# Warren Wang

(213) 886-4199 | usc430vazw@protonmail.com | linkedin.com/in/yjwarrenwang | github.com/crimsonpython24

## Education

**University of Southern California** — Los Angeles, CA
*Master of Science in Computer Science* — *Jan. 2026 – May 2027*
*Bachelor of Science in Computer Science* — *Aug. 2023 – May 2027*

## Technical Skills

**Languages**: JavaScript, TypeScript, Python, C++, C, Golang, SQL, HTML/CSS
**Web Technologies**: React, Node.js, Django, Flask, GraphQL, REST APIs, WebSocket, SEO
**Infrastructure**: Docker, Redis, PostgreSQL, Git, Linux, AWS (Lambda, OpenSearch, CloudFormation, Bedrock)

## Experience

**Software Engineer Intern** — May 2026 – Aug. 2026
*Amazon Web Services* — *East Palo Alto, CA*
- Grounded an on-call AI assistant in 160K+ indexed events, providing a sourced audit trail and cutting triage time by **60%**
- Cut 20-min deployment reviews to <5 minutes by pre-generating plain-English change summaries, tested by **300+** engineers
- Load-tested indexing and APIs under stress test with a 0.1% error rate and **zero downtime**, clearing it for product launch
- Kept search ranking accurate at **50x** production index size by fixing term-weighting drift, holding search precision at >95%
- Bootstrapped the team's CloudFormation deploy pipeline and OpenSearch cluster, writing quick-start guides for teammates

**Full-Stack Tech Lead** — May 2025 – Apr. 2026
*ArchNexus (Startup)* — *Irvine, CA*
- Built a WebSocket streaming backend for LLMs up to 175B parameters, serving concurrent users at **<2 ms added latency**
- Implemented rate limiting on LLM endpoints, capping inference spend at **$1,000/month** with zero downtime over 6 months
- Secured payments with signature checks and idempotency keys on Link, stopping forged and duplicate charges
- Led daily scrums for a 5-engineer team, shipping customer-requested features into beta in **3 days** on average (30+ in total)

**Teaching Assistant** — Sep. 2023 – Apr. 2026
*USC Radical Play Lab | EICOP and Activision Summer Program* — *Los Angeles, CA*
- Grew lab project website traffic **15%** by reconfiguring robots.txt and site structure for Google and DuckDuckGo crawlers
- Deduplicated and standardized the lab's research records, removing 25% redundant entries that would have skewed analysis
- Taught algorithm design to 25 students and wrote 14 C++ data-structure modules, raising assessment scores by **15%**

**Full-Stack Developer Intern** — Mar. 2024 – Apr. 2025
*Crypto-Arsenal* — *Remote*
- Built a WebSocket feed streaming live prices to 25 traders' charts in **under 5 ms** so they can act on market moves instantly
- Cut page load time **33%** by caching hot API responses and HTML resources; user transaction volume rose **20%** after release
- Designed a GraphQL API to add 5 trading platforms without backend rewrites, cutting each integration to **5 days**
- Automated build and test in GitHub Actions to **halve release time** and reduce overhead across 3 repositories

**Software Systems Team Member** — Aug. 2023 – Jul. 2024
*USC Racing – Formula SAE Team* — *Los Angeles, CA*
- Halved control unit latency using hash table lookups and treeshaking unneeded calls, avoiding a **$1,000** control-unit upgrade
- Modeled the car's hardware network as a graph to trace faulty signal paths, catching **20+ bugs** before competition

## Projects

**Document Classification Service** | *Python, Django, TypeScript, SQLite, Azure Foundry* — Aug. 2026 – Present
- Classified **118 of 120** county forms with no LLM calls by matching printed form codes against a 121-form catalogue
- Parsed text-layer PDFs directly and ran local OCR only on scans, reading **93%** of 192 degraded images at no API cost
- Built the Django upload API with SHA-256 duplicate detection and staff review for unmatched files, backed by **73** tests

**Local DNS Server Infrastructure** | *DNSCrypt-Proxy, Unbound, Santa, AIDE, Little Snitch, Bash* — Jan. 2026 – Present
- Load-balanced encrypted DNS over 73 European resolvers with auto-failover, halving uncached lookups from **20 to 10 ms**
- Isolated proxy and cache traffic with Unbound and OS firewall rules, running for 10 months with **98% cache hit rate**
- Hardened the host with a **222-rule** CNSSI-1253 baseline, Santa binary authorization, and AIDE file integrity checks

**Stack Overflow Contributor** | *Django, PostgreSQL, Python, REST APIs* — Dec. 2019 – Jan. 2023
- Ranked **top 0.3%** of Django answerers (Q3 2020); answers on backend architecture and middleware viewed 200K+ times
- Kept a 60%+ acceptance rate across 169 answers over 3 years on Django, PostgreSQL, ReactJS, and REST API questions