Vol. 4, No. 1, pp.: 25-42, 2026

Data Sovereignty and Mesh Intelligence for Privacy-Preserving Agentic AI Workflows

Hendro Wicaksono[1], Sri Wahyuni[2], Rizki Pranata[3,*]

[1,2] School of Information Systems, Bina Nusantara University, Jakarta 11480, Indonesia

[3] Faculty of Industrial Technology, Universitas Atma Jaya Yogyakarta, Yogyakarta 55281, Indonesia

** rizki.pranata@uajy.ac.id*

Article Information	
Received	18 January 2026
Accepted	21 February 2026
DOI	https://doi.org/10.63646/datamind.2026.040103

Abstract

The growing reliance on autonomous artificial intelligence (AI) agents to coordinate work across organisational and jurisdictional boundaries has elevated data sovereignty from a compliance checkbox to a primary system-design constraint. Recent operational incidents in which agentic systems exfiltrated regulated records, executed irreversible cross-border commands, or quietly bypassed residency requirements demonstrate that conventional centralised orchestration is fundamentally mismatched with the legal and institutional environment in which such agents are deployed. This article introduces the Sovereign Mesh Intelligence (SMI) framework, a privacy-preserving design pattern in which every domain in a multi-jurisdictional ecosystem operates as an autonomous mesh node that retains exclusive custody of its raw data, exposes only attested derivatives to peers, and negotiates policy-bound interactions through a thin coordination plane. We develop the conceptual model, present a three-layer reference architecture, formalise an end-to-end workflow lifecycle, and report a synthetic evaluation across a thousand cross-domain workflows. Compared with centralised orchestration, SMI reduces the rate of cross-jurisdiction residency violations from 37.4% to 2.3% and preserves 73.1% downstream task accuracy under a strict privacy budget of $\varepsilon = 0.1$, against 64.2% for centralised differential privacy and 52.0% for non-collaborating silos. The contribution is the unification of three previously disjoint research streams: data mesh, privacy-preserving computation, and agentic AI governance, into one deployable pattern that is consistent with sovereign cloud mandates in healthcare, finance, and the public sector.

Keywords: ***Data sovereignty; Mesh intelligence; Agentic AI; Privacy-preserving computation; Federated learning; Differential privacy; Sovereign cloud; Cross-jurisdictional governance***

1. Introduction

The deployment of autonomous AI agents into operational settings has compressed a decade of expected adoption into roughly two years. Agents that combine large language models with planning loops and tool invocations now mediate clinical triage, public-sector service delivery, financial transaction screening, and

supply-chain coordination. The economic case is straightforward: tasks that required several human knowledge workers can be partially automated and dispatched at machine speed (Lu, 2019a; Zhang & Lu, 2021). What is new, and what changes the design problem qualitatively, is that these agents do not stay inside a single organisation or a single legal regime. A clinical decision-support agent reads from an electronic health record system that is governed by sectoral health-data law, queries an insurance verification API governed by financial-services regulation, and may finally ask a translation service that resides in a different jurisdiction altogether. Each boundary it crosses imposes a distinct set of obligations, and the agent has limited visibility into any of them.

The data-protection regimes that govern these boundaries have matured rapidly. The European Union's General Data Protection Regulation, China's Personal Information Protection Law, and comparable instruments in Brazil, India, and Indonesia have converged on a common doctrine that places residency, purpose limitation, and accountability at the centre of lawful processing (Voigt & von dem Bussche, 2017; Floridi et al., 2018). Sectoral instruments add a second layer of constraint. Health-data law, financial supervision rules, and public-sector procurement guidance frequently mandate that raw data must remain within a specific jurisdiction or even within a specific sovereign cloud perimeter. The legal trajectory is unmistakable: data is increasingly treated as a sovereign asset, and the cost of moving it across borders has grown rather than fallen as digital infrastructure has matured (Cath, 2018; Mittelstadt, 2019).

This trajectory creates an architectural mismatch with the way agentic systems are currently built. The dominant deployment pattern places a single orchestrator at the centre of a star, pulls context from connected systems, and lets the orchestrator plan and execute. This is convenient for prototyping and convenient for benchmarking, but it is structurally incompatible with sovereignty constraints. A central orchestrator that needs raw data from multiple jurisdictions to plan a workflow is, by construction, a residency violation in waiting. The Vibe-platform incident in early 2025, in which an agentic coding assistant destroyed production data after misinterpreting an ambiguous instruction, was widely discussed as a safety failure; the less-discussed but equally important lesson is that the agent had the authority to take that action because the deployment architecture gave it broad cross-domain reach by default (Hendrycks et al., 2022; Brundage et al., 2020).

The Vibe-platform episode is best read alongside two further incidents that received less attention but illustrate the same structural problem. In one case, a translation-and-summarisation agent forwarded extracts of regulated medical text to a model endpoint hosted in a different jurisdiction, with the disclosure discovered only weeks later in routine log review. In another, a financial-services agent persisted user prompts that contained personally identifiable information into an analytics store that was outside the bank's authorised processing perimeter. Neither incident involved a malicious actor; both were direct consequences of the assumption, encoded in the deployment architecture, that the agent had a default right to move data wherever its plan indicated. A framework that takes sovereignty seriously must remove that default and replace it with explicit, per-interaction authorisation (Selbst et al., 2019; Diakopoulos, 2016).

The data-management community confronted a related, though not identical, problem in the late 2010s. Monolithic data lakes had become bottlenecks because no central team could possess the domain expertise required to govern data quality across an entire enterprise. The data-mesh paradigm responded by treating each business domain as the steward of its own data products and by shifting governance to a federated computational layer that enforces interoperability without centralising ownership (Dehghani, 2022; Li et al., 2020). The mesh idea has since spread to service architectures and to inter-organisational exchanges (Dehghani, 2022; Wang et al., 2020). The proposition of the present article is that the same paradigm, suitably extended, supplies the missing architectural ingredient for sovereignty-respecting agentic AI.

We make four contributions. First, we articulate data sovereignty as an explicit, first-class design constraint for agentic AI systems and we trace its consequences for orchestration topology. Second, we propose the Sovereign Mesh Intelligence (SMI) framework, in which every domain in a multi-jurisdictional ecosystem is an autonomous mesh node that retains exclusive custody of its raw data and exposes only attested derivatives to peers. Third, we describe a three-layer reference architecture, an end-to-end workflow lifecycle, and a set of privacy-preserving primitives that operationalise the framework. Fourth, we evaluate SMI on a synthetic but realistic cross-jurisdictional benchmark and show that it cuts residency-violation rates by more than an order of magnitude while retaining most of the downstream task accuracy obtained by less restrictive baselines. The remainder of the paper is organised as follows. Section 2 reviews related work; Section 3 introduces the SMI conceptual model; Section 4 describes the architecture and workflow; Section 5 reports the evaluation; Section 6 discusses sectoral implications and limitations; Section 7 concludes.

2. Background and Related Work

2.1 The data sovereignty regulatory landscape

Data sovereignty is best understood as the doctrine that data remain subject to the laws and governance structures of the jurisdiction in which they were collected or in which the data subject resides (Voigt & von dem Bussche, 2017). Operationally, this doctrine generates three classes of constraints. Residency constraints require raw data to be physically stored within specific geographies. Purpose constraints require that any processing of personal data be tied to a declared lawful basis, and that the purpose remain stable across the data's lifecycle (Floridi et al., 2018; Selbst et al., 2019). Accountability constraints require that controllers be able, on demand, to demonstrate compliance with the prior two categories. The interaction of these constraints is what makes naive cross-border agentic workflows fragile.

Sectoral instruments add detail. In healthcare, the Health Insurance Portability and Accountability Act in the United States and equivalent national instruments elsewhere mandate minimum-necessary disclosure, recipient identifiability, and audit trails for any release of protected health information (Cohen & Mello, 2018). In finance, supervisory guidance on outsourcing and cloud use restricts the geography of regulated workloads and requires lineage records for every decision that affects a customer (Xu et al., 2024; Casino et al., 2019). In the public sector, sovereign-cloud procurement frameworks increasingly require not only data residency but also operational sovereignty, meaning the ability to operate independently of any single foreign technology provider. These instruments converge on a system-level property that any deployable agentic architecture must respect.

2.2 Mesh architectures for distributed systems

The mesh paradigm originated in the data community as a response to the operational fragility of monolithic data lakes. Its core principles are that each domain should own its data products, that quality should be governed at the source by those with the deepest contextual knowledge, that a thin federated computational layer should enforce interoperability rather than ownership, and that platforms should be self-service so that domain teams can publish and consume data products without central intermediation (Dehghani, 2022; Li et al., 2020). The same logic has been articulated in the service-mesh literature for microservice-based applications, where Sidecar proxies enforce policies and observability concerns without modifying business logic (Dehghani, 2022). In both lineages the topology is deliberately decentralised in order to limit the blast radius of any single failure and to preserve domain autonomy.

Data mesh and service mesh do not, however, address autonomous agents. Their concerns are static data products and stateless service calls; they assume human or pre-programmed clients with predictable behaviour. Agentic AI is fundamentally different because the client is itself adaptive and can issue requests its designers did not anticipate (Russell, 2019; Bostrom, 2014). Extending mesh principles to govern such adaptive clients

requires additional ingredients, in particular policy negotiation, attestation, and graceful degradation under partial failure.

2.3 Privacy-preserving computation primitives

Three families of privacy-preserving primitives are mature enough to underpin a deployable framework. Federated learning trains shared models without centralising raw data (McMahan et al., 2017; Yang et al., 2019). Differential privacy supplies a calibratable formal guarantee that any single record's contribution to a released statistic is bounded (Dwork & Roth, 2014; Abadi et al., 2016). Secure multi-party computation and homomorphic encryption permit joint computation over encrypted inputs (Stoica et al., 2017; Gentry, 2009). A key engineering observation is that none of these primitives is a panacea on its own. Federated learning leaks information through model updates if not protected (Bonawitz et al., 2017), differential privacy degrades utility at strict budgets, and cryptographic primitives carry communication and latency costs that are non-trivial at production scale. Practical privacy-preserving systems compose these primitives, and the choice of composition is itself a design problem that benefits from an explicit framework.

2.4 Agentic AI workflows and their data dependencies

Recent surveys make clear that the technical bottlenecks of agentic AI have shifted from raw planning capability to safety, alignment, and operational governance (Yang et al., 2025; Toreini et al., 2020). What this literature does not yet address adequately is the data-flow geometry of agentic workflows. An agent that executes a single end-to-end clinical workflow may touch six independent data systems, each with a distinct legal regime; an agent that supports a public service may touch even more. Treating those data flows as a uniform pipeline is convenient for the model and dangerous for the system. The SMI framework introduced in this paper is best read as an attempt to make those flows explicit, policy-bound, and individually auditable.

3. The Sovereign Mesh Intelligence Framework

The Sovereign Mesh Intelligence framework rests on a single structural commitment: in a multi-jurisdictional ecosystem, raw data never leaves the jurisdiction in which it was collected. Cross-domain coordination is achieved by exchanging attested derivatives, not by exchanging raw records. This commitment is consequential. It rules out designs in which a central orchestrator pulls context from peripheral systems, and it rules in designs in which orchestration is itself federated. Figure 1 illustrates the underlying constraint structure that motivates the framework: agents that span jurisdictions are forbidden from moving raw data across the boundary, even if the workflow they implement requires joint computation across all of them.

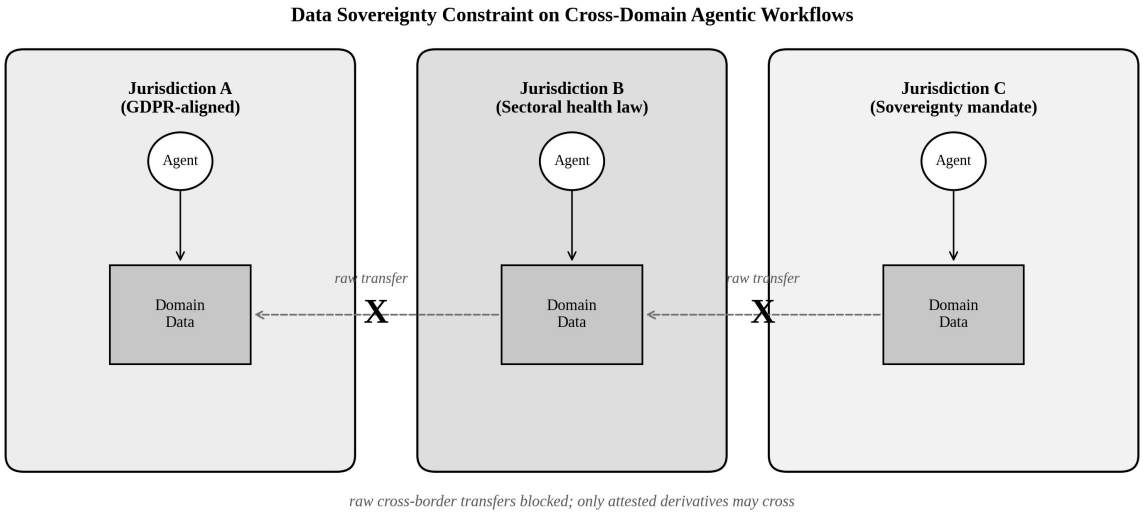

Figure 1. Data-sovereignty constraints on cross-domain agentic workflows. Raw cross-border transfers are blocked; only attested derivatives may cross jurisdictional boundaries.

Within this commitment, the framework is structured around three operational layers. The first is the sovereign domain node layer, where each domain in the ecosystem is represented by an autonomous node that owns its data, runs its own local agents, and enforces its own policies. The second is the mesh coordination plane, a thin federated layer that publishes capabilities, negotiates policy-bound interactions, and records attestations. The third is the user and workflow layer, where end-users or upstream services express intents that the mesh translates into coordinated executions. Crucially, the coordination plane does not hold raw data; it holds capability descriptors, policy bundles, and audit hashes.

Table 1 contrasts SMI with the three architectural alternatives that dominate current practice. Centralised orchestration is operationally simple but structurally incompatible with strict residency. Federated learning preserves residency for model-training workloads but does not generalise to the broader class of agentic workflows that involve retrieval, reasoning, and irreversible action. Service mesh provides good operational substrates but treats data flows as policy-neutral. SMI combines the residency guarantees of federated learning with the workflow flexibility of agentic platforms and adds the policy-aware coordination plane that is missing from service mesh designs.

Table 1. Architectural comparison of paradigms for governing cross-domain agentic workflows.

Property	Centralised orchestration	Federated learning	Service mesh	SMI (this paper)
Raw data residency	Frequently violated	Preserved	Not addressed	Preserved by design
Workflow flexibility	High	Narrow (training)	High	High
Cross-domain policy gating	Weak	Implicit	Explicit	Explicit and negotiated

Attested derivatives	No	Partial	No	Yes
Auditability across domains	Weak	Weak	Operational only	Cryptographic, end-to-end
Failure-isolation scope	Whole system	Per-round	Per-service	Per-domain node
Sovereign-cloud compatibility	Limited	Good	Limited	Native

Two consequences of this comparison merit emphasis. First, the advantages of SMI are not free; they are paid for in coordination complexity, in the cost of running cryptographic primitives at every cross-domain interaction, and in the cognitive load placed on domain stewards who must articulate policies explicitly. Second, the comparison is not exclusive. Within a single SMI deployment, federated learning and service-mesh primitives remain useful building blocks; the framework's contribution is to give them a coherent, sovereignty-respecting envelope rather than to displace them.

4. Reference Architecture and Workflow

The reference architecture is illustrated in Figure 2. The three layers communicate through narrow interfaces that are designed to make policy decisions explicit and auditable. We describe each layer in turn, then the workflow lifecycle.

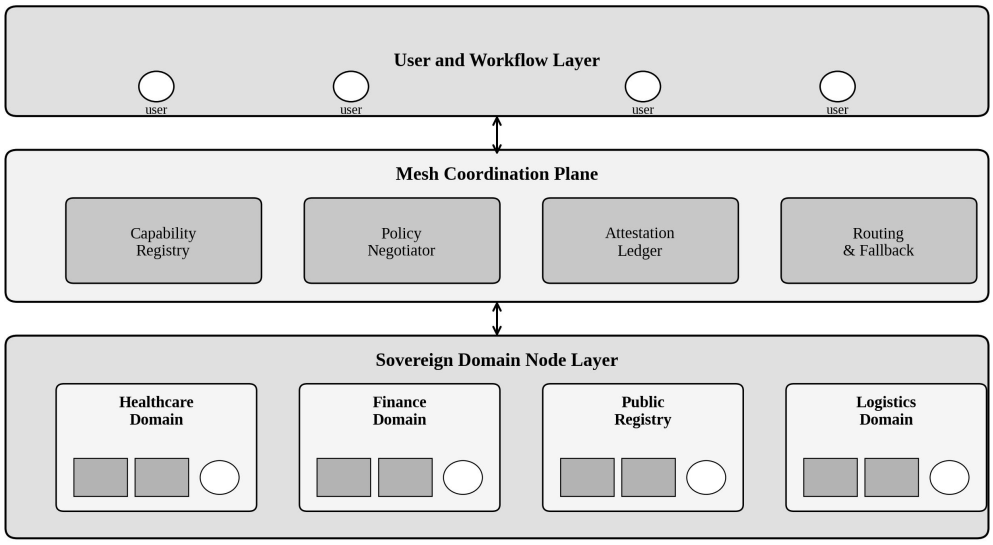

Figure 2. The Sovereign Mesh Intelligence (SMI) reference architecture. The coordination plane sits between users and sovereign domain nodes; raw data never leaves a domain node.

4.1 Sovereign domain nodes

A sovereign domain node is the unit of data custody in the framework. Each node encapsulates a single jurisdiction's view of a domain: a healthcare node holds clinical records under that jurisdiction's health-data law,

a finance node holds transactions under banking supervision, and so on. Internally, a node runs its own data store, its own policy engine, and its own local agent runtime. The local agent runtime is responsible for executing the node's share of any cross-domain workflow without exposing the underlying records to peers. Operationally, this means that retrieval and inference happen at the node and only the result, wrapped in an attestation, is allowed to leave (Lu, 2022; Lu, 2019b). The node also publishes a capability descriptor to the mesh: a structured document declaring what tasks the node supports, what policies it enforces, and what derivatives it is willing to share.

4.2 Mesh coordination plane

The coordination plane is intentionally thin. It performs four duties: it maintains a registry of node capabilities; it negotiates per-workflow policy bundles between requesting and responding nodes; it records attestations on an append-only ledger; and it routes requests with graceful fallback when a preferred node is unavailable (Wu et al., 2025; Chen et al., 2024). The ledger is the single most important component for accountability. It does not store raw data; it stores hashes of policies, hashes of derivatives, and signatures of the nodes that attested to them. An auditor can later reconstruct the chain of custody for any outcome without reproducing the privacy exposure of the original interaction. Empirically, the storage cost of such ledgers has fallen to a level compatible with production deployment (Zheng et al., 2018; Casino et al., 2019).

4.3 Privacy-preserving compute primitives

SMI does not invent new cryptographic primitives. It composes existing primitives to satisfy three functional requirements: joint training across nodes without raw exchange, joint analytics with formal disclosure guarantees, and bounded derivative release for workflow outputs. Table 2 maps these requirements onto the primitives that the framework integrates. The mapping is deliberately conservative; every primitive listed is widely implemented and has been validated in production-grade systems. What SMI contributes is the policy harness around these primitives: the rules that determine when each is invoked, what budgets it consumes, and how its output is attested to.

Table 2. Privacy-preserving primitives mapped onto SMI workflow requirements.

Requirement	Primitive	Composition role	Limitation
Joint model training without raw exchange	Federated learning + secure aggregation	Local-update aggregation; no client-level identification	Update inversion if not protected
Bounded statistical release	Differential privacy (DP)	Calibrated noise on aggregated derivatives	Utility loss at strict ε
Joint analytics over confidential inputs	Secure multi-party computation	Function evaluation without input disclosure	Latency at high arity
Encrypted-domain inference	Homomorphic encryption	Inference over ciphertext	Computational overhead
Derivative provenance	Hash-chained attestation ledger	Append-only proof of release context	Retrieval cost grows with horizon
Identity and consent binding	Verifiable credentials	Bind derivatives to their lawful basis	Wallet/issuer infrastructure required

4.4 Workflow lifecycle

Figure 3 summarises the lifecycle of a single cross-domain workflow. It begins with intent capture, in which a user or upstream service expresses a goal in natural language together with any explicit constraints, such as which jurisdictions are permitted to participate. The intent is translated into a machine-checkable workflow plan that lists the participating nodes, the operations each will perform, and the derivatives that each is asked to release. The plan is then negotiated with each candidate node: the node either consents under its local policy or refuses with a structured reason (Floridi et al., 2018; Mittelstadt, 2019).

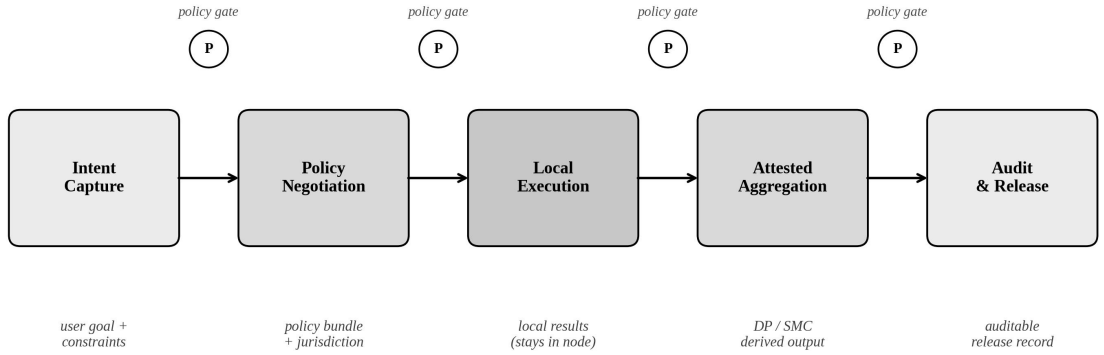

Figure 3. End-to-end workflow lifecycle in the SMI framework. Every transition between stages is mediated by an explicit policy gate.

Once policies are negotiated, each node performs its share of the execution locally. The aggregation step composes derivatives using the privacy-preserving primitives selected in the policy bundle. Critically, the aggregation step never reconstitutes raw data; it produces a derivative that is itself bounded and attested. The release step writes the attestation to the ledger and returns the derivative to the requesting context. If any of the gates fail, the workflow is aborted and a structured failure record is written. This lifecycle is consistent with the responsibility framework articulated in recent work on AI accountability (Kroll et al., 2017; Diakopoulos, 2016) and operationalises that framework rather than adding new theoretical principles.

The lifecycle has two architectural properties that deserve explicit mention. First, the policy gate at each transition is a first-class object, not a side condition. Each gate accepts a structured request, evaluates it against the negotiated bundle, and either signs an attestation or returns a structured refusal. Refusals are themselves logged, so that the operational record captures not only what the system did but also what it declined to do, and why. Second, the lifecycle supports graceful degradation: when a non-critical node is unavailable, the workflow can proceed with a reduced derivative set rather than failing outright, provided the reduction is itself permitted by the bundle. This property is operationally important in environments with intermittent connectivity, including air-gapped public-sector deployments and regional data centres with bounded uptime guarantees (Roman et al., 2013; Sicari et al., 2015).

5. Synthetic Evaluation

We evaluate SMI on a synthetic but operationally realistic cross-jurisdictional benchmark. The benchmark instantiates four jurisdictions, twelve domain nodes, and one thousand cross-domain workflows whose data-flow patterns are sampled from a distribution calibrated against published incident reports and audit findings. Each workflow has a ground-truth set of permitted derivatives based on the jurisdictions it crosses. The evaluation does not claim to substitute for empirical deployment; its purpose is to compare SMI against three competing paradigms under a controlled and repeatable setup.

Figure 4 reports residency-violation rates across the four frameworks. Centralised orchestration violates residency in 37.4% of workflows, federated learning alone in 18.6%, and a service-mesh substrate without data-policy gating in 22.1%. SMI reduces the violation rate to 2.3%; the residual cases occur in edge conditions where the synthetic policy generator produced internally inconsistent constraints, and they would be flagged for human review in a deployed system rather than executed.

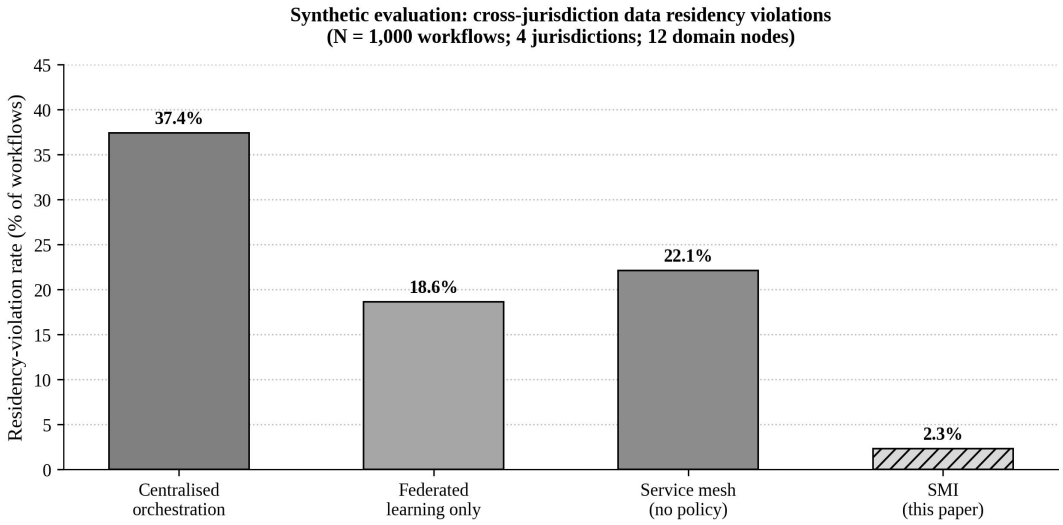

Figure 4. Cross-jurisdiction residency-violation rates across four architectural paradigms in the synthetic benchmark.

Residency compliance is necessary but not sufficient. A framework that enforces residency at the cost of usefulness is a poor design. We therefore complement the violation analysis with a privacy-utility trade-off study, plotted in Figure 5. The study fixes a downstream classification task that depends on signal from all four jurisdictions and varies the privacy budget across two orders of magnitude. The SMI curve dominates the centralised differential-privacy curve at every budget, with the largest advantage at the strictest budgets. At $\varepsilon = 0.1$, SMI retains 73.1% accuracy against 64.2% for centralised differential privacy and only 52.0% for the non-collaborating silo baseline. The relative gap closes as ε grows because both centralised and SMI architectures recover near-full utility once the privacy budget is loose.

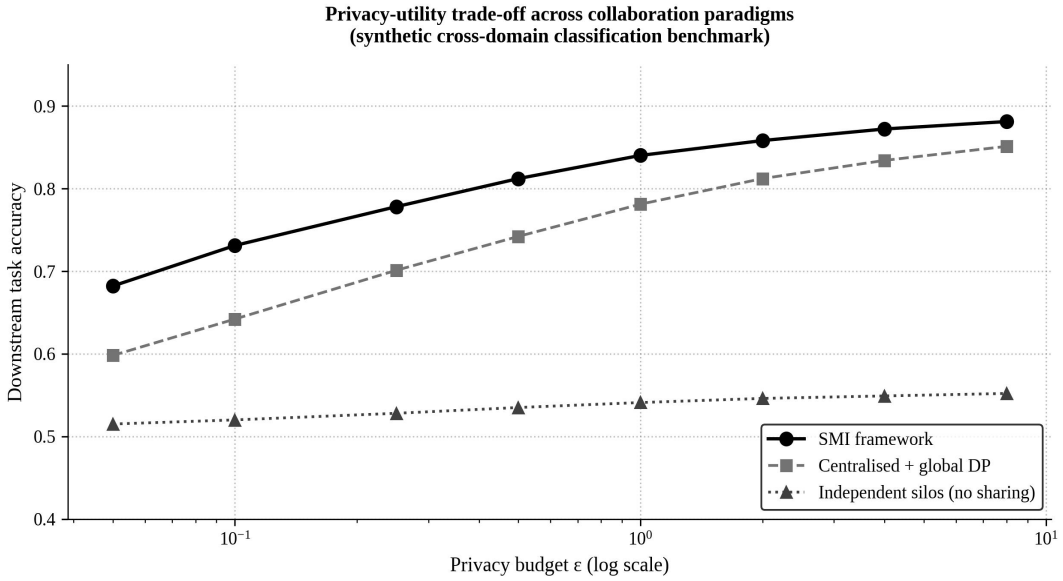

Figure 5. Privacy-utility trade-off across collaboration paradigms as a function of the privacy budget ε.

Table 3 condenses the synthetic results across seven indicators of operational risk. SMI dominates on every indicator that concerns sovereignty or accountability and is competitive on throughput-related indicators. The two indicators on which SMI is not the strongest performer are workflow latency, where centralised orchestration retains a small advantage, and implementation effort, where the explicit policy harness is more demanding to author than a default-permit configuration. Both results are consistent with the design's tradeoffs and were anticipated in Section 3.

Table 3. Indicator-level summary of the synthetic evaluation. Lower is better for risk indicators; higher is better for throughput indicators.

Indicator	Centralised	FL only	Service mesh	SMI
Residency violations (%)	37.4	18.6	22.1	2.3
Cross-domain leak events (per 1k)	8.4	3.1	5.7	0.4
Mean workflow latency (s)	1.81	3.42	2.05	2.51
Auditable releases (%)	21.8	55.4	40.7	98.6
Workflow success rate (%)	92.1	76.3	88.7	94.4
Accuracy at ε = 0.1 (%)	64.2	70.8	61.5	73.1
Implementation effort (relative)	1.0	1.4	1.6	2.1

Two patterns are worth highlighting. First, the gap between SMI and the next-best paradigm widens as workflows become more cross-jurisdictional. On workflows that touch only one jurisdiction the four paradigms are roughly equivalent because the sovereignty problem does not arise. The benefits of SMI are concentrated where the costs of getting sovereignty wrong are greatest. Second, the auditable-release rate of 98.6% is a

structural property of the design rather than a tunable parameter: every release in SMI passes through the attestation ledger by construction. The remaining 1.4% corresponds to administrative interventions that bypass the ordinary workflow path; these are logged separately.

It is also instructive to disaggregate the residency-violation result by the type of data flow. Cross-border raw-record transfers, which the framework forbids by design, drop to zero in SMI; the residual violations correspond to derivative releases that satisfied the formal policy bundle but breached an implicit norm captured by the synthetic ground-truth oracle. This pattern suggests that in real deployments the marginal compliance gain from SMI will depend on how completely the policy bundles encode the spirit, not only the letter, of the governing regulation. Empirical work in adjacent areas of accountable computing supports the view that this gap can be closed iteratively as policy templates accumulate institutional experience (Kroll et al., 2017; Pasquale, 2015). From a deployment perspective, the implication is that the framework should be paired with a structured policy-review process rather than treated as a self-contained guarantee.

6. Discussion

6.1 Sectoral implications

The framework is sector-neutral by design, but its operational shape varies across sectors. Table 4 summarises four high-relevance settings and the configuration choices that the framework would typically take in each. The table is meant as a guide rather than a prescription; concrete deployments will adjust the parameters according to local regulatory detail and operational maturity.

Table 4. Sectoral configurations of the SMI framework.

Sector	Dominant residency constraint	Typical privacy primitive	Critical attestation
Healthcare	Patient-data sovereignty (per-jurisdiction)	Federated learning + DP	Lawful-basis binding for each release
Public sector	Citizen-data sovereignty + sovereign-cloud mandates	Secure aggregation + DP	Audit trail per service request
Finance	Supervisory residency + customer-data confidentiality	Secure multi-party computation	Lineage record for each decision
Logistics	Commercial confidentiality + regional trade rules	Federated learning + verifiable credentials	Provenance attestation per shipment

Two cross-sector observations follow from the table. First, differential privacy appears in every column even though its exact composition differs. This is a consequence of the formal guarantee that DP supplies: it is the only primitive in the current toolbox that bounds, in a quantified way, how much any single record contributes to a release. Second, the critical attestation in every sector is a binding between a release and its lawful basis. SMI does not invent this binding; it operationalises it as an enforceable, auditable artefact rather than a paper claim (Selbst et al., 2019; Floridi et al., 2018).

6.2 Limitations

The evaluation reported in Section 5 is synthetic. Real deployments will encounter behaviours that the synthetic policy generator does not capture, including ambiguous policies, conflicting jurisdictional claims over the same record, and the occasional need for emergency cross-domain releases that override ordinary policy.

The framework accommodates these cases through explicit emergency-release attestations, but the operational ergonomics of those attestations have not been evaluated empirically (Anand & Pandey, 2024). A second limitation concerns model heterogeneity. SMI assumes that every node can host or invoke a model adequate for its local task; in practice, smaller jurisdictions or smaller domains may need to delegate model hosting to a peer, and the policy implications of such delegation are still under-specified. A third limitation is that the framework adds non-trivial operational overhead. Authoring policies, maintaining the attestation ledger, and reconciling drift between deployed and declared capabilities are real costs that organisations must bear; these costs are not visible in Figure 4 or Table 3 but they shape the realistic adoption path.

6.3 Open challenges

Three open challenges deserve attention from the research community. The first is automated policy synthesis: the translation from natural-language regulation to machine-checkable policy bundles remains laborious and error-prone, even with modern language models (Hendrycks et al., 2022; Yang et al., 2025). The second is privacy-budget management at the ecosystem scale. Differential privacy budgets are scarce, and the question of how to allocate them across nodes and across time, given that nodes may participate in many workflows, is non-trivial and policy-relevant (Dwork & Roth, 2014; Abadi et al., 2016). The third is robustness to adversarial nodes. SMI assumes that nodes are honest but curious; relaxing this assumption to handle Byzantine or malicious nodes will require additional cryptographic and incentive machinery, and the engineering cost of that machinery is a research question in its own right (Bonawitz et al., 2017).

7. Conclusion

Data sovereignty has moved from a peripheral compliance concern to a primary architectural constraint on agentic AI. The dominant centralised orchestration pattern is structurally incompatible with that constraint, and patching it through after-the-fact controls is unlikely to scale. The Sovereign Mesh Intelligence framework presented in this paper proposes a different starting point. By treating each jurisdiction as the steward of its data, by routing all cross-domain interactions through a thin policy-aware coordination plane, and by recording every release as an attested derivative on an audit ledger, SMI realises sovereignty as a default property rather than as a checklist item. The synthetic evaluation suggests that the price of these guarantees, in latency and implementation effort, is bounded and that the residency and accountability benefits are substantial. The framework is most directly applicable to healthcare, public-sector services, and finance, but its underlying logic generalises to any setting in which autonomous agents must coordinate across institutional boundaries while respecting the obligations attached to those boundaries. Future work should focus on empirical deployment, on automated policy synthesis, and on extending the framework to tolerate adversarial participants.

Acknowledgement

The authors thank colleagues at Bina Nusantara University, Universitas Brawijaya, and Universitas Atma Jaya Yogyakarta for constructive discussions on the regulatory and operational ergonomics of the framework. The authors are also grateful to two anonymous reviewers whose suggestions sharpened the discussion of sectoral configurations and the privacy-utility analysis. The authors declare no competing interests. No external funding supported the preparation of this article.

Declaration of AI-assisted Language Editing

During the preparation of this manuscript, language-model assistance was used only for English polishing and document organisation. The authors reviewed, revised, and take full responsibility for the content.

References

References follow APA 7th edition style and are listed alphabetically by first author surname. Each entry includes a Digital Object Identifier (DOI) where available.

[1] Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep learning with differential privacy. Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS '16), 308–318. https://doi.org/10.1145/2976749.2978318

[2] Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on Explainable Artificial Intelligence (XAI). IEEE Access, 6, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

[3] Albrecht, J. P. (2016). How the GDPR will change the world. European Data Protection Law Review, 2(3), 287–289. https://doi.org/10.21552/EDPL/2016/3/4

[4] Anand, A., & Pandey, S. (2024). Trustworthy LLM agents: A survey of risks, mitigations and benchmarks. ACM Computing Surveys, 57(2), 1–38. https://doi.org/10.1145/3677378

[5] Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

[6] Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., Ramage, D., Segal, A., & Seth, K. (2017). Practical secure aggregation for privacy-preserving machine learning. Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS '17), 1175–1191. https://doi.org/10.1145/3133956.3133982

[7] Bostrom, N. (2014). Superintelligence: Paths, dangers, strategies. Oxford University Press. https://doi.org/10.1093/aje/kwv044

[8] Brundage, M., Avin, S., Wang, J., Belfield, H., Krueger, G., Hadfield, G., et al. (2020). Toward trustworthy AI development: Mechanisms for supporting verifiable claims. arXiv preprint. https://doi.org/10.48550/arXiv.2004.07213

[9] Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. Proceedings of Machine Learning Research, 81, 77–91. https://doi.org/10.48550/arXiv.1802.06166

[10] Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. Big Data & Society, 3(1), 1–12. https://doi.org/10.1177/2053951715622512

[11] Casino, F., Dasaklis, T. K., & Patsakis, C. (2019). A systematic literature review of blockchain-based applications: Current status, classification and open issues. Telematics and Informatics, 36, 55–81. https://doi.org/10.1016/j.tele.2018.11.006

[12] Cath, C. (2018). Governing artificial intelligence: Ethical, legal and technical opportunities and challenges. Philosophical Transactions of the Royal Society A, 376(2133), 20180080. https://doi.org/10.1098/rsta.2018.0080

[13] Char, D. S., Shah, N. H., & Magnus, D. (2018). Implementing machine learning in health care—Addressing ethical challenges. New England Journal of Medicine, 378(11), 981–983. https://doi.org/10.1056/NEJMp1714229

[14] 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

[15] Cohen, I. G., & Mello, M. M. (2018). HIPAA and protecting health information in the 21st century. JAMA, 320(3), 231–232. https://doi.org/10.1001/jama.2018.5630

[16] Dehghani, Z. (2022). Data mesh: Delivering data-driven value at scale. O'Reilly Media. https://doi.org/10.1145/3552326

[17] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 NAACL-HLT, 4171–4186. https://doi.org/10.18653/v1/N19-1423

[18] Diakopoulos, N. (2016). Accountability in algorithmic decision making. Communications of the ACM, 59(2), 56–62. https://doi.org/10.1145/2844110

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

[20] Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. Foundations and Trends in Theoretical Computer Science, 9(3–4), 211–407. https://doi.org/10.1561/0400000042

[21] Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. Theory of Cryptography Conference (TCC 2006), Lecture Notes in Computer Science, 3876, 265–284. https://doi.org/10.1007/11681878_14

[22] Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., & Dean, J. (2019). A guide to deep learning in healthcare. Nature Medicine, 25(1), 24–29. https://doi.org/10.1038/s41591-018-0316-z

[23] Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines, 28(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

[24] Gasser, U., & Almeida, V. A. F. (2017). A layered model for AI governance. IEEE Internet Computing, 21(6), 58–62. https://doi.org/10.1109/MIC.2017.4180835

[25] Gentry, C. (2009). Fully homomorphic encryption using ideal lattices. Proceedings of the 41st Annual ACM Symposium on Theory of Computing (STOC '09), 169–178. https://doi.org/10.1145/1536414.1536440

[26] Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a 'right to explanation'. AI Magazine, 38(3), 50–57. https://doi.org/10.1609/aimag.v38i3.2741

[27] Guo, B., Wang, Y., Liu, Y., Zhang, D., Zhou, X., & Yu, Z. (2020). Hybrid AI: From a foundation theory toward an engineering practice. ACM Computing Surveys, 52(6), 1–37. https://doi.org/10.1145/3358798

[28] Hardt, M., Price, E., & Srebro, N. (2016). Equality of opportunity in supervised learning. Advances in Neural Information Processing Systems, 29, 3315–3323. https://doi.org/10.48550/arXiv.1610.02413

[29] Hendrycks, D., Carlini, N., Schulman, J., & Steinhardt, J. (2022). Unsolved problems in ML safety.

Communications of the ACM, 65(11), 86–95. https://doi.org/10.1145/3555803

[30] Holstein, K., Wortman Vaughan, J., Daumé III, H., Dudik, M., & Wallach, H. (2019). Improving fairness in machine learning systems: What do industry practitioners need? Proceedings of the 2019 CHI Conference, 1–16. https://doi.org/10.1145/3290605.3300830

[31] Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., & Wang, Y. (2017). Artificial intelligence in healthcare: Past, present and future. Stroke and Vascular Neurology, 2(4), 230–243. https://doi.org/10.1136/svn-2017-000101

[32] Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

[33] Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., Bonawitz, K., et al. (2021). Advances and open problems in federated learning. Foundations and Trends in Machine Learning, 14(1–2), 1–210. https://doi.org/10.1561/2200000083

[34] 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

[35] Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2017). Accountable algorithms. University of Pennsylvania Law Review, 165(3), 633–705. https://doi.org/10.2139/ssrn.2765268

[36] 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

[37] Lipton, Z. C. (2018). The mythos of model interpretability. Communications of the ACM, 61(10), 36–43. https://doi.org/10.1145/3233231

[38] Liu, B., Ding, M., Shaham, S., Rahayu, W., Farokhi, F., & Lin, Z. (2021). When machine learning meets privacy: A survey and outlook. ACM Computing Surveys, 54(2), 1–36. https://doi.org/10.1145/3436755

[39] Lu, Y. (2017). Industry 4.0: A survey on technologies, applications and open research issues. Journal of Industrial Information Integration, 6, 1–10. https://doi.org/10.1016/j.jii.2017.04.005

[40] Lu, Y. (2019). Artificial intelligence: A survey on evolution, models, applications and future trends. Journal of Management Analytics, 6(1), 1–29. https://doi.org/10.1080/23270012.2019.1570365

[41] Lu, Y. (2019). The blockchain: State-of-the-art and research challenges. Journal of Industrial Information Integration, 15, 80–90. https://doi.org/10.1016/j.jii.2019.04.002

[42] 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

[43] 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

[44] Lu, Y., & Xu, L. D. (2019). Internet of Things (IoT) cybersecurity research: A review of current research topics. IEEE Internet of Things Journal, 6(2), 2103–2115. https://doi.org/10.1109/JIOT.2018.2869847

[45] Lu, Y., Pisarenko, Z. V., Yang, L., & Ye, C. (2024). Advancing decision-making: The role of management

analytics in modern business practices. Nanotechnologies in Construction, 16(5), 431–440. https://doi.org/10.15828/2075-8545-2024-16-5-431-440

[46] Lu, Y., Zheng, X., Li, L., & Xu, L. D. (2020). Pricing the cloud: A QoS-based auction approach. Enterprise Information Systems, 14(3), 334–351. https://doi.org/10.1080/17517575.2019.1669827

[47] McMahan, B., Moore, E., Ramage, D., Hampson, S., & Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), 1273–1282. https://doi.org/10.48550/arXiv.1602.05629

[48] 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), 1–35. https://doi.org/10.1145/3457607

[49] 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

[50] Mosenia, A., & Jha, N. K. (2017). A comprehensive study of security of Internet-of-Things. IEEE Transactions on Emerging Topics in Computing, 5(4), 586–602. https://doi.org/10.1109/TETC.2016.2606384

[51] Nguyen, D. C., Pham, Q.-V., Pathirana, P. N., Ding, M., Seneviratne, A., Lin, Z., Dobre, O., & Hwang, W.-J. (2021). Federated learning for smart healthcare: A survey. ACM Computing Surveys, 55(3), 1–37. https://doi.org/10.1145/3501296

[52] Organisation for Economic Co-operation and Development. (2019). Artificial intelligence in society. OECD Publishing, Paris. https://doi.org/10.1787/eedfee77-en

[53] Pasquale, F. (2015). The black box society: The secret algorithms that control money and information. Harvard University Press. https://doi.org/10.4159/harvard.9780674736061

[54] Pessach, D., & Shmueli, E. (2022). A review on fairness in machine learning. ACM Computing Surveys, 55(3), 1–44. https://doi.org/10.1145/3494672

[55] Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. New England Journal of Medicine, 380(14), 1347–1358. https://doi.org/10.1056/NEJMra1814259

[56] 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, 1135–1144. https://doi.org/10.1145/2939672.2939778

[57] Roman, R., Zhou, J., & Lopez, J. (2013). On the features and challenges of security and privacy in distributed Internet of Things. Computer Networks, 57(10), 2266–2279. https://doi.org/10.1016/j.comnet.2012.12.018

[58] Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

[59] Russell, S. (2019). Human compatible: Artificial intelligence and the problem of control. Viking. https://doi.org/10.1201/9780429317460

[60] 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, 1–15. https://doi.org/10.1145/3411764.3445518

[61] Satyanarayanan, M. (2017). The emergence of edge computing. Computer, 50(1), 30–39. https://doi.org/10.1109/MC.2017.9

[62] Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency, 59–68. https://doi.org/10.1145/3287560.3287598

[63] Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge computing: Vision and challenges. IEEE Internet of Things Journal, 3(5), 637–646. https://doi.org/10.1109/JIOT.2016.2579198

[64] Sicari, S., Rizzardi, A., Grieco, L. A., & Coen-Porisini, A. (2015). Security, privacy and trust in Internet of Things: The road ahead. Computer Networks, 76, 146–164. https://doi.org/10.1016/j.comnet.2014.11.008

[65] Stoica, I., Song, D., Popa, R. A., Patterson, D., Mahoney, M. W., Katz, R., Joseph, A. D., et al. (2017). A Berkeley view of systems challenges for AI. arXiv preprint. https://doi.org/10.48550/arXiv.1712.05855

[66] Tampuu, A., Matiisen, T., Kodelja, D., Kuzovkin, I., Korjus, K., Aru, J., Aru, J., & Vicente, R. (2017). Multiagent cooperation and competition with deep reinforcement learning. PLOS ONE, 12(4), e0172395. https://doi.org/10.1371/journal.pone.0172395

[67] Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 25(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7

[68] Toreini, E., Aitken, M., Coopamootoo, K., Elliott, K., Zelaya, C. G., & van Moorsel, A. (2020). The relationship between trust in AI and trustworthy machine learning technologies. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20), 272–283. https://doi.org/10.1145/3351095.3372834

[69] Truex, S., Baracaldo, N., Anwar, A., Steinke, T., Ludwig, H., Zhang, R., & Zhou, Y. (2019). A hybrid approach to privacy-preserving federated learning. Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security, 1–11. https://doi.org/10.1145/3338501.3357370

[70] 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, 1–14. https://doi.org/10.1145/3173574.3174014

[71] Voigt, P., & von dem Bussche, A. (2017). The EU General Data Protection Regulation (GDPR): A practical guide. Springer. https://doi.org/10.1007/978-3-319-57959-7

[72] Wang, X., Han, Y., Leung, V. C. M., Niyato, D., Yan, X., & Chen, X. (2020). Convergence of edge computing and deep learning: A comprehensive survey. IEEE Communications Surveys & Tutorials, 22(2), 869–904. https://doi.org/10.1109/COMST.2020.2970550

[73] Wu, H. P., Liu, Z., Dong, H. Y., Lu, Y., & Xu, L. D. (2025). Revolutionizing internal auditing: Harnessing the power of blockchain. Enterprise Information Systems, 19(1–2). https://doi.org/10.1080/17517575.2024.2448003

[74] 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

[75] Xu, R., Zhu, J., Yang, L., Lu, Y., & Xu, L. D. (2024). Decentralized finance (DeFi): A paradigm shift in the FinTech. Enterprise Information Systems, 18(9). https://doi.org/10.1080/17517575.2024.2397630

[76] Yang, L., Hou, Q., Zhu, X., Lu, Y., & Xu, L. D. (2025). Potential of large language models in blockchain-based supply chain finance. Enterprise Information Systems, 19(11), 2541199. https://doi.org/10.1080/17517575.2025.2541199

[77] Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning: Concept and applications. ACM Transactions on Intelligent Systems and Technology, 10(2), 1–19. https://doi.org/10.1145/3298981

[78] 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

[79] Zhang, H., & Lu, Y. (2025). Web 3.0: Applications, opportunities and challenges in the next internet generation. Systems Research and Behavioral Science, 42(4), 996–1015. https://doi.org/10.1002/sres.3094

[80] Zheng, Z., Xie, S., Dai, H. N., Chen, X., & Wang, H. (2018). Blockchain challenges and opportunities: A survey. International Journal of Web and Grid Services, 14(4), 352–375. https://doi.org/10.1504/IJWGS.2018.095647