Jiaqi Duan

Fremont, CA | +1 (408) 438-7247 | jd.victoria.work@gmail.com
linkedin.com/in/jiaqi-duan | github.com/Victoriakaey | victoria-duan.vercel.app

Education

University of California, Santa Cruz
Computer Science and Engineering, Master of Science (M.S.) — *Expected Dec 2025*
Computer Science and Engineering, Bachelor of Science (B.S.); Psychology, Bachelor of Art (B.A.) — *Dec 2022*

Skills

- Programming Languages: TypeScript, JavaScript (ES6+), Python, Java
- Frontend Development: React, React Native (Expo), Next.js, Redux Toolkit, Tailwind CSS, NativeWind, Vite, HTML5, CSS3, Sass, SolidJS, Axios
- Backend & Databases: FastAPI, Django, Express.js, Node.js, GraphQL, RESTful API & gRPC design, PostgreSQL, Firebase (Firestore), Supabase, Redis, Kafka, RabbitMQ, Microservices Architecture
- Infrastructure & Cloud: Google Cloud Platform (GCP), AWS, Docker, Kubernetes, Terraform, Vercel, Postman
- DevOps & Testing: Git, GitHub Actions, CI/CD pipelines, Pytest, Jest, Jenkins, Unit Testing, Integration Testing, End-to-End (E2E) Testing, Test Automation
- Data & Visualization: Pandas, NumPy, Matplotlib, Seaborn, Plotly, Chart.js, Prometheus, Grafana, Logging & Monitoring
- Generative AI & Tooling: PyTorch, TensorFlow, SpaCy, LoRA, QLoRA, OpenAI API, Gemini API, Claude, Hugging Face Transformers, Ollama, AutoGen, LangChain, LlamaIndex, Chroma, Pinecone, LastMile AI, Weights & Biases (W&B), Cursor

Experience

Founding Engineer @ Ripplet | *June 2024 – Aug 2025*

- Developed and deployed Ripplet's marketing site with **Next.js**, **React**, **Tailwind CSS**, and **Vercel**, achieving pixel-perfect brand consistency, **responsive design**, and optimized performance with **SEO** best practices
- Integrated **multilingual support** using **i18next**, enhancing accessibility and usability for our diverse user base
- Built responsive dashboards with **Chart.js** to visualize **real-time** client sentiment trends, progress milestones, and engagement levels, helping therapists review sessions more effectively and tailor future treatment plans
- Architected and implemented a **HIPAA-compliant**, **multi-agent system** with **retrieval-augmented generation (RAG)** pipelines to analyze session transcripts and map client narratives to evidence-based psychological frameworks
- Implemented a secure API layer for session data ingestion and retrieval, optimizing database access with **PostgreSQL** indexing and **Redis** caching, **reducing query latency by 120ms (30%)** in testing compared to unoptimized baselines
- Integrated asynchronous processing with **Kafka** and **RabbitMQ** for large-volume sentiment analysis and contextual resource retrieval, maintaining stable performance under simulated high-load conditions in staging

Full Stack Engineer @ Tech4Good Lab | *June 2023 – Jan 2025*

- Led a **cross-functional** team of 10 engineers and designers to develop Pathways, an **AI-powered** self-directed learning platform, coordinating backend integration and pilot testing with **over 100 university students**
- Delivered the core application using **Solid.js**, **Express.js**, and **Firebase**, implementing **server-side rendering**, **lazy loading**, and **cached API responses**, **reducing average page load times by 35%** in staging tests
- Containerized services with **Docker** and configured **CI/CD pipelines** via **GitHub Actions** for automated build, test, and deployment to staging environments, **cutting manual deployment time from hours to under 10 minutes**
- Developed and maintained **unit and integration tests** for backend services using **Jest** and **Pytest**, **achieving over 85% test coverage** and ensuring **stable performance under simulated high-load scenarios**
- Designed and refined LLM prompts with **different prompting strategies**, validated via **A/B testing**, **reducing hallucinations** and **improving course recommendation relevance by 15%** as measured by completion rates

Backend Developer Intern @ WayOps | July 2021 - August 2021

- Aligned backend deliverables with business goals by collaborating with engineers and stakeholders, ensuring data infrastructure changes met product timelines and operational requirements
- **Increased query performance by 40%** in a production relational database by applying **indexing**, **query restructuring**, and **strategic caching**, reducing latency and supporting high-throughput analytical workloads
- Enhanced database scalability by **refactoring schema** to remove redundant fields and reorganize table relationships, **reducing duplication by 30%** and improving write throughput under higher volumes and concurrent access

Coding Instructor @ Code For Fun | *Feb 2023 - Feb 2024*

- Taught programming to cohorts of 1–300 students (ages 6–18), fostering an interactive and inclusive environment
- Developed adaptive, project-based curriculum with **90%** parent satisfaction, guiding students to build websites, data-driven applications, and automation scripts to strengthen **real-world problem-solving skills**

Projects

Large Language Models (LLMs) are Autonomous Cyber Defenders (ACD) | Python

- Researched LLMs as autonomous cybersecurity agents, co-authoring an **IEEE CAI 2025** paper on **explainability and decision transparency**, presented at the conference and **published on arXiv: 2505.04843**
- Extracted and embedded **500+ action–reason statements** using **OpenAI's Embeddings API**, converting LLM-generated rationales into high-dimensional vectors to enable downstream clustering and pattern analysis
- Applied **unsupervised ML algorithms (K-Means, DBSCAN, PCA)** with **feature standardization and dimensionality reduction**, uncovering 5+ interpretable behavioral clusters in autonomous agent decision-making
- Built a reasoning summarizer with **OpenAI GPT-4o** that transformed clustered behaviors into human-readable defense strategies via advanced prompting, **improving explainability and transparency** for LLM-driven autonomous systems.

Travel Agent | React Native, NativeWind, Redux, AutoGen, FastAPI, PostgreSQL, Redis, GCP

- Designed and built a **full-stack mobile travel app** with a **multi-agent system**, integrating real-time web search, API tool calling, itinerary generation, and a multi-turn critic to deliver personalized and constraint-aware plans
- Built a robust agentic web scraping module using **Perplexica** for search discovery, **Playwright** for dynamic rendering, and **Trafilatura** for clean content extraction, **improving relevant content retrieval accuracy by 35%** over baseline
- Built a **rule-based and semantic-relevance filtering** layer to evaluate **open-domain** scraped content for freshness, factual accuracy, and domain relevance, **increasing usable content by 35%** over baseline
- **Improved travel plan quality beyond a GPT-4o baseline** by integrating a multi-turn critic agent to review filtered web content, search results, and generated itineraries, producing more structured, detailed, and user-aligned plans
- Implemented **Redis**-backed state management for cross-agent memory and caching, deployed on **GCP** with containerized services and managed **PostgreSQL**, ensuring low-latency queries and high availability in production

Yi | React, Next.js, Tailwind CSS, Django, REST API, PostgreSQL

- Built a **B2B SaaS application** for small businesses to unify business management and financial analytics, featuring interactive dashboards and visualizations for real-time performance tracking
- Implemented secure **large-file upload** workflows with **Google Cloud Storage signed URLs** and **resumable upload handling**, providing **real-time progress feedback** and ensuring transactional integrity for financial record verification
- Engineered a **reusable component library** with built-in **theme switching** using **Tailwind CSS** configuration, ensuring consistent UI patterns, improving codebase **maintainability**, and facilitating faster feature development
- Developed **authentication** and **role-based access control (RBAC)** with **Django REST Framework** and **JWT**, enabling differentiated permissions for staff, admins, and guests to securely access relevant data and features
- Integrated **Redux Toolkit** and **RTK Query** for centralized **state management** and efficient backend integration, utilizing **caching** and **request deduplication** to improve performance and support scalable data flows

NoteGrid | React, Next.js, Tailwind CSS, Supabase

- Built a note-taking web application with switchable modes for rich text or code editing, integrated with a calendar that categorizes and color-tags notes for chronological tracking
- Built a **TipTap**-based text editor in **Next.js** with real-time font and style controls powered by **state-driven UI updates**, storing content as structured JSON for consistent **cross-device rendering** and **fast page-load rehydration**
- Integrated **Monaco Editor** for in-browser code editing with real-time execution via the **Piston API**, implementing multi-language syntax highlighting, and custom themes to improve performance and user experience
- Implemented an **AI-powered** note analysis and summarization feature using the **OpenAI APIs** with **advanced prompting** to generate concise summaries and extract actionable insights directly from user notes