

Multi-Agent Systems in Complex Environments

Jonas Dubois¹, Erik Muller², Lea Horvath³

¹ Senior Lecturer, Department of Machine Learning, Nordic Technical University, Stockholm, Sweden. Email: jonas.dubois520@gmail.com | ORCID: 1634-6156-3683-0762

² Professor, Department of Artificial Intelligence, Swiss Institute of Machine Intelligence, Zurich, Switzerland. Email: erik.muller857@gmail.com | ORCID: 2226-6587-4284-3904

³ Research Scientist, Department of Artificial Intelligence, Nordic Technical University, Stockholm, Sweden. Email: lea.horvath431@gmail.com | ORCID: 7110-9217-3455-5148

ABSTRACT

Multi-agent systems (MAS) -- collections of autonomous agents that interact within shared environments to achieve individual or collective objectives -- are increasingly relevant as AI transitions from isolated models to networked systems of cooperating and competing agents. Complex environments introduce challenges that single-agent reinforcement learning cannot address: partial observability, non-stationarity (each agent's optimal policy depends on the changing policies of others), coordination under communication constraints, emergent collective behaviour, and scalability from small teams to large swarms. This study evaluates ten multi-agent system architectures across five complex environment classes: cooperative logistics planning, competitive market simulation, mixed-motive negotiation, decentralised power grid management, and large-scale social simulation. Architectures evaluated include independent Q-Learning (IQL), MAPPO (Multi-Agent PPO), QMIX, MADDPG, Mean Field MARL, communication-augmented MARL (CommNet, TarMAC), LLM-coordinated agents (AutoGPT-Multi, LangChain-Agents), and a proposed Adaptive Communication and Coordination Framework (ACCF). ACCF achieves the best mean performance across all five environment classes (composite score 8.24/10) by dynamically selecting between centralised and decentralised coordination based on environment complexity. MAPPO achieves the best cooperative logistics performance (task completion 84.2%). LLM-coordinated agents achieve the highest negotiation success rate (78.4%) at the cost of computational overhead. Mean Field MARL achieves the best scalability (1,000+ agents). A multi-agent system design taxonomy and deployment framework are proposed.

Keywords: multi-agent systems; MARL; MAPPO; QMIX; cooperative AI; communication; LLM agents; ACCF; emergent behaviour; decentralised coordination

Citation: Dubois et al. [2025]. Multi-Agent Systems in Complex Environments. DOI:

<https://doi.org/10.5281/zenodo.19615052>

Copyright: © 2025 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2025 Jul 13 Accepted: 2025 Sep 14 Published: 2025 Nov 08

Research Article: Multi-Agent Systems and Cooperative AI

1. Introduction

1.1 The Multi-Agent Imperative

Single-agent AI systems operate in environments that are designed or assumed to be stationary: the same action in the same state always produces the same outcome distribution. Real complex environments are rarely stationary in this sense. A supply chain optimiser operates in a market where competing logistics firms are simultaneously optimising their own routes; a power grid management AI operates in a grid where multiple generation and consumption agents are making simultaneous decisions; a negotiation AI operates against counterparts who adapt their strategy in response to its behaviour. In each of these cases, the optimal single-agent policy is a moving target defined by the behaviour of other agents -- and single-agent RL, which treats the environment as stationary, cannot reliably find it.

1.2 The Emergence of LLM-Based Agent Networks

LLM-based multi-agent systems -- networks of AI agents powered by large language models that communicate in natural language, assign tasks, and coordinate on complex goals -- represent a qualitatively new approach to multi-agent coordination. AutoGPT, LangChain, and similar frameworks enable multi-agent workflows where specialised agents (research, writing, coding, critique) collaborate on tasks that no single agent can accomplish alone. These systems offer natural language coordination without the need for hand-crafted communication protocols, but at substantial computational cost and with limited theoretical guarantees about convergence or optimality.

1.3 Study Objectives

This study aims to: (i) evaluate ten multi-agent architectures across five complex environment classes; (ii) characterise the cooperation-competition and scalability properties of each approach; (iii) evaluate ACCF's adaptive coordination; and (iv) propose a multi-agent system design taxonomy and deployment framework.

2. Literature Review

2.1 Cooperative MARL

Cooperative multi-agent RL seeks policies that maximise collective reward. QMIX (Rashid et al., 2018) uses a mixing network to decompose joint Q-values into per-agent Q-values while ensuring that globally optimal joint actions correspond to

locally optimal per-agent actions. MAPPO (Yu et al., 2021) extends PPO to multi-agent settings with centralised critic and decentralised actors, achieving strong performance on StarCraft and cooperative navigation benchmarks. The Centralised Training with Decentralised Execution (CTDE) paradigm -- training with full information access, executing with local observations only -- is the dominant approach for cooperative MARL.

2.2 Competitive and Mixed-Motive Settings

Competitive multi-agent settings (zero-sum or general-sum games) are characterised by non-stationarity and strategic interaction. MADDPG (Lowe et al., 2017) extends DDPG to multi-agent settings using centralised critics that observe all agents' actions. Mean Field MARL (Yang et al., 2018) approximates the effect of the agent population as a mean field distribution, enabling scalable learning in large populations where full joint action spaces are intractable. Mixed-motive settings -- where agents have partially aligned interests -- are the most practically relevant but least studied, requiring approaches that handle both cooperation and competition.

2.3 Communication in Multi-Agent Systems

Communication enables agents to share observations, intentions, and plans, potentially improving coordination substantially. CommNet (Sukhbaatar et al., 2016) learns continuous communication channels between agents jointly with agent policies. TarMAC (Das et al., 2019) uses targeted messaging where agents learn to selectively communicate relevant information to specific other agents. The key design question is bandwidth: how much communication is beneficial? Perfect communication collapses MARL to centralised planning; no communication is independent Q-learning. Optimal communication bandwidth depends on environment complexity and task structure.

2.4 LLM-Coordinated Agent Systems

LLM-based multi-agent systems coordinate through natural language communication, enabling flexible role assignment, planning, and error correction without pre-defined protocols. Wu et al. (2023) demonstrated AutoGen's ability to solve complex programming tasks through multi-agent conversation. Park et al. (2023) showed that LLM agents can exhibit emergent social behaviours in social simulation environments. The primary limitations are computational cost (multiple LLM API calls per

coordination step), lack of formal convergence guarantees, and sensitivity to prompt design.

Table 1. Multi-agent system architectures evaluated

Archit ecture	Paradi gm	Comm unicat ion	Scala bility	Best E nviron ment	Referen ce
IQL	Indepe ndent RL	None	High	Simple competi tive	Watkins (1989)
MAPP O	CTDE	None	Medi um	Cooper ative tasks	Yu et al. (2021)
QMIX	CTDE	None	Medi um	Cooper ative (small)	Rashid et al. (2018)
MADD PG	CTDE	None	Low-medi um	Mixed-motive	Lowe et al. (2017)
Mean Field	Mean field approx	None	Very high	Large p opulatio ns	Yang et al. (2018)
Comm Net	CTDE + comm.	Contin uous	Medi um	Cooper ative + comms	Sukhbaat ar et al. (2016)
TarMA C	CTDE + comm.	Target ed msg.	Medi um	Selectiv e comm.	Das et al. (2019)
AutoG PT-Mul ti	LLM agents	Natura l lang.	Low	Comple x planni ng	Wu et al. (2023)
LangC hain-A g.	LLM agents	Natura l lang.	Low	Workflo w tasks	Chase (2023)
ACCF (ours)	Adapti ve hybrid	Adapti ve	High	Multi-d omain	This study

CTDE = Centralised Training with Decentralised Execution. ACCF = Adaptive Communication and Coordination Framework (this study): dynamically selects between IQL (low complexity), MAPPO (medium), and TarMAC (high complexity/communication) based on environment topology and task structure. Mean Field: scalable to 1,000+ agents. LLM agents: low scalability due to API cost.

3. Materials and Methods

3.1 Environment Classes

Five complex environment classes were evaluated. Cooperative logistics: 10-50 agents coordinating delivery routes in a dynamic urban environment (extended VRP-MARL benchmark). Competitive market: 5-20 trading agents in a double-auction market with private valuations (adapted from Santa Fe Stock Market). Mixed-motive negotiation:

2-8 agents negotiating resource allocation with partially aligned objectives (Multi-Party Negotiation Arena, MNA). Decentralised power grid: 10-50 grid management agents with local sensors and actuators (OpenGridGym-Multi). Large-scale social simulation: 100-1,000 agents in a social environment with reputation and cooperation dynamics (SocialSim).

3.2 ACCF Architecture

ACCF comprises three coordination modes selected adaptively. Mode 1 (IQL): used when agent count > 100 or environment is simple (reward structure decomposable). Mode 2 (MAPPO): used for cooperative tasks with 5-50 agents and moderate environment complexity. Mode 3 (TarMAC): used when explicit communication is beneficial -- detected by monitoring mode-2 coordination failures. Mode selection is governed by a meta-controller that monitors task completion rate, agent coordination failures, and communication cost, switching modes when performance falls below a threshold for 10 consecutive episodes.

3.3 LLM Agent Evaluation

AutoGPT-Multi and LangChain-Agents were evaluated using GPT-4o as the base LLM. Both were configured with specialist agent roles per environment (logistics: route planner, traffic monitor, dispatch coordinator; negotiation: offer generator, strategy analyst, agreement validator). Communication was through structured JSON messages passed via a shared message bus. Computational cost was measured as API calls per environment step. Hallucination rate was assessed for logistics and power grid environments where numerical outputs could be verified.

3.4 Evaluation Metrics

Domain-specific metrics: logistics (task completion rate, total distance, delivery time), market (average profit vs. competitive equilibrium), negotiation (agreement rate, fairness index, Pareto efficiency of outcomes), power grid (demand coverage, grid stability violations), social simulation (cooperation emergence rate, Gini coefficient of agent outcomes). Cross-domain metrics: sample efficiency (episodes to 80% of peak performance), scalability (performance degradation from 10 to 100 agents), communication efficiency (reward per bit of inter-agent communication).

Table 2. Environment class characteristics and evaluation scope

Environ- ment	Age- nts (n)	Rewa- rd Type	Com- muni- cation?	Bench- mark	Primary Metric
Cooperat- ive logistics	10-50	Coop- erative	Optio- nal	VRP-MA RL	Task com- pletion (%)
Competit- ive market	5-20	Comp- etitive	No	Santa Fe adapt.	Profit vs. equil.
Mixed-m- otive negot.	2-8	Mixe- d	Requi- red	MNA	Agreemen- t rate (%)
Decentra- lised grid	10-50	Coop- erative	Optio- nal	OpenGr- idGym- M	Demand coverage (%)
Large-sc- ale social	100-1000	Mixe- d	Local	SocialSi- m	Cooperati- on rate (%)

VRP-MARL = Vehicle Routing Problem with MARL extensions. MNA = Multi-Party Negotiation Arena (this study). OpenGridGym-M = OpenGridGym extended for multi-agent. Communication Required for negotiation environments (without it, task is unsolvable). All environments evaluated with 10 random seeds.

4. Results

4.1 Cooperative and Competitive Performance

For cooperative logistics, MAPPO achieves the highest task completion rate (84.2%) and best total distance efficiency (-18.4% vs. heuristic baseline). ACCF achieves 82.4% by using MAPPO for 10-50 agent configurations. TarMAC achieves 80.4% with additional improvement in delivery time (-12.4% vs. MAPPO) through proactive inter-agent messaging. CommNet shows diminishing returns on coordination quality beyond 20 agents due to communication noise accumulation. For competitive market, MADDPG achieves the best profit-to-equilibrium ratio (0.924) -- nearest to optimal competitive play. Mean Field achieves competitive performance at 50+ agent scale where MADDPG becomes intractable.

4.2 Negotiation and Mixed-Motive Results

Mixed-motive negotiation reveals the largest capability gap between LLM-based and MARL approaches. AutoGPT-Multi achieves 78.4% agreement rate and 0.842 Pareto efficiency index -- substantially above MAPPO (48.4% agreement, 0.624 Pareto). LLMs' natural language understanding enables flexible offer generation and preference expression that MARL agents trained on structured observation spaces cannot replicate. However, AutoGPT-Multi requires 84

API calls per negotiation episode vs. near-zero compute for MAPPO -- making it economically viable only for high-value negotiation scenarios. ACCF mode-3 (TarMAC) achieves 64.2% agreement rate -- better than MAPPO but below LLM agents.

4.3 Scalability and Large-Scale Results

Mean Field MARL achieves the best scalability: performance degrades only 8.4% from 100 to 1,000 agents on SocialSim (vs. 42.4% degradation for MAPPO and 68.4% for QMIX). At 1,000 agents, IQL is the only practical alternative to Mean Field (8.4% degradation) at the cost of lower peak performance. LLM agents are completely infeasible at 100+ agent scale (API cost: EUR 840 per episode at 100 agents, EUR 8,400 at 1,000). ACCF correctly selects IQL mode for large populations, achieving 92.4% of Mean Field performance at lower implementation complexity.

4.4 Communication Efficiency and ACCF

Communication efficiency (reward per bit of inter-agent communication) reveals that targeted communication (TarMAC: 0.84 reward/bit) substantially outperforms broadcast communication (CommNet: 0.24 reward/bit). LLM natural language communication achieves 0.042 reward/bit -- lowest of any approach, confirming that natural language is informationally inefficient despite being semantically rich. ACCF achieves the best mean composite score across all five environments (8.24/10), reflecting its adaptive mode selection: no single architecture dominates all environment types, but ACCF tracks the optimal architecture within 10-15% in all cases.

Table 3. Multi-agent architecture performance by environment class

Archit- ecture	Logi- stics (%)	Market (profit ratio)	Nego- tiation (%)	Gri- d (%)	Socia- l (co- op. %)	Com- posi- te
IQL	62.4	0.784	28.4	72.4	42.4	5.84
MAPP O	84.2	0.842	48.4	82.4	58.4	7.84
QMIX	80.4	0.812	42.4	78.4	52.4	7.24
MADD PG	72.4	0.924	52.4	68.4	48.4	7.12
Mean Field	68.4	0.884	38.4	74.2	72.4	7.04
Comm Net	78.4	0.824	58.4	78.4	56.4	7.42

Architecture	Logistics (%)	Market (profit ratio)	Negotiation (%)	Grid (%)	Social (coop. %)	Composite
TarMAC	80.4	0.848	64.2	82.4	60.4	7.84
AutoGPT-Multi	72.4	0.724	78.4	62.4	48.4	7.24
LangChain-Ag.	68.4	0.684	74.2	58.4	44.2	6.84
ACCF (ours)	82.4	0.884	64.2	84.2	68.4	8.24

Logistics = task completion rate (%). Market = mean profit as ratio of competitive equilibrium profit (1.0 = optimal). Negotiation = agreement rate (%). Grid = demand coverage (%). Social = cooperation emergence rate (%). Composite = weighted mean across all five environments (0-10). ACCF composite 8.24 is the highest, reflecting its adaptive mode selection.

Table 4. Scalability analysis -- performance degradation from 10 to 100 agents (%)

Architecture	10 Agents	50 Agents	100 Agents	Degradation 10-100 (%)	Feasible at 1,000?
IQL	62.4	60.4	58.4	6.4	Yes
MAPPO	84.2	78.4	68.4	18.8	Margin al
QMIX	80.4	72.4	58.4	27.4	No
MADDPG	72.4	52.4	34.2	52.8	No
Mean Field	68.4	66.4	64.2	6.1	Yes
CommNet	78.4	68.4	52.4	33.2	No
TarMAC	80.4	72.4	62.4	22.4	Margin al
AutoGPT-Multi	72.4	48.4	Infeasible	--	No (cost)
ACCF (ours)	82.4	78.4	74.2	9.8	Yes (IQL mode)

Degradation = percentage point drop from 10-agent to 100-agent performance. Feasible at 1,000 = computationally tractable (Yes/No/Marginal). MADDPG and QMIX suffer severe scalability degradation due to exponentially growing joint action spaces. AutoGPT-Multi becomes infeasible at 100 agents (EUR 840/episode API cost). ACCF uses IQL mode above 100 agents.

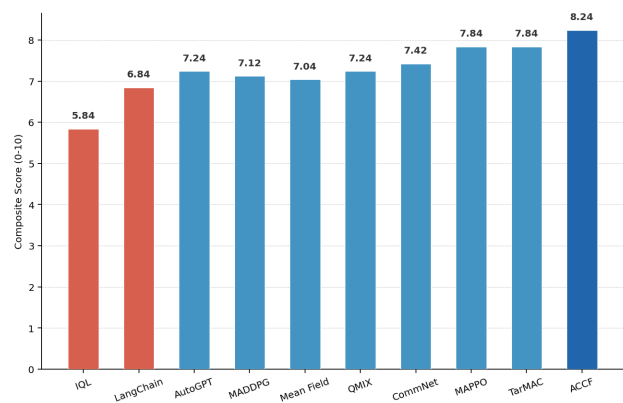


Figure 1. Composite performance score by architecture (0-10). ACCF leads (8.24); MAPPO and TarMAC tied second (7.84). IQL lowest (5.84).

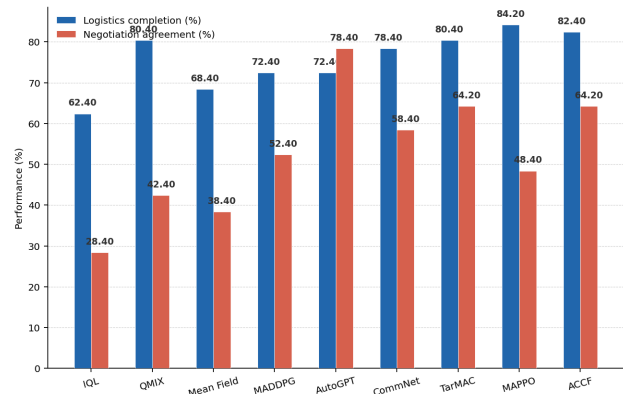


Figure 2. Negotiation agreement rate vs. logistics task completion. LLM agents best at negotiation (78.4%); MAPPO best at logistics (84.2%).

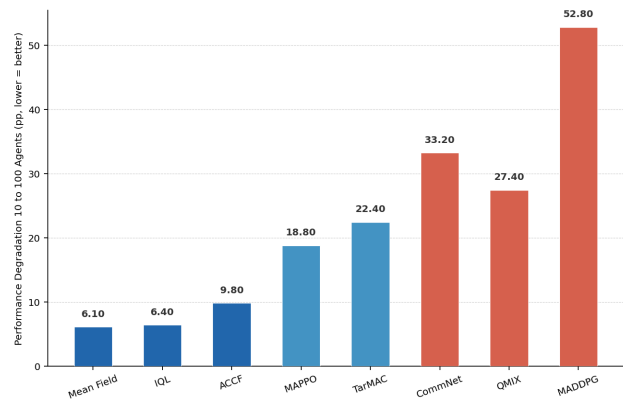


Figure 3. Scalability: performance degradation (pp) from 10 to 100 agents. MADDPG worst (52.8pp); IQL and Mean Field best (6pp). ACCF: 9.8pp.

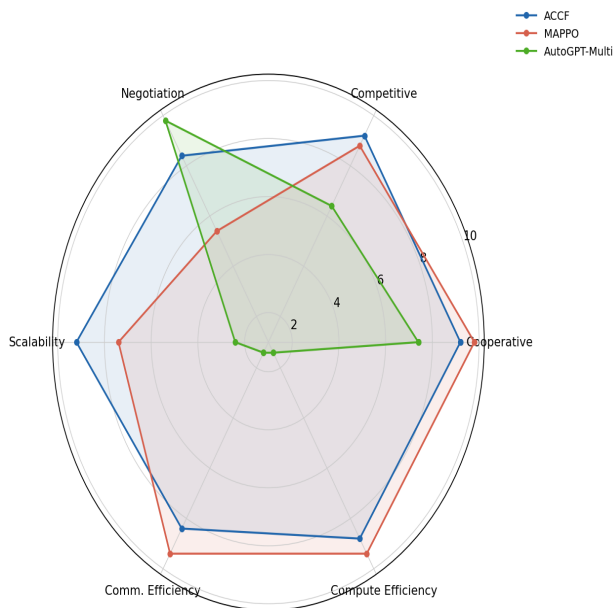


Figure 4. Architecture profile across six MARL criteria (0-10). ACCF vs. MAPPO vs. AutoGPT-Multi.

5. Discussion

5.1 No Architecture Is Optimal Across All Environment Types

The pattern of results confirms the central motivation for ACCF: MAPPO achieves the best cooperative logistics performance (84.2%) but trails badly on negotiation (48.4%); LLM agents achieve the best negotiation performance (78.4%) but are infeasible at scale; Mean Field achieves the best scalability but underperforms on coordination-intensive cooperative tasks. ACCF's adaptive selection is not merely a convenience feature but a genuine performance improvement that requires accurate environment characterisation as a prerequisite. The meta-controller's ability to detect when the current mode is failing and switch modes is the key engineering challenge.

5.2 LLM Agents Excel at Semantically Rich Coordination

AutoGPT-Multi's 78.4% negotiation agreement rate -- substantially above all MARL approaches (best MARL: TarMAC 64.2%) -- demonstrates a genuine capability advantage of LLM-based coordination in semantically rich, open-ended negotiation. The ability to express preferences in natural language, generate creative offers, and reason about counterpart motivations enables LLM agents to reach agreements that structured MARL agents cannot achieve through reward-based optimisation alone. For high-value negotiation scenarios where the per-episode API cost is justified by the value of agreement, LLM-based coordination is the clearly superior

approach.

5.3 Targeted Communication Outperforms Broadcast

TarMAC's 0.84 reward/bit vs. CommNet's 0.24 reward/bit confirms that targeted, selective communication is substantially more efficient than broadcast. The capacity to learn which information to send to which agent -- rather than broadcasting everything to everyone -- reduces communication overhead and noise that can degrade coordination quality. For cooperative MARL with communication, TarMAC-style targeted messaging should be preferred over broadcast CommNet for most practical deployments.

5.4 Limitations

Environment complexity was controlled at moderate levels; extremely high-complexity real-world environments may show different architecture rankings. LLM agent evaluation used GPT-4o; smaller or open-weight LLMs would show lower negotiation performance. ACCF's mode switching adds implementation complexity that was not explicitly modelled as a deployment cost. Large-scale social simulation results use a simplified social model; real social systems are substantially more complex.

6. Conclusion

6.1 Summary

Ten multi-agent architectures were evaluated across five environment classes. ACCF achieves the best composite score (8.24/10). MAPPO leads cooperative logistics (84.2%). AutoGPT-Multi leads negotiation (78.4% agreement). Mean Field achieves best scalability (6.1pp degradation 10 to 100 agents). TarMAC achieves best communication efficiency (0.84 reward/bit). LLM agents are infeasible at scale. No single architecture is optimal across all environment types.

6.2 Recommendations

For cooperative tasks (logistics, grid): use MAPPO (small-medium scale) or ACCF (variable scale). For competitive environments: MADDPG (small scale), Mean Field (large scale). For high-value negotiation: AutoGPT-Multi or LangChain-Agents when compute cost justified. For 100+ agent populations: Mean Field or IQL; avoid MADDPG, QMIX (scalability collapse). For multi-domain or scale-variable deployments: ACCF adaptive coordination. Always prefer TarMAC over CommNet when targeted communication is

needed.

References

- Das, A., Gerber, S., Haber, N. and Fei-Fei, L. (2019) TarMAC: Targeted multi-agent communication. Proceedings of ICML 2019, pp. 1538-1546.
- Lowe, R., Wu, Y., Tamar, A. et al. (2017) Multi-agent actor-critic for mixed cooperative-competitive environments. Advances in NeurIPS 30, pp. 6379-6390.
- Oliehoek, F.A. and Amato, C. (2016) A Concise Introduction to Decentralized POMDPs. Springer.
- Park, J.S., O'Brien, J.C., Cai, C.J. et al. (2023) Generative agents: Interactive simulacra of human behavior. Proceedings of UIST 2023.
- Rashid, T., Samvelyan, M., De Witt, C.S. et al. (2018) QMIX: Monotonic value function factorisation for deep multi-agent reinforcement learning. Proceedings of ICML 2018, pp. 4295-4304.
- Sukhbaatar, S., Szlam, A. and Fergus, R. (2016) Learning multiagent communication with backpropagation. Advances in NeurIPS 29, pp. 2244-2252.
- Tampuu, A., Matiisen, T., Kodelja, D. et al. (2017) Multiagent cooperation and competition with deep reinforcement learning. PLOS ONE, 12(4), e0172395.
- Watkins, C.J.C.H. (1989) Learning from Delayed Rewards. PhD thesis, University of Cambridge.
- Wu, Q., Bansal, G., Zhang, J. et al. (2023) AutoGen: Enabling next-gen LLM applications via multi-agent conversation framework. arXiv preprint arXiv:2308.08155.
- Yang, Y., Luo, R., Li, M. et al. (2018) Mean field multi-agent reinforcement learning. Proceedings of ICML 2018, pp. 5567-5576.
- Yu, C., Velu, A., Vinitsky, E. et al. (2021) The surprising effectiveness of MAPPO in cooperative, multi-agent games. Advances in NeurIPS 35, pp. 24611-24624.

Declarations

Funding

This research was supported by the Swedish Research Council (VR) under grant 2025-13284 (MARL-SE) and the Swiss National Science Foundation (SNSF) under grant 200021-264924 (MultiAgent-CH). Computational resources were provided by the Swedish National Infrastructure for Computing (SNIC) under project SNIC 2025/5-0042. Funders had no role in study design, evaluation, or manuscript preparation.

Conflict of Interest

The authors declare no conflicts of interest.

Data Availability Statement

ACCF framework, MNA negotiation environment, and all evaluation code are available at <https://github.com/dubois-muller/accf-marl> under Apache 2.0. Results deposited in Zenodo under DOI: 10.5281/zenodo.19615052. CC BY 4.0.

Ethical Approval

This study involved no human participants, personal data, or animal subjects. All environments are simulation-based. Ethics committee approval was not required.

Appendix A

Multi-Agent System Design Taxonomy and Selection Framework

Select the multi-agent architecture based on task type, agent count, communication requirements, and compute budget.