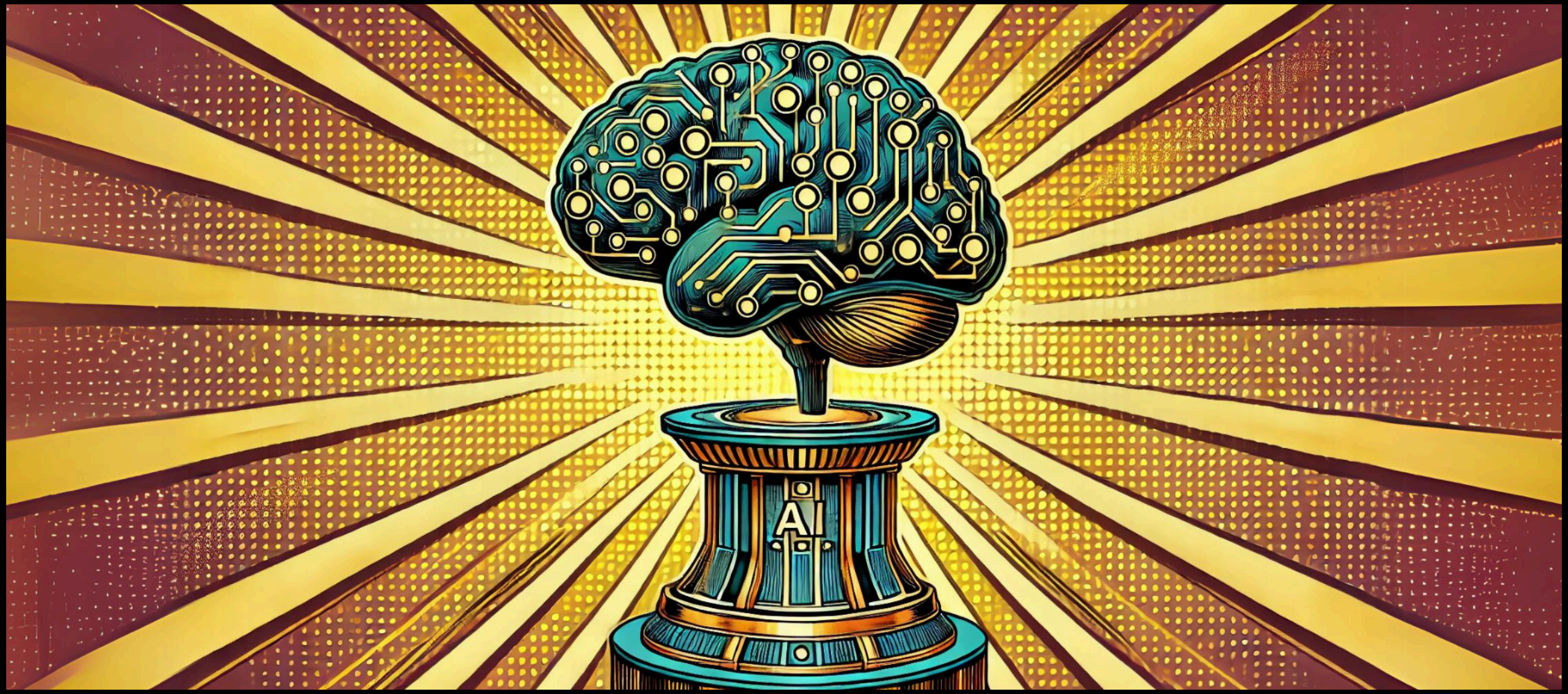
LLM Engineering

MASTER AI & LARGE LANGUAGE MODELS



WEEK 3 DAY 3

Time for the model

What you can now do

- Confidently code with Frontier Models
- Build a multi-modal AI Assistant with Tools
- Use HuggingFace pipelines and tokenizers

After today you will have essential new skills

- Work with HuggingFace lower level APIs
- Use HuggingFace models to generate text
- Compare the results across 5 open source models

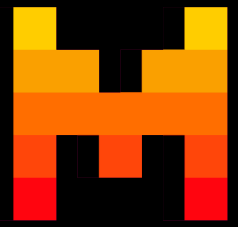
We will use these models

We will try Llama 3.1, Phi and Gemma, and you should try Mixtral and Qwen2

Llama 3.1 from Meta

Phi 3 from Microsoft

Gemma from Google



Mixtral from Mistral

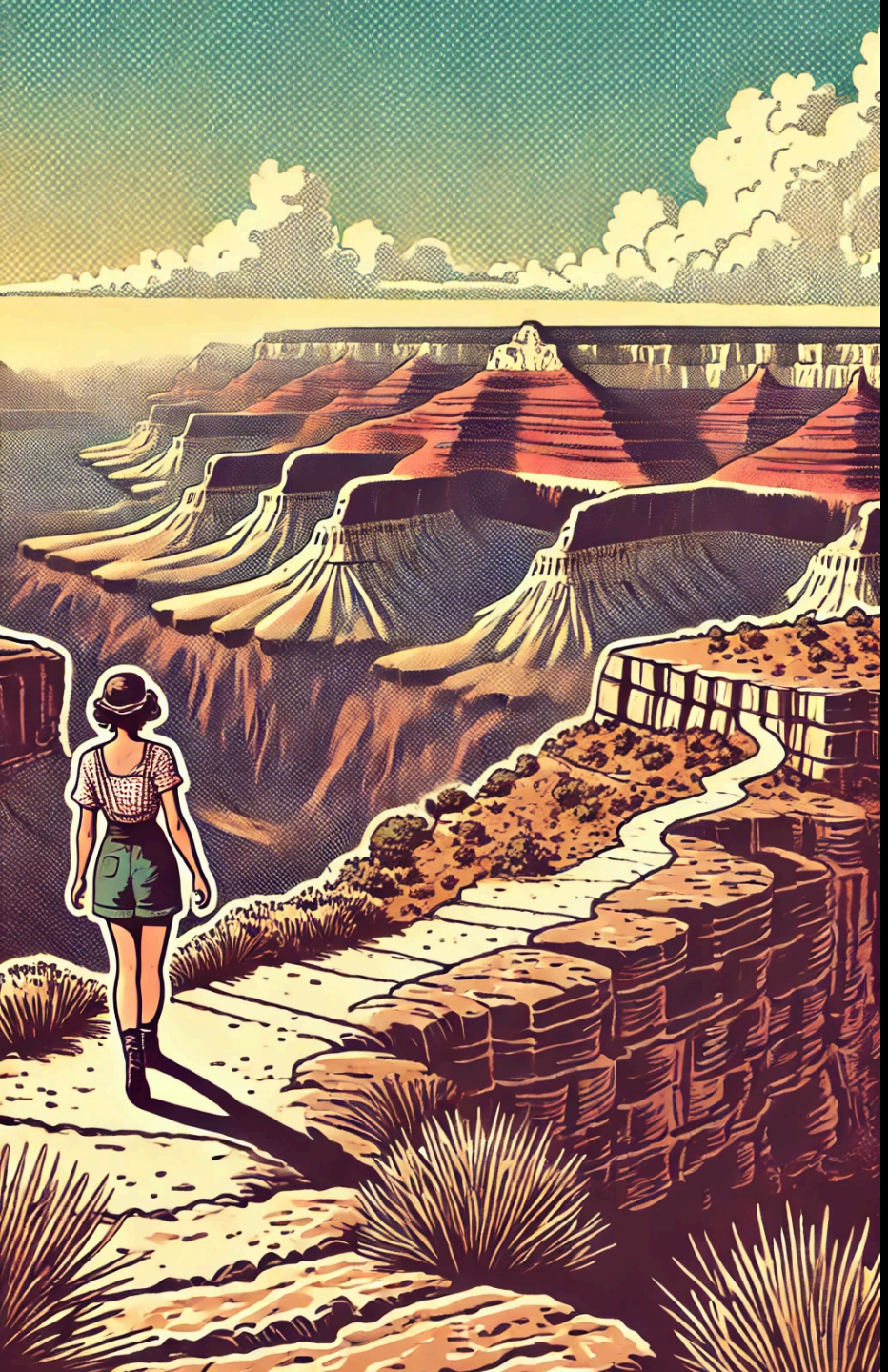
Qwen 2 from Alibaba Cloud

We will also cover...

Quantization

Model Internals

Streaming



WEEK 3 DAY 4

Practice, practice, practice

What you can now do

- Confidently code with Frontier Models
- Build a multi-modal AI Assistant with Tools
- Use HuggingFace pipelines, tokenizers and models

After the next session you will be experienced with models

- Confidently work with tokenizers and models
- Run inference on open-source models
- Implement an LLM solution combining Frontier and open-source models