

EDUCATION

Degree/Certificate	Institute/Board	CGPA/Percentage	Year
B.Tech	Indian Institute of Technology, Guwahati	7.32/10	2020–2024
Senior Secondary	CBSE Board	96.8%	2020
Secondary	CBSE Board	94.4%	2018

EXPERIENCE

- OTSUKA CORPORATION

Oct 2024 — Present

AI Engineering — Full Time

Tokyo, Japan

– Led bilingual, domain-adapted LLMs on **multi-node GPU infrastructure (H100, A100, A6000)**; built LoRA-based training and deployment and RAG for proprietary data.

– Orchestrated pretraining, SFT, instruction tuning, and RLHF (**PPO, DPO, ORPO**) using **Axolotl** and **LLaMA Factory**; compared **LoRA/QLoRA** vs. full fine-tuning to establish reproducible baselines.

– Set up unified evaluation with **DeepEval**; executed MT-Bench, TruthfulQA, and custom evaluators; Japanese-centric models reached 85+ composite scores.

– Architected enterprise-grade RAG chatbots with custom **MCP tools** and **Elasticsearch**, including database connectors, ingestion pipelines and Japanese-English analyzers.
- OTSUKA CORPORATION

June 2023 — July 2023

AI Engineering — Internship

Tokyo, Japan

– Implemented **“Generative Agents: Interactive Simulacra of Human Behavior”**; simulated multi-agent interactions via the Werewolves of Miller Hollow scenario.

– Built a real-time generative-agent pipeline (planning/acting/conversing) using **LangChain, GPT-3.5 Turbo**.

– Developed a business application enabling dynamic group conversations with **multiple AI agents**.

PROJECTS

- DATA-AUGMENTOR: LLM-DRIVEN SFT & ALIGNMENT DATASET BUILDER

2025

Group Project

GitHub

– Built a Python toolkit that converts Markdown notes into **SFT / RLHF datasets** with chunking, and LLM-based checks.

– Designed a concurrent pipeline with **specialized agents (Chunker, QA Generator, Validator, ORPO Generator)**.

– Enabled multi-turn/single-turn **QA generation and ORPO-style preference data** for fine-tuning custom domain-specific LLMs.
- STOCK ETL - DATA COLLECTION PIPELINE

2025

Personal Project

Github

– Built an **Airflow DAG** that scrapes Yahoo Finance’s “Most Active” list, capturing the top 30 tickers every day.

– Cleans, enriches, and upserts price, volume, and P/E data into **PostgreSQL** with conflict-aware logic.

– Delivers a fully automated **ETL pipeline for real-time market-data analytics**.

TECHNICAL SKILLS

- **Programming:** Python, C, C++
- **ML/AI:** NumPy, Pandas, Pytorch, Fine-Tuning, LangChain, LangGraph, RAG, MCP, Transformers, DeepLearning
- **Databases:** MySQL, MongoDB, Elasticsearch
- **MLOps:** MLflow, DeepEval, LangFuse, Apache Airflow, Azure AI Foundry
- **Miscellaneous:** Docker, Git, Prompt Engineering

ACHIEVEMENTS

- JLPT N4**, Certified basic proficiency across vocabulary/grammar, reading, and listening.

2025
- INTERNSHIP COORDINATOR**, Successfully co-ordinated the interview procedure for 300+ students

2023
- AISSCE 2020**, Scored 100% in English Core (top 0.1%)

2020

KEY COURSES TAKEN

- Undergrad: Introduction to Computing (Theory & Lab), Object Oriented Programming (OOP), Database Management Systems (DBMS), Data Structures & Algorithms

EXTRACURRICULARS

- PUBLIC SPEAKING**, Member, Anchoring Club
- EVENT MANAGEMENT**, Organised E-Cell events; hosted various talks