James Moore

jame.moore24@gmail.com jamesmoore24 James Moore jmoore.info

Education

Massachusetts Institute of Technology (MIT) Cambridge, MA
Candidate for B.S. in Artificial Intelligence and Decision Making *Expected May 2025*

- **GPA:** 4.7/5.0
- **Coursework:** Data Structures & Algorithms, Distributed Systems, Computer Vision, Deep Learning, Linear Algebra, Inference, Operating Systems, Networking, Databases, Computer Architecture, Software Engineering

Experience

Capital One June 2024 – September 2024
Software Engineer Intern, Cyber Infrastructure Team *McLean, VA*

- Created a recommendation system and interface using MongoDB, Express.js, React, Node.js (MERN) and machine learning which helps prevent 1,700 cyber threats and secure 106M customer accounts per year
- Led team of 7 engineers in designing a tester scheduling algorithm and interface which integrates with AWS Fargate, Docker containers and AWS API Gateway which saves up to $2M per cyber threat
- Created internal availability dashboard which resulted in a customer satisfaction increase of 25% over a month

MIT Department of Electrical Engineering and Computer Science February 2024 – Present
Machine Learning (6.3900) and Inference (6.3800) Lab Assistant and Grader *Cambridge, MA*

- Supported over 530 students by leading class-wide efforts to help modernize and create curriculum, resolve learning platform issues, give personalized feedback in weekly office hours and grade assignments.
- Taught weekly concepts such as gradient descent, neural networks, autoencoders, CNNs, transformers, bayesian statistics, sampling algorithms, reinforcement learning, decision trees, nearest neighbors and MDPs.

Chevron June 2023 – September 2023
Software Engineer Intern, Developer Infrastructure Team *San Ramon, CA*

- Redesigned an automated API deployment service and interface using Azure Functions, Azure API Management Console, TypeScript, Node.js and Angular.js which saves up to $200K yearly
- Helped develop features for a custom enterprise-wide API specification linter tool for developers based on OpenAPI and Swagger which resulted in 30% faster API specification acceptance

Night Owl June 2022 – August 2022
Software Engineer Intern *Cambridge, MA*

- Developed an events recommendations dashboard using XCode, React Native, GCP BigQuery and GraphQL which resulted in a 55% increase in session length and a 35% increase in daily active users over a month span
- Optimized BigQuery queries using GraphQL which led to cost savings of up to 65% for each user recommendation

Projects

OmniRoute Current

- Leader of a 3 person research and development team aiming to build a unified LLM interface with an low-latency model router which beats existing commercial providers by up to 30%
- Deployed model router and build on an Ubuntu Linux VPS running nginx with Github actions for CI/CD. Model router implemented and evaluated with PyTorch and training done with Google Colab
- Improved OOD performance by using contrastive learning techniques and led to 6% improvement over baseline

PureRecall December 2024

- Built a private, hands-off, meeting transcription service leveraging AWS Transcribe Streaming for speech-to-text and OpenAI for embeddings and summaries.
- Engineered an optimized, custom hybrid RAG search pipeline using semantic embeddings and pgvector in PostgreSQL with RPCs and metadata which improved search result relevancy by 70%
- Designed transcription processing using distributed system design with serverless edge functions which sped up transcription processing by 10x compared to previous sequential implementation

FitLink January 2024

- Designed a "Strava for weightlifting" based social media web application for an MIT-wide web development competition using MongoDB, Express.js, React, Node.js (MERN) which resulted in an "honorable mention"
- Created chatbot which integrates with a RAG pipeline using an embedding model and ChromaDB's vector database and a corpus of fitness literature curated from various online sources to create informed answers to user questions

Misc.

Skills: Python, C/C++, Typescript, SQL, React, Node.js, PyTorch, Git, Docker, AWS, Databases, CI/CD, Linux
Interests: Piano Composition, MIT Leadership Training Institute President, NCAA Varsity Baseball, ML Research