

# Shubham Gupta

SOFTWARE ENGINEER · ML ENGINEER

✉ shubhamga2208@gmail.com | 🌐 www.shubhamg.in | 📷 goodhamgupta | 📄 shubhamgupta2208

## Summary

Software Engineer with eight years of experience in developing and delivering advanced solutions into many business aspects. Excellent ability to pick up relevant technologies for the task and implement end-to-end solutions to generate business value.

## Work Experience

### Temus

Singapore

SOFTWARE ENGINEERING MANAGER, DATA AND AI

Oct. 2022 - Present

- Implemented a financial assistant with LLMs and RAG, merging insights from structured and unstructured data via SQL AST parsing. Enhanced accuracy and relevance to 70%, reduced due diligence time by 20%, and cut inference costs by 40% with an on-premises LLM for structured queries.
- Implemented a Thai Automatic Speech Recognition (ASR) system by integrating fine-tuned OpenAI Whisper with Voice Activity Detection and Speaker Diarization. Leveraged Llama for transcript correction, achieving a significant increase in intent detection accuracy from 11% to 17%.
- Developed LinguaForge, a multi-cloud platform using Kubernetes to manage batch jobs, train, and serve LLMs, reducing serving costs by 40%. Integrated SkyPilot, vLLM, and S-LoRA orchestrated LoRA Adapters for improved model performance.

### Aaqua

Singapore

APPLIED RESEARCH ENGINEER

Jun. 2021 - Sep. 2022

- Implemented a content-moderation system based on the OpenAI CLIP model to detect policy violating content on the platform, thereby decreasing manual moderation effort by 30%.
- Implemented a system to detect spam messages, based on the Microsoft MiniLM model and an open SMS spam dataset. The model was deployed using Sagemaker and FastAPI.
- Implemented a data platform to collect, transform and load data from multiple sources, using Airflow for orchestration and Datahub for data management.

### RADICALi

Singapore

SENIOR DATA SCIENTIST

Oct. 2019 - May. 2021

- Implemented an engine to identify similar documents and topics using LM Jelinek Mercer Smoothing, BERT and Query Re-ranking in Elasticsearch achieving a silhouette score of 0.45.
- Implemented an unsupervised summarization model to find missing topics between two documents using GPT-2 and documents-as-graphs, leading to accuracy of 65%. Distilled and Quantized model to a LSTM leading to an increase in inference speed by 20x.
- Implemented multi-label risk classification for regulations using MultiFit, achieving an average recall of 87% across major classes. Deployed end-to-end pipeline using Flask, AWS Sagemaker achieving P95 inference time of under one second.
- Implemented a Knowledge graph to identify key topics and entities between upcoming legislations, helping users discover and visualize hidden relationships and build custom ontologies. Deployed system using Flask, Neo4J and Streamlit.
- Implemented a system to identify and cluster metadata such as Entities, Obligations using USE Embeddings, Flask and FAISS.
- Implemented a serverless scraping architecture using Scrapy, AWS Lambda, and Fargate scraping ability and increasing efficiency by 45%.

### Scripbox

Bangalore, India

PRODUCT ENGINEER AND DATA SCIENTIST

Oct. 2016 - Oct. 2019

- Implemented a hierarchical forecasting pipeline to predict the revenue and the number of customers in the future months using Prophet and Redash. Forecasts have been 93% accurate.
- Implemented a Hierarchical Bayesian CLV model using BG/NBD and Gamma-Gamma models which helped identify and segregate customers thereby increasing revenue by 10%.
- Developed a document processing pipeline using the YOLO DL model with OCR to help extract information from images thereby improving the customer on-boarding process by 50%.
- Implemented a real-time event sourcing pipeline using Snowplow, Kinesis, and Redshift which has helped improve customer and application analytics.
- Implemented near-real-time data pipeline to massage and aggregate data from multiple sources using Amazon EMR, Hive, HDFS and Airflow. This project is currently in the process of being open-sourced.
- Implemented the portfolio management platform from scratch which displays investments made, return percentages, etc. in Elixir thereby increasing performance by 10x.
- Implemented a data and analytics system to aggregate marketing campaigns spends and customer metrics from social platforms thereby helping reduce costs by 6%.

### Zopper

Bangalore, India

SOFTWARE ENGINEER

Jan. 2016 - Oct. 2016

- Implemented a software database AESOP(LinkedIn) which helped process changes from one type of data store to another reliably. Ingested order processing efficiency by 13%.
- Implemented 'Catalog Extension' allowing retailers to source products from national distributors thereby increasing overall sales by 20%.

## Selected Projects

---

### Presentation on State-Space Models

[HTTPS://BIT.LY/MLSG-SSM](https://bit.ly/mlsg-ssm)

- Presented State-Space-Models, an alternative to attention based LLMs in the ML Singapore Meetup

Singapore

Feb. 2024

### End-to-End Attention based Image Captioning

[HTTPS://BIT.LY/ATTENTION-IMAGE-CAPTIONING](https://bit.ly/attention-image-captioning)

- Implemented as part of a course project at NUS, and finished in top 10% of the leaderboard in the Kaggle contest.

Singapore

Apr. 2021

### Segmentation via Hierarchical CLV Models

[HTTPS://BIT.LY/HIERARCHICAL-CLV](https://bit.ly/hierarchical-clv)

- Presented research on Hierarchical CLV, implemented at Scripbox, at DeepBayes.

Moscow, Russia

Apr. 2019

### Open-Source Contributor

[HTTPS://GITHUB.COM/GOODHAMGUPTA](https://github.com/goodhamgupta), [HTTPS://BIT.LY/ELIXIR-PACKAGES](https://bit.ly/elixir-packages)

- Open source contributor to Nx, OpenMined and Snowplow Analytics. Packages have over **200k** downloads

Singapore

Aug. 2016 - Present

## Education

---

### NUS(National University of Singapore)

MASTERS IN COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE

CGPA: 4.0/5.0

Singapore

Aug. 2020 - Dec. 2022

### Amrita School of Engineering

BACHELOR OF TECHNOLOGY IN COMPUTER SCIENCE

CGPA: 7.96/10.0

Bangalore, India

Aug. 2012 - May. 2016