

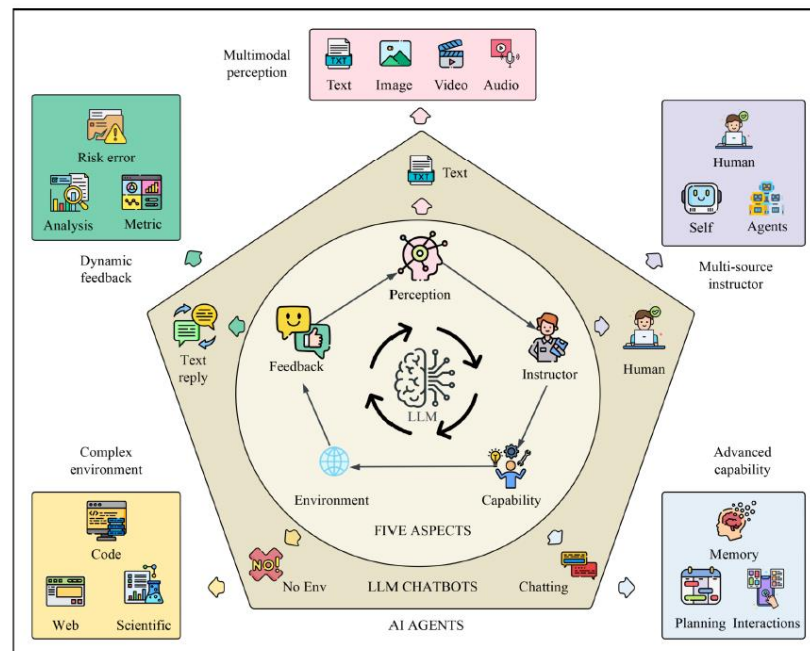
# Evolutionary Perspectives on the Evaluation of LLM-Based AI Agents: A Comprehensive Survey

Jiachen ZHU, Menghui ZHU, Renting RUI, Rong SHAN,  
Congmin ZHENG, Bo CHEN, Yunjia XI, Jianghao LIN,  
Weiwen LIU, Ruiming TANG, Yong YU, Weinan ZHANG

Frontiers of Computer Science, DOI: [10.1007/s11704-026-51590-2](https://doi.org/10.1007/s11704-026-51590-2)

# Problems & Ideas

- The problem:
  - *Given the rapid advancement of AI agents, how can we systematically evaluate LLM-based AI agents from an evolutionary perspective?*
- Ideas:
  - Distinguish AI agents from LLM chatbots from five key aspects: environment, instructor, feedback, perception, and capability.
  - Categorize existing evaluation benchmarks for AI agents along two critical axes: external environments and internal capabilities.



# Main Contributions

- Contributions:
  - We analyze the evolutionary trends in evaluation benchmarks across four aspects: environment, agent, evaluator, and metric.
  - We discuss future directions and offer practical guidelines on selecting appropriate benchmarks, supporting continued progress in agent evaluation research.

