

Samir Musali

Staff Engineer | Full-Stack | Platform | DevEx | LLMOps
samir.musali@gmail.com | GitHub: smusali | LinkedIn: samirmusali

SKILLS

AREAS

- Agentic AI & LLM Infra
- API Development & Integrations
- Backend/Full-Stack Engineering
- Developer Experience & Internal Tooling
- Healthcare Data Integrations (EHR/HIE)
- LLMOps & MLOps Engineering
- Open-Source Development
- Software Product Management
- SRE/DevOps Engineering
- Web & Data Scraping/Mining

LANGUAGES

- Node.js/Python
- JavaScript/TypeScript
- Golang/Ruby

TOOLS

- AWS: EC2/CloudWatch/λ/S3
- CI/CD: CircleCI/Jenkins/GitLabCI
- DB: MongoDB/PostgreSQL/CosmosDB
- Deployment: Ansible/Chef/Terraform
- EHR/HIE: FHIR/HL7/CDA
- GCP: GKE/GCR/Cloud Monitoring
- LLM: OpenAI API/Gemini API
- Node.js: Express/Restify/Fastify
- Python: Django/FastAPI/GraphQL/Flask
- Testing: Mocha/Jest/Tap/PyTest

EDUCATION

UNIVERSITY OF WATERLOO BACHELOR IN COMPUTER SCIENCE

June 2016 | Waterloo, ON, Canada

PROJECTS

LLM TELEPHONE INFORMATION DEGRADATION ANALYSIS

LOGDNA LOGSPOUT LogDNA DOCKER INTEGRATION

LOGDNA S3 LogDNA S3 LAMBDA FUNCTION

LOGDNA ORB LogDNA CIRCLECI INTEGRATION

EXPERIENCE

PERLE | Staff Platform Engineer

August 2025 - present | San Francisco, California, United States | Full-Time

- Architected platform-wide migration from role-based to permission-based access
- Designed mobile recording capabilities expanding platform accessibility
- Eliminated race conditions in Temporal workflows & implemented dynamic workflow config system
- Built bulk export/import infrastructure with upload/download functionality
- Optimized frontend by reducing API calls by 60-70% & implemented code quality improvements
- Implemented healthcheck system improving deployment reliability

EVIDENTLY | Senior Software Engineer

June 2024 - August 2025 | Bay Area, California, United States | Full-Time

- Migrated 12 GraphQL endpoints to FastAPI, cutting p95 latency by 35%
- Implemented trace mechanisms across the logs to improve observability
- Built Gemini-powered OCR+summarizer pipeline that auto-labels >10k pages/day
- Scripted SQL-based weekly reporter for more than a million HL7 records
- Implemented middleware to capture and handle exceptions gracefully
- Built smoke and load testing tools using Apache Bench to improve performance monitoring
- Built HIE Integration to fetch patient data in batch in near-real-time
- Deployed Cloud Function that triages error spikes & pages on-call only for actionable incidents

AUTHENTIK SECURITY | Senior Python Backend Engineer

March 2023 - June 2024 | Remote | Full-Time

- Implemented custom error handling in integration, enhancing robustness
- Updated development setup documentation, improving developer onboarding
- Developed configurable denial message feature, reducing hard-coded strings by 80%
- Shipped GitLab OAuth provider, growing SSO coverage to 85% of the customer base
- Analyzed access logs using OpenAI API for anomaly detection and security insight extraction
- Optimized Redis caching strategy for authentication tokens, increasing hit rate by 94%

BENCHSCI | Software Engineer

November 2021 - March 2023 | Toronto, Ontario, Canada | Full-Time

- Built SLO-based Alerting Policies around some critical API endpoints
- Instrumented SLO dashboards for top 12 APIs; error budget alerts cut 40%
- Built Availability & Latency Dashboards around some critical API endpoints
- Automated provisioning GCP Monitoring resources using Terraform
- Paired with the Search team to profile queries, saving \$18k/month
- Helped the Infrastructure team to optimize CI/CD processes

MEZMO | Software Engineer

December 2017 - November 2021 | Toronto, Ontario, Canada | Full-Time

- Led the projects for building, architecting, and improving new integrations
- Created tools to automatically install and configure LogDNA Agent
- Integrated LogDNA with external tools and applications
- Maintained customer-facing public tools and libraries of LogDNA
- Built internal and external CI/CD pipelines using CircleCI, cutting build times by 60%
- Maintained and improved Audit Logging for Web and API Services

NURALOGIX CORPORATION | Software Developer

May 2016 - December 2017 | Toronto, Ontario, Canada | Full-Time

- Debugged and optimized Python and Matlab scripts to improve system performance
- Implemented simple API calls using Flask, Tornado, Meteor, and Express to optimize data retrieval
- Worked on gRPC calls with protobufs and webRTC within Kurento for real-time communication
- Implemented PostgreSQL, Redis, and StatsD wrappers in Python, streamlining data processing
- Prototyped and documented backend architecture on AWS to establish ground for development
- Authored end-to-end test harness in Mocha/Chai covering 85% of endpoints