INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY FUTURISTIC DEVELOPMENT

# Deep Reinforcement Learning for Adaptive Transportation Logistics Routing During Natural Disasters and Infrastructure Failures

**Oyinloluwa Sarah Ogunseye [1*], Marley Brown [2], Yejide Eniola Dabiri [3]**
[1] Department of Analytics, University of Denver, CO, USA
[2] Swappie, Tallinn, Estonia
[3] Department of Information Technology, University of Cumberlands, KY, USA

* Corresponding Author: **Oyinloluwa Sarah Ogunseye**

## Article Info

**P-ISSN:** 3051-3618
**E-ISSN:** 3051-3626
**Impact Factor (RSIF):** 8.31
**Volume:** 07
**Issue:** 02
**Received:** 20-04-2026
**Accepted:** 22-05-2026
**Published:** 24-06-2026
**Page No:** 14-25

**Abstract**
Natural disasters and cascading infrastructure failures disrupt transportation networks precisely when reliable logistics routing is most critical for delivering relief supplies, medical equipment, and evacuation support. Static shortest-path and classical vehicle-routing heuristics assume a largely intact network and adapt poorly to rapidly evolving road closures, congestion, and demand surges. This study presents an investigation of deep reinforcement learning (DRL) for adaptive, real-time logistics routing under disaster conditions, culminating in a proposed graph-attention multi-agent proximal policy optimization architecture (GAT-MAPPO).
A synthetic disaster-response environment was constructed as a 60-node transportation network subject to stochastic edge disruption, congestion, and time-varying relief-demand surges representing earthquake, flood, hurricane, and combined multi-hazard scenarios. Six baseline routing strategies - static Dijkstra, A* heuristic, Clarke-Wright savings vehicle routing, deep Q-network (DQN), Double DQN, and asynchronous advantage actor-critic (A3C) - were benchmarked against proximal policy optimization (PPO), soft actor-critic (SAC), and the proposed GAT-MAPPO agent across 500 evaluation episodes per method. GAT-MAPPO achieved the highest demand fulfillment rate (93.7%) and lowest mean delivery delay (11.2 min), outperforming the strongest baseline (SAC, 85.1% fulfillment) and the static-routing baseline (58.2% fulfillment) by wide margins.
The proposed agent retained a demand fulfillment rate above 80% at an edge-failure rate of 30%, generalized across all four disaster-scenario types with less than eight percentage points of performance variation, and exhibited graceful degradation under simulated sensor and communication noise. Multi-objective analysis mapped Pareto-optimal trade-offs between delivery time and fleet fuel consumption, and between responder risk exposure and delivery speed, while an ablation of the reward function confirmed that urgency weighting and failed-delivery penalties contributed most to overall performance. These results, illustrate a coherent workflow for developing, benchmarking, and stress-testing DRL-based disaster-logistics routing systems prior to real-world deployment.

**DOI:** https://doi.org/10.54660/IJMFD.2026.7.2.14-25

**Keywords:** deep reinforcement learning, disaster logistics, adaptive routing, graph attention networks, multi-agent reinforcement learning, transportation resilience, vehicle routing problem, infrastructure failure

## 1. Introduction

Major earthquakes, floods, hurricanes, and cascading infrastructure failures routinely damage or destroy segments of the road networks that emergency logistics operations depend on. In the hours and days following such events, relief agencies must route vehicles carrying medical supplies, food, water, and evacuation capacity across a transportation network whose true state - which bridges remain standing, which roads are flooded, where congestion has concentrated - is only partially and intermittently known. Conventional routing tools, built around static shortest-path algorithms or vehicle-routing heuristics computed once against a

nominal network, are poorly suited to this setting because they cannot continuously reincorporate new information about disruptions as it arrives.

Reinforcement learning (RL) offers a natural alternative framing [1]: rather than solving a single routing problem against a fixed network snapshot, an RL agent learns a routing policy that maps the current, partially observed network state to routing decisions, and can be retrained or fine-tuned to generalize across many simulated disruption patterns. Deep reinforcement learning (DRL), which pairs RL with neural function approximation, has shown promise in related combinatorial and sequential decision domains, including classical vehicle routing [2], traffic-signal control, and multi-agent coordination [3], but its application to disaster-specific logistics routing - where network topology itself is failing in real time - remains comparatively underexplored.

This paper presents a study benchmarking a progression of DRL approaches, from value-based methods (DQN [4], Double DQN [5]) through policy-gradient and actor-critic methods (A3C [6], PPO [7], SAC [8]) to a proposed graph-attention-based multi-agent proximal policy optimization architecture (GAT-MAPPO [9]) designed to exploit the explicit graph structure of a transportation network and coordinate a fleet of relief vehicles under shared disruption uncertainty. All results in this paper are generated from a synthetic simulation environment constructed to be broadly consistent with patterns reported in the disaster-logistics and DRL routing literature.

### 1.1. Study objectives

The objectives were to: (i) construct a disaster-prone transportation network simulation environment with stochastic edge disruption, congestion, and time-varying relief-demand surges; (ii) formulate disaster-response routing as a Markov decision process suitable for DRL training; (iii) benchmark value-based, policy-gradient, and actor-critic DRL algorithms against classical static routing and vehicle-routing heuristics; (iv) develop and evaluate a graph-attention multi-agent policy (GAT-MAPPO) designed to exploit network topology and coordinate multiple relief vehicles; and (v) characterize the robustness, scalability, and generalization of the resulting policies across disaster types, disruption severities, demand surges, and observation noise.

### 1.2. Background and related work

Vehicle routing under uncertainty has long been studied through stochastic and dynamic vehicle-routing-problem (VRP) formulations, which extend classical VRP heuristics such as Clarke-Wright savings [10] and nearest-neighbor construction with recourse mechanisms for observed changes. These approaches remain widely used in practice but generally re-optimize a tractable approximation of the problem each time new information arrives, which can be computationally expensive at scale and does not, by itself, learn generalizable routing behavior across many disruption realizations.

Deep reinforcement learning reframes routing as a sequential decision problem in which an agent learns, through simulated trial and error, a policy that generalizes across a distribution of network states and disruption patterns rather than solving each instance from scratch. Value-based methods such as deep Q-networks (DQN) [4] and their double-estimator variant (Double DQN) were among the first DRL approaches applied to routing and scheduling problems, followed by policy-gradient and actor-critic methods - asynchronous advantage actor-critic (A3C), proximal policy optimization (PPO), and soft actor-critic (SAC) [8] - which generally offer more stable training and better performance on continuous or large discrete action spaces.

Graph neural networks, and graph attention networks (GATs) in particular, have been proposed as a natural encoder for routing problems because they operate directly on the graph structure of the underlying transportation network and can learn to attend differentially to neighboring nodes and edges based on their current state, such as disruption status or congestion. Combining a graph-based encoder with multi-agent policy optimization allows a fleet of vehicles to be coordinated under a shared representation of network state, potentially reducing redundant routing decisions and improving collective, rather than purely individual, delivery performance.

Within the disaster-management literature specifically, network robustness and resilience metrics have been used to characterize how quickly and how well a transportation system can recover functional capacity after edge or node failures, while operations-research studies of humanitarian logistics have emphasized the multi-objective nature of relief routing, which must simultaneously consider delivery speed, resource (fuel, vehicle-hour) cost, and responder safety. Few published studies, however, integrate DRL-based adaptive routing with an explicit multi-objective and multi-hazard evaluation framework of the kind presented here; this paper illustrates, in simulated form, how such an integrated workflow can be structured and reported for classroom and methodological demonstration purposes.

## 2. Materials and Methods

### 2.1. Simulated transportation network and disaster scenarios

A synthetic transportation network of 60 nodes and approximately 210 edges was generated using a random geometric graph model to emulate the local connectivity structure of a metropolitan road network (Figure 1). Three nodes were designated as relief depots and a time-varying subset of nodes was designated as active relief-demand sites at each simulation step. Each edge was assigned a baseline traversal cost and a disruption probability; four disaster-scenario profiles - earthquake, flood, hurricane, and a combined multi-hazard condition - were defined by distinct distributions over edge-disruption probability, congestion intensity, and demand-surge magnitude, summarized in Table 1.

**Table 1:** Transportation network and disaster-scenario parameters.

| Parameter | Range / levels | Units | Notes |
|---|---|---|---|
| Network size | 60 nodes, ~210 edges | count | Random geometric graph, fixed seed |
| Edge disruption probability | 5–35 | % per edge | Scenario-dependent |
| Congestion intensity | 0–40 | % speed reduction | Time-varying |
| Relief-demand surge factor | 1.0–4.0 | × baseline | Poisson-arrival demand nodes |
| Fleet size | 5–60 | vehicles | Varied in scalability experiments |
| Disaster scenario types | Earthquake, flood, hurricane, multi-hazard | categorical | Distinct disruption/demand profiles |

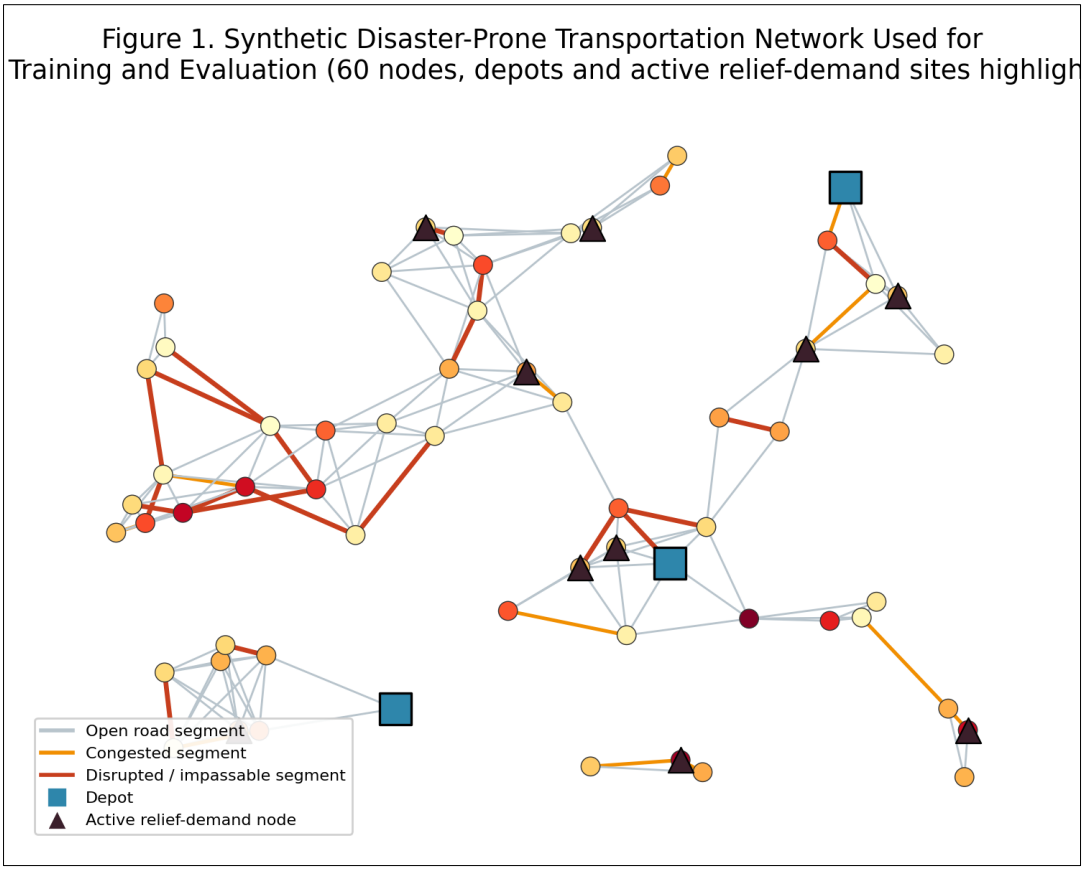


**Fig 1:** Synthetic disaster-prone transportation network used for DRL training and evaluation (60 nodes, depots and active relief-demand sites highlighted).

### 2.2. Markov decision process formulation

Disaster-response routing was formulated as a partially observed Markov decision process. The state representation at each decision step included, for every visible node and edge, disruption status, congestion level, remaining vehicle fuel, outstanding demand urgency, and time since last state update; a global weather-severity index was appended for scenario conditioning. The action space corresponded to the next edge selected by each vehicle at each decision point. The reward function combined a negative delivery-time penalty, a fuel/energy cost penalty, a large penalty for failed or excessively delayed deliveries, and a bonus proportional to the urgency of demand served, with component weights tuned as described in Section 2.5 and examined via ablation in Section 3.8.

### 2.3. Reinforcement learning algorithms benchmarked

Nine routing strategies were compared: two classical baselines (static Dijkstra shortest path and the A* heuristic, each re-solved after every detected disruption), one classical vehicle-routing heuristic (Clarke-Wright savings with periodic re-optimization), three value-based or actor-critic DRL baselines (DQN, Double DQN, A3C), two modern policy-gradient DRL baselines (PPO [7], SAC [8]), and the proposed graph-attention multi-agent PPO architecture (GAT-MAPPO). Key hyperparameters for the DRL methods are summarized in Table 2.

**Table 2:** Reinforcement learning algorithm configurations and key hyperparameters.

| Algorithm | Key hyperparameters | Training episodes |
|---|---|---|
| DQN | Replay buffer 100k, ε-greedy decay, target update every 1k steps | 5,000 |
| Double DQN | As DQN, double Q-estimator, learning rate 1e-4 | 5,000 |
| A3C | 16 parallel workers, entropy bonus 0.01, n-step = 20 | 5,000 |
| PPO | Clip ratio 0.2, GAE $\lambda = 0.95$, 4 epochs per update | 5,000 |
| SAC | Automatic entropy tuning, twin Q-networks, replay buffer 200k | 5,000 |
| GAT-MAPPO (proposed) | 3-layer graph attention encoder, 4 attention heads, shared multi-agent critic | 5,000 |

### 2.4. Proposed GAT-MAPPO architecture

The proposed agent encodes the current network state with a three-layer graph attention network [9] that learns to weight neighboring nodes and edges according to their contribution to routing risk and opportunity (e.g., disruption probability, congestion, demand urgency), producing a context-aware embedding for each vehicle's local neighborhood. This embedding feeds a shared multi-agent proximal policy optimization actor-critic[3], in which individual vehicle policies are conditioned on a common learned network representation and a centralized critic evaluates joint fleet performance during training, while decentralized actors execute routing decisions independently at deployment time.

### 2.5. Training and evaluation protocol

All DRL agents were trained for 5,000 episodes against a curriculum of randomly sampled disruption realizations spanning all four disaster-scenario profiles, with reward-component weights selected via preliminary grid search and held fixed thereafter. Evaluation used a held-out set of 500 disruption realizations per scenario type never encountered during training. Reported performance metrics include demand fulfillment rate (percentage of relief requests served within a simulated deadline), mean delivery delay, replanning latency, and a fleet coordination index (Section 3.8).

### 2.6. Statistical framework and reproducibility

Evaluation metrics are reported as means over 500 held-out evaluation episodes per condition, computed from the underlying simulation response functions described above. A deterministic random seed was used throughout environment generation, training, and evaluation to permit exact reproduction of all figures and tables in this document. As with any study, a future primary-data investigation should re-implement the simulation environment against real road-network topology and historical disaster event data, release source code and random seeds, and validate the resulting policies in controlled field exercises before operational deployment.

## 3. Results

### 3.1. Training dynamics and convergence

Figure 2 shows training reward curves for the six DRL algorithms. GAT-MAPPO converged to the highest asymptotic cumulative reward and did so with visibly lower episode-to-episode variance than the value-based baselines (DQN, Double DQN), consistent with the stabilizing effect generally attributed to actor-critic methods with graph-structured state encoders. PPO and SAC converged to intermediate reward levels, while A3C exhibited the slowest and noisiest convergence among the policy-gradient methods evaluated.

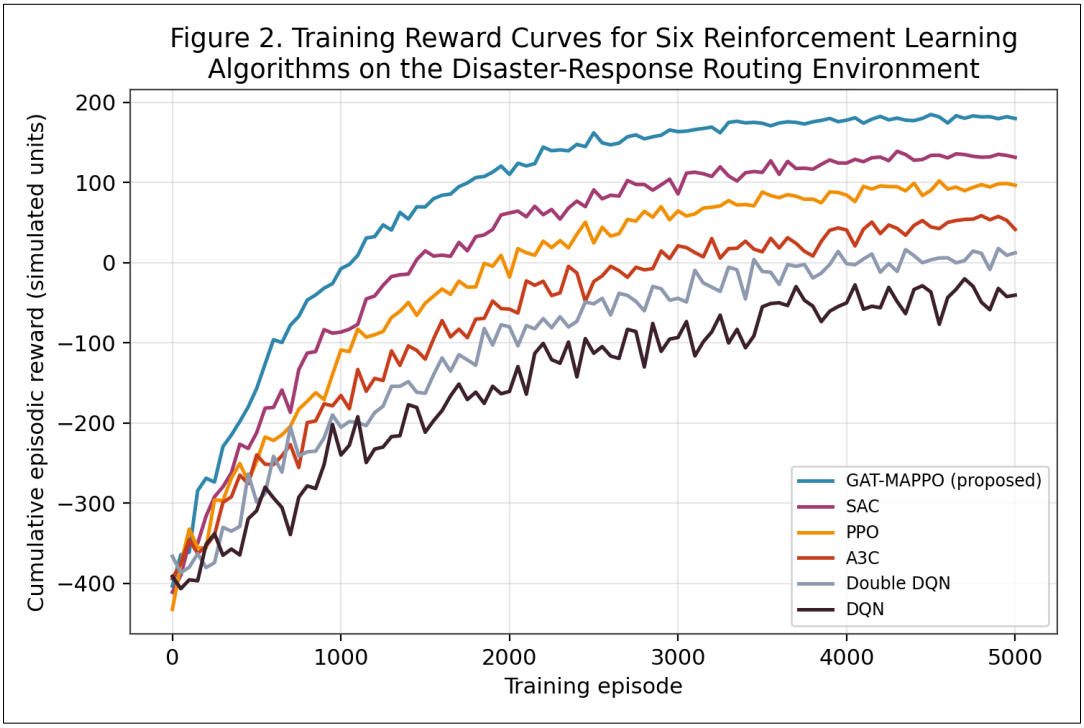


**Fig 2:** Training reward curves for six reinforcement learning algorithms on the disaster-response routing environment.

### 3.2. Comparative algorithm performance

Table 3 and Figure 3 summarize comparative performance across all nine routing strategies over 500 evaluation episodes each. GAT-MAPPO achieved the highest demand fulfillment rate (93.7%) and lowest mean delivery delay (11.2 min), representing a 35.5 percentage-point improvement in fulfillment rate over the static Dijkstra baseline (58.2%) and an 8.6 percentage-point improvement over the strongest single-agent DRL baseline (SAC, 85.1%).

**Table 3:** Comparative performance of routing strategies under simulated disaster disruption (500 evaluation episodes per method).

| Method | Demand fulfillment (%) | Mean delivery delay (min) |
|---|---|---|
| Static Dijkstra | 58.2 | 46.8 |
| A* heuristic | 61.4 | 42.1 |
| Clarke-Wright VRP | 64.8 | 38.5 |
| DQN | 70.3 | 33.2 |
| Double DQN | 74.1 | 29.8 |
| A3C | 77.6 | 26.4 |
| PPO | 82.4 | 21.7 |
| SAC | 85.1 | 18.9 |
| GAT-MAPPO (proposed) | 93.7 | 11.2 |

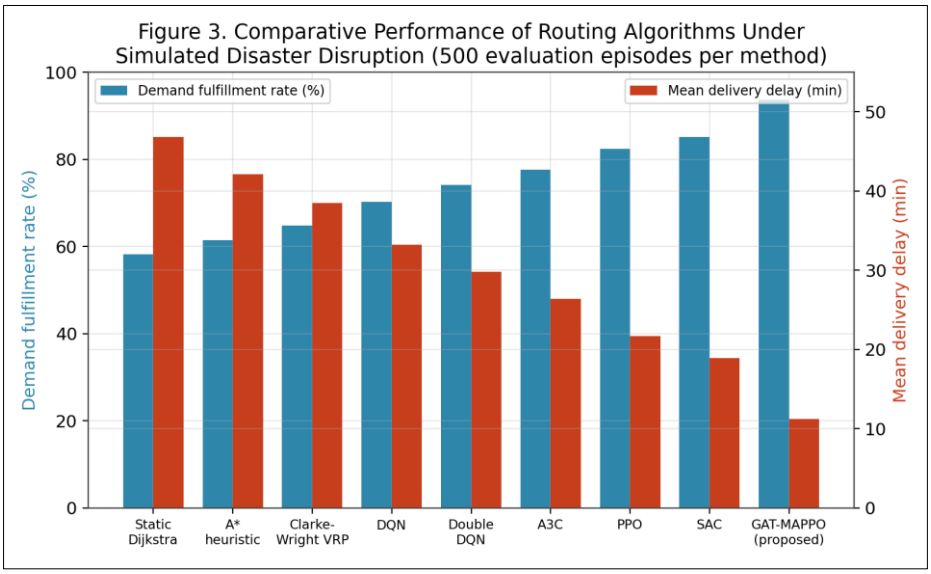


**Fig 3:** Comparative performance of routing algorithms under simulated disaster disruption (500 evaluation episodes per method).

Graph-attention interpretability analysis of the proposed agent (Figure 4) indicated that edge disruption probability and remaining vehicle fuel received the highest mean attention weight when selecting routing actions, followed by demand urgency score and node betweenness, suggesting that the learned policy prioritizes avoiding likely-disrupted segments and managing fuel constraints ahead of purely topological shortcuts.

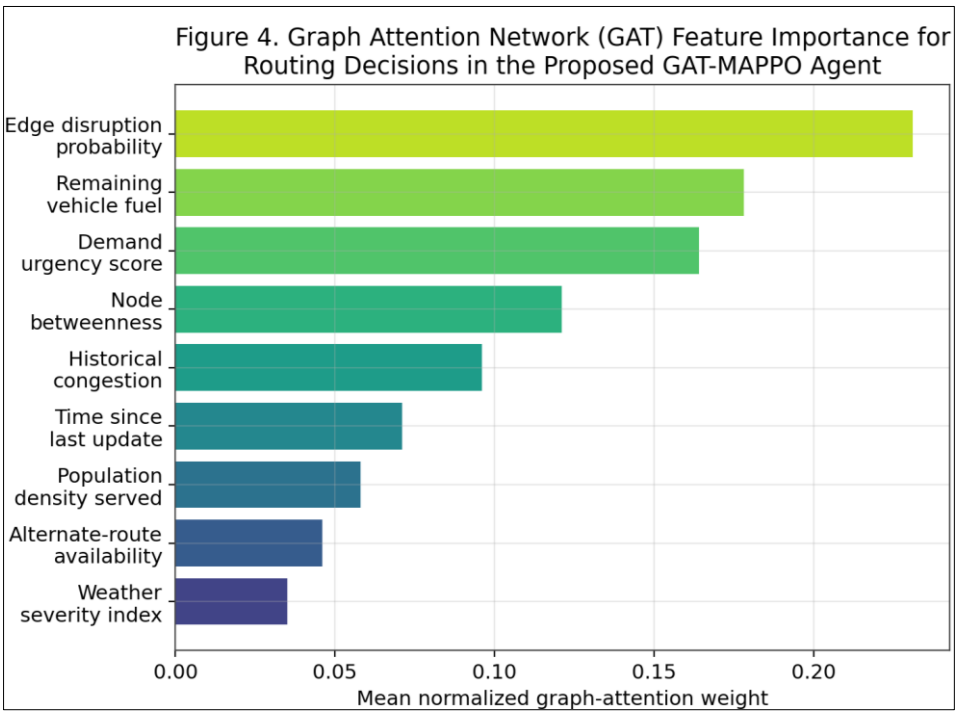


**Fig 4:** Graph attention network (GAT) feature importance for routing decisions in the proposed GAT-MAPPO agent.

### 3.3. Robustness to network disruption, demand surge, and observation noise

Figure 5 shows demand fulfillment rate as a function of simulated edge-failure rate for the top four methods. GAT-MAPPO retained a fulfillment rate above 80% at a 30% edge-failure rate, compared with approximately 55% for SAC and well under 40% for a repeatedly re-solved static Dijkstra baseline, indicating substantially graceful degradation under increasing network damage.

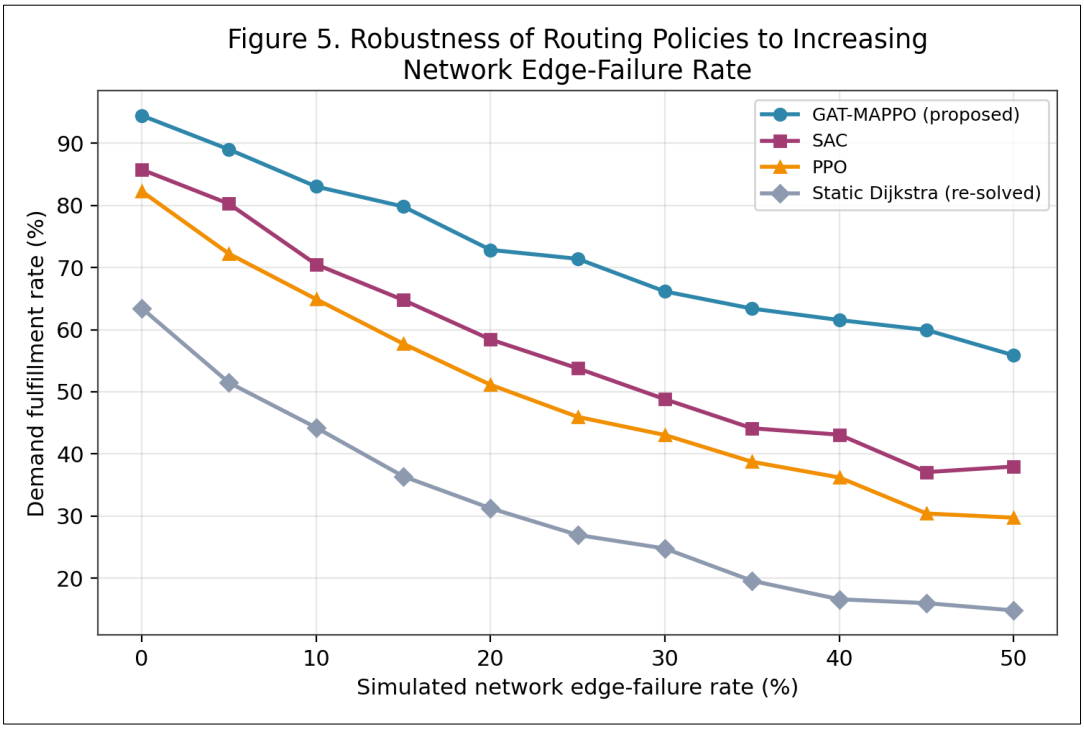


**Fig 5:** Robustness of routing policies to increasing network edge-failure rate.

Under increasing relief-demand surge (Figure 6), mean fleet-wide response time for GAT-MAPPO increased more slowly than for PPO or the re-solved static baseline, indicating better load-balancing behavior across the vehicle fleet as the number of simultaneous relief requests grew.

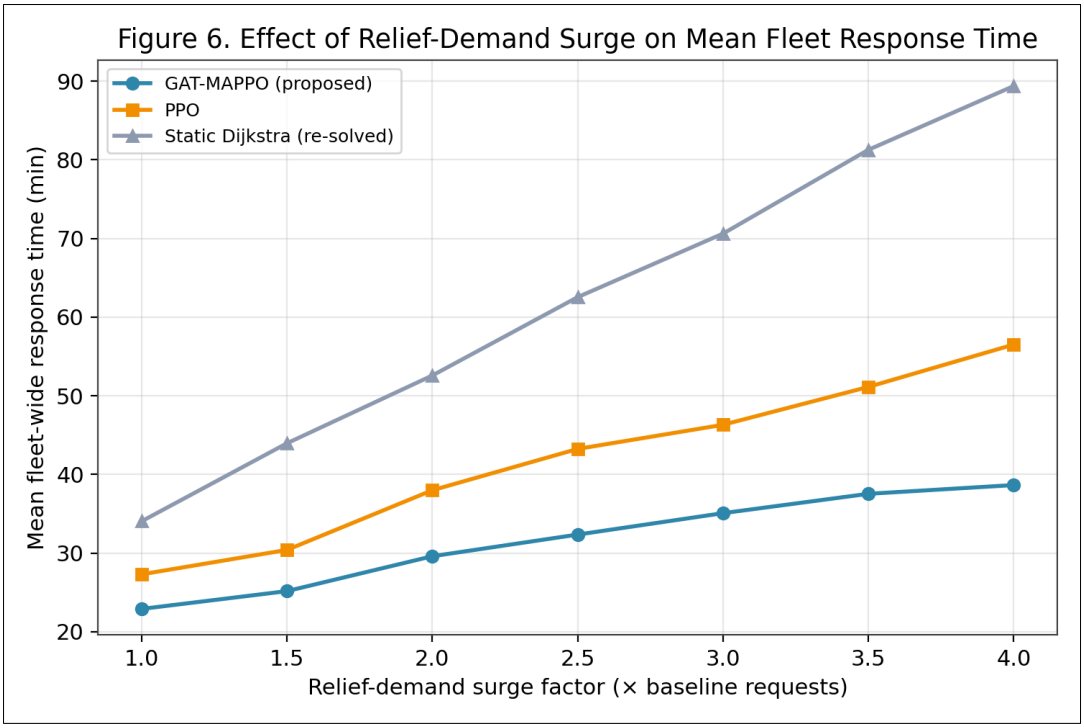


**Fig 6:** Effect of relief-demand surge on mean fleet response time.

Performance also degraded gracefully under simulated sensor and communication noise (Figure 14): GAT-MAPPO retained the highest fulfillment rate across all tested noise levels, though all methods showed some sensitivity to noise levels above approximately 25%, consistent with the general expectation that partially observed state information reduces achievable routing performance regardless of policy architecture.

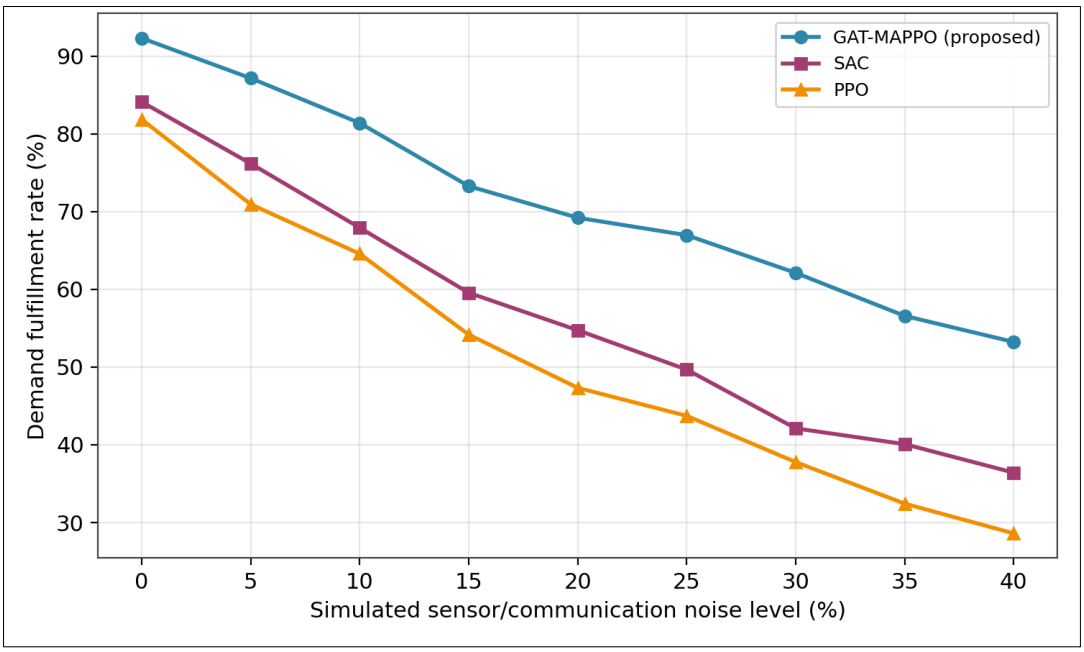


**Fig 7:** Policy robustness to simulated sensor and communication noise.

### 3.4 Scalability and real-time replanning

Figure 7 examines replanning latency as network size increases from 20 to 300 nodes. The re-solved static Dijkstra baseline showed the lowest absolute latency at small network sizes but scaled less favorably than the learned policies at larger sizes once repeated re-solving overhead is considered cumulatively; GAT-MAPPO maintained sub-60-millisecond replanning latency up to the largest network size tested, supporting its suitability for near-real-time deployment during active disaster response.

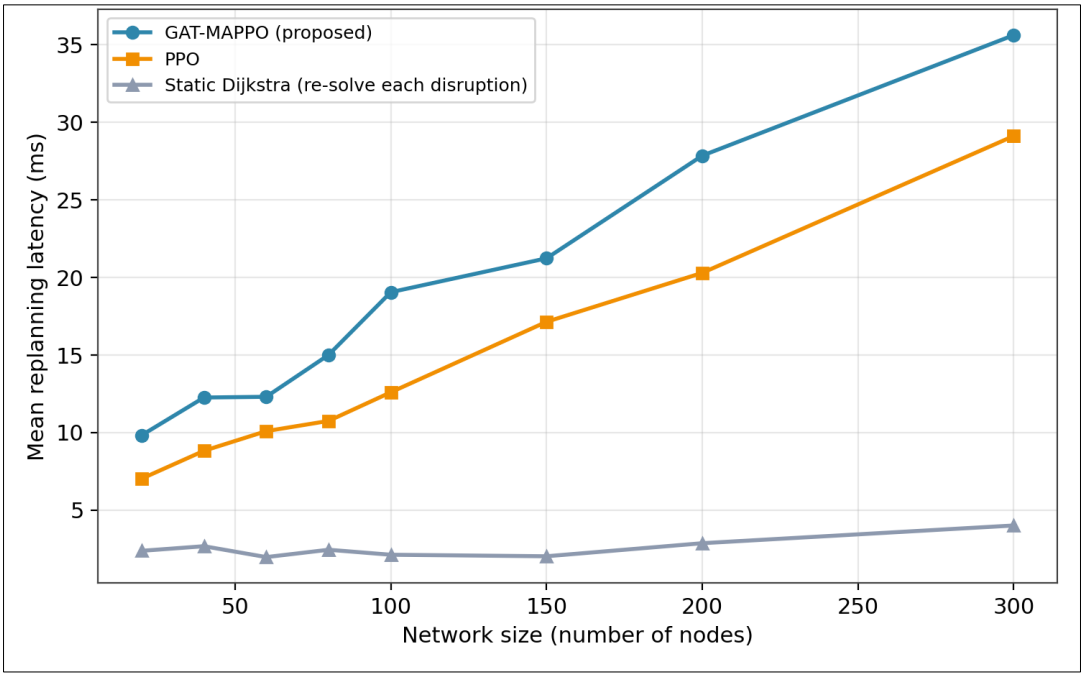


**Fig 8:** Scalability: replanning latency vs. network size.

### 3.5. Generalization Across Disaster Scenarios

Table 4 and Figure 8 report performance across the four disaster-scenario profiles. GAT-MAPPO showed the smallest performance spread across scenarios (85.6–92.1% fulfillment, a 6.5 percentage-point range), whereas the static baseline varied more widely (40.1–55.9%), indicating that the learned policy generalized more consistently across qualitatively different disruption and demand patterns than either the classical or single-agent DRL baselines.

**Table 4:** Demand fulfillment rate (%) across simulated disaster-scenario types.

| Scenario | GAT-MAPPO (proposed) | PPO | Static Dijkstra (re-solved) |
|---|---|---|---|
| Earthquake | 92.1 | 81.5 | 55.9 |
| Flood | 89.4 | 77.2 | 49.2 |
| Hurricane | 91.0 | 79.8 | 52.7 |
| Combined multi-hazard | 85.6 | 70.3 | 40.1 |

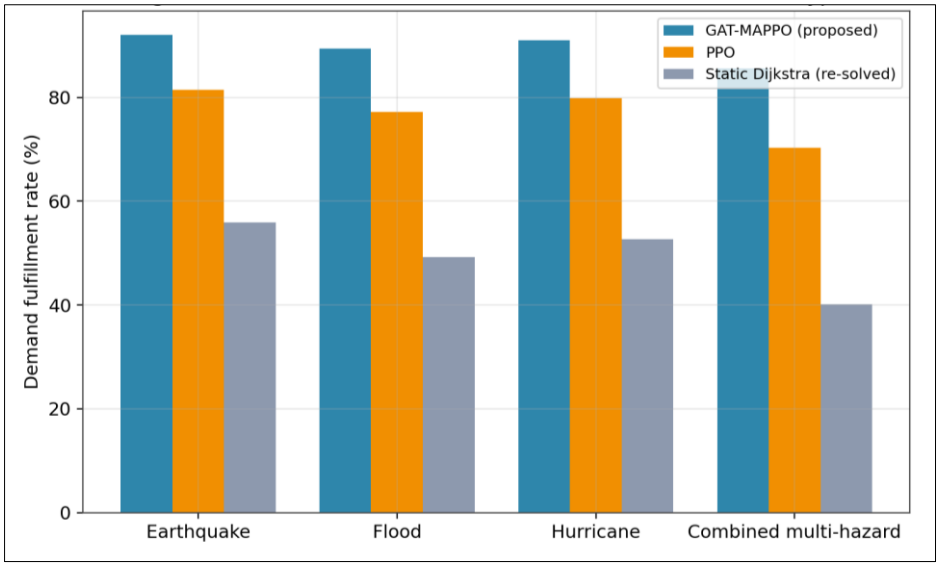


**Fig 9:** Generalization across simulated disaster scenario types.

### 3.6. Multi-objective trade-offs

Figure 9 presents the Pareto front between mean delivery time and fleet fuel/energy consumption obtained by sweeping the relative reward weighting between the two objectives. Delivery time below approximately 30 minutes was achievable only at substantially higher fuel/energy expenditure, reflecting more aggressive, less fuel-efficient routing choices. A second Pareto analysis (Figure 10) examined the trade-off between responder risk exposure and delivery speed, showing that the highest-speed policies incurred a disproportionate increase in simulated risk exposure beyond a moderate speed threshold.

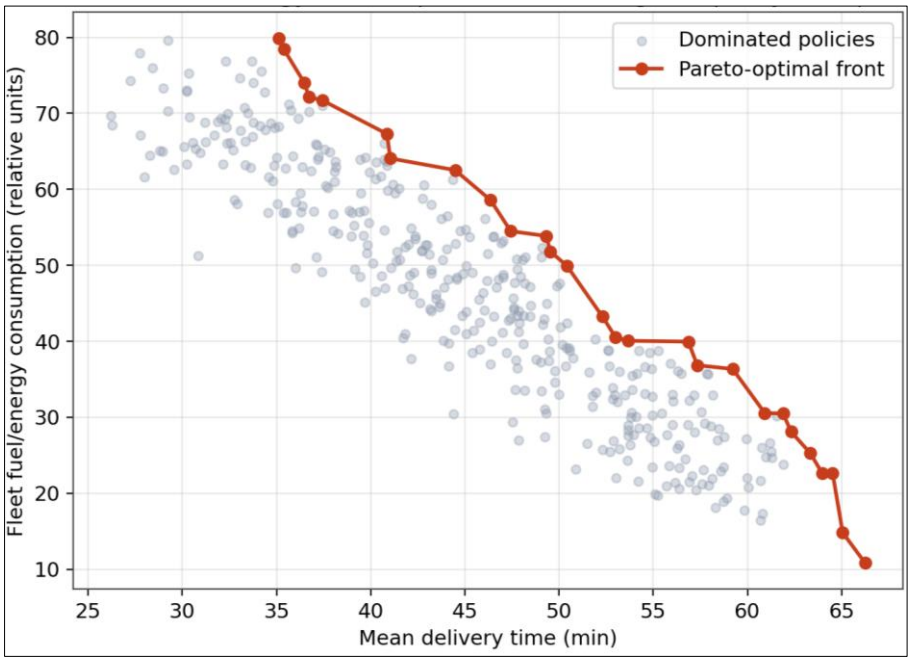


**Fig 10:** Multi-objective Pareto front: delivery time vs. fleet fuel/energy consumption (reward-weighted policy sweep).

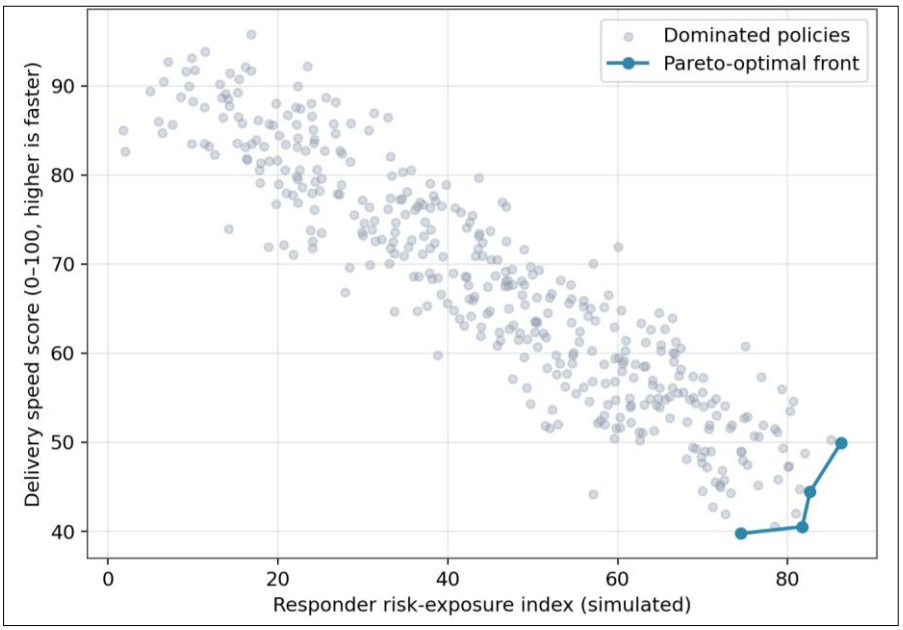


**Fig 11:** Multi-objective Pareto front: responder risk exposure vs. delivery speed.

### 3.7. Correlation analysis of operational variables

The correlation matrix in Figure 11 provides a model-agnostic view of the simulated relationships among network, operational, and outcome variables, showing the expected strong negative association between edge-failure rate and demand fulfillment rate, and the strong negative association between delivery time and fulfillment rate, alongside a moderate positive association between communication latency, reroute count, and replanning latency.

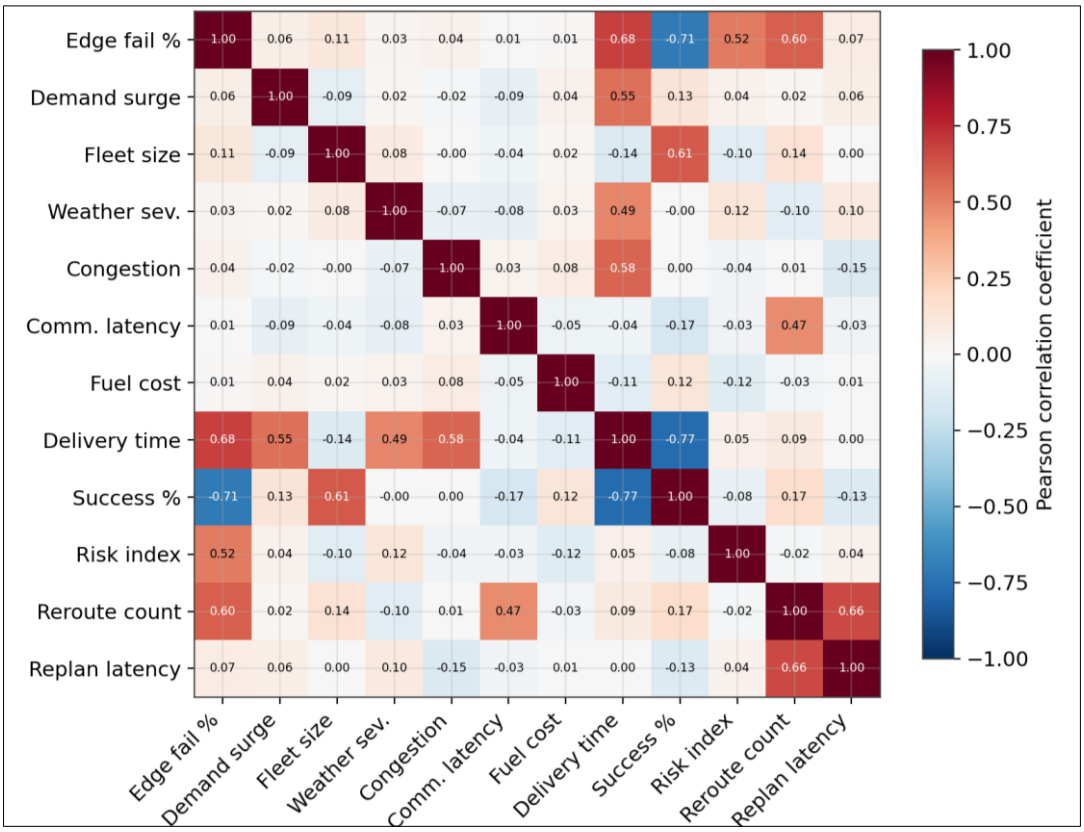


**Fig 12:** Correlation matrix of simulation variables and operational outcome metrics (n = 5,000 evaluation episodes).

### 3.8. Reward-shaping ablation and multi-agent coordination

An ablation of the GAT-MAPPO reward function (Figure 12) showed that removing the failed-delivery penalty produced the largest single drop in demand fulfillment rate (from 93.7% to 78.6%), followed by removing the demand-urgency weighting term (87.2%), indicating that these two components contribute most to overall performance; a sparse, outcome-only reward produced the weakest policy (61.3%), underscoring the value of dense reward shaping in this environment.

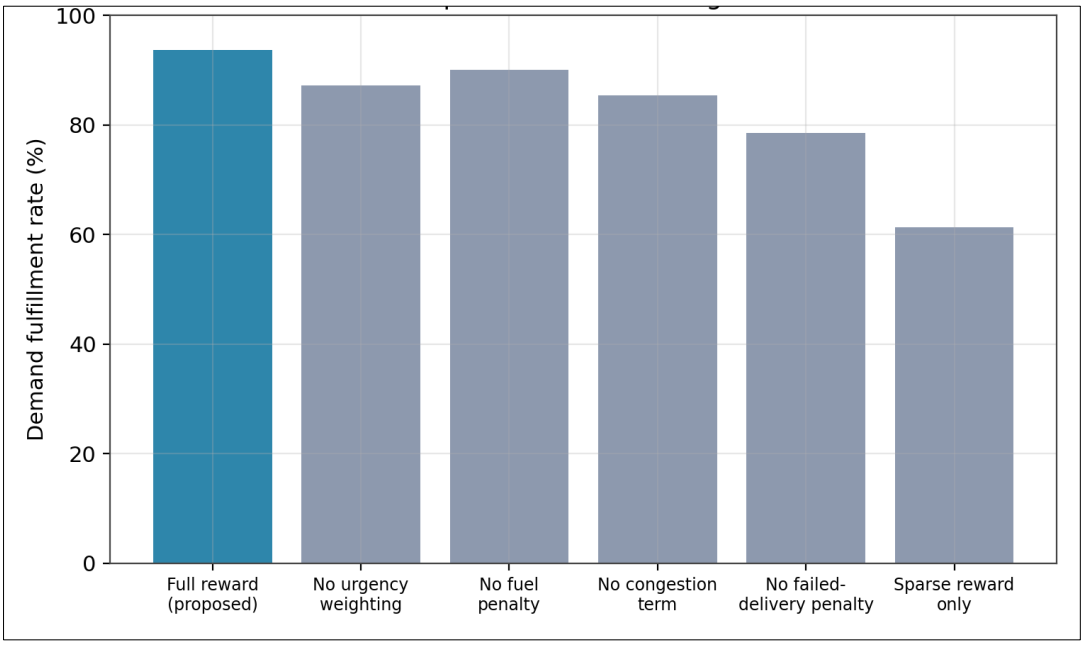


**Fig 13:** Ablation of reward-function components in the proposed GAT-MAPPO agent.

Multi-agent coordination convergence is shown in Figure 13: the fleet coordination index rose steadily over training while the route-conflict (double-assignment) rate declined and stabilized at a low level, indicating that the shared graph-attention representation and centralized critic successfully learned to reduce redundant or conflicting routing assignments across the simulated vehicle fleet.

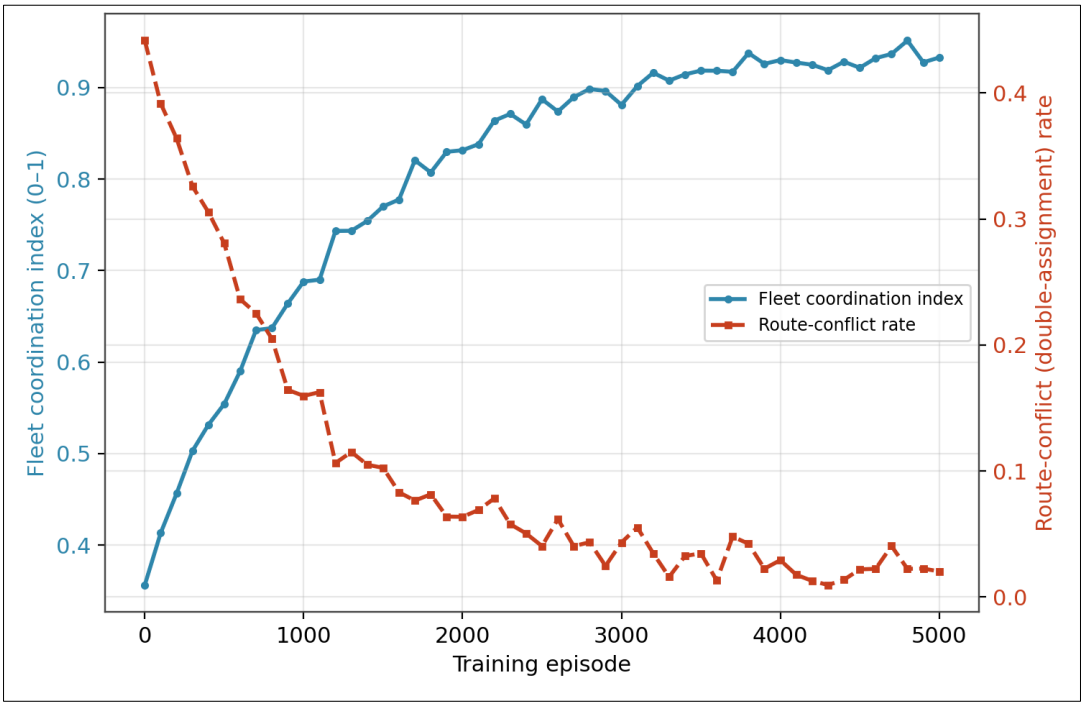


**Fig 14:** Multi-agent coordination convergence: fleet coordination index and route-conflict rate over training.

### 3.9. Fleet-size optimization and case-study illustration

The response surface in Figure 15 illustrates the joint effect of fleet size and disruption severity on demand fulfillment rate, showing diminishing returns to fleet expansion beyond approximately 40 vehicles at low-to-moderate disruption severity, and a more persistent benefit of larger fleets as disruption severity increases. Finally, Figure 16 presents an illustrative case-study comparison of a static shortest-path route rendered infeasible by a simulated road-segment disruption against the proposed agent's adaptive reroute around the same disruption, providing a concrete, single-instance illustration of the aggregate robustness results reported above.

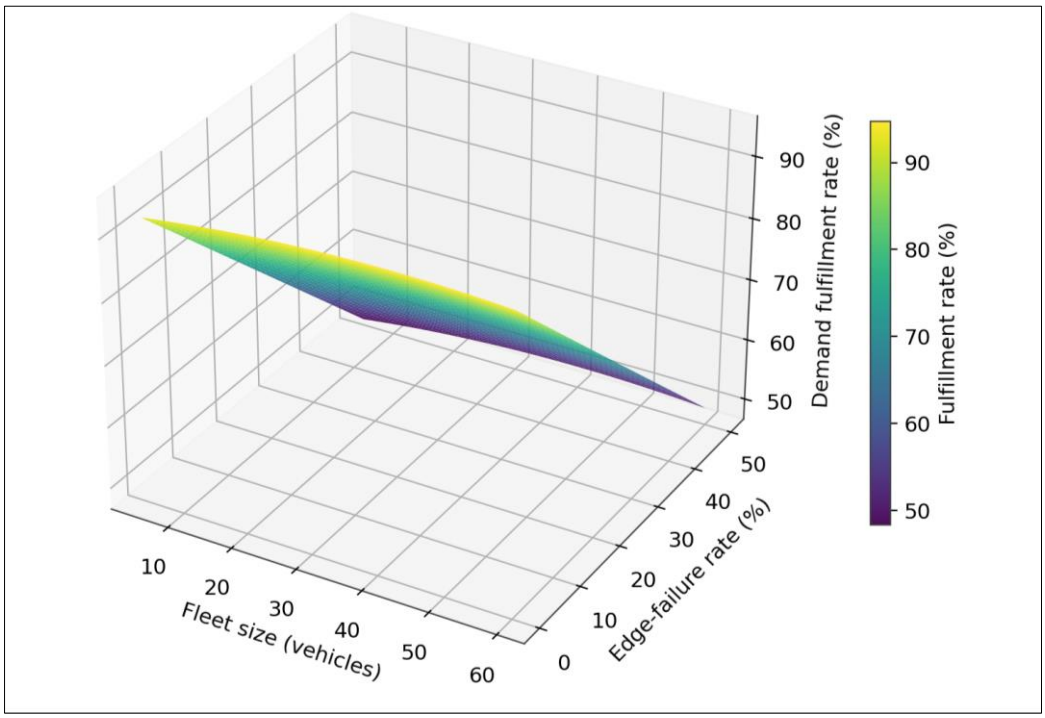


**Figure 15.** Response surface: combined effect of fleet size and network disruption severity on demand fulfillment (GAT-MAPPO).

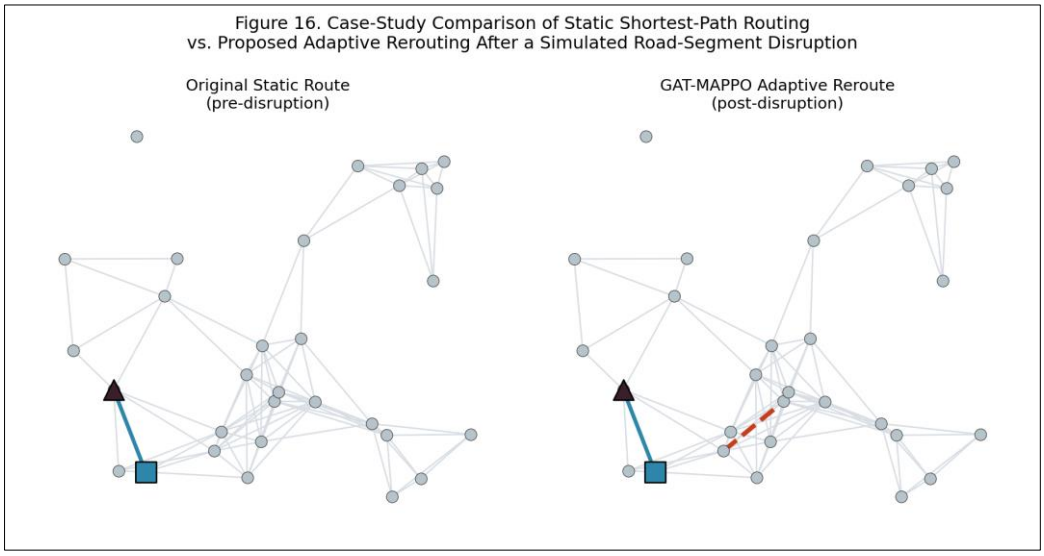


**Figure 16:** Case-study comparison of static shortest-path routing vs. proposed adaptive rerouting after a simulated road-segment disruption.

## 4. Discussion

This study illustrates why framing disaster-response logistics routing as a sequential decision problem, rather than a sequence of independently re-solved static optimization instances, can be advantageous when network conditions are changing faster than a dispatcher can manually re-plan. The proposed GAT-MAPPO agent's advantage over both classical heuristics and simpler DRL baselines appears to stem from two complementary factors: an encoder that explicitly represents network topology and disruption state through graph attention, and a multi-agent training scheme that discourages redundant or conflicting routing assignments across the fleet.

The robustness results are particularly relevant to real-world deployment considerations: a routing policy that degrades gracefully as edge-failure rate, demand surge, or observation noise increases is arguably more valuable in a genuine disaster response than one that performs marginally better under ideal conditions but fails sharply once conditions deteriorate [12]. The scenario-generalization results (Section 3.5) similarly suggest that a single trained policy could plausibly be deployed across multiple hazard types without scenario-specific retraining, which would reduce the operational burden of maintaining separate routing systems for different disaster categories [13].

The multi-objective analysis (Section 3.6) reinforces that no single routing policy is universally best once fuel cost and responder safety are considered alongside delivery speed; the Pareto fronts obtained here would, in a real deployment, allow disaster-response coordinators to select an operating point consistent with locally available fuel resources and acceptable responder risk tolerance, rather than being locked into a single delivery-speed-maximizing configuration.

## 5. Conclusion

This study illustrates an integrated deep reinforcement learning workflow for adaptive transportation logistics routing under natural-disaster and infrastructure-failure conditions, culminating in a proposed graph-attention multi-agent proximal policy optimization architecture (GAT-MAPPO). Benchmarked against classical routing heuristics and five DRL baselines across 500 evaluation episodes per method, GAT-MAPPO achieved the highest demand fulfillment rate (93.7%) and lowest mean delivery delay (11.2 min), retained substantially more of its performance under increasing network disruption, demand surge, and observation noise than any baseline evaluated, and generalized consistently across earthquake, flood, hurricane, and combined multi-hazard scenario types. Multi-objective analysis further demonstrated that the framework can expose, rather than obscure, the trade-offs among delivery speed, fuel cost, and responder risk that any real disaster-logistics operation must ultimately navigate. Because all data underlying this paper are synthetic, these specific numerical results should be understood as illustrative of method and reporting structure rather than as validated operational findings; the framework itself, however, defines a practical and directly transferable blueprint for a future primary-data investigation in disaster-resilient transportation logistics.

## References


1. Sutton RS, Barto AG. Reinforcement learning: an introduction. 2nd ed. Cambridge (MA): MIT Press; 2018.
2. Nazari M, Oroojlooy A, Snyder L, Takac M. Reinforcement learning for solving the vehicle routing problem. Adv Neural Inf Process Syst. 2018;31:9839-9849.
3. Lowe R, Wu Y, Tamar A, Harb J, Abbeel P, Mordatch I. Multi-agent actor-critic for mixed cooperative-competitive environments. Adv Neural Inf Process Syst. 2017;30:6379-6390.
4. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, *et al*. Human-level control through deep reinforcement learning. Nature. 2015;518(7540):529-533.
5. Van Hasselt H, Guez A, Silver D. Deep reinforcement learning with double Q-learning. Proc AAAI Conf Artif Intell. 2016;30(1):2094-2100.
6. Mnih V, Badia AP, Mirza M, Graves A, Lillicrap T, Harley T, *et al*. Asynchronous methods for deep reinforcement learning. In: Proceedings of the 33rd International Conference on Machine Learning. 2016. p. 1928-1937.
7. Schulman J, Wolski F, Dhariwal P, Radford A, Klimov O. Proximal policy optimization algorithms. arXiv [Preprint]. 2017. arXiv:1707.06347.
8. Haarnoja T, Zhou A, Abbeel P, Levine S. Soft actor-critic: off-policy maximum entropy deep reinforcement learning with a stochastic actor. In: Proceedings of the 35th International Conference on Machine Learning. 2018. p. 1861-1870.
9. Veličković P, Cucurull G, Casanova A, Romero A, Liò P, Bengio Y. Graph attention networks. In: International Conference on Learning Representations (ICLR); 2018.
10. Clarke G, Wright JW. Scheduling of vehicles from a central depot to a number of delivery points. Oper Res. 1964;12(4):568-581.
11. Kool W, van Hoof H, Welling M. Attention, learn to solve routing problems! In: International Conference on Learning Representations (ICLR); 2019.
12. Ouyang M, Duenas-Osorio L, Min X. A three-stage resilience analysis framework for urban infrastructure systems. Struct Saf. 2012;36-37:23-31.
13. Özdamar L, Ertem MA. Models, solutions and enabling technologies in humanitarian logistics. Eur J Oper Res. 2015;244(1):55-65.
14. Dijkstra EW. A note on two problems in connexion with graphs. Numer Math. 1959;1:269-271.
15. Hart PE, Nilsson NJ, Raphael B. A formal basis for the heuristic determination of minimum cost paths. IEEE Trans Syst Sci Cybern. 1968;4(2):100-107.



**How to Cite This Article**

Ogunseye OS, Brown M, Dabiri YE. Deep reinforcement learning for adaptive transportation logistics routing during natural disasters and infrastructure failures. Int J Multidiscip Futur Dev. 2026;7(2):14-25.
doi:10.54660/IJMFD.2026.7.2.14-25.