

# >> AI CONF

Milano

17 GIUGNO 2024

## Generative AI and Foundation Models, what's in it for us, developers?

François Bouteruche  
Senior Developer Advocate @ Amazon Web Services

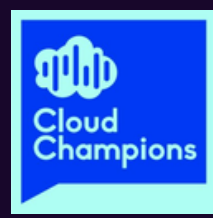


improve

# >> AI CONF

17 GIUGNO 2024  
4a EDIZIONE

adesso.it | software **one**



# Introduction to generative AI

## Image



Generated with Amazon Titan Image Generator G1

“A happy golden retriever wearing glasses and a hat in a portrait painting”

## Text/Chat

> Question: What is generative AI?

Generative AI refers to a type of artificial intelligence that is capable of creating new and original content. It uses algorithms and machine learning to generate text, images, music, and other forms of media based on patterns and rules learned from existing data.

Generated with Amazon Titan Text G1 - Express

# What are Language Models?

# Towards Large Language Models

Next word prediction with simple **word embedding**

Sentence

? ? ? ? ? c ? t ? lazy \_  
person.

# Towards Large Language Models

Next word prediction with 'memory' such as **recurrent neural networks (RNN)**

Sentence

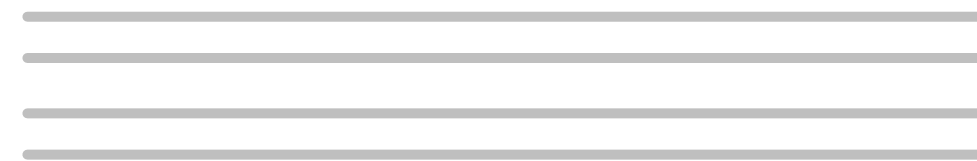
? ? ? fox jumps over the lazy \_  
~~person.~~  
rabbit.

# Transformers – Attention Is All You Need (2017)

---

## Attention Is All You Need

---



[“Attention is all you need”](#)

Vaswani, Ashish & Shazeer, Noam & Parmar, Niki & Uszkoreit, Jakob & Jones, Llion & Gomez, Aidan & Kaiser, Lukasz & Polosukhin, Illia. (2017)



# Transformers – Attention Is All You Need (2017)

## The Transformer model architecture

- “Attention” enabled models to scale the understanding of relationships between words
- Efficiently use parallel computing

Ashish Vaswani et al., [Attention Is All You Need](#).

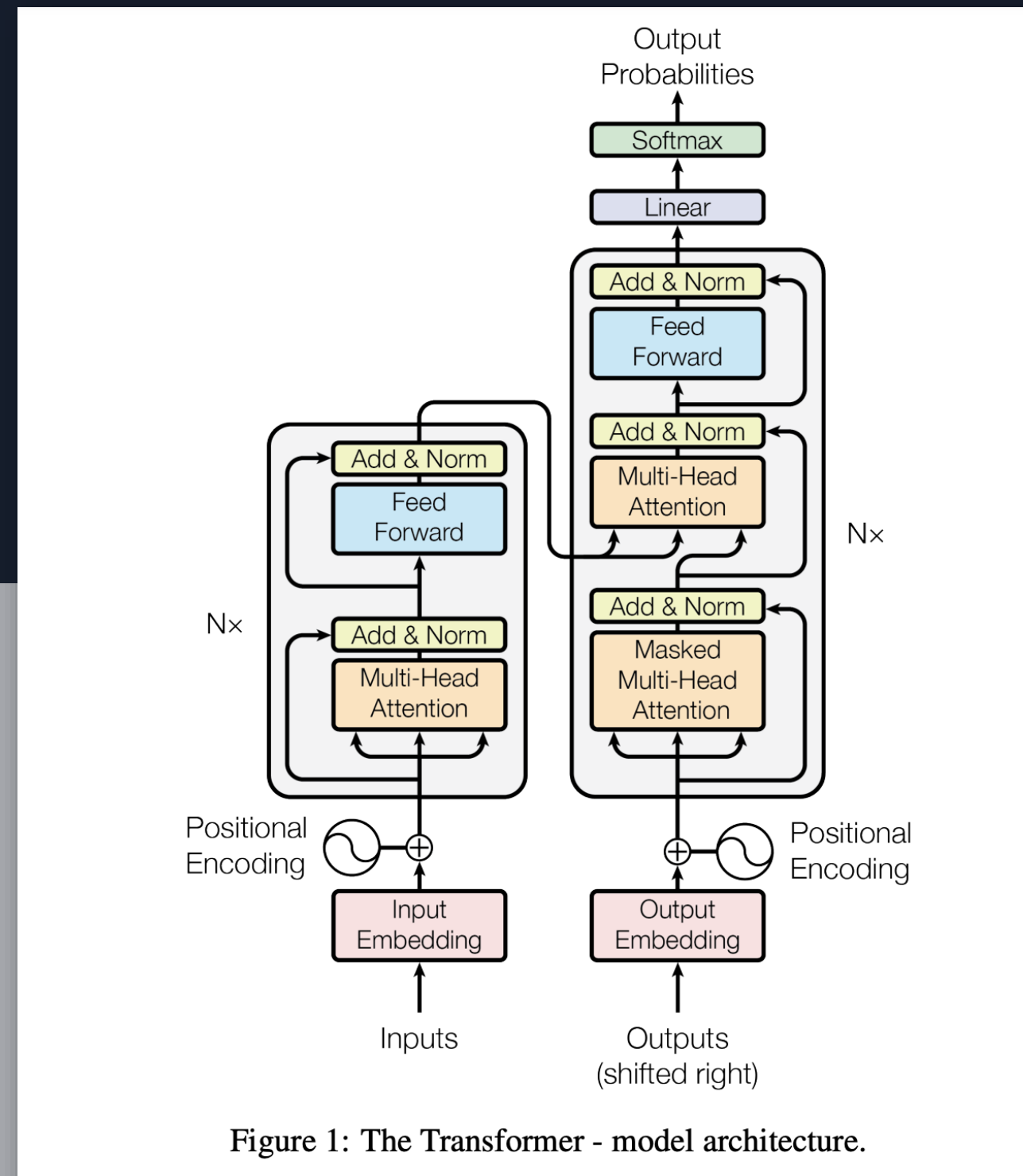


Figure 1: The Transformer - model architecture.



# Towards Large Language Models

Prediction using a **transformer based network**

Sentence

The quick brown fox jumps over the lazy \_

# Towards Large Language Models

Prediction using a **transformer based network**

Sentence

The quick brown fox jumps over the lazy dog.

...

It appears to be referring to the classic English pangram "The quick brown fox jumps over the lazy dog".

# How does text generation work?

# Generating text with transformers

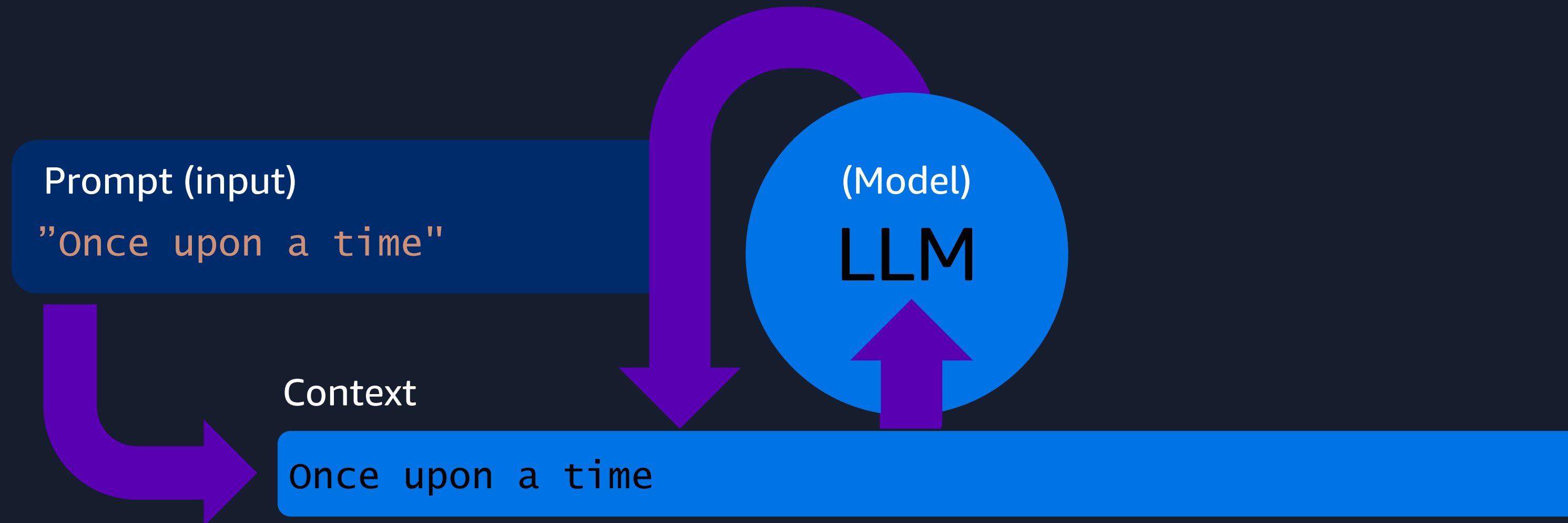
Prompt (input)

"Once upon a time"

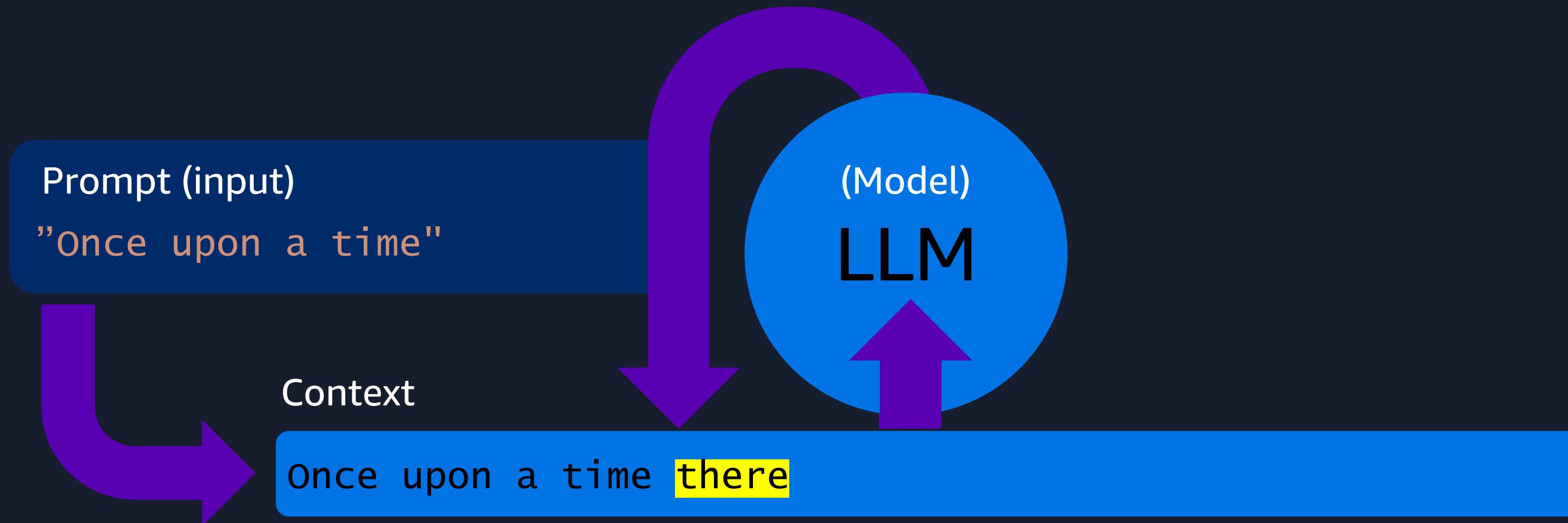
(Model)

LLM

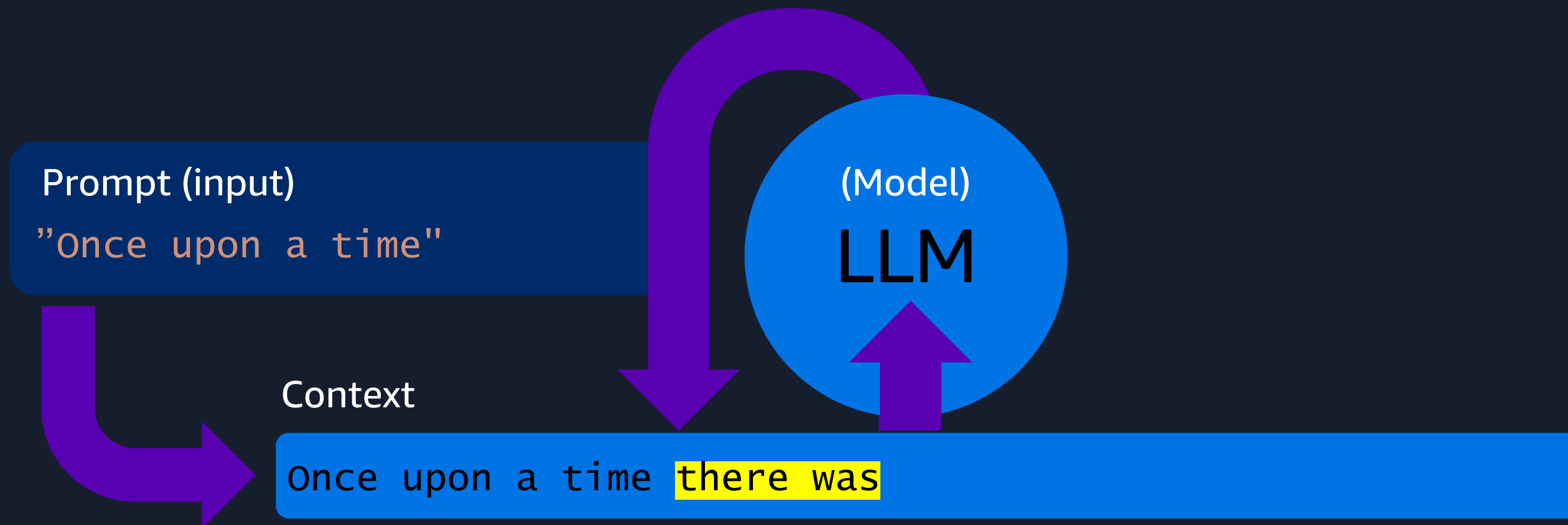
# Generating text with transformers



# Generating text with transformers

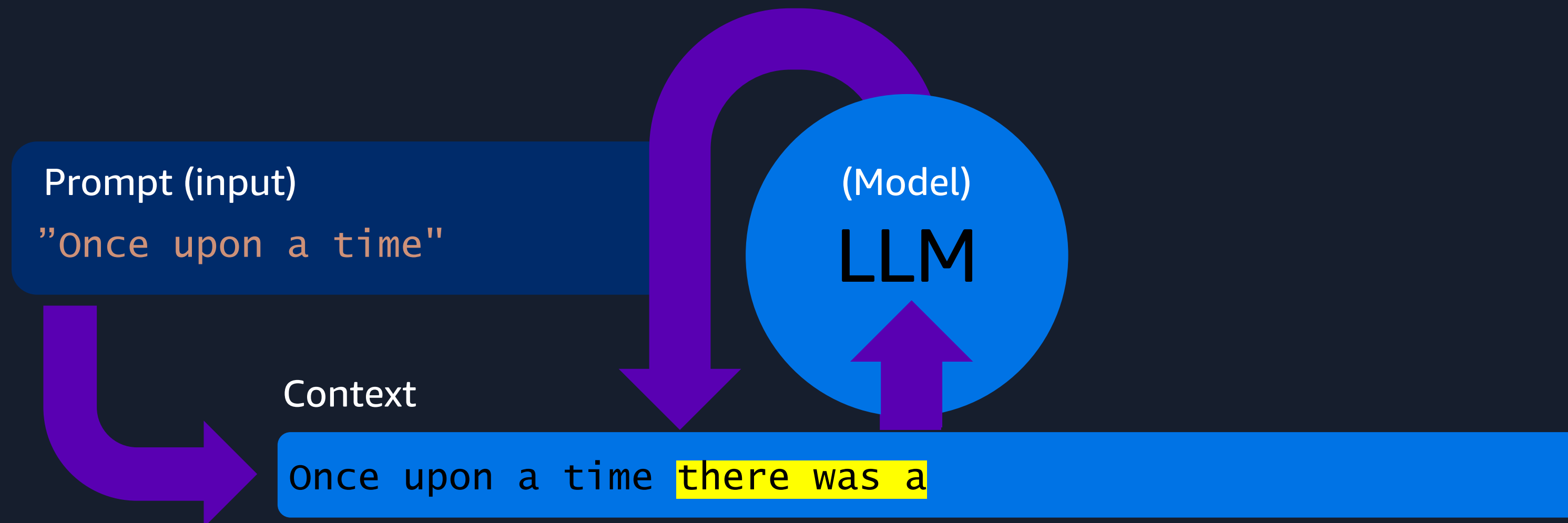


# Generating text with transformers

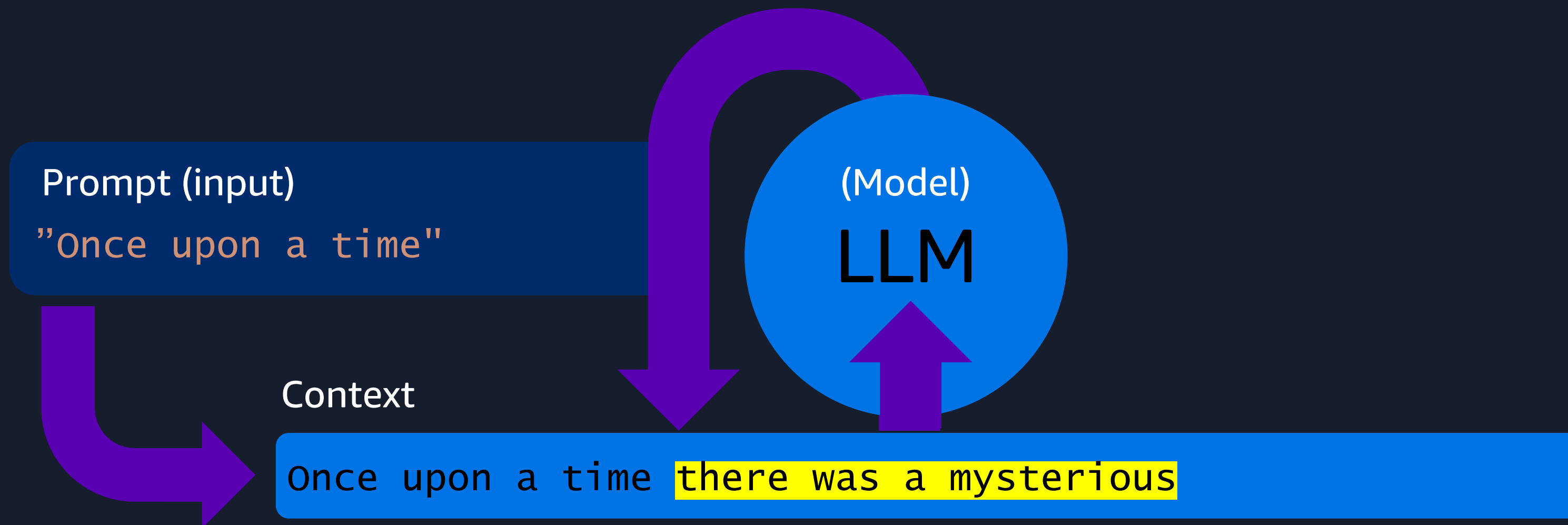




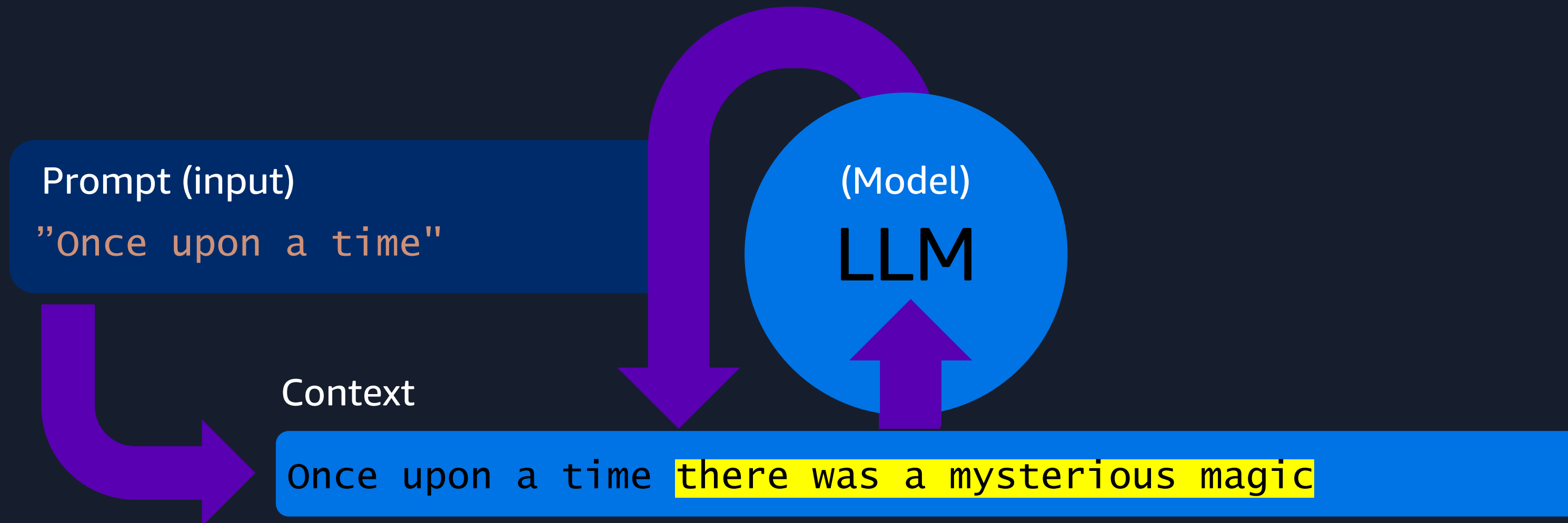
# Generating text with transformers



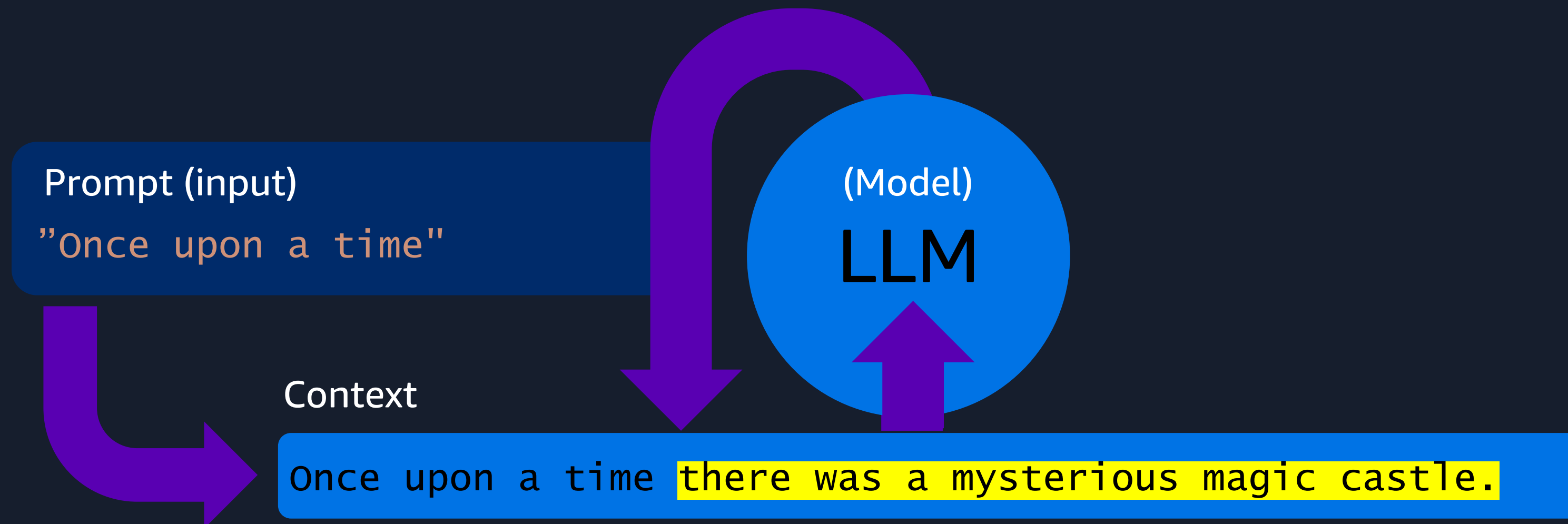
# Generating text with transformers



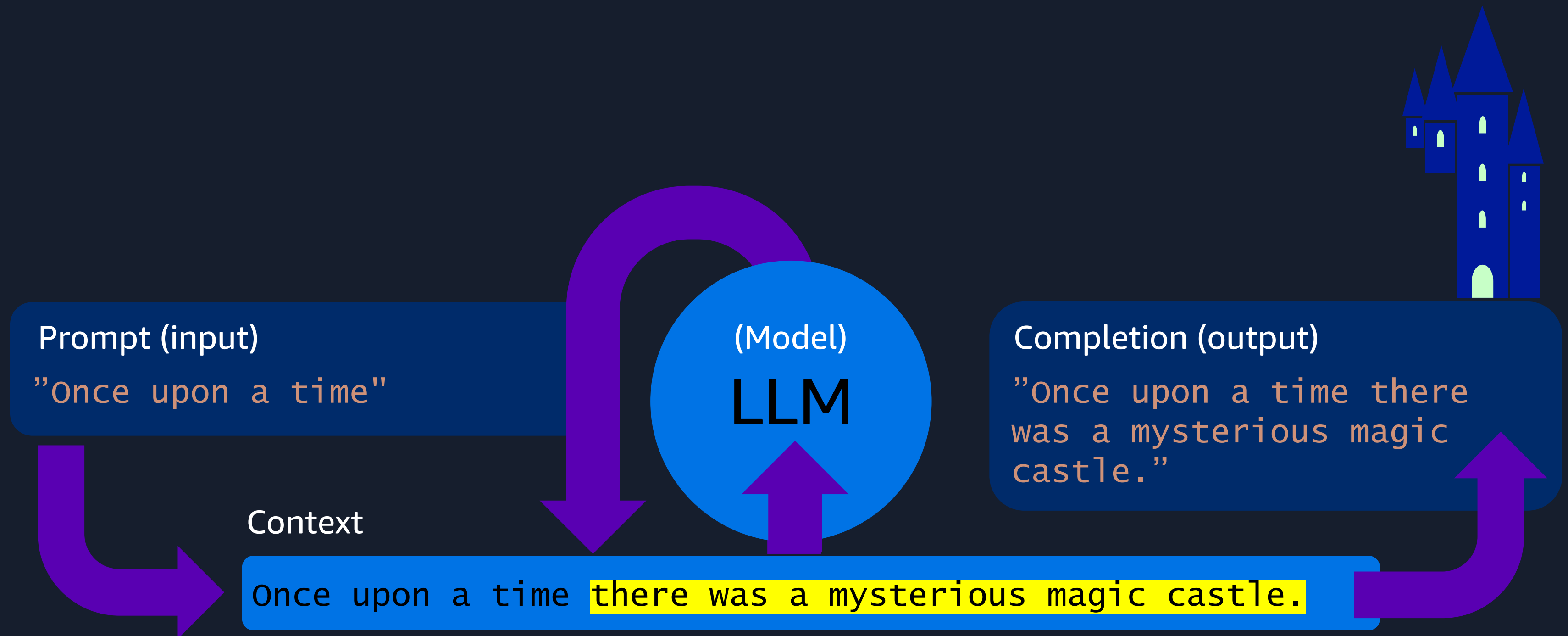
# Generating text with transformers



# Generating text with transformers



# Generating text with transformers



# What are Foundation Models?



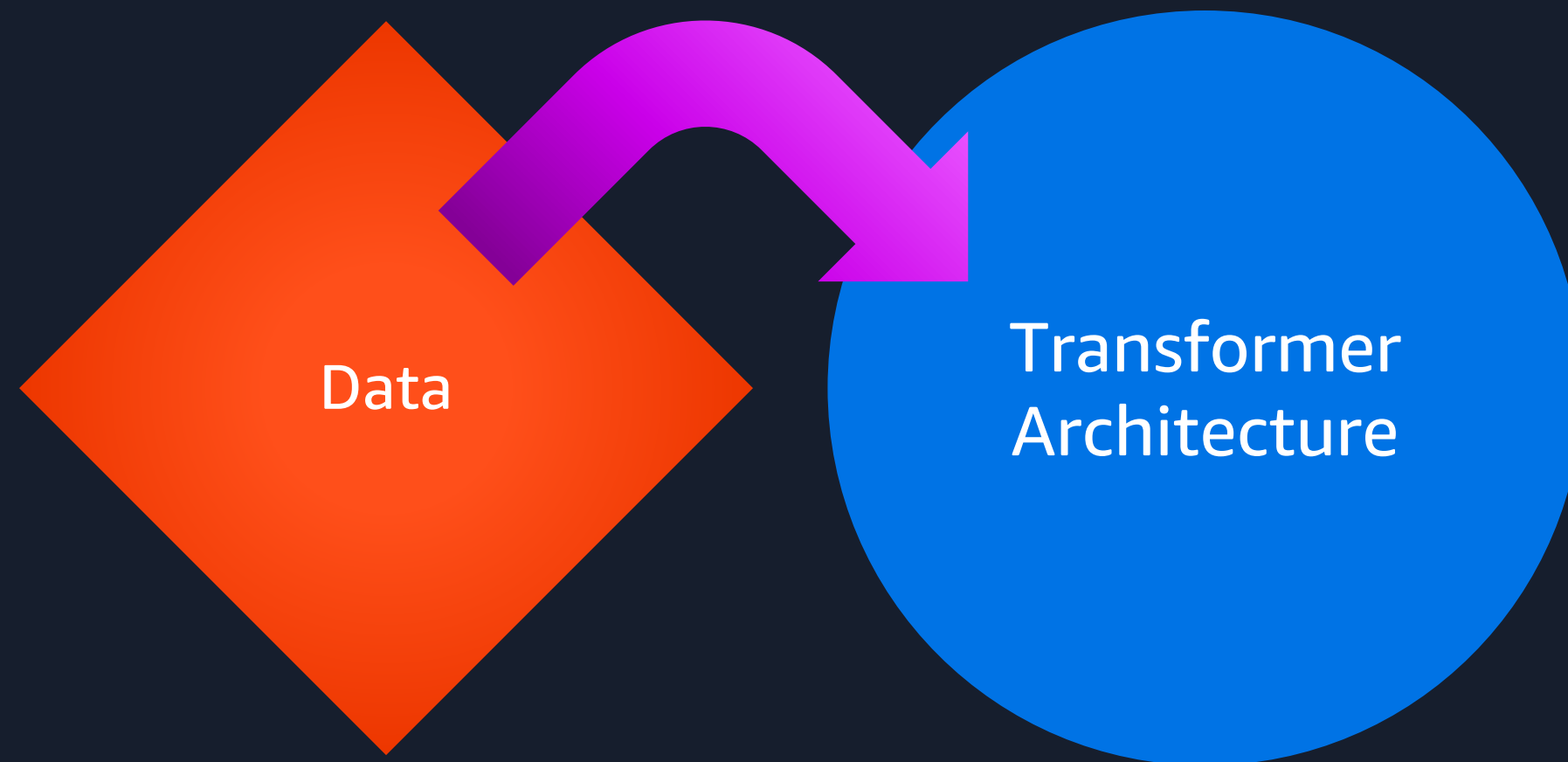
# Foundation models



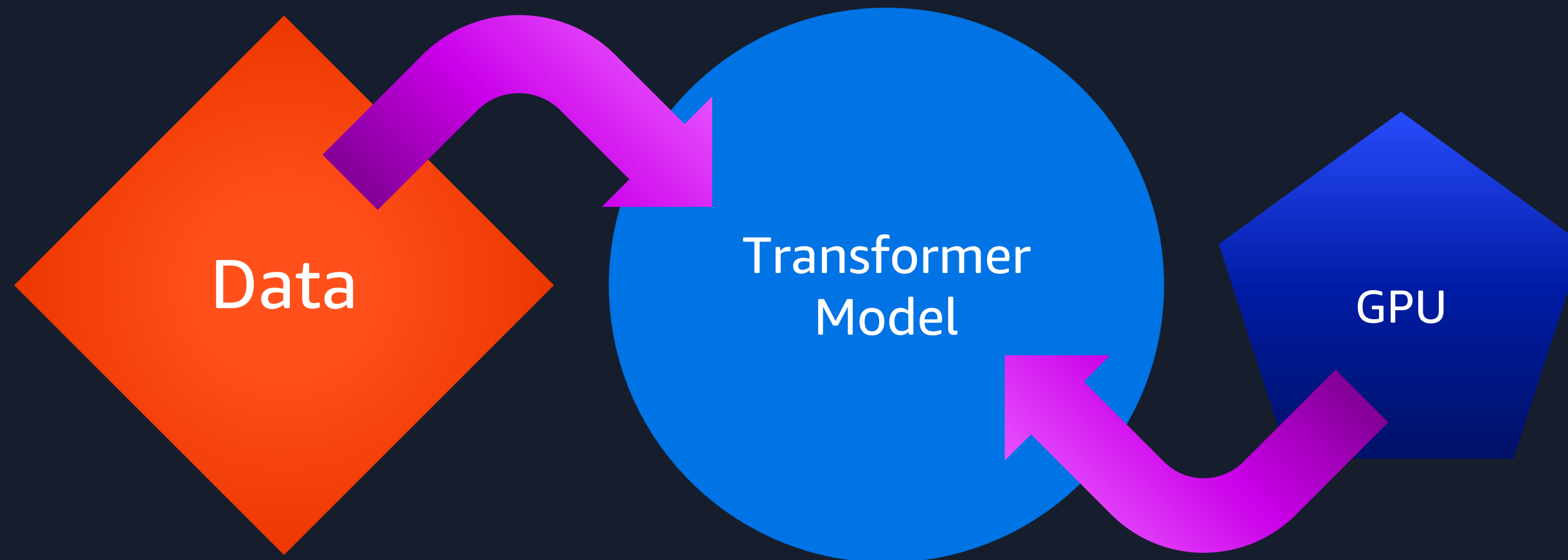
Transformer  
Architecture



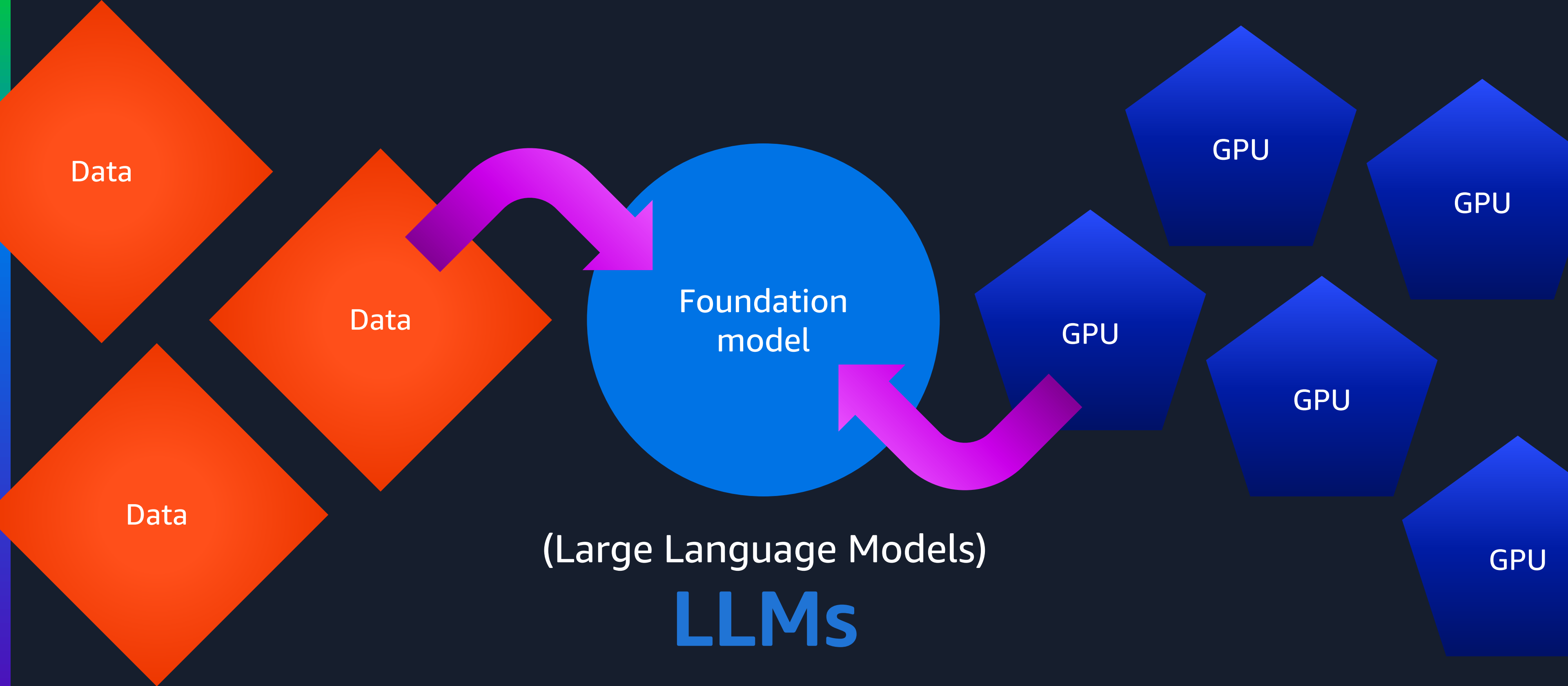
# Foundation models



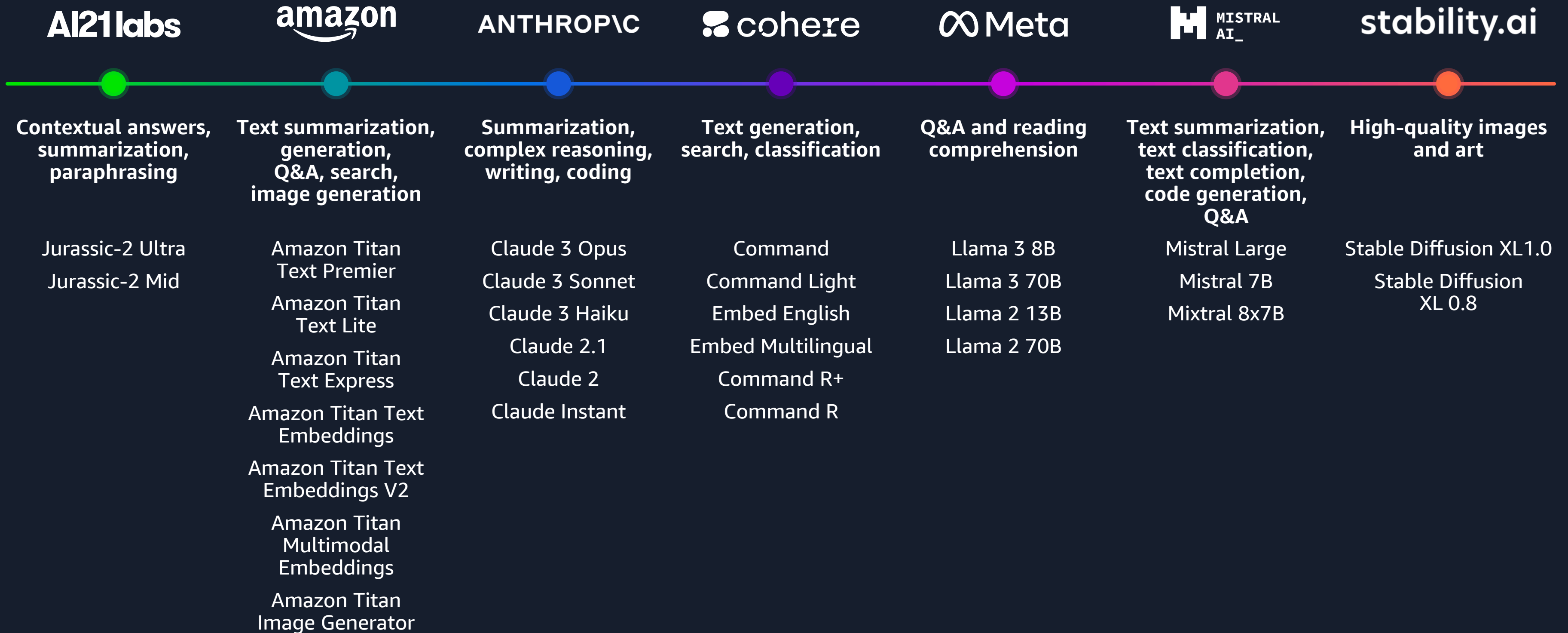
# Foundation models



# Foundation models



# Foundation Models



# Let's discuss about real things

# Wrap up

# Key takeaways

WE pretend Foundation Models to be smart as we do

Foundation Models have no memory, they are stateless

Foundation Models may be optimized for a specific prompt format

Foundations Models share common parameters and have their own

YOU can build smart applications by mixing different AI services and FMs

YOU can build smart applications by creating AI Agents that are specialized



# Feedback is a gift ❤️



# Thank you!

**François Bouteruche**

[linkedin.com/in/francoisbouteruche](https://www.linkedin.com/in/francoisbouteruche)



improve

---

17 / 06 / 2024

>> **AI CONF**

---

4ª EDIZIONE