Computer Fraud and Security
ISSN (online): 1873-7056

Data Privacy: Strategies for Protecting Sensitive Data for OT using Artificial Intelligence

Manam Karthik Babu,
PhD Student, School of Computer and Information Sciences, *University of the Cumberlands,Williamsburg, KY, USA*
kmanam65453@ucumberlands.edu

Yugandhar Suthari,
Research Student, School of Computer and Information Sciences, *University of the Cumberlands,Williamsburg, KY, USA*
ysuthari27952@ucumberlands.edu

***Abstract*—** The exponential rise of data in the digital age has enabled huge possibilities for innovation in every field. Although the data surge has its implications with respect to the protection of sensitive information, it also contains high privacy concerns on account of it. In this paper, we discuss how Artificial Intelligence (AI) can be used to better develop privacy-preserving techniques. Finally, we provide a general overview of how differential privacy, federated learning, homomorphic encryption, and anonymization techniques are being used to protect sensitive data with AI-driven mechanisms.

Through an analysis of existing frameworks and case studies, we illustrate the effectiveness of these AI strategies in mitigating privacy risks while maintaining data utility for analytical purposes. We support these AI strategies via an analysis of existing frameworks and case studies to prove the efficacy of these AI strategies in reducing privacy risk while maintaining data utility for analytic purposes. In the meantime, we also tackle the open challenges of proper trade-off between data privacy and AI property, i.e. computational overhead, algorithmic factuality and accountability. In this way, our study examines them and offers valuable insights as well as directions for future research about the privacy-preserving landscape. This document aims to contribute to the ongoing torsion on AI and privacy of data while proposing actionable strategies for AI researchers, AI practitioners, and AI policymakers seeking to ensure that sensitive data is in an increasingly connected world.

Keywords—component, formatting, style, styling, insert

I. INTRODUCTION

In this contemporary digital age, data has flourished and has unlocked tremendous power to innovate in variety of fields from the field of healthcare and finance to social media and e-commerce[1]. The large amount of data generated or collected on a daily basis allows organisations to gain valuable insights from data, to improve the decision-making processes and also to improve the user experience. Yet, this makes it uncharted territory with related privacy concerns such as protecting personal and sensitive information[1-3].

This growth in the need for such privacy-preserving techniques arises from the fact that data breaches and privacy violations happen all too often. The traditional data protection methods tend to be inadequate in the presence of sophisticated attacks and dynamic usage types of data. And in this regard, it becomes a powerful ally in striving for data privacy in this context. AI-driven solutions provide a higher level of sophistication in the area of data security by being able to detect potential threats in real-time, make the changes necessary to be ready to adapt to other current, and still emerging threats, as well as effectively manage and protect vast databases[4-5][7].

There are many use cases of how AI can help to ensure data privacy. For instance, in the healthcare sector federation learning has been used to protect the data while executing collaborative research. An important example is how in practice, Google applied federated learning for training machine learning models at multiple institutions

with no sensitive patient data transmission between these institutions [1]. Moreover, this enables diverse datasets to be used in order to improve the quality of the models as well as provide some privacy.

Differential privacy techniques are also being integrated into social media for providing user-based experiences while maintaining their privacy. For instance, Apple employs differential privacy to monitor its user's use while preventing individual data from being recognised [2][3][8]. This technique forces the data noise so you can analyse without revealing private information. Homomorphic encryption also has another major application consisting of secure data processes in the financial sector. One of the pioneers in the use of homomorphic encryption is companies like IBM to perform computations on encrypted data whilst keeping the data confidential throughout the processing cycle [3][7].

In spite of these, the integration of AI in data privacy presents a number of challenges. There are some problems with computational overhead, algorithmic transparency, and ethical considerations as others. AI solutions can be resource-intensive and opaque, making it difficult to ensure transparency and accountability, especially given some AI algorithms. Moreover, ethical issues such as AI system bias and the consequences of automated decision-making should be well handled to prevent harmful effects.

In this paper, we study these AI-driven privacy-preserving techniques and explain all such current methodologies in a comprehensive way. We propose to explain the impacts of these strategies by studying the applicability of these approaches in mitigating privacy threats within data utility. Second, since we also propose future research directions to overcome the challenges and develop the privacy-preservation landscape[3], we believe that the details of our privacy-preserving approach should aid in achieving these goals.

There are manifold advantages to employing AI for data privacy. It makes scaling possible and enables protection of large datasets that are out of reach for traditional methods. It allows real-time monitoring, using data security, in a proactive way. In addition, AI systems can also learn and adapt to the changes in privacy threats, thus data protection measures stay adaptive to the new challenges too[2].

It contributes to the ongoing debate by providing researchers, practitioners, and policymakers with some possible ways to achieve data privacy in the context of AI. By making use of what AI can do, we can create more forceful privacy-preserving techniques to ensure that sensitive info is not doing that world.

II. Data Privacy as a Concern

A. About Data Privacy

Data privacy is the protection of the personal information of individuals and the assurance of how this data is being processed in a secure and confidential manner. It includes both the rights of individuals to control how an organisation or entity collects, uses, shares and stores their personal information[11][15].

Key aspects of data privacy include:

Consent: Personal data must be collected and processed only if the individual who has provided that personal data gave their explicit consent for it to be so. The informed, freely given, and revocable consent should be given.

Transparency: Transparency is key in organizations and they should put attention on their data practices, including how they gather, use and share personal information. People should be allowed to know what kind of information is being collected about and for what reasons.

Data Minimization: Personal data should only be collected and kept by organisations to the extent necessary for the purposes corresponding to the collection. Data minimization lessens the risk of a data breach and data exposure without affecting relevant operations.

Security: Security should be exercised in the processing of personal data so that it is not disclosed, that it is not available for unauthorised access, alteration, disclosure, or destruction. That includes encryption and access controls, and regularly doing security audits.

Accuracy: Organisations must take appropriate steps to ensure the accuracy of personal data and to allow an individual to have his or her personal data rectified.

Data Subject Rights: You have the right to access, ask for correction, deletion of your data, and object to the processing of your data for some reasons.

Accountability: It is charged upon organisations to comply with the data privacy laws and regulations and must have a scheme to show compliance. It covers the appointment of data protection officers, privacy impact assessments as well as records about the data processing activities.

Data privacy is vital for involving people in the digital space, ensuring the innovation, and protecting fundamental human rights to privacy and data protection. Inappropriate handling of personal data is often associated with high monetary penalties, loss of image and customer trust. Hence, dealing with data privacy becomes an obligation for organisations and they should implement effective data protection measures to ensure personal information.

B. Why AI for OT Security

Advanced, proactive, and adaptive security offerings that an AI can provide can bring massive value to the security of Programmable Logic Controllers (PLCs) and other industrial controllers. The following are the main reasons why AI is helpful for PLCs and controllers protection [12][15]:

Anomaly Detection

Behavioral Analysis: AI is able to learn the normal operating types of PLCs and controllers. This allows for any deviation from these patterns to be safely detected as potential security threats like cyber attacks or malfunctions.

Real-Time Monitoring: Real time Data monitoring is possible with AI algorithms, capable of continuously monitoring data in real time to find out if there are any unusual activities which might be a security breach.

Predictive Maintenance

Early Warning: The potential failures can be predicted or vulnerabilities can be identified by AI, analyzing trends, historical data, etc. It means failure can be intercepted early so there is little risk of breach related to equipment failure.

Health Monitoring: AI can be used to monitor the continuous health of the programmes and provide an indication of wear and tear before it becomes a security hazard [14].

Threat Intelligence

Pattern Recognition: By knowing and understanding the patterns of known cyber-attacks, networks can detect and respond to them much faster using mechanisms that AI can process and analyze.

Adaptive Defense: Though new and evolving targets, they can be changed based on previous attacks and experience to create threat detection models that the new AI systems can learn from.

Automated Response

Incident Response: As an application of AI, the detected threats can be automated in response such as isolating the affected components, notifying the operators, or even mitigating the threat itself.

Reduced Reaction Time: This way, threats will be dealt with automatically, eliminating the time in which an attacker has to flourish.

Intrusion Detection

Network Traffic Analysis: The analysis of network traffic to and from PLCs can be done with AI that can detect unusual patterns associated with an intrusion.

Deep Packet Inspection: Deep packet inspection using the AI can recognize malicious payloads and command injected for PLCs.

Data Integrity

Integrity Checks: PLCs could signal AI that the data they are currently processing is in the process of being tampered with, and AI can constantly verify that the data never gets tampered with.

Secure Configurations: PLC configurations can be made assured to be consist and ensured, and unsuitable changes will be detected and corrected automatically and uninterruptively by the application of AI.

Access Control

User Behavior Analytics: Such anomalies could also manifest in the behaviour of users trying to access PLCs, for example, and AI can help detect such anomalies, showing traffic anomalies in access behaviour to PLCs, for example that might identify trojaned and other compromised credentials, as well as insider threats.

Dynamic Access Control: Depending with the result of the analysis of the user's behaviour and the current threat, the level of permissions will change as frequent as possible with the aid of AI.

Cyber-Physical System Security

Physical Security Integration: Through integration of the physical and cyber security measures, AI can analyse data from physical sensors (e.g. cameras, motion detectors) to improve PLCs with security.

Holistic View: AI gives the ability to broadly see the security landscape by combining data from the cyber and the physical spheres.

Scalability

Handling Large Volumes of Data: Industrial controllers' data volumes are processed and analysed and security threats, which may be missed by manual analysis, are identified by the AI systems.

Efficient Resource Utilization: AI will allow security resource allocation that delivers maximum efficiency and effectiveness of security resources.

III. Feature Extraction

The number of features that will be able to be extracted for a Programmable Logic Controller (PLC) or other industrial controllers used for AI such as a PLC, varies based on the use case, type of data available and goals of the analysis. Following are some of the common categories of the features which can be extracted:

Operational Data:

Sensor Readings: Temperature, pressure, flow rates, levels, etc.

Actuator States: Valve positions, motor speeds, actuator positions.

Input/Output States: Digital and analog input/output values.

Cycle Times: Time taken for different phases of operation.

Performance Metrics:

Efficiency Metrics: Energy consumption, production rates, downtime.

Error Rates: Frequency of faults or errors, types of faults.

Cycle Counts: Number of operational cycles, start/stop counts.

Diagnostic Data:

Error Logs: Historical error logs, types of errors, frequency.

Maintenance Logs: Maintenance records, scheduled vs. unscheduled maintenance.

Self-Diagnostic Results: Built-in self-test results, diagnostics data.

Environmental Data:

Ambient Conditions: Temperature, humidity, vibration levels.

Power Quality: Voltage levels, power supply stability, frequency deviations.

Operational Parameters:

Control Setpoints: Desired values for controlled variables (e.g., temperature setpoint).

Controller Gains: PID controller parameters (Proportional, Integral, Derivative gains).

Communication Data:

Network Traffic: Data packets sent/received, communication errors.

Latency: Communication delays, response times.

Historical Data:

Trend Data: Historical trends of sensor readings, operational parameters over time.

Pattern Recognition: Recognizable patterns in historical data (e.g., periodic maintenance needs).

Custom Data:

Application-Specific Data: Any custom or application-specific data points relevant to the operation of the PLC.

Example Use Case: Predictive Maintenance

For a predictive maintenance application, you might extract features such as:

- ✓ Historical sensor readings (vibration, temperature)
- ✓ Operating hours
- ✓ Maintenance history
- ✓ Error/fault logs
- ✓ Power consumption patterns
- ✓ Load conditions
- ✓ Example Use Case: Process Optimization

For process optimization, relevant features might include:

- ✓ Real-time sensor readings
- ✓ Actuator positions and states
- ✓ Cycle times and efficiency metrics
- ✓ Control setpoints and controller gains
- ✓ Environmental conditions

Practical Considerations:

Data Availability: Not all features may be available on all PLCs or controllers. The choice of features depends on the specific hardware and its capabilities.

Relevance: Features should be chosen based on their relevance to the specific AI application. For example, vibration data might be crucial for predictive maintenance but less relevant for process optimization.

Data Quality: The quality and granularity of the data affect the usefulness of the features. Data that are high frequency may give us more but also need more storage and processing power.

Dimensionality: High dimensionality is the result of having too many features and can complicate the model and they can even need dimensionality reduction techniques.

Additionally, Genetic Algorithms (GAs) also have great power for data cleanup in AI and machine learning applications, especially when our dataset is complex and multi-dimensional. Natural evolution serves as an inspiration for GA's, which is made in evolving a population of probable solutions over several subsequent

generations using the principle of selection, crossover, and mutation. We start with a generation of initial population of the cleanup strategies, whereby each one is represented as an individual or as a chromosome. Some of these strategies could be handling missing values, reduction of noise or outlier detection. In this way, GAs can be used to optimise cleanup processes to handle noise, misses or outliers. Thus, for example, different imputation techniques for missing values, philtre techniques to cut out noise, and statistical methods to recognise outliers can be evolved and modified. The result is a high-quality dataset that significantly enhances the performance and reliability of AI models used in predictive maintenance and other industrial applications.

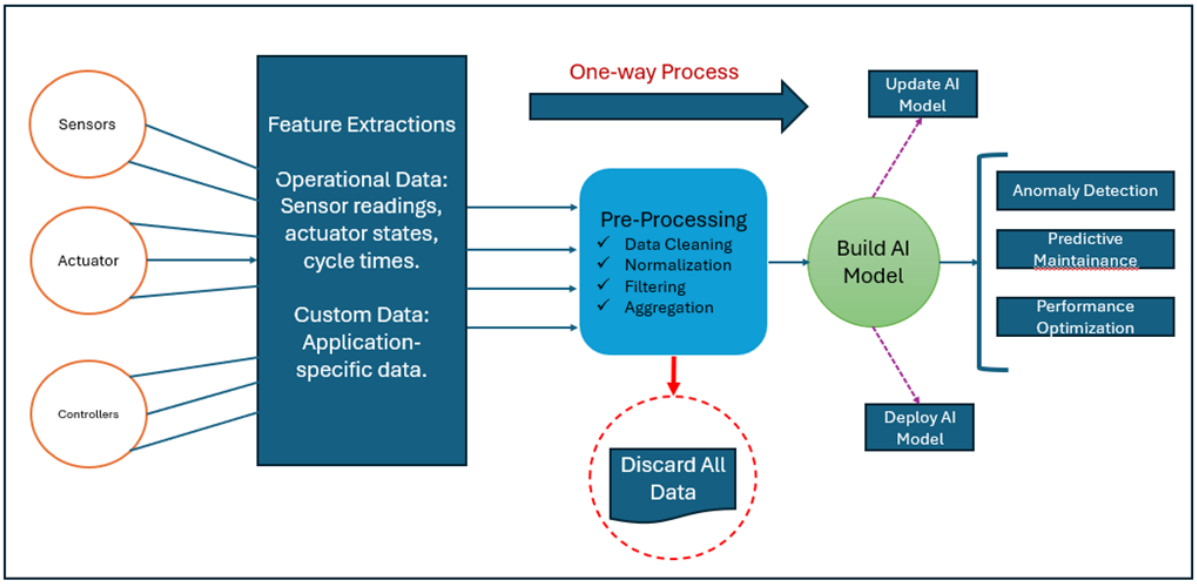

Fig. 1: Proposed Approach to Ensure Data Privacy in OT Components

IV. Proposed Approach

Programming logic controllers (**PLCs**) and industrial controllers assist in machine and process automation. Finally, we envision that these controllers will combine AI to increase the quality of these systems as the data generated by these controllers will eventually be more meaningful. This process transforms raw data into actionable insights. They take data that may not be as useful as they are and find a way to convert it into something that's useful in deciding operations, maintenance, security and even security operations. Fig. 1 shows the high-level architecture of proposed approach for data privacy of OT components.

1. Data Sources

Data Sources are the various inputs from which data is collected. In the context of **PLCs** and controllers, these typically include:

Sensors: Devices that measure physical parameters like temperature, pressure, flow rates, etc.

Actuators: Components that move or control mechanisms or systems, providing data like valve positions, motor speeds, etc.

System Logs: Logs generated by the system that include error logs, maintenance logs, and diagnostic data.

Network Traffic: Information on data packets sent/received, communication errors, and latency.

Environmental Data: Data about ambient conditions such as temperature, humidity, vibration levels.

Historical Data: Past data that includes trend data, historical sensor readings, and operational parameters.

2. Data Collection

Data Collection involves gathering data from all the available sources.

Steps:

Connect to Data Sources: Establish connections to sensors, actuators, system logs, and other data sources.

Data Acquisition: Use protocols such as OPC (OLE for Process Control), Modbus, or MQTT to collect data.

Data Storage: Store the collected data in a centralized database or data lake for further processing.

Tools/Technologies:

SCADA Systems: Supervisory Control and Data Acquisition systems that gather and analyze real-time data.

Data Loggers: Devices or software that record data over time from various sources.

IoT Platforms: Ways that use platforms such as AWS IoT, Azure IoT Hub, Google Cloud IoT to connect and monitor IoT devices.

3. **Data Preprocessing**

Data Preprocessing is the step in which raw data will be cleaned and transformed to make it, qualitative, and quantitative.

Steps:

Data Cleaning: Remove noise, handle missing values, and correct errors.

Normalization: Rescale the data to a consistent range usually in [0,1] or standardise it to have a mean of 0 and a standard deviation of 1.

Feature Engineering: New features can be derived from the existing data, like moving averages, trends, etc derived metrics.

Tools/Technologies:

Python Libraries: Pandas for data manipulation, NumPy for numerical operations.

ETL Tools: Extract, Transform, Load tools like Apache NiFi, Talend, or custom ETL scripts.

Data Cleaning Tools: *OpenRefine*, *Trifacta* for data cleaning and transformation.

4. Feature Extraction

In Feature Extraction, we extract relevant features from the preprocessed data which will be used as inputs to our AI models.

Steps:

Feature Selection: A technique such as correlation analyse, mutual information or domain knowledge is used to determine the relevant features.

Dimensionality Reduction: Other strategies include; lower the number of features to a more manageable and acceptable level through techniques such as; PCA (Principal Component Analysis) and LDA (Linear Discriminant Analysis).

Transformation: Mathematical transformation on the features like log transformation or polynomial features.

Tools/Technologies:

Python Libraries: Scikit-learn for feature selection and dimensionality reduction.

Feature Engineering Platforms: Feature tools for automated feature engineering.

5. AI Models

There are also AI Models which are the procedures that uphold the various features in an endeavour to learn the patterns and come up with the right decision.

Anomaly Detection:

Purpose: Anticipate the possible deviations from normal behaviour which might be a sign of threat to security or malfunction.

Algorithms: Isolation Forest, *One-Class SVM*, *Autoencoders*.

Implementation: Use Scikit-learn/*Keras*/*TensorFlow* to build moreover train anomaly detection models.

Predictive Maintenance:

Purpose: Expect likely failures and compliance to prevent breakdowns and other problems.

Algorithms: *Random Forest*, *Gradient Boosting*, *Neural Networks*.

Implementation: Use *Scikit-learn*, *XGBoost*, or PyTorch to build predictive maintenance models.

Performance Optimization:

Purpose: Introduce factors necessary for fine-tuning for optimum operational parameters for operations efficiency and performance.

Algorithms: Reinforcement Learning, Genetic Algorithms, Optimization Techniques.

Implementation: The different libraries that can be employed are OpenAI Gym for reinforcement learning, DEAP for genetic algorithms, and SciPy for optimization algorithms.

Implementation Steps for Each AI Model

Anomaly Detection

Collect and preprocess data: Ensure data quality and consistency.

Feature extraction: The selected features indicating anomaly are chosen.

Model training: It finds train anomaly detection model based on historical data.

Evaluation: Metrics to evaluate model's performance are precision, *F1-score*, recall, etc.

Deployment: Monitoring real time data for anomalies can be deployed in production environment.

Predictive Maintenance

Collect and preprocess data: The current log should include sensor data, operational metrics and maintenance logs.

Feature extraction: Generate features that correlate to equipment health as well as failure modes.

Model training: Train a failure on historical train data.

Evaluation: Utilize metrics like *ROC-AUC*, accuracy, along with confusion matrix for evaluating model.

Deployment: Provide model for predicting future maintenance needs in real time, as well as a proactive maintenance schedule.

Performance Optimization

Collect and preprocess data: Operate compile performance indicators and operational data.

Feature extraction: Identify features which impacts performance.

Model training: Train optimization models for finding best operational parameters.

Evaluation: Utilize performance metrics for evaluating optimization model.

Deployment: Applying model for optimizing parameters dynamically in live environment.

V. EXPERIMANTAL SETUP AND TESTING

Generally, when developing and deploying AI models, testing is an integral part of the process especially when the data comes from Programmable Logic Controllers (PLCs) and other industrial controllers. The AI models must

be accurate, reliable, and secure, and are thoroughly tested. Among other things, it means that data privacy has to be preserved and the models have to perform, and it means regulatory compliance.

Hardware Requirements:

PLCs and Controllers:

Industrial-grade Programmable Logic Controllers (*PLCs*) and controllers form the backbone of automation systems, interfacing with sensors, actuators, and machinery.

Sensors and Actuators:

Various types of sensors (e.g., temperature, pressure, flow) and actuators (e.g., motors, valves) collect real-time data and control physical processes.

Edge Computing Devices:

Edge computing devices facilitate data processing and analysis at or near the source of data generation, reducing latency and bandwidth requirements.

Servers and Cloud Infrastructure:

Servers and cloud infrastructure are used for data storage, processing, and hosting AI models, enabling centralized management and scalability.

Types of Testing:

Cross-Validation

Ensures robustness by training and validating the model on multiple data subsets.

Hyperparameter Tuning

Optimizes model performance by finding the best set of hyperparameters.

Model Evaluation Metrics

Uses appropriate metrics such as accuracy, precision, recall, and *F1-score* to evaluate model performance.

Ensemble Methods

Combines multiple models to improve overall performance and robustness.

Data Anonymization

Ensures that personal or sensitive information is anonymized to protect data privacy.

Access Control Testing

Verifies that only authorized users have access to sensitive data, ensuring data security.

Data Encryption

Ensures data is encrypted in transit and at rest to protect against unauthorized access.

Compliance Check

Ensures adherence to relevant data protection regulations (e.g., GDPR, CCPA).

Data Minimization

Ensures only necessary data is collected and stored to reduce the risk of data breaches.

VI. Conclusion

Testing the new AI data privacy test is vital in industrial applications involving PLCs and controllers to ensure the model's performance and compliance with regulatory policies. Data anonymization, access controls, encryption protocols, and compliance cheques can all be implemented so that organisations can protect sensitive information

from unauthorised access and limit the possibility of the event of a data breach. These measures are tested through methodologies that focus on data privacy, validating that personal or proprietary data is kept safe during its life.

It also matters more in the era of increasingly sensitive data powered by AI systems as they use trained and perform inference on it. While validating AI data privacy is a technical issue of checking the encryption and access control mechanisms, it is also a compliance test regarding GDPR, CCPA, and industry content rules. Integrating these testing practices, proves organisations are taking active measures to maintain and protect an individual's privacy rights and trust in AI-supported industrial solutions. Ultimately, data privacy in AI applications enhances data stewardship practises, supports responsible AI deployment, and builds the confidence of the stakeholders in the integrity of industrial systems.

REFERENCES

[1] B. C. M. Fung, K. Wang, A. W.-C. Fu, and P. S. Yu, "Privacy-Preserving Data Mining Techniques: A Review," ACM Transactions on Intelligent Systems and Technology (TIST), vol. 1, no. 1, pp. 1-33, 2010.

[2] R. Luh, et al., "Privacy in Industrial IoT: Overview, Challenges, and Solutions," in 2021 IEEE International Conference on Industrial Internet (ICII), 2021.

[3] Y. Qian, W. Shi, J. Yin, and G. Xiao, "Data Privacy and Security in Cyber-Physical Systems: A Survey," IEEE Transactions on Industrial Informatics, vol. 14, no. 7, pp. 1-1, 2018.

[4] R. Shokri and V. Shmatikov, "Privacy-Preserving Deep Learning," in ACM CCS, 2015.

[5] M. D. I. A. M. M. Abdul Momen, R. H. C. Yap, M. A. R. Ahamed, "Robust Security and Privacy Preserving Techniques in Industrial IoT: A Review," IEEE Access, vol. 8, pp. 181665-181681, 2020.

[6] G. Xu, J. Yu, Z. Cheng, and H. Zhu, "A Privacy-Preserving Approach to Secure Industrial IoT Platforms: A Case Study of the Smart Factory," IEEE Access, vol. 9, pp. 38401-38410, 2021.

[7] K. Sangaiah, K. Yang, X. Guizani, and S. Ramakrishnan, "A Survey of Security and Privacy Issues in Industrial Internet of Things," IEEE Communications Surveys & Tutorials, vol. 22, no. 3, pp. 2020-2056, 2020.

[8] R. Zhang, M. A. Al Faruque, "Privacy-Aware Federated Learning for Industrial IoT Systems," in 2020 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton), 2020.

[9] M. Y. Aalsalem, A. H. Abdullah, M. K. A. Aziz, and M. Z. A. A. Aziz, "A Review of Privacy Preservation Techniques in the Internet of Things," IEEE Access, vol. 9, pp. 11055-11077, 2021.

[10] T. Harada, K. Sakurai, and S. Uda, "A Privacy-Preserving Data Collection Scheme in Industrial IoT Environments," in 2018 IEEE 43rd Conference on Local Computer Networks (LCN), 2018.

[11] D. N. Kosti and V. Stankovic, "Privacy-Preserving Data Processing in Industrial IoT: A Survey," IEEE Transactions on Industrial Informatics, vol. 16, no. 7, pp. 4535-4545, 2020.

[12] S. Chen, X. Xu, M. Ma, and J. Chen, "Secure and Privacy-Preserving Data Sharing and Collaboration in Industrial IoT," IEEE Transactions on Industrial Informatics, vol. 16, no. 5, pp. 3353-3362, 2020.

[13] R. Yahya, H. Chen, A. Al-Fuqaha, M. Guizani, and B. H. Kang, "Privacy-Preserving Machine Learning in Industrial Internet of Things: Challenges, Solutions, and Opportunities," IEEE Network, vol. 35, no. 6, pp. 52-58, 2021.

[14] Q. Wang, R. Yin, S. Zhang, X. Lin, and Y. Zhan, "Privacy-Preserving Inference for Industrial IoT: A Survey," IEEE Transactions on Industrial Informatics, vol. 17, no. 1, pp. 571-580, 2021.

[15] C. Zhan, Q. Liu, X. Zhu, X. Wang, and L. Zhang, "A Privacy-Preserving Machine Learning Framework for Industrial IoT: A Federated Learning Approach," in 2021 IEEE 47th Annual Conference on Local Computer Networks (LCN), 2021.