How JP Morgan Built An AI Agent for Investment Research with LangGraph | LangChain Interrupt

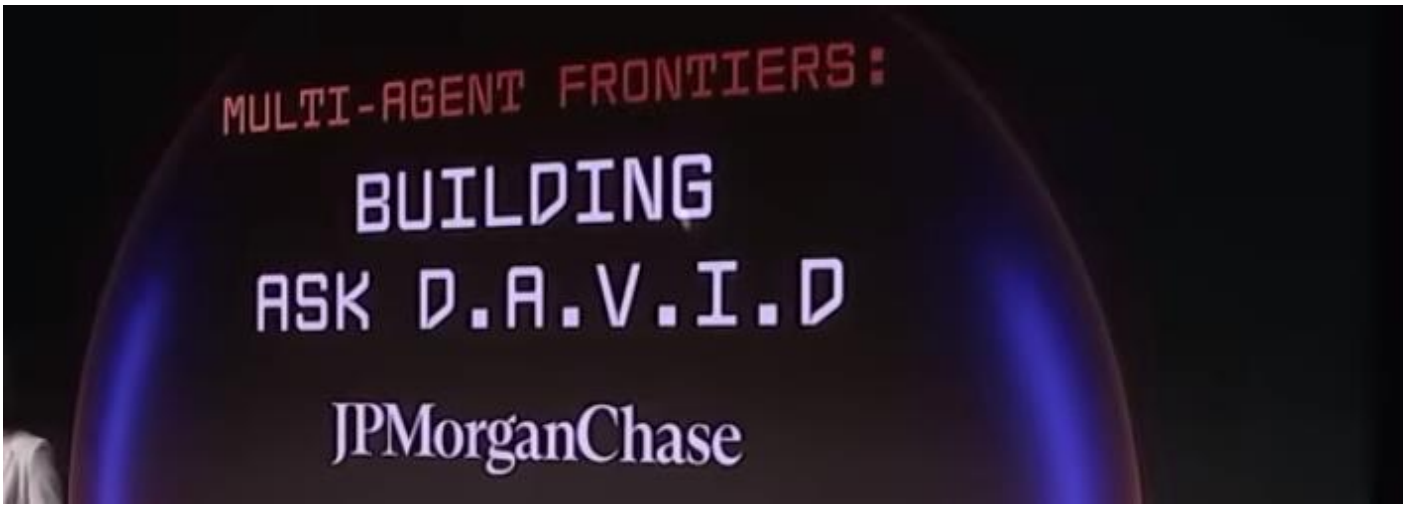
LangChain
134K subscribers

Subscribe

2.1K | Share | Clip | Save | ...

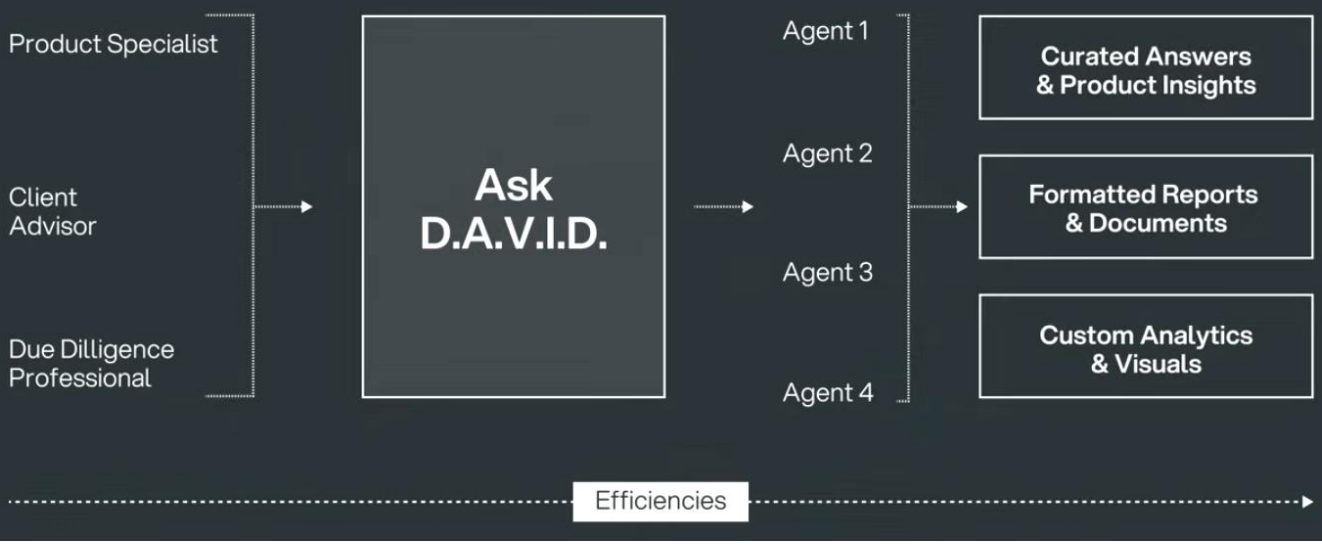
96,241 views May 29, 2025 #LangChain #LangGraph #MultiAgent
Watch all of our recorded sessions from Interrupt here: https://interrupt.langchain.com/video...

David Odomirok and Zheng Xue from JP Morgan Chase Private Bank reveal how they built "Ask David" - a sophisticated multi-agent AI system designed to automate investment research for thousands of financial products. With billions of dollars in assets at stake, this isn't just another chatbot - it's an enterprise system built with human oversight for high-stakes financial decisions.



Tomorrow

J.P.Morgan



Domain Knowledge

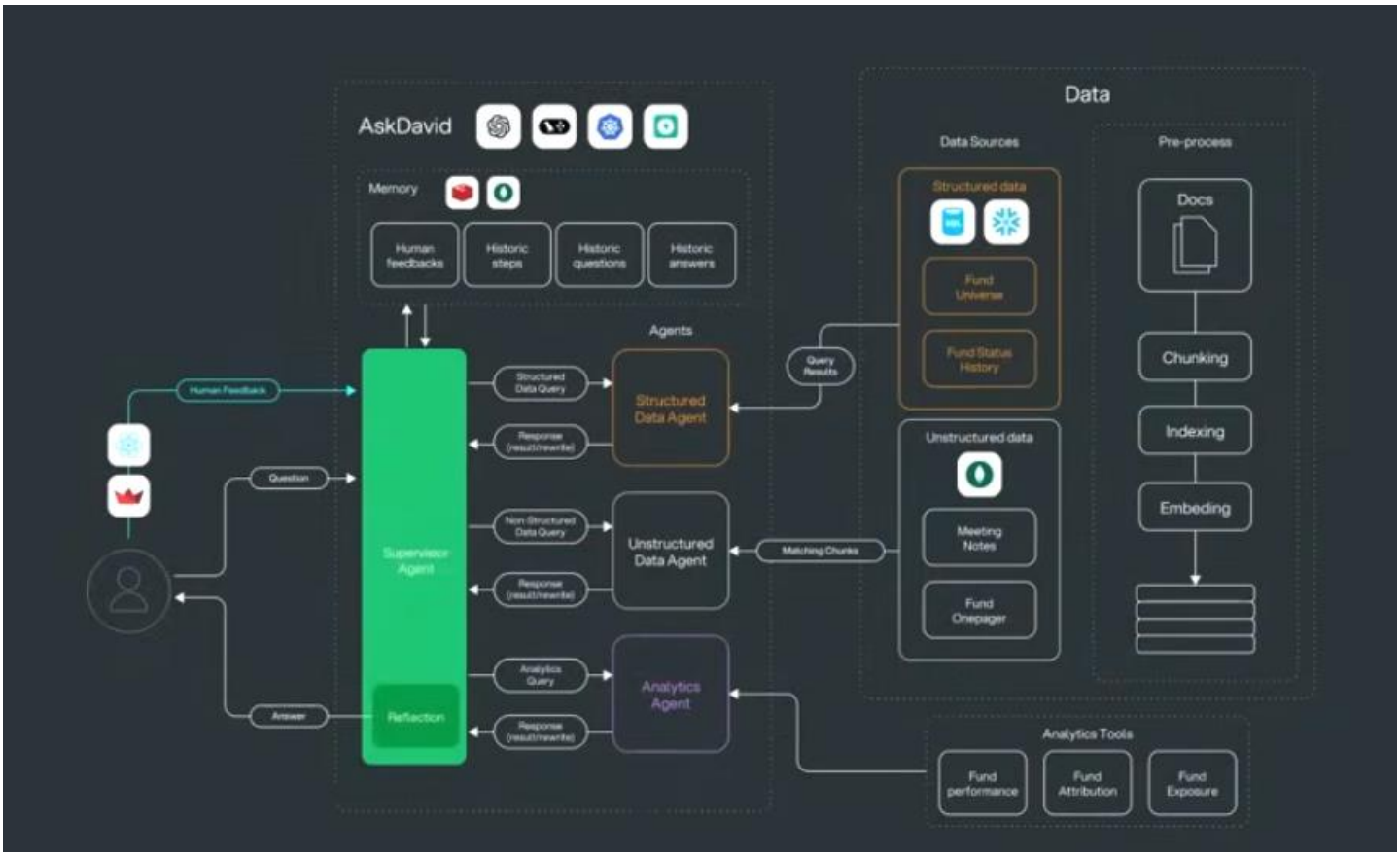
Structured Data →
RDS, Data mesh

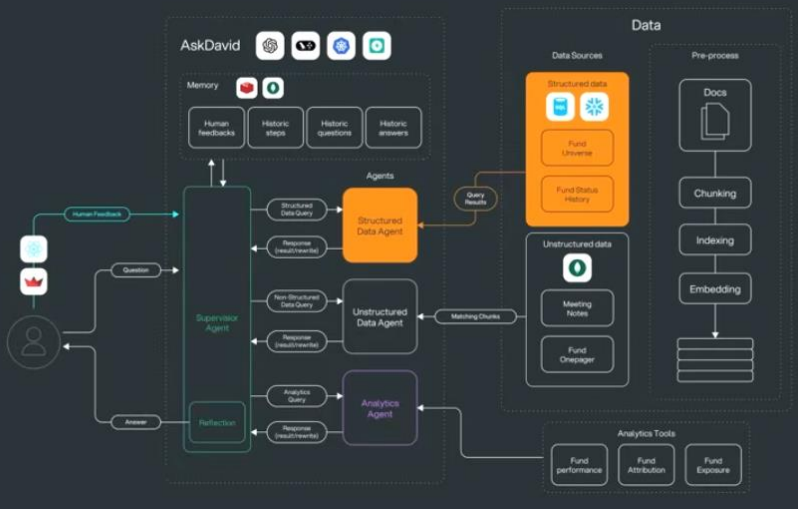
Unstructured Data →
Emails, Meeting notes, Zoom

Proprietary Models and Analytics →
APIs, Libraries

The Approach

J.P.Morgan

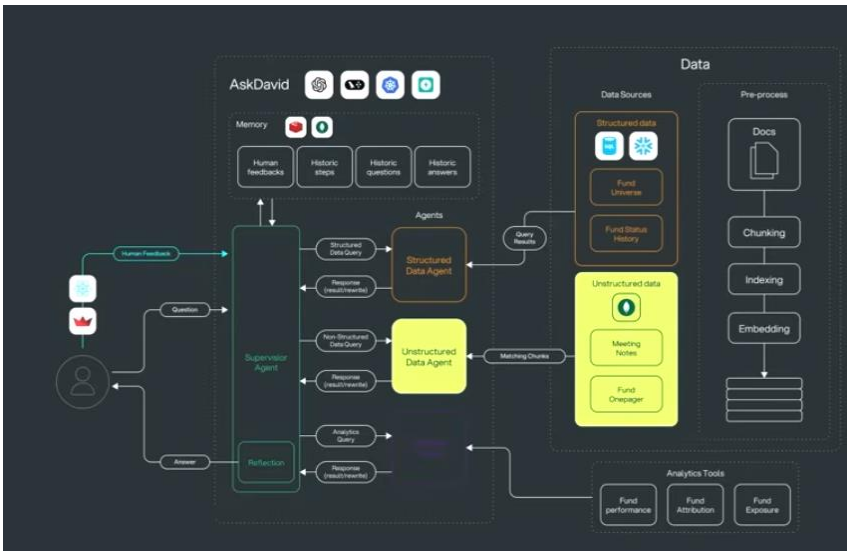




Supervisor Agent:

- Acts as the orchestrator
- It receives questions from users and determines a plan to provide answers
- This may involve engaging one or more of the specialized sub-agents, depending on the nature of the query
- It incorporates human in the loop to ensure the accuracy and reliability
- It also has access to long term memory

Structured Data AI Agent:
This agent translates natural language queries into SQL or GraphQL to retrieve answers from databases or APIs, and then uses a large language model (LLM) to summarize the results.

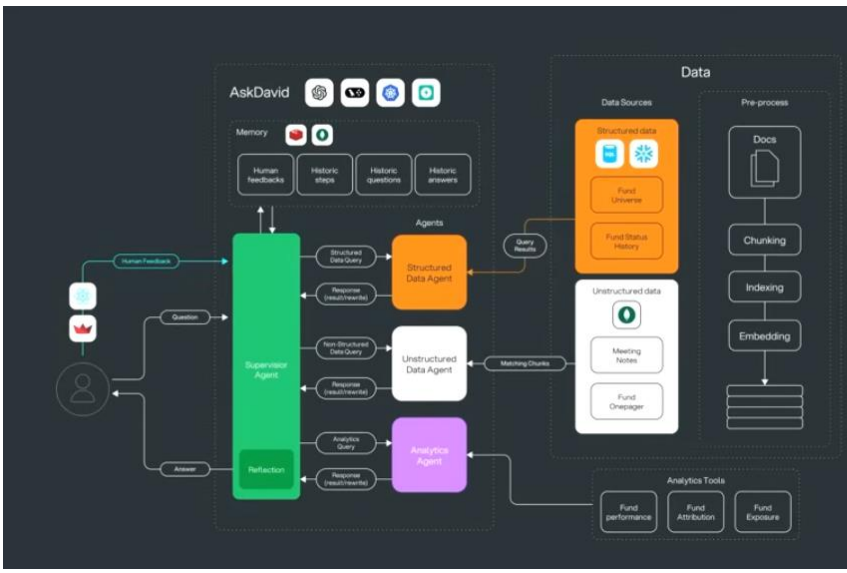


Supervisor Agent:

- Acts as the orchestrator
- It receives questions from users and determines a plan to provide answers
- This may involve engaging one or more of the specialized sub-agents, depending on the nature of the query
- It incorporates human in the loop to ensure the accuracy and reliability
- It also has access to long term memory

Structured Data AI Agent:
This agent translates natural language queries into SQL or GraphQL to retrieve answers from databases or APIs, and then uses a large language model (LLM) to summarize the results.

Unstructured Data AI Agent:
This RAG agent utilizes a vector database and hierarchical index retrieval to store documents and search for relevant information chunks, followed by summarization using an LLM.



Supervisor Agent:

- Acts as the orchestrator
- It receives questions from users and determines a plan to provide answers
- This may involve engaging one or more of the specialized sub-agents, depending on the nature of the query
- It incorporates human in the loop to ensure the accuracy and reliability
- It also has access to long term memory

Structured Data AI Agent:
This agent translates natural language queries into SQL or GraphQL to retrieve answers from databases or APIs, and then uses a large language model (LLM) to summarize the results.

Unstructured Data AI Agent:
This RAG agent utilizes a vector database and hierarchical index retrieval to store documents and search for relevant information chunks, followed by summarization using an LLM.

Analytics Agent:
ReAct agent with access to analytics APIs as tools.

Graph: Workflow

Example: Why was a fund terminated?

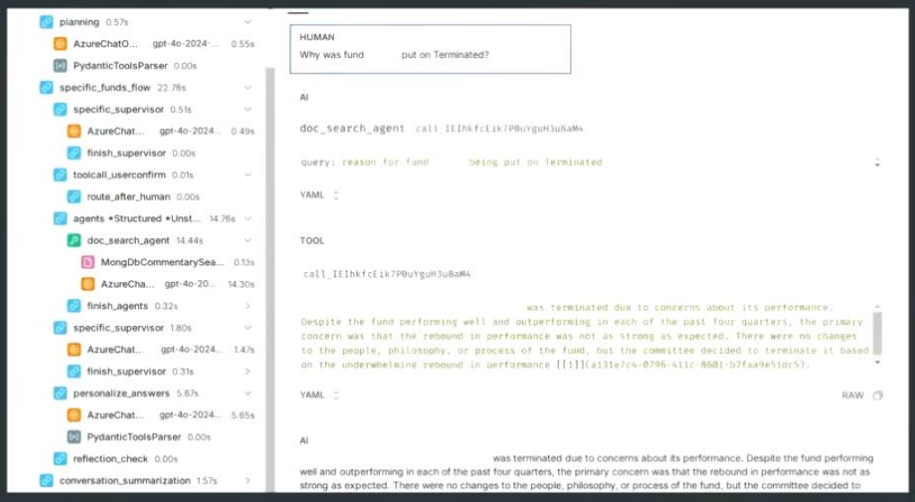
J.P.Morgan

Planning
- Determines the intention
- Invokes the corresponding subgraph

Subgraph - Specific Funds Flow
- delegates to doc_search_agent

Doc Search Agent
- Conducts mongoDB search to retrieve relevant information

Supervisor
- Personalization based on the role of human
- Reflect to ensure the accuracy and relevancy of answers
- Summarize conversation, update memory, and return answer



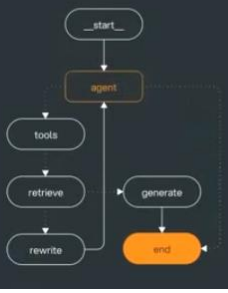
Lessons Learned

J.P.Morgan

Lesson 1: Start Simple and Refactor Often

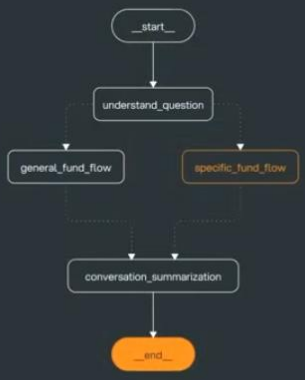
1

2



3

4



J.P.Morgan

Lesson 2: Evaluation Driven Development

→ GenAI has much shorter dev phase than traditional AI projects, which enables rapid prototyping and iteration

→ Domain Confidence requires extensive evaluation and refinement

- Independent Evaluation of Subagents: Assess subagents and the whole flow separately.
- Diverse Metrics for Agent Types: Tailor metrics to the agent design.
- Evaluation with/without Ground Truth: Not all metrics need ground truth; valuable insights can be gained without it.
- LLM as Judge + Human SME Review: Use automated evaluations to scale and reduce SME workload.

J.P.Morgan

Lesson 3: Human SME in the Loop

Initial Accuracy	Quick Improvements	Additional Improvements	Last Mile
Apply GenAI directly	• Chunking strategy • Search methodology • Prompt engineering	• Workflow enhancements • Sub-graphs	• Hardest • May be not achievable • Human SME in the loop

Iterate fast, evaluate early, include Human SME in the loop

J.P.Morgan

Thank you!

jane.xue@jpmchase.com
david.j.odomirok@jpmorgan.com

May 14th, 2025

@ Interrupt,
The AI Agent Conference
by Langchain