

Junran Tao

551 359 1768 | gracejrtao@gmail.com | [linkedin.com/in/junrantao](https://www.linkedin.com/in/junrantao) | junrantao.com | github.com/HuflattichGra
NY/NJ/CA/WA | On-site/Remote | Immediate Full-time Available

Education

Stevens Institute of Technology

Master of Science in Computer Science

Hoboken, NJ
Sep 2024 - May 2026

Publication: [Training Acceleration of Heterogeneous Circuit Graph Neural Network on GPUs](#)

University of Shanghai for Science and Technology

Bachelor of Engineering in New Media Technology (Computer Science Concentration)

Shanghai, China
Sep 2019 - Jun 2023

Experience

Stevens Institute of Technology

Machine Learning Research Assistant | Python, PyTorch, CUDA, PyTorch Profiler, Slurm

Hoboken, NJ
Sep 2024 – May 2025, Part-time

- Optimized CUDA kernels for sparse aggregation in heterogeneous GNN training and quantified impact with PyTorch Profiler, achieving 3–5× kernel speedups and up to 39% end-to-end training time reduction.
- Ran large-scale distributed experiments on a Slurm cluster using 8× NVIDIA A100 GPUs over ~100GB circuit-graph data, producing key paper results and visualizations with traceable logs and experiment documentation.

EarthView Image Inc.

Full-stack Developer | C#, Vue, Node.js, REST APIs, PostgreSQL, CesiumJS, Docker, Alibaba Cloud

Sichuan, China
Jul 2023 - Jun 2024, Full-time

- Delivered 3 public-sector SaaS systems (permitting & case management, disaster damage assessment, underground utility network management) in a 13-person FE/BE/DB team, shipping via internal GitHub and Agile releases.
- Built a decoupled, reusable component system for security-sensitive workflows, implementing RBAC, a full approval state machine (submit/return/transfer/co-sign/withdraw/archive), file uploads (PDF/images), and audit-trail views.
- Optimized data-heavy UIs (virtualized lists, lazy loading, throttling/debouncing, incremental rendering, caching) and supported real-time visualization via REST + WebSocket; automated cloud delivery through CI/CD with Docker image packaging and deployment to Alibaba Cloud.
- Built an AI-assisted Windows EXE inner tool for pipeline inspection, training and integrating a ResNet18-based defect classifier with OpenCV preprocessing on pipeline footage, reducing repetitive manual tagging, and reached 87% accuracy on held-out monitoring videos.

Frontend Developer | C#, JavaScript, Vue, CesiumJS, Element UI, ECharts

Dec 2022 - Mar 2023, Internship

- Built the frontend foundation for a city approval SaaS and shipped 30+ reusable UI components to speed delivery and maintain UI consistency.
- Improved rendering stability for a disaster-response 3D map under high-density, real-time point updates through model and client-side optimizations.

Baosight Software

Backend Developer | C#, ASP.NET Core, SQL Server, IIS, JavaScript

Shanghai, China
May 2022 - Sep 2022, Internship

- Owned a legacy C#/ASP.NET + SQL Server retrieval system (100K+ records), adding full-text search and PDF import/export for Shanghai Theatre Academy users.
- Deployed and operated production on a Windows VM with IIS configuration and log triage, supporting on-call fixes and direct client requirement discussions.

Projects

Convention & Travel Community Platform | JavaScript, React, Node.js, Express.js, MongoDB, Redis, Stripe Sep 2024 – May 2025

- Built a social community platform with user profiles (avatar/background uploads), follow graph, and interactive content workflows with authenticated CRUD, role checks, pagination/sorting, and XSS/sensitive-content safeguards.
- Implemented a secure Stripe funding module (PaymentIntent/Checkout + webhook signature verification) with Redis caching and media optimizations (Redis/ImageMagick), deployed via Cloudflare (frontend) + Railway (backend), cutting media load times by 35%.

Low Risk Capital Management | Python, MongoDB Atlas, Selenium, Pandas, Transformers, scikit-learn, Ollama Oct 2025 – Dec 2025

- Built a local LLM-assisted long-document distillation stage (Ollama + chunked summarization) to extract high-signal company facts from large Wikipedia pages and reduce downstream embedding noise for search and clustering.
- Developed an embedding evaluation framework across 14 model/input configurations and shipped an idempotent, restartable MongoDB vectorization pipeline with indexing/resume logic and resource-aware batching, enabling rapid A/B testing and safe promotion of the winning embedding to production search.

Skills

Languages: C#, TypeScript / JavaScript, Python, SQL, Java, C/C++, HTML5/CSS3

Backend / Data: Node.js, .NET, PostgreSQL, MySQL, MongoDB, Redis

Frontend: React, Vue, [Next.js](#)

Infra / Tools: AWS, Alibaba Cloud, CI/CD, Git, Docker, WebSocket, Cloudflare, Railway, Slurm, PyTorch Profiler