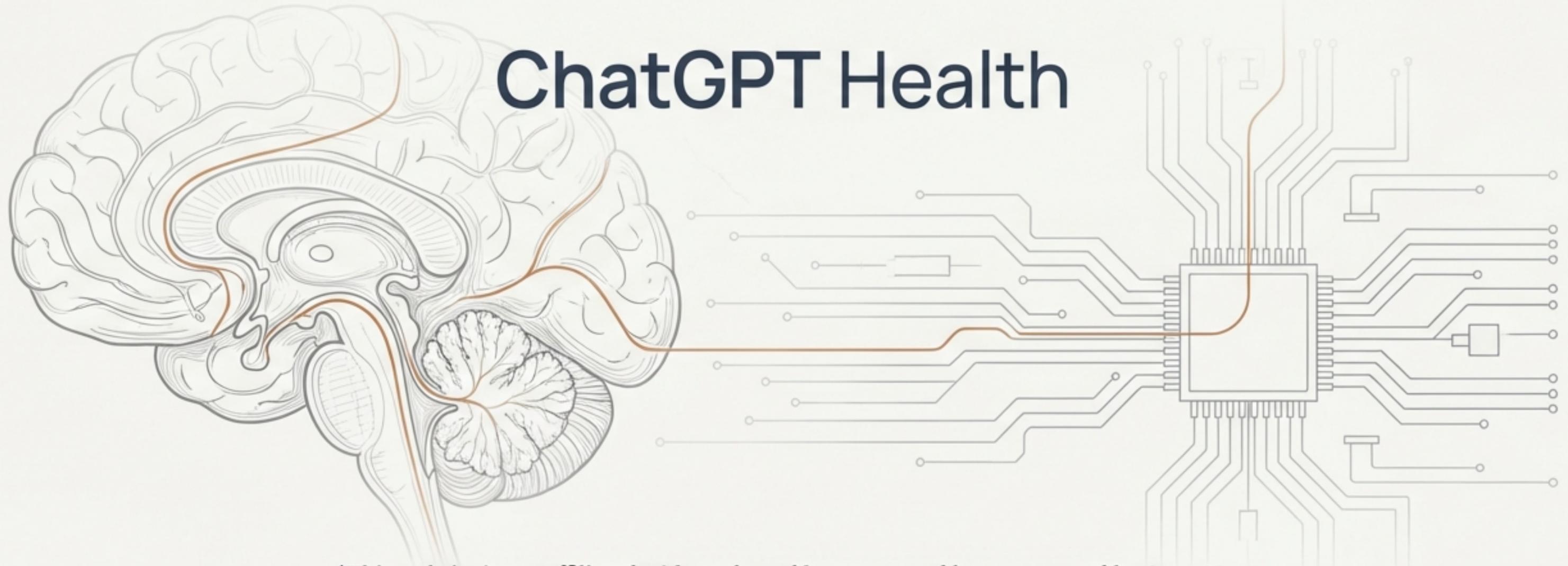# AI Assists. Humans Decide.

A physician-led educational platform exploring how AI systems like ChatGPT™ should be used responsibly, governed carefully, and overseen by humans in healthcare.
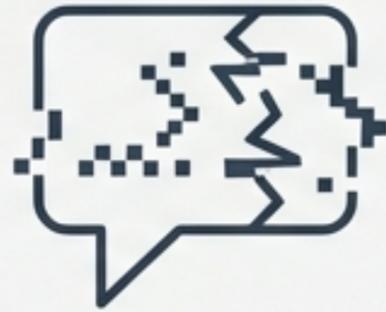
ChatGPT Health

The Risks of Unchecked AI Are Measured and Present and Present

38% → 70%

AI Adoption Growth

Among physicians from 2023 to 2024, highlighting the urgent need for governance.

1.47%

Hallucination Rate

Found in LLMs used for clinical note generation (2025 analysis), generating plausible but false information.

15%+

Error Rates

Documented in even state-of-the-art medical AI models on certain analytical tasks.

The Core Challenge: Moving Beyond Hype to Address Blind Automation

Healthcare does not need more hype. It needs guardrails. As Large Language Models (LLMs) enter clinical environments, the risk of "blind automation" grows. This leads to critical vulnerabilities:

Automation Bias

Over-reliance on AI outputs, leading to errors in clinical judgment. Clinicians may accept flawed AI recommendations and cease searching for confirmatory evidence.

Fluent Errors

A key characteristic of LLM hallucinations. The AI sounds confident and plausible while being factually wrong. Education on identifying these errors is critical for patient safety.

Our Mission: A Corrective Force for Human-Centered Care

ChatGPT Health exists to ensure that human judgment remains central to patient care.

We provide the educational resources to help clinicians, leaders, and policymakers question, validate, and govern these powerful tools responsibly. This platform serves as a guardrail, not a hype engine—a corrective force to ensure that clinicians are seen and protected.

A Human Accountability Layer
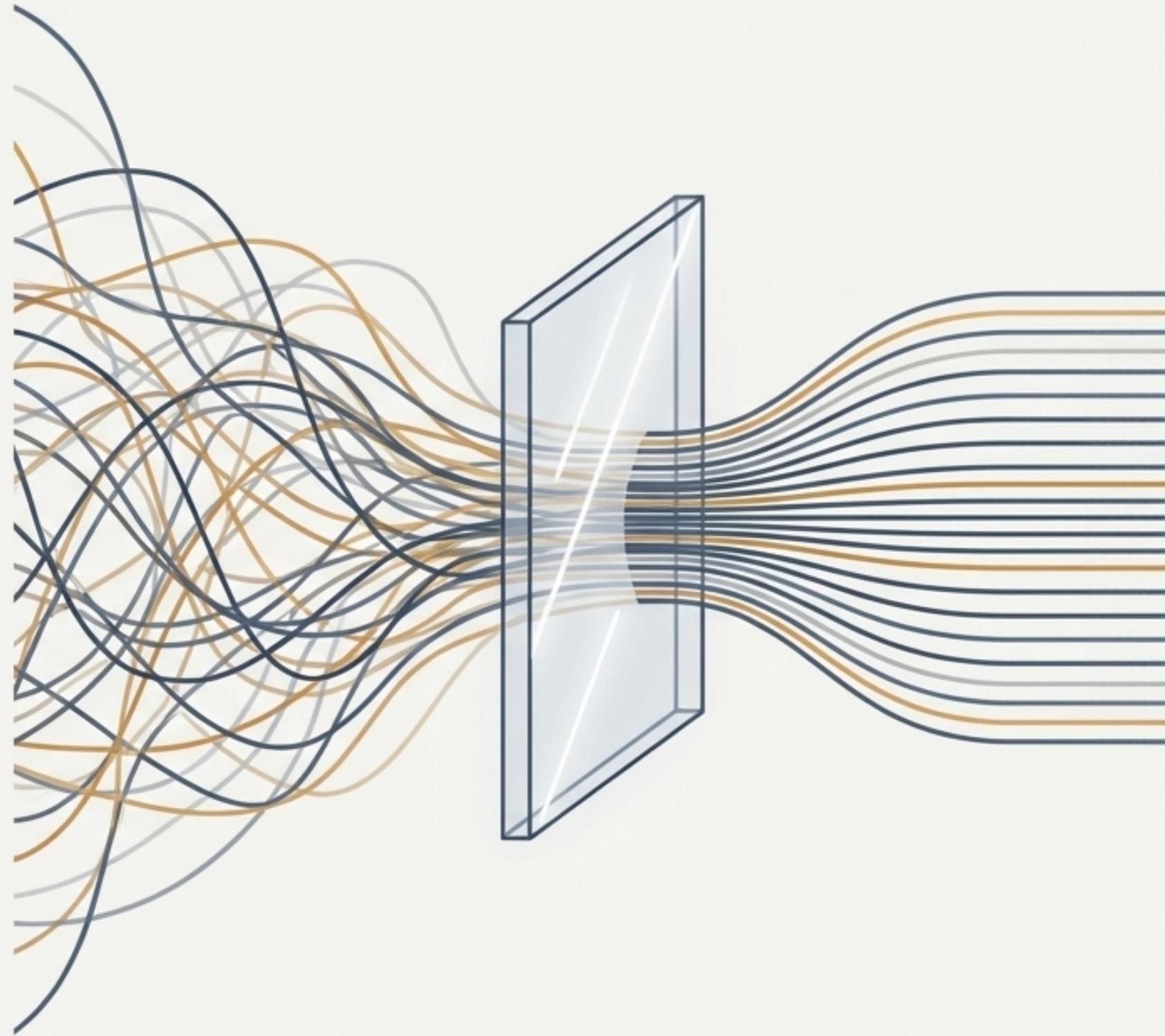
The Guardrails of Care: A Framework for Safe Innovation

Governance is not a barrier to innovation; it is the foundation of safety.
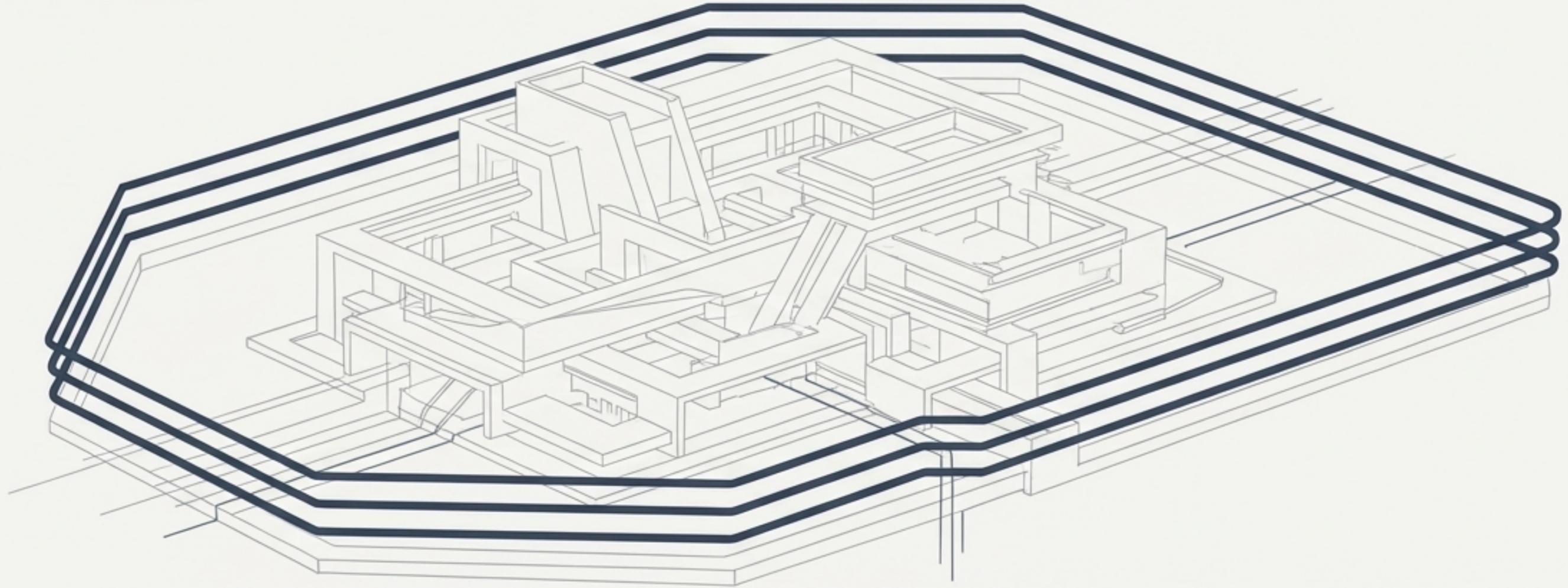
We provide educational content on the critical importance of human-in-the-loop design and robust accountability frameworks. Our focus is on establishing clear lines of responsibility when AI fails, ensuring that human oversight remains central to patient safety.

Our Four Guiding Principles for Responsible AI

Human-in-the-Loop

AI should never be the final decision-maker. We advocate for 'Human-Verified Output' where every AI suggestion is auditable.

Accountability Frameworks

Who is responsible when AI fails? Liability cannot be outsourced to an algorithm. Clear governance policies are essential.

Mitigating Bias & Hallucination

LLMs can sound confident while being factually wrong. Education on identifying 'fluent errors' and addressing data bias is critical.

Mandating Auditability

Systems must include frameworks for tracking and reviewing AI decisions to maintain transparency and trust.

A Comparative Guide to Leading AI Governance Frameworks

Framework	Core Focus	Key Recommendations
World Health Organization (WHO)	Ethics, human rights, and governance for AI in health.	Protect autonomy, ensure transparency, promote equity, and establish mandatory audits.
American Medical Association (AMA)	"Augmented intelligence" to support clinicians and improve patient care.	Implement risk-based oversight, establish clear liability for developers, and avoid mandatory AI use without clinical validation.
NIST AI Risk Management Framework	Voluntary, cross-sector framework for managing AI risks.	Adopt four-function core (Govern, Map, Measure, Manage) to create trustworthy and responsible AI systems.
EU AI Act	Risk classification with strict requirements for high-risk medical AI.	Prohibits high-risk AI without human oversight in medical contexts; mandates transparency and accountability.
HHS AI Strategy	Ethical directives for U.S. health departments.	Focus on patient safety, data security, and responsible deployment across federal health programs.

Identifying and Mitigating Key AI Risks in Clinical Settings

Risk Category	Description	Mitigation Strategy
Hallucination & Inaccuracy	AI generates plausible but false information. (Studies show rates of 1.47% in note generation and over 15% in some analytical tasks.)	Human-in-the-Loop (HITL) validation, robust testing, and chain-of-thought reasoning.
Automation Bias	Over-reliance on AI outputs, leading to clinicians accepting flawed recommendations.	Clinician training on AI limitations, accountability frameworks, and designs that encourage critical evaluation.
Data Bias & Health Equity	Models perpetuate disparities due to biased training data.	Diverse data sourcing, fairness audits, external validation across populations, and continuous monitoring.
Liability & Accountability	Lack of clarity regarding who is responsible for AI-related errors.	Clear governance policies defining liability for developers, institutions, and clinicians, as advocated by the AMA.

Building Clinical AI Literacy for Every Stakeholder

We argue that in the context of patient care, strong governance matters more than model size or capability. We offer targeted educational resources for:

For Clinicians

Capability vs. reliability, failure modes, automation bias, and prompt engineering for safety.

For Health System Leaders

Organizational readiness, deployment strategies, risk management, and system-level governance.

For Policymakers & Educators

Regulatory frameworks, accountability standards, patient protection, and educational curricula.

Inquire for Executive Briefings

The Vision: Agentic AI and Physician Digital Twins

We envision a future where the knowledge, clinical reasoning, and communication style of expert physicians can be replicated in AI agents. Such agents could provide autonomous consultations, offer real-time support to clinicians, and scale medical expertise globally.

This represents the ultimate expression of augmented intelligence, where AI acts as a **force multiplier for human expertise**, enhancing the capacity of the healthcare system.

This vision underscores our core principle: AI should augment, not replace, human judgment.
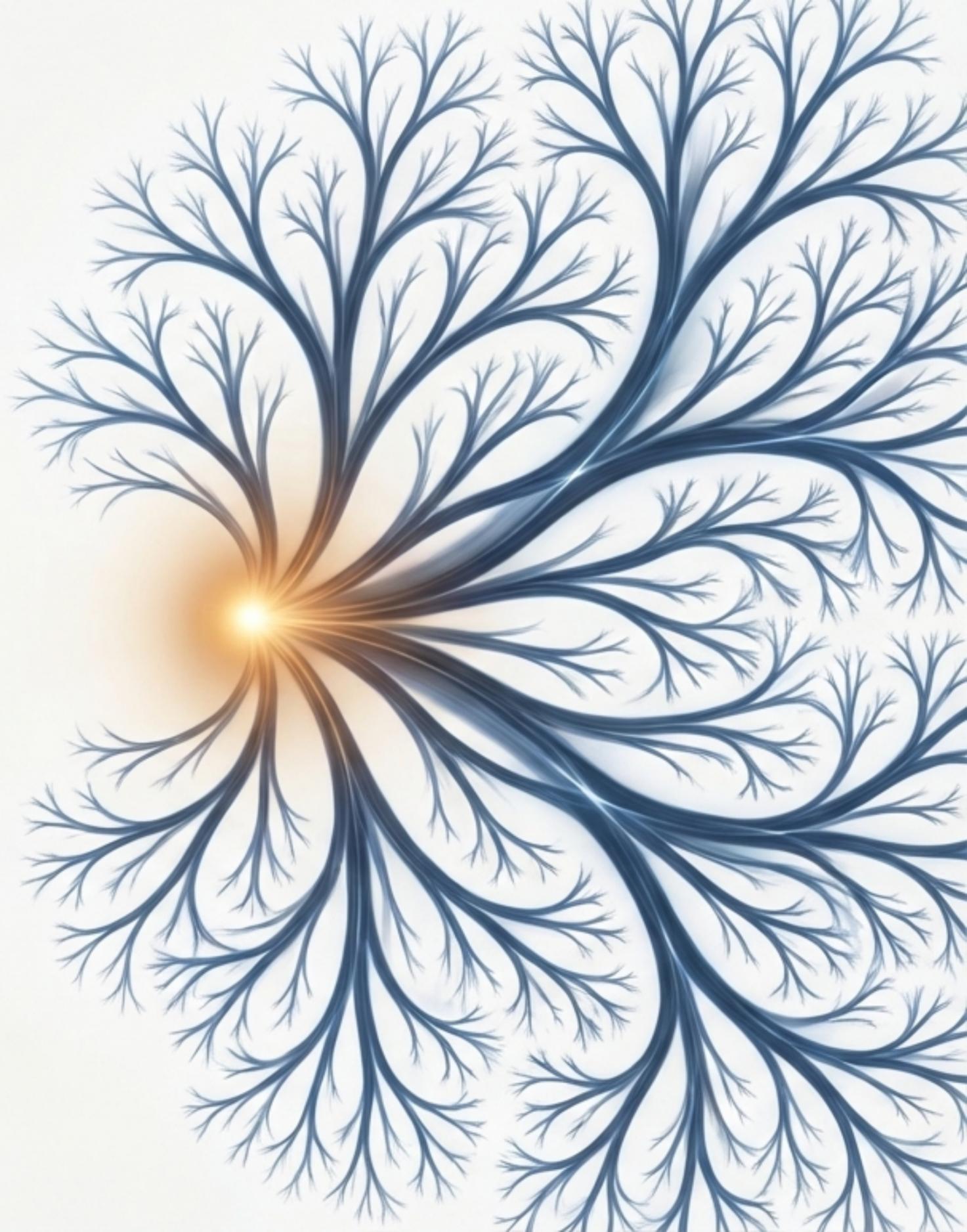
Simulating Systems, Not Patients

'Digital Twins' in this context refer to the simulation of hospital workflows, supply chains, and administrative processes—not individual patients.

We explore using AI to:

- Model ER wait times and staffing needs.
- Run 'virtual drills' to test system response to surge capacity.
- Design human accountability layers within these complex systems.
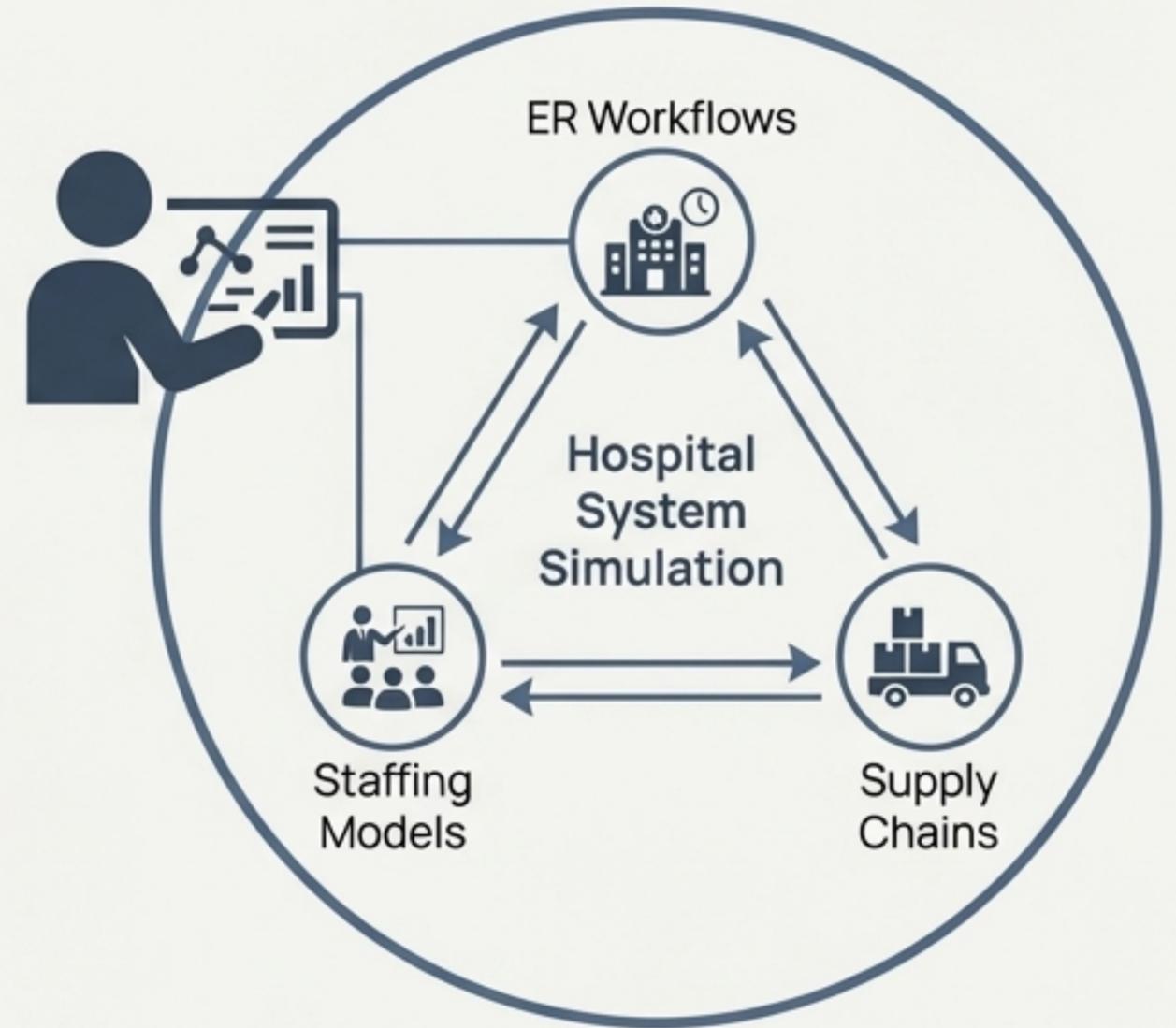
ER Workflows

Hospital System Simulation

Staffing Models

Supply Chains

Proven Impact and Leadership

21M — **Shaping Policy for 21M Citizens**
Serves on the Singapore Ministry of Health's Regulatory Advisory Panel (MOH RAP).

Advising 55,000+ Physicians
AI expert on the Texas Medical Association's Health Information Technology Committee.

40+ ERs Using AI Blueprints
Resulting in an 18% faster door-to-doctor time.

Distinguished Member, ASFAI
The most prestigious, invitation-only AI association in the United States.

Board Member
The American Journal of Healthcare Strategy.

5+ TEDx Talks
Global speaker on AI, healthcare, and human judgment.

Chief Medical AI Officer, Helpp.ai
Tackling the $34B fall-injury problem.

Early Thought Leadership in Responsible AI

Books & Publications

His books are foundational texts calling for strict human oversight and patient-centric design since early 2023.

TEDx & Media

Dr. Harvey Castro has delivered multiple TEDx talks on AI, healthcare, accountability, and the irreplaceable role of human judgment.

- SAFE-AI: Transforming Healthcare with AI's Ripple Effect
- How AI and humans can revolutionize medicine together
- Thriving Before the Emergency: The Future of Medicine with AI

Governance First. Education First.

This site exists to **support clinicians, policymakers,** and the public in **thinking critically about AI in healthcare**—not to replace human care or judgment. We ensure that AI systems are deployed responsibly, with **human accountability** at the center.

Executive Education:

Leadership programs for healthcare executives and decision-makers. Building safe AI strategies for health systems.

Speaking & Advisory:

Dr. Castro is available for keynotes, executive briefings, and strategic advisory work.

harveycastromd.com | in LinkedIn | X X (Twitter) | ▶ YouTube