

# MCP - Model Context Protocol

1. WHY should I care about AI and MCP? (organizational and individual)
2. WHAT is MCP?
3. HOW to implement your first MCP server in JS
4. Challenges in using MCP
5. Where to go next & QA

# How we consume info changes ... again!



The next internet will be voice based and website-less.

The user just walks into his apartment after a hard day of work.

😊 Hey ChatGPT! What matches are on TV this evening?

🤖 There is Real Madrid vs Barcelona and Man United vs Chelsea.

😊 Great! What is the offering for Real Madrid winning?

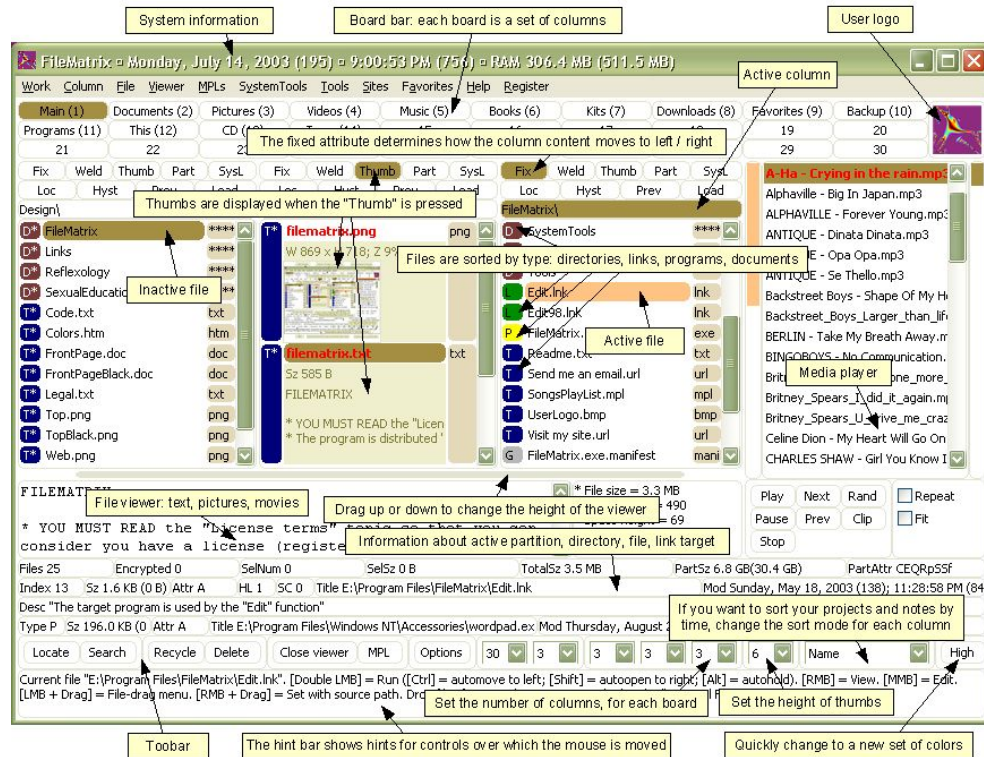
🤖 It's 1.2.

😊 Fantastic! Place a bet for \$100 on Real Madrid.

🤖 Sure, done.

*No clumsy UI & voice-based authentication.*

I just want to change my email address ....

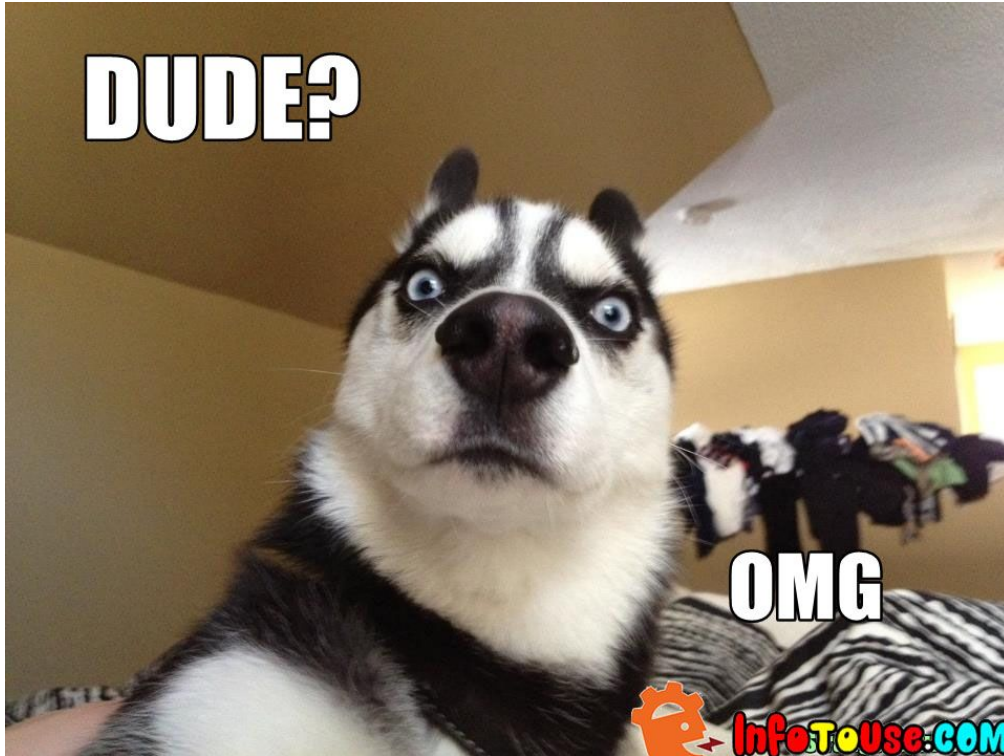


The human speech is our default user interface to send instructions.

# Why should I care? as a developer

**ML !== AI Engineering**

# Me discovering Chat GPT



- This is not if-then-else!
- How does it work?

## TRAINING PARAMETERS

Model Type:

MyNet ▾

# of training epochs:

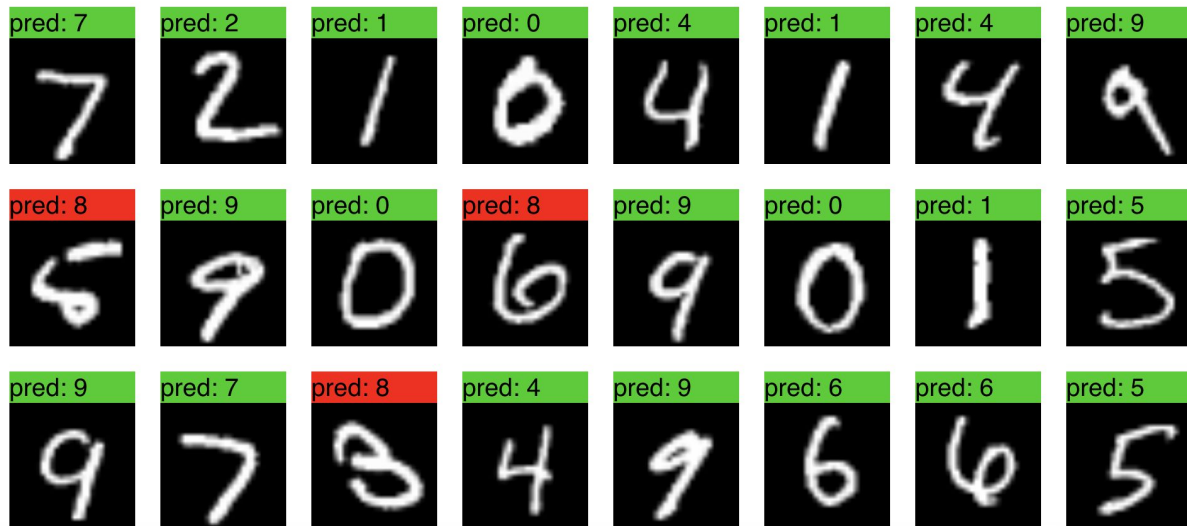
3

Load Data and Train Model

## TRAINING PROGRESS

Final test accuracy: 85.4%

## INFERENCE EXAMPLES



8 mo later

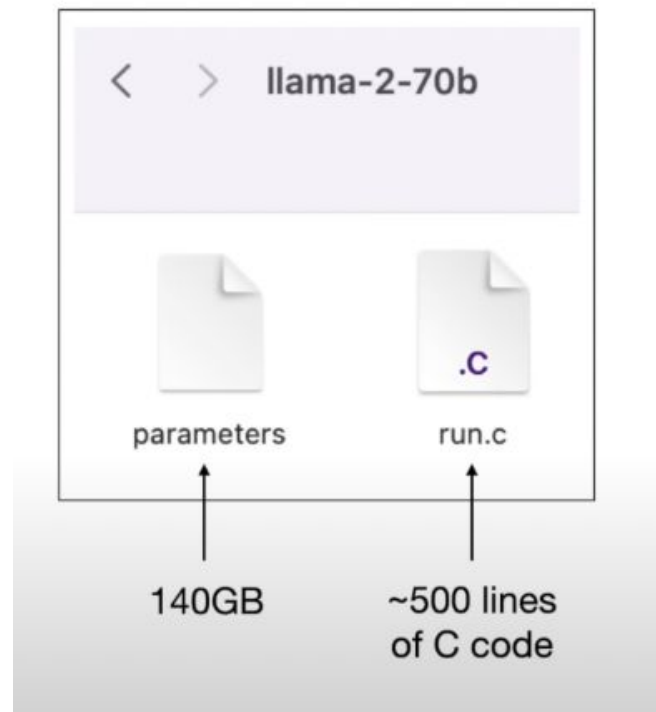


# Fundamental shift

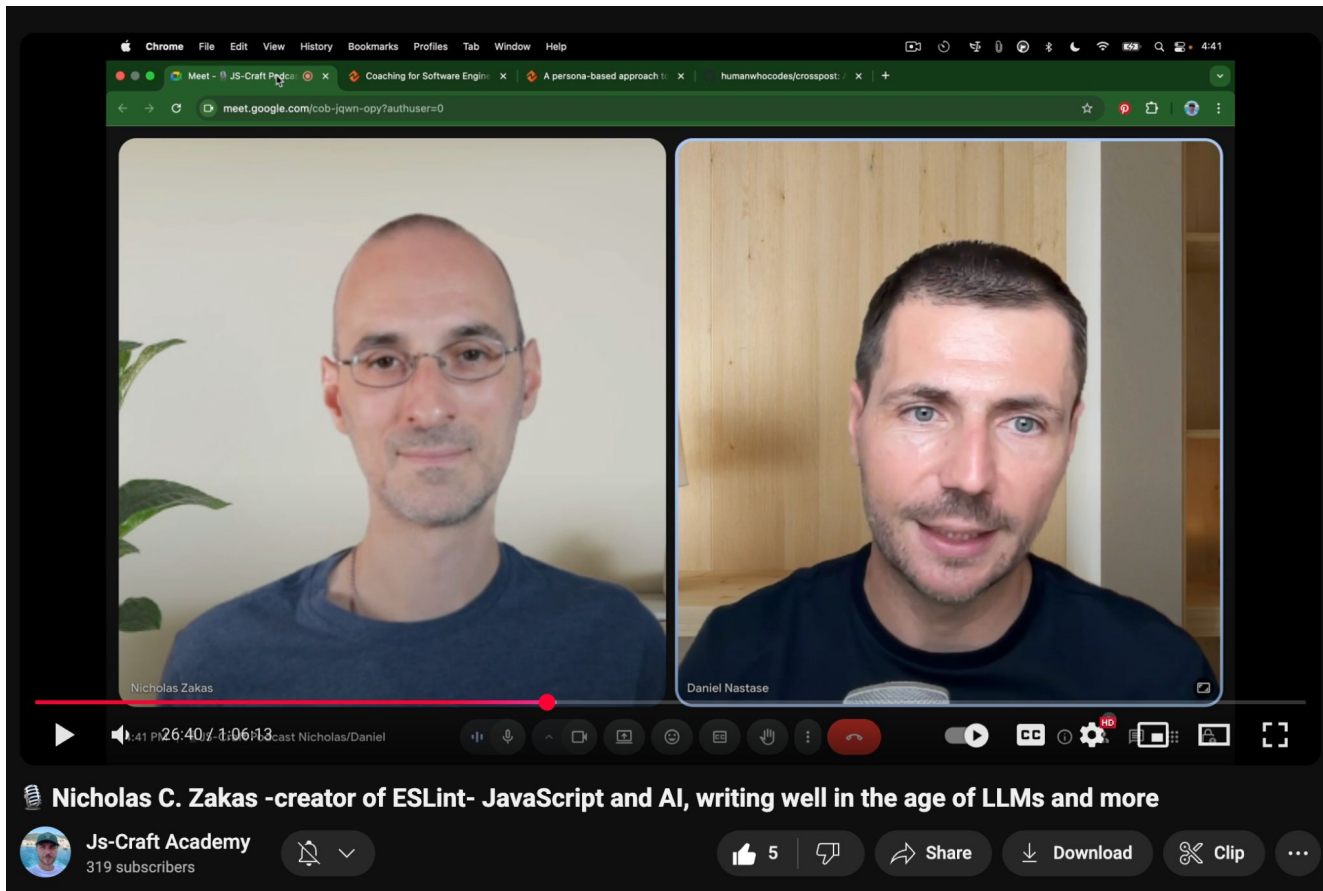
Algorithms, Code, and Intelligence

VS

Data and pure raw calculation power







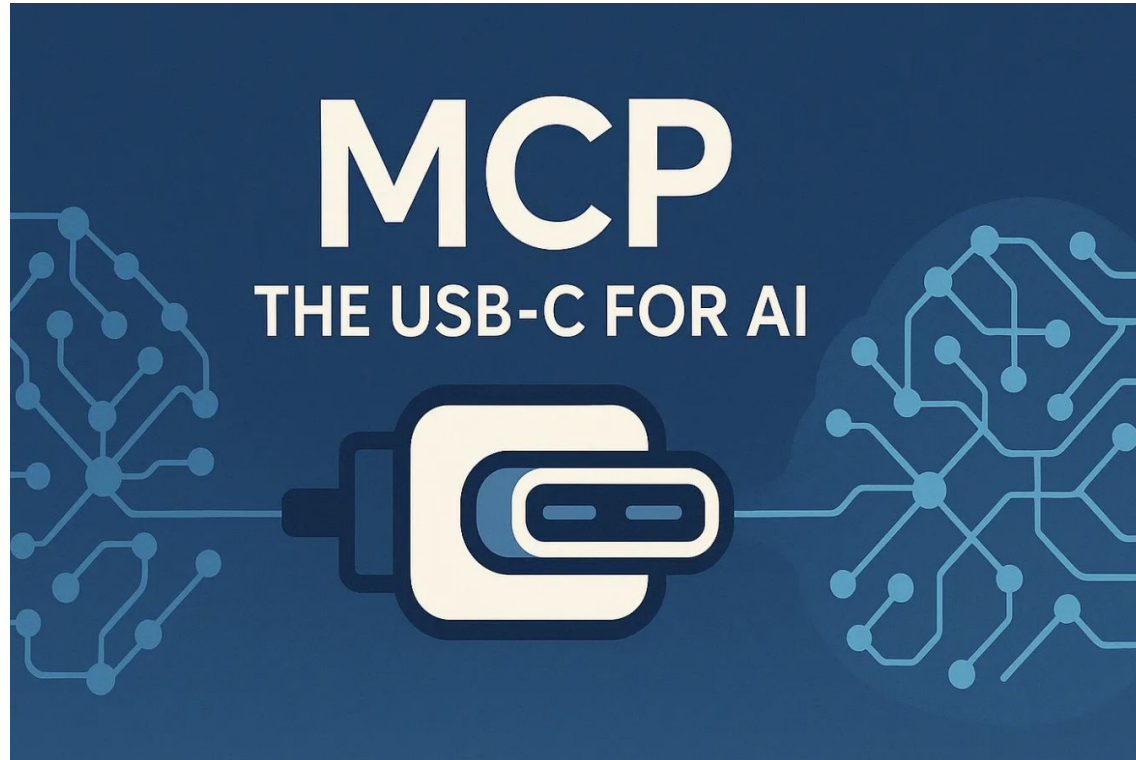
**I don't care - as much - how AI works!**

# AI Engineering tasks

- Managing Long-Term and Short-Term Memory
- Data Streaming
- Building AI agents, architectures, and flows.
- Model Selection, Prompt engineering, Evals performance
- Context Engineering
- Human-in-the-loop
- Model Serving Infrastructure
- Observability & Monitoring, Cost Optimization
- **& MCP**

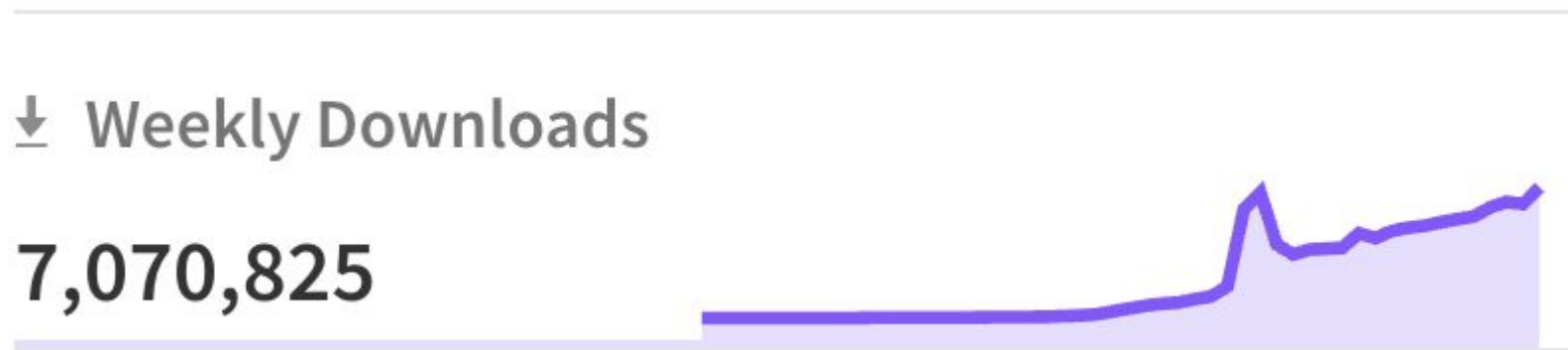


Latent Space




... actually, it's much more about providing context!

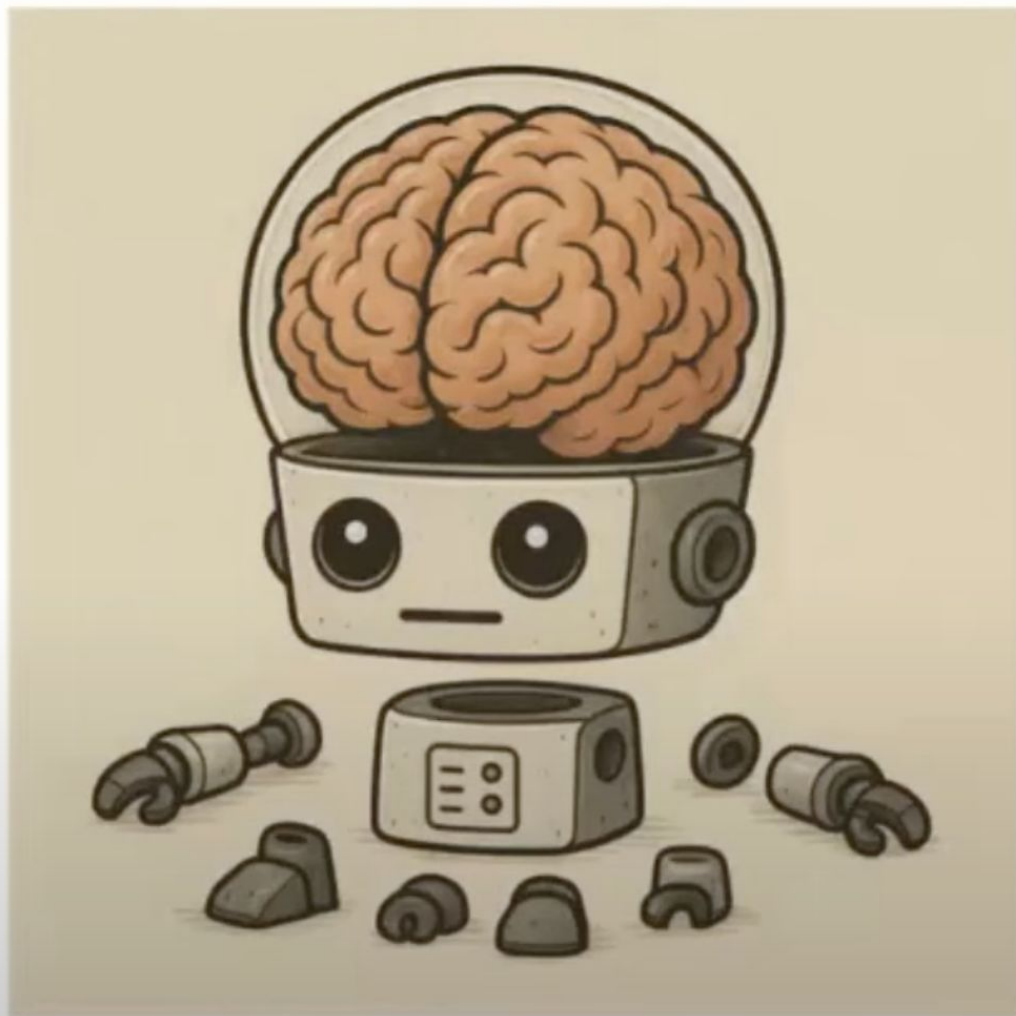
# Search for the evergreen tech



*MPC stats on NPM*

*... it took [Vue.js](#) 11 years to get to the same values*

Let's see a “pure”  
LLM in action  
... meet Llama 



**LLMs + Tools = ❤️**

# MCP exposes context via:



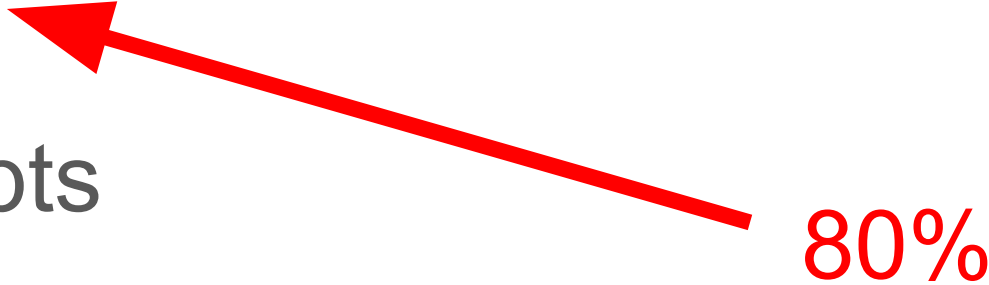
Resources



Tools



Prompts





# Anatomy of a LLM tool - body

```
server.setRequestHandler(CallToolRequestSchema, async request => {  
  if (request.params.name === "get_matches") {  
    const { weekday } = request.params.arguments  
    return {  
      content: [{  
        type: 'text',  
        text: `On ${weekday}, there is Real Madrid vs Barcelona at 9pm.`,  
      }],  
    }  
  }  
  throw new McpError(ErrorCodes.ToolNotFound, "Tool not found")  
})
```



plain JavaScript  
code

# The anatomy of a LLM tool - schema

```
server.setRequestHandler(ListToolsRequestSchema, async () => {  
  return {  
    tools: [{  
      name: "get_matches",  
      description: "Returns the matches for a given weekday",  
      inputSchema: {  
        type: "object",  
        properties: {  
          weekday: {  
            type: 'string',  
            description: 'the weekday to check the matches for',  
          }  
        },  
        required: ["weekday"]  
      },  
    }  
  ],  
})
```

word instructions for the LLM,  
parsed via NLP:

1. when to call the function
2. parameters to extract



**Andrej Karpathy** ✓

@karpathy

The hottest new programming language is English

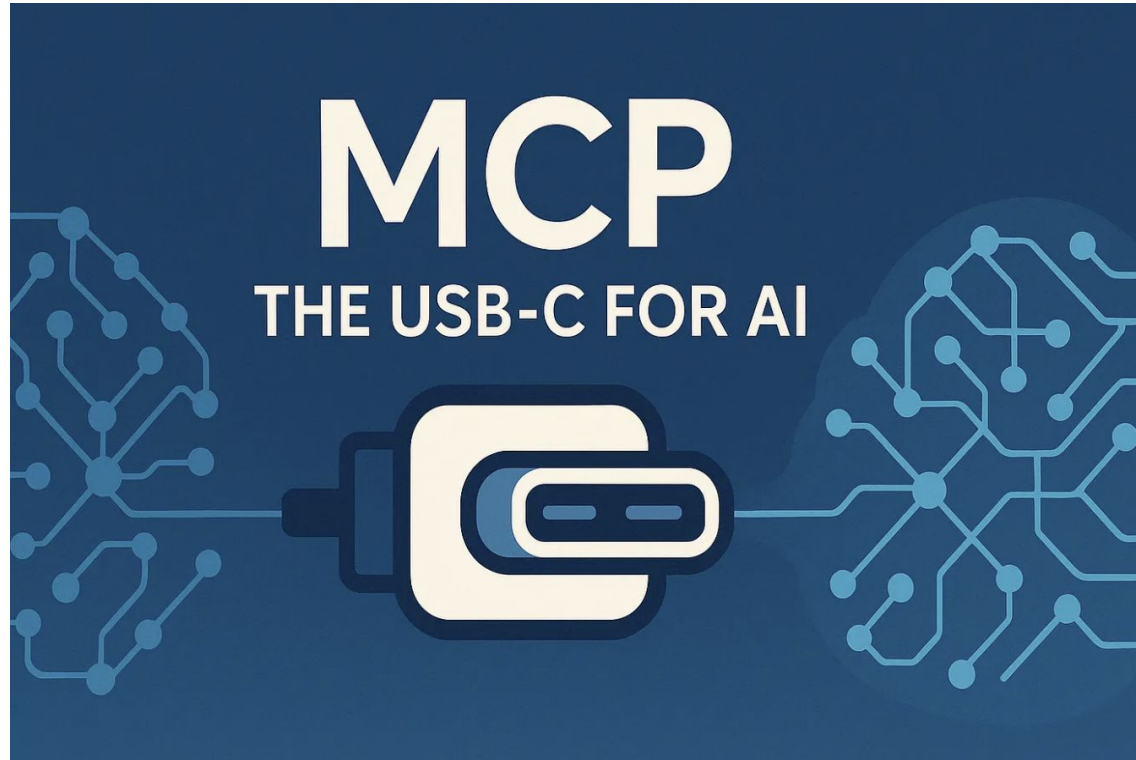
12:14 PM · Jan 24, 2023 · **3.8M** Views

Github repos are becoming code + natural language

DEMO TIME

---





... actually, it's much closer to HTTP or the REST API

The screenshot shows the npm package page for `@modelcontextprotocol/sdk`. The package is version 1.18.0, published 4 days ago, and is public. It has 12 dependencies, 12258 dependents, and 52 versions. The package is licensed under MIT. The page includes a 'Table of Contents' with links to Overview, Installation, Quickstart, What is MCP?, and Core Concepts (Server, Resources, Tools, Prompts). The 'Install' section shows the command `npm i @modelcontextprotocol/sdk`. The 'Repository' section links to the GitHub repository `github.com/modelcontextprotocol/typ...`. The 'Homepage' section links to `modelcontextprotocol.io`. The 'Weekly Downloads' section shows 7,070,825 downloads.


MCP

The screenshot shows the FastMCP documentation page. The page is titled 'Welcome to FastMCP 2.0!' and describes it as 'The fast, Pythonic way to build MCP servers and clients.' It explains that the Model Context Protocol (MCP) is a new, standardized way to provide context and tools to your LLMs, and that FastMCP makes building MCP servers and clients simple and intuitive. The page includes a 'Get Started' section with links to Installation, Quickstart, and a 'Welcome!' message. The 'Servers' section includes links to Overview, Core Components, Advanced Features, Authentication, and Deployment. The 'Clients' section includes links to Essentials, Core Operations, Advanced Features, and Authentication. A code snippet is shown in a dark-themed editor, demonstrating how to use FastMCP to create a simple MCP server that adds two numbers.

FAST-MCP

# Challenges

Microsoft Build



Microsoft

```
mcp-server-neon/src/tools.ts
2077 Lines (1882 loc) · 61.5 KB
Code Blame Row

67 {
68   name: 'run_sql' as const,
69   description: '
70   <use_case>
71   Use this tool to execute a single SQL statement against a Neon database.
72   </use_case>
73
74   <important_notes>
75   If you have a temporary branch from a prior step, you MUST:
76   1. Pass the branch ID to this tool unless explicitly told otherwise
77   2. Tell the user that you are using the temporary branch with ID [branch_id]
78   </important_notes>
79   ',
80   inputSchema: runSqlInputSchema,
81 },
82
83 {
84   name: 'run_sql_transaction' as const,
85   description: '
86   <use_case>
87   Use this tool to execute a SQL transaction against a Neon database, should be used for multiple
88   </use_case>
89
90   <important_notes>
91   If you have a temporary branch from a prior step, you MUST:
92   1. Pass the branch ID to this tool unless explicitly told otherwise
93   2. Tell the user that you are using the temporary branch with ID [branch_id]
94   </important_notes>
95   ',
96   inputSchema: runSqlTransactionInputSchema,
97 },
```

8:13 / 14:43

CC Settings Full Screen

Your API is not an MCP | DEMFP786

**What's next?**





# Timeless knowledge:

1. Demystify the Magical Black Box of LLMs (and better understand the research and be able to call BS )
2. MCP (see NMP Graph)



# Timeless knowledge:

3. AI Agents (Tools, memory management, do-while loop / cyclical structure architecture)

4. Evals (UT for AI models + how good the prompts are doing. They have their own challenges eq: what is a A good summary.)

AI Engineer

World's Fair

FULL WORKSHOP

Input

Tokenization (BPE)

Text & Position Embeddings

HOW LLMS WORK FOR WEB DEVS:

**GPT in 600 lines  
of Vanilla JS**

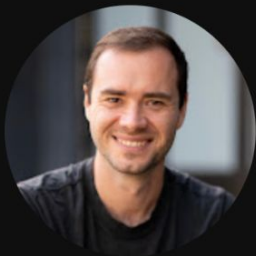
code\_for\_matrix\_add  
(code)



Andriy Burkov

# THE HUNDRED-PAGE MACHINE LEARNING BOOK





# Andrej Karpathy

@AndrejKarpathy · 1.04M subscribers · 17 videos

More about this channel ...more

[x.com/karpathy](https://x.com/karpathy) and 3 more links

Subscribe

Home

Videos

Playlists

Posts



Latest

Popular

Oldest



2:11:12

How I use LLMs

1.9M views · 6 months ago



3:31:24

Deep Dive into LLMs like ChatGPT

3.5M views · 7 months ago



4:01:26

Let's reproduce GPT-2 (124M)

893K views · 1 year ago



2:13:35

Let's build the GPT Tokenizer

902K views · 1 year ago

O'REILLY®

# AI Engineering

Building Applications with Foundation Models



Chip Huyen