

WEEK 3 REPORT:

1. Research on Quantization

<https://learn.deeplearning.ai/courses/quantization-fundamentals/lesson/1/introduction>

- Linear Quantization
- Quanto Library from Hugging Face
- Model Loading in float16, bfloat16

2. Research on Prompt Engineering

<https://learn.deeplearning.ai/courses/prompt-engineering-with-llama-2/lesson/1/introduction>

- Learning prompt engineering of llama family models

3. Research on PEFT library

<https://huggingface.co/blog/peft>

- Learning PEFT for efficiently finetuning a large scale model
- Exploring LORA and its in-built functionality in PEFT

4. Synthetic Dataset Generation

- To generate synthetic data, currently learning about data generation pipeline from paper self instruct
- <https://arxiv.org/pdf/2212.10560>