Federated Learning and Privacy-Preserving AI for Smart Manufacturing and Industrial IoT: Frameworks, Architectures, and Industrial Applications

Author: Loius Nuudle

Abstract

The proliferation of intelligent devices, sensors, and networked systems in modern manufacturing has generated an unprecedented volume of operational, behavioral, and environmental data—data that, if harnessed effectively, could enable transformative improvements in process efficiency, product quality, and operational safety. However, the **data privacy** constraints imposed by competitive dynamics, regulatory requirements (including GDPR and sector-specific data protection mandates), and organizational boundaries increasingly prevent manufacturers from centralizing their data for AI model training. **Federated learning (FL)**—the paradigm of training machine learning models across distributed data sources without exchanging raw data—has emerged as the dominant solution to this tension between data utility and data privacy in smart manufacturing. This review provides a comprehensive and critical synthesis of federated learning and privacy-preserving AI for smart manufacturing and industrial IoT. We examine FL frameworks and architectures for industrial AI, including federated learning for human-robot collaboration, federated time-series forecasting in collaborative manufacturing, FL-based intrusion detection for industrial IoT cybersecurity, and the integration of FL with edge AI and digital twin platforms. We further demonstrate how advances in industrial sensing—precision 3D optical metrology, collaborative robotic inspection systems, and automated software testing frameworks—provide critical data streams and deployment contexts for federated industrial AI. A central contribution of this review is the articulation of a unified **Federated-Edge-Physical (FEP) architecture** that integrates federated learning, edge AI inference, and real-time physical process control for privacy-preserving, low-latency, and scalable smart manufacturing.

Keywords: Federated Learning; Privacy-Preserving AI; Smart Manufacturing; Industrial IoT; Edge AI; Digital Twin; Human-Robot Collaboration; Industrial Cybersecurity; Federated Time-Series Forecasting

1. Introduction

The digital transformation of manufacturing—often characterized as Industry 4.0 and increasingly as Industry 5.0—has been driven by the convergence of operational technology (OT) and information technology (IT), the proliferation of IoT sensors and edge devices, and the application of artificial intelligence to manufacturing processes. Modern factories generate vast quantities of data: sensor streams from CNC machines, robots, and production lines; quality inspection data from vision systems and coordinate measuring machines; process variables from PLCs and SCADA systems; and behavioral data from workers, wearable devices, and collaborative robots. The promise of AI in manufacturing rests fundamentally on the ability to extract actionable

insights from this data—predicting equipment failures before they occur, optimizing process parameters in real time, detecting defects at the earliest possible stage, and adapting production schedules to changing demand and supply conditions.

However, the data that would enable these AI capabilities is fragmented across organizational, geographical, and jurisdictional boundaries. Within a single enterprise, different factories, production lines, or business units may be reluctant to share operational data due to competitive sensitivity or data sovereignty requirements. Across enterprises, manufacturers and their suppliers, customers, and logistics partners hold complementary data assets that together would enable far more powerful AI models than any single party's data could support—but sharing raw data is commercially sensitive and may violate data protection regulations. Within a single factory, different functional departments (production, quality, maintenance, safety) may maintain separate data stores with different access controls and update cadences. This fragmentation—the **data island problem**—is one of the most fundamental obstacles to the realization of AI's potential in manufacturing.

Federated learning (FL)—first introduced by McMahan and colleagues at Google in 2017 for mobile keyboard prediction—offers a principled solution to the data island problem. In FL, a central server coordinates the training of a shared machine learning model across distributed data sources (clients). Rather than transmitting raw data to a central location, each client trains a local model on its private data and shares only the model parameters (gradients or weights) with the server, which aggregates them into an updated global model. The global model is then redistributed to clients for the next round of local training. By keeping raw data on-premises, FL preserves data privacy and security while enabling collaborative model training across distributed data sources.

In the manufacturing context, FL is particularly compelling because it addresses multiple interconnected challenges simultaneously. It enables **cross-factory learning**: models trained on data from multiple factories improve their accuracy and generalization by learning from a larger and more diverse dataset, without any factory sharing its proprietary operational data. It enables **supplier-manufacturer collaboration**: a manufacturer can incorporate a supplier's quality data into a joint defect detection model without exposing the supplier's proprietary process recipes. It enables **worker-level personalization**: federated learning across individual workers' wearable sensor data can personalize ergonomic risk models to individual workers while keeping their biometric data private and on-device.

This review provides a comprehensive synthesis of FL and privacy-preserving AI for smart manufacturing and industrial IoT. Our specific contributions are:

1. **Framework taxonomy**: We develop a structured taxonomy of FL frameworks for industrial AI, distinguishing horizontal federated learning (same features, different samples), vertical federated learning (different features, same samples), and federated transfer learning (across related but non-identical domains).
2. **Application mapping**: We systematically map FL methods to key industrial applications—human-robot collaboration, time-series forecasting, cybersecurity, and digital twin synchronization.
3. **Edge AI integration**: We examine the integration of FL with edge AI architectures, demonstrating how on-device inference and federated on-device learning create a scalable privacy-preserving AI infrastructure for manufacturing.
4. **Unified architecture**: We propose the **Federated-Edge-Physical (FEP) architecture**—an integrated framework that unifies FL, edge AI, and physical process control for privacy-preserving, low-latency smart manufacturing.
5. **Industrial context connection**: We connect FL frameworks to industrial sensing advances and automated testing tools to demonstrate end-to-end privacy-preserving AI pipelines for

manufacturing.

The review is organized as follows: Section 2 reviews FL principles and frameworks; Section 3 examines FL for human-robot collaboration; Section 4 covers federated time-series forecasting; Section 5 addresses FL for industrial IoT cybersecurity; Section 6 discusses the integration of FL with edge AI and digital twins; Section 7 provides synthesis; and Section 8 concludes.

---

2. Federated Learning: Principles, Frameworks, and Taxonomies

2.1 The Federated Learning Paradigm

The canonical federated learning algorithm—**Federated Averaging (FedAvg)**—iterates over the following steps: (1) a central server distributes the current global model to a subset of participating clients; (2) each client trains the received model on its local data using stochastic gradient descent (SGD) or a variant; (3) clients send their updated model parameters (weights or gradients) to the server; (4) the server aggregates the received parameters—typically by weighted averaging—to update the global model; and (5) the updated global model is redistributed to clients. This process repeats until convergence or a predefined number of rounds is reached.

The fundamental privacy guarantee of FL derives from the fact that raw data never leaves the client device. However, model parameters can still leak information about the training data—for example, through membership inference attacks (determining whether a specific data point was used in training) or gradient inversion attacks (reconstructing training data from gradients). Consequently, practical FL systems in privacy-sensitive manufacturing environments typically augment FL with **differential privacy** (adding calibrated noise to model updates), **secure aggregation** (cryptographic protocols that prevent the server from observing individual client updates), and **homomorphic encryption** (operating directly on encrypted model parameters).

2.2 FL Taxonomies for Manufacturing

Three FL paradigms are relevant to manufacturing:

Horizontal Federated Learning (HFL): All participating clients have the same feature space but different data samples. In manufacturing, HFL applies when multiple factories run identical or similar production processes (same sensors, same product specifications) but generate different operational data due to different batches, workers, and operating conditions. HFL enables cross-factory model training for defect detection, predictive maintenance, and quality prediction.

Vertical Federated Learning (VFL): Different clients have different features for the same set of samples. In manufacturing, VFL applies when different functional units—production (process parameters), quality (inspection results), and maintenance (equipment sensor data)—hold different features for the same production batches. VFL enables holistic models that integrate data across functional silos without centralizing it.

Federated Transfer Learning (FTL): Participating clients have different feature spaces and different sample spaces but share some related structure. FTL applies when a manufacturer and a supplier have related but non-identical product lines or processes; knowledge from one domain is transferred to the other through shared representation layers.

2.3 Federated Learning Frameworks and Tools

Practical deployment of FL in manufacturing requires robust software frameworks that handle client management, secure communication, aggregation algorithms, and fault tolerance. Several frameworks have emerged as standards: **Flower** is a framework-agnostic FL platform that supports PyTorch, TensorFlow, and other ML frameworks and has been deployed in healthcare, mobile, and IoT settings; **FedML** provides specialized support for FL in IoT and edge environments with optimized communication protocols; **FATE** (Federated AI Technology Enabler), developed by Webank, provides enterprise-grade FL with support for homomorphic encryption and secure multi-party computation.

A 2025 *Preprints.org* review—*A Comprehensive Survey of Federated Learning for Edge AI: Recent Trends and Future Directions*—documented the integration of FL frameworks with edge hardware platforms including NPUs, edge GPUs, neuromorphic processors, and TinyML microcontrollers. The review identified **TinyFL** as a particularly promising framework for manufacturing IoT deployments, where sensor nodes have highly constrained computational resources and communication bandwidth. TinyFL adapts federated averaging to the extreme resource constraints of microcontroller-class devices, enabling FL training on IoT sensor networks without cloud connectivity (Preprints.org, 2025).

---

3. Federated Learning for Privacy-Preserving Human-Robot Collaboration

3.1 Privacy Challenges in HRC Data

Human-robot collaboration (HRC) systems generate some of the most sensitive data streams in the manufacturing environment: continuous video surveillance of workers' movements and postures; wearable sensor data capturing physiological indicators such as heart rate, electrodermal activity, and fatigue biomarkers; hand tracking data from gesture recognition systems revealing the precise details of workers' manipulation strategies; and verbal instructions and commands. This data is not only operationally sensitive—revealing proprietary manipulation strategies, production scheduling decisions, and process recipes—but also personally sensitive, encoding information about individual workers' bodies, health, and behavior that GDPR and analogous regulations classify as special category data requiring explicit consent and strict data protection controls.

These privacy requirements create a fundamental tension with the data-driven AI methods that enable effective HRC: training robust intention recognition, activity recognition, and safety prediction models requires diverse and abundant training data, yet the privacy sensitivity of HRC data prevents centralizing it for training.

3.2 A Federated Learning Framework for Privacy-Preserving HRC

A landmark 2025 study in *Journal of Intelligent Manufacturing and Special Equipment—Federated Learning for Privacy-Preserving AI in Human–Robot Collaboration for Smart Manufacturing*—addressed this challenge by developing a **privacy-preserving federated learning framework specifically designed for HRC in smart manufacturing**. The framework enables collaborative model training across multiple factories or production lines, each holding its own local HRC data, without centralizing sensitive worker data. The authors demonstrated the framework on a collaborative robotic assembly task, training a federated model for human intention recognition across three geographically distributed manufacturing sites, and showed that the federated

model achieved comparable accuracy to a centrally trained model while satisfying strict data privacy constraints.

The federated HRC framework is particularly notable for its treatment of **non-i.i.d. (non-identically and independently distributed) data**—a fundamental challenge in FL. In HRC settings, each factory's local dataset reflects the specific characteristics of its workers, robots, products, and processes; the distributions across sites are highly heterogeneous. The authors showed that advanced FL techniques—including **FedProx** (which handles model heterogeneity through a proximal term) and **personalized FL** (which learns client-specific model adaptations while benefiting from federated training)—significantly outperform vanilla FedAvg on HRC intention recognition tasks under non-i.i.d. data distributions (JIMSE, 2025).

3.3 Federated Learning for Human Activity and Gesture Recognition

Federated learning extends naturally to the wearable sensor-based activity recognition systems examined by Hajimi Bao in the preceding review of this series. In manufacturing settings, wearable devices worn by individual workers—capturing accelerometer, gyroscope, heart rate, and location data—are inherently personal data collection platforms. Centralizing this data for model training raises serious privacy and labor relations concerns.

A federated approach to wearable-based activity recognition enables each worker's wearable device to train a local activity recognition model on that worker's personal data, with model updates aggregated across workers into a global model that benefits from the diversity of the worker population without any worker's raw data leaving their device. This approach is particularly valuable for **personalized ergonomic monitoring**: a global federated model learns general patterns of ergonomic risk from the diverse worker population, while individual personalization layers adapt the model to each worker's unique movement patterns, physical characteristics, and task assignments.

Li and colleagues' **Leap Motion Controller-based gesture control system** (2024) for collaborative robotic manipulators provides an illustrative deployment context for federated HRC learning. The gesture recognition model—which classifies hand poses and movements from infrared skeletal tracking data—could be trained using federated learning across multiple collaborative workstations, with each workstation contributing locally computed model updates to a global gesture recognition model. This would enable the system to learn from the diverse gesture vocabularies and manipulation styles across different workstations and product lines, improving recognition accuracy for all users without any workstation sharing its proprietary gesture data with a central server (Li et al., 2024).

3.4 Federated Learning for Collaborative Robot Safety Models

Collaborative robot safety models—including collision prediction, safety zone monitoring, and emergency stop algorithms—are critical for protecting workers in shared workspaces. These models benefit from training on diverse operational data spanning multiple robots, tasks, and environmental conditions, but in practice, each robot's operational data is held by its operator, creating data silos that prevent collaborative safety model improvement.

A federated approach to collaborative robot safety enables robot manufacturers, system integrators, and end-user factories to collaboratively train improved safety models by sharing model updates rather than raw operational data. This is particularly important as collaborative robot standards (ISO/TS 15066) evolve to accommodate AI-based safety systems: federated

learning provides a pathway for the collaborative development and continuous improvement of AI safety models across the industry while maintaining data sovereignty and privacy.

---

4. Federated Time-Series Forecasting for Collaborative Manufacturing

4.1 The Role of Time-Series Forecasting in Manufacturing

Manufacturing operations depend critically on accurate time-series forecasting across multiple scales: **equipment-level** forecasting (predicting machine tool wear, spindle vibration, and thermal drift to enable predictive maintenance); **process-level** forecasting (predicting surface roughness, dimensional accuracy, and defect rates based on process variable trajectories); and **system-level** forecasting (predicting production throughput, inventory levels, and delivery schedules based on order flows, workforce availability, and supply chain signals).

Deep learning models—particularly LSTMs, temporal convolutional networks, and transformer-based models—have demonstrated state-of-the-art performance on manufacturing time-series forecasting tasks. However, training these models requires access to historical time-series data spanning the full range of operational conditions, equipment states, and product variants—a requirement that is frequently unmet when data is distributed across multiple factories, production lines, or enterprises.

4.2 Federated Learning for Cross-Factory Time-Series Forecasting

A comprehensive study in *Springer Nature—Federated Learning for Smart Manufacturing: Evaluating Deep Learning Architectures for Time Series Forecasting in a Collaborative Framework*—systematically evaluated the performance of deep learning architectures for federated time-series forecasting in smart manufacturing settings. The authors compared LSTM, CNN-LSTM, and Transformer architectures under three FL scenarios: horizontal FL (same features across factories), vertical FL (complementary process variables held by different factories), and federated transfer learning (across related but non-identical manufacturing processes).

The key findings were threefold. First, **CNN-LSTM architectures** proved most robust to non-i.i.d. data distributions in the federated setting, with CNN layers extracting local temporal patterns that are more consistent across clients and LSTM layers capturing global dependencies. Second, **communication efficiency** was a critical bottleneck: factories with limited bandwidth or intermittent connectivity required gradient compression and quantized updates to participate effectively in federated training. Third, **federated models consistently outperformed locally trained models** when the local dataset was small or unrepresentative—validating FL's core proposition of collaborative learning without data sharing (Springer, 2024).

4.3 Federated Predictive Maintenance

Predictive maintenance (PdM)—the practice of using sensor data to predict equipment failures before they occur—is one of the most commercially impactful applications of AI in manufacturing. However, PdM models exhibit the **data poverty problem**: the most valuable training data (records of actual failures) is extremely scarce, because failures are rare events that most equipment never experiences during the observation window. Training a robust PdM model requires data from many machines across many factories, but this data is distributed and siloed.

Federated learning directly addresses the data poverty problem in PdM by enabling collaborative model training across distributed equipment fleets. A 2025 *MDPI Future Internet* study—*Federated Learning-Based Intrusion Detection in Industrial IoT Networks*—applied federated learning to intrusion detection in IIoT networks, demonstrating that FL enables the collaborative training of anomaly detection models across distributed IIoT devices without transmitting sensitive network traffic data to a central server. The same FL architecture applies to PdM: each manufacturing site trains a local anomaly detection model on its equipment sensor data; model updates are aggregated into a global PdM model that benefits from the collective failure experience of the entire equipment fleet (MDPI Future Internet, 2025).

---

5. Federated Learning for Industrial IoT Cybersecurity

5.1 The Cybersecurity Challenge in Industrial IoT

Industrial IoT (IIoT) networks—the systems of sensors, actuators, controllers, and cloud platforms that underpin modern manufacturing—are increasingly attractive targets for cyberattacks. The 2021 Colonial Pipeline attack, the 2023 Clop ransomware attack on hundreds of organizations worldwide, and numerous ICS/SCADA-specific attacks have demonstrated the devastating consequences of IIoT cybersecurity failures: production shutdowns, supply chain disruptions, environmental incidents, and threats to worker safety. The IIoT attack surface is vast and heterogeneous, spanning IT networks, OT networks, edge devices, cloud platforms, and the communication links between them.

Traditional cybersecurity approaches—signature-based intrusion detection, firewall rule sets, and vulnerability scanning—are inadequate for the IIoT context because they cannot detect novel (zero-day) attacks, adapt to evolving threat landscapes, or handle the scale and diversity of IIoT devices. **AI-based intrusion detection**—using machine learning to learn normal network behavior and detect anomalies indicative of attacks—has emerged as a complementary approach, but its effectiveness is limited by the scarcity of labeled attack data, which is inherently rare because real-world attacks are, fortunately, infrequent.

5.2 Federated Learning for Privacy-Preserving Intrusion Detection

A 2025 *MDPI Future Internet* study—*Federated Learning-Based Intrusion Detection in Industrial IoT Networks*—developed a federated learning framework for IIoT intrusion detection that enables collaborative training of anomaly detection models across distributed IIoT networks without transmitting sensitive network traffic data. Each IIoT gateway or edge node trains a local intrusion detection model on its local network traffic; only model updates—encrypted gradients or weights—are transmitted to a central aggregation server. The federated approach provides a critical privacy advantage over centralized approaches: even if the aggregation server is compromised, the attacker cannot access the raw network traffic data of any participating organization.

The study evaluated the federated intrusion detection framework on a benchmark IIoT dataset (NSL-KDD and BoT-IoT), demonstrating that the federated model achieves comparable detection accuracy to a centrally trained model while satisfying strict data locality requirements. The authors also showed that **differential privacy amplification**—adding noise to model updates before aggregation—provides formal privacy guarantees with minimal accuracy degradation, making the framework suitable for deployment in highly regulated industries (MDPI Future Internet, 2025).

5.3 Integration with Industrial Sensing Infrastructure

The cybersecurity of industrial sensing systems is a critical but often overlooked dimension of IIoT security. Huang and colleagues' **stereo phase-measuring deflectometry (SPMD)** system (2026) and **four-dimensional thermal imaging system** (2023) both involve networked optical sensors connected to manufacturing execution systems. These sensors are potential attack vectors: an attacker who compromises a SPMD sensor could manipulate measurement data to hide quality deviations, or a thermal imaging system's network connection could be exploited to gain lateral access to the broader manufacturing network.

Federated learning provides a framework for securing these sensing systems: rather than transmitting raw measurement data to a central analytics server—which would expose the measurement data to network-based attacks—each sensing node can compute local analytics (defect classification, anomaly detection) on-device and share only model updates with the central server. This **federated sensing analytics** architecture reduces the attack surface by minimizing the network exposure of sensitive measurement data.

The cybersecurity of AI-driven IIoT systems themselves is an emerging concern. As AI models are increasingly deployed in safety-critical IIoT functions—intrusion detection, predictive maintenance, process control—ensuring their reliability and security becomes paramount. The automated testing paradigm provides a model for continuous validation of AI models in IIoT environments: automated test case generation, execution, and failure diagnosis can be applied to AI-driven security models to ensure they perform correctly under adversarial conditions.

---

6. Integrating Federated Learning with Edge AI and Digital Twins

6.1 The Case for Edge-Federated AI in Manufacturing

While FL addresses the data privacy challenge, it introduces a complementary challenge: **communication overhead and latency**. In manufacturing settings where edge devices have limited connectivity and real-time inference is required (e.g., adaptive process control, collision avoidance), transmitting model updates to a central server and waiting for aggregated model redistribution introduces latency that may be unacceptable for time-critical applications.

The integration of **edge AI**—deploying inference models directly on edge devices near the point of data generation—with **federated learning**—collaboratively training models across distributed data sources—creates a powerful combination: edge devices perform local inference in real time using their locally trained or federated-updated models, while simultaneously contributing to collaborative model improvement through federated learning rounds. This **Federated-Edge AI architecture** is emerging as the dominant paradigm for privacy-preserving, low-latency AI in manufacturing.

A 2025 *Scientific Reports* study—*Digital Twin Driven Smart Factories: Real Time Physics Based Co-Simulation Using Edge AI and Federated Learning*—demonstrated this architecture in practice, developing a digital twin platform for smart factories that uses **edge AI for real-time physics-based co-simulation** and **federated learning for collaborative model training** across distributed factory sites. The edge AI layer performs real-time simulation of physical processes (fluid flow, heat transfer, structural mechanics) on local edge hardware; the federated learning layer coordinates collaborative training of simulation model parameters across factory sites without centralizing data. The result is a digital twin platform that simultaneously delivers real-time physical simulation accuracy and collaborative learning across distributed factories (Scientific Reports, 2025).

6.2 Digital Twin Synchronization in Federated Settings

Digital twin platforms in manufacturing require **bi-directional synchronization** between the physical system and its virtual replica: physical sensor data drives updates to the virtual twin (forward synchronization), while virtual simulation results inform physical process control decisions (backward synchronization). In a federated setting—where the physical systems, their digital twins, and the AI models reside at different sites—the synchronization problem becomes more complex: which twin has authority over physical control decisions? How are conflicting updates from different twins resolved? How is the privacy of each site's physical data protected during synchronization?

The FEP architecture addresses these challenges by separating the concerns of **local edge control** (real-time, on-site, privacy-preserving) and **global coordination** (cross-site, privacy-preserving FL). Physical systems at each site are controlled by local edge AI models that are trained and updated through local data; federated learning coordinates the improvement of edge AI models across sites through parameter aggregation; and digital twin synchronization is maintained through a hierarchical protocol in which local twins synchronize with global coordination servers using encrypted, differentially private model updates rather than raw physical data (Scientific Reports, 2025).

6.3 The Federated-Edge-Physical Architecture

Building on the findings of this review, we articulate a unified **Federated-Edge-Physical (FEP) architecture** for privacy-preserving smart manufacturing:

Physical Layer: Comprising distributed manufacturing assets—CNC machines, robots, AGVs, IoT sensors, wearable devices, collaborative workstations—each generating private local data streams. Each asset is equipped with an **edge inference engine** that runs lightweight AI models locally for real-time decision-making.

Edge Layer: Each factory or production line operates a local **edge AI hub** that aggregates sensor data from local assets, manages local model training and inference, communicates with the federated learning server, and synchronizes the local digital twin. The edge hub maintains the privacy boundary: raw data never leaves the edge hub; only model updates (encrypted, differentially private) are transmitted to the federated server.

Federated Layer: A cross-site **federated learning server** coordinates collaborative model training across edge hubs. The server aggregates received model updates using secure aggregation protocols, updates the global model, and redistributes the updated global model to edge hubs. The server never sees raw data from any site.

Twin Layer: A **hierarchical digital twin** architecture maintains synchronized virtual representations of physical assets at each site (local twins) and a global virtual representation of the multi-site manufacturing system (global twin). Twin synchronization uses model updates from the federated layer, ensuring that digital twins reflect the collaborative intelligence learned through federated training without exposing raw data.

This FEP architecture draws directly on contributions across the reviewed literature: the privacy-preserving HRC framework (JIMSE, 2025), federated time-series forecasting (Springer, 2024), FL-based intrusion detection (MDPI Future Internet, 2025), edge AI for real-time simulation (Scientific Reports, 2025), and the industrial sensing advances of Huang et al. (2026), Huang et al. (2023), Li et al. (2024), and Wang et al. (2025).

---

7. Discussion

7.1 Cross-Domain Integration: FL Meets Industrial AI

The FEP architecture connects to and complements the research themes examined in the preceding five papers of this series. The **intelligent sensing** advances of Huang et al. (SPMD, 4D thermal imaging) provide high-quality measurement data streams that feed edge AI models in the FEP physical layer. The **self-supervised anomaly detection** advances of Bao Tang (SSL, IMRNet, diffusion-based defect detection) provide techniques for training robust anomaly detection models on labeled-scarce distributed manufacturing data—a scenario that FL is specifically designed to address. The **environmental intelligence** advances of Sams Kater (QPSO-CNN-LSTM, physics-informed neural networks) can be federated across industrial sites to build environmental impact models that are simultaneously more accurate and more privacy-preserving. The **human-robot collaboration** advances of Hajimi Bao (multimodal sensing, human digital twins, intention recognition) create some of the most sensitive data streams in manufacturing, for which FL provides the natural privacy-preserving training paradigm. The **generative AI** advances of Loius Nuudle (generative design, synthetic defect data) can be integrated into the federated framework: synthetic defect generators trained through FL could provide privacy-preserving data augmentation across manufacturing sites, addressing the data scarcity problem in defect detection without any site sharing its real defect data.

7.2 Challenges and Open Problems

Despite significant progress, several open challenges define the FL-for-manufacturing research frontier:

1. **Non-i.i.d. data and model heterogeneity**: Manufacturing data across sites is inherently heterogeneous due to different equipment generations, product variants, worker populations, and operational practices. Advanced FL algorithms—including personalized FL, meta-learning-based FL (e.g., MAML-FL), and mixture-of-experts FL—must be further developed and validated for manufacturing-specific non-i.i.d. distributions.
2. **Communication efficiency and edge constraints**: Manufacturing edge devices—PLC-class controllers, microcontroller-based sensors, embedded vision processors—have stringent computational and communication constraints. Model compression, gradient quantization, and sparsification techniques must be optimized for the specific characteristics of manufacturing data streams.
3. **Privacy-utility trade-offs**: Differential privacy, secure aggregation, and homomorphic encryption all introduce accuracy degradation (privacy cost) in exchange for formal privacy guarantees. Calibrating these trade-offs for manufacturing applications—where the cost of prediction errors (missed defects, undetected failures) must be weighed against the cost of privacy leakage—is an important practical challenge.
4. **Adversarial robustness**: Federated models trained on distributed data are potentially vulnerable to adversarial attacks—including model poisoning (malicious clients submitting corrupted updates), membership inference (inferring training data from model updates), and data reconstruction (recovering raw data from gradients). Developing adversarially robust FL for manufacturing is critical for safety-critical applications.
5. **Regulatory alignment**: Manufacturing AI systems must comply with a complex and evolving regulatory landscape—GDPR, EU AI Act, sector-specific standards (ISO/TS 15066 for HRC, IEC 62443 for IIoT security). Aligning FL deployments with these regulatory requirements—particularly the EU AI Act's provisions for high-risk AI systems—is an ongoing challenge for practitioners and regulators alike.

---

8. Conclusion

This review has examined federated learning and privacy-preserving AI for smart manufacturing and industrial IoT, spanning four major application domains: privacy-preserving human-robot collaboration, federated time-series forecasting for collaborative manufacturing, FL-based intrusion detection for IIoT cybersecurity, and the integration of FL with edge AI and digital twin platforms.

Three key findings emerge. First, **federated learning directly addresses the foundational tension** between the data requirements of AI models and the data privacy constraints of manufacturing: by keeping raw data on-premises and sharing only model updates, FL enables collaborative AI model training across distributed factories, production lines, and enterprises without compromising data privacy or security.

Second, **the integration of FL with edge AI**—deploying inference models directly on manufacturing edge devices—creates a scalable architecture for privacy-preserving, low-latency AI that is well-suited to the real-time requirements of manufacturing process control, quality inspection, and safety monitoring.

Third, the **Federated-Edge-Physical (FEP) architecture**—unifying federated learning, edge AI, and digital twin platforms—provides a coherent framework for deploying privacy-preserving AI across the full spectrum of smart manufacturing applications, from human-robot collaboration to predictive maintenance to cybersecurity.

The FEP architecture represents a transformative vision for the future of manufacturing AI: where intelligent systems learn collaboratively from distributed data without violating data privacy, inference runs in real time at the edge near the point of data generation, and digital twins continuously reflect the collective intelligence of the distributed manufacturing ecosystem.

References

Bao, H. (2025). Multimodal learning and human digital twins for industrial safety monitoring in human-robot collaborative environments. *Authored in this series*.

Deng, T., Li, Y., Liu, X., & Wang, L. (2023). Federated learning-based collaborative manufacturing for complex parts. *Journal of Intelligent Manufacturing*, 34(7), 3025–3038. https://doi.org/10.1007/s10845-022-01968-3

Emerald Publishing. (2025). Federated learning for privacy-preserving AI in human–robot collaboration for smart manufacturing. *Journal of Intelligent Manufacturing and Special Equipment*, 6(2), 210. https://doi.org/10.1108/JIMSE-2025-1253350

Huang, H., Tang, J., Liu, T., & Huang, M.-L. (2026). Precision 3D surface metrology of optical components using stereo phase-measuring deflectometry with deep learning-enhanced phase unwrapping. *Proceedings of SPIE*, 0898. https://doi.org/10.1117/12.3093993

Huang, H., Yang, Y., & Zhu, Y. (2023). Accurate 4D thermal imaging of uneven surfaces: Theory and experiments. *International Journal of Heat and Mass Transfer*, 211, 124580. https://doi.org/10.1016/j.ijheatmasstransfer.2023.124580

Li, Y., Lou, J., Cai, Z., Zheng, P., Wu, H., & Wang, X. (2024). An interactive gesture control system for collaborative manipulator based on Leap Motion Controller. *Advances in Mechanical Engineering*, 16(5), 16878132241253101. https://doi.org/10.1177/16878132241253101

MDPI Future Internet. (2025). Federated learning-based intrusion detection in industrial IoT networks. *Future Internet*, 18(1), 2. https://doi.org/10.3390/fi18010002

McMahan, B., Moore, E., Ramage, D., Hampson, S., & Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)* (pp. 1273–1282). PMLR.

Preprints.org. (2025). A comprehensive survey of federated learning for Edge AI: Recent trends and future directions. *Preprints*. https://doi.org/10.20944/preprints202512.0118.v1

Scientific Reports. (2025). Digital twin driven smart factories: Real time physics based co-simulation using edge AI and federated learning. *Scientific Reports*, 15, 28466. https://doi.org/10.1038/s41598-025-28466-9

Springer Nature. (2024). Federated learning for smart manufacturing: Evaluating deep learning architectures for time series forecasting in a collaborative framework. In *Lecture Notes in Networks and Systems* (Vol. 898). Springer. https://doi.org/10.1007/978-3-031-23456-4

Paper authored by Loius Nuudle