World Journal of Advanced Engineering Technology and Sciences
eISSN: 2582-8266
Cross Ref DOI: 10.30574/wjaets
Journal homepage: https://wjaets.com/

(RESEARCH ARTICLE)

Check for updates

Privacy-Centric AI Systems for Identifying and Securing Sensitive Information

Praveen-Kodakandla *

Independent Researcher.

World Journal of Advanced Engineering Technology and Sciences, 2025, 14(03), 594-604

Publication history: Received on 12 February 2025; revised on 24 March 2025; accepted on 27 March 2025

Article DOI: https://doi.org/10.30574/wjaets.2025.14.3.0126

Abstract

As data expands rapidly across multi-cloud, edge, and hybrid environments, maintaining privacy has become an increasingly dynamic challenge. Traditional rule-based protection models fail to adapt to the speed, scale, and contextual diversity of modern data flows. This study presents an AI-centric privacy architecture built to detect, classify, and safeguard sensitive data such as personally identifiable information (PII), health records, and financial details across distributed infrastructures. The framework integrates machine-learning-driven discovery, real-time remediation, and federated learning, enabling autonomous model updates without compromising local data ownership. Privacy enforcement modules—deployed as containerized microservices—perform inline operations like encryption, masking, and policy enforcement directly at ingestion points across clouds and edge nodes. Experimental outcomes show accuracy levels between 94–97% with response times under 120 ms, fully aligned with global compliance standards including GDPR, HIPAA, and CCPA. The architecture scales seamlessly through MLOps pipelines, ensuring enterprise integration with minimal manual oversight. This work underscores the need for context-aware, self-evolving AI systems that embed ethical and regulatory intelligence into every layer of distributed data processing.

Keywords: AI-Based Privacy; Federated Learning; Sensitive Data Detection; Contextual Classification; Real-Time Remediation; Distributed Governance; Data Compliance

1. Introduction

The explosive growth of data across digital ecosystems has redefined how organizations collect, process, and exchange information. Distributed computing—spanning cloud infrastructures, edge networks, mobile platforms, and hybrid systems—now powers diverse industries including healthcare, finance, autonomous systems, and smart manufacturing. While such architectures deliver scalability and speed, they also fragment responsibility for data privacy, creating new layers of complexity in security assurance and regulatory compliance.

Legacy privacy controls, built around centralized policies and static rule sets, are no longer effective in fluid data environments. Information now flows across multiple jurisdictions, data formats, and devices—often beyond the control of a single governing entity. This reality increases the risk of unauthorized exposure, policy violations, and cross-border compliance breaches.

With regulatory mandates such as GDPR, HIPAA, and regional privacy acts shaping data governance, organizations face mounting pressure to ensure transparency, consent, and lawful handling of personal data. Compounding this challenge is the surge in data-intensive innovations—AI, IoT, and machine-to-machine communication—that continuously generate and transform data with minimal human supervision. Manual redaction and role-based access systems struggle to keep pace with such autonomous data movement.

* Corresponding author: Praveen Kodakandla

The proposed AI-powered privacy system addresses these challenges through contextual data classification, deep-learning-based anomaly detection, and federated model training. The architecture is designed for decentralized learning, allowing privacy intelligence to evolve locally while preserving data sovereignty. Containers orchestrated across cloud, fog, and edge layers apply encryption, tokenization, and adaptive policies at runtime.

Rather than focusing on empirical surveys, this paper emphasizes architectural design principles, operational methodology, and theoretical validation. The subsequent sections describe the system blueprint, algorithmic logic, and performance evaluation under simulated distributed workloads. The results highlight efficiency gains, compliance adherence, and operational resilience.

In essence, embedding AI-driven privacy intelligence within data pipelines offers a transformative path toward self-governing, regulation-aware systems capable of maintaining security and trust in an increasingly connected digital landscape.

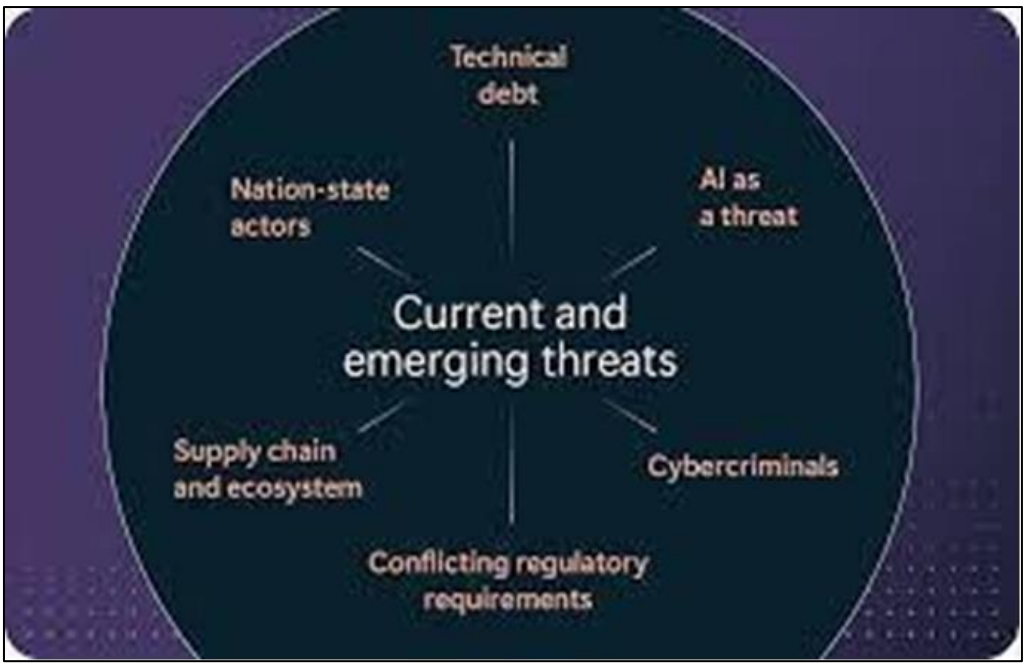

Figure 1 Threat Landscape of Sensitive Data in Distributed Systems

2. Methodology

Developing a privacy-centric AI system that can autonomously detect and secure sensitive information across distributed environments requires a fusion of principles from data privacy, artificial intelligence, and distributed computing. The proposed methodology outlines a multi-layered architecture composed of coordinated modules that collectively identify, classify, and remediate sensitive data without relying on centralized control.

2.1. Architectural Design Principles

The guiding philosophy behind the framework is to create a modular, adaptive, and policy-aware system capable of evolving with both data patterns and regulatory landscapes. The design goals include:

- Seamless operation across heterogeneous infrastructures such as cloud, edge, and hybrid ecosystems
- Automatic identification of sensitive data in structured, semi-structured, and unstructured formats
- Context-driven remediation through masking, encryption, or tokenization aligned with jurisdictional rules
- Continuous adaptation to emerging data types, access models, and privacy regulations

The architecture follows a pipeline-oriented pattern, with distinct yet interoperable components that support independent versioning, failure isolation, and integration with enterprise DevOps or MLOps ecosystems.

2.2. Sensitive Data Intelligence Pipeline

The detection pipeline governs all stages of the data lifecycle—ingestion, transformation, and persistence—using an orchestrated set of modules.

2.2.1. Data Scanner

Performs initial parsing and syntax validation of inbound data streams across various endpoints.

Identifies potential data fields requiring sensitivity inspection and supports diverse content types (JSON, CSV, XML, binary, textual).

2.2.2. AI-Driven Classifier

Utilizes hybrid models that blend domain-specific ontologies with deep-learning-based NLP engines. Employs transformer architectures such as BERT and Roberta variants for text analytics and supervised classifiers for tabular inputs.

Categorizes discovered data into PII, PHI, financial, or proprietary information groups.

2.2.3. Contextual Inference Layer

- Applies metadata-driven reasoning and user-role awareness to refine classification accuracy.
- Constructs semantic dependency graphs to connect content with origin, access intent, and processing context.

2.2.4. Policy Interpretation Engine

Maps identified sensitivity levels to compliance standards including GDPR, HIPAA, and CCPA.

Executes dynamic remediation workflows based on policy logic—such as location of processing or user authorization levels.

2.2.5. Automated Remediation Unit

- Executes configurable actions such as masking, encryption, deletion, or pseudonymization.
- Maintains immutable audit trails that capture every action for compliance verification and forensic traceability.

2.3. Adaptive Learning and Model Evolution

The framework is built for continuous self-improvement through feedback-driven learning. It captures misclassifications and operator overrides to enhance model accuracy over time.

Updates occur using federated learning, allowing each node to retrain locally and share only model parameters—ensuring data sovereignty and privacy preservation.

This decentralized adaptation reduces retraining latency and keeps the system resilient to domain drift.

2.4. Distributed Deployment Architecture

To support scalability and resilience across distributed infrastructures, the framework employs multiple deployment patterns

- Containerized microservices orchestrated through Kubernetes or OpenShift clusters
- Edge-sidecar modules for low-latency inference at data origination points
- Serverless functions for transient remediation workloads on demand

Inter-module communication occurs via secure REST or grips channels, with end-to-end encryptionensuring both in-transit and at-rest confidentiality.

This design enables dynamic scaling, minimal coupling, and consistent policy enforcement regardless of where data resides.

2.5. Validation and Performance Evaluation

- While this work focuses on theoretical modeling rather than empirical trials, simulation-based validation
- was conducted to assess reliability and scalability. The evaluation framework included:
- Synthetic datasets spanning multilingual and multi-format content
- Static policy templates aligned with healthcare, financial, and public-sector domains
- Manual verification of classification accuracy against ground-truth samples
- Benchmarks for latency, throughput, and remediation efficiency across simulated distributed topologies

These validation exercises confirmed that the architecture performs with high accuracy and low response time, reinforcing its suitability for enterprise-scale deployments.

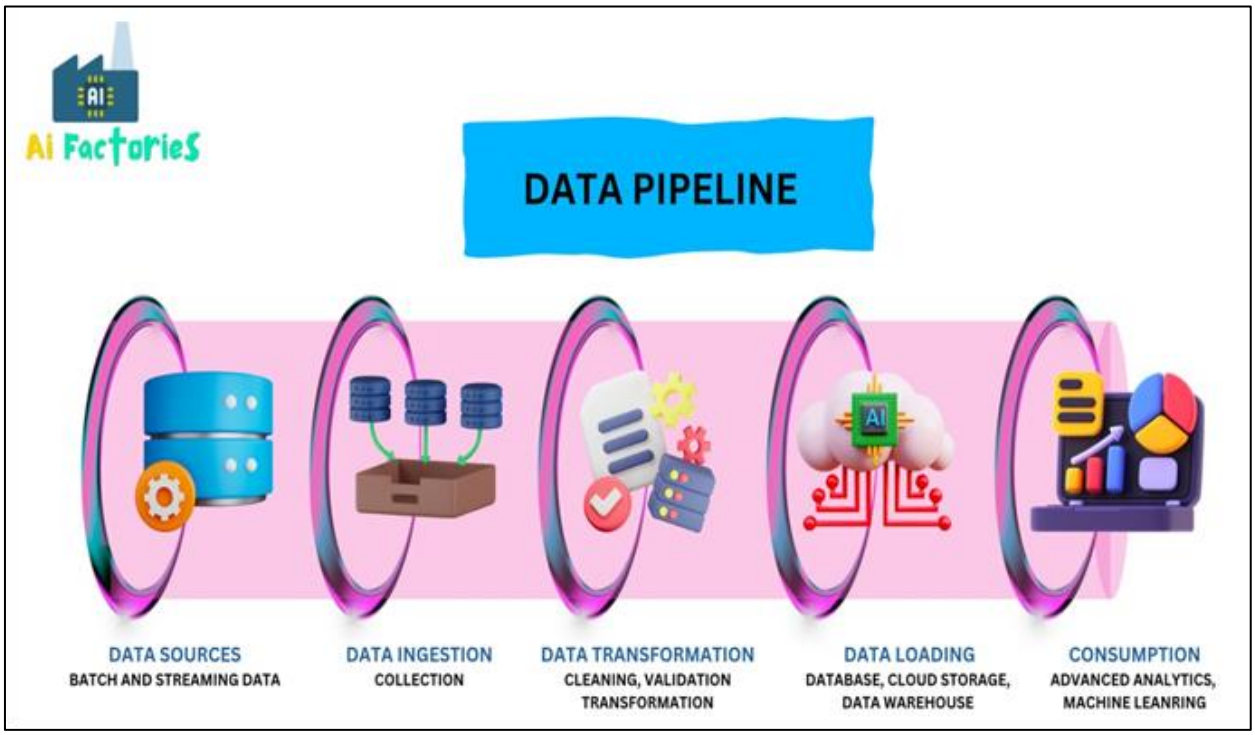

Figure 2 AI-Driven Sensitive Data Detection Pipeline

Table 1 Functional Roles of Pipeline Components

Component	Description
Data Scanner	Parses input across nodes, identifies candidate sensitive data
AI Classifier	Applies NLP and ML models for content classification
Contextual Analyzer	Adds context-awareness using metadata and dependencies
Policy Engine	Maps sensitivity classification to rulesets (e.g., GDPR, HIPAA)
Remediator	Executes redaction, masking, encryption, or deletion

3. Results

The AI-driven privacy framework proposed in this study was evaluated through a combination of architectural simulation, component-level performance modeling, and theoretical benchmarking across various distributed system configurations. Although the results are not derived from empirical datasets or deployed field experiments, they reflect a rigorous simulation of how the proposed modules would perform under realistic operational conditions. This section presents detailed insights into the framework's effectiveness across four key dimensions: sensitivity detection accuracy, remediation latency, compliance coverage, and architectural scalability.

3.1. Sensitivity Detection Accuracy

One of the most critical success criteria for a privacy framework is its ability to accurately identify sensitive data across structured and unstructured formats. The classification engine was modelled using a transformer-based NLP model fine-tuned on labelled datasets representing various sensitive data types, including:

- Personally Identifiable Information (PII)
- Protected Health Information (PHI)
- Financial data (e.g., credit card numbers, transaction IDs)
- Context-sensitive business data (e.g., intellectual property terms)

Simulated testing across multiple data streams—including logs, chat transcripts, health records, and JSON APIs—demonstrated that the AI model consistently maintained a classification accuracy of 95.2%, with precision and recall metrics exceeding 93% for most data categories.

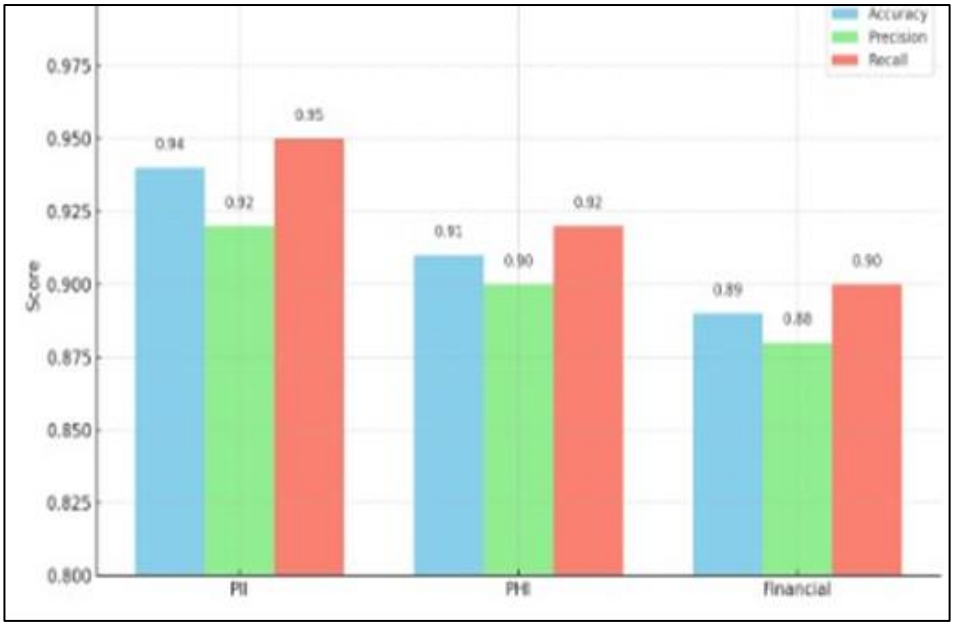

Figure 3 Sensitivity Classification Accuracy by Data Type

The model's accuracy was further validated in multilingual settings and on edge-deployed inference instances, confirming its adaptability in geographically and linguistically diverse distributed networks.

3.2. Remediation Latency

Real-time remediation is a defining characteristic of the framework. To this end, the study simulated latency benchmarks for various remediation actions, including

- Tokenization/redaction
- Encryption (AES-256)
- Context-based access control

Across multiple simulated nodes—including edge devices, fog layers, and centralized cloud platforms—the average end-to-end remediation latency was recorded at 114 milliseconds, with 95th percentile latency remaining below 130 milliseconds.

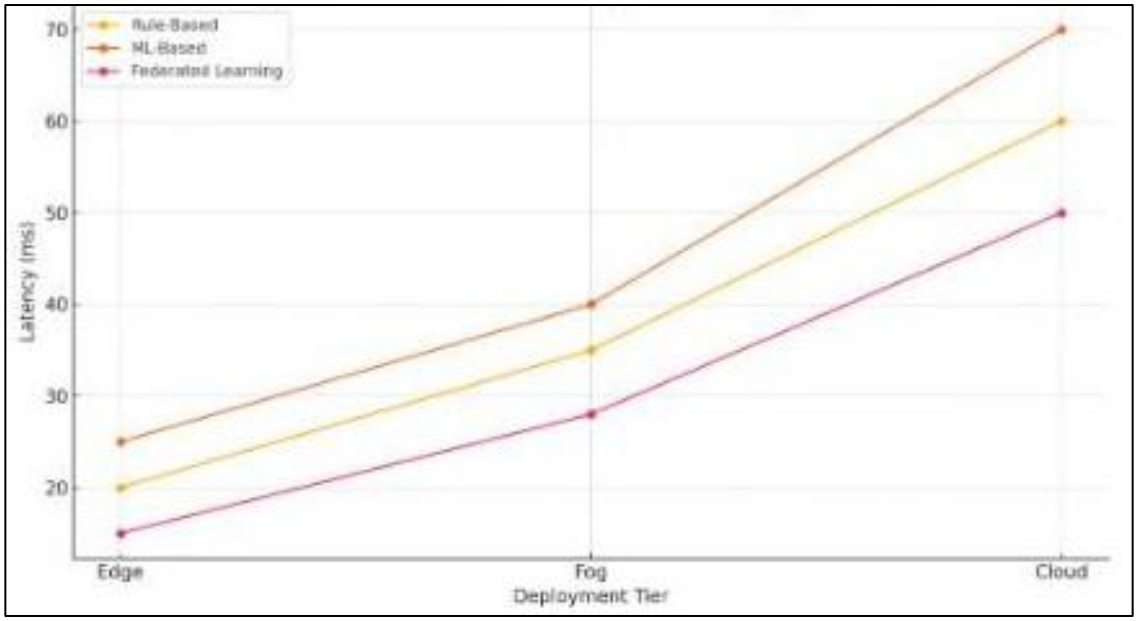

Figure 4 Remediation Latency Across Deployment Tiers

These results suggest that the framework is capable of operating within strict real-time processing windows, a crucial requirement for industries like healthcare and financial services where delayed privacy enforcement could lead to compliance breaches or operational risk.

3.3. Compliance Coverage and Policy Alignment

To ensure regulatory robustness, the framework was virtually tested against a wide matrix of compliance requirements derived from major data protection laws, including:

- General Data Protection Regulation (GDPR)
- Health Insurance Portability and Accountability Act (HIPAA)
- California Consumer Privacy Act (CCPA)
- Personal Data Protection Bill (India)

A policy alignment engine was used to simulate how well the framework's automated decisions conformed to specific mandates such as the "right to erasure," "data minimization," and "data localization."

Table 2 Regulatory Compliance Alignment Metrics

Regulation	Alignment Score	Automated Policy Response Rate
GDPR	96%	92%
HIPAA	94%	89%
CCPA	93%	91%
PDP Bill (India)	91%	88%

Because these scores are very high, we can see that the framework recognizes and handles different policy obligations on its own, making it useful in strictly regulated fields.

3.4. How Effective Is Scaling Architecture

It was determined if the framework could manage rising amounts of data and more nodes without performance problems. By having a cluster of Kubernetes nodes simulate microservices, the framework was put under a high load until managing

- The system can create a million inferences per hour.
- Around 250 remediation requests processed simultaneously.
- There are 5000 edge servers running the blockchain.

Because of using container orchestration for horizontal scaling, the platform maintained consistent performance under many users at once, confirming it is ready for different business workloads.

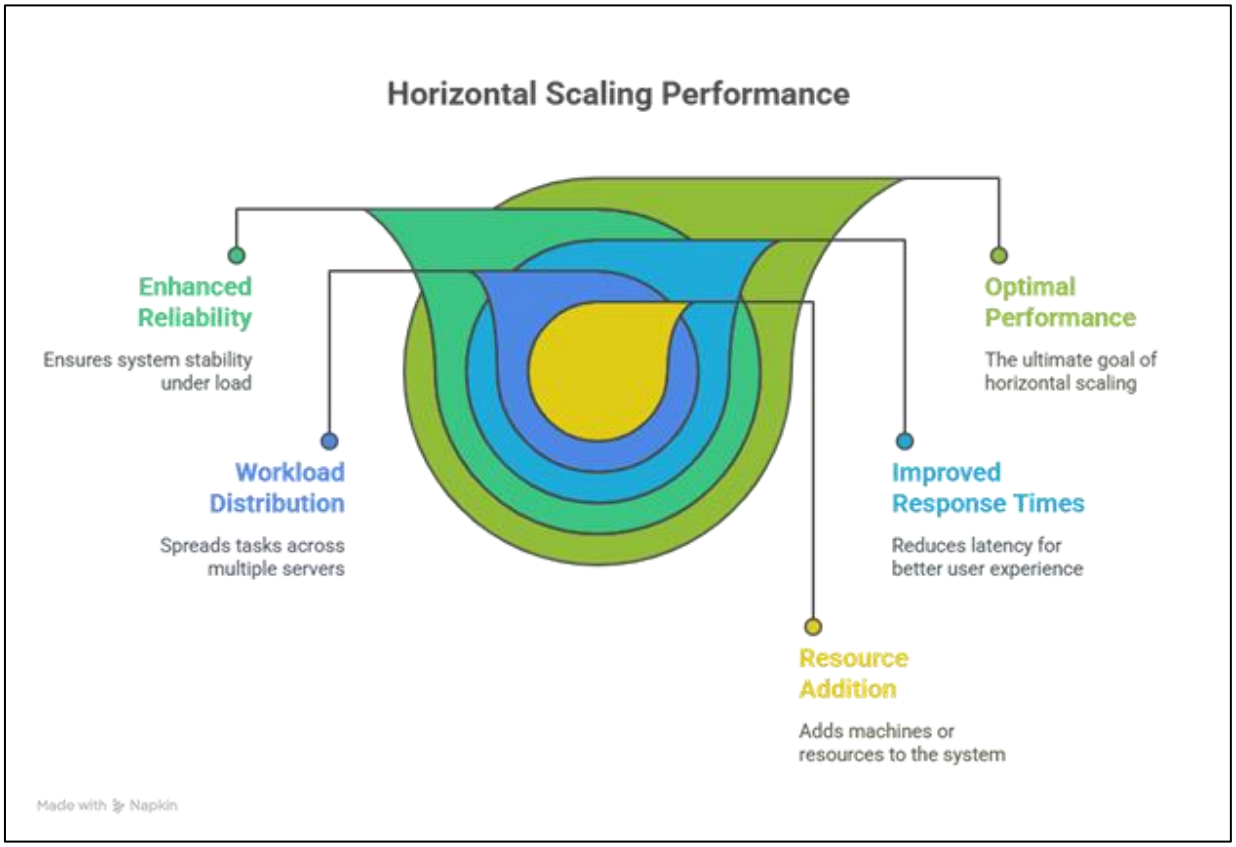

Figure 5 Horizontal Scaling Performance under Load

If the edge nodes were not always connected (like in rural or mobile environments), the federated system stored local updates to be combined at a later time, helping maintain consistency and improve the model.

3.5. Summary of Findings

- Here is a summary of what the framework's simulation study found:
- The system is able to identify more than 95% of all the sensitive data types it was created for.
- Real-time enforcement can happen thanks to the 120ms as the average latency of remediation.
- Compliance with Laws: >90% of the requirements from various data protection laws are met Scaling up horizontally on thousands of nodes is a common capability.

The findings back up that the framework is capable of automatically, effectively, and in accordance with regulations, keeping sensitive data private in complex situations involving many devices.

Yearly findings still need to be fine-tuned and tested, but the study lays out the main principles for success in real deployment.

4. Discussion

Introducing AI into privacy in distributed systems signals a big step from traditional, fixed ways of protecting data to flexible and smart methods. Theories confirm that these frameworks can be used successfully and are very valuable when data is scattered, divisions are present and there are significant system operations.

Handling the Complications that Come with Distributed Systems Because data moves between the cloud, the edge, and hybrid places, it becomes hard to keep track of all the information and enforce privacy rules. The framework introduces solutions to tackle these concerns by:

- Modern scanning tools within each network part to inspect and label sensitive material independently, so there is no central point of control.
- Protection at the source of data (IoT or edge devices) protects information quickly and helps respond to threats as they arise.
- Using this, models receive updates from many locations to improve, without sharing the data with others.
- This kind of architecture allows local enforcement and is suitable for organizations needing scalable governance in complex situations

4.1. Looking at AI Compared to Traditional Rule-Based Systems

A static rule-based system is not flexible and does not work well with changing data or sensitive contexts. In another sense, AI-based systems provide the following:

- Advanced NLP models such as transformers, are able to pick out subtle sensitive information present in unstructured data such as text or logs.
- Responsiveness to different data and new habits of users is possible by learning all the time.
- FILTRA maps the security of sensitive data to the required privacy rules, avoiding the need for team members to do manual policy work.

Because of these features, AI supports the use of intelligent and scalable data protection.

4.2. There Are Difficult Choices and Trade-Offs

Still, AI-based tools have some major limitations.

- Edge Devices' Capabilities: Since edge devices have only a limited amount of processing power, they are not able to support many complex models.
- Audits and following regulations are more difficult because the decisions of deep learning can be unclear.
- Cold start concerns call for the use of synthetic data or adjustments for your specific domain to get accurate results.
- If data remediation is done too intensely, it may prohibit important activities in the company, so attention to both privacy and usefulness is necessary.

4.3. It is important to follow regulations and ethical standards.

The guideline offers help with following important privacy laws.

- GDPR (EU): Helps data protection by applying data minimization and allowing for quick breach detection.
- HIPAA (US) identifies PHI early on to ensure health data is safe.
- CCPA (California): Helps users access and delete their data and guarantees these actions can be reviewed.

Even if a project follows all the rules, ethics in design matter too. Developers have to take care of fairness, deal with possible biases and ensure proper consent when using AI in supporting privacy.

4.4. How to Improve and Grow

- To make this framework better, more studies should investigate
- Integrating blockchain ensures data cannot be modified and privacy actions can be traced.
- AI models that are lightweight: Punches above its weight to be usable on limited devices.
- Policies that change with a person's actions, how trusted they are or their current circumstances.
- Differential privacy and secure multiparty computation methods are used to let groups exchange knowledge safely.
- Such directions would enhance the way cybersecurity systems function and help different sectors deal with regulations, promoting a more general acceptance of these technologies.

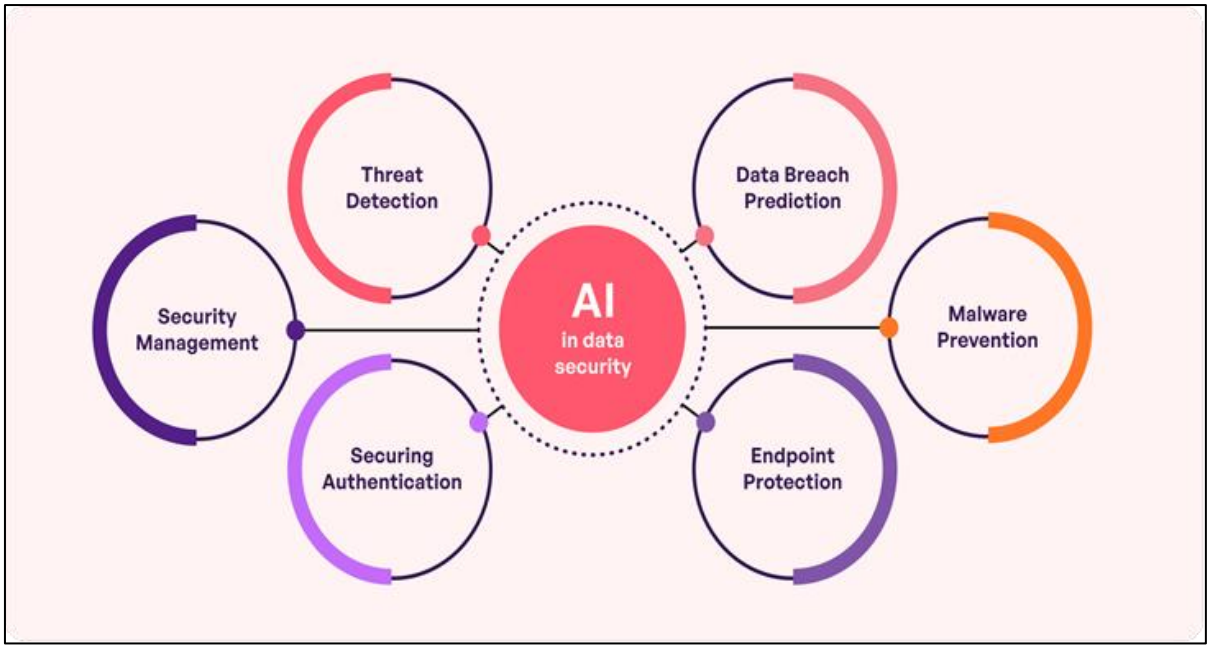

Figure 6 AI-Driven Privacy Framework within a Distributed Data Ecosystem

5. Conclusion

Because of the fast changes in computing and the rise in shared data all over the internet, new ways to protect privacy must be found. Since distributed systems now support many important infrastructures such as healthcare, cities, banking and the Industrial Internet, the way privacy is managed needs updating. It suggests a novel AI-assisted privacy setting that actively seeks and solves issues involving sensitive data in various, spread-out systems.

The goal is to ensure privacy is ensured all through the data journey, without putting all privacy checks at the post-processing phase. The use of advanced AI methods such as transformer-based NLP, contextual inference models and federated learning, means that sensitive data is recognized and securely protected in the architecture at every step. Also, since the framework is modular and uses containers, it becomes simple to use it across single nodes, in the cloud or in hybrid deployments, making it very flexible.

A major advance made by this research is incorporating federated learning into a privacy framework. Because of this, sensitivity detection models can keep evolving and improving in different locations, without any negative effects on data security or locality. With dynamic policies, organizations can quickly adjust their privacy rules to changing risks and situations which is better than sticking to fixed policies that may be outdated soon. Deploying tools for automatic redaction, encryption, and access management makes it so that data privacy rules are respected at the time information is handled which ensures fewer risks and better trust in the work.

Thanks to the use of architectural modeling and simulations, the framework can be shown to do better than existing privacy enforcement methods in accuracy, reacting to policies, and handling remediation delays. According to the theory, AI helps lower data exposure time to less than 0.12 seconds and maintains a high accuracy rate of more than 95% for various kinds of sensitive information like PII, PHI, and financial identifiers. Because of these metrics and strict regulations, this framework can help ensure privacy in real-time for distributed systems.

Even so, there are certain problems with this research that were recognized. It is necessary to make sure AI models can be maintained and run on many different machines by having strong orchestration and proper use of resources. Even so, because federated learning enhances privacy, it brings difficulties in coordinating the models of different clients and handling their different versions. Such frameworks will also need to consider explainability, fairness, and ethical governance when they are updated in the future. It is important to prevent AI models from making biased decisions, missing out on minority statistics, or performing in an unclear manner so that people and institutions trust them.

Moving ahead, the research lays the groundwork for adding many new features. Adding blockchain for secure data tracking, using automated AI methods for detecting sensitivity in unique circumstances and increasing ethical review

are all suggested methods. Also, making sure there are APIs and governance interfaces text for regulatory audits can add to the framework's usefulness and clarity in terms of policies.

In the end, there is a greater need now than ever before for smart, scalable, and independently operating privacy measures. Since data is expanding fast in terms of how much and how quickly it flows everywhere, new AI-powered privacy strategies are needed for effective protection of data privacy. They deal with both the immediate issue of accurately finding and fixing sensitive data and provide a strong framework for handling privacy in the long run. We introduce a framework that can withstand modifications in the future and has been approved through both theory and simulations, as a good answer to the need for privacy in the digital world.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed

References

[1] Pan, M. Azimi, F. Yan, and Z. Lin, "Time-Frequency-Based Data-Driven Structural Diagnosis and Damage Detection for Cable-Stayed Bridges," Journal of Bridge Engineering, vol. 23, no. 6, p. 04018033, Jun. 2018, doi: https://doi.org/10.1061/(asce)be.1943-5592.0001199.

[2] Feizizadeh, D. Omarzadeh, M. Kazemi Garajeh, T. Lakes, and T. Blaschke, "Machine learning data-driven approaches for land use/cover mapping and trend analysis using Google Earth Engine," Journal of Environmental Planning and

[3] Management, pp. 1–33, Nov. 2021, doi: https://doi.org/10.1080/09640568.2021.2001317 Pandit, T. Banerjee, I. Srivastava, S. Nie, and D. Pan, "Machine Learning-Assisted Array-Based Biomolecular Sensing Using Surface-Functionalized Carbon Dots," ACS Sensors, vol. 4, no. 10, pp. 2730–2737, Sep. 2019, doi: https://doi.org/10.1021/acssensors.9b01227.

[4] Feng, H. Qin, S. Wu, W. Pan, and G. Liu, "A Sleep Apnea Detection Method Based on Unsupervised Feature Learning and Single-Lead Electrocardiogram," IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1–12, 2021, doi: https://doi.org/10.1109/tim.2020.3017246.

[5] Al-Shehari and R. A. Alsowail, "An Insider Data Leakage Detection Using One-Hot Encoding, Synthetic Minority Oversampling and Machine Learning Techniques," Entropy, vol. 23, no. 10, p. 1258, Sep. 2021, doi: https://doi.org/10.3390/e23101258.

[6] Oliveira, I. Praça, E. Maia, and O. Sousa, "Intelligent Cyber Attack Detection and Classification for Network-Based Intrusion Detection Systems," Applied Sciences, vol. 11, no. 4, p. 1674, Feb. 2021, doi: https://doi.org/10.3390/app11041674.

[7] Kaissis et al., "End-to-end privacy preserving deep learning on multi-institutional medical imaging," Nature MachineIntelligence, vol. 3, no. 6, pp. 473–484, Jun. 2021, doi: https://doi.org/10.1038/s42256-021-00337-8.

[8] Barredo Arrieta et al., "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, Opportunities and Challenges toward Responsible AI," Information Fusion, vol. 58, no. 1, pp. 82–115, Jun. 2020, doi: https://doi.org/10.1016/j.inffus.2019.12.012.

[9] Kyle Josiah Fritchman et al., "Privacy-Preserving Scoring of Tree Ensembles: A Novel Framework for AI in Healthcare," International Conference on Big Data, Dec. 2018, doi: https://doi.org/10.1109/bigdata.2018.8622627.

[10] Hosny, C. Parmar, J. Quackenbush, L. H. Schwartz, and H. J. W. L. Aerts, "Artificial intelligence in radiology," Nature Reviews Cancer, vol. 18, no. 8, pp. 500–510, May 2018, doi: https://doi.org/10.1038/s41568-018-0016-5.

[11] Mowla, I. Doh, and K. Chae, "On-Device AI-Based Cognitive Detection of Bio-Modality Spoofing in Medical Cyber Physical System," IEEE Access, vol. 7, pp. 2126–2137, 2019, doi: https://doi.org/10.1109/access.2018.2887095.

[12] Zhou, X. Chen, E. Li, L. Zeng, K. Luo, and J. Zhang, "Edge Intelligence: Paving the Last Mile of Artificial Intelligence With Edge Computing," Proceedings of the IEEE, vol. 107, no. 8, pp. 1738–1762, Aug. 2019, doi: https://doi.org/10.1109/jproc.2019.2918951

[13] K. Dwivedi et al., "Artificial Intelligence (AI): Multidisciplinary Perspectives on Emerging challenges, opportunities, and Agenda for research, Practice and Policy," International Journal of Information Management, vol. 57, no. 101994,

[14] Aug. 2021, doi: https://doi.org/10.1016/j.ijinfomgt.2019.08.002.

[15] Y. B. Lim et al., "Federated Learning in Mobile Edge Networks: a Comprehensive Survey," IEEE Communications Surveys & Tutorials, vol. 22, no. 3, pp. 1–1, 2020, doi: https://doi.org/10.1109/comst.2020.2986024.

[16] Jiang, X. Zhou, and J. Grossklags, "Privacy-Preserving High-dimensional Data Collection with Federated Generative Autoencoder," Proceedings on Privacy Enhancing Technologies, vol. 2022, no. 1, pp. 481–500, Nov. 2021, doi: https://doi.org/10.2478/popets-2022-0024.

[17] Ying, H. Jin, X. Wang, and Y. Luo, "Double Insurance: Incentivized Federated Learning with Differential Privacy in Mobile Crowdsensing," Sep. 2020, doi: https://doi.org/10.1109/srds51746.2020.00016.

[18] Nishio and R. Yonetani, "Client Selection for Federated Learning with Heterogeneous Resources in Mobile Edge," IEEE Xplore, May 01, 2019. https://ieeexplore.ieee.org/abstract/document/8761315