

# Generative AI with LLMs

---

In Generative AI with Large Language Models (LLMs), you'll learn the fundamentals of how generative AI works, and how to deploy it in real-world applications.

By taking this course, you'll learn to:

- Deeply understand generative AI, describing the key steps in a typical LLM-based generative AI lifecycle, from data gathering and model selection, to performance evaluation and deployment
- Describe in detail the transformer architecture that powers LLMs, how they're trained, and how fine-tuning enables LLMs to be adapted to a variety of specific use cases
- Use empirical scaling laws to optimize the model's objective function across dataset size, compute budget, and inference requirements
- Apply state-of-the-art training, tuning, inference, tools, and deployment methods to maximize the performance of models within the specific constraints of your project
- Discuss the challenges and opportunities that generative AI creates for businesses after hearing stories from industry researchers and practitioners

Developers who have a good foundational understanding of how LLMs work, as well the best practices behind training and deploying them, will be able to make good decisions for their companies and more quickly build working prototypes. This course will support learners in building practical intuition about how to best utilize this exciting new technology.

## Week 1

Generative AI use cases, project lifecycle, and model pre-training

### Learning Objectives

- Discuss model pre-training and the value of continued pre-training vs fine-tuning
- Define the terms Generative AI, large language models, prompt, and describe the transformer architecture that powers LLMs
- Describe the steps in a typical LLM-based, generative AI model lifecycle and discuss the constraining factors that drive decisions at each step of model lifecycle
- Discuss computational challenges during model pre-training and determine how to efficiently reduce memory footprint
- Define the term scaling law and describe the laws that have been discovered for LLMs related to training dataset size, compute budget, inference requirements, and other factors

[Lab 1 - Generative AI Use Case: Summarize Dialogue](#)

[Week 1 quiz](#)

## Week 2

Fine-tuning and evaluating large language models

## Learning Objectives

- Describe how fine-tuning with instructions using prompt datasets can improve performance on one or more tasks
- Define catastrophic forgetting and explain techniques that can be used to overcome it
- Define the term Parameter-efficient Fine Tuning (PEFT)
- Explain how PEFT decreases computational cost and overcomes catastrophic forgetting
- Explain how fine-tuning with instructions using prompt datasets can increase LLM performance on one or more

[Lab 2 - Fine-tune a generative AI model for dialogue summarization](#)

[Week 2 quiz](#)

## Week 3

Reinforcement learning and LLM-powered applications

### Learning Objectives

- Describe how RLHF uses human feedback to improve the performance and alignment of large language models
- Explain how data gathered from human labelers is used to train a reward model for RLHF
- Define chain-of-thought prompting and describe how it can be used to improve LLMs reasoning and planning abilities
- Discuss the challenges that LLMs face with knowledge cut-offs, and explain how information retrieval and augmentation techniques can overcome these challenges

[Lab 3 - Fine-tune FLAN-T5 with reinforcement learning to generate more-positive summaries](#)

[Week 3 Quiz](#)