Progress Slides II (27th Feb 24)

Oscar Moxon

King's College London

February 28, 2024

Introduction to Multi-Agent Debate System

Developing a multi-agent debate system with narrative judgement to develop c-factor in LLM agent collectives. This agent cluster is part of an operative kernel that could be used to bring embodied agents as a force within the workforce.

1. Goal Identification: Identifying the objective or desired outcome.

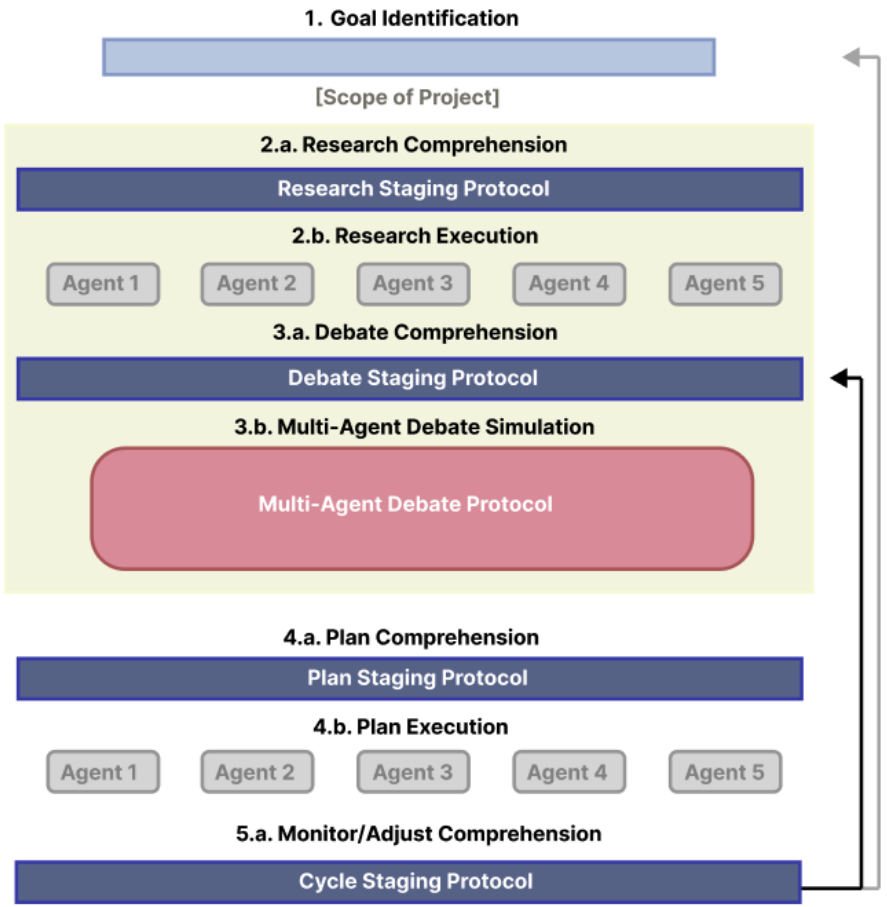
2.a Research & Information Gathering: Collecting relevant data and information.

2.b Research Execution.

3.a Debate Staging.

3.b Multi-Agent Debate Protocol.

4+ Out of scope.



2. Research Staging Protocol

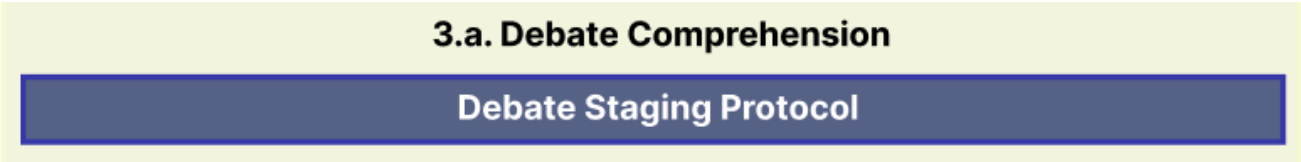
2.a. Task Assignment:

- Agents are assigned specific tasks based on their designed capabilities, such as web navigation, sitemap retrieval, API interfacing, and database querying.
- Web navigation agents use tools like Selenium or Puppeteer for automated browsing.
- Sitemap and API agents may utilise HTTP requests to fetch data from predefined endpoints.

2.b. Parallel Execution:

- Each agent operates in a parallel channel to ensure the gathering phase is concurrent, reducing the overall time for data collection.
- Agents should have asynchronous capabilities to handle I/O operations efficiently.

3.a. Debate Staging Protocol



3.a. Data Handling and QA: Agents preprocess the gathered data to normalise and structure it for downstream use.

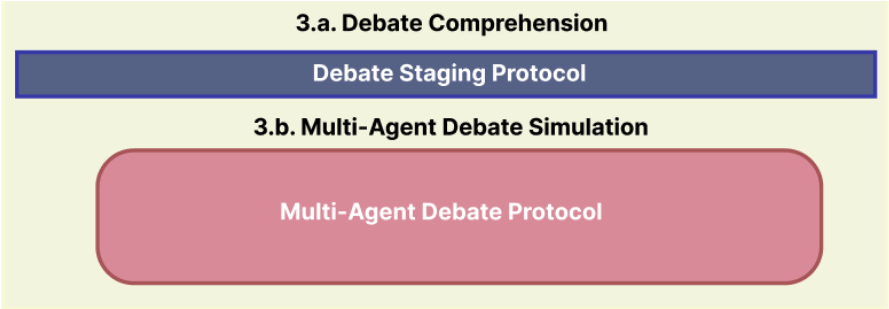
- Extracted data should be stored in a centralised data warehouse with a consistent schema for easy access and analysis.
- Implement checksums, data validation, and verification processes to ensure integrity and accuracy.
- Integrate cross-referencing functions where agents can compare and validate information against multiple sources.

Scope for Debate Coordination

- This stage would establish the framework for the debate, defining the roles of the judge or guide, which would be an overarching model or system designed to moderate the debate.

3.b. Multi-Agent Debate

Exploration into Inquiry dialogues in multi-agent systems



Collective Intelligence "c-Factors":

- **Model Size:** Larger models may provide more nuanced arguments due to their capacity for complex understanding and generating emergent behaviours.
- **Model Persona:** Tailoring personas for different models to represent varied perspectives and expertise.
- **Model Contributions:** Establishing protocols for how models interject, support, or oppose arguments, and how they build upon each other's contributions.

Challenges and Measure of Success

Challenges:

- **Mitigating Groupthink:** Ensuring diversity in model reasoning to avoid uniformity of thought; possibly by using a diverse set of models or incorporating adversarial models.
- **Iterative Improvement:** Keeping the debate dynamic with new information or perspectives; perhaps through real-time updates or iterative rounds of argumentation.
- **Circulation of Information:** Efficiently managing how information is shared among models to inform arguments.

Measure of Success:

- Success measured by the ability of the debate to produce a coherent narrative report that functions like a literature review, documenting debate evolution, supporting final consensus.
- Group c-factors then compared by these reports, with information input as a constant. Allows conclusions to be drawn about quality multi-agent debates.

References I


Shoham, Y. & Leyton-Brown, K.
Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations.
Cambridge University Press, 2008.


Russell, S. & Norvig, P.
Artificial Intelligence: A Modern Approach, 3rd ed.
Prentice Hall, 2010.


Mehrabi, N., et al.
'Bias and Fairness in the Age of Artificial Intelligence'.
Electronics, 8(8):835, 2019.


Sutton, R. S. & Barto, A. G.
Reinforcement Learning: An Introduction, 2nd ed.
The MIT Press, 2018.


Rahwan, I. & Simari, G. R.
Argumentation in Artificial Intelligence.


Bordini, R. H., Hübner, J. F. & Wooldridge, M.
Programming Multi-Agent Systems in Agentspeak using Jason.
Wiley, 2007.

Further Reading I

- Bansal, S. and Weld, D.S. (2021). Chain of Thought Reasoning.
- Hendrycks, D. et al. (2022). Language Models as Few-Shot Common-sense Learners.
- Lu, S. et al. (2022). Enhancing Reasoning in Language Models.
- OpenReview (n.d.). Improving Reasoning in LLMs through Prompting Techniques.
- Prystawski, B. and Goodman, N.D. (2023). The Role of Experience in Reasoning.
- WebGPT, K. et al. (2023). Language Models Learning to Utilize Tools.
- Yao, F. (2022). Tracing GPT's Emergent Abilities.