

진단·변증 스키마 기반 인공지능 표준화 환자 시스템 (CURA) 개발

서수민¹, 김안나¹, 오용택^{2*}

1. 한국한의학연구원 한의약데이터부
2. 우석대학교 한의과대학 진단학교실

Abstract

Development of an AI-Based Standardized Patient System (CURA) Utilizing Diagnostic-Pattern Identification Schema

Sumin Seo¹, Anna Kim¹, Yongtaek Oh^{2*}

¹KM Data Division, Korea Institute of Oriental Medicine

²Dept. of Diagnostics, College of Korean Medicine, Woosuk University

Abstract

This study aims to enhance the clinical competency of Korean medicine students by developing a standardized patient system that integrates the unique diagnostic framework of Korean medicine—pattern identification(辨證) and diagnosis(診斷)—with Large Language Models (LLM), Retrieval-Augmented Generation (RAG), and multi-modal technologies. Using CPX data developed by the National Institute for Korean Medicine Development(NIKOM), we constructed a dataset, derived and designed the clinical workflow, and implemented a standardized patient system named CURA (Clinical Understanding & Response Assistant).

CURA engages in natural conversations with learners while dynamically presenting visual materials, allowing the inclusion of multidimensional diagnostic data such as patient photographs, thermal images, lung scans, and tongue images. This enables learners to interpret and analyze both visual and linguistic information comprehensively in alignment with Korean medical diagnostic logic, fostering clinical reasoning processes similar to real-world practice.

The AI-driven CURA system generates clinical scenarios, interacts with standardized patients, and produces real-time dialogue reports containing patients' main symptoms and medical histories. Furthermore, it incorporates an automated evaluation module based on CPX scoring criteria, providing real-time scoring and feedback derived from conversational and observational data. Learners can immediately identify missing questions or diagnostic errors, facilitating self-directed learning, while educators can assess students' clinical performance using consistent criteria, minimizing evaluation bias.

The CURA system developed in this study offers a potential solution to the shortage of evaluators and ensures equitable clinical training opportunities for all students. It contributes to improving the quality of Korean medicine clinical education and addressing educational resource imbalances. Moreover, by providing diverse simulated clinical scenarios, CURA supports competency-based education for Korean medicine students and offers valuable insights into the future direction of Korean medical education.

Key words

Standardized Patient; cpx; Diagnostic and Pattern Identification Schema; pathway;

* 교신저자 : YongTaek Oh, Department of Diagnostics, College of Korean Medicine, Woosuk University, 443 Samnye-ro, Samnye-eup, Wanju-gun, Jeollabuk-do, 55338, Republic of Korea.

Tel : +82-63-290-9026, E-mail : ohyt@woosuk.ac.kr

• 원고접수일 : 2025.09.17 / 심사완료일 : 2025.10.01 / 게재결정일 : 2025.10.09

I. 서론

최근 의학교육 분야에서는 인공지능(Artificial Intelligence, 이하 AI) 기술을 활용한 교육 혁신이 활발히 이루어지고 있으며, 한의학 교육에서도 임상실습의 질 향상을 위한 새로운 접근법이 요구되고 있다. 특히 가상 환자(virtual patient) 및 표준화 환자(standardized patient, 이하 SP) 시뮬레이션에 AI를 접목하려는 시도가 증가하고 있다.

국외에서는 대화형 AI를 이용한 임상 사례 학습 연구가 다수 보고되었다. 예를 들어, 거대언어모델(Large Language Model, 이하 LLM)을 SP로 활용한 연구에서 학생들은 AI 환자와의 면담을 기존 SP를 보완하는 효과적인 학습 도구로 평가하였다¹⁾. 학생들은 AI와의 상호작용을 통해 진단 능력과 학습 편의성이 향상되었으며, AI 환자가 완전한 대체물은 아니더라도 보조 도구로서 충분한 교육적 가치를 지닌다고 결론지었다²⁾. 또한 터키에서 진행된 연구에서는 인턴 의사들이 ChatGPT 기반 가상 환자와의 임상 사례 연습에 높은 만족도를 보였으며, 안전하고 통제된 환경에서 임상추론 능력을 강화할 수 있다는 점이 확인되었다²⁾. 중국의 연구에서는 ChatGPT-3.5를 활용한 모의 진료에서 전문의 평가 기준으로 평균 90% 이상의 정확도를 보여, 실제 환자와 유사한 대화가 가능함을 입증하였다³⁾. ChatGPT를 AI 기반 SP로 활용한 연구에서는 학생들이 임상 병력청취 과정에서 감정적 요소를 표현하고 응답하는 능력 측면에서도 긍정적 결과가 관찰되었다는 보고가 있다⁴⁾. 또한 의학교육 관점에서 ChatGPT 기반 가상 침상(bedside) 교육을 도입하여 학습 효과 및 경험을 조사한 연구에서는, 다수 학생이 학습 효율과 경험의 증대를 보고하였으며, AI-챗봇이 실시간 보조 및 국제 협력 교육에도 유리하다는 결과가 제시되었다⁵⁾. 더 나아가 GPT 기반 AI를 가상 환자와 혼합현실(Mixed Reality, 이

하 MR) 시나리오에 통합하여 의사소통 훈련을 보조하는 연구도 진행되고 있다⁶⁾.

국내에서도 한의학 교육을 중심으로 AI 기반 SP 연구가 시도되고 있다. 가천대와 부산대 공동연구팀은 GPT-4 API를 이용해 진료 수행시험(Clinical Performance Examination, 이하 CPX) 연습용 챗봇을 시범 개발하였으며, 그 결과 응답의 정확성과 일관성이 우수하다는 평가를 받았다⁷⁾. 또한 가천대에서는 GPT-4와 Claude를 활용하여 CPX 자동 채점 및 피드백 시스템을 구현하였으며, 두 모델의 장점을 결합하여 하이브리드 방식으로 최적화된 시스템을 구현할 수 있었다고 보고하였다⁸⁾. 또한 원광대에서는 챗봇 플랫폼을 기반으로 한 임상실습용 AI 환자 시스템을 개발하였고, 교육 전문가 평가에서 높은 편의성과 교육 효과를 보였다⁹⁾. 동의대에서는 ChatGPT 기반 AI 플랫폼을 활용한 한의학 임상 실습 시스템을 실제 수업에 적용한 사례도 보고되었다. 이 시스템은 학생들이 실제 진료 환경과 유사하게 환자와 대화를 통해 진단을 수행할 수 있게 하며, 학생 만족도 조사가 4점대 이상(5점 만점)으로 높았다는 실증 보고도 있다¹⁰⁾. 또한 Gen-SynDi라는 지식 기반 생성 AI 프레임워크는 피로 증례 중심으로 변증과 질병 진단을 동시에 학습할 수 있도록 설계된 가상 환자 시스템을 제안하였으며, 전문가 평가에서도 시나리오의 현실성, 교육적 유용성 등이 긍정적으로 평가되었다¹¹⁾.

그러나 기존 한의학 교육의 임상실습 과정은 여러 가지 한계를 안고 있다. 첫째, 실제 환자와 유사한 경험을 제공하기 위해 활용되는 SP는 훈련과 관리에 많은 인적·물적 자원이 소요되며, 학생 수 증가에 비해 공급이 제한적이다¹²⁾. 둘째, 한의학 특유의 변증과 진단 스키마는 복잡한 의사결정 과정을 포함하는데, 이를 반복적이고 표준화된 방식으로 훈련할 수 있는 교육 도구는 부족하다¹³⁾. 셋째, 현행 교육 체계에서는 학생들이 충분한 환자 증례를 경험하지 못해 진단·처방 능력

이 불균형하게 발달할 위험이 존재한다.

이러한 맥락에서 최근 발전한 LLM은 방대한 데이터 학습을 통해 인간 수준에 근접한 언어 이해 및 생성 능력을 보이면서 의료 교육에 새로운 가능성을 제시한다^{14,15}. 그러나 LLM은 고정된 학습 데이터에 의존하기 때문에 최신 지식 반영이나 한의학과 같은 특수도메인 활용에는 한계가 있다^{14,15}. 이를 보완하기 위해 제안된 방법이 검색 증강 생성 (Retrieval-Augmented Generation, 이하 RAG)으로, 외부 지식베이스를 활용하여 근거 기반의 신뢰성 있는 응답을 제공할 수 있다¹⁶. RAG는 임상적 타당성을 확보할 뿐 아니라 다양한 시나리오 생성과 표준화된 학습 경험 제공에도 적합하다.

아울러 임상 실습의 현실감을 향상시키기 위해 멀티모달 정보 통합 기술의 도입이 중요하다. 의료현장에서는 문진 정보 외에도 시각 자료(얼굴 사진, 설진 사진, 영상 진단 자료 등), 생체신호, 청각 단서(예: 호흡음, 심음) 등이 복합적으로 활용된다. 최근 가상환자 시스템 연구에서는 이러한 멀티모달 요소를 포함하여 학습자가 복합 정보를 통합해 판단하도록 설계하는 방향이 주목받고 있다. 예컨대, MedDiT 모델은 지식 기반 확산 변환기를 활용해 가상환자 시나리오와 동적으로 연계된 의료 영상을 생성함으로써 복합 진단 훈련을 지원하는 프레임워크를 제안하였다¹⁷. 또 다른 연구에서는 멀티모달 생성 AI 기반의 합성 환자(synthetic patient)를 통해 민감한 대화 상황을 실시간 영상·음성·문자 데이터를 통합하여 시뮬레이션하는 플랫폼을 제시한 바 있다¹⁸. 이러한 멀티모달 접근은 학습자가 실제 임상 환경에서와 유사한 몰입형 정보 해석 훈련을 경험하게 하며, 교육 효과와 현실감 모두를 제고할 수 있다. 따라서 본 연구에서는 한의학의 진단·변증 스키마를 구조화된 지식 체계로 활용하고, CPX 데이터를 기반으로 LLM과 RAG를 적용하여, 멀티모달 기술을 통합한 AI 기반 SP 시스템(Clinical Understanding Response

Assistant, 이하 CURA)을 개발하고자 한다. 이를 통해 교육자 부족 문제 해소, 임상실습 기회의 균등 제공, 한의학 특수성에 맞춘 학습 시나리오 제공이라는 교육적 요구를 충족시키고자 하며, 나아가 한의대생의 임상추론 역량을 강화하고자 한다.

II. 연구 방법

1. 연구 설계 개요

본 연구는 한의학의 고유한 진단 체계인 변증과 진단 스키마를 기반으로, AI 기술을 적용한 SP 시스템 CURA를 개발하는 것을 목적으로 하였다. 연구의 전체 과정은 CPX 기반 데이터 구축, RAG 기반 한의학 지식 통합 모듈 설계, 자동 채점 및 피드백 시스템 개발, 그리고 CURA 시스템의 통합 구현 단계로 구성하여 학습자가 실제 임상 상황과 유사한 환경에서 변증과 진단 과정을 학습할 수 있도록 설계되었다. 이를 통해 임상 추론과 판단 능력을 체계적으로 훈련할 수 있는 기반을 마련하였다.

2. 데이터 구축 및 전처리 과정

데이터 구축에는 한국한의학진흥원(NIKOM)¹⁹의 한의표준임상진료지침 연계 교육도구(CPX 모듈)²⁰를 활용하였다. CPX는 해결해야 할 문제를 가지고 주로 처음 진료실을 찾는 환자를 대상으로 병력청취, 신체진찰 및 환자교육 등을 수행하는 시험으로, 일차 진료 역량 배양을 위한 교육목표를 기반으로 설계되었으며, 상황을 재현하도록 구성되어 있다²⁰. 본 연구에서는 CPX 사례 중 감기 관련 증상 CPX(이하 감기 CPX)와 알레르기 비염 관련 증상CPX(이하 알레르기비염 CPX)를 핵심 자료로 선정하고 데이터를 구축하였다. 각 사례로부터 환자의 인적사항, 감정상태, 주요 증상, 병력, 가족력, 수면, 흡연, 음주, 신체진찰 소견, 문진 문항 등을 추출하였으

며, 이후 한의학의 변증 체계(풍한형, 풍열형, 외감풍열형 등)에 따라 데이터를 재분류하였다. 변증별로 주요 증상과 부증상, 설진, 맥진 등의 항목을 구조화하여 임상적 일관성을 높였다. 이후 각 사례의 내용을 JSON 기반의 질의응답 형식으로 변환하여 LLM과 검색 기반 응답 구조에 활용할 수 있도록 정제하였다. 데이터의 임상적 타당성과 분류의 정확성을 확보하기 위해 경력 15년 이상의 한의사 3인이 모든 데이터를 각각 검토하였으며, 표현상의 모호성과 중복 항목을 수정·보완하였다. 검토 과정에서 변증 분류 기준이나 증상 기술 방식에 대한 의견 불일치가 발생한 경우, 관련 근거문헌 및 임상 가이드라인을 재검토한 후 전문가 합의를 통해 일관된 기준으로 통일하였다. 최종적으로 세 검토자가 모두 동의한 기준에 따라 데이터를 확정함으로써, 자료의 신뢰성과 재현 가능성을 확보하였다. 최종적으로 변증별 대화 시나리오가 구축되어 CURA 시스템의 학습 및 평가 자료로 사용되었다.

3. RAG 기반 한의학 지식 통합 모듈 설계

RAG는 LLM에 외부 지식베이스의 정보를 결합함으로써 사실성과 전문성을 높일 수 있는 방법으로, 최신 지식이나 분야별 정보를 활용한 문맥 생성에 효과적이다²¹⁾. RAG는 LLM이 자체 학습에 없는 외부 지식을 검색하여 참고 한 뒤 답변을 생성하고 한의학 교육에서 동의보감과 같은 고전, 임상 매뉴얼, 최신 진단·치료 지침 등 방대한 한의학 지식을 LLM에 통합하는 유용한 수단이 된다.

대부분의 LLM은 주로 서구 의학 데이터로 학습 되어 한의학 도메인 지식의 밀도가 낮을 수 있는데²²⁾, RAG를 활용하면 별도로 구축한 한의학 지식베이스로부터 필요한 정보를 실시간으로 불러와서 환자 시나리오에 반영할 수 있다. 예를 들어, 한의 변증 시나리오를 생성한다고 가정하면, RAG 시스템은 먼저 해당 질환의 변증 유형과 증후에 관한

한의학 고전의 기술을 검색하여 LLM에 제공한다. LLM은 이를 바탕으로 환자의 주증상, 설진/맥진 소견, 치료경과 등을 한의학 이론에 맞게 구성하여 시나리오를 만들어낸다. 이렇게 하면 모델이 단순히 일반 상식으로 환자를 연기하는 것이 아니라, 한의학적 근거에 기반한 정확한 임상 상황을 묘사하게 된다.

LLM을 한의학에 적용하는 방법에는 대량의 한의학 말뭉치를 사전학습/미세조정하는 접근과 내부 모델 파라미터를 건드리지 않고 프롬프트 엔지니어링이나 RAG로 지식을 주입하는 접근이 있는데, 후자가 기술적 부담이 적고 유연하여 많이 고려된다²²⁾. 한의학 특화 LLM 개발에 관한 보고에 따르면, RAG나 인지 보강 기법을 통해 한의학 지식을 끌어올 경우, 개인맞춤형 교육 콘텐츠 생성, 커리큘럼 개발, 지식 통합, 가상환자 시뮬레이션 등 폭넓은 응용이 가능할 것으로 전망되고 있다²²⁾.

본 연구에서는 수집된 자료는 임베딩(Embedding)을 통해 벡터DB(Vector Database)에 저장하였다. 임베딩 모델로 OpenAI의 text-embedding-3-small을 사용하였고, 벡터DB는 Chroma를 사용하였다. 검색 질의와 문헌 간 의미적 유사도 계산에는 코사인 유사도 알고리즘을 적용하였다. 이 과정을 통해 사용자의 질의가 입력되면, RAG의 Retriever 모듈이 관련 문헌을 벡터 공간에서 검색한다. 검색된 문헌은 GPT-4o-mini 모델에 전달되어, 해당 근거를 반영한 응답을 생성하도록 구성하였다.

또한 풍열감모 등 변증 유형에 따른 문진 흐름을 반영한 프롬프트 템플릿을 설계하여, 응답의 일관성과 진단 논리의 타당성을 동시에 확보하였다. 이러한 구조를 통해 CURA는 단순한 대화형 응답 생성 수준을 넘어, 한의학적 근거에 기반한 변증 논리를 묘사할 수 있도록 구현되었다.

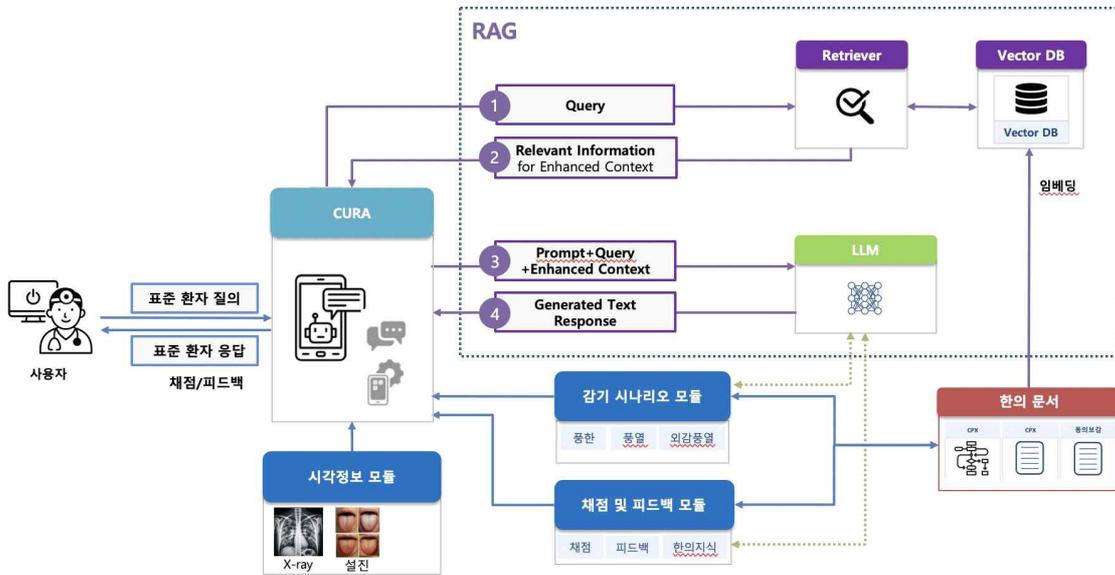


Figure 1. System Architecture of CURA

The diagram represents the interaction architecture among the dialogue module, RAG-based traditional medicine knowledge integration module, visual processing module, and evaluation/feedback module.

4. 자동 채점 및 피드백 시스템 개발

CURA 시스템에는 CPX 채점 기준을 기반으로 한 자동 평가 및 실시간 피드백 모듈이 포함되어 있다.

이 모듈은 학습자의 문진 대화 로그를 분석하여 진단 정확도, 질문의 적절성, 환자 응대 태도, 응답의 완결성 등을 자동으로 평가한다. 채점 알고리즘은 LLM 기반 의미 일치도 분석과 규칙 기반 키워드 매칭 방식을 결합하여 응답을 평가하였다. 이를 통해 학습자가 의도한 질문이 변증 논리에 적합한지를 판단하고, 진단 과정의 완결성을 수치화할 수 있었다. 평가 결과는 항목별 점수와 누락된 질문, 개선이 필요한 영역으로 구성된 피드백 리포트 형태로 즉시 제시되며, 학습자는 자신의 진단 전략을 스스로 점검할 수 있다. 이 시스템은 평가의 일관성과 객관성을 확보하여, 교수자에게는 평가 효율성을, 학습자에게는 자기주도적 학습 기회를 제공한다.

5. CURA 시스템 구현 및 기술적 구성

CURA 시스템의 전체 구조 및 구성 요소 간 상호작용을 설계하였다(Figure 1). 전체 시스템은 대화 모듈, RAG 기반 한의학 지식 통합 모듈, 시각정보 모듈, 채점 및 피드백 모듈로 구성된다. 대화형 언어모델 모듈은 학습자와 자연스러운 대화를 수행하며, 변증 데이터셋을 기반으로 환자의 증상과 반응을 생성한다. RAG 기반 한의학 지식 통합 모듈은 CPX, CPG(한의표준임상진료지침), 동의보감 등의 한의지식을 검색하여, LLM의 응답에 근거를 제공하고 임상 타당성을 높인다. 멀티모달 처리 모듈은 대화 중 환자의 얼굴, 설진 이미지, 체온, 폐 영상 등의 시각 자료를 제시함으로써 학습자의 시각적 진단 추론을 보조한다. 자동 평가 및 피드백 모듈은 대화 로그를 실시간 분석하여 점수와 피드백을 제공하며, 학습자의 진단 수행 능력을 정량적으로 평가한다.

CURA는 Python기반의 웹 프레임워크 Streamlit을 이용하여 구현하였으며, RAG, LLM을 연동하여 대화형 모델을 구성하였다. 이와 같은 통합 구조를 통해 CURA는 한의학 변증 논리에 기반한 임상 시뮬레이션과 자동 평가 기능을 동시에 수행할 수 있는 플랫폼으로 완성되었다.

III. 결과

1. 데이터 구축

감기 CPX를 중심으로 질문-응답을 포함하는 구조화된 데이터셋을 구축하였다. 원본 CPX 자료에서 추출한 환자 환자의 인적사항(연령, 성별, 직업), 활력징후(혈압, 맥박, 호흡, 체온, 체중, 신장), 감정상태, 주요 증상, 병력(약물 복용, 질환, 알레르기), 가족력, 수면, 흡연, 음주, 식사 습관으로 체계화되었으며, 세부 변형 시나리오를 생성하여 다양성을 확보하였다.

감기 질환에 대한 한의학적 변증을 반영하여 데이터를 재분류하고 각 변증별로 특징적인 주증상, 기타증상, 설맥 소견, 문진 요소를 세분화하였다. 구축된 데이터는 LLM과 RAG 시스템에 최적화된 JSON 구조로 변환되었다.

최종 구축된 데이터 셋은 CURA 시스템의 핵심 지식베이스로 활용되어, 실제 한의학 교육 환경에서 요구되는 다양한 임상 시나리오를 지원할 수 있는 기반을 마련하였다.

2. 주소-진단-변증 프로세스

감기와 알레르기비염 CPX를 중심으로, 주소-진단-변증으로 이어지는 단계적 프로세스를 구축하였다(Figure 2). CPX의 진단 스키마, 변증 스키마, 변증별 치법과 처방, 변증별 필수증상, 채점기준 자료를 참고하였다(Table 1). 첫 번째 단계에서는 환자의 주소증을 기반으로 한 기침, 콧물, 발열, 두통 등

문진을 수행하고 증상의 발생 시점에 따라 만성/급성 여부를 적용하였다. 두 번째 단계에서는 문진 결과를 토대로 감염성 상기도염, 폐렴, 부비동염, 알레르기비염 등을 감별진단 경로를 적용하였다. 세 번째 단계에서는 한의학 변증 체계를 적용하여 풍한형, 풍열형, 외감풍한, 외감풍열 등의 변증을 판별하였다. 즉, 주요 증상을 확인하고 감별진단(감염성, 상기도/폐/심장, 알레르기성) 후 변증 감별진단을 한다. 변증 후에 오한, 발열, 가래, 사지부, 설, 맥 등 항목에 대한 필수증상을 참고하여 한의학적 진단 항목을 추가하였으며 임상 교육의 균형성을 강화하였다. 풍한형 감기의 경우 오한, 무한(無汗), 백태 등의 소견이 포함되고 풍열형 감기의 경우 발열, 담조(痰稠), 황조(黃稠), 설태박황(舌苔薄黃) 등의 소견이 대표적인 필수 증상으로 나타난다.

이러한 다차원적 설계는 학습자가 증상을 단순 암기하는 것을 넘어, 실제 진료 상황에서 논리적 추론을 수행하도록 유도하여 자기주도형 임상학습을 가능하게 하였다.

3. RAG를 활용한 한의학 지식 통합 환자 시나리오 생성

시나리오 생성 단계에서는 RAG 기법을 도입하여 한의학 도메인 지식을 통합하였다. 본 시스템 CURA에서는 한의학의 진단·변증 지식 체계(예: 감기의 변증 유형별 증상)를 사전에 데이터베이스화하고, 환자 답변 생성 시 해당 지식을 참조하였다. 그 결과 각 시나리오는 실제 임상 문진 흐름과 유사한 질의응답 구조를 보였다. 예를 들어, 환자가 감기로 내원한 경우 한의사는 열감 유무, 기침 양상 등 표준 문진 항목을 질문하고 환자는 이에 따라 자신의 증상을 답하도록 시나리오가 구성되었다. 또한 시나리오마다 환자의 변증 유형(예: 풍한, 풍열, 외감풍열 등)에 따른 증상 표현과 답변의 일관성을 유지하여, 대화 내용이 한의학 임상 논리에 부합하도록 하였다.

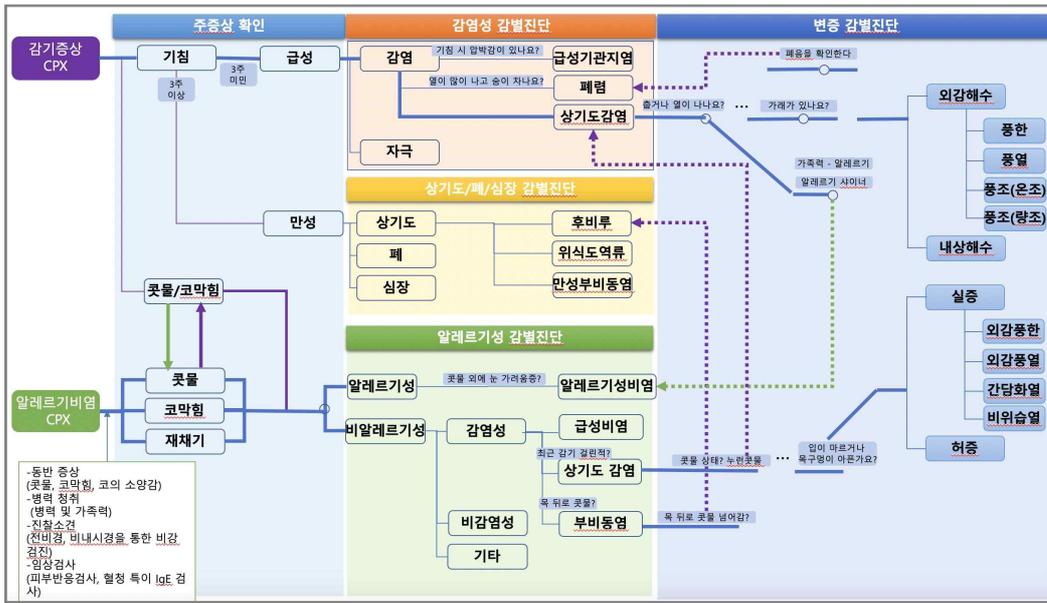


Figure 2. Process of Chief Complaint-Diagnosis-Pattern Identification
 This diagram illustrates the stepwise clinical reasoning process from a patient’s chief complaint to diagnosis and pattern identification in traditional Korean medicine

Table 1. Examples of Key Symptoms by Pattern Identification²⁰⁾

변증	惡寒/發熱	痰	胸腹部	頭面部	四肢部	舌	脈
풍한	惡寒重, 無汗	痰稀薄色白			骨節痠痛	舌苔薄白	脈浮
풍열	身熱	痰稠, 黃稠				舌苔薄黃或薄白	脈浮數
풍조(온조)		咳嗽少痰或咯出難		咽乾, 咽痛		舌苔薄黃而乾, 舌尖紅	脈浮數或無變化
풍조(양조)	惡寒發熱 無汗			喉痒 頭痛鼻塞		舌苔薄白,	脈浮緊
외감풍한	惡寒發熱	嘔嚏 咳嗽, 痰稀		鼻流清涕		舌質紅苔薄白	脈浮 或浮緊
외감풍열	發熱	痰黃粘稠		鼻流黃涕 鼻粘稠 鼻痒嘔嚏 口渴, 咽痛		舌質紅, 苔薄黃	脈浮數
간담화열			煩躁, 易怒	後鼻漏, 嗅覺減退 面紅目赤		舌質紅苔黃	脈弦數
비위습열			惡心, 倦怠 小便黃	鼻塞甚 腥臭/面色黃		舌紅苔黃膩	脈滑數 或濡數

이러한 지식 통합을 통해 생성된 대화형 시나리오는 일반적인 LLM 기반 대화보다 한의학적으로 타당성이 높고, 학습자가 실제 한의 임상 상담에 가까운 경험을 할 수 있게 되었다.

4. 자동 채점과 실시간 피드백 시스템

CURA에는 CPX 채점 기준 및 채점표를 확장 적용한 자동 평가 모듈이 포함되어 있다. 채점표는 각 시나리오마다 16~20개의 평가 항목이 사전에 정의되어 있으며, 항목별로 중요도에 따라 1~2점의 배점이 할당되어 있다(Table 2). 사용자가 CURA와 대화형 시뮬레이션을 수행하면 시스템은 실시간으로 대화 내용을 분석하여 각 평가 항목의 수행 여부를 판정하고 점수를 부여한다. CURA와의 대화를 완료하면 각 평가항목의 수행 여부를 판정하고 총점 및 세부 점수를 산출한다. 또한 즉각적인 피드백 리포트가 제공된다. 예를 들어 한 감기 시나리오의 만점이 16점일 경우, 한 참가자의 대화 결과 12점을 획득하여 75%의 수행도를 보였다고 가정해 보자. 시스템은 이처럼 ‘총점:12점/16점(75%)’이라는 결과와 함께, 어떤 항목에서 점수가 부여되고 누락되었는지를 상세하게 보여준다. 피드백 화면에는 “환자는 기침을 많이 한다고 말하였다(1점)” 등의 세부 결과가 제공된다. 문항별 질문 내용, 점수, 근거까지 포함되어, 학습자의 자기진단을 돕는다. Table 3은 구축된 시나리오 데이터셋과 자동 평가 체계의 주요 내용을 요약한 것으로, 시나리오 수와 평가 항목 구성, 그리고 앞서 언급한 예시 시나리오의 점수 결과를 포함한다. 이와 같은 결과를 통해 CURA는 대화형 진료 시뮬레이션뿐 아니라 정량화된 평가까지 지원하여 학습자에게 객관적인 피드백을 제공할 수 있음이 확인되었다.

Table 2. An excerpt from the CPX Scoring Criteria and Evaluation Rubric (Wind-cold type exogenous cough)²⁰

채점기준	채점표	채점표 점수
어디가 불편하신가요?	환자는 기침을 많이 한다고 말하였다	1점 : 기침을 많이 한다고 말함 0점 : 그렇지 않음
언제부터 기침이 났나요?	환자는 3일전부터 기침이 나기 시작하였다고 말하였다	1점 : 3일전부터 증상이 있었다고 말함 0점 : 그렇지 않음
기침 시 가슴울혈 또는 압박감이 있나요?	환자는 기침 시 가슴 답답함과 압박감이 없다고 말하였다.	1점 : 기침 시 가슴 답답함과 압박감이 없다고 말함 0점 : 기침 시 가슴 답답함과 압박감에 대해 말하지 않음

Table 3. Summary of the Evaluation Components for the Scenario in the CURA System

항목	내용
시나리오당 평균 평가 항목 수	16~20개
평가 항목당 배점	1~2점
시나리오 만점 (총 배점)	16~20점
사용자 득점 (득점/만점)	12점 / 16점 (75%)

5. CURA 시스템 구현

CURA의 표준환자 대화 화면을 구현하였다(Figure 3). 대화창에는 환자의 얼굴 이미지가 표시되며, 학습자는 환자와 대화로 문진을 진행한다. 대화 과정 중에는 실시간으로 체온·혈진 이미지 등이 동적으로 등장하여,

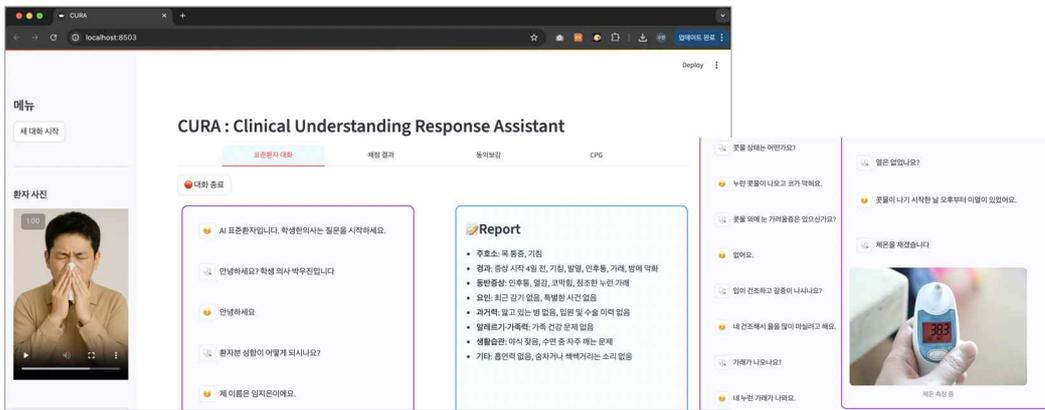


Figure 3. Standardized Patient Dialogue Interface

The CURA system provides a real-time dialogue interface where learners conduct patient interviews with a virtual standardized patient. Dynamic visual elements such as facial images, tongue diagnosis photos, and temperature data are presented during the conversation to support integrated clinical reasoning training.

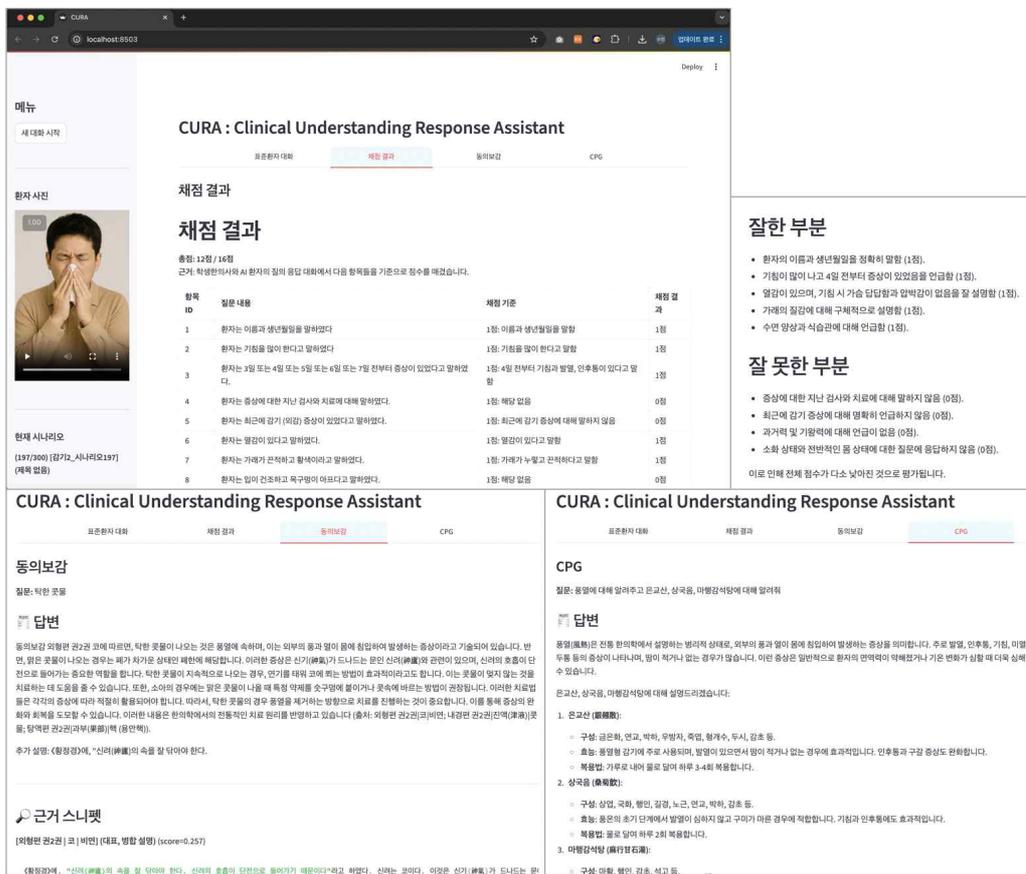


Figure 4. Automated Scoring and Feedback Interface

The CURA system's feedback interface automatically evaluates learner-patient interactions based on CPX scoring criteria. The screen displays overall and itemized scores, identifies missed questions, and provides real-time feedback and improvement suggestions to guide self-directed learning.

학습자는 환자의 반응을 대화와 시각으로 인식하면서 종합적 판단을 수행한다.

자동 채점 결과 화면은 각 항목별 점수와 총점, 피드백 문구가 즉시 제시된다(Figure 4). 하단에는 “향후 개선 권장 항목”이 자동 제시되어, 학습자가 자신의 진단 전략을 수정할 수 있도록 안내한다. 이러한 시각화 기반 피드백은 단순 점수 제시를 넘어, 학습자에게 임상적 사고의 경로를 재구성하도록 돕는다.

IV. 고찰

본 연구에서는 한의학의 진단과 변증 스키마를 기반으로 LLM, RAG, 멀티모달 기술을 통합하여 SP 시스템 CURA를 개발하였다. 이를 통하여 한의대생이 실제 임상 진료와 유사한 환경에서 변증 및 임상추론 능력을 체계적으로 훈련할 수 있는 교육 모델을 제시하였다.

기존 한의학 임상실습 교육은 SP 확보의 어려움, 평가자 간 편차, 실습 기회의 불균형 등 구조적 한계를 지니고 있었다. 본 연구에서 개발한 CURA는 이러한 문제를 기술적으로 보완하고, CPX 기반 자동 평가 및 피드백 기능을 구현함으로써 학습자의 진료 과정을 실시간으로 분석·평가할 수 있도록 하였다. 학습자는 자기주도적으로 학습 내용을 점검할 수 있었으며, 교육자는 일관된 기준 하에 객관적 평가를 수행할 수 있었다.

또한 RAG 구조를 도입하여 LLM의 도메인 지식 한계를 보완하였다. RAG는 한의학 고전과 임상지침 등 외부 지식베이스를 실시간으로 검색하여 LLM의 응답 생성 과정에 반영함으로써, 한의학의 근거 기반 진단 논리를 충실히 구현할 수 있도록 하였다. 이는 기존 LLM이 서양의학 중심 데이터에 편중되어 있는 문제를 완화하고, 한의학의 특수성을 반영한 학습 경험을 제공하였다는 점에서 의의가 있다. 멀티모달 기술을 적용하여 설

진 이미지, 체온 정보, 영상자료 등 시각적 단서를 통합적으로 제시함으로써 학습자가 다양한 감각 정보를 종합적으로 해석할 수 있는 환경을 구축하였으며, 이러한 접근은 실제 임상 현장에서 요구되는 복합적 판단 능력과 통합적 사고력을 향상시키는 데 기여하였다.

CURA 시스템의 설계는 최근 의학교육의 패러다임으로 부상하고 있는 역량 중심 교육의 방향성과 부합하였다. 본 시스템은 학습자가 단순 지식 습득을 넘어 상황 판단, 의사결정, 피드백, 성찰의 순환적 학습 과정을 경험할 수 있도록 설계되었으며, 이를 통해 임상추론 및 의사소통 역량을 강화할 수 있었다. 또한 인적 자원의 제약으로 인해 실습 기회가 제한된 환경에서도 CURA는 공정하고 일관된 학습 환경을 제공함으로써 교육의 형평성을 제고하였다.

그러나 본 연구에는 몇 가지 한계점이 존재하였다. 첫째, 본 연구는 감기 CPX에 한정된 데이터를 활용하였으므로 다양한 질환군으로의 일반화에는 한계가 있었다. 향후 연구에서는 소화기계, 근골격계, 부인과 등 다양한 질환으로 데이터를 확장하여 모델의 범용성을 검증할 필요가 있다. 이를 위해 각 질환군별 CPX 시나리오를 체계적으로 구축하고, 질환 특성에 따른 변증 기준의 차이를 반영하여 데이터 구조화를 정교화할 계획이다. 둘째, CURA의 자동 평가 모듈은 CPX 채점표를 반영하였으나 실제 임상시험 환경에서의 정량적 타당성 검증은 수행되지 않았다. 따라서 향후에는 인간 평가자와의 일치도 및 평가 공정성에 대한 비교 연구가 필요하다. 셋째, 본 연구에서 활용된 멀티모달 데이터는 제한된 표본으로 구성되어 있어 다양한 환자 표현과 영상 자료가 충분히 반영되지 못하였다. 따라서 향후에는 연령, 성별, 체질, 질병 단계 등 다양한 환자 유형을 포괄하는 멀티모달 데이터의 확충과 품질 관리를 통해 모델의 일반화 성능을 강화할 필요가 있다. 특히 설진 이미지나 맥진 등 한의

학적 진단 특성을 반영한 시각·감각 자료의 다양화를 통해 실제 임상 데이터를 보다 충실히 재현하는 것이 요구된다. 넷째, 본 연구는 시스템의 설계 및 기능적 검증에 중점을 두었으며, 실제 교육 현장에서의 장기적 학습 성과 평가는 이루어지지 않았다. 향후에는 CURA를 한의과대학 교육과정에 적용하여 학습자의 임상역량, 자기주도성, 평가 신뢰도 등의 변화를 종합적으로 분석함으로써 교육 효과를 검증할 필요가 있다.

V. 결론

본 연구에서 개발한 CURA는 한의학의 진단 및 변증 과정을 AI 기술과 결합하여 학습자가 실제 환자와 유사한 임상 상황에서 진단과 추론을 학습할 수 있도록 설계된 교육 플랫폼이다. CURA는 CPX 기반 자동 평가와 피드백 기능을 통해 학습자의 진료 과정을 정량화하고, 평가자 간 편차를 최소화함으로써 일관된 학습 경험을 제공하였다. 또한 RAG 기반 지식 통합을 통해 한의학적 근거에 기반한 응답을 생성함으로써, 근거 중심 교육 환경 조성에 기여하였다. CURA는 임상실습의 표준화와 접근성을 향상시켜 평가자 부족 문제를 완화하고, 학습자 간 실습 기회의 형평성을 높였다. 더불어 학습자의 임상추론력, 문제해결력, 의사소통 능력을 통합적으로 향상시키는 역량 기반 교육을 지원한다.

결론적으로 CURA는 한의학 교육의 디지털 혁신의 새로운 방향을 제시하였다. 향후 다양한 질환과 임상 상황으로의 확장, 실제 교육 현장에서의 효과 검증, 정서적 상호작용 강화를 통한 시스템 고도화가 이루어진다면 CURA는 AI 기반 한의 임상교육의 표준 플랫폼으로 발전할 잠재력을 지니고 있다.

감사의 글

본 연구는 한국한의학연구원의 RAG 기반 한의 프라이빗 생성형 AI 서비스 개발 및 임상 실증(KSN2511021)의 지원을 받아 수행되었습니다.

VI. 참고문헌

1. Cross J, Kayalackakom T, Robinson RE, Vaughans A, Sebastian R 외 13명. Assessing ChatGPT's Capability as a New Age Standardized Patient: Qualitative Study. *JMIR Medical Education*. 2025;11:e63353. <https://mededu.jmir.org/2025/1/e63353>
2. Öncü S, Torun F, Ülkü HH. AI-powered standardised patients: evaluating ChatGPT-4o's impact on clinical case management in intern physicians. *BMC Medical Education*. 2025;25:6877-6890. <https://bmcmmededuc.biomedcentral.com/articles/10.1186/s12909-025-06877-6>
3. Liu X, Wu C, Lai R, Lin H, Xu Y, Lin Y, Zhang W. ChatGPT: when the artificial intelligence meets standardized patients in clinical training. *Journal of Translational Medicine*. 2023;21:447. <https://translational-medicine.biomedcentral.com/articles/10.1186/s12967-023-04314-0>
4. Aster A, Ragaller SV, Raupach T, Marx A. ChatGPT as a Virtual Patient: Written Empathic Expressions During Medical History Taking. *Medical Science Educator*. 2025;35:1513-1522. <https://doi.org/10.1007/s40670-025-02342-7>
5. Chan KY, Yuen TH, Co M. Using ChatGPT for medical education: the tec

- hnical perspective. BMC Medical Education. 2025;25(201).<https://doi.org/10.1186/s12909-025-06785-9>
6. Maquilón RG, Uhl J, Schrom-Feiertag H, Tscheligi M. Integrating GPT-Based AI into Virtual Patients to Facilitate Communication Training Among Medical First Responders: Usability Study of Mixed Reality Simulation. JMI R Formative Research. 2024;8:e58623. <https://formative.jmir.org/2024/1/e58623>
 7. Kim JD, Lee HY, Kim JH, Kim CE. Pilot Development of a 'Clinical Performance Examination (CPX) Practicing Chatbot' Utilizing Prompt Engineering. J Korean Med. 2024;45(1):203-214. <https://www.kci.go.kr/kciportal/ci/sereArticleSearch/ciSereArtiView.kci?sereArticleSearchBean.artilId=ART003057037>
 8. Kim JD, Lee HY, Kim JH, Kim CE. Development of an LLM-based CPX Practicing Chatbot for Korean Medicine Education: Implementation of Automated Scoring and Feedback Generation Framework. The Journal of Korean Medicine. 2024;45(4):215-230. <https://jkom.org/journal/view.php?doi=10.13048/jkm.24067>
 9. Han YJ. 한의학 임상실습교육을 위한 AI 기반 환자 챗봇의 사용성과 교육적 효과성. Korean Journal of Acupuncture. 2024;41(1):27-32. <https://www.kci.go.kr/kciportal/ci/sereArticleSearch/ciSereArtiView.kci?sereArticleSearchBean.artilId=ART003062234>
 10. 동의대 AI 기반 한의학 임상 실습 플랫폼 적용 사례. <https://biopharm.deu.ac.kr/www/deu-today.do?articleNo=57165&mode=view&title=AI%EB%A1%9C+%ED%95%9C%EC%9D%98%ED%95%99+%EC%9E%84%EC%83%81+%EC%8B%A4%EC%8A%B5+%EA%B5%90%EC%9C%A1%EC%9D%84+%ED%98%81%EC%8B%A0%ED%95%98%EB%8B%A4>
 11. Lee WY, Han SY, Kim JH, Lee BW, Han YJ, Lee SH. Gen-SynDi: Leveraging Knowledge-Guided Generative AI for Dual Education of Syndrome Differentiation and Disease Diagnosis. Appl Sci. 2025;15(9):4862. <https://www.mdpi.com/2076-3417/15/9/4862>
 12. Cleland JA, Abe K, Rethans JJ. The use of simulated patients in medical education: AMEE Guide No 42. Med Teach. 2009;31(6):477-486. <https://doi.org/10.1080/01421590903002821>
 13. Jeong AR, Lee HY, Shin SW, Im SJ. Challenges and solutions in developing an objective and structured clinical examination for complementary and integrative medicine: A mixed-method approach. Medicine. 2024;103(31):e39058. <https://doi.org/10.1097/MD.00000000000039058>
 14. Bommasani R, et al. On the Opportunities and Risks of Foundation Models. arXiv:2108.07258. 2022. <https://doi.org/10.48550/arXiv.2108.07258>
 15. OpenAI, Achiam J, et al. GPT-4 Technical Report. arXiv:2303.08774. <https://doi.org/10.48550/arXiv.2303.08774>
 16. Lewis P, et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. arXiv:2005.11401. <https://doi.org/10.48550/arXiv.2005.11401>
 17. Li Y, Zeng C, Zhang J, Zhou J, Zou L. MedDiT: A Knowledge-Controlled Diffusion Transformer Framework for Dynamic Medical Image Generation in Virtual Simulated Patient. arXiv:2

- 408.12236. <https://doi.org/10.48550/arXiv.2408.12236>
18. Chu SN, Goodell AJ. Synthetic Patients: Simulating Difficult Conversations with Multimodal Generative AI for Medical Education. 2024.arXiv:2405.19941. <https://doi.org/10.48550/arXiv.2405.19941>
 19. 한국한의약진흥원, <https://nikom.or.kr>
 20. 한의표준임상진료지침 연계 교육도구(CPX) 개발 매뉴얼, https://nikom.or.kr/nckm/html.do?menu_idx=225
 21. [promptingguide.ai](https://www.promptingguide.ai/), <https://www.promptingguide.ai/>
 22. Park SY, Kim CE. 거대언어모델을 활용한 한의학 교육 강화: 교육용 인공지능 개발을 중심으로. J Physiol & Pathol Korean Med. 2023;37(5):134~138. <https://www.kci.go.kr/kciportal/ci/sereArticleSearch/ciSereArtiView.kci?sereArticleSearchBean.artiId=ART003011785>