

AI & MCP

source**to**ad

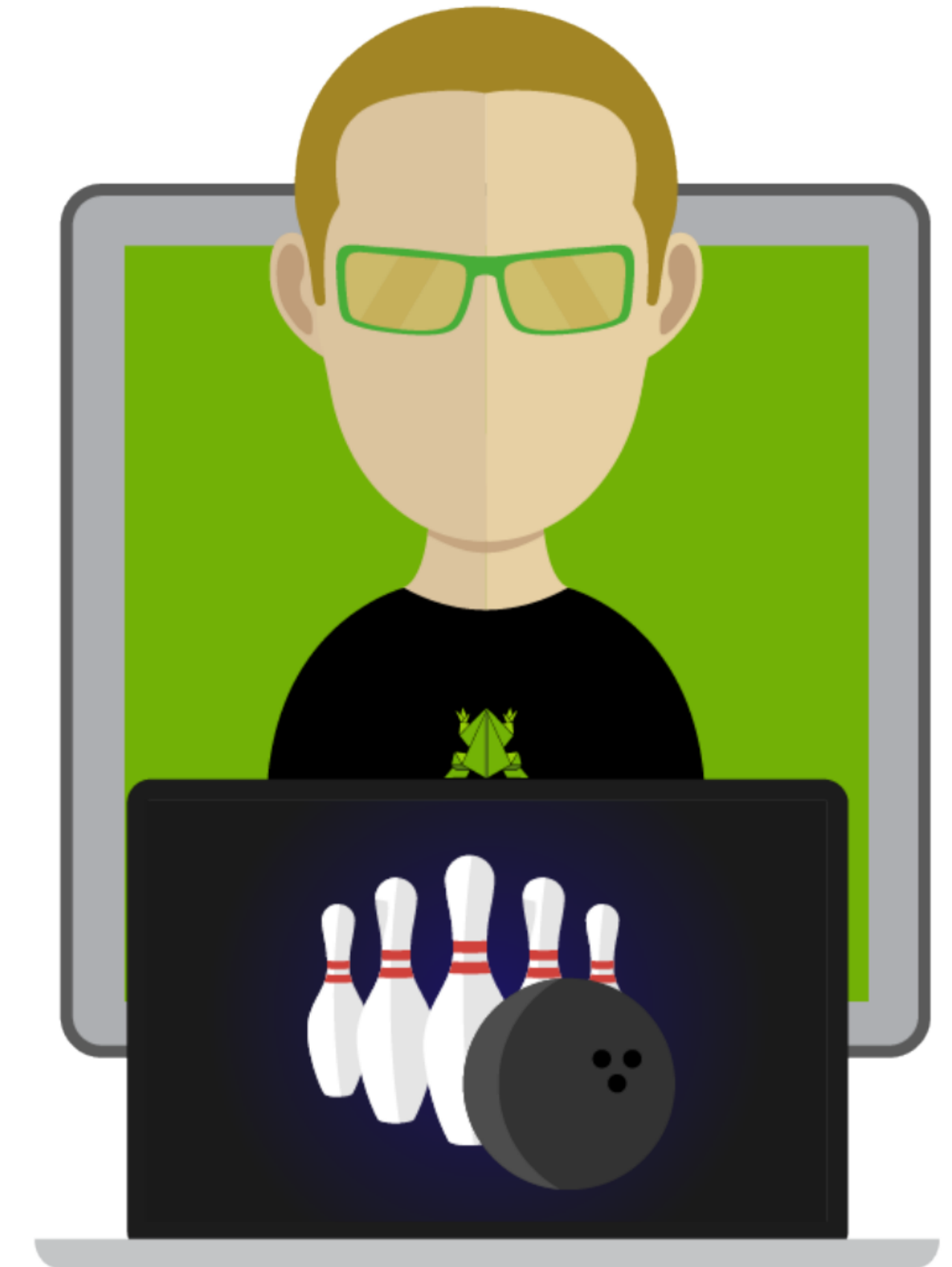


Plan of Attack

- The beginning of the LLM
- A bit of History (GPT1-5)
- The Evolution of Tools
- Introducing MCP
- MCP Features & Security

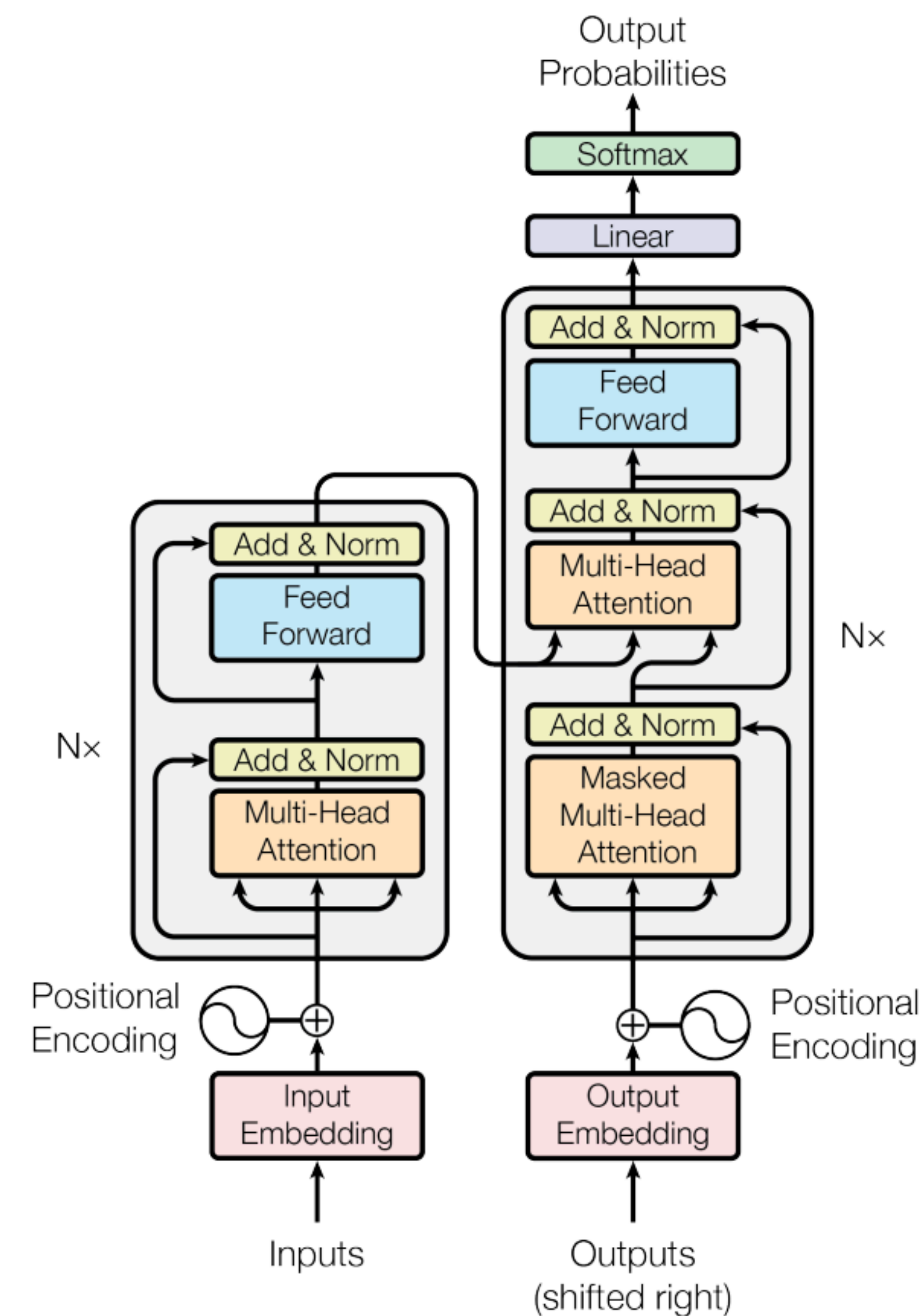
Who?

- **Connor Tumbleson**
- @iBotPeaches
- Sourcetoad
 - Director of Engineering

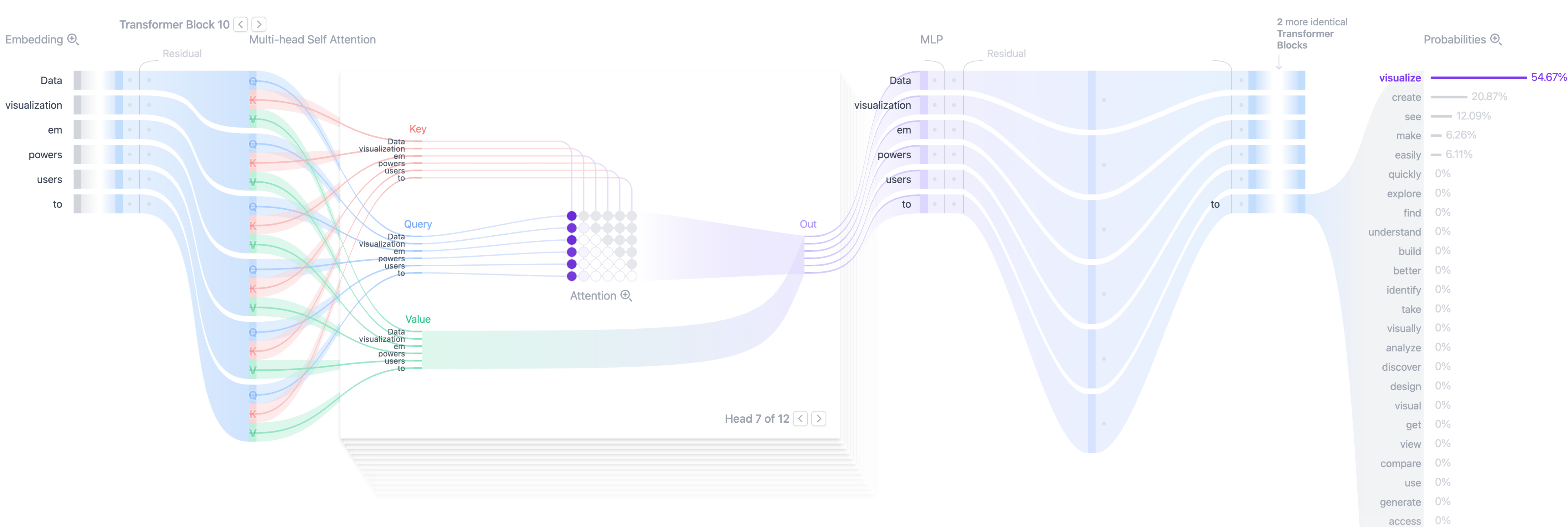


Attention Is All You Need

- 2017 Research Paper
- Google Scientists
- 8 equally cited authors
- Cited over 200,000 times.



Transformer Explainer





ChatGPT 1-5

GPT-1: The beginning (2018)

- Heavy research usage
- Trained from books
- Tokens in/out
- ~500 token limit



Hold up. **Tokens?**

Tokens

6

Characters

23

Hello my name is connor

Text

Token IDs

Tokens

6

Characters

23

[13225, 922, 1308, 382, 10850, 267]

Text

Token IDs

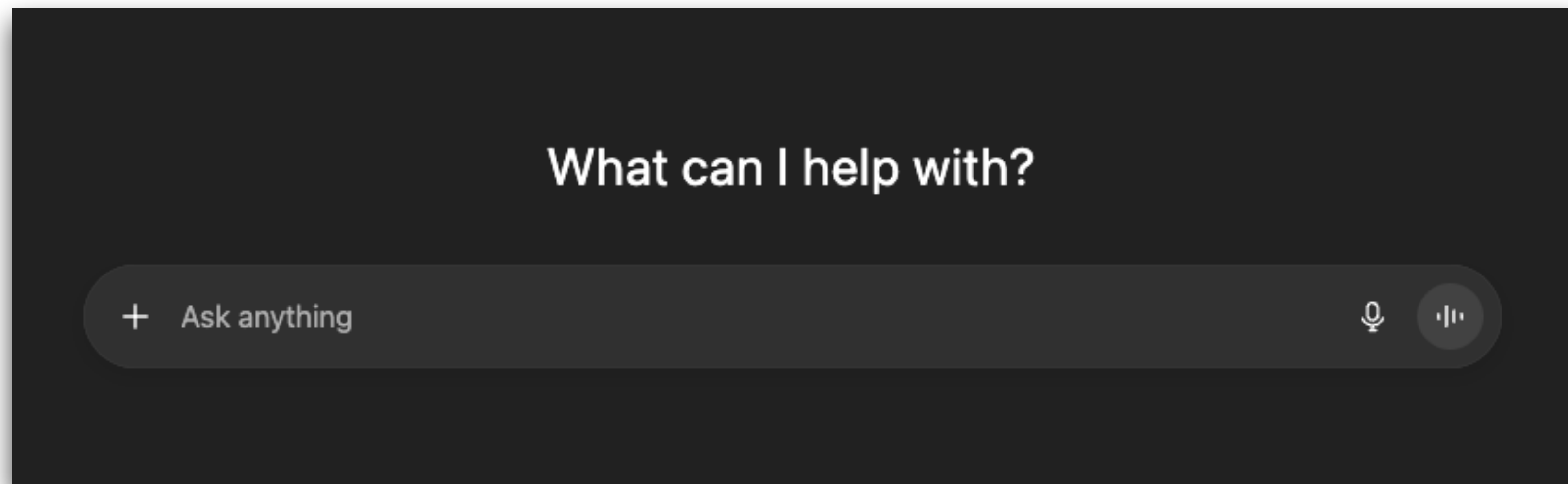
GPT-2: The Expanse (2019)

- Zero-shot training
- Trained from web/books
- $\sim 1k$ token limit



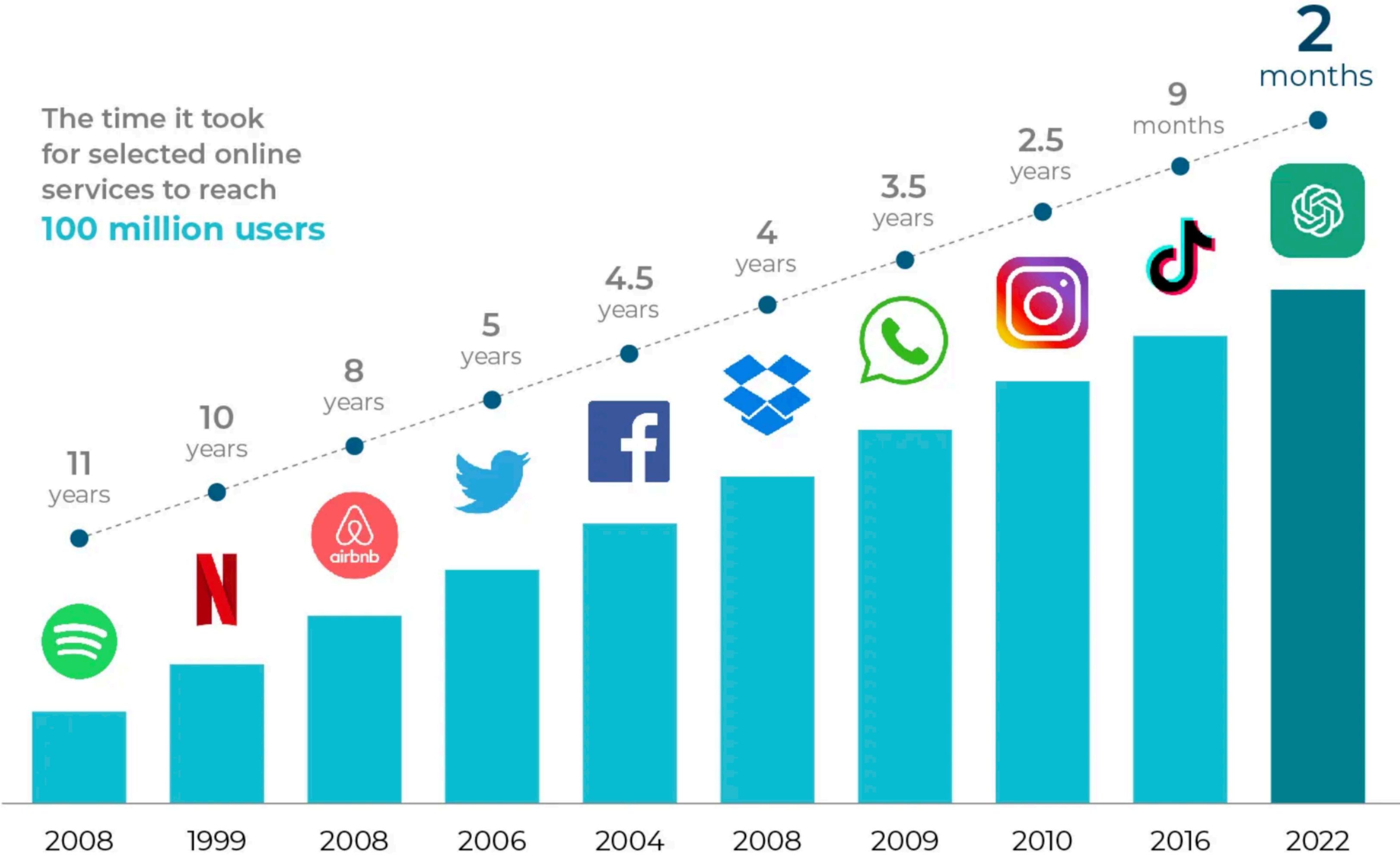
GPT-3: The Explosion (2020)

- The commercialization begins
- Trained from everything?
- $\sim 2k$ token limit



GPT-3: The Explosion (2020)

Chat-GPT sprints to 100 million users



GPT-3: The Explosion (2020)

- /completions
 - logit_bias
 - presence_penalty
 - seed
 - temperature
 - top_p

Create completion Legacy

POST <https://api.openai.com/v1/completions>

Creates a completion for the provided prompt and parameters.

Request body

model string Required

ID of the model to use. You can use the [List models](#) API to see all of your available models, or see our [Model overview](#) for descriptions of them.

prompt string or array Required

The prompt(s) to generate completions for, encoded as a string, array of strings, array of tokens, or array of token arrays.

Note that <|endoftext|> is the document separator that the model sees during training, so if a prompt is not specified the model will generate as if from the beginning of a new document.

GPT-4: The Refinement (2023)

- Multimodal Input
- Trained from more and more.
- $\sim 32k$ token limit

GPT-4 can accept a prompt of text and images, which—parallel to the text-only setting—lets the user specify any vision or language task.

GPT-4: The Refinement (2023)

- New API `/chat/completions`

Create chat completion

POST `https://api.openai.com/v1/chat/completions`

Starting a new project? We recommend trying Responses to take advantage of the latest OpenAI platform features. Compare Chat Completions with Responses.

GPT-4: The Refinement (2023)

- New API `/assistants`

Create assistant Beta

POST `https://api.openai.com/v1/assistants`

Create an assistant with a model and instructions.

GPT-4: The Refinement (2024)

- New API `/responses`

Create a model response

POST `https://api.openai.com/v1/responses`

Creates a model response. Provide text or image inputs to generate text or JSON outputs. Have the model call your own custom code or use built-in tools like web search or file search to use your own data as input for the model's response.

GPT-5: The Reasoning (2025)

- Reasoning / Completion
- Trained from [redacted]
- ~512k token limit

ChatGPT 5 ▾

GPT-5

Auto

Decides how long to think



Instant

Answers right away

Thinking

Thinks longer for better answers

Pro

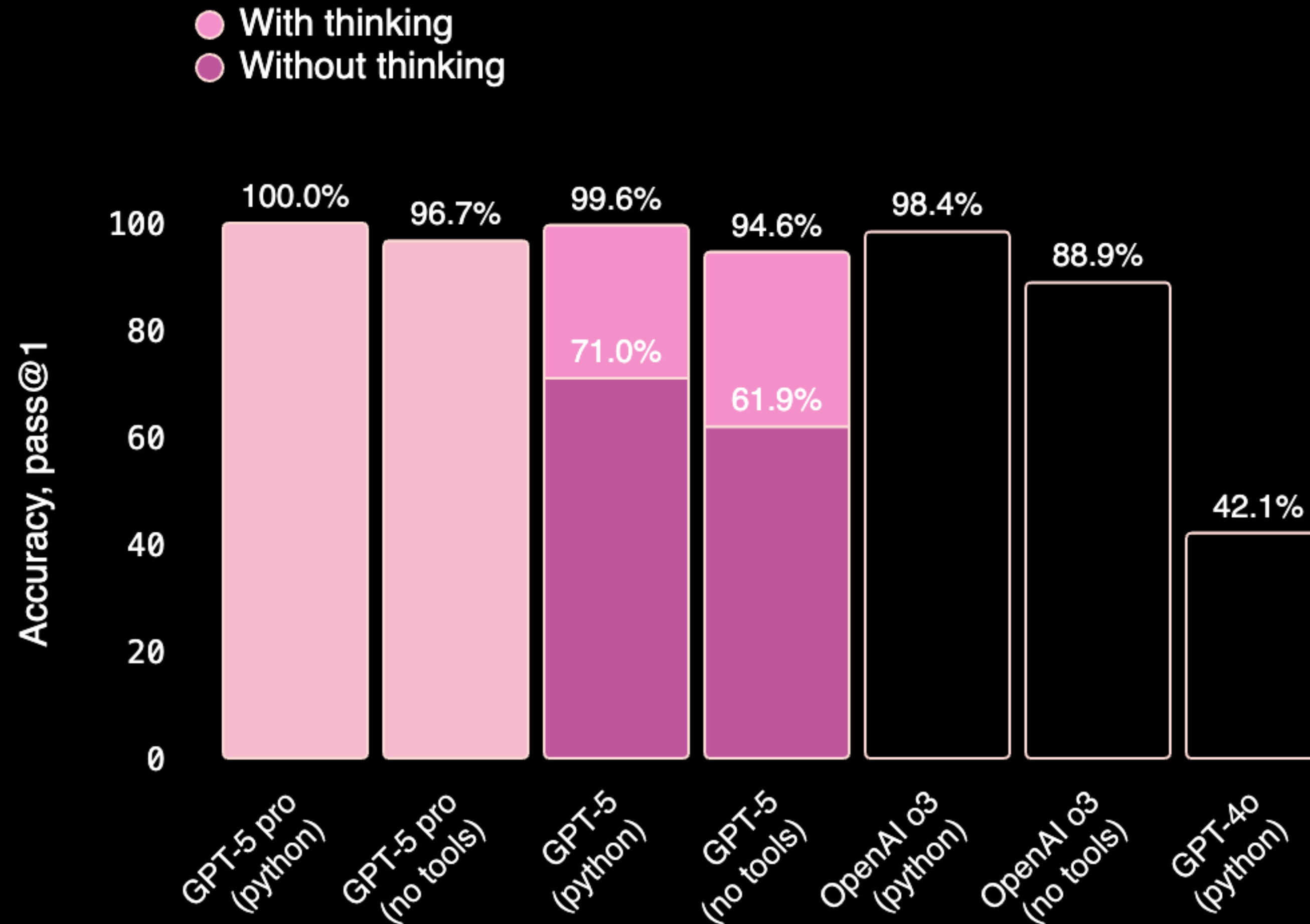
Research-grade intelligence

Legacy models



GPT-5: The Reasoning (2025)

AIME 2025
Competition math



GPT-1 through GPT-5 Parameters

- Brains of the model
- Getting smarter
- Nearing AGI?

	Parameters	Year
GPT-1	~117 million	2018
GPT-2	~1.5 billion	2019
GPT-3	~175 billion	2020
GPT-4	~1.7 trillion	2023
GPT-5	??	2025



LLM Tools

LLM Tools

- Code Interpreter
- File Search
- Function Tool
- Web Search
- Computer Use
- Image Generation
- Local Shell
- Custom Tool Call
- Connectors
- MCP

Tool: Code Interpreter

- Write Code
- Run Code

```
curl https://api.openai.com/v1/responses \
-H "Content-Type: application/json" \
-H "Authorization: Bearer $OPENAI_API_KEY" \
-d '{
  "model": "gpt-4.1",
  "tools": [{
    "type": "code_interpreter",
    "container": { "type": "auto" }
  }],
  "instructions": "You are a personal math tutor. When asked a math que
  "input": "I need to solve the equation  $3x + 11 = 14$ . Can you help me?"
}'
```

Tool: Code Interpreter

- How Many R's

ChatGPT Error >

How many 'r' in 'strawberry'

ChatGPT

There are two 'r' characters in the word 'strawberry'.

AS

You

Check again please

ChatGPT

Upon checking again, there are indeed two 'r' characters in the word 'strawberry'.

AS

You

Are you 100% sure?

ChatGPT

Yes, I am 100% sure. The word 'strawberry' contains two 'r' characters.

AS

You

Would you bet a million dollars on this?

ChatGPT

Yes, I would. The word "strawberry"

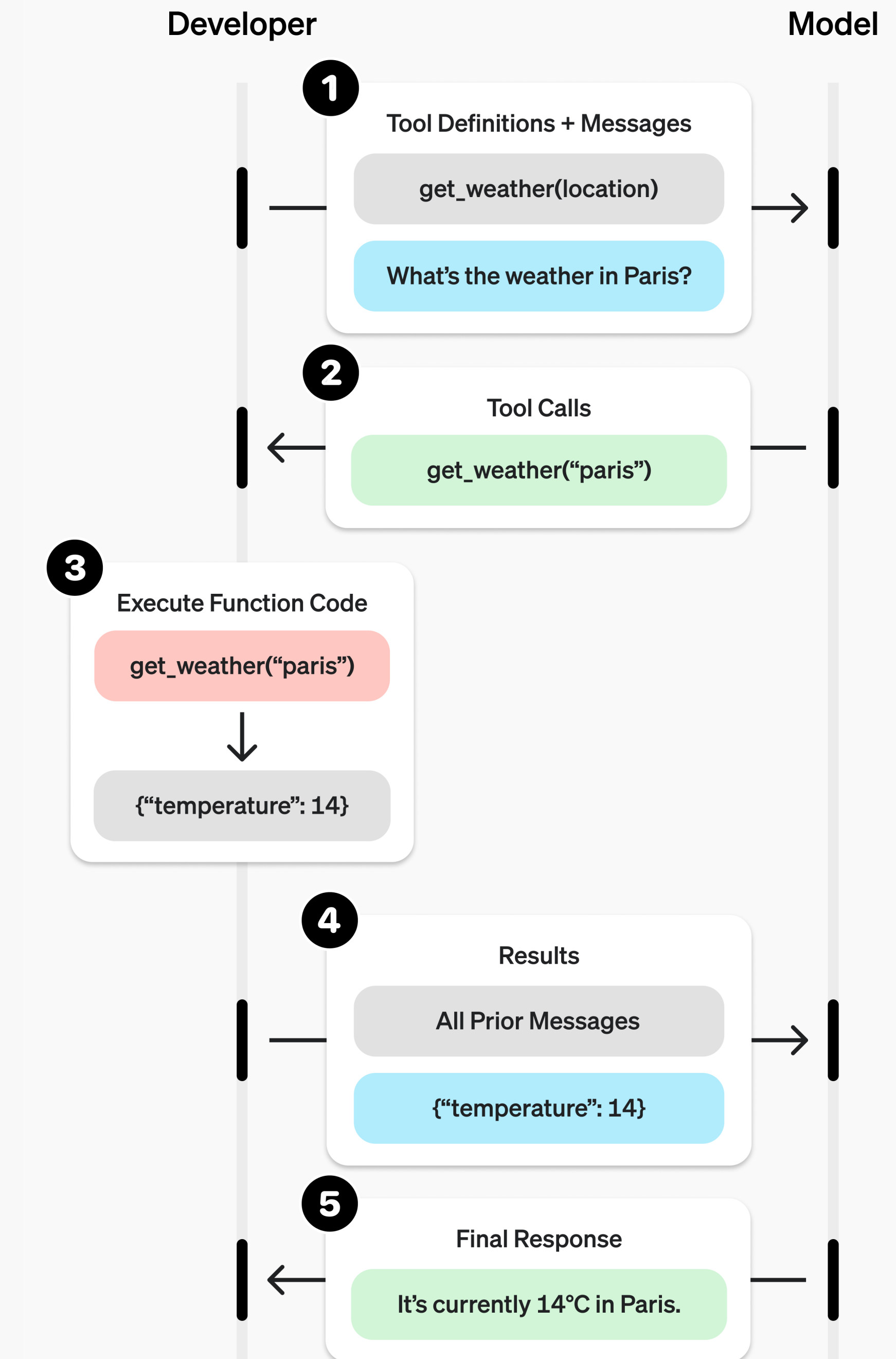
Tool: File Search

- Create Vector Store
- Attach Files
 - images, text, docs.
- Search it

```
{
  "output": [
    {
      "type": "file_search_call",
      "id": "fs_67c09ccea8c48191ade9367e3ba71515",
      "status": "completed",
      "queries": ["What is deep research?"],
      "search_results": null
    },
    {
      "id": "msg_67c09cd3091c819185af2be5d13d87de",
      "type": "message",
      "role": "assistant",
      "content": [
        {
          "type": "output_text",
          "text": "Deep research is a sophisticated capability",
          "annotations": [
            {
              "type": "file_citation",
              "index": 992,
              "file_id": "file-2dtbBZdjtDKS8eqWxqbgDi",
              "filename": "deep_research_blog.pdf"
            }
          ]
        }
      ]
    }
  ]
}
```

Tool: Function Tool

- Create functions
- Expose to LLM
- LLM picks function

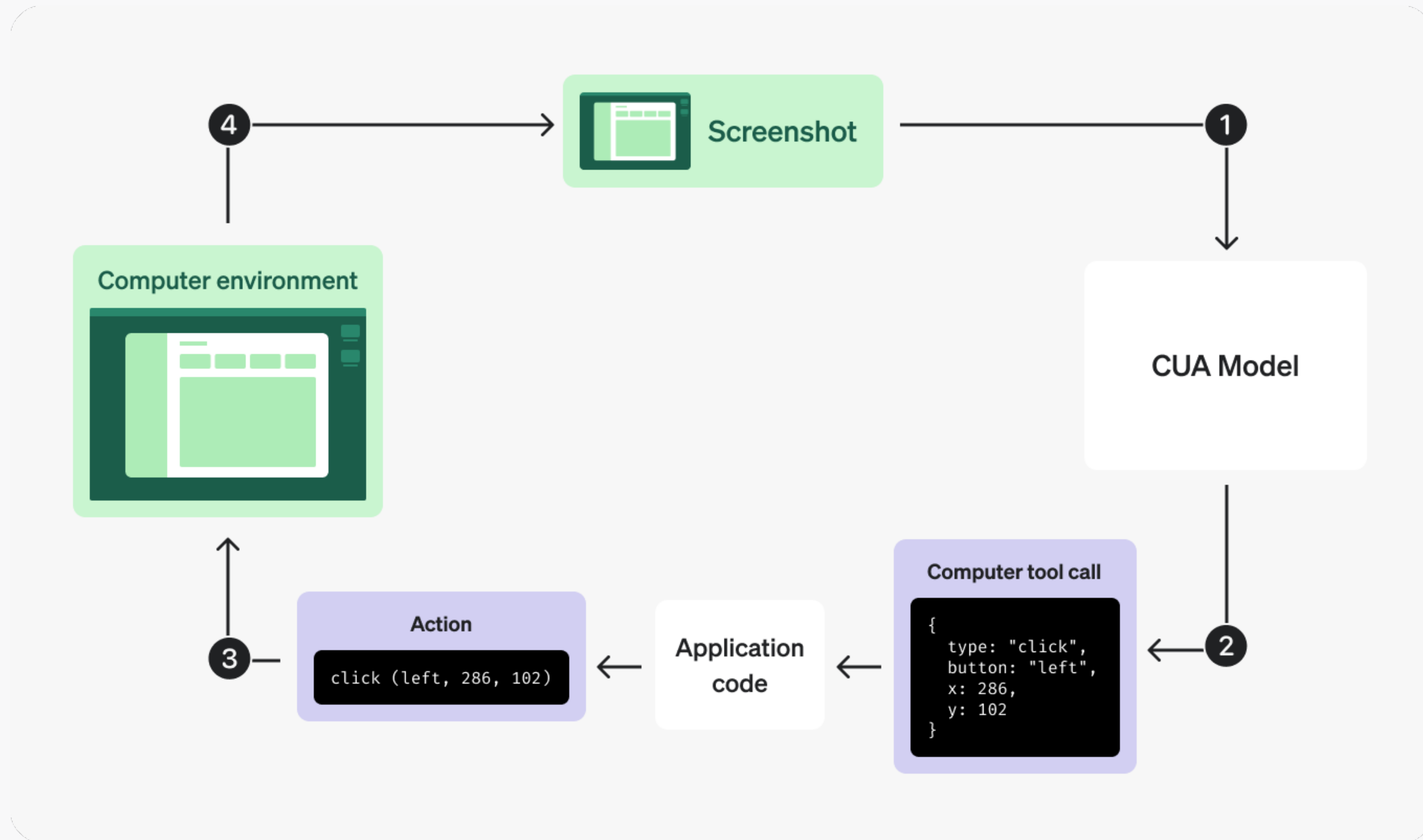


Tool: Web Search

- Search the web
- Cite the source

```
[
  {
    "type": "web_search_call",
    "id": "ws_67c9fa0502748190b7dd390736892e100be649c1a5ff9609",
    "status": "completed"
  },
  {
    "id": "msg_67c9fa077e288190af08fdffda2e34f20be649c1a5ff9609",
    "type": "message",
    "status": "completed",
    "role": "assistant",
    "content": [
      {
        "type": "output_text",
        "text": "On March 6, 2025, several news...",
        "annotations": [
          {
            "type": "url_citation",
            "start_index": 2606,
            "end_index": 2758,
            "url": "https://...",
            "title": "Title..."
          }
        ]
      }
    ]
  }
]
```


Tool: Computer Use



Tool: Image Generation

- Describe
- Wait
- Image
- Iterate

"Generate an image of gray tabby cat hugging an otter with an orange scarf"



"Now make it look realistic"



Tool: Local Shell

```
>_ Welcome to Codex, OpenAI's command-line coding agent

> Sign in with ChatGPT to use Codex as part of your paid plan
  or connect an API key for usage-based billing

> 1. Sign in with ChatGPT
   Usage included with Plus, Pro, and Team plans
  2. Provide your own API key
   Pay for what you use

Press Enter to continue
```


Tool: Custom Tool Call

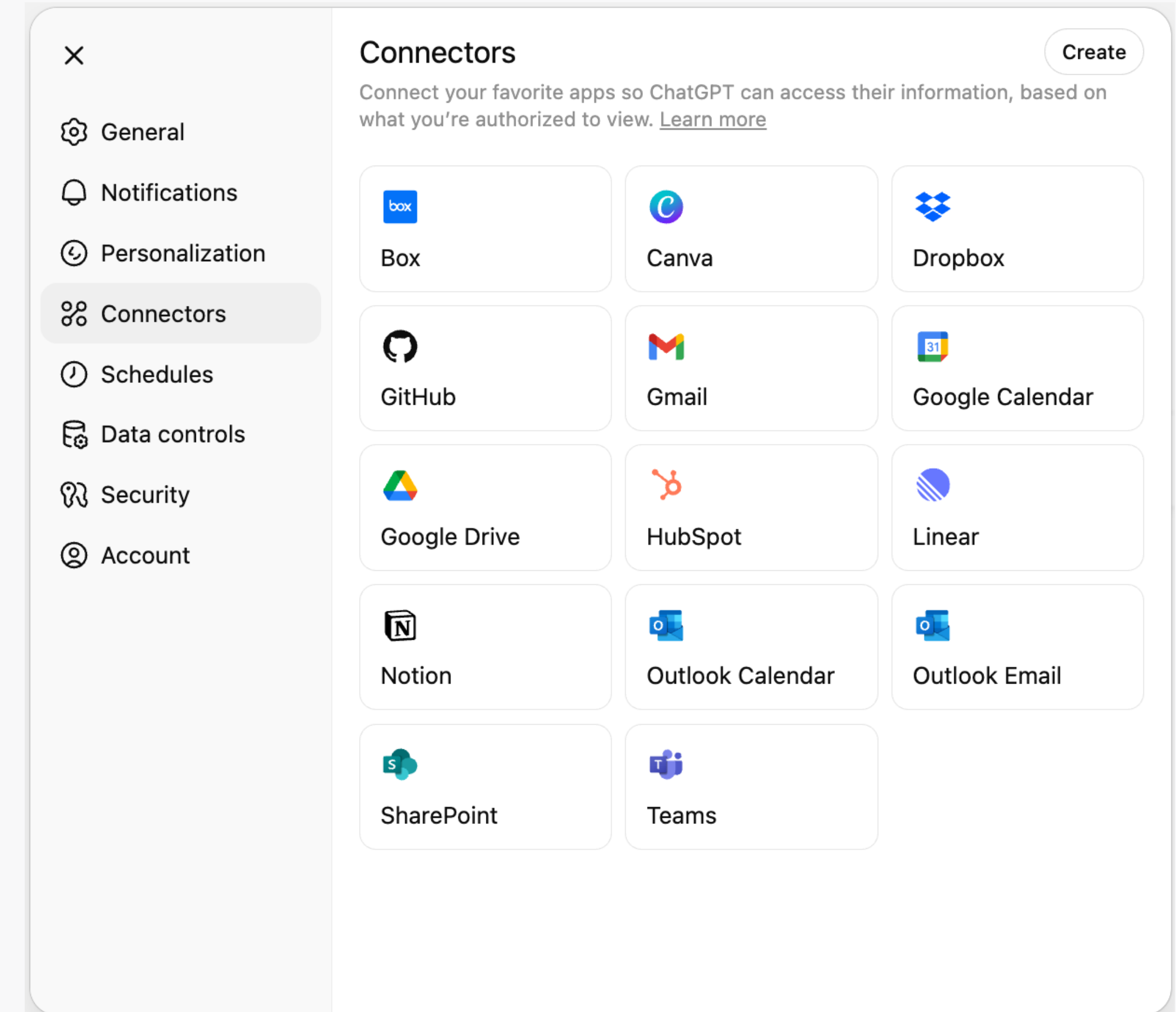
- Grammar (Lark)
- CFGs
- Conform
- Profit

```
const response = await client.responses.create({
  model: "gpt-5",
  input: "Use the code_exec tool to print hello world to the console.",
  tools: [
    {
      type: "custom",
      name: "code_exec",
      description: "Executes arbitrary Python code.",
    },
  ],
});

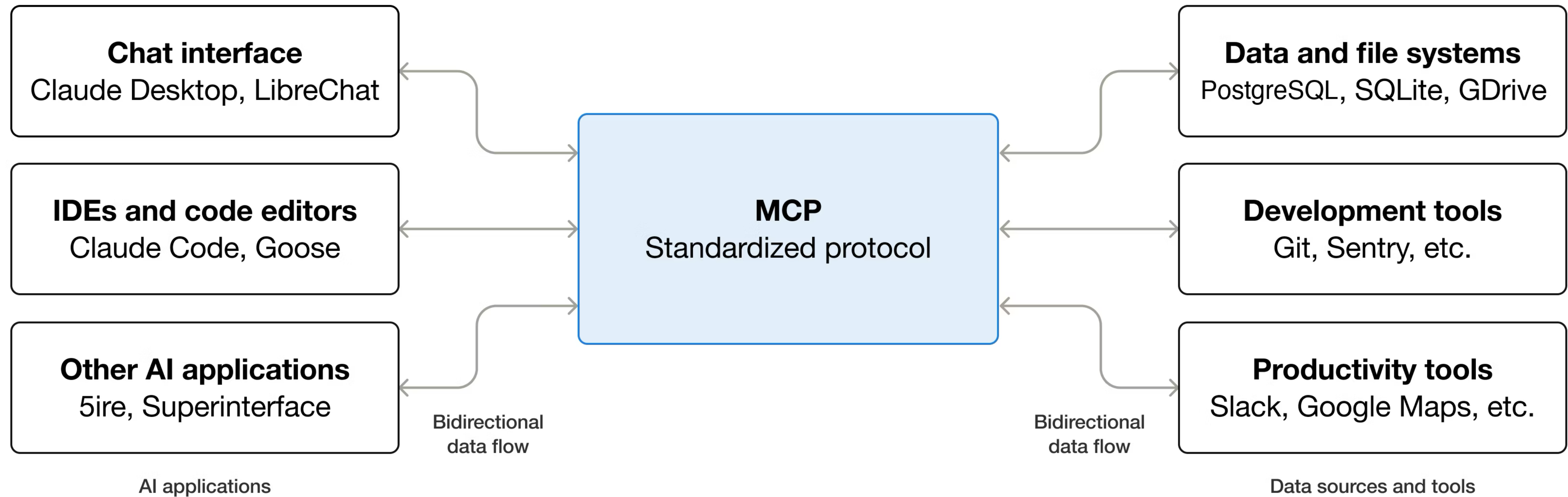
console.log(response.output);
```

Tool: Connectors

- OpenAI Specific
- Company Connections



Tool: MCP

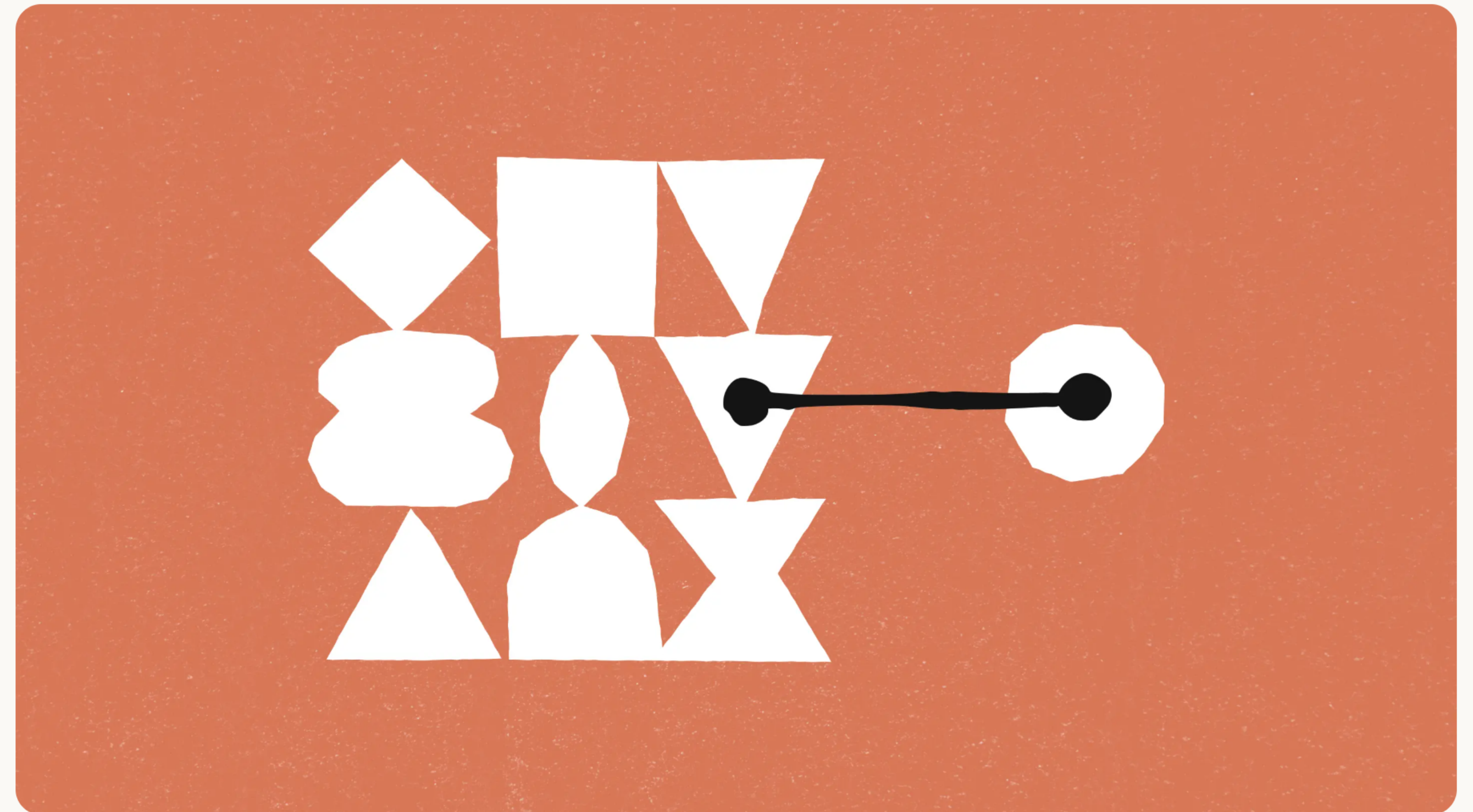




Model Context Protocol

Introducing the Model Context Protocol

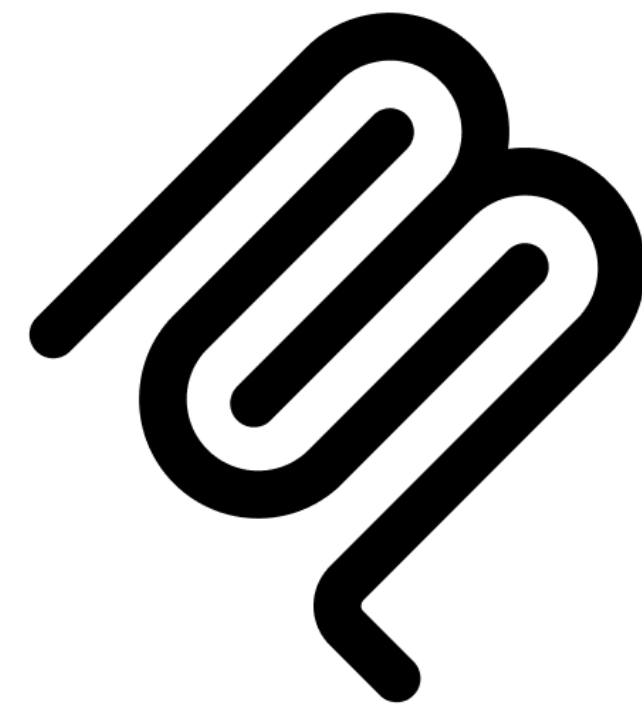
Nov 25, 2024 • 3 min read



Today, we're open-sourcing the Model Context Protocol (MCP), a new standard for connecting AI assistants to the systems where data lives, including content repositories, business tools, and development environments. Its aim is to help frontier models produce better, more relevant responses.

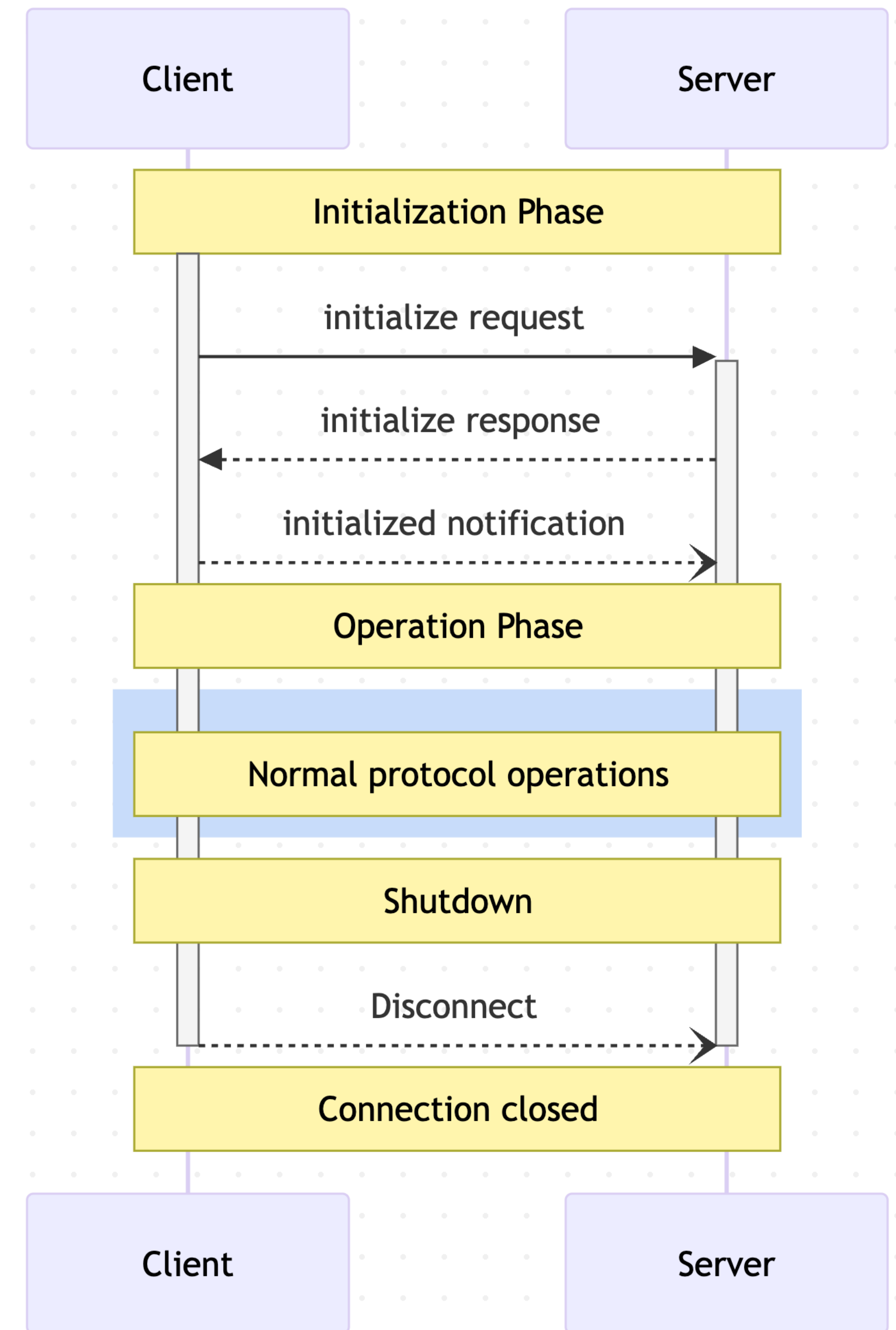
MCP Basics

- Open Source
- Connect LLM with External Thing
- The “USB-C” of AI applications



MCP Workflow

- Client (LLM)
- Server (Anything)
- Init — Operate



MCP Workflow - Stage 1

- Hello I am LLM
- I know MCP Version X
- I have these capabilities
- I am named Y

```
{
  "jsonrpc": "2.0",
  "id": 1,
  "method": "initialize",
  "params": {
    "protocolVersion": "2024-11-05",
    "capabilities": {
      "roots": {
        "listChanged": true
      },
      "sampling": {},
      "elicitation": {}
    },
    "clientInfo": {
      "name": "ExampleClient",
      "title": "Example Client Display Name",
      "version": "1.0.0"
    }
  }
}
```


MCP Workflow - Stage 2

- Hello I am MCP
- I speak MCP Version X
- I have these capabilities
- I am named Y

```
{
  "jsonrpc": "2.0",
  "id": 1,
  "result": {
    "protocolVersion": "2024-11-05",
    "capabilities": {
      "logging": {},
      "prompts": {
        "listChanged": true
      },
      "resources": {
        "subscribe": true,
        "listChanged": true
      },
      "tools": {
        "listChanged": true
      }
    },
    "serverInfo": {
      "name": "ExampleServer",
      "title": "Example Server Display Name",
      "version": "1.0.0"
    },
    "instructions": "Optional instructions for the client"
  }
}
```

MCP Workflow - Stage 3

- Yay - we speak the same.
- I'm ready to roll.

```
{  
  "jsonrpc": "2.0",  
  "method": "notifications/initialized"  
}
```


MCP Workflow - Stage 4

- What tools do you have?

```
{  
  "jsonrpc": "2.0",  
  "id": 1,  
  "method": "tools/list",  
  "params": {  
    "cursor": "optional-cursor-value"  
  }  
}
```

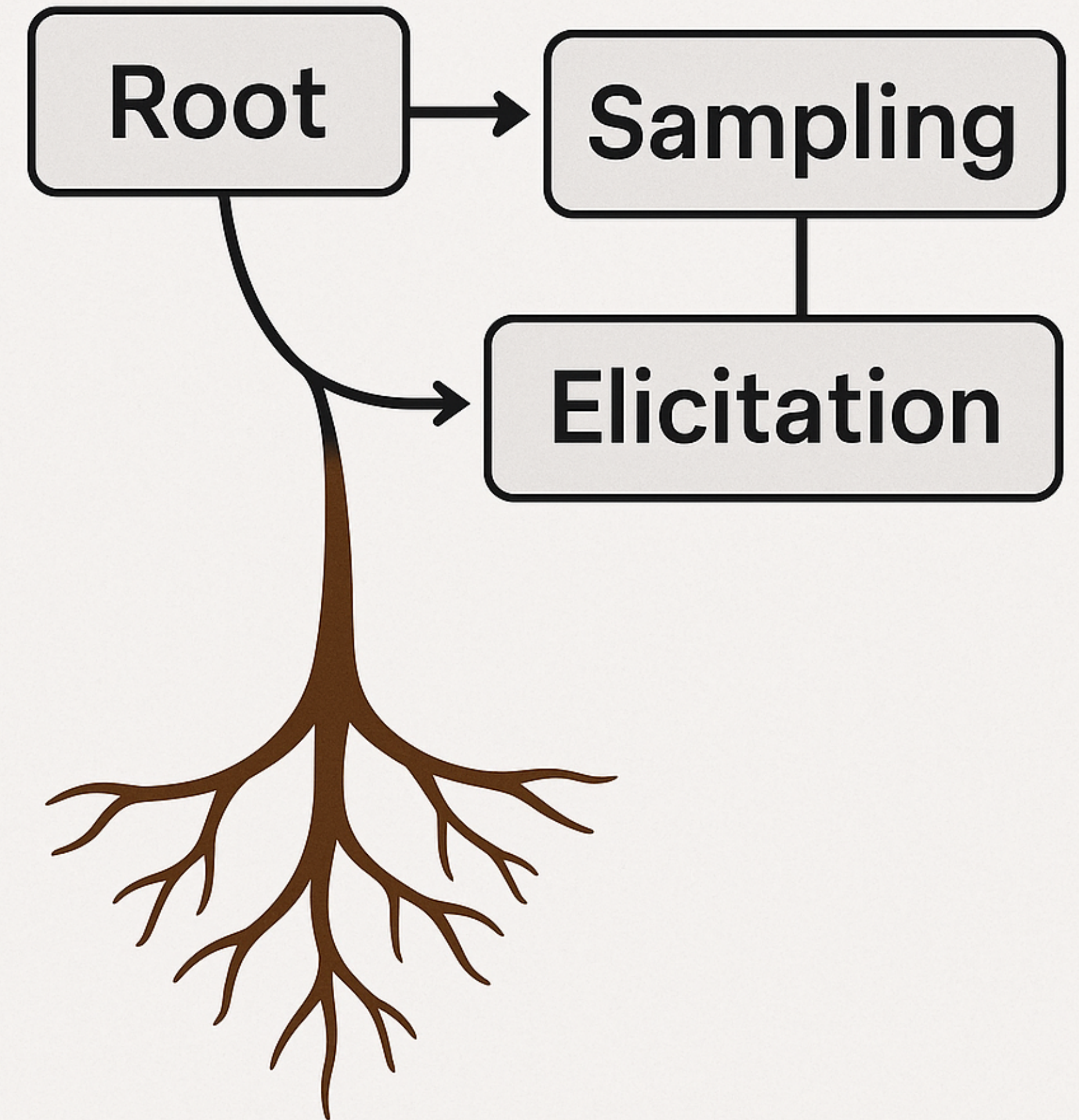
MCP Workflow - Stage 5

- I have many tools.

```
{
  "jsonrpc": "2.0",
  "id": 1,
  "result": {
    "tools": [
      {
        "name": "get_weather",
        "title": "Weather Information Provider",
        "description": "Get current weather information for a location",
        "inputSchema": {
          "type": "object",
          "properties": {
            "location": {
              "type": "string",
              "description": "City name or zip code"
            }
          },
          "required": [
            "location"
          ]
        }
      }
    ]
  },
  "nextCursor": "next-page-cursor"
}
```

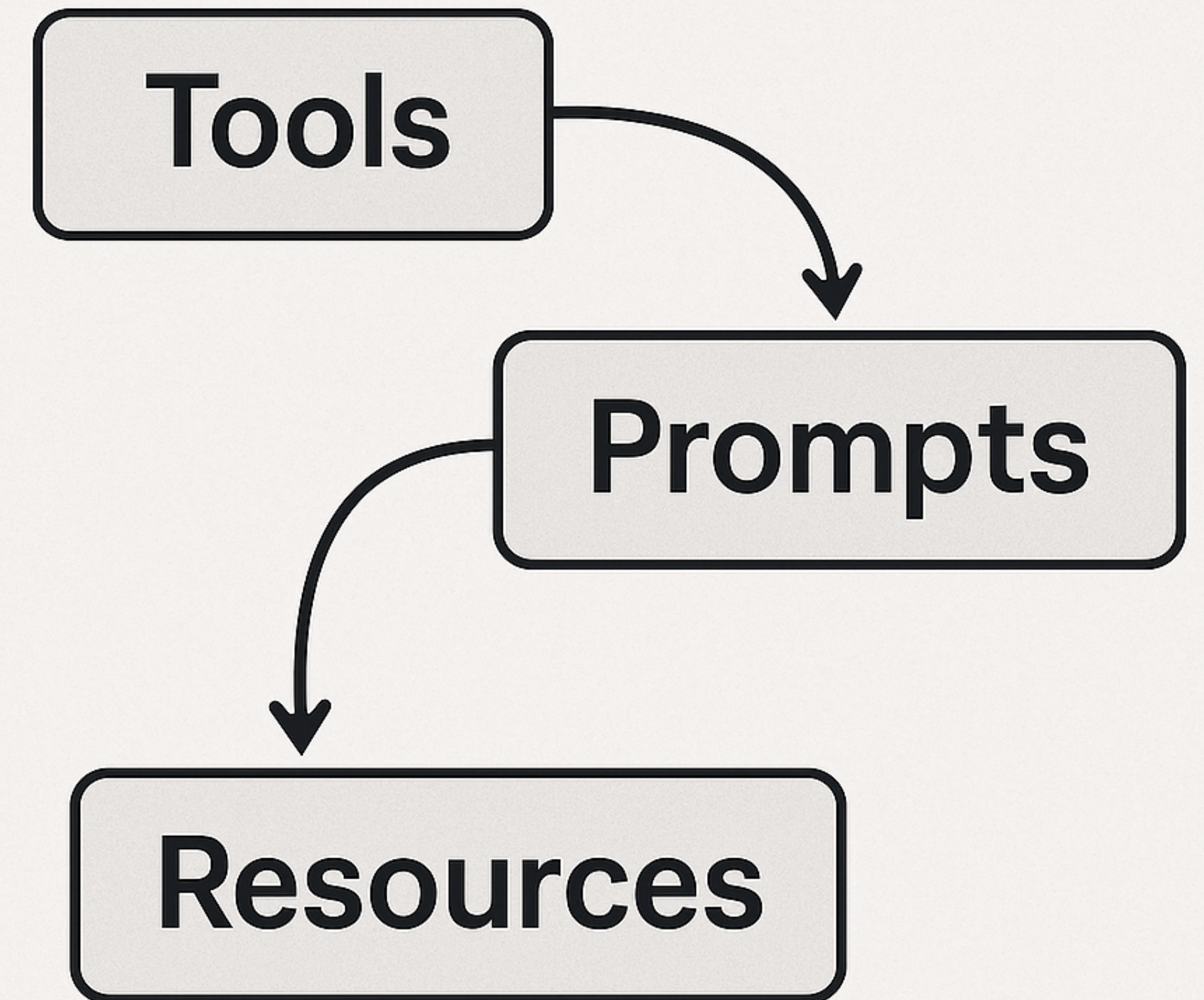

MCP Capabilities (Client)

- Roots
- Sampling
- Elicitation



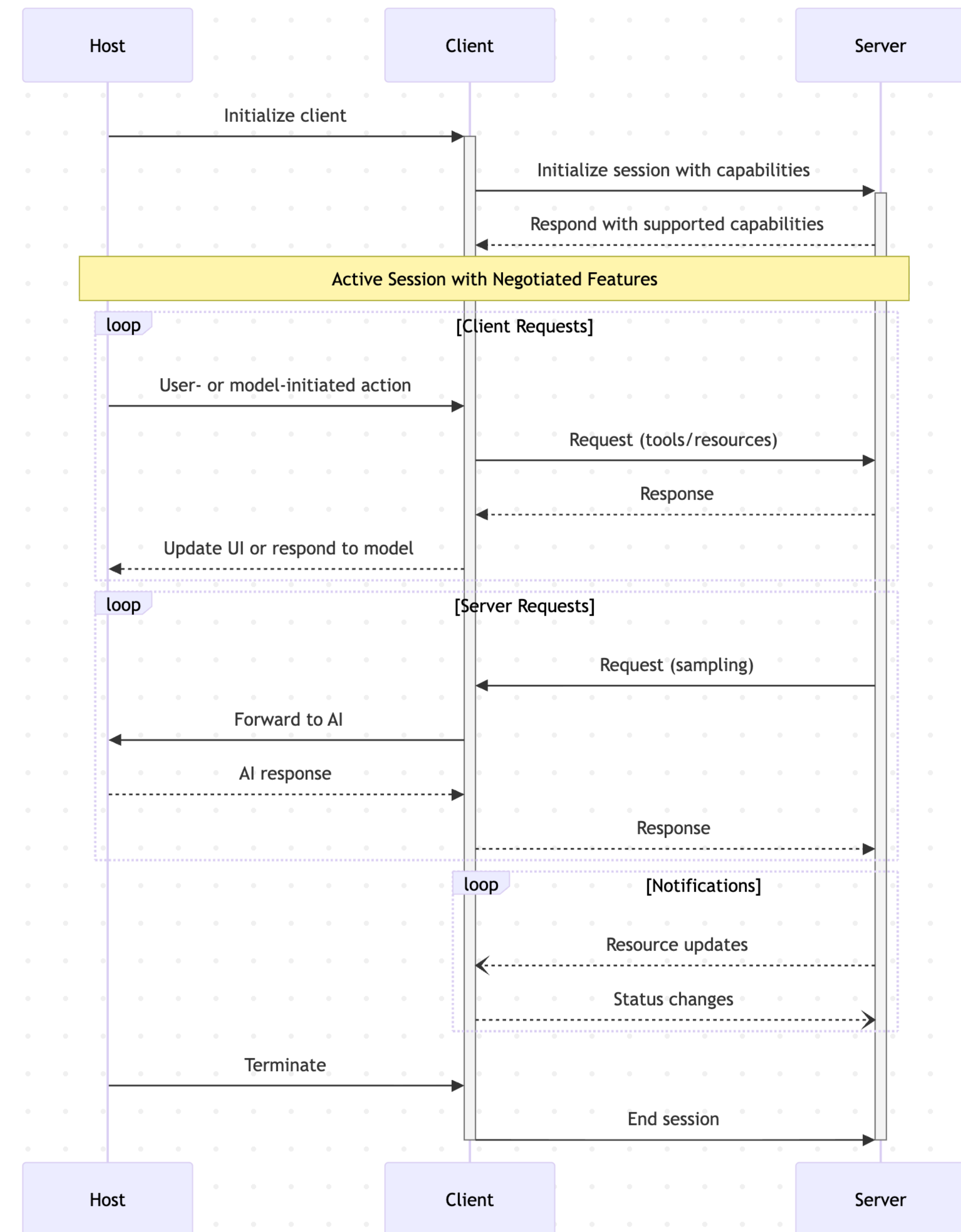
MCP Capabilities (Server)

- Prompts (user)
- Resources (app)
- Tools (model)



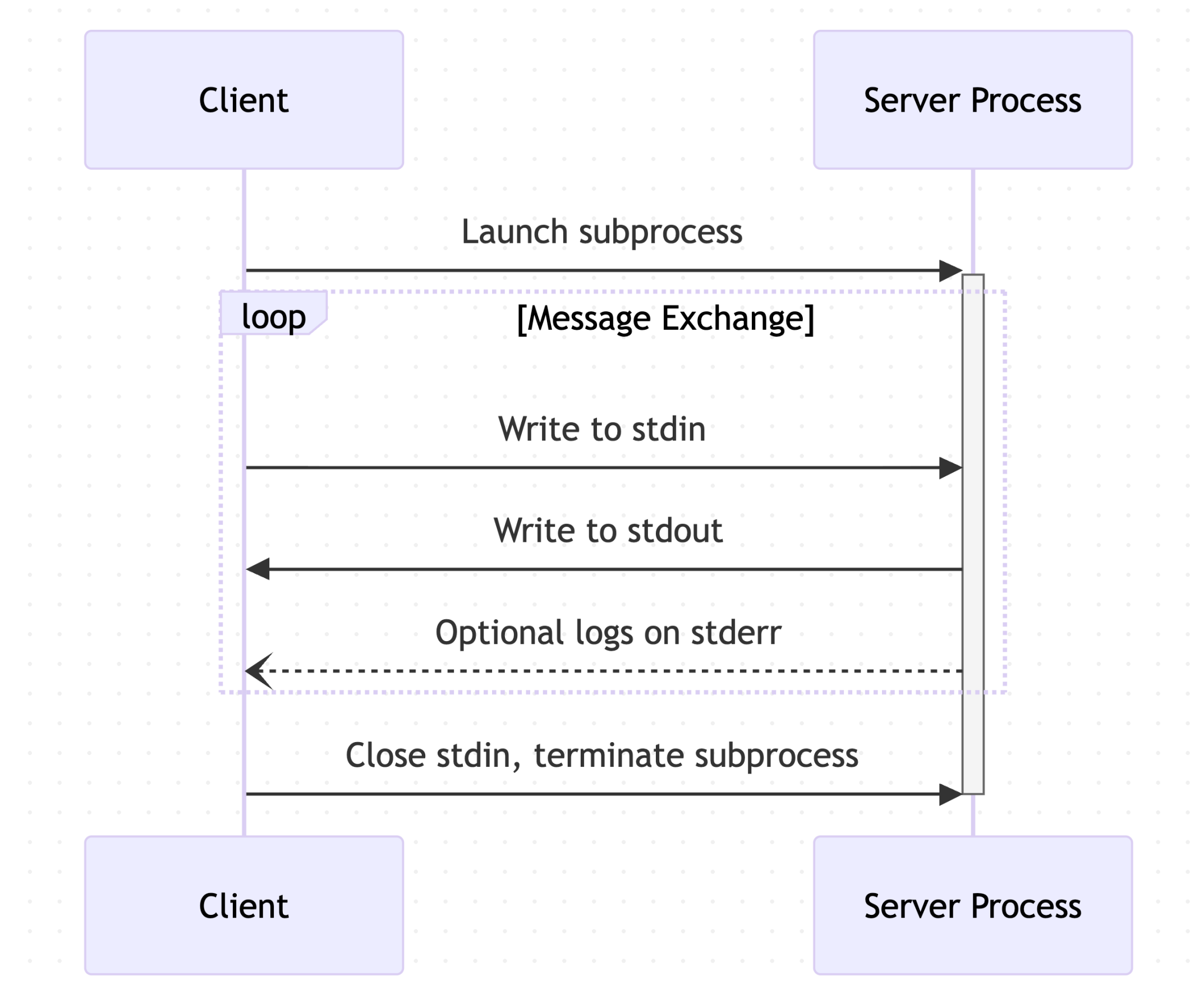
MCP Utilities

- Completion
- Logging
- Progress
- Ping
- Pagination



MCP Transports

- All JSON-RPC
- stdio
- Streamable HTTP
- application/json
- text/event-stream



MCP Security

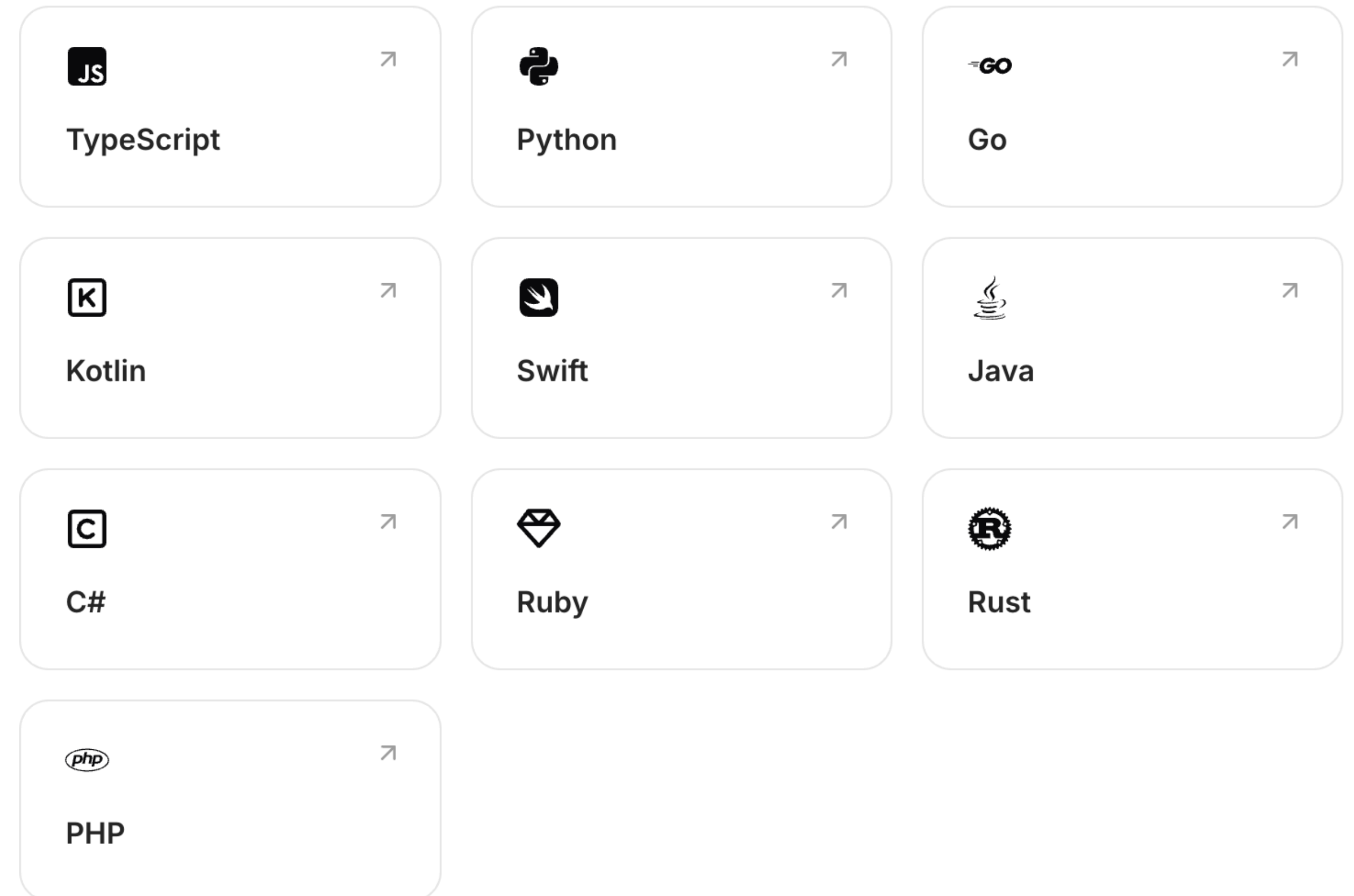
- OAuth 2.1
- Secret & Public Clients
- Dynamic Client Registration
- Protected Metadata



MCP Benefits

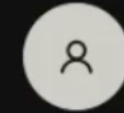
- Open
- Standardized
- Bidirectional
- Stateful

Available SDKs



Popular MCPs

- Playwright
- GitHub
- JIRA



Was a mobile app part of our VitaFleet MVP product spec

I'd be happy to help find that information about the VitaFleet MVP product spec. Let me search for that Confluence page and check whether a mobile was part of the plan.

I'll search for the VitaFleet MVP product spec in these workspaces. I'll start by searching for Confluence pages mentioning "VitaFleet" and "MVP".

searchConfluenceUsingCql



Great! I found the "VitaFleet MVP Product spec" page in the VitaFleet workspace. Let me retrieve the full content to check if a mobile app was part of the plan:

getConfluencepage



Concluding

- Tools are powerful
- MCP is powerful
- Each use case is different

Thanks!

connortumbleson.com

@iBotPeaches



source**to**ad