

Sneha Maurya

New York, NY | +1 (646) 578-0650 | sm5755@columbia.edu | linkedin.com/in/sneha | github.com/sneha1012

EDUCATION

Columbia University <i>Master of Science in Data Science</i>	New York, NY Aug. 2024 – Expected Dec. 2025
Relevant Courses: Statistical Inference & Modelling, Applied ML, DL for NLP, High Performance ML Graduate Teaching Assistant: Data Analysis for Social Science, Databases for Business Analytics, Python for MBAs	

SRM University <i>Bachelor of Technology in Computer Science and Engineering (Merit Based Scholarship)</i>	Chennai, IN Aug. 2020 – May 2024
Relevant Courses: Data Structures, Operating System, Cloud Computing, Artificial Intelligence	

EXPERIENCE

IBM Research	New York, NY Sep. 2025 – Present
<i>Data Science Researcher – Advisor: <u>Atin Sood</u></i>	
<ul style="list-style-type: none">Optimized Qwen2.5-Coder-7B span-infill model with graph-augmented cross-file attention and LLM-based reward shaping, improving edit prediction accuracy by 12% on initial validation runsDeveloped 60K+ multi-file Java edit pairs from large GitHub PRs using tree-sitter parsing and control/data-flow graphs; applied compile validation and LLM-based intent scoring (exact + partial match) for next-action modeling	
NXP Semiconductors	Austin, TX May 2025 – Present
<i>Data Science Intern</i>	
<ul style="list-style-type: none">Built ML based die-level defect screening with KLA, linking layout and ATE signals; cutting false alarms by 25%Led cross-team LLM-powered failure analytics system using Dataiku & BERTopic; clustered 10 K + EMS logs across 400 + tools, saving 6 hrs/ week via automated root-cause insights & Power BI dashboards for engineersStandardized R Shiny ETL converting 200+ Lessons-Learned to Teradata JSON, improving cross-site traceability	
Columbia Business School	New York, NY Jan. 2025 – May 2025
<i>Graduate Research Assistant – Advisor: <u>Prof. Thomas Bourveau</u></i>	
<ul style="list-style-type: none">Developed a multimodal RAG system for ESG compliance analysis under the CSRD, integrating Qwen2.5-VL for visual reasoning, RolmOCR for structured text extraction, and BGE-M3 embeddings for cross-modal retrievalFine-tuned a QLoRA adapter on Qwen2.5-VL-7B to enhance reasoning over table captions, achieving 85% retrieval precision and enabling accurate interpretation of corporate reports through a Streamlit interface	
Metropolis Healthcare	Mumbai, IN May 2024 – Aug. 2024
<i>Data Science Intern</i>	
<ul style="list-style-type: none">Engineered GPT-3.5-based clinical NLP system integrating RAG for retrieval of similar historical cases, generating patient-friendly summaries across 10K+ medical reports and improving patient engagement by 20%Deployed AWS pipeline (S3, Lambda, Bedrock) processing 200+ reports/day with HIPAA complianceCollaborated with marketing to build SQL/Tableau dashboards analyzing 2M+ patient records, tracking acquisition patterns and service utilization across regions, reducing manual reporting by 12+ hours weekly	

TECHNICAL SKILLS

Programming Languages: Python, SQL, Swift, Go, Java, Shell, Groovy, Gherkin, R
Frameworks: PyTorch, TensorFlow, LangChain, FastAPI, Kafka, Spring, Node.js, Flask, Hadoop, Ray, CoreML
Developer Tools: AWS, GCP, Docker, Kubernetes, Maven, Jira, Git, Postman, Jenkins, Splunk, Apache Spark
Libraries: Sklearn, NumPy, Pandas, Hugging Face, Matplotlib, React, Ray, OpenCV, NLTK, Seaborn, Keras, Swagger

PROJECTS

Chronos: Autonomous Scheduling Agent <i>LangGraph, Llama 3, DPO, Gmail API</i>	Nov. 2025
<ul style="list-style-type: none">Fine-tuned Llama 3 8B via LoRA on 3.1K email-action pairs achieving 82% intent classification for draftingApplied DPO on 2.4K preference pairs collected over 6 weeks; improved draft acceptance rate from 61% to 87%Deployed via FastAPI and Chrome extension; scheduled 13 follow-ups, managed 45+ applications autonomously	
Semantic Code Intelligence Engine <i>SPLADE, LightGCN, IPW, Elasticsearch, FAISS</i>	Aug. 2025
<ul style="list-style-type: none">Built multi-stage retrieval over 50+ repos using SPLADE for candidate generation and LightGCN for re-rankingApplied Inverse Propensity Weighting to debias; achieved 0.38 MRR@10 on 500 queries, improved recommendations 28%, deployed with Streamlit	