Junran Tao

551 359 1768 | gracejrtao@gmail.com | LinkedIn | Portfolio | Hoboken, New Jersey - Open to relocate

Education

Stevens Institute of Technology   Hoboken, New Jersey
Master of Science in Computer Science   Sep 2024 - May 2026
Publication: Training Acceleration of Heterogeneous Circuit Graph Neural Network on GPUs

University of Shanghai for Science and Technology   Shanghai, China
Bachelor of Engineering in New Media Technology (Computer Science Concentration)   Sep 2019 - Jun 2023

Experience

Stevens Institute of Technology

Research Assistant | Python, PyTorch, PyTorch Profiler, Pandas, NumPy, CUDA, Slurm   Sep 2024 – May 2025, Part-time

- Designed and productionized compute-efficient research pipelines for large-scale heterogeneous graph datasets, implementing GPU-level parallelism and runtime profiling to accelerate model training by 3.8×–4.7×.
- Developed a reproducible benchmarking framework with PyTorch Profiler and automated data logging, ensuring transparent experiment tracking, version control, and performance governance across multiple model variants.
- Optimized end-to-end training workflows through fine-grained thread scheduling, memory reuse, and resource utilization analysis, achieving 46.8% reduction in training time while maintaining model stability.

EarthView Image Inc.

Full-stack Developer | C#, JavaScript, React, Node.js, WebSocket, Kafka, PostgreSQL, Docker, AWS   Jul 2023 - Jun 2024, Full-time

- Refactored a legacy approval platform into an event-driven SaaS by designing a new RESTful/Kafka backend and 30+ reusable components, cutting request processing time by ~40%.
- Optimized a 24-hour manual data workflow to <5 seconds by engineering a full-stack platform achieving 50+ FPS rendering while streaming 50k+ data points from PostgreSQL through WebSocket to CesiumJS.
- Designed and shipped an AI-assisted inspection feature by deploying a PyTorch model as a client-side plugin, eliminating server costs, hitting 87% accuracy, and boosting inspection efficiency by 50%.

Front-end Developer Intern | JavaScript, React, CesiumJS, Element UI, ECharts, HTML5/CSS3   Dec 2022 - Mar 2023, Internship

- Architected the frontend for a city approval SaaS platform, creating 30+ reusable components that increased code reusability to 40% which the sales team later highlighted as a key differentiator in winning a multi-year contract with the client.
- Solved a critical rendering bottleneck for a government disaster response platform by optimizing the CesiumJS pipeline with client-side clustering, achieving a stable 50+ FPS while rendering 50,000+ dynamic data points and simultaneously leading a 90% UI/UX overhaul.

Yunshang Chuangxiang Intelligent Technology Co. LTD

Backend Developer | Java, Spring Boot, PostgreSQL, JUnit, Mockito, AWS, Jenkins, Docker   May 2022 - Sep 2022, Internship

- Built and deployed a modular data platform supporting CRM operations and cultural archive services for enterprise clients, enabling fast keyword search and download of 100K+ records with full CRUD admin tools.
- Delivered bulk import/export APIs and role-based access control (RBAC) with 99% unit test coverage (JUnit, Mockito), strengthening reliability and security across multi-department workflow.
- Automated deployment pipelines on AWS (CI/CD), ensuring scalable and efficient delivery of frequent client updates, reducing release cycles from weeks to days.

Projects

Low Risk Capital Management | Python, MongoDB Atlas, Selenium, Pandas, Transformers, Scikit-learn   Oct 2025 - Nov 2025

- Engineered a modular, "self-healing" data pipeline using MongoDB and Selenium to ingest and validate data for 1,000+ companies; implemented a multi-pass resolver strategy and automated Data Quality (DQ) checks to achieve and maintain >98% data coverage.
- Deployed a Hybrid Search engine in MongoDB Atlas using Reciprocal Rank Fusion (RRF) , increasing mean search relevance by +24% (Mean RR@10) and eliminating all zero-score failures by unifying the strengths of sparse (BM25) and dense (Vector) search.
- Deployed an "LLM-as-a-Filter" pipeline, classifying 300+ "noisy" search candidates into "core" or "not_relevant" categories , which enabled the creation and 5-year backtesting of 10 final, human-verified investment portfolios.

Convention Platform | Github | React, Node.js, Express.js, MongoDB, Redis, Stripe API   Mar 2025 - May 2025

- Architected and developed the platform's core features, including convention/event lifecycle management, real-time community feeds, and a secure funding module by integrating the Stripe API for sponsorship and payments.
- Implemented a secure Funding module with role-based validation and Redis/ImageMagick optimizations, cutting media load times by 35% and ensuring reliable sponsorship and payment processing during peak traffic.

Skills

Languages: Python, Java, TypeScript / JavaScript, SQL, C#, C/C++, HTML/CSS
Backend / Data: Node.js (Express), Java Spring Boot, .NET, PostgreSQL, MySQL, MongoDB, Redis
Frontend: React, Vue.js, Next.js
Infra / Tools: Docker, AWS, WebSocket, Kafka, PyTorch