

Language Agents and Multi-Hop Reasoning:

Using Language Agent Teams to Research, Identify, and Evaluate “Disruptive Innovations”

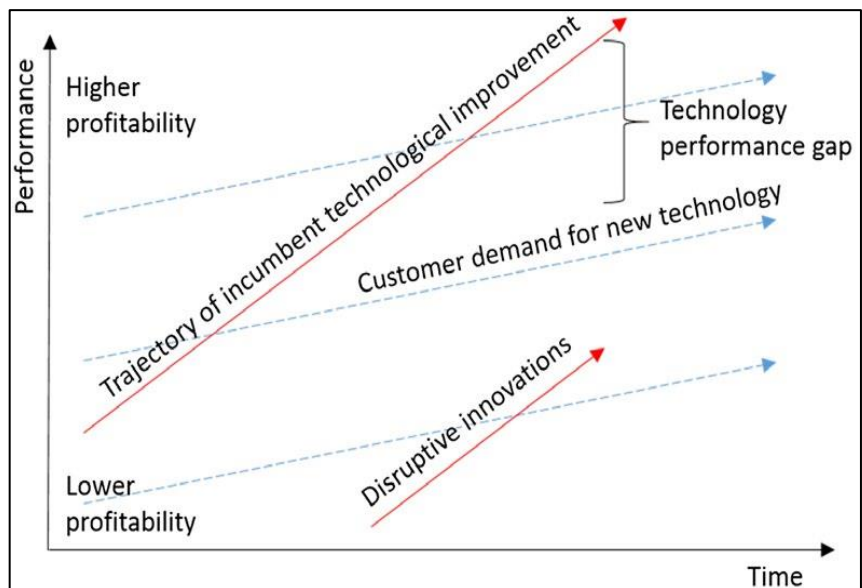
Matt Wise

Summary

- Goal:** We explore the potential for language agent teams to break down and answer questions with multiple reasoning elements or inputs (“multi-hop reasoning”)
- Method:** Given limited input data and paired with effective retrieval, a team of language agents was able to:
 - Break down complex question A into smaller questions
 - Divide out and complete research tasks
 - Re-evaluate and do follow-up research
 - Synthesize findings and generate answer to question A
- Application:** Used for the challenging process of evaluating “disruptive innovation,” in partnership with investment fund Rose Park Advisors (RPA)
- Results:** In blind expert eval, multi-agent model yielded “expert”-level answers >50% of times (vs. 18% for GPT-4)
- Input:** The name of a company
- Outputs:** Detailed report, summary memo for company

Background

- Motivation:** Can language agents break down a complex question, find sub-answers, and work back to a meaningful answer to original question? (see refs. below for related work)
- Flexible, simpler implementation vs. knowledge graph
- Requires understanding context, managing ambiguity, and maintaining coherence across multiple steps
- Applied to the complex business theory of “disruptive innovation” (Christensen, 1997)
 - Theory is usually misunderstood
 - Evaluation is often complicated by marketing jargon
 - Previously considered impossible for computers – “I believed [your] model would never have the insight to articulate disruptive innovation with any degree of accuracy and certainty”



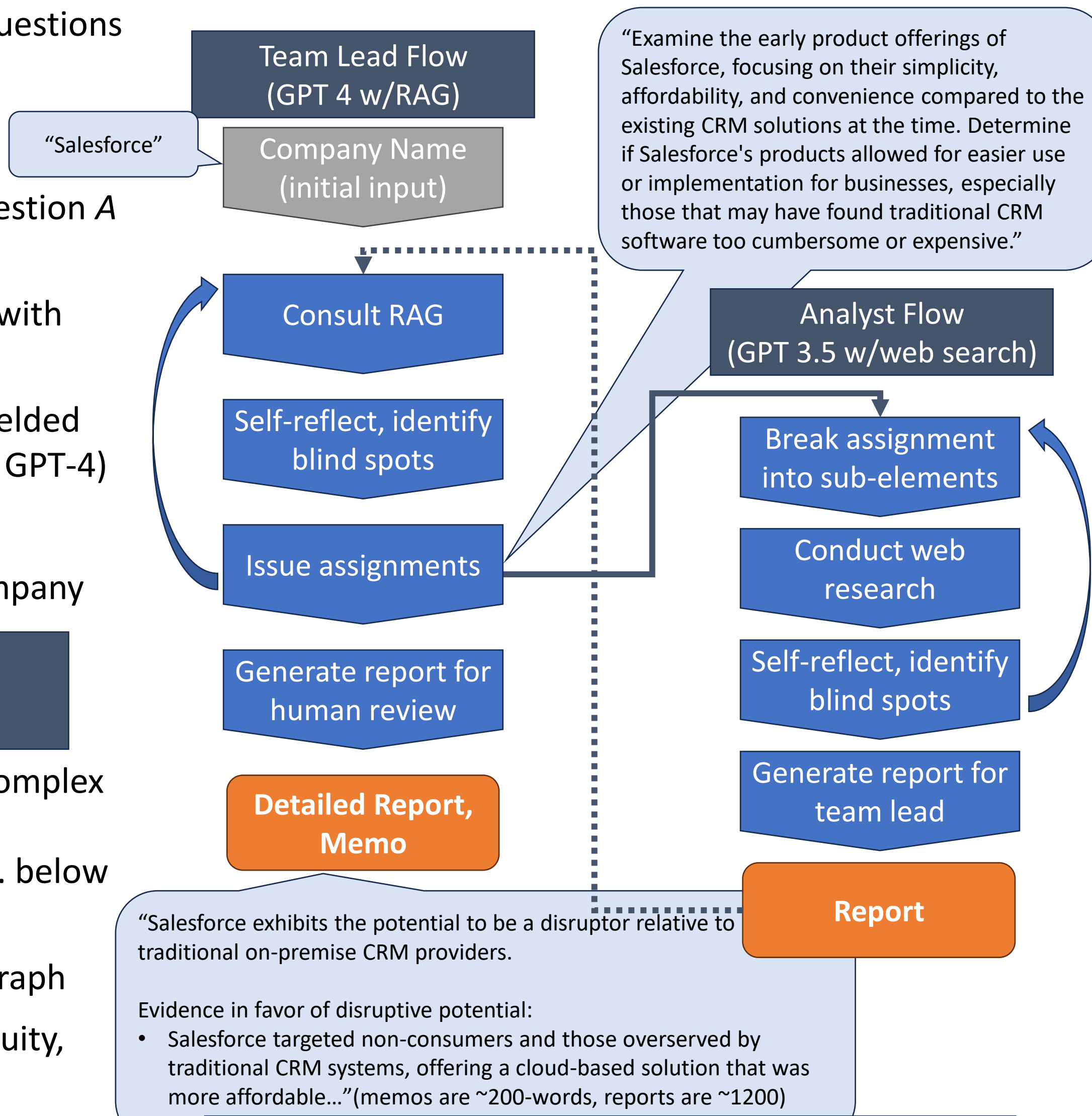
(fig. from Christensen, 1997)

Dataset

- Model evaluated on RPA’s previous investment decisions
- To reflect real-world search for new companies, we elected to use minimal inputs into the model
 - Input is a company name; model does its own research

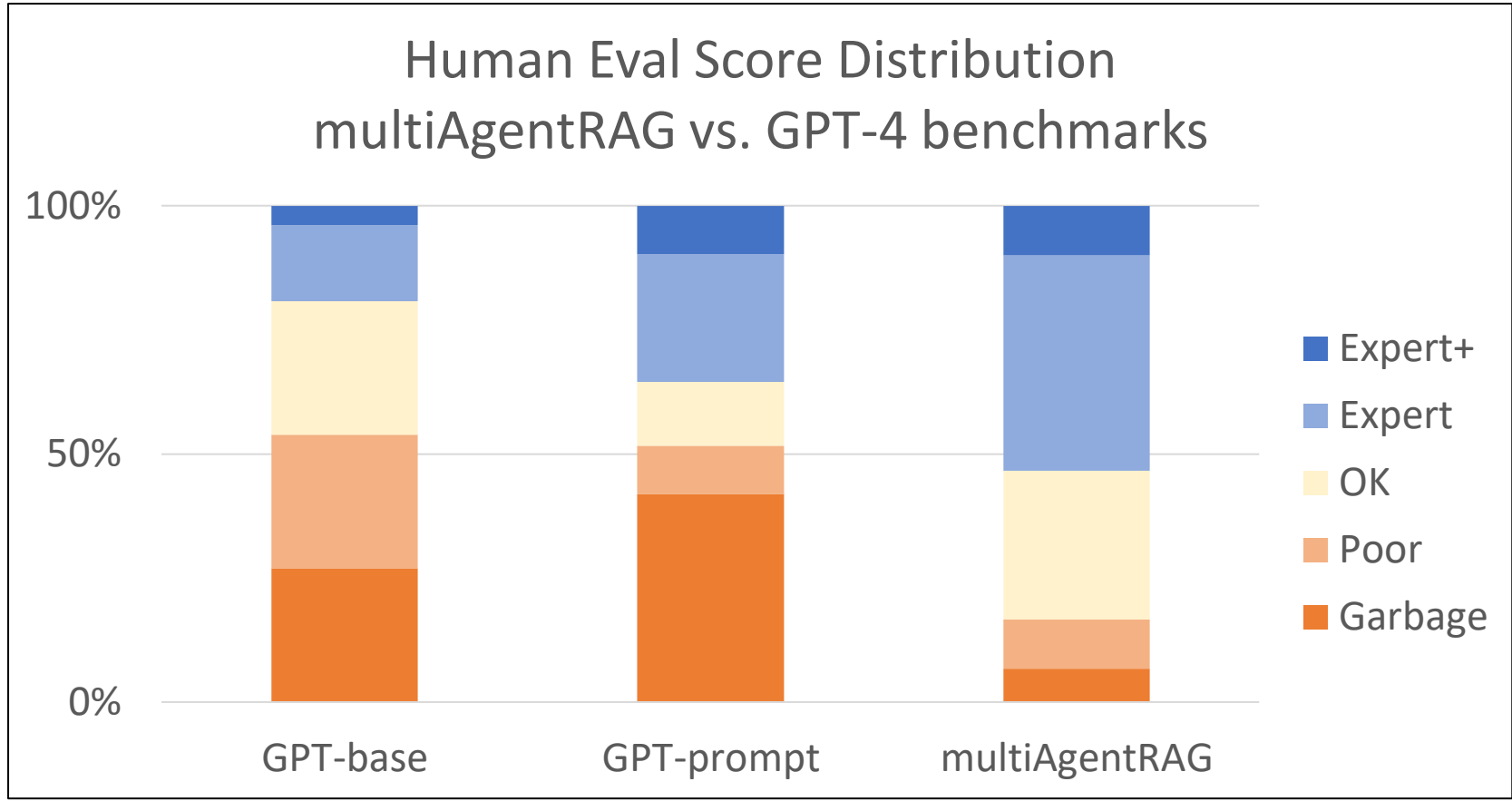
Technical Methods

- Language Agents:** Team Lead + team of analysts
- Retrieval augmented generation (RAG):** Connected to high-quality academic research on disruptive innovation. Used by team lead to for evaluation and decision making
- Web Search:** Used by analyst agents to get relevant data



Evaluation

- Method:** Blind expert human eval from the Chief of Staff to the author of the theory
 - Graded thousands of Harvard MBA papers on disruptive innovation
- Results:** The model outputs were rated at “expert” level 53% of the time (compared with 18% for base GPT-4)



GPT-base: Basic GPT-4 prompt. GPT-prompt: GPT-4 with extensive prompting to match multiAgentRAG

ACKNOWLEDGEMENTS: Thanks to Rose Park Advisors for sharing their time and expertise, especially to Cliff Maxwell, Spencer Christensen, and Chris Calder

KEY REFERENCES
Clayton M. Christensen. *The Innovator's Dilemma: When new Technologies Cause Great Firms to Fail*. HBS Press, 1997
Patrick Lewis, et.al "Retrieval-augmented generation for knowledge-intensive nlp tasks," 2021.
Theodore R. Sumers, Shunyu Yao, Karthik Narasimhan, and Thomas L. Griffiths. "Cognitive architectures for language agents," 2023.
Xing Cao, Yun Liu. "ReLMKG: reasoning with pre-trained language models and knowledge graphs for complex question answering." *Appl Intell* 53, 12032–12046 (2023).