Auditable Institutional Coordination with Episodic Memory for Robust Evacuation under Multi-Source Disruption

Doctoral Consortium

Roman Sultimov
AI Center and IAI MSU, Moscow Independent Research Institute of Artificial Intelligence, NUST MISIS
Moscow, Russia
romansultimov@edu.misis.ru

ABSTRACT

Evacuation is a safety-critical multi-agent coordination problem: it is rare enough that we cannot learn from real incidents, yet the cost of error is extreme. In practice, incident commanders do not steer each vehicle; they act through *institutions*: auditable, enforceable rules such as routing guidance and access control. This dissertation studies *institutional coordination* in the tradition of social laws for multi-agent systems (MAS), grounded in physically realistic traffic simulation (SUMO/TraCI) and classical evacuation dynamics. I focus on the hard regime of *multi-source disruption* where hazards relocate and infrastructure changes simultaneously. My key idea is *episodic institutional memory*: a library of previously successful rule sets, outcomes, and contexts, enabling rapid adaptation without continual end-to-end retraining. In the current prototype, the context cue summarizes hazard origin and a small set of topology features; learning better similarity and shift detection is part of planned work. On a Simulation of Urban MObility (SUMO)-based benchmark with structured distribution shifts, an in-distribution-trained Proximal Policy Optimization (PPO) coordinator collapses under a combined hazard+topology shift (92.5% → 32.7% evacuation rate), while institution-level strategies remain stable (74.3%). Naive memory replay is unsafe under topology change (39.2%), but a lightweight topology-aware repair restores performance (92.4%) *without* gradient updates.

KEYWORDS

multi-agent systems; evacuation; episodic memory

ACM Reference Format:
Roman Sultimov. 2026. Auditable Institutional Coordination with Episodic Memory for Robust Evacuation under Multi-Source Disruption: Doctoral Consortium. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 3 pages. https://doi.org/10.65109/TCMA4845

1 MOTIVATION AND PROBLEM

Evacuations (wildfires, floods, earthquakes) stress-test multi-agent coordination: many agents compete for shared, capacity-limited infrastructure while a hazard evolves. Here, multi-source disruption refers to hazard relocation and infrastructure closures; in the thesis I will extend to additional shifts such as demand surges and partial compliance. Because real events are rare and ethically unfit for experimentation, decision procedures are typically justified through simulation and prior drills rather than online learning. In this work, I use SUMO and its Traffic Control Interface (TraCI) for physically grounded traffic dynamics [3, 4, 17, 30] and connect to broader evacuation modeling traditions [6, 13, 21].

Distribution shift is the default, not the exception. Evacuation plans assume a hazard origin and available infrastructure. Real incidents violate both: ignition can occur on the "wrong" side of the city, and key edges can be unavailable. This motivates structured shift testing (hazard-only, topology-only, and combined) and reporting robustness beyond an in-distribution mean [22, 24].

Why not just retrain? Continual end-to-end retraining after every infrastructure update, population drift, or hazard regime is operationally impractical. It requires repeated simulation runs and compute, complicating trust and certification of a black-box policy. This is a general safety concern in reinforcement learning (RL) and multi-agent learning [1, 8, 9, 32]. In contrast, institution-level rules are compact, auditable, and match how authorities act in practice.

Thesis goal. Design institutional coordination mechanisms that are (i) auditable by humans, (ii) enforceable via routing and infrastructure control, and (iii) robust under multi-source disruption, while avoiding continual end-to-end retraining.

Contributions so far is a physically grounded benchmark for institutional evacuation control with explicit hazard-origin and infrastructure-closure shifts, enabling controlled robustness testing of institution-level coordinators, evaluated by evacuation, casualty, and timeout rates and shift-specific failure modes across 4 scenarios with 20 seeds per scenario (480 episodes). Also we design an episodic institutional-memory coordinator (retrieval to topology-aware repair to validation), enabling fast adaptation under combined hazard+topology shifts where in-distribution PPO can fail, evaluated by stability under the hardest shift (92.5% to 32.7% for PPO vs. ≈ 74% for institutions) and recovery from unsafe naive memory reuse (39.2% to 92.4%) without gradient updates.

2 APPROACH

Institution interface: a small, enforceable action schema. I model a coordinator that periodically selects a bounded, interpretable institutional action rather than controlling each agent directly. In the current benchmark, the institution is a coarse routing template with `bridge` ∈ {`north`, `south`} and `exit` ∈ {`out_sw`, `out_se`}, plus `NOOP` (no-operation). Formally, an *institutional action* (or simply an *institution*) at decision time t selects $I_t \in \mathcal{I} = \{\texttt{NOOP}\} \cup (\mathcal{B} \times \mathcal{E})$, where $\mathcal{B} = \{\texttt{north}, \texttt{south}\}$ and $\mathcal{E} = \{\texttt{out\_sw}, \texttt{out\_se}\}$; applying I_t enforces the corresponding routing/permission rule set for a short

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). https://doi.org/10.65109/TCMA4845

Table 1: Mean evacuation rate (ER) across 20 scenarios.

Coordinator	ID	Hazard OOD	Topology OOD	Combined OOD
Heuristic institutions (rule-based, validated)	87.7	**74.3**	92.4	**74.3**
PPO (end-to-end)	**92.5**	68.6	**92.5**	32.7
Retrieval-only (memory replay)	87.7	72.8	39.2	72.8
Adaptive retrieval (topology repair)	87.7	44.4	92.4	73.7
LLM (schema-constrained)	89.3	46.2	92.4	**74.3**
LLM+Memory (schema-constrained)	89.1	46.5	92.4	**74.3**

horizon. Although the per-step action space is tiny ($|\mathcal{I}| = 5$), the sequence space over a 50-decision episode is $|\mathcal{I}|^{50} \approx 8.9 \times 10^{34}$, so coordination is snontrivial. Example is close/avoid the south bridge and route evacuation flow via the north bridge toward the south-east exit: a rule that a traffic management center could implement by adjusting access permissions and broadcasting detours.

Validation: feasibility as a first-class concern. Each proposed institutional action is passed via a lightweight validator that enforces schema validity and feasibility (e.g., do not route through a blocked bridge). This mirrors shield ideas in safe RL, but at the *institution layer*, where constraints are explicit and auditable [8, 15, 31, 32].

Episodic institutional memory vs. continual retraining. I maintain an episodic memory bank of successful institutions from in-distribution episodes and retrieve candidates via similarity on a compact cue vector. Two key baselines clarify the tradeoff: (i) Retrieval-Only directly replays the institution from the closest episode, and (ii) AdaptiveRetrieval applies a minimal topology-aware repair before validation: if the retrieved institution routes via a closed bridge, deterministically swap the bridge token to an available alternative and re-validate (otherwise, keep the action unchanged). This yields a direct operational contrast: *continual retraining is impractical* (costly and hard to re-certify), whereas *episodic institutional memory enables fast adaptation* via reuse plus repair.

LLM planners as institutional coordinators, not free-form oracles. LLMs are increasingly used for multi-agent planning and coordination [5, 10, 12, 33]. In my setup, an LLM can optionally propose an institutional action *only* within the same fixed schema, optionally conditioned on retrieved episodes. To keep evaluation reproducible, I cache structured decisions and replay them deterministically offline, following recent calls for robust, auditable evaluation pipelines [24]. This design connects naturally to LLM-in-traffic workflows [18, 19] and to evidence-based evacuation [7].

3 PRELIMINARY RESULTS AND INSIGHTS

Scenario suite. I evaluate one in-distribution (ID) scenario and three structured out-of-distribution (OOD) shifts: hazard origin shift, topology disruption, and their combination. Outcomes are reported as evacuation rate (ER); we also track casualty and timeout rates. These initial experiments use a single SUMO city and a compact institutional schema; extending to multiple networks and richer institutions is part of planned work [26, 27].

Episodic memory vs. retraining: the contrast is operational. Topology disruption shows why continual retraining is not the only lever. Naive episodic reuse can be actively harmful (39.2% ER): a once-good institution can route through a now-blocked bridge. A minimal topology-aware repair restores feasibility and recovers to 92.4% ER with no gradient updates, suggesting a practical "fast adaptation" path for institutional coordination.

Compound shift is where brittle generalization appears. PPO remains strong under topology-only shift, but collapses under the combined hazard+topology shift (32.7% ER), primarily through timeouts. Validated institution-level strategies provide a more stable safety floor (74.3% ER) under the same traffic dynamics. On topology and combined OOD, LLM and LLM+Memory match (Tab. 1), suggesting that the fixed scheme plus validation often dominate the final decision; I report memory usage rates to make this explicit.

Hazard relocation exposes negative transfer and similarity design. When the hazard origin shifts, memory-augmented planners can over-trust mismatched episodes and degrade (e.g., LLM+Memory: 46.5% vs. heuristic 74.3% ER). Notably, AdaptiveRetrieval under hazard-only OOD can underperform simple replay (44.4% vs. 72.8% ER) when the repair heuristic makes unnecessary edits in a topology that did not change, underscoring the need for shift-aware gating. Ablations show that increasing the retrieval threshold reduces reliance on misleading similar episodes and yields an improvement of about +9.8 percentage points on hazard-only OOD. This motivates better similarity signals and shift-aware gating.

Auditability yields better diagnosis (not just better scores). Because decisions are made at an institutional layer, I can log and analyze why a coordinator succeeded or failed (e.g., whether it invoked memory, whether repair was triggered, and whether the validator rejected an infeasible proposal). Concretely, each decision step records `(retrieved episode id, similarity score, repair edits, validation outcome)` (e.g., `t=12: retrieve ep#34 (sim=0.81) to repair south to north to valid to apply`). This complements emerging work on explanation and accountability for learning-based decision-making [16, 25, 28, 34].

4 FUTURE WORK AND FEEDBACK

Next steps include (a) *(Near-term) shift-aware memory selection:*learn similarity and gating so the coordinator retrieves only when it helps (and backs off otherwise), reducing negative transfer under hazard relocation; quantify memory usage rates and address anomalies; (b) *(Mid-term) Richer institutions, still auditable*: extend from single-step routing directives to phased controls (staging, dynamic priorities), connecting to dynamic norms and institution synthesis [2, 23, 29]; (c) *(Mid-term) Stronger validation with fast runtime:* tighten feasibility and risk checks at the institution layer, leveraging ideas from safe multi-agent learning [9, 31, 32]; (d) *(Long-term) Generalization across urban dynamics:* study transfer across changing populations and network layouts, and relate to cross-environment traffic control and unseen dynamics generalization [14, 20, 22]. I also believe some of the results might be extended to power control and emergency resource deployment [11].

Feedback from the Doctoral Consortium. I would value feedback on (i) positioning: institutions/social laws vs. safe RL vs. LLM-based planning; (ii) evaluation design for distribution shift in safety-critical MAS; (iii) the right abstraction level for institutions to remain both actionable and auditable; and (iv) principled ways to combine episodic memory, validation, and (optional) LLM proposal generation to guarantee safe degradation.

Acknowledgment. This work is supported by the Ministry of Economic Development of the Russian Federation in accordance with the subsidy agreement (agreement identifier 000000C313925P4H0002; grant No 139-15-2025-012).

REFERENCES

[1] Joshua Achiam, David Held, Aviv Tamar, and Pieter Abbeel. 2017. Constrained Policy Optimization. In *Proceedings of the 34th International Conference on Machine Learning (ICML 2017) (Proceedings of Machine Learning Research, Vol. 70)*. 22–31.

[2] Natasha Alechina, Brian Logan, and Giuseppe Perelli. 2025. Synthesising Minimum Cost Dynamic Norms. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI-25*, James Kwok (Ed.). International Joint Conferences on Artificial Intelligence Organization, 3–11.

[3] Pablo Alvarez Lopez, Angelo Banse, Martin Barthauer, Michael Behrisch, Baptiste Couéraud, Jakob Erdmann, Yun-Pang Fl"otter"od, Robert Hilbrich, Roman Nippold, and Peter Wagner. 2025. Simulation of Urban Mobility (SUMO). Zenodo.

[4] Pablo Alvarez Lopez, Michael Behrisch, Laura Bieker-Walz, Jakob Erdmann, Yun-Pang Fl"otter"od, Robert Hilbrich, Leonhard L"ucken, Johannes Rummel, Peter Wagner, and Evamarie Wießner. 2018. Microscopic Traffic Simulation using SUMO. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2575–2582.

[5] Petr Anokhin, Nikita Semenov, Artyom Sorokin, Dmitry Evseev, Andrey Kravchenko, Mikhail Burtsev, and Evgeny Burnaev. 2025. AriGraph: Learning Knowledge Graph World Models with Episodic Memory for LLM Agents. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI-25*, James Kwok (Ed.). International Joint Conferences on Artificial Intelligence Organization, 12–20.

[6] Carsten Burstedde, Kai Klauck, Andreas Schadschneider, and Johannes Zittartz. 2001. Simulation of pedestrian dynamics using a two-dimensional cellular automaton. *Physica A: Statistical Mechanics and its Applications* 295, 3-4 (2001), 507–525.

[7] Ruxiao Chen, Chenguang Wang, Yuran Sun, Xilei Zhao, and Susu Xu. 2025. From Perceptions to Decisions: Wildfire Evacuation Decision Prediction with Behavioral Theory-informed LLMs. arXiv:2502.17701 [cs.AI]

[8] Jacques Cloete, Nikolaus Vertovec, and Alessandro Abate. 2025. SPoRt - Safe Policy Ratio: Certified Training and Deployment of Task Policies in Model-Free RL. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI-25*, James Kwok (Ed.). International Joint Conferences on Artificial Intelligence Organization, 4976–4984.

[9] Aleksander Czechowski and Frans A. Oliehoek. 2023. Safe Multi-agent Learning via Trapping Regions. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23*, Edith Elkind (Ed.). International Joint Conferences on Artificial Intelligence Organization, 82–90.

[10] Minghong Geng, Shubham Pateria, Budhitama Subagdja, Lin Li, Xin Zhao, and Ah-Hwee Tan. 2025. L2M2: A Hierarchical Framework Integrating Large Language Model and Multi-agent Reinforcement Learning. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI-25*, James Kwok (Ed.). International Joint Conferences on Artificial Intelligence Organization, 99–107.

[11] Joe Gorka, Tim Hsu, Wenting Li, Yury Maximov, and Line Roald. 2024. Cascading blackout severity prediction with statistically-augmented graph neural networks. *Electric power systems research* 234 (2024), 110738.

[12] Taicheng Guo, Xiuying Chen, Yaqi Wang, Ruidi Chang, Shichao Pei, Nitesh V. Chawla, Olaf Wiest, and Xiangliang Zhang. 2024. Large Language Model Based Multi-agents: A Survey of Progress and Challenges. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI-24*, Kate Larson (Ed.). International Joint Conferences on Artificial Intelligence Organization, 8048–8057.

[13] Dirk Helbing and Péter Molnár. 1995. Social force model for pedestrian dynamics. *Physical Review E* 51, 5 (1995), 4282–4286.

[14] Haoyuan Jiang, Ziyue Li, Hua Wei, Xuantang Xiong, Jingqing Ruan, Jiaming Lu, Hangyu Mao, and Rui Zhao. 2024. X-Light: Cross-City Traffic Signal Control Using Transformer on Transformer as Meta Multi-Agent Reinforcement Learner. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI-24*, Kate Larson (Ed.). International Joint Conferences on Artificial Intelligence Organization, 94–102.

[15] Takeshi Kojima, Yaonan Zhu, Yusuke Iwasawa, Toshinori Kitamura, Gang Yan, Shu Morikuni, Ryosuke Takanami, Alfredo Solano, Tatsuya Matsushima, Akiko Murakami, and Yutaka Matsuo. 2025. A Comprehensive Survey on Physical Risk Control in the Era of Foundation Model-enabled Robotics. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI-25*, James Kwok (Ed.). International Joint Conferences on Artificial Intelligence Organization, 10517–10527.

[16] Danil Kononykhin, Mikhail Mozikov, Kirill Mishtal, Pavel Kuznetsov, Dmitrii Abramov, Nazar Sotiriadi, Yury Maximov, Andrey V Savchenko, and Ilya Makarov. 2024. From data to decisions: Streamlining geospatial operations with multimodal globeflowgpt. In *Proceedings of the 32nd ACM International Conference on Advances in Geographic Information Systems*. 649–652.

[17] Daniel Krajzewicz, Jakob Erdmann, Michael Behrisch, and Laura Bieker. 2012. Recent Development and Applications of SUMO – Simulation of Urban Mobility. *International Journal On Advances in Systems and Measurements* 5, 3&4 (2012), 128–138.

[18] Siqi Lai, Zhao Xu, Weijia Zhang, Hao Liu, and Hui Xiong. 2023. LLMLight: Large Language Models as Traffic Signal Control Agents. arXiv:2312.16044 [cs.AI]

[19] Shuyang Li, Talha Azfar, and Ruimin Ke. 2024. ChatSUMO: Large Language Model for Automating Traffic Scenario Generation in Simulation of Urban MObility. arXiv:2409.09040 [cs.HC]

[20] Yilin Liu, Guiyang Luo, Quan Yuan, Jinglin Li, Lei Jin, Bo Chen, and Rui Pan. 2023. GPLight: Grouped Multi-agent Reinforcement Learning for Large-scale Traffic Signal Control. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23*, Edith Elkind (Ed.). International Joint Conferences on Artificial Intelligence Organization, 199–207.

[21] Vahid Mahzoon, Abigail Liu, and Slobodan Vucetic. 2025. Deep Learning-Based Pedestrian Simulation with Limited Real-World Training Data: An Evaluation Framework. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI-25*, James Kwok (Ed.). International Joint Conferences on Artificial Intelligence Organization, 196–204.

[22] Monu Nagar and Debasis Das. 2025. Beyond the Map: Learning to Navigate Unseen Urban Dynamics Using Diffusion-Guided Deep Reinforcement Learning. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI-25*, James Kwok (Ed.). International Joint Conferences on Artificial Intelligence Organization, 8750–8758.

[23] Emery A. Neufeld, Agata Ciabattoni, and Radu Florin Tulcan. 2025. Combining MORL with Restraining Bolts to Learn Normative Behaviour. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI-25*, James Kwok (Ed.). International Joint Conferences on Artificial Intelligence Organization, 4615–4623.

[24] Francesco Percassi, Sandra Castellanos-Paez, Romain Rombourg, and Mauro Vallati. 2025. An Approach to Quantify Plans Robustness in Real-world Applications. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI-25*, James Kwok (Ed.). International Joint Conferences on Artificial Intelligence Organization, 8600–8607.

[25] Daniil Sukhorukov, Andrei Zakharov, Nikita Glazkov, Katsiaryna Yanchanka, Vladimir Kirilin, Maxim Dubovitsky, Roman Sultimov, Yury Maximov, and Ilya Makarov. 2026. Hierarchical AI-Meteorologist: LLM-Agent System for Multi-Scale and Explainable Weather Forecast Reporting. In *Proceedings of the AAAI 2026 Workshop on AI for Earth Sciences (AI4ES)*.

[26] Roman Sultimov, Mikhail Mozikov, Dmitrii Abramov, Mariia Kovalchuk, Maksim Malykh, Ilya Makarov, Andrei Osiptsov, Aleksandr Volkov, and Yury Maximov. 2026. RESPOND: Realistic Environment Simulation of Population and Natural Disasters with LLM-Driven Agents. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI 2026)*.

[27] Roman Sultimov, Aleksandr Volkov, Mile Mitrovic, and Yury Maximov. 2026. LLM-Guided Multi-Agent Evacuation Coordination via Episodic Memory and Cognitive Task Analysis. In *Proceedings of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026)*.

[28] Daria Taniushkina, Aleksander Lukashevich, Valeriy Shevchenko, Ilya S Belalov, Nazar Sotiriadi, Veronica Narozhnaia, Kirill Kovalev, Alexander Krenke, Nikita Lazarichev, Alexander Bulkin, and Yury Maximov. 2024. Case study on climate change effects and food security in Southeast Asia. *Scientific Reports* 14, 1 (7 2024), 16150.

[29] Alexander Tuisov, Evgeny Mishlyakov, Alexander Shleyfman, and Erez Karpas. 2025. Towards a Unified View of Social Laws with Instantaneous Actions. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI-25*, James Kwok (Ed.). International Joint Conferences on Artificial Intelligence Organization, 8661–8668.

[30] Axel Wegener, Michal Piórkowski, Maxim Raya, Horst Hellbr"uck, Stefan Fischer, and Jean-Pierre Hubaux. 2008. TraCI: An Interface for Coupling Road Traffic and Network Simulators. In *Proceedings of the 11th Communications and Networking Simulation Symposium (CNS '08)*. ACM, 155–163.

[31] Wen-Chi Yang, Giuseppe Marra, Gavin Rens, and Luc De Raedt. 2023. Safe Reinforcement Learning via Probabilistic Logic Shields. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23*, Edith Elkind (Ed.). International Joint Conferences on Artificial Intelligence Organization, 5739–5749.

[32] Weiye Zhao, Tairan He, Rui Chen, Tianhao Wei, and Changliu Liu. 2023. State-wise Safe Reinforcement Learning: A Survey. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23*, Edith Elkind (Ed.). International Joint Conferences on Artificial Intelligence Organization, 6814–6822.

[33] Zihao Zhou, Bin Hu, Chenyang Zhao, Pu Zhang, and Bin Liu. 2024. Large Language Model as a Policy Teacher for Training Reinforcement Learning Agents. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI-24*, Kate Larson (Ed.). International Joint Conferences on Artificial Intelligence Organization, 5671–5679.

[34] Rui Zuo, Simon Khan, Zifan Wang, Garrett Ethan Katz, and Qinru Qiu. 2025. Why the Agent Made that Decision: Contrastive Explanation Learning for Reinforcement Learning. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI-25*, James Kwok (Ed.). International Joint Conferences on Artificial Intelligence Organization, 655–663.