☰ Menu

Week 2 Resources

Below you'll find links to the research papers discussed in this weeks videos. You don't need to understand all the technical details discussed in these papers - **you have already seen the most important points you'll need to answer the quizzes** in the lecture videos.

However, if you'd like to take a closer look at the original research, you can read the papers and articles via the links below.

Generative AI Lifecycle

- **Generative AI on AWS: Building Context-Aware, Multimodal Reasoning Applications** ↗ - This O'Reilly book dives deep into all phases of the generative AI lifecycle including model selection, fine-tuning, adapting, evaluation, deployment, and runtime optimizations.

Multi-task, instruction fine-tuning

- **Scaling Instruction-Finetuned Language Models** ↗ - Scaling fine-tuning with a focus on task, model size and chain-of-thought data.

- **Introducing FLAN: More generalizable Language Models with Instruction Fine-Tuning** ↗ - This blog (and article) explores instruction fine-tuning, which aims to make language models better at performing NLP tasks with zero-shot inference.

Model Evaluation Metrics

- **HELM - Holistic Evaluation of Language Models** ↗ - HELM is a living benchmark to evaluate Language Models more transparently.

- **General Language Understanding Evaluation (GLUE) benchmark** ↗ - This paper introduces GLUE, a benchmark for evaluating models on diverse natural language understanding (NLU) tasks and emphasizing the importance of improved general NLU systems.

- **SuperGLUE** ↗ - This paper introduces SuperGLUE, a benchmark designed to evaluate the performance of various NLP models on a range of challenging language understanding tasks.

- **ROUGE: A Package for Automatic Evaluation of Summaries** ↗ - This paper introduces and evaluates four different measures (ROUGE-N, ROUGE-L, ROUGE-W, and ROUGE-S) in the ROUGE summarization evaluation package, which assess the quality of summaries by comparing them to ideal human-generated summaries.