kr

Jong-Bok Kim

Professor
Department of English Linguistics and Literature
Kyung Hee University
26 Kyungheedaero-ro, Dongdaemun-gu,
Seoul, 02447, Korea
E-mail: jongbok@khu.ac.kr

Received: 2026. 02. 03.

Revised: 2026. 07. 21.

Accepted: 2026. 07. 23.