**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.

**Multi-task, instruction fine-tuning**

* [**Scaling Instruction-Finetuned Language Models**](https://arxiv.org/pdf/2210.11416.pdf) - 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**](https://ai.googleblog.com/2021/10/introducing-flan-more-generalizable.html) - 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**](https://crfm.stanford.edu/helm/latest/) - HELM is a living benchmark to evaluate Language Models more transparently.
* [**General Language Understanding Evaluation (GLUE) benchmark**](https://openreview.net/pdf?id=rJ4km2R5t7) - 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**](https://super.gluebenchmark.com/) - 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**](https://aclanthology.org/W04-1013.pdf) - 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.
* [**Measuring Massive Multitask Language Understanding (MMLU)**](https://arxiv.org/pdf/2009.03300.pdf) - This paper presents a new test to measure multitask accuracy in text models, highlighting the need for substantial improvements in achieving expert-level accuracy and addressing lopsided performance and low accuracy on socially important subjects.
* [**BigBench-Hard - Beyond the Imitation Game: Quantifying and Extrapolating the Capabilities of Language Models**](https://arxiv.org/pdf/2206.04615.pdf) - The paper introduces BIG-bench, a benchmark for evaluating language models on challenging tasks, providing insights on scale, calibration, and social bias.

**Parameter- efficient fine tuning (PEFT)**

* [**Scaling Down to Scale Up: A Guide to Parameter-Efficient Fine-Tuning**](https://arxiv.org/pdf/2303.15647.pdf) - This paper provides a systematic overview of Parameter-Efficient Fine-tuning (PEFT) Methods in all three categories discussed in the lecture videos.
* [**On the Effectiveness of Parameter-Efficient Fine-Tuning**](https://arxiv.org/pdf/2211.15583.pdf) - The paper analyzes sparse fine-tuning methods for pre-trained models in NLP.

**LoRA**

* [**LoRA Low-Rank Adaptation of Large Language Models**](https://arxiv.org/pdf/2106.09685.pdf)- This paper proposes a parameter-efficient fine-tuning method that makes use of low-rank decomposition matrices to reduce the number of trainable parameters needed for fine-tuning language models.
* [**QLoRA: Efficient Finetuning of Quantized LLMs**](https://arxiv.org/pdf/2305.14314.pdf) - This paper introduces an efficient method for fine-tuning large language models on a single GPU, based on quantization, achieving impressive results on benchmark tests.

**Prompt tuning with soft prompts**

* [**The Power of Scale for Parameter-Efficient Prompt Tuning**](https://arxiv.org/pdf/2104.08691.pdf) - The paper explores "prompt tuning," a method for conditioning language models with learned soft prompts, achieving competitive performance compared to full fine-tuning and enabling model reuse for many tasks.