Off-Label Use of AI-Enabled Medical Devices

Dr Gabriela Lenarczyk, attorney-at-law
BFP Law firm

Why this, why now, why you?

medical negligence | product liability | patient rights | regulatory compliance

SCALE

FDA lists
~950 AI/ML-enabled devices (June 2024)
98% via 510(k) since 2021.
Radiology Health AI Register:
~214 CE-marked AI products in radiology

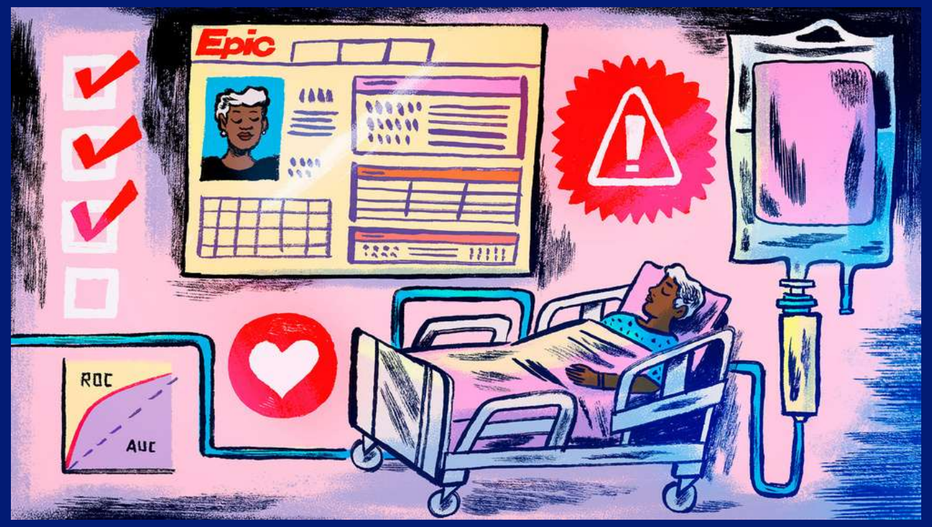
Clinical performance studies?
Paediatric use?

LEGAL ENVIRONMENT

Codes of Medical Ethics
EU AI literacy obligations
Product Liability Directive
MDR+IVDR+AI Act

PEOPIL

BFP



Original Investigation FREE

External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients

Andrew Wong, MD[1]; Erkin Otles, MEng[2,3]; John P. Donnelly, PhD[4]; et al

» Author Affiliations | Article Information

RELATED ARTICLES MEDIA FIGURES SUPPLEMENTAL CONTENT

Key Points

Question How accurately does the Epic Sepsis Model, a proprietary sepsis prediction model implemented at hundreds of US hospitals, predict the onset of sepsis?

Findings In this cohort study of 27697 patients undergoing 38455 hospitalizations, sepsis occurred in 7% of the hosptalizations. The Epic Sepsis Model predicted the onset of sepsis with an area under the curve of 0.63, which is substantially worse than the performance reported by its developer.

Meaning This study suggests that the Epic Sepsis Model poorly predicts sepsis; its widespread adoption despite poor performance raises fundamental concerns about sepsis management on a national level.

AI-enabled medical device

software as a medical device,
or software embedded in hardware

intended for
diagnosis, prediction, monitoring, or treatment decisions

off-label use for medical devices

use outside the approved instructions for use, indications, or intended purpose

note “reasonably foreseeable misuse”

PEOPIL

BFP

Locked vs. Adaptive AI

Locked AI

Frozen model that provides the same result each time the same input is applied to it

Adaptive AI

- Learns/updates in field to potentially change its behaviour using a defined learning process
- Poses greater challenges to regulators, with the EU and US using different governance techniques

US Regulatory Approach

- FDA treats AI/ML SaMD under Food, Drug & Cosmetics Act with no new legislation specifically for medical AI
- Practice of Medicine Exemption: clinicians may use devices off-label but manufacturers cannot promote off-label usage.
- Responding to Adaptive AI: PCCP (Predetermined Change Control Plans) define allowed updates + validation
  - 2025 draft adds bias testing & drift monitoring and plans to detect performance drops outside validated populations.

Liability and on-label use

Was the product defective?
Were the instructions and warnings adequate?
Did the hospital procure and deploy it responsibly?
Did the clinician understand its role and limits?
Did anyone over-rely on it?
Did anyone ignore contradictory clinical signs?
Did the institution provide training?
Did anybody check performance in real-world use?
Was informed consent obtained?

PDL: anyone who substantially modifies a product outside the original manufacturer's control can themselves become a manufacturer for liability purposes.

Was informed consent obtained?

Article 12 of the Polish Code of Medical Ethics

Article 12.

A doctor may use artificial intelligence algorithms in diagnostic, treatment or prevention proceedings under the following conditions:

1) the patient is informed that artificial intelligence will be used in making a diagnosis or in the therapeutic process;

2) informed consent of the patient is obtained for the use of artificial intelligence in the diagnostic or therapeutic process;

3) the artificial intelligence algorithms used are approved for medical use and appropriately certified;

4) the final diagnostic and therapeutic decision is always made by the doctor.

Official commentary:

For AI that materially shapes the clinical service or reduces the real influence of the physician, the patient should know and understand that role.

What does off-label mean for AI?

same product,
different indication?

different indication
different age group
different scanner
different imaging protocol
different ward
different category of staff
different workflow
different level of human oversight

AI performance is relational.

It depends on the interaction between the model, the training data, the local hardware, the workflow, user behaviour, and the clinical environment.

Article | Open access | Published: 28 June 2024

The limits of fair medical imaging AI in real-world generalization

Yuzhe Yang, Haoran Zhang, Judy W. Gichoya, Dina Katabi & Marzyeh Ghassemi

Nature Medicine **30**, 2838–2848 (2024) | Cite this article

47k Accesses | **201** Citations | **208** Altmetric | Metrics

Abstract

As artificial intelligence (AI) rapidly approaches human-level performance in medical imaging, it is crucial that it does not exacerbate or propagate healthcare disparities. Previous research established AI's capacity to infer demographic data from chest X-rays, leading to a key concern: do models using demographic shortcuts have unfair predictions across subpopulations? In this study, we conducted a thorough investigation into the extent to which medical AI uses demographic encodings, focusing on potential fairness discrepancies within both in-distribution training sets and external test sets. Our analysis covers three key medical imaging disciplines—radiology, dermatology and ophthalmology—and incorporates data from six global chest X-ray datasets. We confirm that medical imaging AI leverages demographic shortcuts in disease classification. Although correcting shortcuts algorithmically effectively addresses fairness gaps to create 'locally optimal' models within the original data distribution, this optimality is not true in new test settings. Surprisingly, we found that models with less encoding of demographic attributes are often most 'globally optimal', exhibiting better fairness during model evaluation in new test environments. Our work establishes best practices for medical imaging models that maintain their performance and fairness in deployments beyond their initial training contexts, underscoring critical considerations for AI clinical deployments across populations and sites.

What we expect from informed consent will change…

…and so will the standard of care (think claims for over-reliance and for non-use)

- The human factor and automation bias
- Off-label moves are often invisible at the point of care
- Legal benchmarks are becoming more concrete
- Scholarly literature is moving in a plaintiff-relevant direction

Liability map: US

Medical malpractice

departure from standard of care (e.g., unvalidated off-label AI without consent; blind reliance on AI)

Product liability

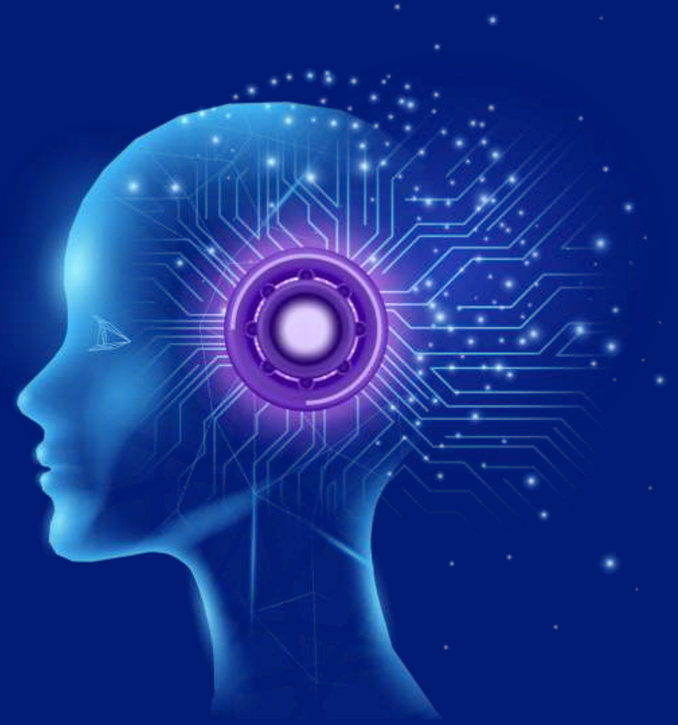
(strict/negligence)
foreseeable misuse → duty to warn/design
unforeseeable misuse → defence

Taylor v. Intuitive: duty to warn/train even when tech works as intended.

Gaps: causation with black-box AI (?); evidentiary burdens.

Policy talk: clinician safe harbours tied to outcome monitoring.

Liability map: EU



New PLD 2024/2853

strict liability now expressly covers software/AI;
lack of updates/fixes can be a defect.

- Foreseeable misuse (ISO 14971): design & warnings must account for realistic off-label patterns.
- National malpractice still governs clinician fault;
  - AI Liability Directive withdrawn (2025) → no EU-wide burden-shifting for now.
- Responsibility gap persists for adaptive systems
  - Who owns drift: maker vs. hospital?

Causation, evidence and PLD

- both parties have a right to request access to evidence
  - instructions for use,
  - the intended-purpose statement,
  - version history,
  - update history,
  - local validation documents,
  - procurement materials,
  - training records,
  - audit logs,
  - override logs,
  - complaints,
  - incident reports,
  - vendor communications,
  - post-market surveillance records, and
  - the institutional policy
- national courts may presume defectiveness or causation, or both, where proof is excessively difficult
- the black box argument

Case patterns to watch

adult-trained AI used in children

the foreseeable-drift case

the update or modification case

The NEW ENGLAND JOURNAL of MEDICINE

CURRENT ISSUE SPECIALTIES TOPICS

ORIGINAL ARTICLE

Survival after Minimally Invasive Radical Hysterectomy for Early-Stage Cervical Cancer

Authors: Alexander Melamed, M.D., M.P.H., Daniel J. Margul, M.D., Ph.D., Ling Chen, M.D, M.P.H., Nancy L. Keating, M.D., M.P.H., Marcela G. del Carmen, M.D., M.P.H., Junhua Yang, M.S., Brandon-Luke L. Seagle, M.D., +6, and J. Alejandro Rauh-Hain, M.D., M.P.H. Author Info & Affiliations

Published October 31, 2018 | N Engl J Med 2018;379:1905-1914 | DOI: 10.1056/NEJMoa1804923

VOL. 379 NO. 20 | Copyright © 2018

FDA In Brief: FDA cautions patients, providers about using robotically-assisted surgical devices for mastectomy and other cancer-related surgeries

February 28, 2019

Media Inquiries

EU Framework

MDR 2017/745

- CE marking to indicate conformity for a defined intended purpose (Art 20)
- Off-label = abnormal use (even where reasonably foreseeable ISO 14971)
- Manufacturers cannot promote misuse (Art 7)
- Post-Market Surveillance must capture abnormal / off-label trends
- A significant change requires a new conformity assessment

AI Act

- Medical AI is high-risk (Art.6, Annex II)
- Requires ongoing risk management, data governance requirements, transparency, human oversight (Arts 9, 10, 13, 14)
- Must have technical documentation in addition to MDR documentation (Art 11)

MDCG Guidance

- Abnormal use, incl. off-label, must be documented & managed" (QMS/PMS)
- Common off-label → regulatory attention

HTAR 2021/2282

- Can pressure evidence generation for expanded indications

PLD 2024/2853

- Explicitly includes AI systems
- Imposes strict liability for certain forms of damage or defectiveness

Thoughts and questions?

Dr Gabriela Lenarczyk
g.lenarczyk@bf.com.pl