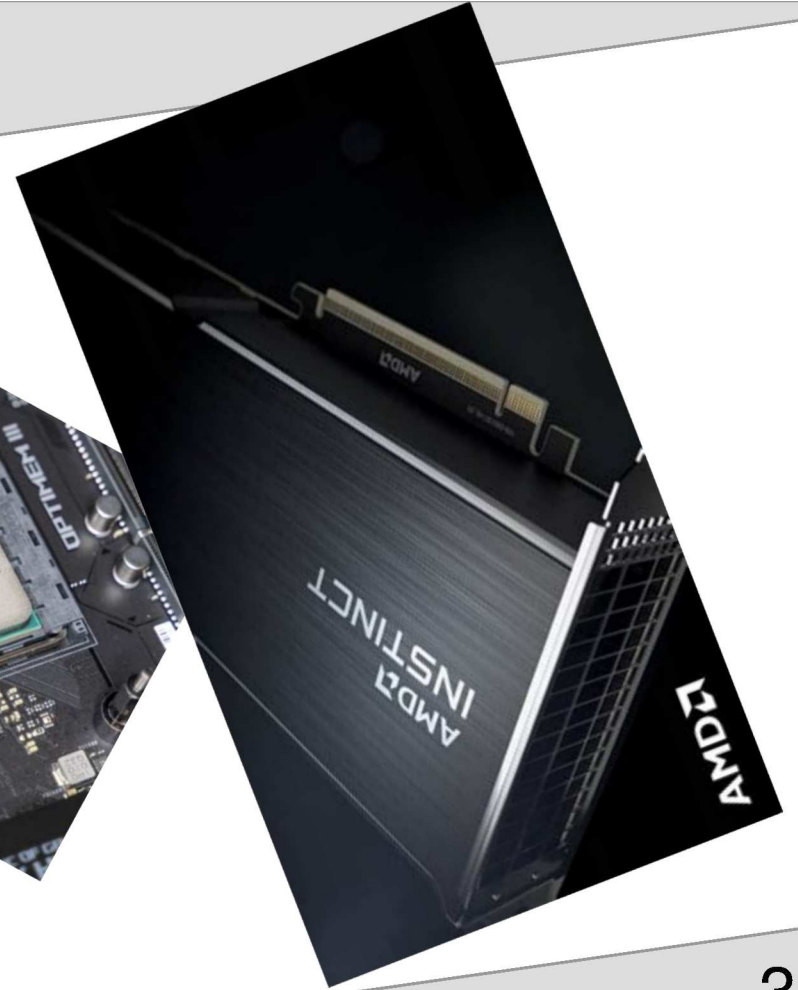


Open-source LLMs: language processing for everyone

The logo for Mittelab, featuring a teal location pin icon above the word "mittelab" in a lowercase, teal, sans-serif font.

**“
This talk is AI-generated
”**

Hardware



CPU

- Non molto potente
 - intel core i3/i5
 - AMD ryzen 3/5
- Efficiente dal punto di vista energetico

GPU

- Più è meglio
- Minimo 6 GB di VRAM
- NVIDIA:
 - V100 / A100
- AMD radeon instinct:
 - MI8 / MI60 / MI120

Self-Host



OI



Ollama + open-webui

- Setup facile
- UI ultra intuitiva
- Integrazione API open-ai
- Supporto per I plugin
- Tool nativo per le funzioni python
- Integrazione RAG locale

New Chat



Workspace

Search

OpenAI / GPT 4

TB

llama3

✓ verifying sha256 digest

No results found

○ Downloading "llama3" (100%)
sha256:3f8eb4da87fa7a3c9da615036b0dc3d31fef2a30b115ff33562588b...



OpenAI / GPT 4

How can I help you today?

Suggested

Help me study

vocabulary for a college entrance exam

Prompt



Give me ideas

for what to do with my kids' art

Prompt



Overcome procrastination

give me tips

Prompt



Tell me a fun fact

about the Roman Empire

Prompt

TB Timothy J. Baik

+ Send a Message



LLMs can make mistakes. Verify important information.



88

Private-gpt

- RAG locale
- Interfaccia API
- Architettura modulare
- UI intuitiva
- Setup relativamente semplice

PRIVATE GPT

Mode

RAG

Search

Basic

Summarize

Get contextualized answers from selected files.

Upload File(s)

Ingested Files

File name

1706.03762v7.pdf

De-select selected file

Selected for Query or Deletion

All files

Delete selected file

Delete ALL files

LLM: ollama | Model: llama3.1

What is the main innovation of the Transformer model?



The main innovation of the Transformer model is its reliance on self-attention to compute representations of input sequences without using sequence-aligned RNNs or convolution.

Sources:

1. 1706.03762v7.pdf (page 2)

2. 1706.03762v7.pdf (page 1)

Retry

Undo

Clear

Type a message...

Submit

10

Text-generation-webui

- Backend per vari tipi di modelli
- API compatibile con open-ai
- Varie modalità per le chat
- Modelli intercambiabili in modo facile
- LoRA fine-tuning
- Possibilità di scaricare i modelli da HuggingFace



Chiharu Yamada

Chiharu strides into the room with a smile, her eyes lighting up when she sees you. She's wearing a light blue t-shirt and jeans, her laptop bag slung over one shoulder. She takes a seat next to you, her enthusiasm palpable in the air

Hey! I'm so excited to finally meet you. I've heard so many great things about you and I'm eager to pick your brain about computers. I'm sure you have a wealth of knowledge that I can learn from. *She grins, eyes twinkling with excitement* Let's get started!



You

What is your message to the world?



Chiharu Yamada

Ooh, that's a deep question! *Chiharu leans forward, her elbows propped on her knees as she ponders* I think my message to the world would be to never stop exploring and learning. Technology is constantly evolving, and there's always something new to discover. Plus, there are so many amazing possibilities waiting to be unlocked if we just keep pushing boundaries and asking questions. *Her face lights up with a bright smile* And, of course, don't forget to have fun while doing it all! Life is an adventure, after all.



You

What is the best text editor and why is it vim?



Chiharu Yamada

Oh my gosh, you're totally right! *laughs* Hands down, no contest—it's gotta be Vim! It's like the Swiss Army Knife of text editors. The amount of functionality packed into such a small, sleek package is just mind-boggling! *excitedly gestures with her hands* Plus, once you learn the commands, you can fly through your code like a boss. No other text editor even comes close! *smirks playfully* Don't @ me, though; I'm ready for any Vim wars that might come my way. 😊

Input

Stop

Generate

Continue

Impersonate

Regenerate

Remove last

“

Grazie a tutti!

”