HEALTH
UNIVERSITY OF UTAH

TECHNICAL CAPABILITIES AND CONSIDERATIONS FOR THE USE OF AI FOR SSA DISABILITY DETERMINATION

AI AND THE MEDICAL RECORD IN THE CONTEXT OF SOCIAL SECURITY DISABILITY EVALUATIONS

APRIL 7, 2026

KENSAKU KAWAMOTO, MD, PHD, MHS, FACMI, FAMIA

THE DR. HELMUTH F. ORTHNER ENDOWED PROFESSOR &
VICE CHAIR FOR CLINICAL AI AND INFORMATICS,
DEPARTMENT OF BIOMEDICAL INFORMATICS
ASSOCIATE CHIEF MEDICAL INFORMATION OFFICER
DIRECTOR, UNIVERSITY OF UTAH HEALTH INNOVATION LAB
DIRECTOR, REIMAGINE EHR INITIATIVE
CO-SENIOR DIRECTOR, DIGITAL HEALTH INITIATIVE

REIMAGINEEHR

DISCLOSURES

- I report honoraria, consulting, sponsored research, licensing, or co-development in the past 24 months with Hitachi, Pfizer, Beckman Coulter, NORC, RTI International, Surescripts, University of Pennsylvania, Yale University, MD Aware, Custom Clinical Decision Support, and the U.S. Office of the National Coordinator for Health IT (via Security Risk Solutions)

SOURCE OF PERSPECTIVES

- **ReImagine EHR:** multi-stakeholder initiative focused on integrating digital health innovations into the electronic health record (EHR)[1]
  - >15 solutions, >$60M in grants, >10,000 local clinician consultations/month
  - Multi-institutional deployments, both patient- and provider-facing tools[2]
- **Health System Innovation Lab:** enterprise innovation hub leveraging AI to catalyze transformational changes across the missions
  - Launched January 2026, with multiple tools already in production use
- **Engagement in standards development and stewardship**
  - Co-chair of Health Level Seven International (HL7) Clinical Decision Support Work Group since 2009
  - Two-term service on US Health IT Advisory Committee
  - Initiative Coordinator for ONC and CMS efforts to develop and harmonize interoperability standards for decision support and clinical quality measurement

1. Kawamoto K et al. *JAMIA Open*. 2021 Jul 31;4(3):ooab041. 2. e.g.: Kawamoto K et al. *JAMA Netw Open*. 2019 Nov 1;2(11):e1915343; Kukhareva PV et al. *JAMA Netw Open*. 2024 Jun 3;7(6):e2415383; Kukhareva PV et al. *JAMA Oncol*. 2026 Feb 1;12(2):167-176.

KEY CAPABILITY 1: HL7 FHIR AND SMART[1]

- **HL7 Fast Healthcare Interoperability Resources (FHIR)[2]**
  - Standard for sharing health data
- **HL7 Substitutable Medical Apps Reusable Technologies (SMART)[3]**
  - Standard for integrating third-party digital tools with the EHR
- Key federal requirements
  - Certified EHRs must support the US Core FHIR profiles[4]
  - Health systems *must* allow patients to access their EHR data via digital tools of their choice

1. Strasberg HR et al. *J Am Med Inform Assoc*. 2021 Jul 30;28(8):1796-1806.
2. https://www.hl7.org/fhir/
3. Mandl KD et al. *J Am Med Inform Assoc*. 2012 Jul-Aug;19(4):597-603.
4. https://build.fhir.org/ig/HL7/US-Core/

KEY CAPABILITY 2: LARGE LANGUAGE MODELS (LLMS)[1]

- **LLMs:** artificial intelligence (AI) models trained on vast amounts of text data to understand, summarize, generate, and predict human-like language
- E.g., OpenAI GPT, Google Gemini, Anthropic Claude
- There are open-source LLMs that are highly capable (e.g., OpenAI gpt-oss-120b[2]), and which enable local LLM hosting
- While earlier LLMs often had significant challenges for use in healthcare (e.g., significant hallucination rates), many of these challenges have been significantly mitigated in recent months (e.g., via reasoning models[3])

1. Thirunavukarasu AJ et al. *Nat Med*. 2023 Aug;29(8):1930-1940.
2. https://openai.com/index/introducing-gpt-oss/
3. https://developers.openai.com/api/docs/guides/reasoning

KEY CAPABILITY 3: TRANSFORMER-BASED PREDICTIONS

- Transformer architecture: same underlying approach to AI that has given rise to LLMs[1]
- Enables rapid development of high-performing prediction models
- Example using 10 years of data (result of 2 days of effort):

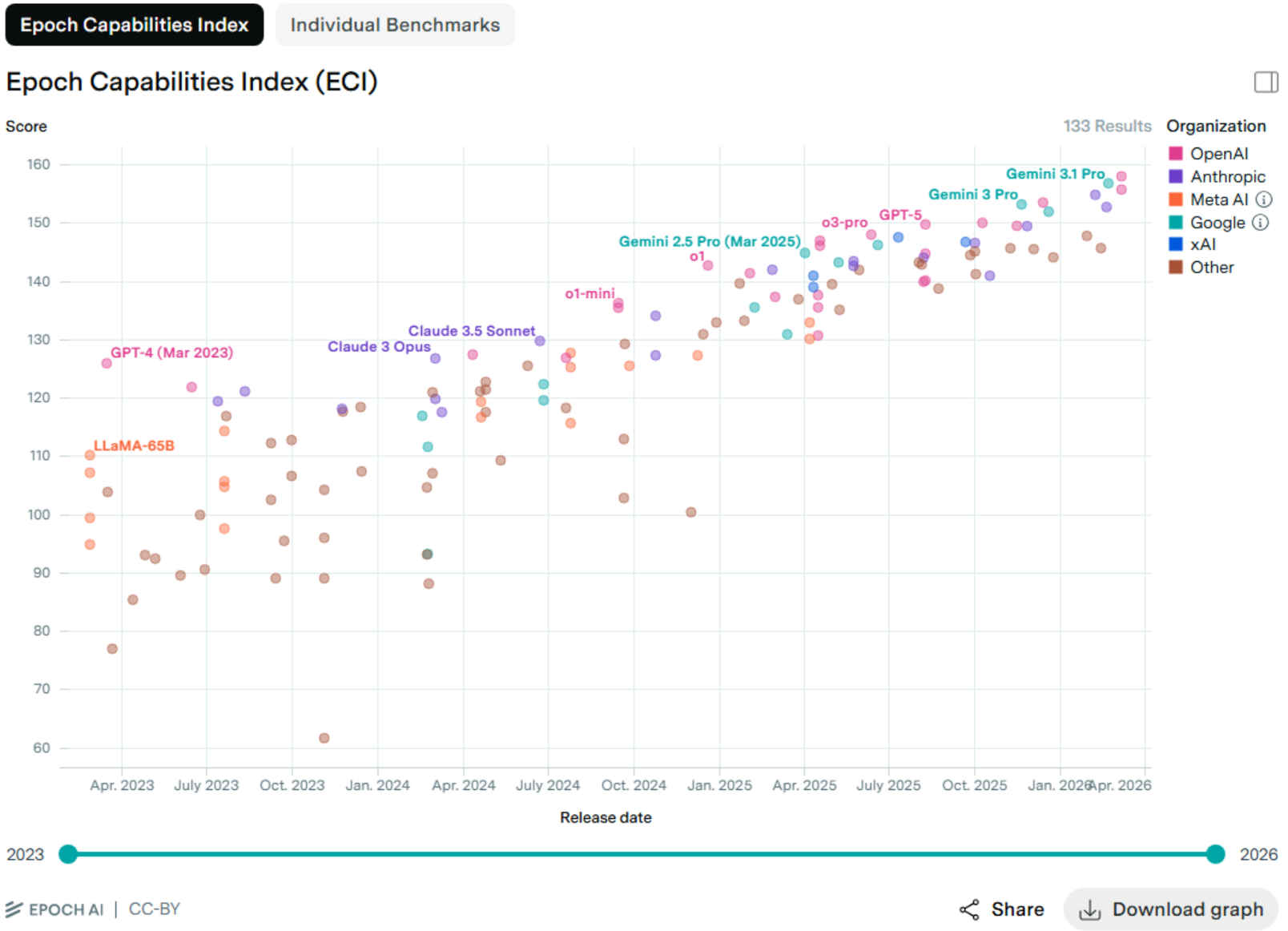
Prediction Target	Prototype Model	Benchmark[2]
30-day readmission	85.3% AUC	79.9%
Length of stay	82.8% AUC	78.7%
In-hospital mortality	96.4% AUC	94.9%

AUC = Area Under the Curve

1. Vaswani A et al. 2017. Advances in neural information processing systems. 5998-6008.
2. Jiang LY et al. *Nature*. 619, 357–362. 2023.

CONSIDERATION 1: RAPID PACE OF CHANGE

- AI is evolving rapidly; what was true recently may be obsolete now
- Staying up-to-date requires monitoring non-traditional sources (e.g., X, Reddit) or working closely with those who do



https://epoch.ai/benchmarks

CONSIDERATION 2: COMPUTE COST

- Many healthcare use cases require the processing of a lot of data
  - E.g., processing years of patient data, including clinical notes, to generate documentation required for SSA disability determination
- This type of processing can get very expensive
- Current approaches to reduce costs include the following:
  - Use smaller models
  - Reduce the amount of data requiring processing
  - Reduce the processing conducted (e.g., degree of double/triple-checking)
  - Run inferences at off-peak times using spare/less expensive compute
  - Use local compute vs. cloud-based computational resources
- Upcoming, next-generation AI servers may provide an up to 10x decrease in inferencing costs[1]

1. https://nvidianews.nvidia.com/news/rubin-platform-ai-supercomputer

CONSIDERATION 3: UNPRECEDENTED SPEED TO SOLUTION

- Problems that used to take large teams years to tackle, or were simply not solvable, can now be developed by small teams in months or even days
  - Leveraging the various capabilities discussed
- The University of Utah Health Innovation Lab has developed prototypes for over a dozen use cases in the last few months, with several of the tools already in production use following validation
- Example on next slide: prototype AI patient intake tool developed and deployed in < 12 hours

AI PATIENT INTAKE PROTOTYPE:

https://tolven.chpc.utah.edu:8443/

Username: **rehr**

Password: **aidoctor2078**

SUMMARY

- Key capabilities
  - HL7 FHIR and SMART
  - LLMs
  - Transformer-based predictions
- Key considerations
  - Rapid pace of change
  - Compute cost
  - Unprecedented speed to solution
- Many opportunities and considerations for leveraging AI for SSA disability determinations

THANK YOU!

Kensaku Kawamoto, MD, PhD, MHS, FACMI, FAMIA

The Dr. Helmuth F. Orthner Endowed Professor &
Vice Chair for Clinical AI and Informatics,
Department of Biomedical Informatics
Associate Chief Medical Information Officer
Director, University of Utah Health Innovation Lab
Director, ReImagine EHR Initiative
Co-Senior Director, Digital Health Initiative

kensaku.kawamoto@utah.edu

AI and the Medical Record:

how health systems validate, monitor and govern AI in real-world deployment

Vincent Liu, MD MS
Chief Data Officer, The Permanente Medical Group/KP
Senior Research Scientist

KAISER PERMANENTE®

The EHR is becoming a ‘human + AI’ document

- Gen AI use cases are rapidly expanding in healthcare
- Ambient AI scribes – technology that listens, synthesizes, and drafts notes – are used in ~2/3rds of hospitals using Epic EHRs[1]
- KP has used ambient AI in >20 million encounters and saved >15K hours in documentation time[2]

“I use it for every visit…and it’s making my visits better. I even had a patient praise the fact that I could listen instead of type.”

-KP Physician

KAISER PERMANENTE®

[1]Yang et al, AJMC 2026
[2]Tierney et al, NEJM Catalyst, 2025

What health AI is being deployed today

AI TYPE	CLINICAL	DIRECT-TO-CONSUMER	BUSINESS OPERATIONS	HYBRID
Description	AI tools used by a clinician to support diagnostic or treatment decisions	AI tools used by a person with health or wellness concerns +/- engaging any healthcare system or professional	AI tools used by a health care system or professional to optimize aspects of care delivery	AI tools that serve multiple purposes (e.g., clinical and business operations)
Opportunity	• Improve patient outcomes • Better access to care • Improved Dx accuracy • Compliance with guidelines • Enhanced risk stratification • More personalized care	• Improved dz recognition & management • Helping people live healthier lifestyles • More prompt, personalized, and/or less costly access to care	• Automating labor-intensive processes • Reduced administrative burden • Reduced waste • Improved revenue generation	• Reduced administrative burden for HCP • Helping HCP execute care tasks • Helping patients navigate healthcare delivery options
Examples	• Imaging AI for autonomous screening for diabetic retinopathy • Embedded software in portable heart ultrasound • EHR-based algorithms for sepsis alerts and treatment prompts	• Smartphone app that evaluates skin conditions • Chatbot that offers mental health wellness support • Algorithms using biosensor data to detect falls or arrythmias	• Algorithm that uses EHR data to optimize coding • Software to optimize supply chain management • Software to optimize patient, staff, or OR scheduling	• Ambient AI • LLMs to reply to patient secure messages • Web-based navigators to help patients schedule appointments based on topic and triage

KAISER PERMANENTE®

[1]JAMA, 2026: “AI, Heath, and Heath Care Today and Tomorrow”

How health systems evaluate AI: Before go-live

”Is AI the right solution for the problem?”

- 3-4% of inpatients on the general hospital ward deteriorate or die unexpectedly
- Patients often give off early signals that their risk of unexpected deterioration is rising
- EHR & streaming data offer an earlier opportunity to unmask rising patient risk

”How does AI perform in a test environment?”

- Using retrospective data incorporating millions of data points, AI demonstrates capabilities to detect rising risk up to 12h before adverse event
- The EDIP model (early detection of impending deterioration) showed acceptable sensitivity, PPV, flag percentages based on estimated risk thresholds

”How does AI perform in silent mode?”

- In silent/live mode, data quality issues are uncovered, checked, and addressed
- Real-time algorithmic performance is validated and risk thresholds are selected
- Alert timing is moved from every 6h calculation to hourly calculation
- Model card generated

”How does AI fit within a workflow?”

- User interfaces are built that maximize value to clinicians, and minimize extraneous information
- Playbooks developed
- Standardized approach to patient Goals of Care
- Virtual RN 1st responder system is developed with local outreach
- Advance Alert Monitor (AAM) is born

KAISER PERMANENTE®

How health systems evaluate AI: After go-live

”What is AI user and workflow adoption? ”

- Dashboards track alert responses, including snooze features, virtual-local communication, and documentation standards
- Iterative refinements are made in the alert interface, thresholds, and workflows to maximize benefit and reduce alert fatigue

”What is the impact of AI in outcomes?”

- Stepped wedge rollout of AAM across 19 non-pilot hospitals
- Analysis of >500,000 hospitalizations demonstrated a 16% decrease in the risk of 30-day mortality[1]
- AAM saves >500 lives/year across 21 hospitals of KP Northern California

”How does AAM performance hold up?”

- Periodic re-assessments algorithm performance to confirm that it continues to identify high-risk patients
- Periodic assessments of equity metrics, evaluating performance across subgroups
- Model card updated

”How do we decide on AAM lifecycle?”

- Ongoing performance evaluations of other EWS systems to capture potential for improvements via AI
- Linkage of AAM workflows to other AI tools to make the entire program more effective – (i.e., genAI to rapidly summarize patient’s chart)

KAISER PERMANENTE®

[1]Escobar et al, NEJM 2020

How does this process work for genAI?

GENAI EXAMPLE	IS AI THE RIGHT SOLUTION?	PRE-GO-LIVE TESTING	POST-GO-LIVE MONITORING	LIFECYCLE
Ambient AI scribes[1]	Documentation burden is profound and strongly associated with clinician burnout	Small pilots with tight feedback loops; Expanded pilots based on various specialties	Independent quality checks of notes using mPDQI-9; In-line feedback mechanisms for real-time feedback; Patient/clinician surveys	KP Northern California has already pivoted between 2 ambient AI providers
Secure message routing[2]	The growth of patient messages via EHR have become overwhelming to clinicians; message routing can improve turnaround time and reduce inbox volume	Validation of NLP/LLM approaches to identify the content of messages for routine; Linkage of paired workflows to address specific message topics	Ongoing performance metrics of routing procedures; iterative improvements in algorithms; development of new routing pathways	KP Northern California is now using 2 AI solutions in tandem as new tools become available
Chart summarization	EHR Information retrieval is challenging and leads to additional cognitive burden	Iterative refinement of genAI prompts and customization; small pilots by Dept Tech Leads across diverse use cases focused on identifying weaknesses	In-line feedback mechanisms for real-time feedback; Dashboard monitoring for usage patterns; Existing technology feedback mechanisms	Recently releasee functionality

KAISER PERMANENTE®

[1]Tierney, NEJM Catalyst 2025; [2]Liu, JAMA Network Open 2023

Why GenAI is harder to monitor at scale

- GenAI is stochastic, not deterministic
- GenAI models are rapidly changing and updating
- Output quality depends on prompt variation and context
- There is a large universe of potential context and edge cases
- ‘Ground truth’ is often ambiguous and can be subjective
- Need multi-layered approach to monitoring and evaluation
  - End-user training and clinician autonomy & agency
  - In-line feedback mechanisms, feedback channels
  - Intermittent independent audits of a sample: human and LLM as a judge
  - Human in the Loop oversight; adversarial testing
  - Prompt, safety, quality, equity monitoring

Practical approaches for AI-assisted EHR data

- Any approach needs to be pragmatic, not burdensome
- Approaches should also be tiered based on potential risk

    - **Lower:** simple labeling indicating AI-assistance
    - **Moderate:** more detailed information on the AI tool/provenance
    - **Higher:** richer provenance, human review/editing trace

- Current practice is that clinicians are still reviewing every draft and editing, as appropriate

Thank you!

vincent.x.liu@kp.org