James Moore

✉ jame.moore24@gmail.com ⬡ jamesmoore24 in 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

AnyStudent   August 2024 – December 2024
Co-founder   *Cambridge, MA*
- Founded AnyStudent, an AI-powered academic assistance platform for teachers, raised $10k from MIT Sandbox Fund in cash and AWS credits and led outreach efforts to teachers in the Greater Boston area.
- Architected and deployed multiple applications (internal and core) with on AWS EC2 instances using TypeScript, Python, Go, Redis, Kafka, PostgreSQL and Docker according to stakeholder preferences.
- Led a team of interns across engineering, product, and design to ship key features including a model routing layer, AI integrations, a RAG tool, and an analytics dashboard.

Capital One   June 2024 – September 2024
Software Engineer Intern, Cyber 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 3 interns 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.

PROJECTS

BallKnower (WIP)   Present
- Developing an open-source sports betting analytics platform using Next.js for frontend, Python (FastAPI) for data ingestion/querying and custom auth, and PostgreSQL for structured data storage.
- Implementing a containerized microservice architecture with Docker Compose orchestrating Go workers, Redis/Kafka messaging queues, ML model inference (PyTorch) and scheduled Cron jobs, providing scalable data processing for thousands of betting odds
- Designing a data scraping system using mitmproxy to capture network traffic from major betting platforms, processing the data with custom parsers and streaming results to production database for real-time analysis

TreeGPT   February 2025
- 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 Amazon Linux EC2 isntance running nginx with Github actions for CI/CD which handles 10K users and 5K requests/day
- Model router implemented and evaluated with PyTorch and training done with Google Colab.

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

BitArray   September 2024
- Implemented a high-performance bit matrix rotation algorithm in C achieving top 3 performance out of 56 teams through AVX-512 vectorization and cache-oblivious traversal, with systematic performance profiling via perf yielding 8x speedup over baseline
- Engineered a multi-level caching strategy combining 7x7 chunk decomposition with 64x64 block processing and optimized byte-swapping operations, reducing cache misses by 85% (verified with valgrind)

MISC.

Skills: Python, C/C++, Typescript, SQL, React, Node.js, PyTorch, Git, Docker, AWS, Databases, CI/CD, Linux
Interests: MIT Leadership Training Institute President, HackMIT, AI@MIT, Piano Composition