# Week 14: Fine-tuning & Post-Training of Language Models

Lecturer: Jordan Hill

GOVERNMENT OF WESTERN AUSTRALIA
TAFE WA North Metropolitan

# Learning Objectives

- Understand the concept of fine-tuning (Transfer Learning) and post-training in LLMs.
- Explore how Hugging Face Trainer API facilitates supervised fine-tuning.
- Segue into practical hands-on fine-tuning of an LLM.
- Visualize how LLMs can store and adapt facts through fine-tuning.

# From Pre-trained to Fine-tuned Models

- **Pre-trained Models (Foundation Models):**
  - Large language models trained on massive datasets.
  - General knowledge and language understanding.
  - Examples: GPT, BERT, LLaMA.
- **Fine-tuning / Post-training:**
  - Adapt the model to specific tasks or domains.
  - Use supervised learning on task-specific data.
  - Improves performance in specific applications like sentiment analysis, medical diagnosis, or customer support.

*Illustration: Foundation model + domain-specific data → Fine-tuned model*

# Why Fine-tune?

- Specialize general knowledge for niche tasks.
- Reduce the need for training models from scratch.
- Achieve better accuracy with less data.
- Save computational resources.

*Question:* How do we efficiently adapt an existing model?

# Introduction to Hugging Face Trainer API

- A high-level interface for training, evaluation, and saving models.
- Supports Transfer Learning / Fine-tuning with few lines of code.
- Handles data loading, tokenization, batching, optimization, and evaluation.

# Demo Focus: Supervised Fine-tuning (SFT) of a Language Model

- Use small domain-specific datasets (e.g., customer reviews, legal texts).
- Leverage Hugging Face’s pre-trained models.
- One code example can adapt a base model to your data.

# Before the Demo: Understand the Key Concepts

- **Transfer Learning:** Starting from a pre-trained model and further training it.
- **What is SFT?** Supervised Fine-Tuning, using labeled data.
- **Post-Training:** Additional training or adjustment after base training.

## Video 1: How might LLMs store facts | DL7 — Theory

<iframe width="100%" height="600" src="https://www.youtube.com/embed/9-Jl0dxWQs8" title="How might LLMs store facts | DL7" frameborder="0" allowfullscreen></iframe>

## Video 2: How AI Models Learn, Explained Visually — Training

<iframe width="100%" height="600" src="https://www.youtube.com/embed/NrO20Jb-hy0" title="How AI Models Learn, Explained Visually" frameborder="0" allowfullscreen></iframe>

# Break

## Let's take a 30-minute break before moving to the practical session

- Stretch your legs
- Grab a coffee or water
- Process what we've learned so far

# Transition to Practical: Using Hugging Face's Trainer API

- We'll load a pre-trained model.
- Fine-tune it on a small dataset.
- Observe how the model adapts to new data.
- Think of this as the "post-training" step: embedding new facts or tasks.

## Next: Hands-On Demo

**Follow along with the Hugging Face course link:**

Hugging Face LLM Course - Chapter 3

- Setup your environment.
- Load a base model.
- Prepare a small dataset.
- Run the Trainer API for fine-tuning.
- Observe the training logs and results.

(Note: The code snippets will be provided in the demo session, or I can prepare an example script here if desired.)

## Summary

- Fine-tuning (SFT) adapts foundation models for specific tasks.
- Hugging Face provides accessible tools to perform transfer learning.
- Visualizations help us understand how models store facts and adapt during post-training.
- The next step involves hands-on practice with real datasets.

# Questions?

- How can fine-tuning improve your AI applications?
- What challenges might you face when adapting large models?
- How does understanding these updates help in your AI projects?

# Let's get started!

*Prepare your environment and laptop for the demo.*

*Make sure your Hugging Face account is ready.*