

# LLM Engineering

## MASTER AI & LARGE LANGUAGE MODELS





WEEK 3 DAY 1

# Welcome to Open Source week

## What you can now do

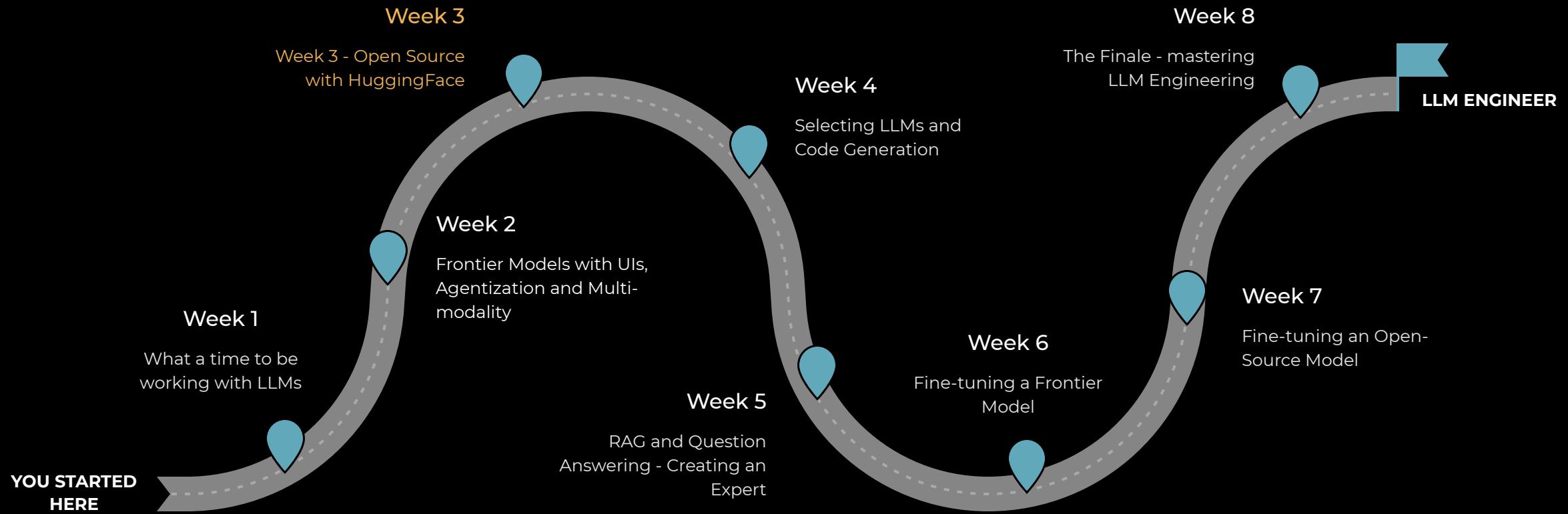
- Describe transformers and explain key terminology
- Confidently code with Frontier Model APIs
- Build a multi-model AI Assistant with Tools

---

## After today you will be able to

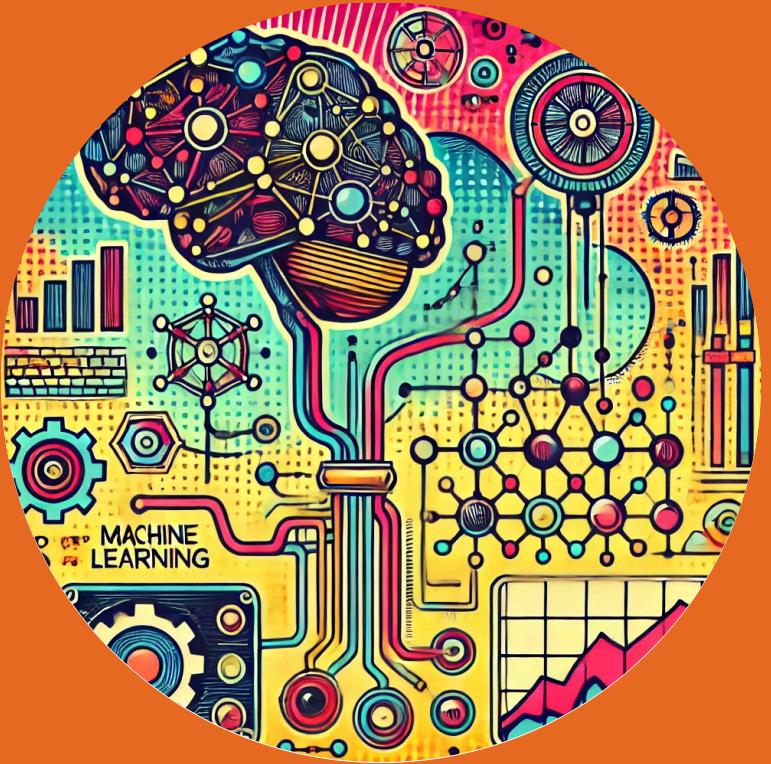
- Describe the HuggingFace platform and everything it offers
- Look at Models, Datasets and Spaces
- Use Google Colab to code on a high spec GPU runtime

# Reminder of the 8 weeks to mastery



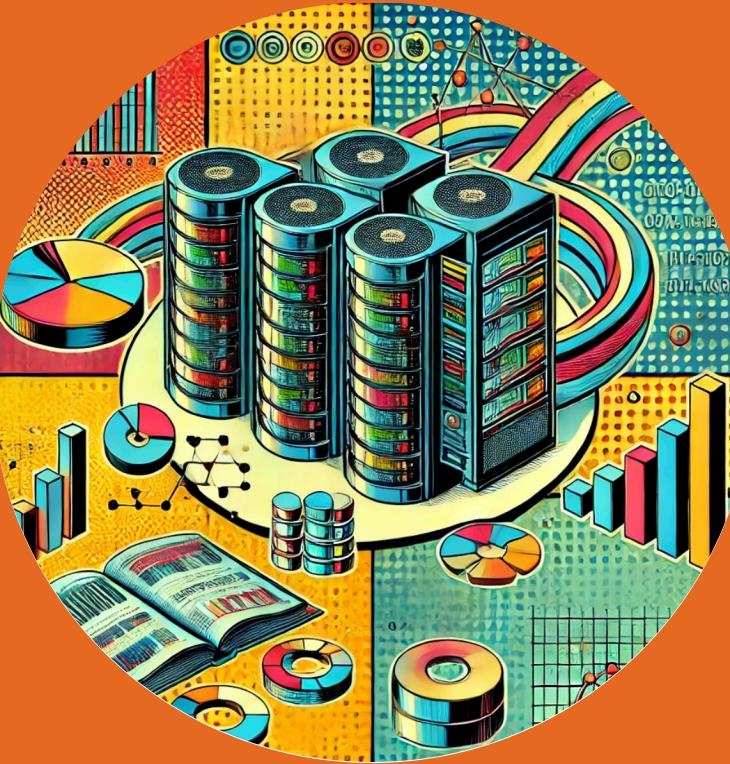
# HuggingFace Platform

The ubiquitous platform for LLM Engineers



Models

Over 800,000 Open Source Models of all shapes and sizes



Datasets

A treasure trove of 200,000 datasets

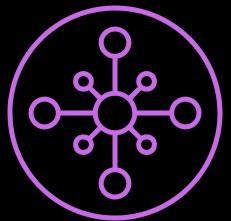


Spaces

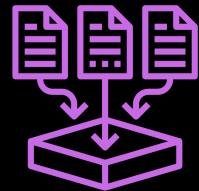
Apps, many built in Gradio, including Leaderboards

# HuggingFace Libraries

And the astonishing leg up we get from them



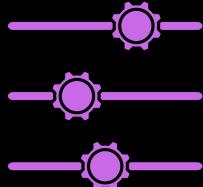
hub



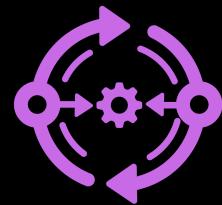
datasets



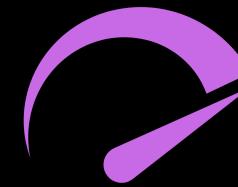
transformers



peft



trl



accelerate

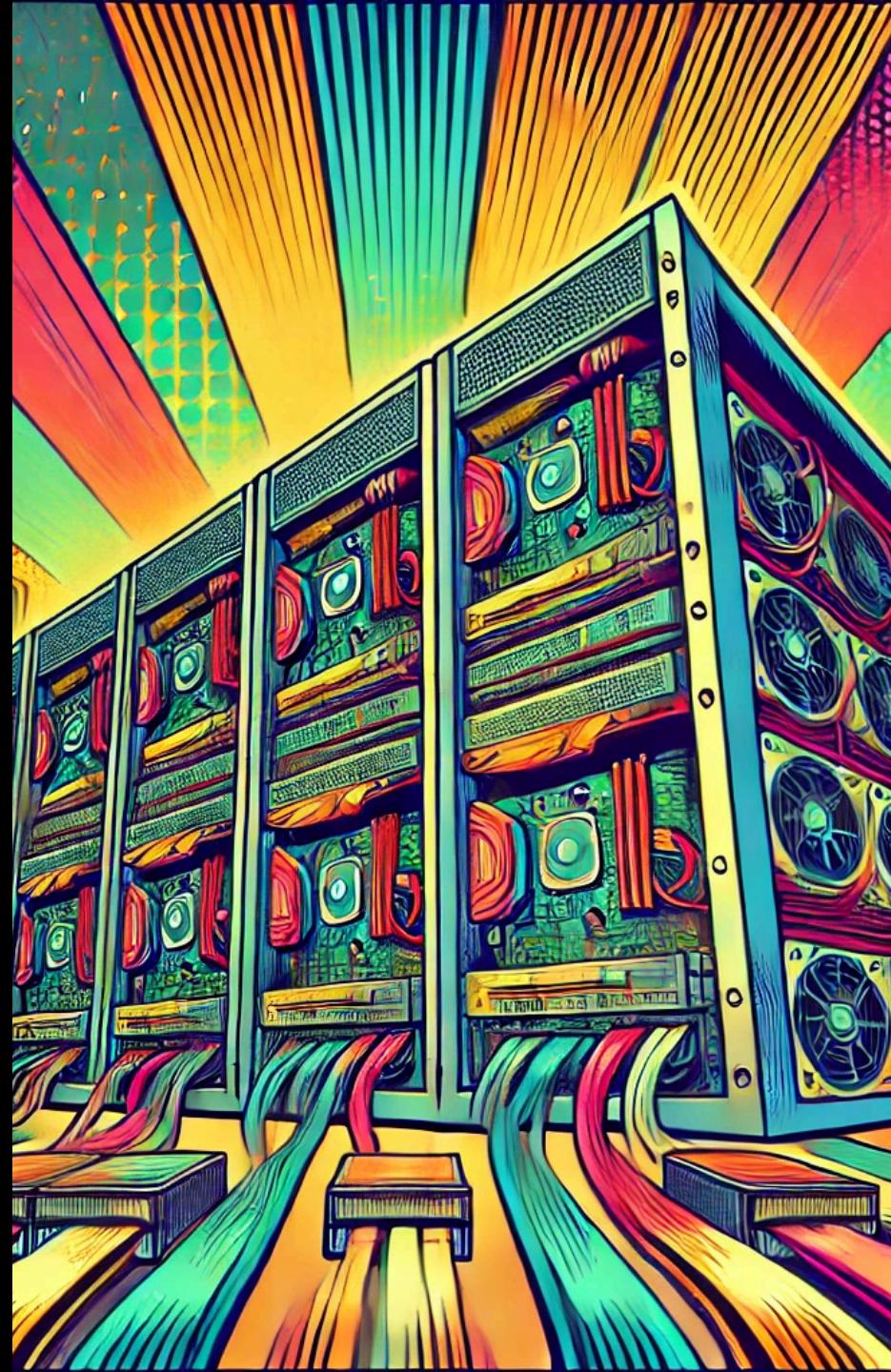
# Code on a powerful GPU box

## Google Colab

- Run a Jupyter notebook in the cloud with a powerful runtime
- Collaborate with others
- Integrate with other Google services

## Runtimes

- <sup>1</sup> | CPU based
- <sup>2</sup> | Lower spec GPU runtimes for lower cost
- <sup>3</sup> | Higher spec GPU runtimes for resource intensive runs



# GPU power in action

```
pipe = FluxPipeline.from_pretrained("black-forest-
labs/FLUX.1-schnell",
torch_dtype=torch.bfloat16).to("cuda")
generator = torch.Generator(device="cuda").manual_seed(0)
prompt = "A futuristic class full of students learning AI
coding in the surreal style of Salvador Dali"

image = pipe(
    prompt,
    guidance_scale=0.0,
    num_inference_steps=4,
    max_sequence_length=256,
    generator=generator
).images[0]

image.save("surreal.png")
```





## PROGRESS

# Ready for open-source coding

### **What you can now do**

- Confidently code with Frontier Models
  - Build a multi-modal AI Assistant with Tools
  - Navigate the HuggingPlace platform; run code on Colab
- 

### **After next time you will be able to run open-source models!**

- Understand the 2 different levels of HuggingFace API
- Use pipelines for a wide variety of AI tasks
- Use pipelines to generate text, images and audio