Agenda

1. Intro to Agents ✓
2. Tools ✓
3. Planning ✓
4. Techniques → CoT → Reflection
5. Good Practices
6. Failures
7. Evaluation
8. Build your own Agent
   - ↳ existing platform
     - Agent.ai
     - crewai.com
   - ↳ Own code

How to build better agents

VC

Domain Research

Intro to Agents

Software Program

Chess Game ⇐ is sensitive about the environment

MiniGame ⇐ and takes up a desired goal

and acts to achieve that goal.

agent to research a particular topic

human agent

internet

act

AI agent

→ S/W program
→ Create a research doc on
→ AI in Education

abilities → visit website, search

Env

internet

file store

S/W systems

→ goal
→ supported f^n
→ achieve goal using the supp f^n

ChatGPT is an agent!!

→ image gen !

→ run code !

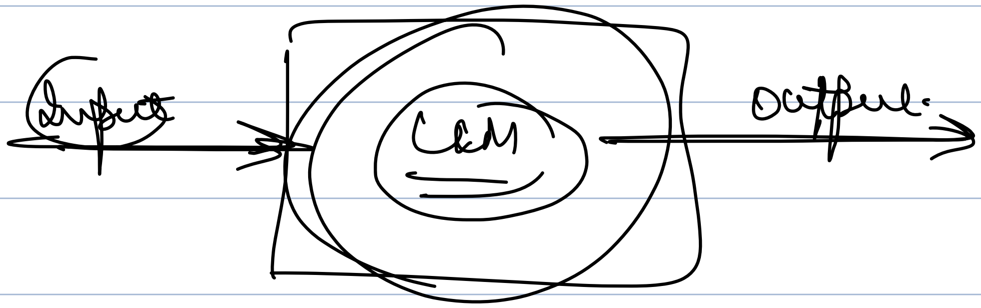
→ read file .

How will we build agent

What are LLM

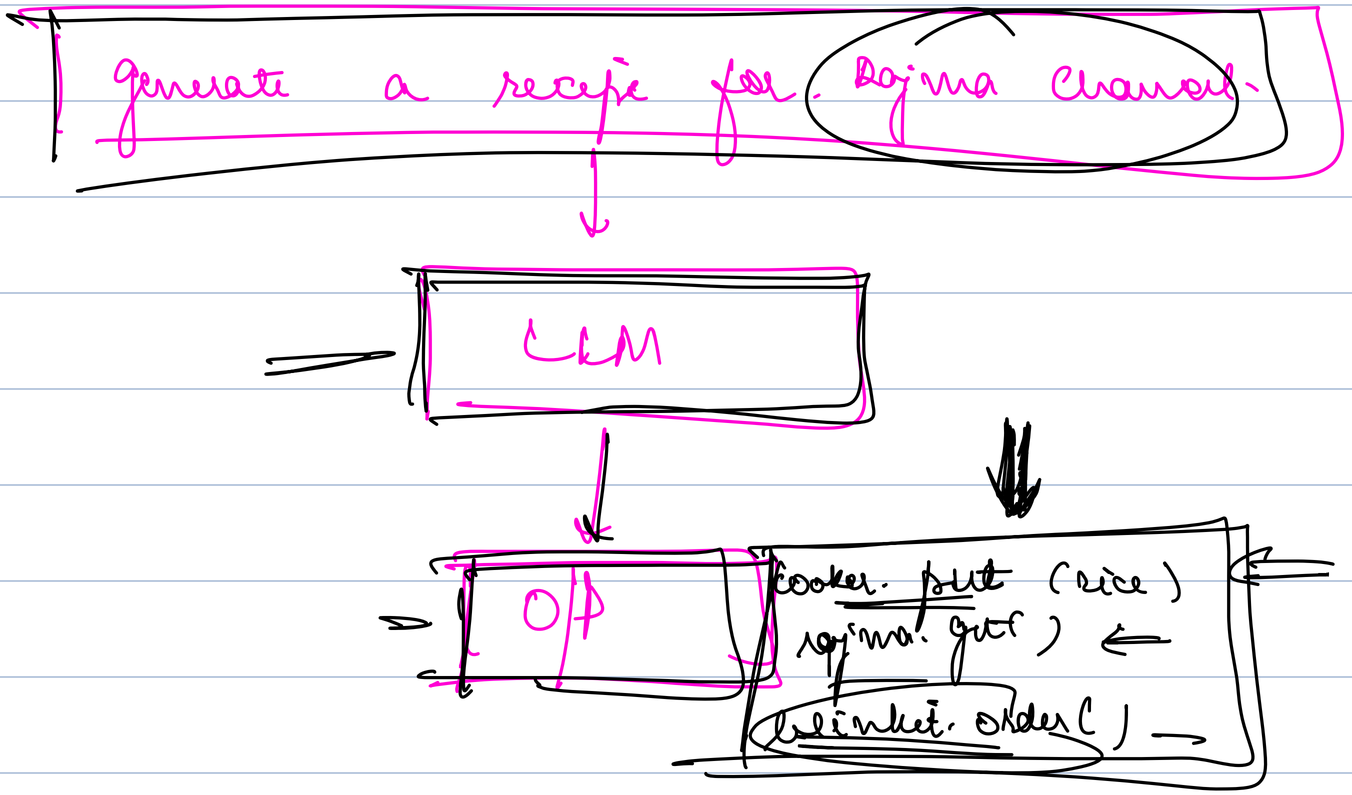
- → Large
- → Language
- → Models.

AI Models that are trained to predict/generate output based on a given input

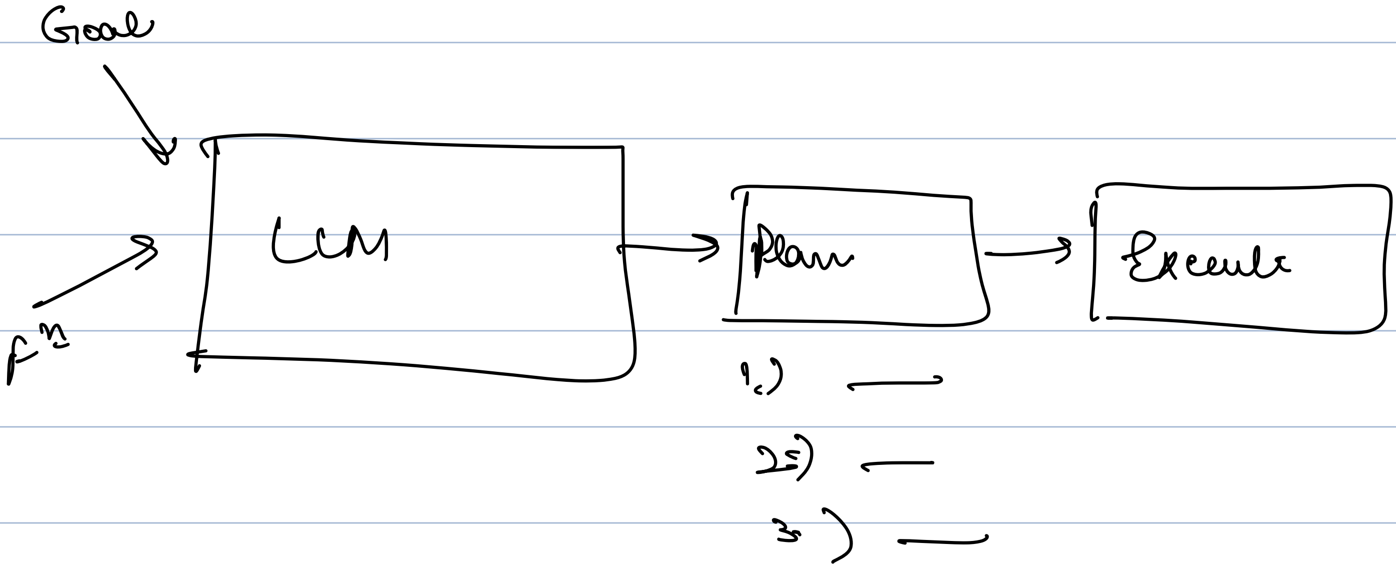


eg: GPT, Claude, Gemini, Deep Seek.

Can we use LLMs to build agents



```
cooker.put (rice)
rajma.get( )
blinkit.order( )
```



Why did we not have agents till now?..

1.) Compounding Mistakes.

2.) High Stakes

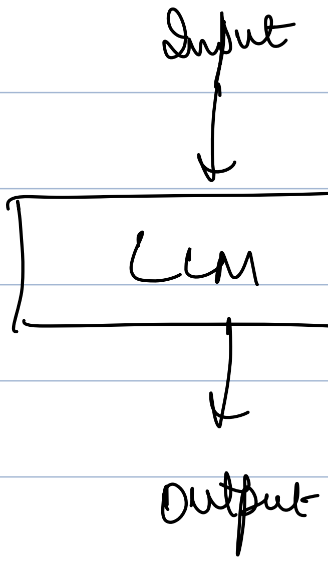
GPT 4o, Claude Sonnet 3.5.

Why Hype around them

AI agents have been considered to be the final AI

AGI Artificial Gen Intelligence

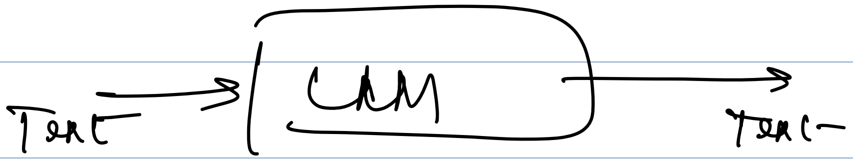
How to build agent



OpenAI's Chat GPT

```
resp = chatgpt . complete(
        "what is the capital of India?"
)
print(resp. text )
```

```
Delhi.
```

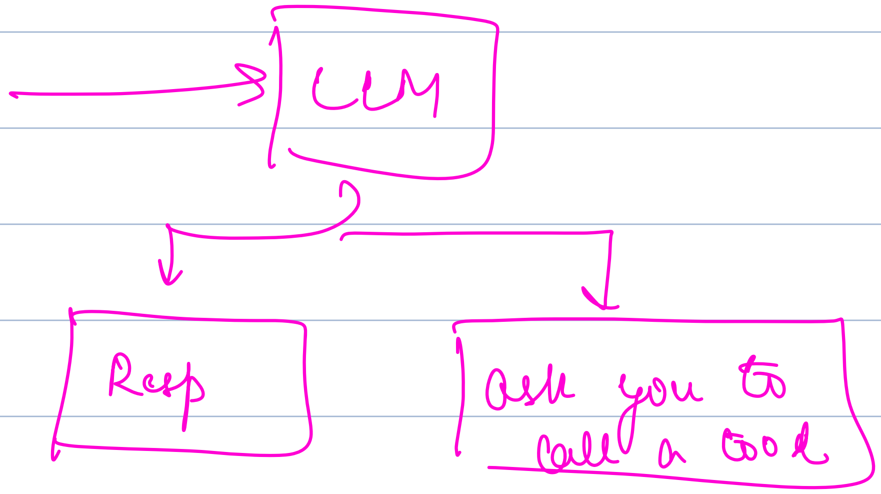


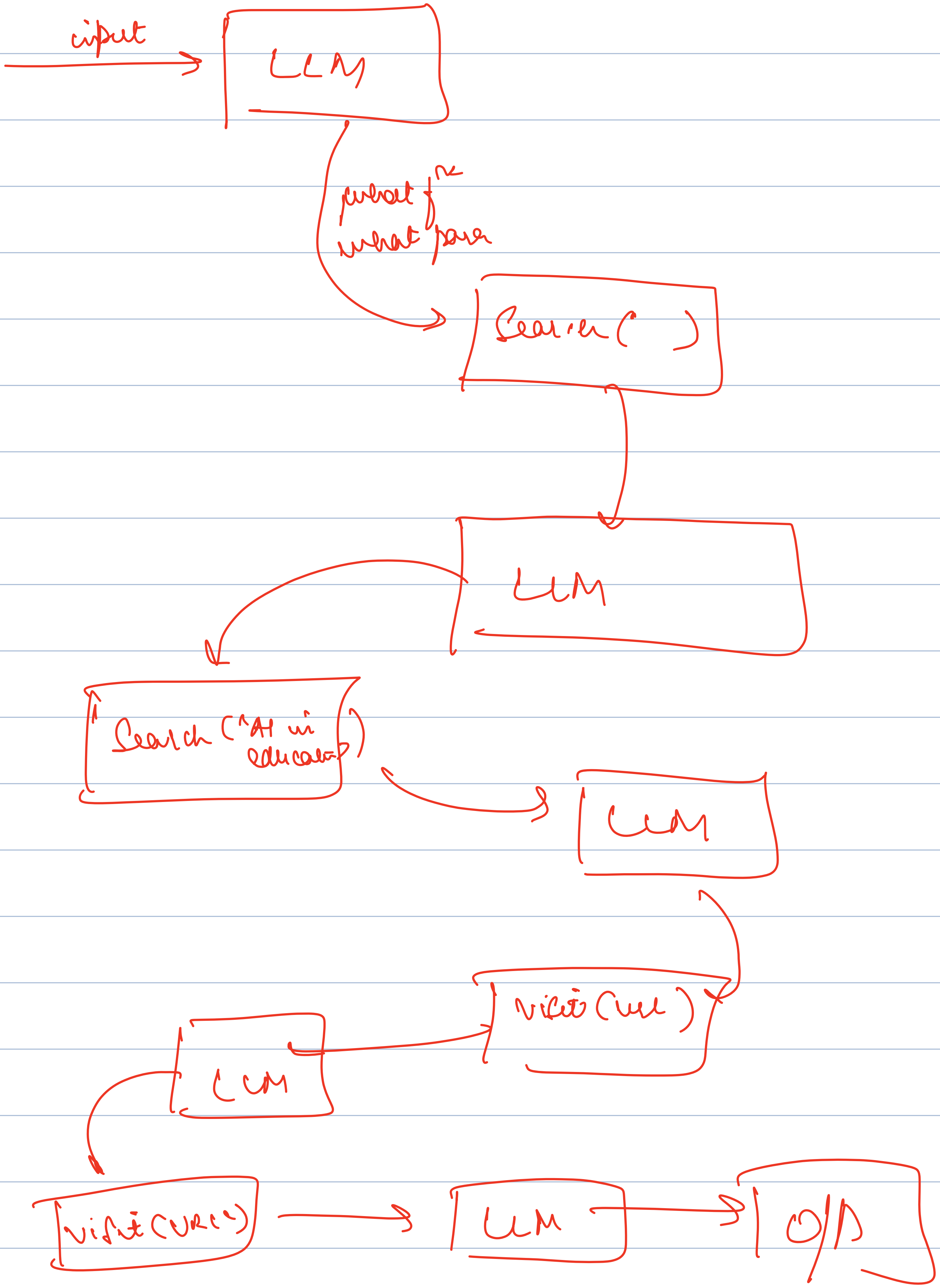
Support for TOOLS in their api

```
Search (query) { }

resp = ChatGPT.complete(
    'create a research document on
     benefits of AI in education',
    tools = [
        search (query),
        visit (anypage),
    ]
)

if (resp == call (AI education)) {
    Search (AI education)
}
```




input
LLM
what f^n
what para
Search( )
LLM
Search("AI in education")
LLM
visit(url)
LLM
visit(URL2)
LLM
O/P

Types of Tools that Exist

- → Context construction.
- → Capability extension
- → Action.

Context Construction

(adding knowledge to LLM)

LLMs are restricted by the knowledge available at time of training them.

- → Search (internet)
- → RAG
  - → Retrieval Aug Generation

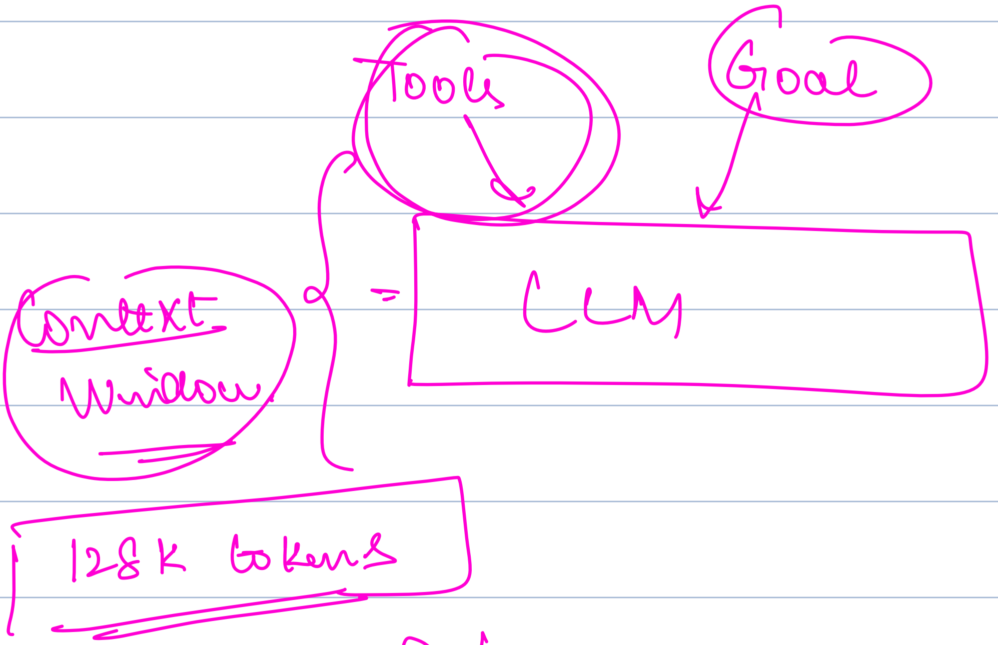
Capability Extension

- → ability to read a file
- → Code executor.
- → Calculator

Act

- → allowing to interact with real provider

Theoretically if I construct tools for everything I can build agent to do whatever I want



- → Bad Quality
- → Exhaust context window
- → Costly

⇒ 1 Big Agent

V/S

Multiple Smaller Agents

eg: [Deep Research]

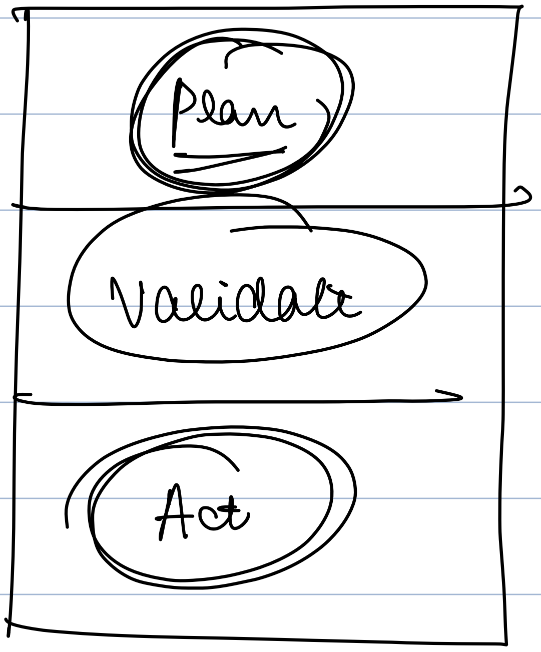
⇒ Weather Prediction Agent

⇒ Search Agent (Perplexity)

⇒ Cursor/ Copilot

How to Build Good Agents

"Project Plan"



Plan

"You are a VC. You have to do a research ___ ___

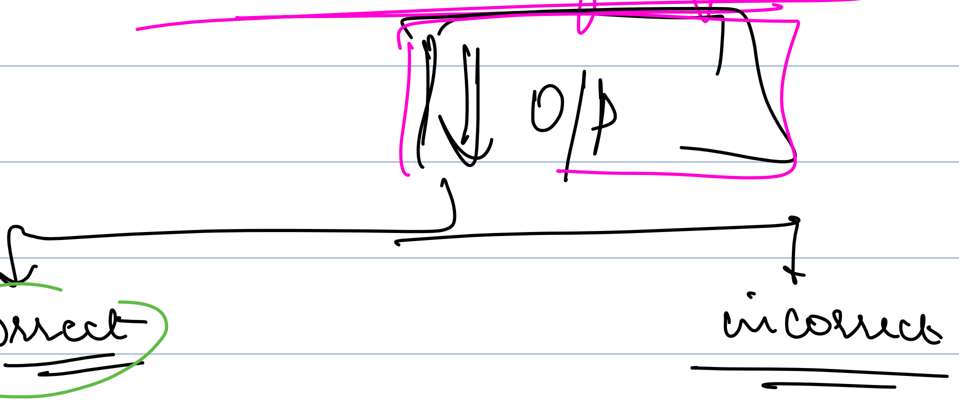
___

___

___

Create a plan step by step on how will you do research"

These are the f^n you have".



→ achieve the goal

→ efficient

→ feasible

Find companies > 1 B USD valuation
but 0 revenue

1000000

① Find all companies with 0 Revenue
② Filter companies with > 1 B valuation.

V/S

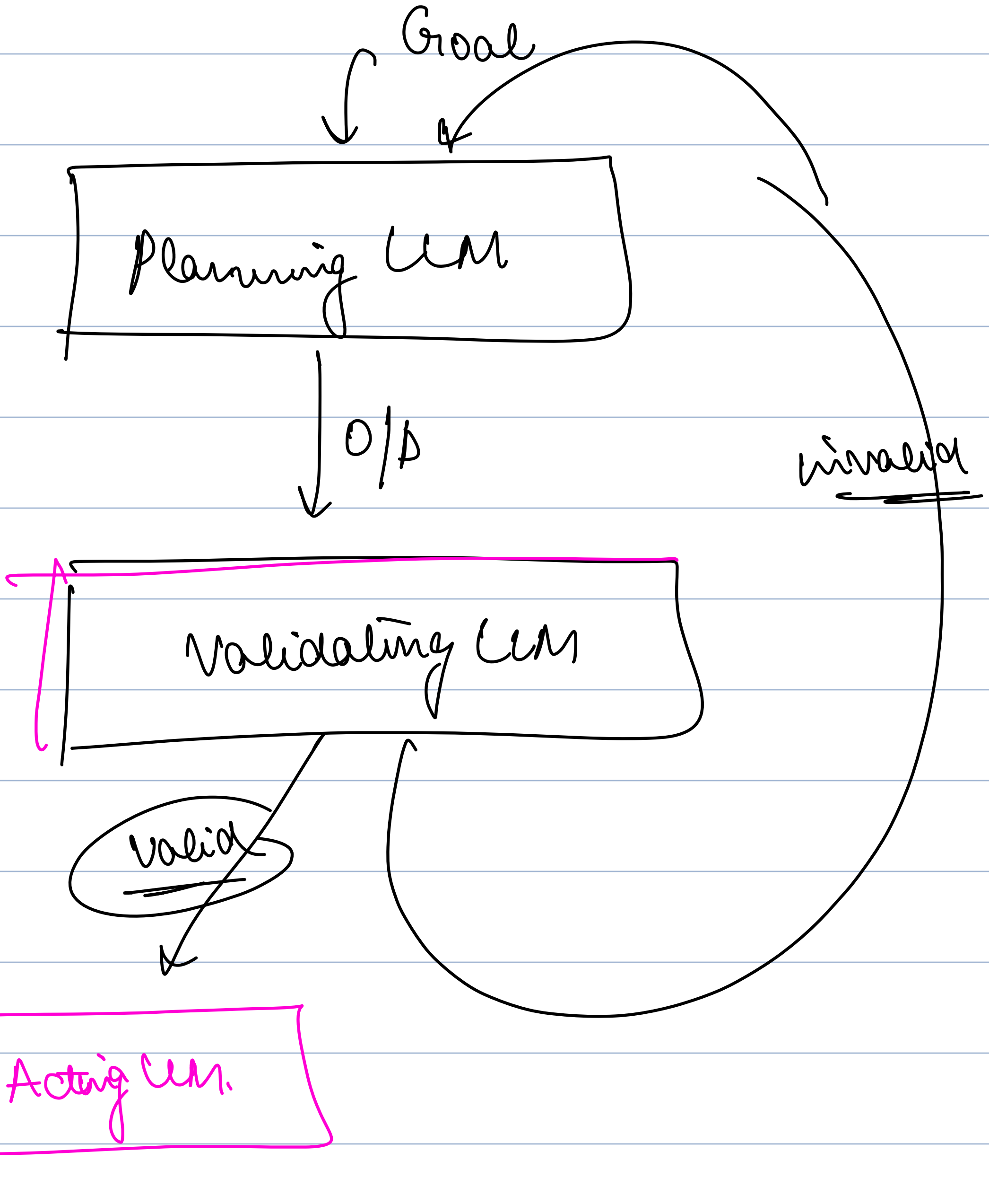
② → Fast

① Find companies with > 1 B valuation
② Filter 0 Revenue

500

Techniques for Better Quality

① CoT Prompting (Chain of Thought Technique)



Break till 8:50

CoT Prompting

Prompt1

what is the output of

⇒ $((8+3)*7)+((2+3)*6))$

x

70% accuracy

Prompt2

what is the output of

First tell me the series of steps you will take to calculate and at end give final result.

result | 92% accura

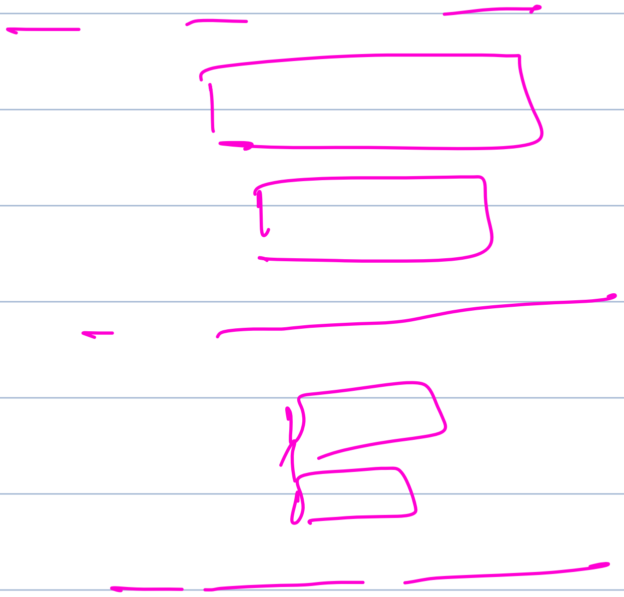
Reflection

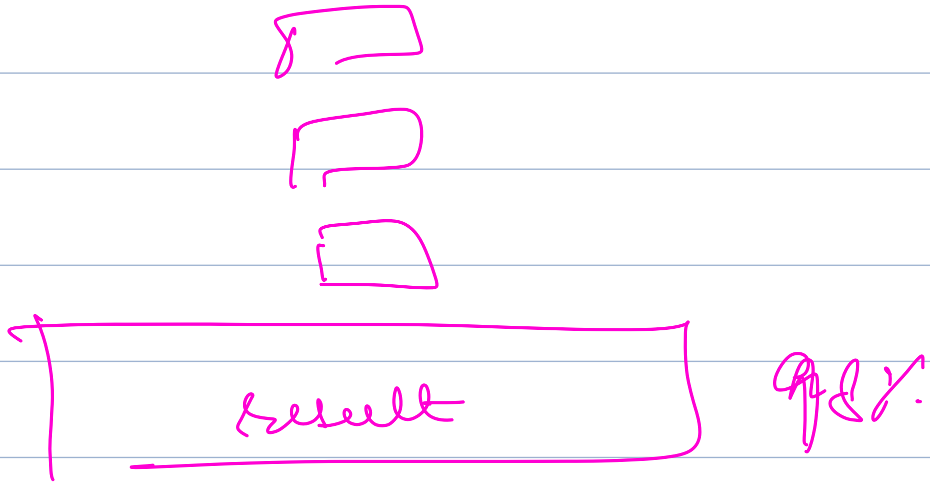
What is the output of

9^n

First think step by step.
For each step first think
if it is correct and then
move next

finally generate the result.





Failures

Search (query)

① Planning

↳ invalid tool

↳ valid tool, invalid params

Search (2, query)

↳ valid tool, valid params, incorrect value

search ("AI cotton").

"Evolution of Everything"

② Goal Failure.

③ Tool Failure.

Evaluation

1. How many Cops,
2. How much cost
3. How much time,

V/S a human

AI Engineering

— Chip Huyen

2 Hr of what we learn in 12 months

Prior coding exp
and
Prior Math comfort

↳ Scaler's SWE program

enhanced multi Gen AI