

Karam Gulati

San Francisco, CA · krmjt98@gmail.com · (530) 979-6163 · github.com/ksg98 · linkedin.com/in/karamjeetsinghgulati

EXPERIENCE

Forward Deployed Engineer — Checksum AI, San Francisco Nov 2025 – Present

- Built a 400 case eval benchmark and RL environment from real enterprise failure modes (network faults, API errors, flaky selectors, UI shifts) to train and measure a self healing LLM test agent, now the team's release quality bar
- Re-architected the agent's context pipeline and failure classification, raising the eval pass rate from 54% to 85% and cutting suite execution time roughly 90% through parallel browser orchestration
- Own the AI testing platform end to end (React, TypeScript, Node.js, PostgreSQL, Playwright) across 20+ enterprise customers, generating 1,200+ tests per deployment and reducing manual QA hours 72%
- Partnered directly with enterprise customers to map production failure patterns into agent skills, eliminating false failure escalations and retaining an at risk account

AI Software Engineer (Founding Engineer) — Hyperlink, San Francisco Jul 2024 – Mar 2025

- Designed an autonomous browser agent platform in TypeScript and Next.js that executes natural language tasks end to end, with multi agent orchestration built on the Model Context Protocol (MCP)
- Shipped a Chrome extension as sole engineer to 400+ monthly active users, cutting manual scripting work 65%
- Profiled and rebuilt an AI video chat pipeline, cutting response latency from 11s to 0.7s by sampling frames only on visual change and streaming speech tokens first
- Built self serve GPU vendor onboarding (Electron installer, K3s, Prometheus, Grafana) that onboarded 150 vendors with 580 GPUs in 40 minutes, replacing a 45 minute manual SSH process per vendor

Software Engineer — Fidelity Information Services (FIS), Bengaluru Feb 2022 – Jul 2023

- Diagnosed legacy million row banking migrations with EXPLAIN plan analysis and rewrote row by row logic into bulk operations, cutting migration time 87% (weekend runs to overnight)
- Built and deployed 3 internal banking web applications (React, RESTful APIs) with containerized CI/CD pipelines on AWS

PROJECTS

fastaf-ide github.com/ksg98/fastaf-ide

- Native multi terminal AI IDE built in Rust and Tauri with a plugin API, running AI coding agents side by side in roughly 300MB of memory versus 2 to 3GB for Electron based equivalents

DataGuardian github.com/ksg98

- Real time multi modal PII anonymization using OpenAI Vision and Microsoft Presidio, detecting text in 6 languages from live video at 94% detection accuracy

EDUCATION

M.S., Computer Science — University of California, Davis Sep 2023 – Sep 2025

- GPA 3.8 · Research: GPU communication library for FSDP training with NCCL, achieving 3.5x faster inter GPU communication across 8+ GPU clusters

B.Tech, Computer Science and Engineering — SRM University, Delhi Jul 2018 – May 2022

SKILLS

Languages	TypeScript, Python, JavaScript, SQL, Rust, Java
AI Engineering	LLM agents, evals and benchmarking, context engineering, RL environments, Claude API, OpenAI API, MCP, LangChain, RAG
Web & Backend	React, Next.js, Node.js, Express.js, FastAPI, REST APIs, Playwright
Infrastructure	PostgreSQL, Redis, Docker, Kubernetes, AWS, CI/CD, Prometheus, Grafana