Muhammad Yousuf

New York Metropolitan Area
732-309-2962 muhammad.y.2305@gmail.com linkedin.com/in/muhammadyousufii github.com/MYousuf3

Education

Georgia Institute of Technology **May 2027**
BS/MS Computer Science (Intelligence and Information/Internetworks) – ***GPA: 4.00 - SAT 1590*** *Atlanta, GA*
Relevant Coursework: Data Structures & Algorithms, Computer Architecture, Discrete Mathematics, Probability & Statistics, Software Development Design, Design & Analysis of Algos, Computer Systems, Differential Equations, Linear Algebra, Artificial Intelligence, Machine Learning, Computer Networking, Operating Systems, Deep Learning, Databases, GPU Hardware Software Codesign Teacher's Assistant for Computer Architecture (C, Assembly, Docker/Autograder), GT Create-X Startup Incubator

Experience

Meta **May 2025 – Aug 2025**
Software Engineering Intern | PyTorch, C++, Triton, FX IR, LLVM *Menlo Park, CA*
- Developed verification tool for Training and Inference Accelerator and integrated it into PyTorch compiler (**torch.compile**) to automate model e2e runtime, accuracy, and compilation testing and validate **Triton kernel** generation for several in-prod models.
- Implemented local telemetry data serialization to reduce network dependency and conducted fused kernel runtime analysis with local caching and deduplication, speeding up **inference runtime** for dozens of models
- Improved operator coverage by 4% and reduced compilation time by 10% through targeted enhancements to the Inductor/Triton backend, including Cpp Pybinds and efficient compiler outlining and inlining, minimizing overhead and optimizing code generation
- Built a tool for round-trip serialization of **torch.nn graph modules** that converts modules (with submodules) into text and reconstructs them back into fully functional graph modules.

John Deere Financial **May 2024 – Aug 2024**
Software and Data Engineering Intern | PySpark, R, AWS, MLFlow *Johnston, IA*
- Used PySpark to develop an Asset Valuation Index spanning 10-year period to illustrate **economic trends**
- Integrated projection models for economic forecasting of 18-months with **LightGBM and XGBoost** for probable values
- Wrote and tested several custom R scripts to automate dataflow of critical files through Databricks, S3, EFS, and DASH
- Implemented simple automatic monitoring and deployment workflow using MLFlow to sustainability beyond summer
- Leveraged Spark to implement **seasonality algorithm** to eliminate extraneous factors, reducing instability by 34%

Johnson & Johnson **May 2023 – Aug 2023**
Software and Data Engineering Intern | Scikit-learn, React, Azure, Javascript *Raritan, NJ*
- Used Python to developed pattern models of EBR data used in the manufacturing process of **CAR-T cancer therapy**
- Identified discrepancies between shift processing times and created POC for solution that improved production workflow
- Worked on quick authentication initiative to streamline data inputting process and speed up production by 7 minutes
- Implemented simple and effective GUI for faster work shift assignment with React, Javascript, and Azure Web Services

Projects

RL Wordle | *Reinforcement Learning, CUDA, Torch* **2025**
- Implemented three algorithms—Naïve Bayes, Monte Carlo Tree Search), and **Q-Learning**—to solve the Wordle game, leveraging probabilistic reasoning, tree-based exploration, and value iteration under an MDP framework.
- Achieved a 97% win rate with the **Reinforcement Learning** agent after training on 1M+ games using epsilon-greedy exploration, custom state-action encoding, and a reward function based on information gain and letter placement feedback.
- Utilized Georgia Tech's PACE high performance computing cluster to parallelize simulation runs and optimize training time with **CUDA kernels** through **batched learning**

Scribe | *MongoDB, React, REST, NEXT.js* **2025**
- Engineered a full-stack **automated changelog** generation platform utilizing Next.js, a custom REST API for backend services, and MongoDB to generate changelogs for project documentation
- Implemented secure GitHub OAuth integration and built advanced **RESTful functionalities** for dynamic log manipulation, including on-the-fly editing, deletion, and search, thereby streamlining release management workflows for developers

Research

Undergraduate Research Assistant **Jan 2025 – Present**
AI Virtual Assistant Lab, Dr. Larry Heck *Atlanta, GA*
- Constructing conversational dataset for improving LLM question/answer responses for research papers in a **multimodal manner**
- Benchmarking performance of responses through **zeroshot and finetuned experiments** on several open source models including LlaMa 3.1 8B and GPT OSS 20B using ROUGE, METEOR, and BERTScore

Skills

Languages: Java, Python, C, C++, SQL, Node.js, PySpark, JavaScript, React, HTML/CSS, PHP, TypeScript, Bash
Tools: PyTorch, Git, Docker, AWS, Apache Spark, Kafka, YAML, REST API, Flask, Spring, Gradle, MySQL, Maven, Postgres, MongoDB, Linux, UNIX, Lambda, DynamoDB, Terraform, TensorFlow, Keras, scikit-learn, Selenium
Skills: Application Dev, Unit Testing, Machine Learning, Distributed Systems, Low Latency Optimization, Web Dev, Data Processing, Cloud Computing, API Dev, Continuous Integration & Deployment, Databases, Backend Dev, Concurrency & Multithreading, Virtualization, Object Oriented Programming, Natural Language Processing, Infrastructure, Agile