THE TRANSFORMATIVE POWER OF GENERATIVE AI: REVOLUTIONIZING LANGUAGE MODELING WITH TRANSFORMER (SESSION 2)

Workshop Instructor:
Dr Christian Von Der Weth
Senior Lecturer & Assistant Dean, Communications
https://www.comp.nus.edu.sg/cs/people/chris/

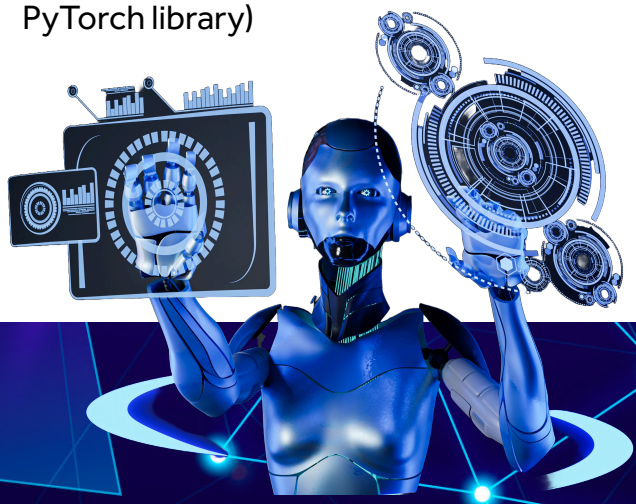
Large Language Models (LLMs) have revolutionized how machines understand and generate human language. All current LLMs are powered by the Transformer architecture. Its power lies in a mechanism called self-attention enabling a deep, nuanced grasp of context, meaning, and even subtle tone.

In this workshop, we will take a deep dive into the Transformer architecture and its use for training LLMs. Beyond the individual components of the Transformer architecture, we will also cover their challenges and recent advances, as well as common strategies for using Transformer-based LLMs in practice (e.g., fine-tuning, RAG)

The workshop is aimed at practitioners, AI engineers, data scientists, etc. looking for a better understanding of how LLMs work under the hood – beyond just using an existing LLM such as ChatGPT as a black box.

Recommended Skill Sets:

- Basic math knowledge (linear algebra, probability theory, calculus)
- Basic proficiency in Python (ideally also in the PyTorch library)



Workshop Outline:

- **Natural Language and its challenges**
- **A brief history to Language Models**
- **Word embeddings: How to represent words?**
- **Transformer Architecture**
  - **Attention**
  - **Encoder & Decoder**
  - **Positional Encoding**
  - **Masking**
- **Common implementations**
  - **Encoder-only: BERT**
  - **Encoder-decoder: T5, BART**
  - **Decoder-only: basically all LLMs**
- **Improving and customizing LLMs**
  - **Fine-tuning**
  - **Retrieval-Augmented Generation**
- **Overview to recent advances (Mixture-of-Experts, KV-Caching, Quantization, Distillation)**

Stay tuned for more information!