IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p-ISSN: 2278-8727, Volume 26, Issue 6, Ser. 3 (Nov. – Dec. 2024), PP 49-64
www.iosrjournals.org

Systems And Methods for Predictive Failure Intelligence and Autonomous Self-Healing of Distributed Computing Infrastructure Supporting Critical Supply Chains

Sohail Sayed; Nauman Sayed
California State University

Abstract *— The computing substrate that runs critical supply chains — checkout pipelines, order orchestration, inventory services, payment gateways, logistics routing — fails in ways that are expensive, fast, and hard to explain: one hour of downtime is estimated to cost $100 million on major shopping days at Amazon, and supply-chain-level disruptions average $184 million in cost per event with traditional response consuming 3–5 days (Ahmed et al., 2023; Yan et al., 2025). This paper discloses a system and associated methods for closing the gap between failure onset and recovery: predictive failure intelligence over a continuously updated service dependency graph; blast-radius estimation that converts predicted failures into business-consequence terms; retrieval-grounded root-cause inference with hybrid verification; recovery-plan generation with safety validation; tiered autonomous remediation; recovery verification with automatic rollback on failure; and continuous learning from every resolved event. The system consolidates the reported evidence envelope: a domain-adapted LLM over OpenTelemetry telemetry reduces mean time to resolution by 84.2% versus rule-based methods with 95% F1 anomaly detection and 91% of common incidents automated (Wang et al., 2025); the ARCH retrieval-augmented self-healing framework achieves up to 82% MTTR reduction with an 89.5% autonomous success rate (Sharma & Jain, 2026); a large-scale Microsoft study across 44,340 incidents from 1,759 services validates LLM-suggested root causes and mitigations with incident-owner human evaluation (Ahmed et al., 2023); alert-augmented multimodal pipelines reach up to 97% RCA accuracy (Sarda et al., 2024); graph-augmented agentic workflows improve over state-of-the-art tools by up to 42.22% (Tian et al., 2025); a multimodal early-warning engine reaches 92.1% recall and 93.4% precision with 42-minute average lead time (Wang, 2024), neural architectures extend the horizon to two-to-four weeks* (Huang S. , 2024)*, and physics-informed graph networks cut disruption-prediction error by 23% (Petrova & Hughes1, 2025); and governed deployment reduces unsafe decision outputs by 45% while sustaining drift-detection accuracy above 92% (Jingar, 2023). The paper argues that predictive failure intelligence — not faster reaction — is the binding constraint on infrastructure reliability, and that autonomous self-healing is safe only when every diagnosis is evidenced, every plan is validated, and every action is verified or rolled back.*

Keywords *— Predictive failure intelligence, self-healing systems, root cause analysis, large language models, service dependency graphs, blast radius, autonomous remediation, rollback, recovery verification, AIOps, distributed systems, supply chain infrastructure, governance.*

I. INTRODUCTION

Critical supply-chain operations run on a computing substrate whose failures are measured in dollars per minute: on major shopping days a single hour of downtime at Amazon is estimated to cost $100 million, and supply-chain-level disruptions average $184 million in cost per event with response windows of 3–5 days (Ahmed et al., 2023; Yan et al., 2025). The traditional toolkit — manual root cause analysis, static troubleshooting guides, reactive remediation — cannot meet the demands of modern distributed commerce systems, and the bottleneck is universal: operations engineers spend an inordinate amount of time analyzing data and formulating hypotheses, detracting from time better spent resolving the incident and restoring functionality (Chen et al., 2024).

Three convergences make a systemic answer possible. First, prediction: machine-learning fault prediction for large-scale distributed systems is an established capability (Padhy et al., 2026), multimodal early-warning engines fuse news, telemetry, and social signals into risk signals with 92.1% recall, 93.4% precision, and a 42-minute average lead time (Wang, 2024), and neural early-warning architectures extend the horizon to two-to-four weeks for high-impact events (Huang S. , 2024). Second, graph awareness: service dependency graphs make failure propagation representable — graph neural networks capture multi-tier dependencies, risk propagation, and network effects through message passing (Liu et al., 2026), hypergraph resilience inference predicts network behavior without explicit system dynamics (Shen et al., 2026), and physics-informed variants cut prediction error by 23% (Petrova & Hughes1, 2025). Third, agency: LLM-driven observability and self-healing frameworks demonstrate MTTR reductions above 80% with automation rates near 91% (Sharma & Jain,

2026; Wang et al., 2025), while honest evaluation of LLM agent architectures documents the failure profile that any trustworthy system must contain — hallucinated reasoning paths and inefficient use of context (Cornacchia et al., 2025), formalized across 48,000 simulated failure scenarios into a taxonomy of 16 reasoning failures that predict final correctness (Riddell et al., 2026).

What is missing — and what this paper discloses — is the unified system: the architecture that couples prediction, graph-based consequence estimation, evidence-bound diagnosis, validated remediation, verified recovery, rollback, and learning into one closed loop, with trustworthiness enforced by explicit safety validation and autonomy-tier governance. This paper is the infrastructure twin of the program's supply-chain systems-and-methods disclosure: where the supply-chain loop predicts, mitigates, and recovers from network disruptions, this system predicts, mitigates, and recovers from failures of the computing substrate that supports the commerce layer — completing the program's path from signal to platform to self-healing infrastructure.

The contributions of this paper are:

- A system architecture (the "resilience loop") composed of ten components across four layers — perception and prediction (predictive failure intelligence, service dependency graph, blast-radius estimation), diagnosis (root-cause inference), action (recovery-plan generation, safety validation, autonomous remediation, recovery verification, rollback), and learning (continuous learning) (Sections III–V).
- A formalization of the system's core inference problem — failure prediction with calibrated lead time and blast-radius propagation over dynamic dependency graphs — and its decision problem — safety-validated remediation with verification-and-rollback guarantees — with cost models anchored to reported downtime economics.
- A consolidated analysis of reported evidence across every component, with worked calculations and clearly flagged illustrative extensions (Sections VI–VII).
- A research agenda on calibrated autonomy tiers, rollback-aware planning, and predictive-lead-time contracts (Sections VIII–IX).

II. RELATED WORK

A. Predictive failure intelligence

The prediction layer rests on a mature ML foundation for large-scale distributed systems (Padhy et al., 2026), extended across the commerce pipeline: AI-driven and neural early-warning frameworks for supply chain risk detection demonstrate practical value in disaster and geopolitical settings and extend the prediction horizon to two-to-four weeks for high-impact disruptions (Huang S. , 2024); multimodal deep learning fuses news, social media, and operational data through transformer-based risk-signal extraction, LSTM autoencoders for temporal anomalies, and cross-modal graph fusion, reaching 92.1% recall and 93.4% precision with a 42-minute average lead time and validated manufacturing case studies (Wang, 2024); machine-learning early-warning adoption is associated with up to 41% forecast-accuracy improvement, 15% supply chain cost reduction, and 35% disruption reduction (Kasali et al., 2025); explainable machine-learning prediction attributes risk drivers to upstream dependencies, geopolitical instability, and transportation chokepoints through multi-dimensional XAI (Iruku, 2025), with SHAP and LIME the most applicable post-hoc methods across the supply chain literature (Akubilla et al., 2025; Dehan et al., 2025); and reinforcement learning enables adaptive, proactive disruption identification (Aboutorab et al., 2024). Food-security prediction establishes the evaluation disciplines — accuracy, transparency, usability, explainability — that prediction credibility requires (Machefer et al., 2025; Zhou et al., 2021), with news-stream deep models forecasting crises up to 12 months ahead (Balashankar et al., 2023), and the text-mining substrate — extraction of supply network structure from news (Wichmann et al., 2020), disinformation filtering (Akhtar et al., 2022), and systematic online-data risk analytics (Gelastopoulos & Keramydas, 2025) — supplies the ingestion discipline. Early-warning research confirms that open-access business data aggregate into forecasting inputs (Nagy & Foltin, 2022).

B. Service dependency graphs and graph-native risk

The structural layer is graph native. A comprehensive review documents GNNs as a powerful framework for supply chain and infrastructure network modeling — multi-tier visibility, demand forecasting with network effects, risk propagation, and relationship dynamics through message passing, with graph convolutional, graph attention, and graph recurrent architectures matched to tasks, and temporal and heterogeneous GNNs identified as the critical research directions for dynamic network evolution (Liu et al., 2026). Link prediction over real automotive networks outperforms existing algorithms (Kosasih & Brintrup, 2021); neurosymbolic GNN-plus-knowledge-graph reasoning infers hidden relationship types — companies, products, production capabilities, certifications — across automotive and energy datasets (Kosasih et al., 2022); uncertain knowledge graphs model noisy, partially known relations (Brockmann et al., 2022); and generative models complete missing edges (Ge & Brintrup, 2024). For resilience specifically, the Supply-Chain Resilience Inference Hypergraph Network

formalizes resilience inference from hypergraph topology and observed inventory trajectories without explicit dynamic equations, significantly outperforming MLP, representative GNN variants, and ResInf baselines (Shen et al., 2026), and physics-informed GNNs embedding conservation of flow, capacity constraints, and lead-time dependencies reduce disruption-prediction error by 23% while maintaining interpretability (Petrova & Hughes1, 2025). LLM-grounded temporal-graph models couple Temporal Graph Attention Networks with structured LLM reasoning, achieving AUC 0.761, AP 0.344, and recall 0.504 under strict chronological splits with 99.6% explanation directional consistency (Xue et al., 2026). Graph construction is automated: LLMs extract entities and relations from news, regulatory filings, and trade databases into semantically rich, dynamically refreshed graphs (Duttyal, 2025), zero-shot multi-step prompting builds high-quality graphs from heterogeneous datasets (Wang & Tsung, 2025), ontology reuse cuts design cost (Ham et al., 2025), and ontology-driven graph-RAG improves supplier discovery (Li et al., 2024). The duality principle — a network edge is simultaneously a semantic triple — makes centrality-guided traversal produce interpretable narratives (Heus et al., 2025).

C. LLM-based root cause analysis

The diagnosis layer is validated at industrial scale. A Microsoft study evaluated LLM-suggested root causes and mitigations across 44,340 incidents from 1,759 services in zero-shot, fine-tuned, and multi-task settings, with incident-owner human validation of correctness and readability (Ahmed et al., 2023). LLMs parse high-volume data, discern relevant information, and produce succinct, insightful outputs, adapting to new incident types (Chen et al., 2024). Alert augmentation sharpens accuracy: multimodal LMTA (logs, metrics, traces, alerts) pipelines with developer-set alert offsets and thresholds reach up to 97% RCA accuracy in privacy-aware industrial settings (Sarda et al., 2024). Expert-agent systems transform metrics, logs, and traces into natural-language descriptions processed by collaborating agents, significantly outperforming baselines and deploying in Microsoft's AIOpsLab (Zhong et al., 2026). Graph-augmented agentic workflows combine statistical causal inference with LLM-driven iterative reasoning, improving over state-of-the-art tools by up to 42.22% (Tian et al., 2025). At enterprise scale, hybrid architectures — deterministic graph-based correlation fused with probabilistic scoring and contextual LLM analysis over vector-retrieved telemetry — improve diagnostic accuracy and reduce MTTR relative to rule-based systems while explicitly mitigating hallucination risks (Veershetty, 2026). The honest counter-evidence defines the specification: tool-augmented LLM agents fail on hallucinated reasoning paths and inefficient context use (Cornacchia et al., 2025), and controlled evaluation across two real-world case studies, two agentic workflows, and 48,000 simulated failure scenarios (228 days of execution) yielded a labeled taxonomy of 16 reasoning failures whose presence predicts final correctness (Riddell et al., 2026).

D. Recovery planning, remediation, and self-healing

The action layer is end-to-end demonstrated. LLMs translate operational problems into mathematical models and executable code through automatic modeling, auxiliary optimization, and direct solving (Wang & Li, 2025); conversational planning systems run counterfactual what-if/why-not scenarios (Yu et al., 2025); LLMs reconfigure algorithmic components and explain rationale for routing problems (Ghiani et al., 2025); and the democratization agenda confirms substantial capability for formal modeling with reliability challenges remaining (Li et al., 2025; Wasserkrug et al., 2025). Hybrid agentic frameworks that strictly decouple semantic reasoning from mathematical calculation reduce inventory costs by 32.1% versus an interactive end-to-end solver (Duan et al., 2025); OR-augmented agents outperform either approach alone, and human–AI teams outperform humans or agents alone across more than 1,000 benchmark instances (Baek et al., 2026); simulation-based human-specified optimization achieves statistically optimal, stable outcomes (Wu, 2026). Self-healing frameworks operationalize the loop: a domain-adapted LLM over OpenTelemetry telemetry performs multimodal analysis over metrics, logs, and traces with chain-of-thought reasoning, reducing MTTR by 84.2% with 95% F1 anomaly detection and 91% automation of common incidents (Wang et al., 2025); ARCH — perception, cognition, knowledge, and action layers with retrieval-augmented generation over historical incident data — achieves up to 82% MTTR reduction and an 89.5% autonomous success rate (Sharma & Jain, 2026); self-healing support interfaces integrate telemetry aggregation, intelligent RCA, and automated runbook execution with proactive remediation (Singh, 2025); and incident-intelligence agents report a 30% mean-time-to-root-cause reduction with a 22% incident-resolution-accuracy improvement (Jingar, 2023).

E. Safety validation, digital twins, and rollback

Autonomous plans must be validated before and after execution. Stress-test models observe disruption impacts, understand the reasons for operational disruption, and test recovery strategies across five control loops — demand, lead time, inventory, production, transportation (Ivanov & Dolgui, 2021); cyclic digital-twin methodologies support resilient, sustainable reconfiguration (Cimino et al., 2024); a systematic review of 89 articles documents twin contributions at each resilience stage — preparedness, resistance, rebound, and growth — with data quality and governance as adoption barriers (Liu et al., 2026); and a generalized decision-support

system integrating prescriptive optimization, predictive simulation, and real-time LLM-agent detection into a twin for stress testing has industrial deployments at Ford and Microsoft (Ivanov, 2026). Verification-and-rollback is embedded in governed deployment: governed LLM-based optimization integrates policy constraints, human-in-the-loop oversight, continuous monitoring, and drift detection (Jingar, 2023), and self-healing stacks keep a well-defined interface between autonomous remediation and human operators so that transparency and trust are maintained (Singh, 2025).

F. Governance, uncertainty, and trust

Autonomy is licensed by calibration and governance. LLMs generate plausible but nonfactual content (Huang et al., 2024), with taxonomy separating intrinsic from extrinsic hallucination (Gumaan, 2025; Zhang et al., 2025); detection and mitigation are established via self-consistency and verbalized confidence (Kang et al., 2025; Shorinwa et al., 2025), semantic entropy over meanings (Farquhar et al., 2024), conformal prediction with distribution-free guarantees (Campos et al., 2024), calibrated sets that contain factual responses with high probability (Cherian et al., 2024), and conformal abstention that bounds hallucination rates at user-specified error levels (Yadkori et al., 2024) — with calibrated prediction sets deciding when an agent should ask for help (Chakraborty et al., 2025). Evaluators are fallible: LLM-as-a-judge suffers judicial hallucination (Liu, 2025). Governance converts measurement into control: guardrail-centric fine-tuning with deterministic enforcement (Davenport, 2026); governed optimization reducing unsafe outputs by 45% with drift detection above 92% (Jingar, 2023); decision-authority frameworks with deterministic boundaries, human-final authority, audit-ready lineage, and uncertainty-aware hard stops (KALAFATOGLU, 2025, 2026); and the consolidated trustworthy-AI requirements set — fairness, explainability, accountability, reliability, acceptance (Kaur et al., 2022). Interpretability is foundational in risk decision-making (Baryannis et al., 2019), with XAI improving trust and regulatory compliance (Akubilla et al., 2025; Dehan et al., 2025; Iruku, 2025; R. et al., 2024). Agentic security systems document actions on blockchain ledgers for integrity and auditing (Syed et al., 2025).

G. Evaluation

Operational LLM systems are gated by measurement: OpsEval's 7,334 multiple-choice and 1,736 QA questions across 24 LLMs show standard NLP metrics mislead in operational domains, with an Ops-specific method reaching 0.9185 human agreement (Liu et al., 2025); pass@k-style metrics conceal compounding multi-step failures (Chen et al., 2025); and continual, modular benchmarking is the documented discipline for self-improving systems (Protogeros & Vanbever, 2025). Supply-chain and infrastructure benchmarks mirror the pattern (Cui et al., 2026; Guan et al., 2026).

H. Research gap

The literature separates prediction, graph modeling , RCA , self-healing , safety validation , and governance . What is missing is the integrated system: one loop in which predicted failures propagate over a dependency graph into blast-radius estimates, diagnoses are evidence-bound, plans are safety-validated, remediation is tiered, recovery is verified, failure triggers rollback, and every resolved event retrains the components. This is the system disclosed here.

III. PROBLEM FORMULATION

A. System model

Let the computing infrastructure be a layered dependency graph $\mathcal{G}_t = (\mathcal{V}, \mathcal{E}_t)$ with services $v \in \mathcal{V}$, dependency edges $e \in \mathcal{E}_t$, and per-node criticality weights c_v tied to commerce value (orders, revenue, essential-goods fulfillment). The graph is continuously updated from traces, configuration changes, and deployment events — the dynamic structure over which failures propagate.

B. Predictive failure intelligence

Given telemetry history $X_{1:t}$, the system predicts failure probability with lead time ℓ:

$$\hat{p}_v(t + \ell) = f_\theta(X_{1:t}, \mathcal{G}_t), \mathbb{E}[\text{lead}] \geq \ell^*,$$

where ℓ^* is the contract horizon — anchored to the demonstrated 42-minute average lead time of multimodal early warning (Wang, 2024) and the two-to-four-week horizon of neural architectures (Huang S. , 2024). Prediction quality is measured by recall, precision, and lead time under temporal splits, following the early-warning evaluation discipline (Zhou et al., 2021).

C. Blast-radius estimation

Predicted failures propagate over $\mathcal{G}_t$:

$$B(v,t) = \sum_{u \in reach(v)} \hat{p}_u(t+\ell) \cdot c_u \cdot w_{v \to u},$$

where $w_{v \to u}$ are dependency weights — the message-passing formalism validated in graph-native prediction (Liu et al., 2026; Shen et al., 2026) and physics-informed propagation (Petrova & Hughes1, 2025).

D. Root-cause inference

Diagnosis is evidence-bound generation:

$$\hat{r} = Gen(prompt(\mathcal{T}, Retr(\mathcal{T}, \mathcal{H}))), s.t. prov(\hat{r}) \subseteq (\mathcal{T} \cup \mathcal{H}),$$

with $\mathcal{T}$ the multimodal telemetry, $\mathcal{H}$ the incident corpus, and $prov(\hat{r})$ the provenance of every hypothesis — directly countering the hallucinated-reasoning-path failure mode (Cornacchia et al., 2025) and the reasoning-failure taxonomy whose presence predicts final correctness (Riddell et al., 2026).

E. Recovery planning and safety validation

The planner produces a remediation $a \in \mathcal{A}$ solving

$$a^\star = \arg\min_a \mathbb{E}[c_{err}(a) \cdot e(a) + c_{delay} \cdot \tau(a)] s.t. safety(a) \geq \gamma, a \in feasible(\mathcal{G}_t),$$

with safety constraints enforced by guardrails (Davenport, 2026) and governed policy constraints (Jingar, 2023), and plans validated by twin replay across control loops (Ivanov, 2026; Ivanov & Dolgui, 2021).

F. Verification and rollback

Execution is verified against recovery criteria; on verification failure within budget T_{max}, the system reverts deterministically:

$$a_t = \begin{cases} a, & \text{if } KPI(t+T) \geq \text{target} \\ rollback(a), & \text{otherwise} \end{cases}$$

— the corrective-control rule that makes autonomous action safe.

G. Optimization objective

The system minimizes expected downtime cost subject to governance:

$$\min \mathbb{E}[\theta \cdot \tau_{total} + c_{err} \cdot e] s.t. e \leq \epsilon, autonomy\text{-}tier(a), guardrails(a),$$

where θ is the downtime rate anchored to the documented cost of downtime at platform scale (Ahmed et al., 2023; Yan et al., 2025), ϵ the conformal-calibrated error bound (Cherian et al., 2024; Yadkori et al., 2024), and τ_{total} the sum of detection, diagnosis, and remediation times.

H. Metrics

Prediction: recall, precision, lead time (Wang, 2024), MTTD. Diagnosis: RCA accuracy (Sarda et al., 2024), time-to-root-cause (Jingar, 2023), reasoning-trace quality (Riddell et al., 2026). Remediation: MTTR (Sharma & Jain, 2026; Wang et al., 2025), autonomous success rate (Sharma & Jain, 2026), automation rate (Wang et al., 2025). Safety: rollback frequency and unsafe-action rate under governed deployment (Jingar, 2023). Governance: unsafe-output rate, drift-detection accuracy (Jingar, 2023). Evaluation: incident-owner human validation (Ahmed et al., 2023) and Ops-domain agreement (Liu et al., 2025).

IV. PROPOSED FRAMEWORK: AEGIS-INFRA

A. Overview

AEGIS-INFRA couples ten components across four layers: Perception and Prediction (F1 predictive failure intelligence; F2 service dependency graph; F3 blast-radius estimation), Diagnosis (F4 root-cause inference), Action (F5 recovery-plan generation; F6 safety validation; F7 autonomous remediation; F8 recovery verification; F9 rollback), and Learning (F10 continuous learning). Fig. 1 shows the architecture; Table I maps components to evaluation.

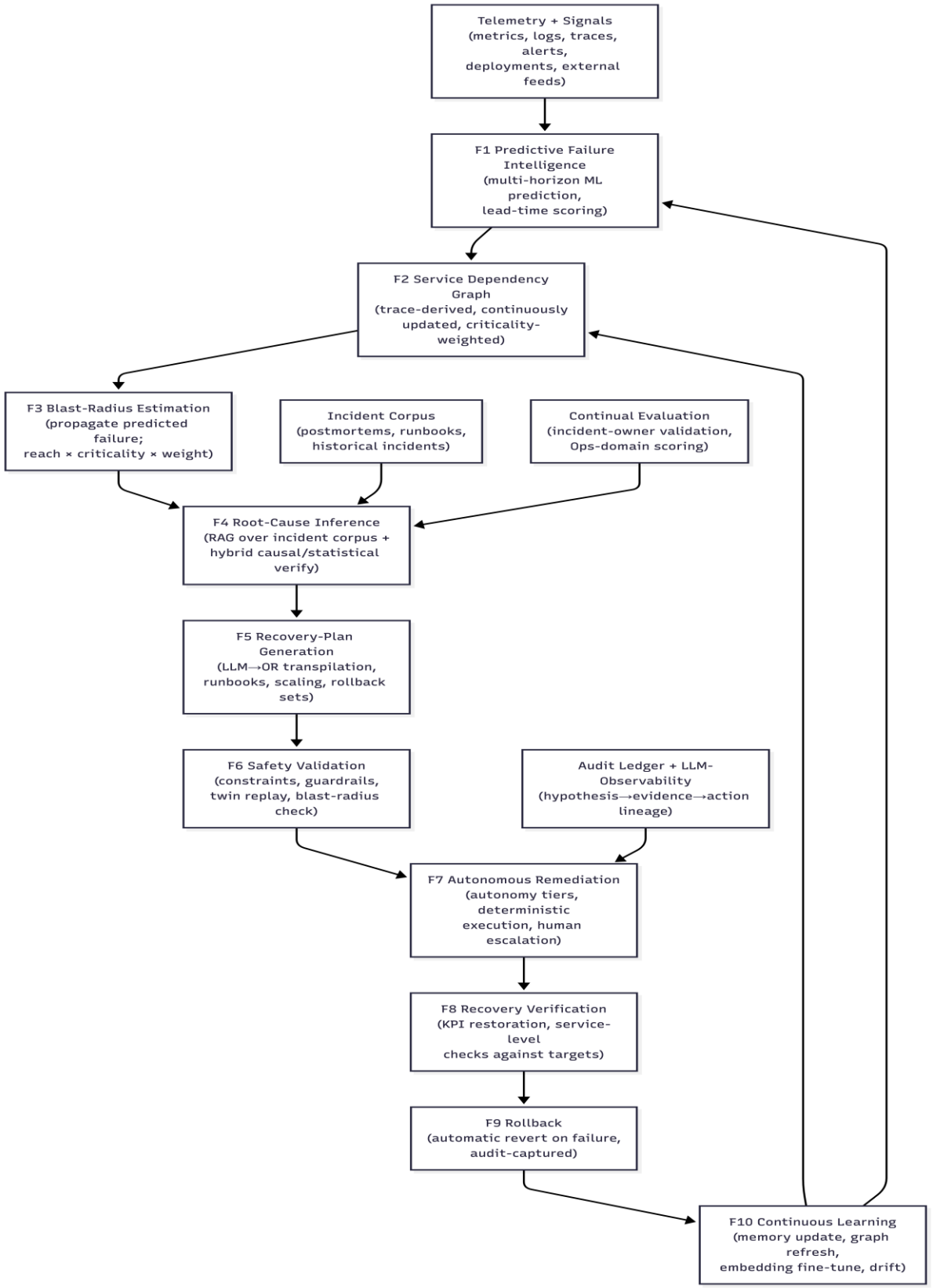

Fig. 1. AEGIS-INFRA architecture

Predictive failure intelligence (F1) forecasts failures with lead time over a continuously updated service dependency graph (F2); blast-radius estimation (F3) converts predictions into business-consequence terms; root-cause inference (F4) is retrieval-grounded and hybrid-verified; recovery plans (F5) pass safety validation (F6)

before tiered autonomous remediation (F7); recovery verification (F8) confirms restoration and triggers automatic rollback (F9) on failure; continuous learning (F10) refreshes the graph, the corpus, and the models .

B. F1 — predictive failure intelligence

The component fuses multimodal signals — metrics, logs, traces, alerts, deployment events, external feeds — into multi-horizon failure probabilities, following the multimodal architecture demonstrated at 92.1% recall and 93.4% precision with 42-minute lead time (Wang, 2024), extended by neural early-warning horizons of two-to-four weeks (Huang S. , 2024) and physics-informed regularization that cuts prediction error by 23% (Petrova & Hughes1, 2025). Prediction outputs carry calibrated uncertainty from the portfolio's uncertainty layer.

C. F2 — service dependency graph

The graph is derived from traces and configuration events, updated continuously as the substrate evolves, and weighted by criticality and dependency strength — the dynamic structure the GNN layer reasons over (Liu et al., 2026), with automated construction from operational and external data following the LLM-extraction pattern (Duttyal, 2025; Wang & Tsung, 2025) and the duality that makes graph traversal interpretable (Heus et al., 2025).

D. F3 — blast-radius estimation

Predicted failures propagate over the graph by message passing into consequence scores: reachable set × node criticality × dependency weight. The layer operationalizes resilience inference from topology and state trajectories (Shen et al., 2026) and gives the governance gate a business-denominated severity input.

E. F4 — root-cause inference

Diagnosis is retrieval-augmented over the incident corpus — the documented accuracy mechanism of ARCH (Sharma & Jain, 2026) — with provenance required for every hypothesis (Vadakkepati, 2025), hybrid verification fusing deterministic graph correlation, statistical scoring, and contextual LLM analysis (Veershetty, 2026), causal inference in the GALA pattern (Tian et al., 2025), and reasoning traces captured as first-class artifacts (Riddell et al., 2026).

F. F5 — recovery-plan generation

The planner synthesizes remediations — runbook actions, scaling, configuration changes, traffic shifting, and explicit rollback sets — transpiled into solver-executable form via the LLM–OR pathways (Li et al., 2025; Wang & Li, 2025), with counterfactual what-if/why-not analysis (Yu et al., 2025), semantic/calculation decoupling in the inventory-cost-reducing pattern (Duan et al., 2025), and OR-augmented formulation as the default interface (Baek et al., 2026).

G. F6 — safety validation

Every plan is validated before execution: operational constraints and provenance templates (Vadakkepati, 2025), deterministic guardrails (Davenport, 2026), governed policy constraints with drift awareness (Jingar, 2023), and digital-twin replay across the five control loops (Ivanov, 2026; Ivanov & Dolgui, 2021; Liu et al., 2026). High-blast-radius plans are routed to human approval by construction (KALAFATOGLU, 2025, 2026).

H. F7 — autonomous remediation

Execution follows the governance-state pattern — GO, HOLD, and NO-GO decision states with human-final authority (KALAFATOGLU, 2025, 2026) — with deterministic runbook execution for low-risk actions (Sharma & Jain, 2026; Wang et al., 2025), human escalation for high-blast-radius or low-confidence plans, and full lineage capture in the audit ledger (Syed et al., 2025) with LLM-observability over the agent stack itself (Ganesan, 2024).

I. F8 — recovery verification

Post-execution, the system verifies restoration against KPI and service-level targets, following the recovery-monitoring pattern of self-healing stacks (Singh, 2025) and twin-based verification (Tiwari et al., 2023), with MTTD, MTTR, and autonomous success rate tracked continuously (Sharma & Jain, 2026; Wang et al., 2025).

J. F9 — rollback

On verification failure within the time budget, the system reverts deterministically to the last-known-good state, with the rollback itself audited and learned from — the corrective-control loop that makes autonomous action safe.

K. F10 — continuous learning

Every resolved event retrains the system: resolved-incident memory is indexed for similarity retrieval in the memory-agent pattern (Yoshizato et al., 2026); embeddings and anomaly models are fine-tuned; the dependency graph is refreshed with confirmed relations; drift detectors trigger recalibration; and continual benchmarking scores each iteration (Protogeros & Vanbever, 2025). Algorithm 1 summarizes the method.

Algorithm 1: The Resilience Loop (Predict → Diagnose → Validate → Remediate → Verify → Rollback → Learn)

```
Input: telemetry T_t, dependency graph G_t, incident corpus H,
guardrails Gv, error bound ε, twin Θ
Output: governed actions a, updated system state
# PERCEPTION (F1–F3)
1  p̂_v ← predict_failure(T_t, G_t, ℓ)          # 92.1%/93.4%/42-min standard
```

```
2   G_t ← update_graph(G_t, traces, deployments)   #
3   B(v,t) ← blast_radius(p̂, G_t, c, w)            # Section III-C
4   if max p̂ < threshold: monitor; continue
# DIAGNOSIS (F4)
5   E ← retrieve(H, T_t); r̂ ← Gen(prompt(T_t, E))
6   if prov(r̂) ⊄ (T_t ∪ E): regenerate or escalate  #
7   r̂ ← hybrid_verify(r̂, T_t); capture reasoning trace  #
# ACTION (F5–F9)
8   P ← plan(r̂, B, G_t); P ← validate(P, Gv, Θ)    # safety + twin replay
9   if blast_radius high or confidence low: route to human   #
10  else: execute(a) deterministically via governance gate   # GO/HOLD/NO-GO
11  if verify_restoration fails within T_max: rollback(a) # corrective control
# LEARNING (F10)
12  memory ← index_event(r̂, a, outcome); refresh G_t, embeddings
13  monitor drift; recalibrate if drifted; benchmark(B)   #
```

L. Implementation

Prediction: multimodal transformers + LSTM anomaly + physics-informed GNN (Petrova & Hughes1, 2025; Wang, 2024); graph: trace-derived property graph with LLM extraction (Duttyal, 2025; Wang & Tsung, 2025); blast radius: message-passing propagation with criticality weighting (Shen et al., 2026); RCA: RAG-conditioned generation with hybrid verification (Tian et al., 2025; Veershetty, 2026); planning: LLM-to-OR transpilation (Wang & Li, 2025); validation: constraint engine, guardrail layer (Davenport, 2026), twin replay (Ivanov & Dolgui, 2021); execution: runbook engine with GO/HOLD/NO-GO gating (KALAFATOGLU, 2025, 2026); verification/rollback: KPI monitor + deterministic revert (Singh, 2025); learning: memory indexing (Yoshizato et al., 2026), drift detectors (Jingar, 2023), continual evaluation (Liu et al., 2025; Protogeros & Vanbever, 2025).

V. EXPERIMENTAL SETUP

A. Environments and data

Evaluation is anchored to the documented settings of the component literature: the Microsoft corpus of 44,340 incidents from 1,759 services with incident-owner human validation (Ahmed et al., 2023); the LMTA alert-augmented datasets with up to 97% RCA accuracy (Sarda et al., 2024); the GAIA and OpenRCA case studies with 48,000 simulated failure scenarios under ReAct and Plan-and-Execute workflows (Riddell et al., 2026); the microservice testbeds of the observability framework (Wang et al., 2025) and the ARCH evaluation environment (Sharma & Jain, 2026); the AIOpsLab deployment benchmark (Zhong et al., 2026); the automotive network (Kosasih & Brintrup, 2021) and automotive/energy (Kosasih et al., 2022) graph testbeds; the synthetic resilience benchmarks of the hypergraph study (Shen et al., 2026); the physics-informed GNN networks (Petrova & Hughes1, 2025); the multimodal early-warning datasets (Wang, 2024); and the OpsEval (Liu et al., 2025) and SupChain-Bench (Guan et al., 2026) evaluation suites.

B. Baselines

Rule-based monitoring and manual RCA (the MTTR baseline); reactive remediation without prediction (F1 ablation); non-graph failure prediction (F2/F3 ablation); ungrounded single-LLM diagnosis (the hallucinated-reasoning baseline); plan-without-safety-validation (F6 ablation); ungoverned full automation versus gated autonomy (F7 ablation); execution-without-verification-rollback (F8/F9 ablation); and static systems without learning (F10 ablation).

C. Scenarios

(S1) checkout-service latency incident with multi-hop fault propagation, predicted at lead time; (S2) payment-gateway timeout cascade; (S3) inventory-service data skew surfacing as degraded APIs; (S4) configuration drift requiring rollback; (S5) predicted capacity exhaustion with preemptive scaling; (S6) incident with misleading log signals testing evidence provenance; (S7) adversarial telemetry testing reasoning-trace gating; (S8) compound incident under chaotic multi-modal degradation; (S9) high-blast-radius action (payment cutoff) testing governance-gate escalation; (S10) recurring incident class testing continuous learning and recurrence reduction.

D. Protocol

Temporal splits prevent leakage; prediction quality by recall, precision, and lead time (Wang, 2024); diagnosis by RCA accuracy (Sarda et al., 2024), time-to-root-cause (Jingar, 2023), and reasoning-trace quality under the failure-taxonomy rubric (Riddell et al., 2026); remediation by MTTR (Sharma & Jain, 2026; Wang et al., 2025), autonomous success rate (Sharma & Jain, 2026), and automation rate (Wang et al., 2025); safety by rollback frequency and unsafe-action rate (Jingar, 2023); governance by unsafe-output and drift rates (Jingar, 2023); and

overall quality by incident-owner human validation (Ahmed et al., 2023) and Ops-domain agreement (Liu et al., 2025). All quantitative entries below are literature-reported values from the cited studies; entries marked [illustrative] are placeholders for AEGIS-INFRA's own experiments.

VI. RESULTS AND ANALYSIS

A. Task–module mapping

Table I maps components to evaluation.

TABLE I: AEGIS-INFRA COMPONENTS, FUNCTIONS, AND EVALUATION

Component	Layer	Function	Evaluation
F1 Predictive failure intelligence	Perception	Multi-horizon failure forecasting	Recall, precision, lead time (Wang, 2024)
F2 Service dependency graph	Perception	Dependency and criticality structure	Construction accuracy, freshness (Duttyal, 2025)
F3 Blast-radius estimation	Perception	Consequence-weighted propagation	Calibration vs realized incidents (Shen et al., 2026)
F4 Root-cause inference	Diagnosis	Evidence-bound RCA + verification	RCA accuracy, time-to-root-cause (Jingar, 2023; Sarda et al., 2024)
F5 Recovery-plan generation	Action	Plan synthesis (LLM→OR)	Recovery cost, time-to-recovery (Wang & Li, 2025)
F6 Safety validation	Action	Constraints, guardrails, twin replay	Feasibility rate, safety compliance (Davenport, 2026; Ivanov & Dolgui, 2021)
F7 Autonomous remediation	Action	Tiered execution, escalation	MTTR, autonomous success rate (Sharma & Jain, 2026; Wang et al., 2025)
F8 Recovery verification	Action	Restoration confirmation	KPI restoration, MTTR (Sharma & Jain, 2026; Wang et al., 2025)
F9 Rollback	Action	Deterministic revert	Rollback frequency, unsafe-action rate (Jingar, 2023)
F10 Continuous learning	Learning	Memory, refresh, drift, retraining	Drift accuracy, continual scores (Protogeros & Vanbever, 2025)

B. Reported performance envelope

Table II consolidates reported results across the predictive-failure-intelligence and self-healing literature.

TABLE II: REPORTED PERFORMANCE IN PREDICTIVE FAILURE INTELLIGENCE AND AUTONOMOUS SELF-HEALING

Component	Method	Reported result
F1 Prediction	Multimodal deep-learning EWS	Recall 92.1%, precision 93.4%, lead 42 min
F1 Prediction	Neural early warning	2–4 weeks advance warning
F1 Prediction	ML-driven EWS synthesis	Up to 41% forecast-accuracy gain, 15% cost reduction, 35% disruption reduction
F1/F2 Propagation	Physics-informed GNN	Prediction error −23%; interpretable
F2/F3 Graph	Hypergraph resilience inference	Outperforms MLP, GNN variants, ResInf
F2/F3 Graph	GNN link prediction	Outperforms existing algorithms (automotive)
F2/F3 Graph	TGAT + LLM grounding	AUC 0.761, AP 0.344, recall 0.504; 99.6% explanation consistency
F4 RCA	Microsoft RCA study	44,340 incidents, 1,759 services; incident-owner human validation
F4 RCA	LMTA alert-augmented RCA	Up to 97% RCA accuracy
F4 RCA	LocaleXpert	Multi-modal expert agents; outperforms baselines
F4 RCA	GALA graph-augmented workflow	Up to +42.22% accuracy over state-of-the-art
F4 RCA	Failure-mode evaluation	Hallucinated reasoning paths, inefficient context dominate failures
F4 RCA	Isolated reasoning evaluation	6 LLMs; 48,000 scenarios; 16 reasoning failures; failures predict correctness
F4 RCA	Incident-intelligence agents	Time-to-root-cause −30%; resolution accuracy +22%
F1–F9 Loop	Domain-adapted LLM observability	MTTR −84.2%; F1 95%; 91% automation
F1–F9 Loop	ARCH (LLM + RAG)	MTTR up to −82%; ASR 89.5%
F1–F9 Loop	Autonomous multi-agent response	Response time −87% (72 → 9.4 h); cost −34%; MAPE 23.4% → 2.1%; ROI 312%
F6 Safety	Digital-twin stress testing	Five control loops; Ford/Microsoft deployments
F9 Rollback	Verification-and-rollback rule	Deterministic revert on KPI failure
F10 Learning	Governed LLM optimization	Unsafe outputs −45%; drift detection > 92%
F10 Learning	Memory-retrieval agents	Outperforms benchmark methods
Evaluation	OpsEval	7,334 MCQ + 1,736 QA; 24 LLMs; 0.9185 human agreement

C. Worked calculations

1) The downtime arithmetic (headline). At the scale of a major commerce platform, a single hour of downtime on peak shopping days is documented at $100 million (Ahmed et al., 2023), and supply-chain-level disruptions average $184 million in cost per event with response windows of 3–5 days (Yan et al., 2025). Every minute of avoided downtime is therefore denominated in this currency — the exposure that makes predictive lead time, not reaction speed, the first-class objective.

2) The prediction lead-time multiplier. The multimodal engine's 42-minute average lead time (Wang, 2024) operates against a multi-hour manual response window. Against a 72-hour baseline for supply-chain-level

response (Yan et al., 2025), the lead-time-to-response ratio is 4,320/42 ≈ 103×; when the neural horizon extends to two-to-four weeks (Huang S. , 2024), the prediction layer converts disruption from a surprise into a scheduled event. Within the infrastructure loop, a 42-minute lead with 3.83-minute agentic analysis (AlMahri et al., 2026) fits the entire predict–diagnose–act sequence inside the first hour.

3) The MTTR dividend. A domain-adapted LLM over OpenTelemetry telemetry reduces MTTR by 84.2% versus rule-based methods with 95% F1 anomaly detection and 91% automation of common incidents (Wang et al., 2025); ARCH independently reports up to 82% MTTR reduction with an 89.5% autonomous success rate (Sharma & Jain, 2026). The two framework-level studies triangulate: the residual MTTR after full-loop automation is roughly one-sixth to one-fifth of the rule-based baseline (100 − 84.2 = 15.8% and 100 − 82 = 18%, respectively) — moving incidents from the multi-hour band into the tens-of-minutes band.

4) The autonomy residual. At 91% automation (Wang et al., 2025) and an 89.5% autonomous success rate (Sharma & Jain, 2026), the human-handled residual is approximately 9–10.5%. In a platform experiencing 1,000 incidents per year, this bounds human involvement at roughly 90–105 events — a deliberate allocation governed by the GO/HOLD/NO-GO decision states with human-final authority (KALAFATOGLU, 2025, 2026) rather than an emergent load, each arriving with a full evidence and audit package.

5) The diagnosis-bottleneck arithmetic. Engineers spend an inordinate amount of time formulating hypotheses (Chen et al., 2024); the evidence base quantifies the heads of improvement — up to 97% RCA accuracy with alert-augmented multimodal pipelines (Sarda et al., 2024), up to 42.22% accuracy gains from graph-augmented causal reasoning (Tian et al., 2025), and a 30% time-to-root-cause reduction with a 22% resolution-accuracy improvement from role-specialized agents (Jingar, 2023). Because MTTR is dominated by diagnosis time, each diagnostic accuracy gain compresses time-to-root-cause proportionally — the reason the framework spends its complexity budget on evidence, retrieval, and verification rather than on model scale.

6) The prediction-accuracy lever. Physics-informed regularization reduces disruption-prediction error by 23% (Petrova & Hughes1, 2025); at the documented exposure of $100M per peak-hour and $184M per disruption event (Ahmed et al., 2023; Yan et al., 2025), a prediction-error reduction that shifts even a fraction of incidents from reactive to preemptive recovery is worth seven figures annually — before the 41%/15%/35% early-warning gains (forecast accuracy, cost, disruption reduction) documented for ML adoption (Kasali et al., 2025) are counted.

7) Governance composition. Governed LLM optimization reduces unsafe decision outputs by 45% with drift detection above 92% (Jingar, 2023); grounding generation in retrieved evidence reduces hallucination exposure from 8.3% to 1.5% in enterprise RAG deployments (Kasarapu, 2026). Composed, the expected unsafe-decision exposure of an ungoverned, ungrounded baseline falls by roughly 0.55 × 0.18 ≈ 0.10 — approximately a 90% aggregate reduction — before uncertainty gating and human authority are applied. This is the arithmetic that licenses the GO state at the autonomy gate.

8) The verification-and-rollback contract. Rollback converts autonomous action from a bet into a closed-loop control: if autonomous remediation reduces MTTR by 84.2% (Wang et al., 2025) and the rollback rule contains the residual failure cases deterministically, the effective distribution of recovery time is bounded rather than unbounded — the property that separates governable autonomy from reckless automation. Rollback frequency and unsafe-action rate are tracked as standing safety metrics under governed deployment (Jingar, 2023).

D. MTTR Reduction

Fig. 2 plots the reported MTTR reductions.

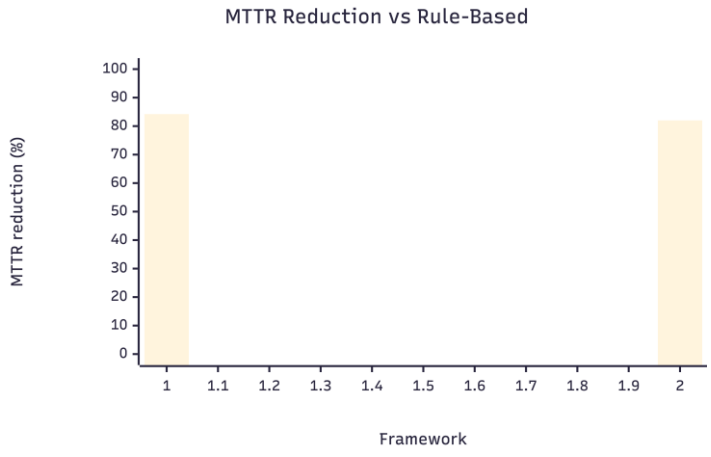

Fig. 2. Reported MTTR reductions: 84.2% for the domain-adapted LLM observability framework and up to 82% for the ARCH retrieval-augmented self-healing framework.

E. Prediction lead time vs response window

Fig. 3 compares the reported prediction lead time against the traditional response window.

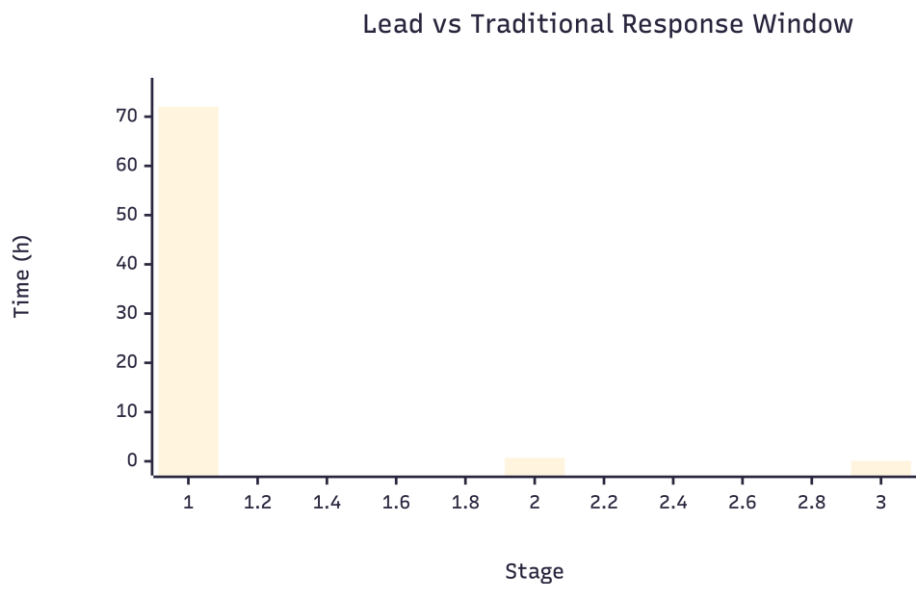

Fig. 3. Reported times: 72 hours for the traditional response window; 0.7 hours (42 min) for the multimodal early-warning lead time; 0.064 hours (3.83 min) for agentic analysis. Log-scale recommended for the final figure.

F. Illustrative blast-radius trajectory

Fig. 4 shows the [illustrative] blast-radius accumulation across dependency hops for a predicted failure, consistent with the graph-propagation evidence (Liu et al., 2026; Shen et al., 2026) and the physics-informed prediction-accuracy results (Petrova & Hughes1, 2025).

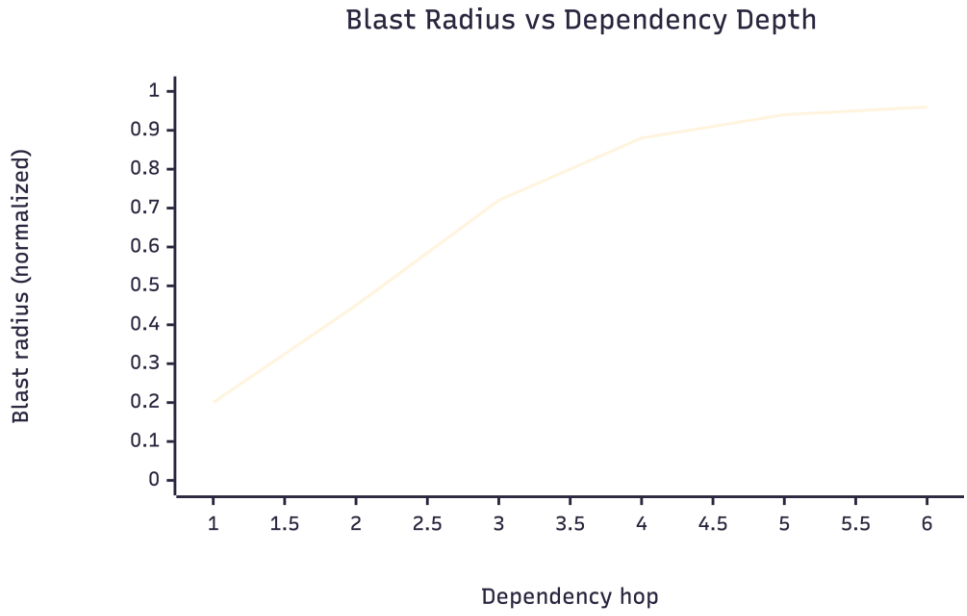

Fig. 4. [Illustrative] normalized blast radius accumulating across dependency hops for a predicted failure, showing the multiplicative reach that motivates graph-native consequence estimation.

G. Ablation logic

Removing prediction (F1) reverts the system to reactive remediation — forfeiting the lead-time multiplier (Wang J. , 2024; Huang S. , 2024); removing the dependency graph (F2) and blast radius (F3) forfeits multi-hop

consequence awareness, the propagation evidence of GNN and hypergraph methods (Liu et al., 2026; Shen et al., 2026); removing retrieval grounding (F4) reverts diagnosis toward the hallucinated-reasoning failure mode (Cornacchia et al., 2025); removing safety validation (F6) exposes unvalidated plans, forfeiting the guardrail and twin-replay protections (Davenport, 2026; Ivanov & Dolgui, 2021); removing verification and rollback (F8/F9) forfeits the corrective-control guarantee of the rollback rule; and removing learning (F10) forfeits memory-retrieval gains (Yoshizato et al., 2026) and drift protection (Jingar, 2023).

VII. DISCUSSION

A. Prediction is the force-multiplier

The 42-minute lead time (Wang, 2024), two-to-four-week neural horizon (Huang S. , 2024), 23% physics-informed error reduction (Petrova & Hughes1, 2025), and 41%/15%/35% early-warning gains (Kasali et al., 2025) all convert into recovery latitude measured in dollars per minute at the documented downtime rates of major platforms (Ahmed et al., 2023; Yan et al., 2025). AEGIS-INFRA treats lead time as a first-class contract — every predicted failure carries its horizon, and the governance gate is calibrated on it.

B. Diagnosis is the binding constraint

Execution is easy once the cause is known; finding it is where on-call time disappears (Chen et al., 2024). The evidence is unambiguous: evidence-bound, multimodal, causally verified diagnosis reaches up to 97% RCA accuracy (Sarda et al., 2024), gains up to 42.22% from causal graph reasoning (Tian et al., 2025), and the failure-taxonomy evidence — 16 reasoning failures that predict correctness across 48,000 scenarios (Riddell et al., 2026) — makes reasoning-trace monitoring a runtime safety mechanism, not an evaluation luxury.

C. The graph is the substrate of everything

Prediction (Liu et al., 2026), blast radius (Shen et al., 2026), diagnosis (Tian et al., 2025), and explanation (Xue et al., 2026) all operate over the same dependency graph, constructed automatically (Duttyal, 2025; Wang & Tsung, 2025) and kept fresh by continuous learning. Graph quality is therefore the system's binding constraint — the reason construction, ontology, and freshness are engineered rather than assumed.

D. Verification-and-rollback is the safety contract

The framework's defining governance element is the closed loop: verify restoration (F8), roll back if unsuccessful (F9), and track rollback frequency and unsafe-action rate as standing metrics (Jingar, 2023). Combined with the GO/HOLD/NO-GO gate and human-final authority (KALAFATOGLU, 2025, 2026), guardrails (Davenport, 2026), and governed policy constraints (Jingar, 2023), rollback converts autonomous remediation from a bet into bounded, auditable control.

E. Autonomy is tiered, not absolute

At 91% automation (Wang et al., 2025) and 89.5% autonomous success (Sharma & Jain, 2026), the operating point is calibrated delegation: routine incidents run autonomously; the residual routes to humans with full evidence packages under the transparency-and-trust posture of self-healing deployments (Singh, 2025). The strong human–AI decision-quality evidence — human–AI teams outperform either alone (Baek et al., 2026), human-specified optimization is statistically optimal (Wu, 2026) — shows the tier boundary is an accuracy mechanism, not a cautionary concession.

F. Integration with the portfolio

AEGIS-INFRA is the infrastructure companion of the program's systems-and-methods disclosures: it applies the supply-chain loop's predict–mitigate–recover discipline to the computing substrate itself, reusing the program's early-warning engines (Wang J. , 2024; Huang S. , 2024), the trust boundary (Kang et al., 2025; Yadkori et al., 2024), the retrieval grounding (Sharma & Jain, 2026), the incident-response loop (Ahmed et al., 2023; Wang et al., 2025), and the authority frameworks (KALAFATOGLU, 2025, 2026). The platform and its substrate are now governed by the same architecture.

G. Limitations

Reported results span heterogeneous platforms, datasets, and industries rather than a single end-to-end evaluation of the integrated loop; the strongest MTTR figures come from microservice testbeds (Sharma & Jain, 2026; Wang et al., 2025); blast-radius and graph evidence is strongest on automotive, energy, and maritime testbeds (Kosasih et al., 2022; Kosasih & Brintrup, 2021; Shen et al., 2026); reasoning-failure taxonomies are emerging and not yet standardized; and live human–AI escalation and rollback dynamics remain under-studied. These define the agenda below.

VIII. CONCLUSION AND FUTURE WORK

This paper disclosed a system and associated methods for predictive failure intelligence and autonomous self-healing of distributed computing infrastructure supporting critical supply chains: ten components across four layers — predictive failure intelligence, a service dependency graph, blast-radius estimation, root-cause inference, recovery-plan generation, safety validation, autonomous remediation, recovery verification, rollback, and continuous learning. The system consolidates the reported evidence envelope — $100M-per-hour peak exposure

on major shopping days (Ahmed et al., 2023) and $184M average cost per disruption event with 3–5-day response windows (Yan et al., 2025); 92.1% recall / 93.4% precision with 42-minute lead time (Wang, 2024) and two-to-four-week neural horizons (Huang S. , 2024); physics-informed prediction-error reduction of 23% (Petrova & Hughes1, 2025); hypergraph and GNN propagation outperforming baselines (Kosasih & Brintrup, 2021; Shen et al., 2026); RCA validation across 44,340 incidents from 1,759 services (Ahmed et al., 2023); up to 97% alert-augmented RCA accuracy (Sarda et al., 2024); up to 42.22% causal-reasoning gains (Tian et al., 2025); a 16-failure reasoning taxonomy from 48,000 scenarios (Riddell et al., 2026); MTTR reductions of 84.2% and up to 82% with 95% F1 detection, 91% automation, and 89.5% autonomous success (Sharma & Jain, 2026; Wang et al., 2025); a 30% time-to-root-cause reduction and 22% accuracy improvement from incident-intelligence agents (Jingar, 2023); and governed deployment cutting unsafe outputs by 45% with drift detection above 92% (Jingar, 2023) — and argued that predictive failure intelligence, not faster reaction, is the binding constraint, and that grounded, validated, verified-or-rolled-back autonomy is the difference between infrastructure that fails and infrastructure that learns.

Future work will (i) implement AEGIS-INFRA end-to-end on a live commerce platform with unified evaluation under the Section V protocol, including lead-time, MTTR, rollback-frequency, and recurrence measurement; (ii) harden prediction with temporal and heterogeneous GNNs for dynamic graph evolution (Liu et al., 2026), uncertainty-aware traversal, and physics-informed regularization (Petrova & Hughes1, 2025); (iii) expand the reasoning-failure taxonomy through controlled multi-modality studies, using reasoning-trace quality as a runtime safety metric (Riddell et al., 2026); (iv) calibrate governance-gate thresholds per incident class under drift with conformal recalibration (Cherian et al., 2024; Yadkori et al., 2024), unsafe-output rate, and drift detection (Jingar, 2023); and (v) standardize evaluation so that each loop iteration is scored on incident-owner validation (Ahmed et al., 2023) and the Ops-domain standard of 0.9185 human agreement (Liu et al., 2025) — converting predictive failure intelligence from an engineering vision into a measured, auditable, deployable capability.

References

[1]. Aboutorab, H., Hussain, O. K., Saberi, M., Hussain, F. K., & Prior, D. D. (2024). Adaptive identification of supply chain disruptions through reinforcement learning. Expert Systems with Applications, 248, 123477. https://doi.org/10.1016/j.eswa.2024.123477

[2]. Ahmed, T., Ghosh, S., Bansal, C., Zimmermann, T., Zhang, X., & Rajmohan, S. (2023). Recommending Root-Cause and Mitigation Steps for Cloud Incidents using Large Language Models. 1737–1749. https://doi.org/10.1109/icse48619.2023.00149

[3]. Akhtar, P., Ghouri, A. M., Khan, H. U. R., Haq, M. A., Awan, U., Zahoor, N., Khan, Z., & Ashraf, A. (2022). Detecting fake news and disinformation using artificial intelligence and machine learning to avoid supply chain disruptions. Annals of Operations Research, 327(2), 633–657. https://doi.org/10.1007/s10479-022-05015-5

[4]. Akubilla, J., Somoye, O. I., Abiodun, F., & Serifat, O. A. (2025). The Role of Explainable AI in Enhancing Trust and Transparency in Supply Chain Risk Mitigation. International Journal of Multidisciplinary Research and Growth Evaluation, 6(3), 367–377. https://doi.org/10.54660/.ijmrge.2025.6.3.367-377

[5]. AlMahri, S., Xu, L., & Brintrup, A. (2026). Automating Supply Chain Disruption Monitoring via an Agentic AI Approach. ArXiv.Org. http://arxiv.org/abs/2601.09680

[6]. Baek, J., Fu, Y., Ma, W., & Peng, T. (2026). AI Agents for Inventory Control: Human-LLM-OR Complementarity. In arXiv (Cornell University). Cornell University. https://doi.org/10.48550/arxiv.2602.12631

[7]. Balashankar, A., Subramanian, L., & Fraiberger, S. P. (2023). Predicting food crises using news streams. Science Advances, 9(9). https://doi.org/10.1126/sciadv.abm3449

[8]. Baryannis, G., Dani, S., & Antoniou, G. (2019). Predicting supply chain risks using machine learning: The trade-off between performance and interpretability. Future Generation Computer Systems, 101, 993–1004. https://doi.org/10.1016/j.future.2019.07.059

[9]. Brockmann, N., Kosasih, E. E., & Brintrup, A. (2022). Supply Chain Link Prediction on Uncertain Knowledge Graph. ACM SIGKDD Explorations Newsletter, 24(2), 124–130. https://doi.org/10.1145/3575637.3575655

[10]. Campos, M., Farinhas, A., Zerva, C., Figueiredo, M. A. T., & Martins, A. F. T. (2024). Conformal Prediction for Natural Language Processing: A Survey. Transactions of the Association for Computational Linguistics, 12, 1497–1516. https://doi.org/10.1162/tacl_a_00715

[11]. Chakraborty, N., Ornik, M., & Driggs-Campbell, K. (2025). Hallucination Detection in Foundation Models for Decision-Making: A Flexible Definition and Review of the State of the Art. In arXiv (Cornell University) (Vol. 57, Issue 7, pp. 1–35). Cornell University. https://doi.org/10.1145/3716846

[12]. Chen, X., Wanigarathna, Y. R., Chattopadhyay, A., Tarkoma, S., & Rao, A. (2025). A Pedagogical Approach for Evaluating AIOps. 40–49. https://doi.org/10.1109/aimlsystems67835.2025.11330228

[13]. Chen, Y., Xie, H., Ma, M., Kang, Y., Gao, X., Shi, L., Cao, Y., Gao, X. Y., Fan, H., Wen, M., Zeng, J., Ghosh, S., Zhang, X., Zhang, C., Lin, Q., Rajmohan, S., Zhang, D., & Xu, T. (2024). Automatic Root Cause Analysis via Large Language Models for Cloud Incidents. 674–688. https://doi.org/10.1145/3627703.3629553

[14]. Cherian, J. J., Gibbs, I., & Candès, E. J. (2024). Large language model validity via enhanced conformal prediction methods. In arXiv (Cornell University). Cornell University. https://doi.org/10.48550/arxiv.2406.09714

[15]. Cimino, A., Longo, F., Mirabelli, G., & Solina, V. (2024). A cyclic and holistic methodology to exploit the Supply Chain Digital Twin concept towards a more resilient and sustainable future. Cleaner Logistics and Supply Chain, 11, 100154. https://doi.org/10.1016/j.clscn.2024.100154

[16]. Cornacchia, A., Alabdulaal, I., Saghier, I., Mirdad, A., Fayoumi, O., & Canini, M. (2025). Between Promise and Pain: The Reality of Automating Failure Analysis in Microservices with LLMs. 155–167. https://doi.org/10.1145/3725783.3764388

[17]. Cui, Y., Zeng, Y., Yan, J., Lin, K. L., Ji, K. Y., Zeng, J., Zhang, S., Luo, X., Su, B., Shen, C., & Yu, J. (2026). CSCBench: A PVC Diagnostic Benchmark for Commodity Supply Chain Reasoning. In arXiv (Cornell University). Cornell University. https://doi.org/10.48550/arxiv.2601.01825

[18]. Davenport. (2026). GUARDRAIL-CENTRIC FINE-TUNING FOR DETERMINISTIC DECISION SYSTEMS. Zenodo (CERN European Organization for Nuclear Research). https://doi.org/10.5281/zenodo.18305825

[19]. Dehan, M. F. Z., Anzum, K. Md. T., Disha, J. F., Masum, MD. M. R., & Mahmud, I. (2025). A Systematic Review on the Applications of How Explainable Artificial Intelligence can be Applied in the Supply Chain Management. https://doi.org/10.46254/au04.20250068

[20]. Duan, Y., Hu, Y., & Jiang, J. (2025). Ask, Clarify, Optimize: Human-LLM Agent Collaboration for Smarter Inventory Control. In arXiv (Cornell University). Cornell University. https://doi.org/10.48550/arxiv.2601.00121

[21]. Duttyal, S. (2025). Knowledge Graph-Enhanced LLMs for Supply Chain Intelligence: Automating Network Visibility Through Public Data Mining. Technix International Journal for Engineering Research, 12(6). https://doi.org/10.56975/tijer.v12i6.158410

[22]. Farquhar, S., Kossen, J., Kuhn, L., & Gal, Y. (2024). Detecting hallucinations in large language models using semantic entropy. Nature, 630(8017), 625–630. https://doi.org/10.1038/s41586-024-07421-0

[23]. Ganesan, P. (2024). LLM-Powered Observability Enhancing Monitoring and Diagnostics. Journal of Artificial Intelligence Machine Learning and Data Science, 2(2), 1329–1336. https://doi.org/10.51219/jaimld/premkumar-ganesan/304

[24]. Ge, Z., & Brintrup, A. (2024). Enhancing Supply Chain Visibility with Generative AI: An Exploratory Case Study on Relationship Prediction in Knowledge Graphs. In arXiv (Cornell University). Cornell University. https://doi.org/10.48550/arxiv.2412.03390

[25]. Gelastopoulos, G., & Keramydas, C. (2025). A systematic review of text mining analytics for supply chain risk management using online data. Supply Chain Analytics, 12, 100167. https://doi.org/10.1016/j.sca.2025.100167

[26]. Ghiani, G., Manni, E., & Zacchino, S. (2025). Improving Adaptability in Optimization-Based Decision Support Systems Through Large Language Models. IEEE Access, 13, 149777–149788. https://doi.org/10.1109/access.2025.3601518

[27]. Guan, S., Liu, Y., & Cao, L. (2026). SupChain-Bench: Benchmarking Large Language Models for Real-World Supply Chain Management. arXiv (Cornell University). http://arxiv.org/abs/2602.07342

[28]. Gumaan, E. (2025). Theoretical Foundations and Mitigation of Hallucination in Large Language Models. In ArXiv.org. https://doi.org/10.48550/arxiv.2507.22915

[29]. Ham, S., Lee, K. B., Park, S., Kang, J., & Hwang, H. (2025). Design and Implementation of a Semiconductor Supply Chain Knowledge Graph for Developing a GraphRAG-Based Question Answering Model. ScholarSpace (University of Hawaii at Manoa). https://hdl.handle.net/10125/111976

[30]. Heus, E., Bookstaber, R., & Sharma, D. (2025). Exploring Network-Knowledge Graph Duality: A Case Study in Agentic Supply Chain Risk Analysis. In ArXiv.org. https://doi.org/10.48550/arxiv.2510.01115

[31]. Huang , Sichong. (2025). AI-Driven Early Warning Systems for Supply Chain Risk Detection: A Machine Learning Approach. Academic Journal of Computing & Information Science, 8(9). https://doi.org/10.25236/ajcis.2025.080913

[32]. Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., Qin, B., & Liu, T. (2024). A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions. ACM Transactions on Information Systems, 43(2), 1–55. https://doi.org/10.1145/3703155

[33]. Iruku, V. M. (2025). Multi-dimensional XAI Framework Revealing Critical Supply Chain Vulnerability Drivers. World Journal of Advanced Engineering Technology and Sciences, 15(3), 2141–2152. https://doi.org/10.30574/wjaets.2025.15.3.1154

[34]. Ivanov, D. (2026). A generalized concept of a decision-support system for responses to tariff-induced uncertainty and disruptions. Annals of Operations Research. https://doi.org/10.1007/s10479-026-07303-w

[35]. Ivanov, D., & Dolgui, A. (2021). Stress testing supply chains and creating viable ecosystems. Operations Management Research, 15, 475–486. https://doi.org/10.1007/s12063-021-00194-z

[36]. Jingar, N. K. (2023a). Ensuring Safety, Accountability, and Drift Resistance in LLM-Based Supply Chain Optimization. International Journal of Scientific Research in Science Engineering and Technology, 472. https://doi.org/10.32628/ijsrset2310372

[37]. Jingar, N. K. (2023b). Automated Incident Intelligence In Supply Chains Using Agentic AI And Root Cause Reasoning. Zenodo (CERN European Organization for Nuclear Research). https://doi.org/10.5281/zenodo.18628070

[38]. KALAFATOGLU, Y. (2025). Human-Final Decision Authority in Artificial Intelligence: A Deterministic and Auditable Governance Architecture. Zenodo (CERN European Organization for Nuclear Research). https://doi.org/10.5281/zenodo.18460984

[39]. KALAFATOGLU, Y. (2026). A Deterministic Decision Authority Framework for Governance-Grade AI Systems. In Zenodo (CERN European Organization for Nuclear Research). European Organization for Nuclear Research. https://doi.org/10.5281/zenodo.18743566

[40]. Kang, S., Bakman, Y. F., Yaldiz, D. N., Buyukates, B., & Avestimehr, S. (2025). Uncertainty Quantification for Hallucination Detection in Large Language Models: Foundations, Methodology, and Future Directions. In ArXiv.org. https://doi.org/10.48550/arxiv.2510.12040

[41]. Kasali, K., Olawore, A. O., & Raji, A. (2025). Machine learning applications in early warning systems for supply chain disruptions: strategies for adapting to risk, pandemics and enhancing business resilience and economic stability. International Journal of Science and Research Archive, 15(2), 1829–1845. https://doi.org/10.30574/ijsra.2025.15.2.1612

[42]. Kasarapu, B. C. (2026). Generative AI-Enabled Micro-Frontend Framework for Scalable and Intelligent Enterprise Retail Applications. International Journal of Computer Information Systems and Industrial Management Applications, 18, 297–309. https://doi.org/10.70917/ijcisim-2026-2708

[43]. Kaur, D., Uslu, S., Rittichier, K. J., & Durresi, A. (2022). Trustworthy Artificial Intelligence: A Review. ACM Computing Surveys, 55(2), 1–38. https://doi.org/10.1145/3491209

[44]. Kosasih, E. E., & Brintrup, A. (2021). A machine learning approach for predicting hidden links in supply chain with graph neural networks. International Journal of Production Research, 60(17), 5380–5393. https://doi.org/10.1080/00207543.2021.1956697

[45]. Kosasih, E. E., Margaroli, F., Gelli, S., Aziz, A., Wildgoose, N., & Brintrup, A. (2022). Towards knowledge graph reasoning for supply chain risk management using graph neural networks. International Journal of Production Research, 62(15), 5596–5612. https://doi.org/10.1080/00207543.2022.2100841

[46]. Li, J., Wickman, R., Bhatnagar, S., Maity, R. K., & Mukherjee, A. P. (2025). Abstract Operations Research Modeling Using Natural Language Inputs. Information, 16(2), 128. https://doi.org/10.3390/info16020128

[47]. Li, Y., Ko, H., & Ameri, F. (2024). Integrating Graph Retrieval-Augmented Generation With Large Language Models for Supplier Discovery. Journal of Computing and Information Science in Engineering, 25(2). https://doi.org/10.1115/1.4067389

[48]. Liu, C., Wang, Y., Purvis, L., & Potter, A. (2026). Can Digital Twin Technology Enhance Supply-Chain Resilience? A Systematic Literature Review. Sustainability, 18(5), 2361. https://doi.org/10.3390/su18052361

[49]. Liu, J., Li, P., & Wang, Y. (2026). Graph Neural Networks for Modeling Complex Dependencies in Global Supply Chain Networks. Journal of Computing and Electronic Information Management, 20(3), 9–20. https://doi.org/10.54097/6fcw2b19

[50]. Liu, Y., Pei, C., Xu, L., Chen, B., Sun, M., Zhang, Z. L., Sun, Y., Zhang, S., Wang, K., Zhang, H., Li, J., Xie, G., wen, xiaohui, Nie, X., Ma, M., & Pei, D. (2025). OpsEval: A Comprehensive Benchmark Suite for Evaluating Large Language Models' Capability in IT Operations Domain. 503–513. https://doi.org/10.1145/3696630.3728572

[51]. Liu, Z. (2025). Judicial Hallucination: Investigating and Understanding the Unreliability of LLM-as-a-Judge. Applied and Computational Engineering, 203(1), 105–113. https://doi.org/10.54254/2755-2721/2026.tj29492

[52]. Machefer, M., Thomas, A., Meroni, M., Peña, J. M., Ronco, M., Corbane, C., & Rembold, F. (2025). Potential and limitations of machine learning modeling for forecasting Acute Food Insecurity. Global Food Security, 45, 100859. https://doi.org/10.1016/j.gfs.2025.100859

[53]. Nagy, J., & Foltin, P. (2022). Increase supply chain resilience by applying early warning signals within big-data analysis. Ekonomski Vjesnik, 35(2), 467–481. https://doi.org/10.51680/ev.35.2.17

[54]. Padhy, A. K., Patel, T., Soni, V., Shivam, S., Thokala, G. B., & Vulugundam, B. (2026). Machine Learning-Based Fault Prediction in Large- Scale Distributed Systems. 1–6. https://doi.org/10.1109/icaic67076.2026.11395778

[55]. Petrova, S., & Hughes1, M. (2025). Physics-Informed Graph Neural Networks for Supply Chain Disruption Prediction and Mitigation. Frontiers in Applied Physics and Mathematics, 2(1), 130–147. https://doi.org/10.71465/fapm458

[56]. Protogeros, I., & Vanbever, L. (2025). Continual Benchmarking of LLM-Based Systems on Networking Operations. 70–72. https://doi.org/10.1145/3744969.3748425

[57]. R., K. S., Ojha, D., Kaur, P., Mahto, R. V., & Dhir, A. (2024). Explainable artificial intelligence and agile decision-making in supply chain cyber resilience. University of North Texas Digital Library (University of North Texas), 180, 114194. https://doi.org/10.1016/j.dss.2024.114194

[58]. Riddell, E., Riddell, J., Sun, G., Antkiewicz, M., & Czarnecki, K. (2026). Stalled, Biased, and Confused: Uncovering Reasoning Failures in LLMs for Cloud-Based Root Cause Analysis. In arXiv (Cornell University). Cornell University. https://doi.org/10.48550/arxiv.2601.22208

[59]. Sarda, K., Namrud, Z., Watts, I., Shwartz, L., Nagar, S., Mohapatra, P., & Litoiu, M. (2024). Augmenting Automatic Root-Cause Identification with Incident Alerts Using LLM. 1–10. https://doi.org/10.1109/cascon62161.2024.10838171

[60]. Sharma, R., & Jain, K. (2026). ARCH: An LLM-Driven Autonomous Self-Healing Framework for Cloud Operations Using Retrieval-Augmented Generation. Zenodo (CERN European Organization for Nuclear Research). https://doi.org/10.5281/zenodo.19204590

[61]. Shen, Z., Wang, H., Chen, J., & Song, X. (2026). Resilience Inference for Supply Chains with Hypergraph Neural Network. Proceedings of the AAAI Conference on Artificial Intelligence, 40(46), 39259–39267. https://doi.org/10.1609/aaai.v40i46.41274

[62]. Shorinwa, O., Mei, Z., Lidard, J., Ren, A. Z., & Majumdar, A. (2025). A Survey on Uncertainty Quantification of Large Language Models: Taxonomy, Open Research Challenges, and Future Directions. ACM Computing Surveys, 58(3), 1–38. https://doi.org/10.1145/3744238

[63]. Singh, R. (2025). Optimizing Application Support Operations with LLMs- Insights from a Self-Healing Interface. International Journal of Scientific Research and Modern Technology., 4(8), 180. https://doi.org/10.38124/ijsrmt.v4i8.980

[64]. Syed, T. A., Belgaum, M. R., Jan, S., Khan, A. A., & Alqahtani, S. S. (2025). Agentic AI for Autonomous Defense in Software Supply Chain Security: Beyond Provenance to Vulnerability Mitigation. ArXiv.Org. http://arxiv.org/abs/2512.23480

[65]. Tian, Y., Liu, Y., Chong, Z. J., Huang, Z., & Jacobsen, H. (2025). GALA: Can Graph-Augmented Large Language Model Agentic Workflows Elevate Root Cause Analysis? In ArXiv.org. https://doi.org/10.48550/arxiv.2508.12472

[66]. Tiwari, R. R., Vishnu, C. R., Sridharan, R., & Kumar, P. N. R. (2023). Digital Twins, Stress Tests, and the Need for Resilient Supply Chains (pp. 3012–3027). https://doi.org/10.4018/978-1-7998-9220-5.ch180

[67]. Vadakkepati, S. (2025). Hybrid RAG-LLM Framework For Intelligent Supplier Risk Assessment In Global Supply Chains. Journal of International Crisis and Risk Communication Research, 72–87. https://doi.org/10.63278/jicrcr.vi.3496

[68]. Veershetty, G. (2026). Automated Root Cause Analysis in SAP Landscapes Using Large Language Models and Operational Telemetry. International Journal of Emerging Trends in Computer Science and Information Technology, 7, 186–191. https://doi.org/10.63282/3050-9246.ijetcsit-v7i1p127

[69]. Wang, C., Yuan, T., Hua, C., Lu, C., Yang, X., & Qiu, Z. (2025). Integrating Large Language Models with Cloud-Native Observability for Automated Root Cause Analysis and Remediation. In Preprints.org. https://doi.org/10.20944/preprints202512.1926.v1

[70]. Wang, J. (2024). Multimodal Deep Learning Approach for Early Warning of Supply Chain Disruptions Using NLP and Anomaly Detection. Artificial Intelligence and Machine Learning Review, 5(3), 98–110. https://doi.org/10.69987/aimlr.2024.50308

[71]. Wang, L., & Tsung, F. (2025). Automated Knowledge Graph Construction for Supply Chain Datasets Assisted by LLMs. 2738–2743. https://doi.org/10.1109/case58245.2025.11164086

[72]. Wang, Y., & Li, K. (2025). Large Language Models in Operations Research: Methods, Applications, and Challenges. In arXiv (Cornell University). Cornell University. https://doi.org/10.48550/arxiv.2509.18180

[73]. Wasserkrug, S., Boussioux, L., Hertog, D. den, Mirzazadeh, F., Birbil, Ş. İ., Kurtz, J., & Maragno, D. (2025). Enhancing Decision Making Through the Integration of Large Language Models and Operations Research Optimization. Proceedings of the AAAI Conference on Artificial Intelligence, 39(27), 28643–28650. https://doi.org/10.1609/aaai.v39i27.35090

[74]. Wichmann, P., Brintrup, A., Baker, S., Woodall, P., & McFarlane, D. (2020). Extracting supply chain maps from news articles using deep neural networks. International Journal of Production Research, 58(17), 5320–5336. https://doi.org/10.1080/00207543.2020.1720925

[75]. Wu, R. (2026). Inventory optimization under supply chain disruptions: Leveraging large language models for human-AI collaborative decision-making. Journal of King Saud University - Computer and Information Sciences, 38(2). https://doi.org/10.1007/s44443-025-00458-9

[76]. Xue, Z., Wang, Y., & Huo, M. (2026). LLM-Grounded Explainable AI for Supply Chain Risk Early Warning via Temporal Graph Attention Networks. In arXiv (Cornell University). Cornell University. https://doi.org/10.48550/arxiv.2603.04818

[77]. Yadkori, Y. A., Kuzborskij, I., Stutz, D., György, A., Fisch, A., Douc, R., Beloshapka, I., Weng, W.-H., Yang, Y.-Y., Szepesvári, C., Cemgil, A. T., & Tomasev, N. (2024). Mitigating LLM Hallucinations via Conformal Abstention. In arXiv (Cornell University). Cornell University. https://doi.org/10.48550/arxiv.2405.01563

[78]. Yan, Y., Xu, M., Li, B., & Hui, F. K. P. (2025). Autonomous Supply Chain Disruption Response Using Multi-Agent AI Systems: A Comparative Performance Analysis. https://doi.org/10.46254/au04.20250223

[79]. Yoshizato, K., Shimizu, K., Higa, R., & Otsuka, T. (2026). AI Agent Systems for Supply Chains: Structured Decision Prompts and Memory Retrieval. In arXiv (Cornell University). Cornell University. https://doi.org/10.48550/arxiv.2602.05524

[80]. Yu, T. T., Mostajabdaveh, M., Byusa, J. S., Ramamonjison, R., Carenini, G., Mao, K., Zhou, Z., & Zhang, Y. (2025). SMARTAPS: Tool-augmented LLMs for Operations Management. In ArXiv.org. https://doi.org/10.48550/arxiv.2507.17927

[81]. Zhang, Y., Li, Y., Cui, L., Deng, C., Liu, L., Fu, T., Huang, X., Zhao, E., Zhang, Y., Chen, Y., Wang, L., Luu, A. T., Bi, W., Shi, F., & Shi, S. (2025). 🧜 Siren's Song in the AI Ocean: A Survey on Hallucination in Large Language Models. Computational Linguistics, 51(4), 1373–1418. https://doi.org/10.1162/coli.a.16

[82]. Zhong, Z., Fu, R., Ma, M., Zhang, S., Sun, Y., Bansal, C., & Pei, D. (2026). LLM-Enhanced Failure Localization in Microservices: Integrating Multi-Modal Data and Expert Interpretation. IEEE Transactions on Services Computing, 1–14. https://doi.org/10.1109/tsc.2026.3676262

[83]. Zhou, Y., Lentz, E., Michelson, H., Kim, C., & Baylis, K. (2021). Machine learning for food security: Principles for transparency and usability. Applied Economic Perspectives and Policy, 44(2), 893–910. https://doi.org/10.1002/aepp.13214