CMPE492 - Week 5

Zeynep Buse Aydın - Asım Dağ

The Rise and Potential of Large Language Model Based Agents: A Survey

a general framework for LLM-based agents

- The Brain
  - primarily an LLM
  - stores crucial memories, knowledge, info
  - essential tasks of information processing, decision-making, reasoning, planning
- The Perception Module
  - like the sensory organs
  - expands the agent's perceptual space from text-only to a multimodal space that includes diverse sensory modalities like text, sound, visuals, touch, smell, and more.
- The Action Module
  - possess textual output, take embodied actions, and use tools

Harnessing AI for Good

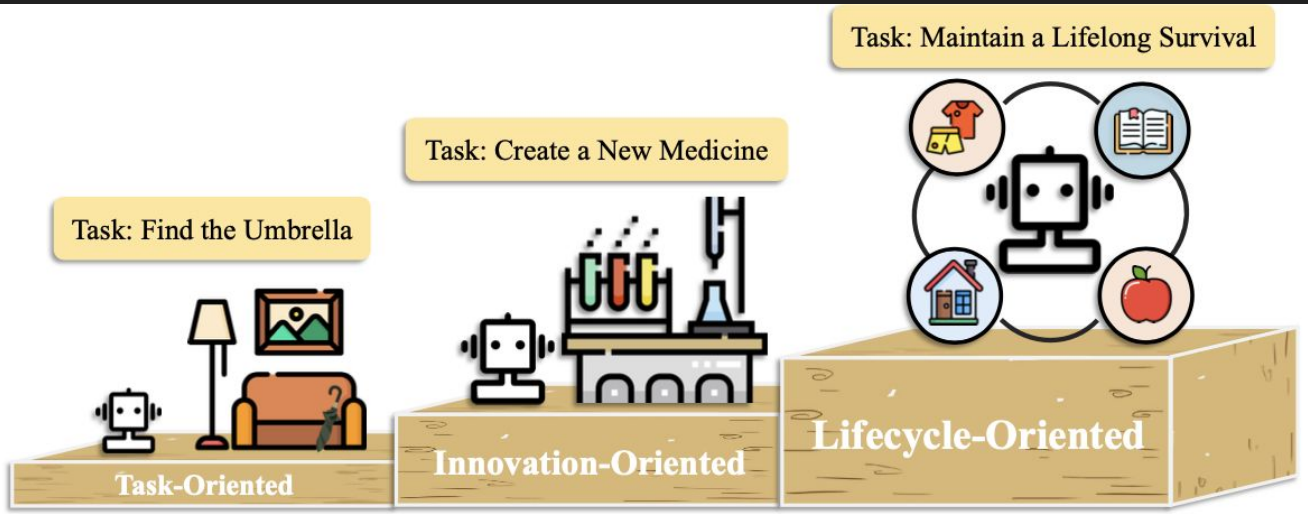


Figure 8: Practical applications of the single LLM-based agent in different scenarios. In **task-oriented deployment**, agents assist human users in solving daily tasks. They need to possess basic instruction comprehension and task decomposition abilities. In **innovation-oriented deployment**, agents demonstrate the potential for autonomous exploration in scientific domains. In **lifecycle-oriented deployment**, agents have the ability to continuously explore, learn, and utilize new skills to ensure long-term survival in an open world.

The Ways LLMs Interact with Each Other

Cooperative Interaction for Complementarity

- Disordered
- Ordered

Adversarial Interaction for Advancement

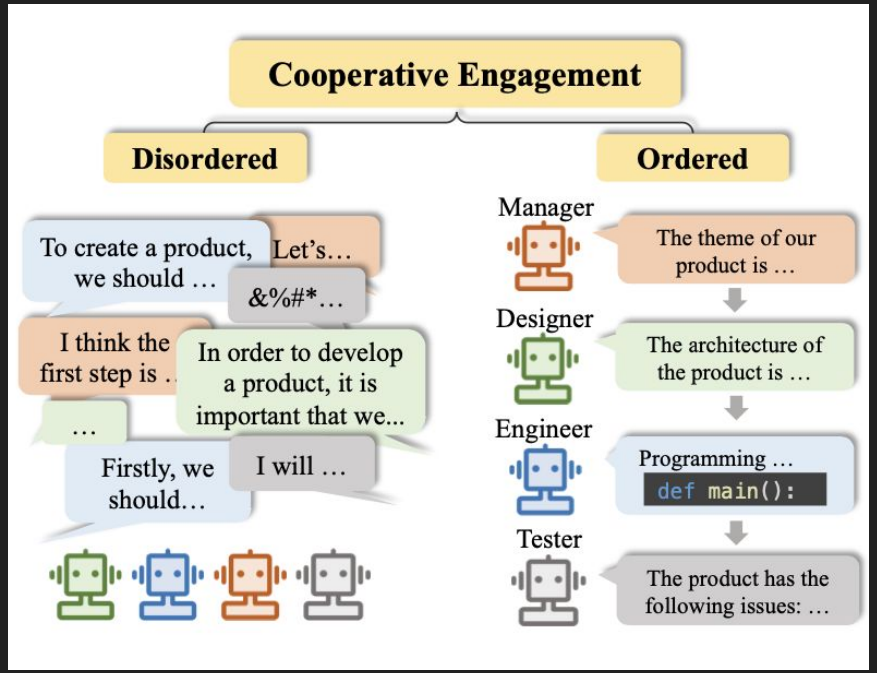
Cooperative Interaction for Complementarity

Disordered

- introducing a dedicated coordinating agent in multi-agent systems, responsible for integrating and organizing responses from all agents, thus updating the final answer,
- majority voting (limited research that integrates this module)

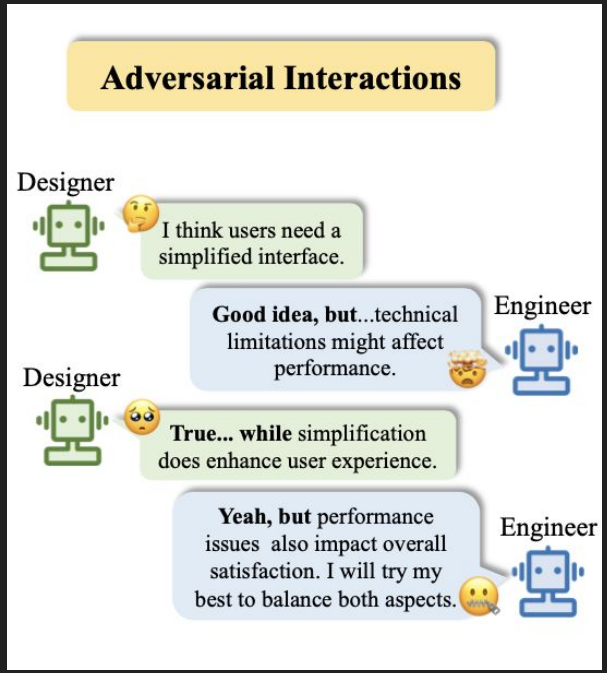
Ordered

- Meta GPT - waterfall model (However, during MetaGPT's practical exploration, a potential threat to multi-agent cooperation has been identified. Without setting corresponding rules, frequent interactions among multiple agents can amplify minor hallucinations indefinitely. For example, in software development, issues like incomplete functions, missing dependencies, and bugs that are imperceptible to the human eye may arise. Introducing techniques like cross-validation or timely external feedback could have a positive impact on the quality of agent outputs.)



Adversarial Interaction for Advancement

- In competitive environments, agents can swiftly adjust strategies through dynamic interactions, striving to select the most advantageous or rational actions in response to changes caused by other agents **competition**, **argumentation**, and **debate**
- By abandoning rigid beliefs and engaging in thoughtful reflection, adversarial interaction **enhances the quality of responses.**
- when multiple agents express their arguments in the state of “tit for tat”, one agent can receive substantial external feedback from other agents, thereby **correcting its distorted thoughts.**
- high-quality responses and accurate decision-making



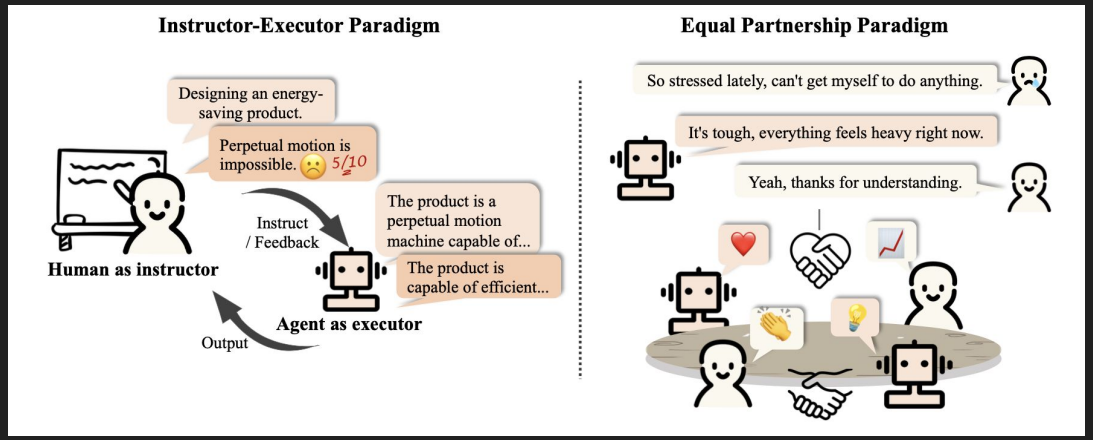
However,

- With prolonged debate, LLM’s limited context cannot process the entire input.
- In a multi-agent environment, computational overhead significantly increases.
- Multi-agent negotiation may converge to an incorrect consensus, and all agents are firmly convinced of its accuracy.

The development of multi-agent systems is still far from being mature and feasible. Introducing human guides when appropriate to compensate for agents’ shortcomings is a good choice to promote the further advancements of agents.

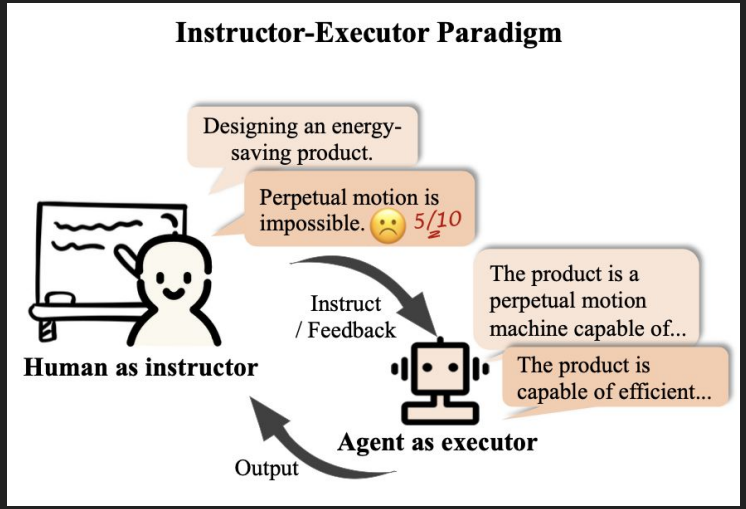
Human-Agent Interaction

- Instructor-Executor
  - (unequal interaction)
- Equal Partnership
  - (equal interaction)



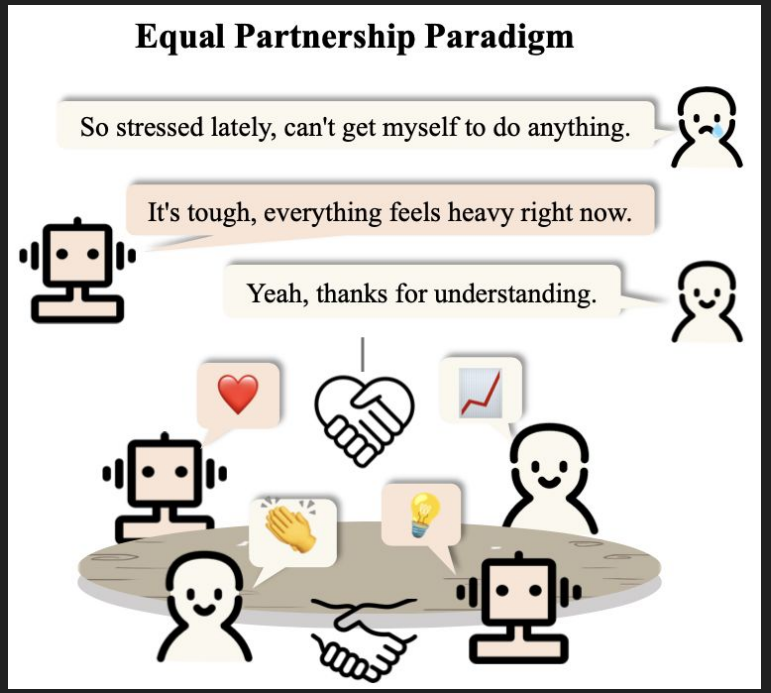
Instructor-Executor

- unequal interaction
- the agent responds to each human instruction, refining its action through alternating iterations to ultimately meet human requirements
- significant demands on humans. It requires a substantial amount of human effort and, in certain tasks, might even necessitate a high level of expertise.
  - Quantitative feedback
    - absolute evaluations like binary scores and ratings, or scores
    - easy to collect, but sometimes it may oversimplify user intent
  - Qualitative feedback
    - text feedback
    - Humans provide advice on how to modify outputs generated by agents, and the agents then incorporate these suggestions to refine their subsequent outputs
    - better convey human intention compared to quantitative feedback, it might be more challenging for the agents to comprehend
  - combining multiple types of feedback can yield better results.



Equal Partnership

- equal interaction
- Empathetic communicator
  - delving into the empathetic capacities of agents
  - enhance user satisfaction
  - make significant progress in fields like healthcare and business marketing
  - Unlike simple rule-based conversation agents, agents with empathetic capacities can tailor their interactions to meet users' emotional needs



Fields

- education
  - early childhood education
  - mathematics
- medicine
  - mental health
  - online communication with adolescents on the autism spectrum, analyzing users' speech and facial expressions in real-time to engage them in multi-topic conversations and provide instant feedback regarding non-verbal cues.
  - contextualized language generation approaches to provide tailored assistance for users who seek support on diverse topics ranging from relationship stress to anxiety.
- business
  - provide automated services or assist humans in completing tasks
  - software development tasks
- scientific research

Agent Societies

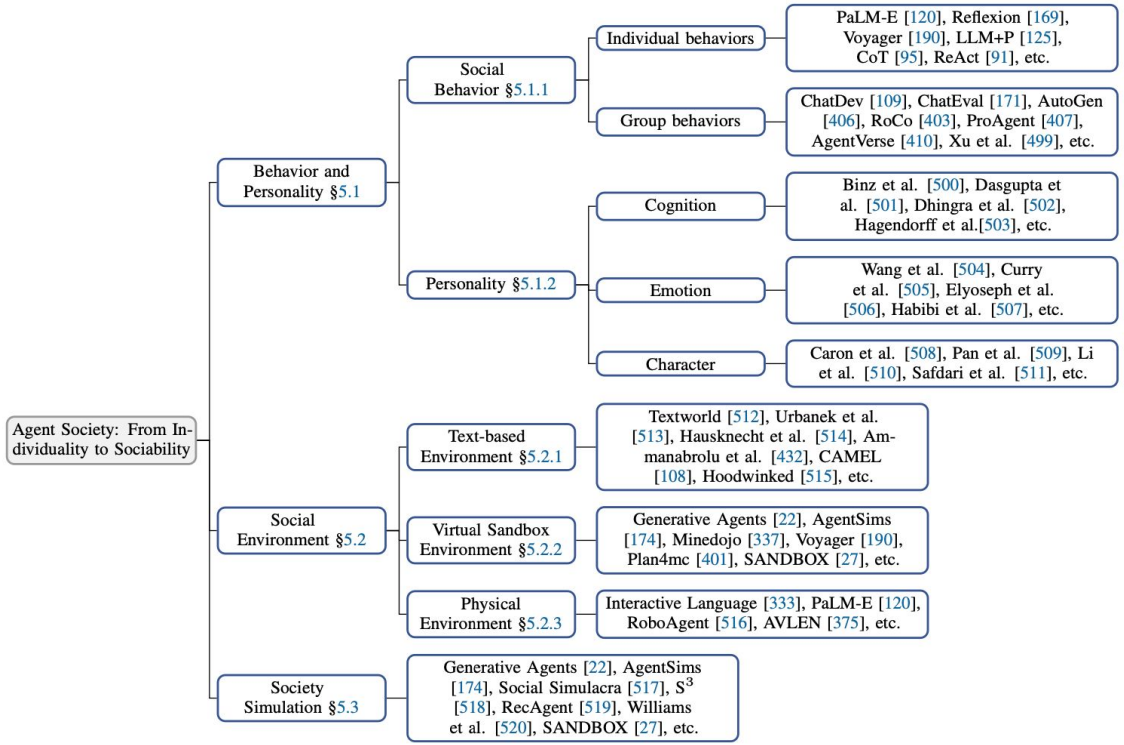


Figure 11: Typology of society of LLM-based agents.

Society Simulation with LLM-based Agents

- micro-level simulation
- **open**
  - Agents have the flexibility to enter or leave the environment without disrupting its operational integrity. The environment can be expanded by adding or removing entities in the virtual or physical world, along with adaptable resources like tool APIs. Humans can also participate in societies by assuming the role of an agent or serving as the “inner voice” guiding these agents.
- **persistent**
- **situated**
  - The situated nature of the society emphasizes its existence and operation within a distinct environment.
- **organized**

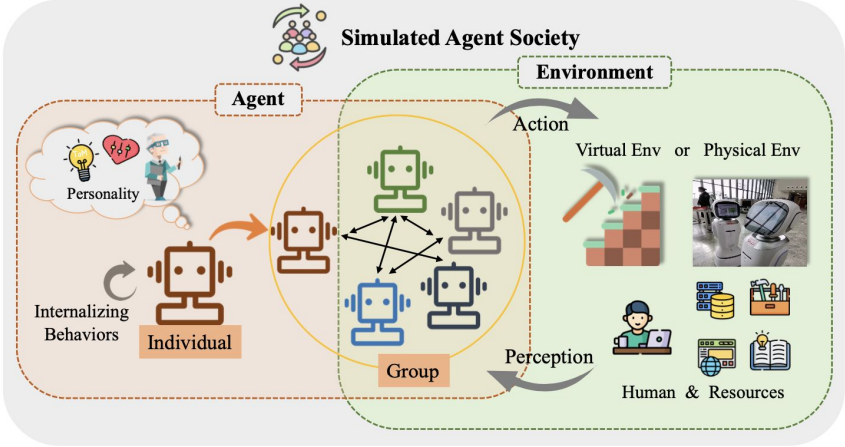


Figure 12: Overview of Simulated Agent Society. The whole framework is divided into two parts: the **Agent** and the **Environment**. We can observe in this figure that: (1) **Left:** At the individual level, an agent exhibits internalizing behaviors like planning, reasoning, and reflection. It also displays intrinsic personality traits involving cognition, emotion, and character. (2) **Mid:** An agent and other agents can form groups and exhibit group behaviors, such as cooperation. (3) **Right:** The environment, whether virtual or physical, contains human actors and all available resources. For a single agent, other agents are also part of the environment. (4) The agents have the ability to interact with the environment via perception and action.

Insights from Agent Society

Organized productive cooperation.

Propagation in social networks.

Ethical decision-making and game theory.

Policy formulation and improvement.

Organized productive cooperation.

- within this simulated society, the integration of diverse experts introduces a multifaceted dimension of individual intelligence. When dealing with complex tasks, such as software development or consulting, the presence of agents with various backgrounds, abilities, and experiences facilitates creative problem-solving.
- Efficient communication also plays a pivotal role in such a large and complex collaborative group. (communication styles with reference to standardized operating procedures (SOPs))

Propagation in social networks.

As simulated social systems can model what might happen in the real world, they can be used as a reference for predicting social processes. Unlike traditional empirical approaches, which heavily rely on time-series data and holistic modeling, agent-based simulations offer a unique advantage by providing more interpretable and endogenous perspectives for researchers. Here we focus on its application to modeling propagation in social networks.

The first crucial aspect to be explored is the development of **interpersonal relationships** in simulated societies. For instance, agents who are not initially connected as friends have the potential to establish connections through intermediaries. Once a network of relationships is established, our attention shifts to the dissemination of information within this social network, along with the underlying attitudes and emotions associated with it. Proposes a user-demographic inference module for capturing both the number of people aware of a particular message and the collective sentiment prevailing among the crowd. This same approach extends to modeling **cultural transmission** and **the spread of infectious diseases**. By employing LLM-based agents to model individual behaviors, implementing various intervention strategies, and monitoring population changes over time, these simulations empower researchers to gain deeper insights into the intricate processes that underlie various social phenomena of propagation.

Ethical decision-making and game theory.

Simulated societies offer a dynamic platform for the investigation of intricate decision-making processes, encompassing decisions influenced by ethical and moral principles. Taking Werewolf game and murder mystery games as examples, researchers explore the capabilities of LLM-based agents when confronted with challenges of deceit, trust, and incomplete information. These complex decision-making scenarios also intersect with game theory, where we frequently encounter moral dilemmas pertaining to individual and collective interests, such as Nash Equilibria. Through the modeling of diverse scenarios, researchers acquire valuable insights into how agents prioritize values like honesty, cooperation, and fairness in their actions. In addition, agent simulations not only provide an understanding of existing moral values but also contribute to the development of philosophy by serving as a basis for understanding how these values evolve and develop over time. Ultimately, these insights contribute to the refinement of LLM-based agents, ensuring their alignment with human values and ethical standards.

Policy formulation and improvement.

The emergence of LLM-based agents has profoundly transformed our approach to studying and comprehending intricate social systems. However, despite those interesting facets mentioned earlier, numerous unexplored areas remain, underscoring the potential for investigating diverse phenomena. One of the most promising avenues for investigation in simulated society involves exploring various **economic and political states** and their impacts on societal dynamics. Researchers can simulate a wide array of economic and political systems by configuring agents with differing economic preferences or political ideologies. This in-depth analysis can provide valuable insights for policymakers seeking to foster prosperity and promote societal well-being. As concerns about environmental sustainability grow, we can also simulate scenarios involving resource extraction, pollution, conservation efforts, and policy interventions. These findings can assist in making informed decisions, foreseeing potential repercussions, and formulating policies that aim to maximize positive outcomes while minimizing unintended adverse effects.

Scaling Number of Agents

Through increasing the number of agents in the society simulation, humans can draw better experiences and insights to improve the harmony of real-world societies.

- Predetermined scaling
- Dynamic scaling

As the number of agents increases, the challenges of communication and message propagation become quite formidable. This is because the communication network of the entire system becomes highly complex.