Stigma and AI in Healthcare

Annika M Schoene, PhD

05/28/2026

Northeastern University

A little about me …

📌 Research Interests

- Natural Language Processing
- AI Safety & Robustness
- Responsible AI
  - Algorithmic audits and fairness, bias detection
  - AI governance and ethics in mental health applications

a.schoene@northeastern.edu

Overview

I. What is Artificial Intelligence?

II. AI in Healthcare

III. AI Safety as a Pillar of Responsible AI

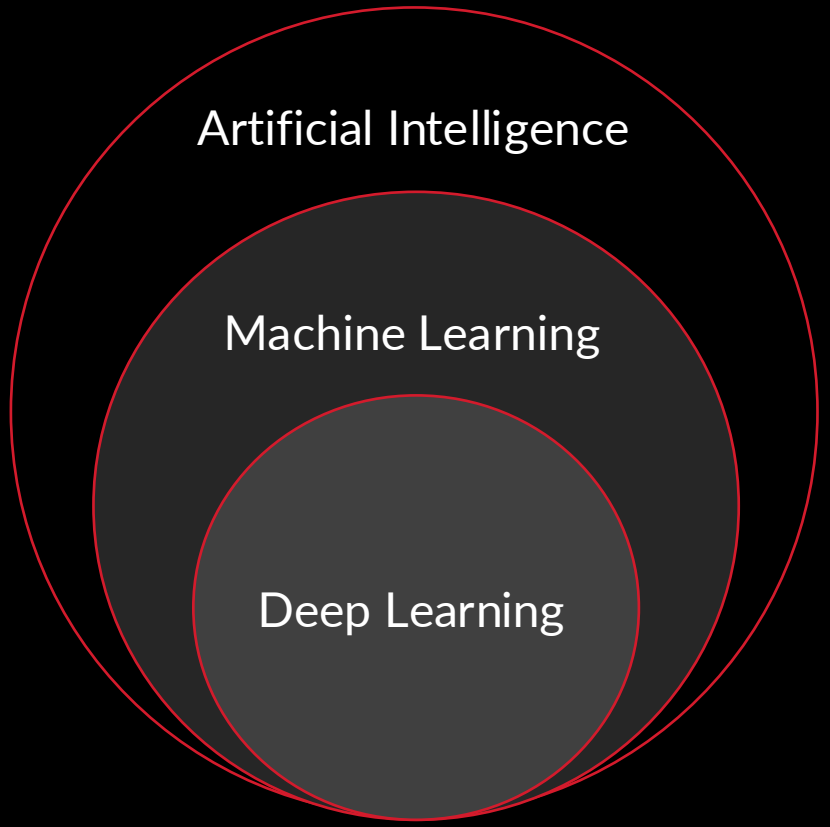
IV. AI & Stigma

V. Key Takeaways

What is Artificial Intelligence?

Artificial Intelligence

AI is the study and design of computational systems capable of perceiving their environment, reasoning about it, learning from experience, and taking actions to achieve specific goals



Fundamental AI Applications

- Machine Learning
- Recommender Systems
- Computer Vision
- Robotics
- Natural Language Processing

(Russell & Norvig, 2021; Poole et al., 2017)

AI in Healthcare

AI in Healthcare

- The global market for AI in healthcare is projected to exceed $868 billion by 2030, reflecting unprecedented growth and adoption trajectories
- Health systems face persistent pressures, where both patient care inconsistencies and workforce shortage can threaten quality and safety of care delivery
- AI is increasingly positioned as a tool to address these inefficiencies and inequities
- Oversight is often inconsistent across tools and institutions

Ambient Listening

- tools listen to visits and draft notes, summaries, or other documentation
- documentation burden is huge, and early studies suggest these tools can reduce note-writing time and perceived burnout

AI Nurse

- remote nurses monitor patients by video, help with admissions / discharges, watch vital signs, and can call in bedside staff when needed
- each virtual nurse watches about 10 patients

Medical Imaging

- Common uses include detecting abnormalities, prioritizing scans, image segmentation, screening,
- help clinicians interpret studies like chest imaging, mammography, retinal imaging, and CT

AI in Incidents in Healthcare

AI Sepsis Model

- randomized vignette study found that exposure to systematically biased AI predictions significantly reduced clinician diagnostic accuracy
- common explanation methods did not offset that harm

AI can worsen clinician decision[1]

- randomized vignette study found that exposure to systematically biased AI predictions significantly reduced clinician diagnostic accuracy
- common explanation methods did not offset that harm

Racial bias in a care-management algorithm

- randomized vignette study found that exposure to systematically biased AI predictions significantly reduced clinician diagnostic accuracy
- common explanation methods did not offset that harm

AI Safety as a Pillar of Responsible AI

What is Responsible AI?

Responsible AI

- Fairness → AI should not systematically disadvantage certain people or groups
- Transparency → There should be clear responsibility for decisions, oversight, and harms
- Accountability → Assigning clear responsibility for AI decisions and outcomes, including mechanisms for rectifying when systems fail or cause harm
- Autonomy → AI should support human judgment, not replace meaningful human control

AI Safety

Fairness	Systems trained on biased datasets can reinforce inequalities, potentially leading to disparities in care
Transparency	Help providers understand, evaluate, and trust AI-generated recommendations
Accountability	Promote clear accountability structures
Autonomy	AI should support human judgment, not replace meaningful human control

AI in Substance Use

NarxCare

- A commercial risk-scoring algorithm embedded in state prescription drug monitoring programs
- Generates an "Overdose Risk Score" plus "Narx Scores" for narcotics, sedatives, and stimulants

AI screening for OUD in hospitalized adults

- an AI screener that flags hospitalized adults at risk for OUD and prompts an addiction-consult referral
- Used to expand access to addiction medicine rather than restrict prescribing

PROVIDENT

- A state-wide, population-based randomized controlled trial using a machine-learning model to forecast census-block-group overdose risk
- Forecasts are handed to community organizations to target naloxone distribution and outreach

AI & Stigma

Stigmatizing language in LLM-generated clinical text

- Mass General Brigham audit tested 14 LLMs on clinically relevant prompts about alcohol use disorder, alcohol-associated liver disease, and SUD
- More than 35% of LLM responses contained stigmatizing language

Rotterdam welfare risk model

- Algorithm scored every welfare recipient on fraud risk using 315 features; mental-health/SUD history was an input that raised the score
- the system operationalizes a “discrediting” attribute that lowers your status as a mathematical feature

Stigma in LLM "therapists"

- Audited general LLMs) and live commercial therapy bots using validated social-distance stigma instruments and crisis-symptom prompts
- Models showed equal or greater stigma toward alcohol dependence and schizophrenia than toward depression in crisis prompts

Challenges of using AI to detect Stigma

- There is no shared, theoretically grounded definition of stigma across AI research; "bias" and "stigma" are routinely used interchangeably
- Most work focuses on a small set of identities, with little attention to intersectional or marginalized populations
- Stigma is contextual and gradient, but AI typically flattens it into a single label or score
- Very few systems are audited for stigma or fairness after deployment; even fewer are re-audited over time
- Proprietary data and commercial tools limit reproducibility and outside scrutiny

Key Takeaways

- AI in healthcare involves significant challenges and tradeoffs, and we can no longer afford to focus only on scale and efficiency. We need to focus on what improves health outcomes while maintaining patient safety.

- AI in healthcare is not just a technical challenge; it is also a clinical, ethical, and governance challenge.

- We need truly interdisciplinary research to build meaningful AI tools for healthcare and to establish how they should be governed.

Questions?

Thank you!