

A Look Into Applied GenAI

Bernhard Krüpl-Sypien
bernhard@became.ai

22 October 2025

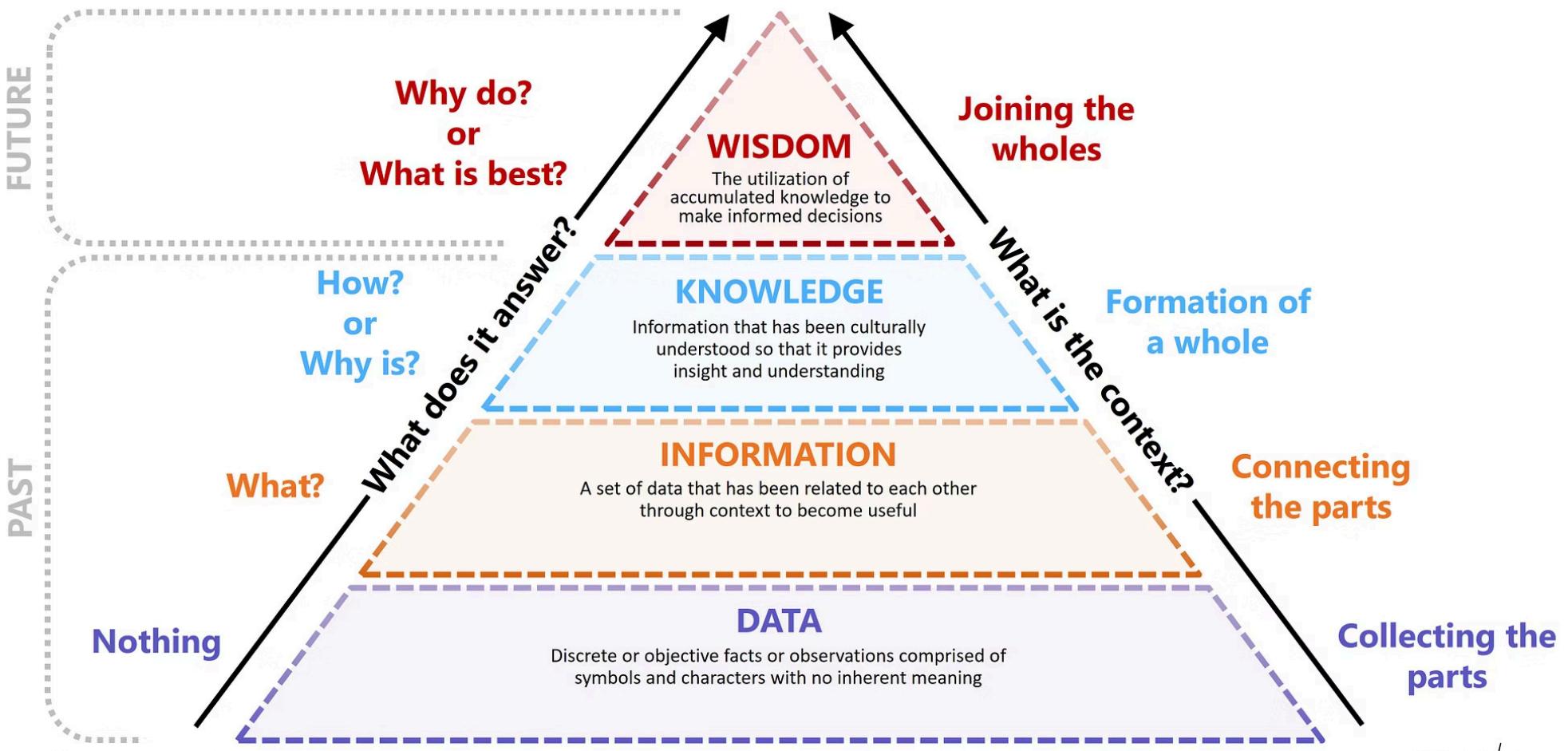
Goals for today

- Present group projects
- Provide pointers to tools that can help you to work on the projects
- Show that it is almost trivial to interact with LLMs programatically
- Make the LLM ecosystem more accessible to you
- Provide a birds-eye overview of the (generative) AI space
- Introduce some concepts around AI

Group Projects Motivation –

What Are You Doing Right Now?

DIKW Pyramid



Source: Adaptation from DIKW pyramid R. Ackoff, 1989

Source: jeffwinterinsights.com / Russell Ackoff 1989

PERSPECTIVE

A Very Brief History Of Artificial Intelligence

Brief History of Artificial Intelligence

1 1950s-1960s

Birth of AI: Logic, problem solving and early neural networks

2 1970s-1980s

Expert systems and knowledge based approaches

3 1990s-2000s

Machine learning taking off, data driven AI

4 2010s-today

Deep learning revolution, large language models





The Birth Of Artificial Intelligence



Dartmouth Workshop 1956

Starting AI research supported by public funding

Participants

John McCarthy, Marvin Minsky, Claude Shannon, Nathaniel Rochester et al.



Term "Artificial Intelligence"

Coined during the conference – probably as a marketing gag

Goal

Simulation of human intelligence with machines

Source: <https://caimi.org/machines2023/program/thomas-haigh-slides.pdf>
(Thomas Haigh 2023)



Hype and disillusionment, repeat



High expectations

AI research starts with great optimism

Bold predictions

Marvin Minsky (1967): AI problem will be solved within a generation
Herbert Simon (1965):
Machines will be capable, within twenty years, of doing any work that a man can do

Disillusionment

"AI was harder than we thought" (John McCarthy)

AI winter

Long period of stagnation during two AI winters:
1974–1980 (perceptrons),
1987–2000 (LISP/expert systems)

The Two Main Strands of Artificial Intelligence

Symbolic AI

Logic based approach for Problem solving

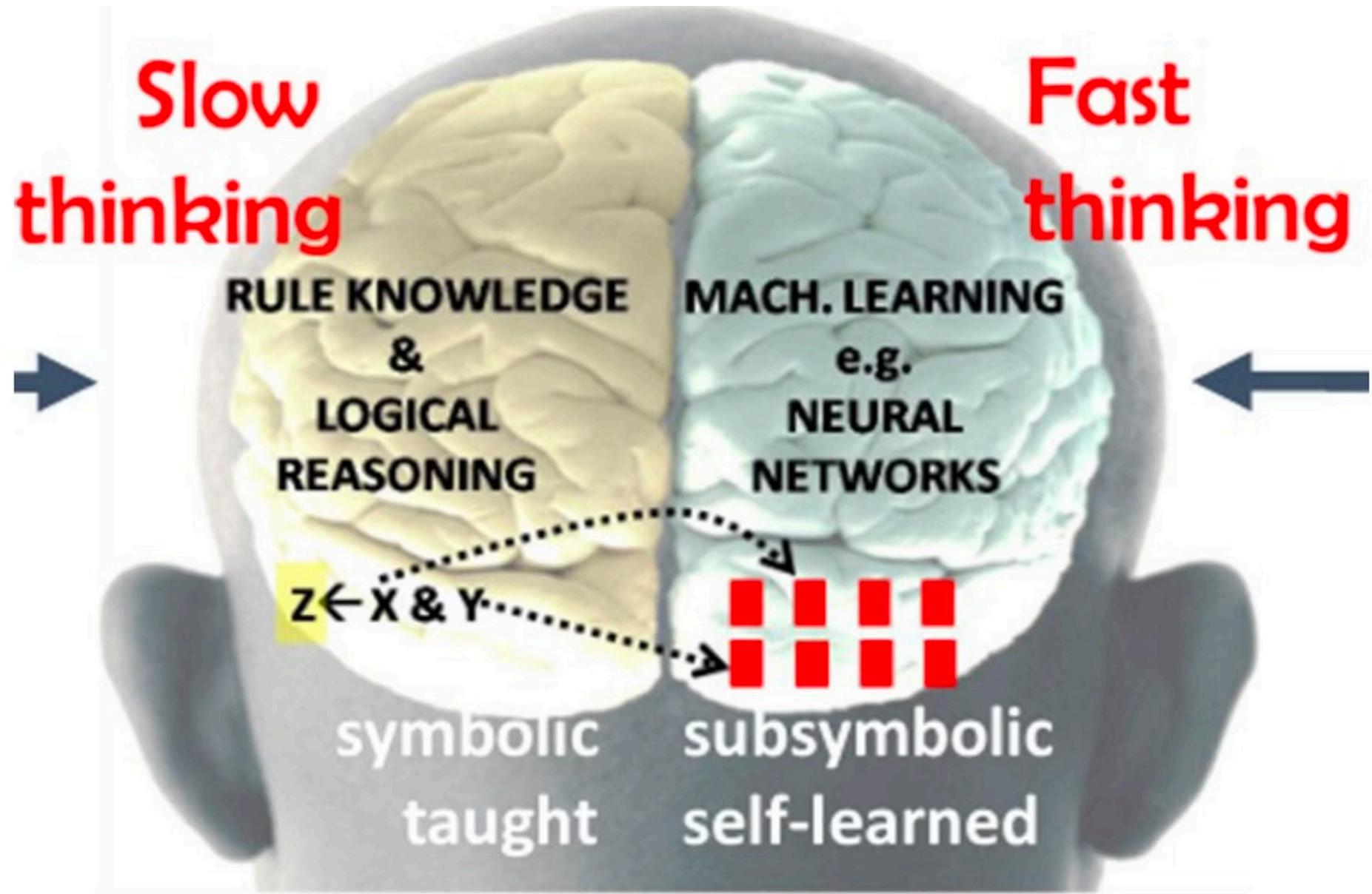
Aims to understand problems

Subsymbolic AI

Data based approach for pattern recognition

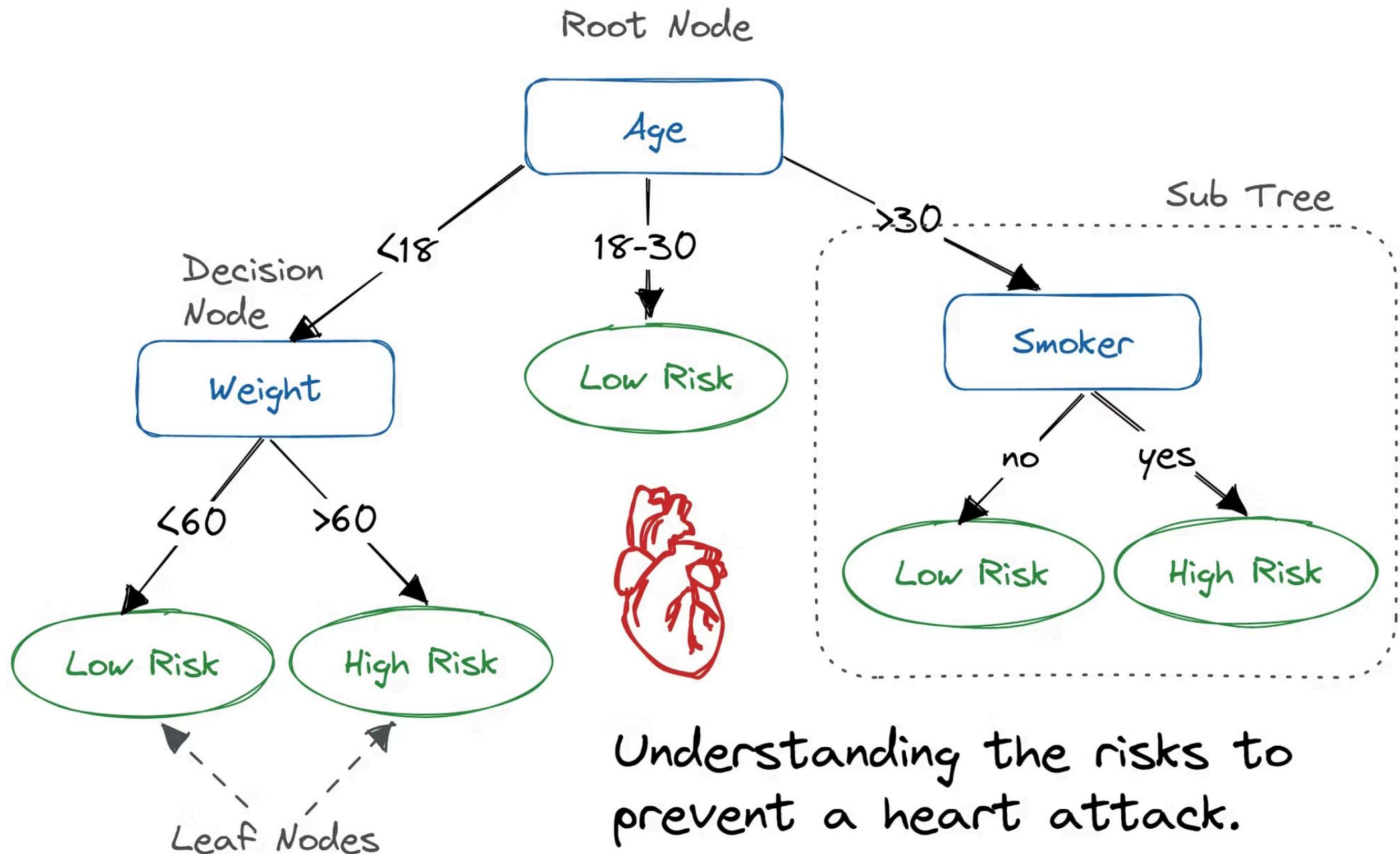
Aims to learn statistical relationships

System 1 and System 2 (Kahneman)



Source: Kahneman, D. (2011). Thinking, Fast and Slow

Explicit Models



Vs Giant Matrices



Two Great Inventions



Who are these two people and what did they invent?

Two Great Inventions

The World Wide Web

Tim Berners-Lee invented the **World Wide Web** in 1989 while working at CERN. He developed the three fundamental technologies that underpin the Web:

- **HTML** (HyperText Markup Language): The formatting language for web pages.
- **URI** (Uniform Resource Identifier): The address system for web resources.
- **HTTP** (Hypertext Transfer Protocol): The communication protocol for data transfer.

The Internet

Vinton Cerf, jointly with Robert Kahn, co-designed the **Transmission Control Protocol** and **Internet Protocol (TCP/IP)** in the early 1970s.

These protocols establish fundamental rules for packet-switched communication across heterogeneous networks, ensuring reliable data transmission and interoperability, thereby constituting the architectural foundation of the modern Internet.

Data As The New Oil



Common Crawl



Common Crawl August 2025 Crawl Archive (CC-MAIN-2025-33)

The August 2025 crawl archive contains 2.44 billion pages, see the [announcement](#) for details.

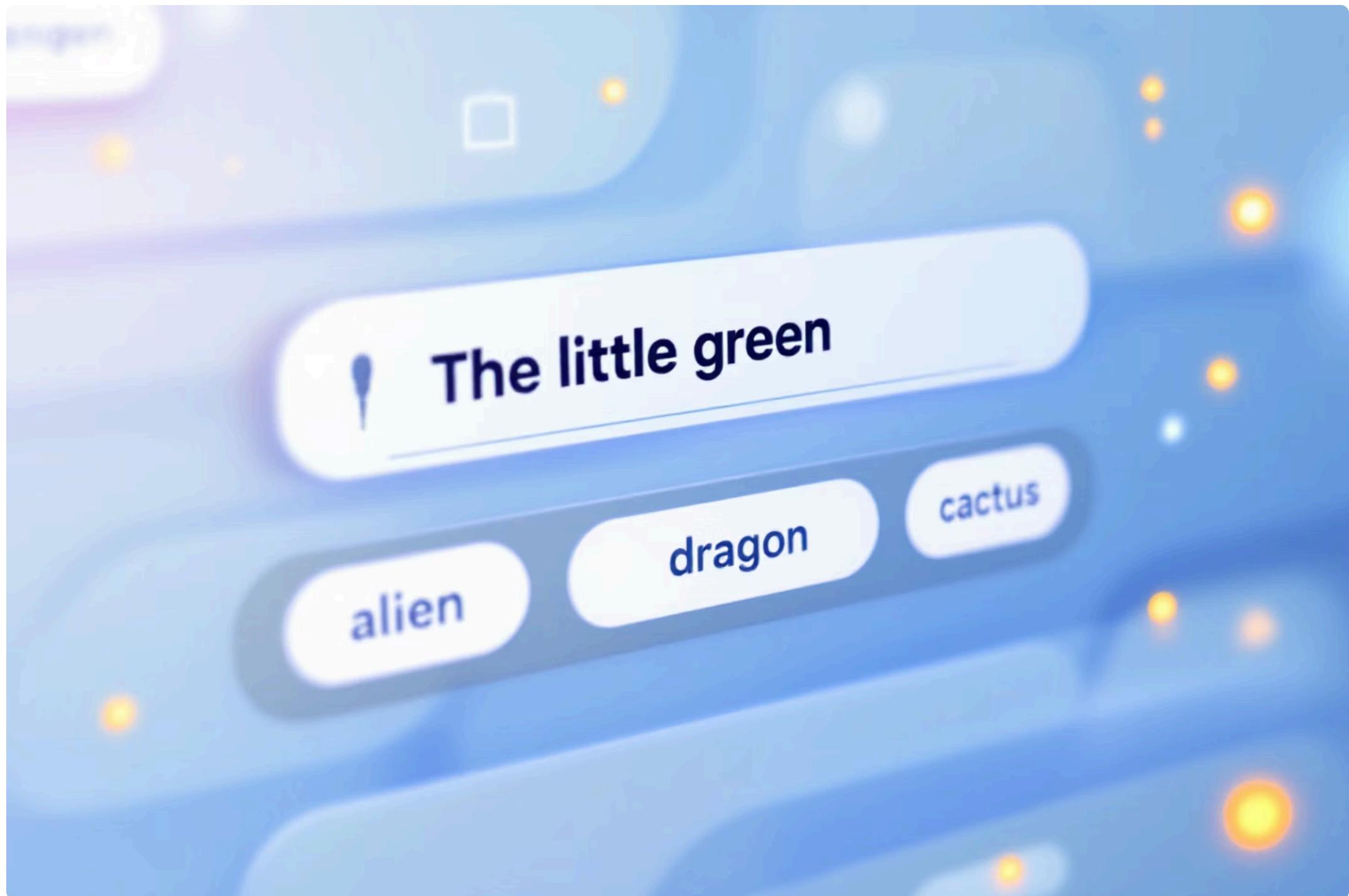
Data Size and File Listings

Data Type	File List	#Files	Total Size Compressed (TiB)
Segments	segment.paths.gz	100	
WARC	warc.paths.gz	100000	88.24
WAT	wat.paths.gz	100000	16.71
WET	wet.paths.gz	100000	6.63
Robots.txt files	robotstxt.paths.gz	100000	0.15
Non-200 responses	non200responses.paths.gz	100000	2.97
URL index files	cc-index.paths.gz	302	0.19
Columnar URL index files	cc-index-table.paths.gz	900	0.21

[Common Crawl](#) is a California 501(c)(3) registered non-profit organization. Hosting of [Common Crawl data](#) is covered by [Amazon Web Services' Open Data Sponsorship Program](#).

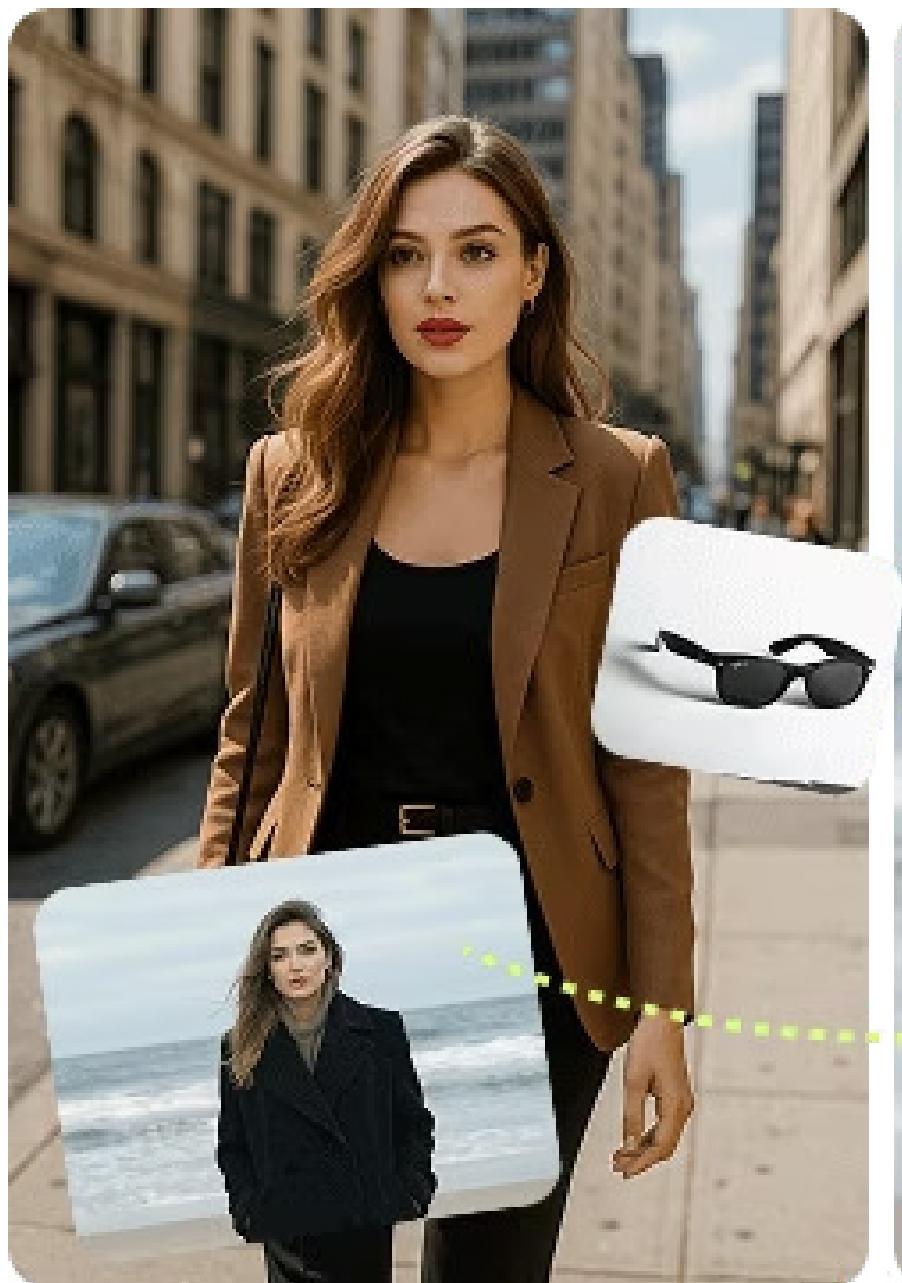
[Terms of Use](#) [Privacy](#)

All Trained For Autoregression / Autocomplete



Promethean Moment in Artificial Intelligence





Source: gemini.google.com

Popular GenAI Frontier Model Providers



Open AI



Anthropic



Google



Meta



xAI



Alibaba



DeepSeek



Mistral

Foundation and Frontier Models

These terms describe the scale and impact of AI models. **Foundation models** are large models, trained on broad data, designed to be adapted to a wide range of downstream tasks. **Frontier models** represent the most advanced and powerful foundation models, pushing the boundaries of AI capabilities at any given time.

AI Model Taxonomy: Openness

1 Closed-Source Models

Proprietary models where weights, inference code, and training data remain confidential. Access is typically via API, and internal workings are not disclosed.

2 Open-Weight Models

Model weights and inference code are released publicly, allowing for local execution and fine-tuning. However, training data and methodology often remain proprietary.

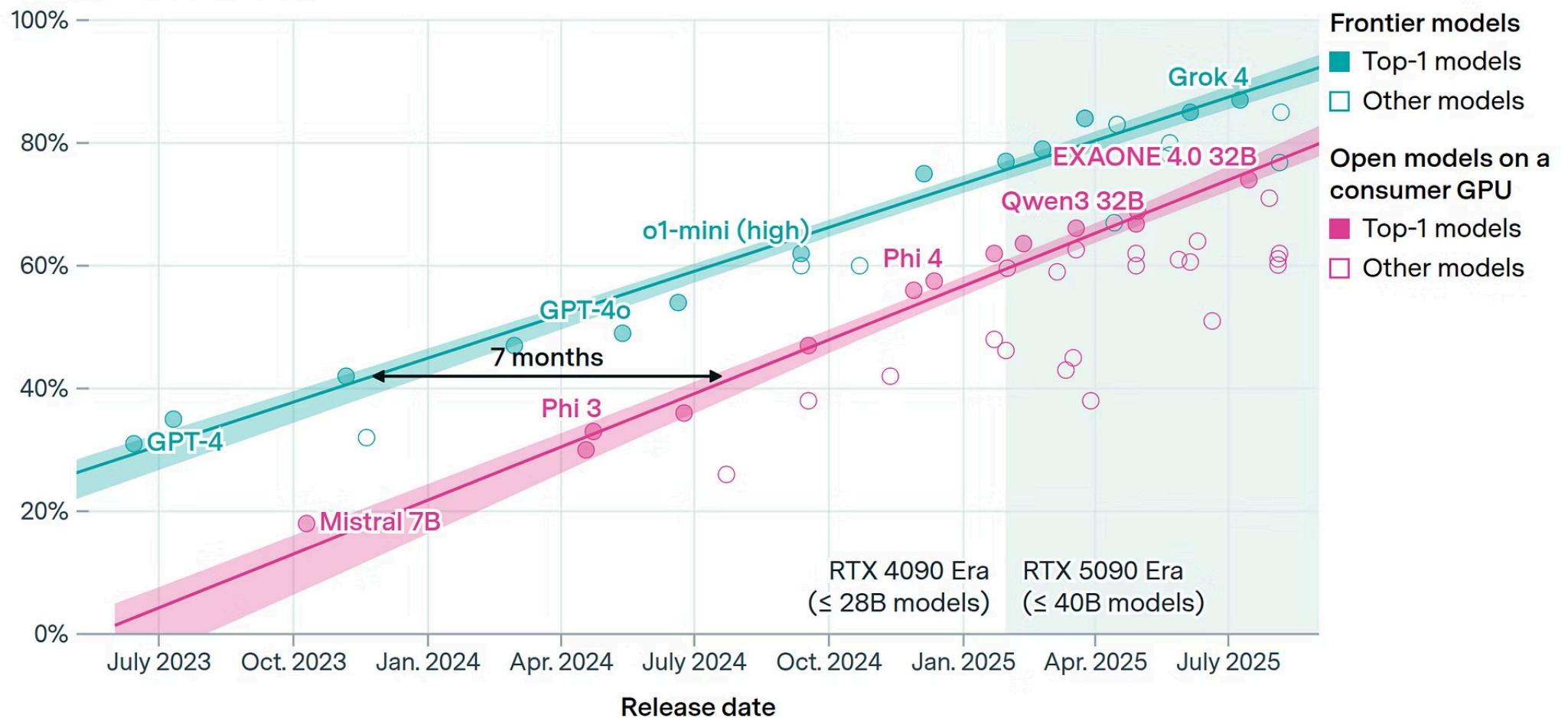
3 Open-Source Models

Complete transparency, with weights, inference code, training code, and data all made public. This enables full scrutiny, reproduction, and extensive community collaboration.

Open Models Performance

Models that fit on a single consumer GPU trail the absolute frontier by less than a year. 

GPQA-Diamond accuracy



huggingface.co/spaces/lmarena-ai/lmarena-leaderboard

Spaces | **lmarena-ai / lmarena-leaderboard** | like 4.65k | Running

App Files Community 74

Overview Text WebDev Vision Text-to-Image Image Edit Search Text-to-Video Image-to-Video Copilot Start Voting

Rank (UB) ↑ Rank (Style Control) ↓ Model ↑ Score ↑ 95% CI (±) ↑ Votes ↑ Organization ↑ License ↑

1	1	G gemini-2.5-pro		1465	±4	54.087	Google	Proprietary	
2	3 ↓	qwen3-max-preview		1440	±6	18.078	Alibaba	Proprietary	
2	10 ↓	glm-4.6		1438	±9	4.401	Z.ai	MIT	
3	2 ↑	chatgpt-4o-latest-20250326		1429	±4	40.013	OpenAI	Proprietary	
3	18 ↓	mistral-medium-2508		1427	±5	23.844	Mistral	Proprietary	
3	18 ↓	glm-4.5		1426	±5	22.612	Z.ai	MIT	
3	11 ↓	qwen3-vl-235b-a22b-instruct		1426	±8	6.312	Alibaba	Apache 2.0	
3	11 ↓	deepseek-r1-0528		1426	±6	19.284	DeepSeek	MIT	
3	1 ↑	claude-sonnet-4-5-20250929-thinking-32k		1423	±8	6.287	Anthropic	Proprietary	
4	14 ↓	grok-3-preview-02-24		1424	±4	34.154	xAI	Proprietary	
4	29 ↓	longcat-flash-chat		1421	±6	11.667	Meituan	MIT	
4	11 ↓	deepseek-v3.2-exp-thinking		1420	±9	4.320	DeepSeek AI	MIT	
4	24 ↓	qwen3-next-80b-a3b-instruct		1420	±6	12.793	Alibaba	Apache 2.0	
4	11 ↓	deepseek-v3.1		1419	±6	15.380	DeepSeek	MIT	

Source: <https://huggingface.co/spaces/lmarena-ai/lmarena-leaderboard>

Training Effort



"Computing Power and the Governance of Artificial Intelligence"
Sastry, Heim, Belfield, Anderljung, Brundage, Hazell, O'Keefe, Hadfield et al., 2024

AI Energy Consumption

The New York Times

Three Mile Island Plans to Reopen as Demand for Nuclear Power Grows

The infamous plant, closed since 2019, is getting a new lease on life after Microsoft agreed to buy its electricity to supply a growing fleet of data centers.



Listen to this article · 5:33 min [Learn more](#)



Share full article

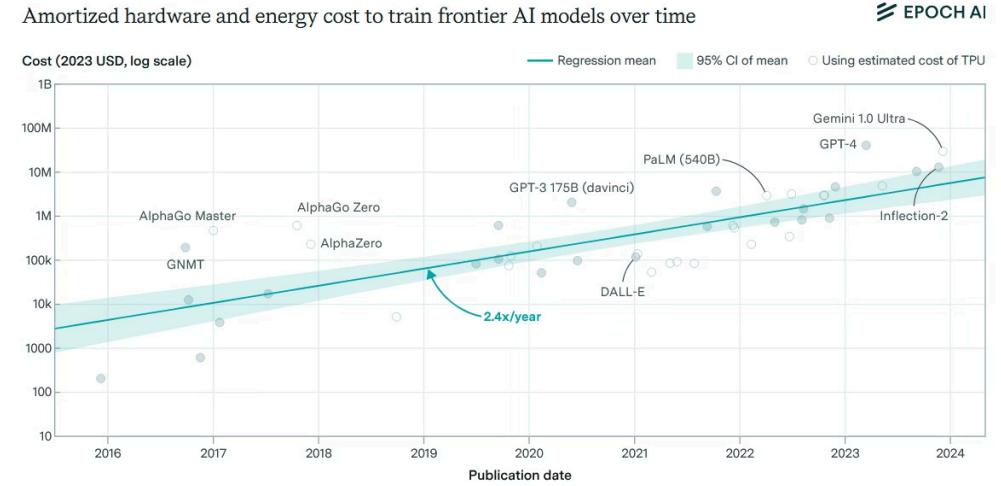


The Three Mile Island nuclear plant in Londonderry Township, Pa. Andrew Caballero-Reynolds/Agence France-Presse — Getty Images



By Brad Plumer

Sept. 20, 2024, 7:00 a.m. ET



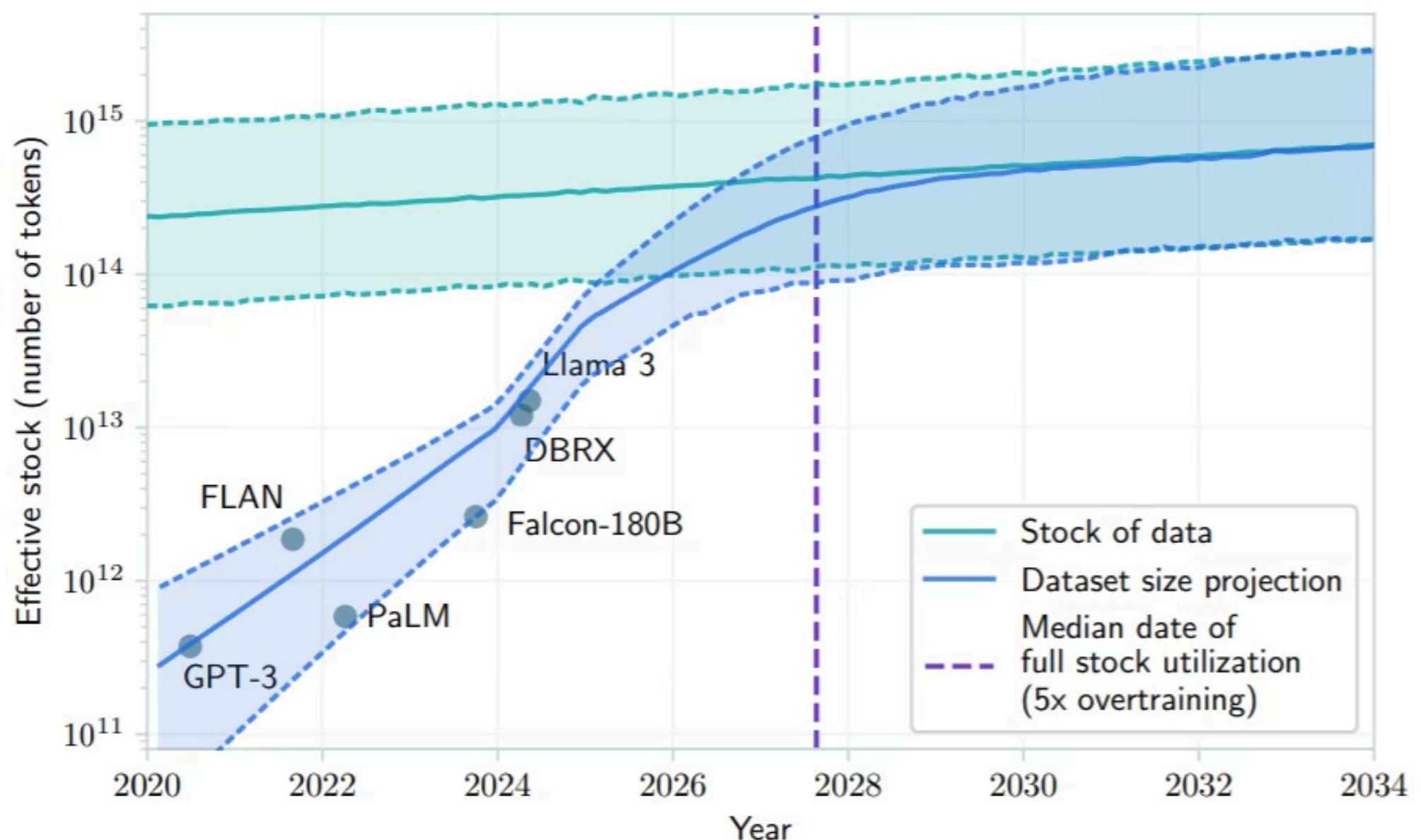
AI is expected to need 35 gigawatts by 2030 (that's over 300 TWh per year)

Eric Schmidt's controversial take on climate change:

"We're not going to hit the climate goals anyway"

"I'd rather bet on AI solving the problem than constraining it and having the problem"

Will We Run Out Of Data?



Will we run out of data? Limits of LLM scaling based on human-generated data; P. Villalobos et al., 2024 – <https://arxiv.org/pdf/2211.04325>

Popular AI Level Terminology

AI (Artificial Intelligence) General term for systems designed to perform tasks that typically require human intelligence. This includes both symbolic AI (rule-based, explicit reasoning) and subsymbolic AI (statistical pattern recognition, like machine learning).

AGI (Artificial General Intelligence) Intelligence with the ability to understand, learn, and apply knowledge across a wide range of tasks at human level or beyond. Unlike narrow AI systems that excel at specific tasks, AGI would have flexible, general problem-solving capabilities and transfer learning across domains. Building comprehensive world models is often considered crucial for AGI.

ASI (Artificial Superintelligence) Intelligence that surpasses human capabilities across virtually all domains. The concept suggests systems that could recursively improve themselves, potentially developing world models far more sophisticated than human understanding.

These terms are important to know, but many AI researchers try to avoid them.

How important is it to build models of the world?

Science. Human understanding. Explainability. Alignment. Safety.

THE TURING TEST

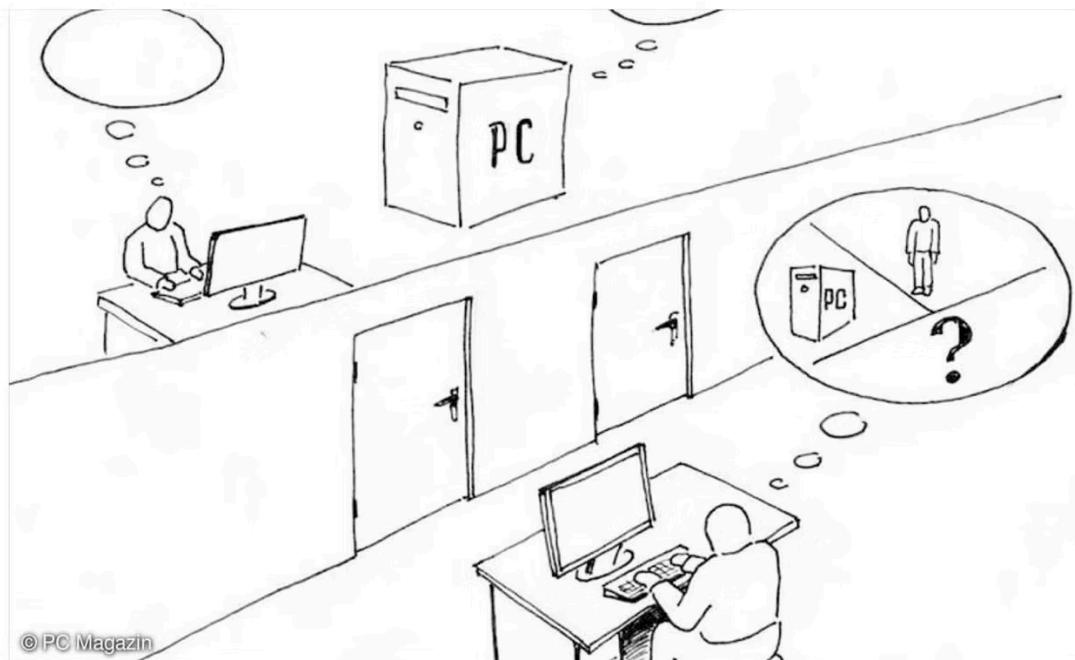
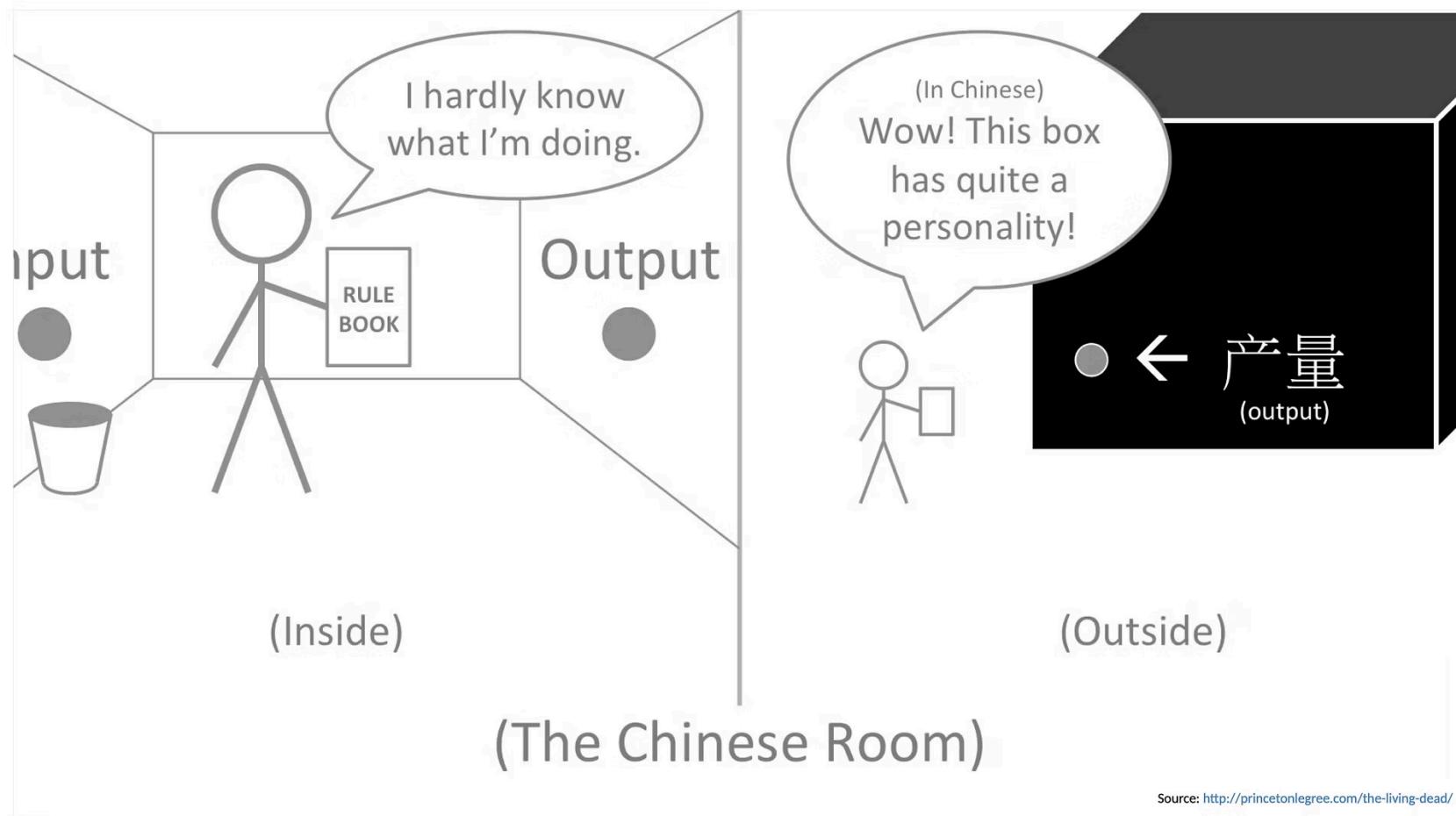


Image: PC Magazin

Further reading: <https://plato.stanford.edu/entries/turing-test/>



Further reading: <https://plato.stanford.edu/entries/chinese-room/>

The Scaling Hypothesis

"The scaling hypothesis regards the blessings of scale as the secret of AGI: intelligence is 'just' simple neural units & learning algorithms applied to diverse experiences at a (currently) unreachable scale. As increasing computational resources permit running such algorithms at the necessary scale, the neural networks will get ever more intelligent."

Source: Gwen Branwen, 2020. <https://gwern.net/scaling-hypothesis>

RGlobal data center demand could reach 68 gigawatts by 2027 and 327 gigawatts by 2030.

Source: RAND. https://www.rand.org/pubs/research_reports/RRA3572-1.html

The Bitter Lesson

Rich Sutton

March 13, 2019

The biggest lesson that can be read from 70 years of AI research is that general methods that leverage computation are ultimately the most effective, and by a large margin. The ultimate reason for this is Moore's law, or rather its generalization of continued exponentially falling cost per unit of computation. Most AI research has been conducted as if the computation available to the agent were constant (in which case leveraging human knowledge would be one of the only ways to improve performance) but, over a slightly longer time than a typical research project, massively more computation inevitably becomes available. Seeking an improvement that makes a difference in the shorter term, researchers seek to leverage their human knowledge of the domain, but the only thing that matters in the long run is the leveraging of computation. These two need not run counter to each other, but in practice they tend to. Time spent on one is time not spent on the other. There are psychological commitments to investment in one approach or the other. And the human-knowledge approach tends to complicate methods in ways that make them less suited to taking advantage of general methods leveraging computation. There were many examples of AI researchers' belated learning of this bitter lesson, and it is instructive to review some of the most prominent.

LLM Ecosystem for computer and data scientists

aka Tools You Should Know or At Least Should Have Heard About

- ⚠ Disclaimer: We have no affiliation with the providers of the mentioned tools and services and neither endorse nor oppose their use. Names of tools and providers will not be part of exams.

Where To Deploy an LLM

1

AI-as-a-Service (AlaaS)

E.g. OpenAI, Google, Anthropic

2

Infrastructure-as-a-Service (IaaS)

E.g. AWS, Runpod, Paperspace

3

Bare-Metal-as-a-Service (BMaaS)

E.g. Hetzner, OVHcloud

4

Self-Hosting or On-Premise AI

E.g. Server setups in own rack space or local

How To Programmatically Interact with an LLM

Using a framework

- Abstraction across various model providers
- Modularity
- Simplified workflow for complex chains
- Prebuilt integrations
- Memory management

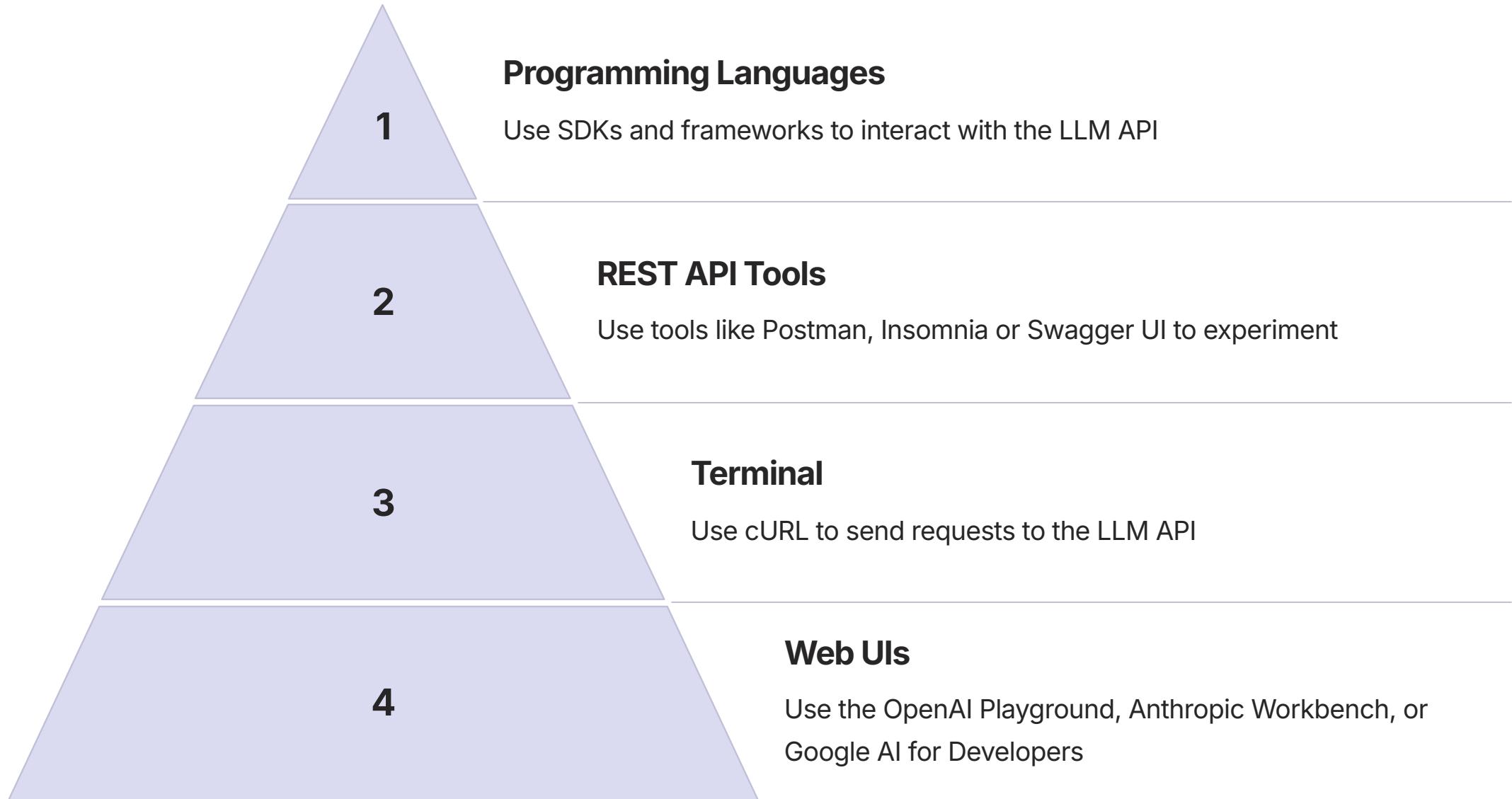
Using APIs directly

- Easier to understand what's actually going on
- Complete control
- No external dependency (bugs)
- Custom logic
- Lower latency

It's a bit like comparing a higher programming language with assembler

Interaction Paradigms

How To Talk To An LLM Via API



Accessing anything on the Internet

```
% telnet www.dbai.tuwien.ac.at 80
Trying 128.131.111.3...
Connected to www.dbai.tuwien.ac.at.
Escape character is '^].
GET / HTTP/1.0
```

```
HTTP/1.1 302 Found
Date: Mon, 20 Oct 2025 10:46:14 GMT
Server: Apache/2.4.25 (Debian)
Location: https://www.dbai.tuwien.ac.at/
Content-Length: 301
Connection: close
Content-Type: text/html; charset=iso-8859-1
```

```
<!DOCTYPE HTML PUBLIC "-//IETF//DTD HTML 2.0//EN">
<html><head>
<title>302 Found</title>
</head><body>
<h1>Found</h1>
<p>The document has moved <a href="https://www.dbai.tuwien.ac.at/">here</a>.</p>
<hr>
<address>Apache/2.4.25 (Debian) Server at www.dbai.tuwien.ac.at Port 80</address>
</body></html>
Connection closed by foreign host.
```

Accessing anything on the Internet /2

```
% curl http://www.dbai.tuwien.ac.at/
<!DOCTYPE HTML PUBLIC "-//IETF//DTD HTML 2.0//EN">
<html><head>
<title>302 Found</title>
</head><body>
<h1>Found</h1>
<p>The document has moved <a href="https://www.dbai.tuwien.ac.at/">here</a>.</p>
<hr>
<address>Apache/2.4.25 (Debian) Server at www.dbai.tuwien.ac.at Port 80</address>
</body></html>
```

Accessing anything on the Internet /3

```
% telnet localhost 11434
Trying 127.0.0.1...
Connected to localhost.
Escape character is '^].
POST /api/chat HTTP/1.0
Host: localhost
Content-Type: application/json
Content-Length: 81

{"model":"phi3","messages":[{"role":"user","content":"Who are you?"}]}

HTTP/1.0 200 OK
Content-Type: application/x-ndjson
Date: Mon, 20 Oct 2025 11:10:07 GMT

{"model":"phi3","created_at":"2025-10-20T11:10:07.697742Z","message":{"role":"assistant","content":"I","done":false}}
 {"model":"phi3","created_at":"2025-10-20T11:10:07.713761Z","message":{"role":"assistant","content":"","done":false}}
 {"model":"phi3","created_at":"2025-10-20T11:10:07.729493Z","message":{"role":"assistant","content":"m","done":false}}
 {"model":"phi3","created_at":"2025-10-20T11:10:07.745335Z","message":{"role":"assistant","content":" Ph","done":false}}
 {"model":"phi3","created_at":"2025-10-20T11:10:07.761405Z","message":{"role":"assistant","content":"i","done":false}}
 {"model":"phi3","created_at":"2025-10-20T11:10:07.777169Z","message":{"role":"assistant","content":":),"done":false}}
 {"model":"phi3","created_at":"2025-10-20T11:10:07.793192Z","message":{"role":"assistant","content":" an","done":false}}
 {"model":"phi3","created_at":"2025-10-20T11:10:07.808877Z","message":{"role":"assistant","content":" A","done":false}}
 {"model":"phi3","created_at":"2025-10-20T11:10:07.824614Z","message":{"role":"assistant","content":"I","done":false}}
 {"model":"phi3","created_at":"2025-10-20T11:10:07.841167Z","message":{"role":"assistant","content":" developed"},"done":false}
 {"model":"phi3","created_at":"2025-10-20T11:10:07.856975Z","message":{"role":"assistant","content":" by"},"done":false}
 ...
...
```

Calling the OpenAI API

```
curl https://api.openai.com/v1/chat/completions  
-H "Content-Type: application/json"  
-H "Authorization: Bearer <YOUR_API_KEY>"  
-d '{ "model": "gpt-4", "messages": [ { "role": "system", "content": "You are a helpful assistant." }, { "role": "user", "content": "Explain  
how to use curl to make an API request." } ], "max_tokens": 100, "temperature": 0.7 }'
```

→

```
{ "id": "chatcmpl...", "object": "chat.completion", "created": 1693505825, "model": "gpt-4", "choices": [ { "message": { "role":  
"assistant", "content": "You can use curl to interact with APIs by sending HTTP requests..." }, "finish_reason": "stop", "index": 0 } ],  
"usage": { "prompt_tokens": 20, "completion_tokens": 32, "total_tokens": 52 } }
```

Three Interaction Paradigms

Interacting with Large Language Models (LLMs) programmatically typically boils down to three primary methods, each offering distinct advantages depending on the application and developer preference.



Direct API Calls (cURL)

The foundational method, allowing direct interaction with LLM endpoints using HTTP requests. Ideal for testing, rapid prototyping, and language-agnostic integration.



Python

The de-facto standard for AI development, offering rich libraries, frameworks (e.g., LangChain, LlamaIndex), and a vast ecosystem for building complex LLM applications.



JavaScript (TypeScript)

Essential for web-based applications and Node.js backend services. Growing LLM support with libraries and frameworks like OpenAI's JS SDK and LangChain.js.

Quite often, online documentation shows how to interact with AI models either directly over API or using the two prevalent programming languages – Python for backend and Typescript for frontend.

gpt-oss-120B API Usage

cURL

Python

TypeScript

ENDPOINT

openai/gpt-oss-120b



```
1 curl -X POST "https://api.together.xyz/v1/chat/completions" \
2   -H "Authorization: Bearer $TOGETHER_API_KEY" \
3   -H "Content-Type: application/json" \
4   -d '{
5     "model": "openai/gpt-oss-120b",
6     "messages": [
7       {
8         "role": "user",
9         "content": "What are some fun things to do in New York?"
10      }
11    ]
12  }'
```

Noteworthy LLM Frameworks

1 LangChain

Arguably the best-known
LLM framework

Application development

<https://langchain.com/>

2 Llamaindex

Arguably the best-known
framework for RAG

Data retrieval

<https://llamaindex.ai/>

3 CrewAI

One of the best-known
frameworks for agents

Multi-agent systems

<https://crewai.com/>

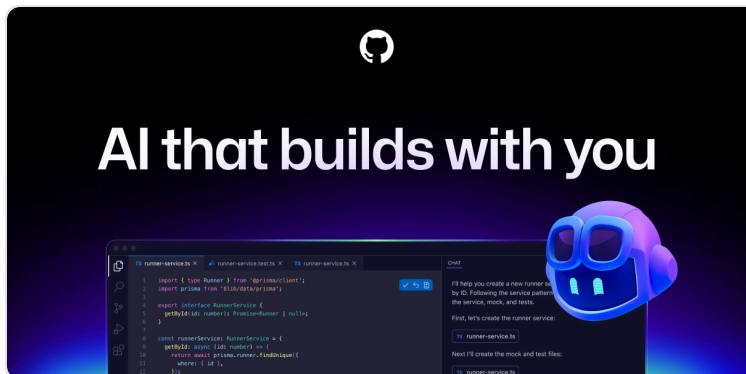


This is a purely subjective selection.

AI Assisted Development

AI Assisted Development

1 GitHub Copilot



GitHub Copilot · Your AI pair programmer

GitHub Copilot works alongside you directly in your editor, suggesting whole lines or entire functions for you.



2 Cursor



Cursor: The best way to code with AI

Built to make you extraordinarily productive, Cursor is the best way to code with AI.



3 Windsurf

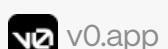
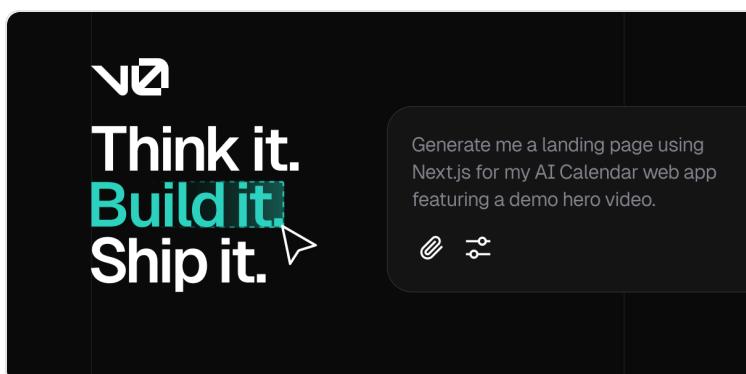


Windsurf – The best AI for Coding

Windsurf is the world's most advanced AI coding assistant for developers and enterprises. Windsurf Editor — the first AI-native IDE that keeps developers in...



4 Vercel v0 (specialized in frontend UI code)



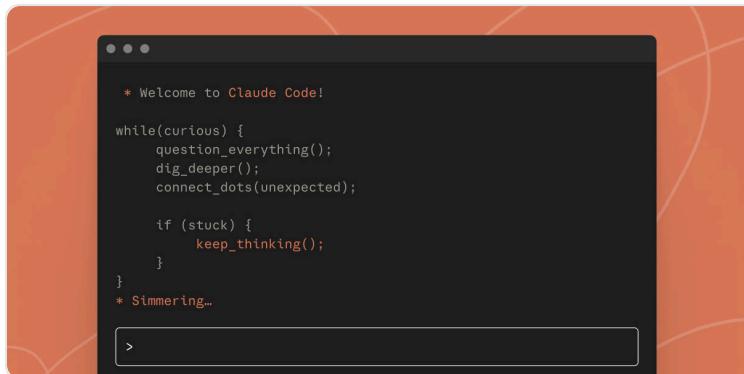
v0 by Vercel

Your collaborative AI assistant to design, iterate, and scale full-stack applications for the web.



AI Coding Assistants

1 Claude Code

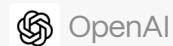


Claude Code | Claude



Your code's new collaborator. Embed Claude directly in your terminal or IDE to manage million-line codebases and implement changes.

2 OpenAI Codex



OpenAI Codex



Codex is OpenAI's series of AI coding tools that help developers move faster by delegating tasks to powerful cloud and local coding agents.

3 Cline



Cline – AI Coding, Open Source and Uncompromised



Cline is the open-source AI coding agent that gives you direct access to frontier models with complete transparency. Plan Mode, MCP integration, and zero...

Low Code / No Code Development

1 Replit Agent

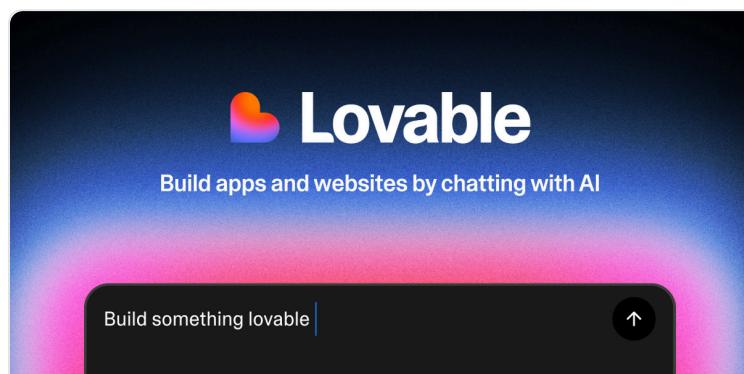


Replit – Build apps and sites with AI

Replit is an AI-powered platform for building professional web apps and websites.



2 Lovable

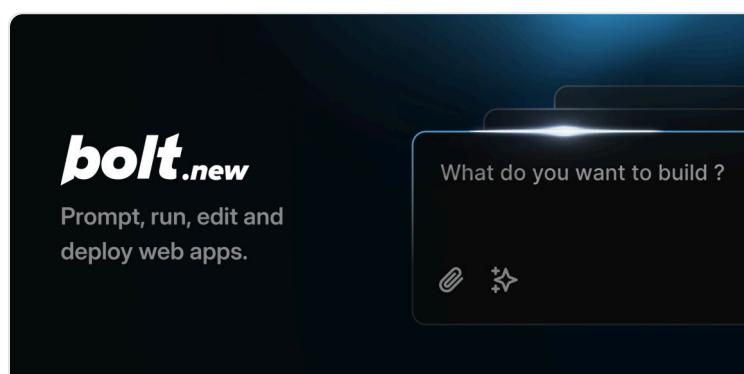


Lovable

Build software products, using only a chat interface



3 Bolt



Bolt AI builder: Websites, apps & prototypes

Prompt, run, edit & publish apps



Notable AI Model Platforms



Hugging Face

Largest repository of open-source models, datasets, and spaces.

Comprehensive model platform including distribution, tooling, fine-tuning and inference.



Together AI, Groq

Access to a range of models for fine-tuning (Together AI) and inference (both).



AWS, Microsoft Azure, Google Cloud AI

Cloud hyperscalers offering extensive foundation model management, training, deployment, and API ecosystems.



Openrouter

Single API for accessing multiple AI providers.

Model aggregator, unified API for inference.

The screenshot shows the OpenRouter AI Models page at openrouter.ai/models. The page features a sidebar on the left with filter categories: Input Modalities (Text, Image, File, Audio), Output Modalities (Text, Image), Context length (4K, 64K, 1M), Prompt pricing (FREE, \$0.5, \$10+), and Series (GPT). The main content area displays three AI models:

- Andromeda Alpha**: 1,25B tokens. Description: This is a smaller reasoning model that has been trained to be really good at image and visual understanding. This is a cloaked model provided to the community to gather feedback. Note: All ... by [openrouter](#) | 128K context | \$0/M input tokens | \$0/M output tokens.
- IBM: Granite 4.0 Micro**: 33,6M tokens. Description: Granite-4.0-H-Micro is a 3B parameter from the Granite 4 family of models. These models are the latest in a series of models released by IBM. They are fine-tuned for long context tool calling. by [ibm-granite](#) | 131K context | \$0,017/M input tokens | \$0,11/M output tokens.
- Deep Cogito: Cogito V2 Preview Llama 405B**: 44,2M tokens. Description: Cogito v2 405B is a dense hybrid reasoning model that combines direct answering capabilities with advanced self-reflection. It represents a significant step toward frontier intelligence with dense ... by [deepcogito](#) | 33K context | \$3,50/M input tokens | \$3,50/M output tokens.

OpenRouter

A router for LLMs and other AI models



OpenRouter

A router for LLMs and other AI models

[OpenRouter](#)



How to Discover & Experiment with LLMs

- **Ask an LLM:** "What are the best LLMs to experiment with?"
- **HuggingFace:** Founded as a teen chatbot platform in 2018. Initially gained popularity for providing PyTorch-friendly implementations of BERT and GPT-2, leading to the creation of the transformers library. Currently hosts about 2 million models and is the go-to place for accessing state-of-the art, pre-trained models.
- **Ollama:** A tool for running and managing LLMs locally, eliminating cloud reliance. Features an interactive shell for direct LLM interaction, a REST API for programmatic requests, Python library integration for easy development, local model management, and GPU acceleration. Supports a wide range of models, including such as Llama 3.1/3.2, Mistral, Code Llama, Phi-3, Qwen3, Gemma 2, and DeepSeek-R1, and allows for custom model creation.

 HuggingFace – <https://huggingface.co/>

Ollama – <https://ollama.com/>

huggingface.co/deepseek-ai/deepseek-v3

Hugging Face Search models, datasets, users...

Models Datasets Spaces Posts Docs Enterprise Pricing

deepseek-ai / DeepSeek-V3 like 1.48k Follow DeepSeek 5.78k

Safetensors deepseek_v3 custom_code fp8 arxiv:2412.19437

Model card Files and versions Community 46 Edit model card

YAML Metadata Warning: empty or missing yaml metadata in repo card (<https://huggingface.co/docs/hub/model-cards#model-card-metadata>)

deepseek

DeepSeek Homepage Chat DeepSeek V3 Hugging Face DeepSeek AI
Discord DeepSeek AI WeChat DeepSeek AI Twitter deepseek ai
Code License MIT Model License Model Agreement

Paper Link

1. Introduction

We present DeepSeek-V3, a strong Mixture-of-Experts (MoE) language model with 671B total parameters with 37B activated for each token. To achieve efficient inference and cost-effective training, DeepSeek-V3 adopts Multi-head Latent Attention

Downloads last month 74,084

Safetensors Model size 685B params Tensor type BF16 · F8_E4M3 · F32

Inference API Unable to determine this model's library. Check the docs.

Model tree for deepseek-ai/DeepSeek-V3 8 models
Adapters 8 models
Finetunes 9 models
Quantizations 12 models

Spaces using deepseek-ai/DeepSeek-V3 6
Akshayram1/data_visualization_ai_excel_togetherai_e2b
shaktibiplab/deepseekv3 sapthesh/deepseekv3
Heuehneje/new-space failtowin/new-space
Sujitha/DreamWeaver-AI

The screenshot shows a web browser window displaying the Ollama website at ollama.com. The page features a large llama logo in the center. Below the logo, the text reads: "Get up and running with large language models." It also mentions that the software is available for macOS, Linux, and Windows. At the bottom of the page, there is a footer with links to various sections like Blog, Docs, GitHub, and Meetups.

© 2025 Ollama

Blog Docs GitHub Discord X (Twitter) Meetups Download

Personal Knowledge Management

Some Hierarchical PKM Tools

1 Notion



The AI workspace that works for you. | Notion

Build custom agents, search across all your apps, and automate busywork. The AI workspace where teams get more done, faster.



2 OneNote



Microsoft OneNote



3 Evernote



Best Note Taking App – Organize Your Notes with Evernote

Our note taking app helps you capture and prioritize ideas, projects and to-do lists, so nothing falls through the cracks. Start your free trial today!



Some Network PKM Tools

1 Roam Research



 Roam Research

Roam Research – A note taking tool for networked thought.

As easy to use as a word document or bulleted list, and as powerful for finding, collecting, and connecting related ideas as a graph database. Collaborate with...

2 Obsidian

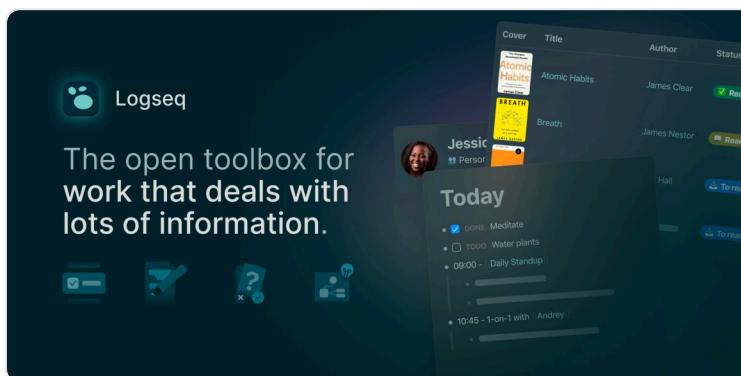


 Obsidian

Obsidian – Sharpen your thinking

The free and flexible app for your private thoughts.

3 Logseq



 logseq

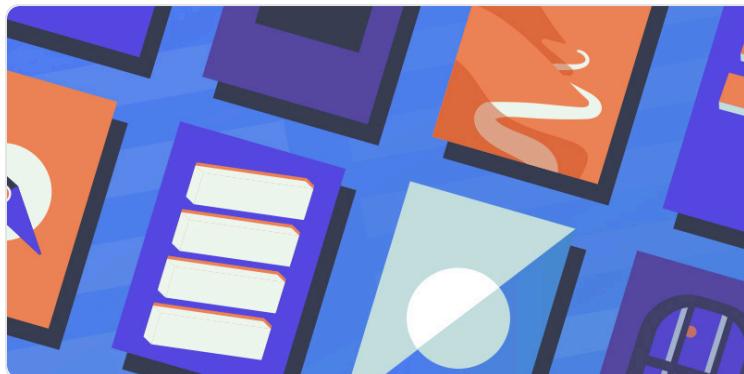
A privacy-first, open-source knowledge base

A privacy-first, open-source platform for knowledge management and collaboration.

Some Capture Tools

1

Readwise



 readwise.io

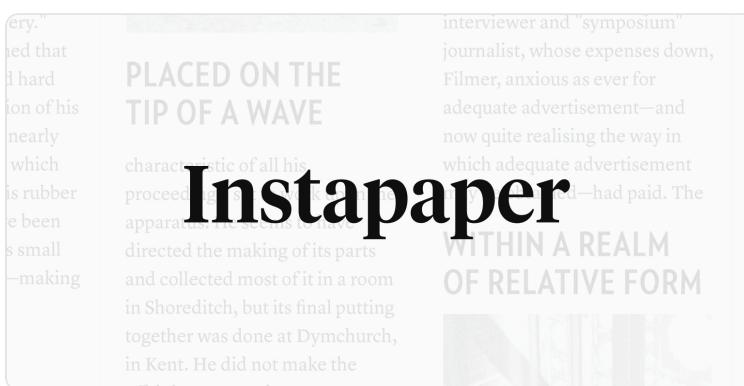


Readwise

Grow wiser and retain books better: Readwise sends you a daily email resurfacing your best highlights from Kindle, Instapaper, iBooks, and more.

2

Instapaper



 www.instapaper.com



Instapaper

A simple tool for saving web pages to read later on iOS, Android, computer, Kindle or Kobo eReader.

3

Raindrop



 raindrop.io



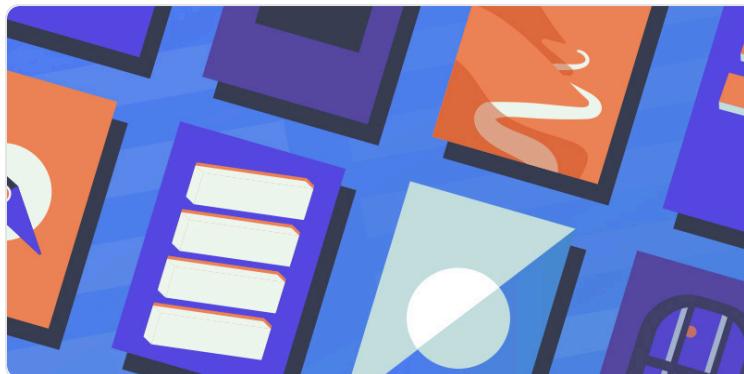
Raindrop.io — All in One Bookmark Manager

All in One Bookmark Manager. For your inspiration, read later, media and stuff.

Some Capture Tools

1

Readwise



 readwise.io

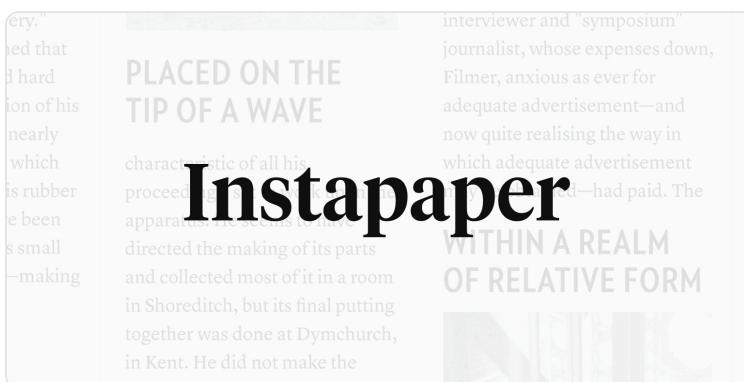


Readwise

Grow wiser and retain books better: Readwise sends you a daily email resurfacing your best highlights from Kindle, Instapaper, iBooks, and more.

2

Instapaper



 www.instapaper.com



Instapaper

A simple tool for saving web pages to read later on iOS, Android, computer, Kindle or Kobo eReader.

3

Raindrop



 raindrop.io



Raindrop.io — All in One Bookmark Manager

All in One Bookmark Manager. For your inspiration, read later, media and stuff.

More PKM Tools

1 Tana

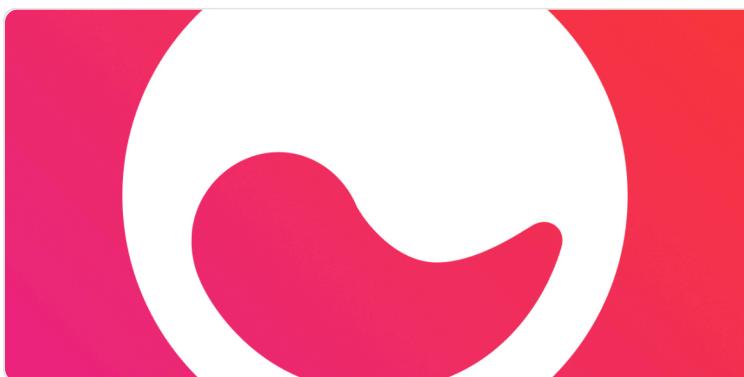


Tana



Stay on top of everything – without the busywork. An AI-native workspace that gives you an unfair advantage.

2 mem.ai



Mem



Let AI organize your team's work—from meeting notes, projects, to knowledge bases. All instantly searchable and readily discoverable.

Modes of Personal Knowledge Management

We propose the following mental model of modes when working with PKM:

- **Collect**
 - Collect data / information (/ knowledge?)
- **Surface**
 - Find collected thoughts, either active (search) or passive (agentic)
- **Connect**
 - Connect related ideas, consider perspective
- **Synthesize**
 - Synthesize knowledge from collection
- **Reflect**
 - Analyze thinking patterns and knowledge gaps