Progress Slides I (5th Feb 24)

Oscar Moxon

King's College London

February 5, 2024

Overview of Multi-Agent Debate Systems

Contents:

- Exploration of Multi-Agent Debate (MAD) systems to facilitate scientific inquiry dialogues [1].
- Motivation: Integration of debate mechanics to enhance reasoning capabilities of language models [3, 4, 6].
- Human-in-the-loop as a critical component for monitoring judgements and completing tasks requiring embodiment [8].

KING'S College LONDON

Making New Discoveries with LLMs

LLMs and Scientific Discovery

- Increasingly apparent that LLMs can produce novel insights about notorious scientific problems.
- Pairing of pretrained LLM with a systematic evaluator to iterate on solutions to the cap set problem.
- LLM generates constructions that are interpretable and falsifiable.
- Leads to the largest improvement in 20 years.

Refinement and Feedback Loop

- Role of evaluator is to guard against confabulations and incorrect ideas.
- Best-shot prompting to enhance algorithms, creating a positive feedback circuit.

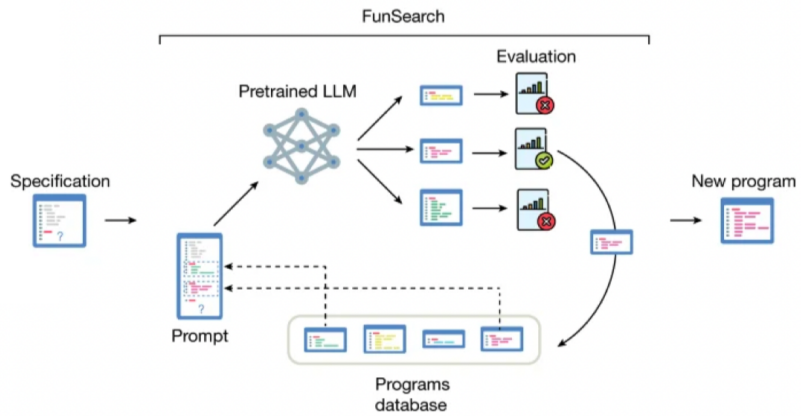


Figure: FunSearch.

Capabilities of Autonomous Agents

- Web search capabilities are crucial for agents to process and synthesize current information.
- Fostering unique personas promotes specialization and varied viewpoints.
- Emphasis on specific expertise integrates diverse perspectives for richer outcomes.
- Human-like evaluation dynamics ensure adaptability and robust responses to novel information.
- Agents' multimodal abilities, such as those provided by Visual Transformers (ViT), enhance interaction fidelity.

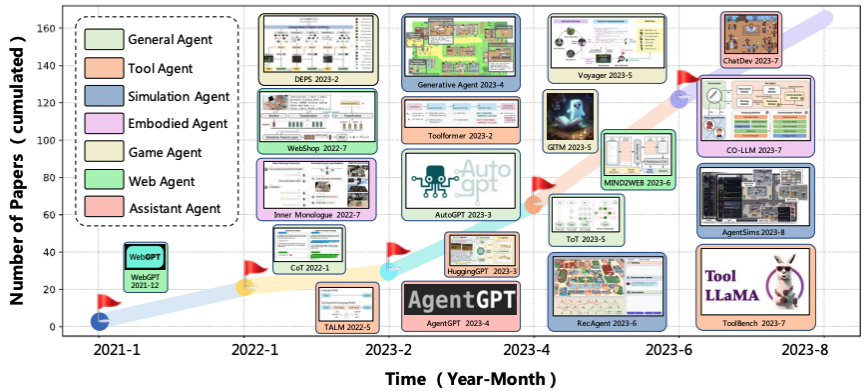
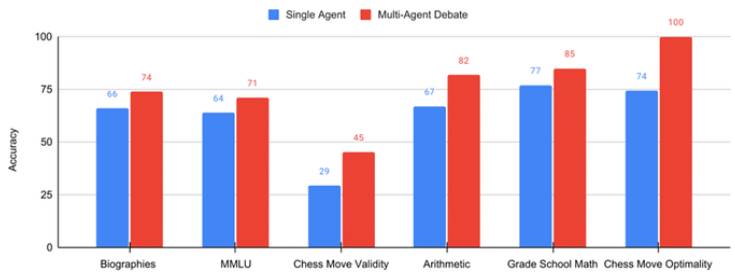


Figure: Map of Autonomous Agents Research in 2023.

Reasoning and Debate Mechanics

- Giving agents the ability to self-reflect and Chain of Thought (CoT) enables advanced reasoning beyond one-shot prompts.
- Overcoming Degeneration-of-Thought (DoT) with feedback cycles for agent cross-evaluation.
- Formal debate dynamics through multiple rounds, with judging agent or voting protocol in place.
- Embracing the Society of Minds concept.
- Collective intelligence factor significantly boosts reasoning in agent systems across benchmarks.
- Multi-Agent Debate (MAD) promotes divergent thinking.



Multi-Agent System for Debate and Inquiry

Project Objectives

- Construct a multi-agent system framework aimed at simulating debate and inquiry processes to enhance knowledge discovery.
- Develop an interpretable process with a judge that evolves discussion productively.
- Equip agents with ability to evaluate research, supporting arguments or challenging opposing views.
- Enable human interaction to further positive feedback.

Implementation Timeline

- By 27th Feb: Have plan for implementing this system, supported by multi-agent debate papers and open source projects.

Literature Overview

1. Encouraging Divergent Thinking in Large Language Models Through Multi-Agent Debate
2. A Survey on Large Language Model based Autonomous Agents
3. Collective Intelligence Factor in Groups
4. ChatEval: multi-agent debate towards better LLM-based evaluators
5. FunSearch
6. Igniting Language Intelligence
7. Large Language Models Cannot Self-Correct Reasoning Yet
8. The Rise and Potential of LLM Agents: A Survey