Part V: Final Remarks and Outlook

Future Directions

🤖 Planning

🤖 Synthetic Data

🤖 Continual Learning

🤖 Safety

🤖 Agent-computer interface

Future directions: planning

- How to do hierarchical planning? Is it just a matter of prompting an LLM, or is there more to it?
- How far can (tree) search go?
- How to evaluate (partial) plans? Value functions? Reward models?
- How to make model-based speculative planning work for generalist agents?

Future directions: synthetic data

- Agents need to learn **perception-decision-execution** capabilities

- Data on the Internet is mostly artifacts from such processes, not capturing the processes *per se*

- Synthesizing data with LLMs provides a possibility to uncover (some of) these hidden processes

Future directions: continual learning

- Currently, the field is transitioning from prompting to behavior cloning / supervised fine-tuning

- Behavior cloning is probably insufficient for generalist agents; they need to explore the environments and **learn from trial and error**

- Challenges from open action space, reward model, and safety

Future directions: safety

- Agent safety research is far behind agent development and deployment

- Language agents
  - Inherent all the safety risks of LLMs (e.g., *bias, fairness, hallucination, privacy, transparency*)
  - amplify some of them (e.g., *workforce displacement*)
  - and bring many new ones (e.g., *irreversible actions*)

Future directions: Agent-computer interface

- Human computer interface -> agent computer interface + human agent interface?
- Human-agent collaboration
  - What's the best way AI and humans work together
  - Devin vs Cursor vs Github copilot
- Most agent benchmark assumes autonomous setup: remove human element
- ACI design inspired by HCI design

Relevant agent workshops and talks

- ICLR 2024 Workshop on LLM Agents
- Trustworthy Multi-modal Foundation Models and AI Agents (TiFA) ICML 2024
- Multi-modal Foundation Model meets Embodied AI ICML 2024
- NeurIPS 2024 Workshop on Open-World Agents
- NeurIPS 2024 Workshop on Towards Safe & Trustworthy Agents
- Princeton PLI Workshop on Useful and Reliable AI Agents
- CMU Agent Workshop 2024
- CoRL 2024 Workshop on Language and Robot Learning
- CoRL 2024 Workshop on X-Embodiment Robot Learning
- Berkeley Course on Large Language Model Agents
- FAccT 2024 Tutorial on LM Agents: Prospects and Impacts

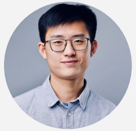
https://github.com/ranpox/awesome-computer-use
https://github.com/showlab/Awesome-GUI-Agent

EMNLP 2024

Language Agents: Foundations, Prospects, and Risks

Yu Su, OSU, @ysu_nlp

Diyi Yang, Stanford, @diyi_yang



Shunyu Yao, OpenAI, @ShunyuYao12

Tao Yu, HKU, @taoyds