KOENIG
step forward

Building Transformer-Based Natural Language Processing Applications Training

Course Outline

Course Introduction

Introduction to Transformers

- Explore how the Transformer architecture works in detail:
  - Build the Transformer architecture in PyTorch.
  - Calculate the self-attention matrix.
  - Translate English to German with a pre-trained Transformer model.

Self-Supervision, BERT, and Beyond

- Learn how to apply self-supervised Transformer-based models to concrete NLP tasks using NVIDIA NeMo:
  - Build a text classification project to classify abstracts.
  - Build a NER project to identify disease names in text.
  - Improve project accuracy with domain-specific models.

Inference and Deployment for NLP

- Learn how to deploy an NLP project for live inference on NVIDIA Triton:
  - Prepare the model for deployment.
  - Optimize the model with NVIDIA TensorRT.
  - Deploy the model and test it.

Conclusion

- Review key learnings and answer questions.
- Learn how to set up your own environment and discuss additional resources and training.