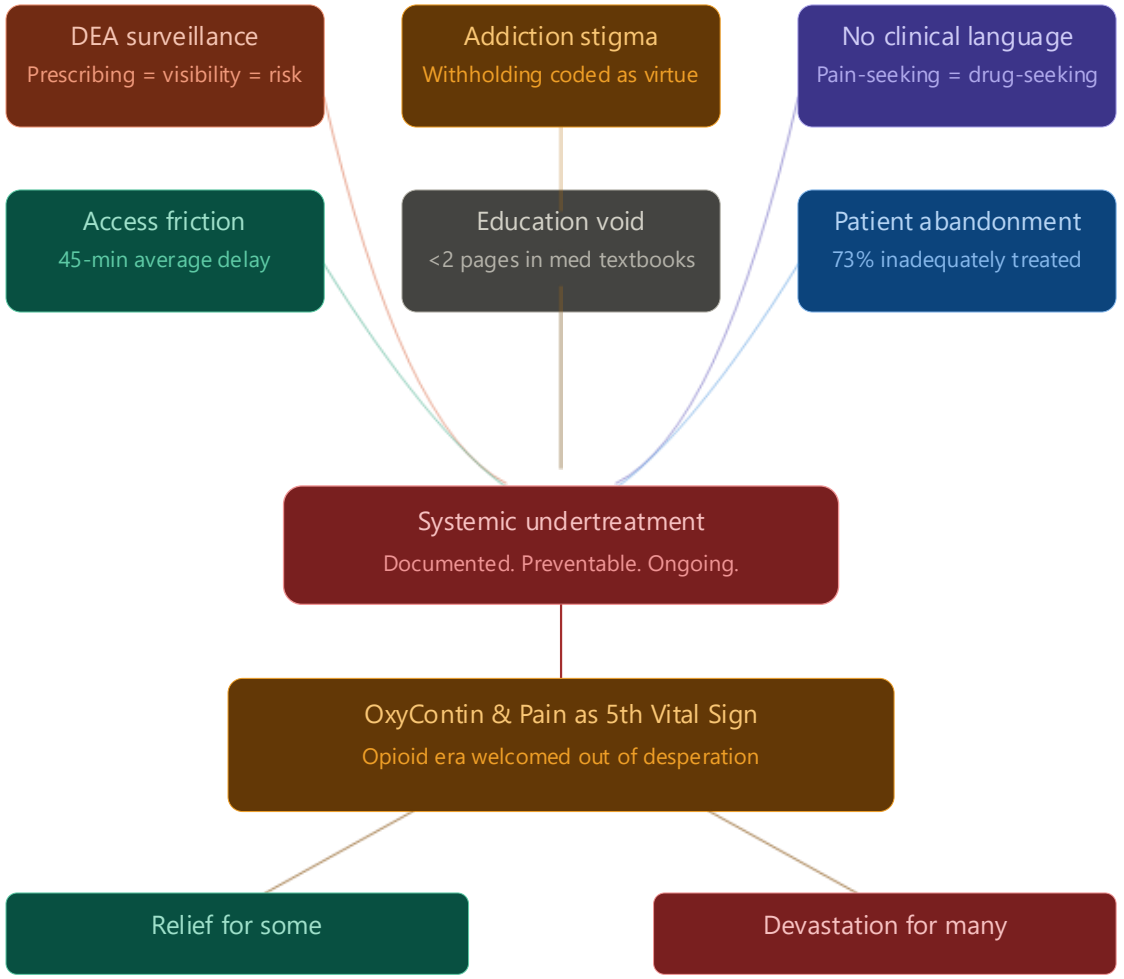
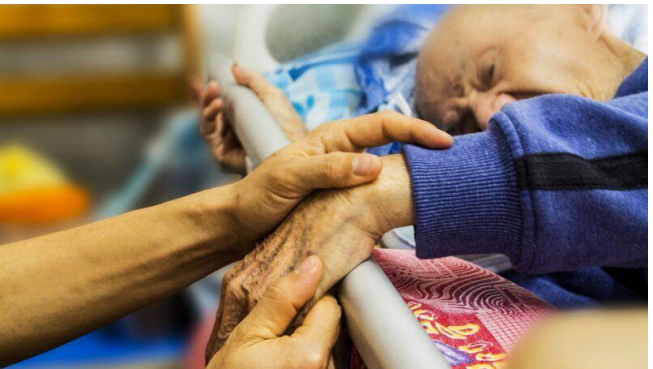
There's Always a Tradeoff :
The Risks Hiding Behind the Hope that AI will Save Healthcare

Josh Russell, MD, MSc, FCUCM, FACEP

Chief Medical Officer, UCP-Merchant Medicine/Intellivisit

1980-1995: From Pain Crisis to Opioid Crisis





"Remember your destination. This will help you to distinguish between an opportunity to be seized and a temptation to be resisted."

The Two Journeys (Bamidbar, Life-Changing Ideas, Covenant & Conversation)

What is our Destination?

-Rabbi Lord Jonathan Sacks

Disclosures

Chief Medical Officer – UCP/Merchant Medicine/Intellivisit

- No financial incentives/COI

AI Integration:
The PAR-Q Framework

1. Procedure
2. Alternatives
3. Risks
4. Questions

AI TOOLS

FOR CLINICAL PRACTICE

A curated set of AI assistants relevant to clinical work — from diagnosis support to medical literature synthesis.

Use with clinical judgment. AI augments, not replaces, physician expertise.

C

Claude

Anthropic

✓

GP

ChatGPT

OpenAI

✓

G

Gemini

Google

✓

P

Perplexity

Perplexity AI

✓

DS

DeepSeek

DeepSeek AI

✓

G

Glass Health

Glass

✓

OE

OpenEvidence

OpenEvidence

✓

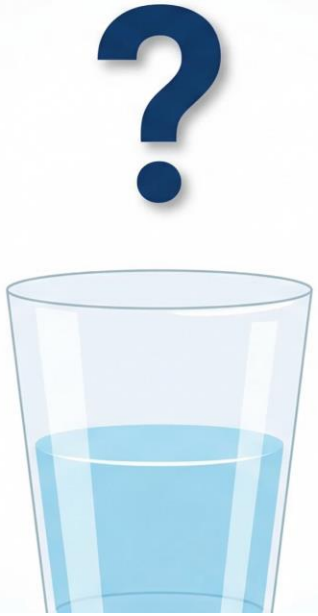
M

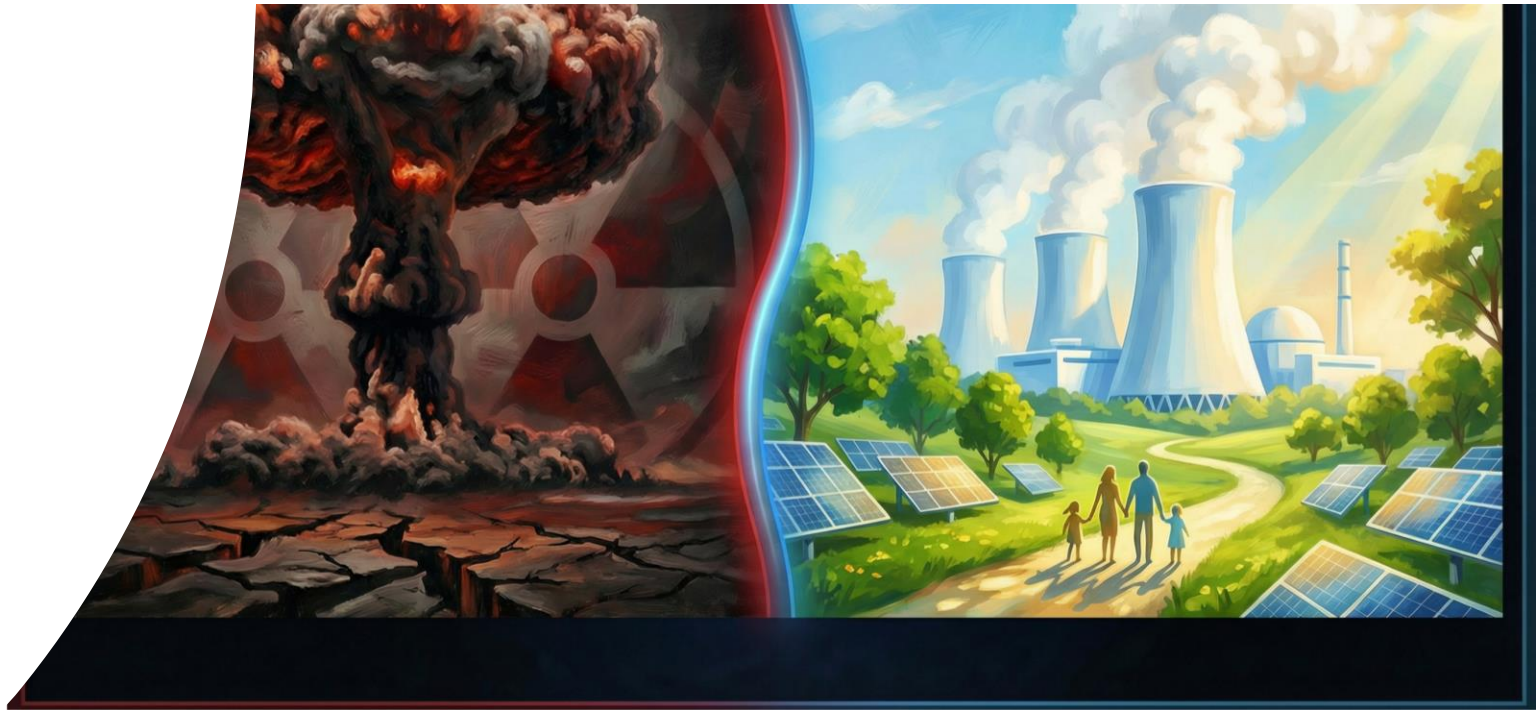
Manus

Manus AI

✓

Risks? : A matter of perspective





AI Risk(s)?

Broad Classes of Artificial Intelligence

Machine Learning

Large Language Models (LLMs)

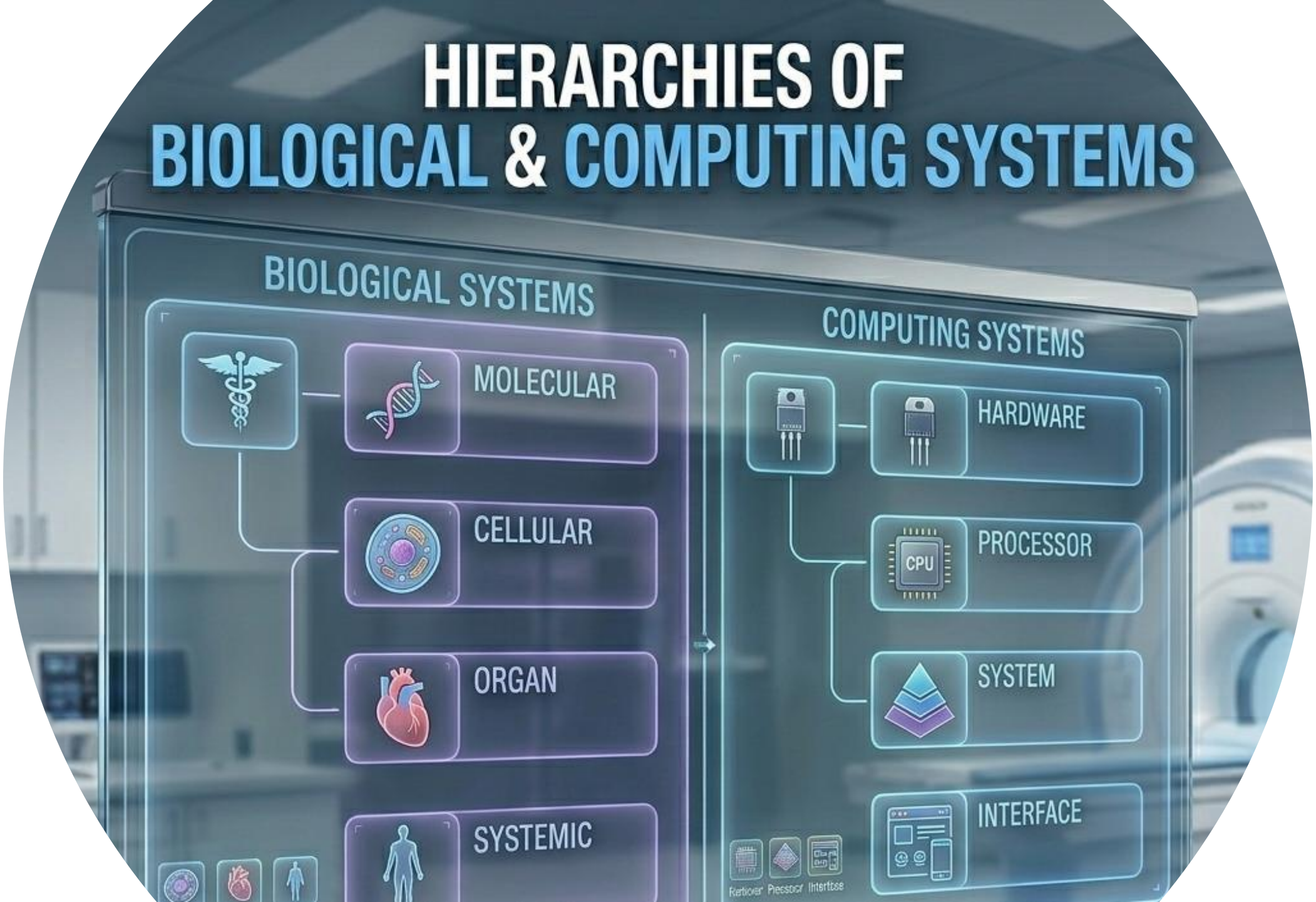
Computer Vision

Natural Language Processing

Robotics & Automation

Expert Systems

From perception to reasoning — AI spans diverse domains.


HIERARCHIES OF
BIOLOGICAL & COMPUTING SYSTEMS
BIOLOGICAL SYSTEMS
MOLECULAR
CELLULAR
ORGAN
SYSTEMIC
COMPUTING SYSTEMS
HARDWARE
CPU
PROCESSOR
SYSTEM
INTERFACE

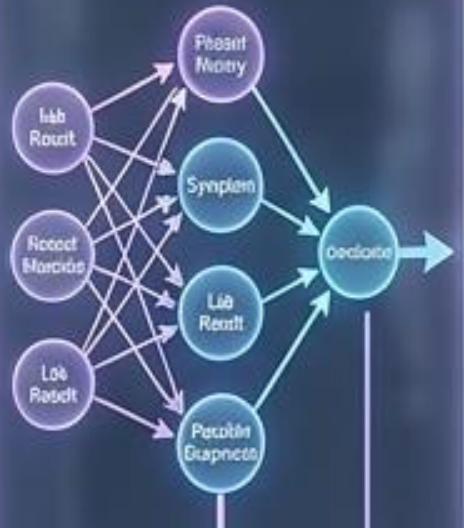
CLASSES OF LLM TASKS

1.) CATEGORIZATION

Sorting data and cases.

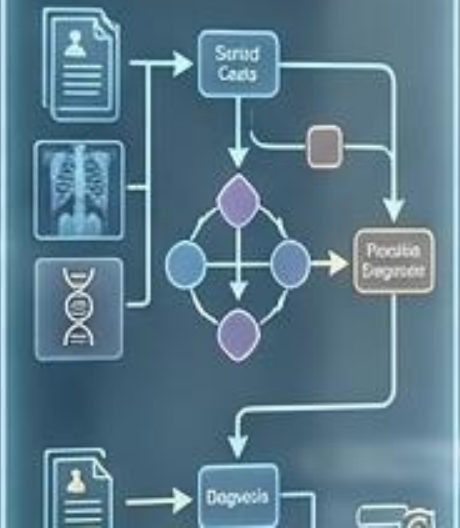
2.) REASONING

Logical multi-step deduction.



3.) MIXED

Combines sorting and logic.



4.) GENERATION

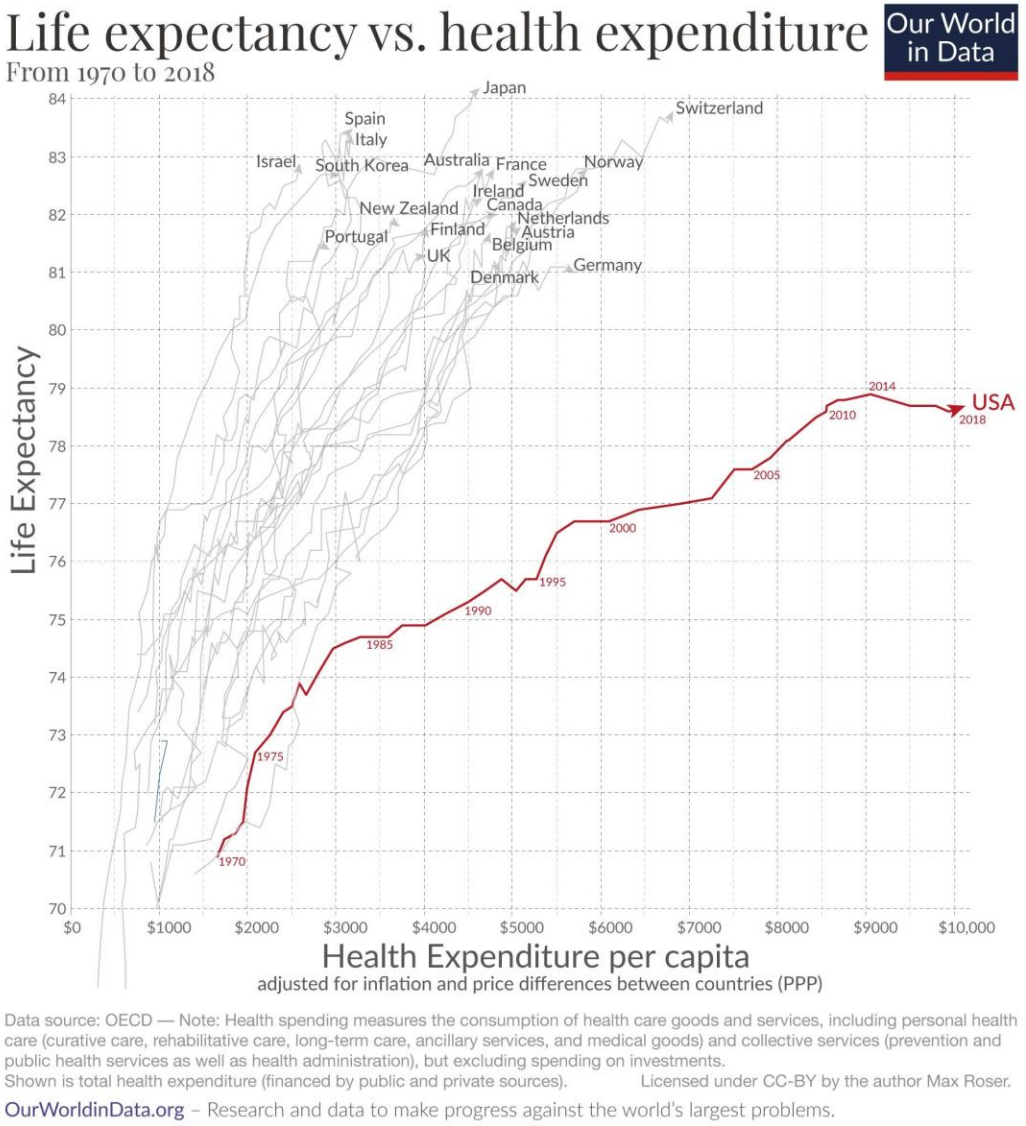
Creating new reports and notes.

Alternatives: What's the Risk of the Status Quo?

Status Quo

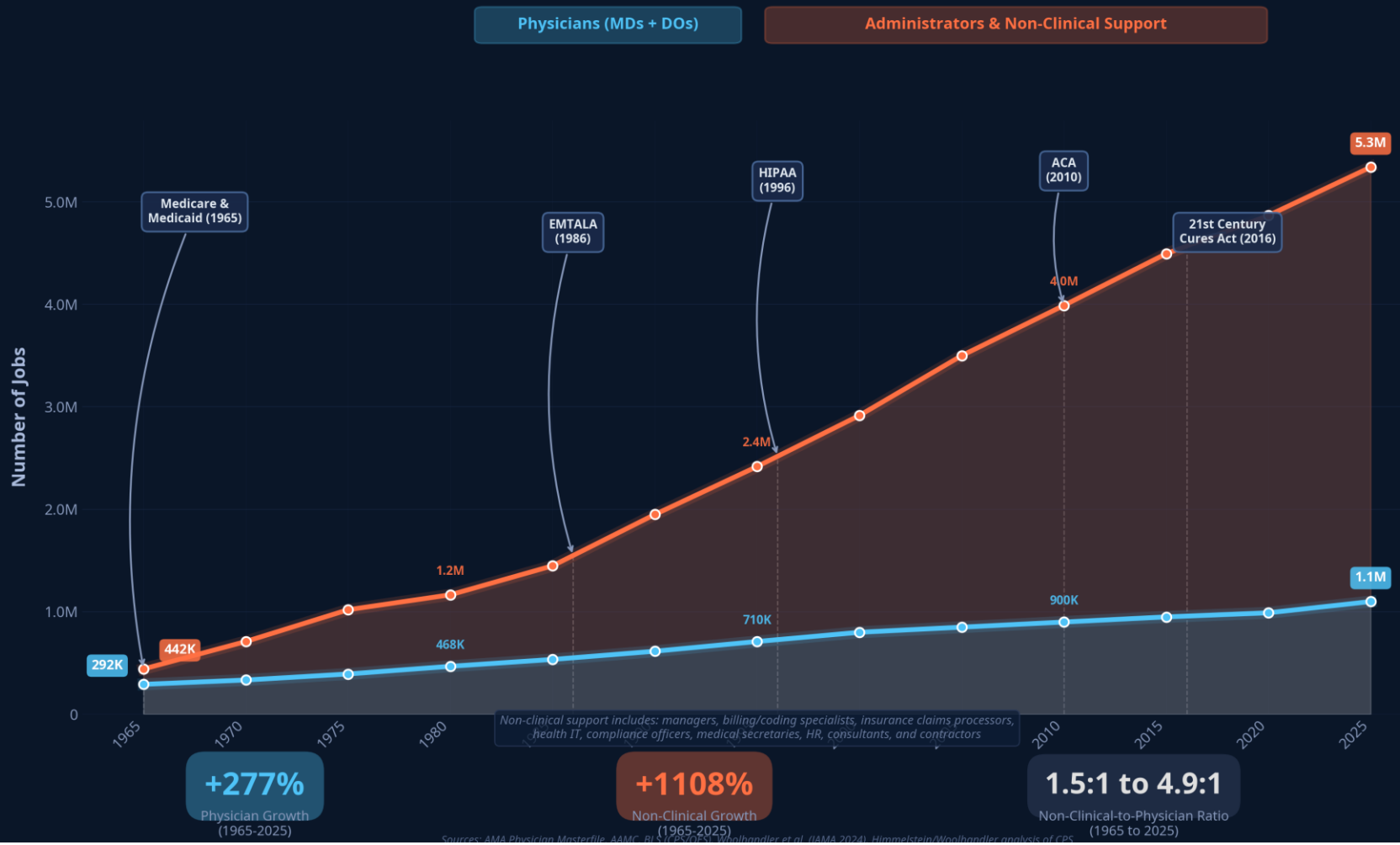
Total U.S. Healthcare Spending:

- 2015 - $3.2 Trillion (18% GDP)
- 2025 - $5.6 Trillion (20% GDP)



U.S. Healthcare Workforce

Physicians vs. Administrators & Non-Clinical Support (1965-2025)



Sources: AMA Physician Masterfile, AAMC, BLS (CPS/OES), Woolhandler et al. (JAMA 2024), Himmelstein/Woolhandler analysis of CPS

FIGURE 3. Administrative and Direct Patient Care Expenditures at U.S. Hospitals, 2011–2023

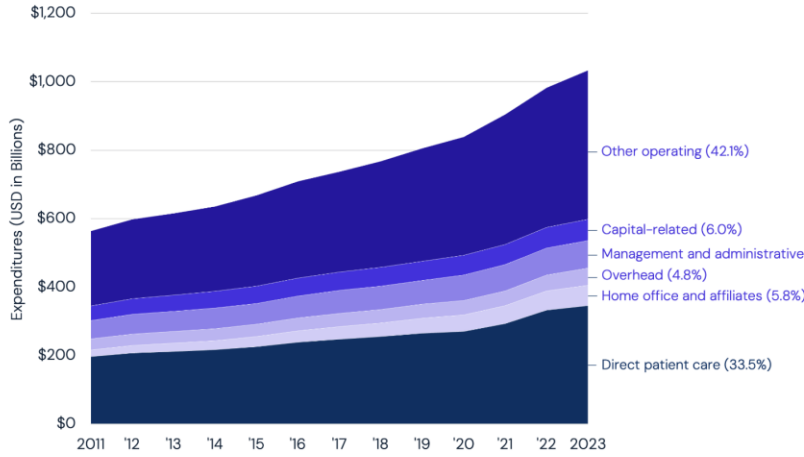
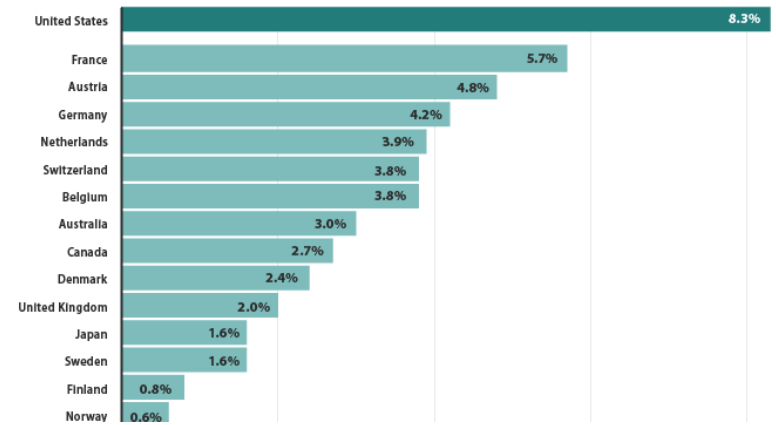


FIGURE 1

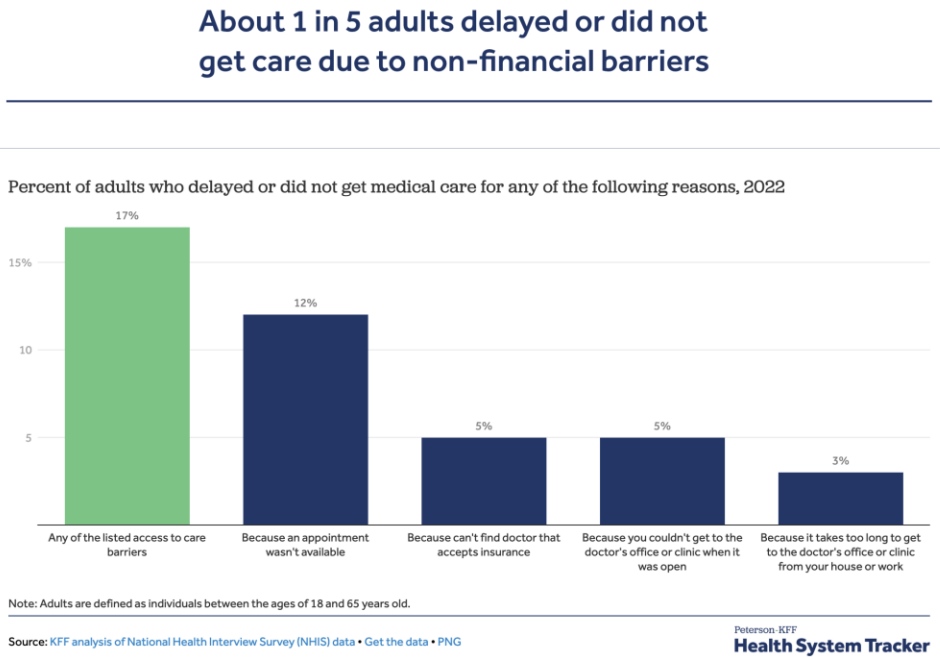
Administrative costs comprise a larger share of health care spending in the United States than in other high-income countries

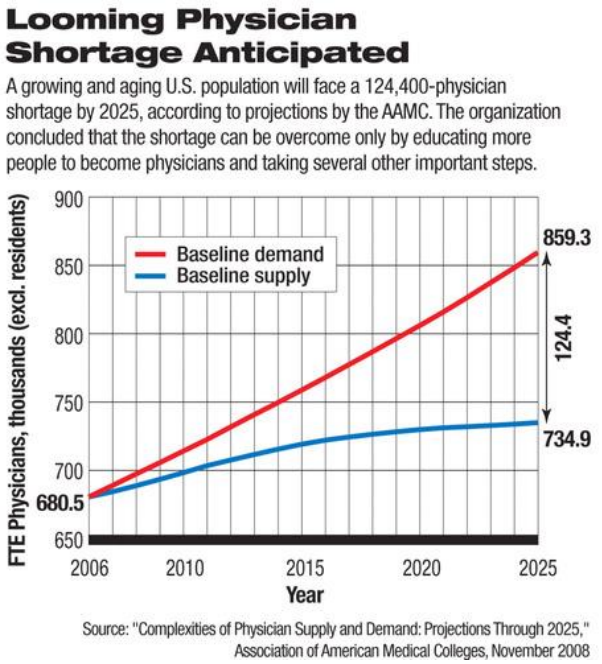
Administrative spending as a percentage of total health expenditures, 2016*



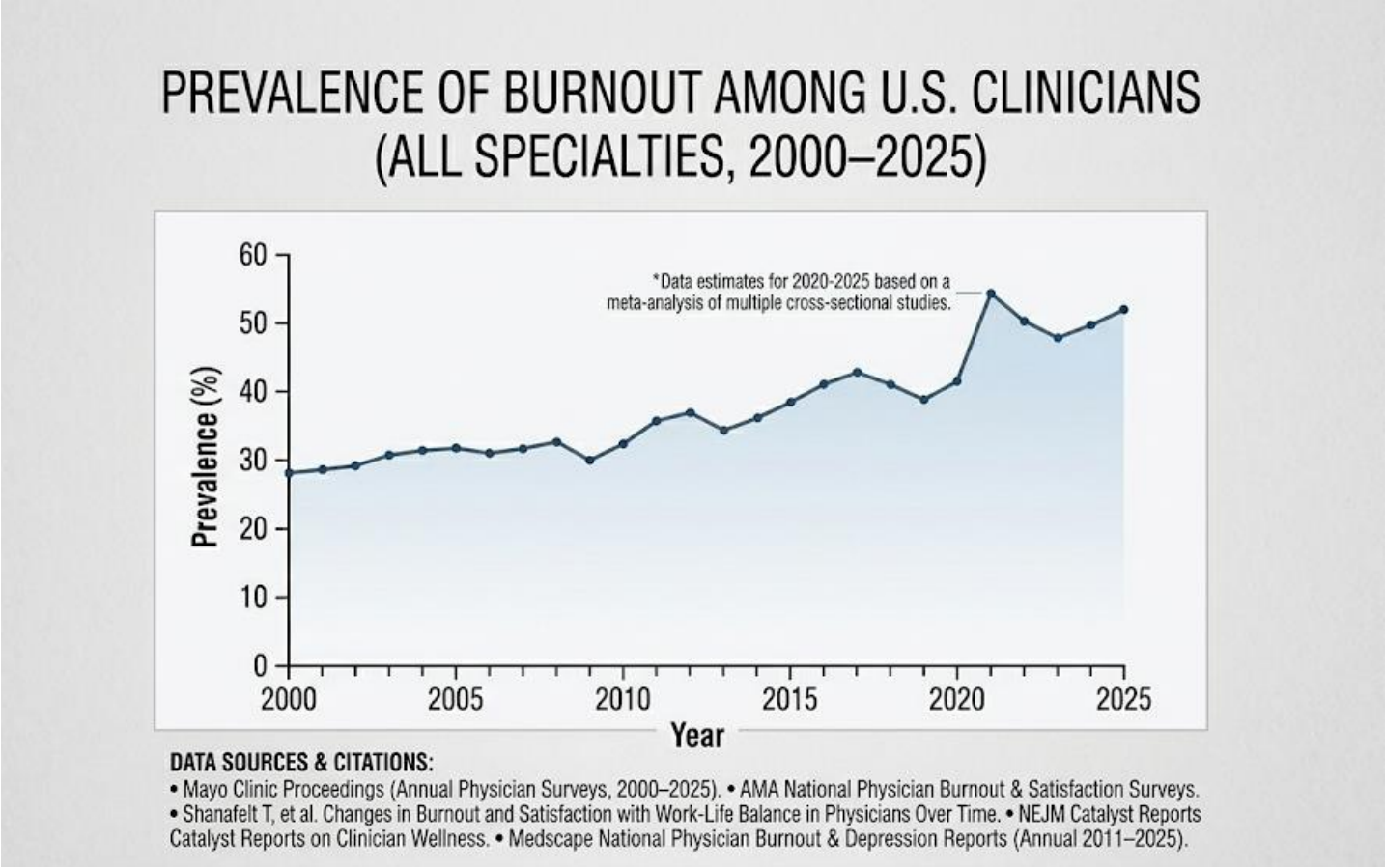
Accessing Care:

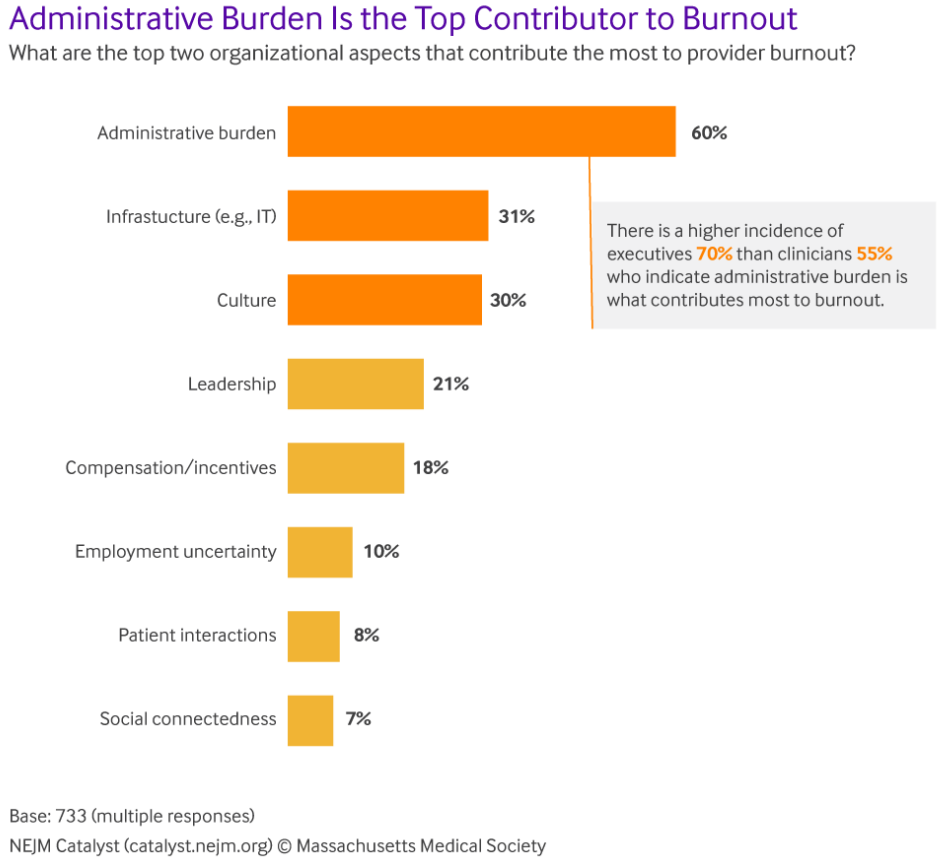
What Options Exist if Scaling the Human Workforce Can't Keep up?



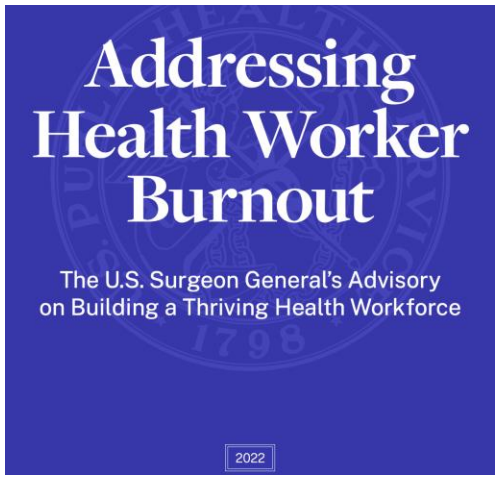


Existing Workforce at Risk



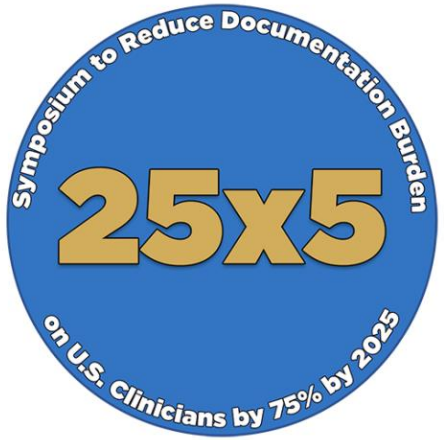


Addressing EHR Burden for Clinicians



Reduce *administrative burdens* contributing to health worker burnout.

- **Partner with health care delivery organizations, professional associations, and other stakeholders to reduce documentation burden by 75% by 2025.**[166, 210]



SUMMARY REPORT

December 2021

25 by 5 Initiative to Reduce Documentation Burden on U.S. Clinicians by 75% by 2025

Summary Report

Abstract

Documentation burden is the stress imposed by the excessive work required to generate clinical records of healthcare-related interactions, occurring as a result of the imbalance between the usability and satisfaction of systems of health records keeping with clinical and regulatory demands of entering and consuming health records data. Examples include high time requirements for creating clinical documents, confusing user interfaces, and externally-imposed requirements for specific content that do not directly align with the needs of immediate clinical care activities. Documentation burden existed long before the implementation of electronic health records (EHR), with the impacts gaining increasing attention over the last few years as EHR use has increased. More recently, documentation burden has been linked to reduced workplace wellness, increased rates of clinician burnout, and a reduction in clinical care quality. The National Library of Medicine (NLM) funded the *25x5: Symposium to Reduce Documentation Burden on US Clinicians by 75% by 2025* with the overarching goal of developing a unified national action plan focused on short, medium, and long-term approaches to achieving this goal.(1) The Symposium activities were informed by one key theme: clinician documentation is for patient care delivery and clinician-patient communication.

Funded by the National Library of Medicine (1R13LM013581-01)

AMIA INFORMATICS PROFESSIONALS. LEADING THE WAY. NIH U.S. National Library of Medicine VANDERBILT UNIVERSITY MEDICAL CENTER COLUMBIA COLUMBIA UNIVERSITY DEPARTMENT OF BIOMEDICAL INFORMATICS

EHR Documentation & Administrative Burden for U.S. Clinicians (2020–2025)

(1) 2020–2021

Rapid increase in after-hours EHR time and inbox volume and during COVID-19 surges

More than half of physicians report documentation as a top burnout driver

(2) 

2022

Burden remains high despite stabilization of visit volumes

National reports highlight persistent EHR clicks, inbox load, and prior auth work

(3) 

2023

Early adoption of scribes, better EHR configuration, and ambient/AI tools

Incremental reductions in documentation time reported in some systems

(4) 

2024–2025

Burden still rated as a major contributor to burnout

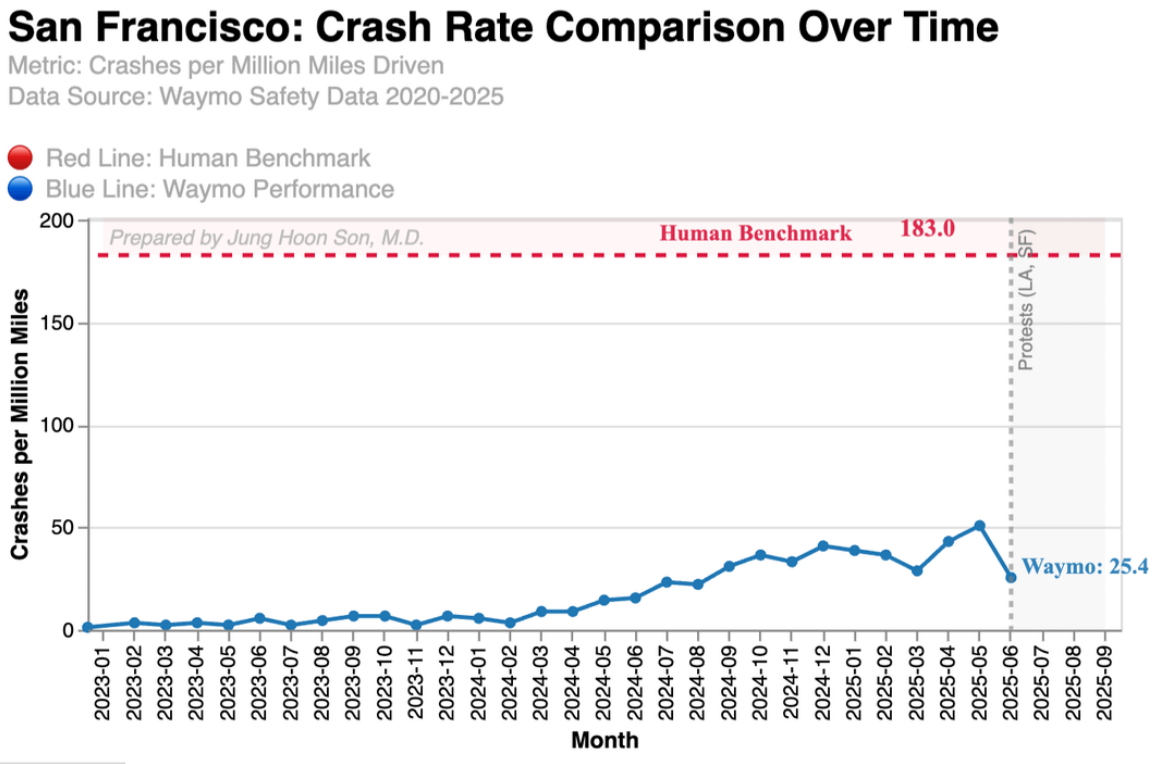
Most clinicians report only small improvements; admin work remains above pre-COVID levels

- Documentation and inbox work remain a top burnout driver
- Some systems show 5–15% reductions in EHR time with workflow redesign and AI tools
- No evidence of large national reduction yet

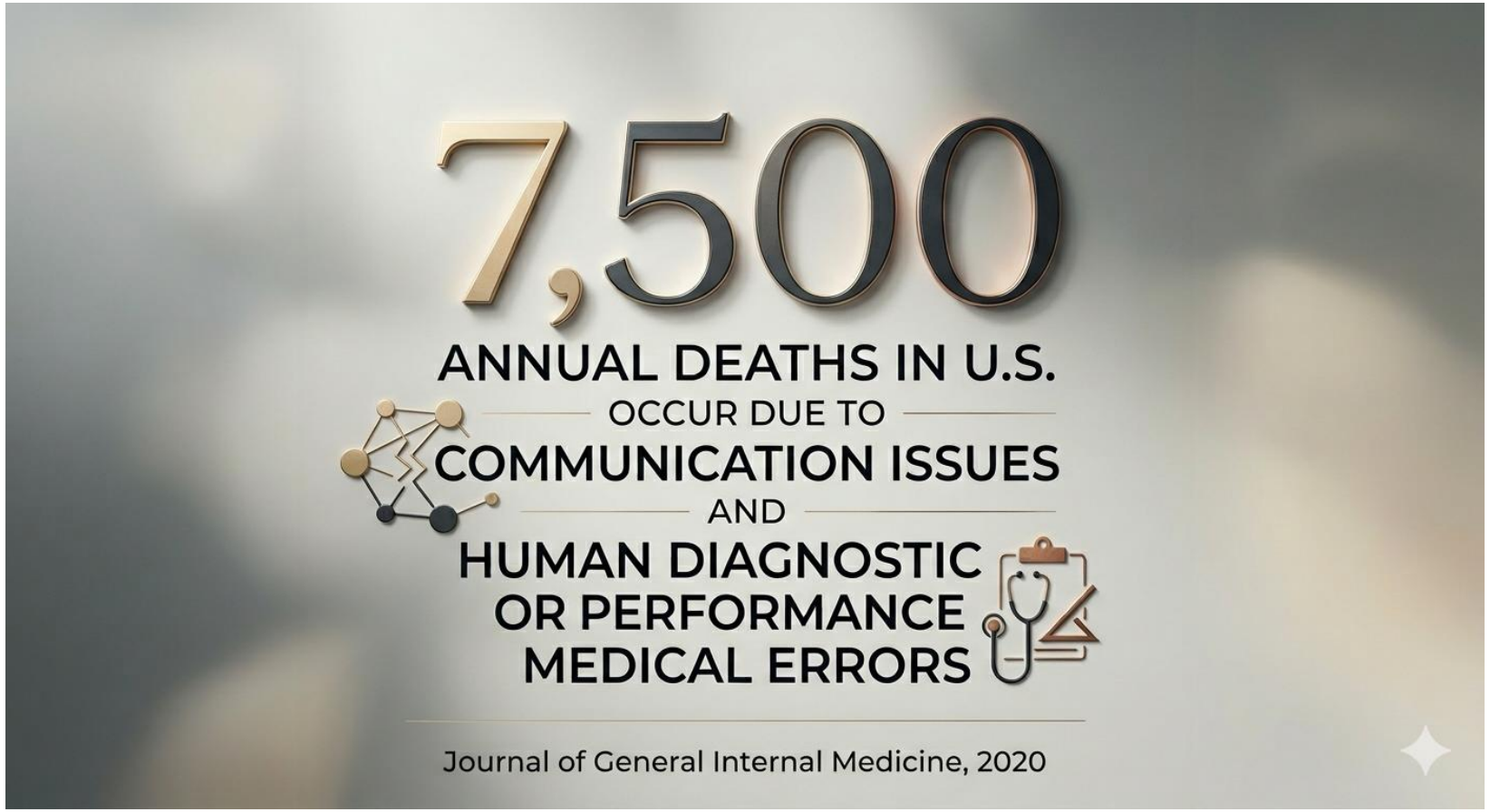
Sources: AMA, JAMA Network Open, NASEM, CDC NIOSH, 2020–2025.

The Status Quo Bias is *Strong* in Medicine

Algorithm Aversion



Human Factors

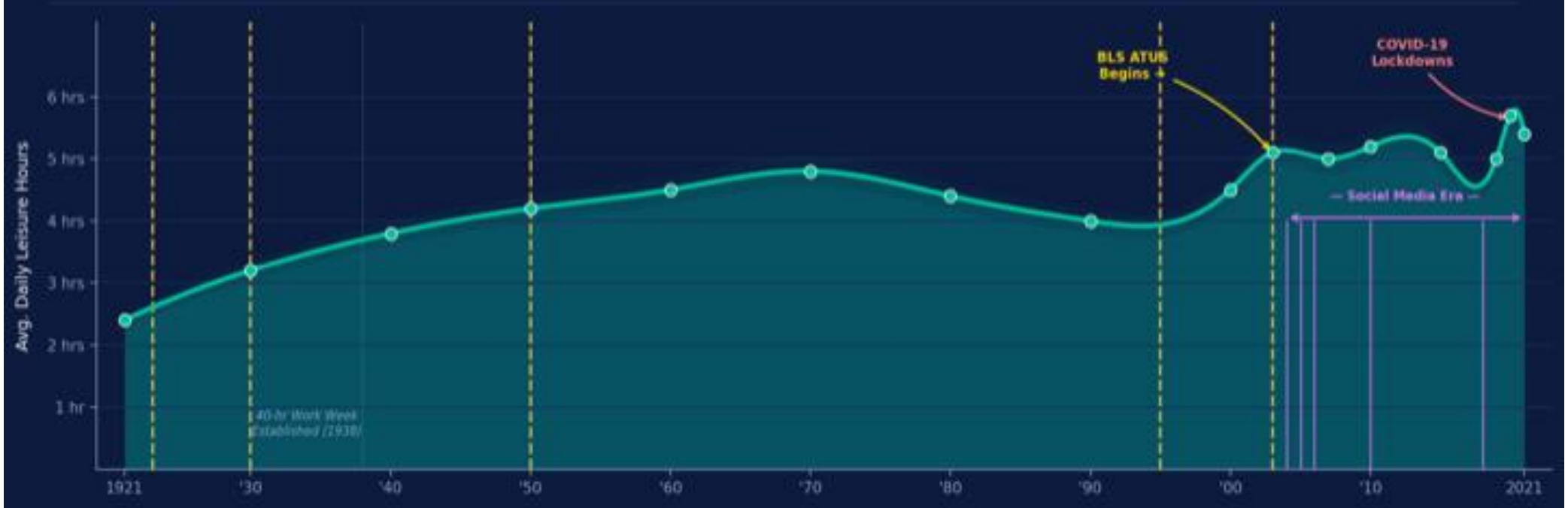


Rodwin BA, Bilan VP, Merchant NB, Steffens CG, Grimshaw AA, Bastian LA, Gunderson CG. Rate of Preventable Mortality in Hospitalized Patients: a Systematic Review and Meta-analysis. J Gen Intern Med. 2020 Jul;35(7):2099-2106.

AI Adoption for Clinical Tasks is Optional (Today)

Where Did the Time Go?

U.S. Adult Average Daily Leisure Time, 1921 – 2021 | Technology & Social Media Adoption Milestones



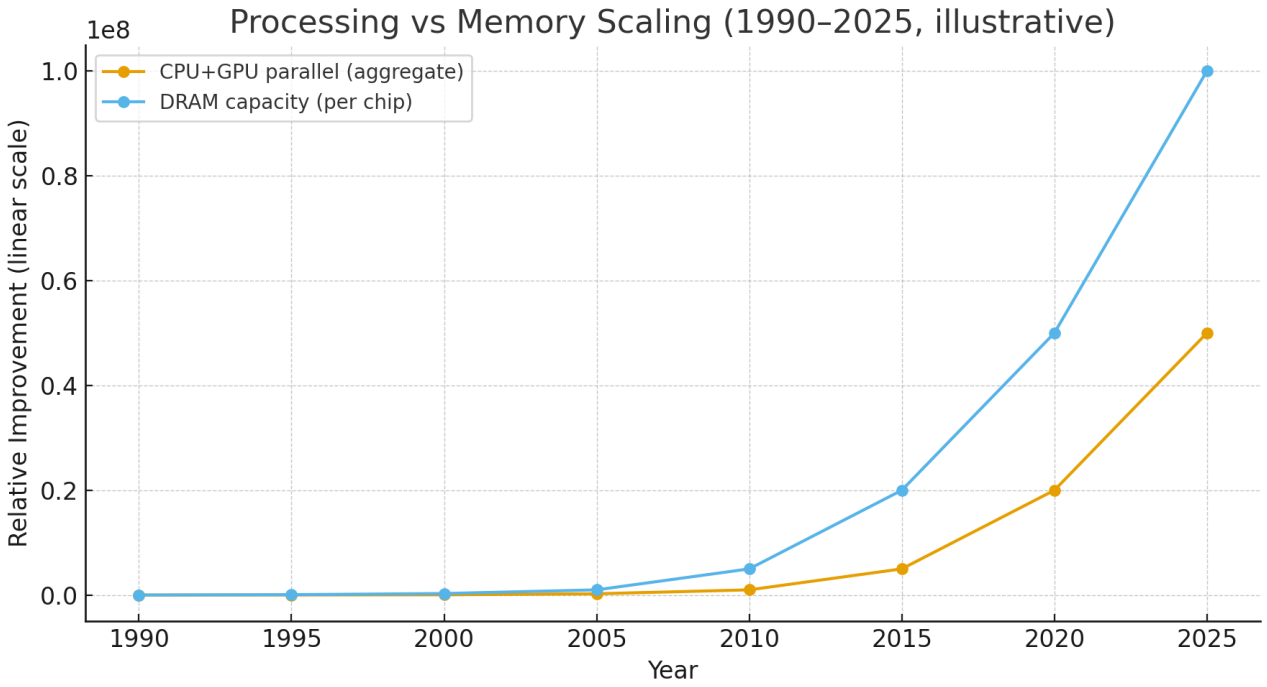
① Automobile (1923) ② Telephone (1930) ③ Appliances (1950) ④ Email (1995) ⑤ Cell Phones (2003)

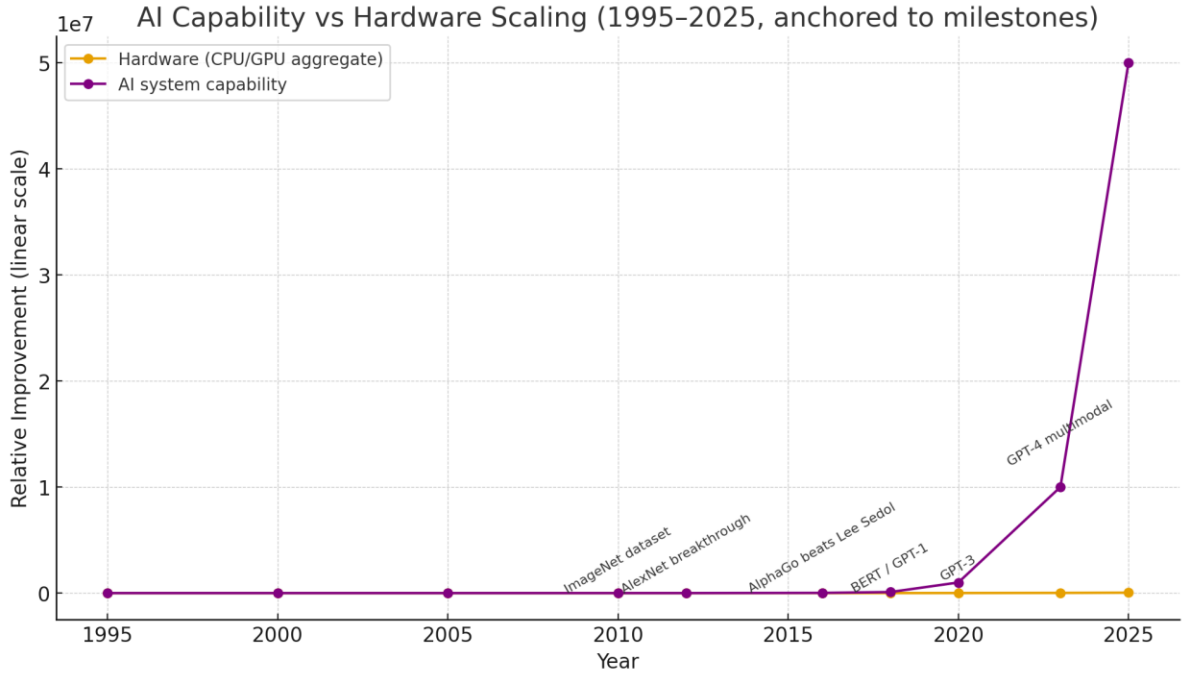
⑥ Facebook (2004) ⑦ YouTube (2005) ⑧ Twitter (2006) ⑨ Instagram (2010) ⑩ TikTok (2018)

Sources: Bureau of Labor Statistics, American Time Use Survey (ATUS), 2003-2021 | Pre-2003 scholarly estimates: Schor (1991), Ramey & Francis (2009), Roberts (2006)

Implications of Exponential Growth

No *Moore*: Pace of Performance Improvements

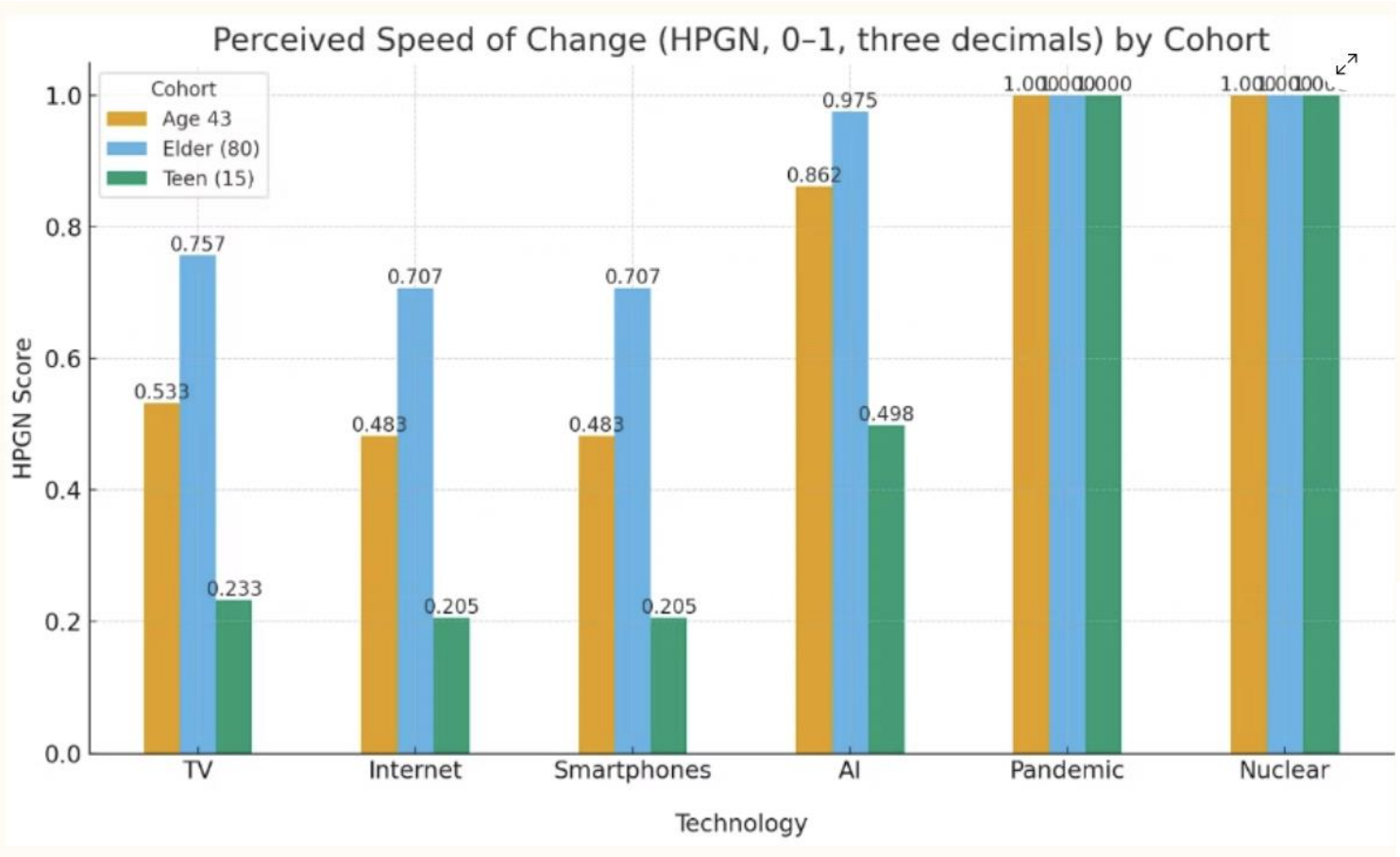




Moore's Law = Transistors/IC double every 2 years (CAGR ~ 41%)

AI Training-Compute = Doubling every 3-6 months (CAGR 300-1000%)

Keeping Up with the Pace of Progress



$$HPGN(a, T^{def\ f}, w) = min(1, (1 - e^{(-\kappa \cdot (a/T^{def\ f})^{\beta})}) \cdot (1 + 0.1w))$$

Interpretation of HPGN

Values range from 0 (imperceptible) to 1 (practically instantaneous).

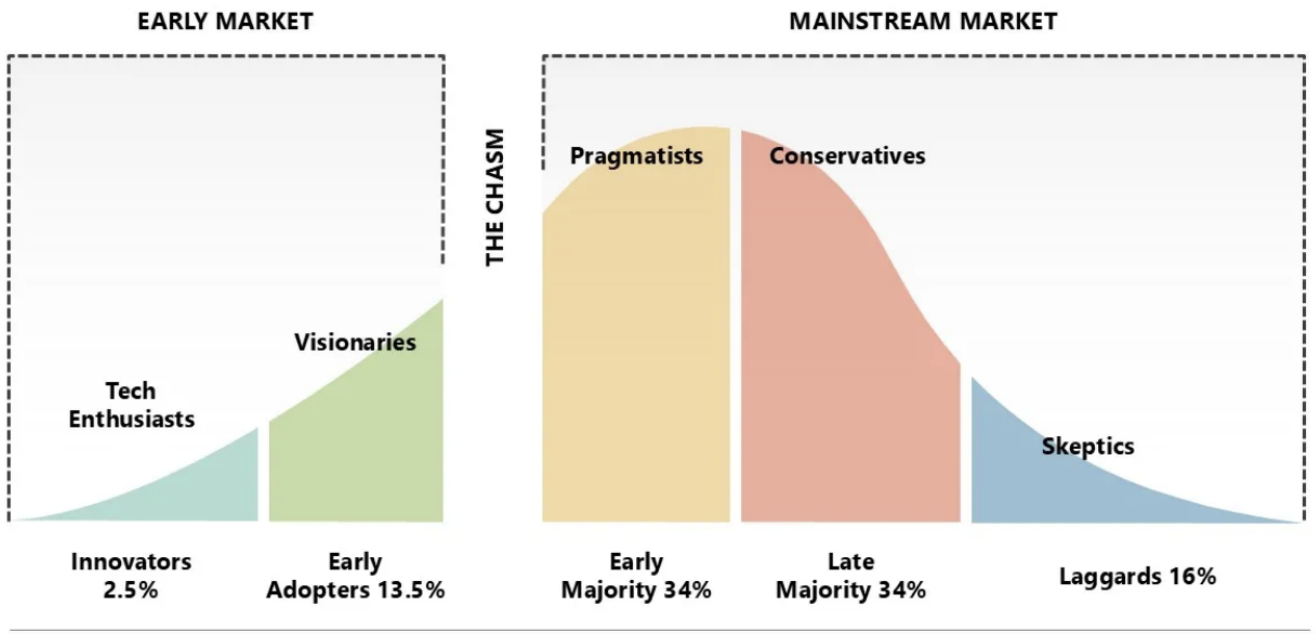
- ***HPGN=0.2–0.4**: change is rapid but manageable for most*
- ***HPGN=0.6–0.8**: change feels unsettlingly fast and disorienting; overwhelm is expected*

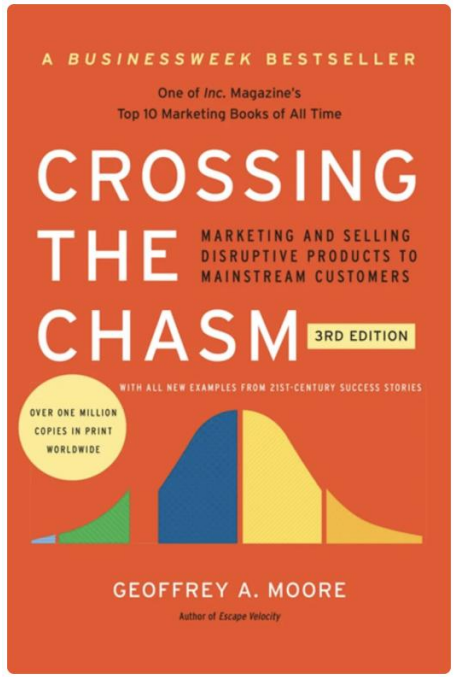
***HPGN=0.9–1.0**: change saturates perception and feels infinite; resignation is natural*

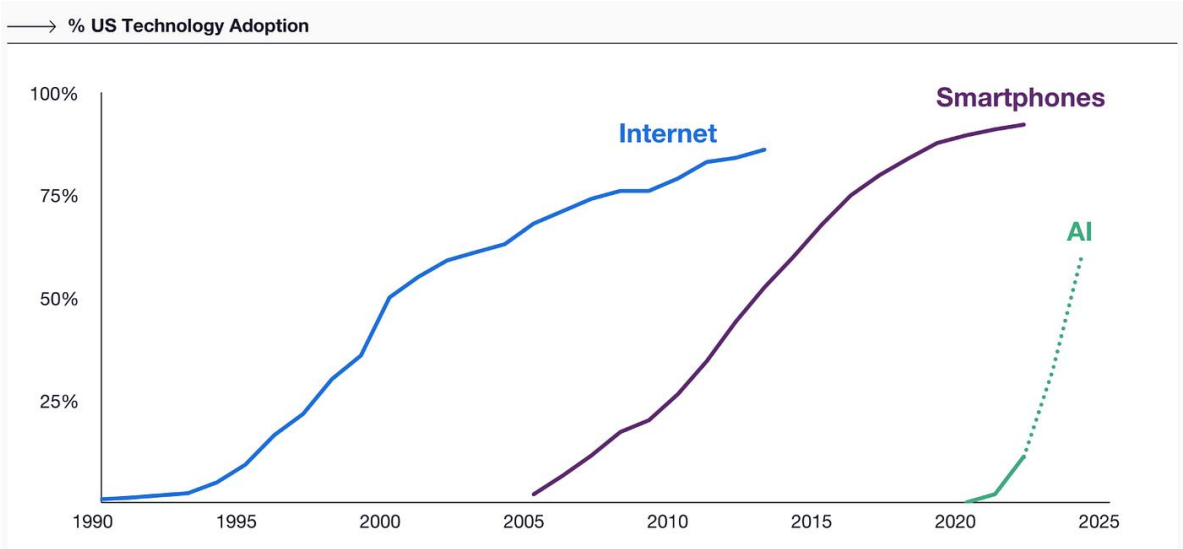
Ref: www.thefocusedexam.com, 9/28/2026

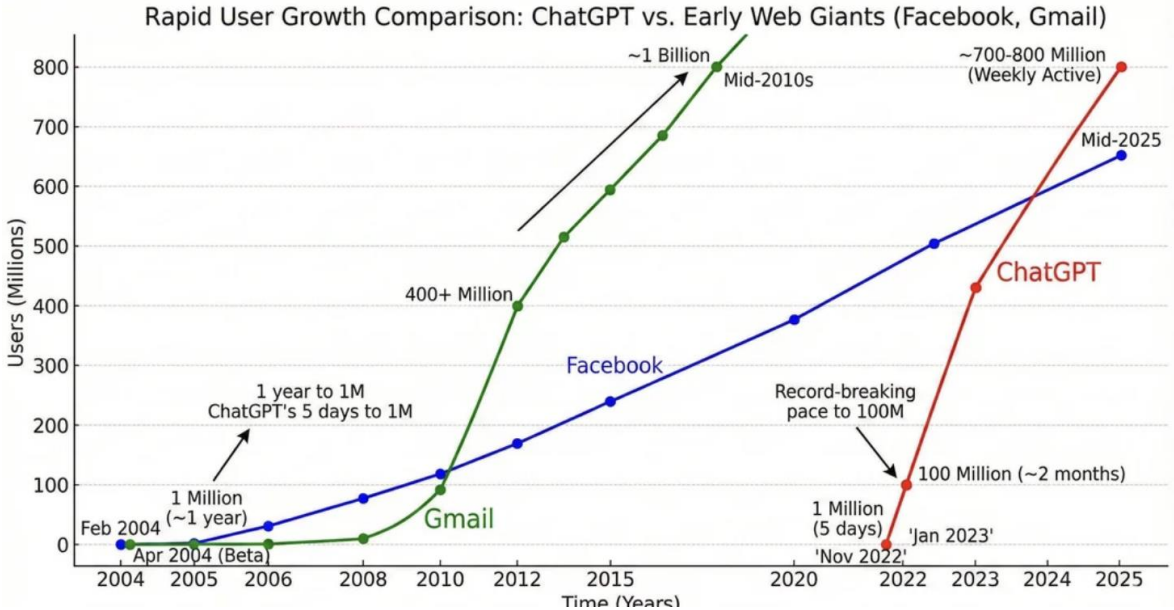
Based on this model and pace of AI improvement (~0.5-year doubling time) the HPGN values would range from 0.498 (teen) to 0.841 (age 43) to 0.975 (age 80).

Crossing the Chasm Segment Selection

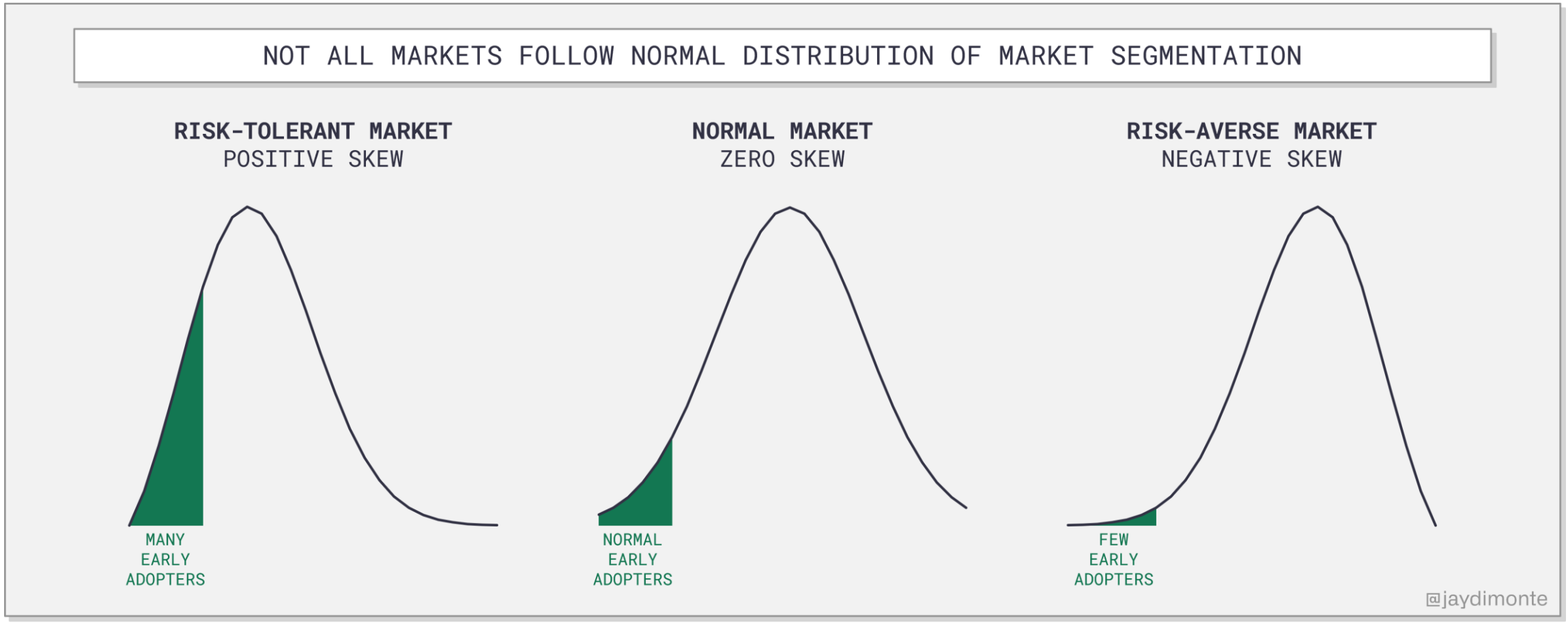








Pace of Adoption of Various Technologies



Differentiation means staying ahead of the curve

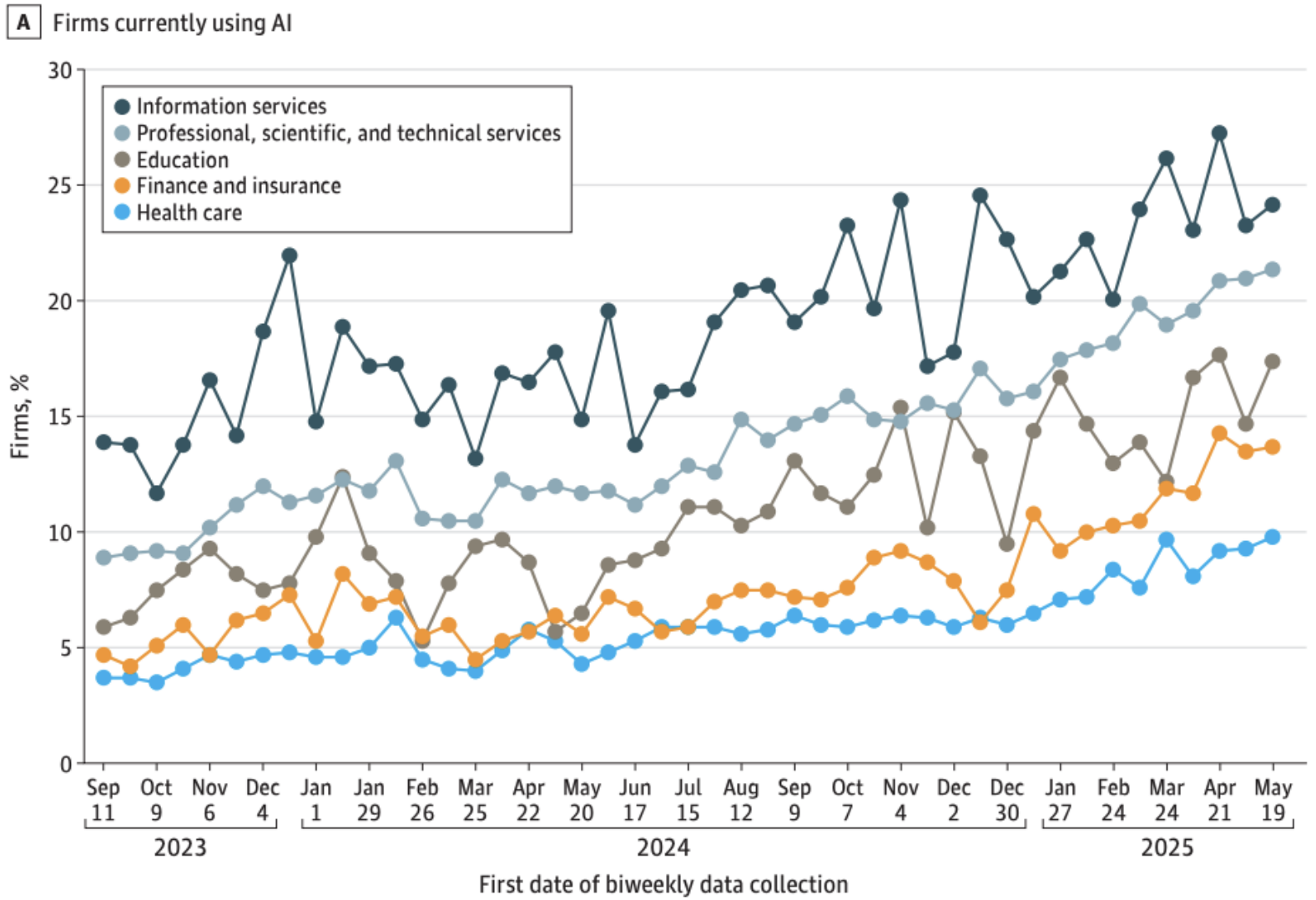
JAMA Health Forum™

Research Letter | AI in Health Policy

Adoption of Artificial Intelligence in the Health Care Sector

Thuy D. Nguyen, PhD; Christopher M. Whaley, PhD; Kosali Simon, PhD; Neil Mehta, BS; Hao Yu, PhD; Ryan K. McBain, PhD, MPH; Ateev Mehrotra, MD, MPH; Jonathan H. Cantor, PhD

Figure. Trends in Artificial Intelligence (AI) Use in Health Care vs Non–Health Care Sectors



DIFFERENT PATHS, SIMILAR FRICTION

HIPAA vs. FDA Regulatory Burdens

HIPAA & FDA approval serve different goals—data security versus patient safety—but both exert significant drag on healthcare progress and innovation.

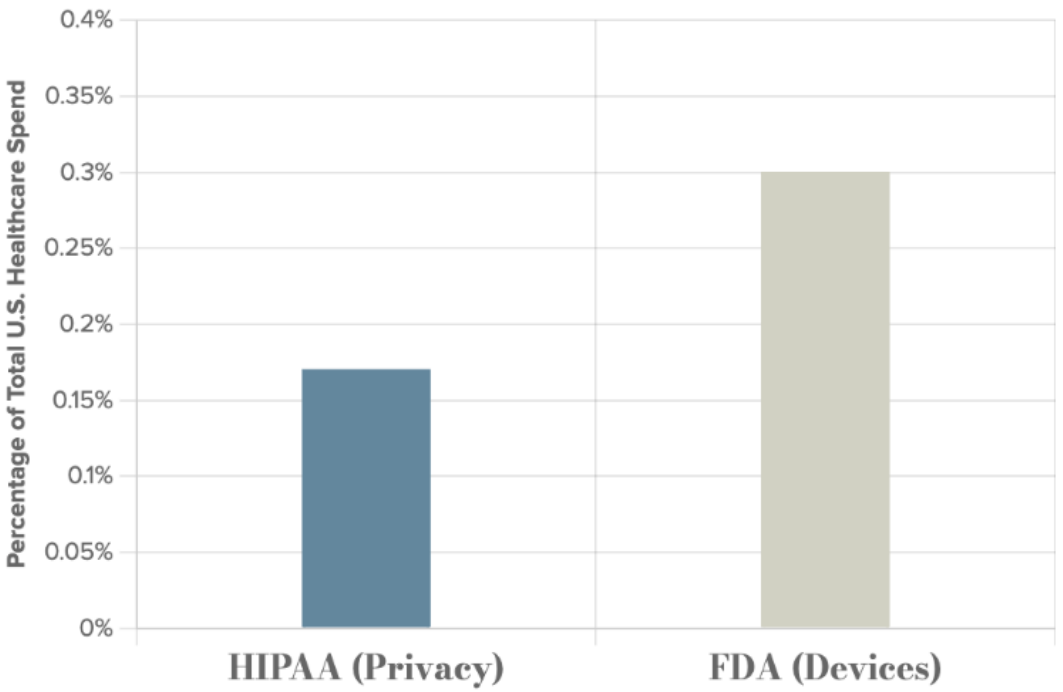
Contributing a combined regulatory cost of ~0.5% of U.S. healthcare spending ($25B/yr)

HIPAA Drag

Innovation barrier: Interoperability.
Time to market impact: 6-12 months.

FDA Drag

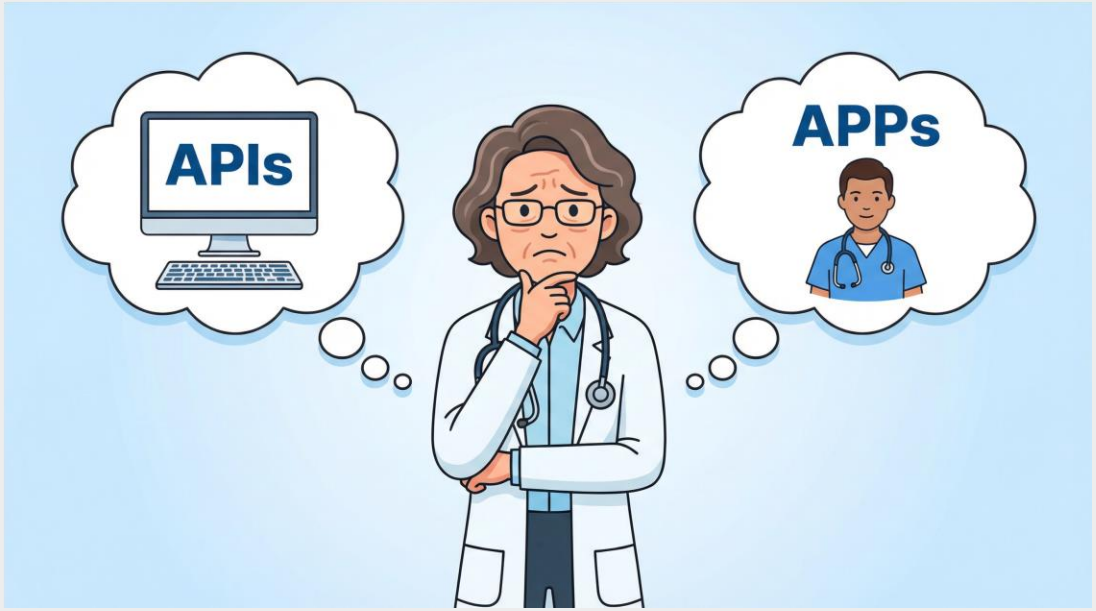
Innovation barrier: Capital
Time to market impact: 3-7 years.



Comparative Analysis of Annual Sector Costs as % of Total Healthcare Expenditure

The Baseline Context: Consumer Ease vs. Clinical Scrutiny

General Population		Clinical Workflows
Frictionless search & creation	**Primary Use Case**	High-stakes diagnostic decision-making
Mild inconvenience or bad copy	**Cost of Failure**	Severe patient harm or death
Unrestricted sandbox	**Regulatory Environment**	Highly regulated, FDA-scrutinized ecosystem
Enthusiastic early adoption	**User Sentiment**	Deeply ambivalent, burnout-driven skepticism
Broadly protected by terms of service	**Legal Liability**	Immediate, asymmetrical medical malpractice risk



Procedure

AI in Clinical Workflows: To Integrate or Not to Integrate

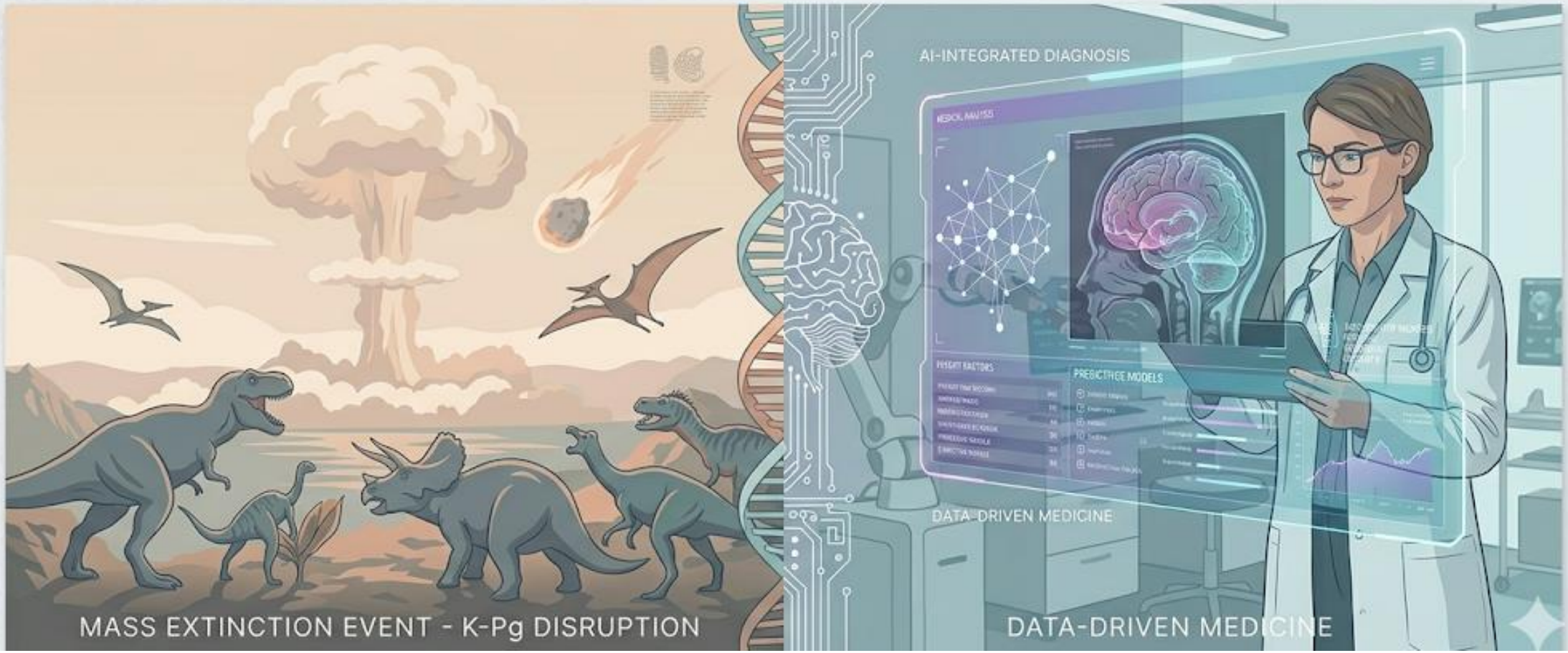
> It is clear to me that AI will never replace physicians — but physicians who use AI will replace those who don't.

— Jesse Ehrenfeld, MD

AMA President

July 2023

WHEN THE WORLD CHANGES ABRUPTLY, IT IS THE MOST ADAPTABLE WHO THRIVE IN THE NEW VERSION OF REALITY.



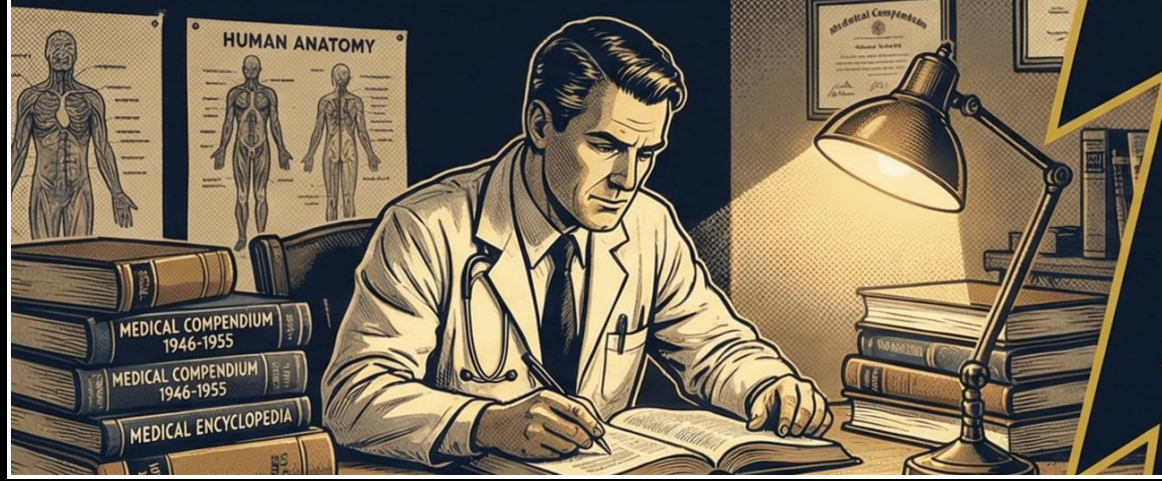
“Platform Risk” for Clinicians?

Value in Diversification

Two Paradigms of Medical Knowledge

JIC – Just-in-Case Knowledge

- Memorized during training
- Broad foundational base
- Textbooks & rote learning

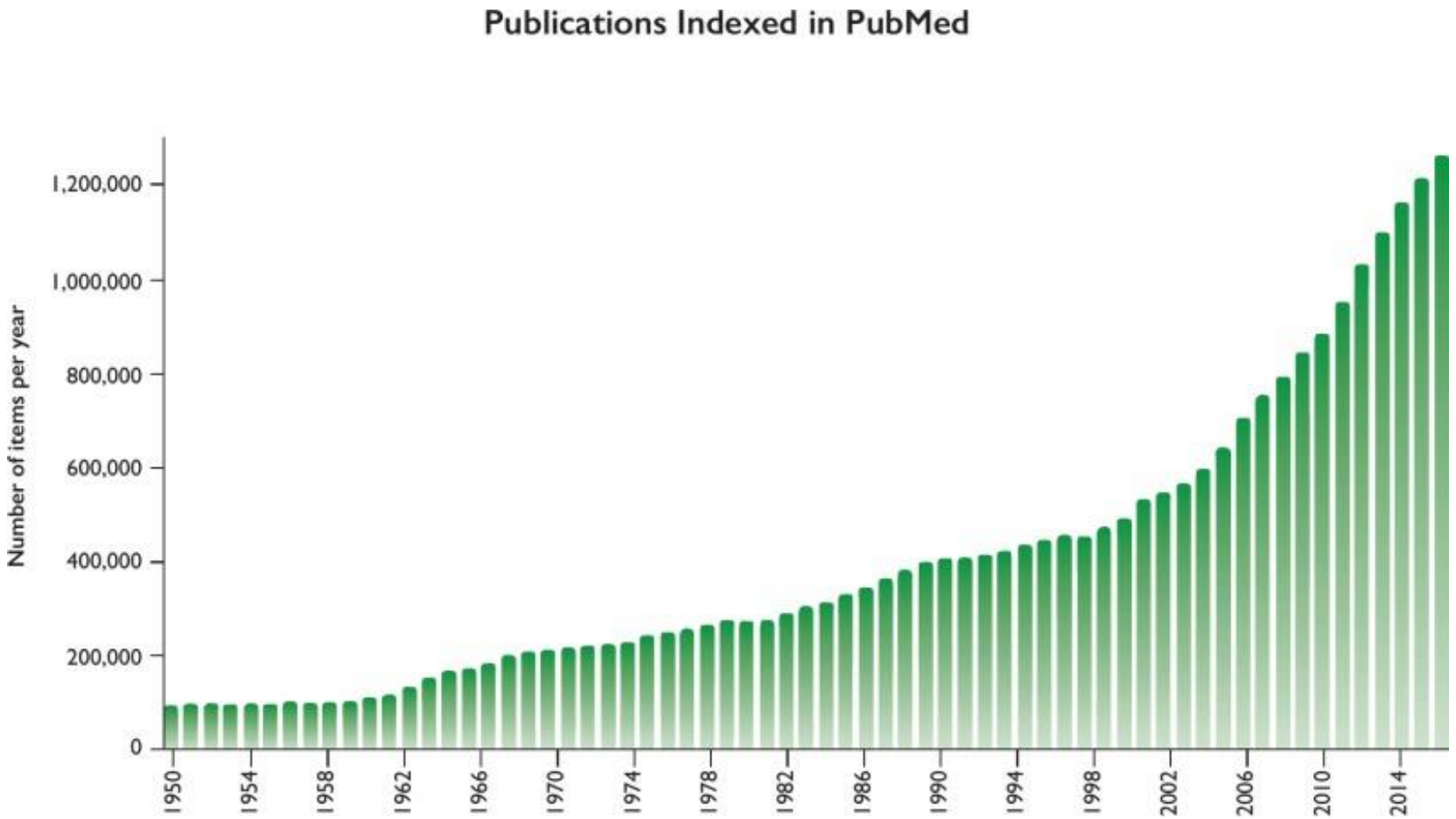


JIT – Just-in-Time Learning

- Retrieved at point of care
- Targeted & immediately actionable
- Apps & AI tools

Commoditization of Information & Shifting Value from Knowledge to Application

Progress Adds Complexity



Psychological Inertia

Medicine is built on tradition & hard-won expertise.

AI 'assistance' feels intrusive and threatening.

Practice patterns are ingrained over years & hard to change.

Identity Threat

The perception that AI diagnostic tools diminish the value of clinical expertise.

Muscle Memory

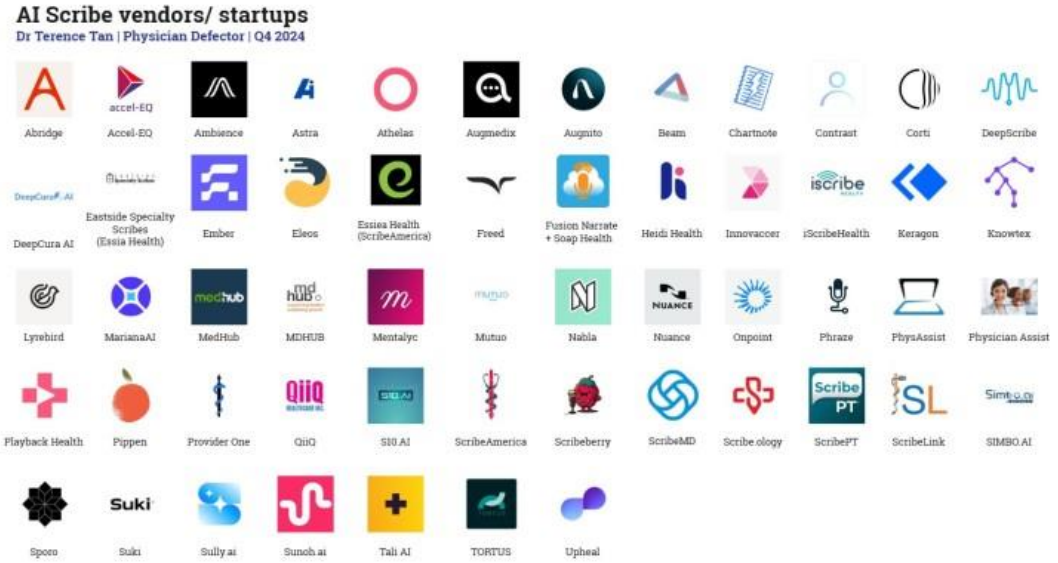
Resistance to tools that disrupt established, efficient clinical physical habits.

Tradition Bias

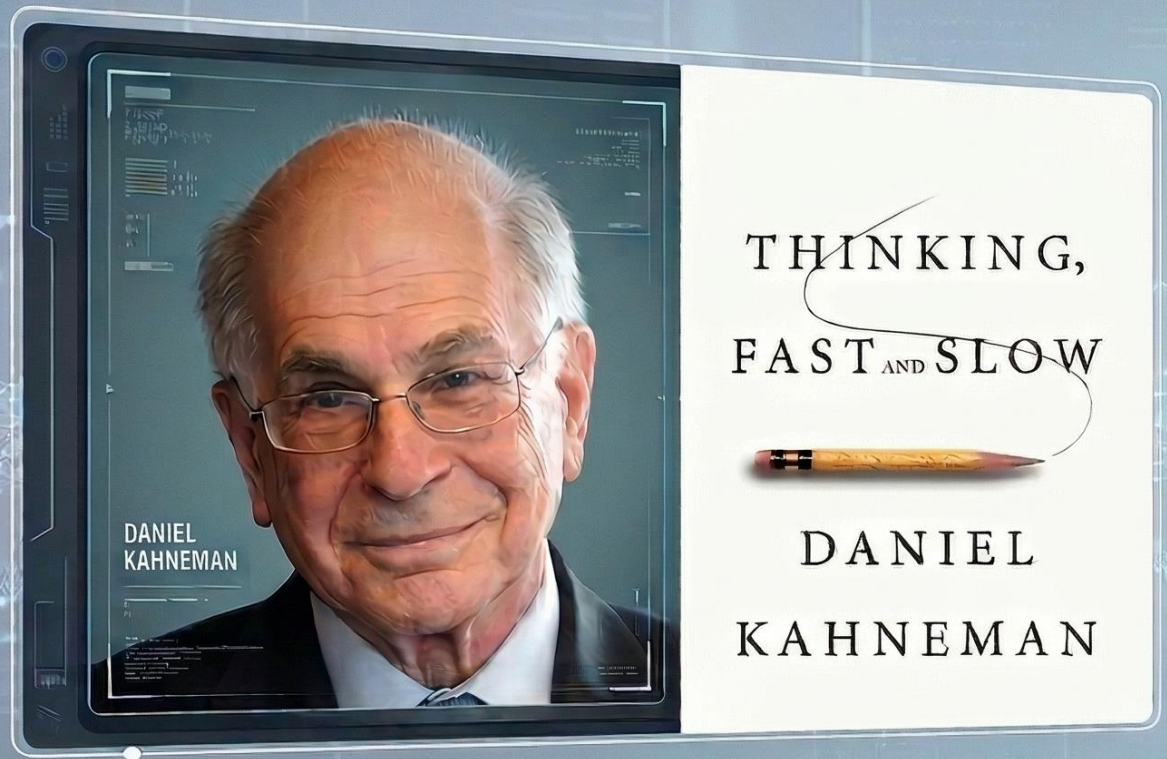
A deep-seated preference for proven, human-led methods over digital alternatives.

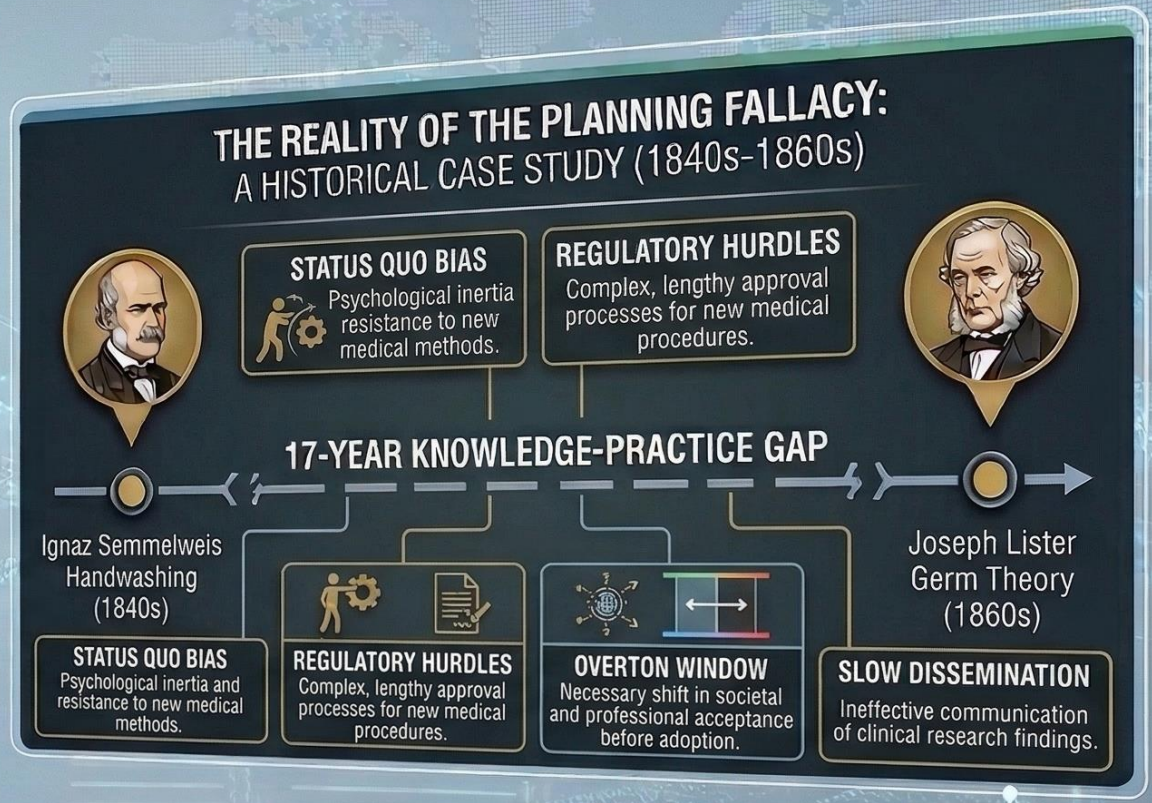
Empathy Gap

Skepticism regarding AI's ability to replicate the nuanced human side of care.



Paradox of Choice

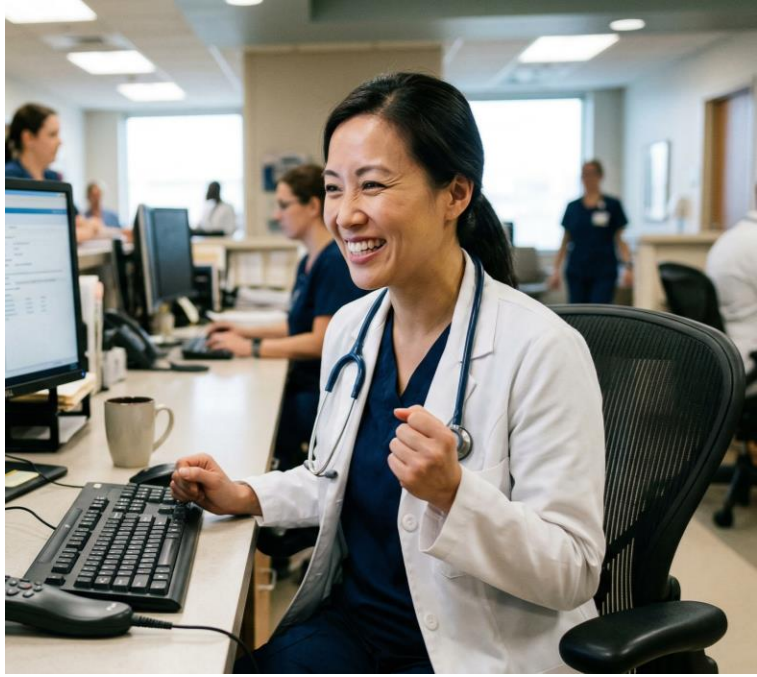




Things always take longer than we expect... Just as Real

- Knowledge-Practice Gap
- Enthusiasm-Practice Gap

Fool Me Once, Shame on You




FYRE FEST 2.0 HAS
FFICIALLY BEEN POSTPONE
VHO WOULD HAVE THOUGH


FYRE
FOR IMMEDIATE RELEASE
FEBRUARY 24, 2025
FYRE
FESTIVAL 2
IS REAL

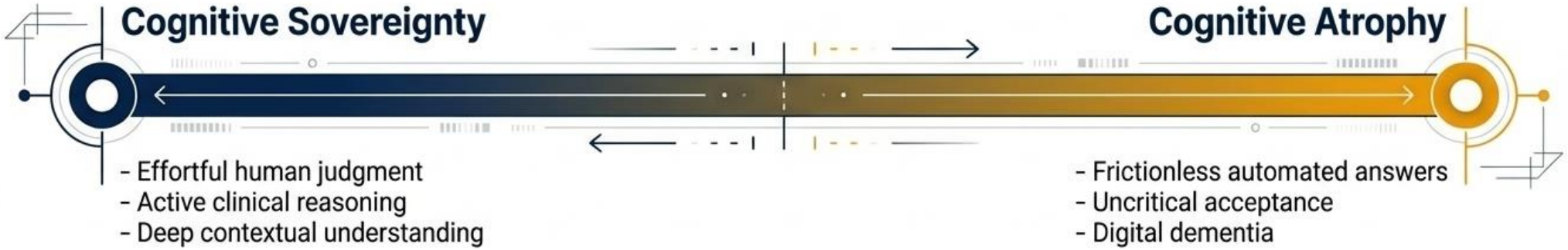
Digital Dementia: Dangers of Human Nature

The cognitive crisis is real — and clinicians are not immune

"We're rapidly losing the ability to think deeply at all — regardless of how much space we can find in our schedules."

-Cal Newport

The Sovereignty Trap and the Fear of Deskilling



The Professional Threat

Medicine requires intense cognitive engagement. Clinicians are deeply wary of a reallocation of human effort where they transition from active diagnosticians to passive reviewers of algorithmic outputs.

The Central Question

What remains uniquely human in diagnosis when answers arrive faster than understanding?

THE CYCLE OF DIGITAL DEMENTIA: HOW COGNITIVE OFFLOADING REWIRES THE BRAIN

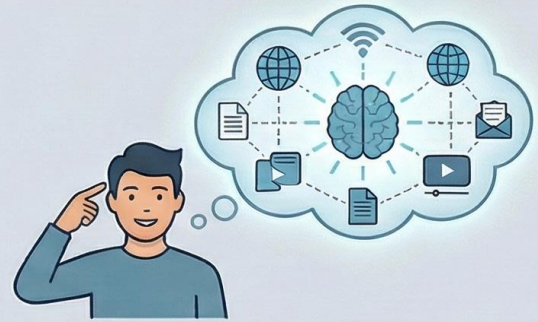
1. THE LEAP TO COGNITIVE OFFLOADING

Individuals delegate memory, calculations, and reasoning to external AI systems and digital tools rather than internalizing information.

2. THE 'GOOGLE EFFECT' RECONFIGURATION

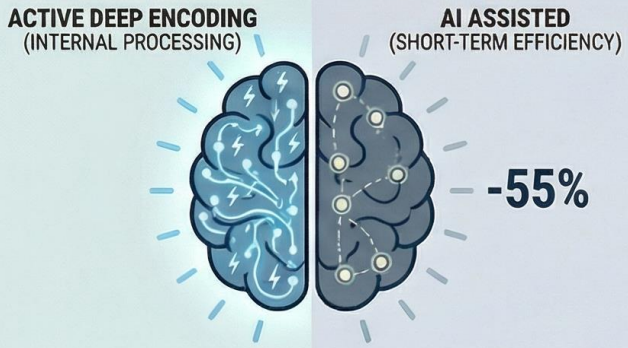
The brain shifts from storing actual content to storing the "pathway" or location of where to find that information.

3. THE ILLUSION OF INTERNAL KNOWLEDGE



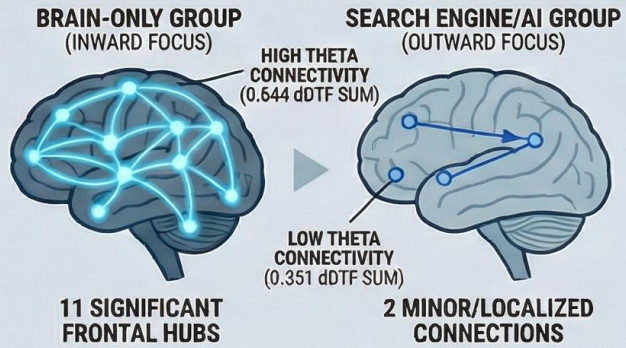
Frequent searching creates a metacognitive distortion where users conflate the internet's knowledge with their own internal abilities.

4. 55% CORTICAL ACTIVITY DROP



Writing with AI assistance can lead to a 55% reduction in cortical activity, purchasing short-term efficiency at the cost of deep encoding.

5. DIMINISHED FRONTAL THETA POWER



Interaction with LLMs reduces theta band connectivity, a hallmark of reduced engagement in working memory and integrative reasoning.

6. THE EMERGENCE OF "CLIP THINKING"

Prolonged digital dependency results in a preference for fragmented information (short videos/texts) and a decreased ability to think critically.

Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task△

Nataliya Kosmyna [1]
MIT Media Lab
Cambridge, MA

Eugene Hauptmann
MIT
Cambridge, MA

Ye Tong Yuan
Wellesley College
Wellesley, MA

Jessica Situ
MIT
Cambridge, MA

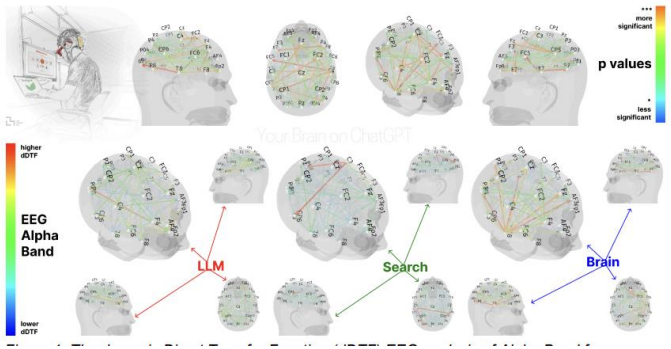
Xian-Hao Liao
Mass. College of Art and Design (MassArt)
Boston, MA

Ashly Vivian Beresnitzky
MIT
Cambridge, MA

Iris Braunstein
MIT
Cambridge, MA

Pattie Maes
MIT Media Lab
Cambridge, MA

United States



Group		Session 1	Session 2	Session 3	Session 4
		18 participants per group, 54 total. Choice of 3 SAT topics per session, 9 topic options total			18 participants total, choice from previously written topics, reassignment of participants: Brain-to-LLM and LLM-to-Brain.
LLM	NLP	Homogenous ontology. Common n-grams shared with Search group. Frequent location and dates NERs. Some participants used LLM for translation. Impaired perceived ownership. Significantly reduced ability to quote from their essay.	Slightly better ontology structure. Much less deviation from the SAT topic prompt. Heavy impact of person NER: like 'Matisse' in ART topic.	Low effort. Mostly copy-paste. Not significant distance to the default ChatGPT answer to the SAT prompt. Minimal editing. Impaired perceived ownership.	Better integration of content compared to previous Brain sessions (Brain-to-LLM). More information seeking prompts. Scored mostly above average across all groups. Split ownership.
	EEG	Initial integration. Baseline.	Higher interconnectivity. Smaller than in the Brain group. High integration flow.	Lower interconnectivity due familiar setup, consistent with a neural efficiency adaptation. Low effort visual integration and attentional engagement.	High memory recall. Low strategic integration. Higher directed connectivity across all frequency bands for Brain-to-LLM participants, compared to LLM-only Sessions 1, 2, 3.

Search Engine	NLP	Mid size essay. 50% to 100% lower use of NER compared to LLM group. High perceived ownership. High quoting ability.	Some topics show the likely impact of search optimizations like focus on 'homeless' n-gram in PHILANTHROPY topic. Split perceived ownership.	Highly homogenous to other topics written using Search Engine.	N/A
	EEG	Initial integration. Baseline.	High visual-executive integration to incorporate visual search results with cognitive decision making. High interconnectivity.	Lower interconnectivity, likely due to familiar setup, consistent with a neural efficiency adaptation.	
Brain only	NLP	Shorter essays. High perceived ownership. High quoting ability.	More concise essays. Scored lower on accuracy by AI judge and human teachers within the group.	Distance between essays written in the Brain group is always significant and high compared to LLM or Search Engine groups.	Used n-grams from previous LLM sessions. Scored higher by human teachers within the group. Split ownership.
	EEG	Initial integration. Baseline.	Robust increases in connectivity in all bands.	Peak beta band connectivity.	High memory recall. High strategic integration. Session 4's brain connectivity did not reset to a novice (Session 1, Brain-only) pattern, but it also did not reach the levels of Session 3, Brain-only. Mirrored an intermediate state of network engagement. Connectivity was significantly lower than the peaks observed in Sessions 2, 3 (alpha) or Session 3 (beta), yet remained above Session 1.

From "magical" to "expected or unexpected"

Once you see the mechanism, surprising behavior becomes classifiable instead of mystical.

EXPECTED

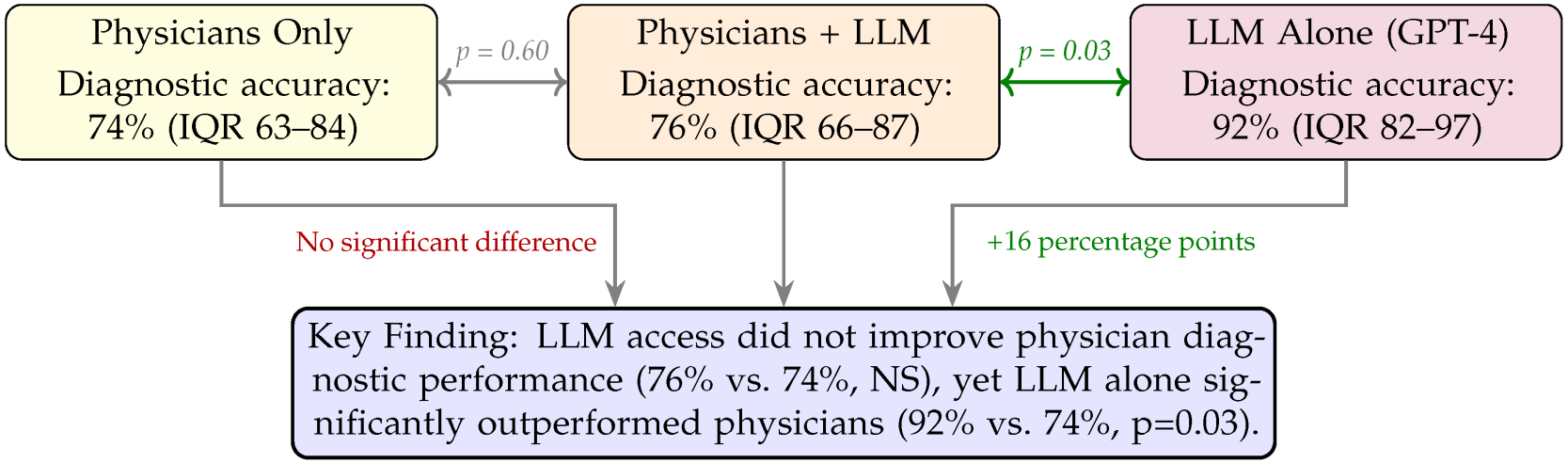
Fast drafting and rewriting
Summaries over large volumes
Pattern completion in familiar formats
Tone, structure, and reading-level shifts
Strong performance when the task is clearly specified

UNEXPECTED

Confidently invented facts
Different answers when wording changes
Missed unstated context or edge cases
Overconfident tone when caution is needed
Variability from run to run

Treat it like any probabilistic system: manage the failure modes, the controls, and the cost of being wrong - not the myth.


MENU ≡
TIME
SUBSCRIBE
HEALTH
AI
Microsoft's AI Is Better Than Doctors at Diagnosing Disease
ADD TIME ON GOOGLE
by Alice Park
SENIOR CORRESPONDENT
JUL 2, 2025 7:11 AM PT
robotic hand holding stethoscope (concepts of AI medical care) Kilito Chan—Getty Images


Physicians Only
Diagnostic accuracy:
74% (IQR 63–84)
p = 0.60
Physicians + LLM
Diagnostic accuracy:
76% (IQR 66–87)
p = 0.03
LLM Alone (GPT-4)
Diagnostic accuracy:
92% (IQR 82–97)
No significant difference
+16 percentage points
Key Finding: LLM access did not improve physician diagnostic performance (76% vs. 74%, NS), yet LLM alone significantly outperformed physicians (92% vs. 74%, p=0.03).

nature medicine

Brief Communication https://doi

ChatGPT Health performance in a structured test of triage recommendations

Received: 15 January 2026

Accepted: 20 February 2026

Published online: 23 February 2026

Check for updates

Ashwin Ramaswamy[1] ✉, Alvira Tyagi[1], Hannah Hugo[2], Joy Jiang[3,4], Pushkala Jayaraman[3,4], Mateen Jangda[4,5], Alexis E. Te[1], Steven A. Kaplan[1], Joshua Lampert[3], Robert Freeman[3], Nicholas Gavin[6], Ashutosh K. Tewari[1], Ankit Sakhuja[3,4,7], Bilal Naved[3,4,7], Alexander W. Charney[3,4,7], Mahmud Omar[3,7], Michael A. Gorin[1,3], Eyal Klang[3,7,8] & Girish N. Nadkarni[3,4,7,8] ✉

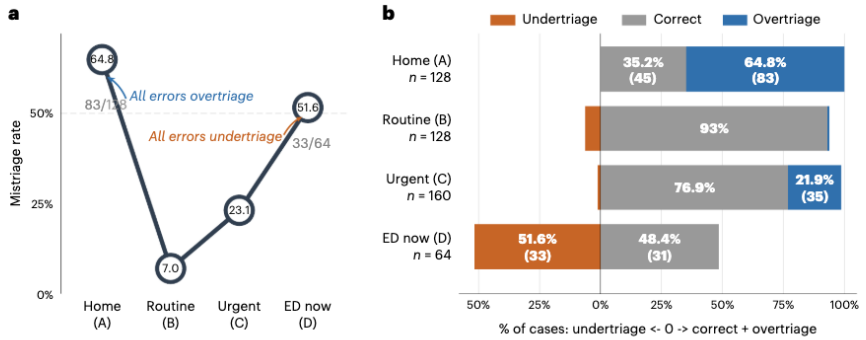


Fig. 1 | ChatGPT Health undertriages emergencies while overtriaging nonurgent cases. Clear vignettes only (single correct gold-standard triage; $n = 480$ responses). Triage levels—A (monitor at home), B (see a doctor within weeks), C (see a doctor within 24–48 h) and D (go to the emergency department now). **a**, Mistriage rate across gold-standard acuity. Mistriage (1 − accuracy) followed a U-shaped pattern, with the highest values at the extremes (A, 64.8%; D, 51.6%) and the lowest at intermediate acuity (B, 7.0%; C, 23.1%). Because A is the least urgent category and D the most urgent, errors at A necessarily represent overtriage and errors at D necessarily represent undertriage (annotations). Dashed line marks 50% mistriage for reference. **b**, Direction of triage outcomes. Within each gold-standard acuity level, stacked diverging bars show the proportion of cases that were undertriaged (recommended less urgent care than gold, left of zero), correctly triaged (gray) or overtriaged (recommended more urgent care, right of zero). Emergencies (D) were undertriaged in 33/64 (51.6%) cases, whereas nonurgent/home-care cases (A) were overtriaged in 83/128 (64.8%).

"Remember your destination. This will help you to distinguish between an opportunity to be seized and a temptation to be resisted."

The Two Journeys (Bamidbar, Life-Changing Ideas, Covenant & Conversation)

What is our Destination?

Clinician AI Fluency

Self-Assessment & Self-Start Guide
Physicians · PAs · NPs · Dentists

10 Domains · 1–5 Scale
Assessment + Reference
Updated: 3/2026

SCAN

PART 1

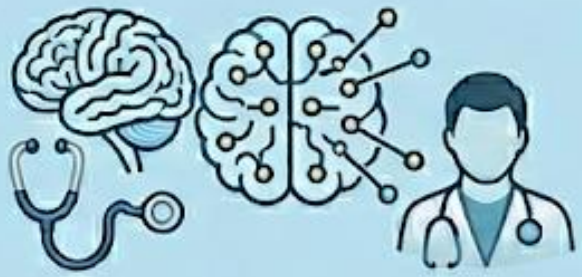
Self-Assessment

Rate each domain 1–5 on *current behavior, not aspiration.* A high score reflects sound judgment about when *not* to use AI. Clinician review remains the final safety check in every domain.

Domain	Description	1 Not yet	2 Rarely	3 Sometimes	4 Usually	5 Reliably
01 Task Triage	Identify which clinical and administrative tasks benefit from AI assistance and which require unassisted clinician judgment, including					

PRACTICING AT THE TOP OF YOUR LICENSE

INCREASING WORKFORCE SATISFACTION BY MAXIMIZING THE IMPLEMENTATION OF EXPERTISE



FOCUS ON COMPLEX CARE

- Direct complex diagnoses and treatment.

APPLY ADVANCED EXPERTISE

- Lead medical management and innovation.

DELEGATE ROUTINE TASKS

- Shift administrative tasks to other staff or technology (e.g., AI Agents)

FOCUS ON HIGH UTILITY UPSKILLING

1.) Maintain cognitive load and challenges. 2. Offload High-Focus, Repeated Admin Tasks.

ADVANCED COMMUNICATION

PRACTICE PERSUASIVE COMMUNICATION

- Refine Humor, Storytelling, and Persuasive Communication.
- Enhance clarity and influence.

COGNITIVE FLEXIBILITY

DEVELOP ANALOGOUS THINKING

- Train analogous thinking for complex challenges.
- Apply diverse thought models.

TECH & TALENT VARIETY

LEVERAGE BROAD SKILLS

- Explore a Variety of Technical Tools.
- Leverage unique IRL (In Real Life) talents.

A MINDSET FOR OUR MOMENT

APOCALOPTIMIST

n. someone who 1.) acknowledges that world is in real trouble; 2.) yet refuses to succumb to denial, paralysis, or cynicism; 3.) instead, they accept reality; 4.) by turning fear into action; 5.) and find meaning by resisting the idea that our feared terrible future is inevitable – '*Counteracting the Self-fulfilled Prophecy*'

Accept Reality Clearly

- AI poses real risks: cognitive atrophy, job disruption, algorithmic error, existential uncertainty
- Attention spans are collapsing; critical thinking measurably declining with AI dependence
- Pretending otherwise is not optimism — it's denial

Act Without Certainty

- You don't need certainty of outcome to act. Uncertainty means **the outcome isn't predetermined**
- Act because it needs doing, not because you know it will make a difference (Burkeman)
- "Pessimism of the intellect, optimism of the will" — Gramsci's insight applied to AI

Actively Engage with AI

- **Don't wait for AI to be "safe"** — engage deliberately today
- **Protect your cognitive skills** while embracing AI for tasks that free your expertise
- **Be the clinician** who shapes how AI is used, not the one it shapes

DOES HISTORY RECORD ANY CASE IN WHICH THE MAJORITY WAS RIGHT?

- ROBERT A. HEINLEIN -

Don’t Be Afraid to Think Differently

Scan any code to explore

PODCAST

Topol & Wachter Discuss AI

Substack | erictopol.substack.com

VIDEO

AI in Clinical Practice

YouTube | youtu.be/TC64U_rzEuI

OPINION

Technology & Mental Fitness

NYT Opinion | Cal Newport

BOOK

"A Giant Leap" by Wachter

Book | amazon.com

RESEARCH

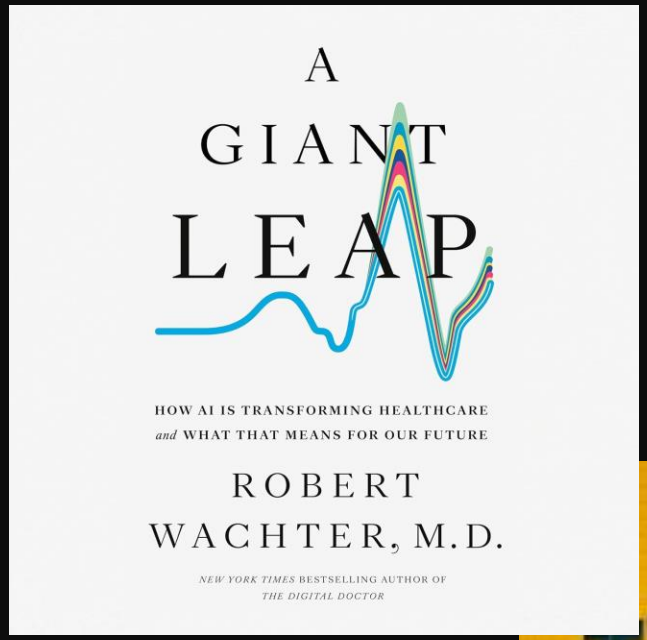
Nature Medicine AI Research

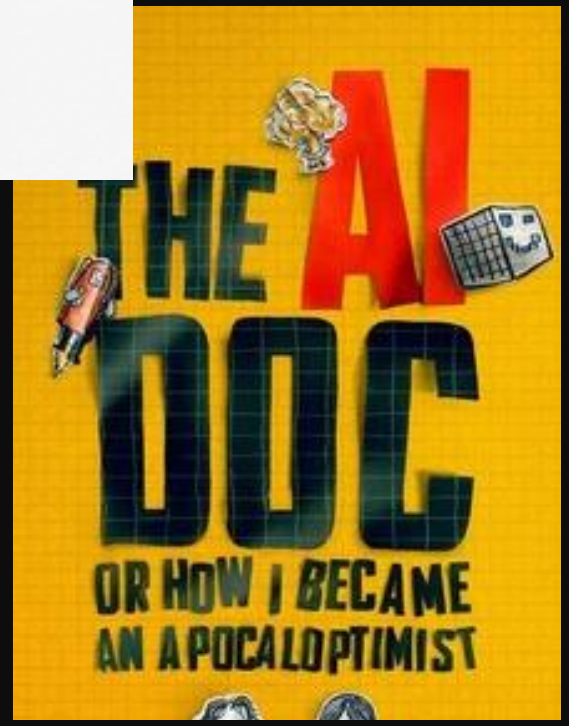
nature.com | Peer-Reviewed

OPTION A

Follow the Controversy.

Become a Thought Leader.





AI TOOLS

FOR CLINICAL PRACTICE

A curated set of AI assistants relevant to clinical work — from diagnosis support to medical literature synthesis.

Use with clinical judgment. AI augments, not replaces, physician expertise.

C

Claude

Anthropic

✓

GP

ChatGPT

OpenAI

✓

G

Gemini

Google

✓

P

Perplexity

Perplexity AI

✓

DS

DeepSeek

DeepSeek AI

✓

G

Glass Health

Glass

✓

OE

OpenEvidence

OpenEvidence

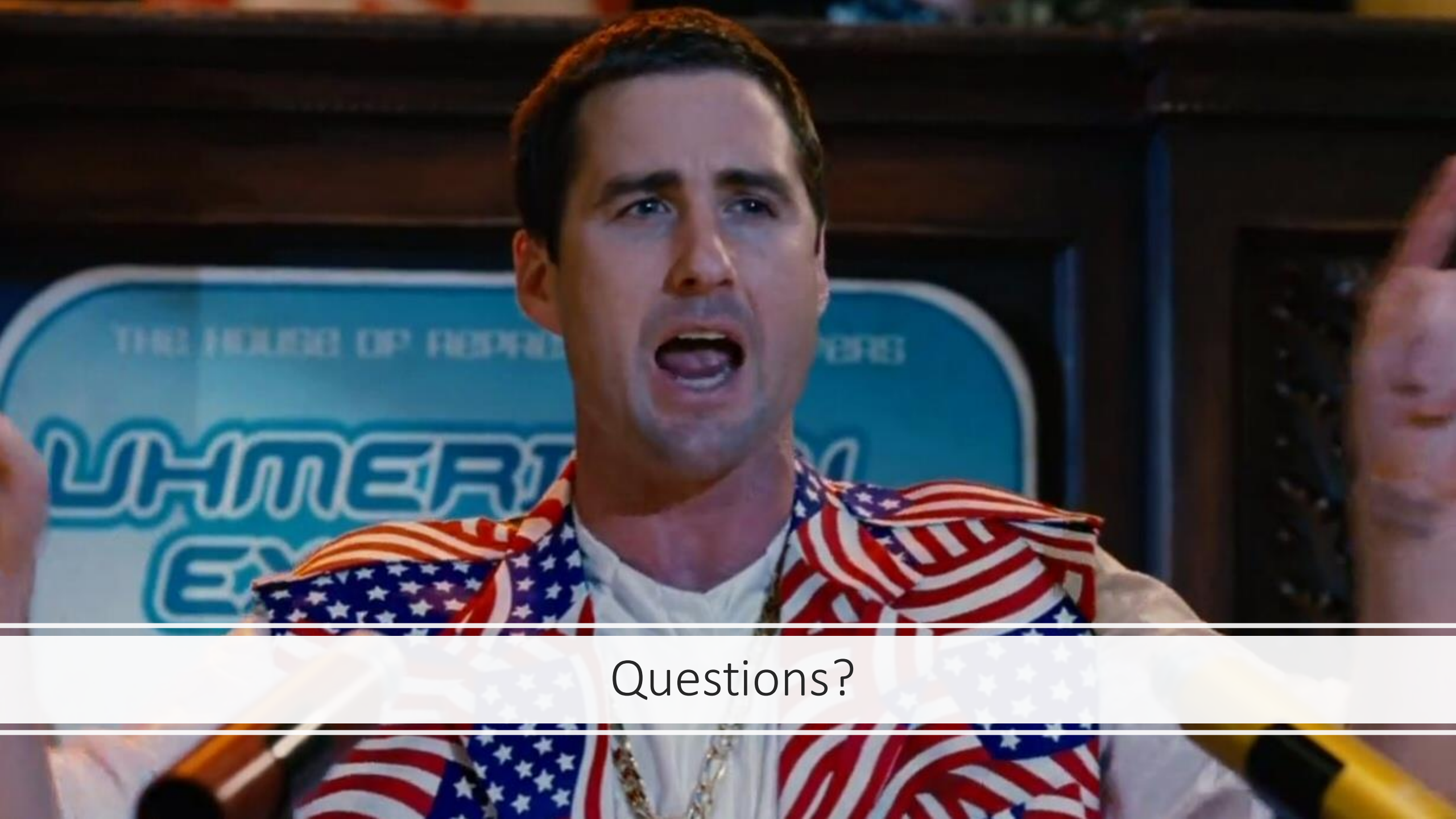
✓

M

Manus

Manus AI

✓



Questions?

I Need Feedback – Scan the QR