# Mohnish Bangaru

Brooklyn, New York | (347) 856-8262 | mohnishbangaru@nyu.edu | LinkedIn | Portfolio

## EXPERIENCE

**Drizz** — **New York, NY**

***AI Engineer*** — **Aug 2025 - Present**

- Built a Mobile Application Testing Agent enabling natural language testing for diverse apps, engineered a **context caching protocol using a context cache manager**, deduplicating static input tokens, resulting in a **75% cost reduction.**
- Designed a comprehensive **library of 14 tools** leveraging the VLM's **native tool calling** capabilities, enabling the agent to interact with the environment with speed and precision.
- Crafted custom **Agent Evaluation Frameworks** tuned to estimate task completion and goal tracking, **improving testing accuracy by 90%** using Action Alignment Score, Goal Completion Rate, and Step Efficiency.
- Developed **test authoring module**, authoring app testing scripts **using Agent-generated metadata**, enabling single-click execution and **reducing test authoring time by 60%.**

Skillset: PyTorch, Vision-Language Models, LangChain, Python.

**New York University** — **New York, NY**

***Graduate Student Analyst*** — **Sep 2023 - May 2025**

- Enhanced data search-ability **using the RoBERTa model**, improving accuracy in **identifying user problems from text comments** by 80%.
- Created and maintained ETL pipelines for University Data Warehouse (5M+ rows), adhering to **strict data governance policies** and **preprocessing malformed data**.
- Introduced a set of **15+ Tableau dashboards** visualizing **key performance indicators** to support data-driven decision-making **across university departments.**

Skillset: Microsoft Excel, Tableau, Qualtrics, Docker, Oracle DB, sci-kit learn, Python.

**KPMG** — **Bangalore, IN**

***Data Scientist*** — **Apr 2022 - May 2023**

- **Automated reporting workflows** using Python scripts, **saving 264+ hours annually** by extracting financial data from unstructured documents using deep-learning techniques.
- Achieved a data extraction **accuracy of 99% using BERT-based models** for **Named Entity Recognition (NER)** on financial documents.

***Analyst*** — **Jan 2021 - Mar 2022**

- **Revamped SQL queries** for real-time data access, cutting query execution time by **30%**.
- Maintained documentation and scripted **REST API's** serving **Forecasting Models** deployed on **AWS.**

Skillset: Alteryx, API Development, AWS, API Documentation, Python.

## EDUCATION

**New York University, Tandon School of Engineering** — **New York, NY**

*Master of Science in Computer Engineering* — **Sep 2023 - May 2025**

**SRM Institute of Science and Technology** — **Chennai, IN**

*Bachelor of Technology in Computer Science and Engineering* — **Jul 2017 – May 2021**

## PROJECTS

**QLoR2C: a low resource parameter efficient fine-tuning framework** – New York University — **Jan 2025 - May 2025**

- Developed an advanced **parameter efficient fine-tuning** technique by integrating QLoR2C (utilizing quantized low-rank adapters) with an adapter management system, applied across **LLaMa 3.2 1B, 3B variants, SmolVLM, and U-net in Stable Diffusion.**
- Optimized hybrid residual connections through quantized low-rank adapters, **decreasing training time per epoch by 90%** across four models and **reduced model size by 25%** for **improved inference** speed.
- Expedited model c**onvergence speeds by an average of 10x** through optimized QLoRA implementation and the framework, **reducing computational costs by 40%.**

**Fine-tuning LLM's using QLoRA** – New York University — **Jan 2024 - Apr 2024**

- **Fine-tuned** Pythia 6.9b and 12b models on the Alpaca instruction tuning data using **QLoRA**.
- Explored and documented the effects of distinct LoRA ranks (1, 2, 4, 8, 16) on Pythia models, observed that a **rank of 8** yielded the **lowest test loss by 8%.**

## COURSEWORK / CERTIFICATIONS

- **Tools / Frameworks:** PyTorch, LangChain, Vertex AI, LM Studio, PineconeDB, Oracle DB, Tableau.
- **Relevant Coursework :** Advanced Computer Vision, Efficient ML and AI Accelerator, Deep-learning, Computing Systems Architecture, Computer Networks, Systems Engineering, Data Science for Business.