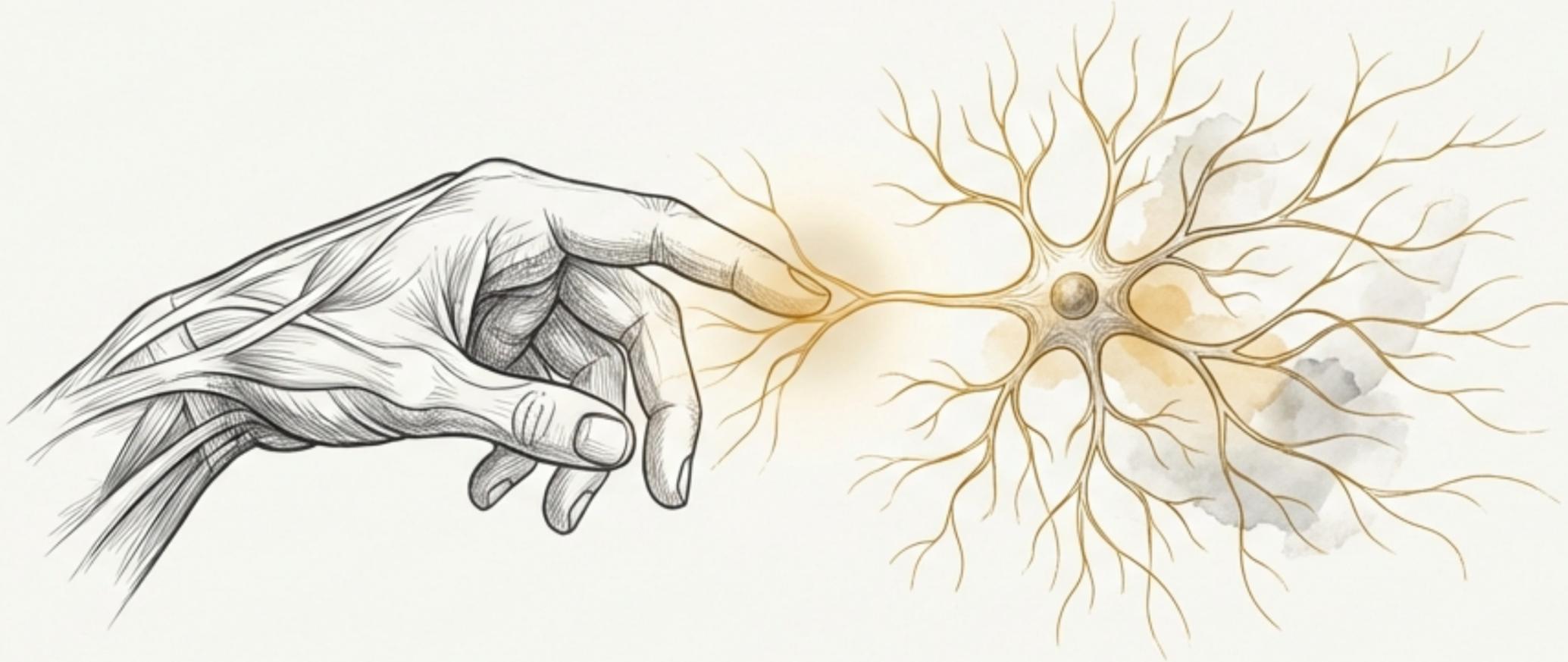# AI Assists. Humans Decide.

Governance and Education for the Responsible Use of AI in Medicine.

ChatGPT Health

A physician-led educational platform from Dr. Harvey Castro.
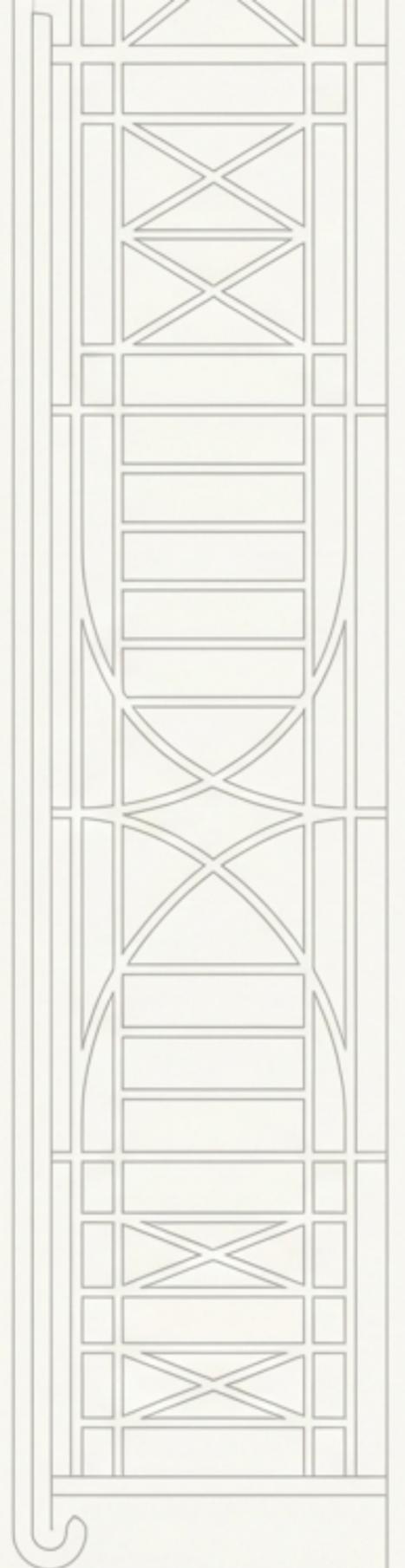
Healthcare Does Not Need More Hype. It Needs Guardrails.

As Large Language Models (LLMs) enter clinical environments, the risk of "blind automation" grows. ChatGPT Health exists to ensure that human judgment remains central to patient care.

We provide the educational resources to help you question, validate, and govern these tools responsibly.

This platform serves as a guardrail, not a hype engine—a corrective force to ensure that clinicians are seen and protected.
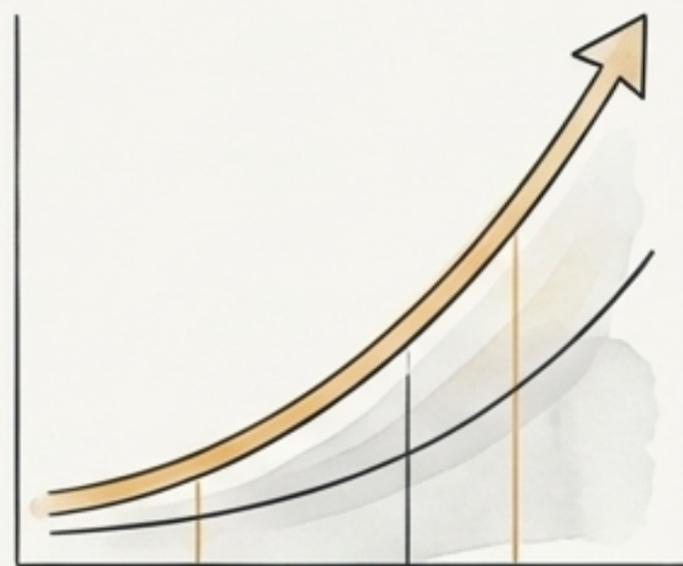
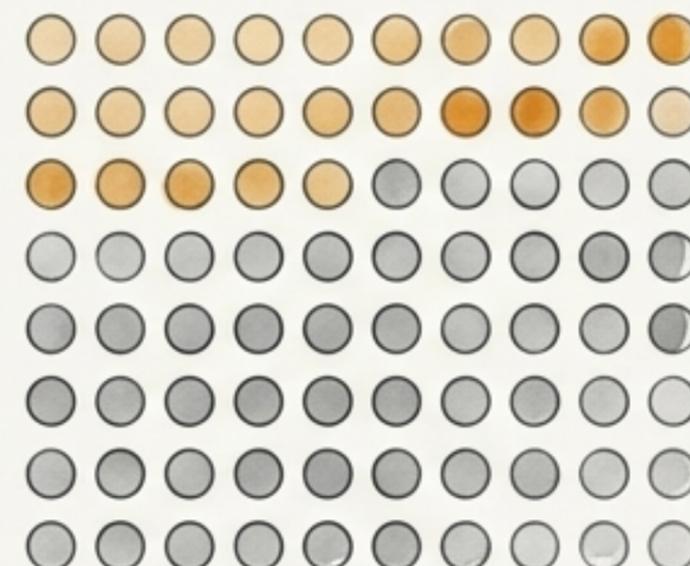# The Unchecked Advance: Quantifying the Risk

38% → 70%

Growth in AI adoption among physicians from 2023 to 2024, highlighting the urgent need for governance.

1.47%

INFORMATION

Measured hallucination rate in LLMs used for clinical note generation (2025 analysis). Defines hallucination as 'fluent errors' – plausible but false information.

15%+

Documented error rates for even state-of-the-art medical AI models on certain analytical tasks.

Rapid adoption without robust governance creates significant risks of over-trust and automation bias.

The Guardrails of Care: Our Guiding Principles

Governance is not a barrier to innovation; it is the foundation of safety.

Human-in-the-Loop

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

Accountability Frameworks

Who is responsible when AI fails? Liability cannot be outsourced to an algorithm. We promote clear lines of responsibility for developers, institutions, and clinicians.

Bias & Hallucination Mitigation

LLMs can sound confident while being factually wrong. Education on identifying "fluent errors" and addressing biased training data is critical to patient safety and health equity.

Mandatory Auditability

All clinical AI systems must have frameworks for tracking and reviewing decisions to maintain transparency, trust, and continuous improvement.

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, establish mandatory audits.
American Medical Association (AMA)	'Augmented intelligence' to support clinicians.	Implement risk-based oversight, establish clear developer liability, avoid mandatory AI use.
NIST AI RMF	Cross-sector framework for managing AI risks.	Adopt four-function core (Govern, Map, Measure, Manage) for trustworthy AI.
EU AI Act	Risk classification with strict rules for high-risk medical AI.	Prohibits high-risk AI without human oversight; 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 programs.

From Theory to Practice: Key Risks & Mitigation Strategies

The Risk

Hallucination & Inaccuracy
AI generates plausible but false information.

Automation Bias
Over-reliance on AI outputs, leading clinicians to accept flawed recommendations.

Data Bias & Health Equity
Models perpetuate disparities due to underrepresented training data.

Liability & Accountability
Lack of clarity on who is responsible for AI-related errors.

The Mitigation Strategy

Human-in-the-Loop (HITL) Validation
Robust testing protocols and chain-of-thought reasoning to enable self-verification by the AI, confirmed by a human expert.

Targeted Clinician Training
Education on AI limitations and system designs that encourage critical evaluation.

Fairness Audits & Diverse Data Sourcing
External validation across different populations and continuous monitoring for bias.

Clear Governance Policies
Define liability for developers, institutions, and clinicians, as advocated by the AMA.

Clinical AI Literacy: Understanding the Tool, Not Just the Hype

In the context of patient care, strong governance matters more than model size or capability.

For Clinicians

Focus on practical skills. Topics include: Capability vs. reliability, identifying failure modes, countering automation bias, and prompt engineering for safety.

For Health System Leaders

Focus on organizational strategy. Topics include: Organizational readiness, deployment strategies, risk management, and system-level governance.

For Policymakers & Educators

Focus on systemic oversight. Topics include: Regulatory frameworks, accountability standards, patient protection, and developing educational curricula.
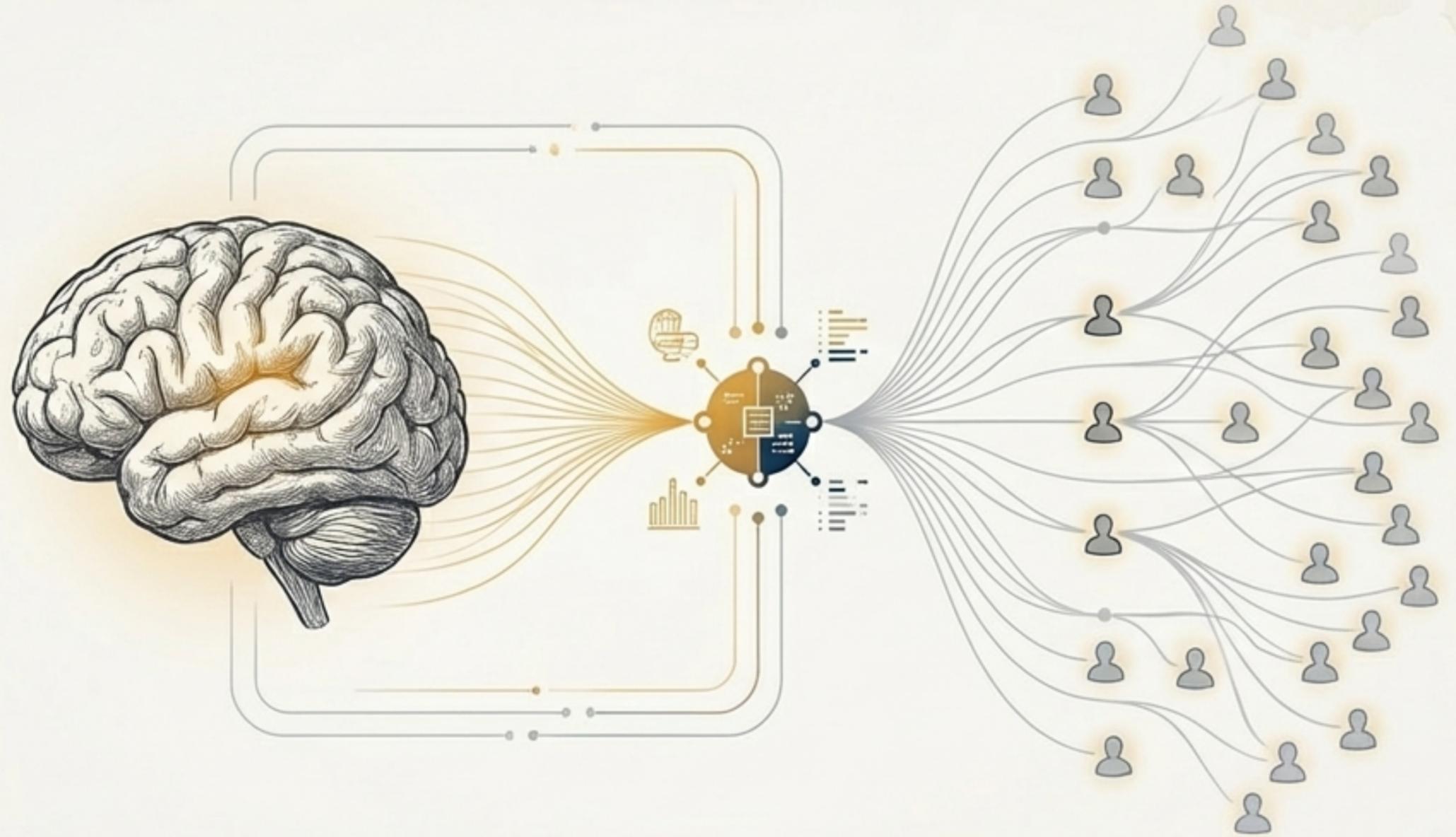
The Vision: Agentic AI & Physician Digital Twins

Core Concept:

A future where the knowledge, clinical reasoning, and communication style of expert physicians can be replicated in AI agents to scale medical expertise.

Key Functions:

- Provide autonomous consultations.
- Offer real-time support to other clinicians.
- Act as a force multiplier for human expertise.

This vision underscores our core principle: AI should augment, not replace, human judgment. It is about making care more accessible while maintaining the irreplaceable role of physician expertise and accountability.
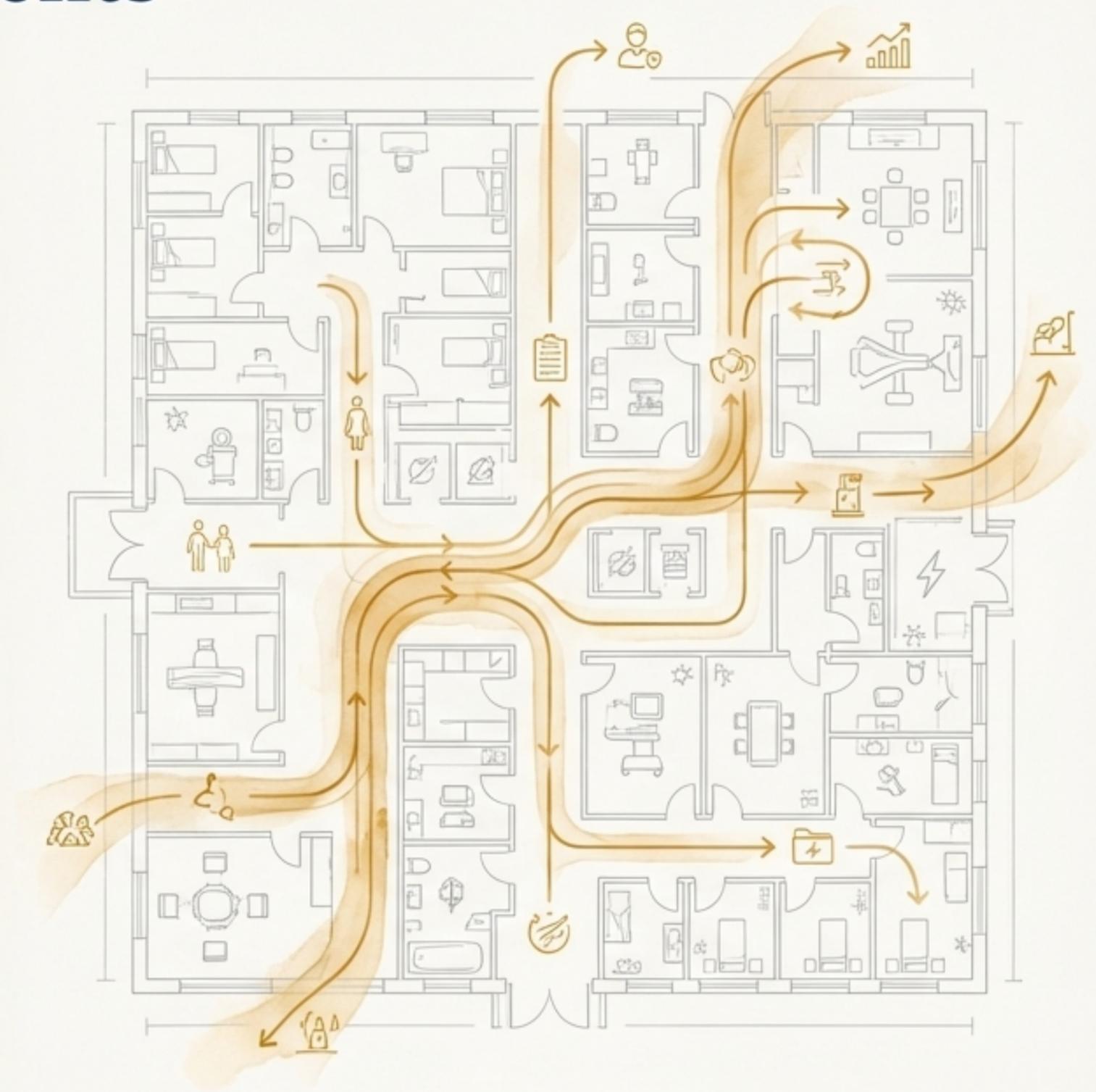
Simulating Systems, Not Patients

We also explore conceptual ideas like digital twins of clinical workflows and systems. This refers to the simulation of hospital workflows, supply chains, and administrative processes—not individual patients.

Practical Examples:

- Modeling ER wait times and staffing needs.
- Running "virtual drills" to test system response to surge capacity.
- Designing human accountability layers within these complex digital systems.

Courses & Executive Education

Executive Education

Leadership programs for healthcare executives and decision-makers.

- AI governance frameworks for healthcare organizations
- Clinical leadership and responsible deployment strategies
- Organizational readiness assessment and planning
- Building safe AI strategies for health systems

Inquire for Executive Briefings

Clinical Courses

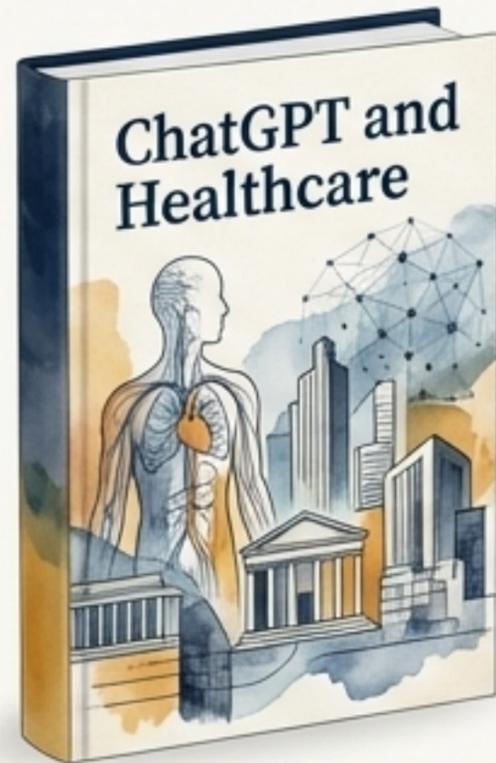Comprehensive modules for healthcare professionals.

In development: Modules on prompt engineering for safety, detecting hallucinations, and ethical AI integration.
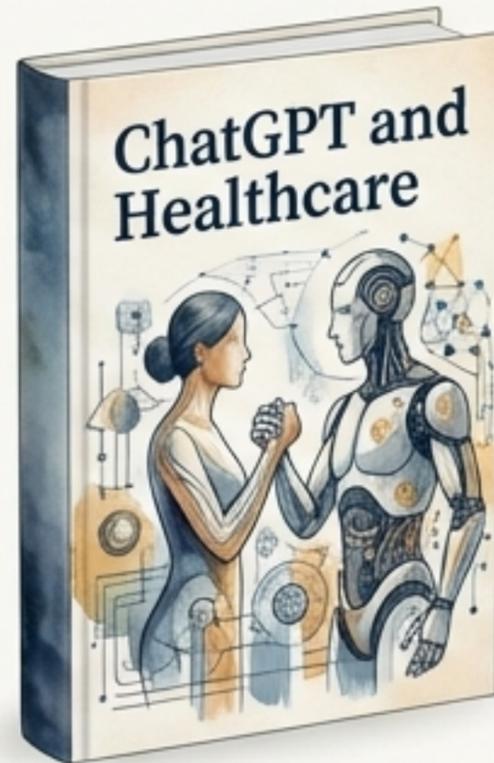
No consumer or patient-facing medical courses offered.

Driving the Conversation: Books & Global Speaking

Early Thought Leadership in Responsible AI

1st Edition

2nd Edition

Foundational texts on the capabilities, risks, and ethical considerations of using LLMs in medicine, calling for strict human oversight and patient-centric design.

View All Books

Keynotes & Advisory

TEDx **CES**

Signature Topics

Governance First
The Human-AI Partnership
The Future of the ER

Credibility Metrics

5+ TEDx Talks
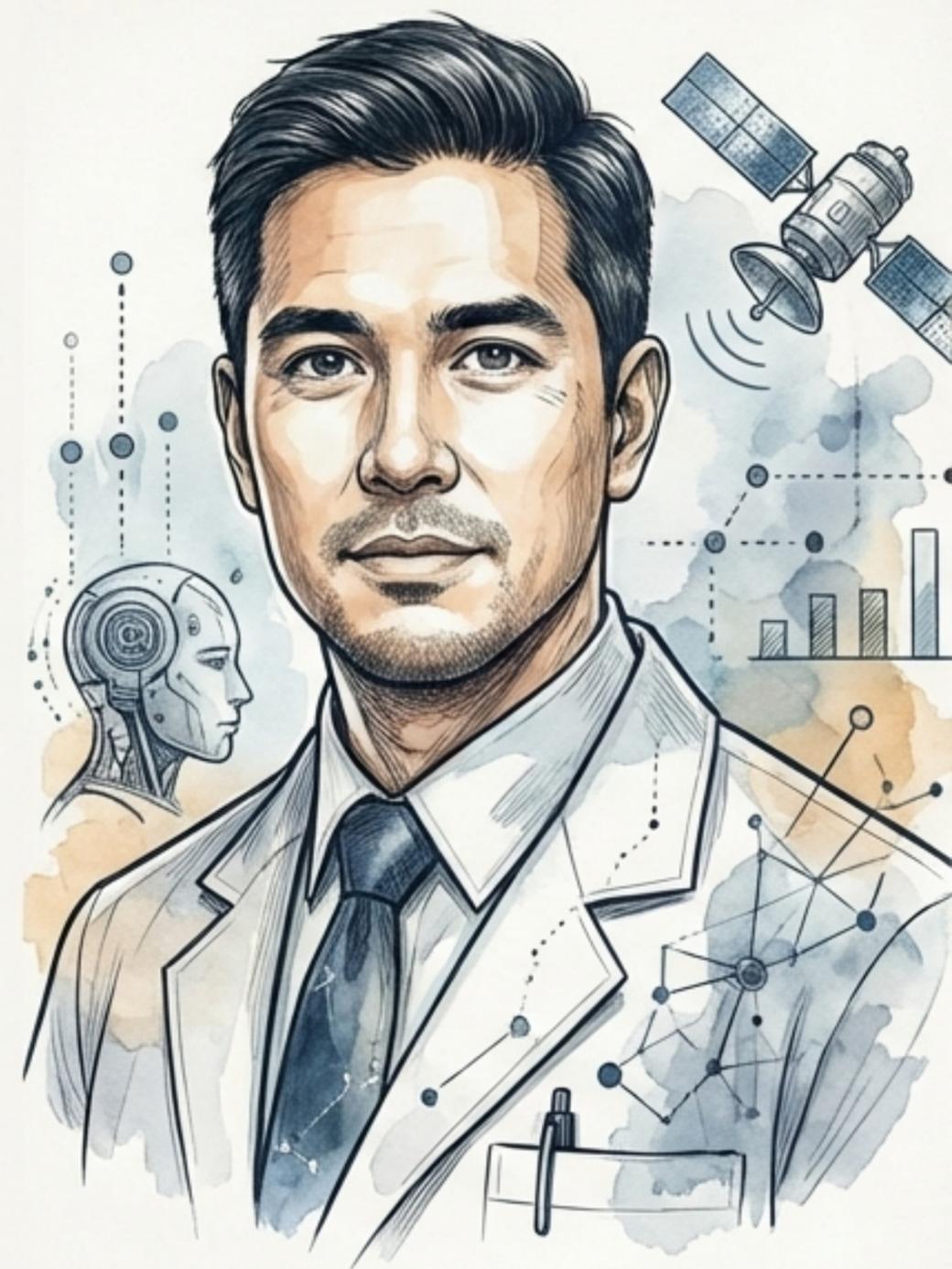
100+ Speaking Engagements

20+ Years Experience

Book Dr. Castro

The Physician-Led Foundation: About Dr. Harvey Castro

Dr. Harvey Castro, MD, MBA (DrGPT) | ER Physician | Chief AI Officer, Phantom Space | TEDx Speaker

"At 2 a.m. in the ER, I rely on seconds; soon I'll depend on satellites. I'm fusing AI, clinical data, and orbital sensors sensors to predict illness before symptoms strike."

- **Policy:** Shaping healthcare policy for 21 million citizens on the Singapore Ministry of Health's Regulatory Advisory Panel.
- **Influence:** AI expert for the Texas Medical Association (55,000+ physicians).
- **Prestige:** Distinguished Member of the American Society for AI (ASFAI).
- **Operational Success:** Blueprints used in 40+ emergency departments, resulting in an 18% faster door-to-doctor time.
- **Innovation:** Chief Medical AI Officer at Helpp.ai, tackling the $34B fall-injury problem.
- **Community:** OpenAI Forum Member, directly engaging with the community shaping ChatGPT's future.

ChatGPT Health: 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 human judgment. We provide education, governance frameworks, and ethical analysis to ensure that AI systems are deployed responsibly, with human accountability at the center.

Website: harveycastromd.com

in   X   ▶