Why AI Is Good for Our Health but May Hurt Our Wallets

Sneha S. Jain*, Stanford University School of Medicine, Morgan Cheatham*, Bessemer Venture Partners, Michael A. Pfeffer, Stanford University School of Medicine, Linda Hoff, Stanford Health, Nigam H. Shah, Stanford University School of Medicine

*These authors contributed equally to the work as co-first authors

Contact: snehashahjain@stanford.edu

Abstract

What is the message? Current regulatory frameworks, reimbursement structures, and business models for AI in healthcare are decoupled, which has created an environment in which AI may significantly increase costs without necessarily improving outcomes. This misalignment stems from inadequate regulatory and business incentives for real-world performance evaluation of AI as well as reimbursement gaps that lead to pricing strategies prioritizing financial gains over improved quality and value to recoup development costs. The authors recommend three key reforms: rigorous pre- and post-deployment evaluation to verify proposed clinical value, development of assessment standards through shared guidelines, and strategic alignment of AI deployment modalities with sustainable business models to ensure these tools enhance care quality while responsibly managing healthcare costs.

What is the evidence? This paper cites over 50 sources, including academic literature, Food and Drug Administration documents and policy, written statements from the Centers for Medicare and Medicaid Services, and industry reports.

Timeline: Submitted: October 19, 2024; accepted after review November 1, 2024.

Cite as: Sneha Jain, Morgan Cheatham, Michael A. Pfeffer, Linda Hoff, Nigam H. Shah.

Listen to an AI-generated podcast with co-author Nigam Shah here

Introduction and Overvie

In 1995, Charlie Munger said "Show me the incentives, and I'll show you the outcome." Thirty years later, this holds true for artificial intelligence (AI) in healthcare. Current approaches to evaluating, regulating, and paying for AI in healthcare incentivizes use of AI in a manner that is likely to increase total cost of care. The problem stems from disconnected regulation and reimbursement approval decisions, anemic health information technology (IT) budgets, and complex revenue structures across stakeholders necessitating creative business models to subsidize adoption of AI tools. Often these business models are crafted independently of, and

without, the necessary workflow redesign to capture the full potential of AI[1].

Three federal agencies, the Food and Drug Administration (FDA), the Office of the National Coordinator for Health IT (ONC), and the Centers for Medicare & Medicaid Services (CMS), regulate AI tools in healthcare. These tools are regulated as software as a medical device

(SaMD)[2], practice-management software under Health Data, Technology, and Interoperability

Certification Program (HTI-1)[3], and the Clinical Laboratory Improvement Amendments (CLIA).

There is also software as a treatment modality, or digital therapeutics (DTx)[4], which are subject to safety and efficacy evaluations. Additionally, other software and tech-enabled services, such as revenue cycle management, are procured directly by healthcare entities outside of regulatory purview and are only subject to enforcement by the Federal Trade Commission (FTC) or the Office of Civil Rights (OCR).

This patchwork of regulatory frameworks leads to the inconsistent evaluation of the clinical impact of AI tools. For example, Wu et al., found that most FDA-cleared medical AI devices were

evaluated pre-clearance through retrospective studies, with many lacking reported number of evaluation sites or sample sizes[5]. El Fassi et al. found that almost half of authorized tools were not clinically validated and were not even trained on real patient data, concluding that FDA authorization is not a marker of clinical effectiveness[6]. In general, it is widely accepted that robust AI testing and validation infrastructure in medicine is lacking[7] and our regulatory regimens need to be updated.

Regulatory approvals examine if AI tools "work", but not whether they create "value" in the form of better quality of care for patients relative to the cost, which is often considered in procurement and reimbursement decisions. They also do not take into consideration how an AI tool will fit into existing or new workflows. This *decoupling* between regulatory approval and reimbursement requires users of AI tools – especially those tools that are used to render medical care – to figure out how to pay for the cost incurred by using a tool based on the value obtained.

Lobig et al. recommend that reimbursement for an AI tool, if separate from the cost of the underlying imaging study, should be decided based on evidence of improved societal outcomes [8]. However, for regulated tools, the assignment of value to a reimbursed AI tool is artisanal at best [9]. Payment rates differ significantly between private vs public payers. For example, Wu et al. found that reimbursement for CPT code 92229 for diabetic retinopathy is approximately 2.8 times higher for private patients than for CMS patients [9]. There is little consistency in how reimbursement for AI tools used for medical care is valued compared to the non-AI alternatives. For example, reimbursement for AI-based interpretation of breast ultrasound is comparable to a traditional breast ultrasound, whereas the reimbursement for AI-powered cardiac CT for atherosclerosis is two to three times the out-of-pocket cost of such a study[9,10]. While mechanisms exist to facilitate reimbursement during the nascent stages of technology adoption, such as New Technology Add-On Payment (NTAP) [11] and Transitional Coverage for Emerging Technologies (TCET), these solutions are unlikely to fully accommodate the rapid growth of AI-based tools in healthcare [12].

For non-regulated technology, adoption depends on market forces to identify high-value solutions, and incumbent vendor platforms in facilitating their use, which may vary by care

setting and reimbursement (i.e., urban vs. rural, fee-for service vs. value-based payment).

Therefore, as Davenport and Glaser note, despite abundant research and startups, very few AI tools have been adopted by healthcare organizations [13]. They attribute this to factors such as regulatory approval, reimbursement, return on investment, integration challenges, workforce education, the need for changing workflows, and ethical considerations, and conclude that new organizational roles and structures will be necessary to successfully adopt these technologies. Many of these challenges stem from needing to pay the hidden deployment costs of AI tools [14].

To address these challenges, Adler-Milstein et al. emphasize the need to couple the creation of equitable tools, their integration into care workflows, and training of health care providers with strong regulatory oversight and financial incentives for adoption in a way that benefits patients [15].

Whether an AI tool is regulated or not, the developers (and users) currently have to conform to existing payment methods for *technology* or *medical care.* Thus, paying for technology ends up being a net-new cost to health IT budgets. For example, a per-user license for ambient scribe technology is not a form of direct 'medical care', and hence brings no new revenue to a provider. Existing ways to pay for the tools as 'medical care' – while having the potential to bring new revenue – is fraught with value judgments and is still a net-new cost to the payers. As a result, adoption of AI tools remains low compared to the hype around them. For example, Wu et al. find that even though the number of devices cleared under the FDA's SaMD exceeds over 500, only two of them – for assessing coronary artery disease and for diagnosing diabetic retinopathy – had over 10,000 CPT claims reimbursed in a four-year period [9].

The lack of suitable payment models for health AI tools has led to prioritizing solutions that offer financial over clinical benefits [16]. In addition, AI developers face high costs driven by compute, data, and large enterprise healthcare sales teams. However, IT budgets are not large enough to sustain the payback assumptions made when investing in the creation of AI tools[17]. The total health IT spend in the US is approximately $46 billion, and approximately 10% of this spend is captured by leading electronic health record vendors [16,18] which does not include the hardware and people needed to run them. The remaining budget includes software (both clinical and

business systems), medical devices, imaging equipment, hardware and networking components, cybersecurity, and salaries for IT personnel, leaving little room to pay startups or incumbents creating AI tools. This results in immense pressure to find non-IT budgetary spend (such as re-allocating salaries and professional services) and for alignment with the way medical care is paid for. This tension has prompted a reevaluation of traditional business models—the overarching strategies companies use to create, deliver, and capture value from their solutions such as software or services—and pricing models, which describe the specific mechanisms by which vendors monetize their solutions, such as recurring subscription fees, pay per use, or

contingency-based pricing [19].

Given these constraints, health AI vendors are developing pricing strategies that align with existing payment paradigms for technology or for medical care. These strategies attempt to cover the high upfront costs of AI implementation, the long-term additive costs associated with 'augment the human' design paradigms, and the uncoupling of technology users from who pays. The challenge becomes clear when we cross-tabulate the pricing strategies (usage- or performance-based) with the payment paradigm (for technology or medical care), as shown in the table below.

Table. Pricing Strategies for Health AI Tools. *Examples of technology and medical AI tools that use either usage-based or performance-based pricing strategies.*

	Technology	Medical
Usage-based	E.g. Ambient scribe	E.g. AI-based screening for diabetic retinopathy
Performance-based	E.g. AI-powered coding and clinical documentation integrity	E.g. Algorithm guided post-acute management

Usage-based pricing charges customers based on volume of utilization of an AI tool. The payment can be a direct payment by the customer (e.g., for a per-user license for ambient scribe) or via reimbursement (e.g., CPT code 92229 for AI-based screening for diabetic retinopathy). The first adds net-new cost to the IT budget while the second generates new revenue for providers but increases costs for payers and carries the risk of overuse via

unnecessary screenings. When AI tools, like ambient scribes, do not generate revenue directly, costs are justified by indirect benefits, such as reduced physician burnout and potentially lower turnover, or downstream benefits such as better documentation for billing. However, in other instances, the expectation is that costs will be covered by having users of the technology see more patients in the time saved. If time is saved, physicians must decide whether to add a patient to their schedule or to keep their normal case load in hopes of providing better care[20]. In some cases, patients bear the cost via out-of-pocket fees, such as with AI-enhanced mammography interpretations [21]. Usage-based pricing can create conflicting incentives, with vendors promoting increased utilization to boost revenue while their customers may limit usage arbitrarily to control costs.

Performance-based pricing distributes financial risk between AI developers and healthcare providers or payers, with payments tied to measurable outcomes (not necessarily *clinical outcomes*). Risk-sharing arrangements range from a base annual fee plus a share of the financial savings to payment solely from created financial savings. For example, an AI-augmented screening program to detect worsening heart failure (HF) may allow for early intervention by the care provider, reducing readmissions[22]. There can be a base fee for access to the software with some percentage of additional revenue to the AI vendor generated by reducing a hospital's readmission rates, and therefore decreasing the associated Medicare reimbursement penalty. Similarly, an AI-powered coding and clinical documentation integrity system can analyze clinical documentation to suggest appropriate diagnostic codes such as Risk Adjustment Factor (RAF) coding and increase compliance by identifying clinical documentation that may be insufficient to support accurate coding for billing. There can be a base fee for access to the software plus some percentage of additional revenue to the AI vendor generated from the improved coding accuracy.

Performance-based pricing has limitations. A vendor may be incentivized to assign codes that reflect higher illness severity than clinically justified, a problem called 'upcoding' that is often reported with tools used by payers offering Medicare Advantage plans[23]. For example, FDA-cleared screening tools were allegedly used to add diagnosis codes to a patient record, even when no further care was rendered [24]. Performance-based pricing, when used to pay for AI-tools that decide on access to post-acute care based on patient needs[25], may override clinicians'

judgment and deny care for seniors to generate 'cost savings' in Medicare Advantage plans[26]. A recent ProPublica investigation reported on the use of an algorithm backed by AI, which can be adjusted to lead to higher denials with the promise of saving $3 for every $1 spent on its use[27].

A Path Forward

The core issues with health AI tools stem from their inadequate evaluation and their decoupled regulatory and reimbursement approval criteria. These challenges are compounded by how technology and medical care are currently paid for. The resulting pricing arrangements reflect the machinations currently necessary to get paid for the use of AI in healthcare either via IT budgets or aligning with either fee-for-service or value-based care paradigms. However, these pricing paradigms reveal a concerning trend: AI can increase the total cost of care without improving healthcare quality, or worse, lead to care denials and possible care disparities in the quest to create "financial savings". To navigate this situation, we make the following suggestions:

Conduct assessments to specify and verify benefits

Regardless of whether an AI tool is under FDA, ONC, or CMS regulation, before adoption, it is necessary to ensure that the use of the tool improves healthcare quality, either through more efficient operations, improved patient experience, or enhanced patient outcomes[28,29]. We need robust estimates of benefits prior to deployment, and then verification of that benefit after deployment, in order to ensure that the use of AI tools improves overall value. For example, healthcare systems can put in place local evaluation regimens to ensure that the use of AI tools is fair, useful, reliable, and monetarily sustainable [28]. An upfront ethics evaluation to assess for unintended consequences is critical to avoid some of the situations detailed in the examples above[30]. Additionally, impact assessments should examine financial sustainability for addressing the disconnect between regulation and current reimbursement for clinical AI tools [31]. In situations where there is no short-term financial benefit – typically defined as return on investment in one year or less – there may be intangible benefits such as improved provider wellness[32] or better long term patient outcomes. If estimated, these can form the foundation for advocating for the adoption of certain AI products. At the minimum, an upfront evaluation of

fairness, usefulness, reliability, and monetary sustainability can prevent organizational waste in the form of *pilotitis* – where hundreds of pilot projects happen and none convert to a broad implementation [33,34]. Finally, given that hundreds of health AI tools have been approved on the basis of limited clinical data[35], there is a related urgent need to institute ongoing evaluation of health AI tools that are already in use[36].

Create consensus on mechanisms of transparent evaluation

Given that value from the use of an AI tool is notoriously difficult to define, and is an interplay of a tool's performance with the care workflows in which it is used, it is necessary to evaluate both[37]. The AI tools (or the underlying models) should be subject to certain manufacturing constraints – as is already the case with FDA-regulated AI tools. For those AI tools that are currently not regulated, consensus best practices are needed around the creation, testing, and reporting of AI tools[38,39]. Given that best practices are typically offered in the form of checklists and reporting guides, adherence to them remains challenging[40]. The necessary next step is to facilitate the routine use of these *desiderata* (as well as verification of a vendor's adherence to them), which can initially be done via a nationwide network of assurance labs[41] and can gradually be transitioned into assurance software that is widely shared for self-service use. For example, Epic Systems has already taken the first step in this direction, with two academic groups contributing code to the software[42,43]. Finally, the construct of the tool in the context of its workflow can only be evaluated in the local setting, for which we need to create consensus assurance guidelines [44], shared open-source software [45], as well as communities of practice (such as Health AI Partnership[46] and RAISE [47]) to develop implementation best practices and centers that can evaluate clinical effectiveness [48]. The creation of common, accepted practices can ensure the *evaluation* process of AI is as cost-efficient as possible.

Align AI implementation with appropriate business models

Strategic alignment of AI implementation with the right business model is crucial for cost-effectiveness and value creation in healthcare. The concept of modality-business model-market fit [49], how the choice of the form AI takes and the business model to support it, determines the

value potential of the resulting solution. By selecting appropriate modalities — such as AI-enabled software, copilots, diagnostic or therapeutic tools — and aligning them with appropriate business models, stakeholders can generate value without unnecessarily inflating costs. For example, AI copilots (ambient scribes being an example) integrated into existing workflows and funded by current budgets, can enhance efficiency without requiring significant infrastructural changes. Hospitals that already allocate resources for human scribes can reallocate that budget to IT for a transition to AI alternatives, improving consistency and quality of clinical documentation while operating within established budgets. An AI-agent conducting a post-discharge follow-up workflow[50], or an agent performing medication titration (such as a voice agent managing insulin dosage using data from a continuous glucose monitor[51]), can off-load work from burnt out and overworked nurse care managers, allowing reallocation of time to other tasks. Other modalities, such as the AI-augmented screening tool for heart failure[52], may require the creation of new workflows in a value-based care setting, so that avoidance of later complications is prioritized in a population health setting.

The value created from AI in healthcare will depend on how well we balance technology innovation, the incentives created by the complex payment structures in healthcare, the payback expectations of those investing in technology creation, the business models adopted by AI vendors, measurable clinical benefit, and associated healthcare costs. We must balance appropriate oversight with flexible infrastructure that continues to support innovation. Current business models of AI tools that impact medical care are square pegs in the two round holes by which medical care is paid for – i.e. fee-for-service and value-based payment methods. For pure technology plays, the available IT budgets might be missing a zero or two. To bridge this gap, we need focused efforts to connect high quality evaluation of benefits, business model choice, regulation, and reimbursement for promising, high-value emerging technologies in order to finally achieve the promise of health IT lowering the cost of healthcare.

References

[1]   Mullangi S, Ibrahim SA, Shah NH, Schulman KA. A Roadmap To Welcoming Health Care Innovation. *Health Affairs Forefront*. doi:10.1377/forefront.20191119.155490

[2]   Center for Devices, Radiological Health. Software as a Medical Device (SaMD). U.S. Food

and Drug Administration. Published August 9, 2024.
https://www.fda.gov/medical-devices/digital-health-center-excellence/software-medical-device-samd. Accessed October 14, 2024

[3]   Health Data, Technology, and Interoperability: Certification Program Updates, Algorithm Transparency, and Information Sharing. Federal Register. Published January 9, 2024. https://www.federalregister.gov/documents/2024/01/09/2023-28857/health-data-technology-and-interoperability-certification-program-updates-algorithm-transparency-and. Accessed October 14, 2024

[4]   What is a DTx? Digital Therapeutics Alliance. Published September 15, 2022. https://dtxalliance.org/understanding-dtx/what-is-a-dtx/. Accessed October 14, 2024

[5]   Wu E, Wu K, Daneshjou R, Ouyang D, Ho DE, Zou J. How medical AI devices are evaluated: limitations and recommendations from an analysis of FDA approvals. *Nat Med*. 2021;27(4):582-584. doi:10.1038/s41591-021-01312-x

[6]   Chouffani El Fassi S, Abdullah A, Fang Y, Natarajan S, Masroor AB, Kayali N, et al. Not all AI health tools with regulatory authorization are clinically validated. *Nat Med*. Published online August 26, 2024:1-3. doi:10.1038/s41591-024-03203-3

[7]   Lenharo M. The testing of AI in medicine is a mess. Here's how it should be done. Nature Publishing Group UK. doi:10.1038/d41586-024-02675-0

[8]   Lobig F, Subramanian D, Blankenburg M, Sharma A, Variyar A, Butler O. To pay or not to pay for artificial intelligence applications in radiology. *NPJ digital medicine*. 2023;6(1). doi:10.1038/s41746-023-00861-4

[9]   Wu K, Wu E, Theodorou B, Liang W, Mack C, Glass L, et al. Characterizing the Clinical Adoption of Medical AI Devices through U.S. Insurance Claims. *NEJM AI*. Published online November 9, 2023. doi:10.1056/AIoa2300030

[10] Yetman D. The Cost of a Coronary Calcium Scan on Your Heart. Healthline. Published November 9, 2022. https://www.healthline.com/health/heart/coronary-calcium-scan-cost.

Accessed October 19, 2024

[11] Website. https://www.cms.gov/medicare/payment/prospective-payment-systems/acute-inpatient-pps/new-medical-services-and-new-technologies

[12] Final Notice — Transitional Coverage for Emerging Technologies (CMS-3421-FN). https://www.cms.gov/newsroom/fact-sheets/final-notice-transitional-coverage-emerging-technologies-cms-3421-fn. Accessed October 9, 2024

[13] Davenport TH, Glaser JP. Factors governing the adoption of artificial intelligence in healthcare providers. *Discov Health Syst*. 2022;1(1):4. doi:10.1007/s44250-022-00004-8

[14] Morse KE, Bagley SC, Shah NH. Estimate the hidden deployment cost of predictive models to improve patient care. *Nat Med*. 2020;26(1). doi:10.1038/s41591-019-0651-8

[15] Adler-Milstein J, Aggarwal N, Ahmed M, Castner J, Evans BJ, Gonzalez AA, et al. Meeting the Moment: Addressing Barriers and Facilitating Clinical Adoption of Artificial Intelligence in Medical Diagnosis. *NAM perspectives*. 2022;2022. doi:10.31478/202209c

[16] Healthcare IT Spending: Innovation, Integration, and AI. Bain. Published September 17, 2024. https://www.bain.com/insights/healthcare-it-spending-innovation-integration-ai/. Accessed October 14, 2024

[17] Cahn D. AI's $600B Question. Sequoia Capital. Published June 20, 2024. https://www.sequoiacap.com/article/ais-600b-question/. Accessed October 14, 2024

[18] Bruce G. Epic's revenue over the past 5 years. https://www.beckershospitalreview.com/ehrs/epics-revenue-over-the-past-5-years.html. Accessed October 14, 2024

[19] Website. https://www.mckinsey.com/industries/healthcare/our-insights/generative-ai-in-healthcare-adoption-trends-and-whats-next#/

[20] American Medical Association. Hattiesburg Clinic doctors say ambient AI lowers stress, burnout. American Medical Association. Published August 15, 2024. https://www.ama-assn.org/practice-management/digital/hattiesburg-clinic-doctors-say-ambient-ai-lowers-stress-burnout. Accessed October 14, 2024

[21] RadNet expects to log upward of $18M in revenue from its AI division this year. Radiology Business. Published June 13, 2023. https://radiologybusiness.com/node/239941. Accessed October 9, 2024

[22] GE HealthCare to acquire Caption Health, expanding ultrasound to support new users through FDA-cleared, AI-powered image guidance. https://www.gehealthcare.com/about/newsroom/press-releases/ge-healthcare-to-acquire-caption-health-expanding-ultrasound-to-support-new-users-through-fda-cleared-ai-powered-image-guidance-?npclid=botnpclid. Accessed October 14, 2024

[23] Geruso M, Layton T. *Upcoding: Evidence from Medicare on Squishy Risk Adjustment*. National Bureau of Economic Research; 2015. doi:10.3386/w21222

[24] Ross C, Lawrence L, Herman B, Bannow T. How UnitedHealth turned a questionable artery-screening program into a gold mine. STAT. Published August 7, 2024. https://www.statnews.com/2024/08/07/unitedhealth-peripheral-artery-disease-screening-program-medicare-advantage-gold-mine/. Accessed October 11, 2024

[25] Herman B, Ross C. Buyer's remorse: How a Medicare Advantage business is strangling one of its first funders. STAT. Published March 13, 2023. https://www.statnews.com/2023/03/13/medicare-advantage-plans-artificial-intelligence-select-medical/. Accessed October 11, 2024

[26] Ross C, Herman B. Denied by AI: STAT series honored as 2024 Pulitzer Prize finalist. STAT. Published May 8, 2024. https://www.statnews.com/denied-by-ai-unitedhealth-investigative-series/. Accessed October 11, 2024

[27] Christian Miller T, Rucker P, Armstrong D. "Not Medically Necessary": Inside the Company

Helping America's Biggest Health Insurers Deny Coverage for Care. #creator. Published October 23, 2024. https://www.propublica.org/article/evicore-health-insurance-denials-cigna-unitedhealthcare-aetna-prior-authorizations. Accessed October 23, 2024

[28] Callahan A, McElfresh D, Banda JM, Bunney G, Char D, Chen J, et al. Standing on FURM ground: A framework for evaluating fair, useful, and reliable AI models in health care systems. *NEJM Catal Innov Care Deliv*. 2024;5(10). doi:10.1056/cat.24.0131

[29] Patel MR, Balu S, Pencina MJ. Translating AI for the Clinician. *JAMA*. Published online October 15, 2024. doi:10.1001/jama.2024.21772

[30] Mello MM, Shah NH, Char DS. President Biden's Executive Order on Artificial Intelligence-Implications for Health Care Organizations. *JAMA*. 2024;331(1). doi:10.1001/jama.2023.25051

[31] Jain SS, Mello MM, Shah NH. Avoiding financial toxicity for patients from clinicians' use of AI. *N Engl J Med*. 2024;391(13):1171-1173. doi:10.1056/NEJMp2406135

[32] Garcia P, Ma SP, Shah S, Smith M, Jeong Y, Devon-Sand A, et al. Artificial intelligence-generated draft replies to patient inbox messages. *JAMA Netw Open*. 2024;7(3):e243201. doi:10.1001/jamanetworkopen.2024.3201

[33] Scarbrough H, Sanfilippo KRM, Ziemann A, Stavropoulou C. Mobilizing pilot-based evidence for the spread and sustainability of innovations in healthcare: The role of innovation intermediaries. *Soc Sci Med*. 2024;340. doi:10.1016/j.socscimed.2023.116394

[34] Kuipers P, Humphreys JS, Wakerman J, Wells R, Jones J, Entwistle P. Collaborative review of pilot projects to inform policy: A methodological remedy for pilotitis? *Aust New Zealand Health Policy*. 2008;5. doi:10.1186/1743-8462-5-17

[35] Rakers MM, van Buchem MM, Kucenko S, de Hond A, Kant I, van Smeden M, et al. Availability of Evidence for Predictive Machine Learning Algorithms in Primary Care: A Systematic Review. *JAMA Netw Open*. 2024;7(9):e2432990-e2432990. doi:10.1001/jamanetworkopen.2024.32990

[36] Shah NH, Pfeffer MA, Ghassemi M. The Need for Continuous Evaluation of Artificial Intelligence Prediction Algorithms. *JAMA Netw Open*. 2024;7(9):e2433009-e2433009. doi:10.1001/jamanetworkopen.2024.33009

[37] Shah NH, Milstein A, Sc BP. Making Machine Learning Models Clinically Useful. *JAMA*. 2019;322(14). doi:10.1001/jama.2019.10306

[38] Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. 2024;385. doi:10.1136/bmj-2023-078378

[39] Bedi S, Jain SS, Shah NH. Evaluating the clinical benefits of LLMs. *Nat Med*. 2024;30(9). doi:10.1038/s41591-024-03181-6

[40] Lu JH, Callahan A, Patel BS, Morse KE, Dash D, Pfeffer MA, et al. Assessment of Adherence to Reporting Guidelines by Commonly Used Clinical Prediction Models From a Single Vendor: A Systematic Review. *JAMA network open*. 2022;5(8). doi:10.1001/jamanetworkopen.2022.27779

[41] Shah NH, Halamka JD, Saria S, Pencina M, Tazbaz T, Tripathi M, et al. A Nationwide Network of Health AI Assurance Laboratories. *JAMA*. 2024;331(3):245-249. doi:10.1001/jama.2023.26930

[42] Fox A. Epic leads new effort to democratize health AI validation. Healthcare IT News. Published May 28, 2024. https://www.healthcareitnews.com/news/epic-leads-new-effort-democratize-health-ai-validation. Accessed October 14, 2024

[43] Seismometer Documentation. https://epic-open-source.github.io/seismometer/. Accessed October 14, 2024

[44] Beavins E. CHAI releases draft framework of quality assurance standards for healthcare AI. FierceHealthcare. Published June 26, 2024. https://www.fiercehealthcare.com/ai-and-machine-learning/chai-releases-draft-rubric-quality-assurance-standards-healthcare-ai. Accessed October 14, 2024

[45] Wornow M, Gyang RE, Callahan A, Shah NH. APLUS: A Python library for usefulness

simulations of machine learning models in healthcare. *J Biomed Inform*. 2023;139. doi:10.1016/j.jbi.2023.104319

[46] Health AI Partnership: an innovation and learning network for health AI software. Duke Institute for Health Innovation. Published December 23, 2021. https://dihi.org/health-ai-partnership-an-innovation-and-learning-network-to-facilitate-the-safe-effective-and-responsible-diffusion-of-health-ai-software-applied-to-health-care-delivery-settings/. Accessed October 14, 2024

[47] Goldberg CB, Adams L, Blumenthal D, Brennan PF, Brown N, Butte AJ, et al. To Do No Harm — and the Most Good — with AI in Health Care. *NEJM AI*. Published online February 22, 2024. doi:10.1056/AIp2400036

[48] Longhurst CA, Singh K, Chopra A, Atreja A, Brownstein JS. A call for artificial intelligence implementation science centers to evaluate clinical effectiveness. *NEJM AI*. 2024;1(8). doi:10.1056/aip2400223

[49] Deakers C. Roadmap: Healthcare AI. Bessemer Venture Partners. Published September 25, 2024. https://www.bvp.com/atlas/roadmap-healthcare-ai. Accessed October 10, 2024

[50] Emma —. Hippocratic AI. https://www.hippocraticai.com/emma. Accessed October 14, 2024

[51] Nayak A, Vakili S, Nayak K, Nikolov M, Chiu M, Sosseinheimer P, et al. Use of Voice-Based Conversational Artificial Intelligence for Basal Insulin Prescription Management Among Patients With Type 2 Diabetes: A Randomized Clinical Trial. *JAMA network open*. 2023;6(12). doi:10.1001/jamanetworkopen.2023.40232

[52] Huang W, Koh T, Tromp J, Chandramouli C, Ewe SH, Ng CT, et al. Point-of-care AI-enhanced novice echocardiography for screening heart failure (PANES-HF). *Sci Rep*. 2024;14(1):1-8. doi:10.1038/s41598-024-62467-4