Bao Tran

571-665-1609 | baotran.swe@gmail.com | linkedin.com/baot1301 | github.com/BaoT1301 | baot1301.com

Education

George Mason University (Expected) May 2027

Bachelor of Computer Science ***GPA: 3.8***

Coursework: Data Structures and Algorithms, Web Development, Databases, Operating System, Object-Oriented Programming (OOP), Artificial Intelligence, Cloud Computing, Computer Networks

Experience

Google July 2025 – Sep 2025

Google For Developers Program

- Built an **ML pipeline on Vertex AI** to classify journal entries using **bert-base-uncased-emotion**, deployed on **Google Cloud** with automated model evaluation and insights visualization for mental health analysis
- Fine-tuned **BigQuery ML model** for emotional state scoring using **structured emotional data**, applying feature engineering and model monitoring in **Gemini API** to improve prediction performance by **18%**

Astrion Bank May 2025 - Aug 2025

Software Engineer Intern (Transaction Platform Team)

- Built a **rate-limited** transaction ingestion API, reducing burst-induced failures by **45%**, by implementing **Java Spring Boot**, **Redis** token-bucket limiting, and **Kafka** backpressure handling under peak load
- Improved transaction service debuggability by **38%**, measured by mean time to detection, by containerizing services with **Docker**, deploying to **Kubernetes**, and instrumenting metrics and traces using **Datadog APM**
- Optimized transaction log storage costs by **30%**, saving **$4,800/month**, by migrating cold data to **AWS S3**, implementing lifecycle policies, and compressing payloads using **Java Spring Boot** with **Snappy** serialization

George Mason University Apr 2025 – Jun 2025

Teaching Assistant, CS425A: Algorithms

- Led weekly discussion sections and office hours for **25+ students**, teaching algorithm implementation in **C++** with emphasis on **STL usage**, **memory management**, and **time/space complexity analysis**
- Designed and graded **C++** programming assignments involving **graph traversal** and **dynamic programming**, providing line-level feedback and debugging support through **GitHub Classroom** and automated test suites

CloudyScale.ai May 2024 – Aug 2024

Software Engineer Intern (Web Development Team)

- Developed internal chatbot platform using **React.js**, **Redux**, and **Bootstrap**, enabling legal teams to retrieve document insights via natural language, increased legal research speed by **40%** and used by **100+ employees**
- Improved backend for chatbot history in **Golang**, **MySQL**, and **Google Cloud Storage**, implementing **timestamped message logging** and **file retrieval API**, increased chatbot usage by **80%** across legal teams
- Wrote **40+ unit and integration tests** across backend and frontend using **Pytest**, **pytest-mock**, and **Selenium**, increasing test coverage to **85%** and reducing manual QA time by **20%** in weekly deployments

Projects

Crypto Pilot - Trading | *MongoDB, React.js, Node.js, Express.js, AWS, Redis, Binance API* May 2025 – Present

- Engineered a real-time trading backend in **Node.js**, **Express.js**, and **Redis** that streamed price updates from the **Binance WebSocket API**, cutting p95 trade-execution latency by 42% through connection pooling
- Implemented secure user authentication with **JWT, bcrypt, and AWS Secrets Manager**, reducing unauthorized API attempts by **52%** and strengthening overall platform security posture
- Improved historical price query performance by **36%**, measured by p95 response time, by designing indexed time-series schemas in **MongoDB** and implementing query-level caching with **Redis** for high-frequency reads

FusionAI | *Python, Next.js, LangChain, Hugging Face, Google Cloud Platform* Jul 2025 – Dec 2025

- Built an AI-powered research chatbot supporting multi-document question answering, used by **10+ users**, by developing a **Python** backend with **LangChain** and integrating a **Next.js** frontend for interactive query workflows
- Deployed scalable inference and retrieval services on **Google Cloud Run**, reducing cold-start latency by **35%**, by containerizing the backend with **Docker** and managing environment secrets and logging via **GCP**
- Fine-tuned a lightweight language model using **Python**, **Hugging Face Transformers**, and domain-specific research data, integrating the model into a **LangChain**-based pipeline for multi-step reasoning and retrieval

Technical Skills

Languages: Python, Java, C++, C#, TypeScript, JavaScript, Go, Rust, C, Bash, HTML/CSS, SQL
Frameworks/Tools: Spring Boot, Flask, FastAPI, .NET, React.js, Next.js, Node.js, Express, Tailwind, Redux, gRPC
Databases: PostgreSQL, MongoDB, Redis, MySQL, DynamoDB, Supabase
Cloud/Infra: AWS (ECS, S3, RDS, Lambda), GCP, Docker, Kubernetes, Nginx, GitHub Actions, Terraform, Linux
ML/AI: PyTorch, TensorFlow, Hugging Face Transformers, scikit-learn, NumPy, LangChain, OpenAI API