Trustworthy Edge Intelligence for Continuous Cardiovascular Monitoring: Combining PPG Foundation Models with Adversarially Secure Medical AI Agents

Grant Bussell
Department of Computer Science, Colorado State University, Fort Collins, CO, USA.
grantrussell349@colostate.edu

Dominik L. Alvarez
Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO, USA.
dominikmail@missouri.edu

Arun L. Subramanian
Department of Computer Science, University of Central Florida, Orlando, FL, USA.
arunmail@ucf.edu

Abstract

Continuous cardiovascular monitoring at the edge of healthcare networks promises to transform the detection and management of arrhythmias, hypertension, and heart failure by shifting computation onto wearable and near-body devices. Recent advances in photoplethysmography (PPG) foundation models have demonstrated remarkable generalization across diverse populations and sensor modalities, yet the deployment of such models within medical AI agent architectures raises profound questions of trustworthiness. This paper examines the system-level integration of PPG foundation models with adversarially secure medical AI agents operating at the network edge. We analyze architectural trade-offs among on-device inference, collaborative edge-cloud splitting, and federated learning topologies, highlighting the tension between model expressivity and resource constraints. A central contribution is a conceptual adversarial security framework that rethinks medical agent hardening in light of emerging models, where input perturbations, model inversion, and prompt-level attacks threaten both diagnostic accuracy and patient data privacy. We explore how statistical-prior informed generative masking strategies within PPG pretext tasks can serve as implicit regularization against distributional shifts and adversarial noise, while structured uncertainty quantification and certified robustness methods bolster agent reliability. The discussion extends into fairness auditing across demographic strata, governance mechanisms for federated model updates, and sustainability considerations for edge hardware lifecycles. By synthesizing cross-domain insights from embedded systems, foundation model training, and adversarial machine learning, we outline a trustworthy edge intelligence paradigm for cardiovascular care that balances clinical safety, data protection, and equitable access.

Keywords

edge intelligence, cardiovascular monitoring, photoplethysmography, foundation models, adversarial security, medical AI agents, trustworthiness.

1. Introduction

The convergence of wearable sensing, low-power computation, and deep learning has opened a new frontier in continuous health surveillance, particularly for cardiovascular diseases that remain the leading cause of mortality worldwide. Optical photoplethysmography (PPG) sensors, now embedded in wrist-worn devices and patches, enable the noninvasive acquisition of pulsatile blood volume signals over extended periods, yielding rich information about heart rate variability, atrial fibrillation burden, and hemodynamic trends. The emergence of edge computing architectures allows signal processing and preliminary inference to be performed locally, reducing reliance on cloud connectivity and minimizing latency in time-sensitive clinical alarms [1]. Large-scale validation studies such as the Apple Heart Study have demonstrated the feasibility of population-level arrhythmia screening via consumer wearables, underscoring both the promise and the scale of data involved [2]. However, traditional deep learning pipelines for PPG analysis often rely on supervised models trained on narrowly curated datasets, limiting their generalizability across skin tones, motion artifacts, and device types. Recent PPG foundation models have sought to address this brittleness by learning universal representations from massive unlabeled data corpora, leveraging self-supervised pretext tasks that capture invariant physiological features [3]. While such models dramatically improve cross-domain robustness, their integration into medical AI agents that execute preventive screening, triage, and chronic disease management at the edge introduces a host of systemic trustworthiness challenges that extend well beyond predictive accuracy.

The vision of a medical AI agent operating on an edge device -- continuously ingesting PPG streams, reasoning about arrhythmic episodes, and interacting with patients through conversational interfaces or automated alerts -- demands a fundamental rethinking of safety, security, and governance. Foundation model size and inference cost must be reconciled with the memory, power, and thermal envelopes of wearable hardware, often necessitating aggressive quantization or model splitting between edge and cloud. At the same time, adversarial vulnerabilities that have been extensively documented in computer vision and natural language processing are equally pertinent in the medical domain, where imperceptible perturbations of a PPG waveform could induce false arrhythmia detections or suppress true pathological signatures [4]. The deployment of large language model agents for medical decision-making further compounds the threat surface, because these agents inherit susceptibility to prompt injection and output tampering, potentially distorting clinical recommendations relayed to a patient or caregiver. Trustworthiness must therefore be engineered as a cross-layer property, encompassing robust representation learning, adversarial hardening, fairness audits, and transparent update mechanisms that preserve patient safety and data sovereignty.

This paper presents a comprehensive systems analysis of trustworthy edge intelligence for continuous cardiovascular monitoring, focusing on the synergistic interplay between PPG foundation models and adversarially secure medical AI agents. We do not propose a single algorithm or architecture but instead develop a conceptual framework that integrates recent foundational advances with adversarial defense strategies, evaluated through the lenses of scalability, privacy, equity, and sustainability. The remainder of the paper is organized as follows. Section 2 positions our work within the broader landscape of edge AI for healthcare and foundation models for time-series signals. Section 3 describes the architectural choices underlying edge-native cardiovascular monitoring systems and their implications for model distribution and inference latency. Section 4 delves into the design principles of PPG

foundation models and their capacity to encode robust physiological representations, including the use of statistical-prior informed masking. Section 5 analyzes the adversarial threat model for medical AI agents and presents a layered defense strategy informed by recent work on adversarial robustness for clinical language models. Section 6 discusses governance, fairness, and lifecycle management in deployed systems. Section 7 concludes with future directions for regulatory science and translational deployment.

2. Background and Related Work

The intellectual lineage of our work draws from three interlocking domains: edge computing and wearable sensing, foundation model paradigms for time-series and medical data, and adversarial robustness in safety-critical AI. Edge computing, originally proposed to push computation closer to data sources in the Internet of Things, has evolved into a rich ecosystem of on-device accelerators and runtime-optimized neural network compilers [1]. In healthcare, this shift has been leveraged to enable real-time electrocardiogram analysis on resource-constrained microcontrollers and on-wrist PPG classification that preserves sensitive data locality. The Apple Heart Study [2] and numerous follow-up investigations have provided clinical validation that consumer-grade PPG sensors can achieve actionable atrial fibrillation detection, albeit with non-trivial false positive rates in ambulatory settings. Deep PPG models, such as the convolutional architectures proposed by Reiss et al., demonstrated that large-scale heart rate estimation from wrist-worn PPG can approach clinical gold standards under controlled motion conditions, but the reliance on supervised labels curtails their adaptation to unseen populations and device characteristics [3]. The need for representation learning across heterogeneous domains thus set the stage for foundation models in time-series health signals.

The foundation model concept, articulated comprehensively by Bommasani et al., describes a paradigm where a single large model pretrained on broad data can be adapted to a multitude of downstream tasks, often without task-specific architectural modifications [4]. In medicine, the generalist medical AI movement has produced models such as those described by Moor et al. that unify imaging, text, and genomic modalities under a self-supervised learning umbrella [6]. For physiological time-series, federated learning strategies have been exploited to harness distributed institutional data while preserving privacy, as elaborated by Rieke et al., enabling models to encounter diverse demographic and clinical populations without centralized data pooling [5]. These developments provide the technical foundation for PPG foundation models that learn from millions of hours of raw optical signals and can be fine-tuned for tasks ranging from blood pressure estimation to sleep apnea screening. Yet the translation of such models to edge-deployed agents introduces new dimensions of risk: model size, inference latency, and the adversarial vulnerability of the resulting decision pipelines demand a security-aware design that has not been central to the foundation model discourse thus far.

Adversarial machine learning research, initially catalyzed by the discovery of imperceptible perturbations that fool image classifiers, has gradually permeated the medical domain. Finlayson et al. demonstrated that health system-level attacks could manipulate imaging models to alter diagnostic outputs, raising alarms about the brittleness of deep learning in clinical workflows [9]. In parallel, work on certified robustness via randomized smoothing and adversarial training has yielded practical tools for bounding worst-case perturbations under certain norm constraints, though these guarantees remain difficult to transfer to complex, temporally correlated signals such as PPG [11]. The intersection of large language models and medical decision-making introduces a qualitatively different attack surface: adversarial prompting, context manipulation, and output steering can subvert a conversational

agent's safety alignment, a problem that has motivated new research on agent hardening [8]. Together, these threads inform our system-level analysis, which situates PPG foundation models within a broader adversarial agent framework and asks how real-time, edge-native cardiovascular monitoring can be made trustworthy in the face of both digital and clinical adversaries.

3. Edge Intelligence Architecture for Cardiovascular Monitoring

Designing a trustworthy edge intelligence system for cardiovascular monitoring requires navigating a complex sociotechnical space of hardware constraints, data distribution strategies, and fault tolerance. A foundational decision is the placement of the model inference pipeline. In a purely on-device paradigm, the entire PPG processing chain -- signal preprocessing, feature extraction, foundation model inference, and agent reasoning -- executes on the wearable or a paired smartphone, ensuring that raw physiological data never leaves the user's personal device. This maximizes privacy and eliminates network dependency for core safety functions such as real-time bradycardia or ventricular tachycardia alarms. The trade-off lies in model capacity: state-of-the-art PPG foundation models can exceed hundreds of millions of parameters, and even aggressively quantized variants demand memory footprints that strain the typical smartwatch's application processor and battery budget. Dynamic voltage and frequency scaling, tiling strategies, and heterogeneous compute offloading to neural processing units offer partial remedies but introduce their own failure modes, including thermal throttling that may delay inference during continuous operation. For sustained monitoring over weeks, the energy envelope becomes a first-class constraint, linking the carbon footprint of edge AI directly to the clinical utility of the system.

A more flexible alternative is a collaborative edge-cloud architecture in which a lightweight embedding encoder runs on-device and transmits a compressed, privacy-preserving representation of each PPG window to an edge server or cloud instance where a larger foundation model performs the bulk of the computation. This paradigm aligns with emerging standards for split learning and inference offloading, where sensitive raw signals are not transmitted but the learned representations may still leak identifiable information if not protected by differential privacy or secure aggregation [18]. The communication channel itself becomes part of the adversarial surface, susceptible to man-in-the-middle attacks that tamper with embeddings or replay outdated signals to deceive the server-side agent. For life-critical arrhythmia detection, any interruption in connectivity must be handled gracefully through a deterministic fallback local classifier that provides a safety net at reduced sensitivity. This layered defense-in-depth approach mirrors well-established principles in avionics and industrial control systems, but its application to consumer health devices remains under-explored in the literature. An emerging body of work on edge-centric computing for healthcare IoT articulates frameworks for quality-of-service orchestration that could be extended to incorporate adversarial anomaly detection at the network layer, though their integration with clinical decision logic is nascent [19].

Federated training represents a complementary architectural pattern wherein PPG foundation models are collaboratively trained across a fleet of wearables without centralizing raw data. This preserves user privacy at the cost of increased communication rounds and vulnerability to model poisoning attacks by compromised clients. Robust aggregation mechanisms, such as trimmed mean or Krum, offer partial mitigation, but the heterogeneity of device hardware, sampling rates, and user behaviors complicates convergence guarantees. A system-level design must therefore weigh the benefits of on-device personalization against the risks of a

coordinated attack that slowly biases a global cardiovascular model across thousands of devices. We argue that trustworthiness at the edge must be woven into every tier: from on-sensor signal quality assessment that rejects adversarial motion-induced transients, through encrypted representation transport, to cloud-side agent orchestration that cross-references multiple physiological signals before issuing a clinical suggestion. This holistic perspective motivates a deep integration of the PPG foundation model's inductive biases with adversarial hardening, a topic we explore in the following sections.

4. PPG Foundation Models and Data-centric Learning

The core promise of PPG foundation models is to distill a universal physiological embedding space from large corpora of unlabeled or weakly labeled photoplethysmographic recordings, thereby equipping downstream tasks with robustness against the myriad sources of variability that confound narrow supervised models. Pretext tasks such as masked signal reconstruction, heart rate perturbation prediction, and cross-device temporal contrastive learning encourage the model to capture fundamental cardiac dynamics -- systolic upstroke morphology, dicrotic notch positioning, and respiratory modulation -- that are conserved across individuals, skin tones, and sensor wavelengths. A recent formulation that leverages statistical-prior informed generative masking architectures, as described by Guo et al. in the SIGMA-PPG framework, demonstrates how clinical knowledge about PPG signal characteristics can be encoded directly into the masking pattern during self-supervised pretraining [7]. Rather than randomly masking time steps, the approach selectively obscures segments that correspond statistically to physiological landmarks, forcing the model to infer plausible atrial filling or arterial compliance behavior from context. This domain-aware regularization not only accelerates convergence but imposes a structured inductive bias that is inherently less sensitive to high-frequency noise and sensor artifacts.

The translational value of such a foundation model lies in its ability to serve as a shared backbone for a heterogeneous suite of downstream clinical tasks, including atrial fibrillation detection, blood pressure trend estimation, cardiac output change monitoring, and sleep-disordered breathing screening. When deployed at the edge, the model's encoder can be frozen and paired with lightweight task heads that are updated periodically via federated fine-tuning, thereby constraining the communication budget and reducing the attack surface associated with transmitting full model gradients. However, representation collapse remains a pressing concern in resource-constrained settings: aggressive quantization can erode the fine-grained temporal features that discriminate between a premature ventricular contraction and an artifact, especially under adversarial perturbations deliberately crafted to exploit quantization error boundaries. Mitigation strategies include stochastic rounding during quantization-aware training, forward-backward pass consistency regularization, and, critically, the integration of uncertainty estimation layers that allow the edge agent to abstain from high-risk predictions. The foundation model paradigm thus offers a pathway to clinically generalizable PPG intelligence, but only when its deployment is paired with a corresponding security architecture that treats the encoder not as infallible but as a probabilistic sensor within a larger decision loop.

5. Adversarial Security in Medical AI Agents

The medical AI agent deployed for cardiovascular monitoring at the edge must be conceptualized as a system that not only processes signals but interacts with patients, clinicians, and electronic health record services through natural language and structured recommendations. This expansion of the interface multiplies the adversarial attack surface.

Beyond classical evasion attacks that add carefully crafted noise to a PPG recording to induce misclassification, the agent is susceptible to prompt-level manipulation if it employs a large language model core for synthesizing explanations or coaching messages [8]. An adversarially injected string, disguised as a plausible user query or a sensor metadata field, could cause the agent to ignore a genuine arrhythmic episode or produce an alarm that triggers unnecessary emergency room visits. Medical AI agents thus demand multi-modal adversarial robustness: the sensor stream, the language prompt, and the fusion logic that combines them must each be hardened.

Traditional adversarial training, which injects perturbed examples into the learning procedure, has demonstrated improved empirical robustness in image and text domains, but its application to PPG signals is complicated by the non-stationary, high-dimensional nature of optical time-series and the clinical imperative to preserve subtle pathological signatures [10]. Adding norm-bounded perturbations during training can inadvertently teach the model to ignore the very low-amplitude morphological details that are diagnostically critical, such as the presence of a wave reflection in arterial stiffness assessment. Certified defenses based on randomized smoothing offer a complementary approach, providing probabilistic guarantees that a model's prediction will remain stable within a radius of the input, but extending these guarantees to temporally correlated sequences with realistic clinical perturbation models -- such as motion artifacts, sensor displacement, or skin temperature shifts -- remains an open research challenge [11]. A promising direction is to marry the structured priors embedded in PPG foundation models with conformal prediction techniques that yield distribution-free confidence sets with finite-sample coverage, allowing the agent to calibrate its decisiveness to the current level of adversarial risk. For instance, if the smoothed prediction set contains both sinus rhythm and atrial fibrillation, the agent can escalate to a higher-fidelity sensor or request a manual confirmation rather than making an ambiguous autonomous decision.

The secure integration of large language model reasoning into the medical agent pipeline requires a distinct set of hardening measures. Hu's recent work on security enhancement for adversarially robust medical decision agents provides a layered strategy that combines input sanitization, context integrity verification, and output consistency checks [8]. Input sanitization can range from regular expression filtering to semantic anomaly detection that flags prompts deviating from expected medical interaction distributions. Context integrity verification enforces that the agent's dialogue state and the sensor-derived clinical facts remain aligned, rejecting generated text that contradicts the most recent physiological evidence. Such measures, while computationally lightweight, become essential when the agent operates on an edge device where full safety alignment via reinforcement learning from human feedback is infeasible due to model size limitations. Adversarial robustness in this setting is not merely a model property but an emergent system behavior that depends on the tight coupling between the PPG encoder's uncertainty signals and the language module's guardrails, a coupling that must be engineered with the same rigor as the physiological signal processing pipeline itself.

6. Trustworthiness Engineering: Robustness, Fairness, and Governance

Trustworthiness in edge-native cardiovascular agents extends beyond adversarial robustness to encompass fairness, interpretability, data governance, and lifecycle accountability. A PPG foundation model trained predominantly on populations with lighter skin tones and low melanin concentrations risks systematic performance degradation for individuals with darker skin, because the optical absorption properties of melanin alter signal-to-noise ratios and

morphology. This form of algorithmic bias has been documented in pulse oximetry and, by extension, in PPG-based heart rate variability and arrhythmia classifiers. Fairness audits must therefore be embedded into the federated fine-tuning loop, monitoring per-subgroup calibration and false negative rates, and triggering model retraining when disparities exceed clinically meaningful thresholds [13]. Explainability techniques, such as integrated gradients over the PPG encoder's temporal attention maps, can further help clinicians understand whether a detected atrial fibrillation episode is driven by genuine p-wave absence patterns or by motion artifact sequences that the model has spuriously correlated with disease [12]. However, explanation fidelity under adversarial conditions must also be scrutinized: an adversary can construct a perturbation that simultaneously flips the prediction and its gradient-based saliency map, creating a convincing but entirely misleading clinical narrative. Trustworthy deployment demands that explanations be treated as auxiliary evidence rather than definitive justifications, a stance consistent with the emerging regulatory guidance on software as a medical device.

Governance frameworks for edge intelligence in cardiovascular care must address several unique tensions. The federated update mechanism, while privacy-preserving in principle, introduces provenance challenges: a model checkpoint may incorporate contributions from millions of devices, each with potentially unverified sensor calibration and label quality. Auditable aggregation ledgers, grounded in verifiable computation or transparent blockchain-based audit trails, can provide post-hoc accountability for model updates without revealing individual training samples [5]. Sustainability is another underappreciated dimension of trustworthiness. The continual retraining and on-device fine-tuning required to maintain fairness and adversarial robustness consume energy and computational resources, raising the total environmental cost of ownership for a wrist-worn monitoring service. Green AI principles advocate for a holistic accounting of the carbon footprint associated with both model training and inference, encouraging the selection of architectures that maximize accuracy per joule rather than accuracy alone [15]. In the cardiovascular context, where monitoring may span years, the embodied energy of device manufacturing and the operational energy of daily model inference cycles must be jointly optimized. This lifecycle perspective aligns trustworthiness with planetary health, reinforcing the ethical obligation of medical AI to avoid exacerbating the environmental determinants of cardiovascular disease.

Policy and regulatory oversight must also evolve to keep pace with the fluidity of edge-deployed medical agents. The distinction between a fixed diagnostic algorithm and a continuously learning agent that adapts its behavior through on-device feedback blurs the traditional premarket approval pathway. Regulators will need to develop frameworks for change control protocols that allow safe, locked-step updates while preventing silent deterioration of model performance [17]. An additional layer of societal concern surrounds the dual-use potential of adversarially hardened medical agents: the same techniques that protect a PPG sensor against malicious tampering could be repurposed to create resilient disinformation bots in health communication channels. Multidisciplinary ethics mapping reviews emphasize that transparency, stakeholder participation, and value-sensitive design must be embedded from the earliest phases of system conception, not retrofitted after widespread deployment [16]. Collectively, these dimensions define a trustworthiness engineering discipline that is intrinsically interdisciplinary, spanning embedded systems, clinical cardiology, machine learning, ethics, and health policy.

7. Deployment, Sustainability, and Policy Implications

Translating the conceptual architecture of trustworthy edge intelligence into real-world clinical workflows requires confronting a series of operational hurdles that reveal deeper structural trade-offs. In resource-limited health systems, the assumption of ubiquitous high-bandwidth connectivity and generous device replacement cycles does not hold, and cardiovascular monitoring agents must be designed for intermittent synchronization and graceful degradation. A pragmatic deployment model involves a tiered service delivery: a baseline arrhythmia detection model that runs entirely on-device with strong adversarial filtering, supplemented by optional cloud-based specialist models accessible only when connectivity and patient consent permit. This tiered approach acknowledges the digital divide and ensures that patients in rural or underconnected regions are not excluded from the safety benefits of continuous monitoring. Validation of such systems across diverse geographic and socioeconomic settings will require adaptive clinical trial designs that capture real-world adherence patterns, sensor longevity, and user trust perceptions, alongside traditional sensitivity and specificity metrics.

Energy sustainability in edge-AI for healthcare demands careful consideration of the dynamic operating modes of the system. A PPG-based agent that performs full foundation model inference every five seconds for years will rapidly deplete the wearable's battery, potentially causing user abandonment. Duty cycling, where the model activates only when preliminary signal quality metrics or heart rate changes cross adaptive thresholds, can reduce energy consumption by an order of magnitude without sacrificing safety. The statistical priors embedded in the PPG foundation model can directly inform these quality metrics, enabling a virtuous cycle where the model's confidence gates its own compute usage. Furthermore, the lifecycle carbon cost can be reduced through model sharing across applications: the same PPG encoder backbone, once trained, can support cardiovascular monitoring, stress quantification, and sleep staging without proportional increases in compute. This multi-task efficiency aligns with the spirit of Green AI and provides a counter-argument to the often-leveled criticism that large foundation models are inherently unsustainable [15].

Policy frameworks for edge-deployed medical AI agents will need to reconcile the tension between continuous safety monitoring and personal autonomy. A patient who wears a cardiovascular monitor may desire the reassurance that an AI agent will detect a dangerous arrhythmic event and trigger an emergency response, yet simultaneously resist the notion that the agent might autonomously contact a physician or emergency services based on a false positive. Consent architectures must therefore be dynamic and granular, allowing users to configure the level of agent autonomy and the communication pathways it is authorized to use. The adversarial security measures we have described -- input sanitization, context integrity checks, and certified robustness envelopes -- serve a dual purpose here: they also function as technical safeguards that enforce the user's declared autonomy boundaries, preventing both external attackers and system malfunctions from violating consent. As medical AI agents move closer to conversational interfaces that coach patients on medication adherence or lifestyle changes, the regulatory boundary around the practice of medicine will be tested, demanding updated licensure and liability frameworks that recognize agency distributed across human clinicians, edge devices, and cloud-based foundation models [17]. The roadmap toward trustworthy deployment is therefore as much a legal and social undertaking as it is a technical one, and the interdisciplinary collaboration must intensify if the promise of continuous, equitable cardiovascular monitoring is to be fulfilled without eroding the trust that undergirds the patient-clinician relationship.

8. Conclusion

This paper has presented a system-level examination of the convergence between PPG foundation models and adversarially secure medical AI agents within the context of continuous cardiovascular monitoring at the edge. We have argued that trustworthiness in such systems cannot be achieved by isolated optimization of model accuracy or adversarial robustness but requires a holistic engineering discipline that spans architectural co-design, representation learning, multi-modal threat modeling, fairness auditing, and sustainable lifecycle management. The integration of statistical-prior informed generative masking, as exemplified by recent PPG foundation model research, provides a path toward generalizable physiological encoders that are intrinsically less vulnerable to high-frequency perturbations. When coupled with layered adversarial hardening for medical decision agents that employ large language model reasoning, the resulting system demonstrates a potential to balance clinical safety with data privacy and autonomy. Yet substantial gaps remain in certified robustness for time-series data, formal verification of agent safety under adaptive attacks, and equitable deployment in under-resourced settings. Addressing these gaps will require sustained collaboration across the machine learning, clinical engineering, regulatory, and ethics communities. The vision of a wrist-worn device that continuously, privately, and securely monitors cardiovascular health and engages the patient in an informed dialogue is within reach, provided that trustworthiness is elevated from a desirable attribute to a non-negotiable design constraint.

References

1. Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge computing: Vision and challenges. IEEE Internet of Things Journal, 3(5), 637–646.

2. Perez, M. V., Mahaffey, K. W., Hedlin, H., Rumsfeld, J. S., Garcia, A., Ferris, T., ... & Turakhia, M. P. (2019). Large-scale assessment of a smartwatch to identify atrial fibrillation. New England Journal of Medicine, 381(20), 1909–1917.

3. Reiss, A., Indlekofer, I., Schmidt, P., & Van Laerhoven, K. (2019). Deep PPG: Large-scale heart rate estimation with convolutional neural networks. Sensors, 19(14), 3079.

4. Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.

5. Rieke, N., Hancox, J., Li, W., Milletarì, F., Roth, H. R., Albarqouni, S., ... & Maier-Hein, L. (2020). The future of digital health with federated learning. npj Digital Medicine, 3(1), 1–7.

6. Moor, M., Banerjee, O., Abad, Z. S. H., Krumholz, H. M., Leskovec, J., Topol, E. J., & Rajpurkar, P. (2023). Foundation models for generalist medical artificial intelligence. Nature, 616(7956), 259–265.

7. Guo, Z., Chen, T., Jiao, Y., Pan, Y., Hu, X., & Ferrario, M. (2026). SIGMA-PPG: Statistical-prior Informed Generative Masking Architecture for PPG Foundation Model. arXiv preprint arXiv:2601.21031.

8. Hu, S. (2026). Research on Security Enhancement Methods for Adversarial Robust Large Language Model Intelligent Agents for Medical Decision-Making Tasks. arXiv preprint arXiv:2605.08257.

9. Finlayson, S. G., Bowers, J. D., Ito, J., Zittrain, J. L., Beam, A. L., & Kohane, I. S. (2019). Adversarial attacks on medical machine learning. Science, 363(6433), 1287–1289.

10. Madry, A., Makelov, A., Schmidt, L., Tsipras, D., & Vladu, A. (2018). Towards deep learning models resistant to adversarial attacks. In International Conference on Learning Representations.

11. Cohen, J. M., Rosenfeld, E., & Kolter, J. Z. (2019). Certified adversarial robustness via randomized smoothing. In International Conference on Machine Learning (pp. 1310–1320). PMLR.

12. Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82–115.

13. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447–453.

14. Majumder, S., Mondal, T., & Deen, M. J. (2017). Wearable sensors for remote health monitoring. Sensors, 17(1), 130.

15. Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). Green AI. Communications of the ACM, 63(12), 54–63.

16. Morley, J., Machado, C. C., Burr, C., Cowls, J., Joshi, I., Taddeo, M., & Floridi, L. (2020). The ethics of AI in health care: A mapping review. Social Science & Medicine, 260, 113172.

17. Char, D. S., Shah, N. H., & Magnus, D. (2018). Implementing machine learning in health care—addressing ethical challenges. New England Journal of Medicine, 378(11), 981–983.

18. Xu, J., Glicksberg, B. S., Su, C., Walker, P., Bian, J., & Wang, F. (2021). Federated learning for healthcare informatics. Journal of Healthcare Informatics Research, 5(1), 1–19.

19. Amin, S. U., Hossain, M. S., & Muhammad, G. (2020). Edge-centric computing framework for healthcare IoT. IEEE Access, 8, 2301–2315.