Tony Nguyen

669-226-8770 | ntony.dev@gmail.com | https://linkedin.com/in/ntony46

Education

University of California, Berkeley — Berkeley, CA
Bachelor of Arts in Computer Science — *May 2025*

Experience

Uber — June 2025 – Nov 2026
Software Engineer — *San Francisco, CA*

- Delivered **$15M+ in fraud-prevention impact** by extending Uber's identity graph to unify user identities and risk signals across third-party platforms, supporting partnerships representing $600M in incremental gross bookings.
- Prevented **$10M+ in measured fraud losses** by designing and deploying an enterprise-risk integration that expanded verification and challenge history modeling from individual customer organizations to complete enterprise hierarchies, enabling graph-derived fraud features.
- Generated **$5M+ in measured fraud reduction** by leading the document-risk graph migration from single-market to multi-jurisdiction status modeling for drivers and couriers worldwide, increasing actionable fraud-signal coverage by 30%.
- Improved reliability and efficiency of downstream Kafka processing by productionizing change-aware event semantics across all streaming graph pipelines, suppressing no-op updates and distinguishing newly created from updated states.

Uber — May 2024 – Aug 2024
Software Engineer Intern — *Sunnyvale, CA*

- Enabled **$4M+ in measured fraud-prevention impact** across Uber's platform by building real-time bot-detection graph analytics, processing 8K peak events/sec into up to 40K graph writes/sec across millions of security-signal relationships.
- Drove **$1M in annual downstream compute-cost savings** by building scalable graph ingestion and query layers that enabled top-100 bot-signature retrieval from endpoint vertices with 1M+ connected edges while achieving 200 ms end-to-end latency.
- Designed Kafka and Cassandra streaming data flows and a temporal graph model across URL endpoints, JA3 TLS fingerprints, and request headers, enabling rolling access-count analytics across 1 to 24-hour windows.

Uber — June 2023 – Sep 2023
Software Engineer Intern — *Sunnyvale, CA*

- Built reusable point-in-time ML infrastructure enabling **$5M+ in modeled annual fraud prevention**, expanding an initial fraud model's candidate feature set by 58% and improving recall by 2%, worth approximately $200K in measured annual impact.
- Productionized historical feature access for fraud-model pipelines at 92K writes/sec, eliminating a 3+ month feature-collection delay and accelerating offline training and experimentation for fraud models.

Projects

Context Engine for AI Coding Agents | *Rust, TypeScript, MCP, SQLite*

- Built developer infrastructure that packages repository evidence into compact, reusable context for coding agents, retrieving and ranking relevant code with exact Git and file-line provenance and serving it through MCP to Claude Code, Codex, and Cursor.
- Architected a version-aware context pipeline that groups and deduplicates code spans, preserves symbol metadata and retrieval rankings, enforces token budgets, and explicitly surfaces omitted evidence.
- Developed an A/B evaluation and retrieval-tracing framework measuring task correctness, searches, file reads, tool calls, token usage, cost, and latency while attributing agent behavior to source evidence with approximately 0.2 ms p50 tracing overhead.

Technical Skills

Languages: Java, Python, C++, C#, JavaScript, SQL
Backend/Infrastructure: Kafka, Cassandra, Pinot, Flink, Hive, Spark, PostgreSQL, FastAPI, Node.js
Developer Tools: AWS, Docker, Git, GitHub, Phabricator