

과학기술학 관점에서 본 헬스케어 AI 윤리와 환자·시민 참여

강미량*

한국과학기술원

Healthcare AI Ethics and Patient and Public Engagement:
An Science and Technology Studies (STS) Perspective

Miryang Kang*

Korea Advanced Institute of Science and Technology, Daejeon, Korea

Received: Nov 3, 2024
Revised: Nov 14, 2024
Accepted: Dec 20, 2024

*Corresponding author

Miryang Kang
Korea Advanced Institute of Science
and Technology, Daejeon, Korea.
Tel: +82-42-350-4841
E-mail: miryang1002@kaist.ac.krCopyright © 2024 Korean Society
for Medical Ethics. This is an Open
Access article distributed under the
terms of the Creative Commons
Attribution Non-Commercial License
([http://creativecommons.org/licenses/
by-nc/4.0/](http://creativecommons.org/licenses/by-nc/4.0/)) which permits unrestricted
non-commercial use, distribution, and
reproduction in any medium, provided
the original work is properly cited.

Conflict of interests

No potential conflict of interest relevant
to this article was reported.

Acknowledgements

Not applicable.

Funding information

Not applicable.

Data availability

Upon reasonable request, the datasets
of this study can be available from the
corresponding author.

ORCID

Miryang Kang
<https://orcid.org/0009-0002-9153-4202>

Author contributions

The article is prepared by a single
author.

Abstract

This article discusses Junhewk Kim's "Patient and Public Involvement Model in Healthcare artificial intelligence (AI) Ethics." I focus on two key aspects of Kim's argument: the particularity of healthcare AI research compared to traditional medical research, and the exclusive legitimacy of patients and citizens as the subjects of external validation. I contend that Kim's current reasoning does not elucidate the unique nature of big data in healthcare AI and risks limiting the democratic potential of participatory models. The field of Science and Technology Studies (STS), which explores the relationship between science and society, could help to strengthen Kim's argument by providing a more focused analysis of big data and citizen participation. I conclude that although public involvement is a procedural condition for creating ethical healthcare AI, it is not sufficient on its own to render healthcare AI ethical. The realization of ethical healthcare AI depends on fostering robust interactions among patients, citizens, and experts that go beyond individualistic notions of privacy and ownership and challenge the rigid divide between healthcare professionals and laypeople.

Keywords: big data; healthcare AI; community participation; citizen science; science and technology studies

I. 서론

제시된 논문에서 김준혁(이하 저자)은 헬스케어 AI(artificial intelligence)가 현 보건의료가 당면한 문제(접근성, 공공성, 의료진-환자 간의 위계)를 해결할 것이라는 기술결정론적인 시각에 반대한다. 논문은 헬스케어 AI가 그 자체로 민주성과 공공성을 담보하지 않으며, 오히려 현 자본주의 생산 체제 내에서 기업의 이윤 추구 도구로 활용되거나 특정 집단에 대한 편향과 배제를 일으킬 가능성이 있음을 지적한다. 헬스케어 AI에서 민주성과 공공성을 성취하기 위해 저자가 제안하는 모델은 시민과 환자를 외부 검증 주체로 참여시키는 것이다. 이러한 저자의 주장은 과학기술학의 수행되지 않은 과학, 시민 과학, 환자 지식 연구 전통과 맞닿아 있으며, 현재 범부처 국가 통합 바이오 빅데이터 구축 사업이 진행되고 있음을 고려할 때 매우 시의적절한 제안이

Ethics approval
Not applicable.

다[1-4]. Shin & Kim은 해당 사업을 분석하며 정부가 데이터 제공자와 관련 종사자를 포함하는 거버넌스 구조를 마련할 것을 촉구한 바 있다[4].

이때 특기할 점은 저자가 환자·시민 참여에서 더 나아가 환자와 시민 만이 “헬스케어 AI 윤리의 틀 안에서 검증의 정당성과 합당성을 획득할 수 있는 유일한 주체라고 주장”한다는 점이다(논평자 강조). 이 논평은 저자의 주장을 독자적 헬스케어 AI 윤리의 필요성과 당사자주의에 기반한 참여적 모델로 분리한 후에 각각을 과학기술학적 관점에서 검토할 것이다. 검토의 목적은 반박이 아닌 저자의 주장을 좀 더 설득력 있게 만들기 위해 논거를 보태는 데에 있다. 특히 저자의 관심사가 윤리적 AI 연구 및 개발을 촉진하기 위한 실질적 참여적 모델을 제안하는 데에 있으므로, 저자의 논지 중 논문의 목적과 다르게 참여의 방식을 축소하거나 악화시키는 결과로 이어질 가능성이 있는 부분을 논할 것이다.

II. 주장에 대한 검토

1. 헬스케어 AI의 독특성과 새로운 윤리의 필요성

이 논문의 대전제는 「보건의료데이터 활용 가이드라인」 등 데이터 가명처리를 명시한 기존 임상 연구윤리 지침만으로는 헬스케어 AI를 적절히 다룰 수 없다는 것이다. 저자는 헬스케어 AI 윤리를 데이터 가명처리 문제로 축소하여 이해하는 경향에 두 가지 문제를 제기한다. 첫째, 가명처리만으로는 전통적인 임상 연구와 헬스케어 AI가 데이터를 수집하고 활용하는 방식이 다르다는 점을 다루지 못한다. 기존 임상 연구가 의학적 판단과 관련된 개별 데이터만을 수집했다면, 헬스케어 AI는 다양한 기술을 활용하여 일상 데이터를 포괄적으로 수집한다. 둘째, 가명처리는 생명윤리의 기본 원칙인 “자율성, 악행금지, 정의에 대한 적극적인 답변이 될 수 없다”. 이는 가명처리가 완전한 익명성을 보장하지 못할 뿐더러, 포괄적 데이터를 알고리즘 개발에 활용하는 과정에서 데이터를 제공한 개인 혹은 그가 속한 인구 집단이 피해를 입거나 특정 개인의 배제로 인한 데이터 편향의 문제가 발생할 수 있기 때문이다.

상기 지적은 살아 있는 인간을 다루는 연구를 전제로 하는 기존 생명윤리 지침이 인간을 데이터 집합으로 만드는 기술을 다룰 때 발생하는 간극을 포착한다. 이 간극을 근거로 저자는 독자적 헬스케어 AI 윤리의 필요성을 역설하는데, 이는 헬스케어 AI를 의료 체계 외부의 것으로 보기보다 “여전히 현대 의료 지식, 제도, 환경 안에서 구현되는 보건의료적 대상”으로 간주하기 위한 시도로서 중요한 정책적 함의를 지니고 있다. 그러나 저자가 헬스케어 AI 연구가 어떤 면에서 전통적 임상 연구와 다른지 구체적으로 서술하지 않는 점이 아쉽다. 저자는 “헬스케어 AI는 병의원의 의무기록뿐만 아니라 일상생활에서 측정, 획득되는 건강 관련 데이터의 분석이나 사회에서 포착, 종합되는 인구 수준의 건강 관련 데이터를 모두 그 대상역(對象役)”으로 한다고 지적하지만, 이러한 포괄성이 왜 기존 임상연구와는 근본적으로 다른 데이터 윤리를 요청하는지에 대한 충분한 설명을 제공하지 않는다. 또한 가명처리의 불완전성이나 연구 대상에 대한 잠재적 위협의 가능성, 데이터 편향의 문제, 데이터 소유권 등은 헬스케어 AI뿐만 아니라 기존 인간대상 연구에서도 제기될 수 있는 문제이다[5-7]. 즉 가명처리의 협소함을 논증하는 것은 헬스케어 AI 윤리가 통상적으로 이해되는 방식에 대한 문제제기로는 적당하나, 새로운 헬스케어 AI 윤리의 필요성 및 참여적 모델의 정당성을 주장하기에는 부족해 보인다.

독자적인 헬스케어 AI 윤리의 필요성은 데이터에 관한 과학사·과학기술학 연구를 통해 보다 정치하게 설득할 수 있다. 과학사학자 브뤼노 스트라세(Bruno Strasser)와 폴 에드워즈(Paul Edwards)는 “무엇이 데이터로 간주되는가”에 대한 답이 역사적으로 변해왔다고 짚는다[8]. 예를 들어, 20세기 중반까지만 해도 기후과학의 데이터란 관측소의 측정값을 의미했으나 1980년대 이후에는 역사적 관측값에 기반한 모델 결과값 또한 데이터로 여겨지게 되었다[9]. 이 질문이 중요한 이유는 질문에 대한 답에 따라 연구 질문, 목적, 설계가 달라지기 때문이다. 저자가 지적한 대로 헬스케어 AI가 기존 의무기록 활용을 넘어 일상생활 데이터(real-world data)를 수집한다면, 문제는 단순히 데이터의 포괄성이 아니다. 보다 근본적인 문제는 이제까지 데이터로 여겨지지 않았던 것을 데이터화(datafication)하려는 시도 자체가 임상 연구 질문, 목적, 범위, 설계, 수행 방법에 근본적인 차이를 만든다는 점이다.

일례로, 일부 헬스케어 AI는 인공지능 모델을 활용하여 일상생활 데이터셋으로부터 의학적 연관성을 도출해내고자 한다. 이때 일상생활 데이터셋은 알고리즘을 통해 분석되기 전까지는 가능성의 상태(‘의학적 함의가 있을 수도 있음’)에 머문다. 알고리즘은 이 가능성을 의학적 사실로 정당화하는 데에 큰 영향을 끼친다. 헬스케어 AI의 여러 요소 중 특히 알고리즘의 불투명성이 문제가 되는 이유가 여기에 있다. 데이터의 기반이 된 사람들에 대한 의학적 추론이 알고리즘이라 불리는 수백, 수천 줄의 코드에 의해 생산되고, 이 과정에서 의학의 영역에 속하지 않았던 현상이 의학적 문제로 새롭게 정의된다. 이 과정이 항상 논리적으로 설명되는 것은 아니지만, 인공지능이 가진 수사적 권력은 생산된 추론을 사실로 만들곤 한다. 즉, 헬스케어 AI에서 설명가능성은 알 권리를 넘어 특정 집단에 대해 생산된 의학적 사실의 구조를 밝힐 수 있는냐의 문제이다. 이렇듯 집단을 선별하고 규정한다는 점에서 헬스케어 AI는 전문가뿐 아니라 시민과 환자의 참여가 필요한 윤리적, 민주적, 정치적 문제가 된다.

더 나아가 헬스케어 AI 윤리는 해당 기술이 활용하는 데이터가 반드시 개인을 준거점으로 삼지 않을 수 있다는 사실도 고려해야 한다. 저자가 “개인의 다면적 특성을 반영하는 (...) 데이터의 포괄 집합”으로 헬스케어 AI 데이터의 특성을 서술할 때, 이는 식별 가능한 개인이 생산하는 데이터의 총합, 혹은 여러 개인이 모인 합으로서 집단이 생산하는 데이터를 상정한다. 그러나 과학기술학자들은 “특정 가능한 것과 개인은 빅데이터 분석에서 더 이상 중심을 차지하지 않는다”¹⁾고 지적했다[10]. 중요한 것은 집단에 대한 확률 및 예측 계산을 가능하게 하는 데이터 모음이기에, 헬스케어 AI는 식별 가능한 데이터를 가명처리하여 활용하는 대신 개인을 비식별한 데이터에서 집단에 대한 분석만을 추구할 수 있다. 혹은 개인과 데이터 간의 연결 관계를 끊기 위해 합성 데이터를 만들 수도 있다. 이러한 유형의 데이터들은 집단에 대한 사실을 생산한다는 점에서 여전히 중요한 윤리적, 정치적 고려의 대상이지만, 개인의 동의, 식별 가능성, 가명처리를 주요하게 논의하는 현 의료윤리 가이드라인으로는 다룰 수 없다[10,11].

2. 환자·시민 참여의 정당성 문제

「보건의료 분야 인공지능 연구자를 위한 연구윤리 지침」은 “연구의 검증·평가는 독립적인

1) Talyor L, Floridi L, van der Sloot B. Chapter 1 Introduction: a new perspective on privacy. In: Talyor L, Floridi L, van der Sloot B, editors. Group Privacy: New Challenges of Data Technologies. Springer International Publishing; 2017, p. 3.

기관을 통하여 정확하고 엄격하게 진행할 것”²⁾을 명기하고 있다[12]. 그러나 저자에 따르면 외부 검증을 실현하기 위한 “윤리적 논의나 절차적 고려는 거의 이루어지지 않으며, 그를 위한 방법론적 검토는 전무한 상황”이다. 이에 저자는 외부 검증 방법으로 환자와 시민 참여 모델을 제안하면서 “이들만이 헬스케어 AI 윤리의 틀 안에서 검증의 정당성과 합당성을 획득할 수 있는 유일한 주체라고 주장”한다. 저자는 두 가지 근거를 들어 이들의 유일한 정당성을 뒷받침한다. 첫째, 외부 전문가는 환자와 시민을 대표할 수 없다. 둘째, 외부 전문가의 관점은 내부 연구자의 관점과 큰 차이가 없으므로 외부 전문가가 검증을 실시한다면 절차적 낭비이다.

논평자는 추상적 원칙으로만 명기된 외부 검증을 환자와 시민 참여라는 모델을 통해 실현하려는 저자의 방향에 동의한다. 이는 전문가 외 더 많은 이해관계자를 헬스케어 AI 논의에 참여시켜 기술의 민주성과 공공성을 증진하려는 시도이다. 그러나 저자의 현 논제가 환자·시민 참여의 정당성을 설득하지 못할 뿐더러, 저자가 목표로 하는 참여적 모델의 실천적 가능성을 제한할 수 있다는 점을 지적하고자 한다. 저자의 두 논거는 모두 환자와 시민이 외부 검증에 어떤 기여를 할 수 있는지 논의하기보다, 외부 전문가가 검증을 효과적으로 수행할 수 없음을 보이고 있다. 본문에서 데이터 소유권과 통제권 이외에는 환자·시민 참여의 정당성 근거를 찾기 어려운데, 이는 저자의 대표성 주장이 오직 데이터를 제공한 사람만이 정당하게 말할 수 있다는 당사자주의에서 비롯된 것임을 시사한다.

이러한 논리는 두 가지 문제를 안고 있다. 첫째, 환자와 시민이 검증에 참여한다고 해서 내부 집단의 대표성 문제가 사라지는 것은 아니다. 헬스케어 AI가 수천, 수만 명의 사람을 대상으로 데이터를 수집한다고 했을 때, 이 중 누가 외부 검증에 참여할 대표가 되어야 하는가? 보다 중요한 문제는 당사자주의에 기반을 둔 대표성 논리가 외부 검증을 위한 자격을 설정한다는 점이다. 앞서 논의했듯이 헬스케어 AI는 특정 집단에 대한 의학적 사실을 생산한다. 그렇다면, 데이터의 일부가 되지 못하는 못했지만 새롭게 생산된 사실에 의해 영향을 받을 것으로 예상되는 사람들은 외부 검증에 참여할 수 있는가? 합성 데이터를 주로 쓰는 헬스케어 AI에 대해 외부 검증을 실시할 수 있는가? 이 질문은 데이터 소유권과 통제권을 근거로 환자·시민 참여가 조직될 때, 참여적 모델의 실천적 가능성이 축소될 수 있음을 보여준다.

그렇다면 환자와 시민은 어떤 기여를 할 수 있는가? 참여적 지식 생산에 관한 과학기술학 연구들은 비전문가들이 세계에 대한 경험적 지식을 형성한다는 점을 보여왔다. 이러한 지식은 일반인 전문성, 환자 지식, 장애 전문성이라는 용어들로 표현된다[2,13,14]. 과학기술학자들이 비전문가의 주관적 경험이나 인식 정도로 여겨지던 것들을 지식과 전문성으로 호명하는 까닭은 근대 과학이 합리성과 객관성을 담보하는 유일한 체계가 아니라는 깨달음 때문이다. 지식의 위치성(situatedness)에 대한 에세이에서 도나 해러웨이(Donna Haraway)는 부분적 관점이라는 테제를 제안하는데, 몸을 가진 모든 사람은 자신의 자리에서 세상을 바라보기에 그 시각은 부분적이라는 주장이다[15]. 전문가라 할지라도 특정 대상에 대한 완전한 지식을 가질 수 없으며, 오로지 그가 살아온 궤적과 그가 받은 학술적 훈련이 허용한 것만 볼 수 있다. 환자·시민 참여가 필요한 본질적 이유는 우리 사회가 인공지능을 매개로 의과학적 지식을 창출하려 할 때, 여러 부분적 관점들을 모아 그 방향과 목적에 대해 숙의할 필요가 있기 때문이다. 이때 환자와 시민의 관점은 전문가의 지식을 보완하고 반박하여 더 나은 결정을 이끄는 데 기여할 수 있다.

2) National Institute of Health. Research ethics guidelines for healthcare AI researchers. 2023, p. 22.

그러나 환자와 시민의 참여를 주장한다고 해서, 그들이 외부 검증의 유일한 주체로 정당화되는 것은 아니다. 저자가 대안적 모델로 제시한 참여적 설계, AI 리터러시, 시민과학은 모두 환자·시민과 전문가 사이의 상호작용이 잘 조직되었을 때 참여가 과학 지식으로 번역되고 활용될 수 있음을 전제로 하고 있다. 환자와 시민을 유일한 외부 검증의 주체로 세우는 것은 이러한 상호작용의 측면을 간과하며, 오히려 전문가에게서 이들과 대화할 책임을 소거하고 참여를 형식적 절차 정도로 축소하는 결과를 낳을 수도 있다. 시민과학에 대한 한 비판적 연구는 시민이 특정 프로젝트의 정치적 정당성을 확보하기 위한 도구로 동원(co-optation)될 수 있음을 보였다[16]. 이는 그 중요성에도 불구하고 환자·시민 참여 자체가 헬스케어 AI의 윤리적 연구개발을 담보하지 않을 수 있음을 암시한다. 저자 또한 “패널이나 포럼과 같은 환자·시민 참여의 전통적인 방법론이 (...) 검토 대상과 외부의 엄격한 분리를 고수하고 있”어 외부 검증이 “형식주의”에 그칠 수 있다고 정확하게 지적하지만, 이러한 지적이 환자와 시민을 외부 검증의 유일한 주체로 세우는 주장과 어떻게 공존할 수 있는지 모호하다. 이런 모호함은 환자·시민 참여가 반드시 외부 검증의 형식을 띠어야 하는가라는 질문과 관련되는 한편, 당사자주의에 기반한 정당성 주장이 예견한 것이기도 하다.

III. 결론

이 논평은 헬스케어 AI의 특이성과 환자·시민 참여의 정당성에 대한 설명을 과학기술학적 관점에서 보완하는 것을 목적으로 했다. 이 과정에서 저자의 논지가 기대고 있는 개인에 대한 윤리와 유일성 주장이 저자가 목표로 하는 민주적 참여의 지평을 오히려 축소할 수 있음을 보였다. 참여는 윤리적 헬스케어 AI 연구개발을 위한 절차적 조건이 될 수는 있지만, 그 자체가 윤리적 문제의 해결책은 아니다. 또한 환자·시민 참여가 꼭 당사자주의에 기반한 정당성과 유일성의 논리 하에 조직될 필요는 없다. 참여적 모델이 필요하다면, 어떤 절차와 어떤 방법론이 가장 적합할 것인가? 이 질문에 대한 답에 따라 환자·시민 참여의 양상과 성격이 달라질 것이다. 

REFERENCES

1. Hess DJ. Undone science: social movements, mobilized publics, and industrial transitions. The MIT Press; 2016.
2. Epstein S. The construction of lay expertise: AIDS activism and the forging of credibility in the reform of clinical trials. *Sci Technol Human Value* 1995;20(4):408-437. <https://doi.org/10.1177/016224399502000402>
3. Brown P. Popular epidemiology and toxic waste contamination: lay and professional ways of knowing. *J Health Soc Behav* 1992;33(3):267-281. <https://doi.org/10.2307/2137356>
4. Shin Y, Kim CJ. The government should play a role greater than that of construction supervisor in the national bio-big data infrastructure. *J Korean Med Soc* 2024;39(36):e262. <https://doi.org/10.3346/jkms.2024.39.e262>
5. Criado Perez C. Invisible women: data bias in a world designed for men. Chatto & Windus; 2019.
6. Eubanks V. Automating inequality: how high-tech tools profile, police, and punish the poor. St. Martin's Press; 2018.

7. Petryna A. *When experiments travel: clinical trials and the global search for human subjects*. Princeton University Press; 2009.
8. Strasser BJ, Edwards PN. Big data is the answer ... but what is the question? *Osiris* 2017;32(1):328-345.
<https://doi.org/10.1086/694223>
9. Edwards PN. *A vast machine: computer models, climate data, and the politics of global warming*. The MIT Press; 2013.
10. Talyor L, Floridi L, van der Sloot B. Introduction: a new perspective on privacy. In: Talyor L, Floridi L, van der Sloot B, editors. *Group privacy: new challenges of data technologies*. Springer International Publishing; 2017, pp. 1-12.
11. Helm P, Lipp B, Pujadas R. Generating reality and silencing debate: synthetic data as discursive device. *Big Data Soc* 2024;11(2):2053951724124944.
<https://doi.org/10.1177/20539517241249447>
12. National Institute of Health. *Research ethics guidelines for healthcare AI researchers*. National Institute of Health; 2023.
13. Pols J. Knowing patients: turning patient knowledge into science. *Sci Technol Human Value* 2014;39(1):73-97.
<https://doi.org/10.1177/0162243913504306>
14. Hartblay C. Disability expertise: claiming disability anthropology. *Curr Anthropol* 2020;61(S21):S26-S36.
<https://doi.org/10.1086/705781>
15. Haraway D. Situated knowledges: the science question in feminism and the privileged of partial perspective. *Fem Stud* 1988;14(3):575-599.
<https://doi.org/10.2307/3178066>
16. Blacker S, Kimura AH, Kinchy A. When citizen science is public relations. *Soc Stud Sci* 2021;51(5):780-796.
<https://doi.org/10.1177/03063127211027662>