Akshay Bhat

+91 6363011529 | bhat7362.com@gmail.com | github/Killua7362 | Leetcode | LinkedIn

EDUCATION

NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA
Bachelor of Technology in Mechanical Engineering, 2019 – 2023

EXPERIENCE

MACHINE LEARNING INTERN | Unify Ivy Jun 2023 – Nov 2023

- Implemented and optimized **Audio Spectrogram transformer** and **Repnet** model for mobile devices using graph compiler.
- Implemented **torch mobile** support for Ivy's proprietary autotuner and added unit tests for it.
- Created a tool to expand variable arguments using Abstract syntax trees for torch compiler.

MACHINE LEARNING RESEARCH INTERN | HALE May 2022 – Aug 2022

- Implemented a domain-specific pre-trained language model **BERT** for classifying self-reporting tweets as potential COVID-19 cases.
- Conducted experiments with various transformers on a dataset comprising over 18,000 tweets, addressing data imbalance through various augmentation techniques and improved the model by 9% from the base model.

PROJECTS

COMPANION | Code | Torch, Transformers, LangChain, ONNX, Redis

- Developed a data analyst companion for Yelp dataset which comprises 15,000 data points. using RAG architecture, incorporating a series of Gemma-7b models to trigger a chain of thought process.
- Integrated ElasticSearch to efficiently store embeddings, enhancing semantic search and experience.
- Implemented bark and whisper ai models for audio output and transcribing.
- Added a redis memory buffer and self query retriever algorithm using Langchain to persist chat history, facilitating seamless interaction.
- Utilized the ONNX backend compiler in conjunction with TensorRT runtime and static quantization to achieve decreased latency and a more compact memory footprint.

REMOVER | Code | Torch, OpenCV, Ultralystics, Transformers, bitsandbytes

- Implemented an image segmentation pipeline to remove the foreground from the image.
- Integrated a two-step segmentation approach by predicting the bounding box prompt for SAM model using YOLOv8 model.
- Employed a Stable Diffusion X1 model with ControlNet to seamlessly fill a segmented part of the image.
- Fine Tuned the pipeline on Stanford Background Dataset of over 4000 outdoor scene images with mixed precision training for reduced training time.

JAILBREAK | Code | Transformers, GPTQ, React, Flask, Jenkins, GCP, Kubernetes, Terraform

- Implemented a black-box prompt attacker architecture to get uncensored output from a LLAMA2 model.
- Utilized Mistral-7B as an attacker to predict prompts for jailbreaking the target model and employed the MAMBA model as a judge for calculating loss between the two models.
- Created a robust backend using flask and frontend using reactjs for seamless user experience.
- Deployed the architecture on GCP using Kubernetes and docker for scalability and reusability.
- Added github workflow to trigger jenkins workflow for continuous deployment of the architecture.

MULTIMODAL CLASSIFIER | Code | Torch, PEFT, PySpark

- Implemented multi modal classification pipeline by fine tuning and optimizing BakLLaVA model.
- Utilized QLORA for fine tuning the model to decrease training time and reduce model weights
- Integrated PySpark to efficiently clean and process fakeddit dataset which comprises 15,000 images and captions of scraped reddit posts for fine tuning of the model.
- Leveraged reinforcement training with DPO technique to align the model with the classification loss.

Technical Skills

- Pytorch, Tensorflow, Jax, React.Js, NextJS, Node.JS, Express, Flask, redux, recoil, Vite, ThreeJS, GraphQL, MySQL, MongoDB, Redis, Jenkins
- Python, Java, JavaScript, TypeScript, SQL, HTML/CSS, C++
- Git, Linux, Docker, Vim, Terraform, Google Cloud platform, Kubernetes, Ansible
- Data Structures and Algorithms, Deep Learning, Machine Learning, Artificial Intelligence