Разработка в NLP

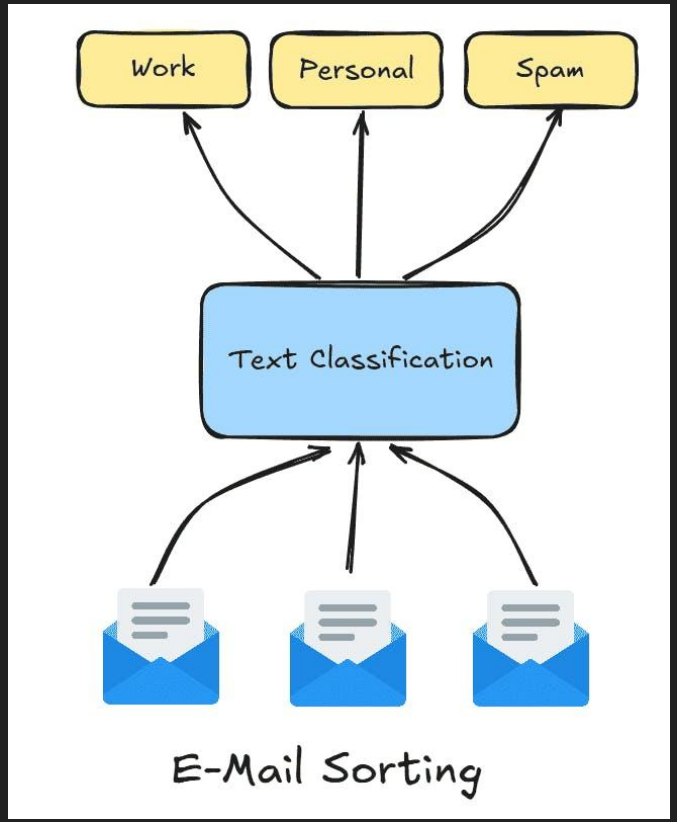
langchain/langgraph

Задачи: NER

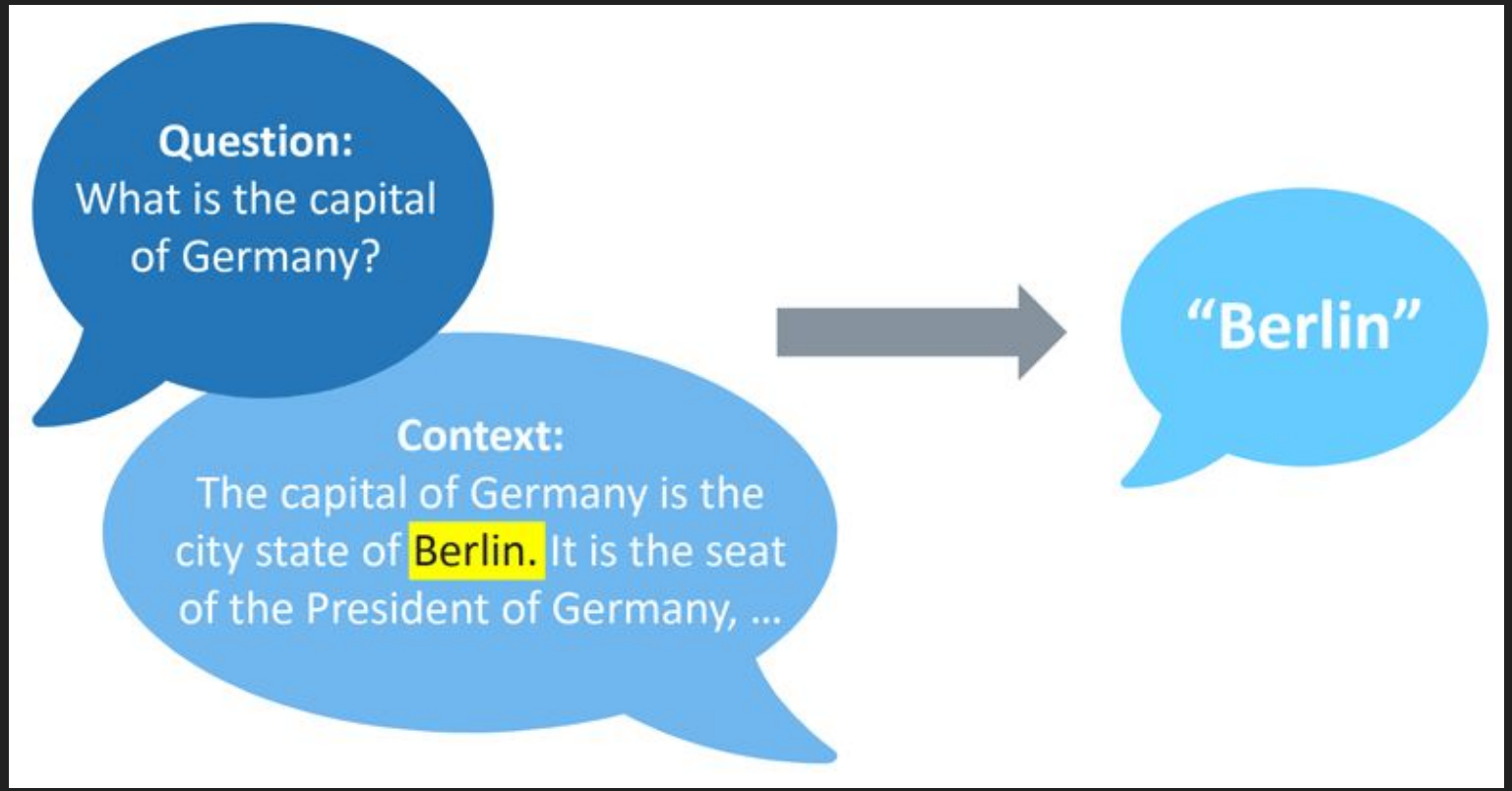
In fact, the Chinese NORP market has the three CARDINAL most influential names of the retail and tech space – Alibaba GPE , Baidu ORG , and Tencent PERSON (collectively touted as BAT ORG), and is betting big in the global AI GPE in retail industry space . The three CARDINAL giants which are claimed to have a cut-throat competition with the U.S. GPE (in terms of resources and capital) are positioning themselves to become the 'future AI PERSON platforms'. The trio is also expanding in other Asian NORP countries and investing heavily in the U.S. GPE based AI GPE startups to leverage the power of AI GPE . Backed by such powerful initiatives and presence of these conglomerates, the market in APAC AI is forecast to be the fastest-growing one CARDINAL , with an anticipated CAGR PERSON of 45% PERCENT over 2018 - 2024 DATE .

To further elaborate on the geographical trends, North America LOC has procured more than 50% PERCENT of the global share in 2017 DATE and has been leading the regional landscape of AI GPE in the retail market. The U.S. GPE has a significant credit in the regional trends with over 65% PERCENT of investments (including M&As, private equity, and venture capital) in artificial intelligence technology. Additionally, the region is a huge hub for startups in tandem with the presence of tech titans, such as Google ORG , IBM ORG , and Microsoft ORG .

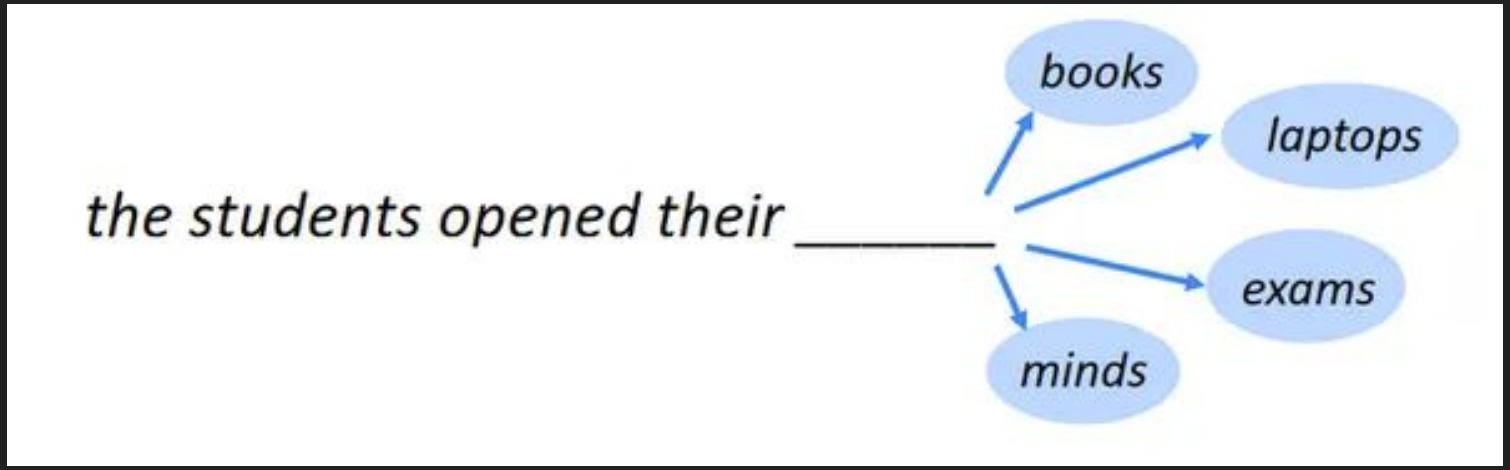
Задачи: классификация текста



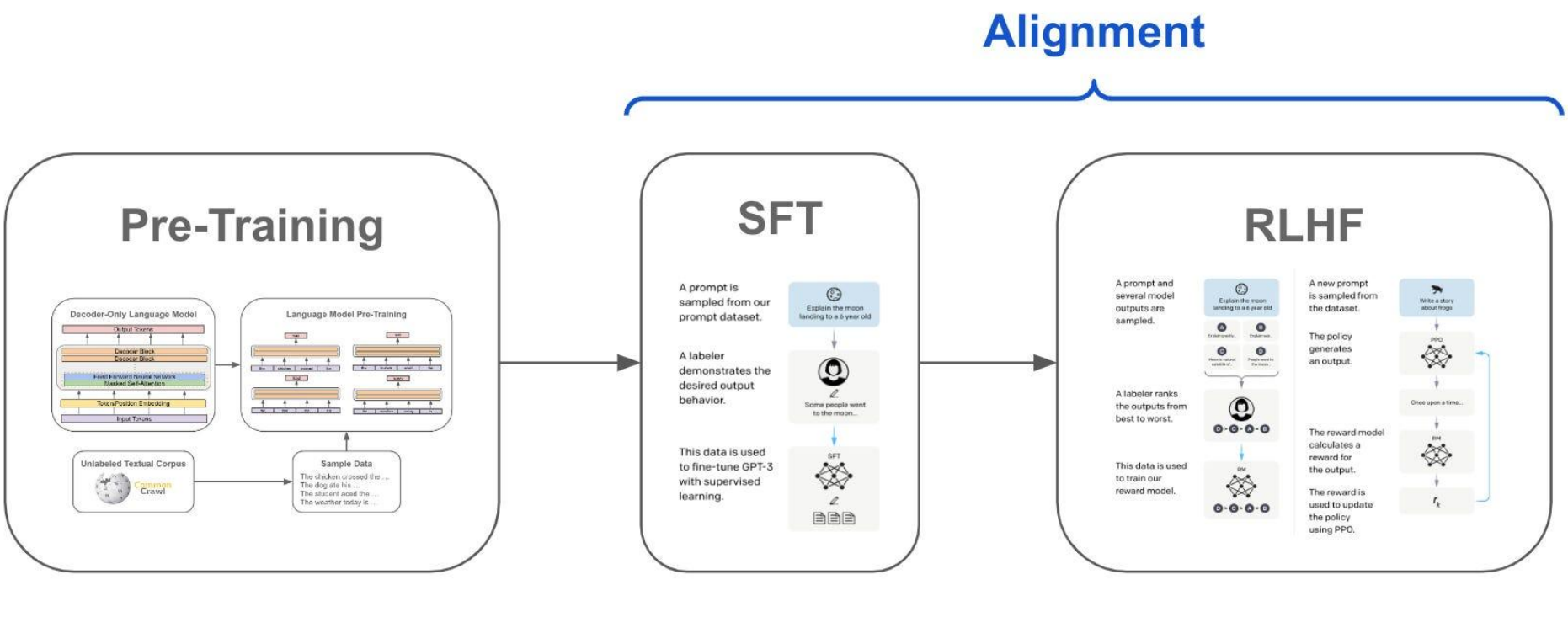
Задачи: QA


Question:
What is the capital of Germany?
Context:
The capital of Germany is the city state of Berlin. It is the seat of the President of Germany, ...
"Berlin"

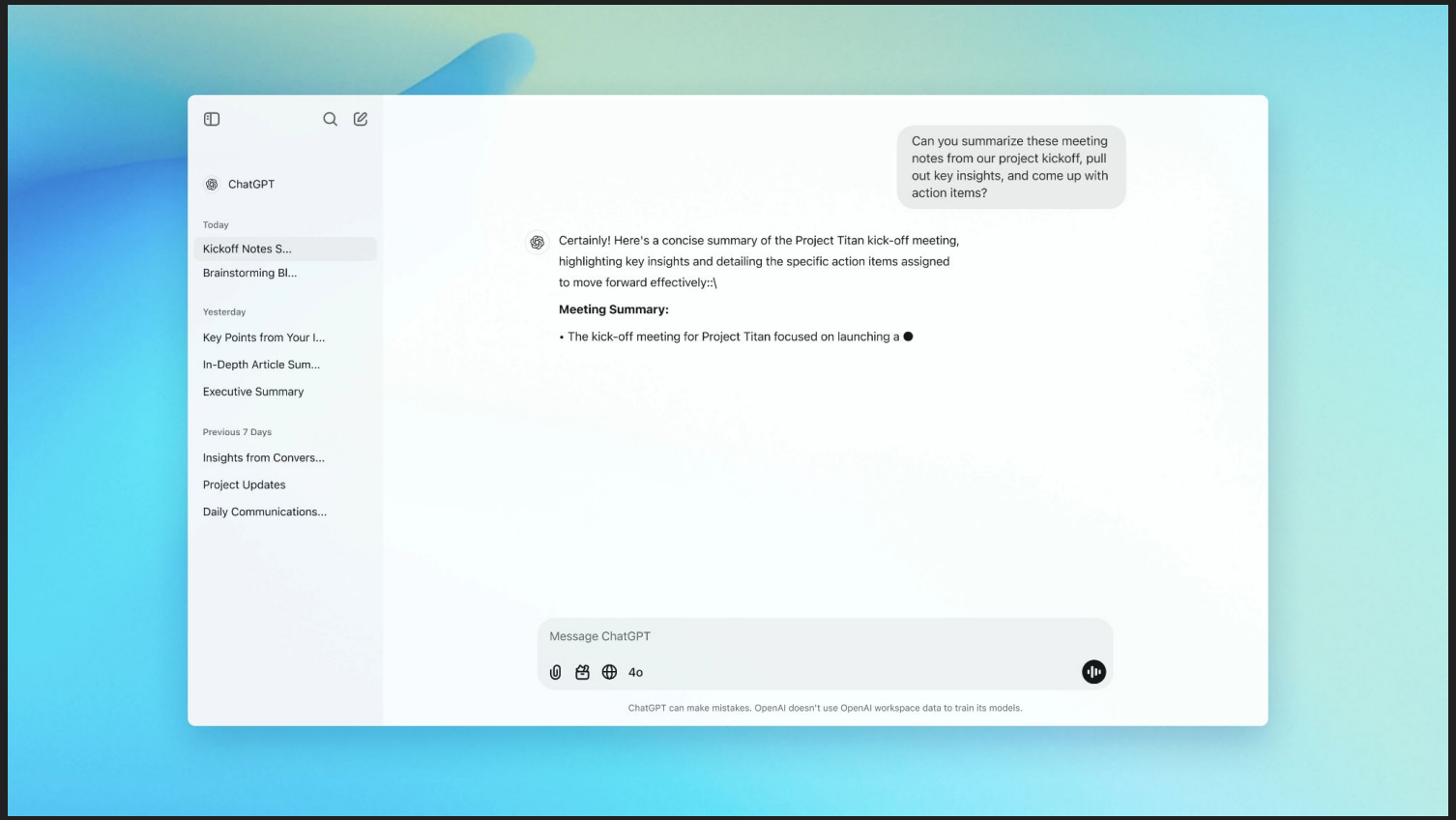
Задачи: авторегрессия



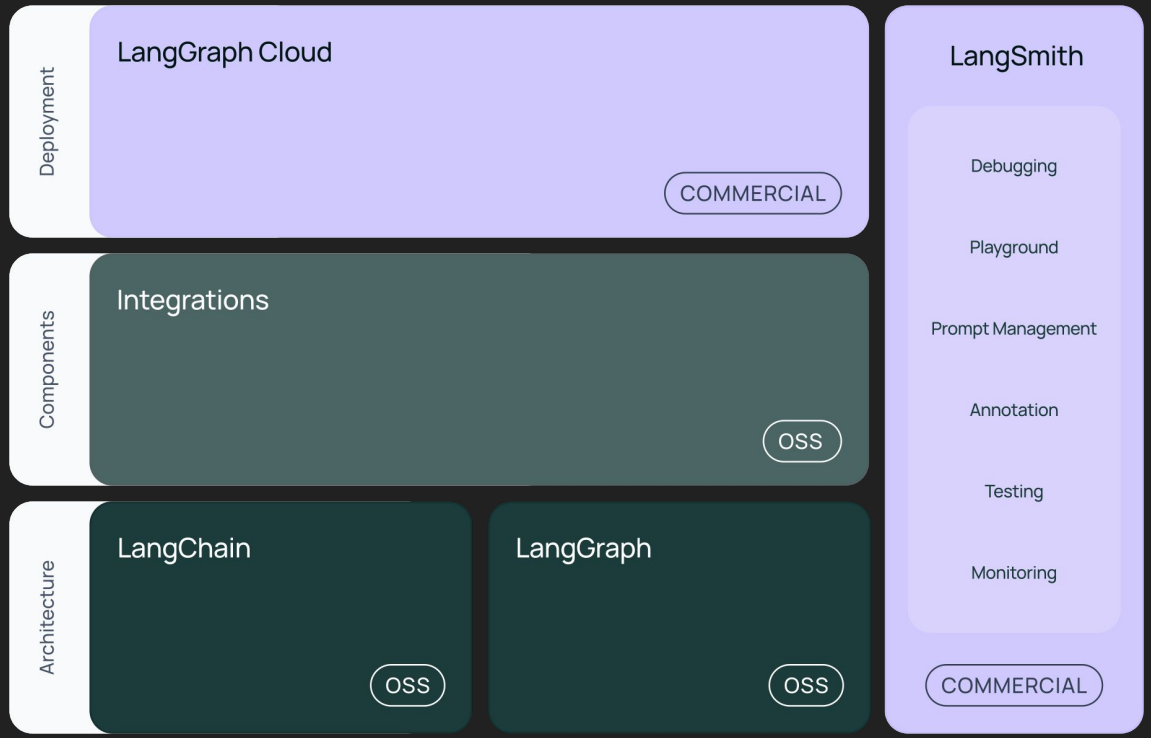
LLMs


Alignment
Pre-Training
Decoder-Only Language Model
Output Tokens
Decoder Block
Decoder Block
Input Tokens
Language Model Pre-Training
Unlabeled Textual Corpus
Common Crawl
Sample Data
SFT
A prompt is sampled from our prompt dataset.
A labeler demonstrates the desired output behavior.
This data is used to fine-tune GPT-3 with supervised learning.
RLHF
A prompt and several model outputs are sampled.
A labeler ranks the outputs from best to worst.
This data is used to train our reward model.
A new prompt is sampled from the dataset.
The policy generates an output.
The reward model calculates a reward for the output.
The reward is used to update the policy using PPO.

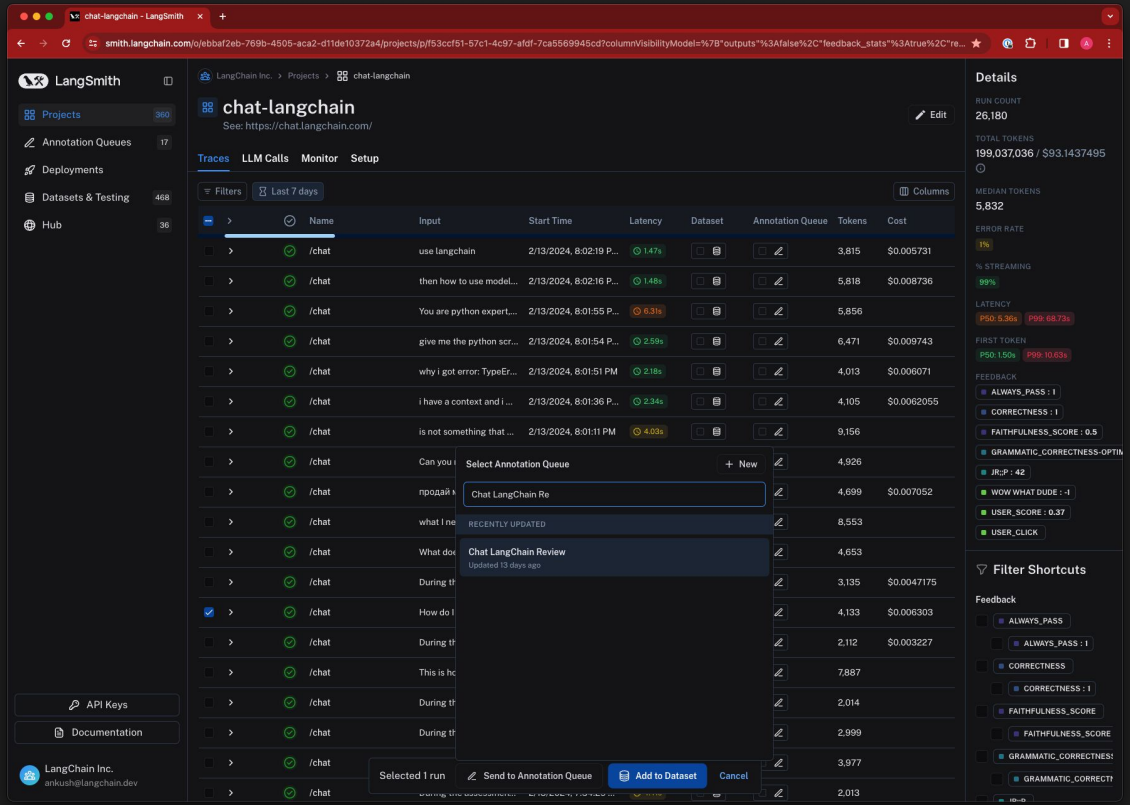
Chat models


ChatGPT
Today
Kickoff Notes S...
Brainstorming Bl...
Yesterday
Key Points from Your I...
In-Depth Article Sum...
Executive Summary
Previous 7 Days
Insights from Convers...
Project Updates
Daily Communications...
Can you summarize these meeting notes from our project kickoff, pull out key insights, and come up with action items?
Certainly! Here's a concise summary of the Project Titan kick-off meeting, highlighting key insights and detailing the specific action items assigned to move forward effectively::\
Meeting Summary:
• The kick-off meeting for Project Titan focused on launching a
Message ChatGPT
4o
ChatGPT can make mistakes. OpenAI doesn't use OpenAI workspace data to train its models.

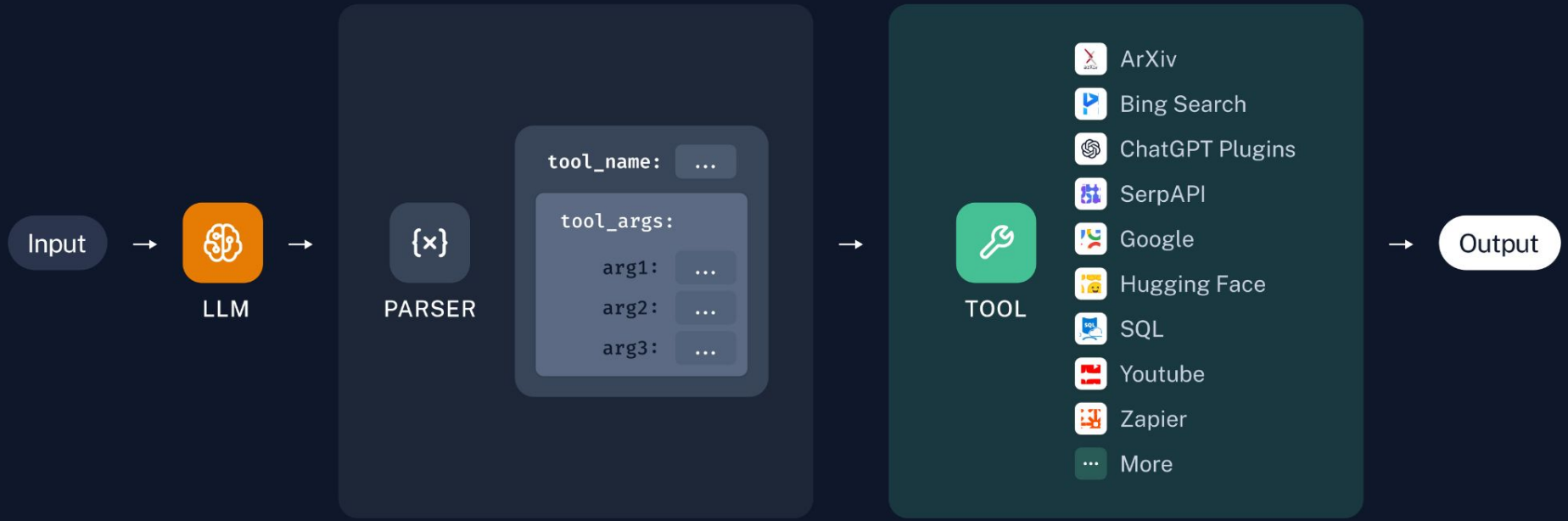
Langchain


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

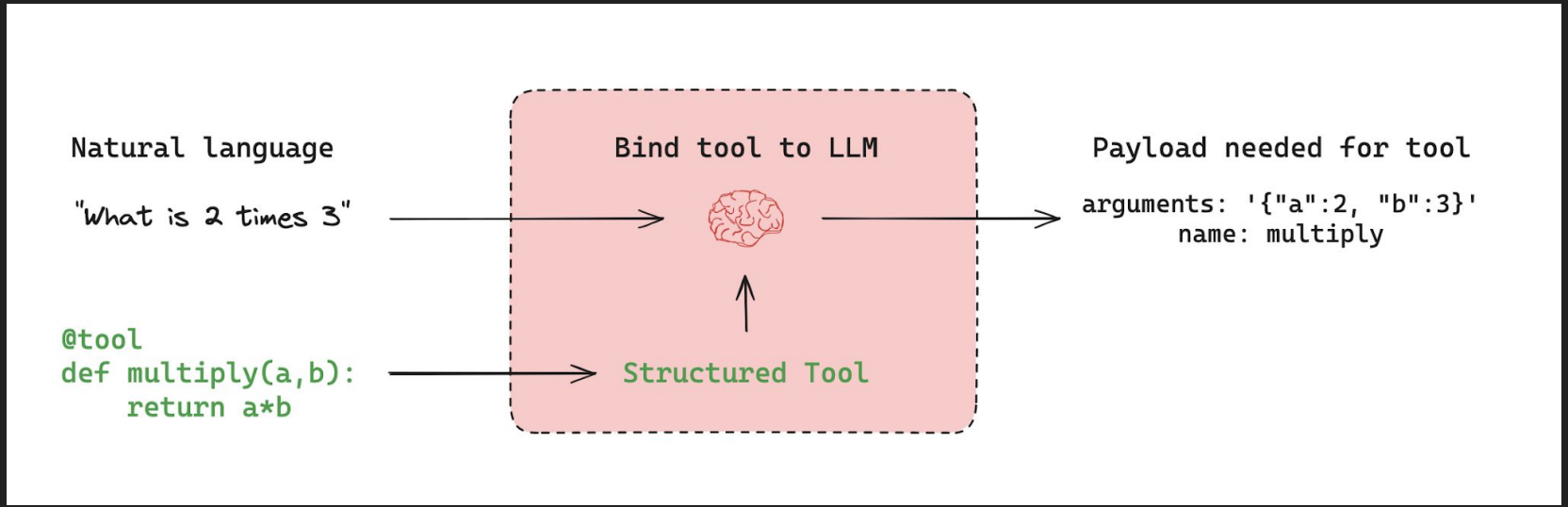
Langsmith


chat-langchain - LangSmith
smith.langchain.com/o/ebbaf2eb-769b-4505-aca2-d11de10372a4/projects/p/f53ccf51-57c1-4c97-afdf-7ca5569945cd?columnVisibilityModel=%7B"outputs"%3Afalse%2C"feedback_stats"%3Atrue%2C"re...
LangSmith
Projects 360
Annotation Queues 17
Deployments
Datasets & Testing 468
Hub 36
API Keys
Documentation
LangChain Inc.
ankush@langchain.dev
LangChain Inc. › Projects › chat-langchain
chat-langchain
See: https://chat.langchain.com/
Edit
Traces
LLM Calls
Monitor
Setup
Filters
Last 7 days
Columns
Name
Input
Start Time
Latency
Dataset
Annotation Queue
Tokens
Cost
/chat
use langchain
2/13/2024, 8:02:19 P...
1.47s
3,815
$0.005731
/chat
then how to use model...
2/13/2024, 8:02:16 P...
1.48s
5,818
$0.008736
/chat
You are python expert....
2/13/2024, 8:01:55 P...
6.31s
5,856
/chat
give me the python scr...
2/13/2024, 8:01:54 P...
2.59s
6,471
$0.009743
/chat
why i got error: TypeEr...
2/13/2024, 8:01:51 PM
2.18s
4,013
$0.006071
/chat
i have a context and i ...
2/13/2024, 8:01:36 P...
2.34s
4,105
$0.0062055
/chat
is not something that ...
2/13/2024, 8:01:11 PM
4.03s
9,156
/chat
Can you
4,926
/chat
продай
4,699
$0.007052
/chat
what I ne
8,553
/chat
What doe
4,653
/chat
During th
3,135
$0.0047175
/chat
How do I
4,133
$0.006303
/chat
During th
2,112
$0.003227
/chat
This is hc
7,887
/chat
During th
2,014
/chat
During th
2,999
/chat
3,977
/chat
2,013
Select Annotation Queue
+ New
Chat LangChain Re
RECENTLY UPDATED
Chat LangChain Review
Updated 13 days ago
Selected 1 run
Send to Annotation Queue
Add to Dataset
Cancel
Details
RUN COUNT
26,180
TOTAL TOKENS
199,037,036 / $93.1437495
MEDIAN TOKENS
5,832
ERROR RATE
1%
% STREAMING
99%
LATENCY
P50: 5.36s
P99: 68.73s
FIRST TOKEN
P50: 1.50s
P99: 10.63s
FEEDBACK
ALWAYS_PASS : 1
CORRECTNESS : 1
FAITHFULNESS_SCORE : 0.5
GRAMMATIC_CORRECTNESS-OPTIN
JR;;P : 42
WOW WHAT DUDE : -1
USER_SCORE : 0.37
USER_CLICK
Filter Shortcuts
Feedback
ALWAYS_PASS
ALWAYS_PASS : 1
CORRECTNESS
CORRECTNESS : 1
FAITHFULNESS_SCORE
FAITHFULNESS_SCORE
GRAMMATIC_CORRECTNESS
GRAMMATIC_CORRECTI

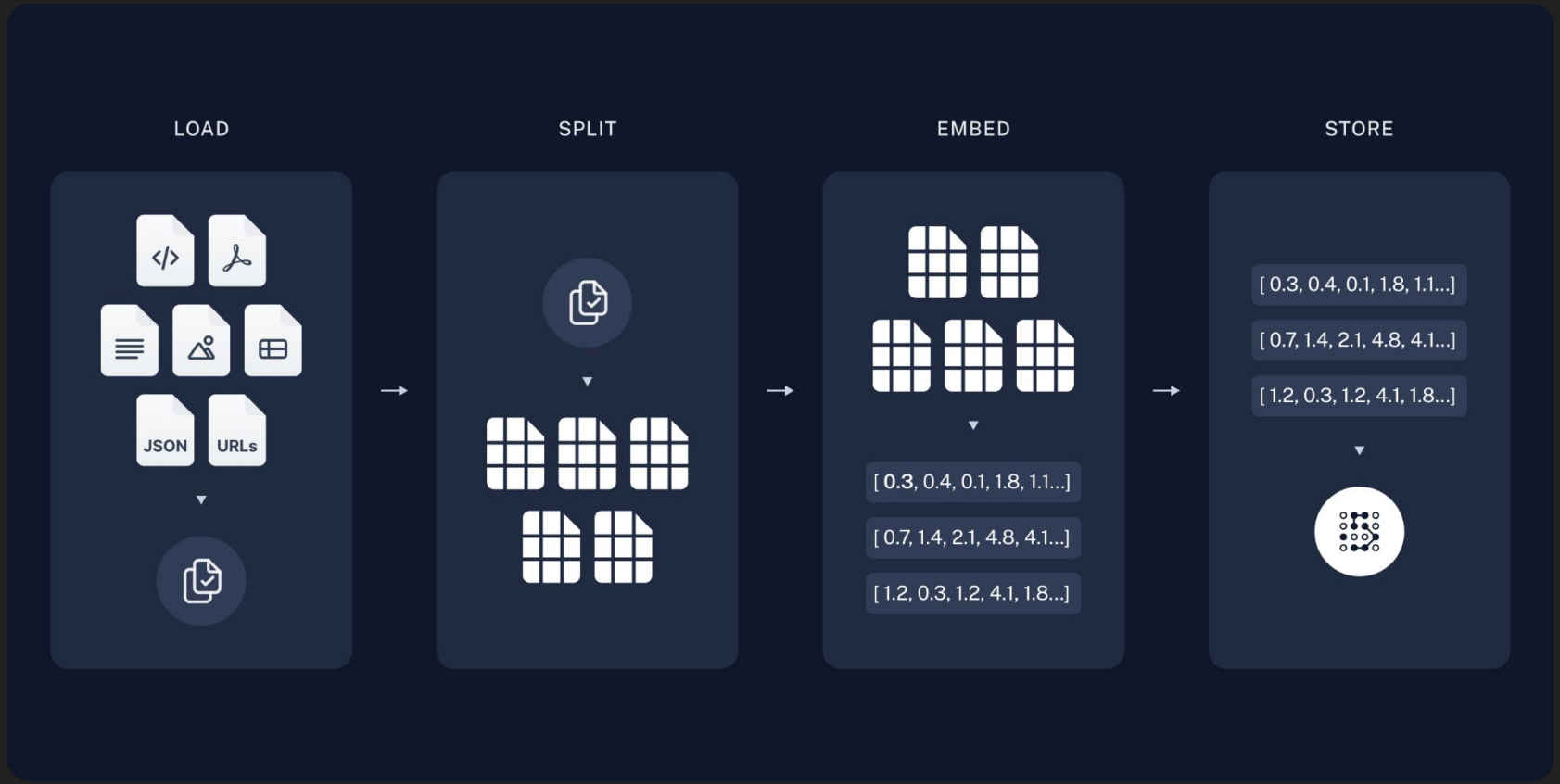
Tools


Input
LLM
{x}
PARSER
tool_name: ...
tool_args:
arg1: ...
arg2: ...
arg3: ...
TOOL
ArXiv
Bing Search
ChatGPT Plugins
SerpAPI
Google
Hugging Face
SQL
Youtube
Zapier
More
Output

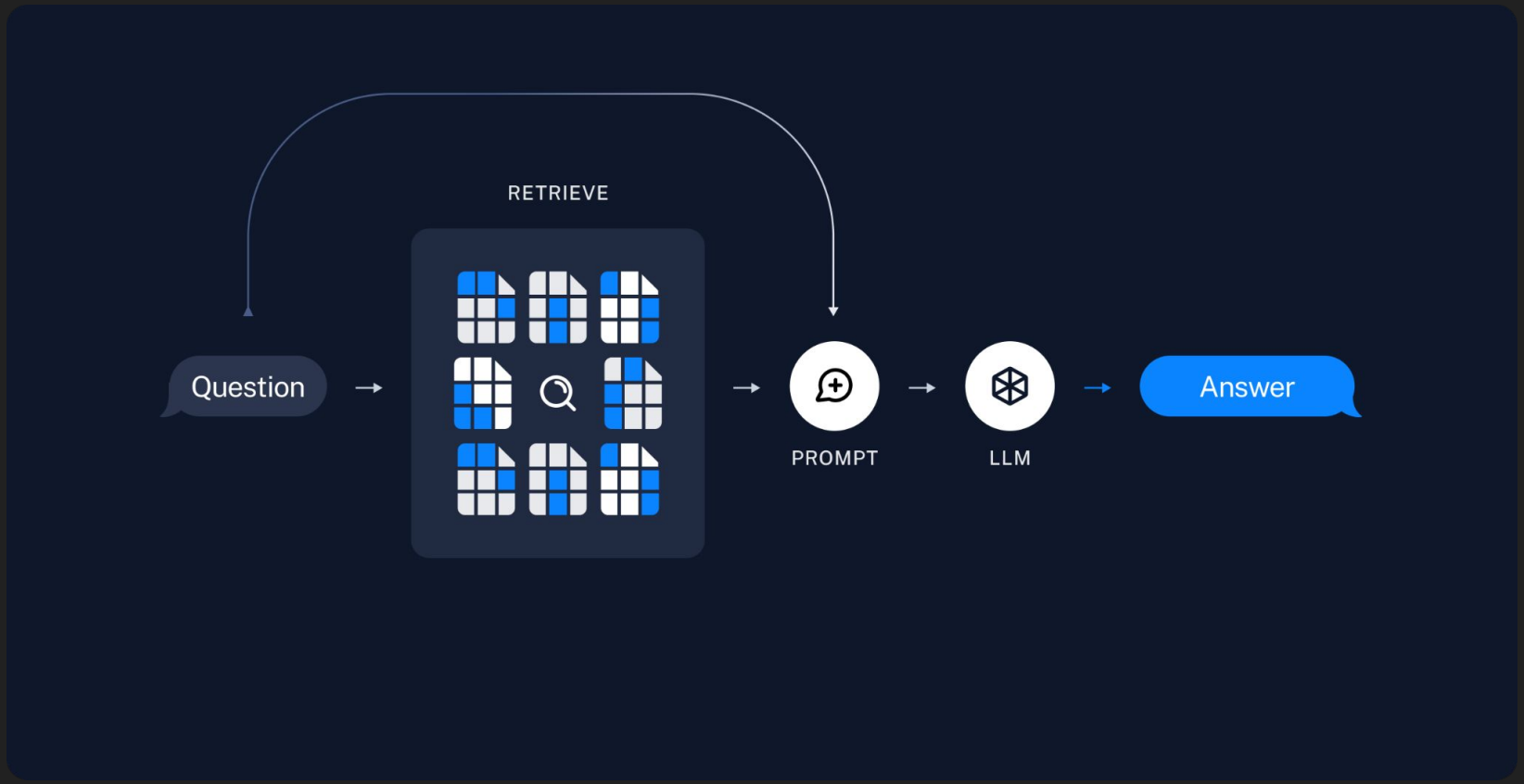
Tools


Natural language
"What is 2 times 3"
@tool
def multiply(a,b):
return a*b
Bind tool to LLM
Structured Tool
Payload needed for tool
arguments: '{"a":2, "b":3}'
name: multiply

RAG: indexing

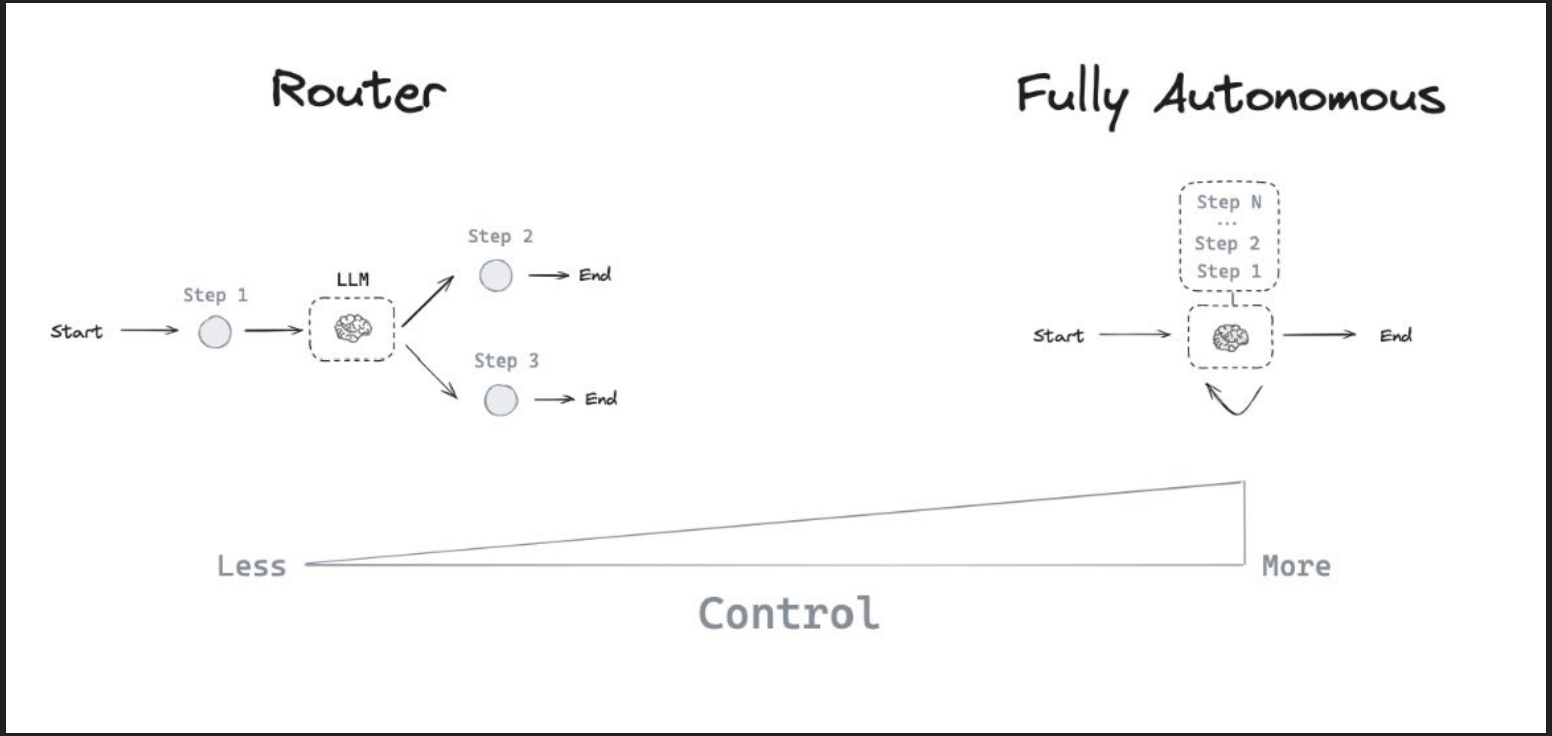


RAG: pipeline

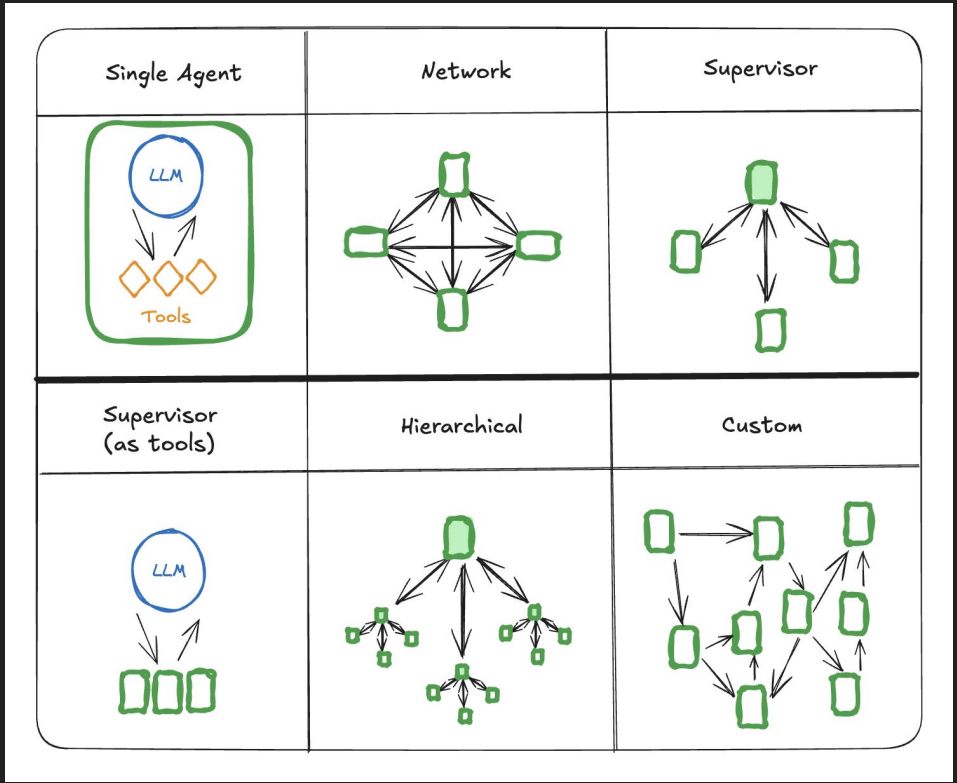


Langgraph

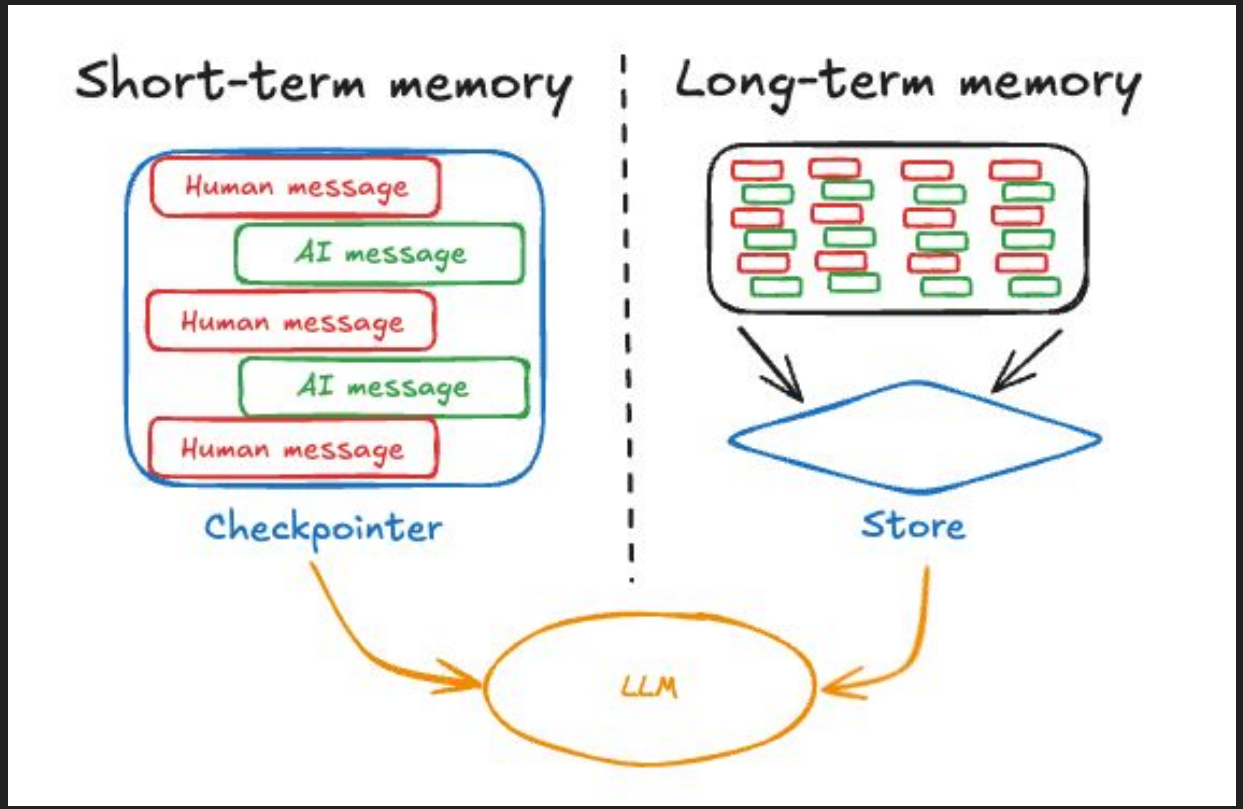
Agents


Router
Start
Step 1
LLM
Step 2
End
Step 3
End
Fully Autonomous
Step N
...
Step 2
Step 1
Start
End
Less
More
Control

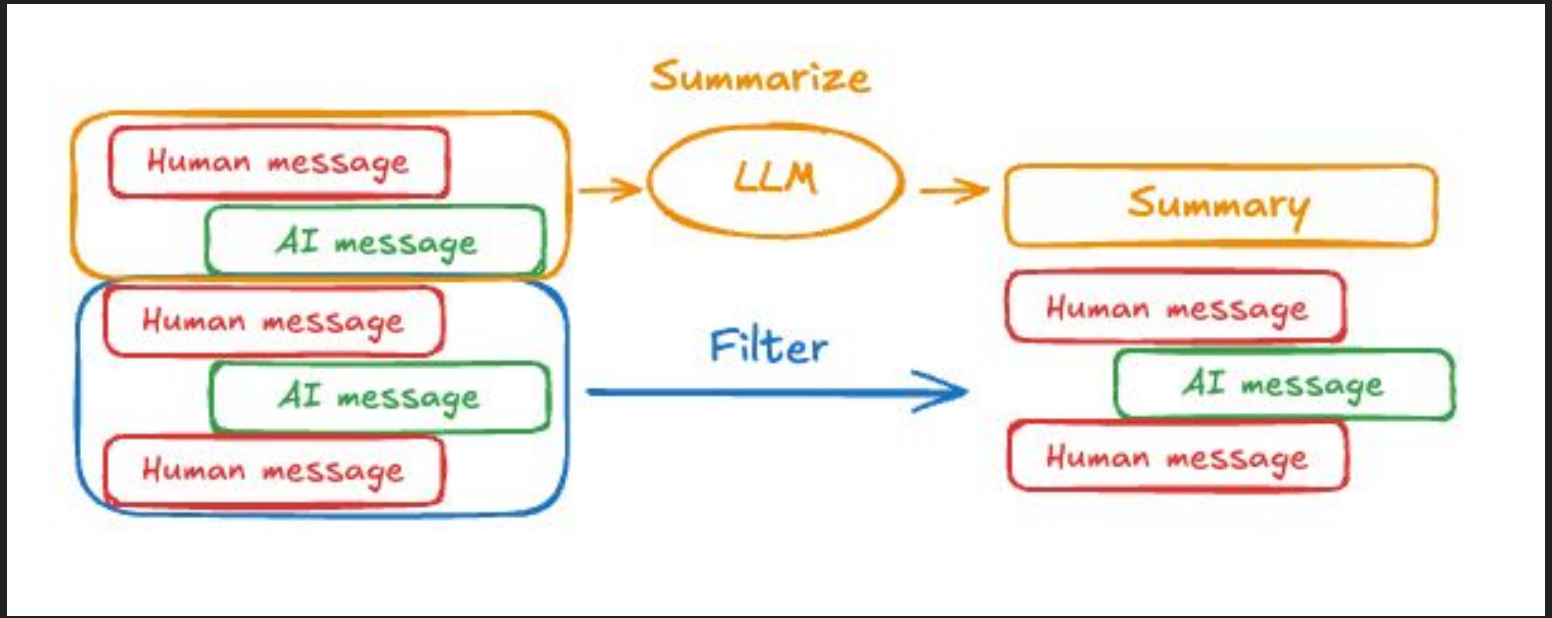
Multi-agents



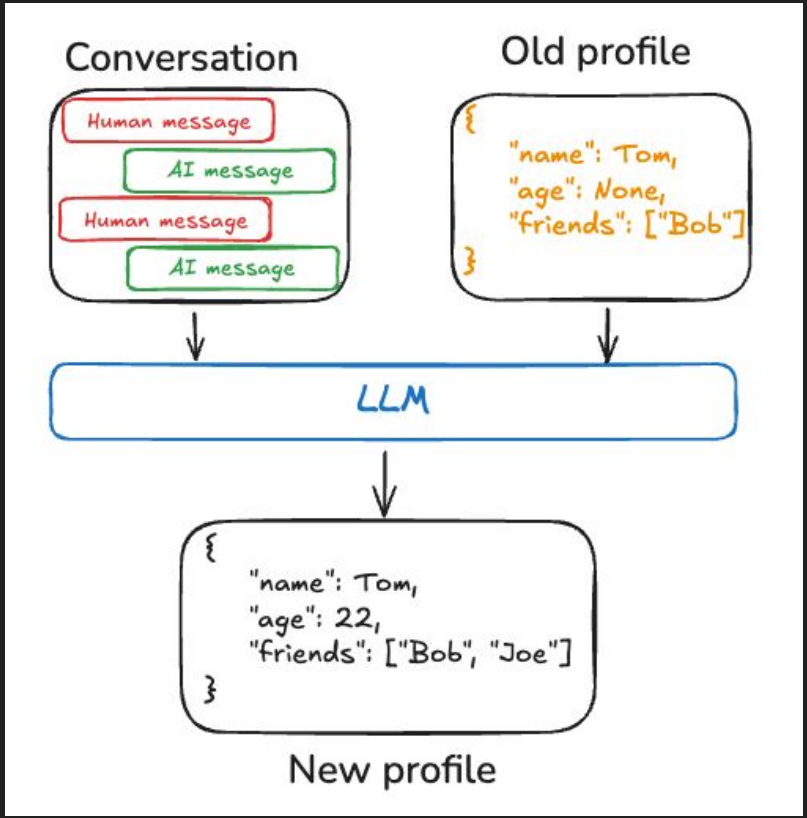
Memory


Short-term memory
Long-term memory
Human message
AI message
Human message
AI message
Human message
Checkpointer
Store
LLM

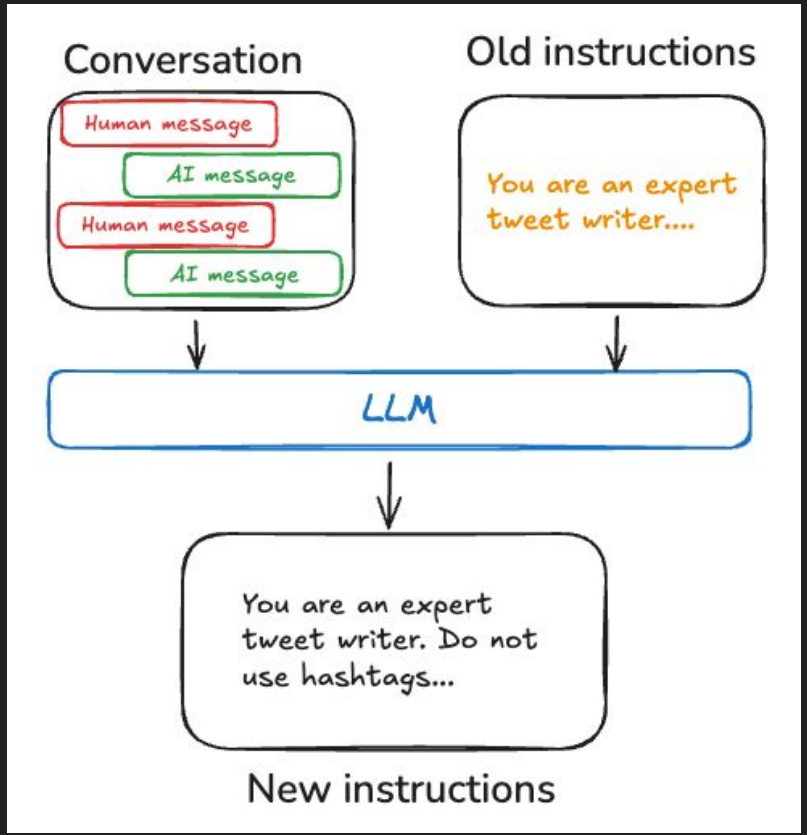
Short-term memory



Long-term memory



Meta-prompting



Meta-prompting