Benefits and Challenges in AI for Primary Care:

Real World Impact & Ongoing Monitoring

Jaky Kueper, PhD

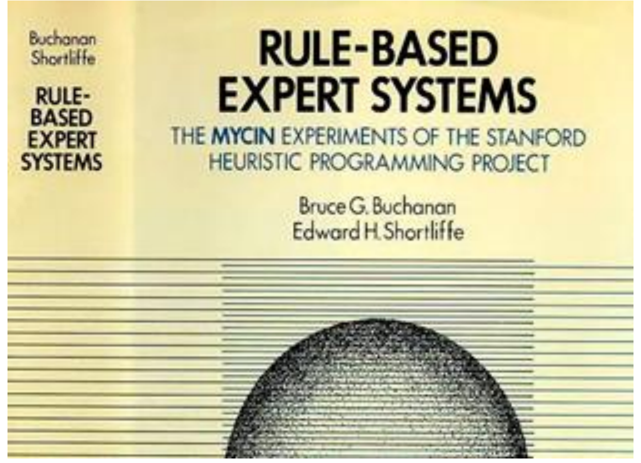
Scripps Research Digital Trials Center

NAPCRG

A *brief* history of milestones in AI methods

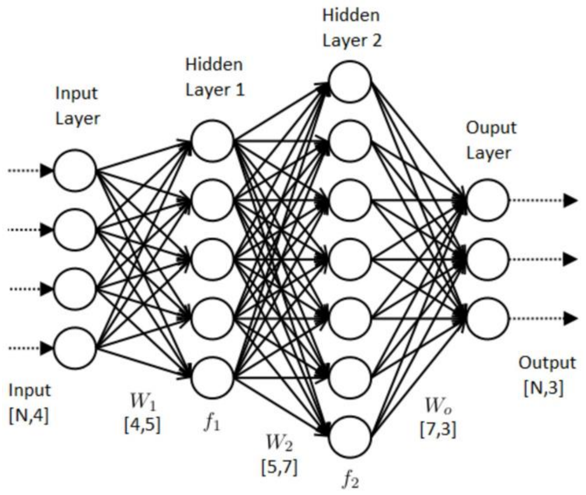
1840s
First Computer Algorithm

1950s
Field of AI Formalized



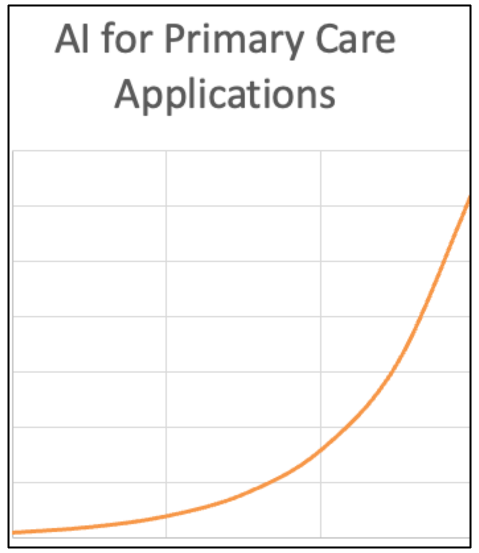
1970s
Rule-Centric Methods
First medical application

2010s
Data-Centric Methods
Deep learning boom



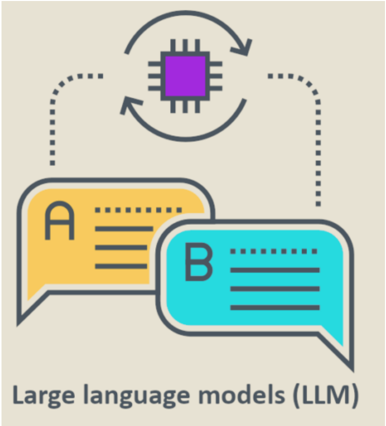
2020s
Narrow AI/ML showing success at health-relevant tasks

"predicting"
"describing"
"recommending"



Large Language Models and Generative AI methods

"creating"

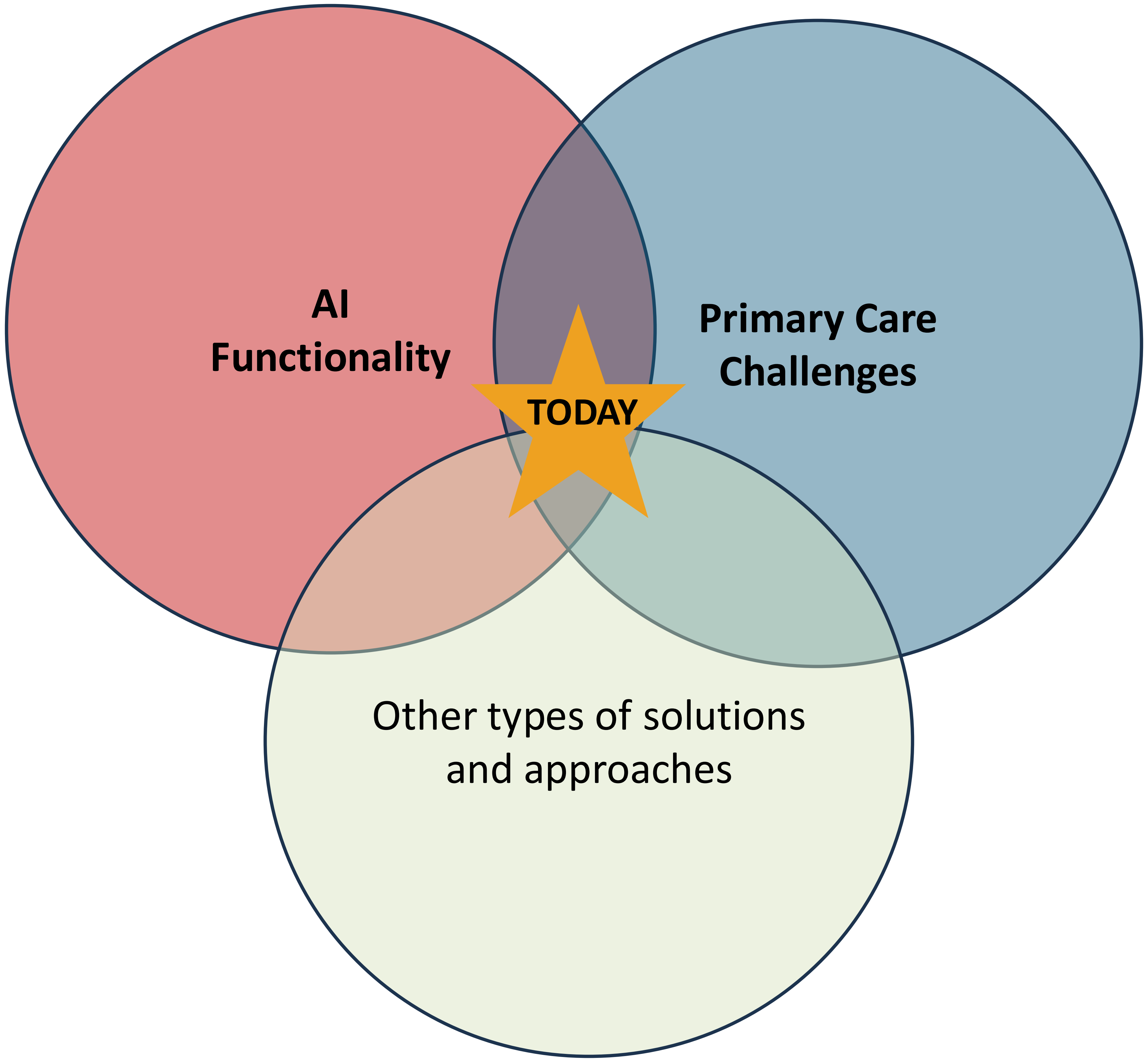


Agentic AI methods

"doing"

NAPCRG

Today: AI methods can now address meaningful primary care challenges



Real world impact starts with a clearly defined goal / problem to solve, while not limiting the vision of what's possible based on past approaches.

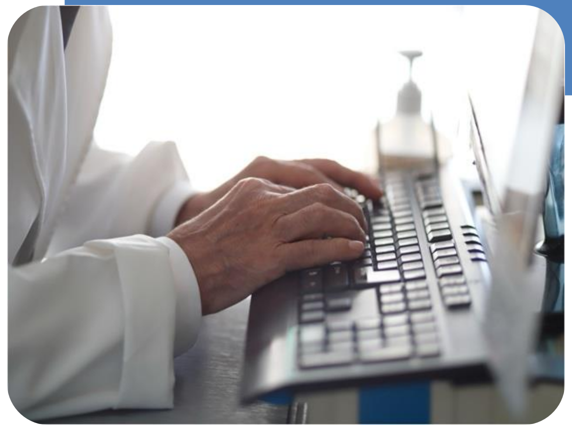
Benefit: Progressing beyond technical milestones to real world impact

Example: AI Scribes

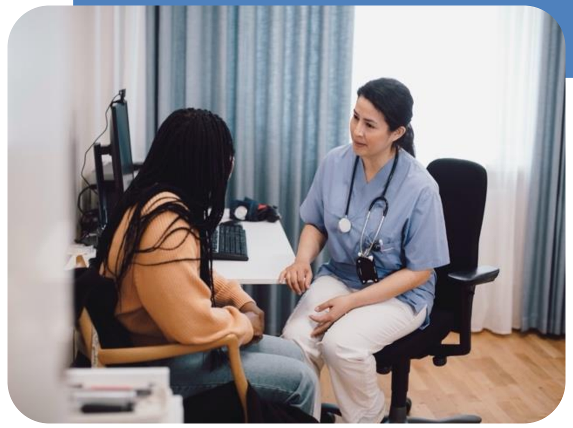
Technical / Immediate AI Output

Direct Impact / Clinical Action

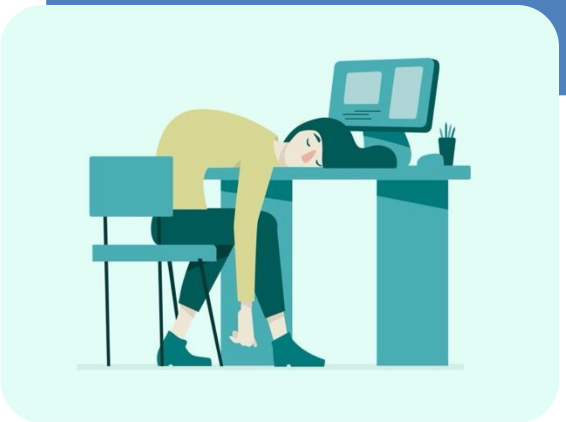
Downstream "Challenging" Outcomes



Example: Accuracy and appropriateness of notes



Example: Engagement with AI scribe, time spent on EMR activities



Example: Care quality, provider job satisfaction and retention

Benefit: Progressing beyond technical milestones to real world impact

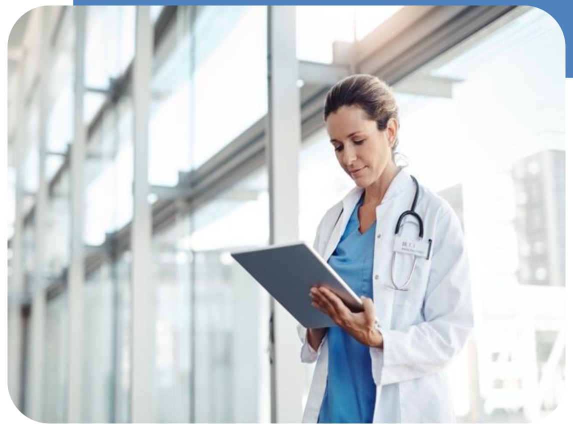
Example: Asthma exacerbation (AE) prevention

Technical / Immediate AI Output

Direct Impact / Clinical Action

Downstream "Challenging" Outcomes

Example: Accuracy of high-risk AE predictions



Example: Physician burden and time for clinical decision with asthma patient

Example: One year occurrence of AE

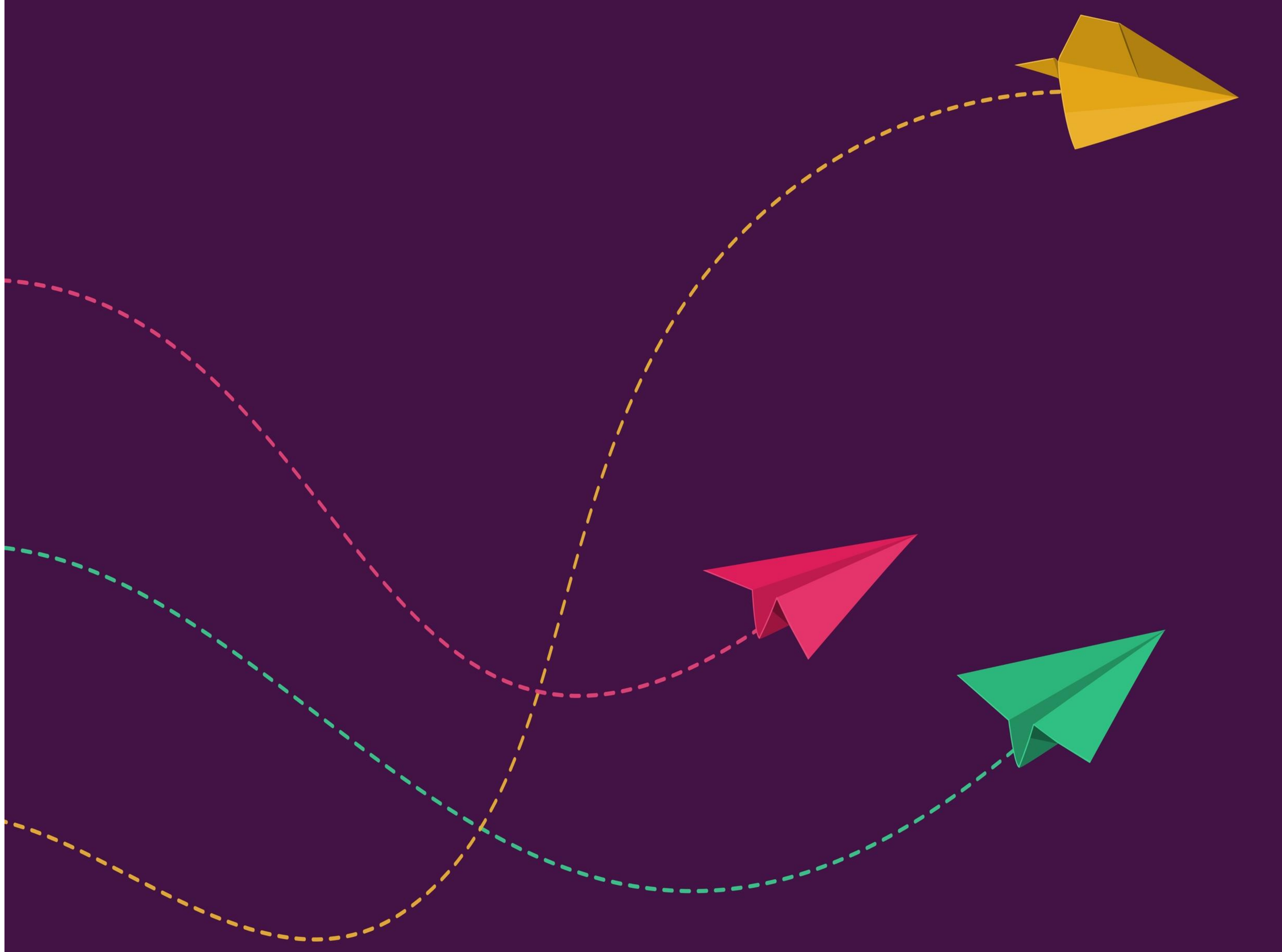
Seol HY, Shrestha P, Muth JF, et al. Artificial intelligence-assisted clinical decision support for childhood asthma management: A randomized clinical trial. *PLoS One*. 2021;16(8):e0255261.

Many Outcomes to Consider

- Patient outcomes
- Provider outcomes
- System-level outcomes
- Care team communication and effectiveness
- Cost-effectiveness
- Safety, security
- Environmental considerations
- Health equity and fairness
  - Do all subpopulations experience similar outcomes
  - Who/where does or does not have access to high-quality, safe AI tools?

Note even a “perfectly” performing AI tool from a technical lens may not have real world impact

- **Need more high-quality primary care prospective evaluation studies!**

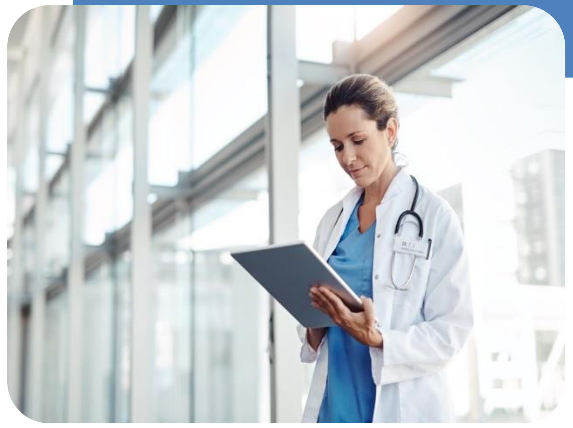


Challenge: AI performance changes across location, time, populations

Need post-deployment ongoing monitoring of performance and impact

Technical / Immediate AI Output

Direct Impact / Clinical Action



Downstream "Challenging" Outcomes

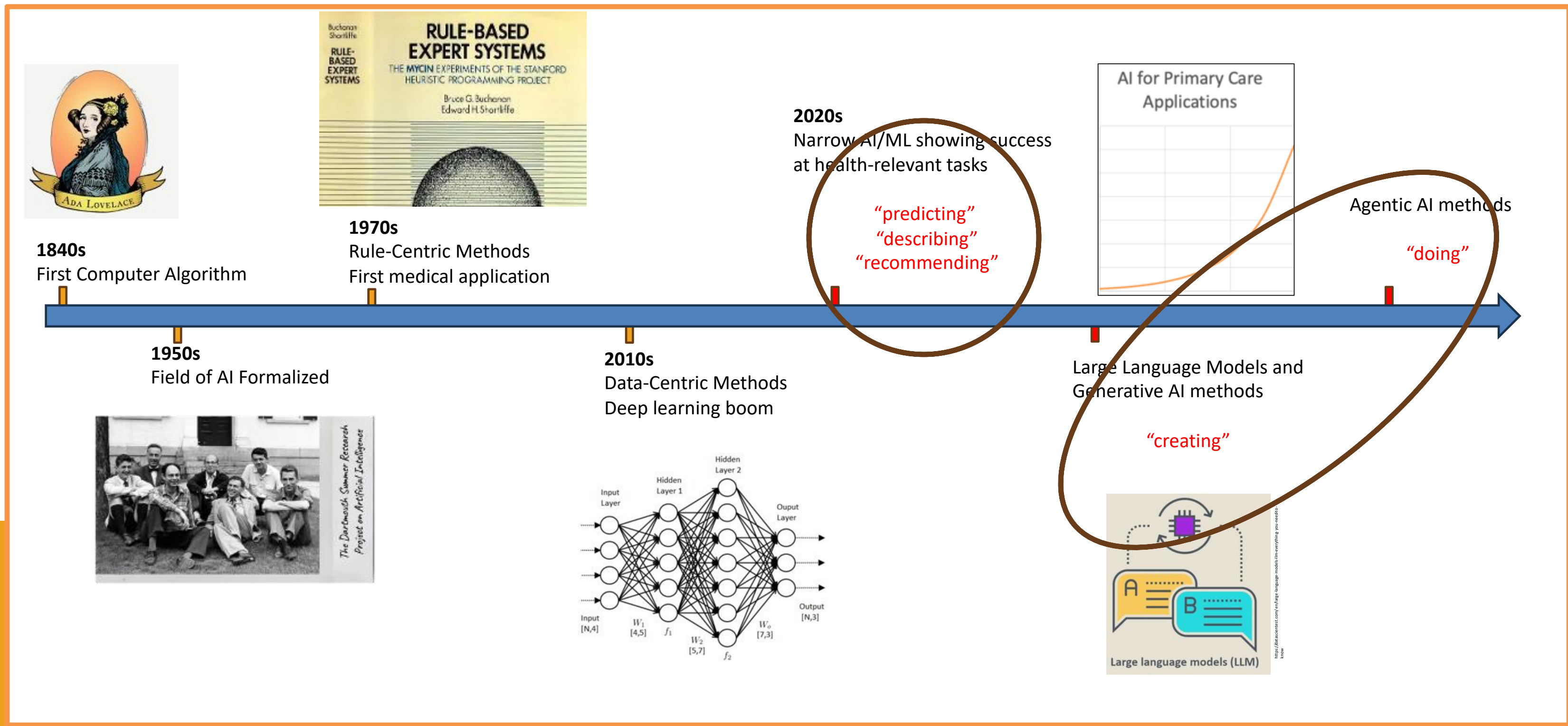
Challenge: Methodological advancements and the availability of (not necessarily high quality) AI tools have outpaced evaluation & monitoring strategies

Methodological advancements & *expected* real world impacts

Rigorous, well-accepted evaluation strategies for single-task “classic” ML tools

Evaluation strategies for multi-purpose generative and agentic AI solutions

Ongoing monitoring and model maintenance over time



Challenge: Beyond single tool/task evaluation & monitoring strategies

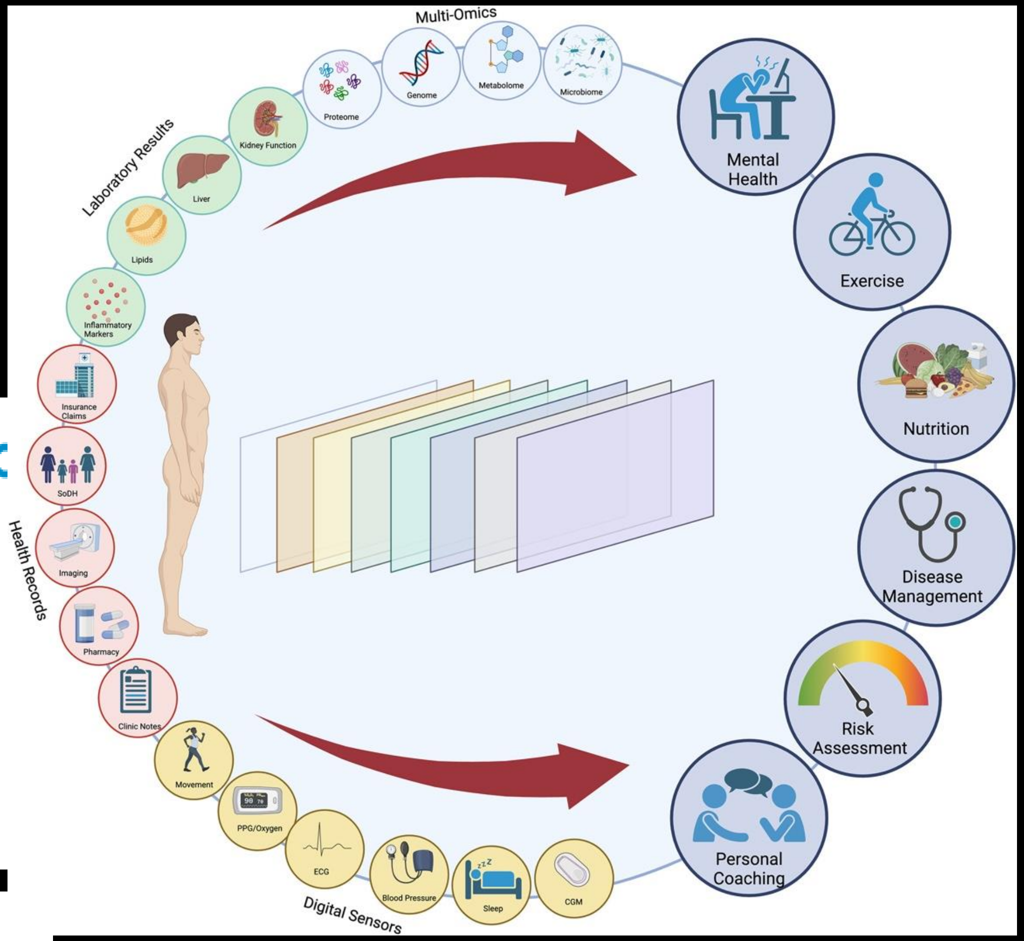
Ongoing monitoring and maintenance

Lots of excitement.
Several challenges.
Largescale success TBD.

CellPress Cell Metabo

Review
Transforming the cardiometabolic disease landscape: Multimodal AI-powered approaches in prevention and management

Evan D. Muse[1,2] and Eric J. Topol[1,2,*]
[1]Scripps Research Translational Institute, Scripps Research, La Jolla, CA 92037, USA
[2]Division of Cardiovascular Diseases, Scripps Clinic, La Jolla, CA 92037, USA



NEJM AI
NEJM AI 2025;2(5)
DOI: 10.1056/AIp2401122

PERSPECTIVE
Why Is Primary Care Different? Considerations for Machine Learning Development with Electronic Medical Record Data

Jacqueline K. Kueper, Ph.D.,[1] Winston Liaw, M.D., M.P.H.,[2] Daniel J. Lizotte, Ph.D.,[3,4] and Sian Hsiang-Te Tsuei, M.D., Ph.D., C.C.F.P.[5]

Receive

Perspective
Foundation models for generalist medical artificial intelligence

https://doi.org/10.1038/s41586-023-05881-4
Received: 3 November 2022

Michael Moor[1,6], Oishi Banerjee[2,6], Zahra Shakeri Hossein Abad[3], Harlan M. Krumholz[4], Jure Leskovec[1], Eric J. Topol[5,7] & Pranav Rajpurkar[2,7]

Towards Generalist Biomedical AI

Tao Tu[*, ‡, 1], Shekoofeh Azizi[*, ‡, 2],
Danny Driess[2], Mike Schaekermann[1], Mohamed Amin[1], Pi-Chuan Chang[1], Andrew Carroll[1], Chuck Lau[1], Ryutaro Tanno[2], Ira Ktena[2], Basil Mustafa[2], Aakanksha Chowdhery[2], Yun Liu[1], Simon Kornblith[2], David Fleet[2], Philip Mansfield[1], Sushant Prakash[1], Renee Wong[1], Sunny Virmani[1], Christopher Semturs[1], S Sara Mahdavi[2], Bradley Green[1], Ewa Dominowska[1], Blaise Aguera y Arcas[1], Joelle Barral[2], Dale Webster[1], Greg S. Corrado[1], Yossi Matias[1], Karan Singhal[1], Pete Florence[2], Alan Karthikesalingam[†, ‡, 1] and Vivek Natarajan[†, ‡, 1]
[1]Google Research, [2]Google DeepMind

NEJM AI
NEJM AI 2024; 1 (8)
DOI: 10.1056/AIp2400289

PERSPECTIVE
Evolution of Future Medical AI Models — From Task-Specific, Disease-Centric to Universal Health

Weizhi Ma, Ph.D.,[1] Bin Sheng, Ph.D.,[2,3] Yang Liu, Ph.D.,[1,4] Jing Qian, Psy.D.,[5] Xiaoxuan Liu, Ph.D.,[6] Jingshan Li, Ph.D.,[7] David Ouyang, M.D.,[8] Haibo Wang, M.B., B.S., M.P.H.,[9] Atanas G. Atanasov, Ph.D.,[10,11] Pearse A. Keane, M.D.,[12,13] Wei-Ying Ma, Ph.D.,[1] Yih-Chung Tham, Ph.D.,[14,15,16,17] and Tien Yin Wong, M.D., Ph.D.[14,1]

Received: March 17, 2024; Revised: May 4, 2024; Accepted: May 16, 2024; Published: July 12, 2024

Perspective

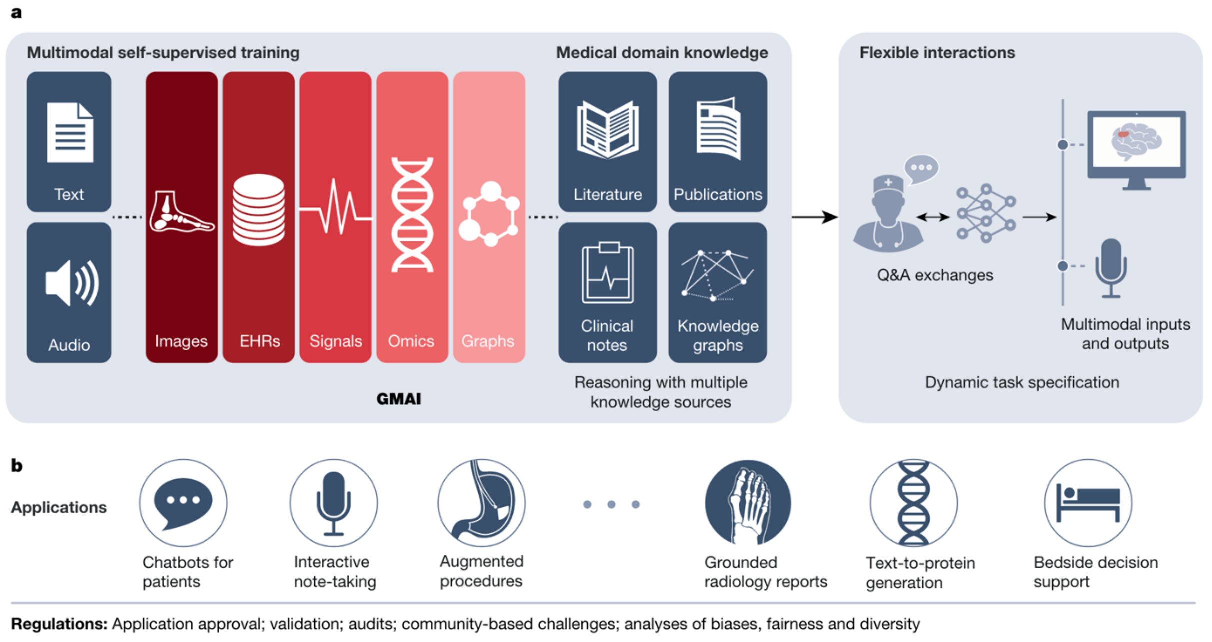


Fig. 1 | Overview of a GMAI model pipeline. a, A GMAI model is trained on ... out tasks that the user can specify in real time. For this, the GMAI model can

JGIM

PERSPECTIVE
Ten Ways Artificial Intelligence Will Transform Primary Care

Check for updates

Steven Y. Lin, MD[1], Megan R. Mahoney, MD[1], and Christine A. Sinsky, MD[2]
[1]Division of Primary Care and Population Health, Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA; [2], American Medical Association, Chicago, IL, USA.

Ethics and governance of artificial intelligence for health: Large multi-modal models

Risks to be addressed		What can be done, and by who	
Development phase		Developer actions	Government actions
Bias	Privacy	Certification/training for programmers	Have and enforce strong data protection laws
Labor concerns	Carbon and water footprints	Data protection impact assessments	Issue target product profiles
		Training data collected with 'best-practice' data protection rules	Mandate outcomes (predictability, interpretability, corrigibility, safety, cybersecurity)
		Training data are refreshed, up-to-date, and context-appropriate	Introduce pre-certification programmes to identify and avoid ethical risks
		Ensure transparency of training data	Conduct audits during early AI development
		Fair wages and support to data workers	Require developers to address carbon and water footprints

NEJM AI
NEJM AI 2025;2(2)
DOI: 10.1056/AIra2400657

REVIEW ARTICLE
Not All Clinical AI Monitoring Systems Are Created Equal: Review and Recommendations

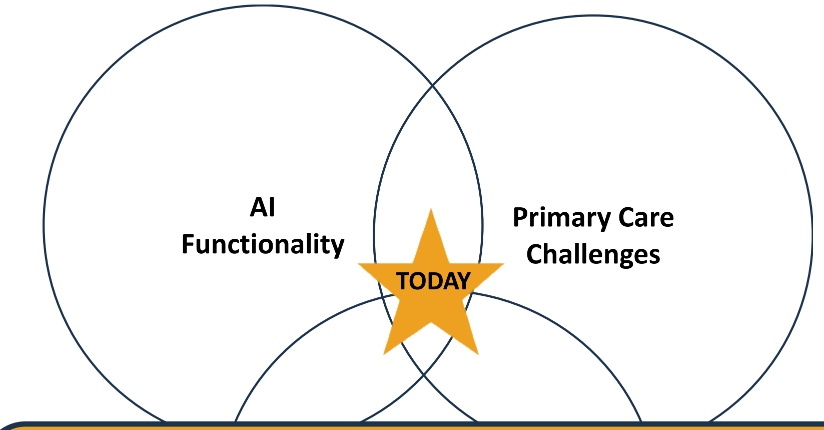
Jean Feng, Ph.D.,[1] Fan Xia, Ph.D.,[1] Karandeep Singh, M.D.,[2] and Romain Pirracchio, Ph.D., M.D.[1]

EDITORIAL

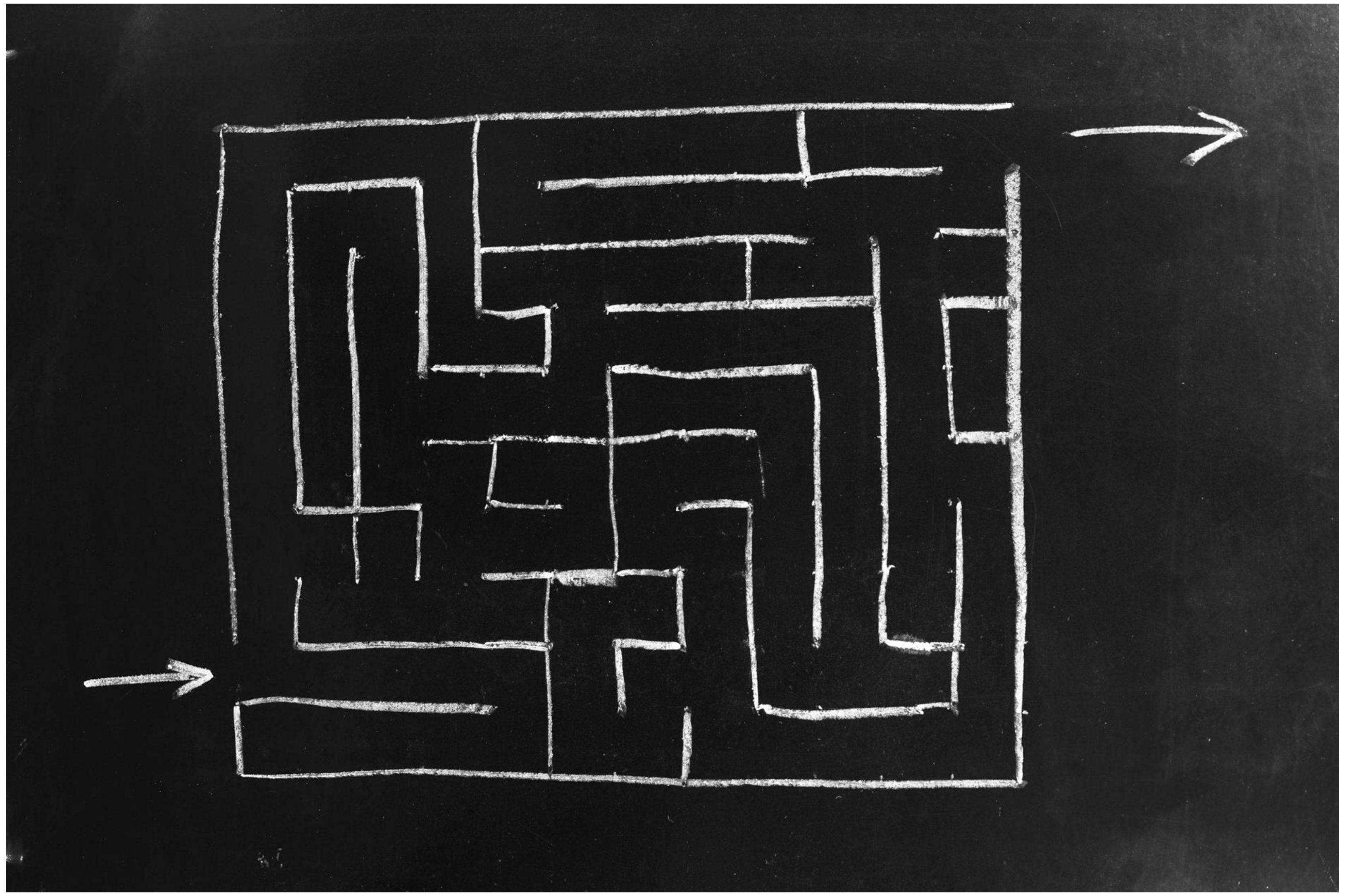
MAYO CLINIC PROCEEDINGS: DIGITAL HEALTH
MAYO CLINIC

Complexities and Questions Toward Artificial Intelligence for Diagnostic Support in Virtual Primary Care

Summary



Benefit: AI systems can now perform meaningful tasks for primary care



Challenge: Evaluation and monitoring strategies to inform AI tool selection and maintenance over time

Longterm Positive Impact

?

Opportunity: Interdisciplinary primary care research teams have a lot to offer in this space!

NAPCRG

Contact:

in /jakykueper

AI in Primary Care

Teaching and Learning

Karim Hanna, MD, FAAFP, FAMIA

NAPCRG

We Are Teachers.

- Benefit: Personalized, On-Demand Learning
- Challenge: The Risk of Misinformation and Overreliance

Board Prep

2024, Volume 56, Issue 9, 555-560, e-ISSN 1938-3800

Family Medicine

ORIGINAL ARTICLE

Performance of Language Models on the Family Medicine In-Training Exam

Rana E. Hanna, BS[a]; Logan R. Smith, BA[a]; Rahul Mhaskar, PhD[b]; Karim Hanna, MD[a,c]

Results: ChatGPT 4.0 scored 167/193 (86.5%) with a scaled score of 730 out of 800. According to the Bayesian score predictor, ChatGPT 4.0 has a 100% chance of passing the family medicine board exam. ChatGPT 3.5 scored 66.3%, translating to a scaled score of 400 and an 88% chance of passing the family medicine board exam. Bard scored 64.2%, with a scaled score of 380 and an 85% chance of passing the boards. Compared to the national average of postgraduate year 3 residents, only ChatGPT 4.0 surpassed the residents' mean of 68.4%.


WE ARE
TEACHERS

npj | digital medicine

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › npj digital medicine › articles › article

Article | Open access | Published: 18 March 2025

Preliminary analysis of the impact of lab results on large language model generated differential diagnoses

Balu Bhasuran, Qiao Jin, Yuzhang Xie, Carl Yang, Karim Hanna, Jennifer Costa, Cindy Shavor, Wenshan Han, Zhiyong Lu & Zhe He

npj Digital Medicine **8**, Article number: 166 (2025) | Cite this article

were created, incorporating demographics, symptoms, and lab data. Five LLMs—GPT-4, GPT-3.5, Llama-2-70b, Claude-2, and Mixtral-8x7B—were tested to generate Top 10, Top 5, and Top 1 DDx with and without lab data. Results show that incorporating lab data enhances accuracy by up to 30% across models. GPT-4 achieved the highest performance, with Top 1 accuracy of 55% (0.41–0.69) and lenient accuracy reaching 79% (0.68–0.90). Statistically significant


OVERRELIANCE
ON AI

Patient Learning

Table 2. Understandability and Actionability of Different LLM Responses Based on PEMAT-Q Rubric

	ChatGPT		Google Bard		Hugging Chat		Claude 2	
	Understandability	Actionability	Understandability	Actionability	Understandability	Actionability	Understandability	Actionability
Hypertension	67%	40%	33%	40%	67%	40%	67%	20%
Hyperlipidemia	67%	40%	56%	20%	67%	60%	67%	60%
Type 2 diabetes	67%	40%	44%	20%	56%	20%	67%	20%
Hypothyroidism	67%	40%	44%	20%	56%	20%	78%	60%
GERD	67%	20%	44%	20%	56%	20%	67%	40%
Atherosclerosis	67%	20%	67%	20%	56%	20%	56%	20%
Vaccination	78%	20%	78%	20%	44%	20%	78%	20%
Average	69%	31%	52%	23%	57%	29%	69%	34%
Standard deviation	4%	11%	16%	8%	8%	16%	8%	19%

Abbreviations: PEMAT-Q, Patient Education Materials Assessment Tool Question; GERD, gastrointestinal reflux disease.

PRiMER
peer-reviewed reports in medical education research
H

LEARNER RESEARCH

Artificial Intelligence-Prom

Mafaz Kattih | Max Bressler | Logan R. Smith | Ant

PRiMER. 2024;8:51.
Published: 9/17/2024 | DOI: 10.22454/PRiMER.2024.91608

Thank You!

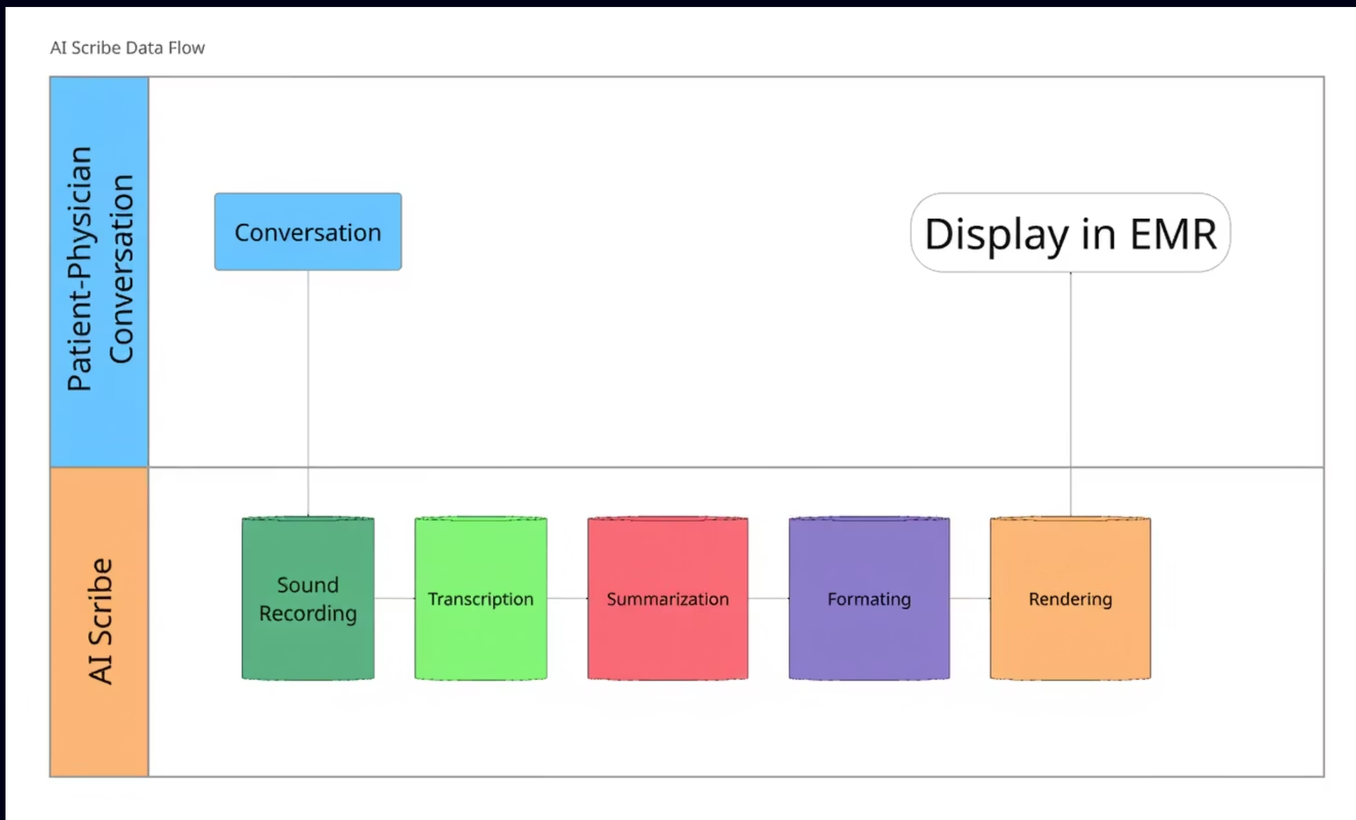
MedEd+AI

AI in Primary Care: Good governance is a challenge

Balancing innovation with governance in clinical settings

Karim Keshavjee MSc, MD, MBAAssistant Professor & Program DirectorMaster of Health InformaticsDalla Lana School of Public HealthUniversity of Toronto, Canada

NAPCRG


AI Scribe Data Flow
Patient-Physician Conversation
Conversation
Display in EMR
AI Scribe
Sound Recording
Transcription
Summarization
Formating
Rendering

The Governance Gap

Small practices lack hospital resources

No legal counsel, privacy officers, or data committees

Clinicians as gatekeepers

Responsible without proper tools or authority

Equity concerns

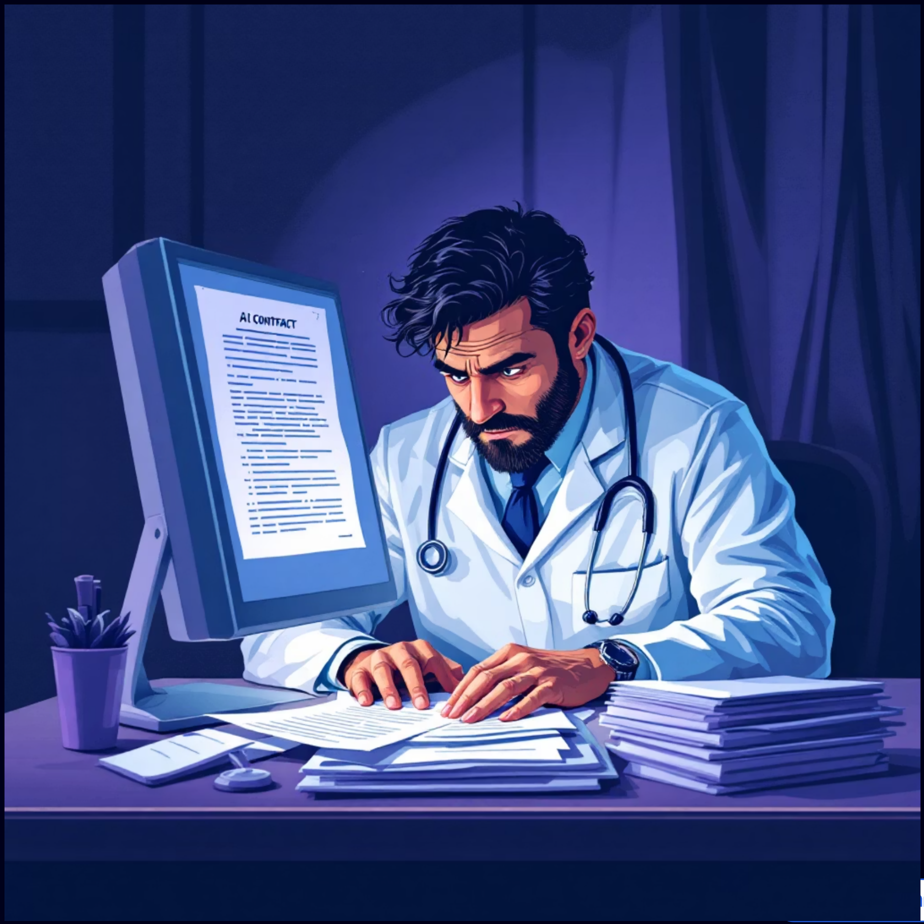
Rural and under-resourced areas face heightened risks

Key Challenge: Risk Without Support

Clinicians remain legally responsible for AI they cannot properly evaluate

AI reaches clinics through informal channels

Oversight is patchy, risk unmanaged



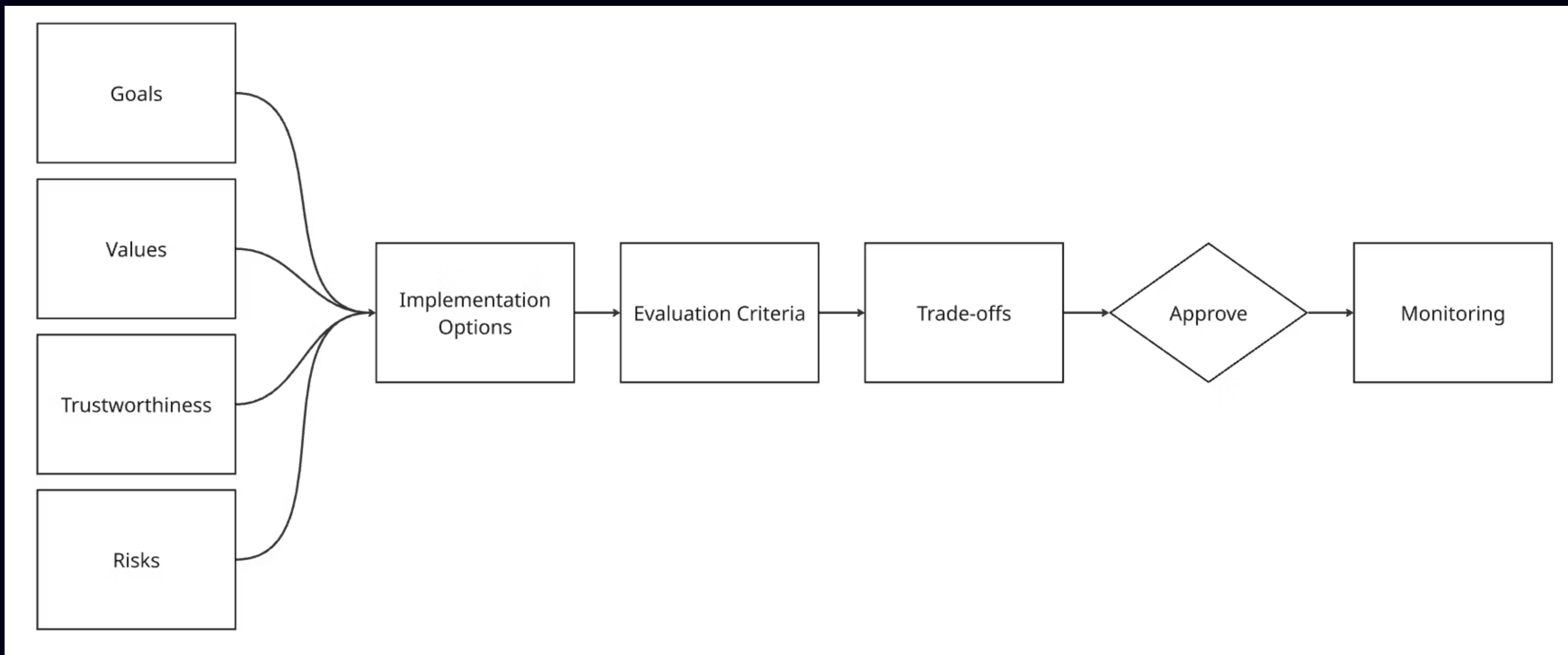
Proposed Solution: Two-Layer Governance

National AI Committee

Certifies tools, sets guidelines, maintains registry

Local AI Governance Boards

Reviews implementation, assesses workflow fit, identifies local risks


Goals
Values
Trustworthiness
Risks
Implementation Options
Evaluation Criteria
Trade-offs
Approve
Monitoring

Key Benefit: Enhanced Patient Care

Sharper Diagnosis

AI improves diagnostic accuracy

Reduced Paperwork

Automated documentation saves time

Tailored Treatment

Personalized care recommendations

The Path Forward

Sustainable Implementation

- Risk-matched governance
- Vendor-paid certification fees
- Standardized education