Aryan Deshpande

Chennai, Tamil Nadu, India

deshpandearyan1@gmail.com — (+91) 89398084000 — linkedin.com/in/aryandeshpande

github.com/Aryan-Deshpande — twitter.com/aryanscript — aryan-deshpande.github.io/tech

SKILLS

Programming: Python, Golang, C, C++, JavaScript

Framework/Tools: Flask, NodeJS, Docker, Kubernetes, Microsoft Azure, Google Cloud Platform, AWS EC2, AWS S3, Github-Actions, MongoDB, MySQL, PostgreSQL, FireBase, LaTeX, ReactJS, Kubeflow, Kafka, Circle-CI

Machine Learning: Deep Learning (PyTorch, TensorFlow), CUDA, Keras, NLTK, OpenCV, Numpy, Pandas, Huggingface, CONDA, Seaborn, Matplotlib

WORK EXPERIENCE

Augrade

Mumbai, Remote

Machine Learning Engineer Intern **April 2023 - August 2023**

- Led efforts in **2D to 3D reconstruction**, utilizing **OpenCV** and **PyMesh** library to create a model reconstructing symmetrical objects.
- Conducted research on Generative models for reconstructing 3D complex shapes from 2D data, **fine-tuning GAN models** with custom outdoor housing datasets.
- Utilized EC2 instances to **train models on Nvidia's V100** Tensor Core GPUs, achieving successful performance on testing datasets.
- Research on environmental-physics based constraints with **3D room reconstruction**, achieving 72% accuracy with a small synthetic training dataset.
- Developed a **Selenium-based web scraper** to gather floor plans and housing information from houseplans.com for fine-tuning a Stable diffusion model in conjunction with Control-Net
- **Trained Low-Ranked Adaptation weights** for fine-tuning parts of the stable diffusion model that are important for generating 2D indoor room features, saving computational resources by 50%

Vigameq Consultancy Services

Bengaluru, Remote

Backend Engineer Intern **May 2022 - December 2022**

Project Link

- Developed a website using **Flask** and **psycopg2** that allows users to sign up for events, and event hosts to register events.
- Created **RESTAPIs** for user booking functionality, event registration, and data retrieval.
- Implemented **containerization**, scalability, and **automation using Kubernetes**, improving development, testing, and deployment processes.
- Used **Github Actions** for **Continuous Integration / Continuous Delivery pipelines** and Pytest for unit testing.
- Introduced **Nginx ingress controller for TLS Termination**, improving application security by allowing only HTTPS traffic and **enabling load balancing**.
- **Applied Role-based access control** auth to increase security and prevent unauthorized access to resources.
- Employed **PostgreSQL relational cloud database** hosted on **Heroku** to manage persistent and session data.

RESEARCH and PROJECTS

Code Debugging Using Large Language Model (GemmaDev)

*** Research Paper accepted to ICACIS 2024 Springer Conference ***

Project Link *Camera Ready Paper* **November 2023 - Current**

- **Fine-Tuned on Google's Gemma Large Language Model** using **Parameter Efficient Tuning** (PEFT)
- **Performed Benchmarking** using **Human-Eval by OpenAI**, obtaining model proficiency of pass@1 = 0.067073170
- **Implemented quantization** using **Low-Rank Adaptation/QLow-Rank Adaptation weights**, by loading the model in a 4-bit (4b) version using normal float-4 (NF4).
- Created Low-Rank Adaptation weights encompassing essential features necessary for fine-tuning the model, and hence improving model's accuracy for it's size.

- **Applied supervised fine-tuning** (SFT) using the **Hugging Face** library, leveraging **NVIDIA A100 Tensor Core GPU with 40GB of VRAM**.
- The final model performed comparably to Code-Parrot's "small model", highlighting that effective tuning on minimal compute proves sufficient for training a code task-specific Large Language Model.
- Developed a website to load models locally and an interface to inference using **VueJS**.

WattWise (Top 100 Finalist Google Solution Challenge Project)

Project Link **April 2023 - July 2023**

- This application was designed to **help citizens monitor electricity consumption**, **reduce your bills**, and **conserve energy**.
- Developed APIs to facilitate data input, storage, and retrieval for electricity consumption statistics.
- Implemented storage mechanisms to securely store user-provided appliance wattage and usage hour data for future reference using **Google's Firestore**
- Integrated **PyTorch LSTM model** for accurate predictions of future electricity usage in wattage.
- **Engineered API functionalities** for **planning energy usage** for the user, adjusting usage to **enhance cost-effectiveness and sustainability**.
- Established a back-end server using **NodeJS**, **ReactJS**, and PyTorch (Python).
- Leveraged **Google's Firebase/FireAuth** for database management and authentication processes.

Latent Diffusion Model for UX Design Generation

Project Link **January 2023 - April 2023**

- **Latent Diffusion AI** (Text to Image model) with a streamlined UI interface that allows users to input text prompts and obtain high-quality images as output.
- Designed and developed a **text-to-image multi-model** using **UNet auto-encoder** and **CLIP guidance** for accurate image generation.
- **Implemented Attention module in the UNet architecture** to **improve segmentation** and **feature aggregation**.

EDUCATION

Hindustan University, Chennai, Tamil Nadu **September 2020 - July 2024**

Degree: Bachelors of Technology - Computer Science and Engineering, **8th Semester**
Thesis: GemmaDev - Code debugging using Large Language Model
Grade: First class with Distinction, **CGPA: 9.5/10.0**

Relevant Coursework: Discrete Mathematics, Applied statistics, Machine Learning, Deep Learning, Computational Linguistics and NLP

HONORS and AWARDS

Google Solution Challenge ' 23 (Project Name: WattWise)
Winners Link **April 2023**

- Participated in Google's Solution Challenge '23, and came Top 100 globally out of thousands of participants

IBM Hackathon 2nd Place **August 2022**

- In-campus hackathon at Hindustan University conducted by IBM, 2nd Runner up

Dan Kohn Scholarship by the Cloud Native Foundation **June 2022**

- Scholarship to attend KubeCon online in 2022, given to only a selected few around the globe

COMMUNITY EXPERIENCE

Google Developers Club Hindustan University Chennai, Tamil Nadu
Core Team - Machine Learning **September 2022 - April 2024**

- Active team member at the Google Developers Club HITS campus.
- Organized and co-hosted a seminar on Deep Learning tailored for students, providing insights into the latest advancements in the field, and live demonstrations to illustrate key concepts.
- Mentoring students across departments and year of study.