

헬스케어 생성형 인공지능 연구에 대한 기관생명윤리위원회의 역할과 생명윤리법 개정 방향: 주제범위 문헌고찰에 기초하여*

김준혁**

[요약]

헬스케어 인공지능, 즉 건강 및 의료와 관련한 인공지능 알고리즘의 빠른 성장은 관련 연구에 대한 거버넌스 체계의 시급한 확립을 요구하고 있으나, 현재 그에 대한 구체적인 대안이 제시되고 있지 못하고 있다. 본 연구는 생명윤리및안전에관한법률이 인간에 대한 연구를 규정하는 한, 헬스케어 인공지능 연구는 인간으로부터 얻은 데이터를 활용하거나 그 적용 대상을 인간으로 하고 있으므로 생명윤리및안전에관한법률이 적용되어야 함에서 출발하였다. 먼저, 지금까지 헬스케어 인공지능과 생명윤리및안전에관한법률의 관계를 다룬 여러 법적, 학술적 논의가 주로 익명화 이슈에 한정되어 있음을 주제범위 문헌고찰을 통해 분석하였다. 헬스케어 인공지능 연구에서 다루어져야 하는 이슈는 익명화만이 아니므로, 관련 연구와 관련하여 기관생명윤리위원회의 심의 역할이 검토되어야 하는바, 여러 문헌을 바탕으로 이를 점검하였다. 또한, 2026년 시행 예정인 인공지능발전과신뢰기반조성등에관한기본법이 윤리원칙 준수를 위해 제안한 민간자율인공지능윤리위원회는 그 기능상 헬스케어 영역에서 기관생명윤리위원회와 역할이 중복되므로, 기관생명윤리위원회가 대신 그 역할을 수행할 수 있도록 생명윤리및안전에관한법률을 정비할 필요성을 제기하였다. 종합하여, 본 논문은 헬스케어 인공지능 연구 거버넌스 체계 정립을 위해 기관생명윤리위원회의 구체적인 역할을 제안하고 이를 위한 생명윤리및안전에관한법률 개정을 촉구하였다.

[색인어]

헬스케어 생성형 인공지능, 생명윤리법, 기관생명윤리위원회, 인공지능기본법, 민간자율인공지능윤리위원회

* 이 논문은 질병관리청 학술연구개발용역과제 '헬스케어·인공지능 연구를 위한 연구윤리 교육 프로그램 운영 및 윤리지침 개선' 및 '디지털 헬스케어 분야 생성형 인공지능 연구 윤리 가이드 개발' 지원을 받아 수행된 연구임(2023-ER0808-00, 2025-ER0801-00).

** 연세대학교 치과대학 치의학교육학교실, 조교수. e-mail: MOLE0619@yuhs.ac

I. 서문

인공지능(이하 AI), 특히 건강과 질병을 다루는 헬스케어 AI의 연구에 있어 감독과 관리, 즉 거버넌스가 필요하다는 것은 굳이 별도의 논의가 필요하지 않을 정도로 분명한 사실이다. 일반적인 AI 연구와 달리 헬스케어 AI는 개인정보이자 민감정보인 환자의 건강 데이터¹⁾를 기반으로 하여 연구·개발되고,²⁾ 환자나 일반인의 건강 및 질병 관련 선택에 영향을 미치거나 치료 및 예방 관련 수행에 직간접적으로 개입하기 때문이다. 이런 헬스케어 AI 연구가 지니는 인간 대상 데이터 습득 및 수행이라는 특징은 직접적으로 인간의 자율성을 저해하거나 해악을 끼치며 불평등한 결과를 유발할 수 있다는 점에서 대응이 필요하다.³⁾ 그러나, 헬스케어 AI 연구 거버넌스를 어떻게 가져갈 것인지에 대한 논의는 아직 분명하지 않은데, 이는 오랜 전통을 쌓아온 생명과학 연구윤리 체계에서 AI 기술이 보이는 이질성과,⁴⁾ 컴퓨터 공학에서 연구윤리 체계가 구체적으로 정립되어 본 적이 없는 역사적 상황에서 기인하는 것으로 보인다.⁵⁾

물론, 이에 대한 조사자료가 있는 것은 아니나 보건의료 연구자가 수행하는 헬스케어 인공지능 연구는 기관윤리생명위원회(이하 IRB)의 심의 아래 진행되고 있다. 그러나, 이 과정에 있어서 오해가 발생하기도 하는데, 개인정보보호위원회가 발표·개정하고 있는 『보건의료데이터 활용 가이드라인(개정안)』⁶⁾은 연구대상자 식별 가능 정보를 이용하는 연구를 인간대상연구에 포함되고 있다고 규정하고 있으면서도, 기관보건의료데이터심의위원회(이하 DRB)의 승인을 받

1) 2025년 현재, 아직 환자 및 일반인으로부터 얻은 것으로서 보건의료 영역 연구에 사용되는 데이터를 가리키는 단어 및 그 규정이 확립되지 않았다. 보건의료 데이터, 건강데이터, 헬스케어 데이터 등 다양한 표현이 사용되고 있으며, 본 논문의 목적상 이들 개념을 엄밀하게 정의하고 분류하기보다는 비슷한 범주를 가리키는 가족 유사성 개념으로 파악하고 논지를 전개한다. 따라서, 이하에서 관련 데이터에 대한 논의를 다루면서 각 문서 및 자료가 사용하는 표현을 그대로 유지하였다.

2) 차현재, 김준혁. “건강데이터 기증 및 공유의 윤리적 접근: 인체유래물 기증 절차를 참조하여”. 『생명, 윤리와 정책』. 제 6권 제2호. 2022. 101-137면.

3) World Health Organization. 『Ethics and Governance of Artificial Intelligence for Health』. Geneva: World Health Organization. 2021.

4) J. Kim, S. Y. Kim, E.-A. Kim, et al. “Developing a Framework for Self-regulatory Governance in Healthcare AI Research: Insights from South Korea”. 『Asian Bioethics Review』. 16(3). 2024. pp.391-406.

5) 정보기술의 윤리에 대한 접근의 시작으로 1985년 제임스 무어(James Moor)의 제안이 꼽히며, 그는 컴퓨터 윤리(computer ethics)를 제안하지만, 그것은 “컴퓨터 기술의 자연·사회적 영향 분석과 그에 따른 기술의 윤리적 활용을 위한 정책의 공식화 및 정당화”로 정의되어 엄밀한 윤리학적 접근보다는 기술정책평가에 가깝다. K. W. Miller, M. Taddeo(Ed). 『The Ethics of Information Technologies』. Oxford, UK: Routledge. 2016; J. H. Moor. “What is Computer Ethics?”. 『Metaphilosophy』. 16(4). 1985. pp.266-275 참조.

6) 개인정보보호위원회. 『보건의료데이터 활용 가이드라인(개정안)』. 2024.12.

은 과제는 7일 이내에 IRB 심의면제 확인서를 발급받을 수 있다고 명시하였다.⁷⁾ 국가생명윤리 정책원에서 발간한 『바이오헬스 데이터 이용 연구에 대한 기관위원회 운영 가이드라인 - 개인정보의 2차 사용을 중심으로 -』 또한 같은 문구를 싣고 있다.⁸⁾ 그러나, 이것은 IRB와 DRB의 역할에 대한 혼동에서 발생한 오류이자,⁹⁾ 기존 연구에 대한 관리 기준만을 따르다 보니 헬스케어 AI 연구의 특수성을 반영하지 못하여 나타난 관리 미흡의 사례다.

먼저, IRB는 「생명윤리및안전에관한법률(이하 생명윤리법)」 제10조에 의거, 연구계획서의 윤리적·과학적 타당성, 연구대상자로부터 적법절차 동의 여부, 연구대상자 등의 안전, 연구대상자 등의 개인정보 보호 대책 등을 심의, 조사·감독하는 역할을 수행하도록 되어 있다.¹⁰⁾ 한편, DRB는 데이터 가명처리의 적정성 검토 등을 수행하도록 규정되어 있다.¹¹⁾ 만약 헬스케어 AI 연구 관리가 데이터에만 국한한다면, DRB로 IRB 심의를 갈음하는 것도 타당한 접근일 수 있으며, 생명윤리법 시행규칙이 명시하고 있는 것처럼 “개인식별 정보를 수집·기록하지 않는 지”만을 따지면 될 것이다.¹²⁾ 그러므로, 가명처리가 충분히 이루어졌다면(따라서, DRB에서 가명처리 적절성을 판단하였다면) IRB 심의면제를 진행하는 것은 적절한 처분이자 중복 규제에의 해소일 수 있다.

그러나, 해당 논의가 진행되던 2020년대 초, 즉 헬스케어 빅데이터 연구에 대한 심의 방안을 검토하던 당시 상황과 현재 상황은 사뭇 다르다. 무엇보다, 지금 논의 대상은 빅데이터 연구가 아닌 AI 연구다. 빅데이터의 패턴 인식 알고리즘의 적정성만을 판단하던 당시의 논의와 달리, 지금 논쟁의 핵심에 놓이는 것은 생성형 AI다. 생성형 AI는 그 훈련 과정에서 데이터를 적절히 가명처리했는지만을 연구 절차의 적절성 및 안전성을 따지는 지표로 삼을 수 없다. 다음 사례를 검토해 보자.¹³⁾

7) 개인정보보호위원회. 위의 글(주 6). 124면.

8) 심의면제 확인서와 관련하여 『보건의료데이터 활용 가이드라인』은 “가명처리 확인 등 데이터심의위원회 승인을 받은 과제”로 별도 제한을 두고 있지 않으나, 『바이오헬스 데이터 이용 연구에 대한 기관위원회 운영 가이드라인』은 “② 연구에 이용하려는 연구대상자 등에 대한 기존의 자료나 문서가 연구대상자들을 식별할 수 없는 경우”에 한정하여 확인서 발급 가능 조항을 귀속하고 있다. 그러나, 본 논문이 이후에서 주장하는 것처럼 기존 생명윤리법의 기관위원회 심의 면제 조항을 인공지능 연구에 그대로 적용하는 것에는 문제의 소지가 있다. 국가생명윤리정책원. 『바이오헬스 데이터 이용 연구에 대한 기관위원회 운영 가이드라인 - 개인정보의 2차 사용을 중심으로 -』, 2024. 37면.

9) 이 조항은 2023년 7월 29일 보건복지부 ‘IRB 제도개선 간담회’와 2023년 8월 8일 개인정보보호위원회의 보건의료 데이터 활용 활성화화를 위한 간담회에서 산업계의 IRB 및 DRB 중복 심의에 의한 제도 개선 요구를 수용하여 결정된 것이나, 각 당사자가 IRB와 DRB를 적절히 이해한 상황에서 논의를 진행한 것인지 재검토가 필요하다.

10) 생명윤리법 제10조 제3항.

11) 개인정보보호위원회. 위의 글(주 6). 29면.

12) 생명윤리법 시행규칙 제13조 제1항.

13) 사례는 다음 논문에서 문제로 지적되었던 부분을 강조하여 작성하였다. N. Norori, Q. Hu, F. M. Aellen, et al. “Ad-

A 의료원은 산하 병원에 축적된 의무기록 및 처방기록을 통합, 대규모 환자 데이터세트를 기반으로 조기 심질환 예측 및 평가 멀티모달 AI 모델을 개발하였다. 개발에 사용한 데이터세트는 충분한 가명처리 절차를 거쳤으며, DRB 심의를 통과하여 별도의 조처가 없는 한 해당 자료에서 연구대상자를 식별할 수 없음을 확인받았다. 데이터 처리의 안전성을 기반으로 의료원은 별도의 심의 없이 모델 개발을 진행하였고 일차 모델이 완성되었다.

그러나, 일차 모델의 내부 평가 과정에서, 의료원 연구진 한 명이 해당 모델이 제시하는 예측 및 평가 설명이 장애 및 희소 질환을 가진 환자에게 더 불리한 결과를 만들어 내었음을 확인하였다. 추가적인 내부 평가 결과, 각 병원에서 수집된 데이터세트가 종종 장애 및 희소 질환 환자를 누락하였으며 이들 요소에 과도한 비중을 두어 예측값을 제시하고 평가문을 기술하고 있음이 확인되었다. 그러나, 의료원은 지금까지 들어간 비용과 노력을 매몰할 수 없다며 해당 보고를 누락하고 모델 출시를 진행하기로 결정하였다.

가명처리가 적절히 이루어졌다고 해도 AI 모델의 안전성을 확보했다고 볼 수 없는 것은 헬스케어 AI가 기존의 약품이나 의료기기와 달리 인간 유사 행위자(human-like agent)¹⁴⁾로 작동하여 기존 보건의료 당사자의 행동을 대체할 수 있기 때문이다. AI, 더구나 생성형 AI가 논의의 핵심에 놓이는 이상 그저 의료인의 ‘손 안의 도구(Zuhandenheit)’였던 의학적 대상을 다루는 방식으로만 연구윤리적 접근을 논의해선 안 된다. AI 도구는 이제 우리 손에서 벗어나고 있으며, 그저 투명한 방식으로 우리가 자유자재로 다룰 수 있는 대상으로 남아 있는 것을 거부하는 것처럼 보인다.

본 논문은 이런 상황 인식 속에서, 헬스케어 생성형 AI 연구 관리 및 감독을 위한 거버넌스를 위한 생명윤리법의 개정을 제안한다. 첫째, 논문은 생명윤리법과 헬스케어 AI에 관한 주제 범위 고찰을 통해, 지금까지의 연구윤리 논의가 데이터 가명처리에만 집중되어 왔음을 확인한다. 둘째, 논문은 데이터 가명처리만으로 헬스케어 AI, 특히 생성형 AI 연구 거버넌스에 한계가

dressing bias in big data and AI for health care: A call for open science”. 『Patterns』, 2(10), 2021. p.100347. 최근 한 학술대회 발표문은 헬스케어 AI의 편향에 대한 체계적 문헌고찰을 진행, 편향의 발생 원인으로 레이블 편향, 대상 구체화 편향, 의사결정 인지 편향, 임상 활용에서 영역 변경, 건강 불평등, 데이터 수집을 지적하였다. A. Moldovan, A. Vescan, C. Grosan. “Healthcare bias in AI: A Systematic Literature Review”. 『20th International Conference on Evaluation of Novel Approaches to Software Engineering(ENASE)』. In press.

- 14) 루치아노 플로리디(Luciano Floridi)는 생물종의 존재 차원인 생태계에 상응하는 인포스피어(Infosphere) 차원에서 정보 존재자 또는 정보 시스템의 행위성 및 피동성에 주목한 바 있다. 이 관점에서 인간과 AI가 수준의 차이는 있으나 모두 정보 존재자라는 차원에서 동등하다. 그러나, 본 논문의 목적상 존재론적 논변을 끌고 들어오는 것은 증명과 논증의 과도한 부담을 지므로, 여기에선 단순화하여 AI가 인간과 유사한 수준의 행위성을 지닌다는 점을 윤리적 접근에서 고려할 필요가 있음을만 다룬다. 루치아노 플로리디. 석기용 역. 『정보철학 입문』. 서울: 필로소픽. 2022.(Floridi, Luciano. 『Information: A Very Short Introduction』. Oxford: Oxford University Press. 2010.)

있음을 사례 기반으로 증명하고 IRB와 DRB의 역할을 다시 정리할 필요가 있음을 제시하며, 연구 과제에 따라 DRB 심의만, IRB 심의만, 또는 DRB와 IRB 심의를 모두 받아야 하는 주제가 있음을 분석한다. 셋째, 논문은 2026년 시행 예정인 「인공지능발전과신뢰기반조성등에 관한기본법(이하 인공지능기본법)」이 그 조항으로 둔 민간자율인공지능윤리위원회의 역할을 보건의료 영역에선 IRB가 맡아야 함을 제안하고, 이를 위해 생명윤리법의 개정이 필요함을 제시한다.

II. 연구방법

본 논문은 인공지능과 관련하여 생명윤리법 개정 요청에 관한 문헌에 있어 주제범위 고찰을 시행하는 것을 그 출발점으로 했다. 주제범위 고찰은 체계적 문헌고찰과 달리 기존 문헌의 포괄적인 검토, 핵심 개념 확인, 연구의 공백을 확인하고자 하며, Arksey와 O'Malley에 의해 구체화되었다.¹⁵⁾ 주제범위 고찰은 해당 주제를 다양한 연구 설계나 구조를 통해 복합적이거나 이질적인 문헌이 다루고 있을 때(즉, 체계적 문헌 고찰을 적용하기 어려울 때) 사용할 수 있으며 이 때 연구의 질을 평가하지 않는다.

연구 질문은 “생명윤리법 개정에 관한 논의에서, 인공지능과 관련된 주제로 어떤 논의가 이루어졌는가?”로 설정하였으며, 따라서 “생명윤리법”과 “인공지능”을 검색 키워드로 설정하여 국내외 데이터베이스에서 문헌을 검색하는 것을 전략으로 설정하였다.¹⁶⁾

국내 학술 검색 데이터베이스 RISS, KCI, Synapse와 국외 학술 검색 데이터베이스 Scopus, WoS(Web of Science), NIH Pubmed, 공통 학술 검색 데이터베이스 Google Scholar에서 “생명윤리법”(또는 “Bioethics and Safety Act”)과 “인공지능”(또는 “Artificial Intelligence”)을 키워드로 하여 논문을 검색하였다. 기간이나 문헌 유형은 한정하지 않았으며, “생명윤리법” 키워드가 이미 범주를 한정하기에 별도로 “헬스케어 인공지능”과 같은 검색어는

15) H. Arksey, L. O'Malley. “Scoping studies: Towards a methodological framework”. 『International Journal of Social Research Methodology』. 8(1). 2005. pp.19-32.

16) 논문이 많이 검색되지 않는 경우 다양한 검색어를 포함하여 검색된 논문의 수를 늘리는 것을 고려할 수 있으나, 본 연구 질문을 인공지능 관련 생명윤리법 개정으로 잡았기에 “생명윤리법”과 “인공지능” 외의 다른 키워드는 더 늘리지 않았다. 즉, 유전 연구, 정밀의료, 빅데이터 등과 생명윤리법의 관련성은 본 연구에서 검토하는 내용과 상이하여 키워드로 사용하지 않았다. 또한, 검색 시 너무 많은 문헌이 잡혔다면 키워드를 추가하여 범위를 좁혔겠지만, 이미 검색된 문헌의 수가 적어 범위를 좁힐 필요성을 확인하지 못했다.

사용하지 않았다.¹⁷⁾ 각 검색 엔진에서 검색식 및 검색 논문의 수를 정리하였다(표 1).

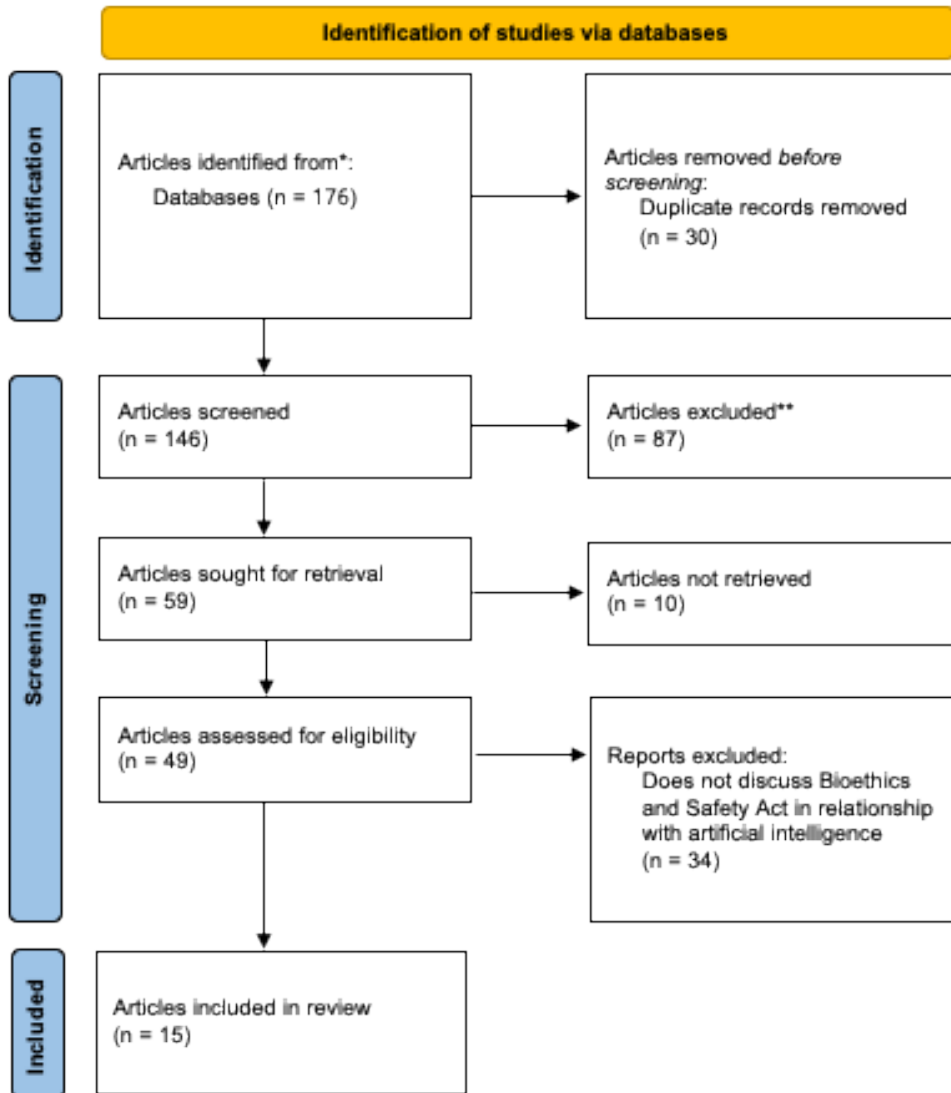
〈표 1〉 “생명윤리법”, “인공지능” 주제범위 검색 결과

데이터베이스	검색 키워드	검색 문헌수
RISS	전체: 생명윤리법 <AND> 전체: 인공지능	32
KCI	KEYALL: (생명윤리법) AND KEYALL: (인공지능)	31
Synapse	(“Bioethics and Safety Act”[ALL]) AND (“Artificial Intelligence”[ALL])	0
Scopus	(ALL (“Bioethics and Safety Act”) AND ALL (“Artificial Intelligence”))	2
WoS	(ALL= (“Bioethics and Safety Act”)) AND ALL= (“Artificial Intelligence”)	0
NIH Pubmed	(“Bioethics and Safety Act”) AND (“Artificial Intelligence”)	0
Google Scholar	“Bioethics and Safety Act” AND “Artificial Intelligence”	111

검색 결과 총 176개의 논문을 확인하였다. 먼저 중복 문헌을 제거하고, 제목과 초록을 기준으로 내용상 무관한 문헌을 배제한 다음, 원문에 접근할 수 없거나 원문이 한글이나 영문이 아닌 문헌을 제외하였다. 여기에서 이미 논문 수가 개별적으로 확인할 수 있을 만큼 줄어들었기에 추가적인 배제 기준을 설정하지 않았다. 다음, 원문을 확인하여 생명윤리법과 헬스케어 인공지능 관련 내용을 다루지 않은 문헌을 제외하였다(그림 1). 최종 15개의 논문을 분석에 포함시켰다.

17) 하단에 기술한 것과 같이, 검색된 문헌의 수가 많지 않아 기간이나 문헌 유형 등을 한정하지 않아도 검색된 문헌을 모두 검토할 수 있었다. “생명윤리법 개정(revision of Bioethics and Safety Act)”이 아닌 “생명윤리법”으로 검색어를 설정한 것 또한, “개정”을 포함하지 않아도 이미 충분히 문헌 수가 적었기 때문이다.

〈그림 1〉 PRISMA 플로우 다이어그램: “생명윤리법”-“인공지능”



III. 연구 결과

분석 대상인 15개의 논문은 윤리 및 법에 관한 연구 논문이 주를 이루었다<표 2>. 논문의 작성 시기가 2018년부터 2024년까지 다양하고(1개 논문을 제외하면 모두 2020년 이후 작성), 건강 데이터 관련 법제의 개정이 2020년에 이루어졌으므로 한 논문이 제언한 것이 이미 다른 논문에선 실행된 것으로 보고 그 다음 단계의 필요성이나 한계를 검토하고 있음을 염두에 두고

논문의 생명윤리법 관련 내용을 살필 필요가 있다. 따라서, 해석의 편의를 위해 발표 시점을 기준으로 논문을 정렬하였다.

〈표 2〉 분석 논문 및 인공지능 관련 생명윤리법 개정 내용

저자	유형	초록 요약	인공지능 관련 생명윤리법 개정 내용	주제
S.-Y. Shin ¹⁸⁾	논평	(초록 없음) 머신러닝을 포함한 AI는 대규모 헬스케어 데이터 분석으로 정밀의료에서 핵심 역할을 수행하고 있음, 그러나 여전히 환자 프라이버시가 중대한 문제로 남음	AI 연구에 활용되는 생명의료 데이터에 적합한 지침을 제공할 수 있는 생명윤리법 개정의 강조: 개인정보, 재식별화 등 용어 명확화, 보건의료 데이터 식별자에 대한 표준 확립, k-익명화 방법 외 생명의료 데이터의 복잡성 다룰 수 있는 프라이버시 보호 기작 도입, 다학제 협력 등	용어 명료화 익명화
C. W. Park et al. ¹⁹⁾	총설	AI 기술의 급속한 발전이 의료 데이터의 광범위한 분석을 가능하게 하나, 안전, 신뢰성, 낮은 수용성의 염려가 잔존함	최근 법제 개선이 건강 데이터 활용을 촉진하는 긍정적인 방향으로 나아가고 있으나, 여전히 한계 있음, 프라이버시와 윤리적 문제에 대한 체계적인 해결책을 제시하기 위한 생명윤리법 지속 개선 필요: 데이터 익명화 강화, AI 대상 윤리 지침 구성, 국제표준에 규제 프레임워크 일치	익명화 AI 윤리지침 규제 프레임워크
S. Oh et al. ²⁰⁾	보고서	한국엔 아직 여러 낡은 규제가 상존함, 최근 보건의료 혁신 사례를 검토하여 신뢰 형성과 당사자 이익균형의 방안을 검토	현재 생명윤리법은 제한적인 동의 요구와 생명윤리위원회의 중앙 통제로 AI 및 데이터 기반 의학 연구의 가능성을 심각히 제한, 유연하고 적응 가능한 규제적 접근을 가능케 하는 생명윤리법 개정의 요청, 포괄 동의 및 데이터 후속 연구에 대한 옵트아웃 제도 활용으로 규제 부담 줄일 것 요구	동의 유연화 적응적 규제
김영국 ²¹⁾	원저	데이터3법 개정을 중심으로 헬스케어서비스 활성화 방안을 논의	현재 생명윤리법이 AI와 관련하여 명확한 한계가 존재하며, 유전 검사 및 치료에 대한 명확한 규제 결여, 생명윤리법 개정 필요: 명확한 유전 치료 규제 프레임워크, 데이터 익명화 정책 검토, 의료·비의료 경계의 명확화	규제 프레임워크 익명화

18) S.-Y. Shin. "Issues and Solutions of Healthcare Data De-identification: The Case of South Korea". 『Healthcare Informatics Research』. 24(1). 2018. pp.3-13.

19) C. W. Park, S. W. Seo, N. Kang, et al. "Artificial intelligence in health care: Current applications and issues". 『Journal of Korean Medical Science』. 35(42). 2020. p.e379.

20) S. Oh, C. Choi, J.-H. Lee. "Regulatory reform for fostering an innovation ecosystem". 『KDI School of Pub Policy & Management Paper』. Forthcoming.

21) 김영국. "헬스케어 활성화를 위한 법정책 소고". 『법률실무연구』. 제8권 제4호. 2020. 175-220면.

저자	유형	초록 요약	인공지능 관련 생명윤리법 개정 내용	주제
H. Ko, S. Park ²²⁾	논평	(초록 없음) EU GDPR과 유사하게, 2020년 데이터법 개정안은 가명처리 데이터의 동의 없는 활용을 허용함	기존 생명윤리법에 대한 보건복지부 해석이 가명처리를 가능하게 함, 2020년 개인정보 보호법을 포함한 유관법 개정을 통해 규제 충돌을 해결하였음, 그러나 데이터 보호 거버넌스에 대한 모호성이 잔존함	익명화/가명처리 데이터 보호 거버넌스
S.-Y. Shin ²³⁾	사설	(초록 없음) 개인정보보호법 개정안은 프라이버시와 데이터 활용의 균형을 잡기 위해 가명처리 가이드라인을 제시, 그러나 규제 불명확성으로 인한 논쟁 지속 중	최근 개인정보보호법이 개정되어 데이터 활용과 프라이버시 사이 균형을 추구하고 있으나 생명윤리법이 여전히 규제적 제한으로 남아 있음, 데이터 익명화와 가명처리 개념 명확화와 IRB 절차의 합리화로 AI 기술 발전의 속도를 따라갈 필요 있음, 윤리와 기술 조화를 위한 기술적 방법(프라이버시 보존 데이터 마이닝, 합성 데이터 활용, 동적 동의 등) 활용 제안	익명화/가명처리 규제 충돌
Y. Kang et al. ²⁴⁾	원저	5개 장기 3,000개 종양 슬라이드 이미지의 주석 작업을 위한 웹 기반 플랫폼 구축, AI 모델 훈련 기반과 데이터 공유 이니셔티브로 작동할 것	현재 인체유래물에 대한 환자 동의의 요구는 현실적 어려움을 초래, 현행 법제 하 프라이버시 규제는 효율적인 AI 연구를 가로막음, 2020년 개인정보보호법 및 관련 규제 개정이 부분적으로 복잡성을 해소하였으나 데이터 공유 규제의 국제적 기준 일치를 위한 관심 필요, 생명윤리법의 프라이버시 보호 프레임워크에 유연성이 요청됨	용어 명료화 익명화
유성희 등 ²⁵⁾	원저	바이오 산업의 엄격한 사전 규제가 기술적 발전을 저해함, 3차 바이오규제 TF에서 발굴된 개선과제와 관련하여 세 가지 개선안을 제시	생명의료법 하 바이오 기술(AI 포함) 연구의 거버넌스와 관련, 세 가지 핵심 문제 제안: IRB(다기관 연구에서 공동IRB 실행의 한계), 임상시험 실시기관(1차 의료기관 참여 문제), AI 기술 규제(신의료기술평가의 과도한 평가 기준, 즉 새로운 정보 제시와 고비용 개입 대체를 제안해야 하는 것이 스타트업의 기술 발전과 상업화를 심각하게 제한)이에 대응하는 별도 규제 절차를 마련할 것 요청	IRB 임상시험 기관 신의료기술 평가

22) H. Ko, S. Park. "How to de-identify personal data in South Korea: An evolutionary tale". 『International Data Privacy』. 10(4). 2020. pp.385-394.

23) S.-Y. Shin. "Privacy protection and data utilization". 『Healthcare Informatics Research』. 27(1). 2021. pp.1-2.

24) Y. Kang, Y. J. Kim, S. Park, et al. "Development and operation of a digital platform for sharing pathology image data". 『BMC Medical Informatics and Decision Making』. 21(1). 2021. p.114.

25) 유성희, 설민, 김현수. "바이오경제 실현을 위한 국내 바이오분야 규제현황과 개선 과제". 『일감법학』. 제49호. 2021. 295-322면.

저자	유형	초록 요약	인공지능 관련 생명윤리법 개정 내용	주제
C. B. Shook, M. Kim ²⁶⁾	원저	병원부터 스타트업에 이르기까지 병리학에서 AI 개발을 검토하여, 복잡하고 불투명한 AI 모델의 임상 판단 활용을 검토하고 혁신에서 발생하는 리스트의 관리 필요를 검토	디지털 병리학 데이터에 대한 현재의 익명화 요구 및 데이터 공유 제한을 개선하기 위해 협력 모델을 검토할 필요 있음, 또한 병원을 비영리 기구로 두는 현재 거버넌스가 혁신 가능성을 제한, 따라서 더 확장적이며 개선된 규제 프레임워크가 필요	익명화 규제 프레임워크
권대철 ²⁷⁾	원저	PACS와 관련한 의료법, 개인정보보호법, 생명윤리법의 충돌을 검토하고, 프라이버시를 보호하는 한편 의료 이미지 활용의 적정성을 확립하기 위한 법적 조화 안 제시	AI 발전을 위해 의료 영상 정보를 활용하기 위한 모호성 해소 필요: 생명윤리법의 익명화 정의 구체화, IRB 승인에서 익명화·가명처리 기준의 정리, 개인정보보호법과의 충돌 해소, 관련법의 충돌 해소를 위한 의료정보보호법 신설 제안	용어 명료화 익명화/ 가명처리 규제 충돌
최용전 ²⁸⁾	원저	인체유래물 연구의 동의 및 면제와 관련한 법적 틀 확인, 공여자 보호와 연구 결과를 통한 공익적 활용 규정 미비를 해결하기 위한 법리 검토	AI 발전이 인체유래물 활용과 수집의 확대를 요청하며, 이는 생명윤리법 하 명확한 거버넌스와 공여자 보호 절차를 요구, 생명윤리법 개정을 통해 충분한 공여자 이해와 동의 절차 보장, 공여자 권리의 명확한 보호, AI 기반 연구 맥락에서 연구 촉진과 자율성 보호와 남용 예방을 위한 구체적 감시 균형	공여자 권리 연구 감시
H. S. Oh ²⁹⁾	원저	고령화 사회의 문제 해결을 위한 디지털 헬스케어 확보를 위해 가명처리를 통한 데이터 활용을 촉구	헬스케어 데이터 활용에 명확한 환자 동의와 익명화를 요구하는 현재 기준은 정보의 실용적 활용을 심각하게 제한, 생명윤리법과 의료법이 만든 위축 상태를 해결하기 위한 제언: 익명/가명 처리에 대한 국제 표준 도입, 가명처리에 대한 명확한 정의와 가이드라인 생성, 비구조화 헬스케어 데이터에 대한 가명처리 절차 확립 등	익명화/ 가명처리 용어 명료화

26) C. B. Shook, M. Kim. "Ethics in multi-modal artificial intelligence and hybrid pathology systems development". 『Asia Pacific Journal of Health Law & Ethics』. 15(3). 2022. pp.49-65.

27) 권대철. "PACS에서 의료영상정보의 익명처리와 관련된 법의 연구". 『한국방사선학회 논문지』. 제16권 제5호. 2022. 627-637면.

28) 최용전. "인체유래물연구에 있어서의 자기결정권과 동의면제 연구". 『국가법연구』. 제18집 제2호. 2022. 111-134면.

29) H. S. Oh. "Examining the utilization of digital healthcare and pseudonymous information". 『Digital Health Research』. 1(1). 2023. p.e2.

저자	유형	초록 요약	인공지능 관련 생명윤리법 개정 내용	주제
H. Park et al. ³⁰⁾	원저	2024년 보건의료데이터 활용 가이드라인의 유전 데이터 가명처리 관련 내용을 분석	AI 기반 보건의료 연구에서 유전 데이터의 확장과 관련하여 생명윤리법의 한계 개정 요청: 유전 데이터 가명처리와 관련한 개념 명료화, 임상·연구 데이터 구분, 가명처리 기반 데이터 접근성 확대, 리스크 평가 프레임워크 강화 등	가명처리 용어 명료화 규제 프레임 워크
R. Kim et al. ³¹⁾	총설	의료 빅데이터 규제는 프라이버시 보호와 임상적 유용성 사이 균형을 추구해야 함, 개인정보보호법 확장을 통해 프라이버시 침해 없이 산업 발전을 추구할 것을 촉구	현재 생명윤리법은 의료 데이터 활용과 AI 혁신을 촉진하기보다는 장애물로 작용하고 있음, 생명윤리법 및 관련법의 개정 촉구: 정의 명확화, 표준 동의 절차, 의료 데이터 2차활용에 관한 세부 지침, EU GDPR과 핀란드 2차활용법 참조한 거버넌스 구조 도입 요청	용어 명료화 동의 절차 2차 활용
김옥주 ³²⁾	논평	2024년 헬싱키 선언 개정안을 검토, 대상자를 연구 참여자로 수정 전체 연구팀의 책임 제시, 취약성 개념 변화 등을 분석	헬싱키 선언 개정안의 AI 관련 연구 내용에 기초할 때 현재 거버넌스의 한계 존재: 인체유래물에만 초점을 맞춘 전통적인 연구윤리 체계의 강화·확장으로 디지털 데이터의 윤리적 거버넌스 포함 필요, AI 포함한 새로운 기술에 대한 연구 참여자 동의 절차 마련(데이터 2차활용 포함), 국내 윤리 거버넌스와 국제 표준의 연결 제안, 규제 개정으로 윤리위원회가 대규모 데이터 및 AI 연구 관리	윤리 거버넌스 동의 절차

다수의 문헌은 주로 건강 데이터 익명화와 가명처리 문제를 다루고 있다. 특히 익명화와 가명처리 개념의 명확화와 가명처리 절차 확립이 생명윤리법-AI 주제 논문에서 핵심 주제로 다뤄져 왔음을 정리에서 확인하였다. 2016년 부처합동 『개인정보 비식별 조치 가이드라인』³³⁾은 금융부터 보건의료 영역에 이르는 다양한 데이터의 비식별 조치를 네 단계의 절차로 취급하도록 제시하였으나, 생명의료 데이터의 다양성과 구조화·비구조화 데이터의 혼합은 해당 가이드

30) H. Park, J. Park, H. G. Woo, et al. "Safe utilization and sharing of genomic data: Amendment to the Health and Medical Data Utilization Guidelines of South Korea". 『Cancer Research and Treatment』. 56(4). 2024. pp.1027-1039.

31) R. Kim, J. Shinn, H.-S. Kim. "Using medical big data for clinical research and legal considerations for the protection of personal information: The double-edged sword". 『Cardiovascular Prevention and Pharmacotherapy』. 6(1). 2024. pp.8-16.

32) 김옥주. "2024년 개정 헬싱키선언의 의의와 함의". 『대한의사협회지』. 제67권 제11호. 2024. 668-673면.

33) 행정자치부. 『개인정보 비식별 조치 가이드라인』. 2016.06.30.

라인을 적용함에 있어 한계를 제시하였으며, 따라서 보건의료 영역의 각 데이터 양식에 대한 구체적인 식별화 처리 기준 및 절차가 요구된 바 있다.³⁴⁾

2020년 개인정보보호법 개정에 기초, 가명처리 기준과 절차가 2020년 『보건의료데이터 활용 가이드라인』 발표³⁵⁾로 정리되었음에도, 이후 논문에서도 계속 익명화·가명처리 개정에 대한 요청이 이어짐을 확인하였다. 예컨대, 아직 의료 데이터에 대한 명확한 정의가 부재하며,³⁶⁾ 가명처리 데이터의 결합 등을 통한 재구축이나 재식별의 위험을 완전히 차단하는 데 한계가 있고 익명화와 가명처리 사이 명확한 선을 긋지 못했으며,³⁷⁾ 심지어 각 헬스케어 데이터 유형에 있어 가명처리와 익명화 중 어느 쪽에 더 적절한지에 대한 연구가 미흡한 상태다.³⁸⁾ 특히, 가명처리를 통한 건강 데이터 활용 방안을 제시하였음에도 생명윤리법 제2조가 익명화 정의를, 제18조 제2항이 익명화 요구를 유지하고 있음이 여전히 문제로 남아 있는 상황을 확인할 수 있으며, 2023~2024년에 발표된 논문도 여전히 가명처리의 기준 및 절차를 개선할 것을 요구하고 있다.

물론, 분석에 포함된 논문은 익명화 문제만을 다룬 것은 아니며, 특히 초기 논문에서 ‘규제 프레임워크 개선’으로 다루어지고 있던 내용은 점차 구체화되어 리스크 평가 요구, 데이터 2차활용 지침, 위원회가 데이터 및 AI 연구의 감시를 수행할 수 있도록 IRB의 절차 및 기반 법제 개선 등의 제안이 등장하고 있음을 확인하였다. 예컨대, 식품의약품안전처 신의료기술평가제도 아래 “① 새로운 정보제공, ② 기존 고가의료행위를 대체”하는 것이 “국내 AI 관련 기업에게는 상당한 부담”으로 작용한다는 분석에 의거, 이에 일반 의료기술과 동일한 평가 기준을 적용할 것을 요청하거나,³⁹⁾ 인체유래물연구(특히, 빅데이터 기반 연구)의 연구결과물 추적성(traceability), “정보공개요구 및 열람청구권, 연구성과의 분배”를 절차화할 것을 요구하며,⁴⁰⁾ 유전 정보 전달 시 보안망을 사용하고 외부적으로 공개할 경우엔 가명처리된 VCF⁴¹⁾를 활용하며 개인 식별이 가능한 유전 정보를 담고 있는 FASTQ⁴²⁾ 데이터의 전달은 제한되어야 하는

34) S.-Y. Shin. 위의 글(주 18). p.3.

35) 개인정보위원회. 『보건의료데이터 활용 가이드라인』. 2020.09.

36) S.-Y. Shin. 위의 글(주 23). p.1.

37) H. Ko, S. Park. 위의 글(주 22). pp.392-393.

38) H. S. Oh. 위의 글(주 29). p.6.

39) 유성희, 설민, 김현수. 위의 글(주 25). 309-317면.

40) 최용진. 위의 글(주 28). 129면.

41) Variant Call Format, 염기서열 변이 정보를 저장하는 파일 형식.

42) 차세대 염기서열 분석 결과를 서열과 정확성 정보로 저장하는 파일 형식.

등⁴³⁾ 구체적인 건강 데이터 활용 지침을 제안하는 논의가 제시되었다. 또한, IRB와 관련하여, 중복 심의를 줄이고 운영 효율성 제고를 위해 공동IRB 운영 또는 상호인증제를 도입할 것,⁴⁴⁾ 대규모 데이터 수집에 있어 2016년 세계의사회 타이페이 선언⁴⁵⁾을 따라 IRB가 모니터링할 것⁴⁶⁾ 등이 제안되었다.

그러나, 전체 논문에서 공통적으로 발견되는 익명화에 대한 문제의식과 가명처리로의 해결, 그리고 해당 개념 및 절차에 대한 지속 개선 요구와 달리, 헬스케어 AI 연구와 관련하여 생명윤리법 및 관련 규정에 대한 일관된 논의나 개선 요청을 찾아보기는 어렵다. 상기에서 확인한 것처럼, 다양한 논의가 제기되었고 최근 논의가 구체화되고 있으나 제도 개선에 있어 구체적인 방향성을 확인하긴 어려운 상태다.

IV. 토의

1. 생명윤리법과 가명처리

먼저 헬스케어 AI 연구 과정의 익명화 및 가명처리에 있어 생명윤리법의 문제를 검토한다.

첫째, 이미 생명윤리법의 익명화 개념은 엄격한 익명화와 가명처리 모두를 포함하고 있다. 생명윤리법 제2조의 익명화 정의는 “개인식별정보를 영구적으로 삭제”(익명화)와 “개인식별정보의 전부 또는 일부를 해당 기관의 고유식별기호로 대체”(가명처리) 양자를 모두 가리키고 있으므로, 엄밀히 말해 생명윤리법은 익명화와 가명처리를 필요시에 따라 활용할 것을 요구하고 있는 셈이다. 무엇보다, 양자 중 택일하여 한쪽만이 필요하다고 보기는 어렵다. 2025년 의료 데이터의 가명처리에 대한 체계적 문헌고찰이 다시 정리하고 있는 것처럼,⁴⁷⁾ 가명처리는 완전한 익명화가 불가능하거나 필요하지 않은 상황에서 프라이버시 적용 설계(privacy-by-design) 원칙 아래 환자 데이터의 수집과 통합을 목적으로 한다. 다시, 이것은 모든 상황에서 가명처리로

43) H. Park, J. Park, H. G. Woo, et al. 위의 글(주 30). p.1037.

44) 유성희, 설민, 김현수. 위의 글(주 25). 312-313면.

45) World Medical Association. 『WMA Declaration of Taipei on Ethical Considerations regarding Health Databases and Biobanks』. Adopted by the 67th WMA General Assembly, Taipei, Taiwan. 2016.10.

46) 김옥주. 위의 글(주 32). 671면.

47) H. A. Attieh, A. Müller, F. N. Wirth, et al. “Pseudonymization tools for medical research: a systematic review”. 『BMC Medical Informatics and Decision Making』. 12(25). 2025. p.128.

충분하다는 것을 의미하는 것이 아니라, 완전한 익명화가 불가능하거나 필요하지 않은 상황의 구별을 그 선제 조건으로 한다. 가도티 등⁴⁸⁾이 정리한 것처럼, 전통적인 익명화 기술은 소수의 속성을 지닌 데이터를 대상으로 적용되고 빅데이터, 즉 고차원의 데이터에 철저히 익명화를 적용하는 것은 불가능하지는 않더라도 어렵다. 그렇다면, 모든 건강 데이터는 모두 고차원의 데이터라서 익명화를 요구하는 것이 무조건 무리한 일인지 먼저 살펴야 한다. 문제는, 아직 이 부분에 대한 제도적 차원의 검토가 진행되지 않았다는 데에 있다.

둘째, 생명윤리법 자체는 익명화 또는 가명처리의 절차에 대한 검토를 제시하지 않는다. 생명윤리법 제18조 제2항은 제3자 제공시 “익명화하여야 한다”라고만 명시할 뿐, 익명화나 가명처리 적정성을 검토하는 조항을 두고 있지 않다. 이미 해당 내용이 개인정보보호법 시행령⁴⁹⁾과 『보건의료데이터 활용 가이드라인』, 그리고 DRB의 가명처리 적정성 심의를 통해 관리되고 있으므로, 모든 인간대상연구가 가명처리를 필요로 하는 데이터를 활용하지 않는 상황에서 추가 조항을 신설할 필요는 없을 것으로 보인다. 오히려, 전술한 것처럼 생명윤리법의 IRB 감독 체계와 보건의료데이터 활용 가이드라인의 DRB 감독 체계는 그 목적이나 검토 대상에서 차이를 보이므로, 여기에서 세 가지 상황을 나누어서 살필 필요가 있다. 즉, 보건의료 및 헬스케어 AI 연구에서 (1) IRB의 검토만이 필요한 연구, (2) DRB의 검토만으로 충분한 연구, (3) IRB와 DRB의 검토가 모두 필요한 연구.

IRB의 검토로 진행할 수 있는 경우는 인간대상연구이되 연구 데이터의 가명처리를 하지 않는 경우 또는 가명처리가 필요하지 않은 경우일 것이다. 즉, 데이터 활용에 대하여 연구대상자의 적법한 동의를 모두 받은 경우는 DRB의 검토가 필요하지 않다. 이는 헬스케어 빅데이터 및 AI 연구에도 마찬가지로 적용된다.

다음, DRB의 검토만으로 충분한 경우다. 물론, 인간대상연구가 아닌 경우가 일차적으로 해당되겠으나 현 검토의 맥락에서 보건의료 분야의 연구이나 인간대상연구가 아닌 경우를 검토하는 것은 연구의 범위를 벗어난다. 건강 데이터의 2차활용으로 가명처리하여 자료를 제공받을 때 DRB의 검토를 받았으나, 이 연구가 생명윤리법 시행규칙 제13조 제1항의 각호에 해당하는 경우, 즉 심의 면제 대상인 경우엔 DRB 검토로 충분하다. 그러나, 헬스케어 AI, 특히 생성형 AI 연구가 시행규칙이 제시하고 있는 “침습적 행위를 하지 않는 연구”, “단순 접촉 측정장비 또는 관찰장비만을 사용하는 연구”, “판매 등이 허용되는 식품 또는 식품첨가물”의 맛·질 평가

48) A. Gadotti, L. Rocher, F. Houssiau, et al. “Anonymization: The imperfect science of using data while preserving privacy”. 『Science Advances』. 10(29). 2024. p.eadn7053.

49) 개인정보보호법 시행령 제29조의5.

연구, “안전기준에 맞는 화장품”의 사용감·만족도 연구, 연구대상자 특정 없고 민감정보를 수집·기록하지 않는 연구, 연구대상자의 기존 자료·문서를 이용한 연구에 포함되는 경우는 많지 않을 것으로 보인다. 인간적·합리적으로 사고·행동하는 AI는⁵⁰⁾ 외과적인 침습 행위를 하는 것이 아니라고 해도 정의상 인간 행동에 개입할 수 있으며, 따라서 이를 침습적 행위라고 규정하거나 그렇지 않더라도 단순 접촉 측정장비 또는 관찰장비 이상의 개입을 제공하는 장비라고 규정하는 것도 가능하기 때문이다. 그러나, 이런 수준의 연구 또한 분명히 있을 것이므로, 어떤 연구가 DRB만으로 충분한지(또는, DRB 심의 후 IRB 면제 승인을 받을 수 있는지)에 대한 명확한 범주 설정이 필요하다.

마지막으로, IRB와 DRB 심의를 모두 받아야 하는 경우다. 사실, 연구 대부분이 이렇게 진행되고 있을 것이며, IRB가 규제 준수를 요구하는 기관인 한 아직 헬스케어 AI 연구에 대한 규제가 불확실한 상황에서⁵¹⁾ IRB가 헬스케어 AI 연구 승인에 어떤 역할을 하고 있는지에 대한 실증 연구가 필요하다. 그러나, 데이터세트 구축에 대한 감독과 모델 설계 및 개발에 대한 감독에 있어 DRB와 IRB 각각의 역할을 구분한 선행 연구⁵²⁾에 기초하여, 본 논문은 가명처리 적절성의 판단 이후에도 IRB가 헬스케어 AI 연구에 대해 감독해야 하는 사항들이 명확히 있음을 제시하고자 한다.

2. 가명처리 이후 헬스케어 AI 연구의 윤리적 이슈

앞서 데이터 편향의 사례를 검토하였으나, 헬스케어 영역에서 AI는 데이터 프라이버시 외에도 여러 윤리적 이슈를 제기하고 있다. 최근 헬스케어 AI의 윤리적 이슈를 재분석한 논문⁵³⁾은 데이터 공유, 접근성 및 프라이버시는 모든 영역의 AI에서 발생하는 문제이므로 헬스케어 AI에서 별도로 검토할 필요가 없으며, 오히려 인식론적 문제(확률적 결과에 의한 오진의 우려, 집단적 부정 평가, 사회적 결정요인에 대한 무지), 규범적 문제(건강과 질병의 개념 변형, 환자-의료인 경험 재형성, 사회적 불평등 강화, 의료 전문직 신뢰 손상), 추적가능성의 문제(환자 비난,

50) S. J. Russell, P. Norvig. 『Artificial Intelligence: A Modern Approach』. 4th ed. Upper Saddle River, NJ: Pearson. 2020.

51) D. Char. “Whether designated as research or not, who resolves ethical considerations emerging with health-care AI?”. 『American Journal of Bioethics』. 24(10). 2024. pp.93-95.

52) J. Kim, S. Y. Kim, E.-A. Kim, et al. 위의 글(주 4). pp.400-403.

53) J. Morley, L. Floridi. “The ethics of AI in health care: An updated mapping review”. 2024.10.12. <<https://ssrn.com/abstract=4987317>>. (마지막 접속일 2025.03.24.).

집단 귀책, 관리·책임의 부재)가 부각됨을 제시하고 있다. 이들 문제는 헬스케어 AI 연구단에서 최소화할 수 있도록 고려되어야 하나, 현재 이를 명시적으로 감시·감독하는 체계는 부재한 상태다. 헬스케어 AI 연구가 IRB 심의를 받는다고 해도, 위와 같이 전통적인 임상 연구 범위 밖에서 발생하는 문제는 쉽게 IRB의 고려 대상으로 들어오지 않는다. 한편, DRB는 가명처리(따라서, 프라이버시)를 중심으로 연구를 심의하며, 따라서 위 문제를 검토하는 일을 DRB가 맡을 수 있는지 애초에 애매한 상태다.

이하에서는 헬스케어 AI 연구 단계에서 고려해야 할 다양한 윤리적·거버넌스적 이슈를 대주제별로 정리하였다.⁵⁴⁾ 또한, 연구 관리 관점에서 핵심 쟁점만을 추려, 헬스케어 AI 연구 과정에서 고려할 점과 이에 상응하는 IRB 감독 필요성을 정리하였다(표 3). 이는 체계적 문헌 고찰의 결과가 아니며, 현재 엄청난 속도로 발전하고 있는 헬스케어 AI 관련 윤리적 이슈 전반을 포괄적 정리를 목표로 한 것을 아님을 밝혀둔다.⁵⁵⁾

먼저, 정확성·안전성의 문제다. 연구윤리 체계가 해악금지를 기반으로 오랫동안 검토해 온 것과 같이, 헬스케어 AI 연구에서 가장 먼저 고려해야 할 문제 또한 알고리즘, 모델, 디바이스 등이 정확성과 안전성을 확보하느냐에 있다. 이는 환자 치료와 건강 증진에 직접적인 영향을 미치므로, 실제 사용 단계가 아닌 연구·개발 단계에서 먼저 검토되어야 한다.

첫째, 알려진 것처럼 생성형 AI, 특히 LLM(Large Language Model)은 환각(hallucination) 현상을 나타내며 해당 모델이 헬스케어 영역에 적용되었을 때 사용자에게 해를 끼치는 결과를 나타낼 수 있다.⁵⁶⁾ LLM은 결과물을 출력함에 있어 사실과는 다른 내용을 ‘꾸며서’ 그럴듯한 언설로 만들어 내는 환각 현상을 보이며, 이로 인하여 사용자가 잘못된 정보에 따라 의료적 의사결정을 내려 건강상 해악을 입을 수 있다.

이를 방지하기 위해서는 먼저 결과물의 불확실성을 정량화하고 이에 따라 실질적인 점검 체계를 마련할 필요가 있다. LLM의 환각은 출력 결과의 불확실성에서 나타나므로,⁵⁷⁾ 불확실성

54) 원 논문은 플로리다의 정보존재론에 기초하여 헬스케어 AI의 윤리적 이슈를 분석하는 체계를 만들어 접근하고 있어, 이에 맞추어 설명하려면 정보존재론에 대한 어느 정도의 설명이 부가되어야 한다. 이는 논문의 범위를 벗어날뿐더러 독자의 이해 부담을 증가시키므로, 여기에선 단순화하여 해석, 연구 관리와 관련된 내용만을 추려 수정한 판본을 제시한다.

55) 이 경우, 각 주제에 대한 별도의 연구가 필요할 것이므로 헬스케어 AI 연구 관련 생명윤리법 개정 필요성을 제시하는 것이 목적인 본 연구의 범위를 벗어난다.

56) L. de Angelis, F. Baglivo, G. Arzilli, et al. “ChatGPT and the rise of large language models: the new AI-driven infodemic threat in public health”. 『Frontiers in Public Health』. 25(11). 2023. p.1166120.

57) S. Farquhar, J. Kossen, L. Kuhn, et al. “Detecting hallucinations in large language models using semantic entropy”. 『Nature』. 630(8017). 2024. pp.625-630.

을 정상화하는 방식⁵⁸⁾ 등을 활용할 수 있다. 또한, 핵심적인 의사결정 상황에는 반드시 인간(의료진 또는 전문가)가 개입(human-in-the-loop)하도록 설계⁵⁹⁾하여 AI의 단독 판단으로 발생하는 위험을 줄여야 한다. 더불어, 생성형 AI를 포함하여 AI 모델은 배포된 이후에도 지속적으로 수정·업데이트될 가능성이 크므로,⁶⁰⁾ 전통적인 연구 단계의 감시(audit)에 더하여 사후 배포단계에서 AI의 안전성과 정확성을 지속 모니터링하고 관리하는 체계가 고려되어야 한다.

둘째, 확률적인 특성에 의하여 LLM은 동일한 입력(예, 동일 증상 또는 동일 병리·방사선 영상)에 대해 다른 진단 결과를 생성할 수 있다.⁶¹⁾ 이러한 특성은 의학적 결정의 일관성을 훼손할 위험성을 지닐뿐더러, 환자에게 불필요한 혼란이나 불안을 초래할 가능성이 있다.

해결책으로 헬스케어 AI 모델이 제시하는 결과에 확률 혹은 신뢰 점수(confidence score)를 함께 표기하도록 하는 방법이 있다.⁶²⁾ 이때 의료진이나 환자는 결과를 해석할 때 모델이 제시한 조언, 진단, 평가 등이 불확실성을 포함함을 가늠하고 그에 주의를 기울일 수 있다. 또한, 연구 단계에서 불확실성에 대한 지침을 미리 명문화할 것을 요구⁶³⁾하여 사용자(의료진 및 환자)가 AI 결과의 신뢰도를 파악할 수 있도록 제도적으로 보완할 필요가 있다.

정리하면, 정확성 및 안전성의 맥락에서 IRB가 다음 사항을 감독할 것을 제안한다.

- 헬스케어 AI 연구 설계 단계에서 환각 및 불확실성을 어떻게 예방하고 완화(mitigation)할 것인지를 고려하였는지 평가
- 환자 및 사용자에게 발생 가능한 해악을 사전에 식별하고 최소화하려는 방안이 연구 프로토콜에 충분히 반영하였는지 평가
- 보건의료 중요 의사결정 영역에서 AI 사용시 해당 체계를 구분하는 지침을 마련하고, 해당 체계로 분류된 경우 인간 개입이 강제되도록 검토

58) E. Fadeeva, A. Rubashevskii, A. Shelmanov, et al. "Fact-checking the output of large language models via token-level uncertainty quantification". 『arXiv』 [preprint]. 2024. arXiv:2403.04696.

59) V. Agarwal, Y. Jin, M. Chandra, et al. "MedHalu: Hallucinations in responses to healthcare queries by large language models". 『arXiv』 [preprint]. 2024. arXiv:2409.19492.

60) 따라서, 학습 헬스learning health system 안에서 healthcare AI development and post-deployment는 지속적인 연구 활동을 수행하는 것으로 가정할 수 있다. 다음 논문을 참고하라. R. Sparrow, J. Hatherley, J. Oakley, et al. "Should the use of adaptive machine learning systems in medicine be classified as research?". 『American Journal of Bioethics』. 24(10). 2024. pp.58-69.

61) S. Harrer. "Attention is not all you need: The complicated case of ethically using large language models in healthcare and medicine". 『EBioMedicine』. 90. 2023. p.104512.

62) Y. Virk, P. Devanbu, T. Ahmed. "Enhancing trust in LLM-generated code summaries with calibrated confidence scores". 『arXiv』 [preprint]. 2024. arXiv:2404.19318.

63) J. Haltaufderheide, R. Ranisch. "The ethics of ChatGPT in medicine and healthcare: A systematic review on Large Language Models (LLMs)". 『NPJ Digital Medicine』. 7. 2024. p.183.

- 연구 및 배포 후에도 지속 모니터링 계획이 구체적으로 마련되어 있는지, 감시 절차가 연구 설계에 포함되어 있는지 점검

다음, 편향(bias) 및 편파성(favoritism)이다. AI 시스템이 학습에 활용하는 데이터세트가 편향을 지니고 있으면, 소수 집단이나 취약 계층을 대상으로 한 부정확한 판단을 내릴 수 있다.

첫째, AI 개발 과정에서 사용하는 데이터 및 데이터세트에 특정 집단이 제대로 반영되지 않거나 지나치게 적은 비중을 차지한다면, 모델이 그 집단의 건강 위험을 적절히 평가하지 못할 수 있다.⁶⁴⁾ 이는 현존하는 불평등을 악화시킬 수 있다는 점에서⁶⁵⁾ 이를 연구 단계에서 검토하지 않는다면 그 자체로 시스템적 부정의로 작동할 수 있다. 이를 해결하기 위해서 데이터세트 편향을 측정할 수 있는 공정성 척도(fairness metric)를 적용하고 그 결과를 평가·보고하는 체계를 갖출 필요가 있다.⁶⁶⁾ 또한, 알고리즘 자체가 불합리한 편향(algorithmic bias)을 일으키지 않는지에 대한 걱정 처리 여부를 검토해야 한다.⁶⁷⁾ 더불어, 대표성 있는 데이터세트를 확보하고 이를 정기적으로 감시하는 거버넌스 차원의 노력을 기울여야 한다.⁶⁸⁾

둘째, 현재 기술을 선도하고 있는 AI 모델, 특히 LLM은 민간 기업에서 개발한 사영 AI 시스템이다. 헬스케어 AI 자체가 이익을 추구하기 위한 목적으로 개발되지 않더라도, 이런 AI가 사영 AI 시스템에 의존하여 그 위의 층위에서 작동할 경우 환자 안전과 건강 이득보다 기업의 수익이 우선시될 위험이 있다.⁶⁹⁾

이에 대한 해결책으로, 기술의 부작용을 감시, 평가, 예방하기 위한 기술 경계 프레임워크(techno-vigilance framework)를 구축할 필요가 있다.⁷⁰⁾ 이에 더하여, 헬스케어 AI 시스템 내에

64) M. D. Abràmoff, M. E. Tarver, N. Noyo-Berrios, et al. "Considerations for addressing bias in artificial intelligence for health equity". 『NPJ Digital Medicine』. 6. 2023. p.170.

65) R. Baumgartner, P. Arora, C. Bath, et al. "Fair and equitable AI in biomedical research and healthcare: Social science perspectives". 『Artificial Intelligence in Medicine』. 144. 2023. p.102658.

66) F. Dehghani, N. Malik, J. Lin, et al. "Fairness in healthcare: Assessing data bias and algorithmic fairness". 『2024 20th International Symposium on Medical Information Processing and Analysis(SIPAIM)』. 2024. pp.1-6.

67) R. J. Chen, J. J. Wang, D. F. K. Williamson, et al. "Algorithmic fairness in artificial intelligence for medicine and healthcare". 『Nature Biomedical Engineering』. 7(6). 2023. pp.719-742.

68) A. S. Bergman, L. A. Hendricks, M. Rauh, et al. "Representation in AI evaluations". 『2023 ACM Conference on Fairness, Accountability, and Transparency(FACCT'23)』. 2023.

69) A. Toma, S. Senkaiahliyan, P. R. Lawler, et al. "Generative AI could revolutionize health care—but not if control is ceded to big tech". 『Nature』. 624(7990). 2023. pp.36-38.

70) F. Pennestrì, F. Cabitza, N. Picerno, et al. "Sharing reliable information worldwide: Healthcare strategies based on artificial intelligence need external validation. Position paper". 『BMC Medical Informatics and Decision Making』. 25. 2025. p.56.

사영 AI를 활용할 때 그 영향과 결과를 검토하기 위한, 외부 검증(external validation)을 의무적으로 거치도록 할 수 있으며 이는 공공 차원의 감시와 참여를 요청한다.⁷¹⁾

이들 논의를 토대로, 편향 및 편파성의 맥락에서 IRB가 다음 사항을 검토할 필요성을 요청하고자 한다.

- 연구 과정에서 데이터 편향을 평가하기 위한 공정성 척도 적용과 알고리즘 편향 검토 여부의 확인⁷²⁾
- 연구 계획에서 데이터세트 대표성(dataset representativeness)과 관련한 고려 있는지 확인(예, 거버넌스 프로토콜 준수, 확립된 대표 데이터세트 활용 또는 모델 검증 과정에서 대표 데이터세트 사용 등)
- 외부 독립 검증을 포함한 모델 모니터링 계획 포함하였는지 점검

이어서, 자율성 침해다. 데이터 프라이버시 침해와는 별도로, AI가 환자 및 의료진을 부당하게 설득하거나 속일 수 있다. 또한, AI가 과도하게 의사결정에 개입하여 헬스케어 맥락에서 개인의 자율성을 저해할 수 있다.

첫째, AI가 설득, 속임, 심리적 해악 등으로 환자 결정에 영향을 미칠 수 있다.⁷³⁾ AI가 ‘의도적으로’ 부정확한 정보를 제공하거나 알고리즘이 의도치 않게 특정 방향으로 환자를 유도하여 환자의 합리적 의사결정을 왜곡할 가능성이 있다.⁷⁴⁾ 예를 들어, 사용자가 AI가 만든 정보와 의료진이 작성한 정보를 구별하지 못하도록 설계된 시스템에서, AI는 최종 결과값을 향상하기 위하여 의도적으로 출력 결과를 의료진이 작성한 것처럼 위장할 수 있다.⁷⁵⁾

이러한 위험을 줄이기 위하여, 사용자를 속일 가능성이 있는 AI 체계를 별도 분류하여 규제 및 결과에 대해 AI가 생성한 것임을 명확히 표시하도록 의무화할 필요가 있다.⁷⁶⁾ 또한, AI

71) 김준혁. “헬스케어 AI 윤리에서 환자·시민 참여 모형: 주제범위 고찰과 방법론적 검토에 기초하여”. 『한국의료윤리학회지』. 27(4). 2024. 177-196면.

72) 이 과정은 현재 연구계획 검토 단계에서 통계적 설계 적절성 검토와 유사한 절차로 이해할 수 있다.

73) L. Floridi. “Hypersuasion—On AI’s persuasive power and how to deal with it”. 『Philosophy & Technology』. 37. 2024. p.64.

74) P. S. Park, S. Goldsetin, A. O’Gara, et al. “AI deception: A survey of examples, risks, and potential solutions”. 『Patterns』. 5(5). 2024. p.100988.

75) 아직 헬스케어 영역의 사례는 없으나, 포커 게임에서 다른 플레이어를 속이는 블러핑을 익히거나 의도적 정보 교란으로 다른 사용자를 속여 최종 결과에서의 이득(즉, 게임 승리)을 취하는 AI의 전략은 이미 여러 사례에서 기술된 바 있다. P. S. Park, S. Goldsetin, A. O’Gara, et al. 위의 글(주 74). Table 1.

76) P. S. Park, S. Goldsetin, A. O’Gara, et al. 위의 글(주 74). Table 5.

가 제시한 의견의 논리적 근거(rationale)를 공개하도록 하는 투명 소통 기준(algorithmic/data/process transparency)을 적용, AI가 결론에 이르는 경로를 이해하고 검토할 수 있도록 할 필요가 있다.⁷⁷⁾ 환자에게는 이를 통해 충분한 정보가 제공되어야 하고, 설문이나 인터뷰를 통해 환자가 실제로 어떻게 결과를 이해하는지 지속하여 확인할 필요가 있다.

둘째, 의사결정에서 AI가 결정 역할을 지나치게 대체하면 의료진의 역할을 축소하여 환자-의료진 신뢰를 약화시킬 수 있다.⁷⁸⁾ 결과값만 제시하는 기존의 도구와 달리 AI는 해석을 통한 결정을 제시하고, 이를 그대로 수용하는 것을 체계가 더 우선하는 경우 의료진은 거수기로 격하할 가능성이 있으며 이는 환자 신뢰 손상과 의료전 전문성의 저해를 초래할 가능성이 있다.

이를 예방하기 위해서, 임상에서 AI가 의사결정자가 아닌 결정 지원 도구의 역할을 하도록 설계 단계에서 검토해야 한다. 예컨대, CDSS(Clinical Decision Support System) 설계를 위한 여러 제안(design principles for AI-supported CDSS,⁷⁹⁾ recommendations for AI-CDS⁸⁰⁾은 인간-AI 상호작용을 검토하여 AI에게 지원 도구로서의 역할을 명확히 설계 단계에서 배정할 것을 요구한다. 또한, 사용자(의료진) 설문을 통해 임상 업무흐름(workflow)을 검토하여 의료진과 환자 모두가 적절하고 만족스러운 의사결정을 내리도록 지원할 필요가 있다[nondisruptive workflow design,⁸¹⁾ Translational Evaluation of Healthcare AI(TEHAI)⁸²⁾].

상기 논의에 따라 자율성 침해와 관련하여, IRB가 다음과 같은 사항을 검토할 것을 제의한다.

- 사용자를 기만하거나 오도할 우려가 있는 AI(deceptive AI)에 대한 식별 지침을 확립할 것을 거버넌스와 학계에 요구하고, 연구 설계에서 해당 AI 체계에 해당되는 경우 결과 검토 계획을 요청

77) A. Singhal, N. Neveditsin, H. Tanveer, et al. "Toward fairness, accountability, transparency, and ethics in AI for social media and health care: Scoping review". 『JMIR Medical Informatics』. 3(12). 2024. p.e50048.

78) G. de Togni, S. Erikainen, S. Chan, et al. "What makes AI 'intelligent' and 'caring'? Exploring affect and relationality across three sites of intelligence and care". 『Social Science & Medicine』. 277. 2021. p.113874.

79) E. R. Burgess, I. Jankovic, M. Austin, et al. "Healthcare AI treatment decision support: Design principles to enhance clinician adoption and trust". 『Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems(CHI'23)』. 2023. pp.1-19.

80) S. Labkoff, B. Oladimeji, J. Kannry, et al. "Toward a responsible future: Recommendations for AI-enabled clinical decision support". 『Journal of the American Medical Informatics Association』. 31(11). 2024. pp.2730-2739.

81) M. Sharma, C. Savage, M. Nair et al. "Artificial intelligence applications in health care practice: Scoping review". 『Journal of Medical Internet Research』. 24(10). 2022. p.e40238.

82) S. Reddy, W. Rogers, V.-P. Makinen, et al. "Evaluation framework to guide implementation of AI systems into healthcare settings". 『BMJ Health & Care Informatics』. 28(1). 2021. p.e100444.

- 의사결정 지원 도구가 독립적 의사결정을 수행할 수 없도록 설계 단계에서 안전 조치를 포함할 것을 요구
- 임상 의사결정 및 개입에 직접 활용되는 AI의 경우, 임상 업무흐름 영향 평가 및 환자 중심성(patient-centeredness) 보장 계획을 확인

마지막으로, 설명 가능성(explainability) 및 추적 가능성(traceability)⁸³⁾의 문제다.

첫째, AI의 추적 가능성 결여가 환자-의료진 신뢰, 충분한 설명에 의한 동의, 효율적인 정보 공유를 저해한다.⁸⁴⁾ 이는 사후 문제 발생 시 책임소재 규명을 더 복잡하게 만드는 원인이 될 수 있다.

설명 가능한 AI 프레임워크는 이러한 문제를 해결하기 위한 대표적 접근이다.⁸⁵⁾ 구체적으로 AI가 의사결정 과정을 보다 명확히 보여주도록 설계 단계에서부터 고려할 필요가 있으며, 이를 통해 환자와 의료진이 AI가 도출한 결과를 이해하고 신뢰를 유지할 수 있도록 도와야 한다.

또한, 데이터와 프로세스의 출처나 변동 이력을 일관성 있게 기록하는 문서 표준(provenance documentation)을 마련하여, 연구와 임상에서 활용하는 AI 알고리즘이 어떠한 과정을 거쳐 개선·변경되었는지 추적할 수 있도록 할 수 있다.⁸⁶⁾ 특정 시점에서 AI가 내린 의사결정 경로를 정기적으로 감사할 필요도 있다. AAA(prospective risk Assessment, operation Audit, system Adherence)와 같은 거버넌스 원칙을 적용하는 것은 헬스케어 현장에서 AI가 불투명하게 작동하지 않도록 하는 규제 방식을 제안한다.⁸⁷⁾

둘째, 헬스케어 AI 연구 현장에서 모델의 정확도나 안전성을 객관적으로 평가할 수 있는 표준화된 벤치마크나 검증 방법론이 아직 충분히 확립되지 않았다.⁸⁸⁾ 이는 AI 연구 결과의 재현성, 신뢰성, 비교 가능성의 문제를 야기하고, 실제 환자 치료에 적용할 때 리스크를 초래한다.

83) AI가 어떻게 하여 결론을 도출했는지 알 수 없음을 가리키며, 타 문헌에선 AI의 블랙박스적 특성으로 언급하곤 한다. 설명가능한 AI(Explainable AI, XAI)를 요청하는 주된 이유가 된다.

84) U. Upadhyay, A. Gradisek, U. Iqbal, et al. "Call for the responsible artificial intelligence in the healthcare". 『BMJ Health & Care Informatics』. 30. 2023. p.e100920.

85) Q. Sun, A. Akman, B. W. Schuller. "Explainable artificial intelligence for medical applications: A review". 『arXiv』 [preprint]. 2024. arXiv:2412.01829.

86) A. Kale, T. Nguyen, F. C. Harris Jr., et al. "Provenance documentation to enable explainable and trustworthy AI: A literature review". 『Data Intelligence』. 5(1). 2023. pp.139-162.

87) G. Falco, B. Shneiderman, J. Badger, et al. "Governing AI safety through independent audits". 『Nature Machine Intelligence』. 3. 2021. pp.566-571.

88) K. Blagec, J. Kariger, W. Frühwirth, et al. "Benchmark datasets driving artificial intelligence development fail to capture the needs of medical professionals". 『Journal of Biomedical Informatics』. 137. 2023. p.104274.

연구 초기 단계부터 AI 성능을 평가하기 위한 벤치마크 및 검증 프로토콜을 마련할 필요가 있다. 이미 공개 플랫폼 구축,⁸⁹⁾ 영상 영역에서 활용할 수 있는 성능 지표(performance metrics) 기반 벤치마크 데이터세트 구현⁹⁰⁾ 등이 제안된 바 있다. 더 나아가, 연구뿐 아니라 AI 모델의 실제 임상 도입 후에도 주기적으로 안전 보고서 발간을 발간하여 모델 성능과 환자 안전 문제를 점검하고 학계와 공유할 것이 요청되었다.⁹¹⁾

설명 가능성 및 추적 가능성의 맥락에서 IRB가 다음 항목을 검토할 것을 제의한다.

- AI 기반 결정이 설명 가능한지 또는 설명 가능성을 고려하고 있는지에 대한 연구 계획 검토
- 문서 표준에 따라 데이터 수집, 훈련, 의사결정 과정을 발표하는 절차를 연구 계획에 포함할 것을 요청
- AI 의사결정 경로 정규 감사 프로토콜 마련 및 준수 모니터링
- 결과 평가를 위한 표준화 벤치마크 및 검증 방법론을 도입하였는지 확인하고 이를 준수할 것을 연구진에 요청
- 모델이 임상 현장에 적용된 이후에도 주기적 안전 보고서 발간 계획을 수립했는지 검토하고, 보고서 발간을 모니터링

〈표 3〉 프라이버시 외 헬스케어 AI 연구에서 고려해야 하는 이슈, 연구 고려사항, IRB 감독

도메인	헬스케어 AI 이슈	연구·개발 차원 고려사항	IRB 감독
정확성 및 안전성	LLM 환각과 사용자 해악	- 불확실성 정량화 - 핵심 의사결정 과정에 인간 개입 강제 - 사후 배포단계에서 결과 감시	- 환각 및 불확실성 예방·완화 조치 검토 - 환자·사용자 해악 사전 식별 및 최소화 방안 프로토콜 반영 여부 검토
	LLM의 확률적 특성으로 인한 다진단·다결과 가능성	- 결과에 확률, 신뢰 점수 출력하도록 강제 - 불확실성에 대한 지침 제시 요구	- 의사결정 분야에서 AI 사용 시 인간 개입 강제 - 지속 모니터링 계획 포함 확인

89) A. Karargyris, R. Umeton, M. J. Scheller, et al. “Federated benchmarking of medical artificial intelligence with MedPerf”. 『Nature Machine Intelligence』. 5. 2023. pp.799-810.

90) N. Sourlos, R. Vliegenthart, J. Santinha, et al. “Recommendations for the creation of benchmark datasets for reproducible artificial intelligence in radiology”. 『Insights into Imaging』. 15. 2024. p.248.

91) K. Lekadir, A. F. Frangi, A. R. Porras, et al. “FUTURE-AI: International consensus guideline for trustworthy and deployable artificial intelligence in healthcare”. 『BMJ』. 388. 2025. p.e081554.

도메인	헬스케어 AI 이슈	연구·개발 차원 고려사항	IRB 감독
편향 및 편파성	편향으로 인한 소수 집단 해악 및 불평등 악화	<ul style="list-style-type: none"> - 데이터세트에 대한 공정성 지수 적용 및 결과 제출 - 알고리즘 편향에 대한 적정 처리 여부 검토 - 데이터세트 대표성 감사 	<ul style="list-style-type: none"> - 공정성 지표, 알고리즘 공정성 적용 검토 - 데이터세트 대표성 준수 확인 - 외부 독립 검증 포함하였는지 확인
	사영 AI 시스템의 역할 강화와 수익 우선	<ul style="list-style-type: none"> - 헬스케어 AI 시스템 내에 사영 AI 시스템을 사용하는 경우, 외부 검증 요구 - 기술 경계 프레임워크 도입 	
자율성 침해	AI 설득, 속임, 조종으로 환자 결정 영향	<ul style="list-style-type: none"> - AI 결과를 AI가 생성하였음을 표시하도록 의무화 - AI 의견의 근거를 공개하도록 하는 투명 소통 기준 적용 - 환자 피드백 설문을 통한 이해도 측정 	<ul style="list-style-type: none"> - 기만 가능 AI의 식별 지침 요청 및 적용 확인 - 설계 단계에서 AI 독립 의사결정 내릴 수 없도록 강제 - 임상 업무 흐름 분석 및 환자 중심성 보장 계획 확인
	의사결정에서 의료진 역할 축소	<ul style="list-style-type: none"> - 임상에서 AI가 의사결정자가 아닌 결정 지원 도구의 역할을 하도록 설계 단계에서 검토 - 사용자(의료진) 설문을 통한 업무흐름 검토 	
설명 가능성 및 추적 가능성	AI 추적 가능성 결여	<ul style="list-style-type: none"> - 설명 가능 AI 프레임워크와 문서 표준을 통한 투명성 강제 - AI 의사결정 경로에 대한 정기 감사 	<ul style="list-style-type: none"> - 설명 가능성 고려에 대한 연구계획 검토 - 데이터 수집·훈련·의사결정 과정 발표에 대한 문서 표준 적용 - AI 의사결정 정규 감사 프로토콜 마련 확인 - 결과 평가 벤치마크, 검증 방법론 도입 여부, 준수 계획 확인 - 정기 안전 보고서 계획 확인
	AI 정확성 및 안전에 대한 측정 지표 결여	<ul style="list-style-type: none"> - AI 성능 평가에 대한 벤치마크, 검증 프로토콜 표준화 - 주기 안전 보고서 발간 요구 	

질병관리청이 헬스케어 AI 연구지침을 발표하였고 자율 검토 형식으로 AI 연구 과정을 검토하도록 요청하였으나,⁹²⁾ 상기 사항을 반영한 IRB 감독 지침은 아직 마련되지 않은 상태다. 이는 생명윤리법에 헬스케어 AI 연구를 고려한 IRB 운영 내용이 포함되지 않고 오로지 익명화와 관련한 규정만 위치하여 현재 어떠한 제도적 강제나 감시를 적용할 수 없는 상황을 반영한다.

92) 질병관리청. 『보건의료 분야 인공지능 연구자를 위한 연구윤리 지침』. 2023.

특히, 생명윤리법 제10조 제3항의 IRB 심의 내용 규정을 기준으로 할 때 연구대상자의 안전을 포괄적으로 해석하는 경우 위 내용들이 포함될 가능성을 검토해 볼 수 있으나, 이는 조항의 과도한 확장으로 여겨질 가능성이 크다.

따라서, 본 논문은 시행 예정인 인공지능기본법 제28조의 내용이 생명윤리법과 함께 헬스케어 AI 연구를 규제할 필요성과 해당 조항이 제시하는 민간자율인공지능윤리위원회의 역할을 IRB와 통합하는 방안을 개략적으로 검토할 것이다.

3. 인공지능기본법과 IRB의 역할

2024년 통과된 인공지능기본법은 AI의 건전한 발전과 신뢰 기반 조성에 필요한 사항을 규정하여 국민의 권익과 존엄성을 보호하고 국민의 삶의 질 향상 및 국가경쟁력 강화를 목적으로 한다.⁹³⁾ 특히, 인공지능기본법은 사람의 생명·안전·기본권에 중대한 영향을 미칠 수 있는 AI를 고영향 AI로 규정하였으며,⁹⁴⁾ 고영향 AI에는 보건의료 제공 및 이용체계의 구축·운영, 그리고 의료기기·디지털의료기기에 활용되는 AI 시스템이 포함된다. 즉, 해당 조항을 따르면 일부 헬스케어 AI 시스템(예, 보건의료 제공: 진단 AI, 의사결정 지원 AI, 의료 상담 AI 등, 이용체계 구축·운영: 환자 평가 및 배정 AI, 의료 자원 분배 AI, 의료 시스템 운영 지원 AI 등, 의료기기·디지털의료기기: AI 기반 진단·치료·예방 의료기기)이 고영향 AI에 포함된다.

동법 제30조에 따르면 사업자는 고영향 AI의 사전 검·인증을 받아야 한다. 제31조는 고영향 AI 또는 생성형 AI 기반 서비스 제공시 AI 기반 운용 사실을 이용자에게 사전 고지할 것 등을 통해 해당 AI 시스템의 결과물이 AI로 생성된 결과임을 고지 또는 표시할 것을 규정하고 있다. 제33조는 사업자가 해당 제품 또는 서비스가 고영향 AI인지를 과학기술정보통신부장관에게 확인 요청할 수 있도록 하고 있으며, 이에 대하여 동법 제33조 제2항은 전문위원회를 설치하여 고영향 AI 해당 여부를 자문받을 수 있도록 하였다. 또한, 제34조는 고영향 AI의 안전성·신뢰성 확보를 위하여 위험관리방안의 수립·운영, 기술적 가능 범위에서 AI 최종 결과, 기준, 학습용데이터의 개요 등에 대한 설명 방안 수립·시행, 이용자 보호 방안 수립·운영, 고영향 AI의 사람 관리·감독, 안전성·신뢰성 조치 내용에 관한 문서 작성 및 보관 등을 이행할 것을 규정하고 있다. 더불어 제35조는 고영향 AI가 사람의 기본권에 미치는 영향을 평가하는 영향평가를 시행하도록 요청하였다.

93) 인공지능기본법 제1조.

94) 인공지능기본법 제2조 제4항.

전술한 것처럼 현행 생명윤리법상 IRB는 인간대상연구의 윤리적·과학적 타당성 검토, 연구참여자의 적법한 동의 확보, 연구 진행 중 참여자의 안전과 권리 보호 감독 등을 주요 기능으로 수행한다.⁹⁵⁾ 헬스케어 AI 연구 또한 인간대상연구로 규정할 수 있으며 그에 따른 윤리적·과학적 타당성을 검토하고 참여자의 동의를 확보하며 안전과 권리의 보호를 감독할 필요성이 발생함에 따라, IRB 체제가 작동하는 것은 타당해 보인다. 이에 헬스케어 AI 연구의 경우 IRB는 연구의 위험성 평가, 개인정보 보호 방안, 윤리적 문제의 사전 예방 등 연구 전반에 걸친 감독 업무를 수행할 것이 요청된다. 또한, 결과 예측이 어렵고 개인정보 및 윤리적 문제 발생 가능성이 높은 특성을 지니는 생성형 AI의 경우, 헬스케어 영역에서 연구·활용됨에 있어 연구 설계부터 실행, 평가 단계까지 IRB가 적극 개입하여 연구의 투명성 및 신뢰성을 확보할 필요가 있다. 이때, 앞 절에서 살핀 헬스케어 AI의 윤리적 고려사항 다수가 이미 인공지능기본법의 고영향 AI 안전성·신뢰성 확보를 위한 조처에 포함되어 있음을 유념할 필요가 있다. 결정 과정에서 인간 개입과 감독, 투명성과 설명 가능성, 적절한 문서 표준의 확립, 해악 식별-리스크 관리, 공정성 지표 및 데이터 대표성-기본 권리 영향평가는 헬스케어 AI 연구에서 고려해야 하는 이슈와 인공지능기본법의 중첩 지점을 잘 보여주는 사안들이다. 인공지능기본법이 고영향 AI 검·인증의 시행 주체를 “단체 등이 인공지능의 안전성·신뢰성 확보를 위하여 자율적으로 추진”하는 것으로 규정⁹⁶⁾하였다. 그렇다면, 이들 사항을 종합하여 헬스케어 AI 연구에 대한 생명윤리법과 인공지능기본법의 교점을 마련할 수 있을 것이다. 이 검토는 헬스케어와 AI라는 중복 영역에 영향을 미치는 두 법이 과다 규율을 하는 것이 아닌 공동 운영의 방식을 제안할 수 있다.

첫째, 인공지능기본법의 민간자율인공지능윤리위원회를 헬스케어 영역에선 IRB로 해석하는 결정이 필요하다. 윤리원칙 준수, 안전 및 인권침해 조사, 절차 및 결과 감독 등의 역할을 수행하는 두 위원회가 별도로 운영되는 것은 그 자체로 과다 규율로 받아들여질 수 있으며, 오히려 관련 감시 및 감독의 빠른 도입이 요구되는 헬스케어 영역에선 이미 그 운영이 잘 확립된 IRB로 민간자율인공지능윤리위원회를 겸하는 것이 적절할 것이다.

둘째, 인공지능기본법의 모든 헬스케어 AI 시스템이 고영향 AI에 포함되지 않을 것이므로, 어떤 시스템이 고영향 AI고 어떤 것은 아닌지에 대한 지침 및 확인이 필요하다. 이미 제2조의 고영향 AI 규정이 다수의 헬스케어 AI 시스템을 포함하고 있으나, 그렇지 않은 연구(즉, 사람의 생명, 신체의 안전 및 기본권에 중대한 영향을 미치지 않거나 위험을 초래할 우려가 없으며 보

95) 생명윤리법 제10조 제3항.

96) 인공지능기본법 제30조 제1항.

건의료 제공 시스템이나 의료기기가 아닌 것)까지 고영향 AI로 규정하고 검·인증을 시행할 필요는 없을 것이다.⁹⁷⁾

셋째, 인공지능기본법이 고영향 AI에서 검토해야 하는 사항과 헬스케어 AI에서 고려해야 하는 사항이 중복되므로, 이를 별도의 기관이 검토하는 것보다 같은 기관이 검토, 관리하는 것이 적절하다. 이미 생명의료법의 IRB 체계가 연구의 윤리적·과학적 타당성, 적법 동의, 안전성 등을 검토하기 위해 운영되고 있으므로, 인공지능기본법이 요청하는 자체 검·인증 수행 기관을 헬스케어 영역에선 IRB로 운용할 수 있을 것이다. 단, 이때 전술한 것처럼 기존 IRB가 수행하지 않았던 역할들이 추가로 요청되므로, 해당 내용을 IRB가 심의, 보완, 감독할 수 있도록 하는 법령 개정 및 지침 추가가 요청된다. 예컨대, 이미 헬스케어 연구는 인간대상연구로서 생명윤리법의 규제 범위 안에 포함되어 있는 것으로서 인정되고 있으므로, 기관위원회의 업무에 대한 규정 보완을 통해 헬스케어 AI 연구 방법, 설계, 프로토콜, 사후 안전 등을 검토할 것을 규정하고 그에 대한 지침을 마련하는 방안을 검토해 볼 수 있다.⁹⁸⁾

넷째, 인공지능기본법이 요구하는 사항보다 헬스케어 AI에서는 추가로 고려해야 할 사항이 많다는 점을 염두에 둔 지침 마련이 필요하다. 인간대상연구에 대한 생명윤리의 규제, 감시는 최소한 나치 인체실험부터 이어진 오랜 역사를 통해 정립된 것이며, 최소한 생명과학 분야의 연구자는 이를 준수하는 것이 당연하다. 그러나, AI에 대한 규제는 이제 막 논의되기 시작한 것으로서 아직 모든 안전성 및 신뢰성에 대한 조치가 완비되었다고 보기 어렵다. 앞선 절에서 검토한 헬스케어 AI 연구의 검토사항들, 정확성 및 안전성, 편향 및 편파성, 자율성 침해, 설명가능성 및 추적가능성은 생명윤리법 및 생명윤리가 추구하는 “인간의 존엄과 가치를 침해하거나 인체에 위해를 끼치는 것을 방지”하는 목적에 부합하며⁹⁹⁾ 먼 미래의 예측이 아니라 이미 현재 기술적 수준에서 고려가 필요한 사항들로서 관리·감독을 요청한다. 따라서, 헬스케어 AI에 대한 IRB를 포함한 연구 거버넌스 마련을 위한 생명윤리법 개정은 인공지능기본법의 내용에 더하여 추가적인 내용을 연구 검토 단계에서 요청하고 모니터링 과정에서 이행할 수 있도록 이루어져야 할 필요가 있다.¹⁰⁰⁾

97) 이 경우를 앞에서 검토한 IRB 심의 면제 대상으로 처리할 수 있을 것이다.

98) 생명윤리법 제10조 제3항 제1호에 고영향 AI로 인정받은 헬스케어 AI 연구 및 개발을 추가하거나, 마목 “그 밖에 기관에서의 생명윤리 및 안전에 관한 사항”에 대한 해석으로 헬스케어 AI 연구 및 개발이 포함되는 것으로 인정하고 그에 대한 지침을 마련하는 것을 예로 들 수 있다.

99) 생명윤리법 제1조.

100) 이를 위하여 생명윤리법에 헬스케어 AI에 관한 새로운 장을 포함하여 선행 연구들이 요청한 개념 정의의 명확성을 추구하는 방안을 검토해 볼 수 있을 것이다.

V. 결론

본 논문은 생명윤리법 기반 헬스케어 AI 연구 규제에 대한 기존 논의가 익명화 및 가명처리에 주로 집중되어 있음을 살피고, 가명처리 외에도 헬스케어 AI 연구에서 관리, 감시되어야 할 여러 사항이 있음을 검토하였다. 이런 감시 역할은 기존의 IRB가 수행할 수 있으며, 신설 인공지능법과의 관계 속에서 IRB가 그 역할을 수행할 수 있도록 생명윤리법의 보완 및 추가 규정이 필요하다고 본 논문은 주장하였다.¹⁰¹⁾

생명윤리법 제정 20년이 흐른 지금, 2005년의 원문에 비하여 생명윤리법에는 많은 내용이 추가되었다. 인간대상연구부터 익명화에 이르는 개념 규정, 연구 참여자 및 기증자의 권리 및 프라이버시 보호 조항, 국가생명윤리심의위원회, IRB, 공용 IRB의 규정과 강화, 단성생식과 배아줄기세포 등 생명공학 기술에 대한 규제 추가, 인체유래물 관리의 철저 등 생명윤리법은 그간 의생명과학의 발전에 대한 대응을 위해 지속적인 개정을 거쳐 왔다. 2025년을 지나는 현재, 보건의료를 포함한 전 세계가 가장 크게 영향을 받고 있는 것은 AI의 대두와 산업 전반에 걸친 영향이며, 헬스케어 AI가 미치는 영향 또한 보건의료 전반을 변혁하고 있다고 일컬어질 만큼 심대하다. 이런 상황에서, 의생명과학 연구 규제 및 안전을 추구해 온 생명의료법이 그에 대응하는 것은 당연해 보인다. 더불어 이미 인공지능기본법이 통과된 상황에서, 두 법이 조응하여 헬스케어 AI 연구를 적절 규제하는 지침을 마련하는 것이 시급하다고 하겠다.

단, 본 논문은 법학적 관점에서 개별 조항의 적절성이나 타당성, 충돌 여부를 검토한 것이 아닌, 의료윤리적 관점에서 법 조항 개정 등의 필요성을 제한한 것이라는 한계를 지닌다. 구체적인 조항 신설, 개정, 규정 마련 등의 논의가 법학적 관점에서 후속 연구로서 철저히 이루어져야 할 필요가 있다.

101) 이미 개별 IRB가 헬스케어 AI 연구에 대한 관리·감독을 수행하고 있으나 그 구체적인 감독 사항이 정리되어 있지 않다는 점, 토의에서 검토한 것처럼 DRB 심의 시 IRB 심의면제에 대한 규정이 이미 운영되고 있으나 그 내용이나 이행이 모호하다는 점이 헬스케어 AI 연구에서 IRB의 역할을 명확히 할 필요성을 방증한다.

[참고문헌]

- 차현재, 김준혁. “건강데이터 기증 및 공유의 윤리적 접근: 인체유래물 기증 절차를 참조하여”. 『생명, 윤리와 정책』. 제6권 제2호. 2022.
- World Health Organization. 『Ethics and Governance of Artificial Intelligence for Health』. Geneva: World Health Organization. 2021.
- J. Kim, S. Y. Kim, E.-A. Kim, et al. “Developing a Framework for Self-regulatory Governance in Healthcare AI Research: Insights from South Korea.” 『Asian Bioethics Review』. 16(3). 2024.
- K. W. Miller, M. Taddeo (Ed). 『The Ethics of Information Technologies』. Oxford, UK: Routledge, 2016.
- J. H. Moor. “What is Computer Ethics?”. 『Metaphilosophy』. 16(4). 1985.
- 개인정보보호위원회. 『보건의료데이터 활용 가이드라인(개정안)』. 2024.12.
- 국가생명윤리정책원. 『바이오헬스 데이터 이용 연구에 대한 기관위원회 운영 가이드라인 - 개인정보의 2차 사용을 중심으로-』. 2024.
- N. Norori, Q. Hu, F. M. Aellen, et al. “Addressing bias in big data and AI for health care: A call for open science”. 『Patterns』. 2(10). 2021.
- A. Moldovan, A. Vescan, C. Grosan. “Healthcare bias in AI: A Systematic Literature Review”. 『20th International Conference on Evaluation of Novel Approaches to Software Engineering (ENASE)』. In press.
- H. Arksey, L. O'Malley. “Scoping studies: Towards a methodological framework”. 『International Journal of Social Research Methodology』. 8(1). 2005.
- 루치아노 플로리디. 석기용 역. 『정보철학 입문』. 서울: 필로소픽. 2022. (Floridi, Luciano. 『Information: A Very Short Introduction』. Oxford: Oxford University Press. 2010.)
- S.-Y. Shin. “Issues and Solutions of Healthcare Data De-identification: The Case of South Korea”. 『Healthcare Informatics Research』. 24(1). 2018.
- C. W. Park, S. W. Seo, N. Kang, et al. “Artificial intelligence in health care: Current applications and issues”. 『Journal of Korean Medical Science』. 35(42). 2020.
- S. Oh, C. Choi, J.-H. Lee. “Regulatory reform for fostering an innovation ecosystem”. 『KDI School of Pub Policy & Management Paper』. Forthcoming.
- 김영국. “헬스케어 활성화를 위한 법정책 소고”. 『법률실무연구』. 제8권 제4호. 2020.
- S.-Y. Shin. “Privacy protection and data utilization”. 『Healthcare Informatics Research』. 27(1).
- Y. Kang, Y. J. Kim, S. Park, et al. “Development and operation of a digital platform for sharing pathology image data”. 『BMC Medical Informatics and Decision Making』. 21(1). 2021.
- H. Ko, S. Park. “How to de-identify personal data in South Korea: An evolutionary tale”. 『International Data Privacy』. 10(4). 2020.

- 권대철. “PACS에서 의료영상정보의 익명처리와 관련된 법의 연구”. 『한국방사선학회 논문지』. 제16권 제5호. 2022.
- 최용전. “인체유래물연구에 있어서의 자기결정권과 동의면제 연구”. 『국가법연구』. 제18집 제2호. 2022.
- H. S. Oh. “Examining the utilization of digital healthcare and pseudonymous information”. 『Digital Health Research』. 1(1). 2023.
- H. Park, J. Park, H. G. Woo, et al. “Safe utilization and sharing of genomic data: Amendment to the Health and Medical Data Utilization Guidelines of South Korea”. 『Cancer Research and Treatment』. 56(4). 2024.
- R. Kim, J. Shinn, H.-S. Kim. “Using medical big data for clinical research and legal considerations for the protection of personal information: The double-edged sword”. 『Cardiovascular Prevention and Pharmacotherapy』. 6(1). 2024.
- 김옥주. “2024년 개정 헬싱키선언의 의의와 함의”. 『대한의사협회지』. 제67권 제11호. 2024.
- World Medical Association. 『WMA Declaration of Taipei on Ethical Considerations regarding Health Databases and Biobanks』. Adopted by the 67th WMA General Assembly, Taipei, Taiwan. 2016.10.
- H. A. Attieh, A. Müller, F. N. Wirth, et al. “Pseudonymization tools for medical research: a systematic review”. 『BMC Medical Informatics and Decision Making』. 12(25). 2025.
- A. Gadotti, L. Rocher, F. Houssiau, et al. “Anonymization: The imperfect science of using data while preserving privacy”. 『Science Advances』. 10(29). 2024.
- S. J. Russell, P. Norvig. 『Artificial Intelligence: A Modern Approach』. 4th ed. Upper Saddle River, NJ: Pearson. 2020.
- D. Char. “Whether designated as research or not, who resolves ethical considerations emerging with healthcare AI?”. 『American Journal of Bioethics』. 24(10). 2024.
- J. Morley, L. Floridi. “The ethics of AI in health care: An updated mapping review”. 2024.10.12. <<https://ssrn.com/abstract=4987317>>.
- L. de Angelis, F. Baglivo, G. Arzilli, et al. “ChatGPT and the rise of large language models: the new AI-driven infodemic threat in public health”. 『Frontiers in Public Health』. 25(11). 2023.
- S. Farquhar, J. Kossen, L. Kuhn, et al. “Detecting hallucinations in large language models using semantic entropy”. 『Nature』. 630(8017). 2024.
- E. Fadeeva, A. Rubashevskii, A. Shelmanov, et al. “Fact-checking the output of large language models via token-level uncertainty quantification”. 『arXiv』 [preprint]. 2024. arXiv:2403.04696.
- V. Agarwal, Y. Jin, M. Chandra, et al. “MedHalu: Hallucinations in responses to healthcare queries by large language models”. 『arXiv』 [preprint]. 2024. arXiv:2409.19492.
- R. Sparrow, J. Hatherley, J. Oakley, et al. “Should the use of adaptive machine learning systems in medicine be classified as research?”. 『American Journal of Bioethics』. 24(10). 2024.

- S. Harrer. "Attention is not all you need: The complicated case of ethically using large language models in healthcare and medicine". 『EBioMedicine』. 90. 2023.
- Y. Virk, P. Devanbu, T. Ahmed. "Enhancing trust in LLM-generated code summaries with calibrated confidence scores". 『arXiv』 [preprint]. 2024. arXiv:2404:19318.
- J. Haltaufderheide, R. Ranisch. "The ethics of ChatGPT in medicine and healthcare: A systematic review on Large Language Models (LLMs)". 『NPJ Digital Medicine』. 7. 2024.
- M. D. Abràmoff, M. E. Tarver, N. Noyo-Berrios, et al. "Considerations for addressing bias in artificial intelligence for health equity". 『NPJ Digital Medicine』. 6. 2023.
- R. Baumgartner, P. Arora, C. Bath, et al. "Fair and equitable AI in biomedical research and healthcare: Social science perspectives". 『Artificial Intelligence in Medicine』. 144. 2023.
- F. Dehghani, N. Malik, J. Lin, et al. "Fairness in healthcare: Assessing data bias and algorithmic fairness". 『2024 20th International Symposium on Medical Information Processing and Analysis (SIPAIM)』. 2024.
- R. J. Chen, J. J. Wang, D. F. K. Williamson, et al. "Algorithmic fairness in artificial intelligence for medicine and healthcare". 『Nature Biomedical Engineering』. 7(6). 2023.
- A. S. Bergman, L. A. Hendricks, M. Rauh, et al. "Representation in AI evaluations". 『2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT'23)』. 2023.
- A. Toma, S. Senkaiahliyan, P. R. Lawler, et al. "Generative AI could revolutionize health care—but not if control is ceded to big tech". 『Nature』. 624(7990). 2023.
- F. Pennestrì, F. Cabitza, N. Picerno, et al. "Sharing reliable information worldwide: Healthcare strategies based on artificial intelligence need external validation. Position paper". 『BMC Medical Informatics and Decision Making』. 25. 2025.
- 김준혁. "헬스케어 AI 윤리에서 환자·시민 참여 모형: 주제범위 고찰과 방법론적 검토에 기초하여". 『한국의료 윤리학회지』. 27(4). 2024.
- L. Floridi. "Hypersuasion—On AI's persuasive power and how to deal with it". 『Philosophy & Technology』. 37. 2024.
- P. S. Park, S. Goldsetin, A. O'Gara, et al. "AI deception: A survey of examples, risks, and potential solutions". 『Patterns』. 5(5). 2024.
- A. Singhal, N. Neveditsin, H. Tanveer, et al. "Toward fairness, accountability, transparency, and ethics in AI for social media and health care: Scoping review". 『JMIR Medical Informatics』. 3(12). 2024.
- G. de Togni, S. Erikainen, S. Chan, et al. "What makes AI 'intelligent' and 'caring'? Exploring affect and relationality across three sites of intelligence and care". 『Social Science & Medicine』. 277. 2021.
- E. R. Burgess, I. Jankovic, M. Austin, et al. "Healthcare AI treatment decision support: Design

- p>principles to enhance clinician adoption and trust". 『Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI'23)』. 2023.
- S. Labkoff, B. Oladimeji, J. Kannry, et al. "Toward a responsible future: Recommendations for AI-enabled clinical decision support". 『Journal of the American Medical Informatics Association』. 31(11). 2024.
- M. Sharma, C. Savage, M. Nair et al. "Artificial intelligence applications in health care practice: Scoping review". 『Journal of Medical Internet Research』. 24(10). 2022.
- S. Reddy, W. Rogers, V.-P. Makinen, et al. "Evaluation framework to guide implementation of AI systems into healthcare settings". 『BMJ Health & Care Informatics』. 28(1). 2021.
- U. Upadhyay, A. Gradisek, U. Iqbal, et al. "Call for the responsible artificial intelligence in the healthcare". 『BMJ Health & Care Informatics』. 30. 2023.
- Q. Sun, A. Akman, B. W. Schuller. "Explainable artificial intelligence for medical applications: A review". 『arXiv』 [preprint]. 2024. arXiv:2412.01829.
- A. Kale, T. Nguyen, F. C. Harris Jr., et al. "Provenance documentation to enable explainable and trustworthy AI: A literature review". 『Data Intelligence』. 5(1). 2023.
- G. Falco, B. Shneiderman, J. Badger, et al. "Governing AI safety through independent audits". 『Nature Machine Intelligence』. 3. 2021.
- K. Blagec, J. Kariger, W. Frühwirth, et al. "Benchmark datasets driving artificial intelligence development fail to capture the needs of medical professionals". 『Journal of Biomedical Informatics』. 137. 2023.
- A. Karargyris, R. Umeton, M. J. Scheller, et al. "Federated benchmarking of medical artificial intelligence with MedPerf". 『Nature Machine Intelligence』. 5. 2023.
- N. Sourlos, R. Vliegthart, J. Santinha, et al. "Recommendations for the creation of benchmark datasets for reproducible artificial intelligence in radiology". 『Insights into Imaging』. 15. 2024.
- K. Lekadir, A. F. Frangi, A. R. Porras, et al. "FUTURE-AI: International consensus guideline for trustworthy and deployable artificial intelligence in healthcare". 『BMJ』. 388. 2025.
- 질병관리청. 『보건의료 분야 인공지능 연구자를 위한 연구윤리 지침』. 2023.

The Role of IRBs in Healthcare Generative AI Research and Directions for Revising the Bioethics Act, Based on a Scoping Review*

Junhewk Kim**

ABSTRACT

The rapid growth of healthcare artificial intelligence (AI)—namely, the various ways of applying AI algorithms in health and medicine—urgently calls for the establishment of a governance system for such research. However, no concrete alternative has yet been proposed. This study begins with the premise that since healthcare AI research uses data obtained from humans or targets humans for application, the Bioethics and Safety Act should apply as long as it regulates research on humans. First, in a scoping review, this study analyzed that previous legal and academic discussions on the relationship between healthcare AI and the Bioethics and Safety Act have been primarily limited to de-identification issues. De-identification is not the only issue in healthcare AI research; therefore, the role of the Institutional Review Board (IRB) in reviewing related research must be examined, which this study assessed based on a variety of literature. Additionally, in the forthcoming AI Basic Act (scheduled to take effect in 2026), an AI Ethics Committee is proposed to ensure compliance with ethical principles. However, its functions in the healthcare domain overlap with IRBs in terms of compliance with ethical principles. Therefore, this paper raises the need to reorganize the Bioethics and Safety Act to allow IRBs to fulfill this role instead. In conclusion, this paper proposes specific roles for the

* This work was supported by the ‘Operation of Education Program and Improvement of Ethics Guidelines for the Use of Artificial Intelligent in Healthcare Research’ and ‘Development of Research Ethics Guide for Generative Artificial Intelligence in Digital Healthcare’ from the Korean National Institutes of Health (2023-ER0808-00 and 2025-ER0801-00).

** Assistant Professor, Department of Dental Education, College of Dentistry, Yonsei University. e-mail: mole0619@yuhs.ac

IRB to establish a governance system for healthcare AI research and urges amendments to the Bioethics and Safety Act to facilitate this.

KEYWORDS

healthcare generative AI, Bioethics and Safety Act, IRB, AI Basic Act, AI ethics committee

