Jiaqi Duan

https://victoria-duan.vercel.app/

jd.victoria.work@gmail.com
408-438-7247

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*

Experience

Founding Engineer @ Ripplet | *June 2024 – Present*

- Collaborated with therapists and domain experts to define real-world use cases, intervention boundaries, and product goals, aligning system behavior with clinical standards and ethical constraints
- Refactored backend schema and optimized query performance with **30%** redundancy reduction and **40%** faster access compared to baseline, enabling scalable, low-latency case retrieval across multi-agent workflows
- Architected a **HIPAA-compliant, multi-agent** cognitive support system integrating **multi-modal retrieval augmented generation (RAG)** to surface decentralized client narratives and psychological research, enabling real-time therapist assistance during high-stakes sessions

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

- Led a cross-functional team of 10 engineers and designers to develop Pathways, an AI self-directed learning platform; increased weekly active users by **25%** in two months via dynamic UI workflows and adaptive content personalization
- Conducted **iterative user research** and **prompt optimization** grounded in self-directed learning literature, improving the quality of LLM-generated content and increasing recommendation relevance by **15%** based on user feedback
- Built a high-performance, real-time platform architecture with **Solid.js**, **Express.js**, **Vite**, and **Firebase (Firestore)**, optimizing data synchronization and load times for a smooth user experience

Projects

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

Explainable AI (XAI) in the Context of Cyber Security | LLMs + RL Agents as Defenders for CAGE 4 Challenge

- Co-authored a peer-reviewed research paper presented at IEEE CAI 2025; published on **arXiv:2505.04843**
- Designed and implemented an adversarial simulation agent to stress-test the robustness of **LLMs** and **Reinforcement Learning (RL) based** autonomous cyber defense systems against real-time service disruptions
- Extracted and embedded action-reason statements using **OpenAI's Embeddings API**, converting LLM-generated rationales into high-dimensional vectors for downstream clustering
- Applied **unsupervised machine learning (K-Means, DBSCAN, PCA)** with feature standardization and dimensionality reduction to uncover interpretable behavioral clusters in agent decision-making
- Built a reasoning summarizer driven by **OpenAI GPT-4o** that converts clustered behavior into human-readable defense strategies via **advanced prompting strategies**, advancing explainability and transparency in LLM-driven autonomous systems

Travel Agent | React Native, NativeWind, Redux, FastAPI, PostgreSQL

Mobile Full-Stack Multi-Agent Travel Planner with Tool Use & Multi-Turn LLM Reasoning

- Architected a modular **multi-agent LLM pipeline** using **AutoGen** to generate travel itineraries based on content that is accurate, up-to-date, semantically relevant, and aligned with user preferences and constraints
- Built a robust **agentic web scraping** module by integrating **Perplexica** for search-based discovery, **Playwright** for dynamic content rendering, and **Trafilatura** for clean content extraction
- Combined **LLM-as-Judge** with **natural language processing (NLP), machine learning (ML)**, and **rule-based filtering** to evaluate content quality and ensure alignment with user travel needs
- Curated and annotated a dataset of real travel queries and content chunks to supervise and evaluate judgment layers across multiple quality dimensions, including **constraint-fulfillment and user preference alignment**
- Fine-tuned the Critic Agent using **LoRA** to condition LLM outputs on user travel needs, reducing hallucinations and improving personalization across both structured JSON schemas and rich-text itinerary narratives

Skills

- **Programming Languages:** Python, TypeScript, JavaScript, Java, C, C++
- **Web Development:** React Native, React, Next.js, SolidJS, Node.js, FastAPI, Django, Django REST Framework, Express.js, Tailwind CSS, NativeWind, Vite, HTML, CSS
- **LLM Systems & AI Tooling:** PyTorch, TensorFlow, Keras, Scikit-learn, Hugging Face Transformers, OpenAI APIs, Ollama, LlamaIndex, AutoGen, Pinecone, Perplexica, Weights & Biases (W&B)
- **Data & Visualization:** Pandas, NumPy, Matplotlib, Seaborn, Plotly, Chart.js
- **Databases & Infrastructure:** PostgreSQL, Supabase, Firebase, MongoDB