

Oscar Rodriguez Moscota

oscar.rodmosc@gmail.com | <https://github.com/dmosc> | <https://www.linkedin.com/in/osdavrm>

EDUCATION

Graduate Certificate in Artificial Intelligence, Stanford (01/2026 - current)

B.S. in Computer Science, Tec de Monterrey (Grade: 94/100) (08/2018 - 12/2022)

English (TOEFL-IBT 109/120), German (Goethe B1), Spanish (Native)

SKILLS

Programming languages: Python, JavaScript, C++, Bash, SQL (All strong).

Areas of knowledge: Machine Learning, Compilers, CI/CD, Microservices, Databases, Systems Design, Testing, Linux systems, Scripting, Application development (web/ mobile).

EXPERIENCE

Software Engineer III, Search TPU Efficiency, Google (09/2025 - current)

- I work on improving runtime efficiency of ML training workloads running on Google's pool of TPUs. I focus on infrastructure optimization, observability and policy enforcement to ensure that all ML training workloads run as efficiently as possible to maximize resource utilization. To date, across a handful of optimizations I've released almost 100+ SWE years worth of savings.

Software Engineer III, Search Infrastructure, Google (04/2024 - 09/2025)

Software Engineer II, Search Infrastructure, Google (04/2023 - 04/2024)

- Productionized Google Search AI services to process the new AI experience workloads we're servicing across the Search products like AI Overviews & AI Mode. This fleet of services currently processes 100K+ QPS across thousands of production jobs.
- Worked on rollout of anti-scraping technology to guard Google Search against abusive scraping in production overall reducing downstream query processing costs by up to 25% across the stack (i.e. now queries costs to Google ¾ of what they used to).
- Led the implementation of a standardized suite of observability, detection, and mitigation tools to catch production regressions across the lifecycle of major binaries in the Google Search ecosystem driving Mean-Time-to-Mitigation & Mean-Time-to-Detection down +90% from historical trends yielding a lower bound estimate of \$4B in cost savings for a trailing year by reducing downtime.

Software Reliability Engineer Fellowship, MLH (01/2023 - 04/2023)

- Intensive SRE training to understand how production systems are built and what techniques large projects implement to securely and efficiently scale to millions of users. I took courses on Linux systems, scripting, systems design, databases, CI/CD pipelines, monitoring, and troubleshooting.

Software Engineer Intern, Search Web Architecture, Google (06/2022 - 08/2022)

Software Engineer Intern, Search Fulfillment, Google (08/2021 - 12/2021)

STEP Intern, Cloud Run, Google (06/2020 - 08/2020)

- Did 3 internships with Google every summer during my undergrad years working with different Google teams across the stack in teams like Google Cloud Run (Kubernetes containers in the cloud), Search Fulfillment (assembling responses for Search), and Search Web Architecture (infrastructure to serve Javascript on the browser fast!).

Software Engineer Intern, Rentals, Lyft (01/2022 - 04/2022)

Software Engineer Intern, Rentals, Lyft (06/2021 - 08/2021)

- Did 2 internships with Lyft Rentals working on their pricing algorithms for rental rides to incorporate pricing branching based on profile types as the company was introducing "Lyft Rentals for Businesses".

Software Engineer Intern, Fresenius Medical Care, Kalypso - acq by Rockwell (01/2020 - 05/2020)

- Wrote all test suites for a medical equipment lifecycle management system to continuously evaluate the rollout and deprecation of medical equipment healthcare companies to manage the lifecycle of their products. The client purchased the project from Kalypso for \$10MM. I got to practice writing all sorts of testing suites like individual, functional, DiRT and E2E to name a few.