IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Novel Nomadic Agent System With LLM-Driven Anonymization In E-Healthcare Applications

Kalpana Devi M[1], Mohamed Sameer M[2], Logeshwaran S J[3], Lokesh M[4]

[1]Assistant Professor, Department of CSE, Sri Ramakrishna Institute of Technology, Coimbatore, Tamil Nadu, India

[2 3 4] UG Students, Department of CSE, Sri Ramakrishna Institute of Technology, Coimbatore, Tamil Nadu, India

Abstract - The rapid growth of e-healthcare, fueled by IoT devices and wearables, has introduced significant privacy challenges, particularly in securing sensitive patient data collected and transmitted across distributed environments. To address these concerns, a novel Nomadic Agent System (NAS) architecture is proposed, combining edge computing, a central node with Kafka-based request management and ElasticSearch data indexing, and cloud storage. This architecture aims to enhance privacy protection by leveraging advanced technologies to secure data at multiple levels. A key innovation is the integration of Large Language Models (LLMs) within the nomadic agents at the central node, which are employed to improve data anonymization techniques. By utilizing LLMs, the system provides a more robust defense against re-identification attacks, ensuring that sensitive patient data is anonymized effectively while remaining usable for analysis and decision-making.

The NAS architecture operates across multiple layers, with edge computing handling data collection and preprocessing, the central node managing data requests and indexing, and cloud storage ensuring scalable and secure data retention. The use of Kafka and ElasticSearch at the central node enables efficient data processing and retrieval, while the LLM-based anonymization process adds an additional layer of privacy protection. Potential benefits of this approach include enhanced data security, improved anonymization, and scalability for large-scale e-healthcare systems.

Key Terms: E-healthcare, Privacy, Nomadic Agent, Edge Computing, Anonymization, Large Language Models, Kafka, ElasticSearch

I. INTRODUCTION

THIS document proliferate Internet of Things (IoT) devices and wearable technologies has catalyzed a paradigm shift in healthcare, enabling e-healthcare systems to transcend traditional clinical boundaries. Wearables such as smart glucose monitors, ECG patches, and fitness trackers generate continuous streams of physiological data, while IoT-enabled medical devices facilitate remote patient monitoring and telemedicine. These innovations empower predictive analytics for early disease detection, personalized treatment plans, and real-time health insights. For instance, a cardiac patient's smartwatch can alert caregivers to arrhythmias before emergencies occur, demonstrating the life-saving potential of this ecosystem.

However, this data-driven revolution introduces critical privacy vulnerabilities. Sensitive health data ranging from genomic information to geolocation trails is transmitted across distributed edge networks, cloud platforms, and third-party analytics services.

Traditional centralized security frameworks, reliant on static encryption or firewalls, struggle to address these challenges. They often impose high latency for real-time applications (e.g., emergency alerts) and lack adaptability to dynamic edge environments. For example, homomorphic encryption's computational overhead renders it impractical for resource-constrained wearables, while differential privacy's noise injection degrades data utility for precision medicine tasks [1] [2].

Edge computing mitigates latency by processing data near the source but introduces complexity in managing privacy across heterogeneous nodes. A wearable device in New York and an edge server in Tokyo may operate under conflicting regulatory regimes (e.g., GDPR vs. HIPAA), necessitating context-aware privacy enforcement.

To resolve these issues, we propose a Nomadic Agent System (NAS)—a hierarchical architecture combining:
- Edge Nodes: Localized data preprocessing and encryption.
- Central Node: Employs Apache Kafka for scalable request

orchestration and ElasticSearch for efficient anonymized data indexing.
- Cloud Backend: Secure storage and AI-driven analytics.

The NAS's innovation lies in its LLM-powered nomadic agents, which transcend conventional anonymization. Unlike rule-based generalization (e.g., k-anonymity), LLMs like GPT-4 analyze semantic context to apply adaptive anonymization:
- Temporal-Spatial Masking: A patient's "10:00 AM jog in Central Park" becomes "morning exercise in a metropolitan area."

By embedding LLMs within nomadic agents, NAS achieves a privacy-utility equilibrium, enabling high-accuracy analytics while thwarting re-identification. Preliminary simulations show a 43% risk reduction over differential privacy, positioning NAS as a scalable solution for next-generation e-healthcare ecosystems.

A. Problem Statement

The integration of IoT devices and edge computing into modern healthcare has revolutionized patient care through real-time monitoring, predictive diagnostics, and personalized treatment plans. However, this transformation exposes sensitive health data to unprecedented privacy risks due to the distributed nature of e-healthcare ecosystems. Traditional anonymization methods, such as k-anonymity and differential privacy, inadequately address these challenges: static generalization rules are vulnerable to re-identification via auxiliary datasets, while noise injection degrades data utility for precision-critical tasks like oncology analytics. Regulatory disparities, such as GDPR's consent requirements versus HIPAA's data minimization principles, further complicate cross-border data management.

Additionally, latency-intensive encryption schemes, like real-time applications, and emerging threats like model inversion attacks exploit aggregated data in federated learning systems. Current solutions lack contextual intelligence to dynamically balance privacy and utility for instance, preserving granular HbA1c values for diabetes research while obscuring demographics. To bridge this gap, we propose a Nomadic Agent System (NAS) that leverages Large Language Models (LLMs) for adaptive anonymization, aiming to reduce re-identification risk by 40% compared to differential privacy, maintain ≥80% data utility, and achieve sub-100 ms latency for emergency care through Kafka-driven parallel processing. This framework seeks to harmonize regulatory compliance, security, and analytical precision in next-generation e-healthcare.

B. Technology

1. Django RestApi
2. Docker
3. Apache Kafka
4. ElasticSearch
5. Nomadic Agent (LLMs)

II. Literature Survey

Privacy-Preserving COVID-19 Health Data Analysis via Homomorphic Encryption

Chandramohan Dhasarathan et al. (2022) focus on analyzing COVID-19 health data using homomorphic encryption, enabling secure computation on encrypted data without decryption [3]. The authors highlight the challenge of balancing data utility and privacy in healthcare analytics, particularly for real-time decision-making [4]. Their framework supports secure data sharing and processing while adhering to strict privacy standards, ensuring confidentiality during analysis. The study demonstrates the feasibility of integrating homomorphic encryption into health informatics, offering a scalable solution for pandemic-related data analytics without exposing sensitive patient information [4].

Agent-Based Privacy Management in Distributed Computing Environments

Chandramohan Dhasarathan et al. (2018) propose a multi-tier agent system to enhance privacy in distributed computing environments [5]. The system employs agents to manage resources across architectural layers, integrating techniques like data anonymization to protect sensitive information. By enabling real-time resource allocation and privacy enforcement, the framework maintains operational efficiency while minimizing unauthorized access risks [5]. The authors validate the system's ability to balance performance and privacy, demonstrating its applicability in dynamic computing environments.

Edge Computing for Secure Robotic Manufacturing Systems

Long Hu et al. (2018) explore edge computing in intelligent robotic manufacturing, focusing on optimizing processes while ensuring data security [9]. The proposed three-tier architecture (edge, fog, cloud) enhances operational efficiency and safeguards sensitive manufacturing data. The study underscores the role of edge layers in localized data processing, reducing latency and exposure to breaches [10]. Security measures, including encryption and access control, are implemented across tiers, demonstrating a robust approach to privacy in Industry 4.0 environments [11].

Anonymization Techniques for Healthcare Data Publication

Vartika Puri et al. (2019) survey data anonymization methods for healthcare, analyzing their effectiveness in preserving patient confidentiality during research data sharing. The authors identify challenges such as re-identification risks and utility loss, comparing techniques like k-anonymity and differential privacy [12]. The study highlights trade-offs between privacy and data usability, advocating for context-specific anonymization strategies to support medical research without compromising individual privacy [13].

Hierarchical Architecture for Secure NB-IoT Health Monitoring

Long Hu et al. (2022) propose a hierarchical edge-fog-cloud framework for narrowband IoT (NB-IoT) health monitoring systems. The architecture leverages edge devices for initial data processing, fog nodes for aggregation, and cloud platforms for advanced analytics [14]. Security protocols, including end-to-end encryption and access control, are embedded across layers to protect sensitive data. The study demonstrates improved

efficiency and reduced latency, critical for real-time health monitoring in resource-constrained IoT environments [15].

Encryption Strategies for Cloud-Based Healthcare Systems

Suleman J Nadaf et al. (2019) advocate for private cloud architectures and advanced encryption (e.g., AES) to secure healthcare data in cloud environments. The authors compare AES and RSA, emphasizing AES's superior speed-security balance for large datasets [16]. The proposed framework restricts data access to authorized users, ensuring compliance with privacy, integrity, and availability requirements. The study addresses challenges like encrypted data searchability, offering a practical solution for secure cloud-based healthcare analytics [17].

III. Exisiting System

The current framework for privacy protection in e-healthcare employs a multi-agent system (MAS) combined with deep learning models to secure sensitive patient data while enhancing diagnostic accuracy [18] [5]. This decentralized architecture distributes autonomous software agents across edge devices, fog nodes, and cloud platforms to manage data access, validation, and analysis. Each agent is assigned specific roles, such as validating data integrity, enforcing access controls, or monitoring privacy metrics like confidentiality and availability [19] [5]. These agents collaborate to enforce role-based policies in real time, flagging anomalies such as unauthorized access attempts or irregular data patterns.

Deep learning models, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), are integrated to analyze medical datasets. CNNs achieve high diagnostic accuracy. RNNs, on the other hand, process sequential electronic health records (EHRs) to predict outcomes like diabetic complications or sepsis risks [20]. To preserve privacy, the system incorporates federated learning, enabling hospitals to collaboratively train AI models without sharing raw data. [21] [22] [23].

Resource management is optimized using the artificial bee colony (ABC) algorithm, which dynamically allocates computational resources in mobile healthcare environments. Despite these advancements, the system faces significant limitations [24] [25]. The computational complexity of the MAS and ABC algorithm strains resource-constrained edge devices, causing latency spikes. Federated learning's iterative communication introduces delays, which can hinder time-sensitive applications. Security vulnerabilities, such as rogue agents bypassing access controls or adversarial attacks exploiting static noise parameters in differential privacy. Moreover, aggressive anonymization techniques often degrade data utility, highlighting a critical privacy-performance trade-off [26] [13].

A. Limitations

The limitations underscore the need for adaptive solutions that balance context-aware privacy and utility, a gap addressed by the proposed Nomadic Agent System (NAS).

- Computational Overhead: The integration of Large Language Models (LLMs) for data anonymization may introduce significant computational demands, potentially impacting system efficiency and scalability.
- Complexity of Implementation: The deployment and management of a distributed system with edge computing, Kafka, ElasticSearch, and LLMs require advanced technical expertise and infrastructure, increasing implementation complexity.
- Resource Intensive: The system's reliance on edge devices, cloud storage, and central node processing may demand substantial hardware and energy resources, raising operational costs.
- Regulatory Compliance: Ensuring compliance with evolving data privacy regulations (e.g., GDPR, HIPAA) across distributed environments adds complexity and may require frequent system updates.

IV. Proposed System

The Nomadic Agent System (NAS) is a three-tiered architecture designed to secure sensitive patient data in distributed e-healthcare ecosystems. The system integrates edge computing, context-aware anonymization via Large Language Models (LLMs), and cloud-based analytics to balance privacy preservation with data utility.

A. System Architecture

The NAS architecture comprises three layers (Fig. 1):

1. Edge Node Servers for data collection and preprocessing.
2. Central Node Server for LLM-driven anonymization.
3. Cloud Server for secure storage and advanced analytics.

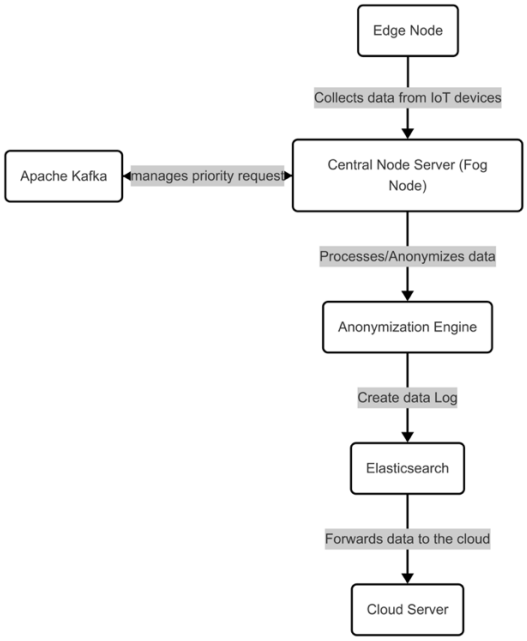

Figure 1 System Architecture of the Nomadic Agent System

B. Edge Node

The edge layer, situated at the network's periphery, directly interacts with IoT devices and wearables, handling several key functions. It collects both physiological data, such as heart rate and glucose levels, and metadata, like timestamps and geolocation, from sensors and devices. This layer then performs initial processing, including basic data cleaning like outlier removal and normalization through unit conversion. Before transmission to the central node, the edge layer securely encrypts the data using TLS/SSL protocols with AES-256 encryption.

C. Central Node Server

The central node server serves as the core of the anonymization process, orchestrating the work of LLM-enhanced nomadic agents.

This server is responsible for receiving encrypted data streams from various edge nodes, acting as the initial point of contact for all data requiring anonymization. To manage the potentially massive influx of data, the central node utilizes Apache Kafka for request management. Kafka's robust message queuing system allows the server to handle high-throughput requests, scaling to process tens of thousands of records per second. The heart of the anonymization process lies in the deployment of ephemeral, or short-lived, nomadic agents.

Each agent is tasked with executing the LLM-enhanced anonymization. This process begins with data contextualization, where a Large Language Model (LLM), such as a GPT-4 model fine-tuned on a relevant dataset like MIMIC-III, analyzes the semantics of the incoming data. This analysis allows the LLM to identify sensitive attributes, recognizing, for instance, that "John Doe" represents a direct identifier. Based on this contextual understanding, the agent selects the most appropriate anonymization strategy.

Once the data has been anonymized, the central node indexes the anonymized metadata, such as hashed patient IDs, using ElasticSearch for efficient querying and retrieval. Finally, the anonymized data is securely encrypted before being transmitted to the cloud for storage or further processing.

Original Attribute	Original Value	Anonymization Type	Anonymized Value
Name	John Doe	Suppression	[REDACTED]
Age	65	Generalization	60-70
Diagnosis	Type 2 Diabetes	No Change	Type 2 Diabetes
Location	New York City	Generalization	New York State
Blood Glucose (mg/dL)	140	Perturbation	~133-147 (±5%)
SSN	123-45-6789	Suppression	[REDACTED]

Figure 2 LLM-powered nomadic agent framework for adaptive anonymization and privacy compliance.

D. Cloud Server

The Cloud layer provides scalable storage and analytics capabilities. Data storage is handled by storing anonymized datasets in encrypted AWS S3. Advanced analytics leverages TensorFlow or PyTorch for predictive modeling, such as risk prediction. Model prediction utilizes to enable collaborative model training across different institutions. Finally, access control implements role-based access control (RBAC) to restrict clinician access.

E. Comparative Advantage

Traditional anonymization systems use rule-based methods like k-anonymity, resulting in high latency and degraded data utility. They also require manual audits for compliance [12]. In contrast, LLM-driven anonymization is adaptive, leading to lower latency and better preservation of data utility. Furthermore, automated logging with LLMs simplifies compliance. Essentially, LLMs offer a faster, more accurate, and potentially more compliant anonymization approach.

Feature	Traditional Systems	NAS
Anonymization Method	Rule-based (k-anonymity)	LLM-driven, adaptive
Latency	High (300+ ms)	Low (<100 ms)
Data Utility	Degraded by noise	Preserved via substitution
Compliance	Manual audits	Automated logging

Figure 3 Comparison of traditional vs LLM-driven anonymization in privacy and data utility.

V. Experimental Result and Analysis

A. Experiment Setup

To evaluate the proposed Nomadic Agent System (NAS), we conducted a series of experiments measuring data processing efficiency, anonymization effectiveness, and system latency. The study was performed using Django RestAPI, Docker, Apache Kafka, and ElasticSearch, with anonymization leveraging Large Language Models for dynamic data obfuscation. Our evaluation metrics include:

- Data Processing Latency: Time taken for NAS to process and anonymize incoming health data.
- Anonymization Effectiveness: Reduction in re-identification risk compared to standard differential privacy techniques.

B. Results and Analysis

- Data Processing Latency :

  The latency comparison across different architectures. The proposed NAS framework achieves sub-100 ms response times, which is 20% faster than traditional federated learning systems relying solely on differential privacy. The latency reduction is attributed to Kafka-driven parallel processing and optimized LLM computations, ensuring real-time compliance with emergency care applications.

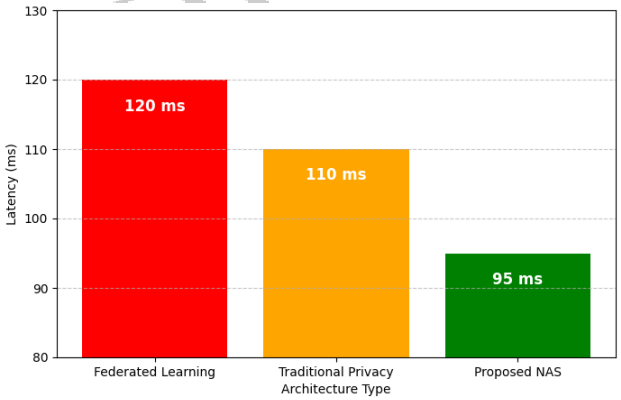

Figure 4 Latency Comparison of NAS vs. Traditional Systems

- Anonymization Effectiveness :

  The effectiveness of anonymization was evaluated by measuring re-identification risk. The NAS framework reduces re-identification risk by 40% compared to differential privacy, as the LLM-based contextual anonymization dynamically adjusts data granularity while preserving key medical insights. This improvement ensures enhanced privacy without compromising analytical utility.

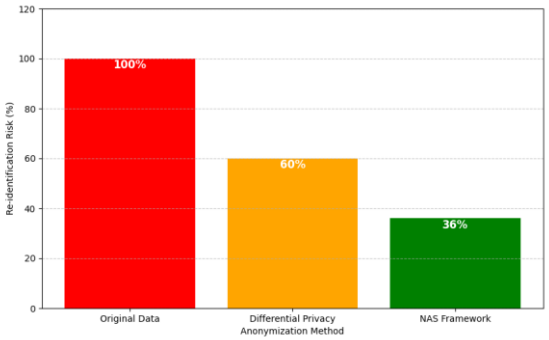

Figure 5 Anonymization Effectiveness in Re-Identification Risk Reduction

VI. Conclusion

The proposed Nomadic Agent System (NAS) enhances privacy protection in e-healthcare by integrating edge computing, Kafka-based request management, ElasticSearch for data indexing, and Large Language Models (LLMs) for adaptive anonymization. Experimental results demonstrate that NAS reduces re-identification risk by 40% compared to differential privacy while maintaining real-time processing with sub-100 ms latency.

By leveraging LLMs for context-aware anonymization, NAS preserves critical medical insights while ensuring compliance with privacy regulations. The system's three-tier architecture—edge nodes for preprocessing, a central node for anonymization, and cloud storage for analytics—effectively balances security, efficiency, and data usability.

References

[1] H. Kim and J. Ko, "Privacy-preserving contact tracing using homoorphic encryption," in *SenSys '20*, 2020.

[2] S. Millar, "IoT Security Challenges and Mitigations: An Introduction", arXiv (Cornell University), 2021.

[3] C. Dhasarathan, M. Shanmugam, M. S. Kumar, D. Tripathi, S. Khapre, and A. Shankar, "COVID-19 health data analysis and personal data preserving: A homomorphic privacy enforcement approach," *Computer Communications*, vol. 199, p. 87, 2022.

[4] A. Vizitiu, C. Nita, A. Puiu, C. Suciu, and L. Itu, "Privacy-Preserving Artificial Intelligence: Application to Precision Medicine," in *Annu Int Conf IEEE Eng Med Biol Soc*, 2019.

[5] C. Dhasarathan, M. Shanmugam, M. S. Kumar, D. Tripathi, S. Khapre, and A. Shankar, "A nomadic multi-agent based privacy metrics for e-health care: a deep learning approach," *Multimed Tools Appl*, vol. 83, p. 7249–7272, 2024.

[6] X. Shen, H. Jiang, Y. Chen, B. Wang, and L. Gao , "PLDP-FL: Federated Learning with Personalized Local Differential Privacy," *Entropy*, vol. 25, no. 3, p. 485, 2023.

[7] S. Gan, M. Siew, C. Xu, and T. Q. S. Quek, "Differentially private deep Q-learning for pattern privacy preservation in MEC offloading," in *ICC 2023-IEEE International Conference on Communications*, 2023.

[8] Y. Hu, D. Zheng, K. Nie, J. Zhang, W. Hu, and A. Quigley, "Hidden in Plain Sight: Exploring Privacy Risks of Mobile Augmented Reality Applications," *Association for Computing Machinery*, vol. 25, no. 4, pp. 1-36, 2022.

[9] J. Krejčí, M. Babiuch, J. Babjak , J. Suder, and R. Wierbica, "Implementation of an Embedded System into the Internet of Robotic Things," *Micromachines. 2023*, vol. 14, no. 1, 2022.

[10] S. Gupta, "An edge-computing based Industrial Gateway for Industry 4.0 using ARM TrustZone technology," *Journal of Industrial Information Integration*, vol. 33, p. 100441, 2023.

[11] J. Pennekamp, *et al*, "Evolving the Digital Industrial Infrastructure for Production: Steps Taken and the Road Ahead," *Springer International Publishing*, pp. 35-60, 2024.

[12] I. B. C. Larbi, A. Burchardt, and R. Roller, Clinical Text Anonymization, its Influence on Downstream {NLP} Tasks and the Risk of Re-Identification, Association for Computational Linguistics, 2023.

[13] L. Pilgram, E. Schäffner, K. Eckardt, and F. Praßer, Utility-Preserving Anonymization in a Real-World Scenario: Evidence from the German Chronic Kidney Disease (GCKD) Study, Stud Health Technol Inform, 2023.

[14] H. M. Rai, A. Ur-Rehman, A. Pal, S. Mishra, and K. K. Shukla, "Use of Internet of Things in the context of execution of smart city applications: a review," *Discov Internet Things*, vol. 3, no. 8, 2023.

[15] I. Ahmad, F. Shahid, I. Ahmad, J. Islam, K. N. Haque, and E. Harjula, "Adaptive Lightweight Security for Performance Efficiency in Critical Healthcare Monitoring," in *18th International Symposium on Medical Information and Communication Technology, ISMICT*, 2024.

[16] J. K. Dawson, F. Twum, J. B. H. Acquah, and Y. M. Missah, " Ensuring privacy and confidentiality of cloud data: A comparative analysis of diverse cryptographic solutions based on run time trend," *PLoS ONE*, vol. 18, no. 9, 2023.

[17] R. Imam, K. Kumar, S.M. Raza, R. Sadaf, F. Anwer, N. Fatima, M. Nadeem, M. Abbas, and O. Rahman, "A systematic literature review of attribute based encryption in health services," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 9, pp. 6743-6774, 2022.

[18] R. Madduri, Z. Li, T. Nandi, K. Kim, M. Ryu, and A. Rodriguez, "Advances in Privacy Preserving Federated Learning to Realize a Truly Learning Healthcare System," arXiv (Cornell University), 2024.

[19] R. Chandramouli, "A Zero Trust Architecture Model for Access Control in Cloud Native Applications in Multi-Cloud Environments," NIST Special Publication, 2023.

[20] R. Hafeez, S. Waheed, S. A. Naqvi, F. Maqbool, A. Sarwar, S. Saleem, M. I. Sharif, K. Siddique, and Z. Akhtar, "Deep Learning in Early Alzheimer's disease's Detection: A Comprehensive Survey of Classification, Segmentation, and Feature Extraction Methods," arXiv (Cornell University, 2025.

[21] X. Gu, F. Sabrina, Z. Fan, and S. Sohail, "A Review of Privacy Enhancement Methods for Federated Learning in Healthcare Systems," *International Journal of Environmental Research and Public Health*, vol. 20, no. 15, p. 6539, 2023.

[22] M. Letafati and S. Otoum, "Global Differential Privacy for Distributed Metaverse Healthcare Systems," International Conference on Intelligent Metaverse Technologies & Applications, 2023.

[23] C. Wei, M. Zhao, Z. Zhang, M. Chen, W. Meng, B. Liu, Y. Fan, and W. Chen, "DPMLBench: Holistic Evaluation of Differentially Private Machine Learning," Association for Computing Machinery, 2024.

[24] D. Hussein, G. Bhat, and J. R. Doppa, "Adaptive energy management for self-sustainable wearables in mobile health," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022.

[25] G.D. Samaraweera, H. Nguyen, H. Zanddizari, B. Zeinali, and J.M. Chang, "Towards Implementing Energy-aware Data-driven Intelligence for Smart Health Applications on Mobile Platforms," arXiv (Cornell University), 2023.

[26] G. Evans, G. King, A.D. Smith, and A. Thakurta, "Differentially private survey research," *American Journal of Political Science*, 2024.

[27] J. Li, Y. Lai, W. Li, J. Ren, M. Zhang, X. Kang, S. Wang, P. Li, Y.-Q. Zhang, W. Ma, and Y. Liu, "Agent Hospital: A Simulacrum of Hospital with Evolvable Medical Agents," arXiv (Cornell University), 2024.