

OPERATING CASE STUDY / DATA, ANALYTICS & AI

Predictive Operations & Decision Support

Build the forecast-to-action contract before scaling the operating response

An anonymized project case study of an operational forecasting engine: designing prediction, data-source and alerting controls while separating runnable software from validated live decisions and customer outcomes.

PROJECT CONTEXT / OPERATIONS, ANALYTICS & IMPLEMENTATION

BLUF / BOTTOM LINE UP FRONT

Authorize live shadow validation.
Test the response before scaling.

RECALCULATED FINDING

The listed forecast cells account for 86.98% of the headline output. Every recorded requested-live run used a fallback source. Neither finding invalidates the prototype; both require a clear evidence gate before operational authorization.

CONTRIBUTION IN THE SOURCE RECORD

Architecture, product standards and evidence gates.
Provisional registry-level hoop support; agent-assisted implementation.

EVIDENCE STATUS

Registry-based design; originals unrecovered. Live predictive validity, sustained adoption and customer benefit are unestablished.

BLUF: analytical inference. Source contribution: historical record. Client and internal identities anonymized.

EXECUTIVE ABSTRACT

Authorize the evidence needed for a forecast-led operating decision

[AI] The registered design combined forecasting with explicit controls on source quality and unsupported features. The next investment decision is whether live forecasts improve a fixed-capacity operating response.

CONTEXT | Short-horizon risk needed an operating interface

[HR] The registered project connected station-level forecasts, source adapters, thresholds, driver explanations and alert delivery. Seed data enabled development while access to usable live outcome histories remained unresolved.

OBJECTIVE | Move from a prediction to a defensible response

[PD] Qualify the forecast on a frozen live population and observation clock; then test whether its priority queue outperforms the usual queue using the same intervention capacity and action package.

APPROACH | Build controls around the product

[HR] The author directed architecture, metric identity, acceptance standards, goal thresholds and feature gates. [AI] These contributions receive provisional registry-level hoop support. Agent-assisted implementation and separate model, application and action-library projects retain their own attribution.

FINDINGS | Output, lineage and decision validity do not coincide

[RC] Listed forecast cells explain 86.98% of the headline output, leaving 13.02% unreconciled. [HR] All recorded requested-live runs used fallback data despite successful job completion. Outcome-label provenance did not establish a complete live validation panel.

DECISION VALUE | Fund the missing test before the scale decision

[PD] The COO should assign data and operations owners to silent live validation: zero seed or anchor outcomes counted as live truth; all mandatory lineage fields valid; and a positive lower 95% confidence bound on prespecified incremental skill, with calibration and tail tolerances met. Then test the queue at fixed capacity.

IMPLICATIONS | Credit the design without borrowing the outcome

[AI] The registered requirements demonstrate product judgment and explicit evidence boundaries. A successful run, seed benchmark, asserted deployment or downstream citation does not establish sustained use, an effective intervention or realized savings.

S1: protected project register and calculation ledger. HR historical record; RC recalculated; AI analytical inference; PD proposed design. Live validation and outcome impact remain unestablished.

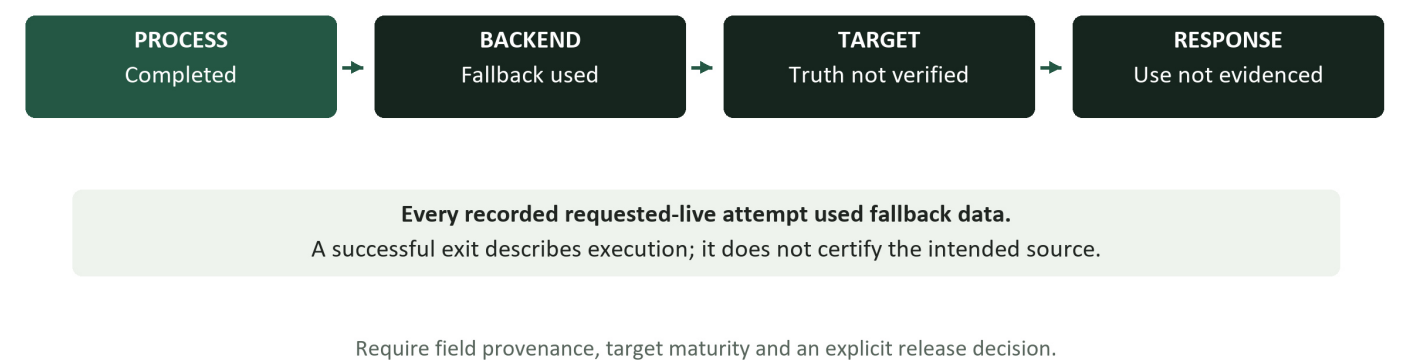
EXHIBIT 1 / RUNTIME AND OPERATING AUTHORIZATION

Treat successful completion and live forecasting as separate contracts

[HR] Every recorded requested-live run used fallback data while completing successfully. [AI] That is a runtime state, not permission to describe its forecasts as live-validated.

SUCCESSFUL EXECUTION DOES NOT AUTHORIZE LIVE DECISIONS

Historical register and proposed control • validate each state separately



[HR] The design exposed a degraded-source status instead of concealing fallback behavior. The register also describes a source adapter requiring an injected payload that the scheduled execution context did not provide. [AI] This is a recorded integration boundary; unrecovered code cannot establish permanent technical impossibility.

Contract	Evidence required
[PD] Job and source	The job finishes; requested and realized sources, field freshness, payload identity and fallback reason remain visible.
[PD] Operating release	Required live fields and target labels meet the frozen contract. A business owner accepts the permitted use and the fallback response.

[AI] A fallback can preserve testing and service continuity. It cannot silently inherit the authorization of a live model. Run volume also mixes requested-source modes, so the degraded share must use the requested-live population rather than every logged run.

PLATFORM OWNER ACTION

[PD] Separate completion, source integrity and operating release in the run manifest. Route a source-gate failure to an explicit hold or approved fallback policy.

[HR/AI/PD] S1:M12/D03/D08/E07/E12. The requested-live subset defines the fallback claim. Runtime and source-adapter statements are register-mediated; no recovered code was executed for this case.

EXHIBIT 2 / OUTPUT RECONCILIATION

Reconcile the forecast population before interpreting its scale

[RC] With headline output indexed to 100, the four listed forecast cells sum to 86.98. The remaining 13.02 is an accounting difference that needs an explicit mapping.

RECONCILE OUTPUT GRAIN BEFORE CLAIMING COVERAGE

Recalculated from reported counts • headline total indexed to 100



The remaining share has no reconciled row-grain explanation.
Unlisted rollups are one possibility; invalid records are not established.

Match version, entity type, horizon, metric and unique key before authorizing a coverage claim.

[HR] The source reports a headline prediction total and separate business-line/geography counts. [AI] It does not establish that those listed cells exhaust the headline. The difference may represent omitted categories or rollups; arithmetic alone cannot classify it as invalid, missing or duplicated predictions.

Reconciliation question	Required record
[PD] What is one output?	A unique entity, metric, horizon, issue time, target period and model/source version; identify aggregates separately.
[PD] Which rows belong?	A mutually exclusive population map connecting every headline row to a declared cell or an explained additional category.

[AI] Prediction rows are not interchangeable with physical sites, independent observations or eligible customer exposures. More horizons or rollups can increase output count without expanding the operating footprint or the amount of validated information.

ANALYTICS OWNER ACTION

[PD] Require a row-level reconciliation before publishing scale or coverage claims. Keep the unreconciled share visible until its identity is established; do not silently rescale the listed cells.

[HR/RC/AI/PD] S1:M09/E04/E08 and private arithmetic ledger. Headline prediction count=100. The 13.02 difference is unallocated in the available breakdown, not adjudicated erroneous output.

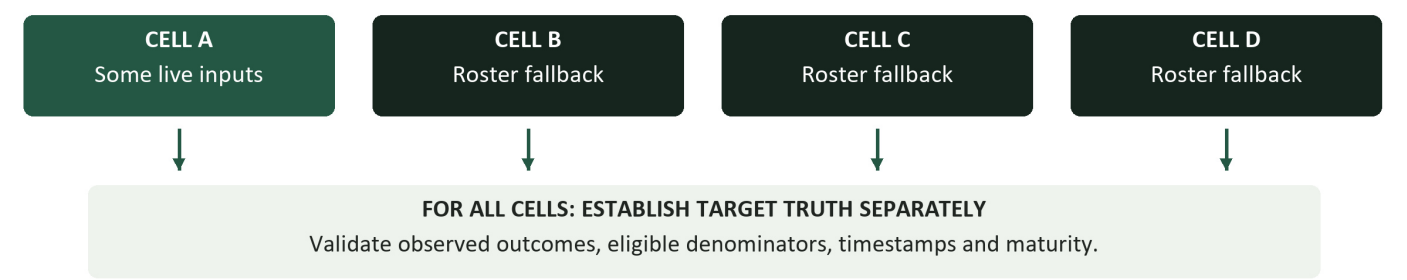
EXHIBIT 3 / FIELD-LEVEL TRUTH

Qualify the target labels even when some input fields are live

[HR] The provenance view identifies realized live columns in only one listed cell. Target histories also include anchored levels, even where genuine predictor fields exist. [AI] A run-level source label cannot certify every field.

A LIVE INPUT IS NOT A LIVE CUSTOMER-OUTCOME LABEL

Historical register • anonymous geography / business cells; no raw roster counts



Neither a source label nor geographic roster completeness proves forecast validity.

[HR] The project distinguished accessible scorecard inputs, roster fallbacks, trend or mix information and anchored outcome levels. [AI] Feed availability, successful capture and actual model consumption are different facts; a real trend feed does not prove that the forecast target is a mature live label.

Field role	Release check
[PD] Predictor	Was this exact value available at issue time, on the declared entity and refresh clock, without later revisions?
[PD] Target and exposure	Is the outcome linked to the eligible cohort with a stable denominator, adjudicated definition and completed follow-up window?

[AI] A hybrid panel can be useful for integration and scenario testing. It cannot establish live calibration merely because one feature family is live. Validation must retain source provenance by field and row, and report performance separately where source regimes differ.

DATA OWNER ACTION

[PD] Deliver an as-of feature snapshot and a later, linked outcome panel. Keep anchored or substituted targets outside the live-validity estimate; disclose excluded populations and missingness.

[HR/AI/PD] S1:M06/M09/D03/E08/E11. Source availability and realized provenance refer to their recorded snapshots. No complete live target panel is authenticated by the available register.

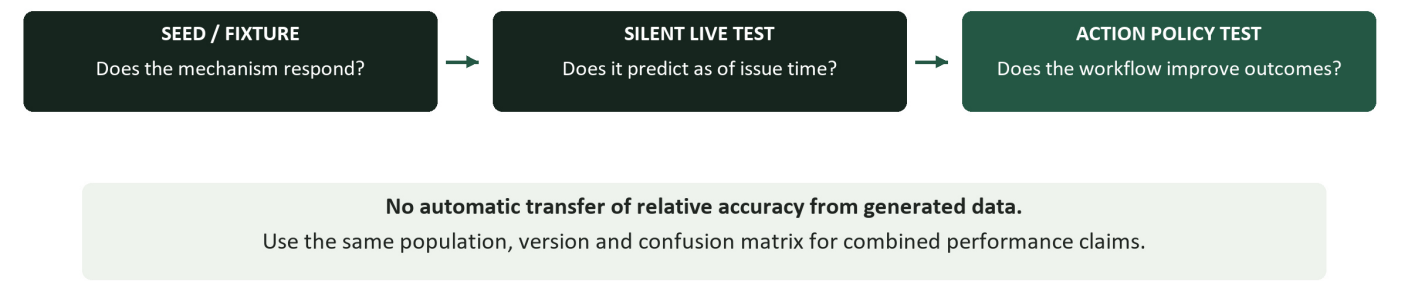
EXHIBIT 4 / VALIDATION AND TRANSPORT

Require fresh live validation before transferring a seed-data advantage

[AI] A benchmark can show that a pipeline learns an injected signal or beats a comparator on its seed panel. Neither establishes that its relative advantage will survive live operations.

EACH VALIDATION STAGE EARNS A DIFFERENT CLAIM

Proposed evidence sequence • later claims cannot be borrowed from earlier tests



Keep the production forecast path and evaluation path identical, then freeze the decision rule.

[HR] The registered design kept unsupported leading features off and used a positive control to test sensitivity to an injected lead. [AI] That is useful diagnostic discipline. Failure on a seed panel does not prove that no live leading relationship exists; success on an injected signal does not prove field predictability.

Comparison safeguard	Decision consequence
[PD] Frozen score identity	Match model, population, horizon, threshold, maturity and averaging rule. Precision, recall and F1 from different versions cannot form one confusion-matrix claim.
[PD] Frozen live evaluation	Score untouched future cohorts using only as-of information. Compare against the same operationally available baseline and report calibration, discrimination and uncertainty.

[AI] Model selection gains belong to the separate model-development project. This case concerns the engine that consumes a selected model and enforces its operating contract. A minimum history-length code gate is not a statistical sufficiency or maturity guarantee.

MODEL OWNER ACTION

[PD] Predeclare live acceptance criteria by horizon and operating regime. Require the lineage gate and predictive gate together before exposing an operational priority queue.

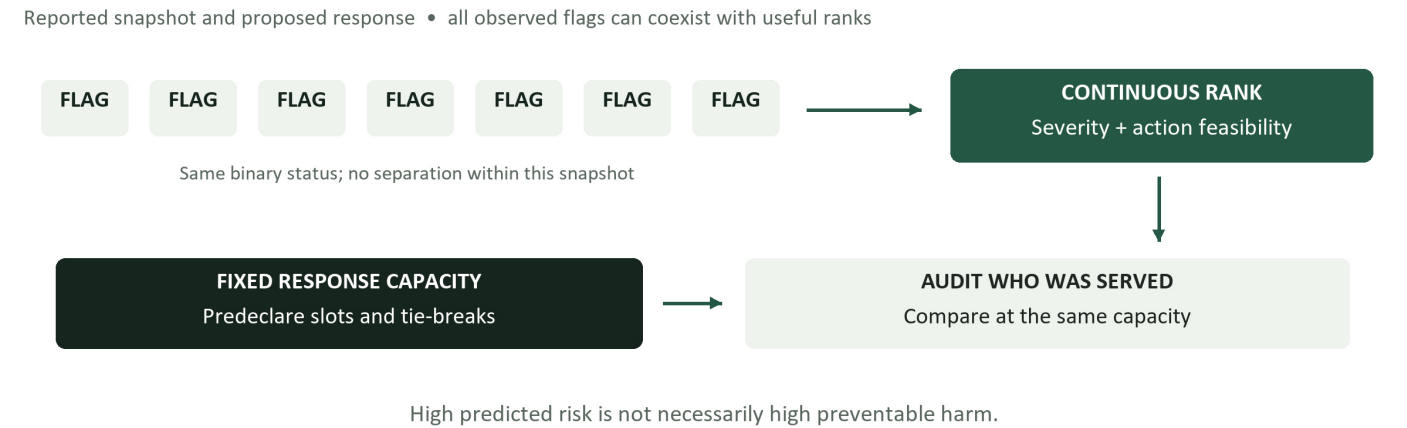
[HR/AI/PD] S1:M01/M04/M05/M07/D02/D06/E03/E05/E06. Benchmark claims remain source- and version-specific. No automatic relative-skill transport or original model-development credit is asserted.

EXHIBIT 5 / GOALS AND SCARCE CAPACITY

Preserve the goal gap while defining who receives the next intervention slot

[HR] Under the recorded fixed goal, every scored station cell in one snapshot breached the threshold. [AI] The binary flag therefore supplied no within-snapshot separation; it still conveyed a systemic goal gap.

SEPARATE ABOVE-GOAL STATUS FROM CAPACITY ALLOCATION



[HR] The author replaced a moving relative threshold with an explicit business goal. [AI] That makes the goal reference clearer. It does not by itself choose among simultaneously flagged sites or establish which response will prevent the most mature customer harm.

Operating question	Separate decision rule
[PD] Are we meeting the goal?	Retain the absolute goal and its eligible population. Do not improve its appearance by moving the reference to the current median.
[PD] Where should capacity go?	Use validated continuous risk, expected preventability and execution feasibility under a fixed slot budget. Preserve a tie-break and an overflow policy.

[AI] Universal binary alerts do not make continuous scores or severity tiers useless. Conversely, a high-risk entity need not be the entity with the greatest response to the available action. Retiering, changed populations and changed alert units prevent a simple before/after alert-count performance claim.

OPERATIONS SPONSOR ACTION

[PD] Authorize the capacity and action package before evaluating queue quality. Compare forecast and usual priorities at equal capacity; judge mature outcomes across every eligible unit.

[HR/AI/PD] S1:M10/D04/E09/E10. Universal flagging applies only to the specified recorded snapshot. Goal-setting contribution: provisional registry-level hoop support, not measured alert effectiveness.

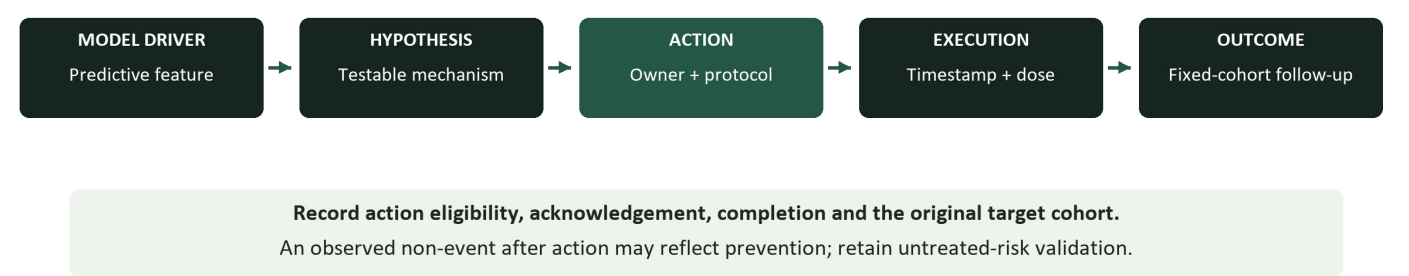
EXHIBIT 6 / PREDICTION AND INTERVENTION

Route a forecast driver as a hypothesis until its remedy is verified

[HR] The product design connected driver explanations to owners and candidate actions. [AI] A model-associated driver does not establish a causal diagnosis, an appropriate accusation or an effective remedy.

A FORECAST DRIVER DOES NOT AUTHORIZE A CAUSAL REMEDY

Proposed operating sequence • preserve the distinction between explanation and intervention



Classification tags and model-linkage fields do not establish treatment effects.

[HR] The author required domain-aware routing and distinguished categories of model linkage. A later simplification directive called for risk, a short driver explanation and a small action set. [AI] These are product requirements; the record does not demonstrate their completed adoption or that recipients executed the recommended work.

Evidence boundary	Owner responsibility
[PD] Risk explanation	Identify the observed association and competing explanations. Check population, time ordering, source quality and action relevance.
[PD] Authorized action	Confirm eligibility, capacity, safeguards and the actual package. Record acceptance, execution, exceptions and linked mature outcomes.

[AI] Reusing an existing production model may avoid a redundant build, but access and integration remain distinct from ownership. Related application, model-selection and action-library work stays separate. A cited insight or claimed deployment is not evidence of sustained field use.

ACTION OWNER GATE

[PD] Keep explanation, authorization and execution as separate records. Test the queue policy with a common action package before crediting any driver-specific operational effect.

[HR/AI/PD] S1:D05/D07/D09/M14/E14/E15/E18. Architecture and routing requirements: provisional registry-level hoop support. Child projects and other teams retain their own implementation and outcome claims.

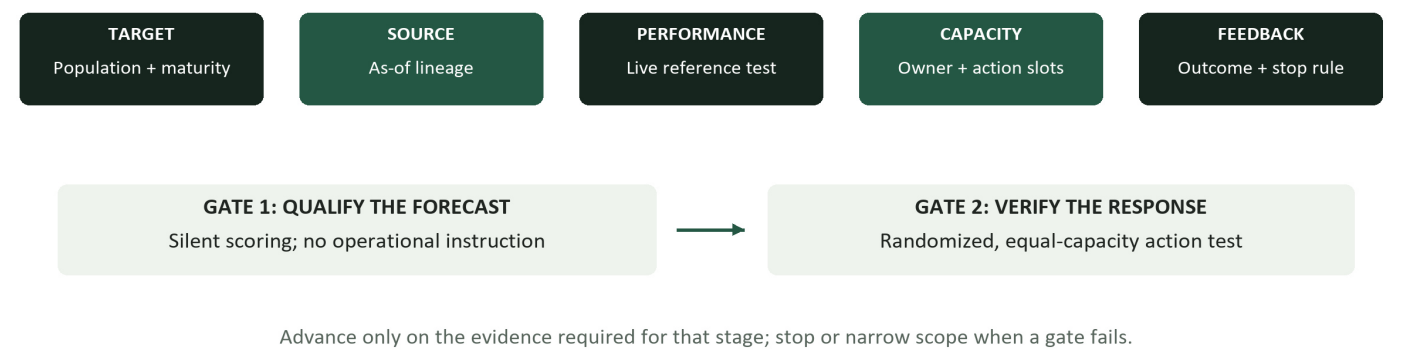
EXHIBIT 7 / FORECAST-TO-ACTION CONTRACT

Release the product through evidence gates owned by data and operations

[PD] The proposed Forecast-to-Action Contract joins target truth, source timing, predictive performance, operating capacity and outcome feedback. Each field constrains the permitted use of a forecast.

THE FORECAST-TO-ACTION CONTRACT LINKS RELEASE TO PROOF

Proposed FAC framework • each contract field has an accountable owner



Decision	Accountable acceptance
[PD] Silent live validation	Data and model owners freeze eligibility, as-of snapshots, target maturity and the comparator. Predictions remain silent until prespecified quality and performance criteria pass.
[PD] Limited action test	The operations sponsor fixes capacity, action package and escalation safeguards. Randomized independent operating units compare forecast and usual queues.
[PD] Scale or revise	The COO reviews a mature outcome estimate, uncertainty, operational guardrails and the same-population economic threshold before expanding the permitted use.

[AI] The contract preserves the project’s positive contribution: a coherent product architecture with explicit feature, source and goal decisions. Its next step is a testable operating proposition, not a retrospective claim that the prediction already changed customer outcomes.

COO DECISION

[PD] Sponsor the live evidence gate and name the operating owner. Hold scale authorization until the action test can distinguish useful prioritization from extra capacity, different treatment or a changed denominator.

[PD/AI] Proposed FAC; S1:D01–D09 and implementation register. Prospective model assumptions, power, economics and information value are separate hypothetical calculations, not historical project results.

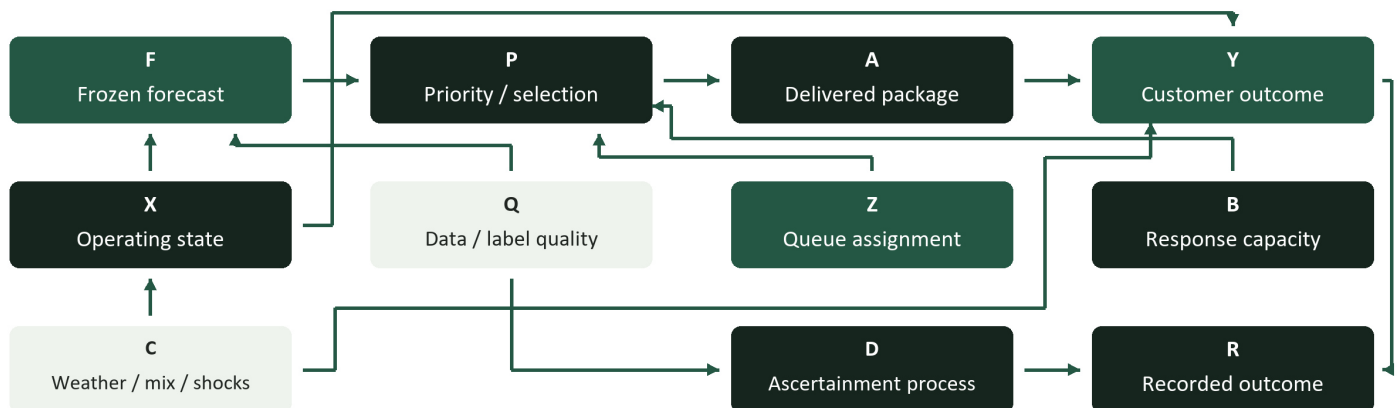
TECHNICAL FOUNDATION · THEORY OF CHANGE

A forecast creates customer value only through a delivered intervention

[PD] The Forecast-to-Action Contract separates source truth, predictive information, capacity-constrained choice, delivered action and the customer outcome. Each transition requires different evidence.

FORECAST VALUE REQUIRES AN ACTION PATH

Proposed FAC graph · randomized queue policy · no historical causal effect



Estimate assignment $Z \rightarrow P \rightarrow A \rightarrow Y$; the outcome is 42-day risk in the proposed trial, not ultimate maturity.

$Q \rightarrow F$ and $Q \rightarrow D \rightarrow R$ separate source truth and recording; $X \rightarrow F$ and $X \rightarrow Y$ encode shared predictive information.

$C \rightarrow X$ and $C \rightarrow Y$ are common operating causes. B constrains selection; shared response teams require larger clusters.

Later A may change future X and forecast calibration. Do not evaluate untreated-risk prediction on altered labels blindly.

[PD] Z assigns the queue policy; F is the frozen forecast; P selects an eligible response; A is the delivered package; Y is the customer outcome. Q is data quality, X the operating state, C common weather/case mix/capacity, B the response budget, and R the recorded outcome through ascertainment process D, including delay and classification.

[AI] The registered run report describes execution around F; it does not validate $Q \rightarrow F$. A useful forecast requires $X \rightarrow F$ information about future Y. Customer benefit requires $P \rightarrow A \rightarrow Y$; the source does not establish this causal chain. Driver contributions to a prediction are not estimates of $A \rightarrow Y$.

[PD] Randomization identifies $Z \rightarrow Y$ under the stated interference and ascertainment contract. Do not condition the primary comparison on receiving an alert, accepting a recommendation or completing an action. Those are post-assignment mechanisms.

[AI] Unexplained residuals can reflect missing operating factors, altered labels or delayed reporting. Prediction error alone cannot identify which path is missing. After deployment, outcomes affected by A cannot be used uncritically to judge an untreated-risk forecast.

Proposed causal graph, not an authenticated historical implementation. Perdomo et al. motivates separating prediction targets from outcomes changed by policy.

TECHNICAL FOUNDATION • ESTIMANDS & CLOCKS

Specify the target before interpreting accuracy or economic value

[AI] Forecast accuracy, response allocation and customer prevention are different quantities. The case needs a separate population, clock and comparison for each.

Quantity	Exact proposed estimand
Live predictive performance	[PD] Mean declared score of frozen as-of forecasts versus a frozen usual baseline, over all eligible station/cohort/horizon records with verified labels. Report source coverage and missing-label bounds alongside scores.
Capacity-constrained policy effect	[PD] Usual-minus-forecast risk of at least one prespecified service concession within 42 days of original scheduled dispatch, across all originally eligible customer units. Assignment is to a queue, not guaranteed action.
Delivered-package effect	[AI] Not identified by the engine record or by comparing completed actions with untreated cases. Separate randomized package variation or defensible identification is required.
Indexed economic effect	[PD] Absolute policy risk difference divided by one fixed pretrial reference risk. This is not a moving percentage of whichever arm performs worse.

[PD] The trial enrolls cohorts for 20 days and closes each cohort’s outcome window 42 days after its original scheduled dispatch. These are synthetic planning choices. A 42-day endpoint remains a 42-day endpoint even if further claims arrive later.

[PD] Freeze eligibility upstream of pre-dispatch checks. Include held, cancelled and unfulfilled units; otherwise the intervention can change its own denominator. Preserve assignment through rerouting and missing data, with declared best/worst outcome bounds.

[AI] Station-level daily rates, aggregate monthly fits and seeded station-day forecasts cannot be exchanged as if they estimate the same target. No historical field effect is reconstructed from their juxtaposition.

Source target and maturity restrictions; proposed trial estimands. The child model-comparison work remains a separate project.

Successful fallback is an engineering state, not live outcome validation

[HR] The registry describes a runtime that completed requests while falling back to seed data. The source-state labels preserved an important distinction; they did not supply the missing live outcome series.

Field / metric	Required validity contract
Outcome risk	[PD] Unique eligible unit, target event taxonomy, original schedule, exposure and ascertainment horizon; validate duplicates, reroutes and denominator stability across networks.
Source state	[PD] Record requested mode, delivered mode, source timestamp, lineage and outcome provenance separately. A live input column cannot certify a live outcome label.
Coverage	[PD] Reconcile headline records to the full cell partition. Geographic presence, populated rows and label authenticity measure different things.
Action status	[PD] Separate suggestion, owner acceptance, eligible target, delivery receipt and verified completion. A library entry does not establish execution.

[RC] The listed source-cell partition leaves about 13% of the reported prediction total unallocated. That unresolved remainder prevents a complete coverage claim; it does not identify which records are missing or false.

[AI] Repeated fallback can preserve availability while preventing accumulation of the live history needed for calibration. Adding successful runs of the same seed is not adding independently matured evidence. An injected payload and autonomous source access have different operating dependencies.

[PD] For taxonomy reliability, double-code a stratified pilot of 200 eligible labels across outcome, network and source-state classes before adjudication. Report confusion matrices, agreement and kappa with cluster-aware intervals. This proposed precision budget is not a proven adequate sample; agreement alone does not establish truth.

S1/private runtime, lineage and coverage reconciliation. Original code, logs and live outcome panel were not recovered in this pass.

TECHNICAL FOUNDATION • PREDICTIVE VALIDATION

A synthetic holdout tests its generating process, not the field

[AI] A seeded test can expose software defects and demonstrate response to a planted signal. It cannot establish that live inputs contain the same signal or that relative model rankings transport.

Validation layer	Permitted conclusion
Deterministic parity	[PD] The same frozen input/version produces the same output. This is reproducibility, not outcome validity.
Synthetic positive control	[AI] Detecting a deliberately planted lead demonstrates sensitivity to that construction; it neither excludes every implementation defect nor verifies a real lead.
Silent live temporal holdout	[PD] Freeze training, tuning, imputation, source cutoffs and baseline before future labels mature. Evaluate the intended target and population without operational use of candidate alerts.
Capacity-matched field trial	[PD] Test whether the queue changes customer risk through the same intervention budget; prediction scores alone cannot replace this test.

[PD] For release from silent validation, require mandatory population/as-of/source fields and no seed outcomes represented as live truth. The lower 95% interval for baseline loss minus candidate loss must exceed zero under a prespecified proper score; calibration, tail, missingness and safety tolerances need explicit owner approval.

[PD] Use rolling calendar cutoffs, grouped uncertainty by operating unit and checks for shared-horizon overlap. Select tuning using earlier folds; a repeatedly inspected test set becomes development data. Report probability/interval scores and calibration by horizon and group.

[AI] A reported error-ratio increase may mean worse performance; define its direction. Precision, recall and F1 from different versions or populations cannot form a common confusion matrix. Aggregate correlation with another target is not a forecasting validation.

Gneiting–Raftery proper scoring; Kapoor–Narayanan leakage taxonomy. These motivate prospective controls, not a diagnosis of unavailable historical code.

TECHNICAL FOUNDATION · CAPACITY & PREVENTABILITY

The best risk forecast can still prioritize the wrong response

[AI] A universal goal-breach flag conveys system stress but cannot distinguish which flagged case receives a scarce response. Continuous severity or ranking may still supply information.

[PD] For candidate i , let untreated risk be p_i , causal package benefit τ_i , value per prevented outcome v_i , cost c_i and resource need b_i . Choose binary x_i to maximize $\sum_i x_i(\tau_i v_i - c_i)$, subject to $\sum_i b_i x_i \leq B$. Ranking p_i is equivalent only with additional assumptions about preventability and value.

Potential mismatch	Required operating control
All cases exceed an absolute goal	[PD] Publish the common goal gap separately from the priority queue; define capacity, tie-breaks and unserved eligible cases.
Driver share versus action effect	[PD] Treat decomposition as a diagnostic hypothesis. Verify the proposed lever’s mechanism, target population, timing and preventability.
High risk outside response authority	[PD] Route to the correct authority or keep an explicit unresolved record. A recommendation cannot confer permission or capacity.
Model use changes the label	[PD] Retain assignment and action logs. Distinguish untreated-risk validation from post-action customer outcomes.

[PD] The synthetic model holds response capacity equal. When high-risk cases are responsive, better high-risk capture helps. When they are unresponsive but lower-risk cases can benefit, the same improved ranking makes outcomes worse by displacing useful work.

[PD] Usual high-risk capture matters: at .25/.45/.55 versus forecast .65, synthetic benefits are 11.429/5.714/2.857 fixed-reference points. Net values are 2.686/−.657/−2.329. A stronger usual queue already defeats the base economic case; measure its actual rule before planning.

Elmachtoub–Grigas constrained decision loss and Athey–Wager treatment-policy learning. FAC does not claim to implement either published algorithm.

TECHNICAL FOUNDATION • VARIATION & SELECTION

Repeated seed runs do not create a stable operating outcome series

[RC] The private audit retains six ordered screens and twenty mechanical benchmark/profile screens. None supplies an eligible stable operating phase or a forecast-caused customer effect.

[PD] For a suitable sequence, $MR_t = |x_t - x_{t-1}|$, tentative individual limits = $\text{mean}(x) \pm 2.66 \text{ mean}(MR)$, and $MR \text{ upper} = 3.268 \text{ mean}(MR)$. Check population, interval, metric, phase and dependence before interpreting an excursion.

Available sequence	Interpretation boundary
Dated forecast snapshots	[RC] Changing model/source versions and forecast values, not realized outcomes. The mechanical limits cannot establish operational stability.
Sparse alert-count excerpts	[RC] Unequal gaps and missing intervening rows prevent a regular full-history analysis. No invented intermediate points.
Constant output / status sequences	[RC] Zero moving range is degenerate; constant rows or successful exit codes do not establish live fulfillment or future reliability.
Model and horizon profiles	[RC] Comparisons across designs or horizons are mechanical shape checks, not temporal natural-process charts.

[PD] No repeated verified event-count/exposure panel supports Laney p' . Do not turn rounded forecast rates into observed denominators. Record late-label revisions separately from temporal operating movement.

[AI] No fixed treated worst-N cohort and untreated follow-up panel establish historical recovery. In a proposed selected cohort, equal-variance untreated reliability p implies initial excess indexed 100 becomes expected $100p$. At $p = .25/.50/.75$, regression alone yields 75/50/25 points of reversion; this is a sensitivity illustration, not an estimated correction.

[PD] Test network, node type, weather regime, label delay and package responsiveness as prespecified moderators. More forecast rows do not substitute for independent operating units.

S1/private raw screen inventory; NIST XmR formulas. No historical signal/stability or regression-to-the-mean effect is claimed.

TECHNICAL FOUNDATION · CONTRIBUTION & ADOPTION

Architecture ownership and field execution require different proof

[HR] The registry credits the author with product scope, architecture, evidence standards, integration direction and gating decisions. It does not make the author the producer of every code module, upstream dataset or downstream action.

Contribution claim	Process-tracing grade
Engine scope and architecture	[HR/AI] Provisional hoop: recorded direction supports ownership of the design problem and operating specification; original instructions and artifacts are not reopened.
Integration and evidence standards	[HR/AI] Provisional hoop: source-state labeling, target correction and gate decisions are attributed in the register. Shared metadata is not independent corroboration.
Production code and data access	[AI] Do not claim personal authorship or authority absent evidence. Directed agent work differs from writing the code; runtime integration does not grant upstream permissions.
Operational adoption / customer benefit	[AI] Unestablished. A runnable app or proposed library does not show users acknowledged, acted or achieved a mature outcome.

[AI] No smoking-gun or doubly decisive personal-attribution test is claimed from the registry. Recover contemporaneous requests, versioned decisions, review receipts and usage logs before upgrading contribution grades.

[HR] The source records a request to simplify the output and an unresolved priority/adoption question. That is product feedback; it is not evidence the simplified workflow was implemented or rejected by all potential users.

[PD] Record received alert, eligible action, owner acknowledgment, attempted response, verified execution and disposition. Keep this fidelity ladder separate from causal outcome analysis. Do not select only completed actions when estimating assigned-policy effects.

[AI] The supported achievement is a specified forecast-to-action architecture with exposed dependencies. The separate benchmark project and later field actions retain their own attribution boundaries.

S1 role, direction, deployment and feedback descriptions. Registry-mediated process tracing; no independently authenticated implementation or customer effect.

TECHNICAL FOUNDATION • PROSPECTIVE EVALUATION

Test incremental allocation value after live validation passes

[PD] Parallel randomization preserves concurrent controls through 42-day follow-up without crossover into legacy cohorts. It does not solve cross-pod spillover. Assign independent operating pods, not alerts competing for one team.

Design element	Preregistered contract
Eligibility / assignment	[PD] Freeze upstream customer units, pod boundaries and equal response capacity before assignment. Keep controls on usual priority through every enrolled cohort’s final ascertainment.
Package and comparator	[PD] Same approved cohort-specific pre-dispatch package and full slot utilization. Usual priority is measured before the trial; the toy unresolved-tie comparator must be replaced if real practice is stronger.
Primary analysis	[PD] Assigned-policy difference in 42-day customer risk; equal exposures imply equal-pod weights. Use randomized assignment, not per-protocol completion, for the primary contrast.
Missing / spillover	[PD] Preserve missing and rerouted cases; report risk bounds. Shared teams, persistent station remedies or cross-pod capacity transfers require a larger cluster or another design.

[PD] Audit baseline and event-time patterns; nonsignificant pretrends do not prove exchangeability. A Callaway–Sant’Anna group-time alternative requires credible conditional parallel trends, no anticipation, overlap and untreated comparisons. It is not powered by this parallel-trial calculation.

[PD] Preregister safety, hold/cancel/unfulfilled, fairness, workload and missingness margins. Independent blinded outcome ascertainment is common to arms; symmetric urgent safety disclosures are logged. Primary economic power does not size these guardrails.

[PD] Recruitment and variance inputs are hypothetical. If independent pods cannot be recruited, change the design and re-power or leave the funding question unresolved. Extra correlated rows cannot satisfy the sample requirement.

Proposed field experiment, preceded by silent temporal validation. Cluster Student planning and direct queue/outcome simulations are disclosed in the model pages.

The Forecast-to-Action Contract makes five claims testable

[PD] FAC records target/population, as-of source, predictive gate, capacity/owner and feedback/stop rule. It connects a forecast product to an accountable decision without assuming that prediction creates impact.

Proposition	Test and boundary
P1 • Truth before volume	[PD] Mandatory provenance checks prevent seed/anchor outcomes from passing live validation. A single concealed source-state substitution falsifies the specified zero-tolerance control.
P2 • Prospective predictive value	[PD] Frozen live forecasts improve prespecified loss over the usual baseline. An upper confidence bound below zero contradicts improvement; an inconclusive interval leaves it unresolved.
P3 • Benefit-aware capacity	[PD] At equal response capacity, the queue improves customer risk only where priority correlates with preventable benefit. Unresponsive high-risk selection can reverse the result.
P4 • Delivered ownership	[PD] Explicit authority and receipts improve faithful execution without harmful delays or displaced work. A populated owner field alone does not support this claim.
P5 • Economic sufficiency	[PD] Causal risk reduction covers incremental cost. A lower interval endpoint above break-even supports the value gate; an upper endpoint below contradicts sufficiency.

[AI] Gneiting–Raftery disciplines forecast scores; Elmachoub–Grigas shows why decision loss differs from prediction loss. Athey–Wager locates allocation value in identified treatment effects and constraints. This case extends those distinctions into an explicit operating release contract, not a new proved theorem.

[AI] Perdomo et al. motivates retaining the outcome policy when scoring forecasts. Kapoor–Narayanan motivates temporal separation and reproducible lineage. The available case corroborates the need for these controls; it does not estimate their effectiveness.

[PD] Each proposition needs its own evidence. Passing the customer-risk test cannot excuse invalid source labeling or a failed safety guardrail.

Primary methodological literature; proposed FAC propositions, not external certification. Statistical inconclusiveness is distinct from refutation.

TECHNICAL FOUNDATION · LIMITATIONS & REPRODUCTION

Keep engineering evidence, predictive evidence and impact evidence separate

[AI] This is an operating-design and evidence-reconciliation case. It supports a bounded architectural contribution and a prospective test, not historical forecasting superiority or realized financial impact.

Boundary	Consequence
Source originality	[AI] Original code, logs, app and evaluation panel were not recovered. Registry quotations and descriptions remain traceable leads rather than authenticated execution evidence.
Target / source mismatch	[AI] Live input availability does not certify live outcomes; aggregate correlations and anchor-filled forecasts cannot validate station-level future risk.
External validity	[AI] Synthetic level or relative skill does not establish field transport. No-lead construction cannot show that real levers lack information or effect.
Capacity / adoption	[AI] Broad alerts can express a system gap while leaving action ranking unresolved. No verified user-action-outcome chain establishes adoption or prevention.
Proposed model	[PD] Risk composition, capture, prevention, frailty, independence, costs and available pods are assumed. Persistent effects, stronger usual practice or severity shifts require redesign.

[PD] Reproduce the private source ledger and version arithmetic first. Retain model hashes, input provenance, as-of cutoffs, metric definitions and source conflicts. Then run conditional moments, effect-path power inversion, complete-process simulations and Bayes decision enumeration. Independent checks validate these declared calculations.

[PD] Keep raw source values, people, sites, systems and identity mappings private. Public numeric views are relative or indexed; avoid companion anchors that reveal confidential denominators. Proposed model probabilities and sample counts are synthetic and clearly separated.

[AI] The decisive new interpretation is that engineering availability can coexist with a blocked learning loop, while better risk ranking can coexist with worse customer allocation. Closing those gaps requires authenticated live labels and a capacity-matched action test.

Consolidated evidence and model limits. Author/agent-assisted review does not authenticate missing artifacts or certify actual field benefit.

PROPOSED DECISION MODEL · FINITE CAPACITY

Compare priority queues at the same response capacity

[PD] This synthetic trial compares a forecast queue with an unresolved-tie usual queue. Both receive the same safe package and exactly one response slot per pod per day. These are prospective assumptions.

Assumption	Usual queue	Forecast queue
Capture highest-risk catchment	0.25	0.65
Customer risk: high / lower	0.020 / 0.004	Same
Selected risk prevented	50%	50%

[PD] Each hypothetical pod contains four equal catchments; one is higher risk each day and its location rotates. Each enrolls 250 eligible units daily for 20 days. Freeze units before response; retain holds and cancellations. Measure at least one prespecified service concession within 42 days of original scheduled dispatch.

[PD] Each independent pod has risk multiplier 0.8 or 1.2 with equal probability. Conditional on pod type, daily selections and customer outcomes are independent. High-risk capture is a stipulated queue property, not an observed model accuracy.

[PD] Exact binomial-mixture moments give usual risk 0.0070. With that reference indexed 100, the assumed policy reduction is 11.4286 points. This is the effect of allocating identical capacity differently under the stated prevention model.

[PD] Better risk ranking is not better treatment ranking when preventability differs. The package is cohort-specific and pre-dispatch; persistent site-wide remedies or stronger usual prioritization require a new model. The measured horizon does not certify ultimate outcome maturity.

Wholly synthetic field-policy design; no historical engine accuracy, source footprint or realized customer benefit enters the calculation.

PROPOSED DECISION MODEL · POWER AT BREAK-EVEN

Size the trial against the economic boundary, not the large-effect variance

[PD] Randomize independent operating pods in parallel. Exact queue-selection and binomial moments retain shared pod frailty; the candidate policy is tested against equal-capacity usual practice.

[PD] Pod variance= $\text{Var}(E[\text{daily risk} | \text{frailty}]) + E(\text{Var}[\text{daily risk} | \text{frailty}]) / \text{days}$. Use $SE = \sqrt{[(V_{\text{usual}} + V_{\text{forecast}}) / n]}$ and Student $df = 2n - 2$. For every effect, vary high-risk capture from 0.25 upward and recompute its variance.

Per arm / total	MDE, 80%	P(lower>BE)	Joint gate
161 / 322	6.656	50.6%	Fail
323 / 646	4.733	80.0%	Pass
404 / 808	4.240	88.0%	Pass

[PD] Two-sided alpha 0.05; require $MDE \leq 6.8376$ and at least 80% chance the lower two-sided 95% interval exceeds that boundary at the assumed 11.4286-point benefit. The 646-pod minimum is conditional, not a recruitment commitment.

[PD] Across 10,000 complete-process simulations per scenario: null rejection 5.34%; break-even rejection 98.42%; full-effect economic clearance 80.00%. Monte Carlo intervals are retained privately.

[PD] The 808-pod buffered candidate still depends on pilot evidence. Shared teams, persistent actions and unequal exposure can invalidate it. Availability is unknown; extra correlated stations or daily rows cannot replace independent pods. Guardrails need separate sizing.

Approximate cluster inference, checked against the actual bounded generative model at null, break-even and full effect. No achieved field power is claimed.

PROPOSED DECISION MODEL · ECONOMIC SUFFICIENCY

Risk reduction creates value only when the targeted risk is preventable

[PD] $\text{Net} = 100r \times a \times q - K$. The same cohort's usual loss pool is 100, avoidable share $a = .65$, exclusive share $q = .90$ and incremental cost $K = 4$. Break-even = 6.8376% of reference risk.

Risk reduction r	$a = .40$	$a = .65$	$a = .90$
4%	-2.56	-1.66	-0.76
8%	-1.12	0.68	2.48
12%	0.32	3.02	5.72

High / low preventability	Policy effect, points	Net value
50% / 50%	11.429	2.686
20% / 50%	2.857	-2.329
0% / 50%	-2.857	-5.671

[PD] Both grids use the same fixed reference pool and $q = .90$, $K = 4$. Higher-risk capture stays 0.65 in the second grid. Selecting unpreventable high-risk work displaces preventable lower-risk work, reversing benefit. Forecast accuracy alone cannot rule out that failure.

[PD] K includes incremental data access, monitoring, training and production verification over the same enrollment/ascertainment pool. Equal package volume and cost cancel only under equal utilization. Research-only evaluation costs remain a separate study budget.

[PD] Constant expected loss per affected unit and unchanged severity mix are required. Otherwise use a cost-weighted outcome and re-power. Avoidability removes nonreleasable cost; exclusive credit removes overlap with other programs. Neither establishes cash realization.

[PD] With frailty spread increased from 20% to 30%, the buffered design fails its joint gate. The private 3×3 recruitment grid varies spread and enrollment days; assumptions require a pilot.

Hypothetical sensitivity, not confidence intervals. Historical alert volume, driver shares and seed accuracy are not substituted for causal risk reduction.

PROPOSED DECISION MODEL · INFORMATION VALUE

Buy evidence about liveness and actionability before buying scale

[PD] A restricted hypothetical decision has three states: live and actionable risk; live signal with weak preventability; or no usable live outcome truth. Prior probabilities are .45/.35/.20.

Action / net utility	Actionable	Weak action	No live truth
Fund gated validation / trial	18	-12	-24
Repair data / keep usual queue	4	8	-4
Defer expansion	0	0	0

[PD] The proposed diagnostic returns the correct state at the stated accuracy; remaining probability splits equally across the other labels. Update the prior by Bayes’ rule and choose the maximum expected utility action. Payoffs already include implementation cost.

Correct-class probability	EVSI	Study cost	Net EVSI
45%	0.000	0.500	-0.500
70%	2.760	1.500	1.260
90%	5.610	3.000	2.610

[PD] $EVSI = E[\max_a E(U_a | \text{signal})] - \max_a E(U_a)$. Perfect information is worth 7.100; every diagnostic respects that ceiling. Subtract acquisition cost once, independently of action cost.

[PD] Funding authorizes gated validation or a limited trial, never bypass of mandatory source, target, owner and safety gates. Expand the states before use. Do not add this information utility to trial benefits pricing the same consequences.

Synthetic priors, error model and payoffs. The evidence decision is distinct from the forecast policy’s assumed customer effect.

Read the engine as a product design with explicit evidence gates

[AI] The project connects forecasts to an operating workflow. Its design can be assessed independently of the unestablished claims that live alerts were consistently consumed or that the recommended actions improved customer outcomes.

Claim class	Meaning in this case
[HR] Historical record	A protected register representation of source artifacts, requirements, run reports or contribution. Original code, logs and payloads remain unrecovered; repeated summaries are not independent authentication.
[RC] Recalculated	Arithmetic from the registered breakdown or an explicitly hypothetical prospective model. Prediction rows are not sites, independent trials or customer exposure counts.
[AI] Analytical inference	A bounded interpretation of source lineage, version compatibility, goal thresholds and operating authority. An unallocated total is not adjudicated erroneous data.
[PD] Proposed design	The Forecast-to-Action Contract, silent live validation, capacity-matched randomized action test and synthetic economics. These are future decision designs, not historical achievements.

[RC] The listed forecast cells account for 86.98% of headline output. [AI] Resolve the unallocated remainder before treating that total as validated coverage. Requested and realized sources, target maturity and benchmark definitions constrain every performance statement.

[AI] Positive contribution rests on architecture, product standards, source and feature gates, goal decisions and integration direction. Those claims receive provisional registry-level hoop support. Separate model, application and action-library projects retain their own scope.

[PD] Reproduce the register arithmetic from the frozen extract and cell-linked ledger. Recover code, input snapshots, issue-time manifests, outputs and linked mature labels before rerunning the original engine or validating historical predictions. Synthetic parameters remain separate from project evidence.

DISTRIBUTION

[PD] Share this anonymized case. Retain raw metrics, identities, internal paths, source mappings and the reproducibility annex privately.

S1: protected project register and calculation ledger. Public historical quantities are indexed, relative or qualitative. The proposed evaluation has its own hypothetical population and parameter contract.

REFERENCES / PRIMARY METHODS AND SUBSTANTIVE LITERATURE

Separate prediction quality from the quality of its operating decision

These sources inform forecast qualification, operating decisions and the proposed evaluation. They do not authenticate the historical code, establish live predictive performance or demonstrate customer impact.

Elmachtoub, A. N., & Grigas, P. Smart “Predict, then Optimize.” Author manuscript, arXiv:1710.08005, version 5 (2020). Prediction error and downstream constrained decision loss are different objectives. The case adapts that distinction; it does not claim the engine implemented the authors’ optimization or training method.

Gneiting, T., & Raftery, A. E. (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, 102, 359–378. Author full text. Proper scores support honest probabilistic forecasts. A good score must still be paired with correct targets, observation timing and a defined decision use.

Gneiting, T., Balabdaoui, F., & Raftery, A. E. (2007). Probabilistic Forecasts, Calibration and Sharpness. *Journal of the Royal Statistical Society B*, 69, 243–268. Author full text. Calibration concerns forecasts and observed outcomes together; narrow intervals or aggregate coverage alone do not establish reliable conditional uncertainty.

Shmueli, G. (2010). To Explain or to Predict? *Statistical Science*, 25, 289–310. Author preprint. Explanatory and predictive tasks require different evaluation choices. A forecast driver is not automatically a causal remedy; an explanatory association does not establish untouched future-cohort prediction.

Perdomo, J., Zrnic, T., Mendler-Dünner, C., & Hardt, M. (2020). Performative Prediction. *Proceedings of Machine Learning Research*, 119, 7599–7609. Full proceedings paper. Decisions induced by predictions can change later outcomes. Monitor the deployed policy and its feedback rather than assuming the untouched-data distribution persists.

Kapoor, S., & Narayanan, A. (2023). Leakage and the Reproducibility Crisis in Machine-Learning-Based Science. *Patterns*, 4, 100804. Primary full text. The proposed validation freezes issue-time information and future targets to prevent contamination. This is a design safeguard, not a finding that historical leakage occurred.

Sculley, D., Holt, G., Golovin, D., et al. (2015). Hidden Technical Debt in Machine Learning Systems. *Advances in Neural Information Processing Systems*, 28. Full proceedings paper. Source dependencies, configurations and feedback loops matter beyond model code. A successful process exit cannot certify the wider operating system.

W3C (2013). PROV-DM: The PROV Data Model. Official recommendation. Entity, activity, agent and derivation records make lineage inspectable. The case applies that principle to realized fields, model versions and issued predictions; provenance documents a relationship, not predictive or causal validity.

Primary research, author manuscripts and official guidance. Access scope is stated. No external effect size, named software implementation or historical result is imported into this case.

REFERENCES / PRIMARY METHODS AND SUBSTANTIVE LITERATURE

Keep the measurement system distinct from the operating result

These sources inform forecast qualification, operating decisions and the proposed evaluation. They do not authenticate the historical code, establish live predictive performance or demonstrate customer impact.

Wang, R. Y., & Strong, D. M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, 12(4), 5–33. Full original article. Contextual fitness, representation and accessibility matter. A live input column cannot establish the fitness of anchored target labels for live forecast validation.

The W. Edwards Deming Institute. The Funnel Experiment. Institutional account of Deming's experiment. Process adjustment without understanding variation can worsen performance. Distinguish goal breaches from process signals, and do not attribute subsequent improvement among selected high-risk sites to the alert alone.

NIST/SEMATECH. e-Handbook of Statistical Methods: Individuals Control Charts. Official formula guidance. Moving ranges concern comparable ordered observations. Model versions, horizons, geographic cells or mixed runtime modes do not automatically form a stable operational series; mechanical limits cannot repair an invalid observation contract.

Campbell, D. T. (1976). Assessing the Impact of Planned Social Change. Occasional Paper Series, 8. Primary ERIC record. Separate implementation from evaluation and preserve criticism and replication. Published alert volume, a changed tiering rule and the customer effect of the resulting action are distinct objects of evaluation.

Ittner, C. D., Larcker, D. F., & Randall, T. (2003). Performance Implications of Strategic Performance Measurement in Financial Services Firms. *Accounting, Organizations and Society*, 28, 715–741. Author institutional record. Align measurement with the strategic decision; forecast accuracy, operator adoption and customer outcomes cannot substitute for one another.

Barnett, A. G., van der Pols, J. C., & Dobson, A. J. (2005). Regression to the Mean: What It Is and How to Deal with It. *International Journal of Epidemiology*, 34, 215–220. Publisher abstract. Extreme selected observations can moderate without treatment. Evaluate a queue on the original eligible population and an appropriate comparison.

Tucker, A. L., & Edmondson, A. C. (2003). Why Hospitals Don't Learn from Failures. *California Management Review*, 45(2), 55–72. Full original article. Local compensating action and system improvement are distinct. Alerts need ownership, an executable response and learning from exceptions; alert delivery alone does not establish that work.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20, 37–46. Primary indexed article. Report classification agreement with category prevalence and disagreement, not a solitary reliability statistic. Even consistent coding of alert or action categories cannot establish their causal truth.

Primary research, author manuscripts and official guidance. Access scope is stated. No external effect size, named software implementation or historical result is imported into this case.

REFERENCES / PRIMARY METHODS AND SUBSTANTIVE LITERATURE

Evaluate the queue policy on the same eligible exposures and resources

These sources inform forecast qualification, operating decisions and the proposed evaluation. They do not authenticate the historical code, establish live predictive performance or demonstrate customer impact.

Athey, S., & Wager, S. (2021). Policy Learning with Observational Data. *Econometrica*, 89, 133–161. Full author manuscript. Allocating constrained treatment requires identified causal value, not risk prediction alone. The case proposes a randomized policy test; it does not claim implementation of the authors' observational policy-learning estimator.

Hernán, M. A., & Robins, J. M. (2020). Causal Inference: What If. Author-hosted book and materials. Specify eligibility, treatment strategies, assignment and follow-up before selecting completed interventions. Queue-policy assignment differs from action receipt; a total policy effect retains all upstream eligible units.

Prentice, R. L. (1989). Surrogate Endpoints in Clinical Trials: Definition and Operational Criteria. *Statistics in Medicine*, 8, 431–440. Publisher abstract. An intermediate measure cannot automatically substitute for the final outcome. Prediction quality and action completion require separate links to mature customer outcomes; no clinical effect is transferred.

Collier, D. (2011). Understanding Process Tracing. *PS: Political Science & Politics*, 44, 823–830. Primary methodological article. Match the test to the contribution claim and consider alternatives. A registered product directive supplies different evidence from original implementation records, verified field adoption or attribution of outcome impact.

Campbell, M. K., Piaggio, G., Elbourne, D. R., & Altman, D. G. (2012). CONSORT 2010 Statement: Extension to Cluster Randomised Trials. *BMJ*, 345, e5661. Primary indexed guidance. State the randomization unit, eligible-unit flow and clustering assumptions. Shared intervention capacity can create interference across nominally separate sites.

Morris, T. P., White, I. R., & Crowther, M. J. (2019). Using Simulation Studies to Evaluate Statistical Methods. *Statistics in Medicine*, 38, 2074–2102. Author full text. Declare the estimand, generating process, statistical method and Monte Carlo uncertainty. Power near economic break-even needs its own relevant effect and variance scenario.

Strong, M., Oakley, J. E., Brennan, A., & Breeze, P. (2015). Estimating the Expected Value of Sample Information Using the Probabilistic Sensitivity Analysis Sample. *Medical Decision Making*. Primary full text. Value information through explicit states, actions, probabilities and payoffs. The case uses a finite-state decision model, not the authors' regression estimator.

Eldridge, S. M., Chan, C. L., Campbell, M. J., et al. (2016). CONSORT 2010 Statement: Extension to Randomised Pilot and Feasibility Trials. *BMJ*, 355, i5239. Author full text. Establish feasibility before a definitive test: live target availability, independent operating capacity, intervention fidelity and outcome follow-up remain empirical questions.

Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-Differences with Multiple Time Periods. *Journal of Econometrics*, 225, 200–230. Full author version. A possible staggered observational alternative requires credible identification assumptions and separate power planning. The randomized queue model does not validate or size that alternative.

Primary research, author manuscripts and official guidance. Access scope is stated. No external effect size, named software implementation or historical result is imported into this case.

FINAL REFEREE REPORT / INTERNAL, AGENT-ASSISTED REVIEW

The engine earns a clear next test without claiming field impact

[AI] Author-led review and independent parallel agents challenge source interpretation, calculations and evaluation design. This is internal review, not external academic review, client acceptance or certification.

Major issue	Resolution or remaining gate
Output and provenance	The headline difference is retained as unallocated, not invalid output. Job completion, requested source, realized field lineage and live target truth are distinct.
Validation transport	Seed comparisons remain seed-specific. Positive controls do not establish field predictability. Cross-version precision, recall and F1 require a common score definition before combination.
Operating value	Universal binary flags preserve the systemic goal message; prioritization requires capacity and a decision rule. A predictive driver is not a causal diagnosis or an effective action.
Contribution boundary	Architecture, standards and evidence gates receive provisional registry-level hoop support. Agent-assisted code and separate model, application and action-library projects retain their attribution.
Prospective evaluation	Independent calculations reproduce queue moments, effect-specific power, economic thresholds and information value. Numerical and decision-model issues are resolved. Live validation and feasible recruitment remain implementation gates.

Partner critique: would the COO act on page 1?

[AI] Yes: authorize the live evidence gate, assign data and operations owners, and reserve scale authorization for a test of the operating response. The decision preserves useful product work while making the next investment falsifiable.

Positive contribution within the evidence boundary

[AI] The registered architecture specifies source adapters, metric identity, explicit goals, controlled feature activation, alert routing and a planned operator interface. Its value here is the disciplined design of a forecast-to-action pathway, not an invented customer result.

FINAL REVIEW DISPOSITION

[AI] Source, editorial, independent numerical, protocol and rendered-page reviews pass for the supported design case. Historical adoption, live predictive validity and attributable customer benefit remain unestablished.

Internal author-led and independent agent review. Numerical reproduction, protocol closure and actual rendered-page inspection are documented privately. No external certification is implied.