

Joseandres Hinojoza

Phone: +1 (650) 213 7607

Email: jhinojoza@outlook.com

Portfolio: jhinojoza.dev

LinkedIn: [jhinojoza](https://www.linkedin.com/in/jhinojoza)

Github: [Ualabi](https://github.com/Ualabi)

SUMMARY

Site Reliability Engineer with 4+ years of professional experience, highlighting 2.5 years at Google as a Site Reliability Engineer operating services over 100M+ QPS, and recent experience building observability infrastructure for LLM serving at Near AI and contributing to frontier model training pipelines at Mercor. Focused on distributed systems, incident response, and production reliability at scale.

SKILLS

Languages: Python, Pypy, C++, Rust, Java, Bash, SQL, Git
Infrastructure: Borg, Kubernetes, Terraform, Ansible, Docker, Podman, GCP, AWS
Observability: Monarch, Prometheus, Grafana, Datadog, PagerDuty, Alerting Pipelines
Reliability: SLOs/SLIs/Error Budgets, Chaos Engineering, Canary Deployments, Blue/Green, Playbooks
Databases: PostgreSQL, MySQL, Redis, Cassandra, BigQuery, Firebase, Supabase
ML/AI Infra: NLP, Word2Vec, vLLM, TensorFlow Serving, distributed training pipelines
Methodologies: On-call, postmortems, capacity planning, load testing, incident command

WORK EXPERIENCE

Site Reliability Engineer, Google, *San Francisco (CA, United States)*

Nov 2022 -- Mar 2025

- Designed and implemented a global monitoring system covering 300+ clusters and 10,000+ tasks serving 10M+ QPS; built controlled-request health probes and multi-signal alerting pipelines that achieved 99.8% SLO compliance.
- Automated incident response playbooks in Python, eliminating manual remediation steps and reducing mean time to recovery (MTTR) by 30% across services handling millions of daily active users.
- Developed a multi-probe diagnostic framework to trace failures across complex distributed request paths, improving root-cause identification and reducing false-positive alert rate by 40%.
- Built and maintained dashboards and multi-signal alerts that improved real-time visibility for cross-functional infrastructure and product engineering teams.
- Participated in a global 24/7 on-call rotation (Doing 6 hours-long shifts in reverse-shadowing stage, spanning London and San Francisco), ensuring round-the-clock reliability for business-critical services 365 days per year.
- Partnered with infrastructure and product engineers to scale services during peak growth, balancing reliability, observability, and performance at global scale.

Software Expert Engineer, Mercor, *Remote (Worldwide)*

Apr 2025 -- Apr 2026

- Validated and adapted 50+ open-source GitHub repositories to evaluate model behavior for frontier LLMs (Grok 4, Claude); resolved environment configuration issues across Python, SQL, MySQL, PostgreSQL, Docker, and Podman.
- Designed prompt evaluation rubrics and documented model failure modes, contributing to measurable robustness improvements in AI model performance.
- Reviewed and evaluated Pull Requests for correctness, test coverage, and code quality across a distributed team workflow.

Software Engineer Inter, Near AI, *San Francisco (CA, United States)*

Nov 2025 -- Dec 2025

- Designed and deployed a monitoring system for 5 open-source LLMs served via vLLM; implemented health probes, alerting pipelines, and real-time dashboards to ensure high availability and model reliability.
- Configured and optimized NGINX-based load balancing across multiple inference services, improving traffic distribution and reducing tail latency.
- Integrated Datadog observability stack (metrics, logs, alerts) to monitor model performance, inference latency, and service health in production.

Systems Engineer, Knapp, *Vancouver (BC, Canada)*

Apr 2022 -- Sep 2022

- Configured and tested source code for smart warehouse automation systems deployed across multiple global locations.
- Debugged Python-based automation for package detection; verified database updates at each handling step to ensure real-time tracking accuracy.

Software Engineer Intern, Meta, *Mexico City (CDMX, Mexico)*

Aug 2021 -- Nov 2021

- Enhanced ranking algorithms for job recommendation systems, optimizing user engagement and match quality at scale.
- Contributed to large-scale backend systems in Python, C++, PHP, and SQL; participated in code reviews and agile sprint cycles.

EDUCATION**Mechatronics Engineering**, Instituto Politécnico Nacional, Mexico City, *GPA: 92/100*

2014 -- 2020

Mathematics, Universidad Nacional Autónoma de México, Mexico City, *GPA: 82/100*

2018 -- 2021

Academic Exchange, AGH University of Science and Technology, Kraków, Poland, *GPA: 90/100*

2018 -- 2019

PROJECTS & ACHIEVEMENTS

Open Source: Developed and published a PyPI Library named [sbbst](#) (AVL-tree based) -- 7,000+ downloads.

Personal Blog: Built [heyualabi.com](#) with PHP, HTML, CSS, and JavaScript to share engineering tutorials.

Business: Developed a website for a restaurant that is used between waiters and kitchen for the order ([repo](#)).

Olympiad: Gold medal, 27th Mexican Mathematical Olympiad (2013). Volunteer trainer for Team Morelos.

CanSat: Participated in CanSat Competitions 2020 and 2021, classified to the world contest.

Competitive: Over 600 problems solved in [LeetCode](#), 70 problems solved in Kick Start and Code Jam competitions and the top place was 663 in Round H 2020. Round 2 in Facebook Hacker Cup 2021, got the top 1133rd out of 32,699 participants.