

Sohom Ghosh

Machine Learning Engineer at Apple 🍏 (Experience: 9.5+ years)

M.Tech (BITS, Pilani), PhD (Jadavpur University, Kolkata, India)

🌐 github.com/sohomghosh, ☺ hf.co/sohomghosh, 🎓 Google Scholar

📍 Hyderabad, India, ☎ +91-8001734384

✉ Email: sohom1ghosh@gmail.com

🏠 Web: sohomghosh.github.io

🌐 LinkedIn: sohomghosh

SUMMARY

- Delivering impactful **Artificial Intelligence (AI) & Machine Learning (ML)** solutions at scale for 9.5+ years
- **Key Achievements, Research & Engineering Leadership:**
 - Delivered **20+** end-to-end data science **projects**, including **10** solutions deployed in **production** impacting millions of users. Patented solution for Natural Language Processing (NLP)-driven call transcript analytics.
 - Earned **10 awards** & completed **11 advanced certifications** (MOOCs) in AI, NLP, & ML
 - Published **32 peer-reviewed research papers** at top venues: **TheWebConf (WWW)**, **CIKM**, **COLING**, **LREC**, **IEEE Big Data**, **CODS-COMAD**, etc. (🔗Google Scholar: **300+** citations, **h-index: 10**)
 - Co-authored **2 technical books** on NLP, Got **1 US patent** granted for NLP-driven call analytics innovation
 - Technical Focus: Applications of **Large Language Models (LLMs)**, **NLP**, **Generative AI**, & **ML**
 - Key Skills: **Large Language Models** | **Multi-modal AI** | **Multi-lingual NLP** | **Retrieval-Augmented Generation (RAG)** | **MLOps** | **Financial NLP** | **PyTorch** | **AWS** | **AI Agents**

WORK EXPERIENCE

• Apple Inc. 🍏 (NYSE: AAPL)

Sep 2025 - Present

Machine Learning Engineer

Hyderabad, India

- * Building intelligent systems to democratize AI across a wide range of solutions within Apple
- * Driving development & deployment of innovative AI models & systems that impact Apple's products & services
- * Implementing robust, scalable ML infrastructure (data storage, processing, & model serving components)
- * Developing automated ML pipelines for feature engineering, data augmentation, prompt engineering, hyperparameter tuning, & evaluation. Implementing advanced model compression & optimization techniques to reduce resource footprint while preserving performance
- * Brainstorming & designing various AI/ML proof-of-concepts for new or existing enterprise problems
- * Collaborating with business & cross-functional teams for delivering solutions to monitor sales of Apple products
- * Ensuring high-quality governance & delivery of ML models & solutions into production

• Fidelity Investments (FMR LLC)

Jun 2019 - Aug 2025

Senior Analyst → Data Scientist → Senior Data Scientist

Bengaluru, India

- * **Generative AI for analysing Call and Live Chat Transcripts** | 🔗Patented | In Production
 - Designed and managed an annotation pipeline using **Appen**, overseeing a team of 98 annotators to ensure **high-quality labelled datasets** for model training and evaluation
 - Fine-tuned advanced NLP models, including **Bi-LSTM**, **T5**, **BART**, and **LLaMA**, for **multi-theme extraction and summarization** from call transcripts. Optimized model performance to accurately identify and condense key insights, enhancing decision-making and workflow efficiency. Applied **Hierarchical Clustering** recursively to group similar themes and identify patterns within unstructured data. Productionized NLP models. Led end-to-end LLM pipelines (fine-tuning, evaluation, deployment)
 - **Applications:** i) **Automated short note generation** for call transcripts to streamline communication and improve accessibility ii) Analysing **root causes of high call volumes** to identify trends and optimize operational efficiency iii) **Featurization of textual interaction data** to enable advanced analytics and actionable insights
 - **Impact:** i) Identified and resolved login-related issues, resulting in a **10% increase in Customer Satisfaction Score** ii) Updated workflows for key processes, driving an **18-point increase in Net Promoter Score (NPS)** and a **27% improvement in Customer Ease Score** iii) Developed a comprehensive solution to address the student debt crisis, leading to a **60% increase in enrolment**, **59,000 yearly payments** processed, and **\$200M+ loan disbursed** to participants
- * **Predictive Customer Interaction System** | 🔗In Production
 - Experimented with **pattern mining algorithms**, **LLM**, **Learning-to-Rank** frameworks & engineered real-time solutions to predict customer's next-steps during an interaction with a chatbot. Identifying content gap for aiding content creation (**32.5% Click-Through Rate**, Increased Customer Satisfaction Score by **6%**)
 - Developed **statistical & linguistic metrics** to measure effectiveness of automated chat sessions
- * **Investor Behaviour Analytics, Enterprise Search & Knowledge Solutions**
 - Developed a **LightGBM-based predictive model** leveraging user profiles and web activity data to **forecast equity sell-offs**. Enabled proactive decision-making by accurately identifying patterns in user behaviour
 - Experimented with **Retrieval-Augmented Generation (RAG)** to enable accurate and context-aware responses to **business queries from call transcripts**. Explored and implemented **advanced chunking and denoising strategies** to optimize data preprocessing.

Technical Stack: Python, PyTorch, AWS, Snowflake, SQL; **Won 7 Awards, Deployed 4 production ML models, Published 3 research papers, Keywords:** LLMs, RAG, Clustering, Predictive Analytics, NLP

• Times Internet (The Times Group)

Jan 2017 - Jun 2019

Data Scientist

Noida, India

- * Designed and implemented a **Word2Vec-based skill recommendation system** for TimesJobs to enhance search experience of candidates
- * Engineered a scalable **XGBoost**-based predictive model using **PySpark** for email targeting, deployed on a **Hadoop** cluster (achieving a **15% increase** in email open rates)

EDUCATION

- **PhD in Engineering (Computer Science)**, Topic: Financial NLP (📄summary 📄poster 📄thesis 📄slides) 2025
Jadavpur University, Kolkata, India
- **Master of Technology in Software Systems (Data Analytics)** 2019
BITS, Pilani, India
- **Bachelor of Technology in Computer Science & Engineering** 2016
Maulana Abul Kalam Azad University of Technology (Formerly, West Bengal University of Technology), India

SELECT OPEN SOURCE PROJECTS & PUBLICATIONS

🔗GitHub, 🗨HuggingFace, 🎓G-Scholar

- **Generator-Guided Crowd Reaction Assessment** (In TheWebConf (WWW-2024), 📄paper, 📄code, 📄model)
Finetuned FLANG-Roberta with responses from LLMs (Claude, ChatGPT) using cross-encoders for tweet engagement prediction
- **IndicFinNLP: Financial Natural Language Processing for Indian Languages** (In LREC-COLING 2024, 📄paper)
Created datasets in Hindi, Bengali, & Telugu for assessing sustainability, detecting exaggerated numerals & ESG themes
- **Experimenting with Multi-modal Information to Predict Success of Indian IPOs** (📄data, 📄code, 📄pre-print)
Proposed a RAG, NLP, & ML-based approach to predict IPO success by analysing Red Herring Prospectus
- **Indian Financial Narrative Inference Tasks & Evaluations** (In Eval4NLP@AAACL-IJCNLP 2025, 📄data, 📄paper)
Multi-modal Indian earnings call analytics for stock price prediction, Indian Budget assisted sectoral impact ranking, etc.

TECHNICAL SKILLS & INTERESTS

Languages & Libraries : Core (Python, PyTorch, SQL & Cloud (AWS)), ML (Scikit-learn, Pandas, Numpy, XGBoost, LightGBM), NLP (NLTK, SpaCy, Transformers, LangChain), UI (Gradio, Streamlit), Spark

Algorithms & Concepts: Regression (Linear/Logistic), Decision Trees, Random Forest, Gradient Boosting Machine, Clustering, PCA, Neural Networks, Deep Learning, Large Language Models, Prompt Engineering, RAG, Fine-tuning

Others: LaTeX, MS Office (Word, Excel, PowerPoint), Confluence, Git, Jira, Kanban, Mural

Expertise: Natural Language Processing / Understanding / Generation

Industries: Internet, Consulting, Education, FinTech, Consumer Electronics, Financial Services, Technology

Relevant Courses: **Technical - Natural Language Processing Specialization** (Score >90%), **Neural Networks and Deep Learning, Improving Deep Neural Networks: Hyperparameter Tuning, Regularization and Optimization**

(Score>90%), **Prompt Engineering, Large Language Models, Cloud - AWS, Machine Learning Engineering, etc.**, AI Agents (LangGraph, crewAI); **Soft skills - Learnship Business English Level 10, Creating Effective Presentations,**

Creative Thinking; Research - Introduction to Research, NPTEL, Score: 87%, **Rank: Top 1%**,

Areas of Interest: Applied NLP, GenAI, ML; **Soft Skills**: Critical thinking, Communication, Problem Solving

ACHIEVEMENTS & EXTRACURRICULAR ACTIVITIES

Awards:CODS-COMAD-2023 YRS Honourable Mention; **Travel Grants**- ARCS-2025, PIC-2025, CIKM-2024, CODS-COMAD-2024; **Fidelity Investments**- Team Award [2025], **Kudos** [2024], **Eureka Innovation Enablers** [2023], **On the Spot** [2023], **Excellence in Action (2X) & You've Earned It (2X Individual, 2XTeam)**, **Shout Out** [2020, 2021, 2023], **Patent Award** [2021, 2024]; **Times Internet**- Rock Star [2018], **Hackathons**: **Kaggle: Datasets Expert & 7 Bronze**, **TechGig**: CodeGladiators-2018 Finalists, **Analytics Vidhya**: 4 times in top 25
Extracurricular Activities: Playing **Harmonica (Energize Talent Show: Rank-2@2024, Finalist@2025)**, Completed 11 **Treks** (received IndiaHikes Green Getter & Trekker for Life Awards)