April 14, 2025

생성형 AI 의료기기, 안전하게 개발하려면
- 위해요인과 GMLP(우수 기계학습 기준) 지도원칙 중심으로

I. 들어가며

최근 인공지능 기술이 급속도로 발전함에 따라 의료 분야에서도 활발히 활용되고 있습니다. 기존에는 CT, MRI 등 영상의학 분야에서 의료영상 분석 및 진단보조에 인공지능이 주로 활용되었으나, 최근에는 생성형 인공지능 기반 챗봇을 활용한 정신건강 테라피(이른바 "**Therabot**")의 임상시험[1]을 통해 주요우울장애, 불안장애 등에 치료 효과가 있는 것으로 확인되는 등 생성형 인공지능을 활용한 정신건강 치료로의 확대도 기대되고 있습니다. 한편, 생성형 인공지능에는 환각이나 편향 등의 문제가 수반될 수 있으며, 의료 영역에서 이러한 오류는 환자의 생명·신체에 중대한 위해를 끼칠 수 있는 만큼, 그 위해 요인은 체계적으로 통제되어야 할 것입니다.

식품의약품안전처(이하 "**식약처**")는 인공지능을 활용한 의료기기 개발에 있어 2025. 1. 세계 최초로 『생성형 인공지능 의료기기 허가·심사 가이드라인』을 발간하고, 2025. 3. 의료기기 개발을 위한 우수 기계학습 기준(Good Machine Learning Practice) 10대 원칙을 발간하는 등 국제 규범 형성에 적극적으로 참여하고 있습니다. 본 뉴스레터에서는 인공지능 의료기기의 위해요인과 이를 통제하기 위한 식약처의 가이드라인들의 주요내용을 간단히 소개하고자 합니다.

II. 의료기기 해당여부 판단기준 – 사용목적과 위해도

생성형 인공지능을 활용하여 사용자의 건강을 증진하는 헬스케어 앱을 개발하는 경우를 예를 들어 설명해 보겠습니다.

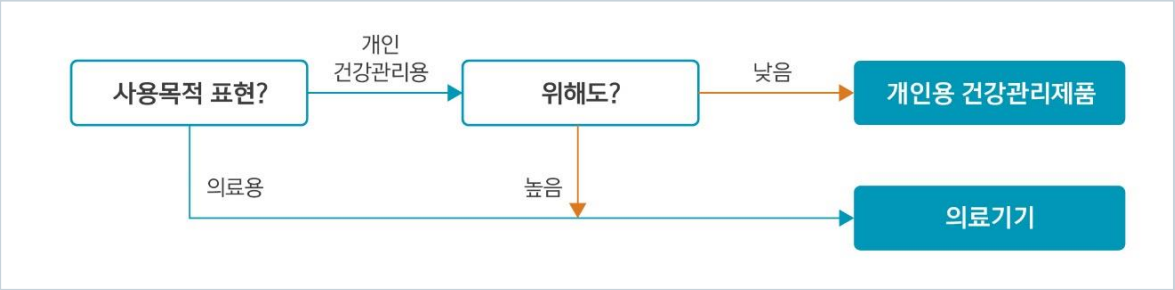
식약처는 해당 헬스케어 앱(AI 소프트웨어)이 의료기기에 해당하는지 여부를 그 **(가) 사용목적과 (나) 위해도**에 따라 판단합니다[2].

---

1 Kumar, S., Zhang, J., Kim, J., et al. (2024). Efficacy of a Generative AI Therapy Chatbot for Mental Health: A Randomized Clinical Trial. NEJM AI, 1(4). https://doi.org/10.1056/AIoa2400802

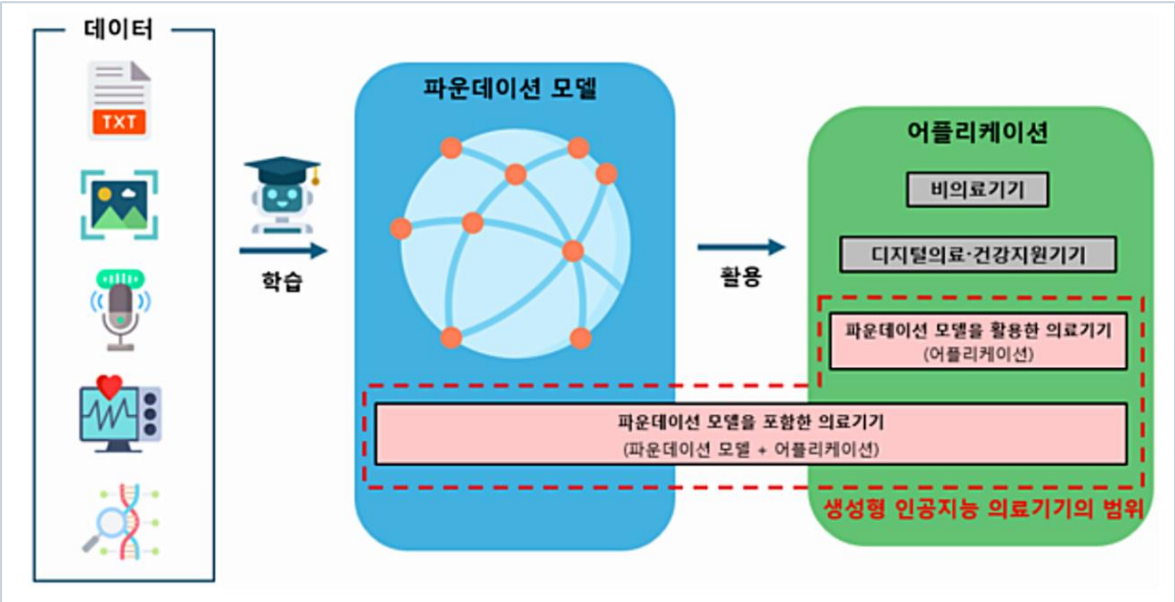
2 식약처, 인공지능 의료기기의 허가·심사 가이드라인, 2022. 5., 9-13면 참조

<의료기기 해당여부 판단기준>[3]



이 중 **(가) 사용목적**이란 제품의 제조자가 의도하여 표방한 기능과 용도를 말하며[4], 제조자가 AI 소프트웨어에 대하여 '질병을 진단, 치료, 예방하는 등의 목적'을 직접적으로 표방하는 경우라면 사용목적상 '디지털의료기기'에 해당할 가능성이 높습니다. 그러나, 생성형 인공지능을 이용하더라도, 비의료기기 또는 디지털의료·건강관리지원기기[5]에 해당하는 경우에는 '생성형 인공지능 의료기기'로 관리되지 않습니다.

<생성형 인공지능 의료기기의 관리범위>[6]



---

3 식약처, 의료기기와 개인용 건강관리(웰니스) 제품 판단기준, 2020. 11., 7면 참조

4 사용목적(intended use)은 구체적으로는 제품 표시사항(규격, 제품설명서, 첨부자료 등)과 광고물 등에서 표현된 제조자의 객관적인 의도를 종합적으로 고려하여 판단합니다{식약처, 의료기기와 개인용 건강관리(웰니스) 제품 판단기준, 2020. 11., 7면 참조}.

5 '디지털의료·건강관리지원기기'란 디지털의료기기에 해당하지 아니하나 의료의 지원 또는 건강의 유지·향상을 목적으로 생체신호를 모니터링·측정·수집 및 분석하거나, 생활습관을 기록·분석하여 식이·운동 등 건강관리 정보를 제공하는 목적으로 사용되는 디지털기술이 적용된 기구·기계·장치·소프트웨어 또는 이와 유사한 제품으로서 식약처장이 지정하는 제품을 말하며(디지털의료제품법 제2조 제4호), 디지털의료·건강관리지원기기의 제조·수입, 성능관리, 유통관리 등에 대한 규정은 2026. 1. 24. 시행될 예정입니다.

6 식약처, 생성형 인공지능 의료기기 허가·심사 가이드라인, 2025. 1., 2면 참조

한편, 제조자가 해당 제품의 사용목적을 개인 건강관리용으로 표방하는 경우라도, 그 **(나) 위해도**가 높다면 의료기기에 해당하여 식약처장의 인허가 대상이 될 수 있습니다. '위해도'는 ① 생체적합성 문제를 야기하는지, ② 침습적인지, ③ 사용의도대로 작동되지 않을 경우 사용자에게 상해, 질병이 발생하는지, ④ 위급한 상황을 탐지하는 기능을 수행하는지, ⑤ 기기의 기능이나 특성을 통제・변경하는지 등의 구체적 판단 요소에 따라 판단됩니다[7].

III. 인공지능 의료기기의 위해요인

앞서 설명드린 바와 같이, 위해도는 의료기기와 비의료기기 내지 디지털의료・건강관리 지원기기를 구분하는 중요한 판단기준이 됩니다. 그런데, 스마트폰에서 사용되는 헬스케어 앱 형태의 제품을 개발하는 경우와 같이, 환자의 신체와 접촉하는 요소(①, ②), 위급 상황 감지 기능(④), 하드웨어 통제 기능(⑤)이 배제된, 인공지능을 이용한 소프트웨어 제품을 개발하는 경우라면, 상기 판단요소 중 **사용의도대로 작동되지 않을 경우 사용자에게 상해, 질병이 발생하는지** 여부(③)가 가장 중요하다고 볼 수 있습니다.

이러한 관점에서, 식약처는 『인공지능 의료기기의 허가・심사 가이드라인』(초판 2017. 11. 제정)에서 인공지능을 이용한 소프트웨어 제품의 위해도 판단 요소로서 **데이터 또는 알고리즘의 오류**, **임상적 근거에 대한 설명가능성(훈련 데이터셋의 출처, 훈련 데이터셋과 결과물 간의 상관관계 등)** 등을 위해도 판단 요소로 설명하고 있습니다.

<의료기기 해당여부 판단기준 – 위해도 판단 요소>[8]

(1) 소프트웨어가 의도한 대로 작동하지 않아 환자에게 위해를 끼칠 가능성이 있는지 여부

"의료용 소프트웨어는 정확도가 담보되지 않으면 국민건강에 위해를 발생시킬 우려가 있다. 예를 들어 **부정확, 부적정한 데이터를 입력・학습**하거나 **알고리즘의 오류**로 인해 질병 유무에 대한 가능성 정도를 잘못 예측할 수 있으며, 이상 부위를 잘못 검출・표시하는 경우에는 진단・치료 결과에 직접적으로 영향을 미친다."

(2) 소프트웨어가 의료인의 임상적 판단을 보장하는지 여부

"의료인이 환자에 대한 임상적 진단이나 치료 방법 등을 결정할 때 (소프트웨어가 제시한) 해당 권장 사항만으로 주요 판단을 내리는 것이 아니라는 것을 알려야 하며, **의료인이 제공된 정보에 대한 임상적 근거를 파악할 수 있도록 훈련 데이터셋의 출처, 훈련 데이터셋과 제공되는 결과물 간 상관관계 등에 대한 충분한 설명을 제공**하여야 한다."

7 식약처, 의료기기와 개인용 건강관리(웰니스) 제품 판단기준, 2020. 11., 8면 참조

8 식약처, 인공지능 의료기기의 허가·심사 가이드라인, 2022. 5., 9면 참조

한편, 식약처는 2025년 1월, 세계 최초로 『생성형 인공지능 의료기기 허가·심사 가이드라인』을 발간하면서, 생성형 인공지능 의료기기에서 고려해야 할 위해요인을 성능, 데이터 품질, 편향, 사용자, 적응형 시스템, 기타의 여섯 범주로 구분하여 구체적으로 제시하였습니다. 이는 미국의료기기진흥협회(AAMI[9])에서 발간한 『의료기기 위험관리 국제표준(ISO 14971) 기반의 인공지능·기계학습 기술보고서[10]』를 참고하였습니다.

<생성형 인공지능 의료기기 위해요인(Hazard) 예시>[11]

구분	대표적인 위해요인(Hazard) 예시
성능 (Performance)	설득력 있는 환각 (Hallucination), 일관성 없음 (Inconsistency), 연관성 없음 (Irrelevancy), 불확실성 척도 부재 (No uncertainty indicator) 등
데이터 품질 (Data Quality)	데이터 오류 (Incorrect data), 이상치 처리 오류 (Incorrect handling of outliers), 데이터 드리프트 (Data drift), 학습/적용 데이터 불일치 (Domain shifted data) 등
편향 (Bias)	선택편향 (Selection bias), 중첩변수 (Confounding variables), 암시적 편향 (Implicit bias), 집단 편향 (Group attribution bias) 등
사용자 (User)	과잉 확신 (Overconfidence), 인지된 위험 (Perceived risk), 사용자 신뢰 차이 (Variation in social trust)
적응형 시스템 (Adaptive System)	연속 학습 (Continuously learning) 등
기타 (Others)	지식 부족 (Lack of knowledge) 등

IV. 우수 기계학습 기준(Good Machine Learning Practice) 주요 내용

식약처는 2025. 3. 26. 『의료기기 개발을 위한 우수 기계학습 기준(Good Machine Learning Practice; "**GMLP**"): 지도 원칙』 가이드라인을 발간하였습니다. 이 가이드라인은 식약처가 국제의료기기규제당국자포럼(IMDRF) 인공지능 기계학습 실무그룹에 2년간 참여하여 공동으로 개발한 국제 공통 문서로, 2025년 1월 IMDRF에 등재되었습니다.

위 가이드라인에는 인공지능 의료기기의 개발, 학습, 검증, 임상 적용 전 과정에서 고려해야 할 10가지 핵심 원칙이 담겨 있으며, 데이터의 독립성, 임상시험 데이터의 대표성, 모델 설계의 적합성, 위험 모니터링 등 신뢰성 확보를 위한 구체적인 기준을 제시하고

9 Association for the Advancement of Medical Instrumentation

10 AAMI TIR34971:2023. Application of ISO 14971 to Machine Learning in Artificial Intelligence – Guide. Association for the Advancement of Medical Instrumentation (AAMI), 2023.

11 식약처, 생성형 인공지능 의료기기 허가·심사 가이드라인, 2025. 1., 8-11면에서 발췌 참조

있습니다. 식약처는 본 가이드라인이 국내 기업의 국제 기준 부합 제품 개발과 글로벌 시장 진출에 기여할 것으로 기대하고 있습니다.

<의료기기 개발을 위한 우수 기계학습 기준: 지도 원칙(Good Machine Learning Practice for Medical Device Development: Guiding Principles)>[12]

주요 내용: 10 대 주요 원칙
① 의료기기의 사용 목적을 명확히 이해하고, 제품의 전체 생명주기 동안 다학제적 전문성을 활용한다.
② 전체 생명주기에 걸쳐 우수한 소프트웨어 공학, 의료기기 설계 및 보안 원칙을 적용한다.
③ 임상 평가 시, 의도된 환자 집단을 대표할 수 있는 데이터셋을 포함한다.
④ 훈련 데이터와 시험 데이터는 독립적으로 설정한다.
⑤ 사용 목적에 부합하는 참조 표준을 선택한다.
⑥ 사용 가능한 데이터와 의료기기의 사용 목적에 맞게 모델을 선택하고 설계한다.
⑦ 의료기기는 단순히 기기 단독의 성능이 아니라, 의도된 사용 환경에서 인간-AI 협업 팀(Human-AI Team)*의 성능을 포함한 인간과 인공지능의 상호작용에 초점을 맞춰 평가한다. * 인간-AI 협업 팀: 인공지능이 분석한 결과를 참고하여 의료인이 최종 진단을 내리는 등의 인간-인공지능 간 협력 체계
⑧ 임상적으로 적합한 조건에서 의료기기의 성능을 시험한다.
⑨ 사용자에게 명확하고 필수적인 정보를 제공한다.
⑩ 배포된 모델의 성능을 지속적으로 모니터링하고, 재훈련의 위험을 관리한다.

V. 시사점

식약처의 가이드라인들을 참고할 때, 생성형 인공지능 기반 의료기기를 개발함에 있어서는 환자의 생명과 신체에 미치는 영향을 고려하여, 위해요인이 통제 가능한 범위 내에서 사용 목적을 구체적으로 설정할 필요가 있습니다. 특히, 임상평가에 적합하고 대표성 있는 훈련 데이터를 확보해야 하며, 실제 의료 환경에서 의료인의 판단과 상호작용이 적절히 개입될 수 있도록 시스템 구조를 설계하는 것이 필수적입니다.

또한, 인공지능 의료기기 분야에서 한국은 세계 최초로 '디지털 의료제품법'을 제정·시행하고, 인공지능 진단보조기기 등 기존 의료기기 허가 경험을 축적해온 바 있으며, 식약처는 국제 인공지능 의료기기 규범 형성 과정에서도 ISO, IMDRF 등 주요 협의체를 통해 적극적인 목소리를 내고 있습니다. 이러한 선도적 입지와 산업 레퍼런스를 고려할 때, 한국이 생성형 인공지능 기반 의료기기의 글로벌 시장 진출을 위한 전략적 테스트베드로서의 역할을 수행할 수 있을 것으로 기대됩니다.

12 식약처 2025. 3. 26.자 보도자료, <의료기기 개발 위한 기계학습국제공통 지침 발간>

이러한 환경에서 기업은 생성형 인공지능 기반 의료기기를 개발할 때, 개발 초기 단계부터 식약처의 규제 요건을 반영하여 구조를 설계하고, 임상평가에 적합한 고품질 데이터를 확보하고 편향 가능성을 사전에 점검하는 데이터 전략이 필수적입니다. 또한 의료인의 판단과 개입이 가능하도록 시스템을 설계하고 실제 임상 환경에 부합하는 사용자 경험을 고려해야 합니다. 더불어, 국내외 규제 변화에 선제적으로 대응할 수 있도록 관련 법령 및 가이드라인을 지속적으로 모니터링하고, 기술 문서 및 품질관리체계를 정비해야 합니다. 마지막으로, 환자 데이터 보호와 함께 환각, 편향 등 생성형 AI의 고유 리스크에 대응할 수 있는 내부 윤리 기준과 운영 체계를 갖추는 것도 중요합니다.

법무법인 태평양의 AI팀은 생성형 인공지능 기반 의료기기 개발시 보다 실효적이고 전략적인 로드맵(Roadmap)을 제공해 드리겠습니다.

관련 구성원

강태욱
변호사
T 02.3404.0485
E taeuk.kang@bkl.co.kr

윤주호
변호사
T 02.3404.6542
E juho.yoon@bkl.co.kr

강정희
변호사
T 02.3404.6480
E jeonghee.kang@bkl.co.kr