The Invisible Bodyguard

Artificial Intelligence as Safety Infrastructure in Healthcare

Summary

Among all sectors of human society ,healthcare stands to gain the highest net positive impact from AI. This case rests on AI's quietest capabilities: synthesizing what fragmented systems scatter, and surfacing what time-pressured encounters miss. In this role, AI acts as an invisible safety infrastructure that supports human judgment rather than replacing it.

To make this argument concrete, the essay examines two applications in women's health, a domain where the gap between what medicine knows and what it actually detects is enormous. First, AI-driven screening for endometriosis, a condition affecting 190 million women worldwide. Endometriosis currently takes an average of 6-7 years after symptom onset to diagnose. This delay occurs because symptoms unfold across visits, providers, and years in ways no single clinician can track. Recent work has demonstrated over 90% accuracy for AI driven endometriosis screening using only routine clinical features already captured in electronic health records. Second, predictive models for pregnancy complications including preeclampsia, hemorrhage, and preterm birth, deployed in a country where over 80% of pregnancy-related deaths are classified as preventable and Black women die at more than three times the rate of white women.

In both cases, AI acts as invisible safety infrastructure: synthesizing longitudinal patterns across fragmented records, flagging risk at the point of care, and reaching patients regardless of their provider's individual expertise or patients' own ability to self-advocate. Early endometriosis diagnosis saves roughly $13,000 per patient in reduced surgical and emergency costs. The societal cost of maternal morbidity for all 2019 U.S. births has been estimated at $32.3 billion, driven by complications that predictive models are designed to catch early. Healthcare is uniquely positioned because it combines immediate individual benefit (such as a specific life saved or a disease caught years earlier), equity multiplier potential (standardized screening that narrows disparities), and technical feasibility with data that already exists in routine clinical systems.

Nevertheless, clinical AI can fail silently: a model that appears accurate overall can quietly underperform for the populations that need it most, as both published case studies and the author's own research have demonstrated. Silent failure is compounded by automation bias, reproductive data privacy vulnerabilities, and a digital divide that risks delivering these tools last to the communities that need them first.

These risks are serious, but must become design requirements instead of reasons for inaction. Each has a known corrective: external validation across diverse populations before deployment, confidence-level reporting that preserves clinician judgment, and open-science frameworks that ensure access is not limited to well-resourced systems. The tools and datasets necessary to implement these safeguards already exist. The alternative to deploying AI with this discipline is not an idealized healthcare system. Rather, it is the current one, in which the United States spends more on healthcare than any wealthy nation yet has the highest maternal mortality rate among them, and women wait years for a diagnosis that pattern recognition could accelerate. The framework described here generalizes beyond women's health to every domain where diagnostic delays, fragmented data, and unequal outcomes persist, building an *invisible bodyguard* for every patient.

The Invisible Bodyguard: Artificial Intelligence as Safety Infrastructure in Healthcare

I. Introduction: When Care Fails Quietly, Safety Infrastructure Matters

The conversation about Artificial Intelligence in healthcare tends to swing between two fantasies. In one, AI replaces healthcare personnel and is "better than doctors":[1] it reads scans faster, diagnoses better, and renders the white coat obsolete. In the other, AI is a threat:[2] a black box making life-or-death decisions that no one can explain or appeal. Between these two extremes lies a role for AI that could fundamentally change the relationship between the healthcare system and the people it serves.

The greatest positive impact of AI in healthcare will not come from replacing clinicians, but from acting as an *invisible bodyguard*: safety infrastructure that monitors, flags, and synthesizes information to support human judgment. In practice, this means machine learning models trained on thousands of patient records to recognize patterns (such as combinations of symptoms, lab values, and clinical histories) that predict a diagnosis or complication before it becomes obvious to any single clinician. Clinicians already work under severe cognitive overload, navigating records fragmented across institutions and time in encounters too brief to surface every emerging risk. When a woman spends almost a decade hearing that her pain is normal before receiving a diagnosis of endometriosis, the failure is not one of medical knowledge, but of system infrastructure.[3] It is a failure of systems unable to recognize patterns that unfold slowly, across visits, and outside narrow diagnostic silos.[4] These are not problems that require Artificial General Intelligence,the kind of all-purpose reasoning system that remains speculative.but only careful, well-validated pattern recognition. The models and the data already exist; what remains is clinical integration, regulatory approval, and equitable deployment.

This essay argues that among all sectors, healthcare stands to gain the highest net positive impact from AI through AI's quietest capabilities: synthesizing what fragmented systems scatter, and surfacing what time-pressured encounters miss. To make this argument concrete, this essay examines women's health: a domain traditionally overlooked by both medicine and technology, where the information needed to do better is already sitting in medical records, unread.

II. Why Healthcare Offers the Fastest Path from Computation to Saved Lives

Among education and climate to finance and scientific research, healthcare stands out as advancements directly s translate to human welfare with rapid and measurable reductions: a disease caught years earlier, a disparity in care narrowed rather than reinforced.

Other sectors, such as climate change prevention or education, have compelling cases but neither sector will deliver an impact of equivalent magnitude in as short a timeframe, as climate's effects will be sustained and amortized throughout centuries, and education currently lacks comparable digital infrastructure to bring about the massive change which we will witness soon in healthcare.

---

[1] Park, Alice, "Microsoft's AI Is Better Than Doctors at Diagnosing Disease," *Time*, July 2, 2025.
[2] ECRI, "Top 10 Health Technology Hazards for 2025," *ECRI*, December 2024.
[3] Frankel, Lexi R., "A 10-Year Journey to Diagnosis With Endometriosis: An Autobiographical Case Report," *Cureus*, January 2022.
[4] Fryer, J. et al., "Understanding diagnostic delay for endometriosis: a scoping review using the social-ecological framework," Health Care for Women International, October 2024.

For over a century, medicine has been reactive; you get sick, you go to the doctor, they treat you. The entire infrastructure is built around this, from hospitals and emergency rooms to specialist referrals and insurance that covers treatment but not prevention. AI will be the first time that we will not just patch this system, as it makes a fundamentally different approach possible for the first time.

In our daily lives there will be continuous, passive monitoring that catches disease before you ever feel a symptom or walk into a clinic. In our occasional fights with new and old sicknesses, stress, or hereditary diseases, AI will be our *invisible bodyguard* standing by to set things in motion when we need protection. There's no other sector which will be re-imagined at such a level in the near future.

Unlike in other sectors where AI applications depend on new sensing infrastructure or speculative scientific breakthroughs, healthcare already generates vast quantities of relevant data through electronic health records, clinical notes, and laboratory systems embedded in routine care.[5]

The invisible bodyguard will arrive in our lives sooner than some might expect. In fact, one in three Americans is already collecting enough data through wearable devices to complement clinical records.[6] Therefore, the challenge evolves from data collection to synthesis. This, therefore, enables evaluation against measurable outcomes, such as lives saved, reduced diagnostic delays, lower costs, and narrowed disparities. Each of these outcomes can be quantified to rigorously assess whether AI systems are truly delivering on their promises.

We can also take a more analytical look at what the healthcare field stands to gain from AI. Currently, medical errors rank among the leading causes of death in the United States.[7] Clinician burnout has reached crisis levels, driven in part by the cognitive demands of synthesizing information across fragmented electronic health record systems.[8] Preventable harm is not an edge case but a structural feature of a system that asks human beings to process more data, across more patients, in less time than any individual can reliably manage. The burden of these failures also falls unevenly, with dramatically different outcomes depending on the race, gender, geography, and insurance status of the patient. Oftentimes, this occurs because of differences in access, attention, and the quality of care delivered. [9]

In few other domains can marginal improvements in pattern recognition so directly and measurably reduce human suffering.

III. Women's Health: Where the Gap Is Widest

---

[5] Office of the National Coordinator for Health Information Technology, "National Trends in Hospital and Physician Adoption of Electronic Health Records," HealthIT.gov, 2021.
[6] Shandhi, M.M.H. et al., "Assessment of ownership of smart devices and the acceptability of digital health data sharing," *npj Digital Medicine* 7, 44 (2024).
[7] Makary, M.A. and Daniel, M., "Medical error — the third leading cause of death in the US," *BMJ* 353:i2139, May 2016.
[8] Asgari, E. et al., "Impact of Electronic Health Record Use on Cognitive Load and Burnout Among Clinicians: Narrative Review," *JMIR Medical Informatics* 12:e55499, April 2024.
[9] Commonwealth Fund, *Advancing Racial Equity in U.S. Health Care: 2024 State Health Disparities Report*, April 2024.

Of all the areas where healthcare is failing, women's health stands out for a telling reason: many of its worst outcomes stem from diseases we already understand but systematically fail to detect in time. Women's health conditions are chronically underfunded,[10] under-researched, [11] and under-taught in medical schools.[12] The result is a system where years-long diagnostic delays, preventable maternal deaths, and persistent racial disparities coexist with the medical knowledge to prevent them. The following two applications aim to illustrate what this could look like in practice.

Endometriosis: From Years of Silence to Early Detection

Endometriosis is a chronic inflammatory condition affecting roughly 10% of reproductive-age women, or 190 million people worldwide. It might sound, therefore, surprising that it takes an average of 6-7 years to diagnose after symptom onset.[13] However, the delay is structural. Initially, a woman presents to her primary care physician with pelvic pain and is told it's a bad period. After two years pass, she is referred to a specialist. Another two and a half years pass before she receives a definitive diagnosis, often only after a surgical procedure requiring general anesthesia, because no non-invasive gold standard exists.[14] At every stage, the same pattern repeats: symptoms are normalized, records are fragmented across providers, and no single clinician sees the full picture.

The physical consequences compound with every year of delay. Endometriosis is a leading cause of infertility, affecting up to half of women with the condition,[15] and prolonged disease progression brings chronic pain that worsens over time. Each year without a diagnosis also means more emergency visits, more imaging, and more ineffective treatments, with costs falling on both the patient and the system. [16]

AI can intervene here because the problem, at its core, is one of pattern recognition across existing data. The symptom histories, laboratory values, and comorbidity patterns that distinguish endometriosis from other conditions are already captured in electronic health records. Therefore, even though a physician might not have the full picture to identify high-risk patients, a model trained across thousands of patients might. Recent publications have shown it works, screening for endometriosis with over 90% accuracy using only routine clinical features.[17] With comprehensive evaluation of these tools and holistic design of the systems, we can integrate the models into clinical workflows, bringing us one step closer to a reality with AI as safety infrastructure for healthcare.

---

[10] Mirin, Arthur A., "Gender Disparity in the Funding of Diseases by the U.S. National Institutes of Health," *Journal of Women's Health*, July 2021.
[11] National Academies of Sciences, Engineering, and Medicine, "A Vision for Women's Health Research: Transformative Change at the National Institutes of Health," March 2025.
[12] Society for Women's Health Research, "Rewriting Endometriosis Education for Providers and Policymakers," April 2024.
[13] Hadjicosta, E. et al., "Time to Diagnose Endometriosis: Current Status, Challenges and Regional Characteristics - A Systematic Literature Review," *PMC*, 2024
[14] Fryer, J. et al., "Understanding diagnostic delay for endometriosis," *Health Care for Women International*, October 2024
[15] Macer, M.L. and Taylor, H.S., "Endometriosis and Infertility," *Obstetrics and Gynecology Clinics of North America*, 2012.
[16] Surrey, Eric et al., "Impact of Endometriosis Diagnostic Delays on Healthcare Resource Utilization and Costs," *Advances in Therapy* 37, 2020.
[17] Bendifallah, Sofiane et al., "Machine learning algorithms as new screening approach for patients with endometriosis," *Scientific Reports* 12, 639, January 2022.

The benefit is concrete: a screening tool like this, deployed at the primary care level, could flag at-risk patients before the referral loop begins. Consequently, this would cut years off the diagnostic pathway and reaching patients regardless of their insurance, their provider's individual expertise, or whether they happen to advocate loudly enough for themselves. Currently, diagnostic delays are longer for patients in public healthcare systems and for those with lower socioeconomic status.[18] With this application, the invisible bodyguard can help support millions of overlooked women.

Pregnancy Complications: Prediction Before Crisis

Maternal health presents the same structural problem as endometriosis in a different form. The United States spends more on healthcare than any other high-income nation, yet has the highest maternal mortality rate among them; Specifically, there were 22.3 deaths per 100,000 live births in 2022, while countries like Norway recorded near zero[19]. Over 80% of those pregnancy-related deaths classified as preventable.[20] Black women die at more than three times the rate of white women even after adjusting for income, education, and clinical risk factors.[21] For every death, approximately 100 women experience severe maternal morbidity, affecting over 60,000 women annually.[22]

The risk factors that precede these outcomes are not hidden. Factors such as hypertensive disorders, gestational diabetes, abnormal lab trends, and patterns in prenatal visit notes are present in the medical record, often months before a crisis occurs.[23] Synthesizing this information from streams of data can unlock preventative tracking of the trajectory of a pregnancy. My own research, using Kaiser Permanente pregnancy records validated against demographically diverse national data, has demonstrated that predictive models integrating these data achieve clinically meaningful early warning scores for preeclampsia, hemorrhage, and preterm birth. Critically, we also noted that external validation across demographically diverse populations is essential, because a model performing well in aggregate can fail specific communities in specific states. [24]

The benefit here is not just earlier warning, but rather is *equitable* earlier warning. A standardized risk model applied at every prenatal visit does not dismiss a Black woman's blood pressure reading, assume a patient on Medicaid is less worth investigating, or forget what happened at a visit three months ago at a different clinic. It synthesizes the full record and flags risk regardless of who the patient is or where she seeks care.

---

[18] Breton, Z. et al., "Endometriosis Diagnostic Delay and Its Correlates," *Journal of Women's Health* 35(2):172-188, 2026.
[19] Gunja, Munira Z. et al., "Insights into the U.S. Maternal Mortality Crisis: An International Comparison," *Commonwealth Fund*, June 2024.
[20] Centers for Disease Control and Prevention, "Working Together to Reduce Black Maternal Mortality," CDC.gov, November 2024.
[21] Hill, Latoya et al., "Racial Disparities in Maternal and Infant Health: Current Status and Key Issues," *KFF*, December 2025.
[22] Howell, Elizabeth A., "Reducing Disparities in Severe Maternal Morbidity and Mortality," *Clinical Obstetrics and Gynecology*, PMC, 2018.
[23] Li, Shilong et al**.**, "Improving Preeclampsia Risk Prediction by Modeling Pregnancy Trajectories from Routinely Collected Electronic Medical Record Data," *npj Digital Medicine* 5, no. 68 (2022).
[24] Information gathered from my research, work under review in a journal

IV. What Changes When Risk Is Flagged Early Enough to Matter

In both cases, the core benefit of AI in healthcare is the same: complications are caught early enough for intervention to change the outcome. Endometriosis is identified at the primary care level before it progresses to surgical severity or preeclampsia is flagged weeks before it becomes an emergency, when low-dose aspirin or adjusted monitoring can still alter the trajectory. For clinicians, these tools reduce cognitive load, synthesizing longitudinal patterns across fragmented records that no single provider can track alone. For patients, the shift is more fundamental: from being a passive recipient of care that arrives too late to being identified, monitored, and acted on early.

The economic case reinforces the moral one. Early diagnosis of endometriosis saves roughly $13,000 per patient compared to the long-delay cost profile, concentrated in reduced surgical interventions and emergency care. [25] The societal cost of maternal morbidity for all 2019 U.S. births has been estimated at $32.3 billion, driven by complications that predictive models are designed to catch early. [26] These figures take on additional urgency in a country that has both the highest maternal mortality rate among wealthy nations and a declining fertility rate.[27] Hence, a system that makes pregnancy dangerous enough to deter it cannot be sustained.

V. The Flipside: A bodyguard for everyone could be no one's bodyguard

The most important thing to understand about clinical AI risk is that it can fail silently. One way AI can fail is due to what academics call a distribution shift. Simply put, different groups of people can exhibit different characteristics. In consequence, a prediction model built on data from one hospital, or one population, does not necessarily work at another. Therefore, the danger of mass adoption of AI models is this: by building AI models for everyone, we could end up with an invisible bodyguard that protects some and abandons others.

This concern has been illustrated from numerous case studies in recent years. For instance, the Epic Sepsis Model, a proprietary early-warning tool deployed at hundreds of U.S. hospitals, was independently tested at Michigan Medicine in 2021 and found to miss two-thirds of sepsis cases. The model performed far below the accuracy its developer had reported, while simultaneously generating alerts for 18 percent of all hospitalized patients who did not have sepsis.[28] Even though the model had been trained and tested internally; when it encountered a different patient population, it quietly degraded.

This kind of silent, uneven failure is what makes every other risk of clinical AI more serious than it initially appears. Clinicians working with decision-support tools tend to defer to algorithmic recommendations. In systematic reviews, providers overrode their own correct judgments in favor of erroneous system output in 6 to 11 percent of cases.[29] If a model is quietly underperforming for a given

---

[25] Soliman, Ahmed et al., "Impact of Endometriosis Diagnostic Delays on Healthcare Resource Utilization and Costs," *PMC*, 2020.
[26] Lalani et al., "Societal Cost of Nine Selected Maternal Morbidities," *PLOS ONE* 17, no. 10 (2022).
[27] Martin JA, Hamilton BE, Osterman MJK, "Births in the United States, 2024," NCHS Data Brief No. 535, July 2025.
[28] Wong et al., "External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients," *JAMA Internal Medicine* 181, no. 8 (2021).
[29] Lyell and Coiera "Automation Bias and Verification Complexity: A Systematic Review," *JAMIA* 24, no. 2 (2017).

population it actively harms that population by replacing a clinician's uncertainty, which might have prompted additional testing, with false confidence that the patient is low-risk.

The data infrastructure these systems require also introduces vulnerabilities specific to women's health. In the aftermath of *Dobbs v. Jackson Women's Health Organization* (2022), reproductive health data has taken on uniquely sensitive legal implications, having been sought for enforcement of state abortion restrictions. This includes at least one documented case where a state health official who compiled patients' menstrual cycles into a spreadsheet for investigative purposes. [30] Any system integrating patient-reported reproductive data requires robust privacy protections, transparent consent, and clear governance against non-medical use, requirements that do not yet exist at the scale deployment would demand.

Finally, AI-powered screening depends on interoperable electronic health record systems, adequate digital infrastructure, and reliable device access for patient reported data. Rural healthcare settings, under-resourced clinics, and low-income populations are likely to receive these tools last. If clinical AI is commercialized as a proprietary product available only to well-resourced health systems, it risks widening the gap between those who already receive adequate care and those who do not.

VI. Why These Risks Demand Better Design, Not Abandonment

The strongest argument for AI in healthcare begins with an honest look at the system it would augment. The failures of the system, such as the high maternal mortality rate and diagnostic delays for endometriosis, are structural problems. AI-augmented care should be evaluated against that system, and when it is, the potential to meaningfully change outcomes becomes clear.

The risks described in the previous section are serious, yet they are solvable if treated as design requirements. The Epic Sepsis Model failed because a proprietary tool was deployed at hundreds of hospitals without independent testing against the populations it was meant to serve. The corrective is straightforward: mandate external validation across multiple health systems, demographic groups, and geographic regions before a model reaches a patient. Datasets built for exactly this purpose already exist. The All of Us Research Program, for instance, deliberately oversamples historically underrepresented populations, offering the kind of diverse validation cohort that responsible deployment demands[31].

The same principle extends to the other risks. Automation bias is mitigated when systems report confidence levels alongside predictions, giving clinicians reason to exercise independent judgment. The digital divide narrows when models are developed through publicly funded, open-science frameworks, ensuring that a rural clinic in Mississippi has access to the same screening capabilities as an academic medical center in Boston. Furthermore, reproductive data privacy requires governance structures that are explicit, enforceable, and built for a legal landscape in which reproductive health data carries criminal implications.[32]

---

[30] Georgetown Law O'Neill Institute, "Emerging Threats to Data Privacy Post-Dobbs" (2024).
[31] Denny et al., "The 'All of Us' Research Program," *New England Journal of Medicine* 381, no. 7 (2019).
[32] Donelson, Rolonda, "Emerging Threats to Data Privacy Post-Dobbs," *Georgetown Law O'Neill Institute*, October 2024.

This is not a tradeoff between two imperfect options. It is an opportunity to hold healthcare AI to a higher standard than the system it augments, and in doing so, to build something which can protect the people it serves. By mandating the safeguards this essay has described and iterating on them through sustained collaboration between researchers, clinicians, and the communities these tools are meant to serve, the United States has an opportunity not just to reduce preventable deaths but to rebuild trust in a healthcare system that has overlooked its patients.

VII. Beyond Women's Health: A Framework That Generalizes

The framework this essay has described is not specific to endometriosis or maternal mortality. It applies wherever the same structural pattern holds: data already present in the medical record, diagnostic delays driven by fragmentation rather than missing knowledge, and outcomes that diverge sharply along lines of race, geography, sex, or other characteristics.

Many other conditions follow a strikingly similar trajectory. Heart disease is the leading killer of women in the United States, yet women are significantly less likely to be diagnosed with acute coronary syndrome at initial presentation, in part because diagnostic criteria were historically developed around male symptom profiles. [33] Rare diseases present the diagnostic odyssey at its most extreme: the average patient waits five to seven years for a correct diagnosis, cycling through multiple specialists, none of whom sees the complete picture.[34] In addition, diabetes management, where racial disparities in outcomes are well documented and the relevant clinical data already sits in EHR systems, represents a domain where this infrastructure could be deployed with relatively low technical barriers and high impact.[35]

Women's health is often treated as a specialty concern, but the conditions described here are the rule in healthcare, not the exception. Every domain discussed in this section shares that same structural failure: data that could prevent harm already exists in the medical record but is not synthesized, surfaced, or acted on in time. Women's health is where this failure is most acute, and therefore where the case for AI-augmented clinical infrastructure is most demanding. If the framework meets that standard, its reach extends far beyond it, and its impact can include health systems where specialist access is not just fragmented but scarce, and where a single screening model could deliver diagnostic capacity that does not yet exist.

---

[33] Haider, Ahmed et al., "Sex and gender in cardiovascular medicine: Presentation and outcomes of acute coronary syndrome," *European Heart Journal* 41(13), December 2019.
[34] Faye, Fatoumata et al., "Time to diagnosis and determinants of diagnostic delays of people living with a rare disease: results of a Rare Barometer retrospective patient survey," *European Journal of Human Genetics* 32, 2024.
[35] Hassan, Saria et al., "Disparities in diabetes prevalence and management by race and ethnicity in the USA: defining a path forward," *Lancet Diabetes & Endocrinology*, 2023.

VIII. Conclusion: Looking Forward

To conclude, the data to catch these failures earlier already sits in our health systems, scattered across fragmented records, clinical notes, vital sign trajectories, and patients' own self-reported symptoms. The models to extract value from this data are not hypothetical. They exist, they have been tested, and they have been validated carefully across institutions, populations, and demographic subgroups.

This essay began by rejecting two fantasies: that AI will replace clinicians, and that it is too dangerous to deploy. The reality is less cinematic and more consequential. AI in healthcare is most valuable as the invisible bodyguard that has been illustrated: infrastructure that synthesizes what no individual clinician can hold in their head and surfaces risks before they become crises. The case studies importantly examine both what that infrastructure looks like in practice, but also the risks section when it is built without discipline.

AI adoption in healthcare is not a future question, considering that over a thousand devices already carry FDA authorization.[36] The question is whether we build these systems as shortcuts around human judgment or as scaffolding that allows human judgment to reach patients it currently fails. The 190 million women living with endometriosis, the mothers whose deaths are classified as preventable, and the communities examined in this essay are not the only stakeholders. Every domain where diagnostic delays, fragmented data, and unequal outcomes persist stands to benefit from the same framework, if we insist on building it right. The question is no longer whether the technology can be ready. It is whether we can.

[36] Mehta et al., "Evaluating transparency in AI/ML model characteristics for FDA-reviewed medical devices," *npj Digital Medicine*, November 2025.