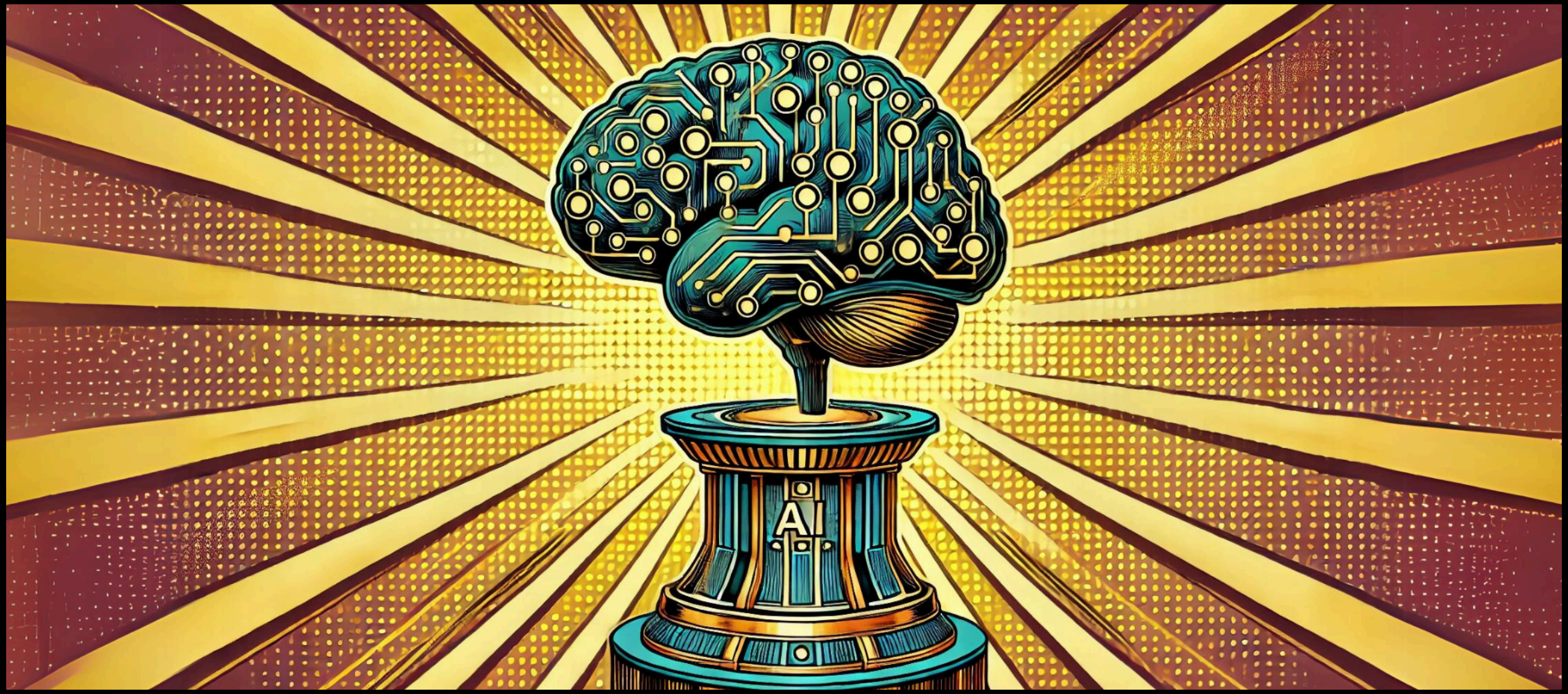
LLM Engineering

MASTER AI & LARGE LANGUAGE MODELS



PROGRESS

This week we get to training

What you can now do

- Generate text and code with Frontier Models including AI Assistants with Tools, and with open-source models with HuggingFace transformers
- Create advanced RAG solutions with LangChain
- Follow a 5 step strategy to solve problems, including dataset curation and making a baseline model with traditional ML and making a Frontier solution

---

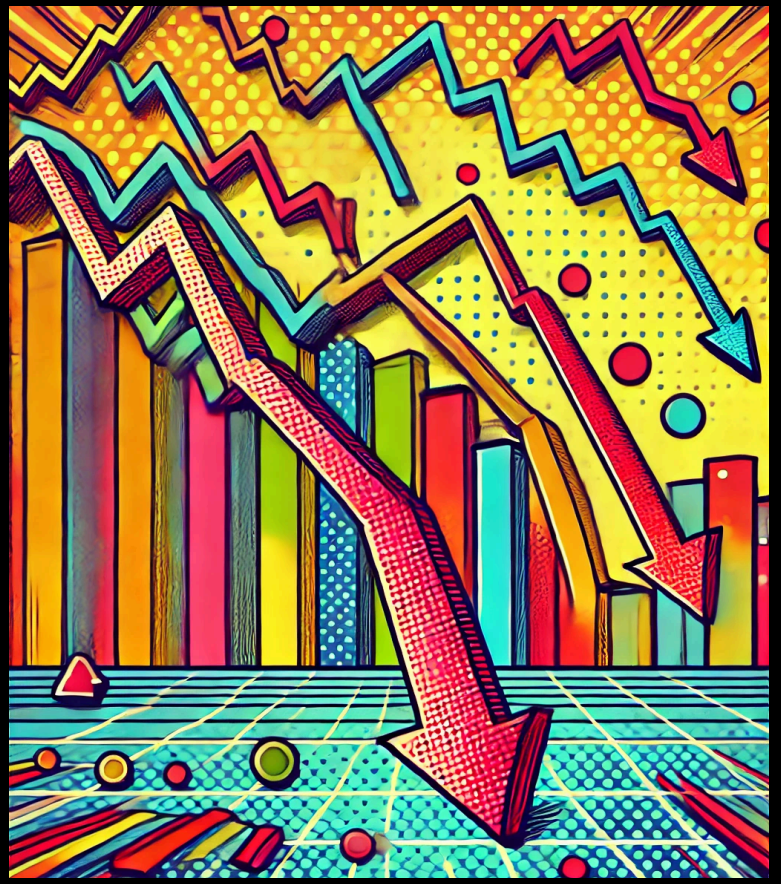
In a few short moments you'll finally be able to

- Understand the process for Fine-Tuning a Frontier model
- Create the fine-tuning dataset and run fine-tuning
- Test a fine-tuned Frontier model

Three Stages

To Fine-Tuning with OpenAI

Create Training Dataset in jsonl format and upload to OpenAI



Run training - training loss and validation loss should decreate

Evaluate results, tweak and repeat

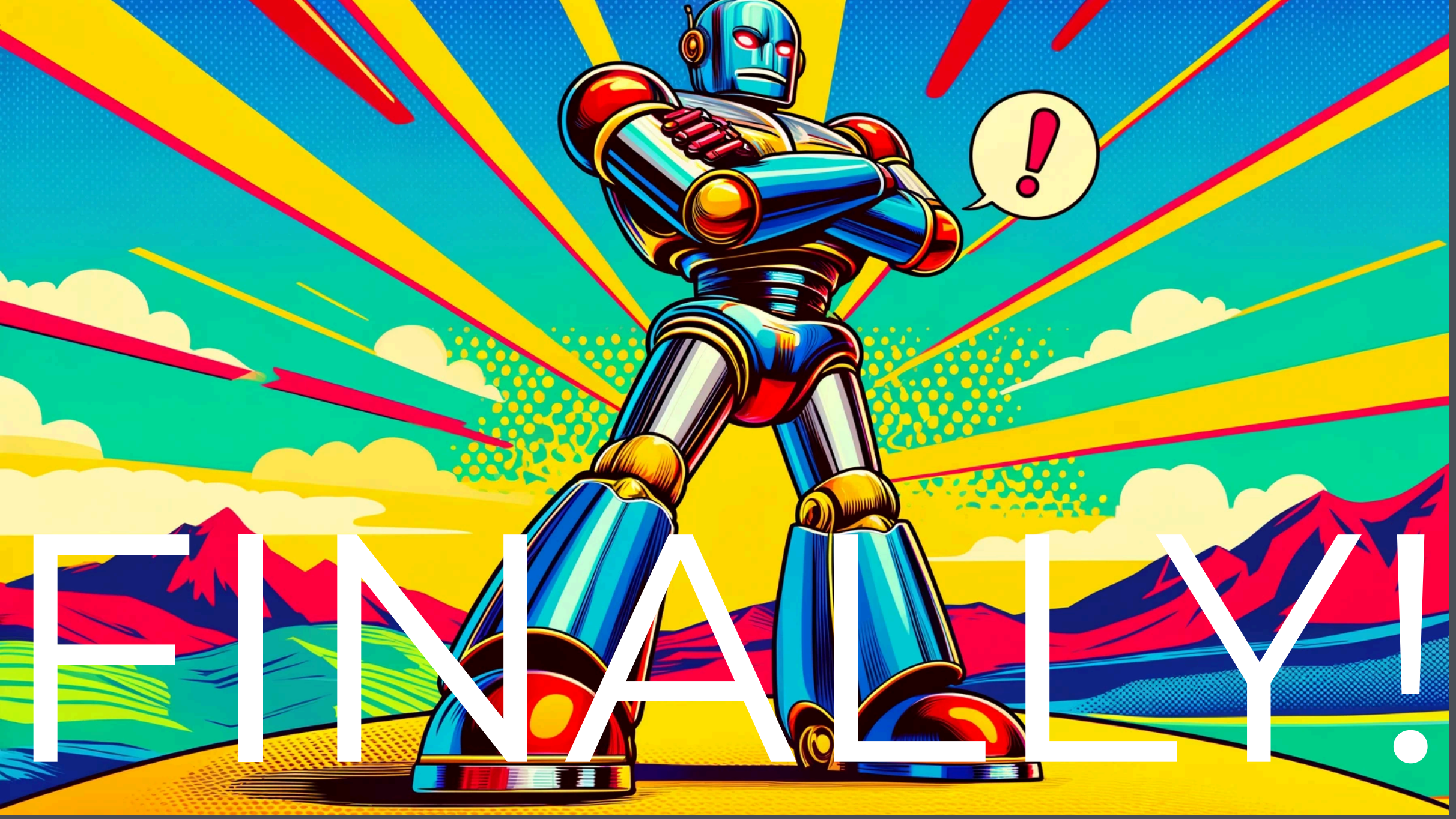
FINE-TUNING

Preparing the Data

OpenAI expects data in JSONL format

Rows of JSON each containing messages in the usual prompt format

```
{"messages" : [{"role": "system", "content": "You
estimate prices..."}]}
{"messages" : [{"role": "system", "content": "You
estimate prices..."}]}
{"messages" : [{"role": "system", "content": "You
estimate prices..."}]}
{"messages" : [{"role": "system", "content": "You
estimate prices..."}]}
{"messages" : [{"role": "system", "content": "You
estimate prices..."}]}
{"messages" : [{"role": "system", "content": "You
estimate prices..."}]}
```


!
FINALLY!

Average prediction error from our models

Well that was disappointing!

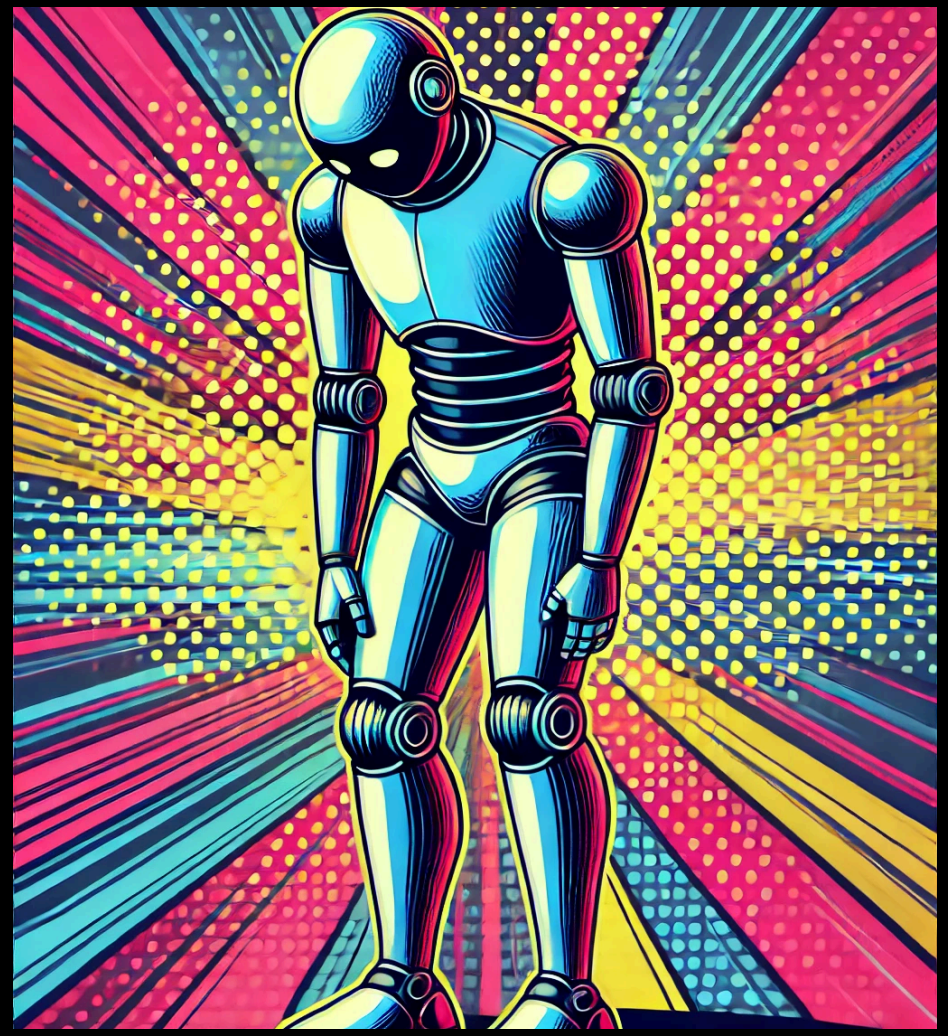
Key Objectives of Fine-Tuning for Frontier models

1. Setting style or tone in a way that can't be achieved with prompting
2. Improving the reliability of producing a type of output
3. Correcting failures to follow complex prompts
4. Handling edge cases
5. Performing a new skill or task that's hard to articulate in a prompt

---

A problem like ours doesn't benefit significantly from Fine Tuning

- The problem and style of output can be clearly specified in a prompt
- The model can take advantage of its enormous world knowledge from its pre-training; providing a few hundred prices doesn't help
- **WEEK 6 CHALLENGE FOR YOU: Experiment with larger training sets and more prompt engineering and BEAT THE CURRENT BASELINE**



PROGRESS

75% TO LLM ENGINEER!!

What you can now do

- Generate text and code with Frontier Models including AI Assistants with Tools, and with open-source models with HuggingFace transformers
- Create advanced RAG solutions with LangChain
- Follow a 5 step strategy to solve problems, including dataset curation and making a baseline model with traditional ML and making a Frontier solution, and fine-tuning Frontier Models

---

Next week we embark on a whole new adventure. You'll be able to

- Explain LoRA for fine-tuning Open Source models
- Describe Quantization and QLoRA
- Select a Base Model that we will use to compete with the Frontier