Soham Mistri

 | in | sohammistriind@gmail.com | +91 913 633 0070 | +91 825 036 7259

EDUCATION

Indian Institute of Technology, Bombay *(2019-2023)*
Bachelors of Technology in Computer Science with Hons., Minor in Machine Learning and Data Science | CGPA: 9.04/10

PROFESSIONAL EXPERIENCE

Applied Scientist *(Aug 2024- present)*
Microsoft India

◇ Worked on **prompt improvements** for the M365 Copilot Web Chat interface to enhance the user experience:

- Iterated changes on the **reasoning prompt** to increase the trigger rate of **Python Code Interpreter** (CI) tool for analytical, file creation and other use cases which require code. Improved trigger rate of the CI tool by about **100 %** which led to an increase in user satisfaction rate by about **150 %** in CI related use cases.
- Iterated changes on the **responding prompt** to introduce **numerical citations** for images and files created by the CI instead of rendering complete links, leading to a **40 %** improvement in P99 and P95 ChatLTR.
- Iterated changes on the **code generation prompt** for the CI tool to introduce the use of `plotly` library and help render **interactive charts** as part of the response and improve user experience.
- Drove the transition to **GPT-5**, worked on **prompt structure optimizations** that boosted **cache efficiency** and reduced ChatFTR metrics by **30%** across performance tiers.

◇ Developed **synthetic data generation pipelines** for fine-tuning reasoning models, creating comprehensive datasets with **queries, ground truth, and evaluation rubrics** across diverse use cases including *code generation, data analysis, web search, ambiguous queries, no-answer cases, and multi-site crawling*, with rigorous quality validation frameworks.

◇ Built **synthetic tenant generation systems** producing realistic enterprise environments with **chat messages, emails, meetings, and calendar events** for custom companies, implementing comprehensive quality validation to ensure production-ready data for agent training and evaluation.

◇ Conducted **training and grader ablation studies** across search and agent pipelines to validate synthetic data quality, implementing **advanced agentic graders** for the **Outlook Agent** that detect hallucinations and measure inbox coverage, uncovering critical limitations in RL training workflows and driving iterative data improvements.

Data Analyst II *(July 2023- July 2024)*
Walmart Global Tech India

◇ Worked in the **Smart Subs** team to **recommend substitute items** to the customer in case an item goes out of stock.

◇ Worked towards **incorporating user preferences** to **personalize substitute item recommendation**. Trained **XG Boost model** and **Neural Network model** to rerank top subs with respect to user preferences from past data

◇ Improved the **quantity ratio** computation logic for anchor item-substitute item pairs to take price of the item into consideration, leading to a **10 % reduction** in average price of substitutes recommended to customers.

INTERNSHIPS

Machine Learning Intern *Jan 2023-May 2023*
KnowDis Data Science

◇ Contributed to the data science team in developing advanced **deep learning models** for tasks such as **background removal** and **human body parsing**, as part of a deep fashion project.

◇ Finetuned **Visual Tansformer based Segformer** model on semantic segmentation datasets for human body parsing.

◇ Trained deep learning-based **image matting networks** specifically for **human image datasets**. Incorporated **Segment Anything** and **Grounding DINO** models for targeted object retention in background removal task.

Quantitative Trader Intern *May 2022-Jul 2022*
Quadeye Securities LLP

◇ Developed **trading strategies** in the **crypto market** using Quadeye's existing tech stack.

◇ **Backtested** the strategy on historic data by running simulations with appropriate **constraints on gross, delta, volumes, positions, orders** etc and ensuring **liquidation** of position upon expiry.

AWARDS AND ACHIEVEMENTS

◇ Achieved **All India Rank 48** in **JEE Advanced** among **240,000** candidates enrolled. *(2019)*

◇ Achieved **All India Rank 42**(**99.99** percentile) in **JEE Mains** among **1.1 million** candidates appeared *(2019)*

◇ Awarded an **AP Grade**(awarded to **top 1%** of the class) in MA105(**Calculus**) course . *(2019)*