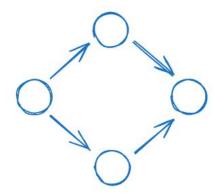


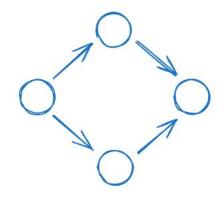
LangGraph

Build controllable agents



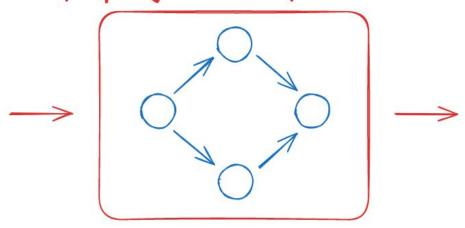
LangGraph

Build controllable agents



LangGraph Platform

Deploy agents to production



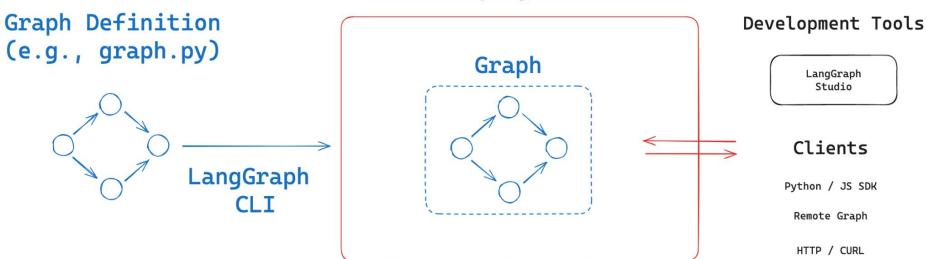
Productionize Existing LangGraph Features

- Streaming
- Checkpoints / short-term memory
- Human-In-The-Loop
- Long-term memory

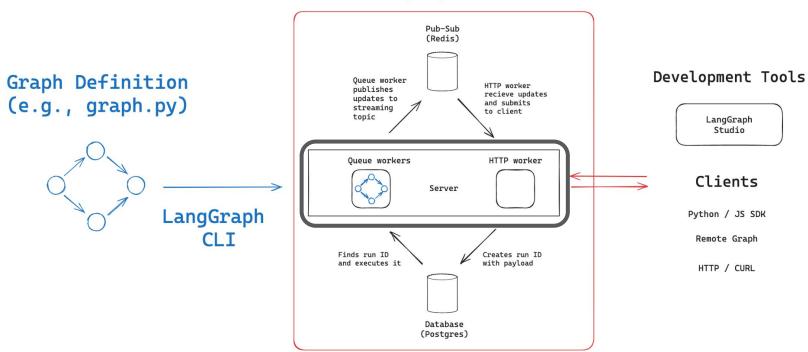
- Streaming
- Checkpoints / short-term memory
- Human-In-The-Loop
- Long-term memory

Also Get New Features For Agent Productionisation

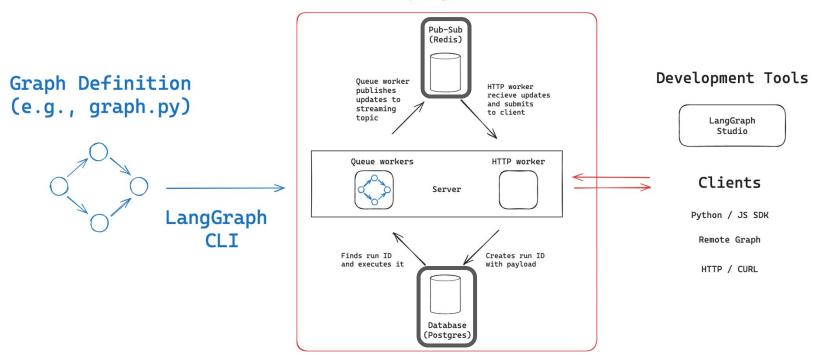
- Agent scheduling: Cron Jobs
- <u>Long-running agents</u>: Background runs / task queue
- <u>Better chat UX</u>: Support for double texting
- Agent configuration: Assistants



- CLI: Builds Docker image w/ code for the server, your graph, and dependencies

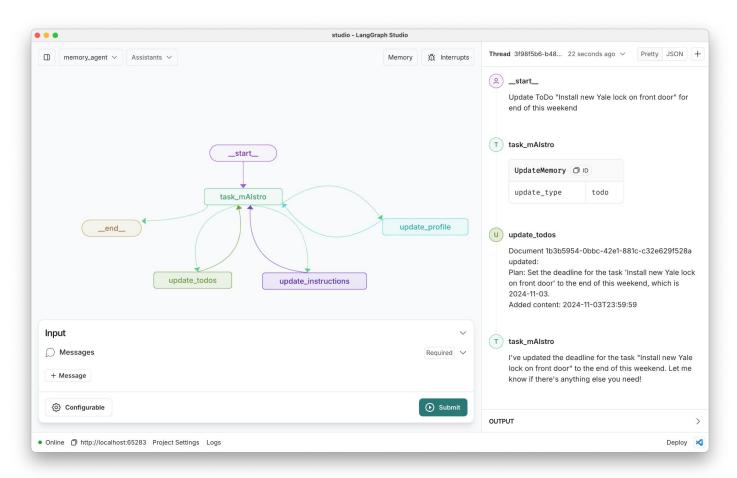


- **HTTP worker**: Client communication, creating run ID for requests to execute graph (e.g., invoke, stream)
- **Queue worker**: Executing the graph



- Redis: Communication between workers (streaming, run cancellation)
- Postgres: Support for short / long-term memory, task queue

- LangGraph <u>Cloud</u> and <u>Studio</u>: Use the <u>CLI</u> to build the Docker image for you



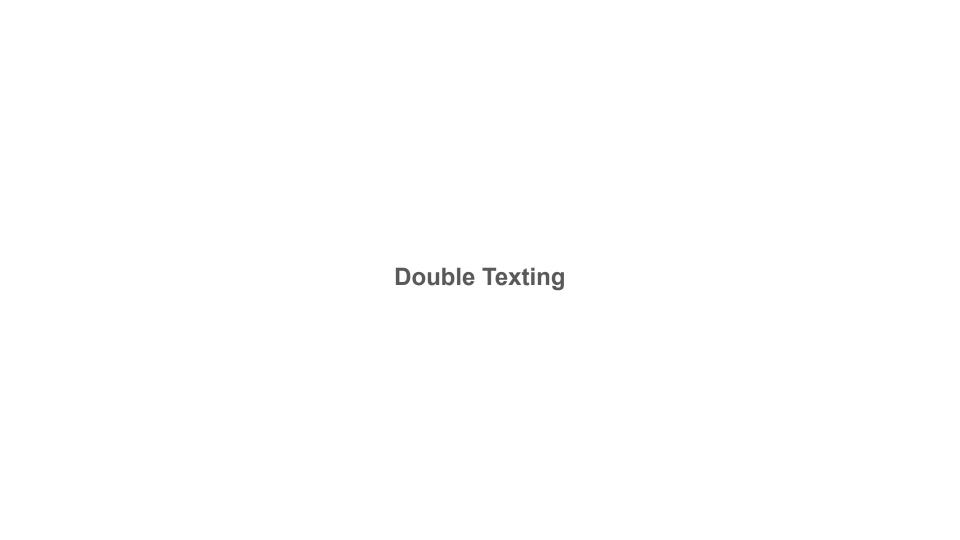
- **Python and JS SDK**: Clients for working with LangGraph server API
- Remote Graph: For working with deployed graphs in LangGraph library

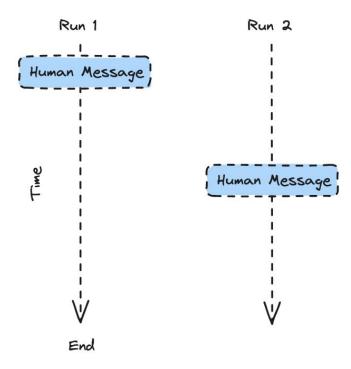
4 Getting started: Self-hosted (Lite)

Self-Hosted Lite

- Free
- Limited to 1M nodes executed / year
- Some missing features
- Quickstart: Just download the CLI!

Option	Manage	Infra	Considerations
LangGraph Cloud	Managed for you	Our infrastructure	Fully managed for easy onboarding
BYOC	Managed for you	Your infrastructure	Data privacy but managed
Self-Hosted (Enterprise)	Managed by you	Your infrastructure	Gives you full control

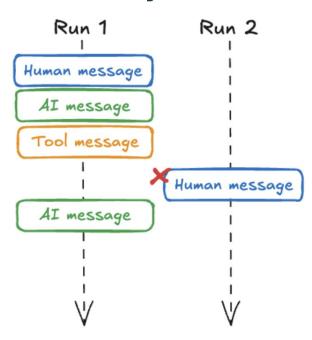




Real-world usage challenge: second request comes before the first one finishes! 4 different approaches.



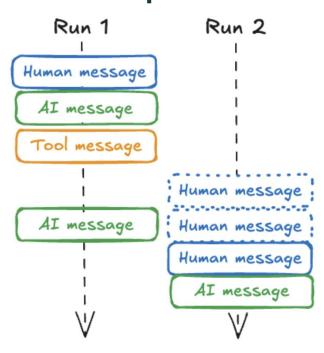
Reject



Rejects any follow up runs and does not allow double texting



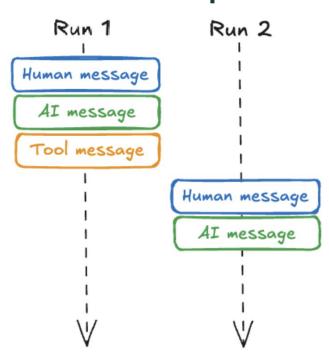
Enqueue



Continues the first run until it completes the whole run



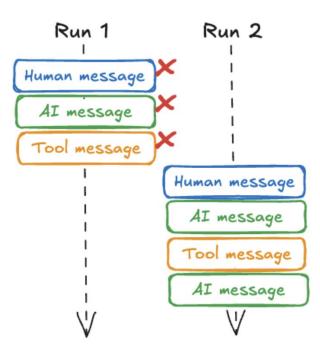
Interrupt



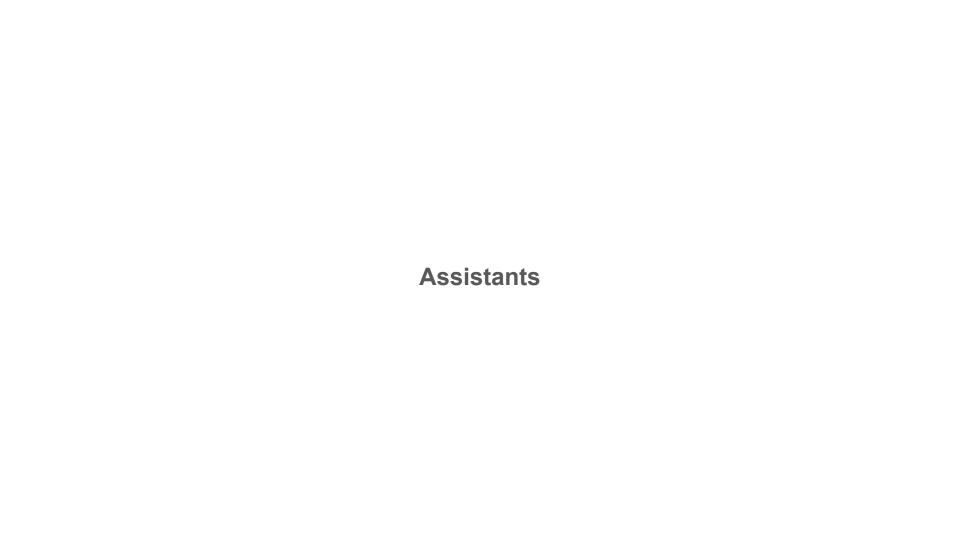
Interrupts the prior run



Rollback



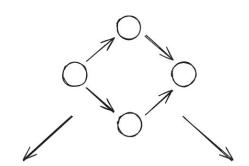
Interrupts and delete the prior run





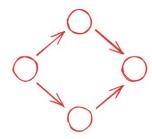


task_mAIstro Graph

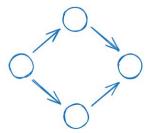


Work Assistant





{user_id: "Lance",
 todo_category: "work"
task_maistro_role: " ... "}



{user_id: "Lance",
todo_category: "personal"
task_maistro_role: ".."}

Deployment cask_mAIstro Graph

