

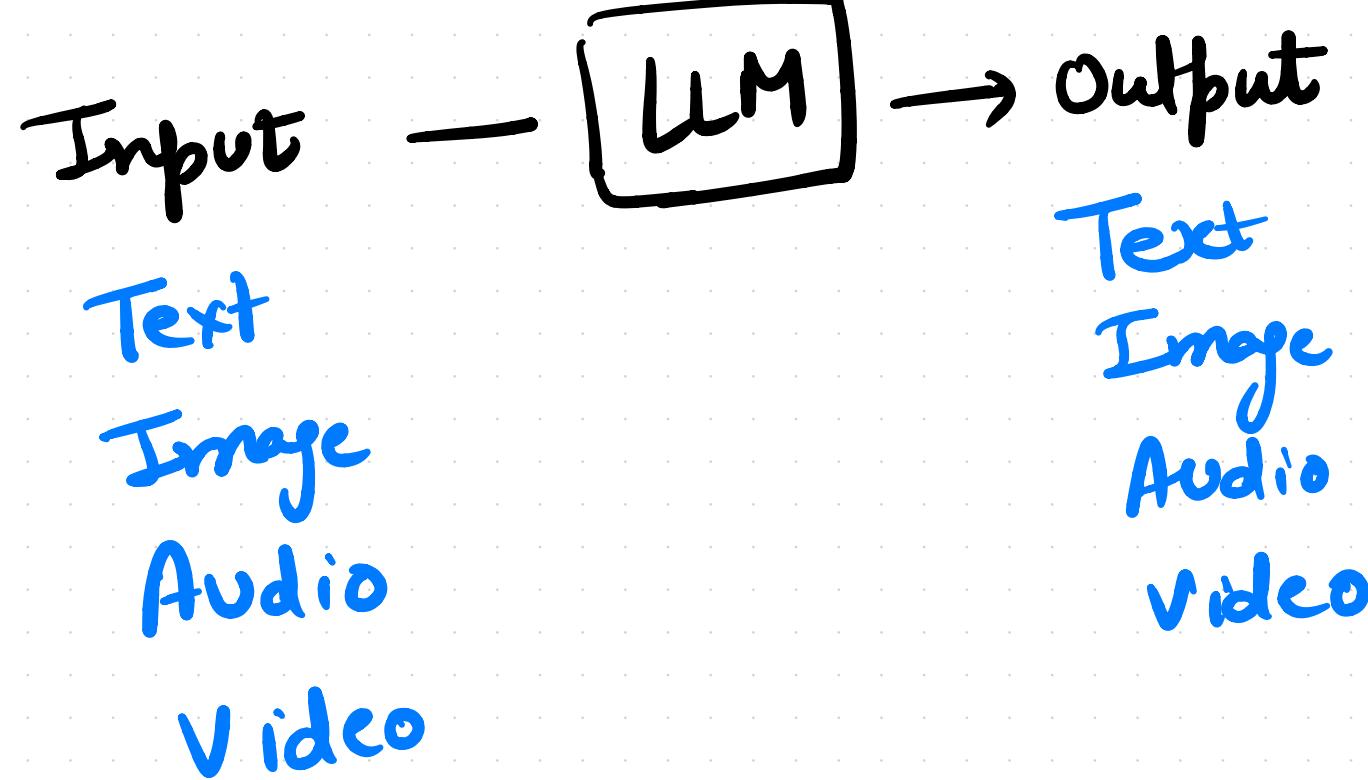
Multimodals

Modalities

- Text
- Image
- Audio
- Video

A hand-drawn diagram on lined paper. On the left, the word "Input" is written above "(Text)". An arrow points from this text to a large rectangular box. The letters "LLM" are written inside the box. Another arrow points from the right side of the box to the word "Output" above "(Text)". Below the main diagram, the words "traditional LLMs" are written in blue ink.

Advanced LLMs

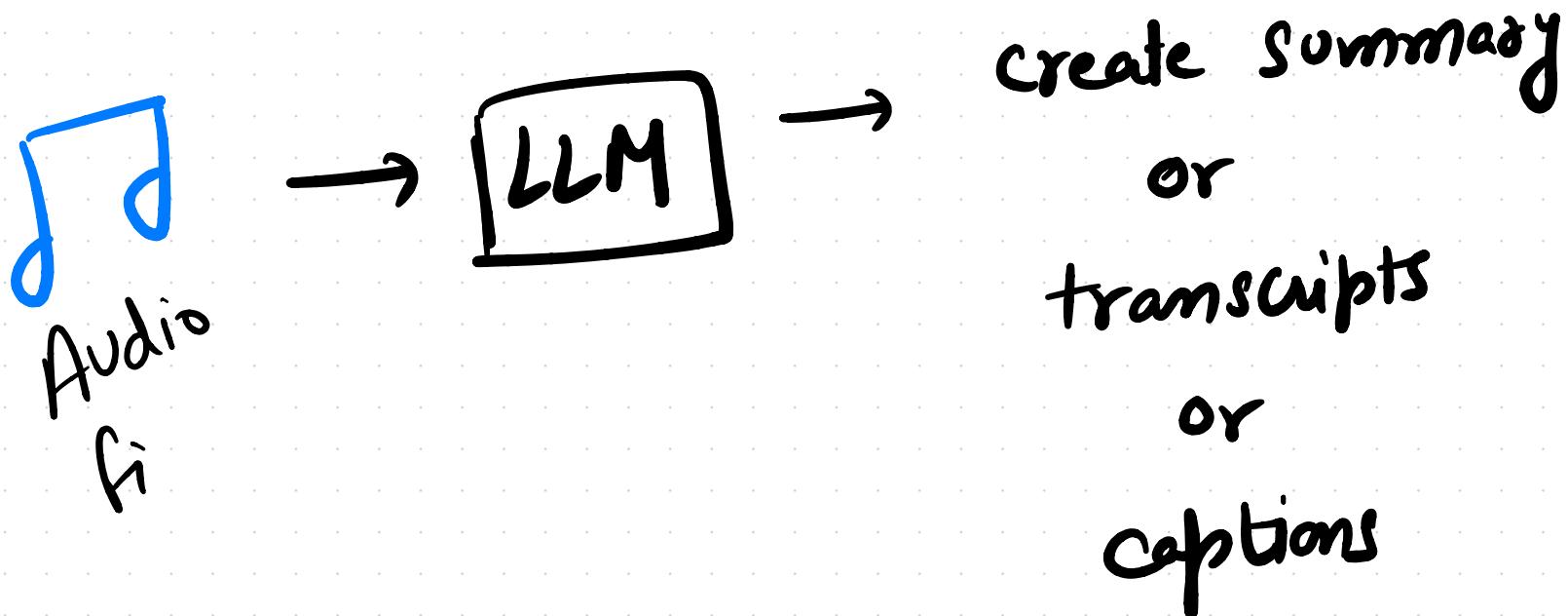


Why ?

Because most of the data in
the World is not just

Text

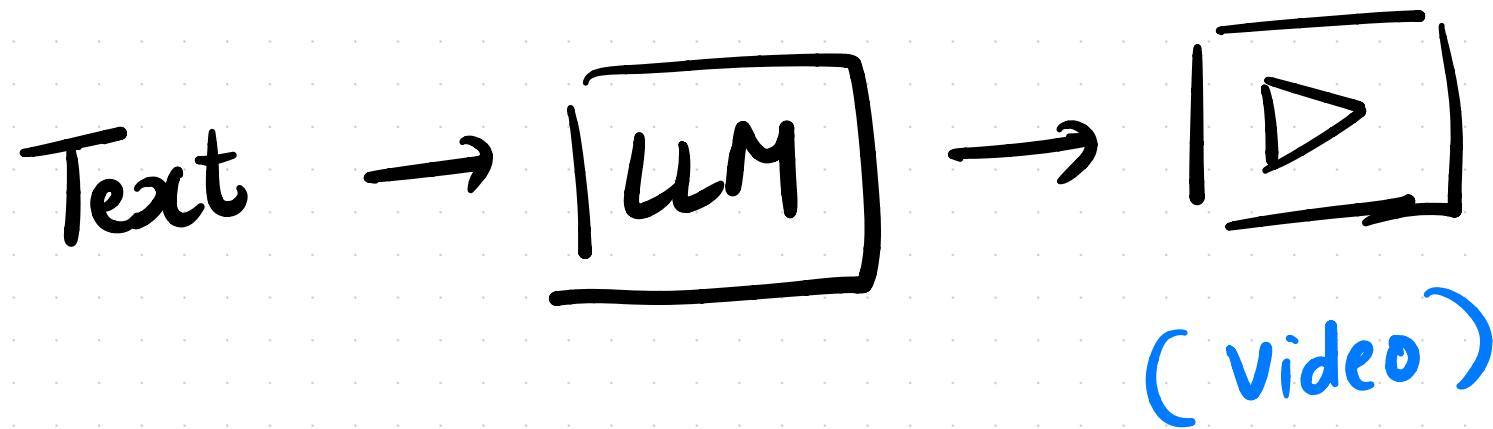
Use Cases



Resolve an Error



Generate a video



We can do a lot more
if LLMs can understand
all kinds of data

Examples

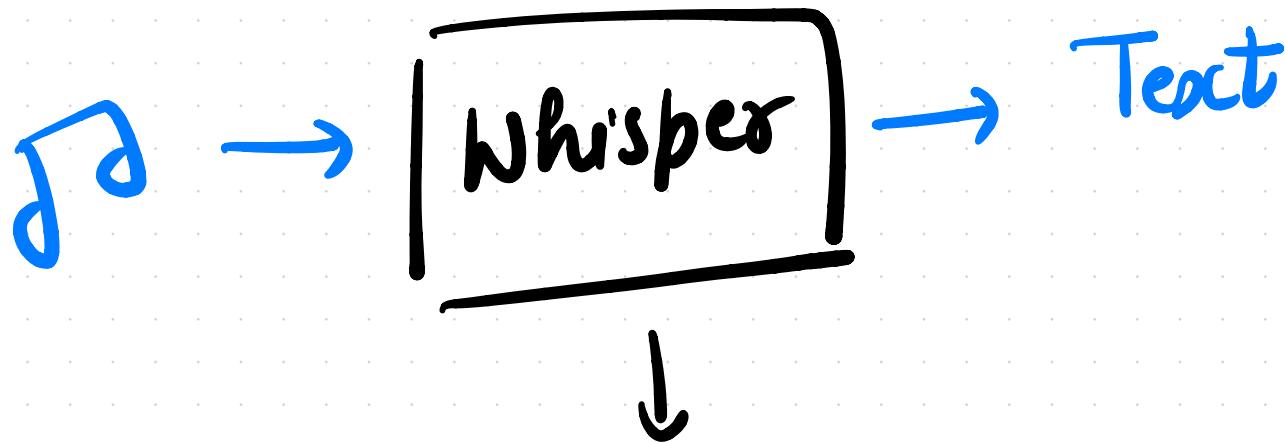
Chat GPT

\geq I/O

Claude

\geq (All modalities)

Gemini

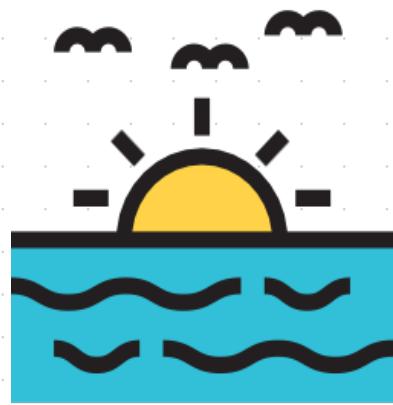


Also used by
Geminiani

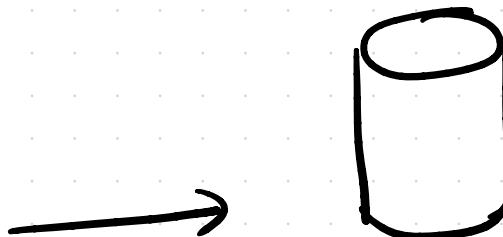
Image → CLIP → Text

(analysing images)

VECTOR DB



(Sunset)



Relational DB

Binary
data

Metadata
(Size, name)

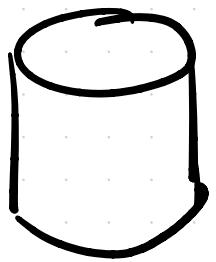
tags
(Sunset
beach)



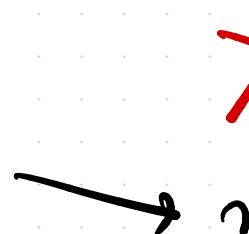
Lacks Semantic Context
warm colors | pink skies

query → 

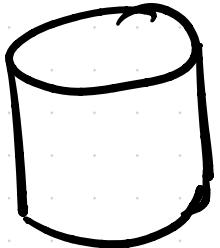
Get images of same color
palette as the sunset
picture



Relational
DB



not good for
Storing unstructured
data

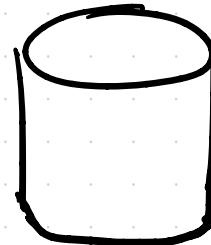
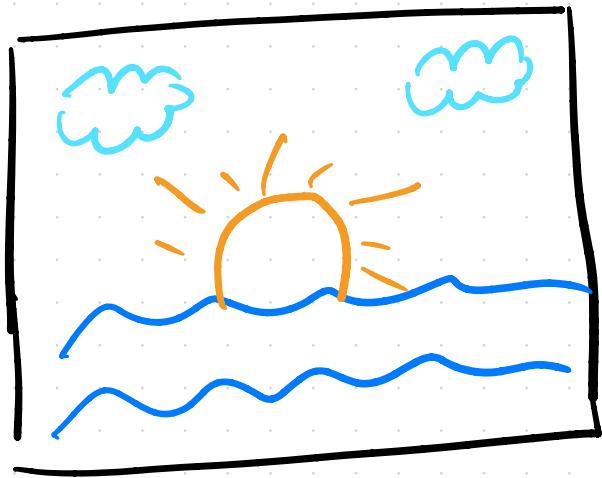


Relational
DB

SEMANTIC GAP



Disconnect between how
Computers store data
Vs how humans
understand



(Vector DB)

Vector dB is used to store
unstructured data like (Images,
Audio, Video)

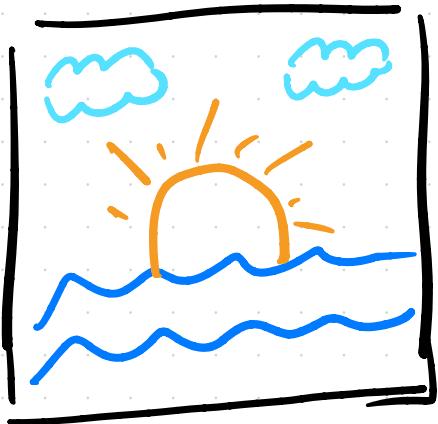
Vector Embeddings

Numerical representation of
the data (Image| Audio| Video)

$$[0.21, 0.15, 0.83 \downarrow \cdots]$$

↓ ↓ ↓

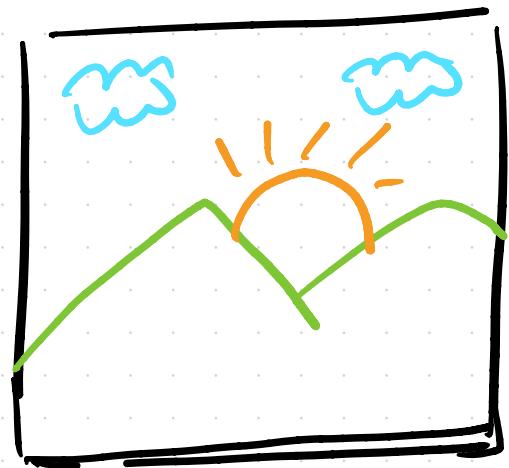
elevation urban worm
 colors



[0.21, 0.15, 0.83 - - -]

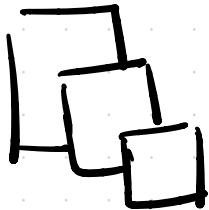
↓ ↓ ↓

elevation urban warm colors



[0.95, 0.20, 0.42 - - -]

Embedding Models



Data

Vectors

(list of numbers)

Text →

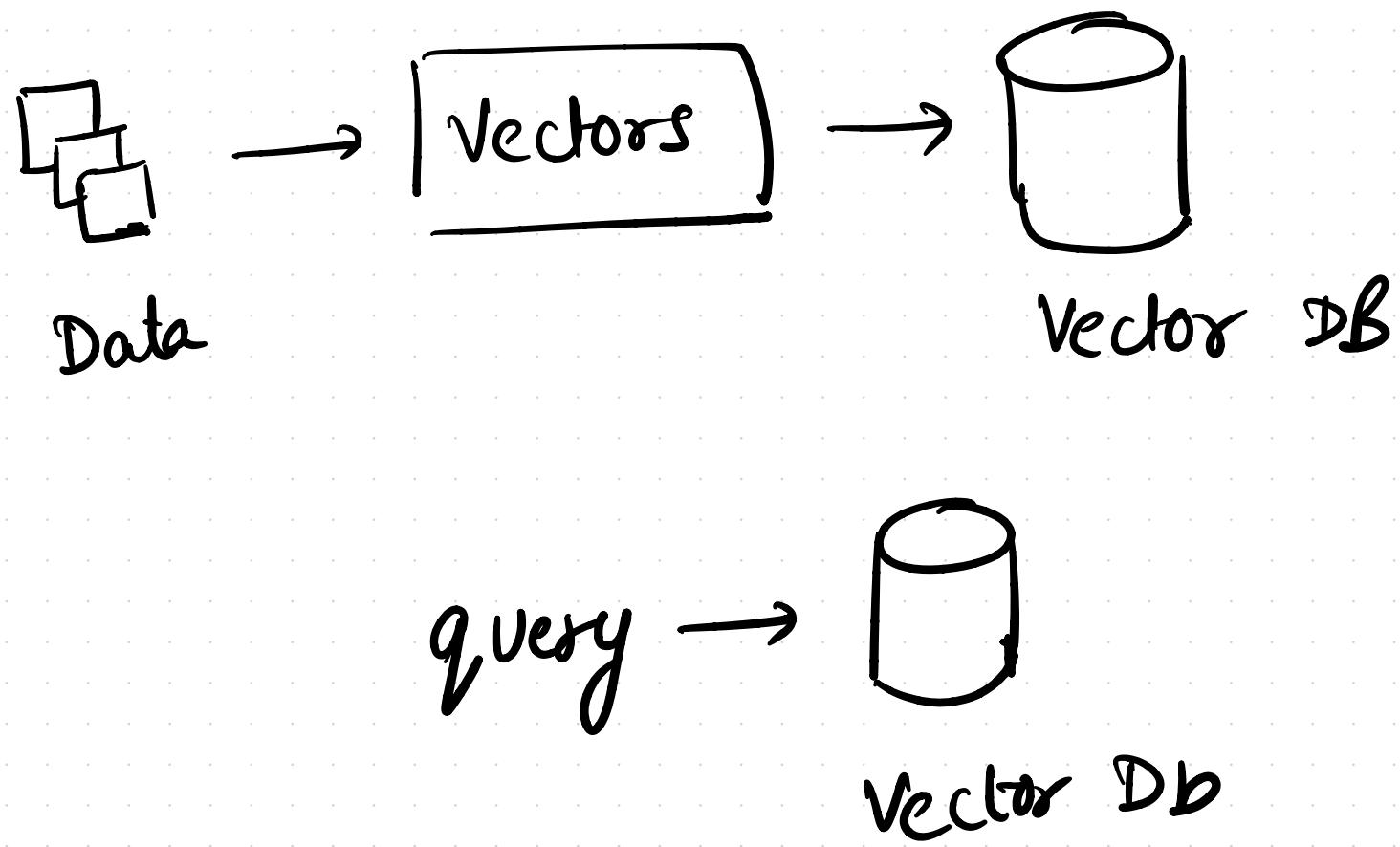
Word/sentence embeddings

Image →

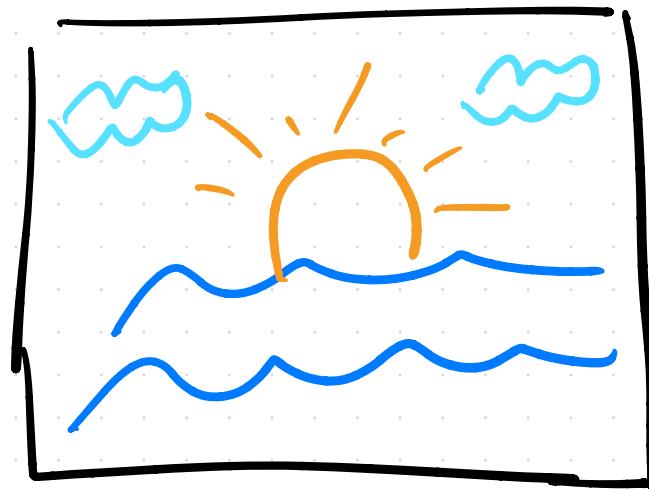
Image embeddings

Audio →

speech embeddings



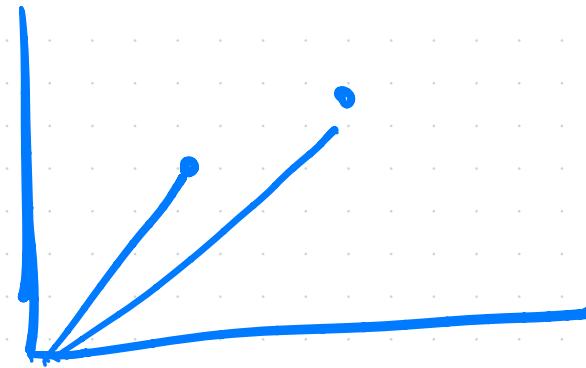
find exact Match



Vector Db

✗ too slow

Vector Indexing



ANN algorithm is used to
find the closest match

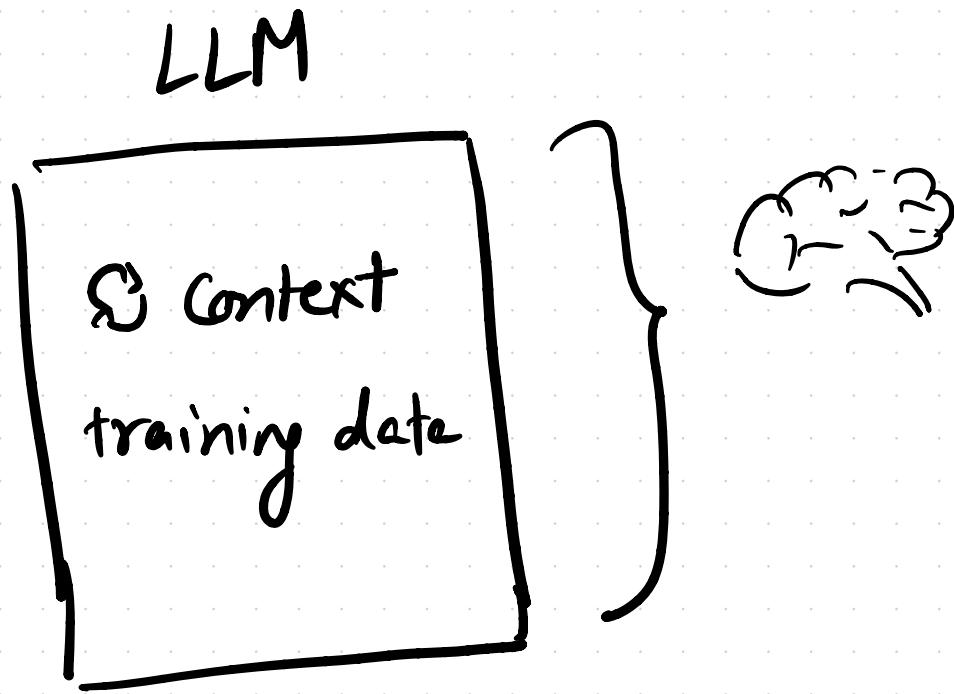
Use Cases

- ✓ Finding painting of sunset
on a beach
- ✓ finding a similar dress
- ✓ Recommendations for similar
hotels.

Popular Vector DBs

- Pinecone
- Weavate
- Chroma
- Redis Vector

RAG



LLM's
Knowledge is
limited to its
training data

What is the latest
version of iphone in
the market ?



Iphone 15 Pro
(Hallucination)

When is my next
O'Reilly training?



LLM



??

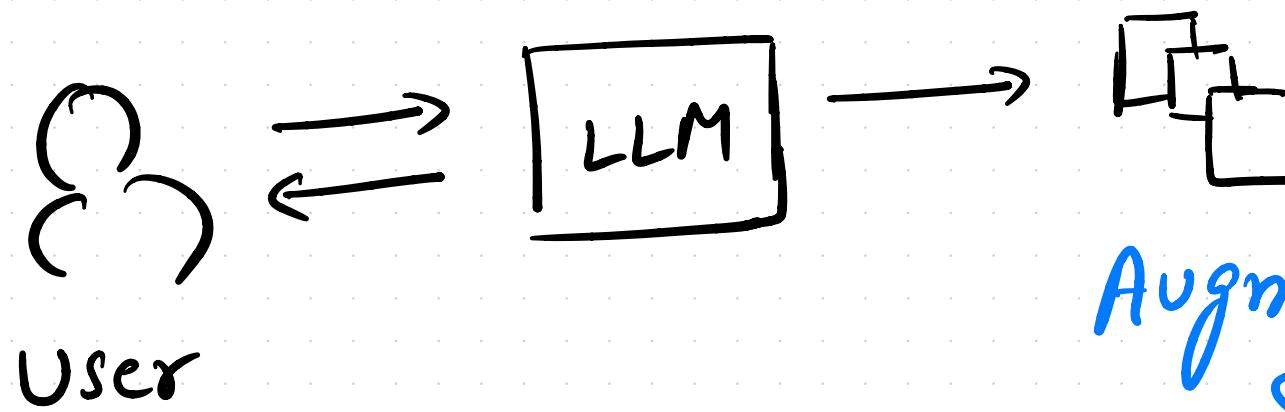
Challenges

① Out of date

② No source

How to extend

LLM's knowledge?



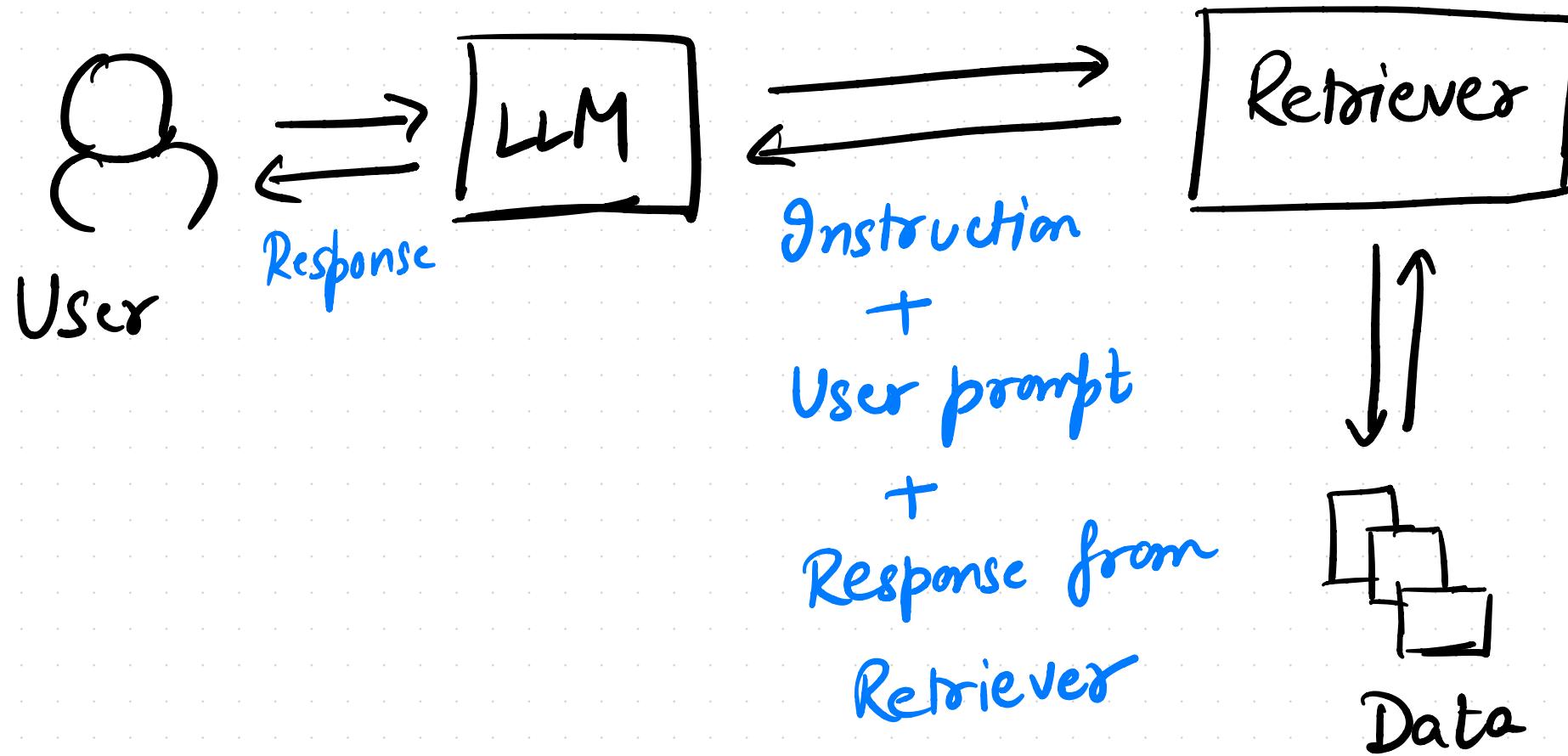
Augmented data
source



✓ documents

✓ Vector DB

✓ Calendar events



By extending the source
LLMs are less likely

to Hallucinate

Pros

- ✓ Latest information from external sources
- ✓ Extended knowledge without extending the size of LLM

Limitation

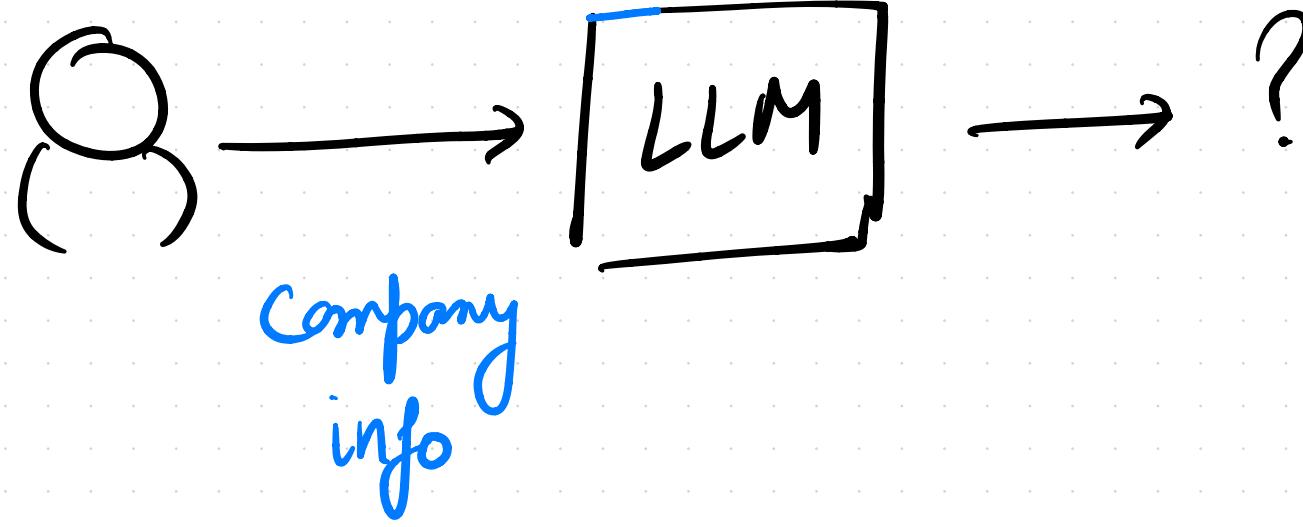
- Two step pipeline

Retriever + LLM → Slower

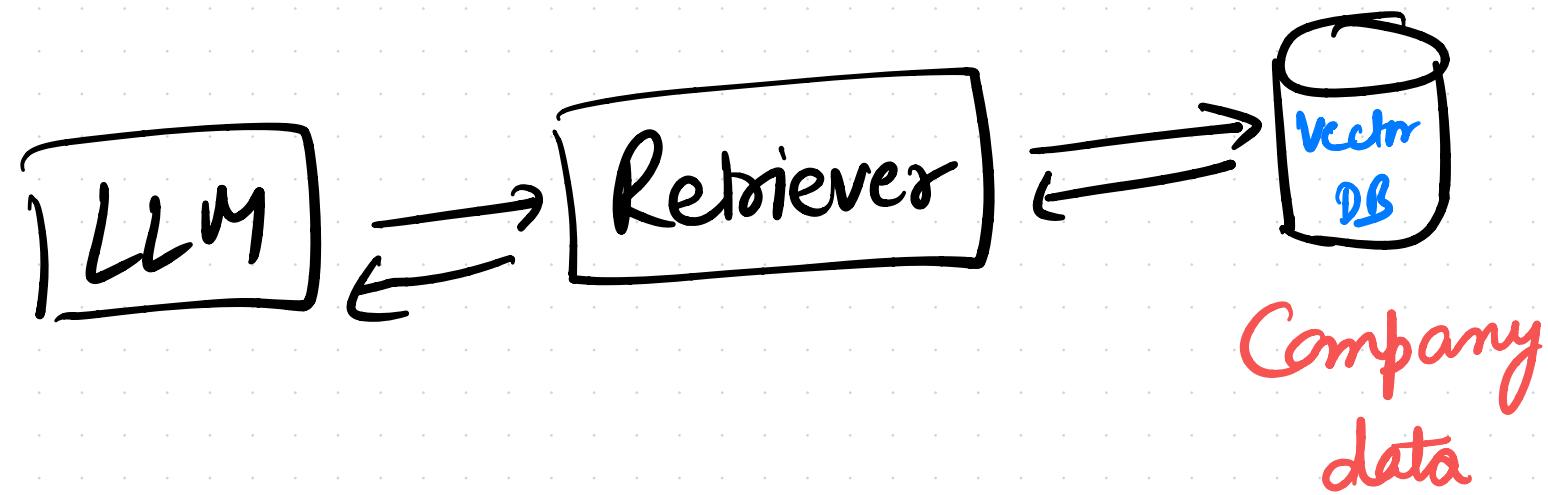
- Errors in Retrieval = bad answers

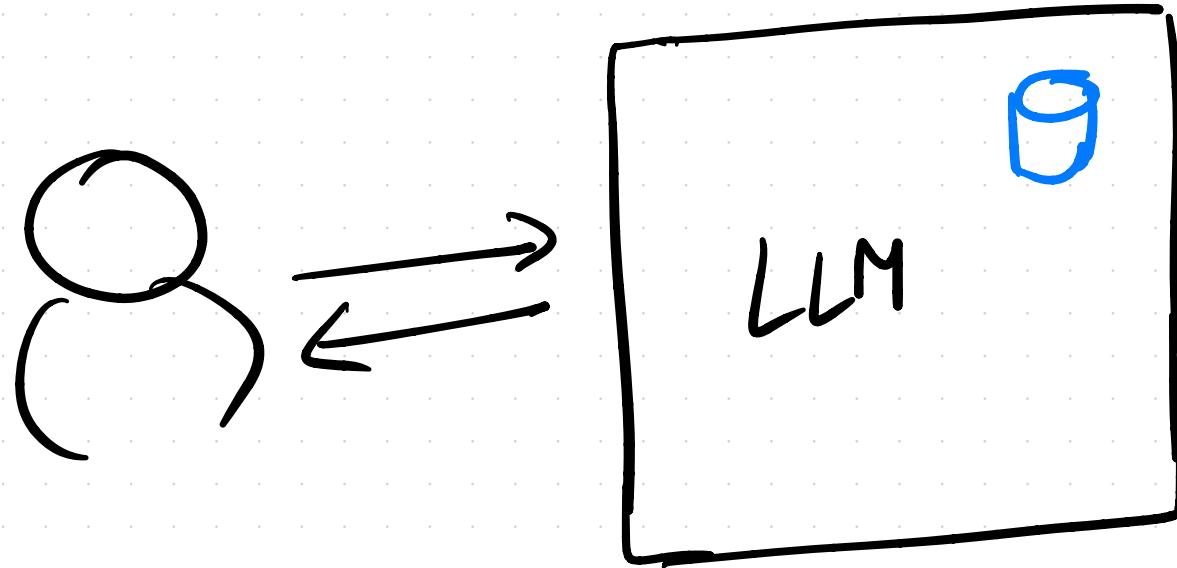
CAG

(Context Augmented
generation)



Using RAG





Provide Company
info directly
to LLM
Context

Instead of calculating data at
runtime, LLM is trained / finetuned
on a specific domain

Pros

✓ faster

✓ Easy to implement on
small knowledge sets

Limitation

- Context size is limited
- Needs frequent updating
if data changes
- Difficult to scale on large
data

Agents

LLMs are limited

- limited to the knowledge 
- They are passive (Wait for user action)
- Can't Act

Create Event

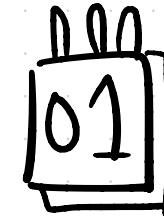


AI Workflows

fetch all my upcoming trainings
from Gmail



Add the events to my



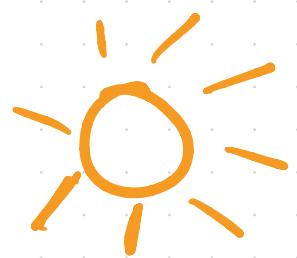
Generate table with columns

name | date/time | training link



Add to my Notes

Workflows extend capabilities of
LLMs but they follow predefined
paths set by humans



→ Weather?

AI Workflow X

Agents

① Reasoning

- Break down the problem
- Make a plan

② Action

- Use external tools like Calendars,
Calculators, PB

③ Iterate

- Iterate the steps if problem was
not solved in one way

Plan → Act → Iterate

Schedule training
for Dec 5



→ check emails & calendar
for existing courses

→ check for course on
Dec 5

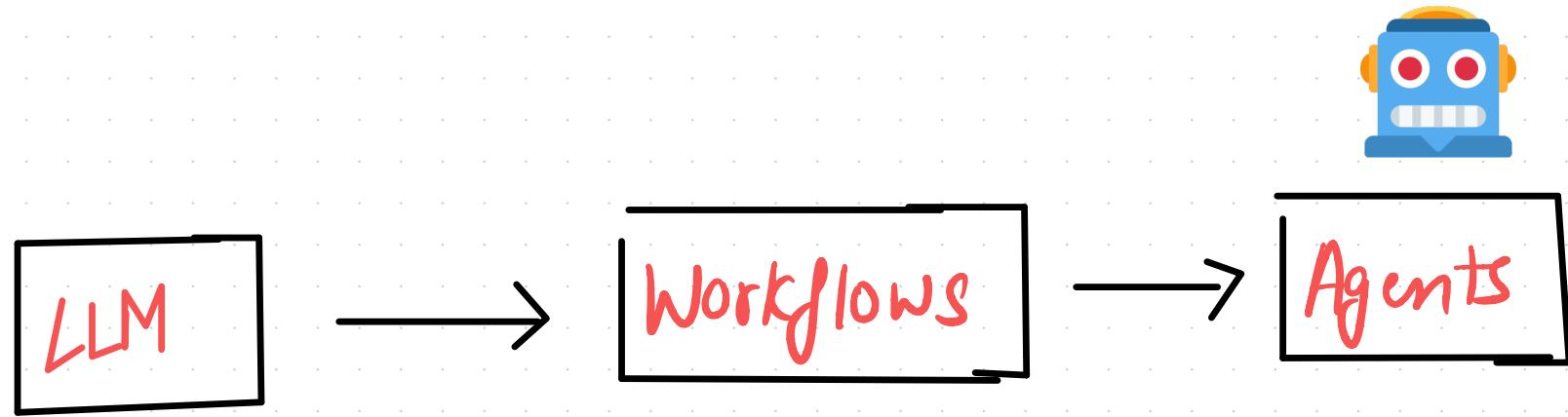


Send conf email to produce ← Add event to my Calendar ← No course found

Agents are flexible, adaptive and
well structured for open ended tasks

Workflows are efficient but follow
a fixed path

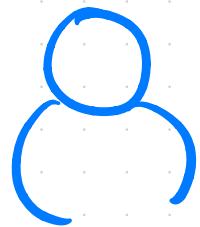
Evolution



MCP
(Model Context Protocol)

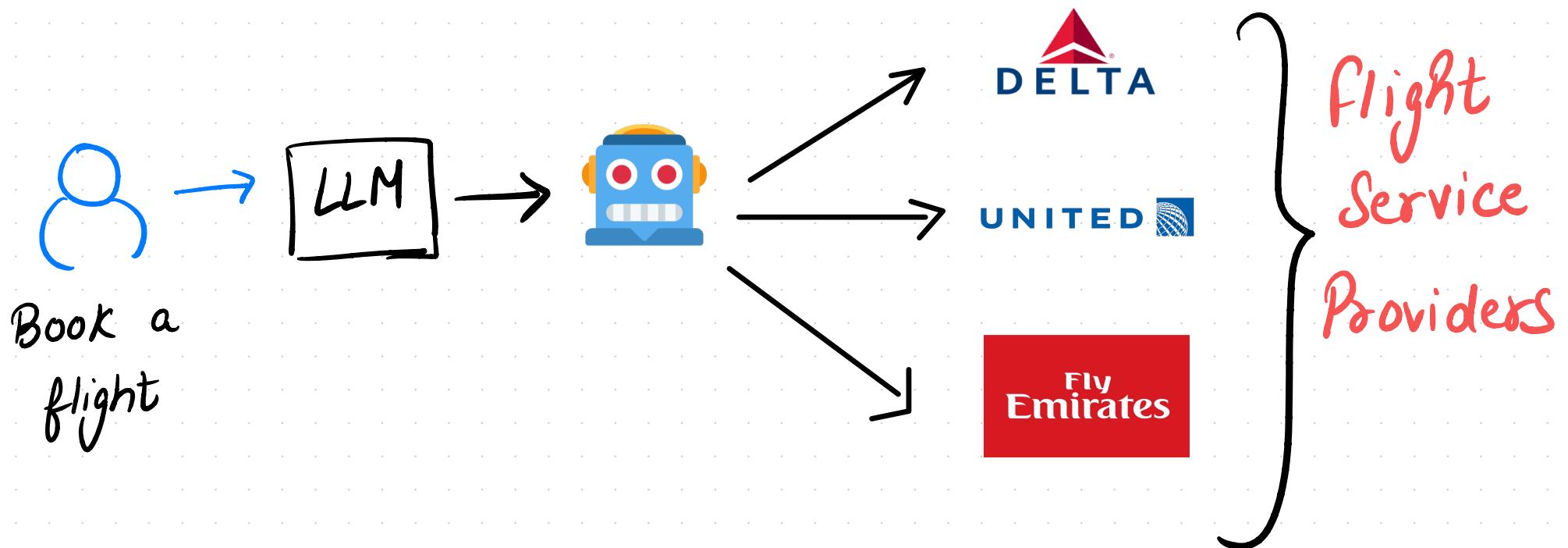


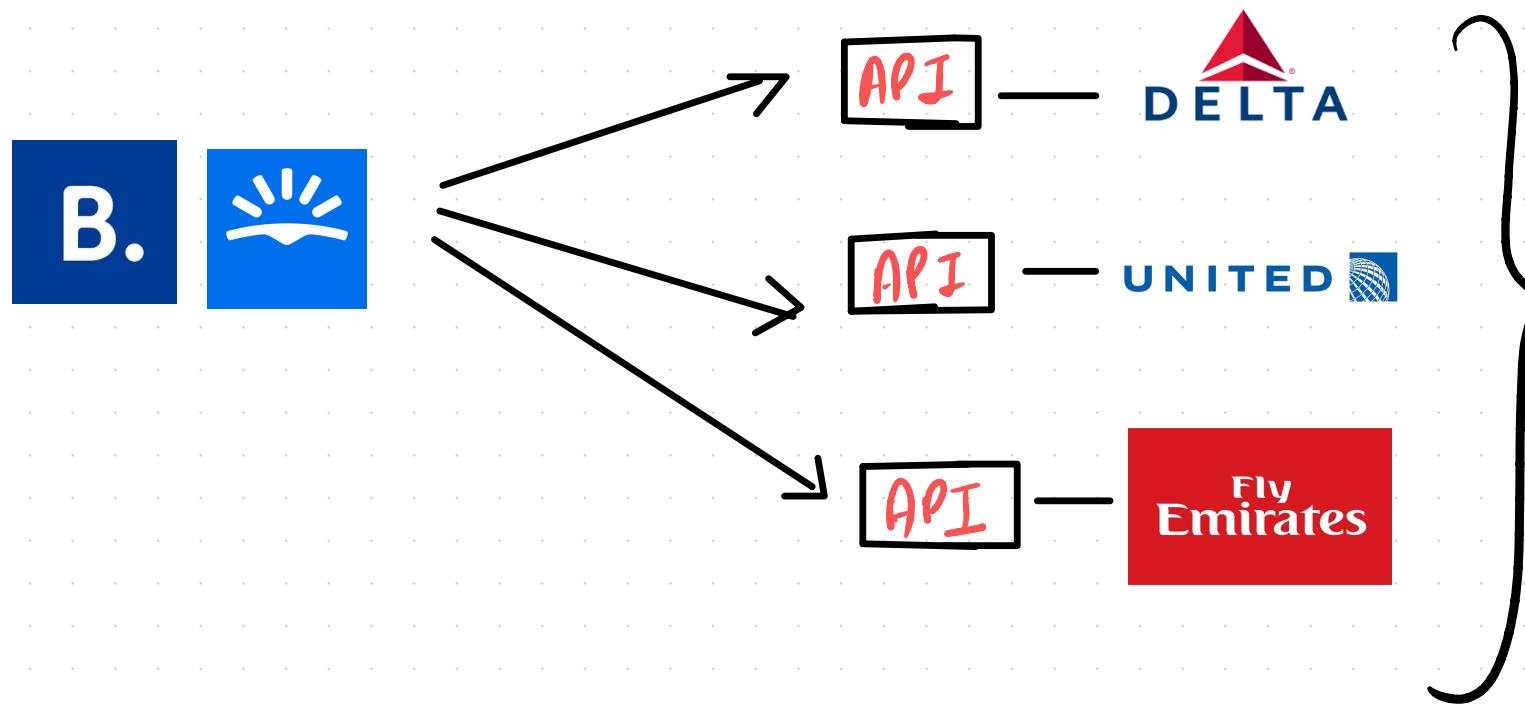
Booking



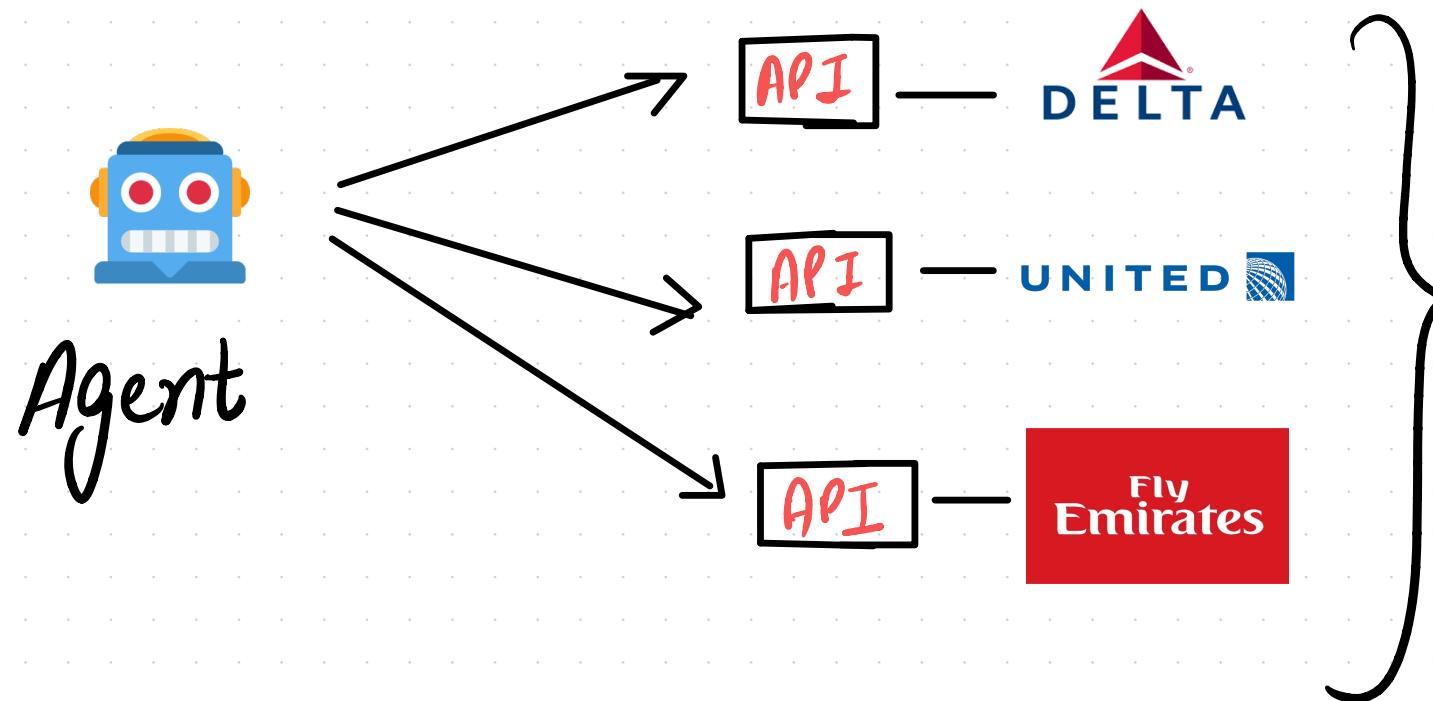
Book flight
from Chicago to
Amsterdam

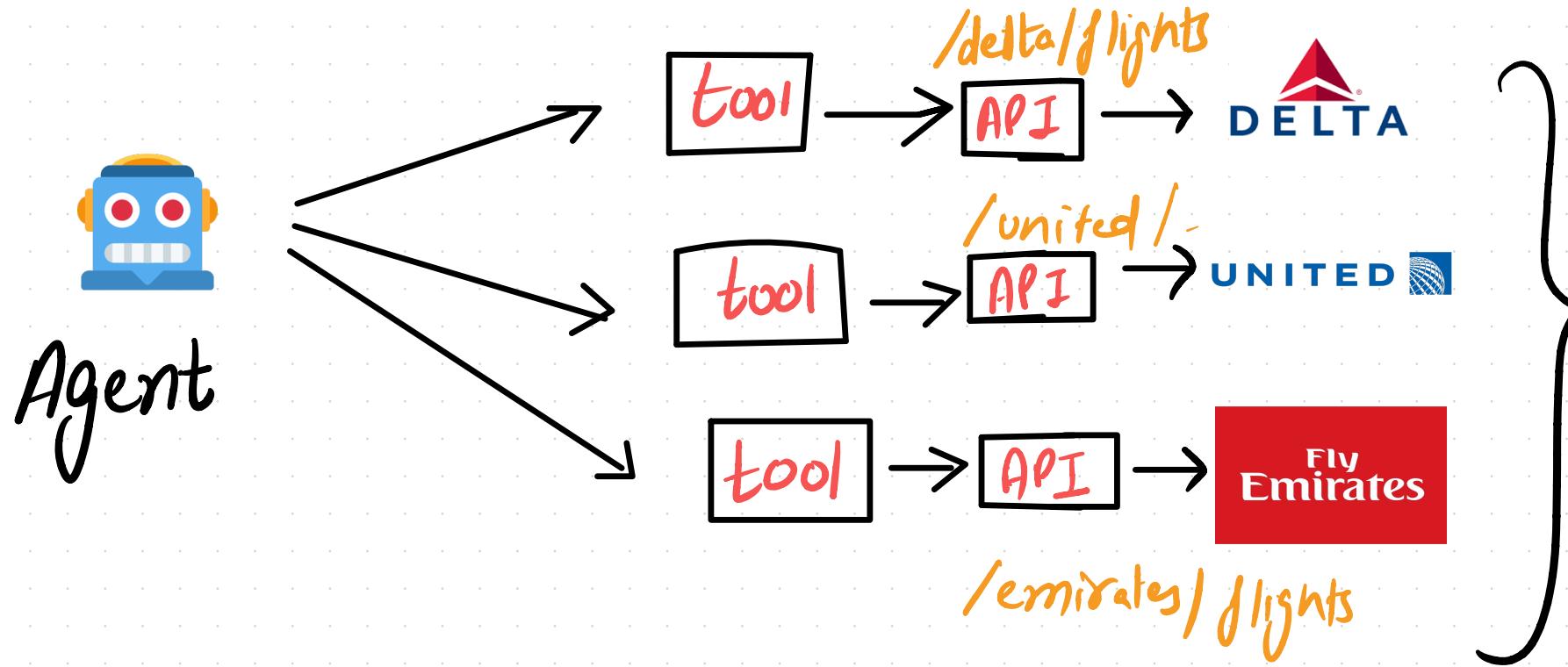
✗ LLM can't book
a flight but
only tells how to
book



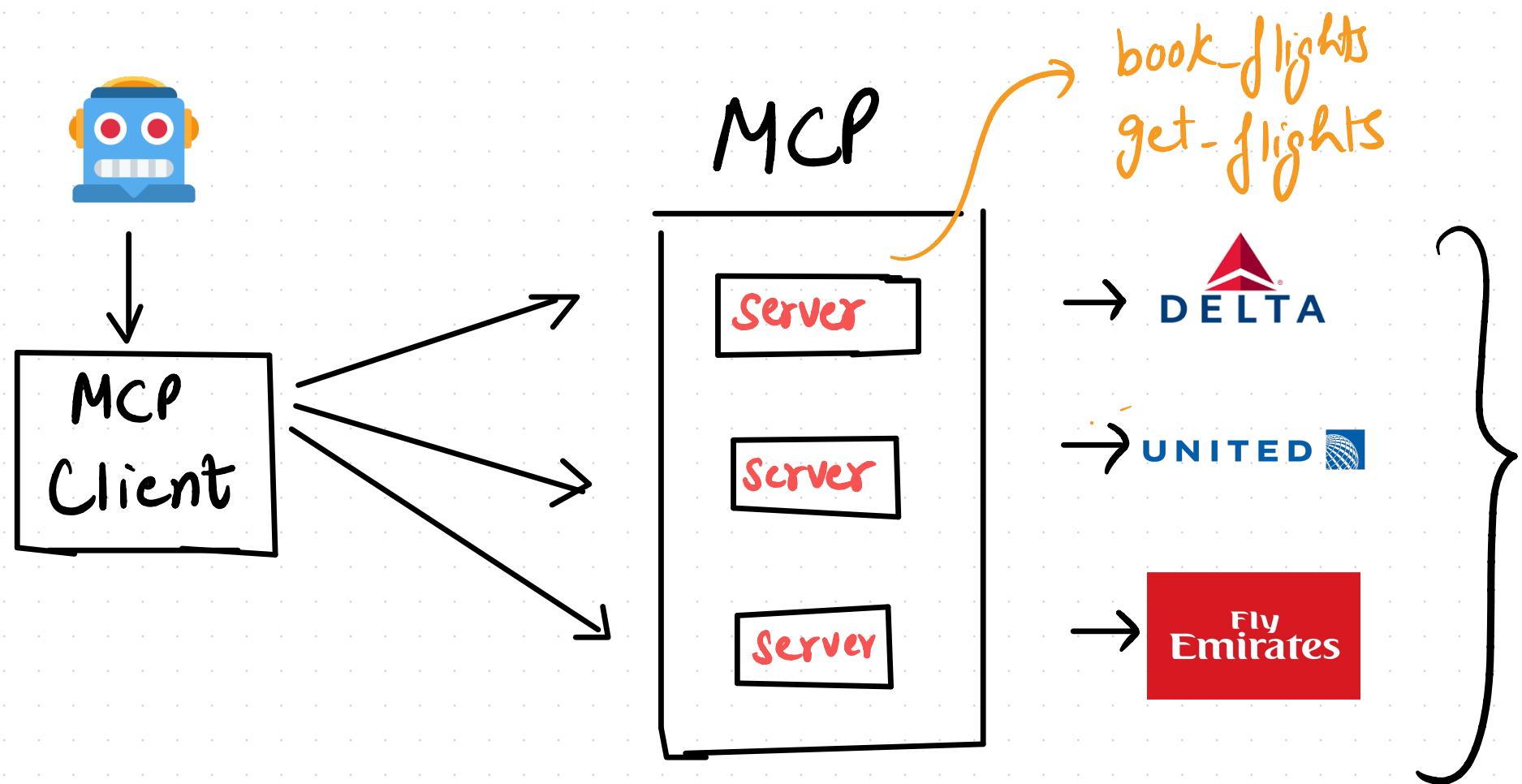


Instead of third party websites, we have Agents booking the flights

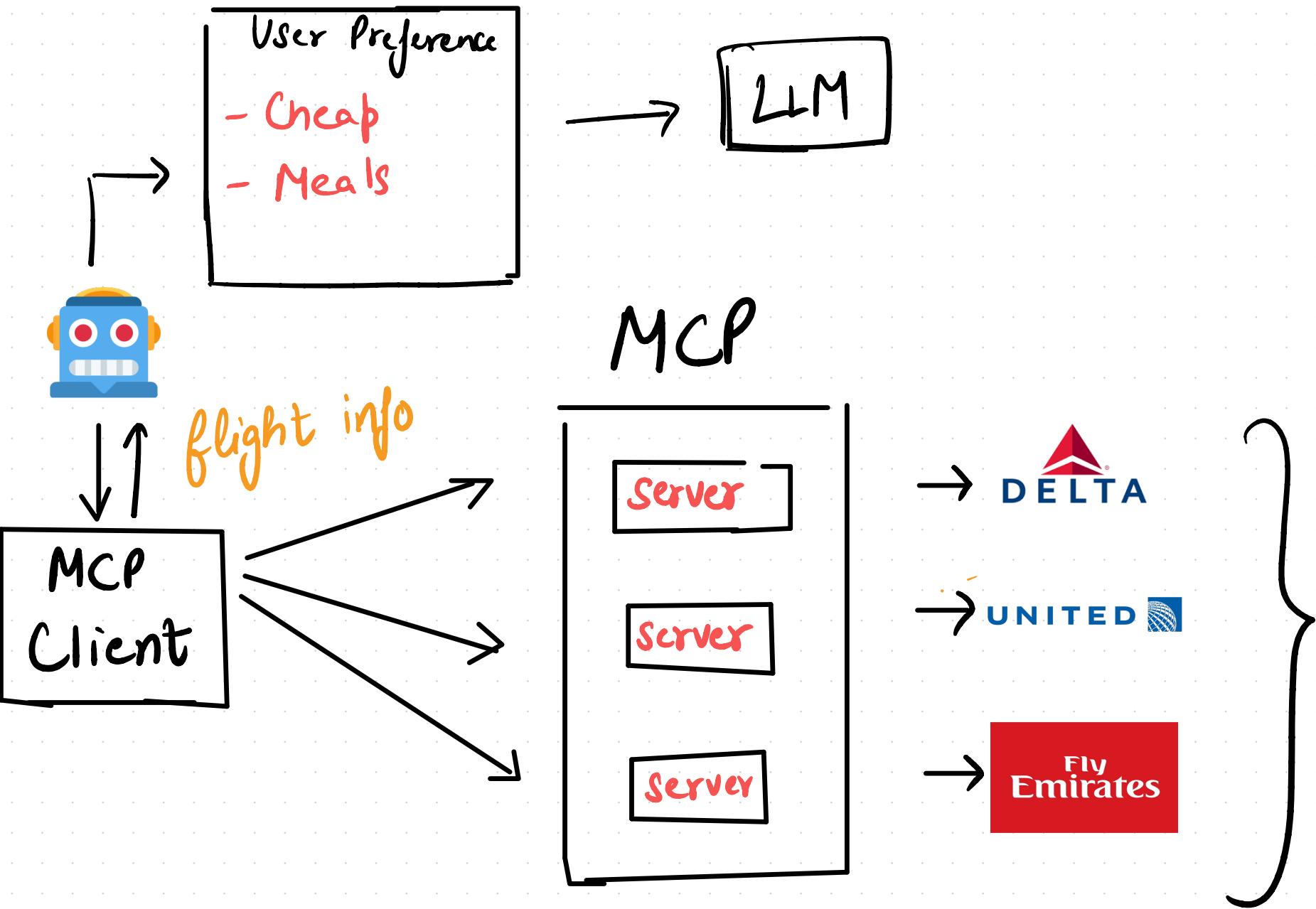


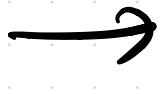
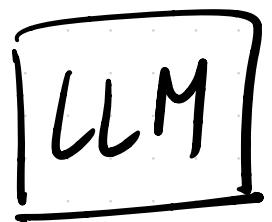


Agent will have to make
3 different API calls



Each website will expose its own MCP server
MCP client will talk to MCP server to get flight details





decides which
flight to book

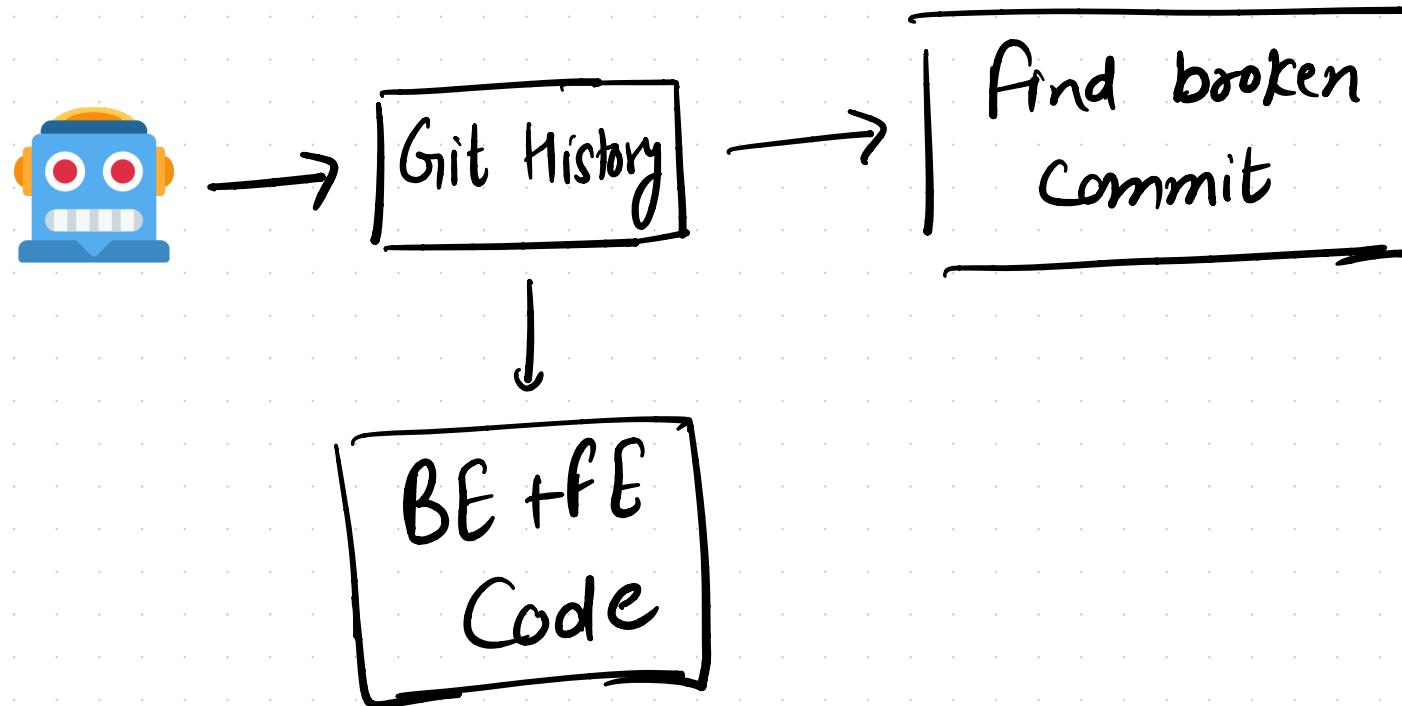


Book the flight

Agent

Real World Scenarios

→ Login functionality is broken on UI



JIRA MCP



→ create issue

→ get issue

→ Start to work on a ticket