

Learn Together, Keep Data Apart: Federated Models for Utilities

A decision framework for collaborative utility AI without centralising sensitive infrastructure data

Authored by Chennakeshav Adya

Independent Researcher

October 2026

ABSTRACT

Utilities hold valuable operating data across control centres, substations, generation fleets, networks, retailers, campuses and customer devices. Centralising these data can be legally restricted, commercially unacceptable, operationally risky or technically expensive. Federated learning offers a different architecture: participating organisations or sites train a shared model by exchanging controlled model information while retaining raw data locally. The architecture can expand the evidence available to a model, but it does not by itself provide privacy, security, accuracy or commercial value. This paper develops a Federated Utility Model Gate for deciding when collaborative learning is preferable to local or centralised alternatives. It examines load and renewable forecasting, asset-health analytics, fault and anomaly detection, non-intrusive load monitoring, distributed-energy coordination and cross-utility benchmarking. The framework tests decision value, data partitions, model portability, non-independent data, communications, privacy leakage, poisoning, secure aggregation, operational-technology boundaries, explainability, validation, contractual rights and total operating cost. The evidence supports bounded pilots rather than automatic deployment. Published studies demonstrate that federated approaches can perform competitively on selected energy datasets, while results remain sensitive to client heterogeneity, aggregation, participation, privacy controls and experimental design. A utility should therefore compare local, centralised and federated options on identical operating cases; conduct privacy and adversarial testing; run in shadow mode; and advance only when attributable operating value exceeds integration, security and governance cost. Three figures and three tables use hypothetical author-developed assumptions to demonstrate the decision mechanics. They are not forecasts, market quotations, vendor-performance claims or investment recommendations.

JEL Classification: C45, L94, O32, Q40, Q48

Keywords: federated learning, utilities, smart grids, privacy, operational technology, load forecasting, distributed energy resources, model governance, cybersecurity, utility strategy

Introduction

Electricity, water and district-energy systems increasingly depend on data produced by assets and customers. Smart meters reveal consumption patterns. Supervisory control and data-acquisition systems record physical states. Advanced distribution-management systems combine network models, telemetry and switching plans. Generators, storage operators and distributed-energy aggregators create forecasts, bids and operating schedules. These data can improve prediction and decision support, but their value is fragmented across organisational and technical boundaries.

The simplest machine-learning architecture collects data in one environment. That design can support uniform cleaning, training and monitoring. It also creates a concentration of sensitive information, a large transfer programme and a demanding entitlement model. Some participants cannot move raw data because of law, licence conditions, contracts, customer expectations, critical-infrastructure controls or competitive concerns. Others can move data but lack a consistent schema or reliable network connection.

Federated learning trains models across distributed data holders. A coordinator provides a model or training task. Participants train locally and send model updates or other controlled information. The coordinator aggregates the contributions and returns an updated model. Variants support cross-silo organisations, large populations of edge devices, horizontal partitions with similar features, vertical partitions with different features for overlapping entities, personalised models and decentralised aggregation.

The central conclusion of this paper is that federated learning is an operating model, not a privacy certificate. Raw data remaining local removes one transfer path while model updates can still reveal information, accept malicious contributions or fail under heterogeneous operating conditions. The utility needs a decision-specific architecture, a credible baseline and a governed evidence chain from data and model to operating outcome.

1 Define the operating decision

A programme should begin with a named decision. “Use data without sharing it” is an architectural intention. “Improve next-day feeder-load forecasts across five distribution regions while customer-level interval data remain within each region” is a testable operating decision. “Detect transformer anomalies across a fleet owned by several operating companies without exporting waveform histories” is another.

The decision record should specify the owner, prediction target, intervention, cadence, tolerance, consequence of error and fallback. A forecasting model may change procurement, reserves or network operations. An asset-health model may change inspection or replacement. A fraud model may trigger an investigation. A model that influences an operational action requires stronger validation and authority controls than one used for exploratory planning.

Decision value should be measured in the native process. Forecast error matters when it changes imbalance cost, reserve procurement, curtailment, service quality or study throughput. A fault alert matters when it changes outage duration, safety exposure or maintenance cost. Technical accuracy without an executable intervention has limited commercial value.

2 Select among local, centralised and federated learning

Federation should earn its place against simpler alternatives. A local model may be sufficient when each participant has enough representative data and cross-participant learning adds little. A centralised model may be preferable when transfers are lawful, secure and economical, and when uniform data engineering materially improves quality. A federated model is attractive when pooled learning has value and raw-data movement creates a binding constraint.

The comparison should preserve the same model family, evaluation periods and operating cases. The local baseline measures what each participant can achieve alone. The centralised baseline estimates the performance ceiling under pooled data, subject to lawful and practical availability. The federated candidate shows how much pooled learning survives distribution, communication limits and privacy controls.

The utility should also test a shared pretrained model with local fine-tuning. This can capture common patterns without continuous federation. It may have lower operating complexity. Synthetic data, clean rooms, trusted research environments and conventional data-sharing agreements provide further alternatives. The decision should reflect value, risk and lifecycle cost rather than architectural novelty.

Table 1. Architecture choice for collaborative utility models

Architecture	Suitable condition	Principal benefit	Principal limitation	Minimum evidence
Local model	Each participant has sufficient representative data	Simple control and limited external dependence	Weak learning for rare events or small datasets	Stable local holdout results and operating value
Centralised model	Raw-data pooling is lawful, secure and economical	Uniform engineering and complete visibility	Transfer, concentration and entitlement exposure	Approved data basis, secure platform and comparative performance
Federated model	Cross-participant learning is valuable and raw-data movement is constrained	Shared learning with local data custody	Update leakage, heterogeneity, poisoning and operating complexity	End-to-end benchmark, privacy test, attack test and cost model
Pretrain and fine-tune	Common historic corpus exists and local adaptation is periodic	Lower coordination burden	Model staleness and weaker continuous learning	Transferability and refresh evidence
Clean room or trusted environment	Joint analysis is allowed under controlled access	Richer analysis than update exchange alone	Governance, platform and query constraints	Contract, entitlement and disclosure controls

The table is an author-developed decision framework. The appropriate design depends on the specific data, law, operating process and risk appetite.

3 Identify the federation structure

Cross-silo federation fits a limited number of utilities, operating companies, plants or network regions with reliable infrastructure and identifiable legal entities. Cross-device federation fits large populations of meters, chargers, inverters or appliances with intermittent participation and constrained hardware. Their threat models, communications and governance differ.

Horizontal federation assumes participants hold similar features for different observations. Several distribution regions may each hold weather, calendar and load histories. Vertical federation assumes participants hold different features for overlapping entities. A utility, retailer and building operator might each hold different information about the same account or site. Vertical designs require entity matching, purpose controls and a more complex protocol.

Personalised federation allows a common representation with local parameters or fine-tuning. This matters because feeder, climate, tariff, asset and customer populations differ. A single global model can underperform in distinctive regions even when average performance improves. The architecture should define what is shared, what remains local and who can change each component.

4 Create a data-partition map

Federated learning does not remove the need for data engineering. It distributes it. Each participant should document sources, sampling, units, timestamps, missingness, topology, events, labels, retention, lawful purpose and quality controls. A model trained across inconsistent definitions can aggregate disagreement rather than knowledge.

The partition map should distinguish observations, features, labels and populations. A load-forecasting federation may divide data by geography. An asset model may divide data by owner or manufacturer. Rare fault labels may cluster at one participant. Renewable output may differ by climate and technology. These differences create non-independent and non-identically distributed data, commonly described as non-IID.

The programme should quantify client imbalance, feature drift and label prevalence. It should record whether every client can calculate the same transformation and loss function. Shared feature definitions, versioned code and local quality reports are essential. The coordinator should not need raw records to know whether a participant trained on the wrong unit, time zone or asset population.

5 Treat non-IID data as an operating fact

Federated Averaging provides an important baseline by combining local updates according to defined weights. Its assumptions can be strained when participant data differ materially. Local training can move models toward each client's distribution, producing client drift and unstable aggregation. Participation frequency and dataset size can allow large clients to dominate.

FedProx adds a proximal term intended to limit divergence from the global model. SCAFFOLD uses control variates to reduce client drift. Clustered and personalised approaches seek separate models for groups or clients. Robust aggregation can reduce the influence of outliers. Each method changes computation, communications, explainability and attack exposure.

The benchmark should report global and client-level results. An average improvement can conceal deterioration for a small region or vulnerable customer group. Tail performance matters for rare weather, outage and asset conditions. A deployment decision should state the minimum acceptable client performance and the process for clients that fail it.

6 Prioritise use cases with shared signal and local constraints

Load forecasting is a natural starting point because regions share calendar, weather and behavioural patterns while retaining local differences. Renewable-generation forecasting can learn across geographically distributed plants. Asset-health models can learn from failure signatures across fleets when individual owners have few events. Fault and anomaly detection can benefit from diverse operating conditions.

Non-intrusive load monitoring seeks appliance-level patterns from aggregate consumption. Research has explored lightweight personalised federated approaches for resource-constrained edge devices. Electricity-theft detection offers another use case, although class imbalance, legal consequences and adversarial behaviour raise

the control standard. Distributed-energy coordination may use federation for prediction or policy learning while keeping device histories local.

Cross-utility benchmarking can use federated analytics without producing one operational model. Participants can calculate common statistics, risk indicators or performance models under a governed protocol. This may be commercially valuable where competitive or regulatory constraints prevent pooling.

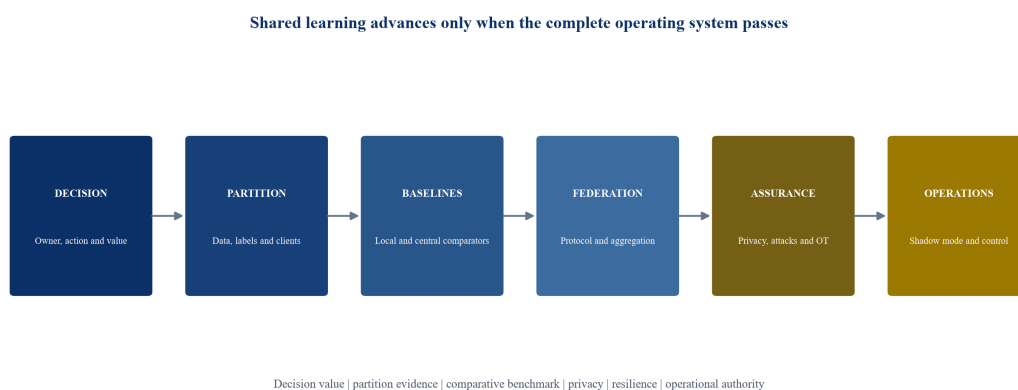
Use-case selection should consider label scarcity. Asset failures, network disturbances and confirmed theft cases are rare relative to normal observations. One utility may have too few events to train a reliable model, while several utilities collectively have a more useful evidence base. Federation can expose the model to a wider range of conditions without moving raw event histories. The benefit depends on compatible labels, equipment and operating regimes. A shared label such as “transformer failure” can conceal distinct causal mechanisms, inspection practices and reporting thresholds.

Forecasting use cases require careful boundary design. A distribution utility can federate across regions, retailers or campuses. A system operator can collaborate with generators or balancing areas. Weather data may already be public or centrally licensed, while consumption and asset data remain local. The architecture should isolate the genuinely constrained features rather than federate every part of the model.

Asset-health federation can be organised by equipment class, manufacturer, operating environment or failure mode. The resulting model should not assume that a fault signature transfers across different sensor configurations or maintenance histories. A hierarchical design can learn common representations at fleet level and retain local calibration. Engineering review remains necessary before an alert changes maintenance.

Water and district-energy utilities present related opportunities. Pump efficiency, leakage, treatment quality, demand and network anomalies can share statistical patterns across sites. The same tests apply: a common decision, comparable measurements, lawful participation, secure interfaces and local engineering validation. A federated architecture should not be presented as electricity-specific when the commercial decision spans several infrastructure operators.

Figure 1. The Federated Utility Model Gate



The framework is author-developed. Every gate requires evidence and accountable approval before operational progression.

7 Separate raw-data locality from privacy

Keeping raw data at a participant reduces direct collection by the coordinator. It does not make the training exchange private. Gradients and model updates can reveal information about training records under some conditions. Membership-inference and reconstruction attacks test whether an adversary can learn whether a record participated or recover sensitive attributes.

Privacy should therefore be expressed through a threat model and tested controls. Secure aggregation can allow the coordinator to learn an aggregate without seeing each participant's clear update. Differential privacy can bound the influence of individual records by clipping and adding calibrated noise. Encryption, trusted execution

environments and multiparty computation can protect selected stages.

Controls have costs. Differential privacy can reduce utility, particularly for small or imbalanced datasets. Secure aggregation may limit inspection of malicious updates. Cryptographic protocols add computation and failure modes. The design should quantify privacy budget, update size, dropout tolerance, key management and recovery rather than use “privacy-preserving” as a general label.

8 Design secure aggregation deliberately

Practical secure-aggregation protocols were developed to calculate sums of participant updates while tolerating some client dropout. A utility implementation must decide which parties hold keys, how identities are authenticated, which minimum cohort size is required and how failed rounds are recovered. Cross-silo deployments may use stronger institutional identity and contractual controls than consumer-device deployments.

Aggregation should preserve auditability. The system needs a signed record of model version, client eligibility, protocol version, round timing, accepted contributions and resulting artefact. Individual update confidentiality may prevent storage of clear contributions. The audit design should document what evidence exists without defeating the privacy objective.

The architecture also needs an availability strategy. A model round should not become a control-system dependency when communications or participants are unavailable. Quorum rules, timeout, retry, previous-model fallback and emergency suspension should be tested.

9 Protect against poisoning and backdoors

A participant can submit an update that degrades the global model or introduces a targeted behaviour. The participant may be malicious, compromised or simply misconfigured. Model-replacement and backdoor research shows that conventional averaging can be manipulated. Utility consequences could include missed anomalies, biased forecasts or unreliable recommendations.

Defences include participant authentication, signed code, attested environments, update clipping, anomaly detection, robust aggregation, reputation, holdout testing and independent validation. No single defence is sufficient. Robust aggregation can fail under collusion or non-IID data, where a legitimate distinctive client resembles an outlier.

The programme should conduct attack simulations before operational use. Tests should cover untargeted degradation, targeted backdoors, label manipulation, replay, sybil participation, compromised coordinator, malicious model distribution and inference from updates. The red team should be independent of the model owner.

10 Preserve operational-technology boundaries

Utility operational technology requires a conservative interface. NIST guidance for operational-technology security emphasises safety, reliability and performance constraints that differ from conventional information technology. A federated-learning service should not create uncontrolled inbound paths to control networks or allow a cloud coordinator to push executable changes directly to field devices.

A suitable pattern can use a demilitarised zone, controlled data diode or broker, signed model packages, allowlisted endpoints and staged import. Local training may occur in a protected analytics zone using replicated data. Operational deployment should pass change control, validation and rollback. Identity, secrets and certificates should be managed through the utility's authorised systems.

The model's authority must remain explicit. A forecast may inform an operator. An anomaly score may prioritise inspection. Closed-loop control requires a separate safety case. Federation does not alter the engineering requirement for deterministic protections and local fallback.

Deployment zones should be drawn before selecting technology. One zone contains source operational data and approved replicas. A second contains local training and validation. A third provides the exchange gateway. The coordinating service resides in a separate governed environment. Only approved artefacts cross each boundary. Logs and model packages follow a controlled retention schedule. This structure allows cybersecurity teams to reason about the additional attack surface.

Local software supply chains also matter. Client libraries, container images, machine-learning frameworks and cryptographic packages can create vulnerabilities across every participant. The operator should maintain a software bill of materials, verify signatures, manage vulnerabilities and test upgrades. A common coordinator cannot compensate for an unpatched client with privileged access.

Incident response should cover both cyber and model events. A compromised credential, unusual update, unexplained performance shift or attempted inference may require suspension of a client or the whole federation. The playbook should define containment, evidence preservation, notification, model rollback and re-entry. Participants need compatible reporting times and contacts.

11 Build a strong comparative benchmark

The benchmark should use a frozen test protocol and relevant operating periods. Local, centralised and federated candidates should receive comparable features and model capacity. Hyperparameter effort should be recorded so one architecture is not favoured by tuning. Results should be repeated across seeds and participation patterns.

Accuracy metrics depend on the decision. Load forecasting may use mean absolute error, root-mean-square error, pinball loss and calibration. Anomaly detection may use precision, recall, false alarms, missed critical events and time to detection. Asset-health models should report lead time and maintenance consequences. All should include confidence intervals or repeated-run distributions where feasible.

System metrics include round duration, bytes transferred, energy and compute, dropout, failed rounds, coordinator availability and recovery. Privacy and security metrics include attack success, epsilon where differential privacy is used, minimum cohort, malicious-client tolerance and audit completeness. Commercial metrics link the result to the operating process.

12 Measure communications and edge economics

Federation exchanges parameters, gradients, activations or compressed representations. A model with millions of parameters can produce material traffic across repeated rounds. Edge sites may have constrained or costly connectivity. Cross-border traffic and cloud egress can affect both governance and cost.

Compression, quantisation, sparsification, fewer rounds and partial participation can reduce traffic. These choices can change convergence and resilience. Local training consumes compute and may compete with operational workloads. Edge hardware, orchestration, patching and support create a distributed cost base.

The total-cost model should include client integration, data preparation, local compute, coordinator service, communications, security, privacy engineering, validation, monitoring, model review and incident response. A centralised alternative should include transfer, storage and access-control cost. A fair comparison uses the complete lifecycle.

Table 2. Minimum federated benchmark scorecard

Dimension	Example measure	Required comparator	Failure signal	Decision owner
Decision quality	Forecast loss, detection value or avoided cost	Local and centralised baselines	Average gain with material client deterioration	Business-process owner
Reliability	Round success, recovery time and fallback availability	Existing production workflow	Dependency on unavailable clients or coordinator	Operations owner
Communications	Bytes per client, rounds and peak bandwidth	Current transfer pattern	Network burden exceeds operating window	Infrastructure owner
Privacy	Tested leakage, privacy budget and minimum cohort	Approved threat model	Reconstruction or membership risk outside tolerance	Privacy and data owner
Security	Poisoning and backdoor attack success	Red-team scenarios	Targeted manipulation survives validation	Cybersecurity owner
Economics	Annual benefit less lifecycle cost	Local, central and non-ML alternatives	Benefit depends on unsupported adoption assumptions	Finance owner

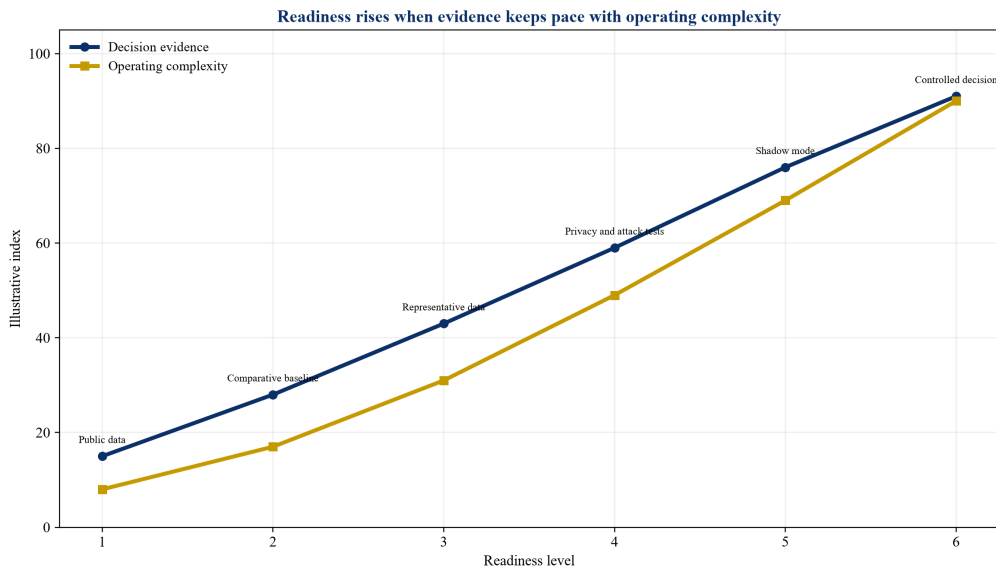
Measures are decision dependent. Thresholds should be approved before testing and should not be retrofitted to results.

13 Use a readiness curve tied to evidence

Technology readiness should measure the complete utility application. Level one reproduces a federated algorithm on a public dataset. Level two compares local, central and federated baselines under controlled partitions. Level three uses utility-representative data and realistic communications. Level four adds secure aggregation, privacy and adversarial tests. Level five operates in shadow mode. Level six supports a controlled human decision with fallback. Higher levels require sustained performance and audited change control.

Progress should require evidence rather than activity. Adding clients, epochs or encryption does not automatically advance readiness. The programme moves when the chosen decision becomes more accurate, resilient, private and economical under realistic conditions.

Figure 2. Hypothetical technology-readiness curve for a federated utility model



Values are hypothetical author-developed assumptions for illustrating the framework. They do not represent a vendor, utility or market forecast.

14 Design the pilot around shadow operation

Shadow operation runs the candidate on live or representative inputs without allowing it to control an asset, market action or customer decision. Local, centralised and federated models can run in parallel. Independent evaluators compare predictions and review the downstream decision.

The pilot should include normal and stressed periods. For load forecasting, tests can include heat, cold, holidays, storms, renewable ramps and new large loads. For asset models, tests should include maintenance changes, sensor faults and rare events. Participation should be varied deliberately to test dropout and regional absence.

Operators should record whether the output was understandable, timely and actionable. A statistically superior result can fail when it arrives late, lacks explanation or conflicts with trusted engineering evidence. The shadow period should include rollback, incident and coordinator-loss exercises.

15 Establish model and data governance

The model register should identify owner, purpose, training code, features, clients, aggregation, privacy controls, validation, approvals, dependencies and current version. Each participant needs a local data register that maps permitted sources and purposes. The coordinator needs a federation register that defines eligibility and protocol.

Change control should distinguish local changes, global model changes and protocol changes. A participant may change a meter population, sensor calibration or preprocessing. The coordinator may change model architecture or aggregation. Either can invalidate prior evidence. Material changes should trigger defined retesting.

The governance body should include the operating decision owner, data owner, cybersecurity, privacy, engineering, model risk, legal, procurement and finance. Independent challenge is particularly important when a technology or innovation team owns the programme budget and success narrative.

16 Contract for rights and responsibilities

A multi-party federation requires a contract or equivalent governance instrument. It should define permitted purpose, eligible data, prohibited inference, intellectual property, model ownership, confidentiality, security, incident reporting, audit, liability, service levels, cost sharing and exit.

The parties should agree whether the global model can be commercialised, transferred or used outside the federation. Local improvements and derived models may need separate treatment. A departing participant may request deletion of its contributions, but removing influence from a trained model can be technically difficult. The contract should state the practical remedy and retention approach.

Vendor terms should preserve portability. The utility needs access to model artefacts, configuration, logs and documentation sufficient to continue, migrate or validate. Subprocessor and hosting changes should require control. Performance claims should remain tied to the agreed benchmark.

17 Evaluate privacy and data-protection duties

Data remaining local can reduce transfers but does not settle legal analysis. Model updates, identifiers, metadata and outputs may still be personal, confidential or regulated information depending on context. The parties need a current assessment of roles, purpose, lawful basis, minimisation, retention, security and individual rights where applicable.

The European Union's General Data Protection Regulation and AI Act can apply according to establishment, market and use case. National privacy and critical-infrastructure regimes vary. Customer contracts, market codes and licence conditions may impose additional restrictions. A technical architecture cannot replace jurisdiction-specific legal analysis.

The privacy impact assessment should describe the complete information flow, including diagnostics and logs. It should test whether a coordinator, participant or attacker can infer sensitive information. Controls and residual risk should be approved by accountable owners.

18 Make explainability local and operational

Federated models can be difficult to explain because the coordinator lacks raw client context and because local distributions differ. A global feature effect may not describe a particular feeder or plant. Explanations should therefore be tested at both global and local levels.

Operators need explanations connected to decisions. A forecast explanation may show weather, calendar, recent load and distributed generation. An asset alert may show waveform features, operating regime and similar confirmed events. Explanations should remain stable enough for use and should not reveal another participant's information.

Human override, escalation and feedback should be designed into the workflow. Feedback becomes training data only through an approved path. Model confidence should not be confused with engineering certainty.

19 Construct the commercial case

The commercial case starts with the decision baseline. A load-forecast model may reduce imbalance or reserve cost. An asset model may reduce inspection or unplanned outage cost. A theft model may improve recoveries while adding investigation and customer-service cost. Benefits should be attributable, observed and approved by finance.

Attribution requires a counterfactual. A lower forecast error does not automatically create cash value. The organisation should show which procurement, dispatch, maintenance or customer action changed; the volume affected; the price or cost basis; and the period over which the result was observed. Finance should reconcile the operating record to the management account. Overlapping initiatives should not claim the same benefit.

Rare-event models create a different case. Their value can arise from improved resilience or avoided loss rather than frequent savings. The team should use scenario analysis, engineering consequence assessment and insurance or outage evidence where available. Management estimates should be identified as assumptions. Low-frequency, high-consequence value should not be converted into certain annual cash.

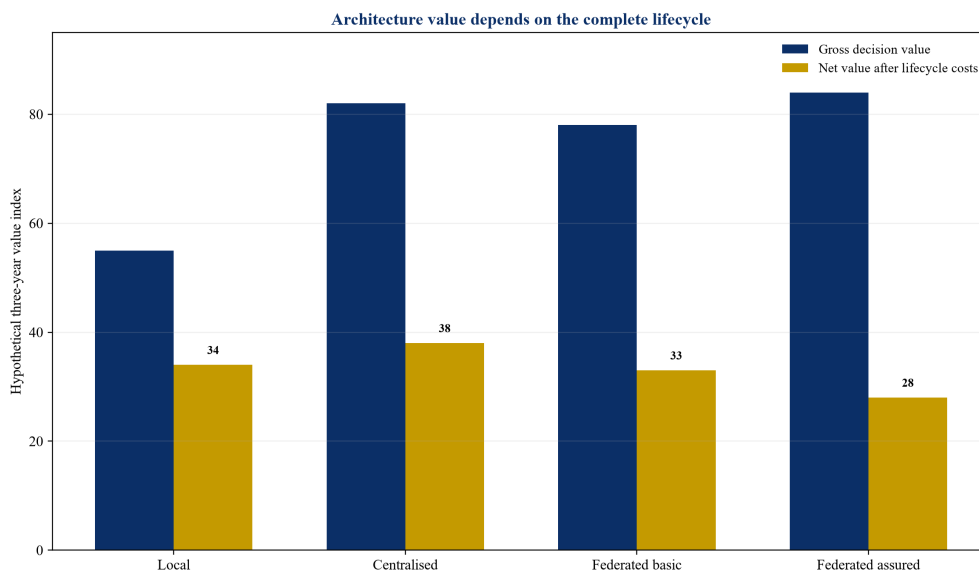
Participant incentives can determine whether the federation survives. A large participant may contribute more data and incur more integration cost. A small participant may benefit disproportionately from the global model. Contribution-based allocation sounds attractive but can reveal sensitive information or reward volume over quality. The consortium should define fees, in-kind obligations, service levels and benefit sharing through a transparent method.

Commercial models include member-funded consortia, utility-owned shared services, vendor subscriptions, outcome-based fees and regulated innovation programmes. Each allocates intellectual property, liability and exit differently. A vendor subscription can accelerate deployment while creating dependence. A utility consortium can preserve control while requiring stronger collective governance. Outcome fees require an auditable baseline and careful treatment of confounding factors.

The federated premium is the incremental value over the best feasible local, centralised or conventional alternative. It may come from access to rare events, faster onboarding of partners, improved regional coverage or reduced data-transfer burden. It should not include the whole value of analytics when a simpler architecture captures most of it.

Costs include pilots that fail, distributed integration and continuing assurance. A business case should state which assumptions are hypothetical, provide ranges and identify the evidence required to replace each assumption.

Figure 3. Hypothetical three-year value bridge for architecture choices



Amounts are hypothetical author-developed assumptions in index points. They are not forecasts, quotations or performance claims.

20 Use scenario valuation rather than one forecast

An investment committee can evaluate three scenarios. The limited case assumes the federated model performs close to local models and is retained as a research capability. The adoption case assumes it improves the named decision for a defined client population. The platform case assumes the architecture supports additional use cases or external participants.

Probability weights are management assumptions and should remain visible. Value should be linked to milestones such as successful attack testing, minimum client performance, shadow-mode adoption and signed participation. Platform value should not be paid upfront without evidence of rights, portability and repeatable onboarding.

For an acquisition or partnership, the buyer should separate current assets from contingent upside. Current assets include contracts, software, validated models, data rights, integration connectors, security evidence and people. Contingent upside can be structured through milestones, earn-outs, options or staged funding, subject to legal and accounting advice.

The limited scenario should still have a defined value. Reusable connectors, data dictionaries, privacy assessments and benchmark infrastructure can improve later analytics programmes. These assets should be valued at replacement cost or supported use value rather than assumed platform revenue. Their portability should

be demonstrated.

The adoption scenario should connect client numbers to operating decisions. A signed participant is not an adopted model. The valuation should distinguish contracted access, integrated clients, completed training rounds, active shadow users and decisions influenced. Revenue recognition, implementation cost and support obligations may differ at each stage.

The platform scenario requires evidence that onboarding becomes repeatable. If every utility needs bespoke data mapping, cyber approval and modelling, scale may remain service intensive. Gross margin should include local integration, validation, assurance and customer success. Expansion into new use cases can require a new legal basis and new evidence even when the technical platform is reused.

Sensitivity should focus on the variables that change the decision: incremental model performance, adoption, client retention, onboarding cost, assurance cost, compute, communication, incident exposure and time to deployment. A wide terminal multiple cannot substitute for these operating drivers. Investment committees should see the break-even number of productive clients and the evidence needed to reach it.

Table 3. Stage-gated commercialisation roadmap

Stage	Core work	Required evidence	Commercial decision	Stop condition
Define	Name decision, clients, baselines and threat model	Approved decision record and partition map	Fund a bounded experiment	No material shared signal or feasible intervention
Prove	Run local, central and federated comparisons	Reproducible performance and system benchmark	Select architecture	Federation does not exceed feasible alternatives
Assure	Add secure aggregation, privacy and attack tests	Residual risks within approved tolerance	Approve representative pilot	Leakage, poisoning or OT control remains unacceptable
Shadow	Run with live cadence and operator review	Timely output, client fairness, fallback and value evidence	Approve controlled use	Poor adoption, reliability or attributable value
Scale	Onboard clients and monitor sustained outcomes	Repeatable onboarding, audited controls and finance-approved value	Expand or commercialise	Lifecycle cost exceeds realised benefit

Budgets and thresholds should be developed for the specific programme. The gates below are author-developed and non-prescriptive.

21 Plan the first ninety days

Days one to fifteen should define the decision and governance. The team records baselines, participants, partitions, legal basis, threat model and stop conditions. It freezes the evaluation protocol and assigns independent reviewers.

Days sixteen to thirty-five should create the local and central comparators, validate data definitions and simulate partitions. The team identifies non-IID conditions and tests whether pooled learning has material value. If it does not, federation should stop.

Days thirty-six to sixty should build the federated candidate, measure communications and test aggregation choices. The programme adds secure aggregation or differential privacy only against a defined threat. It records the performance cost of every control.

Days sixty-one to seventy-five should conduct attack, privacy, resilience and OT-interface tests. Days seventy-six to ninety should review the commercial case and decide whether shadow operation is justified. The gate pack should include evidence, exceptions, owners and next-stage budget.

22 Define credible success conditions

Technical success requires reproducible improvement against relevant baselines, acceptable client-level performance, reliable completion and tested recovery. Privacy success requires a documented threat model, appropriate controls and empirical tests. Security success requires malicious updates and coordinator compromise to be addressed within the approved risk appetite.

Operational success requires timely output, understandable evidence, operator acceptance and a safe fallback. Commercial success requires finance-approved, attributable benefit after lifecycle cost. Strategic success can include reusable data definitions, benchmark harnesses, secure collaboration capability and contract templates.

The programme should state the evidence that would disprove its thesis. This prevents success criteria from moving after results arrive.

23 Recognise common failure modes

The first failure is treating locality as guaranteed privacy. A second is comparing federation with weak local models and omitting a centralised ceiling. A third is reporting only average accuracy. Others include unrecorded client dropout, favourable partitions, insufficient attack tests, unsupported cost savings and direct connectivity into operational networks.

Governance can fail when participant rights are unclear or when a global model disadvantages a region. Commercialisation can fail when every client requires bespoke engineering. Vendor dependence can become material when the coordinator, protocol and model artefacts are not portable.

These risks can be managed through the gate. The programme defines the decision, records partitions, freezes baselines, tests privacy and attacks, protects OT boundaries, runs in shadow mode and measures realised value.

24 Implications for utilities, investors and advisers

Utilities should begin with cross-silo use cases where a limited number of accountable entities share a genuine signal and have a reason to retain raw data. Load forecasting and asset-health analysis provide credible experimental domains. Operational use should follow evidence from the complete process.

Investors should examine data rights, client retention, integration burden, attack resilience, comparative performance and recurring cost. A claim that data never leaves the customer site is incomplete without the protocol, leakage tests and operating model. Valuation should separate current contracted assets from contingent platform potential.

Advisers can connect engineering, data governance, cybersecurity, contracts and finance. A useful mandate defines the decision, structures the benchmark, challenges privacy claims, prices lifecycle cost, designs commercial terms and creates an accountable scale route.

25 Limitations

This paper does not determine legal compliance, cybersecurity sufficiency or engineering safety for a specific utility. Requirements depend on jurisdiction, market, asset, data, participant and use case. The field is developing rapidly and implementations vary.

Published experimental results are not directly comparable when datasets, partitions, models, tuning, threat assumptions and metrics differ. Performance on a public dataset does not establish performance in a live utility environment. Hypothetical figures and tables in this paper illustrate decision methods only.

Secure aggregation, differential privacy, robust aggregation and trusted execution address different risks. Their interaction can create trade-offs that this paper does not quantify for a particular system. Privacy guarantees depend on implementation and parameter choices. Adversarial resilience depends on attacker capabilities and participant populations.

The commercial frameworks are decision aids. They do not establish a market price, forecast revenue or determine accounting treatment. Utility benefits depend on tariff, market design, asset condition, adoption and operating authority. Transaction structures require current legal, tax, accounting, regulatory and technical advice.

Conclusion

Federated learning can help utilities learn across organisational and technical boundaries while retaining raw data locally. Its strongest case arises when shared learning has material decision value and data centralisation faces a genuine constraint. The architecture can support forecasting, asset health, anomaly detection and benchmarking, but it introduces its own privacy, security, communications and governance requirements.

The Federated Utility Model Gate converts the technology into an investment and operating decision. It compares local, centralised and federated options; documents data partitions; manages non-IID behaviour; tests leakage and poisoning; protects operational-technology boundaries; and links performance to an attributable result. Shadow operation and stage-gated funding keep experimentation bounded.

A utility can then decide whether to deploy, redesign or stop using evidence from the whole system. The durable capability is the governed collaboration model: clear rights, reproducible benchmarks, secure protocols, accountable decisions and finance-approved value.

Frequently Asked Questions

Does federated learning guarantee that utility data remain private?

No. Raw data can remain local while model updates, metadata or outputs still reveal information. Privacy requires a defined threat model, controls such as secure aggregation or differential privacy where appropriate, and empirical testing.

When is a federated model preferable to a centralised model?

It is preferable when pooled learning creates material decision value and raw-data movement creates a binding legal, contractual, security, competitive or technical constraint. The choice should be tested against local and centralised baselines.

What is the best first utility use case?

Cross-silo load forecasting or asset-health analysis can be credible starting points when participants use aligned features, have sufficient data quality and share an operating objective. The decision should have a measurable intervention and owner.

How should a utility address different client data distributions?

It should quantify non-IID conditions, report client-level performance, test aggregation and personalisation approaches, and define minimum performance for every material participant. Average performance alone is insufficient.

Can secure aggregation prevent model poisoning?

Secure aggregation protects individual updates from inspection while calculating an aggregate. It can make some forms of malicious-update detection harder. Poisoning requires additional controls such as authenticated participants, trusted code, robust aggregation, holdout tests and red-team exercises.

Should federated models connect directly to operational technology?

A federated-learning service should use controlled interfaces, signed artefacts, staged validation and rollback. Direct operational authority requires a separate safety and control case. Shadow operation is an appropriate early stage.

How should an investor value a federated-learning platform for utilities?

The investor should value current contracts, validated models, integration assets, security evidence, rights and talent separately from future platform upside. Contingent value can be linked to technical, operational and commercial milestones.

When should a utility stop a federated-learning programme?

It should stop or redesign when pooled learning adds insufficient value, federation fails credible baselines, privacy or poisoning risk remains outside tolerance, OT integration is unsafe, client-level outcomes are unacceptable or lifecycle cost exceeds attributable benefit.

References

1. McMahan, H.B. et al., *Communication-Efficient Learning of Deep Networks from Decentralized Data*, AISTATS, 2017. <https://proceedings.mlr.press/v54/mcmahan17a.html>
2. Kairouz, P. et al., *Advances and Open Problems in Federated Learning*, Foundations and Trends in Machine Learning, 2021. <https://doi.org/10.1561/22000000083>
3. Bonawitz, K. et al., *Practical Secure Aggregation for Privacy-Preserving Machine Learning*, ACM CCS, 2017. <https://doi.org/10.1145/3133956.3133982>
4. Li, T. et al., *Federated Optimization in Heterogeneous Networks*, MLSys, 2020. <https://proceedings.mlsys.org/paper/2020/hash/1f5fe83998a09396ebc6477d9475ba0c-Abstract.html>
5. Karimireddy, S.P. et al., *SCAFFOLD: Stochastic Controlled Averaging for Federated Learning*, ICML, 2020. <https://proceedings.mlr.press/v119/karimireddy20a.html>
6. Smith, V. et al., *Federated Multi-Task Learning*, NeurIPS, 2017. <https://proceedings.neurips.cc/paper/2017/hash/6211080fa89981f66b1a0c9d55c61d0f-Abstract.html>
7. Geyer, R.C., Klein, T. and Nabi, M., *Differentially Private Federated Learning: A Client Level Perspective*, 2017. <https://arxiv.org/abs/1712.07557>
8. Zhu, L., Liu, Z. and Han, S., *Deep Leakage from Gradients*, NeurIPS, 2019. <https://proceedings.neurips.cc/paper/2019/hash/60a6c4002cc7b29142def8871531281a-Abstract.html>
9. Bagdasaryan, E. et al., *How To Backdoor Federated Learning*, AISTATS, 2020. <https://proceedings.mlr.press/v108/bagdasaryan20a.html>
10. Blanchard, P. et al., *Machine Learning with Adversaries: Byzantine Tolerant Gradient Descent*, NeurIPS, 2017. <https://proceedings.neurips.cc/paper/2017/hash/f4b9ec30ad9f68f89b29639786cb62ef-Abstract.html>
11. Cheng, X., Li, C. and Liu, X., *A Review of Federated Learning in Energy Systems*, IEEE SmartGridComm Workshops, 2022. <https://arxiv.org/abs/2208.10941>
12. Zhang, Z. et al., *Federated Learning for Smart Grid: A Survey on Applications and Potential Vulnerabilities*, 2024. <https://arxiv.org/abs/2409.10764>
13. Wang, Y. et al., *Advancing Power System Services with Privacy-Preserving Federated Learning Techniques: A Review*, IEEE Access, 2024. <https://doi.org/10.1109/ACCESS.2024.3407468>
14. Lee, J. et al., *FedREP: Towards Horizontal Federated Load Forecasting for Retail Energy Providers*, IEEE APPEEC, 2022. <https://doi.org/10.1109/APPEEC53445.2022.10072290>
15. Li, Y. et al., *Fed-SAD: A Secure Aggregation Federated Learning Method for Distributed Short-Term Load Forecasting*, IET Generation, Transmission & Distribution, 2023. <https://doi.org/10.1049/gtd2.13022>
16. Iqbal, M. et al., *Privacy-Preserving Collaborative Split Learning Framework for Smart Grid Load Forecasting*, IEEE Transactions on Dependable and Secure Computing, 2025. <https://doi.org/10.1109/TDSC.2025.3585297>
17. Pei, Y.-T. et al., *A Daily Load Forecasting Method Based on Federated Learning and Transformer Model*, The Journal of Engineering, 2026. <https://doi.org/10.1049/tje2.70174>
18. Su, H. et al., *Lightweight Federated Learning for On-Device Non-Intrusive Load Monitoring*, IEEE Transactions on Smart Grid, 2025. <https://doi.org/10.1109/TSG.2024.3482363>
19. Zhang, J. et al., *A Privacy-Preserving Heterogeneous Federated Learning Framework with Class Imbalance Learning for Electricity Theft Detection*, Applied Energy, 2025. <https://doi.org/10.1016/j.apenergy.2024.124789>
20. NIST, *Artificial Intelligence Risk Management Framework 1.0*, 2023. <https://www.nist.gov/itl/ai-risk-management-framework>
21. NIST, *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile*, 2024. <https://doi.org/10.6028/NIST.AI.600-1>
22. NIST, *Privacy Framework: A Tool for Improving Privacy Through Enterprise Risk Management*, Version 1.0, 2020. <https://www.nist.gov/privacy-framework>
23. NIST, *Guidelines for Evaluating Differential Privacy Guarantees*, SP 800-226, 2025. <https://csrc.nist.gov/pubs/sp/800/226/final>
24. NIST, *Guide to Operational Technology Security*, SP 800-82 Revision 3, 2023. <https://csrc.nist.gov/pubs/sp/800/82/r3/final>
25. NIST, *Zero Trust Architecture*, SP 800-207, 2020. <https://csrc.nist.gov/pubs/sp/800/207/final>
26. NIST National Cybersecurity Center of Excellence, *Securing Distributed Energy Resources*, SP 1800-32, 2022. <https://www.nccoe.nist.gov/projects/securing-distributed-energy-resources>
27. NIST, *Cybersecurity Framework 2.0*, 2024. <https://www.nist.gov/cyberframework>
28. CISA and international partners, *Principles of Operational Technology Cybersecurity*, 2024. <https://www.cisa.gov/news-events/alerts/2024/10/01/asds-acsc-cisa-fbi-nsa-and-international-partners-release-guidance-principles-ot-cybersecurity>
29. National Renewable Energy Laboratory, *System Architectures to Support Autonomous Energy Systems*. <https://www.nrel.gov/grid/system-architecture.html>
30. National Renewable Energy Laboratory, *Smart-DS: Smart Distribution System Models*. <https://www.nrel.gov/grid/smart-ds.html>
31. United States Department of Energy, Office of Electricity, *Grid Modernization and the Smart Grid*. <https://www.energy.gov/oe/grid-modernization-and-smart-grid>
32. North American Electric Reliability Corporation, *Security Guideline for the Electricity Sector: Supply Chain*. <https://www.nerc.com/pa/comp/Pages/Supply-Chain-Risk-Mitigation-Program.aspx>
33. European Parliament and Council, *Regulation (EU) 2016/679, General Data Protection Regulation*. <https://eur-lex.europa.eu/eli/reg/2016/679/oj>
34. European Parliament and Council, *Regulation (EU) 2024/1689, Artificial Intelligence Act*. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>
35. European Union Agency for Cybersecurity, *Cybersecurity for Smart Grids*. <https://www.enisa.europa.eu/topics/critical-information-infrastructures-and-services/smart-grids>
36. International Electrotechnical Commission, *IEC 62443 Industrial Communication Networks, Network and System Security*. <https://www.iec.ch/cyber-security>
37. International Organization for Standardization, *ISO/IEC 27001 Information Security Management Systems*. <https://www.iso.org/isoiec-27001-information-security.html>

38. International Organization for Standardization, *ISO/IEC 42001 Artificial Intelligence Management System*. <https://www.iso.org/standard/81230.html>
39. International Organization for Standardization, *ISO 31000 Risk Management*. <https://www.iso.org/iso-31000-risk-management.html>
40. IEEE, *IEEE 1547-2018 Standard for Interconnection and Interoperability of Distributed Energy Resources*. <https://standards.ieee.org/ieee/1547/5915/>
41. North American Electric Reliability Corporation, *Reliability Standards for the Bulk Electric Systems of North America*. <https://www.nerc.com/pa/stand/Pages/ReliabilityStandards.aspx>
42. International Energy Agency, *Electricity Grids and Secure Energy Transitions*, 2023. <https://www.iea.org/reports/electricity-grids-and-secure-energy-transitions>
43. International Renewable Energy Agency, *Innovation Landscape Brief: Market Integration of Distributed Energy Resources*, 2019. <https://www.irena.org/publications/2019/Feb/Innovation-landscape-for-a-renewable-powered-future>
44. IFRS Foundation, *IAS 38 Intangible Assets*. <https://www.ifrs.org/issued-standards/list-of-standards/ias-38-intangible-assets/>
45. IFRS Foundation, *IFRS 13 Fair Value Measurement*. <https://www.ifrs.org/issued-standards/list-of-standards/ifrs-13-fair-value-measurement/>

About the Author

Authored by Chennakeshav Adya

Independent Researcher

Chennakeshav (CK) is a corporate finance and investment banking executive with 25+ years of global experience in deal origination, structuring and execution across M&A, growth capital and corporate strategy. He has led value-creation mandates for founders, corporates and funds — bridging the boardroom view to hands-on execution and close.

His career spans Morgan Stanley, HSBC, Lloyds Banking Group, EWEC, ADQ portfolio companies and Emirates Growth Fund, across TMT, real estate, fintech, deeptech, cleantech, infrastructure and energy. He has partnered with C-suite leaders, private equity and venture funds, sovereign wealth funds and family offices to finance complex fund raises and scale-up ventures, and has led M&A due diligence, post-merger integration and business-transformation initiatives to create value.

At Matchpoint Partners he is Managing Partner, leading the firm's corporate finance, M&A and capital-raising practice. He holds an MBA from London Business School, an engineering degree from VTU and a Master of Laws (LLM, in progress) from UCL London.

An active start-up mentor, CK mentors at Techstars, DIFC FinTech Hive, Startup Grind, Founder Institute and IN5, serves as Entrepreneur Mentor in Residence (EMiR) at London Business School, and judges the Entrepreneurship World Cup.

<https://www.linkedin.com/in/ckadya/>

<https://www.matchpoint-partners.com/team/ck-adya.html>

This paper is part of a continuing series on the structure of private and alternative markets. The views expressed are the author's own. The paper is for information only, describes market structure in general terms, and does not constitute investment, legal, tax or regulatory advice or a recommendation in respect of any security, vehicle or counterparty.