Luis Rasgado

Senior AI & Full Stack Engineer

luis@lrasgado.com
(408) 673-8257
San Francisco, CA
https://lrasgado.com

SUMMARY

Senior AI & Full Stack Engineer with **12+ years of experience** turning early-stage ideas into reliable, production-ready software across edtech, healthcare and fintech. Built AI agents, RAG pipelines and **LLM-powered applications** with deep expertise in **Java**, **Python**, **TypeScript** and **React**. Cut $200K a year in AI costs, scaled systems to **75 million users**, and helped support **$2.2B in annual revenue**.

SKILLS

Language
Java, TypeScript, Python, JavaScript, SQL, GraphQL

AI/ML
LLMs, Claude, OpenAI, LangChain, LangGraph, RAG, Agentic AI, MCP

Frontend
React, Redux, Next.js, Vue.js, Angular, Tailwind CSS, Vite

Backend
Spring Boot, Node.js, NestJS, Express.js, Django, FastAPI

Database Management
PostgreSQL, MongoDB, Redis, Elasticsearch, Pinecone

DevOps
AWS, GCP, Docker, Kubernetes, Git, Terraform, Jenkins

WORK EXPERIENCE

Applied AI Engineer May 2023 - Nov 2026
Blackboard, Reston, VA

- Reduced instructor follow-up by **12 hours per week** by building AI agents with LangGraph and Anthropic Claude that identify at-risk students and automate personalized outreach.
- Cut course discovery time from **2 minutes to under 30 seconds** by implementing RAG pipelines using LangChain, OpenAI embeddings, Hugging Face Transformers, and Pinecone.
- Saved $200K annually in LLM spend through prompt engineering, model routing, retrieval tuning, and context optimization that reserved premium models for high-complexity tasks.
- Built secure MCP servers in Python to connect agents with internal content and workflow systems, enabling reliable tool calling and scalable multi-agent orchestration across Java and AI backends.
- Reduced low-confidence **AI responses by 30%** by building guardrails pipelines for hallucination detection, PII filtering, and policy enforcement.

Full Stack AI Engineer Jul 2021 - Apr 2023
Medtronic, Minneapolis, MN

- Built a healthcare platform for personalized insulin dosing using Java, Spring Boot, React, Redux, TypeScript, and GraphQL, supporting **$2.2B in annual revenue** across diabetes care.
- Improved patient-message **classification accuracy to 91%** by fine-tuning healthcare NLP models with PyTorch and Hugging Face Trainer, and by developing pipelines for data extraction, enrichment, and classification from unstructured clinical text.
- Reduced accessibility **QA defects by 95%** and improved UI consistency by creating an ADA-compliant component library and design system with React, TypeScript, Storybook, semantic HTML, and Tailwind CSS.

- Implemented **HIPAA-compliant** observability and security controls using OpenTelemetry, Prometheus, Grafana, OIDC authentication, and encryption, strengthening platform reliability and compliance practices.
- Increased **team productivity by 50%** through code reviews, mentoring, and technical leadership in component design, software architecture, clean code, and SOLID principles.

Senior Software Engineer Oct 2015 - Jun 2021
Boston Consulting Group, San Francisco, CA

- Supported Axle's **$10 million Series A** fundraising by re-architecting a Django monolith into microservices with event-driven architecture using Java, Spring Boot, and Spring Cloud.
- Built real-time ETL data pipelines processing **50 million user events** using Spark Streaming, Hadoop, Hive, and Kafka, powering analytics dashboards for engagement, feature adoption, and subscription insights.
- Drove a **15% increase in mobile conversion** by building responsive web applications with React, Next.js, Material UI, and TypeScript, improving frontend architecture, state management, and rendering performance.
- Reduced **regression issues by 85%** by enhancing code quality with SonarQube and expanding automated unit and end-to-end test coverage using Jest, React Testing Library, JUnit, Mockito, and Cypress.
- Implemented security controls including OAuth 2.0/JWT, RBAC, rate limiting, and audit logging, enabling **SOC 2 compliance** and reducing security incidents.

Software Engineer Jun 2013 - Sep 2015
Spotify, San Francisco, CA

- Reduced inter-service **latency by 80%** by migrating service-to-service communication from REST/JSON to gRPC with Protocol Buffers across distributed backend systems.
- Improved production reliability to support a **99.99% uptime** SLA through CI/CD pipelines, containerized workloads, and monitoring with Git, Docker, Kubernetes, Jenkins, Prometheus, and GCP.
- Scaled backend services supporting **75 million users**, delivering architecture and performance improvements in Java and Node.js through object-oriented design, MVC, and service-oriented architecture.

EDUCATION & CERTIFICATIONS

Bachelor of Science, Computer Science Graduated: 2013
University of San Francisco

Certifications
- Oracle Certified Professional: Java SE 21 Developer
- IBM AI Engineering Professional Certificate