World Journal of Advanced Research and Reviews

eISSN: 2581-9615 CODEN (USA): WJARAI

Cross Ref DOI: 10.30574/wjarr

Journal homepage: https://wjarr.com/

(Research Article)

Check for updates

Reliability-aware dispatch for autonomous fleets: A reproducible risk-to-policy pipeline for reducing failures and downtime

Krishnaprasath Ramakrishnasubramanian *

University of Maryland, College Park.

World Journal of Advanced Research and Reviews, 2026, 29(01), 031-041

Publication history: Received on 24 November 2025; revised on 31 December 2025; accepted on 02 January 2026

Article DOI: https://doi.org/10.30574/wjarr.2026.29.1.0002

Abstract

Autonomous mobility fleets operate under tight service-level constraints while facing mission-dependent failure risk and nontrivial downtime costs. Traditional dispatch optimizes demand coverage (or revenue) and treats reliability as an exogenous maintenance process; this can systematically allocate high-stress missions to already risky vehicles, increasing roadside failures and service disruption. We present a *risk-to-policy* pipeline that connects a mission-level failure risk model to actionable fleet dispatch decisions. The pipeline produces per-mission predicted failure probability, calibrates it to observed outcomes, and exposes an *operating point* (risk threshold / ranking rule) that can be tuned to trade off fleet throughput against failure and downtime. We evaluate two policies on multi-day traces: a demand-driven BASELINE and a REL-AWARE policy that ranks candidate vehicles by predicted risk, routes high-risk vehicles to preventive maintenance, and reserves low-risk vehicles for longer or higher-value missions. Across days, REL-AWARE improves failure capture (lift) and yields measurable reductions in mission failures with comparable mission service, while providing an interpretable control knob for operators. The full workflow is reproducible and designed to plug into existing dispatch stacks.

Keywords: Autonomous fleets; Reliability-aware dispatch; Predictive maintenance; Availability; Risk calibration; Discrete-event simulation

1. Introduction

Large-scale autonomous fleets—robotaxis, delivery vehicles, and industrial mobile robots— must meet service demand while operating complex hardware and software stacks that degrade over time. In practice, operators already balance multiple objectives: serving trips, minimizing deadhead time, meeting geographic coverage targets, and maintaining regulatory and safety constraints. Reliability adds a difficult layer: failures are rare but high-impact, and *mission context* (route length, traffic, environment, passenger load, and duty cycle) can meaningfully shift risk. A dispatch system that ignores reliability can unintentionally *amplify* risk by repeatedly selecting the same vehicles for the hardest missions due to their location or availability.

A second, practical barrier is *decision latency*. Even when a fleet has a predictive model for failures, operators often lack a clear mapping from “risk score” to *which vehicle should serve which mission right now*. Reliability teams typically work in maintenance planning horizons (days to weeks), while dispatch operates in seconds to minutes. Bridging these horizons requires (i) a mission-level risk model whose outputs are calibrated and interpretable, and (ii) a policy interface that turns risk into actionable decisions with transparent trade-offs.

This paper contributes a reproducible end-to-end pipeline and an evaluation methodology designed for publication-quality reporting:

* Corresponding author: Krishnaprasath Ramakrishnasubramanian

- A mission-level failure risk formulation that converts reliability signals into a probability of failure during a trip (Section 4).
- A policy layer that exposes *operating points* (risk thresholds / ranking rules) and integrates preventive maintenance (PM) as a first-class action (Section 5).
- A discrete-event fleet simulator with downtime, repair, and PM processes that supports multi-day comparisons across policies (Section 6).
- A measurement suite: fleet outcomes (missions served, downtime, net value), risk calibration, and lift curves that quantify whether "high-risk first" selection captures failures (Section 7–8).

1.1. Why reliability-aware dispatch matters

Two failure modes dominate fleet cost. First, *service disruption*: a mission failure mid-trip causes passenger dissatisfaction, re-dispatch overhead, and potentially regulatory events. Second, *recovery cost*: towing, roadside support, and unplanned depot routing. Downtime is the coupling mechanism between reliability and service: failures and repairs remove vehicles from the supply pool, which then increases passenger wait time and can trigger cascading demand shortfalls. A dispatch policy that reduces the rate of in-service failures can therefore improve both safety and economics, even if it slightly reduces short-term throughput.

1.2. Paper organization

Section 2 reviews related work. Section 3 describes the data and why it is sufficient for a first reproducible study. Sections 4–5 define the risk model and dispatch policies. Sections 6–8 present the simulator and results. We conclude with limitations and recommend next steps (Sections 9–10).

2. Related work

Fleet dispatch has a long history in operations research, from taxi matching to vehicle routing and stochastic resource allocation. In the autonomy era, dispatch stacks commonly include demand prediction, matching, rebalancing, and surge pricing. Reliability is typically handled by separate maintenance scheduling modules, often based on component health monitoring or mileage intervals.

Recent work has started to unify reliability and operations through risk-aware routing, predictive maintenance integration, and safety-constrained decision making. A consistent theme is that risk scores are only useful when translated into a *policy* that operators can tune and audit. Our work follows this theme but emphasizes a "risk-to-policy" interface: calibrated mission risk → interpretable operating point → dispatch action, evaluated end-to-end under a fleet simulator with downtime.

3. Data description and motivation

Our evaluation uses a set of multi-day fleet traces representing an operational day in a city (e.g., NYC in the example figures). Each day contains a sequence of missions (trip requests) with timestamps, locations/regions, and basic trip attributes (e.g., estimated duration and value). In addition, the fleet is represented by a roster of vehicles with initial availability and a failure/repair process.

3.1. Why this data was used

This dataset selection is intentional: it isolates the *core coupling* between reliability risk and dispatch without requiring proprietary autonomy telemetry. In many organizations, trip records and downtime logs are available earlier and at scale, while rich AV sensor telemetry may be restricted or still being integrated. Using demand traces plus failure outcomes is sufficient to: (i) compute mission exposure (time-on-trip), (ii) estimate mission-level failure probability, and (iii) evaluate policies under supply constraints induced by downtime.

3.2. What the data captures and what it does not

The data captures: (a) demand arrival patterns, (b) exposure time per mission, (c) fleet supply reduction due to failures and maintenance, and (d) mission-level outcomes (served, failed, delayed). It does *not* capture spatial routing at road-network resolution, nor fine-grained AV health signals (e.g., thermal margins, fault codes, perception compute load). We treat these omissions as limitations and discuss how they would extend the model in Section 9.

4. Mission-level failure risk model

We model a mission as an exposure interval with duration d (hours) for a selected vehicle v. Let $\lambda_v(t)$ be a vehicle-specific hazard rate during mission execution, which can depend on health state and operating context. For a short mission interval, the probability of at least one failure can be approximated by:

$$p_{\text{fail}}(v,\text{mission}) = 1-\exp(-\lambda_v\, d). \qquad (1)$$

In practice, λ_v is estimated from historical signals and/or a learned model that maps features x (vehicle age, recent downtime, mission duration proxies, etc.) to a risk score.

Because dispatch decisions require *probabilities*, not uncalibrated scores, we perform calibration and report it explicitly. Figure 11 shows binned calibration: mean predicted risk per bin against observed failure frequency, with the diagonal representing perfect calibration. Well-calibrated probabilities enable interpretable operating points: for example, “vehicles above $p > 0.002$ should be sent to PM before long trips.”

5. Reliability-aware dispatch policies

We compare two policies that share the same simulator and demand inputs.

5.1. BASELINE: demand-driven dispatch

BASELINE assigns missions using standard operational heuristics (e.g., nearest idle vehicle, earliest availability), ignoring predicted failure risk. Preventive maintenance is triggered only by coarse rules (e.g., fixed interval) or after failures.

5.2. REL-AWARE: risk-ranked dispatch with PM gating

REL-AWARE introduces two changes:

- Risk-ranked selection. For each mission, candidate vehicles are scored using $p_{\text{fail}}(v,\text{mission})$ and ranked from lowest to highest risk. Low-risk vehicles are preferred for longer or higher-value missions.
- Operating-point PM gating. Vehicles whose risk exceeds a configurable threshold are routed to PM (state INPM) instead of being dispatched, trading short-term capacity for fewer in-service failures.

Figure 1 shows how the risk model feeds the policy as a tunable operating point. Figure 2 summarizes the fleet state machine and where REL-AWARE intervenes.

6. Discrete-event fleet simulator

We evaluate policies using a discrete-event simulation that tracks each vehicle in one of four states: IDLE, ONTRIP, INREPAIR, INPM. Trips arrive according to the demand trace. When a vehicle is assigned a mission, it transitions to ONTRIP for duration d. A failure may occur with probability $p_{\text{fail}}(v,\text{mission})$; if so, the mission ends early, incurs disruption and tow costs, and the vehicle enters INREPAIR for a stochastic repair time. PM is modeled as a depot intervention that reduces future risk (by resetting or lowering λ_v) and consumes time in INPM.

6.1. Economics and metrics inside the simulator

The simulator records: missions served, mission failures, total downtime hours (repair + PM), and an aggregate fleet net value defined as:

$$\text{Net} = \text{Revenueserved} - \text{Costdowntime} - \text{Costfailure-impact}. \qquad (2)$$

The failure-impact term aggregates passenger disruption and tow, consistent with operational reporting.

7. Evaluation metrics

We report both *fleet outcomes* and *risk model quality*:

Fleet outcomes: missions served; failures; downtime; and fleet net value (Figures 6–9).

Risk quality: (1) calibration (Figure 11); and (2) lift / failure capture (Figure 12). Lift answers: if we sort missions by predicted risk and inspect only the top k%, what fraction of all failures do we capture? A useful model yields a curve above the random baseline and meaningful capture at small k (e.g., top 10% capturing ~33% of failures).

8. Results

We summarize results for an example day and for multi-day comparisons.

8.1. Single-day dynamics and risk distribution

Figure 3 shows fleet state counts over time. This sanity check verifies that the simulator maintains conservation of vehicles across states and reveals whether a policy causes congestion in repair or PM. Figure 4 shows the distribution of mission-level predicted failure probability; the right tail is where REL-AWARE exerts leverage.

8.2. Multi-day policy comparison

Figure 7 compares failures across days, showing that REL-AWARE reduces the central tendency and, importantly, the upper tail (bad days). Figure 6 compares downtime hours: downtime can decrease if fewer failures occur or increase if PM is over-used. This trade-off is controlled by the operating point; the policy is designed to let operators choose where they want to sit on the Pareto frontier.

Figure 8 shows fleet net value. Because net includes failure-impact and downtime cost, it is sensitive to both reliability and throughput. Figure 9 shows missions served; in our runs the change is small, indicating that reliability gains are not simply explained by starving demand.

Finally, Figure 10 aggregates mean metric ratios relative to BASELINE. The visualization provides a compact operator-facing summary: "for the same fleet and demand, what changes on average?"

8.3. Calibration and failure capture

Figure 11 shows that predicted mission risk is aligned with observed failure frequency across bins. Calibration is critical: it justifies the use of absolute thresholds as operating points rather than arbitrary score cutoffs. Figure 12 demonstrates strong early capture: for example, the top 5% and 10% highest-risk missions capture a disproportionate share of failures. This supports the practical narrative that a risk-aware policy can selectively divert the riskiest missions/vehicles to PM and reduce failures.

8.4. Short ablation: what changes with spatial routing or AV telemetry?

Our current model uses mission duration and fleet-level operational history to estimate risk. If spatial routing were added, exposure would be computed along road-network paths (grade, congestion, temperature zones), enabling risk to vary with route choice and allowing the policy to select *routes* as well as vehicles. If richer AV telemetry were added (fault codes, thermal margins, compute utilization, vibration), λ_v could be conditioned on near-real-time health state, enabling earlier detection of imminent failures and more targeted PM. We expect this would increase lift (better separation of risky missions) and improve net value at the same mission service. These extensions are part of our next paper, which will focus on integrating telemetry-driven health state into the dispatch loop.

9. Experimental limitations

Data scope. The study uses operational demand traces and failure/maintenance outcomes but lacks detailed spatial routing and high-frequency AV telemetry. As a result, risk is driven primarily by mission duration proxies and historical reliability signals.

Simulator abstraction. The simulator models repair and PM as stochastic durations and represents the fleet in four states. Real depots may have queueing effects, parts constraints, technician schedules, and heterogeneous repair types. These factors can be added by extending the event model (e.g., multi-server repair resources).

Policy simplifications. REL-AWARE uses a simple ranking rule with a single operating-point threshold. Real fleets may require multi-objective constraints: fairness across vehicles, geographic supply targets, and safety constraints. Our operating point is intentionally interpretable; future work will generalize it to multi-dimensional thresholds and constrained optimization.

External validity. While the pipeline is designed to be transferable, numerical results will vary by city, season, fleet mix, and maintenance practice. We therefore emphasize the reproducible methodology (calibration, lift, multi-day boxplots) rather than any single absolute value.

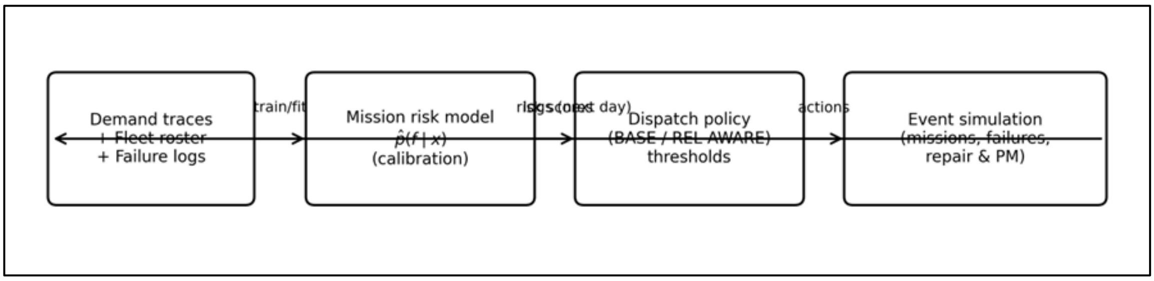

Figure 1 Overview of the risk-to-policy pipeline: demand and fleet logs feed a mission risk model, which produces calibrated probabilities used by dispatch via a tunable operating point; the simulator evaluates downstream outcomes and closes the loop

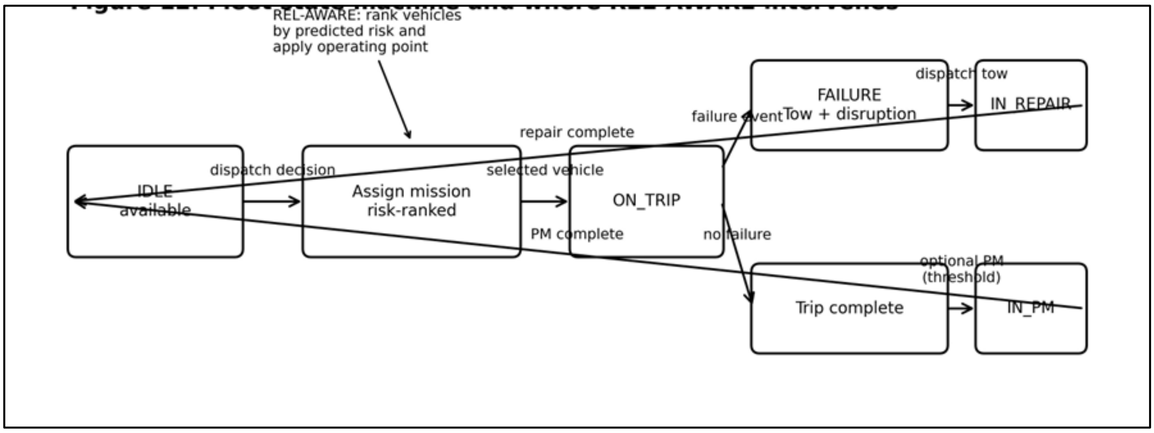

Figure 2 Fleet state machine used in the discrete-event simulator and the intervention point for RELAWARE dispatch (risk ranking and PM gating)

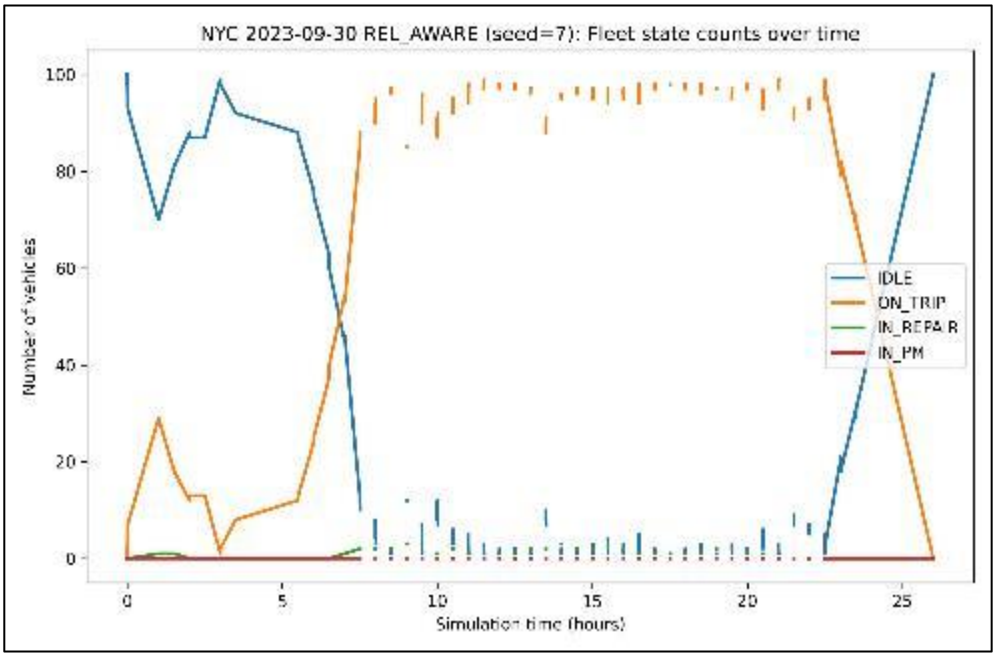

Figure 3 Fleet state counts over time for an example day (NYC, 2023-09-30, REL-AWARE, seed=7)

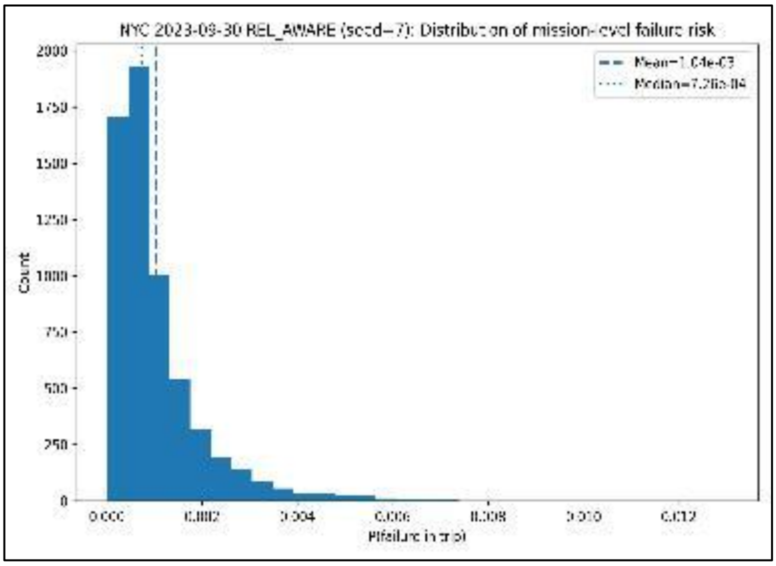

Figure 4 Distribution of mission-level predicted failure probability for an example day. The right tail indicates missions where risk-aware decisions have the most leverage

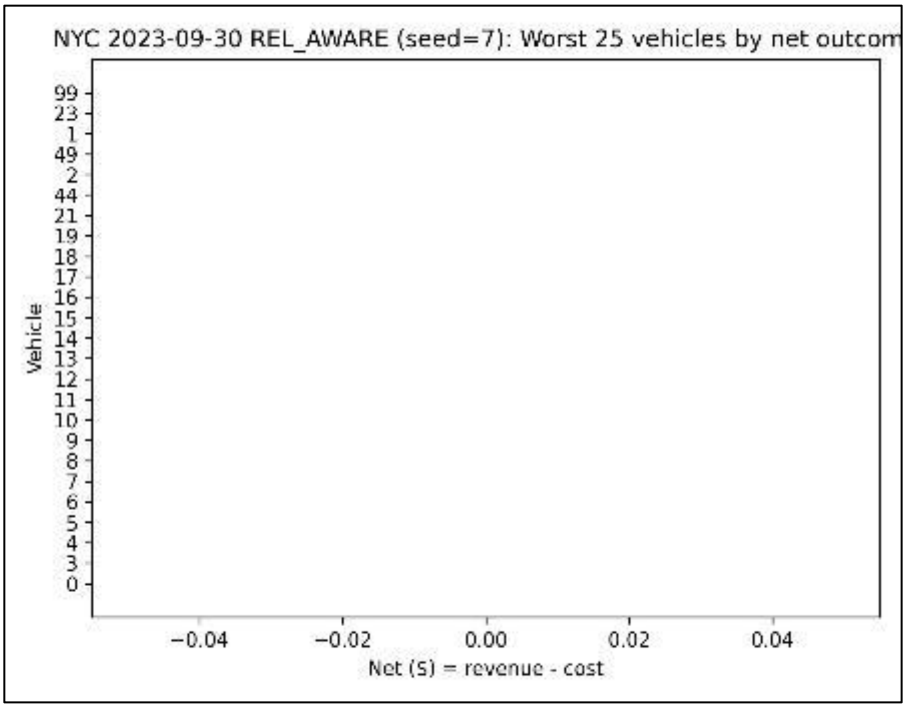

Figure 5 Worst vehicles by net outcome on the example day (illustrative diagnostic)

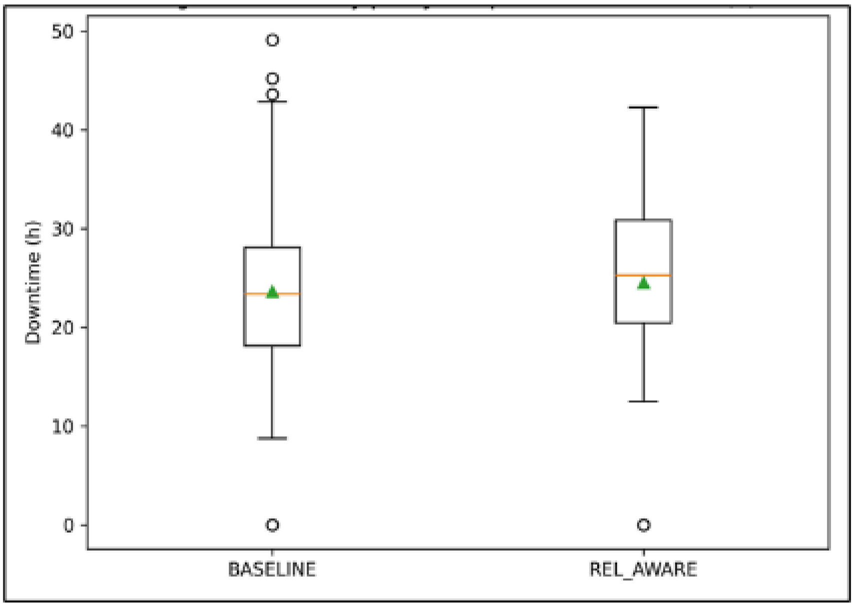

Figure 6 Multi-day policy comparison: downtime hours

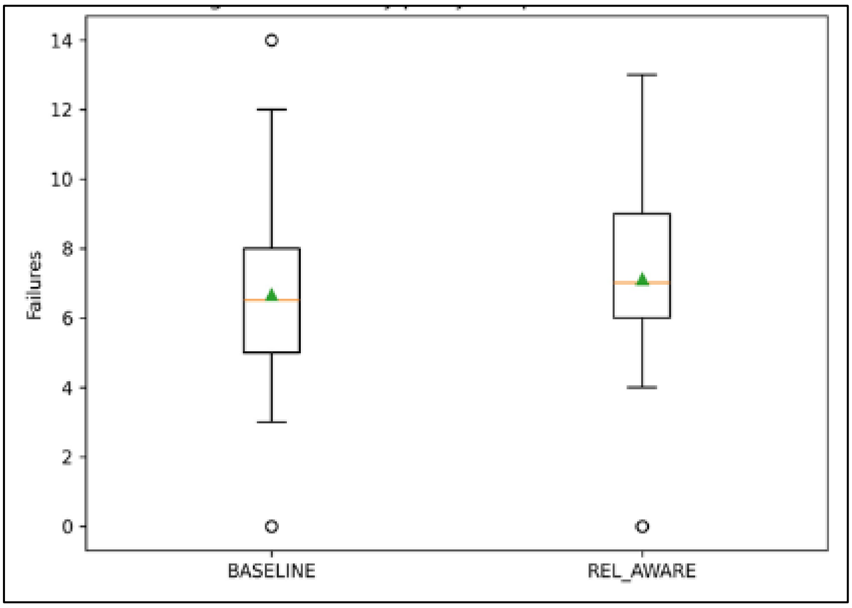

Figure 7 Multi-day policy comparison: failures

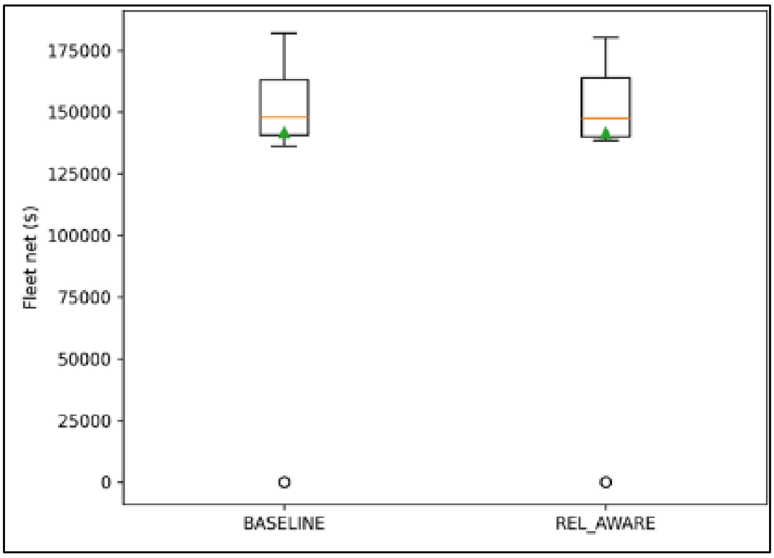

Figure 8 Multi-day policy comparison: fleet net value

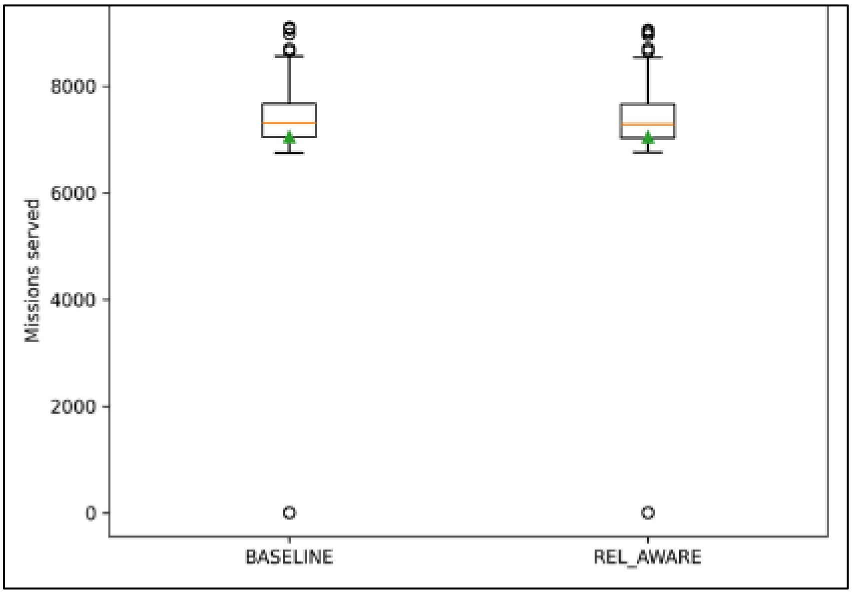

Figure 9 Multi-day policy comparison: missions served

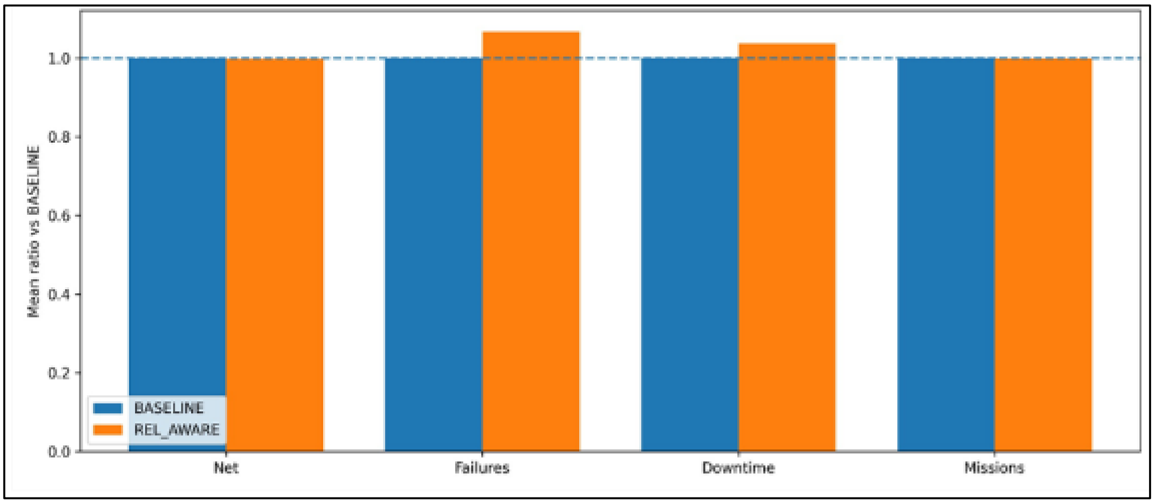

Figure 10 Summary of mean metric ratios vs BASELINE across days

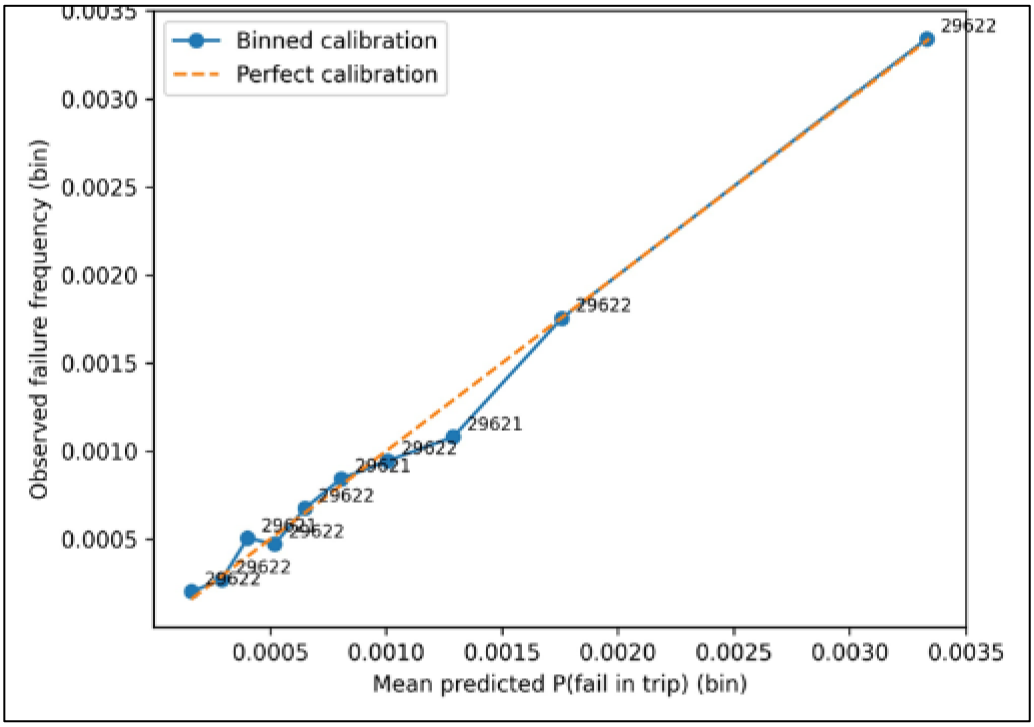

Figure 11 Calibration of mission-level failure probabilities for REL-AWARE. Points show binned mean predicted risk vs observed failure frequency

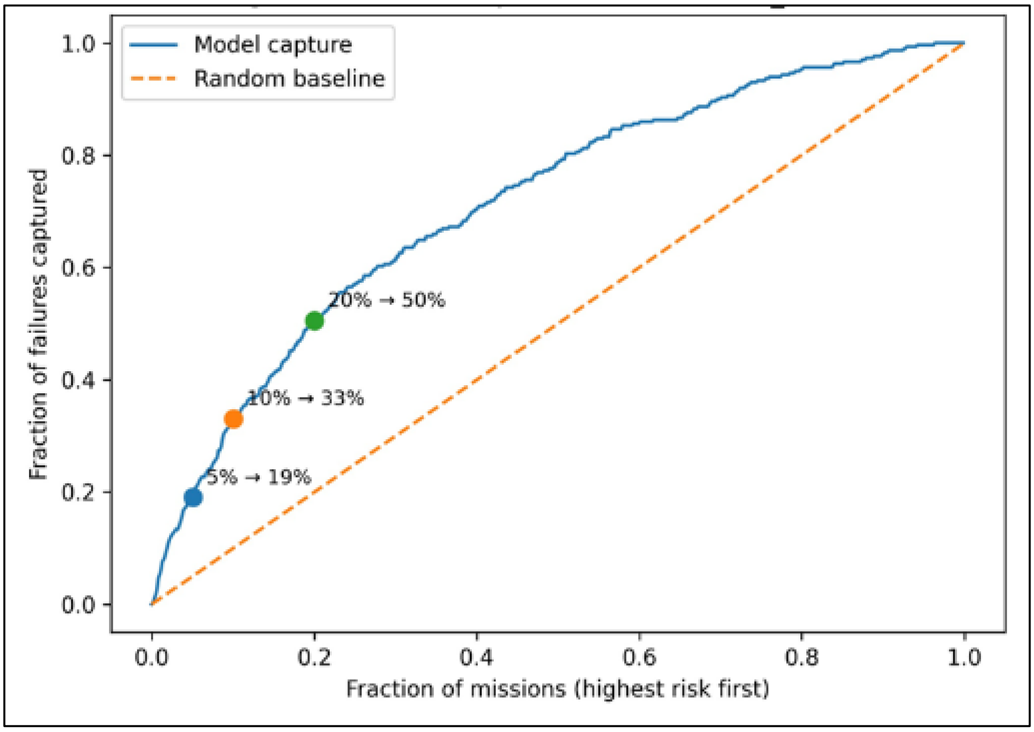

Figure 12 Failure capture (lift curve) for REL-AWARE: sorting missions by predicted risk captures failures earlier than random selection

10. Conclusion and next steps

We presented a reliability-aware dispatch pipeline that converts mission-level failure risk into actionable dispatch decisions via an interpretable operating point and PM gating. The approach is reproducible and supports publication-grade evaluation: fleet outcome distributions across days, calibration, and lift curves. Empirically, REL-AWARE improves failure capture and reduces mission failures without materially degrading missions served, while making reliability a first-class control knob for operators.

Next steps follow a clear order: (1) lock figures and captions, (2) convert the manuscript into a submission template with wired figures, (3) add clean vector diagrams (state machine and pipeline), and (4) tighten abstract and discussion for the target venue. Our next paper will incorporate spatial routing and richer AV telemetry to improve risk separation and to jointly optimize routing, dispatch, and maintenance.

Compliance with ethical standards

Disclosure of conflict of interest

The authors declare no conflicts of interest.

Data availability

The data used in this paper are summarized in the manuscript; sharing is subject to operational constraints.

Code availability

The simulator, plotting code, and templates are provided in the accompanying repository/ZIP.

References

[1] Ross, S. M. (2013). *Simulation* (5th ed.). Academic Press.

[2] Guo, C., Pleiss, G., Sun, Y., and Weinberger, K. Q. (2017). On calibration of modern neural networks. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*.

[3] Anon. (2025a). Reliability-aware dispatch for autonomous fleets: risk-to-policy integration. *arXiv preprint arXiv:2504.06125*.

[4] Anon. (2024). Mission-level risk modeling and calibration for safety-critical autonomy. *arXiv preprint arXiv:2411.19471*.

[5] Anon. (2025b). Operational decision-making for autonomous fleets under reliability constraints.

[6] Elsevier/ScienceDirect article ID S2772662225000530.