

Islam Tayeb

(919) 685-2112 | islam.tayeb@duke.edu | [linkedin.com/in/islam-tayeb](https://www.linkedin.com/in/islam-tayeb) | islamtayeb.dev

EDUCATION

Duke University

B.S. in Computer Science (Software Systems), Minor in Chemistry

Graduation: May 2027

GPA: 3.85

SKILLS

Frontend: TypeScript/JavaScript, Next.js, React, HTML, CSS

Backend: Python, Java, SQL, Flask, Node.js, Docker, AWS, GCP, Git, Redis, PostgreSQL, Supabase, Pinecone, Celery, Selenium

Machine Learning: PyTorch, TensorFlow, LangChain, AutoGen, Google BERT, TF-IDF, RAG, GraphRAG, NumPy, R

EXPERIENCE

Incoming Software Engineer Intern

Soff (YC S24)

May 2025 – Present

San Francisco, CA

- Developing infrastructure for AI-powered supply chain platform that unifies procurement data for manufacturing companies

Software Engineer

Helian

Dec 2024 – May 2025

Durham, NC

- Achieved **80%** speedup and **65%** reduced resources by designing ETL pipeline using **Selenium + ChromeDriver**
- Embedded **25K PDFs** daily to **Pinecone** (<100ms latency) by building a **LangChain + Celery** distributed processor
- Minimized **token usage by 25%** and **response time by 40%** through A/B testing using **FastAPI + Redis** for **RAG**
- Visualized **50K+** **GraphRAG** document connections for **500+** users, creating a knowledge graph using **Neo4j**

ML Engineer Intern

Reveal Genomics

Feb 2025 – May 2025

Durham, NC

- Enhanced biomarker identification by **55%** by designing gene expression pipeline with **Dask + NetworkX** graph algorithms
- Improved detection of non-linear gene correlations by **65%** using visualization algorithms with **PCA + t-SNE** in **Python**
- Reduced analysis time by **85%** for R&D by engineering genetic analysis dashboards with custom data pipelines in **Next.js**

ML Research Assistant

Duke Health

Oct 2024 – Apr 2025

Durham, NC

- Increased domain generalization by **9%** with a protein continual learning model using **PyTorch** (**publication pending**)
- Processed **1M+** protein candidates **20%** faster by optimizing **PyRosetta** antibody pipeline with distributed training
- Improved antibody-antigen prediction by **20%** by ensembling **GearNet** structural data to enhance candidate selection

ML Engineer Intern

Life Edit Therapeutics

Sep 2024 – Jan 2025

Durham, NC

- Achieved **92% accuracy** predicting cell editing mechanism using **Random Forest** classifier on **4 RNA-seq datasets**
- Accelerated gene differential expression analysis for entire department by **40%** through real-time **Streamlit** dashboard
- Processed **10-15** features from **50K+** gene profiles via feature extraction system using **Pandas + PCA + TF-IDF**

Software Engineer Intern

Duke Institute for Health Innovation

Jun 2024 – Aug 2024

Durham, NC

- Processed **250+** **papers** daily into database + newsletters by making automated literature review system with **GROBID**
- Achieved **98%** accuracy and **95%** faster processing by designing paper relevance classifier with **Google Gemini + BERT**
- Served **350+** analysts across **50+** organizations by building grant-writing assistant serving using **AutoGen + Llama**
- Analyzed **10K+** **EHR** records hourly at Duke Health by developing **BERT** VTE detector, replacing rule-based system

Software Engineer Intern

Project: Sapien

Dec 2023 – Jan 2024

Princeton, NJ

- Processed **5,000+** responses, reducing creation time by **25%** by making population health survey builder with **React**
- Reduced analysis time **95%** by developing **Google BERT**-based symptom classification pipeline for **12** categories
- Ensured **HIPAA** patient privacy by implementing data anonymization system using **regex + PII** masking techniques

ML Research Assistant

Saudi Aramco, King Fahd University of Petroleum & Minerals

Jul 2022 – Sep 2023

Saudi Arabia

- Predicted CO₂ capture capacity within **12%** of experiments by performing GC Monte Carlo simulations using **MATLAB**
- Led meta-analysis by analyzing **5K+** papers for Aramco proposals using **SKL + Seaborn** (**3 publications**)
- Processed **36K+** points hourly by developing web interface for real-time breakthrough analysis using **Streamlit**

PROJECTS

Etchr – AI GitHub README Generator | *Next.js, Express, Supabase, Google Gemini, GitHub API, GCP*

- Automated technical documentation, reducing time from **120+ mins** to **5 mins** for **100+** users with **65%** repeat usage
- Optimized file loading by **80%** by developing repository analyzer excluding imported dependencies using **BFS** traversal
- Reduced LLM processing time and costs by **40%** through intelligent file selection using **Google Gemini**

Spotify NLP Text-to-Playlist Generator | *Next.js, FastAPI, Google BERT, Google Gemini, Spotify API, Genius API*

- Extracted **12** custom audio-lyric features alongside Spotify's features by designing generator with **Google BERT**
- Reduced API calls by **85%** and latency by **60%** through caching user's music library with local storage middleware