

Critical AI Health Literacy as Liberation Technology: A New Skill for Patient Empowerment

Hugo Campos, Computational Precision Health Strategic Advisory Board, University of California, Berkeley and University of California, San Francisco; **Liz Salmi, AS**, OpenNotes, Beth Israel Deaconess Medical Center

December 8, 2025

Introduction

Artificial Intelligence (AI) is reshaping health care in ways that affect patients, clinicians, and institutions alike. *Health AI*, as defined in this commentary, encompasses digital tools deployed by health care organizations (e.g., generative AI, ambient scribes, machine learning) as well as tools used independently by patients and their care partners (typically generative AI). In institutional settings, AI is typically deployed to standardize clinical workflows, enforce compliance, manage operational risk, meet financial objectives, and extract value from big data (Allen et al., 2024; Goh et al., 2025; Gonzalez-Smith et al., 2022). For clinicians, AI is framed as a way of reducing administrative burden and improving diagnostic accuracy. Within health care organizations, all staff operate under the constraint of using institutionally approved digital solutions, which may narrow clinicians' discretion, while aligning practice with organizational priorities.

In contrast, patient-directed AI offers new opportunities for personal autonomy and agency that are not constrained by institutional policy. While patients may face barriers such as access, limited digital literacy, and incomplete access to health data, these tools leave room for curiosity and experimentation, allowing them to ask unconventional questions and pursue lines of inquiry institutions may discourage or overlook, a shift some describe as the rise of "AI Patients" (Blumenthal and Goldberg, 2025; Woods et al.,

2025). This commentary explores how AI can evolve from a tool of compliance to one of patient agency, reflection, and liberation, showing how strategic use can sharpen patient reasoning, deepen critical engagement with institutional priorities, and support patients in influencing the systems that shape care, reclaiming ownership of their health narrative.

The Need for Critical AI Health Literacy

In practice, most Health AI systems serve institutional priorities. Clinicians use tools shaped by operational and compliance requirements, insurers deploy algorithms that flag claims for denial, and hospitals implement triage models that can disadvantage marginalized patients through biased training data (Obermeyer et al., 2019; Robeznieks, 2021; Kelley, 2025). These applications are largely opaque to patients yet directly influence their care.

Patient-directed AI may offer an alternative path that shifts control toward the individual. In this way, patients select their own tools, cross-check recommendations, and explore perspectives institutional systems may overlook. The Patient AI Rights Initiative, an AI governance framework led by patient advocates, emphasizes seven foundational rights patients must become literate in and advocate for to challenge institutional priorities in Health AI, including transparency, self-determination, and independent duty to patients (Cordovano et al., 2024).

Yet, with these opportunities come significant risks. Generative AI can hallucinate or embed systemic bias. Commercial platforms lack HIPAA protections, and any data entered may be stored or disclosed through legal discovery. Without critical skills, patients may accept AI outputs uncritically, mistaking AI's confident framing for objective truth.

Critical AI Health Literacy addresses these risks by fostering what the authors call *algorithmic resistance*: the deliberate and informed use of AI to challenge institutional priorities that conflict with patient values. Building on Abel and Benkert's definition of critical health literacy as "the ability to reflect upon health determining factors and processes and to apply the results of the reflection into individual or collective actions for health in any given context," this approach positions AI not as a replacement for expert judgment, but as a catalyst for developing expertise (Abel and Benkert, 2022). Emerging evidence suggests patients often cannot distinguish AI-generated advice from physician-generated advice, and may trust inaccurate AI outputs (Shekar et al., 2025; Karinshak et al., 2023). This is why AI's true promise lies not in providing definitive answers, but in sharpening patients' capacity to question them.

Too often, AI health literacy is narrowly understood as the ability to craft effective prompts. While prompting skills are useful, they represent only the bare minimum necessary to operate generative AI. In contrast, Critical AI Health Literacy goes further: it equips patients and care partners to interrogate the systems behind AI, recognize bias and institutional alignment, and resist outputs that prioritize organizational interests over patient values.

The authors define Critical AI Health Literacy as the ability to strategically use AI tools to analyze health determining factors and power structures, evaluate AI outputs for bias and institutional alignment, and take informed action to advance individual and collective health goals through algorithmic resistance to systems that prioritize institutional priorities over patient values.

Why Critical AI Health Literacy Matters Now

Building Critical AI Health Literacy requires competencies such as evaluating AI outputs for bias and accuracy, cross-checking information across multiple sources, understanding the political and economic forces shaping AI deployment, and using AI as a "thought partner" as opposed to an "oracle." These skills transform AI from the "Dr. Google" of the last decade into a tool for strategic advocacy in the future (Obermeyer et al., 2019; Pascoe et al., 2024).

Patients are already cultivating these skills and fostering collaborative learning by sharing prompting strategies and interpretive approaches that deepen understanding and improve health care navigation. These practices are associated with improved health outcomes, higher quality, and safer care when patients are empowered to meaningfully participate (Greene et al., 2015). Examples of patients strategically using AI include: translating complex clinical jargon while preserving evidence-based meaning, compiling scattered medical records to accelerate symptom analysis, generating successful insurance appeal letters, and creating shareable health summaries that improve coordination for children with rare conditions (Salmi et al., 2025; DeBronkart, 2024). These practices highlight AI as a tool for patient empowerment rarely acknowledged outside of health systems (Sundar, 2025).

AI-Enabled Praxis: A Framework for Critical AI Health Literacy

Brazilian educational theorist Paulo Freire described *praxis* as a cycle of critical reflection, followed by strategic action, leading to liberation through transformative change (Freire et al., 2018). Applied to Health AI, praxis invites patients not only to understand the systems that shape their care but to use that understanding to assert agency and create change. It is not merely about acquiring health information, but critiquing the social, institutional, and political contexts in which that information exists, and then taking purposeful

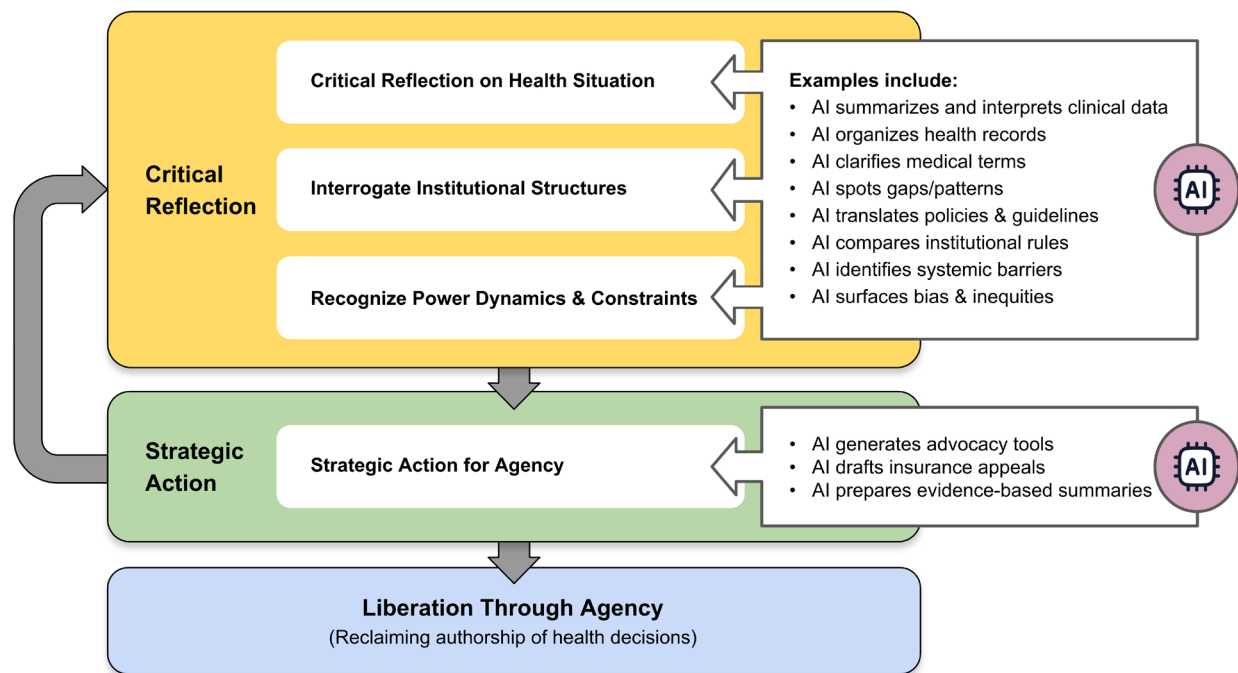


FIGURE 1 | AI-Enabled Critical Health Literacy: Freirean Praxis Cycle

SOURCE: Created by the authors.

steps to act upon it. These steps include critical reflection and strategic action, ultimately achieving liberation through agency (see *Figure 1*).

Critical Reflection: Critical reflection involves analyzing one's health condition, interrogating institutional practices, and uncovering the power dynamics across interactions among patients, clinicians, health care systems, and the technologies that shape their care. AI accelerates this process by helping patients explore treatment options, compare clinical guidelines, review medical literature, and detect patterns in fragmented health data that may otherwise remain hidden (see *Box 1*).

Strategic Action: Strategic action then applies these insights gained to influence decisions and advocate effectively. Patients might craft targeted questions, prepare persuasive insurance appeals, and integrate medical research into summaries for their care team (see *Box 2*).

Liberation through Agency: The final step, liberation through agency, occurs when patients move beyond using AI for surface-level inquiries, such as confirming existing information, and instead engage in an ongoing cycle of reflection and action. In this deeper mode, AI becomes a means of navigating—and when necessary, resisting—institutional priorities while remaining grounded in personal and community-defined values. The goal is not to reject professional expertise, but to break free from disempowerment by becoming an active co-author of one's health narrative. This, in Freire's terms, is liberation (Torre et al., 2017).

AI Limitations

While Critical AI Health Literacy holds promise for expanding patient agency, it faces significant limitations. Access to AI tools is not equitable. Many patients face barriers due to limited digital literacy, financial constraints, or inadequate

BOX 1 | Case Study: Communicating New Clinical Guidelines

This case shows how AI can help bridge gaps between clinician awareness and patient needs. Diagnosed with Restless Legs Syndrome (RLS) in 2019, LS began treatment with ropinirole, a dopaminergic drug commonly prescribed off-label for RLS. At first, the medication relieved her symptoms, but by 2024, she experienced “augmentation,” which is a worsening of symptoms ironically caused by the drug itself. As her quality of life declined, LS entered a cycle of critical reflection, questioning why her symptoms were intensifying and whether better treatments existed.

Turning to an RLS community on Reddit, LS learned about updated Mayo Clinic guidelines from 2021 that no longer recommended dopamine agonists as first-line therapy for RLS (Silber et al., 2021). She recognized a disconnect: her primary care provider (not a sleep medicine specialist) was unaware of new guidelines and recommendations. LS downloaded the updated guidelines and cited studies and sought access to publications located behind paywalls. To distill the information, she used GPT-4o (OpenAI, San Francisco, CA) to generate a one-page summary that could be easily skimmed by a busy doctor.

In a moment of strategic action, LS messaged the AI-generated summary to her primary care provider via the patient portal, noting “no reply” was needed. At the visit, LS discovered that her doctor had printed and read the summary, reviewed the guidelines, and agreed to taper LS off the dopamine agonist and begin alternative therapy. Two months later, LS reported improved sleep and reduced symptoms.

This story demonstrates liberation through AI-enabled praxis: a patient who used AI not just for information retrieval, but to interrogate clinical norms, and demonstrate agency. In this scenario, LS used AI to partner with her doctor—enacting Freire’s vision of liberation through reflective, strategic, and transformative action.

SOURCE: Created by the authors; based on second author’s experience.

internet connectivity. Without addressing these disparities, AI risks widening existing gaps in health care access and outcomes. Privacy is also a significant public concern. Commercial AI tools are not HIPAA-compliant, and any data entered may be stored, analyzed, or shared in ways that are not transparent to the user, potentially surfacing later through legal discovery, marketing analytics, or other unforeseen channels (Downing and Perakslis, 2022). A further risk lies in over-trusting AI outputs. Generative AI presents information persuasively, even when it contains inaccuracies or embedded biases. Patients may mistake fluency for accuracy, and act on misleading or incomplete advice (Hart, 2025; Dober, 2025; Draelos et al., 2025).

Finally, while AI can help individuals, structural inequities in health care cannot be addressed through individual empowerment alone. Equity requires coordinated systemic reform alongside the cultivation of patient skills.

Key Takeaways

When patients adopt Critical AI Health Literacy, they move from consumers of organizational AI to active agents of algorithmic resistance. This unlocks new forms of participation and control over health narratives, shifts long-standing power structures, and opens the possibility of more equitable, patient-centered decision making. Patient communities can play an important role

BOX 2 | Case Study: Using Multiple AI Systems to Manage Care

This case shows how AI can be used to challenge institutional decisions and protect patient preferences. In 2025, HC, living with hypertrophic cardiomyopathy, faced a dispute with his health system over replacement of his implantable cardioverter-defibrillator. His previous devices, both Medtronic, had delivered 18 years of safe outcomes, including shock-free therapy, on-label MRI compatibility, and reliable remote monitoring. Despite this history, new vendor contracts restricted coverage to devices from Boston Scientific and Biotronik, without clinical evidence to justify the change.

This policy created a direct conflict between a clinically supported patient preference and institutional procurement priorities. Refusing to accept a device change driven by contractual agreements, HC turned to generative AI tools, including GPT-4o (OpenAI, San Francisco, CA), Claude Sonnet (Anthropic, San Francisco, CA), Gemini 2.5 Pro and NotebookLM (Google LLC, Mountain View, CA), to build an evidence-based appeal. AI assisted in drafting and refining grievance letters, identifying clinically meaningful differences between device platforms, evaluating cost-effectiveness, and framing the risks associated with losing trusted remote monitoring capabilities. It also supported the careful selection of language that would strengthen the persuasiveness of the case. Through iterative use, HC transformed his appeal into a well-documented challenge to a system-centered decision.

This example illustrates how Critical AI Health Literacy enables patients to move from passive acceptance of institutional determinations to informed, strategic advocacy rooted in evidence and personal values. Using AI, HC reshaped the care decision toward a more individualized and patient-centered outcome by combining reflection on the clinical and institutional dynamics at play with deliberate, targeted action.

SOURCE: Created by the authors; based on first author's experience.

by co-designing AI tools that reflect shared needs, values, and ethics, and by advocating for technology that serves justice as well as efficiency. The AI-enabled praxis framework offers a way to understand the value of patient agency through Critical AI Health Literacy. True empowerment lies not in trusting AI for the right answer, but in learning to think with and beyond it, to question, shape, and challenge its outputs when they fail to serve the patient. In this vision, AI becomes more than a tool of oversight, efficiency, or automation; it becomes a technology of liberation.

References

Abel, T., and R. Benkert. 2022. Critical health literacy: Reflection and action for health. *Health Promotion International* 37(4). <https://doi.org/10.1093/heapro/daac114>.

Allen, M. R., S. Webb, A. Mandvi, M. Frieden, M. Tai-Seale, and G. Kallenberg. 2024. Navigating the doctor-patient-AI relationship—A mixed-methods study of physician attitudes toward artificial intelligence in primary care. *BMC Primary Care* 25(42). <https://doi.org/10.1186/s12875-024-02282-y>.

Blumenthal, D., and C. Goldberg. 2025. Managing patient use of generative health AI. *NEJM AI* 2(1). <https://doi.org/10.1056/Alpc2400927>.

Cordovano, G., D. DeBronkart, A. Downing, Y. Duron, L. K. Glenn, J. Holldren, D. Lewis, M. Murphy, V. Robinson, L. Salmi, C. Sarabu, and C. von Raesfeld. 2024. *Collective AI rights for patients: Written by patient experts and community leaders*. The Light Collective.

- Available at: <https://lightcollective.org/patient-ai-rights/> (accessed November 21, 2025).
- DeBronkart, D. 2024. #PatientsUseAI to fight benefit denials, leveling the playing field. *Patients Use AI* (blog), Substack, October 7. Available at: <https://patientsuseai.substack.com/p/patientsuseai-to-fight-benefit-denials> (accessed November 21, 2025).
- Dober, C. 2025. Using generative AI for therapy might feel like a lifeline—but there's danger in seeking certainty in a chatbot. *The Guardian*, August 3. Available at: <https://www.theguardian.com/commentisfree/2025/aug/03/generative-ai-chatbot-therapy-dangers-risks> (accessed November 20).
- Downing, A., and E. Perakslis. 2022. Health advertising on Facebook: Privacy and policy considerations. *Patterns* 3(9):100561. <https://doi.org/10.1016/j.patter.2022.100561>.
- Draelos, R. L., S. Afreen, B. Blasko, T. L. Brazile, N. Chase, D. P. Desai, J. Evert, H. L. Gardner, L. Herrmann, A. V. House, S. Kass, M. Kavan, K. Khemani, A. Koire, L. M. McDonald, Z. Rabeeah, and A. Shah. 2025. Large language models provide unsafe answers to patient-posed medical questions. Preprint. arXiv, July 25. <https://doi.org/10.48550/arXiv.2507.18905>.
- Freire, P., M. Bergman Ramos, I. Shor, and D. P. Macedo. 2018. *Pedagogy of the oppressed: 50th anniversary edition*. New York: Bloomsbury Academic.
- Goh, E., R. J. Gallo, E. Strong, Y. Weng, H. Kerman, J. A. Freed, J. A. Cool, Z. Kanjee, K. P. Lane, A. S. Parsons, N. Ahuja, E. Horvitz, D. Yang, A. Milstein, A. P. J. Olson, J. Hom, J. H. Chen, and A. Rodman. 2025. GPT-4 assistance for improvement of physician performance on patient care tasks: A randomized controlled trial. *Nature Medicine* 31:1233-1238. <https://doi.org/10.1038/s41591-024-03456-y>.
- Gonzalez-Smith, J., H. Shen, and C. Silcox. 2022. Moving ahead of the pack: Understanding health system priorities on AI-enabled clinical decision support. *Biomedical Instrumentation & Technology* 56(4):119-123. <https://doi.org/10.2345/0899-8205-56.4.119>.
- Greene, J., J. H. Hibbard, R. Sacks, V. Overton, and C. D. Parrotta. 2015. When patient activation levels change, health outcomes and costs change, too. *Health Affairs* 34(3):431-437. <https://doi.org/10.1377/hlthaff.2014.0452>.
- Hart, R. 2025. Chatbots can trigger a mental health crisis. What to know about 'AI Psychosis.' *TIME*, August 5. Available at: <https://time.com/7307589/ai-psychosis-chatgpt-mental-health/> (accessed November 20, 2025).
- Karinshak, E., S. X. Liu, J. S. Park, and J. T. Hancock. 2023. Working with AI to persuade: Examining a large language model's ability to generate pro-vaccination messages. *Proceedings of the ACM on Human-Computer Interaction* 7(16):1-29. <https://doi.org/10.1145/3579592>.
- Kelley, J. 2025. How AI can predict & prevent insurance claim denials. *ENTER*, April 10. Available at: <https://www.enter.health/post/how-ai-can-predict-prevent-insurance-claim-denials> (accessed November 21, 2025).
- Obermeyer, Z., B. Powers, C. Vogeli, and S. Mullainathan. 2019. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 336(6464):447-453. <https://doi.org/10.1126/science.aax2342>.
- Pascoe, J. L., L. Lu, M. M. Moore, D. J. Blezek, A. E. Ovalle, J. A. Linderbaum, M. R. Callstrom, and E. E. Williamson. 2024. Strategic considerations for selecting artificial intelligence solutions for institutional integration: A single-center experience. *Mayo Clinic Proceedings: Digital Health* 2(4):665-676. <https://doi.org/10.1016/j.mcpdig.2024.10.004>.
- Robeznieks, A. 2021. Feds warned that algorithms can introduce bias to clinical decisions. *American Medical Association*, June 23. Available at: <https://www.ama-assn.org/delivering-care/health-equity/feds-warned-algorithms-can->

introduce-bias-clinical-decisions (accessed November 21, 2025).

- Salmi, L., D. M. Lewis, J. L. Clarke, Z. Dong, R. Fischmann, E. I. McIntosh, C. R. Sarabu, and C. M. DesRoches. 2025. A proof-of-concept study for patient use of open notes with large language models. *JAMIA Open* 8(2). <https://doi.org/10.1093/jamiaopen/ooaf021>.
- Shekar, S., P. Pataranutaporn, C. Sarabu, G. A. Cecchi, and P. Maes. 2025. People overtrust AI-generated medical advice despite low accuracy. *NEJM AI* 2(6). <https://doi.org/10.1056/aioa2300015>.
- Silber, M. H., M. J. Buchfuhrer, C. J. Earley, B. B. Koo, M. Manconi, J. W. Winkelman, and P. Becker. 2021. The management of restless legs syndrome: An updated algorithm. *Mayo Clinic Proceedings* 96(7):1921-1937. <https://doi.org/10.1016/j.mayocp.2020.12.026>.
- Sundar, K. R. 2025. When patients arrive with answers. *JAMA* 334(8):672-673. <https://doi.org/10.1001/jama.2025.10678>.
- Torre, D., V. Groce, R. Gunderman, J. Kanter, S. Durning, and S. Kanter. 2017. Freire's view of a progressive and humanistic education: Implications for medical education. *MedEdPublish* 6(119). <https://doi.org/10.15694/mep.2017.000119>.
- Woods, S. S., S. M. Greene, L. Adams, G. Cordovano, and M. F. Hudson. 2025. From e-patients to AI patients: The tidal wave empowering patients, redefining clinical relationships, and transforming care. *Journal of Participatory Medicine* 17:e75794. <https://doi.org/10.2196/75794>.

DOI

<https://doi.org/10.31478/202512a>

Suggested Citation

Campos, H., and L. Salmi. 2025. Critical AI health literacy as liberation technology: A new skill for patient empowerment. *NAM Perspectives*.

Commentary, National Academy of Medicine, Washington, DC. <https://doi.org/10.31478/202512a>.

Author Information

Hugo Campos, is Strategic Advisory Board Member, Computational Precision Health Strategic Advisory Board, University of California, Berkeley and University of California, San Francisco. **Liz Salmi, AS**, is Communications & Patient Initiatives Director, OpenNotes, Beth Israel Deaconess Medical Center.

Acknowledgments

The authors thank **Sue Woods, MD**, for her insightful comments on an earlier version of this manuscript.

Conflict-of-Interest Disclosures

Liz Salmi has received research support from Abridge AI, Inc. All other authors report no competing interests.

Correspondence

Questions or comments should be directed to Hugo Campos at hcampos@researchpartner.org.

Sponsor(s)

No external financial support or grants were received from any public, commercial, or not-for-profit entities for the research, authorship, or publication of this article.

Disclaimer

The views expressed in this paper are those of the authors and not necessarily of the authors' organizations, the National Academy of Medicine (NAM), or the National Academies of Sciences, Engineering, and Medicine (the National Academies). The paper is intended to help inform and stimulate discussion. It is not a report of the NAM or the National Academies. Copyright by the National Academy of Sciences. All rights reserved.