Jeffrey Guo

360-339-1697 | jeffrey.guo00703@gmail.com | linkedin.com/in/jeffrey-f-guo | jeffreyguo.xyz/

Education

Western Washington University — Bellingham, WA
Bachelor of Science in Computer Science – ***GPA: 3.90*** — *Expected June 2027*

Experience

Software Engineer Intern — June 2026 – September 2026
A10 Networks — *San Jose, CA*
- Incoming SWE intern on AI-Application Delivery Controller team

Machine Learning Researcher — July 2025 – November 2025
Algoverse — *Remote*
- Second author on long-context agentic LLM systems research accepted to **NeurIPS 2025 Responsible Foundation Models Workshop**
- Reduced hallucination rates from **52% to 6.8%** by building a **hierarchical memory store** to enforce accurate state management across long-context sessions

Software Engineer Intern — June 2025 – August 2025
Fisher Investments — *Camas, WA*
- Designed and executed a full **data migration of 2M+ data points** from legacy Excel-based systems to Salesforce via nightly **ETL pipeline**, eliminating **15+ hours** of weekly manual spreadsheet processing
- Designed and implemented a Salesforce data model across 6 objects, replacing siloed spreadsheets to enable first-time campaign performance analysis and serve as the foundation for downstream marketing reporting
- Constructed a serverless ingestion architecture using **C# Azure Functions** to handle secure PDF uploads, integrating Microsoft Defender for real-time malware scanning

Projects

Multi-Agent Code Reviewer | *Typescript, Cloudflare Workers, Node.js, SQLite* — March 2026 – April 2026
- Deployed a 4-agent GitHub PR reviewer on Cloudflare Workers, fanning out domain-specialized agents in parallel across a 3-tier ReAct tool pipeline (regex → LLM sub-calls → structured extraction)
- Reduced false positive findings from ∼5 to 0-1 per review via an LLM-as-judge pipeline and chain-of-thought prompt engineering

Medicaid Policy Network Simulator | *C, Python, FastAPI, React* — February 2026 – February 2026
- Built a multithreaded C server with 10 workers writing to a shared **mutex/condvar-synchronized bounded queue**, distinguishing queue-saturation vs. TTL-expiration drop modes to model distinct Medicaid policy failure outcomes
- Streamed live simulation metrics to a React dashboard using a websocket connection in FastAPI, enabling real-time policy parameter reloads without server restart

Serverless Receipt Scanner | *TypeScript, React, AWS, Python* — December 2025 – January 2026
- Enabled direct client-to-S3 uploads by generating **presigned URLs** with embedded websocket connection and file metadata, decoupling upload bandwidth from server
- Designed concurrent OCR pipeline on AWS using per-image Lambda triggers, reducing batch processing time by **43% (30s → 17s)**
- Reduced **time-to-first-result** by **87% (30s → 4s)** by streaming each completed receipt to the client via WebSocket as its Lambda invocation finished

Awards

2nd Place @ OSU Hacks 2026 (Google Gemini Track) · 2nd Place @ InsForge x Qoder AI Hackathon 2025 · Best Use of Cloudflare AI @ DubHacks 2024

Technical Skills

Languages: Python, JavaScript, TypeScript, C/C++, Golang, C#, Java, PostgreSQL
Cloud & DevOps: AWS, Docker, Cloudflare, Linux/Unix, CI/CD, Git
Libraries & Frameworks: FastAPI, Flask, React.js, Next.js, Node.js