# Quantifying and Preventing AI-Aware Privacy Leakage in Large-Scale Data Ecosystems: A Systematic Review

1st Prodip Kumer Das
*Department of Computer Science and Engineering*
*Chandigarh University*
Mohali-140413, Punjab, India
25YCS7037@cuchd.in
prodipkuet@gmail.com

2nd Dr. Amardeep Singh [0000-0002-7112-9042]
*University Institute of Engg.(AIT-CSE)*
*Chandigarh University*
Mohali-140413, Punjab, India
amardeep.e17149@cumail.in
amardeepsingh$_2$6@$yahoo.com$

***Abstract*—Now, AI runs on cloud platforms, edge systems with federated settings, and in large language model(LLM) pipelines or data-sharing services, which create even wider privacy leakage paths beyond classical database disclosure. This paper offers a systematic structured literature review of a curated, cost-effective reference corpus on quantification and prevention of privacy leakage in AI-enabled data ecosystems. The review ties together four strands of research which are often treated, respectively. Firstly, the privacy risk throughout the AI life cycle. Secondly, the measurement of the quantitative leakage. Thirdly, architectures of the privacy-preserving models, and finally, operational governance for real-world deployment. Our analysis demonstrates that state-of-the-art approaches are moving from static mechanisms based on anonymization to metric-aware protections, including information-theoretic leakage scores, cumulative differential privacy accounting, personalized privacy budgets, and benchmark-driven attack evaluation. In parallel, prevention methods are evolving beyond single homomorphic noise injection and instead are multi-layered defenses producing combinations of differential privacy, federated learning, quantization of weights, synthetic data generation, policy-driven automation and LLM controls. The review uncovers four itched-over gaps: fractured assessment metrics, shaky privacy-utility trade-offs, flimsy integration of technological controls and compliance processes, and low cross-context validation across cloud-based, edge computing (data processing at or near the source), federated learning (distributed machine-learning methods) and generative AI systems. The paper concludes by outlining a unified research agenda toward building AI-aware, quantifiable, usable privacy protection stacks.**

***Index Terms*—Privacy leakage, differential privacy, federated learning, large language models, synthetic data, big data ecosystems, systematic review**

## I. Introduction

AI systems increasingly rely on large-scale data ecosystems as opposed to isolated datasets. Across cloud analytics, edge intelligence, federated learning, Internet of Things infrastructures and generative AI workflows, data is collected, cleaned, shared, transformed, modeled and reused. Consequently, privacy leakage is no longer limited to the publication epoch; it can arise in data curation, model training and inference process, parameter hugging and prompt processing at runtime, synthetic release for benchmarking and also compliance operations [1]- [5].

Existing work has surveyed privacy challenges in the big data and AI life-cycle context [1]- [3], as well as examined specific privacy-preserving technologies (e.g., differential privacy, homomorphic encryption, federated learning, secure multi-party computation, synthetic data generation) [4], [5]. However, the literature is fragmented in at least one important respect: leakage quantification and leakage prevention are typically treated as separate problems. In practice, organizations need both. They require methods for quantifying privacy exposure to help prioritize risk, compare alternatives, allocate privacy budgets and justify controls. Technical and procedural mechanisms are also required that minimize leakage while not sacrificing utility, scalability, or regulatory usability [6]- [11].

This review moves beyond reviewing the literature on privacy leakage or attacks to directly address this gap by providing an umbrella where studies can be related around a central question: how can we quantify and prevent the leakage of privacy for AI-aware large-scale data ecosystems? The paper makes three contributions. First, it presents a unified taxonomy of leakage surfaces, measurement methods, and protection strategies over the entire AI pipeline. Second, they compare each quantification approach in a small number of core families ranging from information-theoretic models through utilitarian scoring and cumulative differential privacy accounting. Third, it draws attention to the operational transition in progress across the field: from spectrum-shift/step-disguising protective moats towards context-aware deployment & adaptive defensive stacks that need to accommodate federated, split, edge and LLM-based systems [12]- [18].

## II. Privacy Leakage Landscape in AI Data Ecosystems

A recurring theme in the surveyed literature is that risk to privacy lies all throughout an AI life cycle rather than being localized at any one step. Lifecycle-oriented studies see threats naturally associated to data acquisition, integration, labeling,

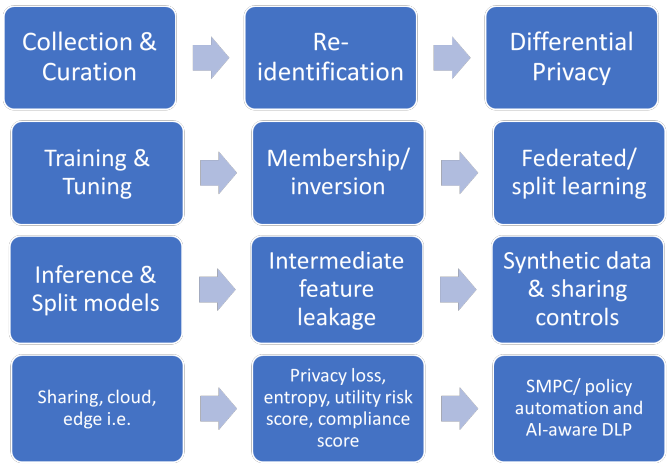


Fig. 1. Integrated taxonomy linking AI lifecycle layers, leakage vectors, quantification metrics, and prevention controls.

training, inference and downstream model use [1], [2]. Leakage in AI systems can happen via direct release of sensitive records, indirect inference of its attributes, reconstructing training examples, leaking intermediate representations and non-safe outputs by generative systems [2], [4], [13], [17], [18].

Three ecosystem trends compound the severity of this risk profile. As a result, while distributed learning methods mitigate raw-data centralization, they do not completely prevent leakage; updates to an ensemble of models, gradients or representations shared in the process can exfiltrate training data or user participation [11], [18]- [21]. Second, large language models produce an own surface in two layers: privacy risks can potentially breach both the training corpus as well as user prompts, context windows and generated responses [4], [13]. Third, edge and IoT deployments bring computation closer to sources of data, but they also increase the attack surface by introducing heterogeneous devices, resource constraints and hybrid edge-cloud coordination [25], [27].

Another key finding is that traditional data leakage prevention tools are becoming increasingly misaligned with AI operations. Arora demonstrates in [16] that traditional DLP solutions built on static rules find it difficult to capture dynamic data flows and the deployment of synthetic data for AI processing. Likewise, Wu claims in his work that risk scoring needs to be a function of user behavior, data sensitivity and contextual signals as opposed to fixed policy checks [14]. With our observations strengthen principles of AI-aware privacy protection that do not occur without a broad lifecycle visibility as shown in Figure 1 and strongly models & continuously assesses risk in attack mode flanking as this may impact all interactions with the data itself.

## III. PRIVACY LEAKAGE QUANTIFICATION

The review of the literature presented here shows a less binary model to quantify privacy, adopting instead more formal and graded measures of leakage. Previously, the idea about quantifying privacy stemmed from information theory and de-identification that defined privacy as uncertainty reduction and disclosure risk [6]. They then focused on recent investigations, followed by entropy-based measures (including group privacy preferences and domain-specific sharing constraints) [9], and split-inference analysis based on Fisher-approximated Shannon information to quantify leakage from intermediate model outputs [17]. And that matters because they treat privacy as an updatable quality of the system instead of a mushy design principle.

A second approach is a family of probabilistic and personalized modeling methods. Meng lays out a Markov-chain and Bayesian based formulation of privacy valuation for individual and group level [8], while Chen discusses secret specification-based differential privacy enabling users to specify what aspects of their information are worth stronger protection [22]. Zhang also establishes a distinct system of user credibility and data sensitivity in medical data publishing, existing alongside the privacy parameter allocation discussion to suggest ways that privacy settings can be effectively adjusted [26]. Such studies suggest a more context-sensitive perspective on privacy, where the same level of protection is not optimal for every user, or every dataset, or every task.

A third stream stresses utility-aware and workflow-aware quantification. Du, in [7], proposes a way to quantify the usability of masked data under generalization and noise addition, thus making the privacy–utility trade-off more explicit. PRIV-ML tackles iterative machine learning workflows and evaluates end-to-end guarantees through composition aware budget tracking over cumulative privacy loss [11]. Guo takes this quantification into the governance layer, implementing LLM automation to expedite the process of risk quantification for compliance processes of big-data platforms [23]. Arshad further extends the notion on privacy using a systematic perspective where it states that while quantifying privacy should reflect its legal and organizational compliance rather than being treated as an inadequately performed academic-mathematical exercise [15].

However, there's still no one measurement stack that the field agrees on. On one hand, information-theoretic methods are appealing for formal reasoning, but can be hard to operationalize over diverse pipelines. Utility-aware metrics facilitate deployment decisions, but may not correspond with adversarial guarantees. Personalized schemes make usability a bit better, but they create complexity, reducing the comparability and benchmarking between assets. The practical take away is that organizations require multi-layer quantification: dataset-level, model-level, workflow-level and governance level measurement should be integrated but not replaced.

## IV. PREVENTION STRATEGIES

The prevention literature reflects a similar pivot from single-mechanism defenses to multipoint protection architectures. Large-scale surveys [3], [4], [24] continue to attest that the main families of AI and machine learning privacy-preserving techniques fall under anonymization, encryption, differential

privacy, synthetic data (in its many forms), homomorphic encryption, federated learning [3], and secure multi-party computation. However, newer studies in the corpus increasingly shift from evaluating these mechanisms in isolation to ways they can be composed in realistic deployment settings.

Differential privacy is still the standard formal mechanism, but recent literature treats it as a design space rather than a recipe. The Kumar's article presents differential privacy as a basic tool to be tested against re-identification threats [10]. Thantharate shows that for long-lived ML pipelines, explicit budget accounting is necessary to prevent accumulation of privacy loss [11]. Han introduces dual-layer and layer-specific differential privacy for federated scenarios to minimize the degradation of performance due to naive noise perturbation [19]. They show how the combination of privacy and quantization can mitigate leakage and communication depends on the federated or edge system (Ardic: & Feng) [20], [21]. Taken together, these works show that the relevant question is no longer whether to employ DP, but rather how one should adjust budget allocation, layer choice and compression strategy for a given deployment context.

Another main prevention stream are synthetic data and privacy preserving data sharing. Osorio-Marulanda reviews privacy mechanisms and evaluation metrics for synthetic data generation, showing that the privacy claims of various approaches are highly dependent on how fidelity and disclosure are quantified [5]. Tang proposes a synthesis of the records for online sharing, without releasing original records directly [12]. Most importantly, these studies are relevant to large-scale ecosystems, where data sharing is operationally inevitable; if managed correctly (as we argue above) synthetic data reduces direct disclosure risk while maintaining analytical availability.

Generative AI and LLM use cases need extra measures of protection. Feretzakis examines privacy-preserving methods specifically curated for generative AI, such as DP, FL, HE, and secure computation, highlighting that complete privacy is ephemeral; it must be integrated with regulatory requirements [4]. Ullah, suggests a LLM framework that separates privacy preservation for data curation from training and contextual use [13]. Largely due to these contributions, it is clear that not all LLM privacy can be addressed as classical database anonymization since the model becomes an additional surface of potential disclosure.

Operational prevention is also driven by governance and detection. Wu uses two methods, anomaly detection and contextual scoring, as part of its framework to gauge leakage risk in enterprise spaces [14]. As a consequence standard DLP is too low-tech for AI pipelines and also need to be refreshed when dealing with synthetic data and fluid model-centric workflows [16]. In the realm of edge and IoT deployments, Khakpoor et al describe how privacy controls must meet such demands to justify usage but that they must also consider their specific environment in terms (context) e.g. latency, heterogeneity, limitations on devices [25], [27]. One clear trend is that the best prevention will be multi-layered and adaptive, combining formal privacy guarantees with architecture-aware defenses and operational monitoring among the prevention families shown in Figure 2.

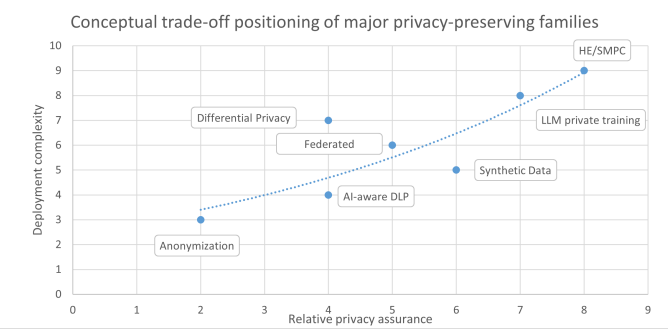


Fig. 2. Conceptual positioning of major prevention families across privacy assurance and deployment complexity.

## V. DISCUSSION AND RESEARCH AGENDA

In this survey literature, we have identified four major open research challenges. The first is evaluation fragmentation. A benchmark study on membership inference attacks by Niu et al [18] demonstrates that comparisons of systems are vulnerable to reversals contingent on the evaluation scenario and metric employed. This implies that privacy claims are high-level, which makes it difficult for practitioners to either compare defenses or determine deployment thresholds.

The second challenge is the unresolved privacy–utility–efficiency triangle. Conveniently as we approach prefix and 1D codes, utility-aware models and optimization methods are improving decision-making [7], [26], while privacy protections remain: imbuing accuracy, latency, or communication costs in federated, edge, and medical settings [19]- [21], [26]. Something that works for centralized analytics may be too expensive (in terms of latency or bandwidth) for edge intelligence, or too frail against the rigors of LLM learning.

The third challenge is different kinds of personalization versus standardization. Personalized privacy budgets and secret specifications more accurately capture user expectations [9], [22], [26], but are harder to audit, benchmark, and interoperability. This is both a technical and regulatory challenge since organizations require controls that are explainable to users and defensible to auditors [1], [15], [23].

A fourth challenge is that governance continues to be weakly integrated with technical privacy engineering. Demand for compliance-aware automation, risk scoring, and policy translation has come from the literature [1], [15], [16], yet most technical publications still evaluate defenses predominantly using task accuracy or attack success. Therefore, future work should focus on unifying the evaluation pipelines to submit privacy leakage, residual attack success, data utility, computational overhead and compliance interpretability in a coherent manner. From the synthesized evidence, a pragmatic research agenda arises as gaps shown in the Figure 3. Benchmarks of the future should harmonize attack models, contexts, and even metrics across centralized systems like federated &

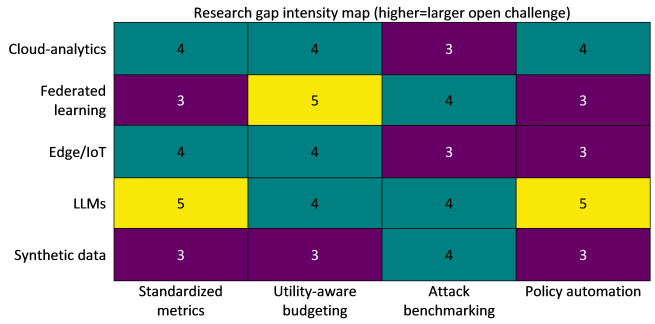


Fig. 3. Research gap intensity map synthesized from the reviewed literature.

split and LLMs. Second, privacy quantification will need to be reported on at multiple layers: input data, model internals, outputs and governance workflows. Thirdly, deployment pattern should dictate prevention stacks, not just algorithms For instance, large scale cloud analytics may focus on automating policies and synthetic data guarding mechanisms whereas edge or federated systems may need an optimization of DP, quantization and communication jointly. Finally, there is a need for dedicated evaluation protocols for LLM systems that address prompt privacy, training-data memorization and response-time leakage.

## VI. CONCLUSION

In this paper, we reviewed the landscape of privacy leakage quantification and mitigation in AI-aware, big data ecosystems through a structured aggregation of the provided bibliography corpus. The evidence suggests that privacy leakage should be understood as a lifecycle and ecosystem issue: it transpires during collection, learning, inference, sharing and governance; it occurs in centralized, federated, edge and generative AI settings.

The literature also indicates that there is an emerging feasibility of measurable privacy, though not yet unified. Known information-theoretic leakage metrics, Bayesian and entropy-based models, cumulative privacy accounting, utility-aware scoring and compliance oriented quantification only cover parts of the problem. On the prevention front, the best approach is a multilayered architecture that incorporates differential privacy, federated and split learning controls, synthetic data, model-aware protections and operational risk monitoring.

On a practical level the take-home message is that privacy engineering for AI can no longer be based on a single tool or a single metric. The future AI systems that preserve privacy will require integrated measurement (built into their design), defenses that are aware of deployment needs and evidence ready for guidance. This is the direction for privacy protection that is both theoretically sensible and operationally credible at ecosystem scale.

## REFERENCES


[1] T. Timan and Z. Mann, "Data Protection in the Era of Artificial Intelligence: Trends, Existing Solutions and Recommendations for Privacy-Preserving Technologies," The Elements of Big Data Value, pp. 153-175, 2021, doi: 10.1007/978-3-030-68176-0_7.

[2] S. Shahriar, S. Allana, S. M. Hazratifard, and R. Dara, "A Survey of Privacy Risks and Mitigation Strategies in the Artificial Intelligence Life Cycle," IEEE Access, vol. 11, pp. 61829-61854, 2023, doi: 10.1109/ACCESS.2023.3287195.

[3] H. Patel, A. Patel, and A. Patel, "A Comprehensive Analysis of Privacy-Preserving Techniques in Machine Learning," 2024 1st International Conference on Advances in Computing, Communication and Networking (ICAC2N), pp. 1836-1841, 2024, doi: 10.1109/ICAC2N63387.2024.10895292.

[4] G. Feretzakis, K. Papaspyridis, A. Gkoulalas-Divanis, and V. S. Verykios, "Privacy-Preserving Techniques in Generative AI and Large Language Models: A Narrative Review," Information, vol. 15, no. 11, p. 697, 2024, doi: 10.3390/info15110697.

[5] P. A. Osorio-Marulanda, G. Epelde, M. Hernandez, I. Isasa, N. M. Reyes, and A. B. Iraola, "Privacy Mechanisms and Evaluation Metrics for Synthetic Data Generation: A Systematic Review," IEEE Access, vol. 12, pp. 88048-88074, 2024, doi: 10.1109/ACCESS.2024.3417608.

[6] Z. Zhang, Z. Lu, and Y. Tian, "Data Privacy Quantification and De-identification Model Based on Information Theory," 2019 International Conference on Networking and Network Applications (NaNA), pp. 213-222, 2019, doi: 10.1109/NaNA.2019.00046.

[7] H. Du and J. Liu, "A Method for Quantifying Data Utility in Privacy-Preserving Scenarios," 2024 IEEE 2nd International Conference on Electrical, Automation and Computer Engineering (ICEACE), pp. 20-24, 2024, doi: 10.1109/ICEACE63551.2024.10898248.

[8] X. Meng, "Privacy Quantification Model Based on Markov Chain and Bayesian Theorem," 2024 7th International Conference on Computer Information Science and Application Technology (CISAT), pp. 648-654, 2024, doi: 10.1109/CISAT62382.2024.10695388.

[9] H. Zhu, X. Huang, P. Yu, and Y. Zhai, "Privacy Protection Intensity Measure for Energy Big Data Based on Information Entropy and Group Privacy Preferences," 2024 7th International Conference on Mechatronics and Computer Technology Engineering (MCTE), pp. 1722-1726, 2024, doi: 10.1109/MCTE62870.2024.11118009.

[10] G. S. Kumar, K. Preethie, S. Madhumitha, R. Sushma, and M. Nivaashini, "Data Privacy Preservation Using Differential Privacy and Re-Identification Attacks," 2024 International Conference on Science Technology Engineering and Management (ICSTEM), pp. 1-6, 2024, doi: 10.1109/ICSTEM61137.2024.10560868.

[11] P. Thantharate, D. A. Todurkar, and A. T, "PRIV-ML: Analyzing Privacy Loss in Iterative Machine Learning with Differential Privacy," 2024 IEEE Cloud Summit, pp. 107-112, 2024, doi: 10.1109/CloudSummit61220.2024.00024.

[12] Y. J. Tang and P. W. Chi, "Theseus Data Synthesis Approach: A Privacy-Preserving Online Data Sharing Service," IEEE Access, vol. 12, pp. 141130-141143, 2024, doi: 10.1109/ACCESS.2024.3467373.

[13] I. Ullah, N. Hassan, S. S. Gill, B. Suleiman, T. A. Ahanger, Z. Shah, J. Qadir, and S. S. Kanhere, "Privacy preserving large language models: ChatGPT case study based vision and framework," IET Blockchain, vol. 4, no. S1, pp. 706-724, 2024, doi: 10.1049/blc2.12091.

[14] X. Wu, J. Li, and W. Ren, "Risk Assessment Framework for Data Leakage Prevention Using Machine Learning Techniques," Artificial Intelligence and Machine Learning Review, vol. 5, pp. 55-66, 2024, doi: 10.69987/AIMLR.2024.50305.

[15] R. Arshad and M. R. Asghar, "Characterization and Quantification of User Privacy: Key Challenges, Regulations, and Future Directions," IEEE Communications Surveys & Tutorials, vol. 27, no. 5, pp. 3266-3307, 2025, doi: 10.1109/COMST.2024.3519861.

[16] S. Arora, V. Manral, D. S, and P. Chakraborty, "AI, Privacy, and Data Leakage: A Study of Current DLP Shortcomings," 2025 6th International Conference on Artificial Intelligence, Robotics and Control (AIRC), pp. 269-273, 2025, doi: 10.1109/AIRC64931.2025.11077549.

[17] R. Deng, Z. Lu, and Q. Duan, "Quantifying Privacy Leakage in Split Inference via Fisher-Approximated Shannon Information Analysis," arXiv, 2025, doi: 10.48550/arXiv.2504.10016.

[18] J. Niu, X. Zhu, M. Zeng, G. Zhang, Q. Zhao, C. Huang, Y. Zhang, S. An, Y. Wang, X. Yue, Z. He, W. Guo, K. Shen, P. Liu, L. Zhang, J. Ma, and Y. Zhang, "Comparing Different Membership Inference

Attacks With a Comprehensive Benchmark," IEEE Transactions on Information Forensics and Security, vol. 20, pp. 6592-6606, 2025, doi: 10.1109/TIFS.2025.3550070.

[19] J. Han, L. Wang, Z. Liu, B. Qin, K. Zhang, and W. Li, "PPFL: Privacy-Preserving Federated Learning Based on Differential Privacy and Personalized Data Transformation," IEEE Internet of Things Journal, vol. 12, no. 20, 2025, doi: 10.1109/JIOT.2025.3595533.

[20] E. Ardıç and Y. Genç, "Enhanced Privacy and Communication Efficiency in Non-IID Federated Learning With Adaptive Quantization and Differential Privacy," IEEE Access, vol. 13, pp. 54322-54337, 2025, doi: 10.1109/ACCESS.2025.3554138.

[21] C. Feng and P. Venkitasubramaniam, "Randomized Quantization for Privacy in Resource Constrained Machine Learning at-the-Edge and Federated Learning," IEEE Transactions on Machine Learning in Communications and Networking, vol. 3, pp. 395-419, 2025, doi: 10.1109/TMLCN.2025.3550119.

[22] J. Chen, C. Hu, Z. Liu, T. Xiang, P. Hu, and J. Yu, "Secret Specification Based Personalized Privacy-Preserving Analysis in Big Data," IEEE Transactions on Big Data, vol. 11, no. 2, 2025, doi: 10.1109/TBDATA.2024.3433433.

[23] Z. Guo, J. Cao, W. Ma, Q. Xu, B. Niu, and H. Li, "LLMQUA:Using LLM Automation to Enhance Risk Quantification in Data Processing for Big Data Platforms," 2025 11th IEEE International Conference on Privacy Computing and Data Security (PCDS), pp. 345-353, 2025, doi: 10.1109/PCDS65695.2025.00054.

[24] Q. Razi, R. Piyush, A. Chakrabarti, A. Singh, V. Hassija, and G. S. S. Chalapathi, "Enhancing Data Privacy: A Comprehensive Survey of Privacy-Enabling Technologies," IEEE Access, vol. 13, pp. 40354-40385, 2025, doi: 10.1109/ACCESS.2025.3546618.

[25] A. Shafee, S. R. Hasan, and T. A. Awaad, "Privacy and security vulnerabilities in edge intelligence: An analysis and countermeasures," Computers and Electrical Engineering, vol. 123, p. 110146, 2025, doi: 10.1016/j.compeleceng.2025.110146.

[26] D. Zhang, H. Xu, P. Li, Y. Sun, W. Jiang, X. A. Wang, and M. Zhou, "Privacy Parameter Setting and Usability Optimization Algorithm for Medical Data," IEEE Transactions on Consumer Electronics, vol. 71, no. 2, 2025, doi: 10.1109/TCE.2025.3569752.

[27] A. Tsouplaki, C. Fung, and C. Kalloniatis, "Enhancing IoT privacy with artificial intelligence: Recent advances and future directions," Internet of Things, vol. 34, p. 101752, 2025, doi: 10.1016/j.iot.2025.101752.