PYCHINA
Python 中国社区

基于 LangGraph 构建可靠的 AI Agent 系统

张海立（沧海九粟），LangChain Ambassador

张海立
Harry ZHANG

LangChain 社区大使

《LangChain实战》作者

webup

zhanghaili0610

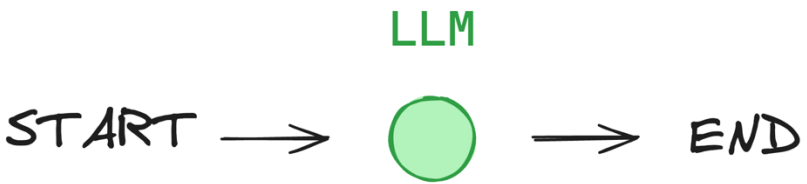
沧海九粟

Applications that can reason.

Powered by LangChain.

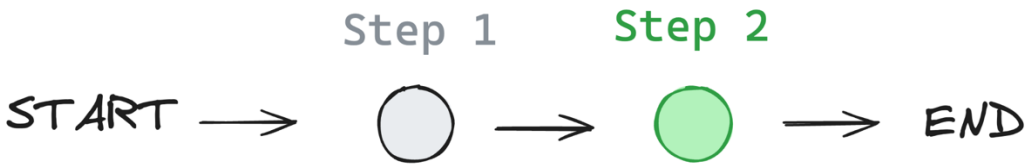
Building AI Systems

A solitary language model is fairly limited…

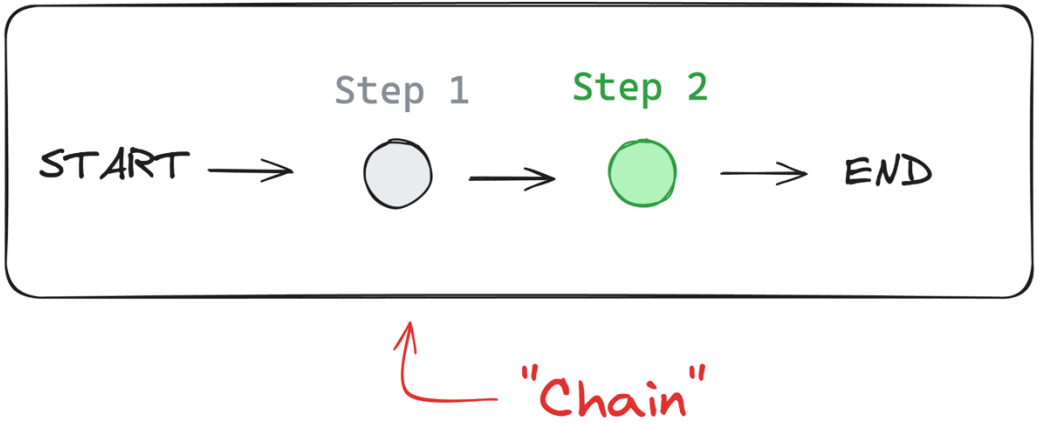


… it lacks relevant context.

Good LLM applications follow a control flow.

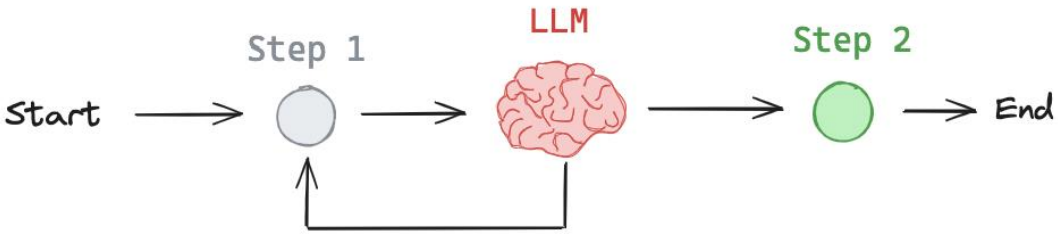


This control flow forms a “chain”



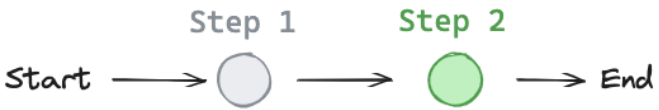
Example: search -> LLM (RAG) chain

Agent ~= control flow defined by an LLM



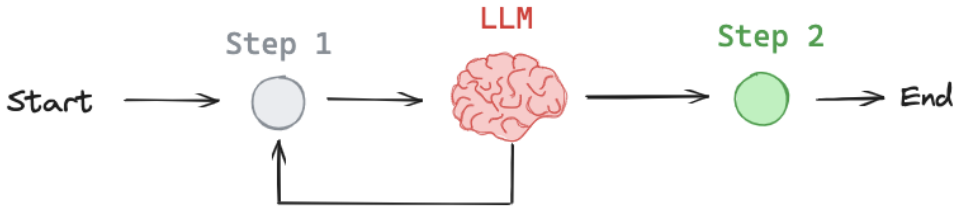
Chain

Developer defined control flow

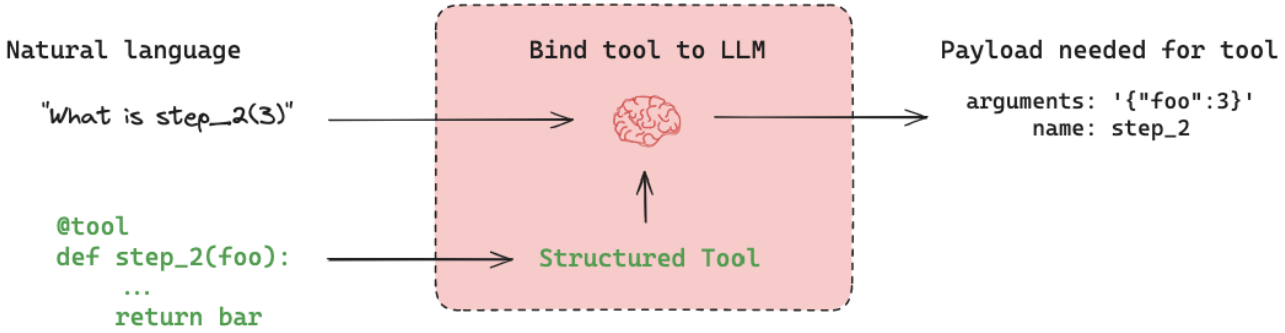


Agent

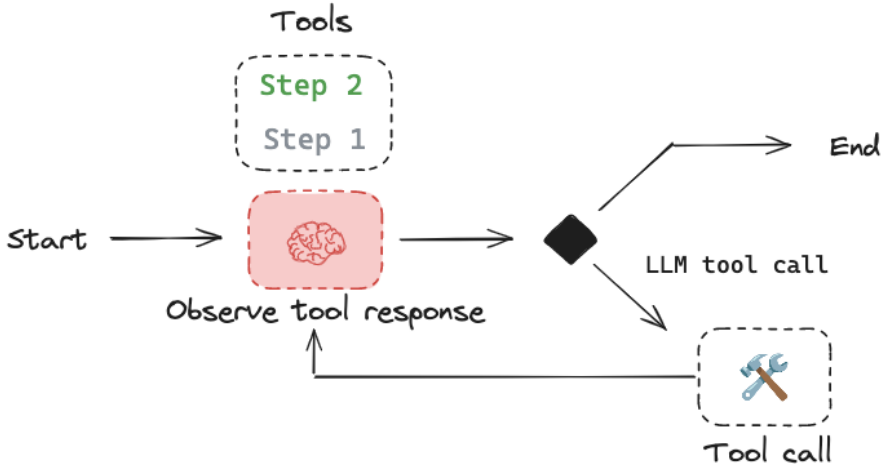
LLM defined control flow



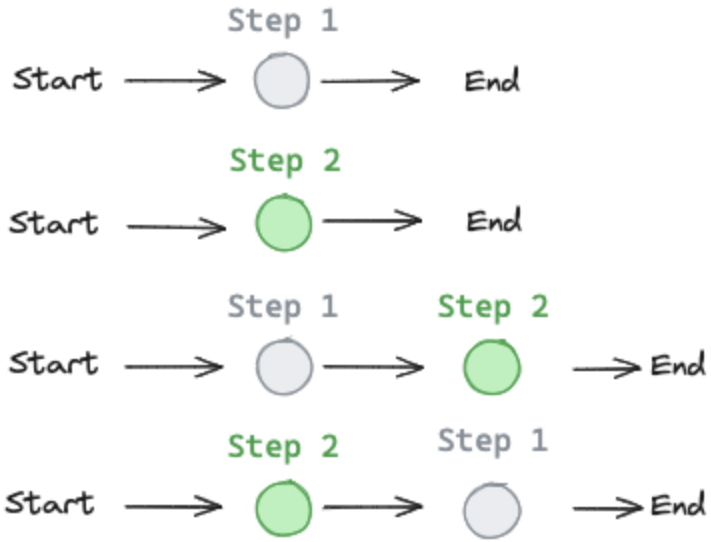
Agents typically use tools-calling to execute steps



Simplest design is a `for loop` (ReAct)



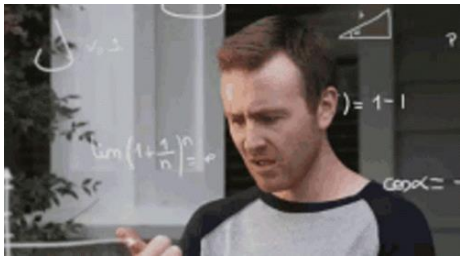
ReAct agents are flexible: any state possible!

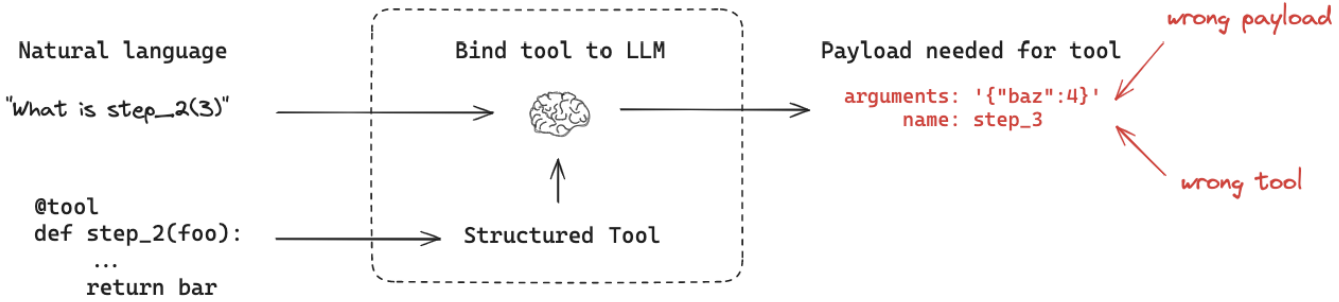


… but they suffer from poor reliability

… often caused by

- Task ambiguity
- LLM non-determinism
- Tool misuse
- Tool dependencies
- … and more!

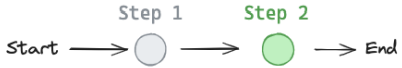




Reference: https://www.reddit.com/r/LocalLLaMA/comments/1648pav/thoughts_on_autonomous_llm_agents/

Can we have both?

Chain



Not flexible

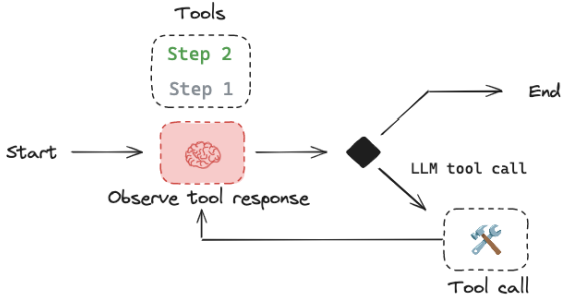
More reliable

?

Flexible

Reliable

Agent (for loop)



Flexible

Less reliable

Introducing LangGraph

What is LangGraph ?

LangGraph applications balance agent control with agency

Its core pillars support:

A **Controllability:** to define both explicit and implicit workflows

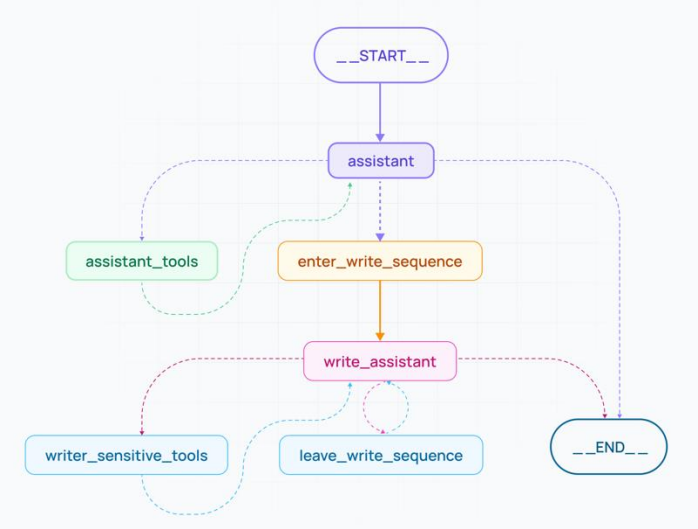
B **Persistence:** to enable human-agent/multi-agent interactions & fault tolerance

C **Human-in-the-loop:** to facilitate human guidance

D **Streaming:** to expose any event (or token) as it occurs

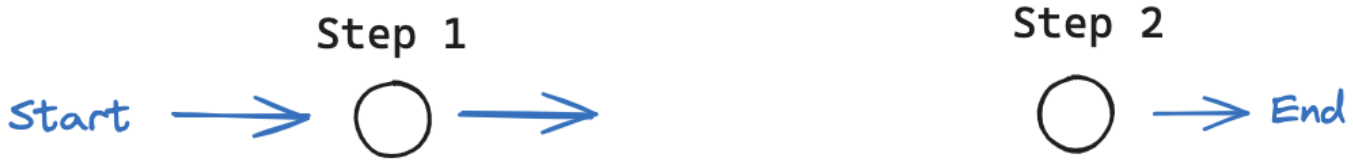
LangGraph also:

- Works with or without LangChain
- Integrates with LangSmith



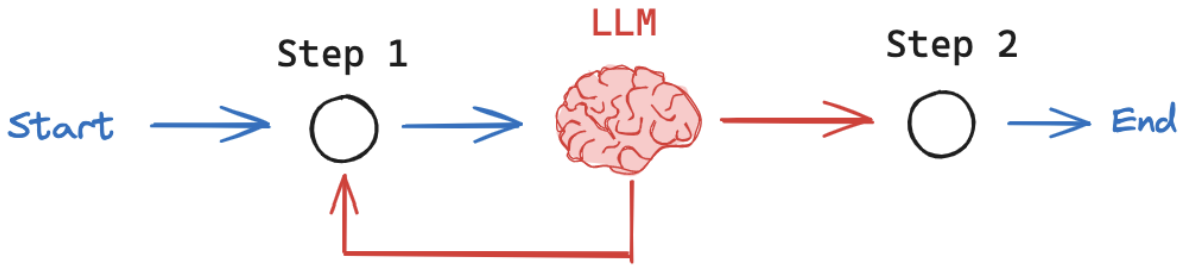
A Controllability

Intuition: Let developer set parts of control flow (reliable)



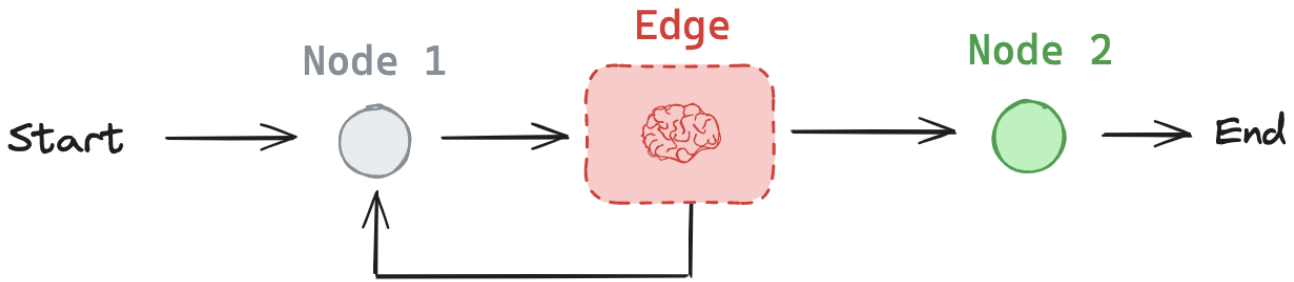
A Controllability

Intuition: Inject LLM to make it an agent (flexible)



A Controllability

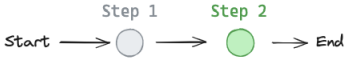
LangGraph : Express control flows as graphs



A Controllability

We can have both!

Chain



Not flexible

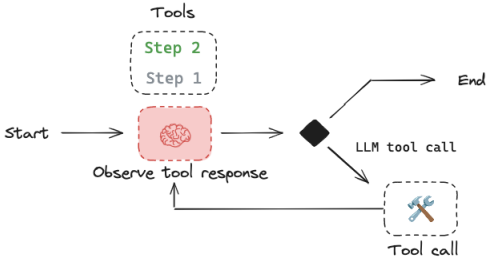
More reliable

LangGraph

Flexible

Reliable

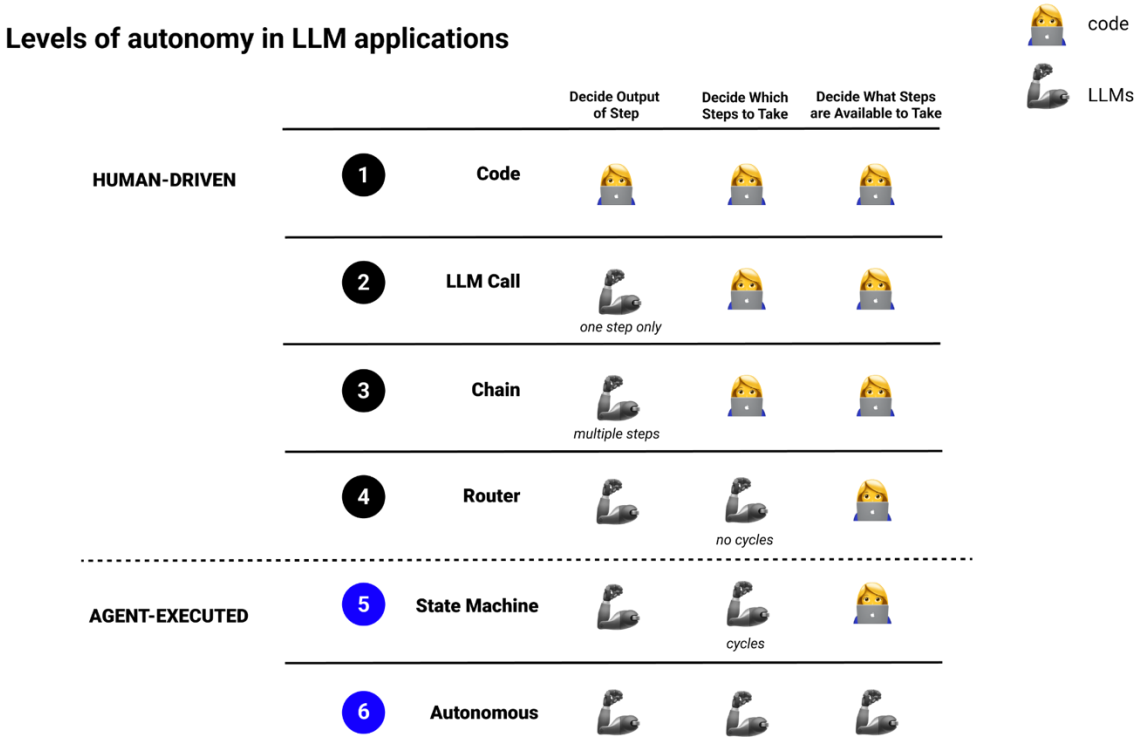
Agent (for loop)



Flexible

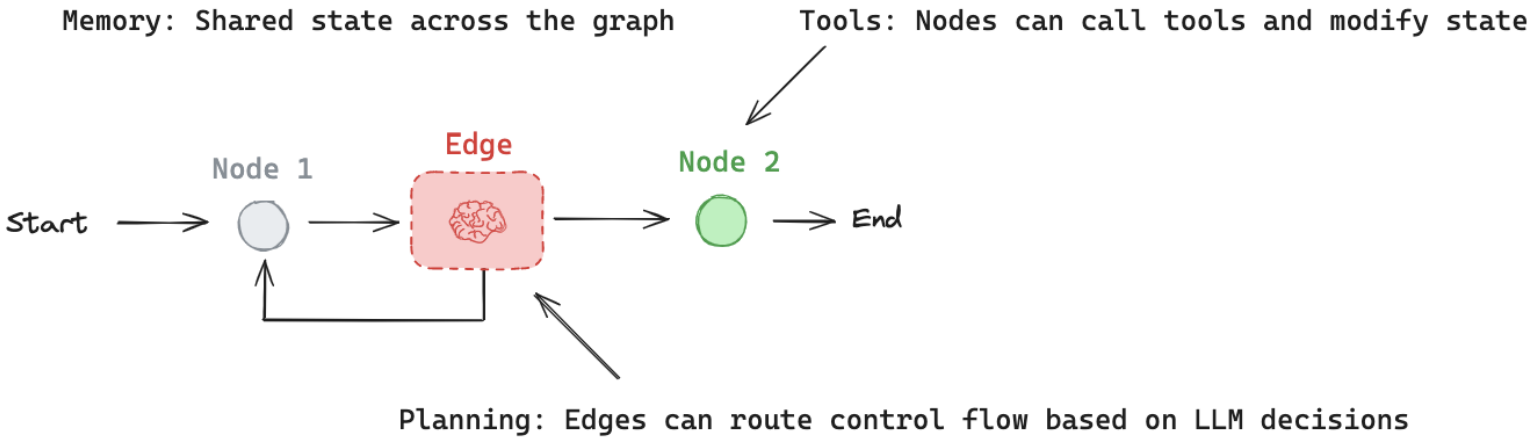
Less reliable

A Controllability



A Controllability

LangGraph Agent



A Controllability

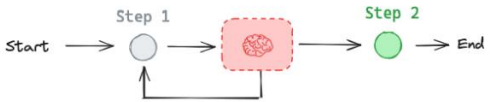
LangGraph allows for developer + LLM-defined control flows

Chain

Not flexible

Most Reliable

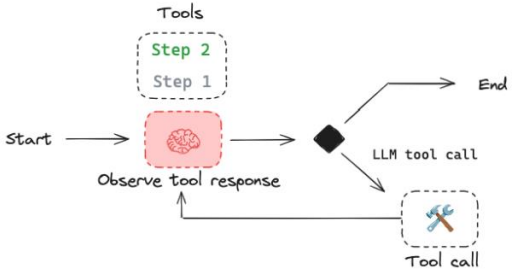
Agent (LangGraph)



More Flexible

More Reliable

Agent (for loop)

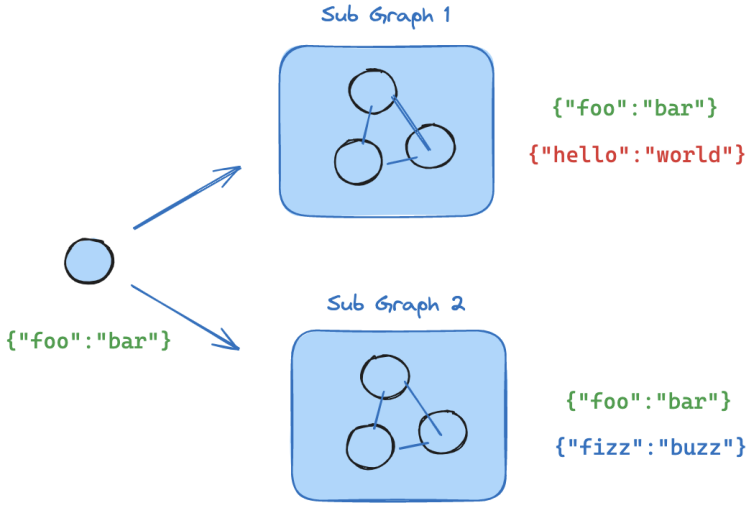


Most Flexible

Least reliable

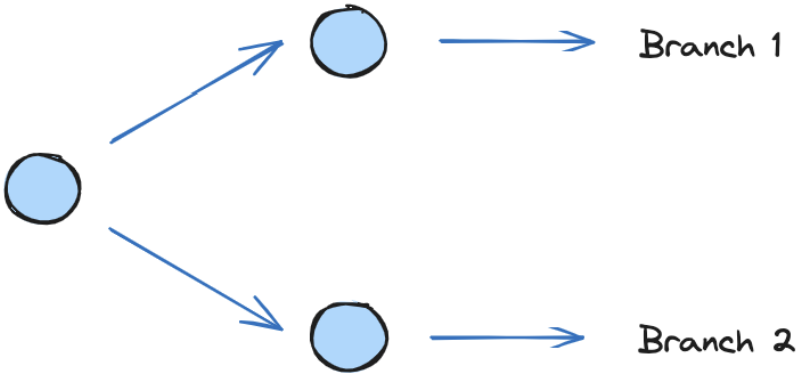
A Controllability

Subgraphs enable complex system design by managing states separately



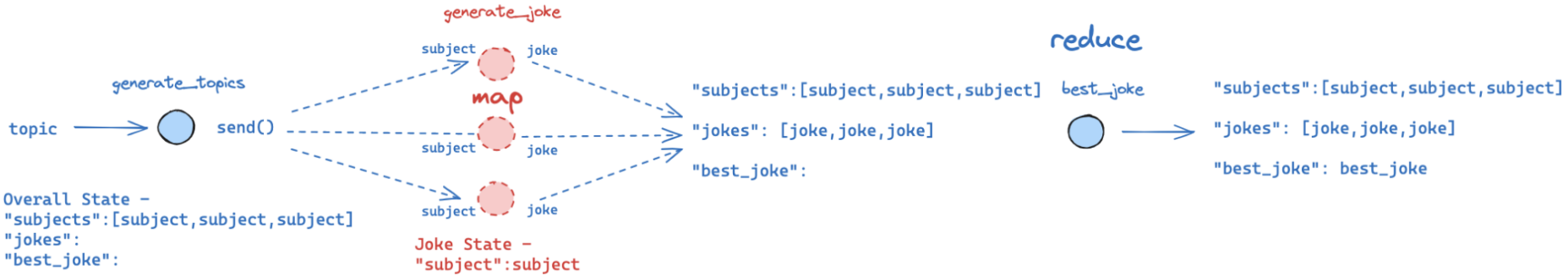
A Controllability

Branches enable parallel execution of nodes to speed up overall graph operation



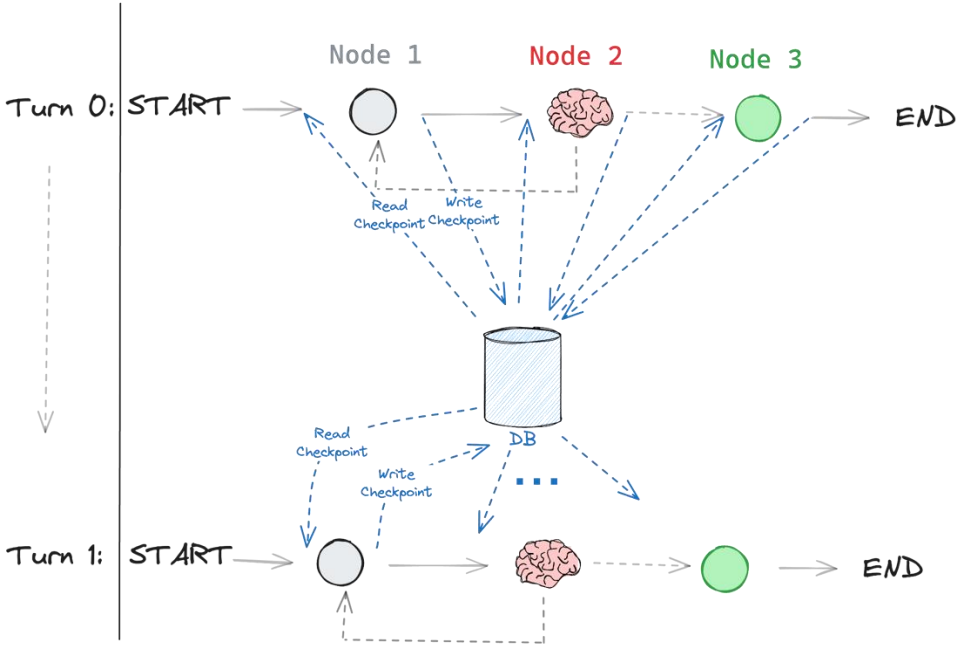
A Controllability

Map-reduce branches enable efficient parallel processing and flexible execution



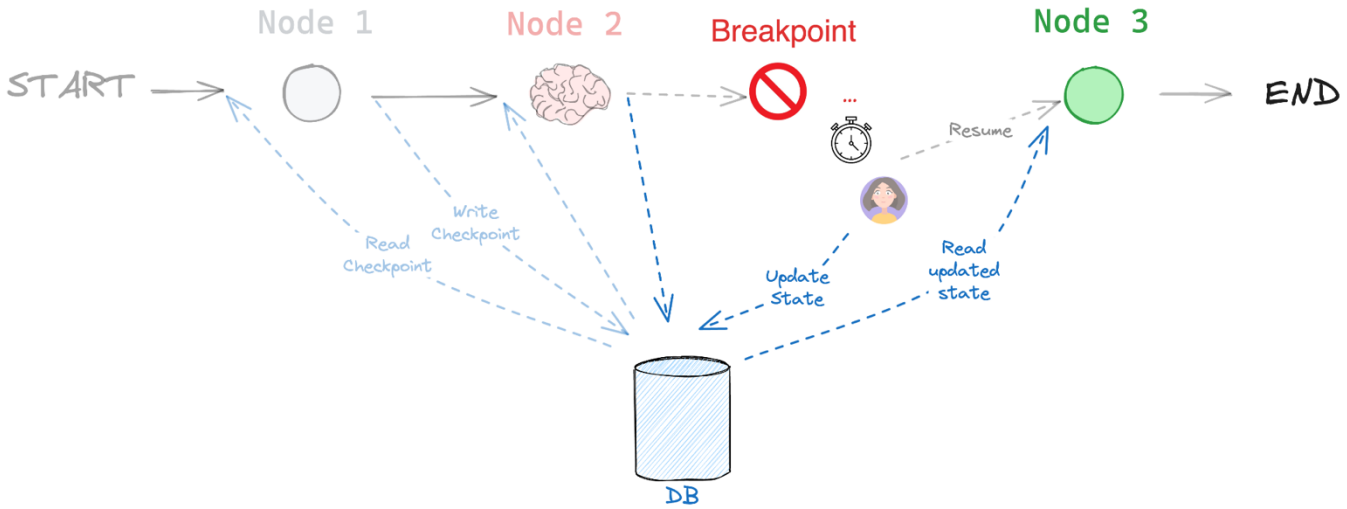
B Persistence

Provides “Memory”



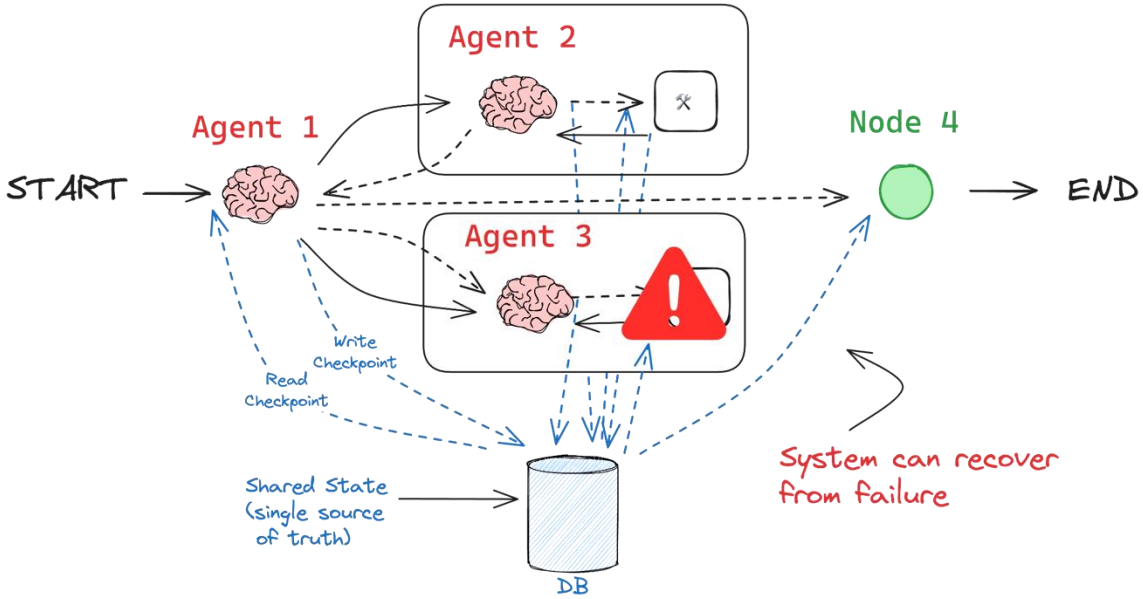
B Persistence

Enables human-agent interactions



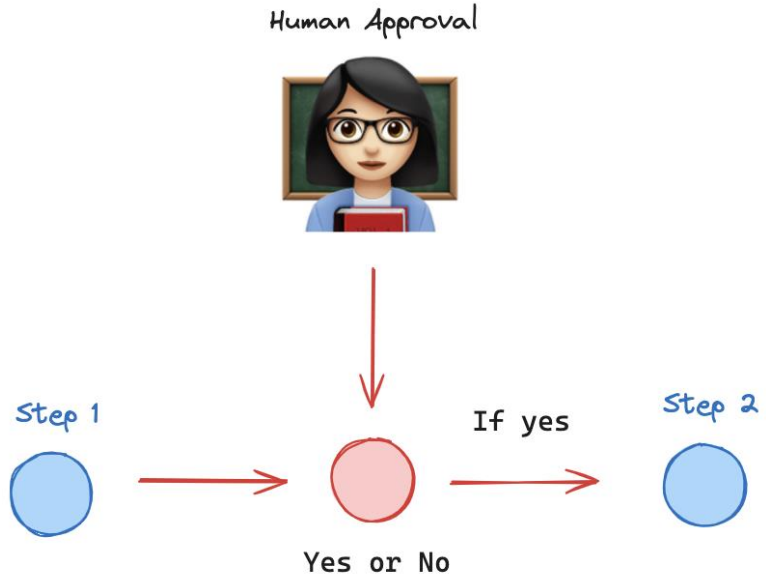
B Persistence

Enables fault-tolerant multi-agent interactions



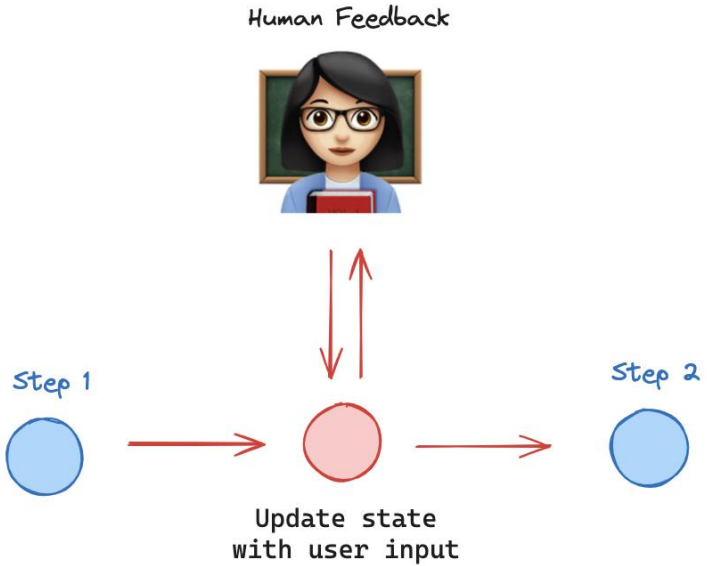
C Human-in-the-loop

Breakpoints enable human-in-the-loop interactions



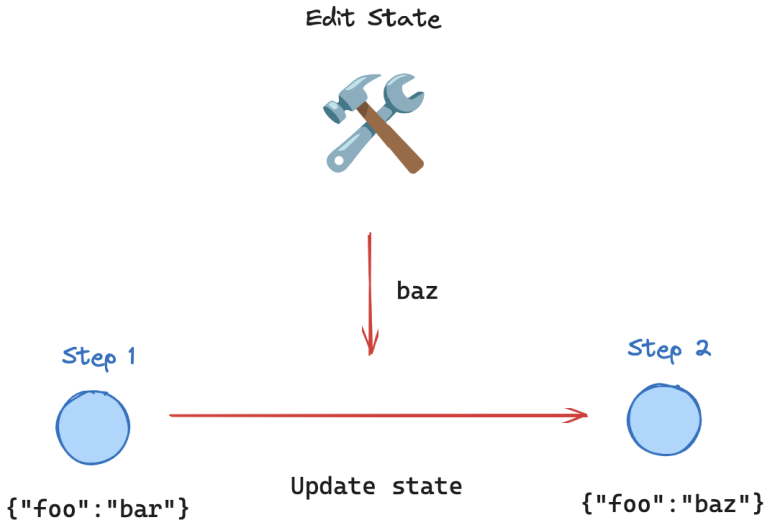
C Human-in-the-loop

At a breakpoint, we can wait for human input and then proceed



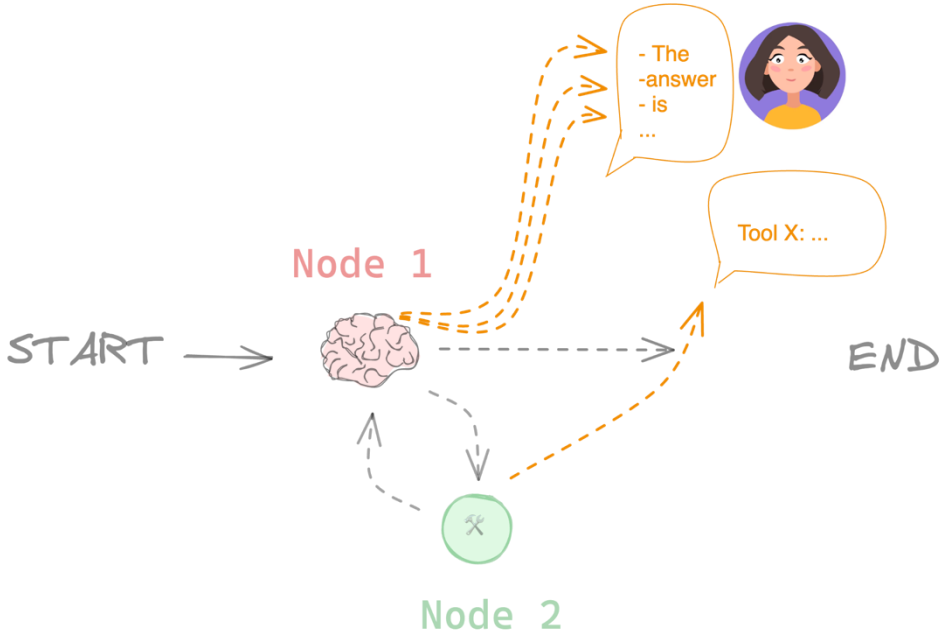
C Human-in-the-loop

Real-time editing to graph state during execution

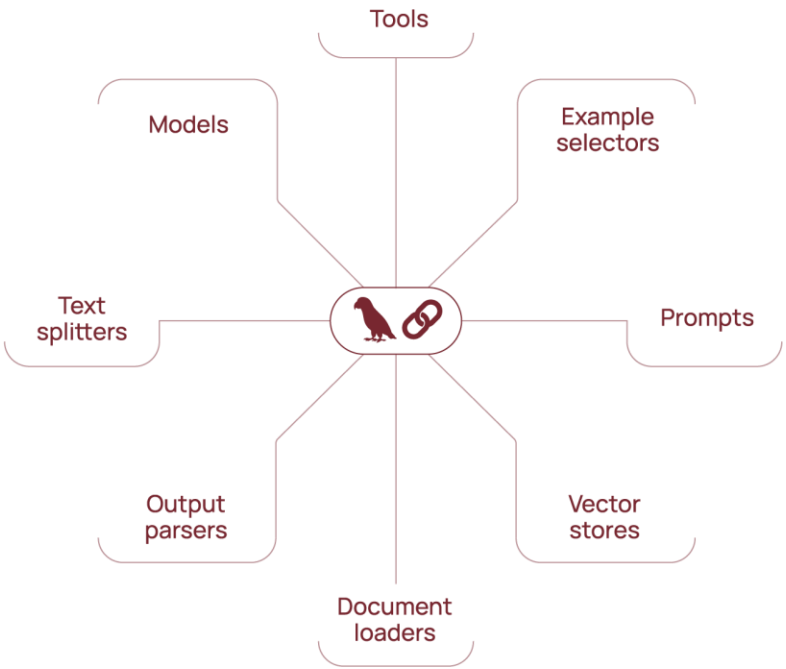


D Streaming

First-class support for token and event-level streaming



LangGraph within LangChain's Ecosystem



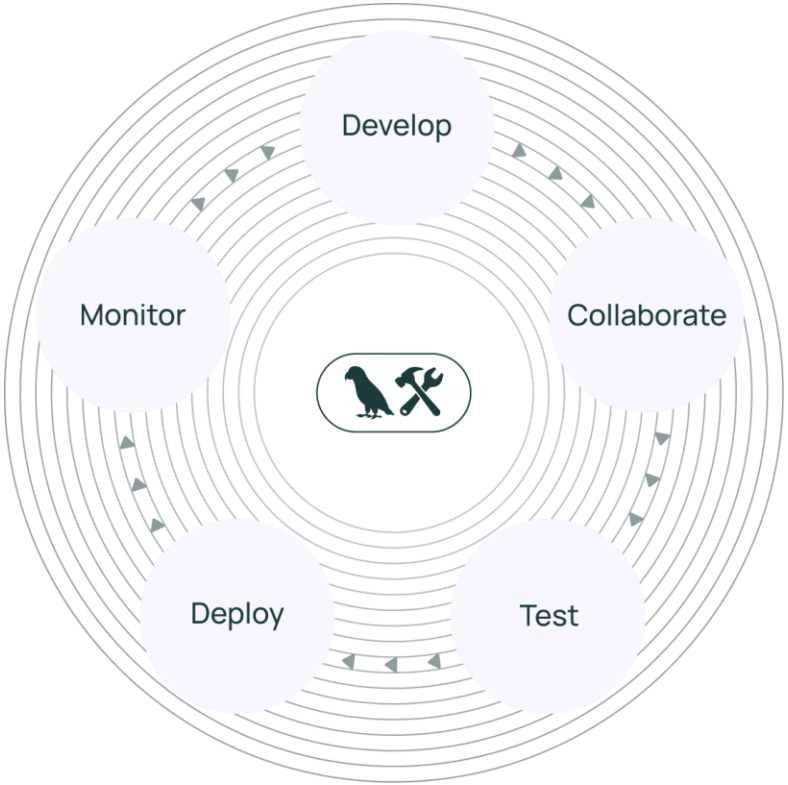
LangChain

LangChain's AI abstractions and integrations make it the #1 choice for developers when building with GenAI

15M+
Monthly Downloads

100K+
Apps Powered

2K+
Contributors



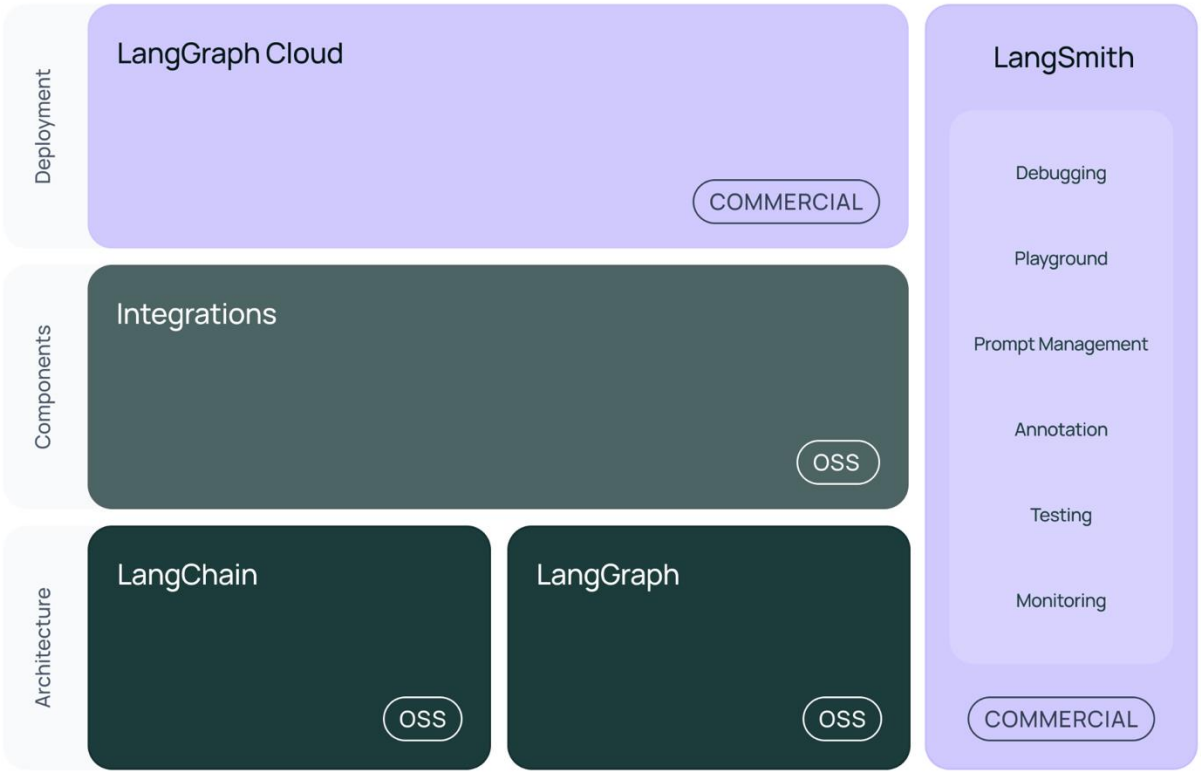
LangSmith

LangSmith is a unified DevOps platform, purpose-built for LLM applications

100K+
Users signed up

300M+
Traces logged

30K+
Monthly active teams


Deployment
LangGraph Cloud
COMMERCIAL
Components
Integrations
OSS
Architecture
LangChain
OSS
LangGraph
OSS
LangSmith
Debugging
Playground
Prompt Management
Annotation
Testing
Monitoring
COMMERCIAL

PyCon China 2024

For Good . For fun.

2024/11/23 中国 上海