

# Nafis Neehal

nafisneehal95@gmail.com | (518) 805-8633 | Github | LinkedIn | Portfolio  
Machine Learning Engineer | End-to-End ML/LLM Systems | MLOps

## TECHNICAL SKILLS

---

**LLM Expertise:** RAG, Fine-tuning (SFT/PEFT), Quantization, Prompt Engineering, Benchmarking, GraphRAG  
**ML/DL/Causal:** PyTorch, DDP, TensorFlow, Scikit-learn, AutoML, OpenCV, EconML, DoWhy  
**MLOps Stack:** MLflow, Docker, CI/CD, ChromaDB, Hopsworks, PySpark, AWS (SageMaker, Lambda, EC2)  
**LLM Frameworks:** LangChain, LlamaIndex, Hugging Face, Axolotl, Unsloth, Autotrain, LangGraph, Opik, Comet  
**Languages & DB:** Python, SQL, R, C++, Neo4j, Google Firestore, MySQL

## RESEARCH & DEVELOPMENT EXPERIENCE

---

### IBM - RPI Research Projects

Feb 2022 – Present

Graduate Research Assistant, Team Lead — (Funding: IBM HEALS Project)

Troy, NY

- Building “TrialBrain” - LLM Augmented clinical trial automation framework: [[ArXiv](#), [Github](#), [HuggingFace](#)]
  - \* Released 4-bit quantized Llama-3.2-3B models fine-tuned (PEFT) on 65k+ clinical trials specializing on feature generation task
  - \* Created “CT-Bench” (1700 trials, 1.6k+ medical conditions) for benchmarking LLMs in feature generation task
  - \* Identified 3 types of Hallucination in our task and implemented novel hallucination-adjusted metrics for GPT-4/LLaMA-70B evaluations (*Under Review in Trustworthy ML4H @ IEEE BigData*)
  - \* Developed end-to-end trial feature generation pipeline achieving 48.5% accuracy improvement over baseline with RAG-based few-shot examples and 0.85 Cohen’s Kappa with human experts (*Under Review in ARR*)
- Architected “FRESCA” - a fairness-aware ML-Based patient matching framework: [[Github](#)]
  - \* Engineered novel ML-based patient recommendation improving treatment effect estimation accuracy by 75-80%
  - \* Implemented dual-adjustment pipeline for bias mitigation improving demographic alignment by 96-99% and achieving patient recruitment cost reduction by 25%
  - \* Published in top ML/Healthcare venues, won awards - [[RecSys’24](#)] [[AMIA’23](#)] [[SCT’23 \(Best Poster Award\)](#)]

### CDPHP - RPI Research Projects

May 2020 – Jan 2022

Graduate Research Assistant — (Funding: CDPHP Industrial Research Grant)

Troy, NY

- Type-2 Diabetes Health Management Program Evaluation using Machine Learning: [[HIMS’22](#)]
  - \* Developed deep autoencoder for patient matching (35% faster, 40% memory reduction) and multi-stage survival analysis for outcome tracking
  - \* Optimized 9M+ patient record processing pipeline using PySpark/AWS achieving 60% faster processing time
- Improving Targeted Intervention using Machine Learning: [[BIBM’22](#)]
  - \* Built hybrid ML framework combining novel PCM clustering algorithm to identify 3 treatment-response subgroups in 350K control vs 1.6K treatment cohort (high class imbalance)
  - \* Implemented nearest-neighbor and exact matching techniques for unbiased treatment effect estimation
- High-risk patient identification using Machine Learning: [[BIBM’21](#)] [[Github \(Non-Proprietary\)](#)]
  - \* Engineered ML pipeline processing 22.5M+ records (87 temporal features) achieving 95% physician agreement and 30% early detection rate
  - \* Created hybrid preprocessing pipeline with PCA reduction and downsampling for extreme class imbalance (0.5% positive class), achieving 200x efficiency gain through temporal windowing and sparse data handling

## OPEN-SOURCE PROJECTS (SELECTED)

---

**BanglaLLM** ([HuggingFace](#)): Developing fine-tuned open-source LLMs for reasoning and factual analysis in Bengali  
**MAMA-GPT** ([Github](#)): GPT-4 powered Bengali Voicebot integrating real-time STT/TTS and bidirectional translation  
**Trade-Mind** ([Github](#), [Demo](#)): End-to-end MLOps pipeline for Bitcoin price prediction using Hopsworks (real-time feature engineering, model registry/serving), and GitHub Actions (CI/CD) for data update and model retraining  
**ChanBOT** ([Github](#)): Fine-tuned Llama3.1-8B using PEFT and 4-bit quantization to mimic a fictional TV character

## EDUCATION

---

### Rensselaer Polytechnic Institute

2019 – 2025 (Expected)

Ph.D. in Computer Science - ML Systems/LLM/Healthcare, GPA: 3.74

Troy, NY

### Rensselaer Polytechnic Institute

2019 – 2021

M.S. in Computer Science - ML Systems/Healthcare, GPA: 3.68

Troy, NY