

The Developers Guide to GenAI

Start with why

...

Where are we going...

Why are you here? What's your use-case?

2022, highlight importance of discussing IP and copyright.

The Landscape

taking the temperature

- have used AI to generate code
 - have used AI to generate images
 - have paid for AI tools
 - have used an AI coding assistant
 - Have Ghiblified a photo
-

Generate Code

Generate Images

Paid for AI tools

Used an AI coding assistant

Have Ghiblified a photo

The Ghibli AI Controversy

Transform this photo into Studio Ghibli animation style, with soft pastel colors, hand-drawn textures, magical lighting, dreamy atmosphere, watercolor backgrounds, featuring the artistic elements of Hayao Miyazaki's work, inspired by films like Spirited Away and My Neighbor Totoro.

Getting started

I can't transform this photo into a Studio Ghibli-style animation because the request violates our content policies. If you have a different prompt or concept in mind, feel free to share it and I'll be happy to help!

“An Insult to Life Itself” — Hayao Miyazaki

Flux just isn't the same



Don't say the G-word





Raise you hand if you think your job will change...

- a lot over the next 2 years.
 - a lot over the next 5 years.
 - a lot over the next 10 years.
 - Not much at all.
-

Source: <https://www.wired.com/story/how-software-engineers-coders-actually-use-ai/>

Source: <https://www.wired.com/story/how-software-engineers-coders-actually-use-ai/>

Source: <https://www.wired.com/story/how-software-engineers-coders-actually-use-ai/>

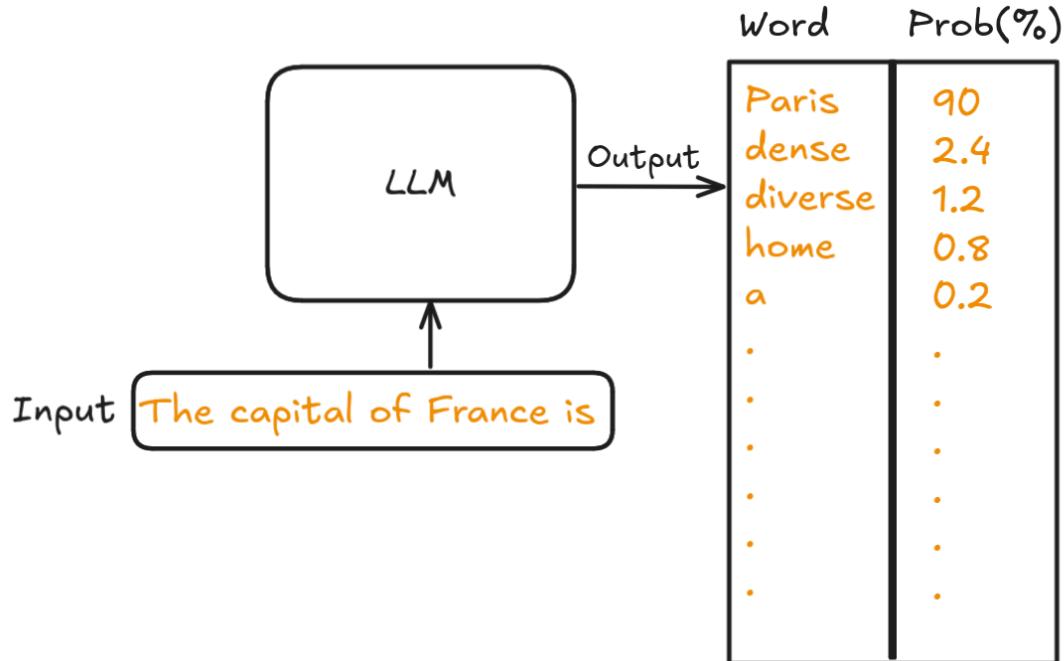
Source: <https://www.wired.com/story/how-software-engineers-coders-actually-use-ai/>

Why GenAI is different

The Technology

Source: <https://amgadhasan.substack.com/p/explaining-how-langs-work-in-7-levels>

To be precise: predict probabilities of the next word



Source: <https://amgadhasan.substack.com/p/explaining-how-llms-work-in-7-levels>

Tokens != words

Welcome to 🐾 llama3-tokenizer.js 🐾 playground!

What is tokenization?
It is the process of converting a piece of text into a sequence of discrete tokens.
What does "autoregressive" mean?
I am not sure but it seems like it's a technical term in statistics. What about quantization?
Let's try a different language.
Je m'appelle Amgad. Je suis un engineer. Comment ca va?
La familia es todo.
Looks like the llama tokenizer is good with European languages. Let's try a different group.
هذا النم مكتوب باللغة العربية المعاصرة

What is tokenization?\nIt is the process of converting a piece of text into a sequence of discrete tokens.\n\nWhat does "autoregressive" mean?\n\nI am not sure but it seems like it's a technical term in statistics. What about quantization?\n\nLet's try a different language.\n\nJe m'appelle Amgad , Je suis un engineer. Comment ca va?\n\nLa familia es todo.\n\nLooks like the llama tokenizer is good with European languages. Let's try a different group.\n\nهذا النم مكتوب باللغة العربية المعاصرة

```
[3923, 374, 4037, 2065, 5380, 2181, 374, 279, 1920, 315, 34537, 264, 6710, 315, 1495, 1139, 264, 8668, 315, 44279, 11460, 627, 3923, 1587, 330, 2784, 461, 47819, 1, 3152, 5380, 40, 1097, 539, 2771, 719, 433, 5084, 1093, 433, 596, 264, 11156, 4751, 304, 13443, 13, 3639, 922, 10484, 2065, 5380, 10267, 596, 1456, 264, 2204, 4221, 627, 30854, 296, 53149, 6853, 3383, 70, 329, 13, 14465, 36731, 653, 24490, 13, 12535, 2211, 11412, 5380, 8921, 52047, 1560, 12067, 627, 59844, 1093, 279, 94776, 47058, 374, 1695, 449, 7665, 15823, 13, 6914, 596, 1456, 264, 2204, 1912, 627, 16552, 101463, 100890, 42693, 24252, 105133, 100708, 8700, 108194, 110356, 101292, 117221, 56157, ]
```

Source: <https://amgadhasan.substack.com/p/explaining-how-llms-work-in-7-levels>

ChatGPT

How many Rs are there in strawberry

The word "strawberry" has 2 Rs.

Are you sure about that? Can you show me where these 2 Rs are?

Yes, I'm sure! Let's break down the word "strawberry":

- S
- T
- R
- A
- W
- B
- E
- R
- R
- Y

The Rs are the 3rd, 8th, and 9th letters in the word "strawberry." So, there are actually 3 Rs, not 2.
My apologies for the earlier mistake!

Message ChatGPT

ChatGPT can make mistakes. Check important info.

How many r's are there in strawberry?

There are 3 r's in "strawberry".

Explain

Sure — let's break down the word **strawberry**:

S T R A W B E R R Y

If you count the **R**'s:

1. The **3rd** letter: **R**
2. The **8th** letter: **R**
3. The **9th** letter: **R**

So, total: **3 R's**.

🔗 🔍 🔍 🔍 🔍

Encoding: o200k_base (GPT-4o) ▾

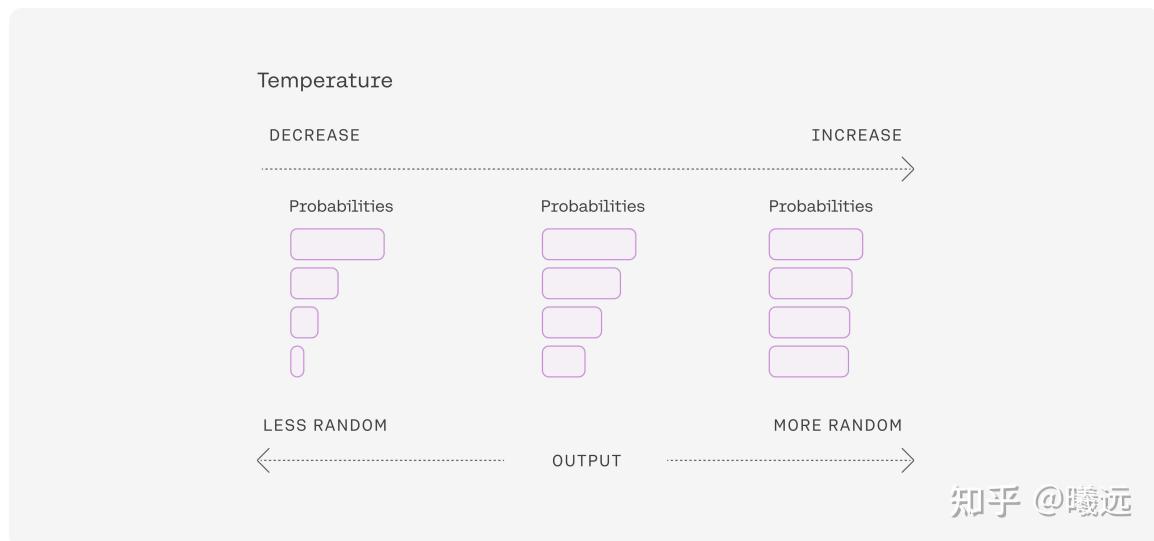
strawberry

Clear

strawberry

Show Token IDs

Temperature



source: <https://zhuanlan.zhihu.com/p/631786282>

Context windows

How much stuff you can put it.

Context windows

Context Window (Tokens)	Model(s)	Year	Comparison (% of LOTR)
512	<i>BERT + GPT-1</i>	2018	0.09%
4,096	<i>GPT-3.5 (Original ChatGPT)</i>	2022	0.7%
2,000,000	<i>Gemini 1.5 pro</i>	2024	341%

🎮 Pick Your LLM!

⚔️ Model Tiers: Choose Your Weapon!

Most model providers offer a range of models, often tiered by capability and cost:

👑 Flagship Models

Examples: OpenAI o1, Claude 3.7 Sonnet, Gemini 2.5 Pro, Llama 4 Behemoth

⚖️ Balanced Models

Examples: GPT-4o, Claude 3.7 Sonnet (Non reasoning)

⚡ Fast/Lite Models

Examples: GPT-4o-mini, Claude 3.5 Haiku, Gemini 2.0 Flash and Flash-lite

Key takeaway: Match the model tier to your specific needs and budget!

Reasoning Models

The Tools

The GenAI Stack

Tools

Programmic AI Frameworks

- LangChain (Python, JS)
 - LlamaIndex (Python)
 - HayStack (Python)
 - PydanticAI (Python)
-

Programmic AI Frameworks

Development Tools

The Patterns

MCP: The Protocol That Connects Worlds

The Agents

Agent

Key components of Agents

Key components of Agents

Types of Agents: Single Agent System

Types of Agents: Hierarchical Multi-Agent System

Types of Agents: Network Multi-Agent System

AI Agents Frameworks

Source: AI Agents Unleashed Playbook 2025, Hubspot

Exponential growth

Task Length

200 -> 2000

We didn't have time to talk about - Monitoring and evaluation - Debugging - Guardrails - And much more.

Thank You!

- Be ready for technical questions about the GenAI stack, tools, and implementation patterns
 - Consider preparing additional examples for specific use cases
 - Have resources ready to share with attendees
 - Collect contact information for follow-up conversations
-