

# Jawad Haider

AI Systems Engineer · Machine Learning Infrastructure · LLMOps

San Jose, CA

## PROFILE

---

AI Systems Engineer with 6+ years of experience building production-grade ML infrastructure, LLM-powered applications, and intelligent automation pipelines. Proven track record delivering high-impact AI solutions for startups and enterprise clients across fintech, SaaS, and B2B. Specialized in translating complex AI capabilities into scalable, cost-efficient systems that drive measurable business outcomes.

## EXPERIENCE

---

### AI Systems Engineer — *Freelance / Contract · Remote* Mar 2023 – Present

*Selected engagements across fintech, SaaS, and enterprise automation*

- ▶ Designed and deployed multi-agent LLM orchestration systems using LangChain and LlamaIndex, cutting manual research workflows by 70% for a Series B fintech client
- ▶ Built a production RAG (Retrieval-Augmented Generation) pipeline on AWS for a B2B SaaS company, improving support ticket deflection by 42%
- ▶ Architected real-time inference infrastructure on Kubernetes with model versioning, A/B testing, and observability — supporting 10M+ monthly API calls
- ▶ Delivered AI-driven lead enrichment automation for a GTM team using Clay + OpenAI, reducing prospecting time from 4 hrs/day to under 30 minutes
- ▶ Clients include YC-backed startups, PE-backed SaaS firms, and Fortune 500 innovation teams

### Senior ML Engineer — *NovaSpark AI · San Francisco, CA (Remote)* Jun 2021 – Feb 2023

- ▶ Led development of a document intelligence platform processing 2M+ pages/month using fine-tuned transformer models (LayoutLM, RoBERTa)
- ▶ Reduced model inference latency by 58% through ONNX quantization and batching optimizations
- ▶ Collaborated cross-functionally with product and sales to build a self-serve AI demo environment that shortened enterprise sales cycles by 3 weeks on average
- ▶ Mentored 3 junior engineers; introduced MLflow-based experiment tracking adopted org-wide

### Machine Learning Engineer — *Helix Data Systems · San Jose, CA* Aug 2019 – May 2021

- ▶ Built NLP pipelines for structured data extraction from unstructured enterprise documents (contracts, invoices, SOWs) — accuracy improved from 61% to 94% over 18 months
- ▶ Developed and maintained feature engineering library used across 5 internal ML teams
- ▶ Integrated model monitoring with PagerDuty alerts, reducing P1 model drift incidents from 8/month to 1

### Data Scientist — *Meridian Analytics · Austin, TX* Jan 2018 – Jul 2019

- ▶ Designed customer churn prediction models reducing annualized churn cost by \$2.1M for a telco client
- ▶ Built automated reporting dashboards in Tableau and dbt, replacing 40+ hrs/week of manual analyst work

## TECHNICAL SKILLS

---

**AI / ML:** LLMs, RAG, Fine-tuning, LangChain, LlamaIndex, Transformers, Hugging Face, OpenAI API, Anthropic API

**Infrastructure:** AWS (SageMaker, Lambda, ECS), GCP Vertex AI, Kubernetes, Docker, Terraform, MLflow, Weights & Biases

**Languages:** Python (expert), SQL, TypeScript, Bash

**Data:** Postgres, Snowflake, BigQuery, Pinecone, Weaviate, Redis, Apache Spark

**GTM / Automation:** Clay, n8n, Zapier, HubSpot CRM, Notion API, Airtable

**Practices:** LLMOps, CI/CD (GitHub Actions), A/B testing, model monitoring, prompt engineering

## EDUCATION

---

**B.S. Computer Science** — *San Jose State University* 2013 – 2017

Concentration: Intelligent Systems & Data Engineering · GPA: 3.7

## CERTIFICATIONS & RECOGNITION

---

- ▶ AWS Certified Machine Learning – Specialty (2023)
- ▶ Google Professional ML Engineer (2022)
- ▶ Deep Learning Specialization – Coursera / deeplearning.ai (2020)
- ▶ Speaker, PyData San Francisco 2024: "Practical LLMOps for Small Teams"