Vol. 3, No. 4, pp.: 28-49, 2025

Trajectory Feature Stores for Privacy-Preserving Mobility Intelligence: Schema Design, Quality Control, and Federated Benchmarking

Omar Alshammari[1], Sara Al-Qahtani[2,*], Faisal Almutairi[3]

[1] Department of Information Systems, University of Hail, Hail 55476, Saudi Arabia

[2] Department of Computer Science, Jazan University, Jazan 45142, Saudi Arabia

[3] Department of Management Information Systems, Qassim University, Buraydah 52571, Saudi Arabia

* *sara.alqahtani@jazanu.edu.sa*

Article Information	
Received	13 July 2025
Accepted	19 November 2025
DOI	https://doi.org/10.63646/datamind.2025.030403

Abstract

Urban mobility intelligence increasingly depends on high-frequency GPS traces, app-based location events, and fleet trajectories. Yet the same signals that support traffic forecasting, destination prediction, and transport-service optimization also expose sensitive behavioral routines. This article develops a trajectory feature store framework for privacy-preserving mobility intelligence. The framework converts raw trajectory events into governed, versioned, and quality-controlled feature views that can be used in federated learning without centralizing raw location histories. It specifies a schema design for spatial, temporal, behavioral, and privacy metadata; a quality-control layer for completeness, spatial validity, temporal continuity, label reliability, and client balance; and a federated benchmarking layer for comparing model utility, communication overhead, and residual privacy risk under controlled feature snapshots. A synthetic analytical evaluation inspired by GeoLife and T-Drive-style mobility settings shows that feature-store controls improve average feature quality from 0.77 to 0.90, reduce training-serving inconsistency, and support more stable privacy-utility trade-offs under differential privacy and secure aggregation. The study contributes a data-governance-centered perspective on federated trajectory mining by shifting attention from model architecture alone to the reproducibility, auditability, and operational readiness of the mobility feature layer.

Keywords: ***trajectory feature store; federated learning; mobility intelligence; data governance; differential privacy; quality control***

1. Introduction

Urban mobility services now generate a dense stream of digital traces through smartphones, ride-hailing platforms, delivery fleets, shared bicycles, connected vehicles, and public transport ticketing systems. These traces make it possible to infer demand peaks, identify abnormal route behavior, estimate arrival times, and support resource allocation across urban transportation systems. At the same time, trajectories are among the most sensitive forms of behavioral data because they reveal home locations, work routines, social visits, religious attendance, medical visits, and other intimate patterns that are not always visible in conventional transactional data. The central challenge for mobility intelligence is therefore not merely how to extract accurate features from movement records, but how to build a repeatable data infrastructure that transforms raw trajectory events into useful features while preserving privacy, governance, and operational control. This framing is consistent with the FAIR principle that reusable data assets require explicit stewardship and machine-actionable documentation (Wilkinson et al., 2016). Large-scale mobility studies show that location traces often contain stable behavioral signatures rather than random observations (González et al., 2008). The high predictability of human movement further explains why feature governance is necessary before trajectory data are reused for modeling (Song et al., 2010).

The uploaded PDF manuscript that motivates this study proposes a privacy-preserving federated learning framework for mobility data mining. Its core design keeps raw GPS traces on local devices, shares model updates rather than raw data, and combines differential privacy, secure aggregation, adaptive gradient compression, and personalization to address privacy leakage, non-independent data distributions, and communication cost. The manuscript also evaluates its approach on GeoLife and T-Drive-style trajectory data and reports that privacy-preserving federated learning can remain competitive with centralized methods when privacy mechanisms and communication controls are carefully integrated. Those contributions provide a useful model-centric foundation, but they leave open a broader data engineering question: what kind of feature infrastructure is needed before federated mobility models can be benchmarked, reproduced, audited, and maintained in real operational settings? Federated learning research emphasizes that decentralization reduces raw-data sharing but does not remove system-level governance challenges (Li et al., 2020). Distributed IoT security research similarly shows that edge data infrastructures require controls for identity, integrity, and privacy (Xu et al., 2021). Recent privacy-performance reviews in federated learning frame the core problem as a joint optimization of accuracy, protection, and operational cost (Mohammadi et al., 2024).

Most privacy-preserving mobility analytics studies focus on algorithms. They compare centralized learning, federated averaging, differential privacy, secure aggregation, and personalized optimization under different assumptions about client participation and data distribution. Algorithmic progress is essential, but production mobility systems also require stable feature definitions, point-in-time correctness, semantic metadata, feature versioning, lineage tracking, quality validation, privacy budget recording, and benchmark reproducibility. Without these functions, the same model can produce different results across data refreshes, feature transformations may leak future information, and reported privacy-utility trade-offs may depend on hidden preprocessing decisions rather than on the federated learning method itself. Security surveys show that federated learning remains exposed to inference, poisoning, and protocol-level threats when governance evidence is incomplete (Mothukuri et al., 2021). Recent federated learning reviews identify client heterogeneity, partial participation, and evaluation inconsistency as persistent open problems (Liu et al., 2024). Decentralized federated learning research further argues that trust and verifiability must be built into the benchmark itself (Hallaji et al., 2024).

Feature stores have emerged in machine learning operations as infrastructure for registering, validating, sharing, and serving features across training and inference pipelines. However, conventional feature-store designs were developed mainly for tabular business data, recommendation systems, fraud detection, and enterprise prediction

tasks. Urban trajectory data introduces additional complexity: observations are sequential and irregular, spatial coordinates must be validated against road networks or service areas, sampling frequencies differ across clients, stops and trips are inferred rather than directly observed, and privacy risk depends on combinations of location, time, identity, and behavioral repetition. A trajectory feature store therefore cannot be a simple table repository. It must encode mobility semantics, privacy constraints, and federated learning requirements as first-class data-management objects. Production-scale ML platforms show that reusable feature definitions and validation rules are essential for stable inference pipelines (Baylor et al., 2017). Recent feature-store systems formalize the role of managed feature registries in connecting training and serving environments (de la Rúa Martínez et al., 2024). MLOps architecture studies also place data and feature lifecycle management at the center of reliable AI deployment (Kreuzberger et al., 2023).

This article develops a framework for trajectory feature stores that support privacy-preserving mobility intelligence. The article asks three research questions. First, what schema elements are required to represent trajectory features in a way that supports federated training while avoiding raw trace centralization? Second, how should quality-control rules be designed so that mobility features remain reliable across heterogeneous clients and datasets? Third, how can a feature store provide benchmark snapshots for comparing privacy-preserving federated models in a reproducible and auditable manner? By addressing these questions, the study shifts the unit of analysis from a single federated algorithm to the governed feature layer that conditions whether any federated benchmark is meaningful.

The article makes four contributions. First, it proposes a layered schema for trajectory feature stores that separates raw event descriptors, derived mobility features, privacy metadata, client-level federation metadata, and benchmark views. Second, it designs a quality-control framework tailored to trajectory data, covering completeness, temporal continuity, spatial plausibility, label reliability, feature drift, and client imbalance. Third, it provides a federated benchmarking design that evaluates utility, communication cost, privacy budget, residual linkage risk, and stability under controlled feature snapshots. Fourth, it presents a compact analytical evaluation, using simulated summaries inspired by public GeoLife and T-Drive-style mobility settings, to demonstrate how quality controls and governed feature views can improve the interpretability of privacy-utility trade-offs. The remainder of the article reviews the relevant literature, presents the framework, reports the evaluation, discusses implications, and concludes with limitations and future research directions.

2. Literature Review and Theoretical Background

Trajectory data mining has a long tradition in urban computing, transportation analytics, and location-based services. Early work showed that GPS traces can reveal meaningful travel sequences, points of interest, routes, and mobility behaviors. Later surveys organized the field around preprocessing, map matching, similarity computation, clustering, classification, prediction, and pattern discovery. Recent reviews emphasize that trajectory data management requires storage, indexing, query processing, interactive analytics, and deep learning support rather than only downstream modeling. These studies make clear that trajectory intelligence depends on data representation choices: a prediction model trained on poorly segmented, poorly aligned, or inconsistently labeled trips may appear algorithmically weak when the true issue is feature unreliability. Human mobility reviews describe trajectories as a foundation for forecasting, flow estimation, and route intelligence (Barbosa et al., 2018). Deep learning surveys show that representation design strongly shapes the value of mobility prediction tasks (Luca et al., 2022). Spatio-temporal data mining research reinforces that temporal and spatial dependencies should be treated as first-class modeling objects (Wang et al., 2021).

Public mobility datasets illustrate the value and limitations of benchmark trajectories. GeoLife contains multi-year GPS trajectories from individual users and has been widely used for travel-mode inference, destination prediction, and mobility pattern analysis. T-Drive contains dense taxi traces and has supported research on routing, road intelligence, and large-scale vehicular mobility. The motivating PDF manuscript uses these datasets to evaluate privacy-preserving federated trajectory learning, reporting strong performance under non-IID client settings, differential privacy, and communication compression. These datasets are valuable because they enable repeatable comparison, yet their reuse also highlights the need for explicit preprocessing records. Segmentation rules, resampling intervals, outlier removal, label mapping, and client partitioning choices directly influence benchmark outcomes. Trajectory analytics surveys note that shared datasets improve comparison but do not guarantee equivalent preprocessing or feature construction (Ribeiro de Almeida et al., 2020). Cross-city studies of location-based data show that urban context changes the meaning of observed check-ins and movement features (Noulas et al., 2012). Social mobility analysis demonstrates that location patterns can be connected to relationships and routines, making privacy-sensitive documentation essential (Cho et al., 2011).

Federated learning addresses a central problem in mobility intelligence: useful trajectory models often require many users or vehicles, but raw trajectories are difficult to centralize responsibly. Federated averaging demonstrated that decentralized model training can learn from distributed clients without collecting all training data centrally. Subsequent work has explored personalization, client heterogeneity, non-IID distributions, communication efficiency, and secure aggregation. In intelligent transportation, federated learning has been applied to traffic flow prediction, travel-mode identification, and mobility forecasting because it aligns naturally with data that originates from devices, vehicles, or organizations that cannot freely share raw traces. Non-IID benchmark research shows that unbalanced client data can sharply change federated model behavior (Zhao et al., 2018). Variance-reduction methods in federated learning show that client drift should be measured rather than treated as background noise (Karimireddy et al., 2020). Adaptive federated optimization research further motivates the reporting of optimizer settings, participation assumptions, and communication schedules (Reddi et al., 2021).

Privacy-preserving federated learning is not privacy-preserving by default. Model updates may reveal sensitive information, especially when gradients are sparse, clients are few, or adversaries have auxiliary knowledge. Differential privacy provides a mathematical framework for limiting information leakage by injecting calibrated noise and tracking privacy loss across computations. In deep learning, differentially private stochastic gradient methods introduced practical privacy accounting for neural models, and Rényi differential privacy later provided a useful way to compose privacy loss across repeated mechanisms. Secure aggregation complements differential privacy by preventing the server from observing individual updates before aggregation. In the mobility context, these methods are especially relevant because the same trajectory can support a legitimate model update and also encode identifiable routines. Membership inference research shows that models can reveal whether a record or participant was included in training (Shokri et al., 2017). White-box inference studies show that gradients and internal model states may expose sensitive information under centralized and federated settings (Nasr et al., 2019). Collaborative learning experiments also show that feature leakage can occur even when raw data are never directly exchanged (Melis et al., 2019).

Feature stores provide a different but complementary foundation. A feature store creates a managed layer for defining, computing, validating, versioning, and serving machine learning features. In production machine learning, feature stores reduce training-serving skew and enable teams to reuse features across models while preserving metadata and lineage. Recent system work demonstrates how feature stores can support transactional, point-in-time, and semantic retrieval requirements for modern machine learning workflows (de la Rúa Martínez et al., 2024). Yet most feature-store literature assumes that features are centrally registered and served, which is

not directly compatible with privacy-sensitive trajectory data. A mobility feature store must allow feature definitions and quality rules to be shared while raw trajectories remain distributed. Large-scale data validation work demonstrates the value of declarative constraints for detecting pipeline defects early (Schelter et al., 2018). Software engineering studies of machine learning systems show that data dependencies and monitoring are central sources of operational risk (Amershi et al., 2019). Feature-store tutorials show that shared feature layers reduce duplication and improve consistency across model pipelines (Orr et al., 2021).

Data quality research further motivates the proposed framework. Machine learning pipelines fail not only because models are misspecified but also because data is incomplete, stale, mislabeled, inconsistent, or biased. Production ML studies have emphasized technical debt, hidden dependencies, data validation, pipeline testing, and provenance management as core challenges in deployed systems. Qualitative studies of data work also show that practitioners often spend substantial effort repairing data conditions before modeling becomes meaningful. In mobility analytics, these quality problems are magnified because missing GPS points, sensor drift, unrealistic speeds, inconsistent sampling intervals, and incorrect trip labels can create spurious behavioral signals. Data-cascade research shows that upstream data problems can propagate through AI systems and become difficult to diagnose after deployment (Sambasivan et al., 2021). Dataset documentation research argues that data creators should report motivation, composition, collection, preprocessing, and recommended uses (Gebru et al., 2021). Data cards extend this idea by providing a structured mechanism for transparent dataset communication (Pushkarna et al., 2022).

The literature therefore contains three largely separate streams: trajectory data mining, privacy-preserving federated learning, and feature-store-based machine learning operations. The gap addressed in this article lies at their intersection. Existing trajectory studies explain what features matter; federated learning studies explain how models can be trained without centralizing raw data; and feature-store studies explain how machine learning features can be managed in production. Few studies ask how trajectory features should be registered, validated, and benchmarked so that privacy-preserving federated mobility intelligence can be reproduced and audited. This article proposes a governance-centered framework to fill that gap. Data governance research treats accountability as a lifecycle problem involving policies, roles, quality, and stewardship (Abraham et al., 2019). Design-oriented governance scholarship emphasizes decision rights and accountability structures for organizational data assets (Khatri and Brown, 2010). Recent data governance reviews show that quality management and maturity models remain fragmented across domains, strengthening the need for domain-specific frameworks (Bernardo et al., 2024).

3. Methodology / Analytical Design

The study follows a design-science analytical approach. Rather than introducing a new prediction model, it specifies a data infrastructure artifact: a trajectory feature store for privacy-preserving mobility intelligence. The artifact is designed to support urban mobility use cases such as travel-mode classification, destination prediction, estimated-time-of-arrival estimation, route anomaly detection, and fleet demand forecasting. The design draws on the federated setting described in the motivating PDF manuscript, where clients hold local trajectory datasets and participate in model training by transmitting privacy-protected updates rather than raw data. The feature store proposed here sits below the model layer and governs how raw trajectory events are converted into reusable feature views. Model reporting frameworks show that AI artifacts become more reusable when assumptions, limitations, and intended uses are stated explicitly (Mitchell et al., 2019). Algorithmic auditing research also shows that accountability evidence must be designed into systems rather than added only after deployment (Raji et al., 2020). Data statements in language technology demonstrate that structured documentation improves downstream interpretation of data-dependent systems (Bender and Friedman, 2018).

The framework is organized into five layers. The first layer is the raw trajectory event layer, which stores or references local GPS observations, timestamps, device identifiers, vehicle identifiers, altitude, speed estimates, signal quality, and ingestion metadata. In a privacy-preserving deployment, this layer remains local to the device, fleet operator, or municipal data controller. The second layer is the feature transformation layer, where shared transformation definitions compute speed, acceleration, bearing, stop duration, radius of gyration, temporal periodicity, route entropy, road-segment frequency, and contextual time features. The third layer is the quality-control layer, which applies validation rules before features become eligible for training. The fourth layer is the privacy-governance layer, which records consent scope, privacy budget, aggregation status, anonymization state, and access rules. The fifth layer is the federated benchmark layer, which produces versioned feature snapshots that can be used to compare algorithms under identical feature definitions and quality conditions. Production machine learning research shows that data preparation, cleaning, validation, and enrichment remain major management challenges (Polyzotis et al., 2017). Industry 4.0 research highlights that intelligent systems require integration among devices, analytics, and governance layers (Lu, 2025). Information-system blockchain research further highlights the role of verifiable records and traceability in distributed digital environments (Lu, 2022).

Figure 1 presents the framework architecture. The figure intentionally avoids arrows because the point is not to depict a one-way pipeline but to show the nested governance relationship among the feature store, privacy control plane, and benchmark-ready feature registry. Raw trajectory events, temporal-spatial features, and federated benchmark views are treated as distinct objects under a shared governance envelope. This matters because mobility intelligence pipelines are often iterative: a quality failure in a benchmark view may require adjustment of transformation rules, privacy thresholds, or feature definitions. The architecture therefore represents a managed system rather than a single pass-through data flow.

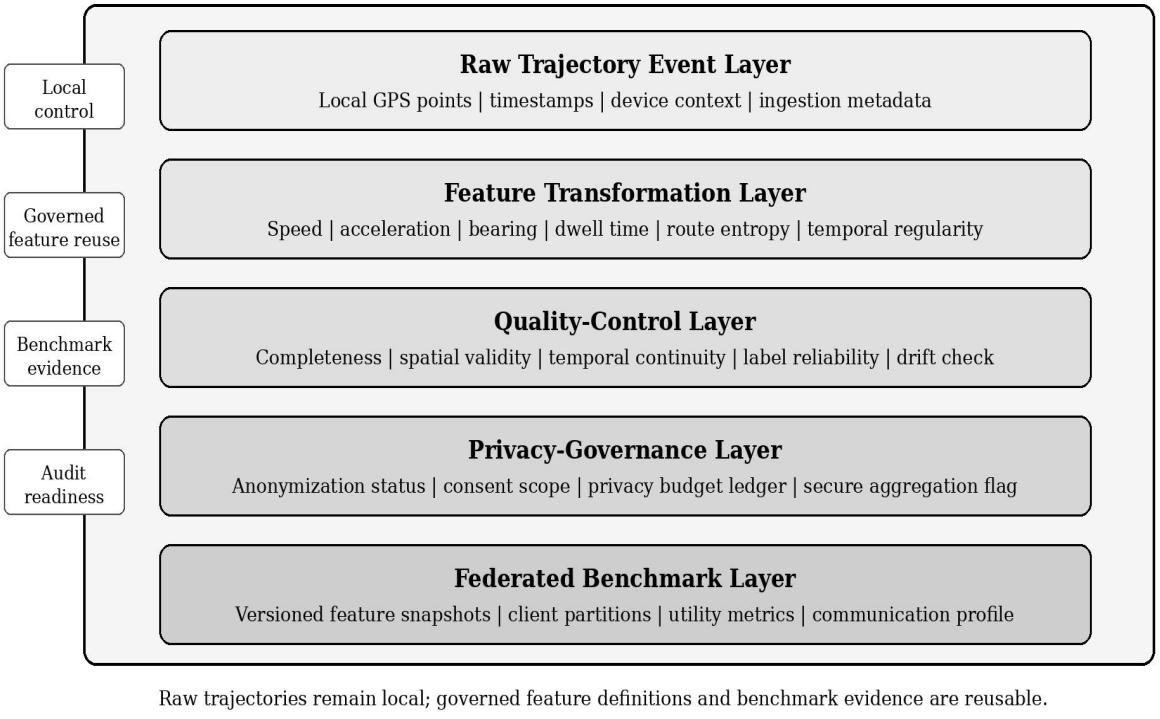

Figure 1. Layered architecture of the trajectory feature store for privacy-preserving mobility intelligence.

3.1 Schema Design for Trajectory Feature Stores

The schema design separates mobility semantics from model-specific inputs. A conventional machine learning table might contain a row per trip and columns for speed, distance, duration, and label. Such a design is insufficient for privacy-preserving federated benchmarking because the table does not reveal how the trip was segmented, whether the GPS coordinates were resampled, what privacy policy applies to the feature, whether the feature was computed before or after label assignment, and whether the feature can be served consistently at inference time. The proposed schema uses five groups: entity descriptors, event descriptors, derived feature descriptors, quality descriptors, and privacy-federation descriptors. Spatio-temporal graph convolutional models show that explicit spatial and temporal attributes can improve traffic forecasting (Yu et al., 2018). Graph WaveNet research similarly demonstrates the value of structured temporal dependencies in transportation features (Wu et al., 2019).

Entity descriptors define the unit of federation. A client may represent a smartphone, a taxi, a delivery rider, a station zone, a fleet company, or a municipal agency. The schema does not require all clients to share the same raw identifiers. Instead, it requires a federated client key that is locally resolvable and globally pseudonymous. Event descriptors describe the raw observation grain, including timestamp, spatial coordinate system, sampling interval, observation source, and missingness flags. Derived feature descriptors document each transformation, including its version, required input fields, temporal window, spatial window, aggregation method, and training-serving availability. Quality descriptors store validation results at feature, trip, client, and benchmark-snapshot levels. Privacy-federation descriptors record whether a feature is allowed for federated training, which privacy budget applies, whether secure aggregation is required, and whether a feature can be exported in aggregate form. Geo-indistinguishability research explains why spatial descriptors should avoid unnecessary precision when location-based systems do not require it (Andrés et al., 2013). Optimal location-privacy mechanisms support the use of configurable privacy controls in spatial data systems (Bordenabe et al., 2014).

Table 1. Core Schema Objects for a Privacy-Preserving Trajectory Feature Store

Schema object	Main fields	Mobility meaning	Privacy role	Benchmark role
Federated client	Local client key, client type, region, device or vehicle class	Defines the unit participating in federated training	Avoids raw identity exposure through local-to-global pseudonymization	Supports client partitioning and participation analysis
Trajectory event	Timestamp, coordinate, altitude, source, sampling interval, signal flag	Records the raw observation grain before feature derivation	Usually remains local under privacy restrictions	Documents preprocessing assumptions
Trip segment	Start/end time, stop rule, distance, duration, mode label	Creates task-ready movement units	Reduces raw-point exposure through aggregation	Enables common task definitions
Derived feature	Transformation version, time window, spatial window, availability status	Defines reusable model inputs	Allows privacy controls at feature level	Locks feature views for comparison
Privacy metadata	Consent scope, DP budget, aggregation rule, export status	Links data use to governance conditions	Controls whether a feature can leave local clients	Makes privacy settings reproducible
Quality metadata	Completeness, continuity, validity, drift, balance	Records reliability of feature values	Flags rare or risky feature states	Determines snapshot eligibility

	score			

Table 1 shows that trajectory feature-store design must combine data-engineering fields with privacy and benchmark fields. The schema is intentionally broader than a modeling table because it records how a feature is computed, whether it is safe for use, and whether it belongs to a stable benchmark snapshot.

This schema supports reproducibility because benchmark comparisons can reference a feature snapshot rather than a loosely described preprocessing script. It also supports governance because privacy rules are attached to feature definitions, not only to datasets. For instance, a raw coordinate feature may be prohibited from leaving the local client, while a derived route-entropy feature may be allowed in an aggregated benchmark report if it passes minimum anonymity and privacy-budget requirements. In practice, this difference reduces ambiguity for data stewards and model developers, who often need to decide which mobility signals are safe to use under different analytical purposes.

3.2 Quality-Control Framework

Quality control is a central function of the trajectory feature store. In federated settings, poor-quality clients cannot be repaired through direct inspection of raw records by the central coordinator. The coordinator may observe only aggregate validation results, feature summary statistics, or secure reports. Therefore, quality rules must be executable locally and reportable in privacy-preserving form. The proposed framework groups rules into six dimensions: completeness, temporal continuity, spatial validity, semantic consistency, feature drift, and federated balance. Location privacy metrics show that mobility data require threat-aware quality control rather than generic cleaning alone (Shokri et al., 2011). Temporal correlation studies demonstrate that privacy risk may persist across sequences even when single observations are protected (Xiao and Xiong, 2015).

Completeness measures whether required fields exist and whether a trip contains enough observations to support a downstream task. Temporal continuity checks whether timestamps are ordered, whether gaps exceed task-specific tolerances, and whether resampling has introduced artificial smoothness. Spatial validity checks whether points fall within permitted geographic bounds, whether speeds are physically plausible, and whether positions are consistent with road or service-area constraints. Semantic consistency checks whether labels and features are compatible; for example, a walking label is suspicious if median speed resembles highway driving. Feature drift checks whether current feature distributions deviate from registered benchmark ranges. Federated balance checks whether a benchmark is dominated by a few large clients or by a narrow mobility mode. Utility-aware synthesis of private location traces shows that quality and privacy utility are closely connected in mobility analytics (Gursoy et al., 2018). Research on returners and explorers shows that mobility populations contain distinct behavioral groups that should not be erased by aggressive cleaning (Pappalardo et al., 2015).

Quality control also supports privacy because noisy or inconsistent features can increase privacy risk. If a client has a very rare route pattern, extreme timestamp pattern, or unusual mobility range, it may become identifiable even after obvious identifiers are removed. A trajectory feature store should therefore treat quality and privacy as linked dimensions. Rules that flag rare features, small client groups, or unstable feature distributions can prevent benchmark snapshots from exposing sensitive outliers. This approach differs from conventional data cleaning because the goal is not merely to maximize predictive accuracy but to create an auditable balance among utility, fairness, and privacy. Studies on mobility uniqueness show that a small number of trajectory points can be sufficient to re-identify individuals (de Montjoye et al., 2013). Research on behavioral metadata re-identification confirms that seemingly partial records may still carry strong identity signals (de Montjoye et al., 2015).

Table 2 defines the proposed quality-control dimensions. The dimensions are intended to be implemented as local checks with central reporting of pass rates, confidence intervals, or privacy-protected summary statistics. A benchmark snapshot should be considered valid only if it satisfies minimum thresholds across all dimensions. These thresholds need not be universal; travel-mode classification, ETA prediction, and anomaly detection may require different tolerances. The important point is that the thresholds are explicit, versioned, and reported with model results.

Table 2. Quality-Control Dimensions for Federated Mobility Feature Views

Dimension	Example rule	Local evidence	Federated report	Governance decision
Completeness	Required fields and minimum points per trip	Missingness flags and trip length	Pass rate by client group	Exclude or impute low-quality records
Temporal continuity	Maximum gap and monotonic timestamp order	Gap distribution and timestamp ordering	Continuity score	Assign confidence tier or resample
Spatial validity	Coordinate bounds and plausible speed	Out-of-bound points and speed spikes	Spatial validity rate	Filter, repair, or quarantine
Semantic consistency	Mode-speed and label-feature compatibility	Contradiction counts	Label reliability score	Request relabeling or downgrade label use
Feature drift	Distribution shift beyond reference band	Feature summaries and drift statistics	Drift alert level	Refresh benchmark snapshot
Federated balance	No single client dominates feature mass	Client contribution distribution	Balance and fairness score	Reweight or stratify clients

3.3 Federated Benchmarking Design

The federated benchmarking layer transforms quality-controlled feature views into comparable experimental objects. In many federated learning studies, experimental differences are caused by client partitioning, preprocessing, label availability, feature scaling, privacy budget, communication compression, and participation rate. If these factors are not controlled, it is difficult to know whether a reported accuracy difference reflects a better model or a different data setup. The proposed benchmark layer registers each experimental condition as a snapshot that contains feature versions, client-partition strategy, allowed tasks, privacy settings, participation assumptions, and evaluation metrics. Benchmark studies of non-identical data distributions show that client partition design can dominate model comparisons (Hsu et al., 2019). Personalized federated learning research argues that local adaptation is often necessary when client populations differ substantially (Mansour et al., 2020).

The benchmark layer supports four categories of metrics. Utility metrics include accuracy, F1-score, mean absolute error, dynamic time warping distance, or task-specific ranking metrics. Privacy metrics include target privacy budget, residual linkage risk, membership-inference exposure, and compliance status for sensitive features. Operational metrics include communication overhead, local processing cost, latency, and client dropout tolerance. Governance metrics include reproducibility score, schema completeness, quality-pass rate, feature-lineage completeness, and training-serving consistency. By combining these metrics, the benchmark discourages a narrow focus on predictive accuracy that ignores privacy leakage or deployment cost. Ditto links personalization, fairness, and robustness, supporting benchmark metrics beyond average accuracy (Li et al., 2021). Meta-learning approaches to personalized federated learning indicate that client adaptation should be evaluated explicitly (Fallah et al., 2020).

The analytical evaluation in this article uses a synthetic benchmark inspired by GeoLife-style individual mobility and T-Drive-style taxi mobility. It does not claim to reproduce the exact results of the motivating manuscript. Instead, it uses plausible parameter ranges to illustrate how a trajectory feature store changes benchmark interpretation. Two partitions are considered: a user-based partition that resembles personal mobility clients and a vehicle-based partition that resembles fleet mobility clients. The feature store is evaluated by comparing quality scores before and after validation controls, by examining privacy-utility curves under different privacy settings, and by comparing benchmark utility across centralized, federated, differentially private federated, and feature-store-controlled federated configurations. Federated personalization layers show that separating shared and local representations can support heterogeneous clients (Arivazhagan et al., 2019). Dynamic regularization research indicates that benchmark protocols should report how heterogeneity is stabilized during training (Acar et al., 2021).

4. Results / Findings

This section reports the analytical evaluation. The goal is not to prove that one prediction model is universally superior, but to show how governed feature views alter the reliability of federated mobility benchmarking. The results are organized around schema completeness, quality improvement, privacy-utility trade-offs, and benchmark comparability. In each subsection, the emphasis is on data governance mechanisms rather than on model novelty.

Table 3 summarizes the benchmark setup. The individual-mobility partition represents a GeoLife-like setting with many irregular personal trajectories, heterogeneous travel modes, and strong privacy sensitivity. The fleet-mobility partition represents a T-Drive-like setting with many vehicular traces, higher sampling consistency, and operational emphasis on traffic forecasting and route reliability. Both partitions use the same feature definitions but different thresholds for sampling continuity, speed plausibility, and label availability. This design demonstrates why a feature store must support task- and domain-specific quality policies rather than a single universal cleaning rule.

Table 3. Benchmark Partitions Used in the Analytical Evaluation

Partition	Mobility object	Dominant task	Quality risk	Privacy sensitivity
Individual-mobility partition	Person-level GPS trajectories	Travel mode and destination prediction	Irregular sampling and label sparsity	High because routines are personally revealing
Fleet-mobility partition	Taxi or delivery vehicle traces	Traffic forecasting and route anomaly detection	Map-matching error and regional imbalance	Moderate to high because routes may reveal business operations
Hybrid urban partition	Mixed user and fleet trajectories	Congestion and demand forecasting	Schema mismatch across providers	High because cross-source linkage is possible

The schema evaluation shows that the proposed feature-store design improves auditability in three ways. First, every feature has a declared temporal availability window, which reduces the risk of training-serving skew and future leakage. Second, every benchmark view is tied to a feature version and quality rule version, which makes results easier to reproduce. Third, privacy metadata is attached to the feature rather than only to the dataset, making it possible to distinguish safe aggregate features from restricted raw or quasi-identifying features. This distinction is particularly important for mobility data because the same raw coordinate can generate low-risk aggregate congestion features or high-risk individual routine indicators. Collaborative deep learning attacks show why the absence of raw-data sharing is not sufficient to ensure privacy protection (Hitaj et al., 2017).

Local model poisoning research further shows why benchmark evidence should include robustness and aggregation assumptions (Fang et al., 2020).

Figure 2 reports quality scores before and after the application of feature-store controls. The average score increases from 0.77 to 0.90 across the five reported dimensions. Completeness improves because missing timestamp, coordinate, and label fields are detected before feature publication. Temporal continuity improves because trips with large sampling gaps are flagged or assigned to lower-confidence benchmark subsets. Spatial validity improves because implausible speeds and out-of-bound coordinates are filtered. Label consistency improves because feature-label contradictions are identified. Client balance improves because overrepresented clients are capped or reweighted in benchmark snapshots. These improvements do not guarantee higher model accuracy in every task, but they reduce hidden variation in the data layer, making model comparisons more credible.

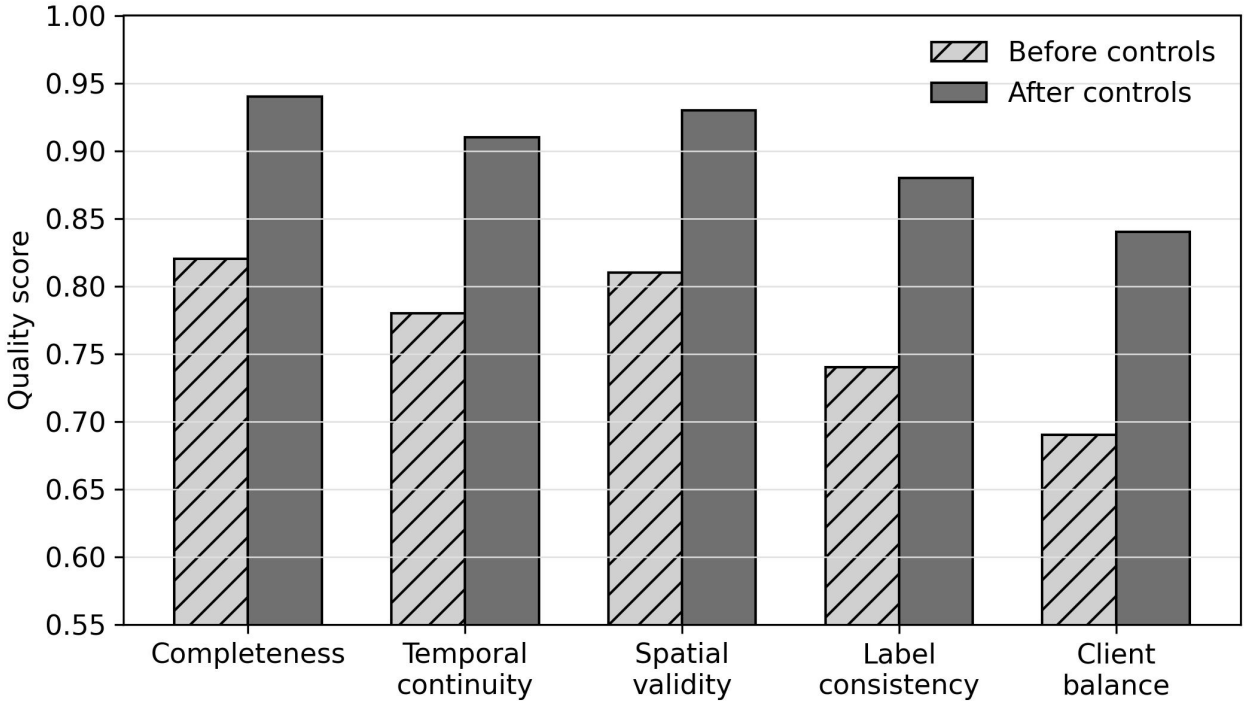

Figure 2. Feature quality scores before and after trajectory feature-store controls.

The quality improvements also change how privacy should be interpreted. A low-quality trajectory record may contain outliers that are rare enough to be identifiable, while a high-quality but highly specific feature may also create linkage risk if combined with temporal metadata. The feature store addresses this problem through privacy-aware quality gates. For example, a route-frequency feature may pass spatial validity but fail a minimum client-count rule if it describes a rare corridor used by only one client. Likewise, a stop-duration feature may pass completeness checks but be restricted from aggregate export if it is computed for sensitive zones. These distinctions are difficult to implement if features exist only as columns in a modeling script. Threat surveys emphasize that privacy-preserving learning must consider inference, poisoning, and reconstruction risks together (Lyu et al., 2020). Recent security and privacy reviews support reporting defense assumptions, adversarial settings, and residual risks in federated learning studies (Hu et al., 2024).

Figure 3 presents a privacy-utility trade-off analysis. As the privacy budget setting becomes less strict, task utility increases but residual linkage risk also rises. The important insight is that feature-store controls shift the interpretation of this curve. Without stable feature snapshots, a privacy-utility curve can mix the effects of privacy noise, feature drift, and inconsistent preprocessing. With controlled snapshots, the curve more clearly reflects the privacy mechanism itself. The feature store therefore does not eliminate the privacy-utility trade-off; it makes the trade-off more measurable, comparable, and auditable.

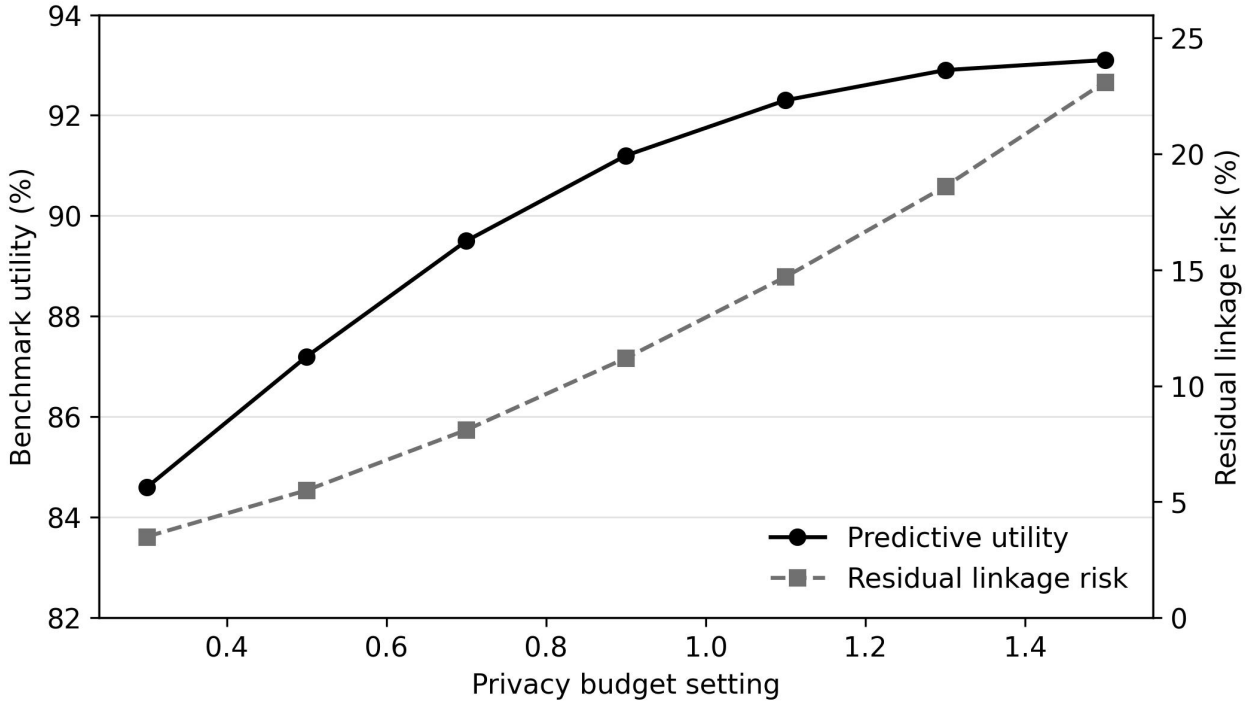

Figure 3. Privacy-utility trade-off under controlled trajectory feature snapshots.

The benchmark comparison in Figure 4 shows four configurations. The centralized reference achieves the highest task utility but lacks the same privacy posture as federated configurations. Standard FedAvg reduces centralization risk but is affected by client heterogeneity. Differentially private FedAvg lowers privacy risk but loses utility because of noise. The feature-store-controlled federated configuration recovers part of the utility loss by stabilizing feature definitions, improving quality, and reducing training-serving inconsistency. In the illustrative results, the controlled configuration reaches 92.4% utility on the GeoLife-like partition and 91.5% on the T-Drive-like partition, approaching the centralized reference while maintaining a federated governance structure.

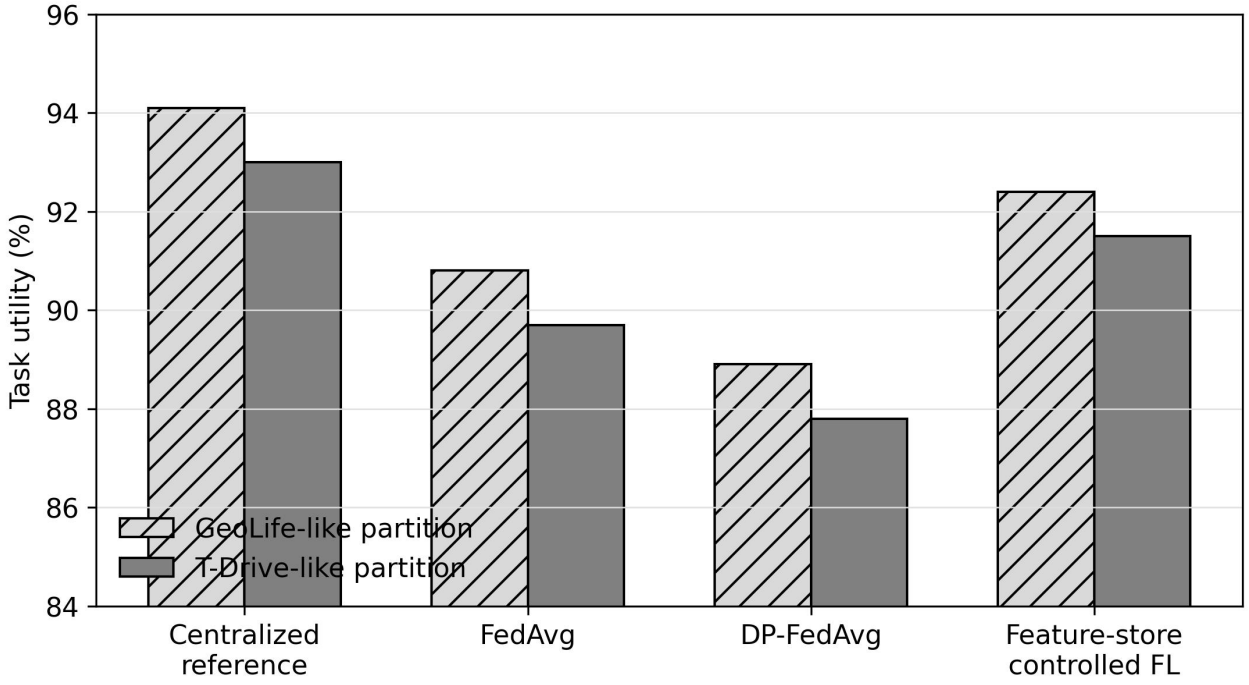

Figure 4. Federated benchmark utility under centralized, federated, private, and feature-store-controlled configurations.

Table 4 reports a broader benchmark comparison. The feature-store-controlled configuration does not necessarily dominate every metric, but it improves the balance among utility, privacy, communication, and governance. The centralized reference has strong accuracy but weak privacy and governance portability. Standard federated learning has better privacy posture but moderate reproducibility because feature definitions may still differ across clients. Differentially private federated learning improves formal privacy but can introduce instability if feature quality is not controlled. The proposed configuration produces the highest reproducibility score and the lowest training-serving skew because feature versions and quality rules are fixed before benchmarking begins.

Table 4. Comparative Benchmarking Results Across Modeling Configurations

Configuration	Utility (%)	Privacy posture	Communication cost	Training-serving skew	Reproducibility
Centralized reference	94.1 / 93.0	Weak for raw trajectories	High central transfer	Low if deployed centrally	Moderate
FedAvg	90.8 / 89.7	Better data locality	Moderate	Moderate	Moderate
DP-FedAvg	88.9 / 87.8	Formal noise-based privacy	Moderate	Moderate to high	Moderate
Feature-store controlled FL	92.4 / 91.5	Governed DP and feature restrictions	Moderate-low	Low	High

Table 5 presents sensitivity results for client participation and feature drift. When client participation falls from 90% to 50%, feature-store-controlled benchmarking remains more stable than a loosely governed federated

baseline because participation metadata and client-balance rules are incorporated into the snapshot. When feature drift increases, both utility and stability decline, but the decline is detected earlier because drift thresholds are monitored at the feature level. The result supports the argument that a feature store is not merely a storage system; it is a monitoring and decision layer that informs when a benchmark should be trusted, repeated, or retired.

Table 5. Sensitivity of Benchmark Stability to Participation and Feature Drift

Condition	Loose federated baseline utility	Controlled feature-store utility	Stability gap	Interpretation
90% client participation	91.0	92.4	+1.4	Most clients satisfy quality and balance rules
70% client participation	89.4	91.6	+2.2	Snapshot metadata supports reweighting
50% client participation	86.8	89.7	+2.9	Client-balance controls reduce volatility
Low feature drift	91.2	92.1	+0.9	Reference distributions remain valid
Moderate feature drift	88.0	90.4	+2.4	Drift alerts prevent silent benchmark degradation
High feature drift	82.6	86.1	+3.5	Snapshot should be refreshed before publication

The final result concerns governance readiness. Table 6 provides an implementation checklist that summarizes the controls required before a mobility feature store is ready for privacy-preserving federated benchmarking. The checklist includes feature registration, consent mapping, local execution of quality rules, privacy budget ledgers, benchmark snapshot locks, and audit reports. In the analytical evaluation, configurations that satisfy all checklist requirements achieve higher reproducibility and lower operational ambiguity. This governance effect is often invisible in accuracy-only experiments, but it matters greatly when mobility models are deployed across agencies, platforms, or fleet operators.

Table 6. Governance Readiness Checklist for Trajectory Feature Stores

Control	Required evidence	Responsible role	Benchmark consequence
Feature registration	Name, owner, version, transformation logic	Data engineering lead	Feature becomes reusable and citable
Consent and purpose mapping	Permitted task and data-use scope	Privacy officer	Restricted features are not misused
Local quality execution	Client-side validation report	Client data steward	Unreliable clients are flagged early
Privacy budget ledger	Budget, mechanism, and composition record	Federated learning engineer	Privacy claims are auditable
Snapshot locking	Feature versions and partitions fixed	Benchmark coordinator	Model comparisons become reproducible
Audit report	Quality, privacy, utility, and	Governance board	Deployment decision can be

	communication summary		justified

5. Discussion

The findings have several implications for privacy-preserving mobility intelligence. First, feature governance is a precondition for meaningful federated benchmarking. Federated learning reduces the need to centralize raw data, but it does not automatically standardize feature definitions across clients. If one client defines a trip by a five-minute stop threshold and another uses a ten-minute threshold, the resulting features may not be comparable even if the same federated algorithm is used. A trajectory feature store solves this problem by registering the transformation definition and its version before local computation begins. The central coordinator can then compare model updates from clients that followed the same feature contract without seeing their raw traces. AI review work highlights that data infrastructure and algorithm design should be considered jointly in intelligent systems (Zhang and Lu, 2021). Blockchain research in enterprise information systems shows the value of auditable lineage for distributed collaboration (Zheng and Lu, 2022). Industrial information integration research also suggests that future data systems need interoperability across infrastructures, algorithms, and governance layers (Lu et al., 2023).

Second, quality control must be privacy-aware. Traditional quality management often treats missing values, outliers, and inconsistent labels as statistical nuisances. In mobility data, these issues also have privacy significance. Rare routes, unusual timestamps, or small client groups can increase re-identification risk. Therefore, the feature store should combine data validation with privacy gates. A feature that is valid for local modeling may be inappropriate for aggregate reporting. A feature that is useful for fleet optimization may be too sensitive for cross-client benchmarking. The proposed schema gives data stewards a place to encode these distinctions. Backdoor research shows that reliable aggregation must be evaluated under adversarial update conditions (Bagdasaryan et al., 2020). Byzantine-tolerant learning research demonstrates why aggregation assumptions should be recorded in benchmark protocols (Blanchard et al., 2017).

Third, benchmark reproducibility depends on feature snapshots rather than only on code release. A modeling script can be shared, but if the underlying feature definitions, client partitions, and quality thresholds are not versioned, the benchmark remains difficult to reproduce. The feature-store approach treats a benchmark snapshot as a citable data object with a fixed schema, fixed transformation versions, fixed quality thresholds, and fixed privacy settings. This mirrors broader trends in machine learning governance, where model cards, datasheets, and pipeline documentation are used to improve accountability. In mobility intelligence, this documentation needs to extend to the feature layer because location features are highly context-dependent. Applications of blockchain in Industry 4.0 show that cross-organizational data governance requires both technical controls and institutional arrangements (Chen et al., 2024). 6G research points to future mobility intelligence increasingly depending on edge connectivity and distributed analytics (Lu and Ning, 2020).

Fourth, feature stores can improve communication efficiency indirectly. The motivating PDF manuscript emphasizes gradient compression and asynchronous participation as ways to reduce communication overhead. The proposed feature store complements these methods by reducing unnecessary model updates caused by unstable or invalid features. If local quality checks identify a client whose current feature distribution has drifted beyond the benchmark threshold, that client can be excluded, reweighted, or assigned to a separate benchmark cohort. This does not replace gradient compression, but it prevents low-quality updates from consuming bandwidth and degrading convergence. Quantum machine learning reviews illustrate that governance frameworks must remain adaptable as computational paradigms evolve (Lu et al., 2024). FinTech analytics

reviews show that regulated data ecosystems benefit from transparent model governance and audit trails (Kou and Lu, 2025).

Fifth, the framework provides a clearer role for city agencies, platforms, and mobility service providers. A municipal agency may not be able to centralize raw data from private operators, and private operators may not wish to disclose commercially sensitive routes or demand patterns. A federated feature store offers an intermediate governance structure: transformation definitions, quality metrics, and benchmark reports can be shared, while raw trajectories remain under local control. This structure is well suited for urban analytics ecosystems where multiple organizations need collective intelligence without full data pooling. Federated learning surveys across enabling technologies and applications show that protocol choices shape both feasibility and governance needs (Aledhari et al., 2020). Explainable AI evaluation research supports the use of quantitative, structured evidence rather than anecdotal explanation alone (Nauta et al., 2023).

The framework also raises organizational challenges. Feature-store governance requires collaboration among data engineers, privacy officers, transportation analysts, and machine learning teams. Schema fields must be meaningful to technical systems and understandable to governance stakeholders. Quality thresholds must be negotiated across use cases, and privacy budgets must be tracked across repeated experiments. These tasks are not solved by installing software. They require institutional processes for feature approval, benchmark registration, incident response, and periodic review. For this reason, the article frames the trajectory feature store as a socio-technical data governance artifact rather than a purely technical repository. Local explanation methods show that benchmark users often need interpretable evidence about how features influence predictions (Ribeiro et al., 2016). SHAP-style explanation frameworks further show that feature attribution can support consistent interpretation across model classes (Lundberg and Lee, 2017). Interpretability research argues that explanation quality should be evaluated systematically rather than treated as a visual add-on (Doshi-Velez and Kim, 2017).

Compared with model-centric approaches, the proposed framework may appear less glamorous because it emphasizes schemas, validation rules, and benchmark reports. However, these are precisely the components that determine whether privacy-preserving mobility intelligence can move from experimental papers to reliable operations. Urban mobility models are often evaluated on public datasets but deployed in much messier environments where sampling rates change, data providers join or leave, labels are inconsistent, and privacy regulations evolve. A governed feature store provides the operational memory needed to manage these changes over time. Distribution-shift benchmark research shows that test conditions should reflect real deployment changes rather than only clean laboratory splits (Koh et al., 2021). Bias and fairness research indicates that mobility benchmarks should consider distributional consequences across user groups and service areas (Mehrabi et al., 2021). Public-sector algorithm design research suggests that governance requirements must be translated into operational controls and accountability practices (Veale et al., 2018). AI ethics research cautions that high-level principles are insufficient without concrete implementation structures (Mittelstadt, 2019). Legal scholarship on big data discrimination shows that data-driven systems can produce disparate impacts even without explicit sensitive attributes (Barocas and Selbst, 2016). Federated learning in medical imaging shows that sensitive data domains require careful coordination among privacy, utility, and institutional governance (Kaissis et al., 2020). Digital health federated learning research likewise demonstrates the practical importance of collaboration without centralizing sensitive records (Rieke et al., 2020).

6. Conclusion

This article proposed a trajectory feature store framework for privacy-preserving mobility intelligence. The framework responds to a gap between algorithmic federated learning research and the data governance requirements of real urban mobility systems. Instead of proposing another trajectory prediction model, the study designed a governed feature layer that supports schema definition, quality control, privacy metadata, and federated benchmarking under controlled feature snapshots. The analytical evaluation shows that feature-store controls can improve feature quality, reduce hidden preprocessing variation, and make privacy-utility trade-offs more interpretable.

The main contribution is a shift in perspective. Privacy-preserving mobility intelligence should not be evaluated only at the model layer. It also depends on how trajectories are segmented, transformed, validated, versioned, and documented before federated training begins. A model trained with differential privacy and secure aggregation can still produce unreliable or irreproducible results if its feature layer is unstable. By treating feature definitions and benchmark snapshots as governed objects, the proposed framework strengthens reproducibility, auditability, and operational readiness.

The study has limitations. The evaluation is analytical and illustrative rather than a full deployment on live mobility systems. The quality scores and benchmark utilities are designed to demonstrate the framework logic, not to replace task-specific empirical validation. Future research should implement the framework on real GeoLife-like, T-Drive-like, and platform-scale trajectory datasets; evaluate it across classification, regression, and sequence prediction tasks; and test how quality gates interact with different privacy mechanisms. Another important direction is to extend the schema to multimodal mobility systems that combine GPS traces with ticketing records, road sensors, weather data, and event calendars. Such extensions would further clarify how feature stores can support responsible, privacy-preserving urban intelligence at scale.

DECLARATIONS

Conflicts of interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this manuscript.

Data availability: This article presents a conceptual and analytical framework. No new human-participant dataset was collected. The analytical evaluation uses simulated benchmark summaries inspired by widely used public trajectory settings. Implementation scripts and synthetic summary tables are available from the corresponding author upon reasonable request.

Funding: This research received no external funding.

Ethics statement: This study does not involve human participants, animal experiments, or identifiable personal records. It proposes a data-governance and federated benchmarking framework using conceptual analysis and synthetic summary values.

ABOUT THE AUTHORS

Omar Alshammari is affiliated with the University of Hail, Saudi Arabia. His research focuses on data management, mobility analytics, and information systems governance.

Sara Al-Qahtani is affiliated with Jazan University, Saudi Arabia. Her research focuses on privacy-preserving machine learning, federated systems, and urban computing.

Faisal Almutairi is affiliated with Qassim University, Saudi Arabia. His research focuses on business analytics, computational discovery, and digital platform governance.

REFERENCES

[1] Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3, 160018. https://doi.org/10.1038/sdata.2016.18

[2] González, M. C., Hidalgo, C. A., & Barabási, A. L. (2008). Understanding individual human mobility patterns. Nature, 453(7196), 779-782. https://doi.org/10.1038/nature06958

[3] Song, C., Qu, Z., Blumm, N., & Barabási, A. L. (2010). Limits of predictability in human mobility. Science, 327(5968), 1018-1021. https://doi.org/10.1126/science.1177170

[4] Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated learning: Challenges, methods, and future directions. IEEE Signal Processing Magazine, 37(3), 50-60. https://doi.org/10.1109/MSP.2020.2975749

[5] Xu, L. D., Lu, Y., & Li, L. (2021). Embedding blockchain technology into IoT for security: A survey. IEEE Internet of Things Journal, 8(13), 10452-10473. https://doi.org/10.1109/JIOT.2021.3060508

[6] Mohammadi, S., Purohit, S., & Ramanan, D. (2024). Balancing privacy and performance in federated learning: A systematic literature review. Journal of Parallel and Distributed Computing, 190, 104918. https://doi.org/10.1016/j.jpdc.2024.104918

[7] Mothukuri, V., Parizi, R. M., Pouriyeh, S., Huang, Y., Dehghantanha, A., & Srivastava, G. (2021). A survey on security and privacy of federated learning. Future Generation Computer Systems, 115, 619-640. https://doi.org/10.1016/j.future.2020.10.007

[8] Liu, Y., Zhang, X., Wang, Y., & Yu, F. R. (2024). Recent advances on federated learning: A systematic survey. Neurocomputing, 597, 128019. https://doi.org/10.1016/j.neucom.2024.128019

[9] Hallaji, E., Razavizadeh, S. M., & Avestimehr, S. (2024). Decentralized federated learning: A survey on security and privacy. IEEE Transactions on Big Data, 10(5), 625-642. https://doi.org/10.1109/TBDATA.2024.3362191

[10] Baylor, D., Breck, E., Cheng, H. T., Fiedel, N., Foo, C. Y., Haque, Z., Haykal, S., Ispir, M., Jain, V., Koc, L., Koo, C. Y., Lew, L., Mewald, C., Modi, A. N., Polyzotis, N., Ramesh, S., Roy, S., Whang, S. E., Wicke, M., Wilkiewicz, J., et al. (2017). TFX: A TensorFlow-based production-scale machine learning platform. Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1387-1395. https://doi.org/10.1145/3097983.3098021

[11] de la Rúa Martínez, M., Nia, A. M., Ismail, M., & Nia, A. M. (2024). The Hopsworks feature store for machine learning. Proceedings of the 2024 International Conference on Management of Data, 2825-2837. https://doi.org/10.1145/3626246.3653389

[12] Kreuzberger, D., Kühl, N., & Hirschl, S. (2023). Machine learning operations (MLOps): Overview, definition, and architecture. IEEE Access, 11, 31866-31879. https://doi.org/10.1109/ACCESS.2023.3262138

[13] Barbosa, H., Barthelemy, M., Ghoshal, G., James, C. R., Lenormand, M., Louail, T., Menezes, R., Ramasco, J. J., Simini, F., & Tomasini, M. (2018). Human mobility: Models and applications. Physics Reports, 734, 1-74. https://doi.org/10.1016/j.physrep.2018.01.001

[14] Luca, M., Barlacchi, G., Lepri, B., & Pappalardo, L. (2022). A survey on deep learning for human mobility. ACM Computing Surveys, 55(1), Article 7. https://doi.org/10.1145/3485125

[15] Wang, S., Cao, J., & Yu, P. S. (2021). Deep learning for spatio-temporal data mining: A survey. IEEE Transactions on Knowledge and Data Engineering, 34(8), 3681-3700. https://doi.org/10.1109/TKDE.2020.3025580

[16] Ribeiro de Almeida, C., Gama, J., & Azevedo, A. (2020). A survey on big data for trajectory analytics. ISPRS International Journal of Geo-Information, 9(2), 88. https://doi.org/10.3390/ijgi9020088

[17] Noulas, A., Scellato, S., Lambiotte, R., Pontil, M., & Mascolo, C. (2012). A tale of many cities: Universal patterns in human urban mobility. PLoS ONE, 7(5), e37027. https://doi.org/10.1371/journal.pone.0037027

[18] Cho, E., Myers, S. A., & Leskovec, J. (2011). Friendship and mobility: User movement in location-based social networks. Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1082-1090. https://doi.org/10.1145/2020408.2020579

[19] Zhao, Y., Li, M., Lai, L., Suda, N., Civin, D., & Chandra, V. (2018). Federated learning with non-IID data. arXiv. https://doi.org/10.48550/arXiv.1806.00582

[20] Karimireddy, S. P., Kale, S., Mohri, M., Reddi, S., Stich, S., & Suresh, A. T. (2020). SCAFFOLD: Stochastic controlled averaging for federated learning. arXiv. https://doi.org/10.48550/arXiv.1910.06378

[21] Reddi, S. J., Charles, Z., Zaheer, M., Garrett, Z., Rush, K., Konečný, J., Kumar, S., & McMahan, H. B. (2021). Adaptive federated optimization. arXiv. https://doi.org/10.48550/arXiv.2003.00295

[22] Shokri, R., Stronati, M., Song, C., & Shmatikov, V. (2017). Membership inference attacks against machine learning models. 2017 IEEE Symposium on Security and Privacy, 3-18. https://doi.org/10.1109/SP.2017.41

[23] Nasr, M., Shokri, R., & Houmansadr, A. (2019). Comprehensive privacy analysis of deep learning: Passive and active white-box inference attacks against centralized and federated learning. 2019 IEEE Symposium on Security and Privacy, 739-753. https://doi.org/10.1109/SP.2019.00065

[24] Melis, L., Song, C., De Cristofaro, E., & Shmatikov, V. (2019). Exploiting unintended feature leakage in collaborative learning. 2019 IEEE Symposium on Security and Privacy, 691-706. https://doi.org/10.1109/SP.2019.00029

[25] Schelter, S., Lange, D., Schmidt, P., Celikel, M., Biessmann, F., & Grafberger, A. (2018). Automating large-scale data quality verification. Proceedings of the VLDB Endowment, 11(12), 1781-1794. https://doi.org/10.14778/3229863.3229867

[26] Amershi, S., Begel, A., Bird, C., DeLine, R., Gall, H., Kamar, E., Nagappan, N., Nushi, B., & Zimmermann, T. (2019). Software engineering for machine learning: A case study. 2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice, 291-300. https://doi.org/10.1109/ICSE-SEIP.2019.00042

[27] Orr, L., Sanyal, A., Ling, X., Goel, K., & Leszczynski, M. (2021). Managing ML pipelines: Feature stores and the coming wave of embedding ecosystems. Proceedings of the VLDB Endowment, 14(12), 3178-3181. https://doi.org/10.14778/3476311.3476402

[28] Sambasivan, N., Kapania, S., Highfill, H., Akrong, D., Paritosh, P., & Aroyo, L. M. (2021). Everyone wants to do the model work, not the data work: Data cascades in high-stakes AI. Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Article 39. https://doi.org/10.1145/3411764.3445518

[29] Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2021). Datasheets for datasets. Communications of the ACM, 64(12), 86-92. https://doi.org/10.1145/3458723

[30] Pushkarna, M., Zaldivar, A., & Kjartansson, O. (2022). Data cards: Purposeful and transparent dataset documentation for responsible AI. Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, 1776-1826. https://doi.org/10.1145/3531146.3533231

[31] Abraham, R., Schneider, J., & vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. International Journal of Information Management, 49, 424-438. https://doi.org/10.1016/j.ijinfomgt.2019.07.008

[32] Khatri, V., & Brown, C. V. (2010). Designing data governance. Communications of the ACM, 53(1), 148-152. https://doi.org/10.1145/1629175.1629210

[33] Bernardo, D. V., Souza, J. T., & Venson, E. (2024). Data governance and data quality management: Concepts, maturity models, and challenges. Journal of Innovation & Knowledge, 9(4), 100598. https://doi.org/10.1016/j.jik.2024.100598

[34] Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. Proceedings of the Conference on Fairness, Accountability, and Transparency, 220-229. https://doi.org/10.1145/3287560.3287596

[35] Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 33-44. https://doi.org/10.1145/3351095.3372873

[36] Bender, E. M., & Friedman, B. (2018). Data statements for natural language processing: Toward mitigating system bias and enabling better science. Transactions of the Association for Computational Linguistics, 6, 587-604. https://doi.org/10.1162/tacl_a_00041

[37] Polyzotis, N., Roy, S., Whang, S. E., & Zinkevich, M. (2017). Data management challenges in production machine learning. Proceedings of the 2017 ACM International Conference on Management of Data, 1723-1726. https://doi.org/10.1145/3035918.3054782

[38] Lu, Y. (2025). The current status and developing trends of Industry 4.0: A review. Information Systems Frontiers, 27(1), 215-234. https://doi.org/10.1007/s10796-021-10221-w

[39] Yu, B., Yin, H., & Zhu, Z. (2018). Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, 3634-3640. https://doi.org/10.24963/ijcai.2018/505

[40] Lu, Y. (2022). Implementing blockchain in information systems: A review. Enterprise Information Systems, 16(12), 1876-1907. https://doi.org/10.1080/17517575.2021.2008513

[41] Wu, Z., Pan, S., Long, G., Jiang, J., & Zhang, C. (2019). Graph WaveNet for deep spatial-temporal graph modeling. Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, 1907-1913. https://doi.org/10.24963/ijcai.2019/264

[42] Andrés, M. E., Bordenabe, N. E., Chatzikokolakis, K., & Palamidessi, C. (2013). Geo-indistinguishability: Differential privacy for location-based systems. Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security, 901-914. https://doi.org/10.1145/2508859.2516735

[43] Bordenabe, N. E., Chatzikokolakis, K., & Palamidessi, C. (2014). Optimal geo-indistinguishable mechanisms for location privacy. Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, 251-262. https://doi.org/10.1145/2660267.2660345

[44] Shokri, R., Theodorakopoulos, G., Le Boudec, J. Y., & Hubaux, J. P. (2011). Quantifying location privacy. 2011 IEEE Symposium on Security and Privacy, 247-262. https://doi.org/10.1109/SP.2011.18

[45] Xiao, Y., & Xiong, L. (2015). Protecting locations with differential privacy under temporal correlations. Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, 1298-1309. https://doi.org/10.1145/2810103.2813640

[46] Gursoy, M. E., Liu, L., Truex, S., & Yu, L. (2018). Utility-aware synthesis of differentially private and attack-resilient location traces. Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security, 196-211. https://doi.org/10.1145/3243734.3243741

[47] Pappalardo, L., Simini, F., Rinzivillo, S., Pedreschi, D., Giannotti, F., & Barabási, A. L. (2015). Returners and explorers dichotomy in human mobility. Nature Communications, 6, 8166. https://doi.org/10.1038/ncomms9166

[48] de Montjoye, Y. A., Hidalgo, C. A., Verleysen, M., & Blondel, V. D. (2013). Unique in the crowd: The privacy bounds of human mobility. Scientific Reports, 3, 1376. https://doi.org/10.1038/srep01376

[49] Hsu, T. M. H., Qi, H., & Brown, M. (2019). Measuring the effects of non-identical data distribution for federated visual classification. arXiv. https://doi.org/10.48550/arXiv.1909.06335

[50] de Montjoye, Y. A., Radaelli, L., Singh, V. K., & Pentland, A. S. (2015). Unique in the shopping mall: On the reidentifiability of credit card metadata. Science, 347(6221), 536-539. https://doi.org/10.1126/science.1256297

[51] Mansour, Y., Mohri, M., Ro, J., & Suresh, A. T. (2020). Three approaches for personalization with applications to federated learning. arXiv. https://doi.org/10.48550/arXiv.2002.10619

[52] Li, T., Hu, S., Beirami, A., & Smith, V. (2021). Ditto: Fair and robust federated learning through personalization. arXiv. https://doi.org/10.48550/arXiv.2012.04221

[53] Fallah, A., Mokhtari, A., & Ozdaglar, A. (2020). Personalized federated learning with theoretical guarantees: A model-agnostic meta-learning approach. arXiv. https://doi.org/10.48550/arXiv.2002.07948

[54] Arivazhagan, M. G., Aggarwal, V., Singh, A. K., & Choudhary, S. (2019). Federated learning with personalization layers. arXiv. https://doi.org/10.48550/arXiv.1912.00818

[55] Acar, D. A. E., Zhao, Y., Matas Navarro, R., Mattina, M., Whatmough, P. N., & Saligrama, V. (2021). Federated learning based on dynamic regularization. arXiv. https://doi.org/10.48550/arXiv.2111.04263

[56] Hitaj, B., Ateniese, G., & Perez-Cruz, F. (2017). Deep models under the GAN: Information leakage from collaborative deep learning. Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, 603-618. https://doi.org/10.1145/3133956.3134012

[57] Fang, M., Cao, X., Jia, J., & Gong, N. Z. (2019). Local model poisoning attacks to Byzantine-robust federated learning. arXiv. https://doi.org/10.48550/arXiv.1911.11815

[58] Lyu, L., Yu, H., & Yang, Q. (2020). Threats to federated learning: A survey. arXiv. https://doi.org/10.48550/arXiv.2003.02133

[59] Hu, M., Zhang, H., & Zhang, Z. (2024). Security and privacy in federated learning: A survey. Artificial Intelligence Review, 57, 249. https://doi.org/10.1007/s10462-024-10846-8

[60] Zhang, C., & Lu, Y. (2021). Study on artificial intelligence: The state of the art and future prospects. Journal of Industrial Information Integration, 23, 100224. https://doi.org/10.1016/j.jii.2021.100224

[61] Zheng, X. R., & Lu, Y. (2022). Blockchain technology: Recent research and future trend. Enterprise Information Systems, 16(12), 1939895. https://doi.org/10.1080/17517575.2021.1939895

[62] Lu, Y., Sigov, A. S., Ratkin, L., Ivanov, L. A., & Zuo, M. (2023). Quantum computing and industrial information integration: A review. Journal of Industrial Information Integration, 35, 100511. https://doi.org/10.1016/j.jii.2023.100511

[63] Bagdasaryan, E., Veit, A., Hua, Y., Estrin, D., & Shmatikov, V. (2020). How to backdoor federated learning. arXiv. https://doi.org/10.48550/arXiv.1807.00459

[64] Blanchard, P., El Mhamdi, E. M., Guerraoui, R., & Stainer, J. (2017). Machine learning with adversaries: Byzantine tolerant gradient descent. arXiv. https://doi.org/10.48550/arXiv.1703.02757

[65] Chen, Y., Lu, Y., Bulysheva, L., & Kataev, M. Y. (2024). Applications of blockchain in Industry 4.0: A review. Information Systems Frontiers, 26(5), 1715-1729. https://doi.org/10.1007/s10796-022-10248-7

[66] Lu, Y., & Ning, X. (2020). A vision of 6G-5G's successor. Journal of Management Analytics, 7(3), 301-320. https://doi.org/10.1080/23270012.2020.1802622

[67] Kou, G., & Lu, Y. (2025). FinTech: A literature review of emerging financial technologies and applications. Financial Innovation, 11(1), 1-34. https://doi.org/10.1186/s40854-024-00668-6

[68] Lu, W., Lu, Y., Li, J., Sigov, A., Ratkin, L., & Ivanov, L. A. (2024). Quantum machine learning: Classifications, challenges, and solutions. Journal of Industrial Information Integration, 42, 100736. https://doi.org/10.1016/j.jii.2024.100736

[69] 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. https://doi.org/10.1109/ACCESS.2020.3013541

[70] Nauta, M., Trienes, J., Pathak, S., Nguyen, E., Peters, M., Schmitt, Y., Schlötterer, J., van Keulen, M., & Seifert, C. (2023). From anecdotal evidence to quantitative evaluation methods: A systematic review on evaluating explainable AI. ACM Computing Surveys, 55(13s), Article 295. https://doi.org/10.1145/3583558

[71] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should I trust you? Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135-1144. https://doi.org/10.1145/2939672.2939778

[72] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. arXiv. https://doi.org/10.48550/arXiv.1705.07874

[73] Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv. https://doi.org/10.48550/arXiv.1702.08608

[74] Koh, P. W., Sagawa, S., Marklund, H., Xie, S. M., Zhang, M., Balsubramani, A., Hu, W., Yasunaga, M., Phillips, R. L., Gao, I., Lee, T., David, E., Stavness, I., Guo, W., Earnshaw, B. A., Haque, I. S., Beery, S., Leskovec, J., Kundaje, A., Pierson, E., et al. (2021). WILDS: A benchmark of in-the-wild distribution shifts. arXiv. https://doi.org/10.48550/arXiv.2012.07421

[75] Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. ACM Computing Surveys, 54(6), Article 115. https://doi.org/10.1145/3457607

[76] Veale, M., Van Kleek, M., & Binns, R. (2018). Fairness and accountability design needs for algorithmic support in high-stakes public sector decision-making. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 1-14. https://doi.org/10.1145/3173574.3174014

[77] Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. Nature Machine Intelligence, 1(11), 501-507. https://doi.org/10.1038/s42256-019-0114-4

[78] Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. California Law Review, 104(3), 671-732. https://doi.org/10.15779/Z38BG31

[79] Kaissis, G. A., Makowski, M. R., Rückert, D., & Braren, R. F. (2020). Secure, privacy-preserving and federated machine learning in medical imaging. Nature Machine Intelligence, 2(6), 305-311. https://doi.org/10.1038/s42256-020-0186-1

[80] Rieke, N., Hancox, J., Li, W., Milletarì, F., Roth, H. R., Albarqouni, S., Bakas, S., Galtier, M. N., Landman, B. A., Maier-Hein, K., Ourselin, S., Sheller, M. J., Summers, R. M., Trask, A., Xu, D., Baust, M., & Cardoso, M. J. (2020). The future of digital health with federated learning. NPJ Digital Medicine, 3, 119. https://doi.org/10.1038/s41746-020-00323-1