LLM-Guided Multi-Agent Evacuation Coordination via Episodic Memory and Cognitive Task Analysis

Demonstration Track

Roman Sultimov
Lomonosov Moscow State University
Moscow, Russia

Aleksandr Volkov
Interdata
Astana, Kazakhstan

Mile Mitrovic
SB AI Lab
Moscow, Russia

Yury Maximov
Interdata
Astana, Kazakhstan

ABSTRACT

We demonstrate [1] an LLM-guided multi-agent platform for coordinating evacuations during wildfires. A lightweight *Commander* agent observes traffic congestion and wildfire spread in a simulated environment and issues high-level routing decisions that mimic emergency management actions. *User* agents receive these decisions, locally observe nearby behavior, and choose how to evacuate. We design the Commander to be safe, interpretable, and intended to transfer across settings (number of agents, geographies, wildfire scenarios). On a synthetic city benchmark with 150 user agents, our reinforcement learning policy achieves strong in-distribution performance (92.6% evacuation rate) but degrades under a zero-shot fire shift to 67.1%. The LLM Commander improves the zero-shot evacuation rate to 74.2% without memory, and up to 82.5% with episodic memory, mainly by reducing agent timeouts. We also introduce a cognitive task analysis perspective that makes the Commander decision making process fully auditable by humans. Our interactive demo illustrates these results and supports debugging, ablations, and future learning for both AI systems and human operators.

KEYWORDS

LLM agents; multi-agent reinforcement learning; evacuation

ACM Reference Format:
Roman Sultimov, Aleksandr Volkov, Mile Mitrovic, and Yury Maximov. 2026. LLM-Guided Multi-Agent Evacuation Coordination via Episodic Memory and Cognitive Task Analysis: Demonstration Track. 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/SEYR5220

1 INTRODUCTION

Large-scale wildfire evacuation is a high-stakes, data-scarce problem with a small number of historical events highly varying across time and geography available. As a result, real emergency management systems often operate in zero- or few-shot settings and must rely on training and testing in synthetic environments. This also creates a key requirement: methods should adapt to distribution shifts (e.g., a new ignition location or a different congestion pattern) while remaining efficient and interpretable to human experts [25].

Our system uses a two-level architecture. First, each vehicle follows a learned stochastic reinforcement learning policy. Second, an LLM-based *commander* performs sparse coordination: it observes congestion and wildfire spread and occasionally issues broad, network-level routing rules [16, 17].

To support human interpretability, we add a cognitive task analysis (CTA) perspective of the commander: for each commander rule, we log key hints behind decision (e.g., fire side, bridge congestion, etc) and summarize them as simple rules of decision tree.

2 RELATED WORK AND CONTRIBUTION

Large-scale evacuation planning has been widely studied using optimization and simulation-based methods, including approaches that reduce route congestion and methods for routing and scheduling under disasters [5, 11, 13]. Recent learning-based approaches also coordinate road conditions adaptation [30] while performance under hazard shifts (e.g. new fire locations) is in question.

LLMs have recently been successfully used as high-level tool for grounded decision making [10], and as software frameworks for multi-agent conversation and tool use [7, 14, 29]. Episodic memory mechanisms are essential to improve recall and adaptation in reinforcement learning [2, 22, 27] and are increasingly seen as a core capability for agents built with LLMs [8, 15, 31]. Finally, cognitive task analysis and the critical decision method can elicit expert hints and turn them into an explicit decision structure [3, 4, 9].

Our demonstration makes a practical contribution: a reproducible pipeline that combines (1) a strong reinforcement learning baseline based on proximal policy optimization, (2) LLM commander mimicking emergency response, and (3) an a combination of episodic memory with cognitive task analysis (CTA) resulting in human-interpretable decision rules that can be employed by practitioners to improve evacuation procedures and save human lives.

3 METHODOLOGY AND RESULTS

Methodology. We simulate evacuation in SUMO [1] on a non-orthogonal city given by a road network with two scenarios: "Fire on the West" and "Fire on the East" . Each episode ends when all agents are either evacuated, become casualties, or time out.

[1]Demonstration video: https://youtu.be/oliDIjvz_Sw

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/SEYR5220

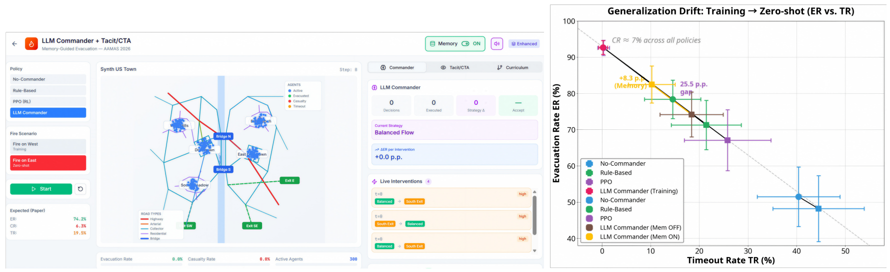

Figure 1: Left: Demo user interface showing the Commander policies, executed actions, and retrieved episodes. Right: generalization change in ER-TR space (training $T \Rightarrow$ zero-shot Z); episodic memory reduces timeouts for the LLM policy.

Table 1: Main results on Synth US Town (mean ± std over 20 random seeds; 150 agents/episode). Fire-on-West is the training distribution; Fire-on-East is zero-shot. *Mem ON* is evaluated only over "Fire on the East".

Policy	ER (Train)	CR (Train)	TR (Train)	ER (Zero-shot)	CR (Zero-shot)	TR (Zero-shot)
No-Commander (Random)	51.5 ± 8.2	7.3 ± 1.1	41.2 ± 7.8	48.2 ± 9.1	7.5 ± 1.3	44.3 ± 8.5
Rule-Based	78.4 ± 5.3	**7.2 ± 0.9**	14.4 ± 4.8	71.3 ± 6.8	7.4 ± 1.0	21.3 ± 6.2
PPO (RL)	92.6 ± 2.1	7.4 ± 0.8	**0.0 ± 0.0**	67.1 ± 8.4	7.6 ± 1.1	25.3 ± 7.9
LLM Commander (Memory Off)	**92.7 ± 1.9**	7.3 ± 0.7	**0.0 ± 0.0**	74.2 ± 6.2	7.4 ± 0.9	18.4 ± 5.8
LLM Commander (Memory On)	–	–	–	**82.5 ± 4.1**	**7.3 ± 0.8**	**10.2 ± 3.9**

We compare four policies: (a) no commander (Random), (b) rule Based (static heuristics), (c) proximal policy optimization (PPO, trained over "Fire on the West" data) [6], and (d) LLM Commander standing for PPO with sparse rules (i.e. accidental road closures). We ablate the commander by enabling/disabling episodic memory; CTA rules are logged for interpretability and do not affect control. For Memory On regime, the memory bank is populated only from episodes in the training scenario ("Fire on the West") and does not include any episodes from the zero-shot test distribution. Rule-based static heuristic that routes all agents away from the fire origin (use the south bridge if the fire is on the west side; use the north bridge if the fire is on the east side). No runtime adaptation.

We adapt reinforcement learning episodic memory mechanisms to retrieve prior strategies under similar directions [2, 22, 26]. Each episode produces a summary vector $e \in \mathbb{R}^d$ encoding scenario data (fire location, congestion patterns) and outcome metrics [19]. The memory stores tuples $m_i = (e_i, s_i, o_i)$, where s_i is the strategy (such as Balanced-Flow, South-Exit-Bias) and o_i the outcome. Given a current situation embedding e_t, we retrieve top-k episodes by cosine similarity, $\mathrm{sim}(e_t, e_i)$. The commander conditions its next intervention on retrieved episodes $\mathcal{N}_k(e_t)$, preferring strategies with higher observed evacuation rate and lower timeout rate under similar cues. We expose the commander internal decision factors as a small set of interpretable rules [12, 24].

We compute a rule-weighted explanation score $Q_{\mathrm{CTA}}(\sigma \mid s_t) = \sum_j w_j\, c_j(\sigma, s_t)$ to summarize which rules supports the selected strategy σ_t. The LLM commander is high-level strategy selector analyzing the current state, retrieves similar episodes from memory and selecting a strategy based on the LLM-generated rationale.

4 RESULTS

We evaluate 20 random episodes with 150 agents each. We report three per-episode agent rates: Evacuation Rate (ER) (percent evacuated), Casualty Rate (CR) (percent killed), and Timeout Rate (TR) (percent not evacuated before the time limit). By definition, ER + CR + TR = 100%. Figure 1 (right) shows generalization drift in ER–TR space from training to zero-shot evaluation, and Table 1 reports mean±std across runs. PPO and the LLM Commander perform similarly during training (ER 92.6–92.7%) but both drop in zero-shot. The LLM Commander generalizes better than PPO (ER 74.2% vs. 67.1%), mainly by reducing timeouts (TR 18.4% vs. 25.3%); CR stays nearly constant at 7–8%. Adding episodic memory (zero-shot only) further improves ER to 82.5% and reduces TR to 10.2%, supporting the claim that memory retrieval and explicit strategy updates matter under distribution shift. In this benchmark, the Casualty Rate (CR) changes only slightly across policies and scenarios (Table 1). As a result, most performance differences show up as a near one-to-one TR → ER shift in the ER-TR plane, Fig. 1. CR is nearly policy-independent: ignoring the fire changes CR only from 7.1% to 7.3%, a 16× faster fire changes CR by at most 1.4%, and all casualties occur within ≤ 7 steps (median 0), before coordination can help. Thus, in Synth US Town, CR is driven mainly by initial proximity to ignition, while policy quality mostly reduces congestion and timeouts (TR). Richer hazard models (e.g., delayed ignition) may make CR more policy-dependent [18, 20, 21, 23, 28].

5 DEMO INTERFACE

The demo provides (i) a live map view with wildfire and congestion dynamics, (ii) the commander rules log and acceptance rate by the users, (iii) episodic memory retrieval of similar past episodes, and (iv) CTA-derived rules, rationales, and decision tree interpretability.

6 CONCLUSION

Since fatalities mainly occur before coordination can take effect, policy gains in this benchmark appear almost entirely as TR→ER conversion; episodic memory moves performance along this trade-off under hazard shift. We present a memory-guided LLM Commander for multi-agent evacuation that improves zero-shot performance under hazard shifts, along with a reproducible pipeline and an interpretable interface for analyzing high-level coordination decisions.

ACKNOWLEDGMENT

The work of Roman Sultimov 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] Michael Behrisch, Laura Bieker, Jakob Erdmann, and Daniel Krajzewicz. 2011. SUMO–simulation of urban mobility: an overview. In *Proceedings of SIMUL 2011, The Third International Conference on Advances in System Simulation*. ThinkMind.

[2] Charles Blundell, Benigno Uria, Alexander Pritzel, Yazhe Li, Avraham Ruderman, Joel Z. Leibo, Jack Rae, Daan Wierstra, and Demis Hassabis. 2016. Model-Free Episodic Control. https://doi.org/10.48550/arXiv.1606.04460 arXiv:1606.04460 [cs.LG]

[3] Olivia Brown, Nicola Power, and Julie Gore. 2025. Cognitive Task Analysis: Eliciting Expert Cognition in Context. *Organizational Research Methods* 28, 3 (2025), 375–404. https://doi.org/10.1177/10944281241271216

[4] Nancy Cooke. 1994. Varieties of knowledge elicitation techniques. *International Journal of Human-Computer Studies* 41, 6 (1994), 801–849. https://doi.org/10.1006/ijhc.1994.1083

[5] Caroline Even, Victor Pillac, and Pascal Van Hentenryck. 2015. Convergent Plans for Large-Scale Evacuations. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI-15)*. 1121–1127. https://dl.acm.org/doi/10.5555/2887007.2887163

[6] Yang Gu, Yuhu Cheng, CL Philip Chen, and Xuesong Wang. 2021. Proximal policy optimization with policy feedback. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 52, 7 (2021), 4600–4610.

[7] Taicheng Guo, Xiyan Hu, Xingzhe Yu, Wenqiao Wang, Kevin Liao, Zhiming Chen, Jun Nie, and Hongxuan Wang. 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)*. 8048–8057. https://doi.org/10.24963/ijcai.2024/890

[8] Kostas Hatalis, Despina Christou, Joshua Myers, Steven Jones, Keith Lambert, Amanda Amos-Binks, Zachary Dannenhauer, and Dustin Dannenhauer. 2024. Memory Matters: The Need to Improve Long-Term Memory in LLM-Agents. In *Proceedings of the AAAI Symposium Series (Fall 2023)*, Vol. 2. 277–280. https://doi.org/10.1609/aaaiss.v2i1.27688

[9] Robert R. Hoffman, Beth Crandall, and Nigel Shadbolt. 1998. Use of the Critical Decision Method to Elicit Expert Knowledge: A Case Study in the Methodology of Cognitive Task Analysis. *Human Factors* 40, 2 (1998), 254–276. https://doi.org/10.1518/001872098779480442

[10] brian ichter, Anthony Brohan, Yevgen Chebotar, Chelsea Finn, Karol Hausman, Alexander Herzog, Daniel Ho, Julian Ibarz, Alex Irpan, Eric Jang, Ryan Julian, Dmitry Kalashnikov, Sergey Levine, Yao Lu, Carolina Parada, Kanishka Rao, Pierre Sermanet, Alexander T Toshev, Vincent Vanhoucke, Fei Xia, Ted Xiao, Peng Xu, Mengyuan Yan, Noah Brown, Michael Ahn, Omar Cortes, Nicolas Sievers, Clayton Tan, Sichun Xu, Diego Reyes, Jarek Rettinghouse, Jornell Quiambao, Peter Pastor, Linda Luu, Kuang-Huei Lee, Yuheng Kuang, Sally Jesmonth, Nikhil J. Joshi, Kyle Jeffrey, Rosario Jauregui Ruano, Jasmine Hsu, Keerthana Gopalakrishnan, Byron David, Andy Zeng, and Chuyuan Kelly Fu. 2023. Do As I Can, Not As I Say: Grounding Language in Robotic Affordances. In *Proceedings of The 6th Conference on Robot Learning (Proceedings of Machine Learning Research, Vol. 205)*, Karen Liu, Dana Kulic, and Jeff Ichnowski (Eds.). PMLR, 287–318. https://proceedings.mlr.press/v205/ichter23a.html

[11] Md. Rifat Islam, Fangfei Chen, Achla Marathe, Henning Mortveit, Samarth Swarup, and Anil Kumar S. Vullikanti. 2024. Strategic Routing and Scheduling for Evacuations: Extended Abstract. In *Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS '24)*. International Foundation for Autonomous Agents and Multiagent Systems, 2312–2314. https://ifaamas.org/Proceedings/aamas2024/pdfs/p2312.pdf

[12] 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.

[13] Chris Kuhlman, Achla Marathe, Anil Kumar S. Vullikanti, Fajar Halim, and Pitu Mozumder. 2020. Increasing Evacuation During Disaster Events: A Multiobjective Artificial Society Approach. In *Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS '20)*. International Foundation for Autonomous Agents and Multiagent Systems, 654–662. https://dl.acm.org/doi/10.5555/3398761.3398840

[14] Xinyi Li, Sai Wang, Siqi Zeng, Yu Wu, and Yi Yang. 2024. A Survey on LLM-Based Multi-Agent Systems: Workflow, Infrastructure, and Challenges. *Vicinagearth* 1, 1 (2024), 1–25. https://doi.org/10.1007/s44336-024-00009-2

[15] Yitao Liu, Chenglei Si, Karthik R. Narasimhan, and Shunyu Yao. 2025. Contextual Experience Replay for Self-Improvement of Language Agents. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 14179–14198. https://doi.org/10.18653/v1/2025.acl-long.694

[16] Aleksander Lukashevich, Vyacheslav Gorchakov, Petr Vorobev, Deepjyoti Deka, and Yury Maximov. 2023. Importance sampling approach to chance-constrained dc optimal power flow. *IEEE transactions on control of network systems* 11, 2 (2023), 928–937.

[17] Aleksander Lukashevich and Yury Maximov. 2021. Power grid reliability estimation via adaptive importance sampling. *IEEE Control Systems Letters* 6 (2021), 1010–1015.

[18] Alexander Marusov, Vsevolod Grabar, Yury Maximov, Nazar Sotiriadi, Alexander Bulkin, and Alexey Zaytsev. 2024. Long-term drought prediction using deep neural networks based on geospatial weather data. *Environmental Modelling & Software* 179 (2024), 106127.

[19] Mile Mitrovic, Aleksandr Lukashevich, Petr Vorobev, Vladimir Terzija, Semen Budennyy, Yury Maximov, and Deepjyoti Deka. 2023. Data-driven stochastic AC-OPF using Gaussian process regression. *International journal of electrical power & energy systems* 152 (2023), 109249.

[20] Mikhail Mozikov, Ivan Lukyanenko, Ilya Makarov, Alexander Bulkin, and Yury Maximov. 2023. Long-term hail risk assessment with deep neural networks. In *International Work-Conference on Artificial Neural Networks*. Springer, 288–301.

[21] Polina Pilyugina, Timofey Chernikov, Maria Smirnova, Alexey Zaytsev, Alexander Bulkin, Evgeny Burnaev, Ilya S Belalov, Nazar Sotiriadi, Albert Efimov, Yury Maximov, et al. 2025. A physics-informed machine learning framework for permafrost stability assessment. *IEEE Access* (2025).

[22] Samuel Ritter, Jane X. Wang, Zeb Kurth-Nelson, Siddhant M. Jayakumar, Charles Blundell, Razvan Pascanu, and Matthew Botvinick. 2018. Been There, Done That: Meta-Learning with Episodic Recall. https://doi.org/10.48550/arXiv.1805.09692 arXiv:1805.09692 [cs.LG]

[23] Valeriy Shevchenko, Aleksandr Lukashevich, Daria Taniushkina, Alexander Bulkin, Roland Grinis, Kirill Kovalev, Veronika Narozhnaia, Nazar Sotiriadi, Alexander Krenke, and Yury Maximov. 2024. Climate change impact on agricultural land suitability: An interpretable machine learning-based Eurasia case study. *IEEE access* 12 (2024), 15748–15763.

[24] 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), Poster Track*.

[25] Roman Sultimov. 2026. Auditable Institutional Coordination with Episodic Memory for Robust Evacuation under Multi-Source Disruption. (May 2026). Doctoral Consortium. To appear.

[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] Egor Suraveikin, Dastan Omirzak, Roman Sultimov, and Yury Maximov. 2026. Efficient Contextual Bandit Learning via Reward-Space Sampling and Online Optimization. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI 2026)*.

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

[29] Qingyun Wu, Gagan Bansal, Jieyu Zhang, Yiran Wu, Shaokun Zhang, Erkang Zhu, Beibin Li, Li Jiang, Xiaoyun Zhang, and Chi Wang. 2023. AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation Framework. https://doi.org/10.48550/arXiv.2308.08155 arXiv:2308.08155 [cs.AI]

[30] Chuan-Zhi Thomas Xie, Qihua Chen, Bin Zhu, Eric Wai Ming Lee, Tie-Qiao Tang, Xianfei Yin, Zhilu Yuan, and Botao Zhang. 2025. Coordinating dynamic signage for evacuation guidance: A multi-agent reinforcement learning approach integrating mesoscopic crowd modeling and fire propagation. *Chaos, Solitons & Fractals* 194 (2025), 116246. https://doi.org/10.1016/j.chaos.2025.116246

[31] Wujiang Xu, Zujie Liang, Ming Zhang, Sheng Zhang, and Xin Zhang. 2025. A-MEM: Agentic Memory for LLM Agents. https://doi.org/10.48550/arXiv.2502.12110 arXiv:2502.12110 [cs.AI]