

# Dhanush Kumar

[dhanushkmr.com](http://dhanushkmr.com) | [dhanushk@protonmail.com](mailto:dhanushk@protonmail.com)

+65 97990948 | [linkedin.com/in/dhanush-manogar](https://linkedin.com/in/dhanush-manogar) | [github.com/dhanushkmr](https://github.com/dhanushkmr)

## EDUCATION

---

### Singapore University of Technology and Design

Sep 2020 – May 2024

*Bachelor of Engineering (Computer Science and Design)*

## EXPERIENCE

---

### Machine Learning Engineer Intern

May 2023 – Sept 2023

*Handshakes by DC Frontiers*

- Experimented fine-tuning **Llama2-7B** with **QLoRa** and **Proximal Policy Optimization** for factuality alignment for **abstractive news summaries** using **Atomic Content Units** as the reward model.
- Implemented an **extractive summarization** model and **context-aware chunking** based on a research paper for news summaries. Conducted **blind-testing** to evaluate the effectiveness of the model.
- Explored **automatic summary evaluation** using **GPT4** to assess quality and coherence and factuality.

### Data Science Intern

Sep 2022 – Jan 2023 & May 2021 – Sep 2021

*Portcast.io - SGINNOVATE SUMMATION & INFINITY Programme*

- Enhanced data provider selection accuracy and saved costs by 30% through developing a technique using Geospatial Data Analysis to compare convex hull areas for over 600000 historical vessel routes.
- Increased the accuracy of **ETA prediction** models by 3-5% through the development of innovative new model features, leveraging concepts from **queuing theory**, and implementing **automated hyperparameter tuning**.
- Launched a live port congestion dashboard covering over 800 ports and 5000 vessels to clarify ETA predictions which was instrumental in communicating supply chain issues with clients and teams within the SLA of 24 hours.

### NLP Research Assistant

Feb 2022 – Feb 2023

*NVIDIA NVAITC Student Ambassador*

- Authored **"Improving Code Generation Behaviour Via Curriculum Training"** for ACL 2023.
- Implemented a custom **AST** parser to accurately detect and reduced code degeneration by 37%.

### Data Science Intern

Jan 2020 – Aug 2020

*Urban Redevelopment Authority*

- Introduced a comprehensive **Tableau** dashboard that streamlined the visualization and analysis of electrical and water consumption of over 600000 entities, identifying key outliers and enabling strategic policy recommendations based on trends discerned through **Gaussian Mixture Model** clustering analysis.
- Introduced an NLP pipeline to perform topic modelling on monthly public feedback and pigeonhole to relevant departments. Achieved a model accuracy of 71% using **Latent Dirichlet Allocation**.

## PROJECTS

---

### Writing a compiler in Rust vs. Scala

Jan 2024

- Rewrote a compiler originally written in **Scala3**, in **Rust** and compared the implementation differences of functional programming and imperative programming, and **LLVM** vs **JVM**.

### Information Retrieval for Finance - Handshakes | SUTD CAPSTONE

Sept 2023 – Present

- Led collaborative team project with Handshakes to retrieve key data from annual reports based on complex **natural language queries**, helping improve workflow efficiency, leveraging **MLOps** and end-to-end development.
- Supported integration of **semantic search** and **Retrieval Augmented Generation** models to boost query comprehension and relevance of results, in a user-friendly web-app.

### Poe Hackathon | 3rd Runner Up

Oct 2023

- Deployed a data analysis chatbot with over 50 monthly users which accepts csv data and can run code to produce tables and plots using **GPT4** which allows users to carry out **Exploratory Data Analysis**.

## TECHNICAL SKILLS

---

**Languages:** Python, Rust, Scala, Javascript, Java, C, MATLAB, SQL, HTML/CSS

**Technologies:** PyTorch, spaCy, scikit-learn, pandas, NumPy, Plotly, React, FastAPI, Selenium

**Developer Tools:** Git, Docker, AWS, Linux, Nvim