Large Language Models Miss the Multi-Agent Mark

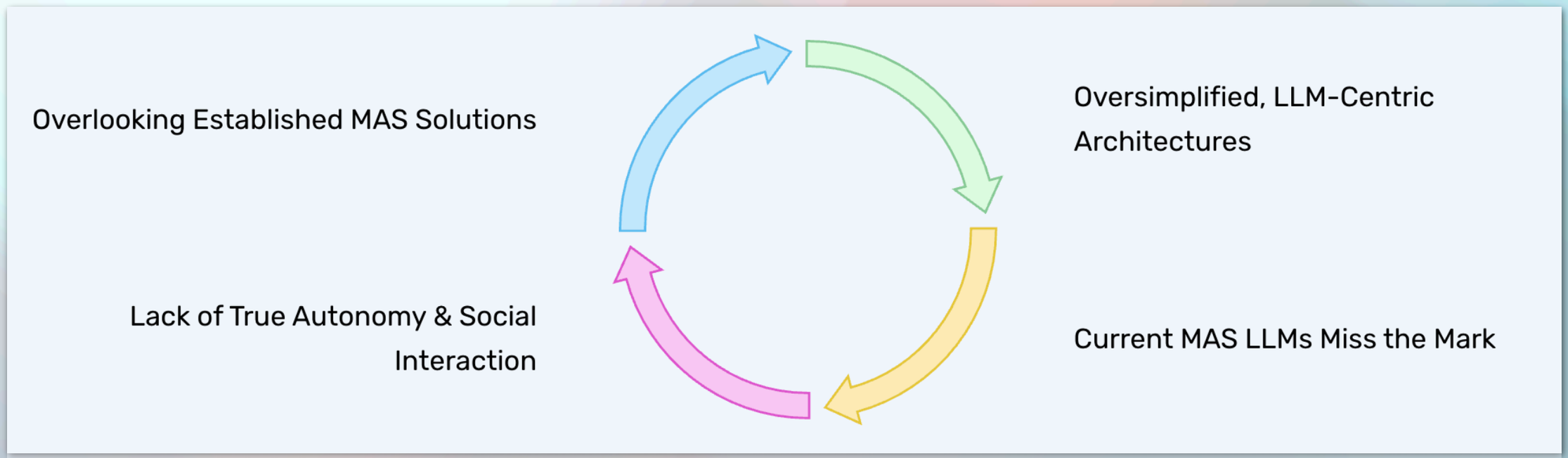
E. La Malfa, G. La Malfa,
S. Marro, J.M. Zhang, E. Black, M. Luck, P. Torr, M. Wooldridge

University of Oxford,
King's College London,
University of Sussex

The Core Position:
A Misappropriation of Terminology

Paper Position

Current **MAS-LLMs often fail to embody fundamental multi-agent system characteristics**, such as autonomy, social interaction, and structured environments, **by overemphasising the role of LLMs and overlooking solutions that already exist in MAS literature.**



Where MAS LLMs Fall Short

1. Social Intelligence

LLM agents ‘cooperation' is often scripted or orchestrated, not a result of inherent social abilities like negotiation or reasoning about others' intentions.

2. Environment Design

Environments are designed specifically for LLMs (text-based). This design choice inherits LLM limitations like non-determinism and hallucinations.

3. Coordination & Communication

Interactions are typically sequential and synchronous, lacking true asynchronicity. Communication relies on inefficient and ambiguous natural language.

4. Emergent Behaviours

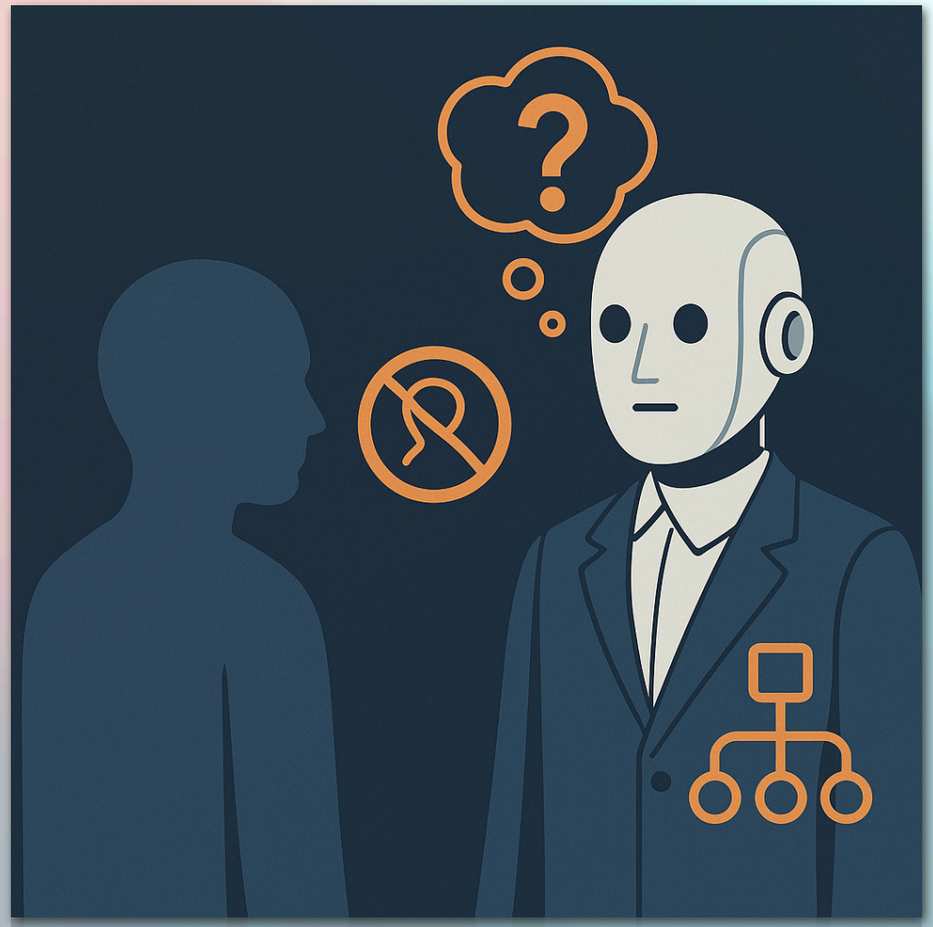
There is no proper definition of emergent behaviour, just observational hype.

Critique 1:
The Illusion of Social Intelligence

A true intelligent agent in MAS is defined by being reactive, proactive, and social. LLMs are reactive (to prompts) and proactive (task initiation).

But they fail on the social dimension.

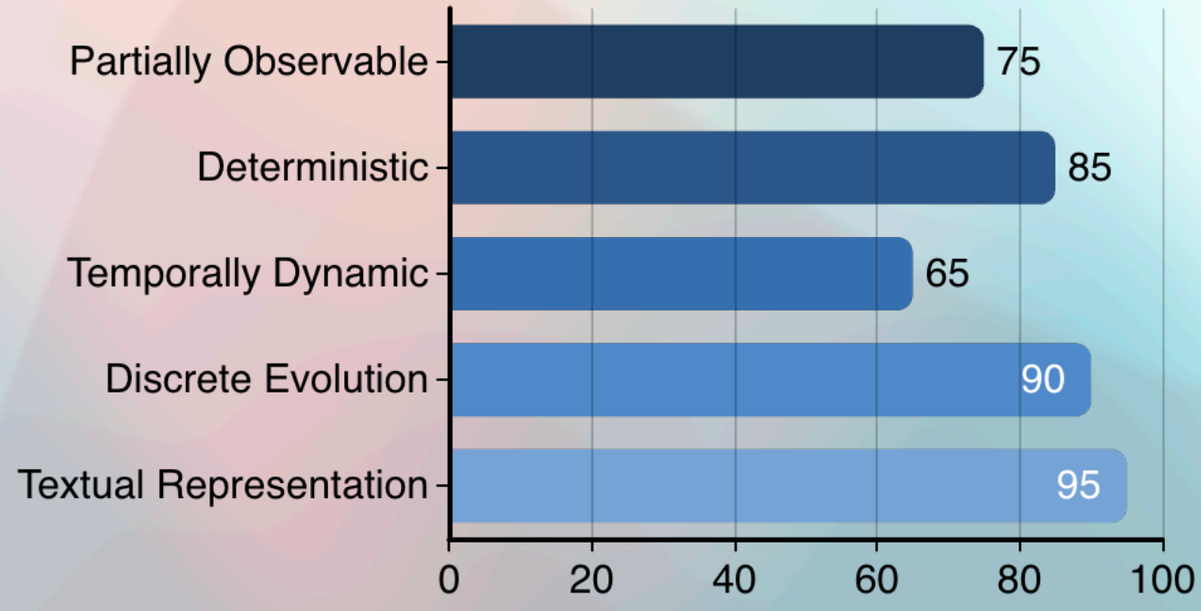
- **Not Natively Trained:** LLMs are trained in isolation to respond to user requests, not to interact, compete, or cooperate with other agents.
- **Poor Theory of Mind:** They consistently struggle to infer the beliefs, desires, and intentions of other agents, a critical social skill.
- **Ensembles, Not Systems:** Many 'multi-agent' setups are simply aggregation mechanisms (like majority vote), not true concurrent systems with complex, emergent interaction strategies.



Critique 2:
The LLM-Centric Environment

Traditional MAS models the environment first, with agent architecture being flexible. MAS LLMs subvert this by designing environments specifically for LLMs that communicate via text. This creates significant problems:

- **Non-Determinism:** LLMs are inherently non-deterministic, making it impossible to guarantee safety or specific outcomes in supposedly controlled environments.
- **Hallucinations & Memory:** Reliance on text as memory exceeds context windows, causing hallucinations and a lack of persistent, consistent state representation.
- **Limited Representation:** Forcing all perception into text is inefficient and loses information, especially for multi-modal tasks.



Analysis of ~110 papers shows a heavy bias towards simplified environments that cater to LLMs but fail to capture real-world complexity.

Critique 3:
Flawed Coordination & Communication

Asynchronicity is Absent

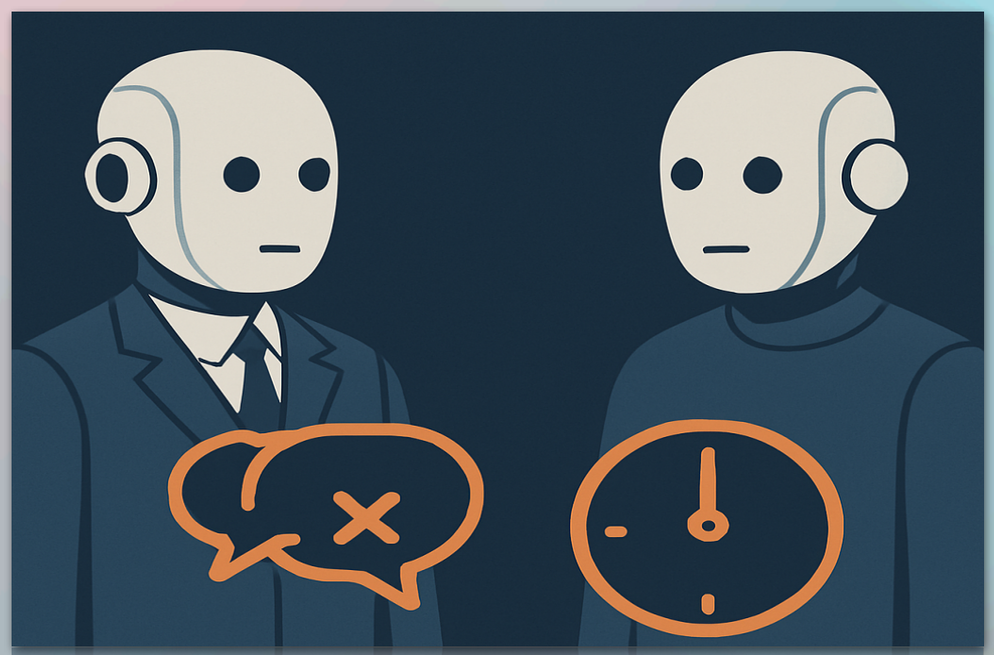
Real-world systems are asynchronous, with agents acting concurrently and at different times. Most MAS LLM frameworks operate in rigid, sequential pipelines or simple parallel execution, failing to model complex, real-time interactions.

Natural Language is Not a Protocol

Relying on natural language for agent-to-agent communication is inefficient, expensive (token cost), and dangerously ambiguous. An utterance like 'a car is coming' can be information, a warning, or a request. This ambiguity leads to failures that are hard to inspect or prevent.

The Established MAS Solution

The MAS field has decades of research on performative communication. Structured languages like KQML or FIPA ACL provide unambiguous, efficient ways for agents to communicate beliefs, commitments, and actions, separating information exchange from requests.

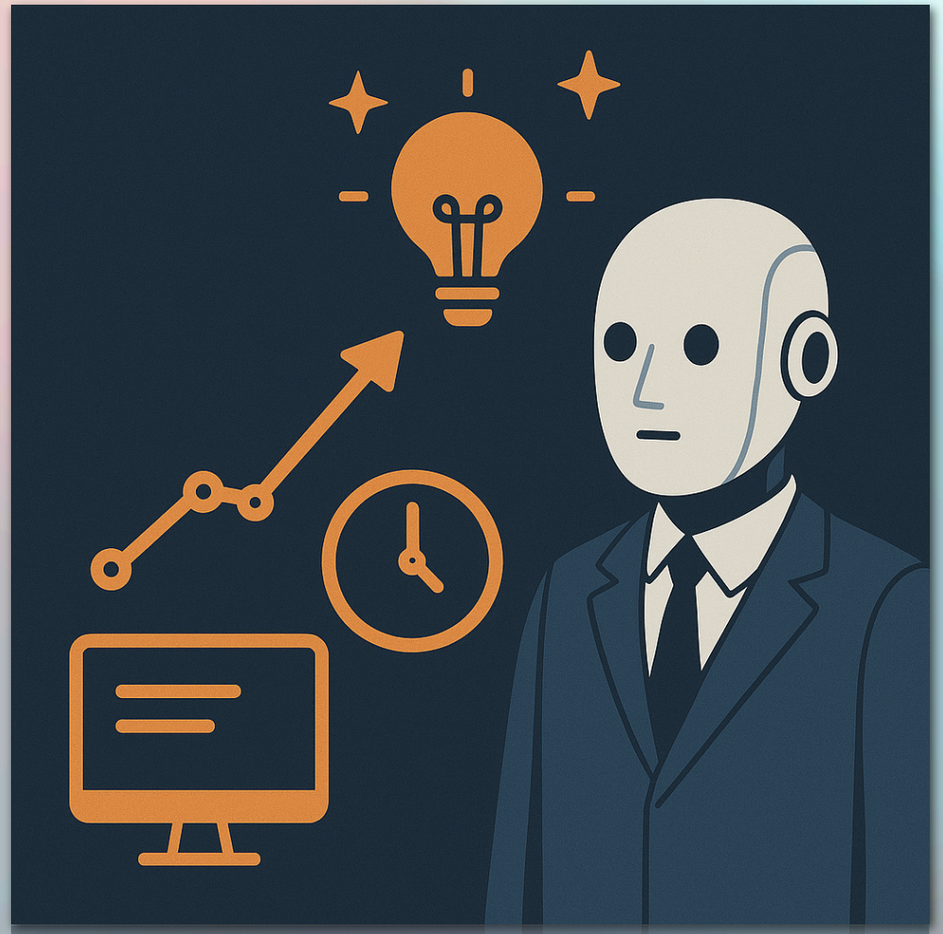


Critique 4:
Emergence as Anecdote, Not Science

Problem: Observational, Not Measurable

Many influential works claim LLM agents produce 'emergent behaviours' in open-ended simulations. However, these claims are almost always based on qualitative observation and anecdotal evidence; researchers let the system run and report interesting outcomes.

- **No Formal Definition**: 'Emergence' is used loosely without a systematic definition or reference to the system being analyzed.
- **Lack of Benchmarks:** There are no agreed-upon metrics to identify, quantify, or falsify claims of emergence.
- **Correlation vs. Causation**: It's difficult to distinguish truly emergent coordination from coincidental outputs or the inherent, pre-programmed capabilities of a powerful general-purpose LLM.



A Path Forward: Integrating Established MAS Principles

Social Pre-Training

Move beyond single-agent fine-tuning. Pre-train LLMs in multi-agent scenarios involving cooperation and competition, using reinforcement learning and interactive feedback to develop native social skills.

Agent-Agnostic Environments

Design open, multi-modal environments that are not LLM-centric. Integrate structured memory systems and formal planners to provide guarantees, handle non-determinism, and reduce hallucinations.

Standardized Communication

Adopt natively asynchronous frameworks and standardized, open-source communication protocols (inspired by KQML, FIPA). This will ensure security, identity, trust, and clarity, reducing cost and ambiguity.

Quantifiable Emergence

Establish clear, falsifiable definitions and benchmarks for emergent behaviour, distinguishing between weak emergence (explainable from parts) and strong emergence (requiring new laws). This moves analysis from anecdote to science.

Acknowledgments

University of Oxford - Department of Computer Science and Department of Engineering

King’s College University of London - Department of Informatics

University of Sussex

IDAI - Institute for Decentralized AI

Thank you