

# LLM 기반 헬스케어 AI 연구 동향

## LLM-Based Healthcare AI: Research Trends and Challenges

정현태 (H.T. Jeong, htjeong@etri.re.kr)	휴먼증강연구실 책임연구원
임정묵 (J.M. Lim, jmlim21@etri.re.kr)	휴먼증강연구실 책임연구원
오세원 (S.W. Oh, sewonoh@etri.re.kr)	휴먼증강연구실 책임연구원
노경주 (K.J. Noh, kjnoh@etri.re.kr)	휴먼증강연구실 책임연구원
정승은 (S.E. Chung, schung@etri.re.kr)	휴먼증강연구실 선임연구원
이선경 (S.K. Lee, sklee2014@etri.re.kr)	휴먼증강연구실 연구원
정규원 (G.W. Jung, gwjung@etri.re.kr)	휴먼증강연구실 박사후연구원

### ABSTRACT

The rapid advancement of large language model (LLM)-based healthcare AI is transforming personalized health coaching, medical data analysis, clinical decision support, and precision medicine. LLM-based healthcare AI is being increasingly applied in clinical document analysis, patient record summarization, medical consultation support, and automated medical documentation. Additionally, the integration of multimodal AI, which combines text, medical imaging, physiological signals, and genomic data, enhances diagnostic accuracy and predictive capabilities. Despite these advancements, critical challenges remain, including data bias, privacy concerns, reliability, transparency, and regulatory compliance. Addressing these issues is essential for the ethical and effective deployment of AI in healthcare. Future research must focus on optimizing LLM performance, developing domain-specific healthcare models, and establishing continuous data updating and maintenance frameworks to ensure sustainable and trustworthy AI adoption. This paper analyzes emerging trends in LLM-based healthcare AI, evaluates opportunities and risks, and proposes strategic directions for the long-term development and integration of AI in the healthcare sector.

**KEYWORDS** AI, Explainable AI, Healthcare, LLM, Multimodal AI, Privacy & Ethical AI

## 1. 서론

인공지능(AI) 기술의 급속한 발전은 헬스케어 분야를 크게 변화시키고 있으며 그 영향력은 계속 확

대되고 있다. 초기 AI 기반 헬스케어 기술은 주로 데이터 학습에 중점을 두어 인간의 판단과 유사한 진단 결과를 도출하는 데 중점을 두었다. 그러나, 최근 기술의 발전으로 대규모 멀티모달 데이터 융합에

\* DOI: <https://doi.org/10.22648/ETRI.2025.J.400305>

\* 본 연구는 한국전자통신연구원 연구운영비지원사업의 일환으로 수행되었음[25ZB1100, 자율성장형 복합인공지능 원천기술 연구].



기반하여 더욱 정교한 의사결정이 가능해졌으며, 추론 과정에 대한 설명 가능성도 향상되었다. 특히, 대규모 언어 모델(LLM: Large Language Model)의 출현은 데이터셋을 신속하게 분석하고 그에 따른 전략이나 통찰력을 제공함으로써 헬스케어 AI의 응용 가능성을 넓히는 데 중요한 역할을 하고 있다. LLM 기반 헬스케어 AI는 임상 문서 및 연구 논문의 분석, 환자 기록 요약, 의료 상담 지원, 자동 진료 기록 정리 등의 영역에서 활용이 늘고 있다. 이러한 기술 발전은 조기 질병 예측, 진단 지원 시스템, 정밀의료를 포함한 다양한 분야에서 혁신적인 변화를 주도하고 있으나, 동시에 데이터 편향(Bias), 개인정보 보호 문제, 신뢰성, 윤리적 고려사항, 규제 준수 및 의료 시스템과의 통합과 같은 도전 과제가 남아 있다.

본고에서는 LLM 기반 헬스케어 AI 기술의 최신 동향에 대해 분석하고, 기술 개발과 관련된 기회와 위험을 살펴본다. 또한, 헬스케어 분야에서 지속 가능한 기술로 발전하기 위한 전략적 방향을 제안한다.

## II. 연구 동향

디지털 헬스케어는 헬스케어 산업과 정보통신기술이 융합되어 개인의 건강과 질환을 관리하고 증진시키는 서비스와 기술 전반을 포함한다[1,4]. 특히, 헬스케어 AI는 빅데이터, 기계학습, 딥러닝, 자연어처리, 컴퓨터 영상처리 기술 등의 다양한 데이터 처리 기술을 활용하여 질병의 진단과 예측, 치료 계획 수립, 신약 개발, 환자 모니터링 등의 분야에서 의료 서비스의 효율성과 정확성 향상을 견인하는 역할을 하고 있다.

### 1. 헬스케어 AI 기술 현황

초기 헬스케어 AI 모델은 주로 의료 전문가의 지

식을 기반으로 한 규칙 기반 엔진을 통해 특정 증상에 대한 진단이나 치료 방법 제안에 관한 것이었다. 이러한 규칙 기반 AI 시스템은 전문가의 지식에 의존하여, 지식의 범위와 정확성에 따라 성능이 제한되는 특징이 있다[2]. 따라서, 새로운 의학 지식이나 기술 발전에 따라 규칙을 계속해서 업데이트하는데 어려움이 있었고, 복잡한 환자 상태를 분석하거나 다중 질환을 다루는 데 한계가 있었다.

현재 헬스케어 AI는 방대한 의료 데이터를 분석하여 질병 예측, 진단, 치료 계획 수립의 과정을 자동화하고 정밀화하는 기술로 발전하고 있다. 특히, 웨어러블 기기와 의료 사물인터넷(IoMT)의 발전으로 개인의 건강 데이터를 지속적으로 수집·분석하여 질병을 조기에 예측하고 예방하는 기술이 주목받고 있다. 이러한 기술은 만성 질환 관리 및 개인 맞춤형 건강관리 서비스의 핵심 요소로 자리 잡으며, 의료의 패러다임을 치료 중심에서 예방 중심으로 전환하는 데 기여하고 있다[3].

최근에는 수백억 개의 매개변수를 학습하여 다양한 언어 처리 작업을 수행할 수 있는 대규모 언어 모델을 헬스케어 분야에 활용하는 연구가 활발하게 수행되고 있다. 환자 기록, 임상 데이터 등 대규모 텍스트 기반 의료 데이터 분석을 넘어, 멀티모달 및 웨어러블 데이터를 활용하는 연구로 확장되고 있다.

이러한 기술 발전은 의료 서비스의 효율성과 정확성을 향상시키며, 궁극적으로 환자 중심의 정밀의료 서비스 환경 구축에 기여할 것으로 기대된다. 표 1은 헬스케어 AI의 대표적인 활용 분야를 나타낸다.

### 2. LLM 기반 헬스케어 AI의 필요성

전자건강기록(EHR)의 보편화와 함께 환자의 의료 기록, 진단 정보, 처방 내역, 의료 영상 등 방대한 의료 데이터가 축적되면서 이를 효율적으로 정리,

표 1 헬스케어 AI 활용 분야

활용 분야	주요 내용
대규모 의료 데이터 분석	의료 AI를 통해 의료 영상(X-ray, CT, MRI), 생체신호(EEG, ECG), 유전자 데이터, 전자건강기록(EHR) 등의 방대한 데이터를 분석하여 패턴을 학습하고 질병을 예측한다.
정밀의료 및 개인 맞춤형 의료	환자의 유전자 정보, 생활 습관, 의료 기록을 분석하여 개인 맞춤형 치료 계획을 수립하는 데 도움을 준다.
의료 영상 분석	CNN과 같은 딥러닝 모델을 통하여 X-ray, CT, MRI, 병리 슬라이드 등의 의료 영상을 분석하여 질병을 탐지한다.
의료 데이터 자동화	의료 문서 및 전자건강기록(EHR) 등을 자동화한다.
실시간 환자 모니터링	스마트워치, 스마트 밴드 등의 웨어러블 장치에서 수집된 생체신호(심박수, 혈압, 수면 데이터 등)를 분석하여 실시간으로 건강 상태를 모니터링하고 질병 위험을 예측한다.

검색, 분석할 필요성이 더욱 커지고 있다. 최근 인공지능(AI)과 자연어처리(NLP) 기술의 급격한 발전으로 헬스케어 AI는 기존의 이미지 분석(CT, MRI) 및 특정 질병 진단 중심의 데이터 처리에서 벗어나, 의료 문서 요약, 진단 지원, 임상시험 문헌 분석 등 더 광범위한 의료 작업을 수행하는 단계로 발전하고 있다. 특히, LLM을 활용하면 의료 데이터의 맥락을 더욱 깊이 이해하고 복잡한 임상 정보를 종합적으로 분석할 수 있어, 의료 현장에서 더욱 정확하고 신뢰도 높은 의사결정을 지원할 가능성이 더욱 커지고 있다. 이에 따라 LLM 기반 헬스케어 AI의 핵심 역할은 다음과 같이 정리될 수 있다.

### 가. 임상 기록 자동 분석 및 요약

전자건강기록(EHR)과 의사 소견서 등 방대한 임상 텍스트를 자동 분석하고 요약하는 기능이 LLM의 대표적인 활용 사례다. 이를 통해 의료진의 문서 작업 부담을 줄이고, 환자 정보를 검토하는 시간을 단축함으로써 더욱 신속하고 정확한 의사결정을 지원할 수 있다. 예를 들어, 응급실 인계문 자동 생성

LLM[5] 및 환자-의사 대화 기반 임상 노트 자동 생성 LLM[6] 등의 연구는 LLM이 실제 의료 현장에서 보조 도구로 유용하게 활용될 수 있음을 입증하고 있다.

### 나. 진단 지원 및 의사결정 보조

LLM은 의료 지식 기반의 질의응답 시스템이나 임상 의사결정 지원 도구로 활용될 수 있다. 질병 증상과 검사 결과를 입력받아 감별 진단을 수행하고, 치료 옵션 및 권고 조치를 제안함으로써 의료진의 의사결정을 도울 수 있다[4]. 특히, 의학 전문지식에 특화된 LLM은 희귀 질환이나 복잡한 증상에 대한 정보를 신속하게 제공하며, 진단의 근거를 논리적으로 제시할 수 있어 임상 의사결정 지원에 유용하다. 또한, 설명된 증상을 기반으로 잠재적 진단 목록을 생성하고, 적절한 검사와 치료 계획을 제안함으로써 진단의 정확성과 속도를 높이고 의료진의 판단을 강화하는 역할을 한다[7].

### 다. 환자 상담 챗봇 및 의료 질의응답

LLM 기반 의료 챗봇은 자연스러운 대화를 통해 환자의 증상이나 궁금증을 분석하고, 방대한 의료 지식을 바탕으로 맞춤형 답변을 생성하여 의료 조언을 제공할 수 있다.

또한, 개인별 지속적인 건강관리와 모니터링 기능을 갖춘 AI 시스템을 통해 의료 서비스의 연속성을 높이고, 환자 중심의 맞춤형 건강관리를 지원할 수 있다[8].

### 라. 신약 개발 및 임상시험 최적화

LLM은 신약 후보 물질 발굴, 약물 상호작용 예측, 임상시험 설계 최적화 등 제약 및 의학 연구 분야에서 중요한 역할을 수행하고 있다. 특히, 방대한 생물학 지식은 기반으로 약물-단백질 상호작용

용, 부작용 예측, 유망한 치료 표적 식별, 기존 약물의 적응증 확대 등에도 활용되고 있다. 또한, 임상시험 데이터 및 프로토콜 문서를 분석하여 최적의 환자 코호트를 선별하고, 임상시험 설계를 최적화함으로써 신약 개발의 전반적인 과정에서 의사결정 속도를 높이고 연구 효율성을 개선할 수 있다[9].

### 3. LLM 기반 헬스케어 AI 데이터셋

LLM 기반 헬스케어 AI 모델을 학습하기 위해서는 고품질의 데이터셋이 필수적이다. 의료 시험 및 문서에서부터 의료 영상, 웨어러블 센서 데이터까지 다양한 유형의 데이터셋이 활용되고 있다.

#### 가. 의료 자연어처리(NLP) 데이터셋

자연어처리 기술은 의료 문서 분석, 질문 응답, 진단 리포트 요약 등 각종 의료 작업에서 활용되고 있다. MIMIC-III[10]는 대규모 전자건강기록 데이터셋으로, 익명화된 중환자실 보고서를 포함한다. 해당 데이터셋은 LLM의 사전학습뿐만 아니라 임상 노트 분류, 의료 리포트 요약 등의 작업에 활용될 수 있다. MedQA[11]는 미국 의사면허시험(USMLE)에서 출제된 객관식 문제로 구성되어 있다. 이 데이터셋은 일반적인 의학 지식을 다루는 미국 의료면허 취득에 필요한 의학적 지식과 논리적 사고 능력을 평가하기 위해 설계되었으며, 의료 QA 시스템 개발 및 평가에 중요한 역할을 하고 있다.

#### 나. 의료 멀티모달 데이터셋

최근에는 텍스트뿐만 아니라 의료 영상을 포함하는 멀티모달 데이터를 활용하는 연구가 증가하고 있다. VQA-RAD[12]는 의료 영상과 연계된 질의 응답 작업을 위한 대표적인 데이터셋이다. 두부, 흉부, 복부 CT, MRI, X-ray 영상과 함께 의료 전문가

가 제공한 정답으로 구성되어 있어 영상 판독을 보조하는 AI 개발에 사용된다. MIMIC-CXR[13]은 흉부 X-ray와 이에 대한 의료 전문가 소견을 기반으로 한 데이터셋으로 의료 영상 설명 생성과 같은 멀티모달 자연어처리 연구에 활용된다. 최근에는 이들을 포함한 12개의 오픈소스 데이터셋으로 구성된 MultiMedBench[12]가 등장했으며, 의료 영상, 유전체 데이터 등 다양한 모달리티와 분류, 요약, 해석 등 14가지 작업을 지원한다.

#### 다. 웨어러블 생체신호 데이터셋

웨어러블 기기를 통한 건강 모니터링이 활성화됨에 따라, 생체신호 데이터 활용의 중요성이 높아지고 있다. PMData[14]와 LifeSnaps[15]는 대표적인 웨어러블 생체신호 데이터셋으로, FitBit과 같은 웨어러블 장치와 설문지를 통해 심박수, 걸음 수, 수면 정보 등을 수집했다. MIT는 PMData와 LifeSnaps가 포함된 4개의 오픈소스 데이터를 바탕으로 정신 건강, 수면품질 등 10개의 건강 예측 문제를 수행할 수 있는 Wearable Sensor 데이터셋을 구성했다[22]. Google은 생체신호를 바탕으로 각각 수면과 운동에서 도메인 전문가를 통한 분석 및 추천이 포함된 데이터셋을 생성했고, 이를 바탕으로 웨어러블 데이터에 대한 LLM의 연구 방향과 사용 가능성을 제시했다[23].

### 4. LLM 기반 헬스케어 AI 모델

빅테크 기업들과 주요 대학 연구진이 LLM 기반 헬스케어 AI 모델 개발에 적극적으로 참여하면서, 텍스트 기반 의료 데이터뿐만 아니라 멀티모달 및 웨어러블 데이터를 활용하는 연구가 점차 확대되고 있다. 연구의 방향 또한 사전학습(Pre-Training) 중심에서 높은 성능을 보이는 모델들을 활용한 미세조

표 2 주요 LLM 기반 헬스케어 AI 모델

모델	최초 등록일	기본 모델	기본 모델 활용 방법	주 용도	주 활용 데이터
GatorTron [16]	22.02	BERT	PT + FT	Medical NLP	EHR
Med-PaLM [17]	22.12	Flan-PaLM	FT	Medical QA	Medical Exam, Literature
Med-PaLM 2 [18]	23.05	PaLM 2	FT	Medical QA	Medical Exam, Literature
LLaVA-Med [19]	23.06	LLaVA	FT	Medical VQA	Biomedical Images
Med-PaLM M [20]	23.07	PaLM E	FT	Multimodal Medical NLP	Multimodal Biomedical Data
HealthAlpaca [21]	24.01	Alpaca	FT	Health Prediction	Wearable Data
Med-Gemini [22]	24.04	Gemini	FT	Multimodal Medical NLP	Multimodal Biomedical Data
PH-LLM [23]	24.06	Gemini	FT	Health Insights Generation	Wearable Data
PHIA [24]	24.06	Gemini	Agent	Health Insights Generation	Wearable Data
PhysioLLM [25]	24.06	GPT-4	Chat	Health Insights Generation	Wearable Data

정(Fine-Tuning) 및 에이전트 기반 접근 방식으로 변화하고 있다. 표 2는 현재 주목받고 있는 주요 LLM 기반 헬스케어 AI 모델을 보여준다.

가. 자연어 기반 모델

초기 LLM 기반 헬스케어 AI 모델 연구는 전자건강기록, 의료 시험 문제와 같은 텍스트 데이터를 활용한 자연어처리 모델 개발에 집중되었다. 대표적인 사례로 NVIDIA와 플로리다 대학교가 공동 개발한 GatorTron은 BERT 구조를 기반으로 방대한 전자건강기록 데이터를 사용하여 사전학습 및 미세조정을 수행하고, 의료 개체 인식, 관계 추출, 문서 분류 등 5가지 핵심 의료 자연어처리 작업을 수행하도록 설계되었다[24]. Google 또한 의료 QA에 특화된 Med-PaLM 시리즈를 개발하면서, 의료 도메인에 최적화된 LLM 연구를 확장해 나갔다. 2022년에 공개된 Med-PaLM은 Flan-PaLM을 의료 도메인에 맞게 개선한 모델로, MedQA 데이터셋에서 67.2%의 점수를 달성하며 최초로 미국 의사면허시험 스타일 질문에 합격 점수를 넘었다[25]. 후속 모델인 Med-PaLM 2는 PaLM 2를 기반으로 의료 도메인 미세조정과 새로운 앙상블 개선 접근법을 적용하여 성능을

향상시켰다. MedQA 데이터셋에서 86.5%의 정확도를 달성하였으며, 장문형 질문에 대한 인간 평가에서도 의사 답변 대비 높은 선호도를 받았다[18].

나. 영상-자연어 멀티모달 기반 모델

LLM 연구가 발전함에 따라 헬스케어 AI도 단순한 텍스트 기반 자연어처리를 넘어 의료 영상, 유전체 데이터 등과 결합한 멀티모달 AI 모델로 발전하고 있다. Microsoft의 LLaVA-Med는 영상-언어 모델 LLaVa를 미세조정된 모델로, 의료 영상(CT, MRI, X-ray 등)을 이해하고 분석할 수 있도록 학습되었다. 이 모델은 의료 영상과 자연어를 결합하여 Medical VQA 작업을 수행하며, 의료 영상 데이터에서 의미 있는 정보를 추출하여 질의응답 형식으로 제공하는데 강점을 가진다[19]. Med-PaLM M은 멀티모달 의료 인공지능으로, 텍스트뿐만 아니라 의료 이미지, 유전체 데이터 등 다양한 의료 데이터를 처리할 수 있다. 각 모달리티를 이미지 형태로 전환하여 하나의 모델로 14가지의 다양한 의료 작업을 수행하며, 일부 작업에서는 특화된 모델보다 우수한 성능을 보였다[20]. Med-Gemini는 임상 추론 능력을 강화하기 위해 자가 학습 및 웹 검색 기능을 도입했으

며, 미세조정과 맞춤형 모달리티 인코더를 바탕으로 멀티모달 성능을 향상시켰다. 14개의 의료 벤치마크에서 평가한 결과, 10개에서 최고 성능을 달성하였다[22].

#### 다. 웨어러블 데이터 활용 모델

최근에는 환자 진료가 아닌 개인 건강관리 측면에서 웨어러블 디바이스 데이터와 LLM을 접목하려는 시도들이 등장하고 있다. 이 연구들은 스마트 워치나 웨어러블 센서로부터 얻는 생체신호 시계열 데이터와 개인 문맥 정보를 LLM에 통합하여, 개인별 맞춤형 건강 예측 및 코칭을 목표로 한다. MIT의 HealthAlpaca는 공개 웨어러블 데이터셋을 기반으로 정신건강, 신체활동, 대사 건강, 수면 평가 등 사용자 건강 예측 작업을 수행한다. 특히 컨텍스트 강화 기법을 적용하여 사용자 프로필, 건강 지식, 시간 정보를 조합함으로써 예측 성능을 23.8% 향상할 수 있음을 보였다[21]. Google의 PH-LLM은 수면 및 운동 관련 개인 맞춤형 건강 코칭을 목표로 Gemini를 미세조정된 모델이며, 코칭 추천, 전문가 지식 평가를 위한 객관식 시험, 환자가 보고한 건강 상태 예측 작업을 수행함으로써 실제 환경에서 건강 이해 및 코칭 역량을 보였다[23]. Google의 PHIA는 개인 건강 통찰을 제공하는 LLM 기반 에이전트로, 코드 실행 및 웹 검색 기능을 활용하여 다단계 추론을 수행한다. 기존의 단순한 문장 생성 방식이 아니라, 코드 생성 기반 데이터 분석과 신뢰할 수 있는 건강 정보 검색을 병행함으로써 더욱 정확하고 실용적인 건강 조언을 생성한다[24]. MIT의 PhysioLLM은 GPT-4를 기반으로 사용자의 건강을 분석하고 맞춤형 인사이트를 제공하는 대화형 AI 시스템이다. 웨어러블 데이터의 심층적인 통계 분석을 통해 의미 있는 패턴을 발견하고, 이를 사용자와 자연어로 소통할 수 있도록 설계되었다[25].

### III. 도전 과제

강력한 디지털 인프라와 양질의 국민 건강 정보를 효과적으로 활용할 수 있는 역량을 갖춘 국가들이 COVID-19 위기에 유연하게 대응하여 빠르게 회복해 낸 이후로[26], 이제 전 세계 각국은 인공지능을 의료 분야에 접목함으로써 개인 맞춤형 건강 관리와 질병 예방 중심의 의료 패러다임 전환을 가속화하고 있다[27-29]. 그러나 헬스케어 분야는 생명과 건강에 직결된 분야로, 단순한 시장경제 논리보다는 사회적·정치적 합의를 통해 발전하는 특성을 지닌다[30]. 국내에서도 복잡한 규제와 시장 참여자 간의 이해관계가 복잡하게 얽혀 있어 LLM을 포함한 AI 기술의 도입이 쉽지 않으며, 헬스케어 AI 시대로의 전환에는 상당한 시간이 소요될 가능성이 크다[31]. 이에 따라, 본 장에서는 헬스케어 AI 도입의 주요 도전 과제를 검토하고, 이를 해결하기 위한 발전 방향을 모색하고자 한다.

#### 1. 법적·윤리적 문제

##### 가. 프라이버시 침해 및 데이터 보안

헬스케어 및 의료 데이터는 환자의 개인정보와 진료 기록 등 민감한 정보를 포함하므로, AI 모델의 학습 및 활용 과정에서 기밀 정보 유출 위험이 존재한다[32,33]. 예를 들어, OpenAI의 GPT-2 모델은 악의적인 질의 프롬프트를 통해 학습에 사용된 개인 식별 정보가 노출된 사례가 보고된 바 있다. 이를 방지하기 위해 가명·익명 처리, 차분 프라이버시(Differential Privacy), 연합학습 등의 기법이 고려되고 있으며[34], 데이터 유출 및 오용 방지를 위한 보안 프로토콜과 개인정보 접근 관리 정책의 강화가 필요하다[35].

**나. 윤리성 및 책임 소재**

헬스케어 AI의 의사결정 오류로 인해 환자에게 피해가 발생할 경우, 책임 소재를 둘러싼 법적 분쟁이 발생할 가능성이 크다. 따라서, 대부분의 AI 기술은 의료진의 최종 판단을 보조하는 역할로 권장되고 있다[36]. 또한, 환자가 의료진을 신뢰하고 치료 과정에 대한 신념을 가질수록 치료 효과가 높아진다는 연구 결과도 보고된 바 있다[37]. 이에 따라, 헬스케어 AI의 활용 과정에서 환자에게 충분한 설명과 사전 동의를 제공하고, 의료 행위의 최종 결정권은 인간 의료진에게 있음을 명확히 해야 한다[38]. 아울러, 사회 윤리적 이슈를 고려하여 의료진과 환자 모두가 AI 기술을 올바르게 이해하고 활용할 수 있도록 교육과 훈련이 필수적이다.

**다. 규제 준수 및 기술 통제**

헬스케어 분야 기술이 빠르게 발전하는 반면, 의료기기의 임상 활용을 위한 규제 지침은 상대적으로 정립 속도가 느리다. 이에 따라, 기존 의료기기 승인 평가제도에서 LLM 기반 헬스케어 AI의 성능과 안전성을 검증하는 과정은 상당한 시간과 자원을 요구할 수밖에 없다. 특히, 생성형 인공지능의 주요 이슈인 환각(Hallucination) 문제를 해결하기 위해 정량적 평가 기준을 마련하고 이를 통제할 수 있는 규제 지침과 기술적 체계를 구축할 필요가 있다.

**2. 기술적 문제**

**가. 신뢰성 및 정확도**

헬스케어 AI의 판단과 예측 결과를 임상에 활용하려면, 모델이 생성하는 정보의 신뢰성과 정확도를 평가할 수 있는 기준과 척도를 사전에 마련해야 한다. 특히, 개방형 인터넷 데이터를 기반으로 학습된 범용 LLM은 의학적으로 부정확한 정보를 답변

에 반영하거나, 근거 없는 내용을 그럴듯하게 생성할 위험이 있다. 문제는 이러한 오류가 신뢰할 수 있는 정보처럼 보일 수 있어, 비전문가가 이를 분별하기 어렵고 치명적인 의료 오류로 이어질 가능성이 크다는 점이다. 따라서, 헬스케어 AI의 신뢰성을 보장하고 정확도를 향상시키기 위한 기술적 해결책 마련이 필수적이다.

**나. 데이터 편향성**

편향된 데이터로 훈련된 인공지능 모델은 특정 인구집단(인종, 나이, 성별, 직업, 거주 지역 등)에 대한 오진 위험을 높이고, 의료 격차를 심화시킬 수 있다. 또한, LLM이 사회문화적 편향을 학습하여 차별적인 답변을 생성한 사례도 보고된 바 있다. 이처럼 데이터 편향으로 인한 오류는 의료 의사결정의 공정성을 저해하고, 특정 환자군에게 불이익을 초래할 가능성이 크다. 따라서, 학습 데이터의 불균형을 해소하기 위한 공정성 검증과 편향 완화 방안을 마련하고, 이를 지속적으로 개선하는 노력이 필요하다.

**다. 설명 가능성 및 투명성**

인공지능 모델의 의사결정 과정이 불투명하다는 점은 의료 분야에서 AI 도입을 가로막는 주요 장애물 중 하나다. 설명 가능성(Explainability)이 부족하면, AI가 잘못된 판단이나 조언을 내렸을 때 이를 즉각 수정하기 어렵고, 의료진과 환자의 신뢰를 얻기도 힘들다. 이를 해결하기 위해, 최근에는 AI의 추론 과정을 명확히 제시하는 LLM 기술이 개발되고 있다. 예를 들어, 생각의 사슬(Chain-of-Thought) 기법을 적용하면 AI가 답변을 생성할 때 여러 단계의 이유를 먼저 나열한 후 최종 답을 도출하도록 설계할 수 있다. 또한, 입력 문장에서 추출한 의료 지식을 구조화하여 사람이 이해할 수 있는 형태로 제공하

는 방식도 제안되고 있다. 이와 함께, AI 모델이 자신의 불확실성을 확률값이나 신뢰 점수 형태로 제공하여, 의료진이 답변의 신뢰도를 직접 평가할 수 있도록 지원하는 연구도 진행되고 있다.

### 3. 헬스케어 도메인 특화 문제

#### 가. 특화 모델 개발

LLM을 활용한 헬스케어 AI 솔루션은 방대한 의료 지식을 기반으로 의료진의 의사결정을 지원하고, 환자 맞춤형 건강 정보를 제공하며, 의료 시스템의 효율성을 향상시킬 것으로 기대된다. 그러나 대부분의 고성능 LLM은 범용 언어 모델로 개발되어 있어, 의료 분야의 특수성을 반영한 맞춤형 LLM 기술 개발이 필수적이다[30]. 의료특화 LLM은 의료 데이터를 활용한 사전학습과 미세조정으로 구분되어 구축될 수 있다. 또한, 의료 데이터의 다양성과 품질을 높여 다양한 의료 상황에 대응할 수 있는 대형 모델로 구축하거나, 특정 의료 분야에만 고도로 전문화된 소형 모델로 구축할 수도 있다.

#### 나. 고성능 연산 자원 요구

현재 LLM을 훈련하고 운영하기 위해서는 수십억~수천억 개의 파라미터를 처리할 수 있는 고성능 서버와 전력 자원이 필요하다. 의료 분야의 특성상 프라이버시 보호 및 데이터 보안 문제로 인해 클라우드 환경이 아닌 자체 구축 서버에서 LLM을 운영하려는 경우, 인프라 투자 비용이 AI 도입의 부담으로 작용한다. 중소 병원이나 의원급 의료기관은 거대한 모델을 운영할 수 있는 전용 서버와 인프라를 갖추기 어려워 LLM 도입에 대한 장벽이 높으며, 결과적으로 대형 의료기관이나 기업 중심으로 활용이 제한될 가능성이 크다.

#### 다. 지속적 학습 및 유지보수

의료 지식은 빠르게 변화하기 때문에, LLM이 최신 의학 정보 및 치료 지침을 반영할 수 있도록 지속적인 업데이트가 필수적이다. 그러나 이미 학습된 지식의 규모가 커질수록, 전체 모델을 반복적으로 재훈련하거나 추가 학습하는 과정은 기술적으로 복잡하며 상당한 비용이 소요된다. 또한, 배포 활용 중인 LLM의 성능을 관찰하면서 사용자 피드백을 반영하여 개선하는 체계를 갖추어 나갈 필요가 있다. 이를 위해서는 LLM 라이프사이클 전반을 고려한 운영 인프라 구축에 대한 고려가 필요하며 이러한 유지보수 비용을 정당화할 만한 효용을 입증하는 것이 헬스케어 AI 도입의 중요한 과제 중 하나이다.

## IV. 결론

헬스케어 AI 기술은 의료 진단, 개인 맞춤형 치료, 자동화된 임상 의사결정 지원 등 다양한 분야에서 혁신을 주도하며, 의료 서비스의 효율성, 정확성, 접근성을 향상시키고 있다. 그러나 개인정보 보호, 윤리적 책임, 법적 규제, 데이터 편향, 신뢰성 확보 등 여전히 해결해야 할 중요한 과제가 남아 있다.

LLM 기반 헬스케어 AI의 발전으로 국내 의료 환경을 반영한 특화 모델의 필요성이 커지고 있으며, 특정 의료 분야에 최적화된 모델 개발이 요구되고 있다. 이러한 모델은 빠르게 변화하는 의료 데이터에 적응할 수 있도록 지속적인 학습과 유지보수가 가능하도록 설계되어야 한다. 또한, 텍스트, 의료 영상, 생체신호 데이터를 통합하는 멀티모달 AI를 활용하여 더욱 정밀한 질병 패턴 분석과 의사결정 지원이 가능하도록 발전해야 한다. 기술 개발뿐만 아니라 표준화된 규제 프레임워크와 강력한 데이터 거버넌스 정책을 마련하는 것이 윤리적 AI 구현에 필수적이다.

이러한 다각적인 노력을 통해, LLM 기반 헬스케어 AI는 신뢰할 수 있는 기술로 자리 잡고, 개인의 건강과 안전을 보장하는 지속 가능한 AI로 발전할 것으로 기대된다.

**약어 정리**

AI	Artificial Intelligence
BERT	Bidirectional Encoder Representations from Transformers
CNN	Convolutional Neural Network
DL	Deep Learning
ECG	Electrocardiogram
EEG	Electroencephalogram
EHR	Electronic Health Record
HIT	Health Information Technology
ICU	Intensive Care Unit
IoMT	Internet of Medical Things
LLM	Large Language Model
ML	Machine Learning
NLP	Natural Language Processing
RL	Reinforcement Learning
USMLE	United States Medical Licensing Examination
VQA	Visual Question Answering
XAI	Explainable AI

**참고문헌**

[1] 박민서, "디지털헬스케어에서의 인공지능 적용 사례 및 고찰," 한국융합학회논문지, 제13권 제1호, 2022, pp. 141-147.

[2] W.J. Clancey, "The epistemology of a rule-based expert system—a framework for explanation," *Artif. Intell.*, vol. 20, no. 3, 1983, pp. 215-251.

[3] 삼정KPMG 경제연구원, "AI로 촉발된 헬스케어 산업의 대전환," *Samjong Insight*, 제89권, 2024.

[4] G.E. Iyawa et al., "Digital Health Innovation Ecosystems: From Systematic Literature Review to Conceptual Framework," *Procedia Comput. Sci.*, vol. 100, 2016, pp. 244-252.

[5] V. Hartman et al., "Developing and Evaluating Large Language Model-Generated Emergency Medicine Handoff Notes," *JAMA Netw Open*, vol. 7, no. 12, 2024.

[6] J. Giorgi et al., "Wanglab at medqa-chat 2023: Clinical note generation from doctor-patient conversations using large language models," in *Proc. Clin. Natural Lang. Process. Workshop*, (Toronto, Canada), Jul. 2023, pp. 323-334.

[7] 최소영, 강재우, "생성형 AI, 헬스케어 산업의 미래," *Bio Economy Brief*, 제193호, 2024.

[8] E.K. Jo et al., "Understanding the Benefits and Challenges of Deploying Conversational AI Leveraging Large Language Models for Public Health Intervention," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, (Hamburg, Germany), Apr. 2023, no. 18, pp. 1-16.

[9] M. Moret et al., "Leveraging molecular structure and bioactivity with chemical language models for de novo drug design," *Nat. Commun.*, vol. 14, no. 114, 2023.

[10] A.E.W. Johnson et al., "MIMIC-III, a freely accessible critical care database." *Sci. Data*, vol. 3, no. 160035, 2016.

[11] D. Jin et al., "What disease does this patient have? a large-scale open domain question answering dataset from medical exams," *Appl. Sci.*, vol. 11, no. 14, 2021.

[12] J.J. Lau et al., "A dataset of clinically generated visual questions and answers about radiology images," *Sci. Data*, vol. 5, no. 180251, 2018.

[13] A.E.W. Johnson et al., "MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports," *Sci. Data*, vol. 6, no. 317, 2019.

[14] V. Thambawita et al., "Pmdata: a sports logging dataset," in *Proc. ACM Multimedia Syst. Conf.*, (Istanbul, Turkey), May. 2020, pp. 231-236.

[15] S. Yfantidou et al., "LifeSnaps, a 4-month multi-modal dataset capturing unobtrusive snapshots of our lives in the wild," *Sci. Data*, vol. 9, no. 663, 2022.

[16] X. Yang et al., "A large language model for electronic health records," *NPJ Digit. Med.*, vol. 5, no. 194, 2022.

[17] K. Singhal et al., "Large language models encode clinical knowledge," *Nature*, vol. 620, 2023, pp. 172-180.

[18] K. Singhal et al., "Toward expert-level medical question answering with large language models," *Nat. Med.*, vol. 31, 2025, pp. 943-950.

[19] C. Li et al., "Llava-med: Training a large language-and-vision assistant for biomedicine in one day," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, (New Orleans, LA, USA), Dec. 2023, pp. 28541-28564.

[20] T. Tu et al., "Towards generalist biomedical AI," *Nejm AI*, vol. 1, no. 3, 2024.

[21] K. Saab et al., "Capabilities of gemini models in medicine," *arXiv preprint*, 2024. doi: 10.48550/arXiv.2404.18416

- [22] Y.B. Kim et al., "Health-llm: Large language models for health prediction via wearable sensor data," in Proc. Conf. Health, Inference, Learn., (New York, NY, USA), Jun. 2024, pp. 522-539.
- [23] J. Cosentino et al., "Towards a personal health large language model," arXiv preprint, 2024. doi: 10.48550/arXiv.2406.06474
- [24] M.A. Merrill et al., "Transforming wearable data into health insights using large language model agents," arXiv preprint, 2024. doi: 10.48550/arXiv.2406.06464
- [25] C.M. Fang et al., "Physiollm: Supporting personalized health insights with wearables and large language models," in Proc. IEEE-EMBS Int. Conf. Biomed. Health Inform., (Huston, TX, USA), Nov. 2024.
- [26] OECD, "Health at a glance 2023: OECD indicators," OECD Publishing, 2023.
- [27] D.H. Lee and S.N. Yoon, "Application of artificial intelligence-based technologies in the healthcare industry: Opportunities and challenges," Int. J. Environ. Res. Public Health, vol. 18, no. 1, 2021.
- [28] 배재권, "스마트 헬스케어 산업의 인공지능 기술 적용에 관한 연구," 로고스경영연구, 제19권 제4호, 2021, pp. 179-196.
- [29] D. Ho et al., "Enabling technologies for personalized and precision medicine," Trends Biotechnol., vol. 38, no. 5, 2020, pp. 497-518.
- [30] 삼일PwC경영연구원, "디지털 헬스케어의 개화: 원격의료의 현주소," PwC Korea Insight Research, 2022. [https://www.pwc.com/kr/ko/insights/insight-research/samilpwc\\_](https://www.pwc.com/kr/ko/insights/insight-research/samilpwc_paradigm-shift-july2022.pdf)  
[paradigm-shift-july2022.pdf](https://www.pwc.com/kr/ko/insights/insight-research/samilpwc_paradigm-shift-july2022.pdf)
- [31] S.B. Lee et al., "A Trend of Artificial Intelligence in the Healthcare," J. Korea Contents Assoc., vol. 20, no. 5, 2020, pp. 448-456.
- [32] R. Yang et al., "Large language models in health care: Development, applications, and challenges," Health Care Sci., vol. 2, no. 4, 2023, pp. 255-263.
- [33] E.J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," Nat. Med., vol. 25, 2019, pp. 44-56.
- [34] S.M. Williamson and V. Prybutok, "Balancing privacy and progress: a review of privacy challenges, systemic oversight, and patient perceptions in AI-driven healthcare," Appl. Sci., vol. 14, no. 2, 2024.
- [35] B. Murdoch, "Privacy and artificial intelligence: challenges for protecting health information in a new era," BMC Med. Ethics, vol. 22, 2021.
- [36] D.D. Luxton, "Artificial intelligence in psychological practice: Current and future applications and implications," Prof. Psychol.: Res. Pract., vol. 45, no. 5, 2014.
- [37] T.J. Kaptchuk and F.G. Miller, "Placebo effects in medicine," N. Engl. J. Med., vol. 373, no. 1, 2015, pp. 8-9.
- [38] S. Gerke et al., "Ethical and legal challenges of artificial intelligence-driven healthcare," Artif. Intell. Healthc., 2020, pp. 295-336.