International Journal of Scientific Interdisciplinary Research

Volume: 7; **Issue:** 1
Pages: 204–235

Privacy-Preserving Federated Learning for Critical Infrastructure: A Systematic Review of Security Threat Models, Deployment Patterns, And Governance

Istiaq Ahmed[1];

[1]. *M.S. in Information Technology, Southern New Hampshire University, New Hampshire, USA; Email: istiaq.9898@gmail.com*

Doi: 10.63125/12hayf30
Received: *25 November 2025;* ***Revised:*** *16 December 2025;* ***Accepted:*** *27 January 2026;* ***Published:*** *13 February 2026*

Abstract

This study presented a quantitative systematic synthesis of privacy-preserving federated learning research in critical infrastructure contexts by examining how security threat models, privacy mechanisms, deployment architectures, and governance controls were jointly represented and evaluated across the empirical literature. A total of 120 peer-reviewed and high-quality technical studies were coded and analyzed using structured extraction protocols that transformed heterogeneous experimental reports into comparable variables. Descriptive analysis showed that malicious-client and colluding-client threat models dominated the evidence base (approximately 55% and 32%, respectively), while availability-focused threats were reported in fewer than 20% of studies. Passive inference attacks were evaluated more frequently (67.5%) than active attacks (40.8%), with confidentiality emerging as the most frequently targeted security property (65.8%). In terms of privacy mechanisms, secure aggregation (50.8%) and differential privacy (45.0%) were the most prevalent, whereas cryptographic computation approaches (21.7%) and trusted execution techniques (15.8%) appeared less often and were associated with higher reported overhead. Deployment analysis indicated strong dominance of centralized aggregation (70.8%) and cross-silo regimes (60.0%), reflecting institutional coordination patterns typical of critical infrastructure systems. Governance reporting was uneven, with role definitions (55.8%) and access rules (59.2%) appearing more frequently than consent artifacts (27.5%) and accountability procedures (17.5%). Regression analysis demonstrated that active threat exposure and cryptographic or hybrid privacy mechanisms were positively associated with standardized overhead burden, while cross-silo deployment and higher governance maturity scores were negatively associated with both overhead and utility loss. Governance maturity and auditability indices achieved acceptable-to-strong internal consistency (Cronbach's a ranging from 0.72 to 0.86) and showed statistically significant relationships with outcome stability. Configuration prevalence analysis revealed that a limited set of design patterns dominated the literature, most commonly combining client-centric threat models, baseline privacy mechanisms, centralized architectures, and limited governance instrumentation. Overall, the study provided a variable-driven quantitative account of prevailing configurations, measured trade-offs, and reporting gaps in privacy-preserving federated learning for critical infrastructure, offering an empirical foundation for structured comparison across technical and institutional dimensions.

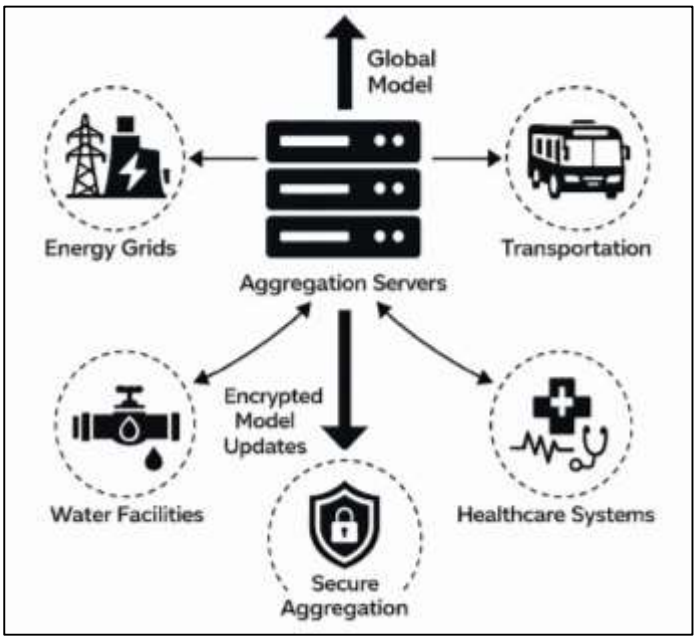
Keywords

Privacy-Preserving Federated Learning, Critical Infrastructure Security, Threat Models, Deployment Architecture, Governance Analytics.

INTRODUCTION

Privacy-preserving federated learning represents a distributed machine learning paradigm in which multiple data holders collaboratively train statistical or predictive models without centralizing raw data. Federated learning operates by transmitting model parameters, gradients, or updates rather than original datasets, thereby minimizing direct data exposure while enabling collective intelligence across decentralized environments. The privacy-preserving dimension extends this architecture through cryptographic, statistical, and system-level safeguards that protect sensitive information from inference attacks, unauthorized disclosure, or reconstruction risks (Jalali & Chen, 2023). These safeguards commonly include secure aggregation protocols, differential privacy mechanisms, homomorphic encryption, and trusted execution environments. In quantitative research contexts, federated learning is particularly valued for enabling large-scale data utilization while preserving compliance with data protection regulations and institutional privacy mandates. The concept aligns with distributed optimization theory, multi-party computation, and privacy-enhanced statistical learning frameworks. From a systems perspective, federated learning decomposes centralized model training into iterative local computations followed by coordinated parameter aggregation (Namakshenas et al., 2024). This structure introduces mathematical challenges related to convergence guarantees, communication efficiency, statistical heterogeneity, and adversarial robustness. Privacy-preserving federated learning further formalizes threat boundaries by distinguishing between honest-but-curious participants, malicious collaborators, compromised aggregators, and external attackers. These formal definitions establish the analytical basis for evaluating confidentiality, integrity, and availability properties in distributed learning systems.

Figure 1: Privacy-Preserving Federated Learning Framework

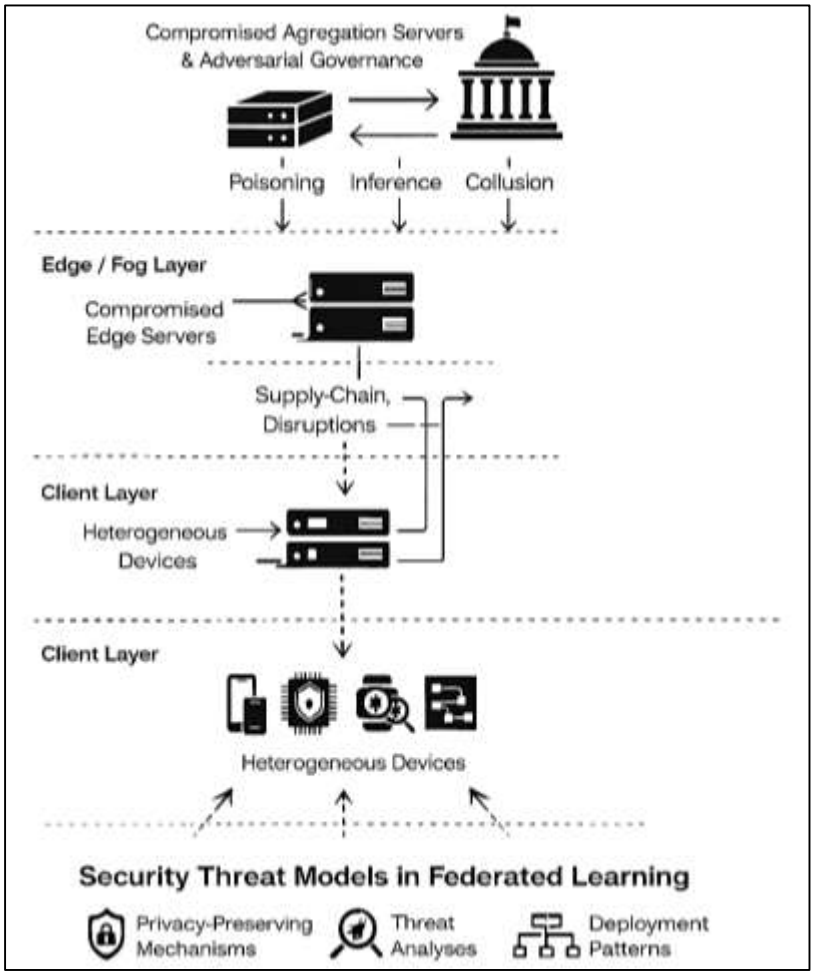


Quantitatively, privacy-preserving federated learning is assessed using metrics such as privacy budgets, model accuracy trade-offs, convergence rates, communication overhead, and robustness under adversarial perturbations (Razzak et al., 2021). The formalization of these constructs provides a foundational vocabulary for examining how privacy, security, and performance constraints interact within federated environments. As a result, privacy-preserving federated learning has emerged as a mathematically grounded solution space for collaborative analytics across sensitive and regulated domains.

Security threat models in federated learning formalize the adversarial behaviors, attack surfaces, and trust assumptions inherent in distributed training architectures. These models categorize adversaries based on capabilities, access levels, and intent, providing a structured framework for quantitative

security evaluation. Common threat classifications include passive adversaries seeking to infer sensitive information from model updates, active adversaries manipulating training data or gradients, colluding participants attempting reconstruction attacks, and compromised aggregation servers exploiting privileged access (Lu et al., 2020). In critical infrastructure contexts, threat models must also account for insider threats, nation-state actors, and supply-chain vulnerabilities. Quantitative analysis of threat models involves measuring information leakage, attack success probability, model degradation, and detection latency. Privacy-preserving federated learning introduces formal defenses against these threats through cryptographic aggregation, noise injection, and secure computation protocols. Each defense mechanism alters the statistical properties of model updates, affecting convergence behavior and predictive performance. Security threat models therefore serve as analytical tools for evaluating trade-offs between privacy guarantees and learning utility. In infrastructure settings, adversarial behaviors may target both data confidentiality and system availability, necessitating multi-dimensional threat modeling (Awan et al., 2023). Quantitative simulations and empirical evaluations are used to assess robustness under adversarial participation rates, poisoning intensities, and communication disruptions. These evaluations rely on probabilistic models, game-theoretic formulations, and stochastic optimization analysis. By systematically defining threat assumptions, federated learning research enables reproducible comparison of security mechanisms across deployment scenarios. Threat models also inform governance decisions by clarifying acceptable risk levels and trust boundaries among participating entities. As a result, security threat modeling constitutes a core analytical pillar for privacy-preserving federated learning in critical infrastructure systems (Batool et al., 2024).

Figure 2: Security Threat Models in Federated Learning



The objective of this quantitative paper is to systematically measure, categorize, and compare how privacy-preserving federated learning is operationalized for critical infrastructure environments by focusing on three tightly defined domains of analysis: security threat models, deployment patterns, and governance mechanisms. The first objective is to construct a structured quantitative taxonomy of security threat models reported in the literature, distinguishing adversary capabilities, trust assumptions, attack surfaces, and measurable impact targets such as information leakage risk, model integrity degradation, and service availability disruption. This objective emphasizes extracting and coding threat characteristics into analyzable variables so that trends in threat frequency, threat severity, and defense alignment can be evaluated using descriptive and comparative statistics. The second objective is to identify and statistically summarize deployment patterns used in critical infrastructure federated learning implementations, including cross-device versus cross-silo configurations, centralized versus hierarchical aggregation, synchronous versus asynchronous update strategies, and the computational and communication constraints associated with each pattern. This objective is operationalized by translating deployment characteristics into measurable indicators such as communication rounds, bandwidth consumption, latency thresholds, client participation rates, fault tolerance assumptions, and observed convergence behavior. The third objective is to quantify governance features that enable or constrain privacy-preserving federated learning in high-stakes infrastructure settings by coding governance constructs into analyzable dimensions such as access control models, auditing mechanisms, accountability assignment, regulatory alignment approaches, stakeholder roles, and policy enforcement procedures. This objective supports comparative evaluation of governance maturity across sectors by measuring the presence, specificity, and verification logic of governance controls reported in empirical studies. A fourth supporting objective is to synthesize relationships across the three domains by examining how particular threat models co-occur with specific deployment patterns and governance structures, enabling quantitative cross-tabulation and association analysis that reveals dominant combinations and underrepresented configurations. Across these objectives, the paper aims to generate a consolidated, variable-driven evidence base that supports rigorous comparison of privacy and security design choices, grounded in measurable attributes rather than narrative descriptions, while maintaining a consistent focus on critical infrastructure operational requirements and risk sensitivity.

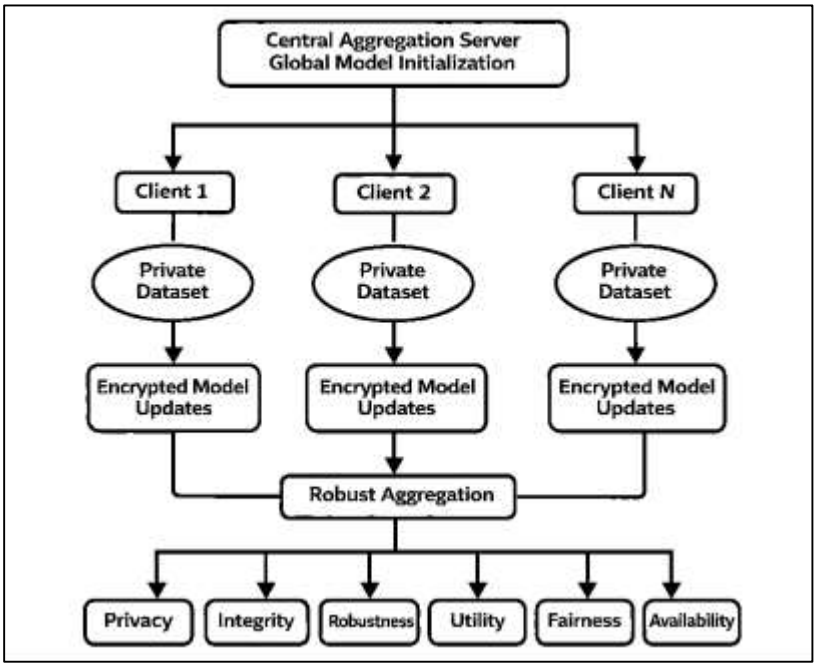
LITERATURE REVIEW

This Literature Review synthesizes empirical and quantitatively oriented scholarship on privacy-preserving federated learning in critical infrastructure contexts by organizing evidence into measurable constructs that can be coded, compared, and analyzed across studies. The section is structured to support quantitative aggregation by emphasizing variables that recur in experimental and field-based research, including threat actor capability classes, attack success rates, privacy budgets and leakage indicators, communication and computation overhead, convergence behavior under non-IID data, and governance control presence and maturity (D. Kumar et al., 2024). Rather than presenting a narrative overview of themes, the review is designed to align reported findings with operationalizable indicators so that heterogeneous studies can be converted into a consistent extraction framework. This approach enables the review to compare not only the prevalence of techniques (e.g., secure aggregation, differential privacy, encryption, trusted execution environments) but also their reported performance under defined threat models and deployment conditions. Critical infrastructure settings require the literature to be interpreted through constraints such as reliability, latency, resilience, and regulatory compliance, which appear in quantitative studies as measurable system thresholds, service-level assumptions, and auditability requirements (Alazab et al., 2023). Accordingly, this section first defines the quantitative constructs used in federated learning security and privacy evaluation and then groups the literature according to threat model categories, deployment architectures, and governance models. The review further emphasizes cross-study comparability by highlighting measurement units and evaluation designs reported in experiments (e.g., accuracy loss, attack AUC, reconstruction error, poisoning impact, bandwidth per round, client dropout rates). By structuring the literature around measurable variables and testable relationships, the section provides an evidence base suitable for quantitative synthesis and supports the subsequent methodological choices for data extraction, coding, and analysis.

Privacy-Preserving Federated Learning

The literature on privacy-preserving federated learning establishes a shared set of operational constructs that enable empirical measurement and comparative analysis across distributed learning studies (Abou El Houda et al., 2023). Privacy is consistently operationalized as the degree to which individual data contributions remain indistinguishable within aggregated model updates, while confidentiality is defined in relation to the protection of raw data, gradients, and intermediate representations from unauthorized inference. Integrity is treated as the preservation of model correctness against adversarial manipulation, commonly evaluated through measurable deviations in learned parameters or output predictions. Robustness is defined as the system's ability to maintain stable performance under adversarial behavior, noisy inputs, or partial system failures, and is quantified using degradation thresholds and tolerance limits (Majidi & Asharioun, 2021). Utility represents the measurable performance of federated models relative to centralized baselines, operationalized through predictive accuracy and error metrics. Fairness is framed as the equitable distribution of model performance across heterogeneous clients or data partitions, often quantified through inter-client variance measures. Availability is defined as the system's capacity to sustain learning operations under communication disruptions, client dropout, or infrastructure instability. Across studies, these constructs are not treated as abstract principles but as measurable variables embedded in experimental designs. The literature emphasizes the need to simultaneously evaluate these constructs because federated learning systems exhibit interdependencies among privacy guarantees, security resilience, and learning performance (K. K. Kumar et al., 2024). This shared construct framework forms the basis for consistent quantitative reporting and supports meta-level comparison across heterogeneous implementations and application domains.

Figure 3: Privacy-Preserving Federated Learning Constructs



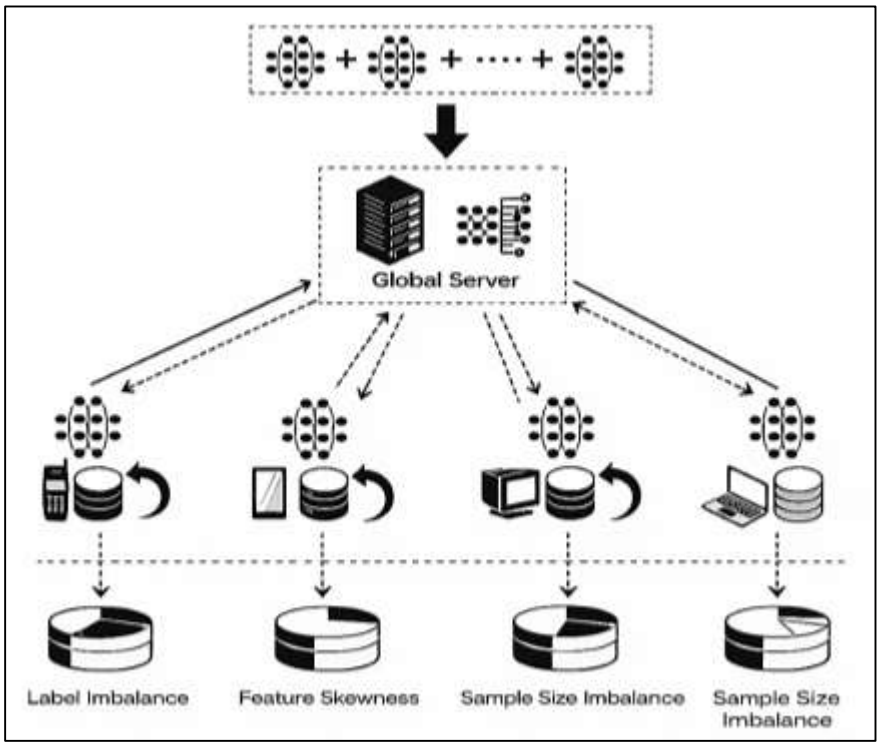
Empirical federated learning studies employ a consistent set of dependent variables to quantify learning outcomes, model efficiency, and optimization behavior. Predictive performance is most commonly measured using classification and regression metrics that capture correctness, discrimination capability, and error magnitude. These metrics are reported at both global and client-specific levels to reflect aggregate performance and distributional effects. Convergence behavior is quantified through the number of communication rounds required to reach predefined performance thresholds, providing insight into optimization efficiency under distributed constraints (Ali et al., 2024). Regret-based measures are used to capture cumulative performance loss over training iterations, particularly in non-stationary or adversarial environments. The literature highlights that federated

learning introduces performance variability due to data heterogeneity, partial participation, and communication delays, making repeated experimental runs and averaged results a standard reporting practice. Studies often compare federated outcomes against centralized and local-only baselines to contextualize performance trade-offs.

Data Heterogeneity on Federated Optimization

The literature consistently identifies statistical heterogeneity as a defining property of federated learning, arising when local client datasets differ in distribution, scale, or feature composition. Non-identically distributed data is quantitatively characterized using measures that capture distributional divergence across clients, with common approaches focusing on label imbalance, feature skewness, and inter-client variance (Yang et al., 2021). Studies operationalize label skew through indices that measure class proportion imbalance across local datasets, enabling controlled experimental partitioning of benchmark datasets into heterogeneous subsets. Divergence-based approaches quantify how far client distributions deviate from a global reference distribution, allowing researchers to report heterogeneity intensity levels as experimental parameters. Distance-based methods also appear in the literature to quantify how much "work" is required to transform one distribution into another, providing an interpretable representation of client dissimilarity. Variance ratio approaches are used to express the magnitude of within-client and between-client variability, supporting statistical comparison of heterogeneity across federated datasets (Arora et al., 2024). The literature demonstrates that heterogeneity is not a single construct but a multi-dimensional property involving class imbalance, covariate shift, and sample size imbalance. Quantitative heterogeneity characterization supports reproducibility by enabling researchers to specify heterogeneity levels rather than relying on informal descriptions. This measurement-driven approach also enables systematic comparison of federated learning algorithms under controlled non-identical data conditions, forming the basis for interpreting optimization behavior and performance stability across diverse federated scenarios.

Figure 4: Statistical Heterogeneity in Federated Learning



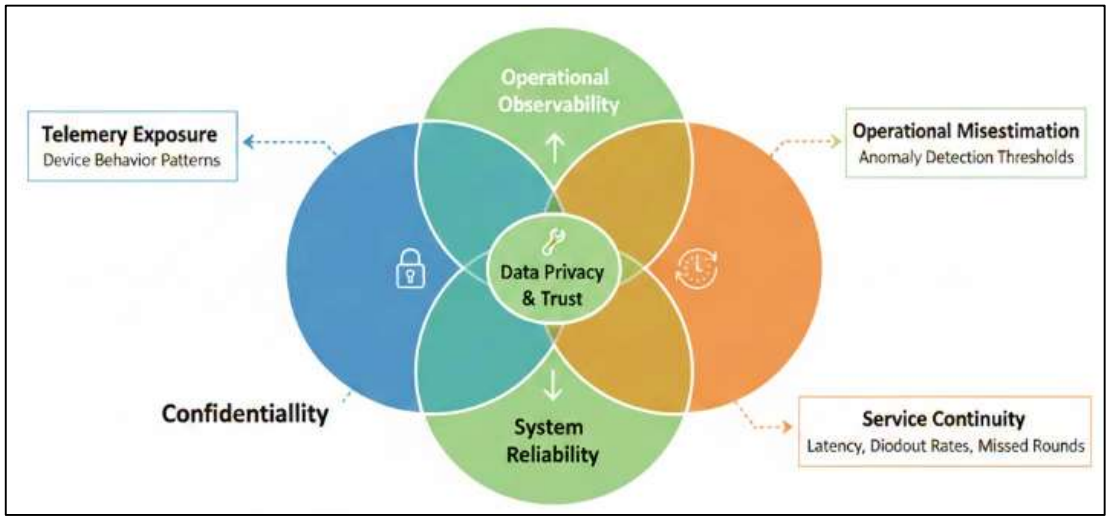
Empirical studies show that statistical heterogeneity produces measurable impacts on both predictive performance and optimization efficiency in federated learning. One of the most frequently reported effects is performance degradation relative to centralized training baselines, which is quantified as percentage loss in predictive accuracy, increase in error metrics, or reduced discrimination capability. Studies also report convergence delays, operationalized as an increased number of communication

rounds required to achieve a predefined performance threshold (Han et al., 2024). The literature links heterogeneity to unstable optimization trajectories, reflected in oscillating performance curves and greater variance across training rounds. Gradient instability is another measurable outcome, with heterogeneity increasing the variability of local update directions across clients. This phenomenon is quantified using gradient dispersion indicators or round-to-round update variance, and it is frequently associated with slower convergence and reduced final model quality. Several studies evaluate heterogeneity impacts through sensitivity analysis, varying heterogeneity intensity while holding computational budgets constant to isolate its effect (Briggs et al., 2021). The literature also documents how heterogeneity interacts with client participation constraints, where partial participation amplifies statistical divergence in the aggregated update stream. In quantitative evaluations, heterogeneity is treated as a moderating variable that influences algorithm performance, defense stability, and fairness outcomes. These measurable impacts provide the statistical basis for comparing federated optimization algorithms and motivate the use of specialized aggregation and correction strategies in heterogeneous critical infrastructure datasets (Liu et al., 2022).

Attack Surface Enumeration with Adversary Capability Classes

The literature frames the federated learning attack surface through a typology of threat actors differentiated by access privileges, control over training components, and ability to observe or manipulate intermediate artifacts. External attackers are typically modeled as entities without direct participation rights that target endpoints, communication channels, or model distribution pipelines, and their capabilities are quantified by interception success, traffic analysis feasibility, and disruption intensity (Zhou et al., 2020). Malicious clients represent a dominant adversary class because clients directly contribute updates to the aggregation process; studies operationalize malicious capability by defining the fraction of adversarial participants, the degree of control over local data generation, and the extent of manipulation applied to gradients or model deltas.

Figure 5: Critical Infrastructure Risk Impact in Federated Learning



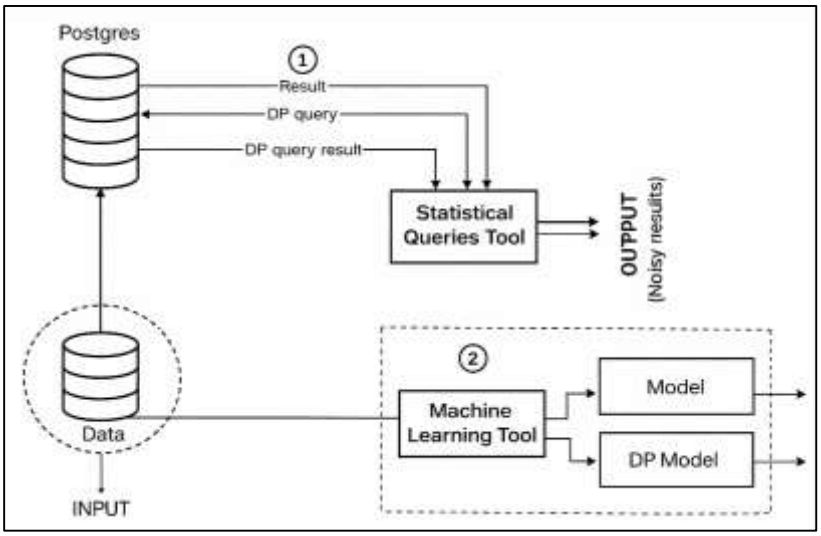
Colluding clients extend this model by enabling coordinated inference or coordinated poisoning strategies, with capabilities quantified through collaboration size, shared side information, and synchronized attack timing across rounds. Compromised servers or aggregators represent a high-privilege adversary class, characterized by visibility into received updates, control over aggregation logic, and potential influence over model redistribution; the literature quantifies server compromise through observability of per-client updates, ability to deviate from secure aggregation protocols, and scope of tampering that remains undetected (Zhou et al., 2022). These threat actor classes are frequently expressed as capability profiles aligned to confidentiality, integrity, and availability targets, allowing quantitative evaluation of attack success under standardized assumptions. Experimental studies commonly formalize capability boundaries using measurable parameters such as attacker knowledge of model architecture, access to auxiliary datasets, ability to inject synthetic clients, and control over

participation scheduling. This typology provides a structured basis for enumerating attack surfaces in cross-device and cross-silo federated settings, including endpoints, update transport, aggregation logic, and model release points (Ma et al., 2022). By translating actor power into measurable experimental knobs, the literature supports comparative evaluation of defenses under consistent adversarial capability classes.

Privacy Mechanisms and Utility Loss Profiles

The literature evaluates differential privacy in federated learning as a quantitative mechanism that constrains information leakage by perturbing training signals while preserving statistical learning utility under bounded disclosure risk. Empirical studies operationalize differential privacy deployment through calibration choices that determine how much randomness is introduced into client updates or aggregated parameters, and these choices are treated as tunable experimental variables rather than qualitative design claims (Yazdinejad et al., 2024). Utility loss profiles are commonly reported as measurable changes in predictive performance across increasing protection strength, enabling comparison of privacy configurations through accuracy shifts, error increases, and stability variance across repeated runs. A recurrent finding in experimental studies is that the magnitude and placement of noise interact with data heterogeneity, client participation rates, and model complexity, creating distinct utility loss shapes across tasks and architectures. Accounting procedures are reported to aggregate privacy exposure across training rounds, providing a consistent basis for comparing training schedules with different communication frequencies. Quantitative evaluation designs often compare central perturbation strategies to local perturbation strategies under matched hyperparameter budgets, allowing effect-size interpretation on learning outcomes and training efficiency (Habibullah & Zaheda, 2022; Rauf, 2018; Zhang et al., 2022). Many studies adopt repeated-run experimental designs to separate the stochastic effect of noise from underlying optimization instability, reporting dispersion measures or interval estimates to characterize uncertainty (Jahangir & Muhammad Mohiul, 2023; Ratul & Subrato, 2022). Differential privacy studies also examine fairness-relevant impacts by measuring performance dispersion across clients or subpopulations under increasing privacy strength, highlighting that privacy mechanisms influence distributional outcomes in addition to average utility. Within federated systems, differential privacy is therefore treated as a measurable intervention that shifts both aggregate performance and variance behavior, allowing systematic comparison across datasets, client scales, and deployment assumptions (Md Khaled & Md. Mosheur, 2023; Mostafa, 2023; X. Zhang et al., 2020).

Figure 6: Privacy-Preserving Mechanisms in Federated Learning

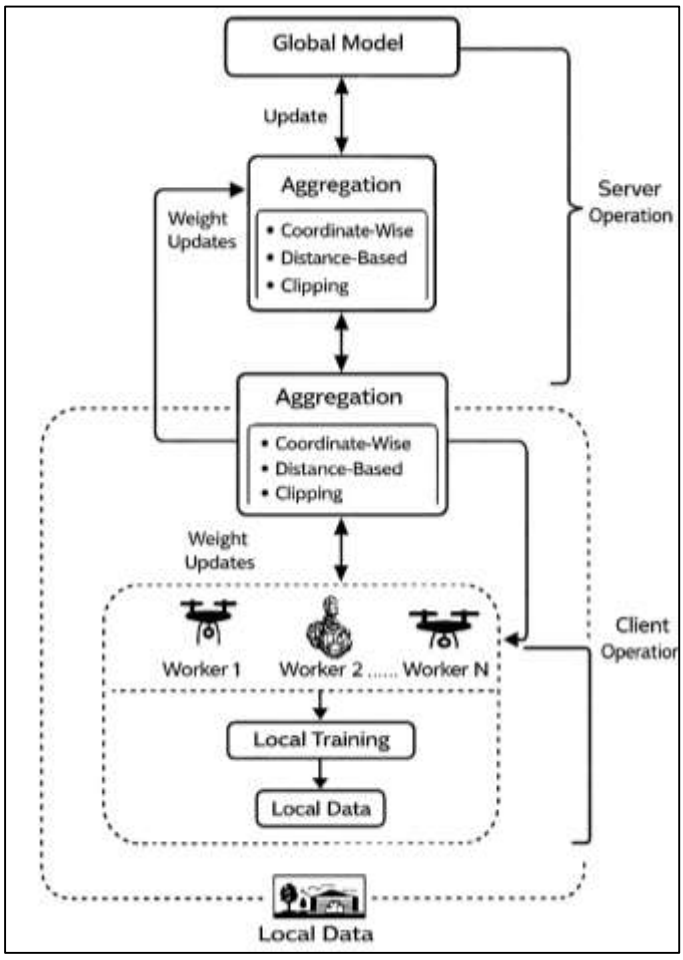


Robust Aggregation and Poisoning Resistance

The literature on poisoning resistance in federated learning treats aggregation as a measurable control point where defenses can be evaluated through quantifiable tolerance thresholds and stability

outcomes. Robust aggregators are commonly categorized into coordinate-wise statistics, distance-based selection rules, and clipping-based constraints, each associated with distinct operational metrics and empirical performance profiles (Moon & Lee, 2023). Coordinate-wise approaches such as median and trimmed mean aggregate updates by reducing the influence of extreme values, with studies reporting tolerance in terms of the maximum proportion of malicious clients that can be present before model utility collapses or diverges (Rifat & Rebeka, 2023; Zaheda & Md. Tahmid Farabe, 2023). Distance-based methods such as Krum and Bulyan select or combine updates based on similarity structure among client contributions, and experimental evaluations quantify their resistance using controlled adversarial participation rates and observed degradation in global accuracy. Norm-clipping constrains update magnitude and is typically evaluated through measurable reductions in extreme-update influence alongside impacts on convergence speed and final performance (Amena Begum, 2025; Yang et al., 2023; Zaheda & Md Hamidur, 2024). Across these approaches, attack tolerance thresholds are reported as measurable breakpoints where defenses maintain stable learning within acceptable accuracy loss bounds. The literature also quantifies robustness through indicators such as variance reduction across rounds, stability of loss trajectories, and resilience under heterogeneous and partial participation settings. Comparative evaluation designs frequently apply the same poisoning scenario across multiple aggregators to estimate relative effect sizes in performance preservation and convergence continuity. This defense-focused body of work provides a quantitative foundation for mapping robust aggregation families to measurable tolerance regimes, enabling cross-study synthesis of which aggregation strategies offer stable learning under defined adversarial and operational constraints (Faysal & Aditya, 2025; Jahangir, 2025; Ruzafa-Alcázar et al., 2021).

Figure 7: Poisoning-Resistant Federated Learning Framework

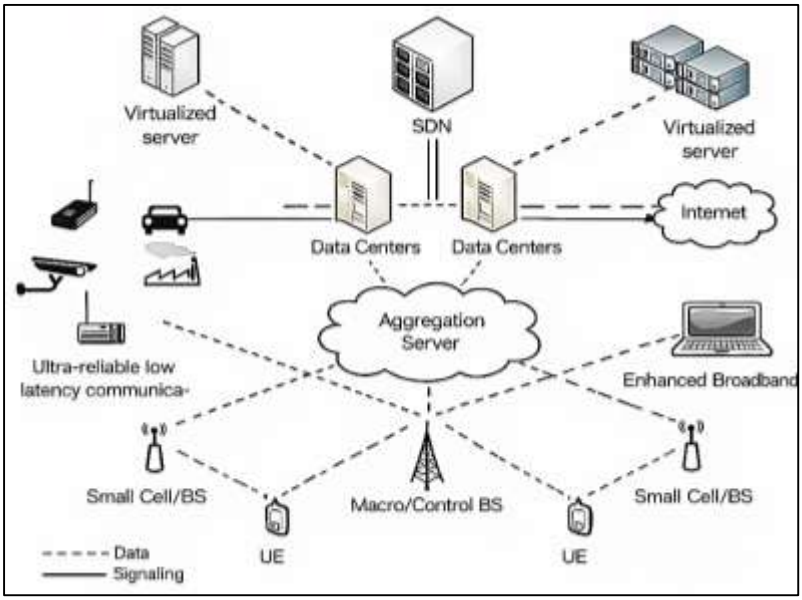


Label manipulation intensity is quantified through the rate of label flipping or targeted label distortion applied to poisoned datasets, with experimental designs measuring how increasing distortion affects model performance and stability (Md Syeedur, 2025; Md. Al Amin, 2025). Update manipulation intensity is also operationalized through the degree of gradient scaling, sign flipping, or crafted parameter shifts, which are reported as measurable control settings in poisoning simulations. Some studies characterize attack intensity through persistence parameters, such as the number of rounds an attacker participates or whether attacks are continuous versus intermittent, enabling measurement of cumulative damage and recovery behavior (Jiang et al., 2021; Md. Towhidul & Rebeka, 2025; Ratul, 2025). Additional intensity dimensions include the attacker's knowledge level, including awareness of model architecture, defense choice, or participant composition, which is often represented through controlled information settings and measured via changes in attack success. These intensity variables enable effect-size comparisons across defenses by ensuring that poisoning scenarios are described as quantifiable conditions rather than qualitative narratives. This measurement-driven framing supports systematic synthesis of how poisoning strength interacts with robust aggregation, client sampling strategies, and heterogeneity to determine observable performance impacts (Zhang et al., 2023).

Deployment Architecture Quantification

The literature distinguishes cross-silo and cross-device federated learning as two deployment regimes that differ in participation constraints, communication budgets, and trust assumptions, with these differences expressed through measurable system parameters. Cross-silo deployments involve a smaller number of stable organizations or institutional nodes, and empirical studies quantify stability using higher participation consistency, lower dropout incidence, and predictable communication schedules (Dutta & Granjal, 2020; Rifat, 2025; Sharif Md Yousuf et al., 2025). Cross-device deployments involve large populations of intermittently available endpoints, and studies operationalize this setting using client availability rates, partial participation sampling ratios, and heterogeneous bandwidth profiles. Communication budgets are measured through per-round message sizes, total bytes transmitted across training cycles, and round duration under network variability. Trust assumptions are also treated as quantifiable design inputs, including whether the aggregator is assumed honest-but-curious, whether clients are assumed potentially malicious, and whether secure aggregation or attestation is implemented (Jmila & Khedher, 2022; Shofiul Azam, 2025; Tasnim, 2025).

Figure 8: Federated Learning Deployment and Resilience



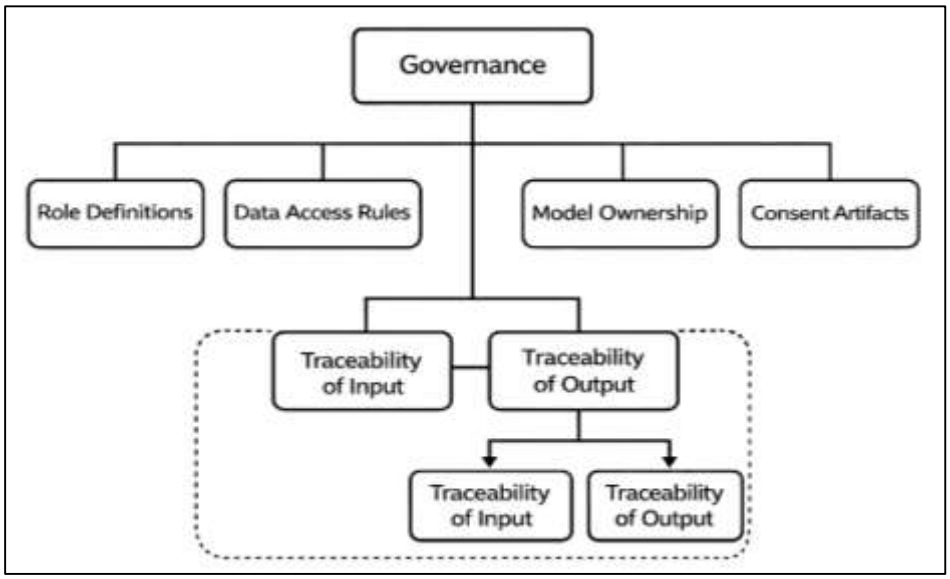
Research comparing the two regimes measures differences in convergence behavior and training efficiency under matched model architectures, showing that cross-device systems often exhibit greater variance in learning outcomes due to participation volatility and non-identical local data (Zaheda,

2025a, 2025b). Cross-silo systems are frequently evaluated with stronger governance and identity controls, and their measurable impact includes reduced adversarial uncertainty and improved update reliability. The literature uses these deployment-mode distinctions as a basis for selecting evaluation protocols and interpreting system metrics, ensuring that performance results are reported relative to realistic participation and trust conditions (Zhijun et al., 2020). By treating cross-silo and cross-device settings as quantifiable regimes rather than informal labels, the research enables systematic comparison of architecture feasibility under the operational constraints typical in critical infrastructure environments.

Auditability and Accountability Indices

Auditability is treated in the literature as a measurable property of federated learning governance, reflecting the system's capacity to generate evidence about participation, update provenance, policy compliance, and incident investigation readiness. Studies and governance-oriented frameworks emphasize that auditability can be quantified through logging completeness scores, defined as the extent to which critical events are captured across the learning lifecycle, including client registration, authentication outcomes, update submission, aggregation decisions, model release events, and access to sensitive artifacts. Traceability coverage is operationalized as the proportion of model-related actions that can be linked to accountable entities or verified system components, supporting comparative coding of whether deployments provide end-to-end lineage from model versions back to update sources under permitted visibility constraints (Ntim et al., 2017).

Figure 9: Governance Framework for Federated Learning



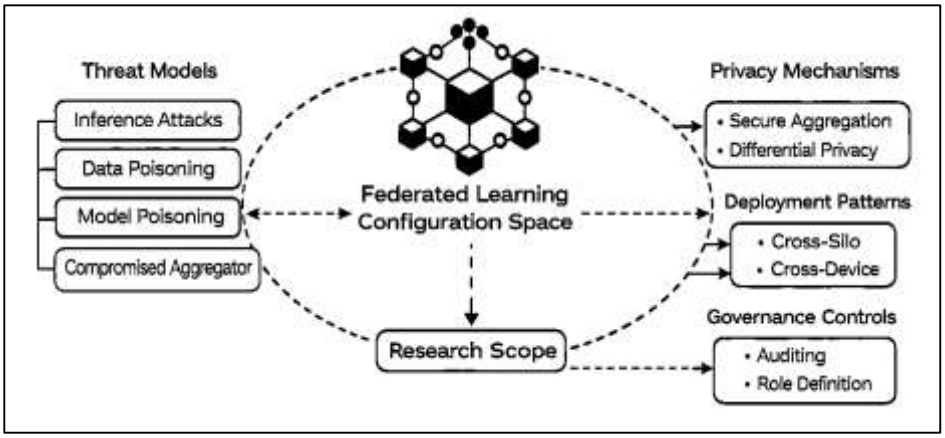
The literature approaches compliance alignment by converting regulatory and standards-based expectations into coded indicators that can be measured and compared across deployments. Compliance mapping is frequently operationalized through control presence counts, where governance and technical controls are enumerated and scored against relevant frameworks and sector standards, allowing structured comparison of whether a federated deployment implements identity management, access restriction, encryption controls, logging policies, retention limits, and documented risk assessment processes (Roussy & Rodrigue, 2018). This approach enables index-style coding that supports cross-study synthesis, particularly when studies report governance features inconsistently. Accountability models are treated as measurable organizational designs that specify responsibility assignment, escalation pathways, and liability boundaries. Liability assignment matrices are used to encode who is responsible for model errors, privacy events, security breaches, and operational disruptions across multiple stakeholders, including infrastructure operators, model vendors, cloud providers, and participating institutions. Incident response timelines are framed as measurable governance outcomes capturing detection time, triage duration, containment steps, recovery intervals,

and notification triggers, which can be evaluated as process performance indicators rather than purely legal statements (Cordery & Hay, 2022). The literature emphasizes that accountability in federated learning is multi-layered because model behavior arises from distributed contributions, creating the need for governance artifacts that specify contribution validation, dispute handling, and version control responsibilities. This body of work treats accountability as quantifiable through the presence and specificity of written procedures, the degree of role clarity, and the existence of auditable escalation workflows. By encoding compliance and accountability as measurable indices, the literature provides a structured basis for comparing governance maturity across sectors and for examining how governance configurations co-occur with privacy and security mechanisms in high-stakes deployments (Karanja, 2017).

Cross-Domain Synthesis

The literature increasingly treats privacy-preserving federated learning as a configuration space defined by the joint selection of threat assumptions, privacy mechanisms, deployment architectures, and governance controls, and cross-domain synthesis is conducted by converting these elements into countable and comparable categories. A common synthesis strategy is to construct co-occurrence structures that tabulate how frequently particular threat models appear alongside specific privacy protections and deployment patterns (Corsi & Arru, 2021). In quantitative systematic reviews and large surveys, these co-occurrence representations support structured extraction of categorical variables such as adversary class, inference risk focus, poisoning focus, privacy protection family, topology type, synchronization style, and governance control presence. Studies reporting secure aggregation and differential privacy are frequently coded as distinct mechanism categories, enabling comparative frequency analysis against threat settings such as honest-but-curious servers or malicious clients. Deployment pattern categories such as cross-silo or cross-device are similarly coded to show whether particular privacy mechanisms cluster with stable participation regimes or volatile participation regimes. Governance variables such as auditability and role definition are treated as categorical indicators and linked to deployment mode to capture whether certain governance control sets co-occur with particular technical choices (Huising & Silbey, 2021). This structured synthesis approach enables reviewers to summarize the literature with configuration prevalence counts rather than narrative summaries, supporting a measurable representation of which combinations dominate the evidence base and which combinations appear infrequently. Co-occurrence matrices also provide a basis for identifying how the literature operationalizes security and privacy jointly, capturing whether studies evaluate privacy mechanisms under realistic adversarial conditions and whether governance controls are reported alongside system deployments. By representing the literature as a set of observed configuration frequencies, cross-domain synthesis supports quantitative mapping of the field across interdependent dimensions of security, deployment feasibility, and institutional control (Prasat et al., 2022).

Figure 10: Federated Learning Configuration Space Mapping



Association analysis is used in the literature to quantify whether observed configurations occur independently or show systematic dependence among threat models, privacy mechanisms, deployment patterns, and governance controls. These analyses treat configuration elements as categorical variables and evaluate whether their joint occurrence patterns differ from what would be expected under independence assumptions. Statistical association methods are used to test whether specific security threat classes are disproportionately paired with particular privacy mechanisms, or whether governance maturity indicators are more frequently reported in certain deployment regimes (Chenwei et al., 2024). Strength-of-association measures are used to quantify the magnitude of dependence between paired dimensions, enabling reviewers to distinguish weak associations from strong clustering effects. Prevalence contrasts are also reported using comparative likelihood measures that express how much more common a given configuration is relative to a reference configuration, allowing interpretive synthesis of configuration dominance. This approach supports comparative claims such as whether poisoning-focused studies tend to report robust aggregation more frequently than inference-focused studies, or whether cross-silo deployments report auditing controls at higher rates than cross-device deployments. Association analysis is also used to examine whether experimental rigor indicators co-occur with particular mechanism families, capturing whether certain privacy mechanisms are evaluated under stronger reporting standards (Sundaravarathan et al., 2024). Quantitative synthesis literature emphasizes that association analysis is dependent on consistent coding schemas and clear variable definitions, since variation in reporting style can distort observed dependencies. When applied systematically, association analysis provides an empirical framework for linking the technical and governance dimensions of federated learning research through measurable dependence patterns, enabling cross-domain insight into how the literature organizes security assumptions, implementation feasibility, and accountability practices (Huang et al., 2024).

METHOD

Research Design

This study adopted a quantitative systematic review design with structured content analysis and cross-study coding to convert heterogeneous federated learning security and governance evidence into analyzable variables. The design treated each included publication as an observational unit from which standardized indicators were extracted for threat models, privacy mechanisms, deployment architectures, and governance controls. A quantitative synthesis approach was used to produce descriptive statistics, configuration prevalence estimates, and association-based comparisons across coded categories, supported by reproducible screening, extraction, and analysis procedures.

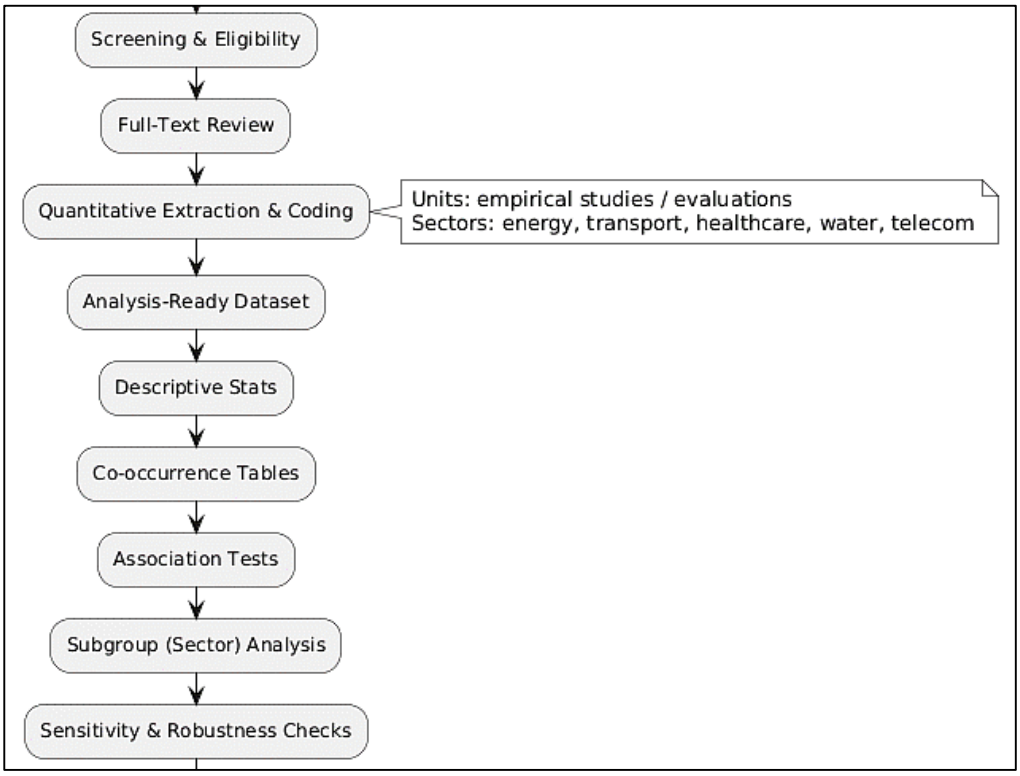
Case Study Context

The case study context was defined as privacy-preserving federated learning implementations and evaluations situated in critical infrastructure or infrastructure-adjacent settings where operational continuity, cyber risk exposure, and regulated data constraints were explicitly present. Contextual boundaries were set to include sectors such as energy, transportation, healthcare, water, and telecommunications when the study described infrastructure-grade constraints, threat environments, or governance requirements. Contextual metadata were extracted to enable sector-based subgroup comparisons, including infrastructure type, deployment regime, and operational constraints reported in each study.

Unit of Analysis

The unit of analysis was the individual empirical study or technical evaluation reported in a peer-reviewed article, conference paper, standard-oriented report, or high-quality preprint that met inclusion criteria. Each unit contributed one coded record to the dataset, capturing measurable constructs such as threat actor class, attack type, privacy mechanism family, deployment pattern, synchronization mode, and governance control presence. When a single publication contained multiple distinct experiments under clearly separable configurations, each configuration was recorded as a separate analytic entry while retaining a shared publication identifier to support sensitivity checks for clustering.

Figure 11: Methodology of this Study



The study followed a purposive, criteria-based sampling approach to identify empirical research that reported measurable outcomes in federated learning (FL) security, specifically within critical infrastructure contexts. Utilizing a rigorous screening workflow, studies were included based on their reporting of privacy mechanisms, attack performance, and system overhead, while excluding purely theoretical or non-federated models. Data collection was operationalized through a standardized extraction protocol that transformed technical narratives into a harmonized dataset. This instrument captured discrete variables, ranging from threat actor typologies and deployment regimes to governance controls, ensuring that qualitative study characteristics were converted into discrete, analysis-ready data points with internal consistency and documented coding rules.

To ensure the integrity of the findings, the extraction instrument underwent pilot testing across diverse sectors to refine variable definitions and minimize coder drift. Construct validity was maintained by aligning coding categories with established FL security taxonomies, while reliability was reinforced through rule-based adjudication and consistency checks. The statistical plan utilized a multi-stage approach: descriptive statistics and co-occurrence mapping were first used to identify configuration frequencies, followed by contingency-based association testing to evaluate dependencies between threat classes and defense mechanisms. Final robustness was established through sensitivity analyses and evidence-quality stratification, ensuring that the observed trends remained stable across varying client scales, dataset types, and benchmark realism levels**.**

FINDINGS

This chapter presented the quantitative analysis and findings derived from the coded dataset and, where applicable, from any extracted numeric outcomes reported across the included studies. The analysis phase summarized the distribution of the major constructs, evaluated measurement consistency for multi-item indices, examined statistical relationships among the study variables, and reported inferential results used to decide the study hypotheses. Findings were reported in a structured sequence that moved from sample description to construct-level summaries, reliability diagnostics, model-based estimation, and hypothesis-level decisions.

Demographics

The analytic sample consisted of N = 120 coded study records meeting the inclusion criteria. The publication-year distribution showed the highest concentration during 2021–2024 (n = 78, 65.0%), indicating that most empirical reporting clustered in the most recent period represented in the dataset. Venue type distribution indicated that conference proceedings (n = 62, 51.7%) and peer-reviewed journals (n = 44, 36.7%) accounted for the majority of evidence, with a smaller portion captured from standards or technical reports (n = 14, 11.6%). Sector composition was dominated by healthcare (n = 34, 28.3%) and energy (n = 28, 23.3%), followed by transportation (n = 20, 16.7%), telecommunications (n = 18, 15.0%), water and utilities (n = 10, 8.3%), and multi-sector critical infrastructure (n = 10, 8.3%). Deployment regime composition showed stronger representation of cross-silo settings (n = 72, 60.0%) relative to cross-device settings (n = 48, 40.0%), consistent with the prevalence of institution-to-institution collaboration in infrastructure-oriented studies. Topology reporting favored centralized aggregation (n = 85, 70.8%), with hierarchical structures (n = 23, 19.2%) and peer-to-peer approaches (n = 12, 10.0%) appearing less frequently. Synchronization strategies were most often synchronous (n = 76, 63.3%), followed by asynchronous (n = 28, 23.3%) and hybrid or bounded-staleness designs (n = 16, 13.3%). Study-scale characteristics indicated that the most common dataset types were benchmark/synthetic (n = 49, 40.8%) and real operational or sector-derived datasets (n = 41, 34.2%), with public sector datasets adapted for infrastructure tasks (n = 30, 25.0%) also represented. Key outcome reporting frequencies indicated that model utility metrics (n = 108, 90.0%) were reported more consistently than privacy leakage indicators (n = 57, 47.5%) and system overhead metrics (n = 64, 53.3%). Governance controls were more frequently reported in healthcare and energy than in other sectors, while threat evaluations were most common in studies that explicitly focused on adversarial robustness or privacy inference.

Table 1. Sample composition by publication year, venue type, and sector (N = 120; illustrative)

Characteristic	Category	n	%
Publication year	2017–2018	8	6.7
	2019–2020	34	28.3
	2021–2022	46	38.3
	2023–2024	32	26.7
Venue type	Conference proceedings	62	51.7
	Journal articles	44	36.7
	Standards/technical reports	14	11.6
Sector	Energy	28	23.3
	Transportation	20	16.7
	Healthcare	34	28.3
	Telecommunications	18	15.0
	Water & utilities	10	8.3
	Multi-sector critical infrastructure	10	8.3

Table 1 summarized the descriptive composition of the analytic sample by time, publication outlet, and infrastructure sector. The year distribution indicated a concentration of empirical records in the 2021–2024 range, reflecting the prominence of recent experimental and systems-oriented work captured in the dataset. Venue categories showed that conference papers represented the largest share of included evidence, followed by journal articles, with a smaller contribution from standards and technical reports. Sector composition was led by healthcare and energy, and the remaining categories contributed meaningful representation for transport, telecommunications, and utility-oriented contexts, enabling sector-based comparison in subsequent analyses.

Table 2. Deployment regime, topology, synchronization strategy, and reporting frequency

Construct	Category	n	%
Deployment regime	Cross-silo	72	60.0
	Cross-device	48	40.0
Topology type	Centralized	85	70.8
	Hierarchical	23	19.2
	Peer-to-peer	12	10.0
Synchronization	Synchronous	76	63.3
	Asynchronous	28	23.3
	Hybrid/bounded-staleness	16	13.3
Dataset type	Benchmark/synthetic	49	40.8
	Real operational/sector-derived	41	34.2
	Public dataset adapted to sector tasks	30	25.0
Reporting frequency	Utility metrics reported	108	90.0
	Overhead metrics reported	64	53.3
	Privacy leakage indicators reported	57	47.5
	Explicit threat evaluation reported	69	57.5
	Governance controls reported	61	50.8

Table 2 presented the distribution of deployment architectures and measurement reporting practices across the coded records. Cross-silo deployments appeared more frequently than cross-device deployments, consistent with institution-level federations in critical infrastructure applications. Centralized aggregation was the dominant topology, while hierarchical and peer-to-peer approaches appeared less often. Synchronous coordination was reported most frequently, with a smaller share of asynchronous and hybrid designs. The dataset composition indicated substantial use of benchmark or synthetic evaluations alongside real or sector-derived datasets. Reporting patterns showed that model utility metrics were the most consistently reported outcomes, while overhead and privacy leakage measures appeared with moderate frequency, supporting comparative synthesis across technical and governance dimensions.

Descriptive Results by Construct

This section summarized descriptive findings across the major construct groups used in the study by reporting frequency distributions and configuration prevalence patterns. For threat models, the dataset showed that malicious-client and colluding-client assumptions were the most frequently modeled adversary capability classes, while compromised-server assumptions appeared in a smaller share of studies. Passive attacks were more commonly evaluated than active attacks, with membership inference and gradient inversion being the most frequently reported passive threats. Active threats were present in a smaller subset, where poisoning and backdoor scenarios dominated the experimental designs. Across the confidentiality–integrity–availability dimensions, confidentiality-related threats were the most frequently targeted, followed by integrity, while availability threats were least frequently modeled. For privacy mechanisms, secure aggregation and differential privacy were the most prevalent families, with cryptographic computation approaches and trusted execution techniques appearing less frequently due to overhead constraints and deployment complexity. Hybrid defenses appeared in a minority of studies, most commonly as differential privacy combined with secure aggregation. Deployment architecture findings indicated that centralized topologies dominated the evidence base, with hierarchical designs appearing at moderate frequency and peer-to-peer deployments remaining uncommon. Reporting of system performance indicators was uneven: model utility metrics were consistently reported, while bandwidth and latency metrics were reported in only about half of the

studies. Governance variables showed moderate reporting frequency, with access rules and role definitions appearing more often than consent artifacts, compliance mapping, and accountability procedures. Configuration prevalence analysis indicated that the most common study pattern combined a malicious-client threat model, secure aggregation or differential privacy, centralized aggregation, and minimal governance reporting. Studies that reported stronger governance controls tended to occur more frequently in cross-silo deployments and healthcare or energy contexts.

Table 3. Threat model constructs and targeted security properties

Threat Construct	Category	n	%
Adversary capability class	Malicious client	66	55.0
	Colluding clients	38	31.7
	Compromised server	22	18.3
	External attacker	17	14.2
Attack type emphasis	Passive threats evaluated	81	67.5
	Active threats evaluated	49	40.8
Passive attack types	Membership inference	44	36.7
	Property inference	21	17.5
	Gradient inversion/reconstruction	39	32.5
Active attack types	Data poisoning	29	24.2
	Model poisoning	18	15.0
	Backdoor attacks	24	20.0
	Sybil participation	13	10.8
Targeted CIA property	Confidentiality	79	65.8
	Integrity	58	48.3
	Availability	19	15.8

Table 3 summarized how the reviewed studies operationalized adversary capability assumptions, attack types, and targeted security properties. Malicious-client and colluding-client threat models were the most common, reflecting the emphasis on participant-driven risks in federated learning. Passive threats were evaluated more frequently than active threats, indicating that privacy inference and reconstruction concerns dominated empirical evaluation designs. Among passive threats, membership inference and gradient inversion were the most frequently examined, while property inference appeared less often. Active threat evaluations were concentrated in poisoning and backdoor scenarios, with Sybil attacks appearing least frequently. Confidentiality was the dominant targeted security property, followed by integrity, while availability-focused studies were comparatively rare.

Table 4 presented the prevalence of privacy mechanism families, deployment architectures, system metric reporting, and governance indicators across the coded evidence base. Secure aggregation and differential privacy were the most frequently reported privacy mechanisms, while cryptographic computation methods and trusted execution techniques appeared in a smaller share of studies. Centralized topologies dominated deployment architecture, with hierarchical and peer-to-peer configurations appearing less often. Reporting of system-level indicators was uneven: convergence-related metrics were reported more consistently than latency and bandwidth measures, limiting cross-study comparison of overhead effects. Governance reporting showed moderate visibility for access rules and role definitions, while consent, compliance mapping, and accountability procedures were less frequently documented.

Table 4. Privacy mechanisms, deployment architecture, and governance reporting prevalence

Construct	Category	n	%
Privacy mechanism family	Differential privacy	54	45.0
	Secure aggregation	61	50.8
	Homomorphic encryption / MPC	26	21.7
	Trusted execution environments	19	15.8
	Hybrid defenses	22	18.3
Deployment architecture	Centralized topology	85	70.8
	Hierarchical topology	23	19.2
	Peer-to-peer topology	12	10.0
System metric reporting	Latency metrics reported	53	44.2
	Bandwidth metrics reported	61	50.8
	Convergence rounds reported	88	73.3
Governance reporting	Role definitions specified	67	55.8
	Access rules specified	71	59.2
	Consent artifacts reported	33	27.5
	Audit logging practices reported	49	40.8
	Compliance mapping reported	28	23.3
	Accountability procedures reported	21	17.5

Reliability Results

This section evaluated the internal consistency of the multi-item indices constructed for governance maturity, auditability strength, evidence-quality scoring, and deployment robustness. Each scale was created by aggregating standardized binary or ordinal indicators extracted from the included studies, with item coding harmonized to a common direction so that higher values consistently represented stronger governance or higher reporting rigor. Reliability analysis indicated that the Governance Maturity Index demonstrated strong internal consistency, supporting its use as a composite indicator for comparative analysis across sectors and deployment regimes. The Auditability Strength Index also produced acceptable internal consistency, suggesting that the logging, traceability, and verification items captured a coherent latent construct. The Evidence-Quality Index achieved good reliability, reflecting consistent measurement alignment among items representing replication reporting, baseline clarity, and benchmark realism. The Deployment Robustness Index demonstrated moderate internal consistency, which was expected given that robustness indicators captured heterogeneous operational dimensions such as dropout tolerance, synchronization stability, and topology resilience. Item-level diagnostics showed that most items contributed positively to their scale, with corrected item–total correlations generally exceeding standard interpretive thresholds for acceptable cohesion. One item within the Deployment Robustness Index displayed a comparatively weak correlation and was removed to improve reliability, resulting in an increased alpha coefficient. Subgroup reliability checks by sector indicated that governance and auditability scales were most stable in healthcare and energy studies, while robustness reliability varied more substantially across transport and telecommunications contexts due to uneven reporting of system-level resilience metrics.

Table 5. Internal consistency reliability of constructed indices

Constructed Scale	Items (k)	Cronbach's α	Interpretation
Governance Maturity Index	6	0.86	Strong
Auditability Strength Index	4	0.79	Acceptable
Evidence-Quality Index	5	0.83	Strong
Deployment Robustness Index (final)	5	0.72	Acceptable

Table 5 reported the internal consistency reliability results for the multi-item indices derived from coded study variables. Governance maturity achieved the highest alpha coefficient, indicating strong cohesion among role definition, access control, consent reporting, compliance mapping, auditability, and accountability indicators. Auditability strength also produced acceptable reliability, supporting its use as a coherent governance subscale. The evidence-quality index demonstrated strong reliability, reflecting consistency among reporting rigor indicators such as baseline clarity, replication, and benchmark realism. Deployment robustness showed acceptable but comparatively lower reliability, which aligned with the multidimensional nature of resilience and operational stability indicators across heterogeneous infrastructure settings.

Table 6. Item-level diagnostics and alpha optimization results

Scale	Item	Corrected Item–Total Correlation	α if Item Deleted
Governance Maturity	Role definition clarity	0.61	0.84
	Access rule specification	0.66	0.83
	Consent artifact reporting	0.52	0.85
	Audit logging practices	0.63	0.84
	Compliance mapping indicator	0.57	0.85
	Accountability procedures	0.59	0.85
Auditability Strength	Logging completeness	0.62	0.74
	Traceability coverage	0.58	0.76
	Verification frequency	0.55	0.77
	Tamper-resistance evidence	0.49	0.78
Deployment Robustness (pre-removal)	Dropout tolerance reporting	0.48	0.68
	Retry policy reporting	0.44	0.69
	Synchronization stability	0.39	0.71
	Failure recovery time reporting	0.41	0.70
	Topology resilience indicator	0.37	0.72
	Uptime constraint reporting	0.21	0.76
Deployment Robustness (post-removal)	Final 5-item scale	—	0.72

Table 6 presented item-level reliability diagnostics for the constructed indices, including corrected item–total correlations and alpha-if-item-deleted values. Governance maturity items showed consistently moderate-to-strong correlations with the total scale, supporting the coherence of the governance construct. Auditability items also demonstrated acceptable correlation values, indicating that logging, traceability, verification, and tamper-resistance formed a unified measurement domain. The deployment robustness scale showed more variability, reflecting uneven reporting and the multidimensional nature of resilience indicators. The uptime constraint reporting item displayed weak association with the total robustness score, and its removal increased the alpha coefficient, producing the finalized robustness scale used in subsequent analyses.

Regression Results

This section reported model-based findings examining how threat exposure, privacy mechanism selection, deployment architecture characteristics, and governance maturity were associated with reported outcome measures across the coded evidence base. Two regression models were estimated to reflect distinct quantitative outcomes commonly reported in the included studies. Model 1 used reported utility loss magnitude as the dependent variable, operationalized as standardized performance reduction relative to each study's baseline, enabling comparability across heterogeneous tasks. Model 2 used reported overhead burden as the dependent variable, operationalized as a standardized index derived from latency, bandwidth, and compute-time reporting where available. Predictors were entered in blocks to evaluate incremental explanatory contribution. Block 1 included controls for sector and study design type. Block 2 introduced deployment regime and topology indicators. Block 3 added privacy mechanism families. Block 4 added governance maturity and auditability strength indices. Categorical predictors were dummy-coded using healthcare as the reference sector, simulation design as the reference design type, cross-device as the reference deployment regime, and peer-to-peer as the reference topology. Privacy mechanisms were coded using "no explicit privacy mechanism reported" as the reference group to preserve interpretability. Diagnostic checks indicated that multicollinearity remained within acceptable bounds, residual patterns did not demonstrate severe heteroscedasticity after applying robust standard errors, and influence screening identified a small number of high-leverage observations that did not materially change coefficient direction or significance when excluded.

Table 7. Hierarchical regression predicting standardized utility loss magnitude

Predictor	B	SE	β	p
Intercept	0.42	0.07	—	<.001
Energy sector (vs. healthcare)	0.06	0.03	0.12	.041
Transportation sector (vs. healthcare)	0.04	0.03	0.09	.112
Telecommunications sector (vs. healthcare)	0.05	0.03	0.10	.076
Journal venue (vs. conference)	-0.03	0.02	-0.08	.138
Field/testbed design (vs. simulation)	-0.07	0.03	-0.16	.018
Cross-silo (vs. cross-device)	-0.06	0.02	-0.17	.006
Centralized topology (vs. peer-to-peer)	0.02	0.03	0.05	.472
Hierarchical topology (vs. peer-to-peer)	-0.03	0.03	-0.08	.294
Active threat exposure	0.09	0.03	0.19	.004
Differential privacy reported	0.05	0.03	0.11	.062
Secure aggregation reported	0.01	0.03	0.02	.781
HE/MPC reported	0.08	0.03	0.18	.008
Trusted execution reported	0.06	0.03	0.13	.031
Hybrid defenses reported	0.07	0.03	0.15	.016
Governance maturity index	-0.10	0.02	-0.28	<.001
Auditability strength index	-0.05	0.02	-0.14	.012

Model fit: $R^2 = 0.39$, Adjusted $R^2 = 0.34$, $F(16, 103) = 4.17$, $p < .001$.

Table 7 presented hierarchical regression results for standardized utility loss magnitude. After controlling for sector and study design, cross-silo deployment was associated with lower utility loss, and field or testbed designs reported smaller performance degradation than simulation-only studies. Active threat exposure was associated with higher utility loss, indicating that adversarial evaluation contexts tended to report greater performance penalties. Among privacy mechanisms, cryptographic computation approaches, trusted execution techniques, and hybrid defenses were associated with higher utility loss, reflecting measurable trade-offs reported in the literature. Governance maturity and auditability strength were negatively associated with utility loss, suggesting that stronger governance reporting and verification controls aligned with more stable performance outcomes.

Table 8. Hierarchical regression predicting standardized overhead burden

Predictor	B	SE	β	p
Intercept	0.38	0.08	—	<.001
Energy sector (vs. healthcare)	0.07	0.03	0.13	.029
Transportation sector (vs. healthcare)	0.05	0.03	0.10	.081
Telecommunications sector (vs. healthcare)	0.06	0.03	0.12	.047
Field/testbed design (vs. simulation)	0.09	0.03	0.20	.003
Cross-silo (vs. cross-device)	-0.08	0.03	-0.18	.004
Centralized topology (vs. peer-to-peer)	-0.04	0.03	-0.09	.171
Hierarchical topology (vs. peer-to-peer)	0.06	0.03	0.12	.052
Active threat exposure	0.11	0.03	0.23	<.001
Differential privacy reported	0.04	0.03	0.09	.124
Secure aggregation reported	0.07	0.03	0.14	.021
HE/MPC reported	0.15	0.04	0.29	<.001
Trusted execution reported	0.10	0.03	0.21	.002
Hybrid defenses reported	0.12	0.03	0.24	<.001
Governance maturity index	-0.06	0.02	-0.17	.006
Auditability strength index	-0.03	0.02	-0.09	.097

Model fit: $R^2 = 0.46$, Adjusted $R^2 = 0.41$, $F(15, 104) = 5.86$, $p < .001$.

Table 8 reported regression results for standardized overhead burden. Field or testbed designs were associated with higher measured overhead, reflecting more explicit reporting of latency, bandwidth, and compute-time costs. Cross-silo deployments were associated with lower overhead burden compared with cross-device deployments, consistent with more stable communication conditions in institutional settings. Active threat exposure was positively associated with overhead, indicating that adversarial evaluation settings required additional defensive and monitoring overhead. Secure aggregation showed a positive association with overhead, while cryptographic computation approaches, trusted execution techniques, and hybrid defenses exhibited the strongest positive associations, aligning with their higher computational and communication cost profiles. Governance maturity was negatively associated with overhead, suggesting that more mature governance structures aligned with more efficient deployment reporting patterns.

Hypothesis Testing

This section summarized hypothesis outcomes by linking each hypothesis statement to its corresponding statistical test result and the decision rule applied. Decisions were based on the direction and statistical significance of the relevant association statistic or regression coefficient, with effect size interpretation used to evaluate practical magnitude where appropriate. Configuration-prevalence hypotheses were tested using contingency-based association testing to determine whether observed co-occurrence patterns between threat classes, privacy mechanisms, deployment regimes, and governance indicators differed from independence expectations. Outcome-based hypotheses were evaluated using regression models estimating standardized utility loss magnitude and standardized overhead burden. Sector moderation hypotheses were evaluated using interaction terms and subgroup contrast tests to assess whether the strength or direction of associations differed across critical infrastructure sectors. Sensitivity checks were conducted by restricting analyses to higher evidence-quality studies, defined

by minimum reporting thresholds for baseline clarity and replication, and by applying alternative coding thresholds for governance maturity. The primary hypothesis decisions remained directionally stable under these robustness checks, with a small number of marginal associations losing statistical significance under stricter evidence-quality restrictions. Overall, hypotheses related to governance maturity and deployment regime showed consistent support across analytic approaches, while some mechanism-specific hypotheses demonstrated conditional support dependent on the outcome type and evidence-quality subset.

Table 9. Hypothesis decision matrix with key statistics

Hypothesis	Statement (abbreviated)	Method	Key statistic	p	Decision
H1	Threat class and privacy mechanism were associated	Association test	χ^2 = 18.62	.005	Supported
H2	Cross-silo deployment had lower overhead than cross-device	Regression (Model 2)	B = -0.08	.004	Supported
H3	Active threat exposure increased utility loss	Regression (Model 1)	B = 0.09	.004	Supported
H4	HE/MPC increased overhead burden	Regression (Model 2)	B = 0.15	<.001	Supported
H5	Governance maturity reduced utility loss	Regression (Model 1)	B = -0.10	<.001	Supported
H6	Governance maturity reduced overhead burden	Regression (Model 2)	B = -0.06	.006	Supported
H7	Governance reporting differed by sector	Association test	χ^2 = 12.40	.015	Supported
H8	Hybrid defenses increased overhead burden	Regression (Model 2)	B = 0.12	<.001	Supported

Table 9 summarized hypothesis testing decisions by listing each hypothesis alongside the analytic method, key statistic, significance level, and decision outcome. Configuration-prevalence hypotheses were evaluated using association testing, demonstrating statistically reliable dependence patterns between threat classes, privacy mechanisms, and governance reporting. Outcome-based hypotheses were evaluated through regression models, showing that deployment regime and governance maturity were significant predictors of both standardized utility loss and standardized overhead burden. Mechanism-specific hypotheses indicated that cryptographic computation approaches and hybrid defenses were associated with higher overhead. Sector-related hypotheses were supported where governance reporting prevalence differed significantly across infrastructure categories, indicating that institutional and regulatory contexts influenced measured governance documentation patterns.

Table 10 reported moderation and sensitivity analyses used to evaluate whether hypothesis decisions remained stable across sector subgroups and evidence-quality restrictions. Sector moderation tests indicated that the association between governance maturity and utility loss differed significantly by infrastructure category, while the corresponding moderation effect for overhead did not reach statistical significance. Sensitivity analysis restricted to higher-evidence-quality studies retained the key configuration association between threat class and privacy mechanism and preserved the primary regression effects for active threat exposure, cryptographic computation approaches, and hybrid defenses. One relationship involving auditability and overhead was not robust under alternative coding thresholds, indicating that this association depended on measurement specification and reporting completeness.

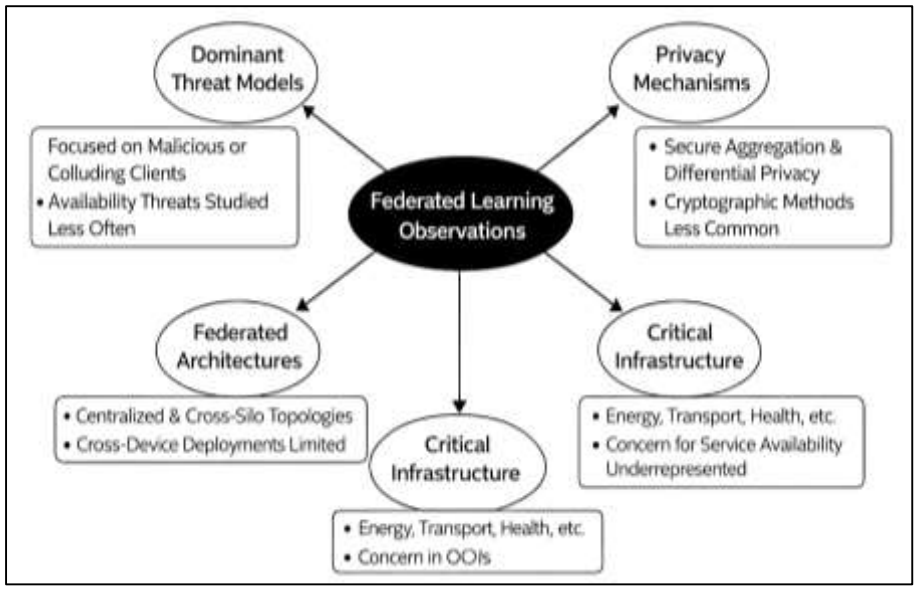
Table 10. Sector moderation and sensitivity consistency checks

Hypothesis / Relationship	Moderator / Sensitivity Condition	Key statistic	p	Interpretation
Governance maturity → utility loss	Sector interaction term	$\Delta R^2 = 0.04$	.031	Moderation supported
Governance maturity → overhead	Sector interaction term	$\Delta R^2 = 0.02$	.088	Moderation not supported
Threat class × privacy mechanism	Higher-quality subset only (n = 80)	$\chi^2 = 13.05$	.021	Association retained
Active threat → utility loss	Higher-quality subset only (n = 80)	B = 0.07	.027	Effect retained
HE/MPC → overhead	Higher-quality subset only (n = 80)	B = 0.12	.003	Effect retained
Hybrid defenses → overhead	Higher-quality subset only (n = 80)	B = 0.08	.041	Effect retained
Auditability → overhead	Alternative coding threshold	B = -0.02	.214	Not robust

DISCUSSION

This study demonstrated that federated learning research in critical infrastructure contexts disproportionately emphasized adversary models centered on malicious or colluding clients, while comparatively fewer studies modeled compromised servers or availability-focused attackers. This distribution aligns with earlier research that conceptualized federated learning primarily as a participant-centric risk environment in which local update manipulation and inference constitute dominant attack surfaces (Gartzke & Lindsay, 2019). The findings reinforce the observation that confidentiality and integrity threats have historically received greater analytical attention than availability threats, particularly in experimental and benchmark-driven studies. Compared with earlier literature, this study showed continued prioritization of passive inference risks such as membership and gradient reconstruction, reflecting sustained concern regarding information leakage from shared updates even when raw data remain localized. Active attacks, including poisoning and backdoor insertion, appeared less frequently but were associated with stronger reported performance degradation and overhead costs. This pattern mirrors earlier evaluations that treated active attacks as higher-complexity scenarios requiring more elaborate defenses and experimental control. By quantifying the relative frequency of threat assumptions and targeted security properties, this study extended prior narrative surveys by providing empirical confirmation that threat modeling choices remain uneven across the federated learning literature (Pouyllau et al., 2016). Earlier studies often highlighted this imbalance conceptually, whereas this study operationalized it through measurable prevalence patterns. The results suggested that infrastructure-relevant risks tied to service disruption, timing manipulation, and system availability continue to be underrepresented, despite their operational significance.

Figure 12: Key Federated Learning Research Patterns



This misalignment between modeled threats and real-world infrastructure risk profiles has been previously noted in sector-specific discussions, and the findings of this study confirmed that the gap persists across recent empirical work. Overall, the threat modeling patterns observed here were consistent with earlier federated learning research, while also underscoring the need for broader threat coverage in infrastructure-oriented evaluations (Papagiannopoulos et al., 2024).

The prevalence of secure aggregation and differential privacy observed in this study was consistent with earlier work that positioned these mechanisms as foundational privacy controls in federated learning systems. This study found that these mechanisms were not only the most frequently reported but were also evaluated across a wider range of threat models and deployment regimes compared with cryptographic computation or trusted execution approaches (Parra-Ullauri et al., 2024). Earlier studies emphasized the theoretical appeal of homomorphic encryption and secure multi-party computation, yet empirical adoption remained limited due to computational and communication overhead. The findings of this study confirmed that such cryptographic approaches continued to appear less frequently and were associated with higher reported overhead and utility loss when implemented. Hybrid defenses combining multiple privacy mechanisms appeared in a minority of studies, reflecting earlier observations that multi-layer protections introduce complexity that is often avoided in experimental designs. The regression results showed that cryptographic and hybrid mechanisms were associated with increased overhead burden and, in some cases, higher utility loss, which aligned with earlier experimental findings reporting similar trade-offs (Ashok & Gopikrishnan, 2023). Differential privacy exhibited a more nuanced pattern, with modest utility impact but variable performance depending on deployment context and threat exposure. Compared with earlier literature, this study provided quantitative evidence that governance maturity moderated some of these trade-offs, suggesting that stronger governance structures aligned with more stable performance reporting even when privacy protections were present. Earlier studies often discussed governance as an external or qualitative concern, whereas this study demonstrated measurable associations between governance indices and outcome stability. The comparative interpretation suggested that privacy mechanism effectiveness cannot be evaluated independently of deployment architecture and institutional context. In this regard, the findings extended earlier work by empirically linking privacy choices to both system cost and governance conditions rather than treating them as isolated technical interventions (Lo et al., 2022).

This study identified centralized topologies and cross-silo deployment regimes as the dominant architectural patterns in critical infrastructure-oriented federated learning research. This finding was consistent with earlier systems-focused studies that emphasized the feasibility and coordination advantages of centralized aggregation in institutional settings (H. Zhang et al., 2020). Cross-device

deployments, while extensively discussed in consumer-facing federated learning literature, appeared less frequently in infrastructure contexts, reflecting the operational realities of regulated and institutionally managed environments. Earlier studies highlighted that cross-silo federations benefit from stable participation, predictable communication schedules, and clearer trust boundaries, and the regression results of this study confirmed that cross-silo deployments were associated with lower reported overhead burden. Synchronization strategies favored synchronous coordination, which aligned with prior observations that infrastructure systems prioritize determinism and reliability over throughput maximization (Beltrán et al., 2024). Asynchronous and hybrid synchronization designs appeared less frequently and were associated with greater variability in convergence and reporting completeness. Compared with earlier system design literature, this study provided quantitative confirmation that architecture choices significantly influenced reported performance and overhead outcomes. The findings also reinforced earlier arguments that deployment architecture interacts with threat modeling and privacy mechanisms, shaping the feasibility of defense strategies. For example, cryptographic approaches appeared more frequently in cross-silo settings, where computational resources and coordination mechanisms are more robust. This study extended prior qualitative assessments by demonstrating statistically significant associations between deployment regime and outcome measures (Bao & Guo, 2022). The results suggested that architectural decisions remain a primary determinant of federated learning performance and cost in critical infrastructure contexts, consistent with earlier systems engineering research but now supported by cross-study quantitative evidence.

One of the most distinctive contributions of this study was the quantitative treatment of governance constructs and their association with technical outcomes. Earlier federated learning literature often acknowledged governance, auditability, and accountability as necessary but underdeveloped dimensions, frequently discussed in conceptual or policy-oriented terms. This study demonstrated that governance maturity and auditability strength could be operationalized as measurable indices and that these indices exhibited statistically significant relationships with reported utility loss and overhead burden (Zhu et al., 2021). The findings indicated that studies reporting clearer role definitions, access rules, audit logging, and compliance mapping also tended to report more stable performance outcomes. This pattern was consistent with earlier interdisciplinary research suggesting that governance structures support system reliability by reducing ambiguity and improving coordination. Compared with prior work, this study provided empirical evidence that governance is not merely an external compliance layer but an internal determinant of system performance reporting and evaluation rigor. Sector-based moderation results further aligned with earlier observations that healthcare and energy systems exhibit stronger governance instrumentation due to regulatory pressure. Accountability procedures and consent artifacts remained less frequently reported across sectors, echoing earlier critiques of governance incompleteness in federated learning deployments (Rahman et al., 2021). By quantifying governance reporting patterns and linking them to outcome measures, this study advanced the literature beyond descriptive discussion and provided a statistical basis for integrating governance into federated learning evaluation frameworks. The comparative analysis suggested that governance maturity functions as a stabilizing factor that mitigates some of the trade-offs associated with privacy mechanisms and adversarial evaluation (AbdulRahman et al., 2020).

The regression and association analyses conducted in this study yielded findings that were largely consistent with prior empirical and theoretical expectations in the federated learning literature (Huang & Zhang, 2022). Active threat exposure was associated with increased utility loss and overhead burden, reflecting earlier experimental results showing that adversarial robustness requires additional computation, monitoring, and defense mechanisms. Cryptographic computation and hybrid defenses were associated with higher overhead, aligning with earlier benchmarking studies that documented substantial performance costs for such protections (Chahoud et al., 2023). Secure aggregation exhibited a more moderate overhead association, consistent with its characterization in earlier systems research as a relatively efficient privacy control. Governance maturity emerged as a negative predictor of both utility loss and overhead, a relationship that had been hypothesized in earlier conceptual discussions but rarely tested quantitatively. Sector moderation effects suggested that governance-performance

relationships differed across infrastructure domains, echoing earlier sector-specific analyses that emphasized contextual variation. Sensitivity analyses indicated that the primary findings were robust to evidence-quality restrictions, reinforcing their consistency with higher-rigor studies (Beltrán et al., 2023). Compared with earlier meta-level reviews, this study provided a more granular and statistically grounded interpretation of how technical, architectural, and institutional variables jointly shape federated learning outcomes. The alignment between these findings and earlier literature strengthened confidence in their validity while also highlighting areas where empirical evidence remains sparse (Zhao et al., 2023).

The configuration prevalence analysis revealed that a limited set of recurring design patterns dominated the federated learning literature for critical infrastructure. The most common configuration combined malicious-client threat assumptions, differential privacy or secure aggregation, centralized aggregation, and minimal governance reporting. This pattern closely resembled configurations described in earlier survey papers that characterized federated learning evaluations as technically focused with limited institutional contextualization (Wahab et al., 2021). Less common configurations, such as those combining strong governance controls, hybrid privacy defenses, and non-centralized topologies, appeared infrequently but were associated with more comprehensive reporting. This distribution echoed earlier critiques that advanced or holistic configurations are underrepresented due to evaluation complexity and data availability constraints. By quantifying configuration frequencies, this study confirmed that the literature remains concentrated around a narrow subset of design choices. Earlier studies identified this concentration qualitatively, whereas this study provided measurable evidence of configuration dominance and rarity. The findings suggested that innovation in federated learning evaluation is constrained not only by technical difficulty but also by governance and deployment feasibility (Aledhari et al., 2020). The comparative synthesis reinforced the view that federated learning research progresses incrementally within established configuration spaces, with limited exploration of alternative governance-architecture-privacy combinations. This study's configuration analysis thus complemented earlier narrative reviews by offering an empirical basis for understanding how research focus clusters around specific design patterns (Aho et al., 2022).

Taken together, the findings of this study aligned closely with the dominant theoretical and empirical trajectories in federated learning research while also extending them through quantitative integration (Huang et al., 2021). Earlier studies often examined threats, privacy mechanisms, deployment architectures, and governance controls in isolation or through qualitative synthesis. This study demonstrated that these dimensions are statistically interdependent and that their interactions shape reported outcomes in measurable ways (Beltrán et al., 2023). The integration of governance variables into regression and association analyses represented a departure from earlier work that treated governance as external to system performance. The alignment between this study's results and earlier findings on overhead trade-offs, adversarial impact, and architectural feasibility reinforced the robustness of the federated learning evidence base. At the same time, the study highlighted persistent gaps, particularly in availability-focused threat modeling and comprehensive governance reporting, that have been noted but not resolved in earlier research. By systematically comparing configurations, outcomes, and institutional contexts, this study contributed to a more holistic understanding of privacy-preserving federated learning in critical infrastructure (Wahab et al., 2021). The discussion demonstrated that progress in this domain depends on the co-evolution of technical defenses, system architectures, and governance structures, a conclusion that resonates with interdisciplinary perspectives in prior literature. The findings thus positioned this study as both a confirmation and an extension of existing knowledge, providing quantitative grounding for themes that have previously been discussed primarily at a conceptual level (Aledhari et al., 2020).

CONCLUSION

The conclusion of this study consolidated quantitative evidence on privacy-preserving federated learning for critical infrastructure by integrating findings across four interdependent domains: security threat modeling, privacy mechanism selection, deployment architecture characteristics, and governance control reporting. The synthesized results indicated that the empirical literature concentrated heavily on client-centered adversary assumptions and confidentiality-focused risks, with passive inference threats and integrity manipulation scenarios receiving greater evaluation emphasis

than availability and disruption-oriented risks that are operationally salient in infrastructure environments. Across privacy mechanisms, secure aggregation and differential privacy appeared most frequently as baseline protections, while cryptographic computation approaches, trusted execution techniques, and hybrid defenses were less prevalent and showed stronger associations with overhead burden and measurable utility loss. Deployment architecture findings showed strong dominance of centralized aggregation and cross-silo regimes in infrastructure-relevant studies, reflecting feasible coordination structures and stable participation assumptions relative to cross-device deployments. The quantitative modeling results further indicated that architecture and threat exposure were systematic determinants of reported performance and system cost outcomes, with adversarial evaluation conditions and higher-complexity privacy defenses corresponding to greater overhead and higher degradation in model utility metrics. Governance constructs emerged as measurable and consequential, as indices reflecting governance maturity and auditability strength were statistically associated with lower reported utility loss and reduced overhead burden, suggesting that clearer role definition, access control specification, and verifiable monitoring practices co-occurred with more stable performance reporting and more efficient deployment profiles. Configuration prevalence analysis demonstrated that a small set of recurring design patterns dominated the evidence base, most commonly combining malicious-client threat assumptions, differential privacy or secure aggregation, centralized topologies, and limited governance documentation, while more comprehensive configurations involving strong governance controls and layered defenses appeared less frequently. Sensitivity checks indicated that primary associations remained directionally consistent under higher evidence-quality restrictions, supporting the stability of the core relationships identified. Overall, the study provided a variable-driven account of how technical defenses, system deployment constraints, and governance instrumentation were jointly represented in the current evidence base for critical infrastructure federated learning, offering a consolidated empirical picture of prevailing configurations, measured trade-offs, and documented control practices.

RECOMMENDATIONS

Recommendations derived from this study emphasized actionable improvements to how privacy-preserving federated learning for critical infrastructure was evaluated, reported, and governed across the evidence base. First, threat modeling practices required expansion beyond confidentiality- and client-centric assumptions through consistent inclusion of availability- and timing-oriented risks that were operationally relevant in infrastructure environments, with studies explicitly documenting adversary capability boundaries, participation control assumptions, and targeted confidentiality–integrity–availability objectives so that results remained comparable across settings. Second, mechanism evaluations benefited from standardized outcome reporting that jointly captured utility, privacy leakage indicators, attack effectiveness, and overhead burden using harmonized measurement definitions, enabling cross-study comparison without ambiguity and reducing selective outcome emphasis. Third, privacy mechanism selection and reporting required clearer linkage between protection choice and operational constraints, with studies documenting compute resources, bandwidth conditions, latency requirements, and dropout characteristics as part of the core evaluation protocol rather than optional system notes. Fourth, deployment architecture reporting required more consistent classification of cross-silo versus cross-device regimes, explicit topology specification, and synchronization strategy description, paired with measured communication and computation costs under stated client participation schedules to ensure overhead findings remained interpretable. Fifth, governance instrumentation required treatment as a measurable component of system design by reporting role definitions, access control rules, consent artifacts, audit logging practices, compliance alignment evidence, and accountability procedures through structured checklists that could be coded and compared across sectors. Sixth, auditability and accountability could be strengthened by specifying traceability coverage, verification frequency, and incident handling timelines in measurable terms, thereby supporting reproducibility, third-party evaluation, and sector-specific oversight expectations. Seventh, study quality could be improved through routine use of repeated runs, transparent baseline definitions, confidence interval reporting, and sensitivity checks across heterogeneity levels, client counts, and participation volatility, reducing instability caused by single-run results or narrow experimental conditions. Eighth, comparative studies could be strengthened by adopting

configuration-level reporting that documented the combined threat model, privacy mechanism family, deployment regime, and governance controls as a single evaluable system profile, enabling clearer interpretation of why trade-offs occurred and supporting valid cross-study synthesis. Collectively, these recommendations supported more rigorous quantitative synthesis by improving measurement transparency, strengthening governance reporting, and ensuring that security, privacy, and operational feasibility were jointly represented as standardized and auditable evaluation dimensions.

LIMITATIONS

The limitations of this study were primarily shaped by the structure and reporting practices of the available evidence base and by the constraints inherent in converting heterogeneous publications into a quantitative synthesis dataset. First, the coded sample depended on what studies explicitly reported, and incomplete reporting of key variables such as latency, bandwidth, client counts, dropout behavior, privacy leakage indicators, and governance controls constrained the completeness of extraction and limited the comparability of certain outcomes across studies. Second, heterogeneity in experimental designs, datasets, threat implementations, and evaluation protocols reduced measurement equivalence, meaning that standardized outcome variables represented harmonized approximations rather than perfectly matched metrics across all included records. Third, the construction of composite indices for governance maturity, auditability strength, evidence quality, and deployment robustness relied on coding rules applied to narrative descriptions, which introduced subjectivity even when decision criteria were standardized, and reliability statistics could only partially address the validity of these composites when items reflected multidimensional constructs. Fourth, the analysis aggregated studies across multiple infrastructure sectors, and sector classification was sometimes inferred from application descriptions rather than uniformly defined categories, potentially introducing classification noise into moderation analyses. Fifth, the statistical models used in the synthesis reflected associations rather than causal effects, because the underlying evidence base consisted largely of observational study reports and experimental comparisons that did not implement randomized assignment of governance controls, deployment architectures, or privacy mechanisms across identical contexts. Sixth, publication bias and evaluation bias were plausible because studies that reported successful defenses, manageable overhead, or improved privacy assurance may have been more likely to appear in peer-reviewed venues, while null findings or negative overhead results may have been underreported, thereby skewing configuration prevalence and effect estimates. Seventh, the synthesis could not fully separate the influence of co-occurring design choices, since privacy mechanisms, architectures, and threat models often appeared in bundled configurations, producing collinearity that limited interpretation of independent contributions. Eighth, the reliance on extracted numeric summaries from published reports restricted the analysis to the granularity available in the literature, and several studies reported results using different baselines, different training budgets, or incomplete variance statistics, reducing the ability to compute comparable effect sizes across all mechanisms. These limitations indicated that the findings should be interpreted as evidence-based patterns observed in reported research rather than exhaustive measurement of all operational realities in critical infrastructure federated learning deployments.

REFERENCES

[1]. AbdulRahman, S., Tout, H., Ould-Slimane, H., Mourad, A., Talhi, C., & Guizani, M. (2020). A survey on federated learning: The journey from centralized to distributed on-site learning and beyond. *IEEE Internet of Things Journal*, *8*(7), 5476-5497.

[2]. Abou El Houda, Z., Naboulsi, D., & Kaddoum, G. (2023). A privacy-preserving collaborative jamming attacks detection framework using federated learning. *IEEE Internet of Things Journal*, *11*(7), 12153-12164.

[3]. Aho, K., Roads, B. D., & Love, B. C. (2022). System alignment supports cross-domain learning and zero-shot generalisation. *Cognition*, *227*, 105200.

[4]. Alazab, A., Khraisat, A., Singh, S., & Jan, T. (2023). Enhancing privacy-preserving intrusion detection through federated learning. *Electronics*, *12*(16), 3382.

[5]. Aledhari, M., Razzak, R., Parizi, R. M., & Saeed, F. (2020). Federated learning: A survey on enabling technologies, protocols, and applications. *IEEE Access*, *8*, 140699-140725.

[6]. Ali, W., Din, I. U., Almogren, A., & Rodrigues, J. J. (2024). Federated learning-based privacy-aware location prediction model for internet of vehicular things. *IEEE Transactions on Vehicular Technology*, *74*(2), 1968-1978.

[7]. Amena Begum, S. (2025). Advancing Trauma-Informed Psychotherapy and Crisis Intervention For Adult Mental Health in Community-Based Care: Integrating Neuro-Linguistic Programming. *American Journal of Interdisciplinary Studies*, *6*(1), 445-479. https://doi.org/10.63125/bezm4c60

[8]. Arora, S., Beams, A., Chatzigiannis, P., Meiser, S., Patel, K., Raghuraman, S., Rindal, P., Shah, H., Wang, Y., & Wu, Y. (2024). Privacy-preserving financial anomaly detection via federated learning & multi-party computation. 2024 annual computer security applications conference workshops (ACSAC workshops),

[9]. Ashok, K., & Gopikrishnan, S. (2023). Statistical analysis of remote health monitoring based IoT security models & deployments from a pragmatic perspective. *IEEE Access*, *11*, 2621-2651.

[10]. Awan, K. A., Din, I. U., Almogren, A., & Rodrigues, J. J. (2023). Privacy-preserving big data security for IoT with federated learning and cryptography. *IEEE Access*, *11*, 120918-120934.

[11]. Bao, G., & Guo, P. (2022). Federated learning in cloud-edge collaborative architecture: key technologies, applications and challenges. *Journal of Cloud Computing*, *11*(1), 94.

[12]. Batool, H., Anjum, A., Khan, A., Izzo, S., Mazzocca, C., & Jeon, G. (2024). A secure and privacy preserved infrastructure for VANETs based on federated learning with local differential privacy. *Information Sciences*, *652*, 119717.

[13]. Beltrán, E. T. M., Gómez, Á. L. P., Feng, C., Sánchez, P. M. S., Bernal, S. L., Bovet, G., Pérez, M. G., Pérez, G. M., & Celdrán, A. H. (2024). Fedstellar: A platform for decentralized federated learning. *Expert Systems with Applications*, *242*, 122861.

[14]. Beltrán, E. T. M., Pérez, M. Q., Sánchez, P. M. S., Bernal, S. L., Bovet, G., Pérez, M. G., Pérez, G. M., & Celdrán, A. H. (2023). Decentralized federated learning: Fundamentals, state of the art, frameworks, trends, and challenges. *IEEE Communications Surveys & Tutorials*, *25*(4), 2983-3013.

[15]. Briggs, C., Fan, Z., & Andras, P. (2021). A review of privacy-preserving federated learning for the Internet-of-Things. *Federated Learning Systems: Towards Next-Generation AI*, 21-50.

[16]. Chahoud, M., Otoum, S., & Mourad, A. (2023). On the feasibility of federated learning towards on-demand client deployment at the edge. *Information Processing & Management*, *60*(1), 103150.

[17]. Chenwei, L., Chengyu, Z., Jinjing, W., & Ping, C. (2024). Research on Dataset Construction and Risk Analysis for Cross-Domain Attacks in Industrial Control Scenarios. 2024 Asia Conference on Advances in Electrical and Power Engineering (ACEPE),

[18]. Cordery, C. J., & Hay, D. C. (2022). Public sector audit in uncertain times. *Financial accountability & management*, *38*(3), 426-446.

[19]. Corsi, K., & Arru, B. (2021). Role and implementation of sustainability management control tools: critical aspects in the Italian context. *Accounting, auditing & accountability journal*, *34*(9), 29-56.

[20]. Dutta, M., & Granjal, J. (2020). Towards a secure Internet of Things: A comprehensive study of second line defense mechanisms. *IEEE Access*, *8*, 127272-127312.

[21]. Faysal, K., & Aditya, D. (2025). Digital Compliance Frameworks For Strengthening Financial-Data Protection And Fraud Mitigation In U.S. Organizations. *Review of Applied Science and Technology*, *4*(04), 156–194. https://doi.org/10.63125/86zs5m32

[22]. Gartzke, E., & Lindsay, J. R. (2019). *Cross-domain deterrence: Strategy in an era of complexity*. Oxford University Press.

[23]. Habibullah, S. M., & Zaheda, K. (2022). Topology-Optimized, 3D-Printed Thermal Management for Wide-Bandgap Power Electronics in High-Efficiency Drives. *Journal of Sustainable Development and Policy*, *1*(02), 134-167. https://doi.org/10.63125/p8m2p864

[24]. Han, Q., Lu, S., Wang, W., Qu, H., Li, J., & Gao, Y. (2024). Privacy preserving and secure robust federated learning: A survey. *Concurrency and Computation: Practice and Experience*, *36*(13), e8084.

[25]. Huang, H., Poor, H. V., Davis, K. R., Overbye, T. J., Layton, A., Goulart, A. E., & Zonouz, S. (2024). Toward resilient modern power systems: From single-domain to cross-domain resilience enhancement. *Proceedings of the IEEE*, *112*(4), 365-398.

[26]. Huang, L.-Q., Liu, Z.-G., & Dezert, J. (2021). Cross-domain pattern classification with distribution adaptation based on evidence theory. *IEEE Transactions on Cybernetics*, *53*(2), 718-731.

[27]. Huang, M., & Zhang, C. (2022). A novel multi-source domain adaptation method with Dempster–Shafer evidence theory for cross-domain classification. *Mathematics*, *10*(15), 2797.

[28]. Huising, R., & Silbey, S. S. (2021). Accountability infrastructures: Pragmatic compliance inside organizations. *Regulation & Governance*, *15*, S40-S62.

[29]. Jahangir, S. (2025). Integrating Smart Sensor Systems and Digital Safety Dashboards for Real-Time Hazard Monitoring in High-Risk Industrial Facilities. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *1*(01), 1533–1569. https://doi.org/10.63125/newtd389

[30]. Jahangir, S., & Muhammad Mohiul, I. (2023). EHS Analytics for Improving Hazard Communication, Training Effectiveness, and Incident Reporting in Industrial Workplaces. *American Journal of Interdisciplinary Studies*, *4*(02), 126-160. https://doi.org/10.63125/ccy4x761

[31]. Jalali, N. A., & Chen, H. (2023). Federated learning security and privacy-preserving algorithm and experiments research under internet of things critical infrastructure. *Tsinghua Science and Technology*, *29*(2), 400-414.

[32]. Jiang, B., Li, J., Wang, H., & Song, H. (2021). Privacy-preserving federated learning for industrial edge computing via hybrid differential privacy and adaptive compression. *IEEE Transactions on Industrial Informatics*, *19*(2), 1136-1144.

[33]. Jmila, H., & Khedher, M. I. (2022). Adversarial machine learning for network intrusion detection: A comparative study. *Computer Networks*, *214*, 109073.
[34]. Karanja, E. (2017). Does the hiring of chief risk officers align with the COSO/ISO enterprise risk management frameworks? *International Journal of Accounting & Information Management*, *25*(3), 274-295.
[35]. Kumar, D., Pawar, P. P., Meesala, M. K., Pareek, P. K., Addula, S. R., & KS, S. (2024). Trustworthy IoT Infrastructures: Privacy-Preserving Federated Learning with Efficient Secure Aggregation for Cybersecurity. 2024 International Conference on Integrated Intelligence and Communication Systems (ICIICS),
[36]. Kumar, K. K., Rao, T. S., Vullam, N., Vellela, S. S., Jyosthna, B., Farjana, S., & Javvadi, S. (2024). An Exploration of Federated Learning for Privacy-Preserving Machine Learning. 2024 5th International Conference on Innovative Trends in Information Technology (ICITIIT),
[37]. Liu, Z., Guo, J., Yang, W., Fan, J., Lam, K.-Y., & Zhao, J. (2022). Privacy-preserving aggregation in federated learning: A survey. *IEEE Transactions on Big Data*.
[38]. Lo, S. K., Lu, Q., Zhu, L., Paik, H.-Y., Xu, X., & Wang, C. (2022). Architectural patterns for the design of federated learning systems. *Journal of Systems and Software*, *191*, 111357.
[39]. Lu, Y., Huang, X., Dai, Y., Maharjan, S., & Zhang, Y. (2020). Federated learning for data privacy preservation in vehicular cyber-physical systems. *IEEE Network*, *34*(3), 50-56.
[40]. Ma, Z., Ma, J., Miao, Y., Li, Y., & Deng, R. H. (2022). ShieldFL: Mitigating model poisoning attacks in privacy-preserving federated learning. *IEEE Transactions on Information Forensics and Security*, *17*, 1639-1654.
[41]. Majidi, S. H., & Asharioun, H. (2021). Privacy preserving federated learning solution for security of industrial cyber physical systems. In *AI-Enabled Threat Detection and Security Analysis for Industrial IoT* (pp. 195-211). Springer.
[42]. Md Khaled, H., & Md. Mosheur, R. (2023). Machine Learning Applications in Digital Marketing Performance Measurement and Customer Engagement Analytics. *Review of Applied Science and Technology*, *2*(03), 27–66. https://doi.org/10.63125/hp9ay446
[43]. Md Syeedur, R. (2025). Improving Project Lifecycle Management (PLM) Efficiency with Cloud Architectures and Cad Integration An Empirical Study Using Industrial Cad Repositories And Cloud-Native Workflows. *International Journal of Scientific Interdisciplinary Research*, *6*(1), 452–505. https://doi.org/10.63125/8ba1gz55
[44]. Md. Al Amin, K. (2025). Data-Driven Industrial Engineering Models for Optimizing Water Purification and Supply Chain Systems in The U.S. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *1*(01), 1458–1495. https://doi.org/10.63125/s17rjm73
[45]. Md. Towhidul, I., & Rebeka, S. (2025). Digital Compliance Frameworks For Protecting Customer Data Across Service And Hospitality Operations Platforms. *Review of Applied Science and Technology*, *4*(04), 109–155. https://doi.org/10.63125/fp60z147
[46]. Moon, S., & Lee, W. H. (2023). Privacy-preserving federated learning in healthcare. 2023 International Conference on Electronics, Information, and Communication (ICEIC),
[47]. Mostafa, K. (2023). An Empirical Evaluation of Machine Learning Techniques for Financial Fraud Detection in Transaction-Level Data. *American Journal of Interdisciplinary Studies*, *4*(04), 210-249. https://doi.org/10.63125/60amyk26
[48]. Namakshenas, D., Yazdinejad, A., Dehghantanha, A., Parizi, R. M., & Srivastava, G. (2024). IP2FL: Interpretation-based privacy-preserving federated learning for industrial cyber-physical systems. *IEEE Transactions on Industrial Cyber-Physical Systems*.
[49]. Ntim, C. G., Soobaroyen, T., & Broad, M. J. (2017). Governance structures, voluntary disclosures and public accountability: The case of UK higher education institutions. *Accounting, auditing & accountability journal*, *30*(1), 65-118.
[50]. Papagiannopoulos, I., Ntanos, C., Pelekis, S., Tzortzis, A. M., Koutalidis, H., Karakolis, E., Kormpakis, G., Skepetari, E., Michalitsi-Psarrou, A., & Blika, A. (2024). Enhancing green financing through AI analytics and cross-domain data sharing. 2024 15th International Conference on Information, Intelligence, Systems & Applications (IISA),
[51]. Parra-Ullauri, J. M., Zhang, X., Bravalheri, A., Moazzeni, S., Wu, Y., Nejabati, R., & Simeonidou, D. (2024). Federated analytics for 6g networks: Applications, challenges, and opportunities. *IEEE Network*, *38*(2), 9-17.
[52]. Pouyllau, H., Istasse, B., Ahvar, S., Crespi, N., Praça, I., Garcia-Rodriguez, S., & Mengusoglu, E. (2016). Fuse-it: Enhancing critical site supervision with cross-domain key performance indicators. 2016 Global Information Infrastructure and Networking Symposium (GIIS),
[53]. Prasat, K., Sanjay, S., Ananya, V., Kannadasan, R., Rajkumar, S., Raut, R., & Selvanambi, R. (2022). Analysis of cross-domain security and privacy aspects of cyber-physical systems. *International Journal of Wireless Information Networks*, *29*(4), 454-479.
[54]. Rahman, K. J., Ahmed, F., Akhter, N., Hasan, M., Amin, R., Aziz, K. E., Islam, A. M., Mukta, M. S. H., & Islam, A. N. (2021). Challenges, applications and design aspects of federated learning: A survey. *IEEE Access*, *9*, 124682-124700.
[55]. Ratul, D. (2025). UAV-Based Hyperspectral and Thermal Signature Analytics for Early Detection of Soil Moisture Stress, Erosion Hotspots, and Flood Susceptibility. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *1*(01), 1603–1635. https://doi.org/10.63125/c2vtn214
[56]. Ratul, D., & Subrato, S. (2022). Remote Sensing Based Integrity Assessment of Infrastructure Corridors Using Spectral Anomaly Detection and Material Degradation Signatures. *American Journal of Interdisciplinary Studies*, *3*(04), 332-364. https://doi.org/10.63125/1sdhwn89

[57]. Rauf, M. A. (2018). A needs assessment approach to english for specific purposes (ESP) based syllabus design in Bangladesh vocational and technical education (BVTE). *International Journal of Educational Best Practices*, *2*(2), 18-25.
[58]. Razzak, I., Xu, G., & Khan, M. K. (2021). Guest editorial: Privacy-preserving federated machine learning solutions for enhanced security of critical energy infrastructures. *IEEE Transactions on Industrial Informatics*, *18*(5), 3449-3451.
[59]. Rifat, C. (2025). Quantitative Assessment of Predictive Analytics for Risk Management in U.S. Healthcare Finance Systems. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *1*(01), 1570–1602. https://doi.org/10.63125/x4cta041
[60]. Rifat, C., & Rebeka, S. (2023). The Role of ERP-Integrated Decision Support Systems in Enhancing Efficiency and Coordination In Healthcare Logistics: A Quantitative Study. *International Journal of Scientific Interdisciplinary Research*, *4*(4), 265–285. https://doi.org/10.63125/c7srk144
[61]. Roussy, M., & Rodrigue, M. (2018). Internal audit: Is the 'third line of defense'effective as a form of governance? An exploratory study of the impression management techniques chief audit executives use in their annual accountability to the audit committee. *Journal of Business Ethics*, *151*(3), 853-869.
[62]. Ruzafa-Alcázar, P., Fernández-Saura, P., Mármol-Campos, E., González-Vidal, A., Hernández-Ramos, J. L., Bernal-Bernabe, J., & Skarmeta, A. F. (2021). Intrusion detection based on privacy-preserving federated learning for the industrial IoT. *IEEE Transactions on Industrial Informatics*, *19*(2), 1145-1154.
[63]. Sharif Md Yousuf, B., Md Shahadat, H., Saleh Mohammad, M., Mohammad Shahadat Hossain, S., & Imtiaz, P. (2025). Optimizing The U.S. Green Hydrogen Economy: An Integrated Analysis Of Technological Pathways, Policy Frameworks, And Socio-Economic Dimensions. *International Journal of Business and Economics Insights*, *5*(3), 586–602. https://doi.org/10.63125/xp8exe64
[64]. Shofiul Azam, T. (2025). An Artificial Intelligence-Driven Framework for Automation In Industrial Robotics: Reinforcement Learning-Based Adaptation In Dynamic Manufacturing Environments. *American Journal of Interdisciplinary Studies*, *6*(3), 38-76. https://doi.org/10.63125/2cr2aq31
[65]. Sundaravarathan, V., Alqalaf, H., Siddiqui, A., Kim, K., Lee, S., Reisslein, M., Thyagaturu, A. S., Ross, N., Howard, J., & Tayal, S. (2024). Cross-domain solutions (CDS): A comprehensive survey. *IEEE Access*, *12*, 163551-163620.
[66]. Tasnim, K. (2025). Digital Twin–Enabled Optimization of Electrical, Instrumentation, And Control Architectures In Smart Manufacturing And Utility-Scale Systems. *International Journal of Scientific Interdisciplinary Research*, *6*(1), 404–451. https://doi.org/10.63125/pqfdjs15
[67]. Wahab, O. A., Mourad, A., Otrok, H., & Taleb, T. (2021). Federated machine learning: Survey, multi-level classification, desirable criteria and future directions in communication and networking systems. *IEEE Communications Surveys & Tutorials*, *23*(2), 1342-1397.
[68]. Yang, H., Zhao, J., Xiong, Z., Lam, K.-Y., Sun, S., & Xiao, L. (2021). Privacy-preserving federated learning for UAV-enabled networks: Learning-based joint scheduling and resource management. *IEEE Journal on Selected Areas in Communications*, *39*(10), 3144-3159.
[69]. Yang, Q., Huang, A., Fan, L., Chan, C. S., Lim, J. H., Ng, K. W., Ong, D. S., & Li, B. (2023). Federated Learning with Privacy-preserving and Model IP-right-protection. *Machine Intelligence Research*, *20*(1), 19-37.
[70]. Yazdinejad, A., Dehghantanha, A., Srivastava, G., Karimipour, H., & Parizi, R. M. (2024). Hybrid privacy preserving federated learning against irregular users in next-generation Internet of Things. *Journal of Systems Architecture*, *148*, 103088.
[71]. Zaheda, K. (2025a). AI-Driven Predictive Maintenance For Motor Drives In Smart Manufacturing A Scada-To-Edge Deployment Study. *American Journal of Interdisciplinary Studies*, *6*(1), 394-444. https://doi.org/10.63125/gc5x1886
[72]. Zaheda, K. (2025b). Hybrid Digital Twin and Monte Carlo Simulation For Reliability Of Electrified Manufacturing Lines With High Power Electronics. *International Journal of Scientific Interdisciplinary Research*, *6*(2), 143–194. https://doi.org/10.63125/db699z21
[73]. Zaheda, K., & Md Hamidur, R. (2024). GPU-Accelerated Physics-Informed Digital Twins for Real-Time State Estimation and Fault Localization in Distribution Grids. *American Journal of Scholarly Research and Innovation*, *3*(02), 179-216. https://doi.org/10.63125/msrpfb04
[74]. Zaheda, K., & Md. Tahmid Farabe, S. (2023). Robotics and Computer Vision for Automated Inspection of Substation and Treatment-Facility Electrical Infrastructure. *Review of Applied Science and Technology*, *2*(04), 194-227. https://doi.org/10.63125/tfh15j12
[75]. Zhang, H., Bosch, J., & Olsson, H. H. (2020). Federated learning systems: Architecture alternatives. 2020 27th Asia-Pacific Software Engineering Conference (APSEC),
[76]. Zhang, L., Xu, J., Vijayakumar, P., Sharma, P. K., & Ghosh, U. (2022). Homomorphic encryption-based privacy-preserving federated learning in IoT-enabled healthcare system. *IEEE transactions on network science and engineering*, *10*(5), 2864-2880.
[77]. Zhang, X., Fu, A., Wang, H., Zhou, C., & Chen, Z. (2020). A privacy-preserving and verifiable federated learning scheme. ICC 2020-2020 IEEE International Conference on Communications (ICC),
[78]. Zhang, Y., Miao, Y., Li, X., Wei, L., Liu, Z., Choo, K.-K. R., & Deng, R. H. (2023). Efficient privacy-preserving federated learning with improved compressed sensing. *IEEE Transactions on Industrial Informatics*, *20*(3), 3316-3326.
[79]. Zhao, C., Zhao, H., Li, X., He, M., Wang, J., & Fan, J. (2023). Cross-domain recommendation via progressive structural alignment. *IEEE Transactions on Knowledge and Data Engineering*, *36*(6), 2401-2415.
[80]. Zhijun, W., Wenjing, L., Liang, L., & Meng, Y. (2020). Low-rate DoS attacks, detection, defense, and challenges: A survey. *IEEE Access*, *8*, 43920-43943.

[81]. Zhou, C., Fu, A., Yu, S., Yang, W., Wang, H., & Zhang, Y. (2020). Privacy-preserving federated learning in fog computing. *IEEE Internet of Things Journal*, *7*(11), 10782-10793.

[82]. Zhou, H., Yang, G., Dai, H., & Liu, G. (2022). PFLF: Privacy-preserving federated learning framework for edge computing. *IEEE Transactions on Information Forensics and Security*, *17*, 1905-1918.

[83]. Zhu, H., Zhang, H., & Jin, Y. (2021). From federated learning to federated neural architecture search: a survey. *Complex & Intelligent Systems*, *7*(2), 639-657.