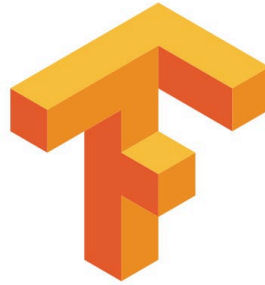Google Summer of Code has played a phenomenal role in my open source journey. I thank the TensorFlow and Keras team for granting me this opportunity. I am grateful to my mentors for their continuous guidance and encouragement. Without them this project would not have been possible. Throughout the summer I contributed across KerasNLP, KerasCore and KerasIO repositories with contributions ranging from tutorials to models and LLM infrastructure.

Goals

KerasNLP is a relatively new library. With the ongoing LLM Rush in the field of Generative AI, KerasNLP is moving fast. I was allowed to work on the tasks that align with the long term goals of the library. This gave me a plethora of opportunities to contribute. This also opened a door for me to contribute to the new multi-backend KerasCore library which was still a work in progress and not released. I principally worked on adding models, writing guides, porting KerasNLP infra to multi-backend and fixing bugs.

Mentors

I was mentored by [Matthew Watson](#) and [Chen Qian](#), two very cool people, who work in the Keras team at Google. Other than helping me with the technical nitty-gritties, they were very accommodating and supportive throughout my journey! I also got the opportunity to interact with other Googlers at Keras team including [Francois Chollet](#), [Jonathan Bischof](#) and [Ramesh Sampath](#).

My Contributions

KerasIO

Description	PR	Status
Add Semantic Similarity Tutorial	[keras-io/#1181](#)	Merged, Approved
Data Parallel Training Guide with KerasNLP	[keras-io/#1395](#)	Merged, Approved

Port Transformer Pretraining Guide to Keras Core	keras-io/#1438	Merged, Approved
Port GPT Text Generation Guide to Keras Core	keras-io/#1490	Under Review

KerasNLP : GPTNeoX

The purpose of this project was to add the GPTNeoX model by EleutherAI to KerasNLP. It is an autoregressive language model trained on the Pile and is a particularly powerful few-shot reasoner. There are various suites of checkpoints (GPT-NeoX-20-B , pythia) freely and openly available to the public through a permissive license. The model ranks high in the OpenLLM leaderboard.

Description	PR	Status
GPT NeoX Backbone	keras-nlp/#1056	Merged, Approved
Add GPTNeoXPreprocessor	keras-nlp/#1093	Merged, Approved
Refactoring GPTNeoXAttention and RotaryEmbedding	keras-nlp/#1101	Merged, Approved
Add GPTNeoXCausalLMPreprocessor	keras-nlp/#1106	Merged, Approved
Add GPTNeoXCausalLM	keras-nlp/#1110	Merged, Approved

KerasNLP : LLaMA 2

Llama2 is a revolutionary LLM developed by Meta. It is currently on top of the OpenLLM leaderboard at time of writing this report.

Description	PR	Status
Add Llama Backbone	keras-nlp/#1203	Under Review

Layers API

Description	PR	Status
Generic RotaryEmbedding Layer	keras-nlp/#1180	Merged, Approved
Add rms_scaling in LayerNormalization layer	keras-core/#726	Merged, Approved

Ops API

Description	PR	Status
Add rsqrt to Keras-core Ops	keras-core/#708	Merged, Approved

Keras Core Porting

Description	PR	Status
Port BeamSampler to Keras Core	keras-core/#1181	Merged, Approved
Port GPTNeoX to Keras Core	keras-nlp/#1137	Merged, Approved
Add compute_output_shape to tokenizer	keras-nlp/#1166	Merged, Approved

Bug Fixes

PR	Status
keras-core/#1204	Merged, Approved
keras-nlp/#1211	Merged, Approved
keras-nlp/#1213	Merged, Approved
keras-io/#1414	Merged, Approved

What's left to do?

1. The Generate workflow for GPTNeoX is facing some bugs and still a work in progress.
2. Adding the Llama Model to KerasNLP is a much bigger project which I will continue working on after GSoC is over.

Challenges

1. Making KerasNLP's model functionality match with those of official implementation while following Keras Philosophy was challenging indeed. This required me to debug the model code step by step and monitor weights and outputs of every layer to ensure the final outputs match.
2. While I was doing my GSoC, the Keras Core was also a work in progress. Learning the new framework without documentation, and making models comply with the new multi-backend ecosystem was fun and challenging at the same time.