International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(23s) 2026 pp. 1707-1720
https://doi.org/10.70917/ijcisim-2026-5723
Article

Physics-Constrained Synthetic Data Generation For Industrial Machine Learning: Enforcing Conservation And Operating-Envelope Constraints, Validating Fidelity Against Sensor Records, And Measures

Arjun Gopichander Ravichander

Independent Researcher, India
Email: arjungr97@gmail.com

Abstract: Synthetic data generators optimise for distributional realism [4][5][6][7]. They match means, variances, and correlations, and they do not match physical law. The consequence for industrial systems is a specific and under-examined failure: a generator can produce sensor traces statistically indistinguishable from real records while violating mass and energy balance, exceeding actuator limits, or exhibiting impossible state transitions such as temperature rising with no energy input. A downstream model trained on such data learns spurious correlations rather than system dynamics, and fails silently at the rare, physically extreme operating points where reliability matters most. This paper proposes a method for generating industrial sensor data under explicit physical constraints, and specifies how it should be evaluated. It identifies four classes of constraint that an industrial generator must respect, covering conservation laws, operating-envelope bounds, monotonicity and irreversibility, and state-transition legality. It compares four enforcement mechanisms, in the loss, in the architecture, by rejection sampling, and by post-processing projection, and states what each costs, including the case against each. It specifies a fidelity validation design going beyond marginal-distribution testing to cover cross-sensor coupling, spectral content, direct measurement of constraint violation rates, and temporal dynamics. It specifies a downstream evaluation with named datasets, arms, seeds, and statistical tests. No implementation exists. No code has been written, no generator built, no data generated, and no validation performed. This is a method proposal, and the paper reports no results because there are none.

Keywords: synthetic data generation, physics-informed machine learning [9], industrial sensor data, conservation constraints, operating envelope, predictive maintenance, remaining useful life, model collapse, rare event detection, fidelity validation, train on synthetic test on real

1. INTRODUCTION

1.1 The gap between statistical and physical realism

Generative approaches for tabular and time-series data optimise for distributional realism. They are trained and evaluated on whether generated samples resemble real ones in their statistical properties, and by that criterion modern generators perform well.

Physical law is not among those properties. A generator can therefore produce sensor traces that pass statistical comparison against real records while describing a machine that could not exist: mass or energy that appears or


Copyright: © 2026 by the authors

vanishes across an equipment boundary, an actuator commanded beyond its physical range, a cumulative wear index that decreases, or a temperature that rises with no energy input.

When such data trains a downstream model, that model learns correlations present in the generated data rather than the dynamics of the real system. In industrial deployment this breaks fault-detection thresholds, corrupts digital twins, and produces predictions that fail silently at exactly the rare and physically extreme operating points where reliability matters most, because those are the regions where statistical resemblance is weakest and physical constraint is most binding

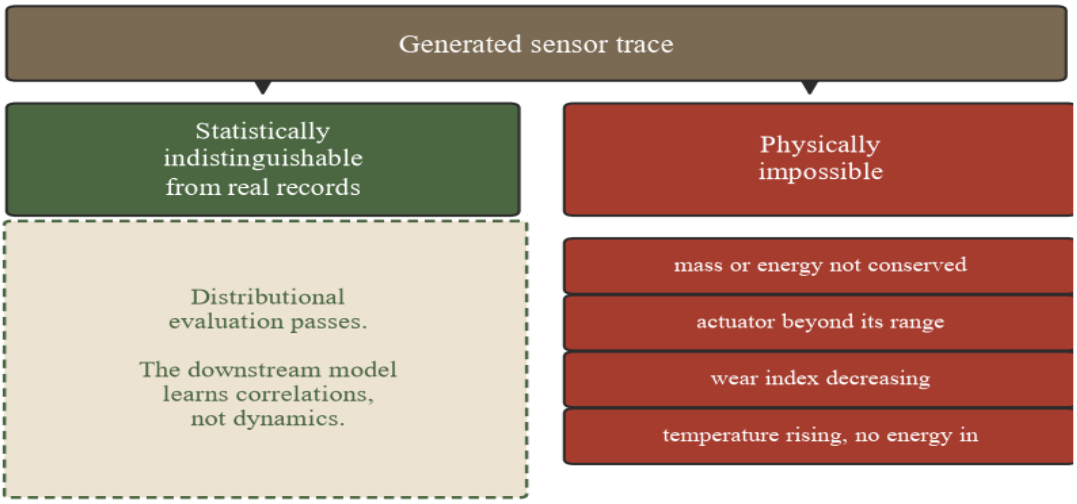

Figure 1. The gap distributional evaluation cannot detect. A trace can be statistically indistinguishable from real records and physically impossible.

1.2 Status of this work

This paper proposes a method. Nothing has been built.

There is no prior implementation, no code, no prototype, and no generated dataset. No validation has been performed against sensor records, no downstream model has been trained on generated data, and no comparison against existing generative baselines has been run. The author holds no industrial dataset for validation at the time of writing.

This is stated at the outset, repeated in Section 13, and reflected in the absence of any results section. The paper's contribution is a specification: which constraints matter, how each might be enforced and at what cost, how fidelity should be measured, and what experiment would establish whether the approach works. Whether it does work is an open question this paper poses rather than answers.

1.3 Contribution

1. A taxonomy of the constraint classes an industrial sensor generator must respect, stated concretely rather than as an appeal to physical plausibility in general.

2. A comparison of four enforcement mechanisms with an explicit account of what each costs, including the argument against each.

3. A fidelity validation design covering the properties that generic tabular validation misses, in particular cross-sensor coupling, spectral content, and direct measurement of constraint violation.

4. A downstream evaluation design specified to the level of arms, seeds, and statistical tests, so that the central claim is falsifiable.

5. A scoping of where the approach does not apply, which is narrower than the literature in this area typically admits.

2. Related Work

Four strands bear on this proposal, and one of them argues against its premise.

Physics-based simulation for industrial failure data. Work on predictive maintenance for industrial robotic manipulators has built physics-based simulators generating multi-sensor synthetic failure data. This is the closest existing approach to a working system of the kind proposed here, and it establishes that the direction is practical. It is a simulator rather than a learned generator, which means it produces what its equations describe and cannot capture behaviour those equations omit.

Thermodynamic constraints as training conditions. Recent work embeds the second law of thermodynamics as a training constraint for machine learning applied to chemical kinetics, with residual-based synthetic augmentation. This is conceptually close to the loss-based enforcement discussed in Section 5, and it is domain-specific to combustion and chemistry rather than to general industrial equipment.

The counterargument, which this paper must address. Recent work on generative modelling for photonic grating-coupler spectra argues something that cuts directly against the premise here: that explicit conservation enforcement is often mathematically redundant when the underlying equations are already consistent, and that only certain physical constraints meaningfully change the outputs. If that generalises, then a substantial part of the machinery proposed in this paper buys nothing.

This paper takes that argument seriously rather than noting it and moving on, and the response has two parts. The redundancy claim is strongest where the generator is already parameterised in terms of the governing equations, which is not the case for a generator learning from sensor records without an explicit forward model. And the claim concerns conservation constraints specifically, which are one of four classes in Section 4; operating-envelope bounds, monotonicity, and transition legality are not implied by equation consistency and would not be redundant even if the argument holds in full. The correct response is to treat constraint value as an empirical question per constraint class, which is why the ablation in Section 7 separates them rather than evaluating the composed system only.

Physics-informed machine learning surveys. Recent surveys describe the general concept at a high level. They do not build sensor-validated industrial systems, which is the gap this proposal addresses.

Model collapse and the tail [13]. A separate concern bears on this proposal and is worth naming, because it interacts with the enforcement choice in Section 5. Generators trained on or augmented with their own output are known to lose distributional tails progressively, converging toward the mode of the distribution they were trained on. In an industrial setting the tail is not a statistical curiosity; it is where the failures are. Any augmentation strategy therefore has to be evaluated on whether it preserves tail mass rather than only on whether it improves aggregate metrics, and rejection sampling is the enforcement mechanism most likely to accelerate the loss.

Adjacent work on operating envelopes. Work on decision support in industrial tasks has generated synthetic manufacturing data described as falling within a feasible operating envelope. Envelope-matching in that setting is statistical range-matching rather than physics constraint enforcement, and fidelity validation is not its focus. The distinction between matching a range and enforcing a constraint is precisely the distinction this paper is built on.

No published work appears to combine constraint enforcement across all four classes with sensor-record fidelity validation and a downstream utility evaluation for general industrial equipment.

3. Constraint Classes

Four classes of constraint must be respected, and they differ in whether they can be expressed in closed form, which determines the enforcement options available in Section 5.

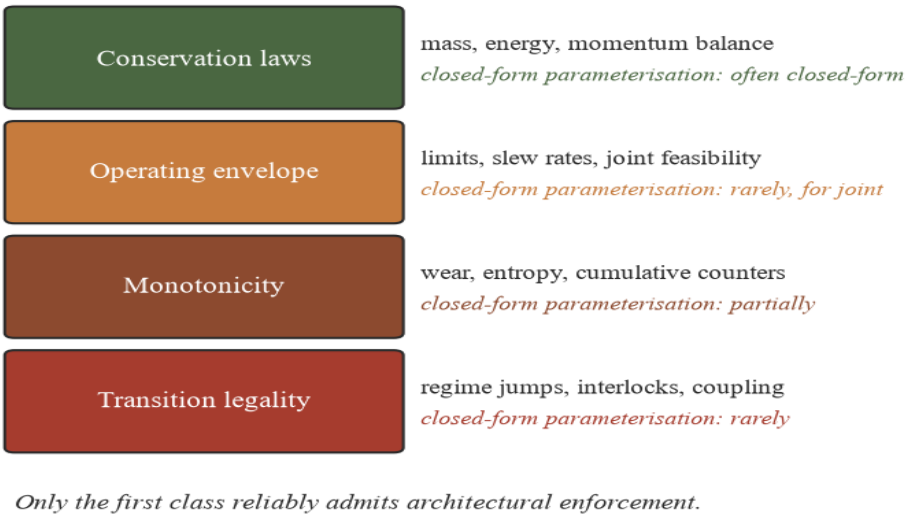

Figure 2. The four constraint classes. Only conservation reliably admits a closed-form parameterisation, which bounds the enforcement options.

Table 1. Constraint classes for industrial sensor generation.

Class	**Constraints**	**Closed-form parameterisation**
Conservation laws	Mass balance, inflow equals outflow plus accumulation; energy balance, first-law consistency between input power, output work, losses, and storage change; momentum and torque balance for rotating equipment	Often available
Operating-envelope bounds	Actuator and sensor physical limits; rate-of-change limits on temperature, pressure, and flow; joint-state feasibility, where values individually in range are jointly impossible	Rarely available for joint feasibility
Monotonicity and irreversibility	Cumulative wear and degradation indicators never decreasing; entropy production non-negative; cumulative counters strictly non-decreasing	Partially available
State-transition legality	No discontinuous jumps between operating regimes without a valid intermediate trajectory; startup and shutdown sequencing with interlocks and dwell times; correlated-failure realism, where sensors that should co-move remain coupled	Rarely available

The third column matters more than it appears. Architectural enforcement, which is the strongest guarantee available, requires a closed-form parameterisation, and only the first class reliably has one. This is why Section 5 concludes that no single enforcement mechanism suffices.

The joint-state feasibility case deserves particular note, because it is the constraint most often omitted. A generator that respects every individual sensor's range can still produce high flow at a fully closed valve, and no per-channel bound detects that. Joint feasibility is a property of the combination, and expressing it requires knowing which combinations the equipment permits.

4. Enforcement Mechanisms and What Each Costs

Four mechanisms are available. Each is presented with the argument against it, because the choice between them is a trade rather than a ranking.

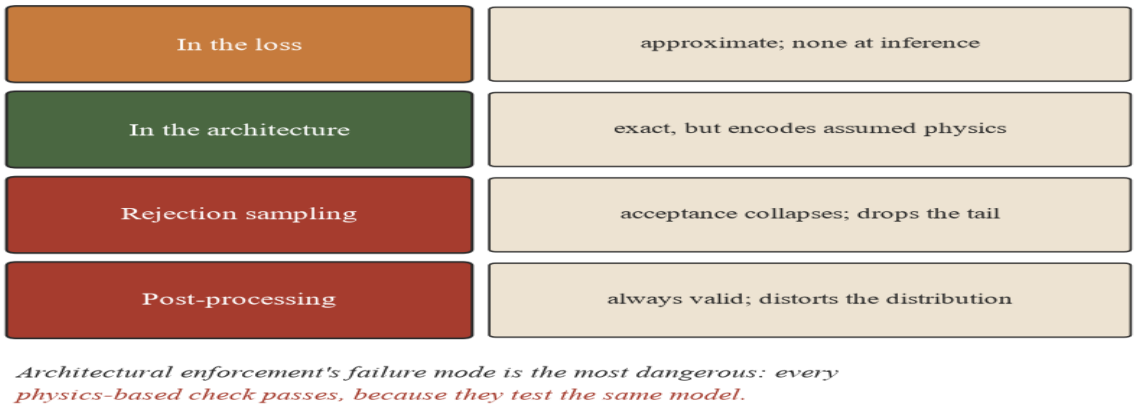

Figure 3. Enforcement mechanisms and what each costs. No single mechanism suffices across the four constraint classes.

Table 2. Enforcement mechanisms compared.

Mechanism	**Guarantee**	**Principal cost**
Penalty in the loss [8]	Approximate; none at inference	Weight trades fidelity against physical correctness; a sample can still violate balance, just less. Adds hyperparameter tuning burden. If the residual is non-differentiable or stiff, gradients become noisy and training destabilises
Architectural construction [10][11]	Exact, by construction	Requires a known closed-form parameterisation. If the assumed governing equations are wrong or incomplete for the real equipment, the architecture enforces the wrong physics with total confidence, which is worse than a soft penalty that lets data correct it
Rejection sampling	As clean as the validity check	Acceptance collapses as constraints stack; five independent constraints each passing 80 percent of samples give roughly 33 percent joint acceptance, worsening exponentially. Rare, physically extreme states are rejected disproportionately, reintroducing tail loss
Post-processing projection	Always produces a valid sample	Silently distorts the learned distribution; the correction is a deterministic patch rather than something learned, so statistical fidelity degrades exactly where correction is heaviest. It also masks generator failure, hiding badly infeasible outputs rather than fixing them

Three observations follow.

Architectural enforcement's failure mode is the most dangerous. A penalty that is wrong produces samples that are imperfectly constrained, which a violation-rate check will detect. An architecture encoding the wrong physics produces samples that are perfectly consistent with an incorrect model, and every physics-based validation check will pass, because those checks test consistency with the same model. Section 9 develops this as a general limit.

Rejection sampling is worst exactly where the method is most needed. The tail events that motivate synthetic augmentation for fault detection are the ones a stacked constraint check rejects most often, so the mechanism removes the samples the exercise exists to produce.

Post-processing undermines the paper's own validation claims. If heavy correction is applied after generation, a low violation rate measures the corrector rather than the generator, and the fidelity evidence in Section 6 becomes circular.

The proposal is therefore a composition rather than a choice: architectural construction for conservation constraints where a parameterisation exists and confidence in the governing equations is high, loss penalties for constraints without closed form, and violation-rate measurement reported on uncorrected output so the generator's own behaviour remains visible. Post-processing is used, if at all, as a final admissibility filter whose action is reported rather than absorbed.

5. Fidelity Validation Design

Time-series sensor records have structure that marginal-distribution testing does not reach, so validation is specified across four categories.

Marginal and joint statistical fidelity. Per-sensor distribution comparison between real and synthetic marginals for each channel, computed per operating regime rather than pooled, since pooling hides regime-specific failure. Cross-sensor correlation structure compared through covariance or copula, which catches the case where individual sensors look correct in isolation while their coupling is wrong. Spectral fidelity for rotating machinery, comparing power spectral density and dominant frequency peaks, since a time-domain distribution test can pass while frequency content is physically nonsensical.

Physical consistency, which generic frameworks omit. Constraint violation rate measured directly on generated samples rather than inferred from training loss, reporting the distribution of residuals rather than the mean, since a low mean can conceal a heavy violating tail. Envelope compliance rate, as the fraction of samples within physical limits and satisfying rate-of-change constraints. Monotonicity checks on cumulative variables, flagging any decrease.

Downstream utility. Training a fault-detection or remaining-useful-life model purely on synthetic data and evaluating on held-out real records, compared against a model trained on real data. The gap between them is the number that matters to a practitioner; statistical similarity is a proxy for it.

Temporal and dynamical fidelity. Autocorrelation and trend comparison over realistic operating cycles covering startup, steady state, and shutdown, since per-timestep tests miss whether trajectories evolve like real ones. Transition-legality audit confirming that synthetic sequences never jump between operating states in ways real equipment cannot.

The second category is the one that distinguishes this design, and reporting the residual distribution rather than its mean is the specific detail that makes it informative.

6. Downstream Evaluation Design

The claim that physics-constrained generation improves downstream performance is empirical, and this section specifies the experiment that would test it.

Task. One supervised task per dataset where real labelled failure examples are genuinely scarce, since scarcity is the premise. Two public benchmarks fit that condition and are proposed here: a turbofan engine run-to-failure simulation corpus [1], whose damage propagation model and generation procedure are documented separately [14], supporting remaining-useful-life regression or failure-imminent classification; and a semiconductor manufacturing dataset [2] whose native pass-to-fail ratio makes it a natural class-imbalance case rather than an artificially rebalanced one. Candidate public datasets are a turbofan degradation dataset for remaining-useful-life regression or binary failure-imminent classification, a semiconductor manufacturing dataset for pass-fail classification with natural class imbalance, and a robotic assembly dataset for fault classification.

Splits. A real test set the generator never sees in any form: not in training the generative model, not in tuning constraint weights, not in setting acceptance thresholds. Leakage here invalidates the entire result. Remaining real

data splits into a small real-train pool simulating the scarce regime, and a real-calibration pool used only to condition the generator's physical parameters.

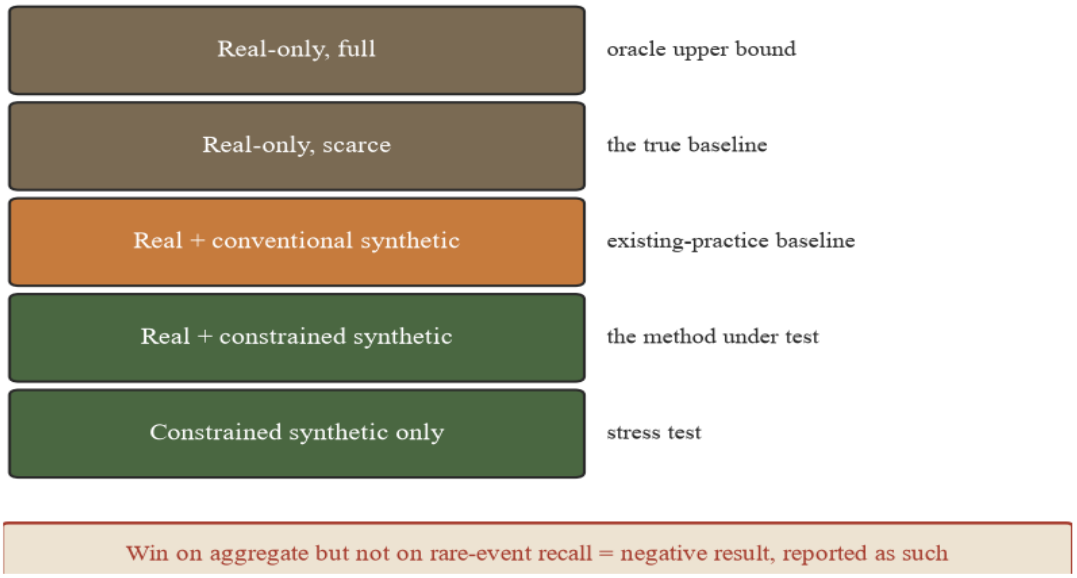

Figure 4. The proposed experimental arms, with acceptance criteria fixed in advance including the condition that would count against the method.

Table 3. Experimental arms.

Arm	**Data condition**	**Purpose**
Real-only, full	All available real data	Oracle upper bound
Real-only, scarce	Reduced failure examples	The true baseline; the realistic condition
Real-scarce plus conventional synthetic	Augmented with an existing generative approach	The existing-practice baseline to beat
Real-scarce plus constrained synthetic	Augmented with the proposed method	The method under test
Constrained synthetic only	No real augmentation	Stress test of how far synthetic alone reaches

The same downstream model architecture is used across arms, fixed rather than tuned per arm, and every arm runs across multiple seeds varying both data sampling and model initialisation. Single-run comparisons cannot support a significance claim.

Protocol. The design follows the train-on-synthetic, test-on-real convention established for generated time series [12]: the downstream model is fitted on generated data and evaluated against held-out real records, so the reported number measures usefulness to a practitioner rather than similarity to a distribution.

Metrics. Primary task metric on the held-out real test set, using error measures for regression and precision-recall based measures for classification rather than accuracy alone given the imbalance. Rare-event recall on real failure cases reported separately from aggregate performance, since that is the tail the method is supposed to protect and where model collapse would appear first. Statistical comparison between the constrained-synthetic and conventional-synthetic arms by paired test across seeds, rather than a difference in point estimates.

Acceptance criteria, stated in advance. The constrained-synthetic arm must beat real-scarce-only by a margin larger than the seed-to-seed noise floor established by the baseline arm's own variance. It must beat the conventional-synthetic arm on the primary metric and specifically on rare-event recall. If it wins on aggregate metric but not on rare-event recall, that is a negative result for the core claim and should be reported as such. The gap to the oracle arm should be reported explicitly, since beating a weak baseline and being usable in practice are different claims.

Ablation. Because the method composes several constraint classes, the evaluation isolates them: constraints in the loss alone, then with collapse-prevention blending, then the full composition. This is what permits a claim about which component drives any improvement, and it also bears directly on the counterargument in Section 2, since an ablation showing conservation constraints contribute nothing would support the redundancy claim rather than refute it.

7. Where Synthetic Data Remains Inappropriate

The honest scope of this method is narrower than the framing of much of this literature, and six boundaries define it.

Unobserved failure modes are an epistemic ceiling. A generator can encode only physics known to govern the failure. It can produce novel combinations of known mechanisms; it cannot produce a failure driven by physics not in its model, such as a fatigue mechanism nobody modelled or a multi-component interaction the equations do not capture. This is a structural limit rather than an engineering gap: generation is interpolation and extrapolation within a known physical model, and genuinely novel failure modes are by definition outside it. Any claim about rare failure coverage must be scoped to rare but mechanistically known modes. Conflating the two is the most common overclaim in this area.

Model misspecification masquerades as data. If the governing equations are wrong, incomplete, or idealised, the generator produces confidently wrong data that passes the constraint checks by construction, because it is consistent with the model rather than with reality. This is worse than statistical mimicry failure, because it is undetectable by the very validation this paper proposes. Physics-constrained fidelity checks validate consistency with a model of the system, not with the system.

Some processes lack mature governing equations. Complex chemical reactions with unknown intermediate pathways, certain composite failure modes, and human-in-the-loop variability do not have equations mature enough to constrain a generator meaningfully. Imposing a residual penalty on a poorly characterised system either does nothing or actively misleads by imposing false structure. The public datasets named in Section 6 were chosen partly because they are well characterised, and that is a scope limitation rather than a convenience.

System changes invalidate calibrated physics. Retrofits, new setpoints, firmware updates, or material substitutions shift real behaviour in ways a model calibrated on historical parameters does not reflect. Data generated under stale parameters will be fluent, constraint-satisfying, and describing a system that no longer exists.

Regulatory acceptance does not follow from fidelity. [3] Even with strong fidelity metrics, regulators and safety engineers are frequently unwilling to certify a system whose training data includes synthetic examples of catastrophic failures that have never occurred, because there is no real-world referent to audit against. Statistical and physical fidelity evidence supports a compliance argument; it does not substitute for real incident data where demonstrated real-world validation is required.

Adversarial failure modes are outside physical law. Cyberattack, sabotage, and intentional misuse produce sensor signatures shaped by an adversary's goals rather than by physics. A conservation-respecting generator has no mechanism to anticipate an adversary spoofing sensors within physically plausible bounds, and constraints may worsen the situation, since adversarial data can be crafted to satisfy the fidelity checks precisely.

Figure 5. What lies outside the method's scope. The defensible claim is narrower than solving industrial data scarcity.

The scoped claim. The method, if it works, reduces the gap between statistical and physical fidelity for failure modes within the modelled physics. It does not solve industrial data scarcity. The defensible statement is narrower: it aims to produce trustworthy synthetic data for known, mechanistically characterised failure modes, and should not be relied upon for unmodelled, novel, or adversarial ones.

8. What Would Have to Be Built

The proposal is easier to evaluate if what it entails is stated concretely rather than as a set of principles. Five components would be required, and they differ substantially in difficulty.

A physical parameter set per equipment class. The constraints in Section 3 are not generic; a mass balance requires knowing which measured quantities constitute inflow, outflow, and accumulation for the specific equipment, and an envelope requires the actual limits. Assembling this is domain work rather than machine learning work, and it is the component most likely to be underestimated, because it cannot be derived from the data and must be supplied by someone who knows the equipment.

A base generator. The proposal deliberately does not commit to one. Whichever is chosen determines which enforcement mechanisms are available: architectural construction requires a generator whose output can be reparameterised, and loss-based enforcement requires differentiable residuals.

A constraint evaluation layer, separable from training. This is the component that makes the validation in Section 5 credible. Constraint violation must be measurable on generated samples independently of the training objective, because a violation rate computed from the training loss measures the optimisation rather than the output. Building this as a standalone evaluator, usable against any generator including baselines, is what allows the comparison in Section 6 to be fair.

A missingness model. Production sensor data does not go missing at random. A generator producing complete traces is producing data unlike the data it will be used alongside, and a downstream model trained on it will not encounter the structured gaps it will meet in deployment.

An evaluation harness. The design in Section 6 specifies arms, seeds, and paired statistical tests. Running it requires the harness to hold the test set genuinely separate across every arm and every tuning decision, which is a discipline more often stated than implemented.

The order matters and is not the order of interest. The parameter set is the least interesting component and the one that gates everything else, since no constraint can be enforced or measured without it. A project beginning with the generator, which is the natural instinct, discovers the parameter problem after the architecture is fixed.

9. What Grounds the Proposal

The method is a proposal, and its motivation is not purely theoretical. Two elements of the author's background inform specific design choices, and they are stated as the origin of a concern rather than as evidence for a claim.

Sensor-to-model pipeline experience. Work in an industrial engine-testing environment, converting raw temperature, pressure, and torque sensor readings into predictive models, is the source of the concern about physical fidelity. Raw test-cell sensor streams carry equipment-specific artifacts including sampling irregularities, calibration behaviour, and cross-dependencies between torque, pressure, and temperature that do not appear in benchmark datasets. A pipeline designed against benchmark data assumes properties that real acquisition does not have.

Production data engineering. Experience with large-scale production pipelines informs the treatment of missingness. Production data fails in structured, non-random ways: outages, schema drift, and silent breakage cluster around specific operational events. A generator and its validation therefore need a missingness model tied to real failure and maintenance events rather than a naive assumption that missing readings are missing at random. This is a modelling choice rather than a preprocessing step, and it is one that only production exposure suggests.

Neither of these is an application of the proposed method. They explain why particular constraints and validation properties were selected, and they are not evidence that the method works.

10. How This Could Fail Without Being Wrong

A method proposal is strongest when it states the outcomes that would count against it, and there are three that are more likely than outright failure.

The constraints could be satisfied and irrelevant. The generator could produce data with a near-zero violation rate, excellent envelope compliance, and monotonic wear indices, and the downstream model trained on it could perform no better than one trained on unconstrained synthetic data. This is the outcome the counterargument in Section 2 predicts, and it would mean the constraints are real, correctly enforced, and not the binding limitation on downstream utility. It is not a failure of implementation and it would be reported as a negative result rather than tuned away.

The improvement could be real and concentrated in the wrong place. The method could improve aggregate downstream metrics while leaving rare-event recall unchanged. Since rare events are the entire motivation, this would be a result that looks like success on the primary metric and fails the claim the paper actually makes. Section 7 fixes the acceptance criteria in advance specifically so this outcome cannot be reported as a win.

The enforcement could reduce diversity more than it improves validity. Constraining a generator narrows what it can produce. If the narrowing removes valid-but-unusual regions along with invalid ones, the generated data becomes more physically correct and less useful, because the augmentation value of synthetic data lies in covering regions the real data does not. This failure would show up as improved violation rates alongside degraded rare-event recall, and it is the most likely mechanism by which a technically successful implementation produces a worse downstream result.

Table 4. Outcomes that would count against the method.

Outcome	What it would mean	Detectable by
Low violation rate, no downstream gain	Constraints are not the binding limitation	Comparison against the conventional-synthetic arm
Aggregate gain, no rare-event gain	The claimed benefit is not the observed one	Rare-event recall reported separately
Improved validity, reduced diversity	Enforcement removed useful tail regions	Rare-event recall falling as violation rate falls

Gains only on one dataset	Result is dataset-specific rather than general	Running all three named datasets
Ablation shows one component carries everything	The composition is unnecessary	The per-tier ablation in Section 7

The last row deserves emphasis. If the ablation shows that a single constraint class accounts for the entire improvement, the correct conclusion is that the method should be that one class rather than the composition, and reporting the composed system as the contribution would overstate it. Designing the evaluation to be able to reach that conclusion is what makes the ablation worth running rather than a formality.

11. Evidence

No implementation exists and no results are reported.

Table 5. Status of every claim class in this paper.

Claim class	**Status**
Implementation of the generator	**Does not exist.** No code has been written
Generated data	**None.** No dataset has been produced
Fidelity validation against sensor records	**Not performed**
Downstream evaluation	**Not performed.** No model trained on generated data
Comparison against existing generative baselines	**Not performed**
Access to industrial sensor data for validation	**None held at the time of writing**
Constraint taxonomy, Section 3	Design proposal
Enforcement mechanism comparison, Section 4	Analysis of published mechanisms and their known properties
Fidelity validation design, Section 5	Specification. Not executed
Downstream evaluation design, Section 6	Specification. Not executed
Scope boundaries, Section 7	Reasoned position
Practitioner motivation, Section 9	Experiential basis for design choices; not evidence of method performance

This is a method proposal in the strict sense. The paper's value, if it has any, lies in specifying the problem precisely enough that the proposed evaluation could be run by someone with access to the data, and in stating the scope boundaries narrowly enough that a positive result would mean something specific.

12. What a Reviewer Should Press On

The proposal has three points where a knowledgeable reviewer will and should apply pressure, and it is better to name them than to leave them to be found.

The redundancy argument. Section 2 states a published position that conservation enforcement is often mathematically redundant, and answers it with reasoning rather than evidence. A reviewer is entitled to treat that answer as insufficient, because it is. The proposal's response is structural: the ablation in Section 7 is designed so that the redundancy claim, if correct, appears as a result rather than being argued away.

The circularity in the validation. The physics-based checks in Section 5 confirm consistency with the assumed model. A reviewer will observe that this cannot detect a wrong model, and the paper agrees in Section 8. The partial answer is that the downstream evaluation on real held-out records is not circular, since it tests against reality rather

than against the model, and it is the only component of the validation with that property. That places more weight on the downstream test than a fidelity-focused framing would suggest.

The absence of a chosen architecture. Section 9 records that no base generator is specified. A reviewer may reasonably ask whether a proposal that does not commit to one has specified enough to be evaluated. The honest answer is that the constraint machinery is intended to be architecture-agnostic and that this generality has not been demonstrated across even two architectures, which is a claim the paper makes and cannot yet support.

Naming these is not a defence. It is an attempt to make the paper's weak points explicit enough that a reviewer's time goes to whether the proposal is worth executing rather than to establishing what it is missing.

13. LIMITATIONS

The constraint set is assumed knowable. Section 3 lists four classes as though the specific constraints for a given equipment class can be enumerated. In practice that enumeration is itself a research activity, frequently incomplete, and its incompleteness is invisible: a generator constrained on three of the four balances that actually govern the equipment will satisfy every check that was written.

No cost is assigned to constraint violation. The validation in Section 5 reports violation rates. It does not weight them, and violations are not equally consequential: a small mass imbalance may be immaterial while a monotonicity breach invalidates a sample entirely. A violation rate that treats these alike is a weak summary of physical validity.

Everything in this paper is untested. The constraint taxonomy may omit classes that matter; the enforcement composition in Section 4 may behave worse than any single mechanism; the validation design may miss failure modes it was constructed to catch. None of this has been exercised.

The central premise is contested in the literature. As Section 2 records, recent work argues that explicit conservation enforcement is often mathematically redundant. This paper offers a reasoned response and no evidence. If the redundancy argument holds broadly, the conservation component of this proposal is unnecessary, and the ablation in Section 6 is designed to detect that rather than to avoid it.

No dataset access. The evaluation in Section 6 names public datasets. Whether they contain the physical structure the constraints depend upon, particularly the balance quantities required for conservation enforcement, has not been verified against the actual data.

The validation is partly circular by construction. As Section 7 states, physics-constrained fidelity checks confirm consistency with the assumed model rather than with reality. Detecting model misspecification requires independent real data covering the misspecified regime, which is the scarcity problem the method exists to address.

The practitioner grounding is motivational only. Section 9 explains where the design concerns came from. It provides no evidence about whether the design addresses them.

The proposal names no generator architecture. Sections 4 and 5 discuss constraint classes and enforcement mechanisms without committing to a base generative model. That is deliberate, since the constraint machinery is intended to compose with several, and it is also a gap: the enforcement options available depend materially on the base architecture, and a proposal that does not choose one has not fully specified what would be built.

No implementation cost estimate. The computational cost of the proposed enforcement composition, and whether it is tractable at realistic sensor dimensionality and sequence length, has not been assessed.

14. FUTURE WORK

The work is entirely ahead, and the order matters.

Verify dataset suitability first. Before implementing anything, confirm that the candidate public datasets carry the quantities the constraints require. A conservation constraint needs measured inflow, outflow, and accumulation,

and a dataset lacking any of them cannot support the method regardless of how well it is implemented. This is the cheapest step and it can invalidate the plan.

Implement the simplest enforcement first. A loss-based penalty on one conservation constraint, with violation rate measured on uncorrected output, would establish whether constraint enforcement changes generated distributions at all. It also directly probes the redundancy counterargument at low cost.

Then run the ablation before the full comparison. Establishing which constraint classes affect the output, before building the composed system, avoids investing in components that contribute nothing.

Run the downstream evaluation as specified. With the acceptance criteria in Section 6 fixed in advance, including the negative-result condition on rare-event recall.

Build the missingness model. As Section 9 argues, tying missingness to operational events rather than assuming it random is a design commitment that has not been implemented and that distinguishes an industrially realistic generator from a benchmark-realistic one.

A note on what would make this worth publishing. The specification in this paper is only useful if someone executes it, and the outcome most worth reporting is not necessarily a positive one. A rigorous negative result, showing that physics constraints are enforced correctly and do not improve downstream rare-event recall, would be more valuable to the field than another paper reporting aggregate gains from a composed system whose components were never separated. The counterargument in Section 2 makes that outcome genuinely plausible rather than a formality, which is why the ablation is specified before the composed evaluation rather than after it.

The order in this section reflects that. Dataset suitability is verified first because it can invalidate the plan cheaply. The simplest enforcement is implemented next because it probes the redundancy question directly. Only then does building the composed system make sense.

15. Conclusion

Synthetic data for industrial systems has a failure mode that distributional evaluation cannot detect: data that is statistically indistinguishable from real records and physically impossible. A model trained on it learns correlations rather than dynamics, and fails at the extreme operating points where the exercise was supposed to help.

This paper specifies a response. Four constraint classes must be respected, and only one of them reliably admits a closed-form parameterisation, which is why no single enforcement mechanism suffices and the proposal composes several. Fidelity validation must reach cross-sensor coupling, spectral content, and directly measured violation rates reported as distributions rather than means. And the downstream evaluation must be specified with its acceptance criteria fixed in advance, including the condition under which the result would count against the method.

The scope is narrower than this area usually claims. The method addresses failure modes within the modelled physics, and it is silent on unobserved mechanisms, misspecified equations, poorly characterised processes, and adversarial signatures. A constraint-satisfying generator built on wrong equations produces confidently wrong data that every physics-based check will pass, and that is a limit no amount of engineering removes.

Nothing here has been built. There is no implementation, no generated data, no validation, and no downstream result, and the paper reports none. What it offers is a specification precise enough to be executed and scoped narrowly enough that executing it would settle something.

REFERENCES

1. United States National Aeronautics and Space Administration, Turbofan Engine Degradation Simulation Data Set (C-MAPSS), NASA Prognostics Center of Excellence Data Repository, Moffett Field, CA, USA.
2. University of California Irvine, SECOM Semiconductor Manufacturing Data Set, UCI Machine Learning Repository, Irvine, CA, USA.

3. European Parliament and Council, Regulation Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act), Article 10, Data and Data Governance.
4. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio, "Generative Adversarial Networks," in Advances in Neural Information Processing Systems 27, 2014. arXiv:1406.2661.
5. D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," in Proc. Int. Conf. on Learning Representations (ICLR), 2014. arXiv:1312.6114.
6. M. Arjovsky, S. Chintala and L. Bottou, "Wasserstein GAN," 2017. arXiv:1701.07875.
7. J. Ho, A. Jain and P. Abbeel, "Denoising Diffusion Probabilistic Models," in Advances in Neural Information Processing Systems 33, 2020. arXiv:2006.11239.
8. M. Raissi, P. Perdikaris and G. E. Karniadakis, "Physics-Informed Neural Networks: A Deep Learning Framework for Solving Forward and Inverse Problems Involving Nonlinear Partial Differential Equations," Journal of Computational Physics, vol. 378, pp. 686-707, 2019. doi: 10.1016/j.jcp.2018.10.045.
9. G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang and L. Yang, "Physics-Informed Machine Learning," Nature Reviews Physics, vol. 3, pp. 422-440, 2021. doi: 10.1038/s42254-021-00314-5.
10. S. Greydanus, M. Dzamba and J. Yosinski, "Hamiltonian Neural Networks," in Advances in Neural Information Processing Systems 32, 2019. arXiv:1906.01563.
11. M. Cranmer, S. Greydanus, S. Hoyer, P. Battaglia, D. Spergel and S. Ho, "Lagrangian Neural Networks," 2020. arXiv:2003.04630.
12. C. Esteban, S. L. Hyland and G. Ratsch, "Real-valued (Medical) Time Series Generation with Recurrent Conditional GANs," 2017. arXiv:1706.02633.
13. I. Shumailov, Z. Shumaylov, Y. Zhao, N. Papernot, R. Anderson and Y. Gal, "AI Models Collapse When Trained on Recursively Generated Data," Nature, vol. 631, pp. 755-759, 2024. doi: 10.1038/s41586-024-07566-y.
14. A. Saxena, K. Goebel, D. Simon and N. Eklund, "Damage Propagation Modeling for Aircraft Engine Run-to-Failure Simulation," in Proc. Int. Conf. on Prognostics and Health Management (PHM), 2008, pp. 1-9. doi: 10.1109/phm.2008.4711414.