

Test the Grid without Exposing It: Synthetic Operational Data

A decision framework for utility control-system testing when operational data are restricted

Authored by Chennakeshav Adya

Independent Researcher

October 2026

ABSTRACT

Electricity, water and district-energy operators need representative operational data to test control systems, train analysts, evaluate vendors and develop artificial-intelligence applications. Real supervisory-control, protection, asset and customer data can reveal critical-infrastructure configurations, operating weaknesses, commercial positions or personal information. Direct access may therefore be restricted even when the intended work is legitimate. Synthetic operational data can widen safe access, but a plausible file is not automatically a valid test environment. This paper develops a Synthetic Operational Data Assurance Gate for deciding when generated networks, time series, events and cyber-physical traces are suitable for a named utility decision. The framework separates structural fidelity, temporal fidelity, physical feasibility, control-response fidelity, tail-event coverage, privacy and security. It examines public synthetic grid cases, simulation, digital twins, software-in-the-loop, controller-in-the-loop, power-hardware-in-the-loop and controlled replay of real data. It also defines the evidence needed to move from development and training to procurement, shadow operation and controlled operational use. Published programmes show that synthetic networks and testbeds can support reproducible research without disclosing an actual grid. They also show that data sets are usually designed for particular applications and may omit dynamic, sequence, protection, market, communications or behavioural detail. A utility should therefore validate synthetic data against the decision it will support, retain a protected real-data benchmark, test rare and adversarial conditions, and prevent test artefacts from acquiring operational authority without an independent safety case. 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: C53, C63, L94, O32, Q40

Keywords: synthetic data, power systems, utilities, operational technology, digital twins, control-system testing, privacy, cyber ranges, hardware in the loop, grid innovation

Introduction

Modern utilities depend on interconnected physical assets, communications, software and human decisions. Control centres receive telemetry, state estimates, alarms, weather, market positions and maintenance information. Distribution systems add smart meters, distributed generation, storage, electric vehicles and flexible demand. Water networks add pumps, reservoirs, treatment stages and quality measurements. These records can support better testing and analysis, but they can also reveal how critical infrastructure behaves and where it is vulnerable.

Access restrictions create a practical problem. A vendor may need realistic data to demonstrate an advanced distribution-management function. A cyber team may need representative traffic and process signals to test detection. A data-science team may need labelled disturbances that are scarce in normal operations. An investment committee may need evidence that a digital product can work beyond a demonstration. Sharing a production extract can expose protected infrastructure, customer information, contractual data or commercial strategy. Removing names and identifiers does not necessarily remove structural or temporal clues.

Synthetic operational data are generated representations of systems, measurements or events that do not directly reproduce a production record. They include fictitious network topologies, simulated load and renewable profiles, generated phasor measurements, cyber-range traffic, fault sequences, market scenarios and digital-twin outputs. Some are built entirely from engineering rules and public information. Others are trained or calibrated using protected real data. Their risk and evidential value differ.

The central conclusion of this paper is that synthetic data are a test instrument. Their fitness depends on the specific decision, the generating process and an evidence chain to protected real-world behaviour. The same synthetic feeder can be sufficient for software integration and inadequate for protection studies. The same generated time series can support user-interface testing and distort extreme-ramp risk. A governed programme therefore needs a decision-specific validation contract rather than a general claim of realism.

1 Define the decision before generating data

A synthetic-data programme should begin with a named decision and an accountable owner. “Create realistic grid data” is a technical intention. “Determine whether a new outage-management interface preserves alarm priority during a multi-feeder storm” is a testable decision. “Compare two vendor state estimators under missing telemetry and topology error” is another. The decision record should identify the system boundary, user, intervention, timing, tolerance, consequence of error and fallback.

The acceptable evidence depends on the decision. Training analysts to navigate a control-room interface can tolerate larger deviations from production behaviour than validating an automatic voltage controller. Screening algorithms before they enter a secure environment requires different evidence from setting protection parameters. Procurement may need repeatable scenarios and transparent scoring. Regulatory assurance may require traceability to standards, approved models and independent witnesses.

The decision should also state what synthetic data cannot prove. Passing a generated scenario does not demonstrate safe performance in every operating condition. Failure on a synthetic scenario can reveal a weakness, while success may only show compatibility with the assumptions embedded in that scenario. This asymmetry is important. Synthetic tests are often strong at finding defects and weaker at proving absence of defects.

2 Distinguish the synthetic-data families

Synthetic operational data cover several different artefacts. A network case represents buses, feeders, lines, transformers, generators, loads and electrical parameters. A temporal case adds load, renewable output, weather, prices, outages and switching over time. A dynamic case adds control models, inertia, converter behaviour and disturbance response. A protection case requires fault types, sequence quantities, relay logic and timing. A cyber-physical case combines process behaviour with communications, identities, commands and adversarial actions.

Generated records also vary by relationship to production data. A fully simulated case can be built from engineering equations and public inputs. A statistically fitted case learns distributions or dependencies from protected observations. A hybrid case combines a synthetic network with calibrated load and event models. A replay case uses transformed or selected real records in a controlled environment. An augmented case adds generated rare events to real normal-operation data. These approaches should not be treated as equivalent.

The programme needs an inventory that records provenance, generator, parameters, source restrictions, intended use, prohibited use, validation evidence, release tier and expiry. Each version should be reproducible. A user should be able to identify the model, random seed, source period, calibration, scenario definition and software environment that produced a result.

Table 1. Synthetic operational data families and appropriate uses

Data family	Typical contents	Appropriate early use	Principal limitation	Minimum evidence
Synthetic network	Fictitious topology and asset parameters	Planning-tool integration and reproducible studies	May omit real operational constraints and dependencies	Structural statistics, feasible power flow and documented scope
Synthetic time series	Load, renewable, price and weather sequences	Forecast development and scenario testing	Can smooth tails or break cross-variable dependence	Marginal, joint, temporal and extreme-event validation
Dynamic and protection data	Disturbances, controls, sequence quantities and relay events	Controller and protection research	Parameter uncertainty can dominate results	Engineering model review and disturbance-response comparison
Cyber-physical traces	Process signals, network traffic, commands and attacks	Detection testing and exercises	Test traffic can reveal unrealistic shortcuts	Protocol fidelity, timing fidelity and adversarial coverage
Privacy-preserving synthetic records	Generated customer or asset-level observations	Data-science access and external collaboration	Privacy and analytical utility can conflict	Disclosure testing, utility testing and approved release tier
Hardware-in-the-loop output	Real controller or equipment response to a simulated system	Pre-production control validation	Laboratory interfaces and scale may differ from field conditions	Calibration, uncertainty budget, repeatability and safe failure tests

3 Use public synthetic grids for reproducible baselines

Public synthetic grids have an important role because researchers and vendors can examine the same system without exposing an actual network. IEEE research on ACTIVSg cases developed statistical validation criteria for geography, substations, connectivity, line structure, electrical feasibility and system complexity. Larger cases extended the approach to thousands of buses and combined transmission and distribution representations.

The Grid Modernization Laboratory Consortium's Reliability Test System updates the earlier IEEE reliability system for production-cost modelling and publishes data in several tool formats. NREL's SMART-DS programme created large synthetic distribution systems with detailed customer, feeder and distributed-energy representations. OPEN COG Grid seeks coherent data sets with the parameters required across a wider range of computational studies.

These resources are valuable baselines, not replicas of a utility. Their documentation states the intended applications and known gaps. A data set built for production-cost modelling may lack the sequence data required for fault studies. A distribution case may represent topology and load while omitting vendor-specific control logic. A synthetic system can be statistically similar to real grids and still differ from a particular operator's asset mix, operating rules or weather exposure.

The first comparison in a procurement or development programme should use a well-documented public case. It creates a reproducible reference, exposes software assumptions and allows independent replication. Utility-specific calibration can follow after the basic workflow works and access controls are in place.

4 Build a fidelity contract

"Realistic" is too broad to govern synthetic operational data. The programme should create a fidelity contract that lists the properties relevant to the decision. Structural fidelity covers network size, degree distribution, impedance, voltage levels, transformer ratios, generation and load geography. Temporal fidelity covers seasonality, ramping, autocorrelation, outages, forecast error and cross-variable dependence. Physical fidelity covers power balance, voltage, frequency, thermal limits and equipment capability.

Control fidelity covers response time, dead bands, saturation, sequencing, interlocks and fallback. Cyber fidelity covers protocols, message timing, identities, network zones, normal traffic and plausible adversarial activity. Human fidelity covers alarm volumes, operating procedures, shift behaviour and escalation. Commercial fidelity covers market intervals, settlement, contractual constraints and cost consequences where these affect the decision.

Every property needs a measure, comparator, tolerance and owner. Some measures compare distributions. Others compare trajectories, event ordering, stability margins or operator outcomes. A useful contract also records features intentionally excluded. Omissions are safer when explicit because users can prevent the data from migrating into an unsupported use.

5 Preserve physical feasibility

Synthetic data should respect the governing physical system. A set of plausible measurements can still be impossible when considered together. Electricity data must satisfy network equations, equipment constraints and operating logic. Water data must respect mass balance, pressure, pump curves and storage. District-energy data must reflect thermal inertia, flows and plant limits.

Physical feasibility can be imposed during generation or tested after generation. Model-based simulation provides strong consistency with its equations, although the equations and parameters may be incomplete. Statistical or generative methods can learn complex patterns, but they may produce combinations that violate physical constraints. Hybrid methods use a learned proposal followed by a physics-based acceptance, correction or optimisation stage.

Validation should cover normal states, stressed states and transitions. A profile that matches daily averages can still miss steep evening ramps. A feasible steady-state dispatch can still produce an implausible dynamic response. A voltage trajectory can remain within limits because the generator has silently ignored reactive-power capability. The test harness should therefore preserve constraint violations as visible outcomes rather than

correcting every difficult scenario away.

6 Validate distributions and dependence

Univariate similarity is an incomplete test. Matching the distribution of load, voltage or frequency separately does not establish that their relationships are credible. Synthetic data should preserve relevant joint distributions, lag relationships, spatial correlation, weather dependence and conditional behaviour. For forecast testing, the relationship between forecasts and actuals matters more than either series in isolation.

The scorecard can include quantiles, autocorrelation, spectral density, cross-correlation, ramp distributions, duration curves, transition probabilities and event-conditioned summaries. Multivariate distances and classifier-based tests can help detect differences, but they should be connected to an engineering interpretation. A good overall score can conceal a weak substation class or weather regime.

The validation should be stratified by season, asset type, operating mode and stress condition. Tail behaviour deserves separate tests. The programme should publish the accepted tolerance before comparing candidate generators. Tuning a generator until it passes an undisclosed composite score creates a risk of overfitting the benchmark.

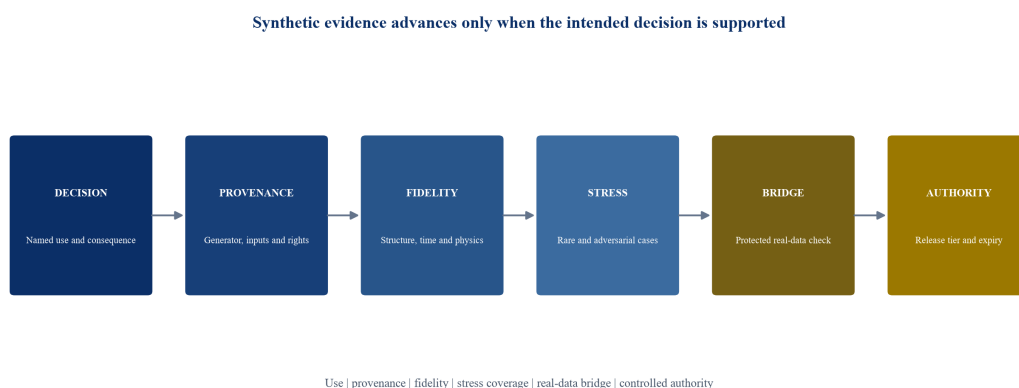
7 Treat rare events as designed experiments

Major disturbances, cascading outages, protection misoperations and cyber incidents are rare in observed data. Their scarcity is one reason to use simulation. It also means that a generative model trained primarily on normal operations may have little evidence for the events that matter most.

Rare-event scenarios should therefore be designed with engineering and threat expertise. The scenario record should identify initiating conditions, event sequence, protection or control response, expected safe state and uncertainty. Examples can include loss of a critical line, islanding, inverter control interaction, forecast failure during extreme weather, bad-data injection, communications loss, delayed operator action and restoration.

Generated variations can explore combinations and timing around the designed event. They should not replace the causal scenario definition. A model that creates dramatic but physically incoherent events can waste testing effort and teach a detection system the wrong signatures. Coverage should be measured against an approved hazard and attack library, not by the number of generated records.

Figure 1. The Synthetic Operational Data Assurance Gate



The framework is author-developed. Each stage requires documented evidence and accountable approval before the data move to a higher-consequence use.

8 Separate confidentiality from privacy

Critical-infrastructure confidentiality and personal privacy overlap, but they are not the same problem. A network model may contain no personal data while revealing substations, dependencies or operational limits. Smart-meter or customer-service records can contain personal information even when network topology is removed. Commercial data can reveal bids, fuel positions, plant economics or vendor performance.

Synthetic generation does not automatically remove these risks. A statistically fitted generator can reproduce unusual source records or allow membership and attribute inference. NIST guidance warns that synthetic data without differential privacy can remain vulnerable to privacy attacks. Differential privacy can provide a measurable bound for individual contribution, but it introduces an accuracy cost and does not protect every infrastructure secret.

The programme should define the protected entities and threat model. Privacy tests can examine exact and near matches, membership inference and attribute disclosure. Infrastructure tests can examine whether topology, asset identity, location, operating weakness or proprietary settings can be recovered. Release approval should consider the recipient, purpose, access environment and combination with other data. Public release needs stronger assurance than use in a controlled utility laboratory.

9 Use tiered release rather than one data product

A practical operating model has several release tiers. A public tier can contain open synthetic networks and documented scenarios. A partner tier can add calibrated behaviour under contract and secure access. An internal development tier can include richer synthetic records, protected mappings and detailed failure cases. A restricted validation tier can hold controlled real data and the bridge tests required to assess synthetic fidelity.

Each tier should have allowed users, tools, outputs, retention, monitoring and approval. Movement between tiers requires a new review because joining data or adding metadata can change disclosure risk. A partner should not infer that receiving a synthetic data set grants rights to the generating model or source statistics. Contracts should address permitted purpose, redistribution, derived models, publication, incidents and deletion.

Tiering also improves efficiency. Most software integration, user training and basic model development can happen on public or partner data. Expensive secure-environment work can focus on the final bridge tests. This reduces the number of people and systems exposed to restricted data while preserving an evidence route to production.

10 Establish a protected real-data bridge

A synthetic environment needs a bridge to reality. The bridge is a limited, controlled comparison against production or high-fidelity operational evidence. It does not require broad disclosure. It can run inside the utility, a trusted research environment or an approved laboratory. External developers can submit code or models and receive scored results without receiving the protected data.

The bridge should use a frozen protocol. It should compare the same decisions, periods, events and metrics across synthetic and protected cases. Results should be stratified by operating regime. The programme should preserve failed comparisons and diagnose which synthetic assumptions caused them. A bridge score should expire when topology, operating practice, data schema or technology changes materially.

The protected benchmark also prevents circular validation. If a synthetic generator and evaluation metric are trained on the same source and assumptions, high similarity can be misleading. Independent holdout periods, events and engineering review reduce this risk. For safety-critical functions, the bridge is followed by shadow operation and controlled field trials.

11 Combine model validation with software verification

Synthetic data can test the data and model layers while leaving software defects undiscovered. A complete programme separates verification from validation. Verification asks whether the implementation solves the specified equations and handles data correctly. Validation asks whether the representation is adequate for the intended real-world decision.

Software verification covers schema, units, time zones, missing values, ordering, numerical tolerances, interfaces and deterministic replay. Model validation covers physical behaviour, distributional fidelity, event response and

decision consequences. Independent code paths can reduce common-mode error. A vendor should disclose the generator version, simulator, solver settings and preprocessing.

Regression tests should freeze representative scenarios and expected outputs. New releases should show which results changed and why. When the generator is learned, changes to training data or model architecture should trigger revalidation. Synthetic data should therefore be governed as a model and a data product, with ownership for both.

12 Move through simulation and hardware in the loop

Testing maturity progresses through several environments. Offline software-in-the-loop testing checks algorithms against generated data and simulated systems. Real-time simulation adds timing, communications and control-loop behaviour. Controller-hardware-in-the-loop connects a physical controller to a simulated plant or network. Power-hardware-in-the-loop connects real power equipment through an interface to a real-time simulation.

NREL describes facilities that combine virtual models, communications, control software and physical assets. Its advanced distribution-management testbed evaluates commercial or precommercial grid-management software using realistic real-time simulations and industry protocols. DOE's National SCADA Test Bed and laboratory cyber ranges provide environments for discovering and addressing control-system security weaknesses.

Each step reveals different failure modes. Offline testing is repeatable and scalable. Real-time testing exposes latency and scheduling. Controller hardware exposes firmware, input-output and device behaviour. Power hardware exposes physical limits and interactions. Progression should follow risk and evidence. A low-consequence analytics tool may stop at software testing and shadow mode. An autonomous controller may require the complete chain plus an operational safety case.

13 Reproduce communications and timing

Control systems depend on timing, message order, quality flags and communications failure. A data set that contains accurate values without realistic timing can overstate performance. The test environment should represent sampling, latency, jitter, loss, retries, buffering, clock error and protocol behaviour relevant to the use.

Cyber-physical testing also needs realistic separation between process and network events. An intrusion-detection system can exploit artefacts of a laboratory generator if attacks always occur at fixed times or use different address patterns from normal traffic. The scenario should vary timing and route while preserving causal effects on the process. Benign maintenance and unusual normal operations should be included to test false alarms.

The environment should use controlled network zones and interfaces. It should never require unrestricted connectivity to production operational technology. NIST guidance emphasises the safety, reliability and performance characteristics of OT. Test systems should preserve safe isolation, controlled import and export, signed artefacts, logging and recovery.

14 Evaluate artificial-intelligence models against decision loss

Synthetic data are often proposed for training forecasting, anomaly-detection or control models. Model accuracy on synthetic holdouts is weak evidence because the generator and learner may share assumptions. Evaluation should include transfer to protected real data, decision consequences and robustness to distribution shift.

For forecasting, measures can include absolute error, probabilistic calibration, ramp performance and downstream imbalance or reserve decisions. For anomaly detection, precision, recall, false-alarm burden, missed high-consequence events and time to detection matter. For control, stability, constraint violations, recovery and safe-state behaviour matter. The programme should report performance by operating regime and event class.

Synthetic pretraining can still be useful. It can teach general structure, provide labelled events and accelerate integration. Fine-tuning or calibration on limited real data may close part of the gap. The evidence should identify which value comes from synthetic pretraining and which requires protected real data. A model should not receive operational authority merely because it performs well on the environment that created its training records.

Table 2. Minimum validation scorecard for synthetic operational data

Dimension	Example measure	Comparator	Failure signal	Decision owner
Structure	Network degree, impedance and asset-class distributions	Public and protected benchmark statistics	Material class or topology mismatch	System planning
Time	Ramps, autocorrelation, duration curves and forecast error	Protected holdout periods	Smoothed tails or broken dependence	Operations analytics
Physics	Balance, limits, stability and equipment capability	Engineering simulator and accepted cases	Hidden corrections or infeasible states	Engineering authority
Controls	Response time, interlocks, saturation and fallback	Approved functional specification	Safe-state or sequencing failure	Control-system owner
Cyber	Protocol, timing, attack coverage and false alarms	Threat model and normal-traffic baseline	Detection relies on test artefacts	Cybersecurity owner
Privacy and confidentiality	Disclosure attacks and protected-feature recovery	Approved threat model	Source record or infrastructure secret can be inferred	Data and security owners
Decision value	Error cost, operator outcome or avoided test effort	Current testing process	Technical gain does not change an approved decision	Business owner and finance

Measures and thresholds should be approved for the named decision before results are produced.

15 Design procurement demonstrations that can fail

Vendor demonstrations often use favourable data prepared by the vendor. A utility should control the evaluation protocol, scenarios, metrics and acceptance thresholds. The vendor can receive development data and a schema, while hidden scenarios remain with an independent evaluator. The process should include normal, stressed, missing-data and adversarial cases.

Acceptance should cover accuracy, reliability, timing, explainability, integration, cyber controls and safe failure. The utility should record configuration effort and manual intervention. A demonstration that requires repeated data cleaning or vendor adjustment may reveal a high implementation cost even when the final score is good.

The contract should distinguish benchmark acceptance from production acceptance. Synthetic-data results can support a down-selection, prototype or conditional award. Final acceptance may require secure real-data validation, shadow operation and field evidence. Payment milestones can be linked to these stages. Intellectual-property terms should preserve the utility's access to test artefacts, results and documentation needed for audit and future competition.

16 Test cyber defences in a cyber-physical range

A cyber range allows teams to exercise attacks and defensive controls without risking a production network. For utility use, the range should connect network activity to physical or simulated process effects. A denial-of-service event, false-data injection or malicious command should influence the simulated grid in a credible way. A purely information-technology range can miss the operational consequence and safety response.

The attack library should draw on current threat intelligence, MITRE ATT&CK for industrial control systems, CISA scenarios and utility-specific hazards. It should cover compromised credentials, malicious remote access, command manipulation, firmware or software supply chain, historian interference, time synchronisation, false measurements and loss of communications. Defenders should also see benign maintenance, device failure and unusual weather-driven operations.

The exercise record should preserve scenario, attacker capability, detection, operator response, containment, recovery and evidence gaps. Results should change controls, playbooks, architecture or procurement. Repeated dramatic exercises without tracked remediation create limited value. Synthetic data and range scenarios should be refreshed when technology, network architecture and adversary behaviour change.

17 Govern generative models used for synthesis

Generative models can create high-dimensional time series and records, but they introduce model risk. The utility should document training data, architecture, objectives, constraints, privacy method, tuning, random seeds and known failure modes. Model cards and data sheets can support this record, although utility assurance needs decision-specific evidence.

The generator can memorise examples, smooth rare events, create impossible combinations or amplify bias. A large language model used to generate event narratives or commands can hallucinate unsupported equipment, protocols or actions. A diffusion or adversarial model can match visual or statistical patterns while breaking causal structure. Physics-informed constraints help only for the equations and limits that are encoded.

Human review should be assigned by domain. Engineers assess physical and control behaviour. Cyber specialists assess protocols and attacks. Privacy specialists assess disclosure. Operators assess workflow and alarms. Model-risk reviewers assess stability and change. A single composite quality score should not replace these approvals.

18 Quantify uncertainty and preserve limitations

Synthetic data add uncertainty from source sampling, parameter estimation, model structure, generation and simulation. Downstream analysis adds further uncertainty. The programme should identify these components and show which decisions are sensitive to them. Where formal propagation is impractical, scenario ranges and structured sensitivity can still expose fragility.

Model discrepancy should remain visible. Calibration can make a simulator match selected outputs while compensating for wrong parameters. Validation against several variables and events reduces this risk. Uncertainty bounds should widen outside the calibrated operating region. Extrapolation to new technologies, weather extremes or control regimes should be treated as a new use requiring evidence.

Documentation should state unsupported uses in reader-facing language. For example, a distribution data set may support hosting-capacity software integration but not protection coordination. A generated customer profile may support interface testing but not tariff design. These statements should travel with the data and appear in access approvals and procurement packs.

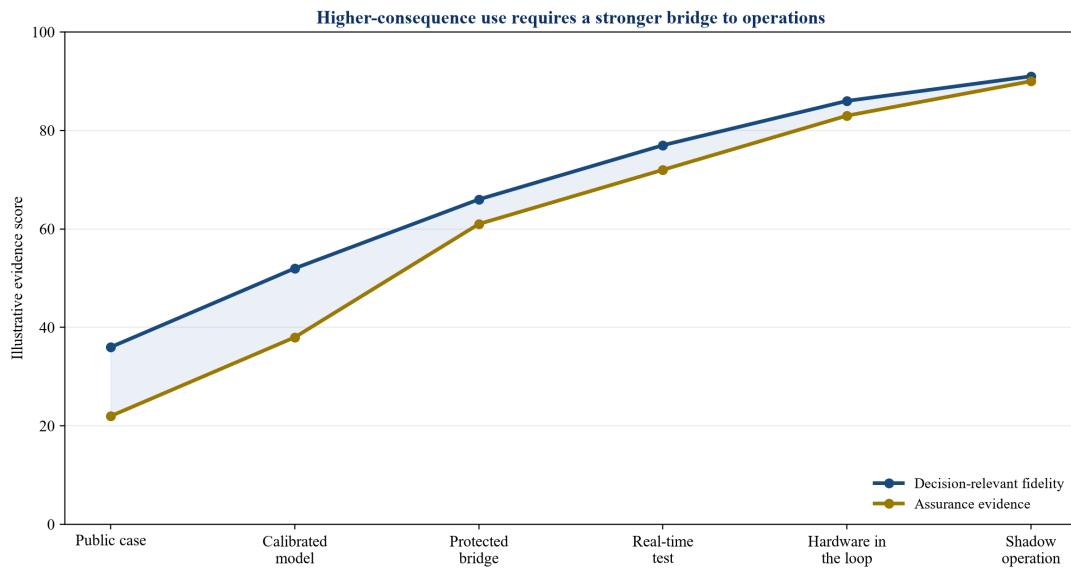
19 Construct the cost and value case

Synthetic operational data can reduce the amount of work performed inside restricted environments, accelerate repeatable testing and enable wider collaboration. It also creates cost. The lifecycle includes source preparation, engineering models, generator development, privacy and security, computing, test infrastructure, protected bridge validation, documentation, support and refresh.

The baseline should be the current testing process. Benefits can include reduced secure-lab time, faster vendor evaluation, fewer integration defects, reusable scenarios, safer training and improved research reproducibility. These benefits should be measured through observed cycle time, defect discovery, avoided rework or approved operating outcomes. A synthetic-data platform should not claim the entire value of the application being tested.

Shared data and test infrastructure can create option value across several use cases. The case should still identify which components are reusable. A network model may be reusable while a control-response model is application specific. Commercialisation requires rights to the generator, documentation, source inputs and validation evidence. A service-heavy bespoke process can limit software margins.

Figure 2. Hypothetical readiness and evidence curve for synthetic grid data



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

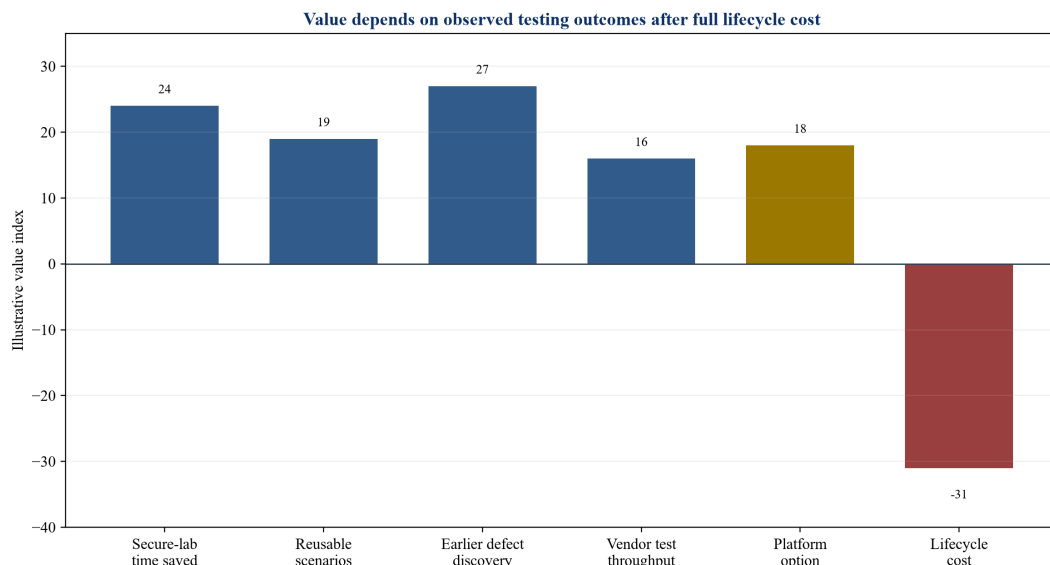
20 Use scenario valuation rather than one forecast

An investment committee can evaluate three scenarios. The utility case assumes the data set supports training, software integration and early vendor screening. The assurance case assumes the programme adds protected bridge testing, real-time simulation and reusable acceptance protocols. The platform case assumes that several utilities, laboratories or vendors adopt the environment under governed commercial terms.

Probability weights, adoption rates and savings are management assumptions and should remain visible. Value should be linked to milestones such as completed fidelity contracts, independent validation, privacy approval, repeatable onboarding and observed reductions in test time or defects. Platform value should not be paid for before rights, portability and customer evidence exist.

For an acquisition or partnership, current assets can include documented data sets, generators, scenario libraries, connectors, validation evidence, laboratory access, contracts and specialist staff. Future applications remain contingent. Earn-outs, staged funding or options can link consideration to evidence, subject to current legal, tax and accounting advice.

Figure 3. Hypothetical three-year value bridge for a synthetic-data programme



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

21 Build a stage-gated operating model

The programme should separate data creation from approval for use. A data-product owner manages provenance, versions and access. Engineering owners define fidelity and physical validation. Cybersecurity and privacy owners define threat models and release controls. The operational decision owner approves the application. Independent reviewers challenge evidence before higher-consequence use.

Stage gates prevent a technically impressive artefact from bypassing controls. The define stage records the decision and fidelity contract. The build stage creates a reproducible generator and baseline. The validate stage tests physics, statistics, events and disclosure. The bridge stage compares against protected evidence. The exercise stage uses real-time or hardware-in-the-loop environments. The operate stage begins with shadow use and monitored authority.

Budgets should be released against evidence. A failed gate should lead to redesign, narrower scope or termination. Keeping these options explicit protects the utility from continuing because of sunk cost. The repository should preserve approved versions, tests, exceptions and decisions.

Table 3. Stage-gated commercialisation and deployment roadmap

Stage	Core work	Required evidence	Commercial or operating decision	Stop condition
Define	Name decision, users, hazards and fidelity contract	Approved scope, prohibited uses and comparator	Fund a bounded build	No material decision or safe test boundary
Build	Create generator, schemas and reproducible scenarios	Versioned artefacts, feasible cases and baseline tests	Fund independent validation	Irreproducible or physically incoherent output
Validate	Test distributions, physics, tails and disclosure	Scorecard within pre-approved tolerances	Approve protected bridge	Material fidelity or disclosure failure
Bridge	Compare with restricted real evidence	Independent transfer results and limitations	Approve real-time or procurement test	Synthetic success does not transfer
Exercise	Run cyber range, real-time or hardware-in-the-loop tests	Timing, fallback, operator and failure evidence	Approve shadow operation	Unsafe interface or unacceptable failure behaviour
Scale	Refresh scenarios and measure realised outcomes	Repeatable onboarding and finance-approved value	Expand, commercialise or retire	Lifecycle cost exceeds realised benefit

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

22 Plan the first ninety days

Days one to fifteen should define the decisions, systems, users, data restrictions and owners. The team writes the fidelity contract, threat model, release tiers and stop conditions. It selects a public synthetic baseline and identifies the protected evidence required for later validation.

Days sixteen to thirty-five should build the data inventory, schemas and reproducible generation pipeline. Engineers verify units, topology, power flow or process balance. Data specialists measure distributions and dependence. Cyber teams define protocols, traffic and attack scenarios. The team creates a versioned scenario library.

Days thirty-six to sixty should run independent validation. Tests cover normal operation, tails, rare events, missing data, timing, privacy and infrastructure confidentiality. Candidate algorithms or vendor systems run against hidden cases. Results are stratified by operating regime and linked to decision loss.

Days sixty-one to seventy-five should execute the protected real-data bridge inside an approved environment. Days seventy-six to ninety should review the evidence, commercial case, contracts, laboratory requirements and next-stage budget. The decision can approve a real-time pilot, narrow the use, redesign the generator or stop.

23 Define credible success conditions

Technical success requires reproducible data generation, documented provenance, physical feasibility and decision-relevant fidelity. Security success requires controlled environments, realistic protocols, adversarial coverage and safe isolation from production. Privacy and confidentiality success require an approved threat model and tests showing that protected records or infrastructure details remain within tolerance.

Operational success requires realistic timing, understandable scenarios, operator acceptance, fallback and a clear authority boundary. Commercial success requires an observed improvement in testing throughput, defect discovery, secure-environment use or another finance-approved outcome after lifecycle cost. Strategic success can include reusable scenarios, stronger procurement, external collaboration and a governed route for artificial-intelligence development.

The programme should state evidence that would disprove its thesis. Examples include failure to transfer from synthetic to protected real data, repeated false confidence under rare events, inability to prevent disclosure, laboratory results that cannot be reproduced, or lifecycle cost above realised benefit. These conditions prevent the success definition from moving after results arrive.

24 Recognise common failure modes

The first failure is using visual plausibility as validation. A chart can look credible while breaking physical or temporal relationships. A second failure is reporting an average similarity score that conceals tail, asset-class or regional weakness. A third is training and testing on data produced by the same generator without an independent bridge.

Other failures include undocumented source rights, exact or near reproduction of sensitive records, missing protection data, unrealistic attack timing, hidden simulator corrections, incomplete control logic and favourable vendor scenarios. Data sets can become stale as topology, distributed resources, markets and operating practices change. Teams can also reuse a data set for a higher-consequence purpose than its evidence supports.

Governance can fail when no owner controls versions, release and expiry. Commercialisation can fail when every customer requires bespoke calibration and laboratory work. Procurement can fail when passing a synthetic benchmark is mistaken for production acceptance. The assurance gate manages these risks by tying use to explicit evidence and authority.

25 Implications for utilities, investors and advisers

Utilities should treat synthetic operational data as reusable test infrastructure. Public cases support common baselines. Utility-calibrated models support decision-specific development. Protected bridge tests establish transfer. Real-time and hardware-in-the-loop environments address timing and physical interaction. Shadow operation protects the authority boundary before production use.

Investors should examine data rights, generator provenance, fidelity evidence, customer-specific work, validation assets, secure-laboratory dependence and recurring refresh cost. A large synthetic data set has limited value when it lacks a named use and protected bridge. A smaller, governed scenario library can be more valuable when it drives procurement, assurance or repeatable customer outcomes.

Advisers can connect engineering, data governance, cybersecurity, procurement and finance. A useful mandate defines the decision, structures the validation contract, tests the commercial model, prices lifecycle cost and creates milestones for funding or transaction terms. Claims about safety, privacy or value should remain tied to evidence from the complete system.

26 Limitations

This paper does not establish that a particular synthetic data set is safe, private or fit for operational use. Requirements depend on the utility, jurisdiction, system, data, model, threat, vendor and decision. Public synthetic grids vary in scope and were developed for different studies. Results from one case do not establish transfer to another network.

Differential privacy addresses defined risks to individual records and can reduce utility. It does not by itself protect critical-infrastructure topology, proprietary operating settings or commercial strategy. Statistical similarity does not establish physical or causal fidelity. Physics-based simulation depends on model structure and

parameters. Hardware-in-the-loop tests depend on interface calibration and laboratory scale.

The figures and tables in this paper are decision aids using hypothetical author-developed assumptions. They do not represent measured performance, market pricing or expected returns. Commercial structures require current legal, tax, accounting, regulatory, insurance, cybersecurity and engineering advice.

Conclusion

Synthetic operational data can help utilities test systems, train people, evaluate vendors and develop analytics without broadly exposing production records. Their value arises from a governed connection to a named decision. Public synthetic networks provide reproducible baselines. Simulation and cyber ranges expose defects. Real-time and hardware-in-the-loop environments test timing and physical interaction. A protected real-data bridge establishes whether synthetic success transfers.

The Synthetic Operational Data Assurance Gate makes that connection explicit. It records provenance, defines decision-relevant fidelity, tests physics and tails, challenges disclosure, preserves operational-technology boundaries and controls release authority. Stage-gated funding keeps experimentation bounded. The resulting capability is a reusable test system with documented limits, accountable decisions and evidence that can support procurement, investment and controlled deployment.

Frequently Asked Questions

Can synthetic operational data replace all real utility data?

No. Synthetic data can support development, testing, training and early procurement. Higher-consequence decisions usually require a protected comparison with real operational evidence, followed where appropriate by shadow operation or field validation.

Does synthetic data guarantee privacy or infrastructure confidentiality?

No. Generators can reproduce source records or reveal protected relationships. The programme needs a threat model, disclosure tests, release tiers and controls for both personal information and critical-infrastructure information.

What is the best first use for a utility?

Software integration, analyst training or vendor screening on a documented public synthetic case can provide an efficient first use. The decision and limitations should be explicit before utility-specific calibration begins.

How should realism be measured?

The utility should create a fidelity contract covering the structure, time dependence, physics, control behaviour, cyber characteristics and human workflow relevant to the named decision. Each property needs a comparator, tolerance and owner.

Why is a protected real-data bridge necessary?

A generator and its synthetic holdout can share assumptions. A controlled comparison with independent protected evidence tests whether the result transfers to real operating conditions without broadly releasing the source data.

When is hardware-in-the-loop testing required?

It is appropriate when physical devices, firmware, communications timing, controller limits or closed-loop interaction materially affect safety or performance. Low-consequence analytics may use software testing and shadow operation instead.

How should synthetic data be used in vendor procurement?

The utility should own the protocol, use hidden scenarios, define acceptance thresholds in advance and separate synthetic benchmark acceptance from production acceptance. Milestones can require protected validation and shadow evidence.

When should a synthetic-data programme stop?

It should stop or narrow scope when output is physically incoherent, transfer to protected real data is weak, disclosure risk remains outside tolerance, interfaces are unsafe, the intended decision does not improve or lifecycle cost exceeds realised value.

References

1. Birchfield, A.B. et al., Grid Structural Characteristics as Validation Criteria for Synthetic Networks, IEEE Transactions on Power Systems, 2017. <https://doi.org/10.1109/TPWRS.2016.2616385>
2. Birchfield, A.B. et al., A Metric-Based Validation Process to Assess the Realism of Synthetic Power Grids, Energies, 2017. <https://doi.org/10.3390/en10081233>
3. Birchfield, A.B. et al., The Power Flow Cases of Systems with Thousands of Buses, IEEE Transactions on Power Systems, 2018. <https://doi.org/10.1109/TPWRS.2018.2849017>
4. Xu, Y. et al., US Test System with High Spatial and Temporal Resolution for Renewable Integration Studies, 2020. <https://doi.org/10.1109/TPWRS.2020.3002093>
5. Xu, Y. et al., Synthetic Test Systems for Transmission and Distribution Networks, IEEE Transactions on Power Systems, 2021. <https://doi.org/10.1109/TPWRS.2020.3037099>
6. NREL, SMART-DS Synthetic Distribution System Data Sets. <https://www.nrel.gov/grid/smart-ds.html>
7. NREL, Synthetic Data Sets Set New Standard in Energy Systems Research, 2020. <https://www.nrel.gov/news/program/2020/synthetic-data-sets-set-new-standard.html>
8. Grid Modernization Laboratory Consortium, Reliability Test System Grid Modernization Laboratory Consortium. <https://github.com/GridMod/RTS-GMLC>
9. Barrows, C. et al., The IEEE Reliability Test System: A Proposed 2019 Update, IEEE Transactions on Power Systems, 2020. <https://doi.org/10.1109/TPWRS.2019.2925557>
10. Aravena, I. et al., OPEN COG Grid: Extendable Coherent Models-Datasets for Cognitive Power Grids, Performance Year 1 Technical Report, 2024. <https://doi.org/10.2172/2484228>
11. Aravena, I. et al., Open Power System Datasets and Open Simulation Engines: A Survey Towards Machine Learning Applications. <https://www.osti.gov/pages/servlets/purl/2571673>
12. NREL, Grid Research Facilities and ARIES. <https://www.nrel.gov/grid/facilities.html>
13. NREL, ARIES and Advanced Distribution Management System Test Bed, 2025. <https://www.nrel.gov/docs/fy25osti/92384.pdf>
14. United States Department of Energy, National SCADA Test Bed. <https://www.energy.gov/oe/national-scada-test-bed>
15. NIST, Guide to Operational Technology Security, SP 800-82 Revision 3, 2023. <https://doi.org/10.6028/NIST.SP.800-82r3>
16. NIST, Guidelines for Evaluating Differential Privacy Guarantees, SP 800-226, 2025. <https://doi.org/10.6028/NIST.SP.800-226>
17. NIST, De-Identifying Government Datasets: Techniques and Governance, SP 800-188, 2023. <https://doi.org/10.6028/NIST.SP.800-188>
18. NIST, Artificial Intelligence Risk Management Framework 1.0, 2023. <https://www.nist.gov/itl/ai-risk-management-framework>
19. NIST, Privacy Framework 1.0, 2020. <https://www.nist.gov/privacy-framework>
20. NIST, SDNist Synthetic Data Report Tool. <https://www.nist.gov/services-resources/software/sdnist-synthetic-data-report-tool>
21. CISA, Compendium of Cybersecurity and Infrastructure Security Agency Technical Investigations, 2024. https://www.cisa.gov/sites/default/files/2024-05/Compendium%20CISA%20Tech%20Investigations%2020240424%20SPP_V2_508c.pdf
22. CISA, Industrial Control Systems Recommended Practices. <https://www.cisa.gov/resources-tools/resources/ics-recommended-practices>
23. CISA, Cybersecurity Scenarios for Critical Infrastructure, 2023. <https://www.cisa.gov/resources-tools/resources/cybersecurity-scenarios>
24. MITRE, ATT&CK for Industrial Control Systems. <https://attack.mitre.org/matrices/ics/>
25. IEEE, IEEE Standard for Synchrophasor Measurements for Power Systems, IEEE C37.118.1-2011. <https://standards.ieee.org/ieee/C37.118.1/4902/>
26. IEEE, IEEE Standard for Interconnection and Interoperability of Distributed Energy Resources, IEEE 1547-2018. <https://standards.ieee.org/ieee/1547/5915/>
27. International Electrotechnical Commission, IEC 61850 Power Utility Automation. <https://www.iec.ch/basecamp/iec-61850-power-utility-automation>
28. North American Electric Reliability Corporation, Critical Infrastructure Protection Standards. <https://www.nerc.com/pa/Stand/Pages/CIPStandards.aspx>
29. Zimmerman, R.D., Murillo-Sanchez, C.E. and Thomas, R.J., MATPOWER: Steady-State Operations, Planning and Analysis Tools for Power Systems Research and Education, IEEE Transactions on Power Systems, 2011. <https://doi.org/10.1109/TPWRS.2010.2051168>
30. Thurner, L. et al., pandapower: An Open-Source Python Tool for Convenient Modeling, Analysis, and Optimization of Electric Power Systems, IEEE Transactions on Power Systems, 2018. <https://doi.org/10.1109/TPWRS.2018.2829021>
31. Coffrin, C. et al., PowerModels.jl: An Open-Source Framework for Exploring Power Flow Formulations, 2018. <https://doi.org/10.1109/PWRS.2018.2832107>
32. NREL, Hierarchical Engine for Large-Scale Infrastructure Co-Simulation. <https://helics.org/>
33. Pacific Northwest National Laboratory, Framework for Network Co-Simulation. <https://github.com/FNCS/fncs>
34. Pacific Northwest National Laboratory, GridLAB-D. <https://www.gridlabd.org/>
35. Electric Power Research Institute, OpenDSS. <https://sourceforge.net/projects/electricdss/>
36. Oyewumi, I.A. et al., Cyber-Physical System Testbed Architectures for Wide-Area Monitoring, Protection and Control, IEEE Access, 2023. <https://doi.org/10.1109/ACCESS.2023.3283271>
37. NREL, Generative AI for Power Grid Applications, 2025. <https://www.nrel.gov/docs/fy25osti/91176.pdf>
38. United States Department of Energy CESER, Stormbreaker Testbed for AI Evaluation in Critical Infrastructure, 2026. <https://www.energy.gov/ceser/articles/ceser-releases-new-testbed-advance-llm-and-agentic-ai-evaluation-critical>
39. Sargent, R.G., Verification and Validation of Simulation Models, Journal of Simulation, 2013. <https://doi.org/10.1057/jos.2012.20>
40. Oberkampf, W.L. and Roy, C.J., Verification and Validation in Scientific Computing, Cambridge University Press, 2010. <https://doi.org/10.1017/CBO9780511760396>
41. NASA, Standard for Models and Simulations, NASA-STD-7009B, 2024. <https://standards.nasa.gov/standard/nasa/nasa-std-7009>
42. European Parliament and Council, Regulation (EU) 2016/679, General Data Protection Regulation. <https://eur-lex.europa.eu/eli/reg/2016/679/oj>
43. European Parliament and Council, Regulation (EU) 2024/1689, Artificial Intelligence Act. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>

44. International Organization for Standardization, ISO/IEC 27001 Information Security Management Systems. <https://www.iso.org/isoiec-27001-information-security.html>
45. International Organization for Standardization, ISO/IEC 42001 Artificial Intelligence Management Systems. <https://www.iso.org/standard/81230.html>
46. International Organization for Standardization, ISO 31000 Risk Management. <https://www.iso.org/iso-31000-risk-management.html>
47. IFRS Foundation, IAS 38 Intangible Assets. <https://www.ifrs.org/issued-standards/list-of-standards/ias-38-intangible-assets/>
48. 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.