PRAKHAR JAIN

AI Evaluation Engineer | Full-Stack Engineer | Co-Founder, Sarveio

+91 7400501247 | prakhargdg@gmail.com | LinkedIn | GitHub | Portfolio

PROFESSIONAL SUMMARY

AI evaluation engineer and full-stack developer who builds RL environments, Terminal-Bench-style tasks and grading rubrics to stress-test frontier LLMs (Claude Opus 5, Claude Fable 5.1), and ships production software: multi-tenant SaaS, real-time systems, 3D/WebGL experiences and cross-platform mobile apps. Co-founder of Sarveio (restaurant SaaS, $2,000+ revenue) and founder of Vernier, an AI benchmarking platform. Skilled in Python, TypeScript, React/Next.js, Node.js/FastAPI, Flutter and cloud delivery (Docker, Vercel, Cloudflare). Hackathon finalist and Algoverse Hackathon winner.

EDUCATION

S.A.T.I (Samrat Ashok Technological Institute), Vidisha *Expected Graduation — July 2028*

Bachelor of Engineering (B.E.), Computer Science & Engineering (Blockchain) · CGPA: 7.6 / 10

TECHNICAL SKILLS

AI Evaluation & LLMs: LLM evaluation, RL environments, benchmark and task design, rubric design, Terminal-Bench, Snorkel, data annotation and QA, preference ranking, prompt engineering, Anthropic Claude API, multi-provider LLM routing (OpenRouter, Gemini, Groq, Cerebras, Hugging Face), vector embeddings, OpenCV, Playwright browser automation, webhook & cron-based automation

Languages: Python, TypeScript, JavaScript, Dart, SQL, GLSL, HTML5, CSS3

Frontend: React 18/19, Next.js (App Router, SSR/SSG, RSC), TanStack (Start, Router, Query), Vite, Redux/Zustand, Tailwind CSS, Radix UI / shadcn/ui, react-hook-form, Zod, Recharts, Framer Motion, GSAP, Three.js / React Three Fiber, WebGL2, WebXR

Backend & APIs: Node.js, Express.js, FastAPI, REST APIs, WebSockets (Socket.IO, Pusher), Server-Sent Events, Celery, MERN Stack

System Design: multi-tenant architecture, idempotent APIs, durable workflows and queues, fair-share scheduling, sandboxed code execution, object storage and signed URLs, usage-based billing, event-driven design, caching and CDNs

Mobile: Flutter, Dart, Riverpod, Capacitor, Firebase (Auth, Firestore, Storage, Cloud Functions, FCM)

Databases: MongoDB (Atlas, Mongoose ODM), PostgreSQL, pgvector, Firebase Firestore, Redis

Geospatial & Maps: Leaflet, OpenStreetMap, Overpass API, Haversine distance and bearing calculations

Cloud, DevOps & Testing: Docker & Docker Compose, Vercel, Cloudflare (CDN, Workers, Pages), AWS S3/R2, GitHub Actions (CI/CD), HTTP/CDN caching (ETag, stale-while-revalidate), Vitest, Playwright, Cloudinary

Web Platform & Performance: Web Workers, Web Audio API, GLSL shaders, profiling, frame and memory budgets, accessibility (ARIA, reduced motion)

Auth & Security: JWT, OAuth 2.0, Clerk, RBAC, bcrypt, Helmet, CSRF protection, rate limiting

Tools & Design: Git/GitHub, Figma, Postman, ESLint/Prettier, VS Code, Cursor

EXPERIENCE

AI Data Annotator | Perit AI *Aug 2026 – Present*

- Drive model training data curation: decide which datasets and examples enter training and for which capability or task, directly shaping model quality.
- Evaluate, rank and annotate LLM responses against detailed guidelines for accuracy, instruction-following, reasoning quality and safety.
- Write prompts and reference answers and run side-by-side model comparisons (preference ranking) that feed training and evaluation.
- Run quality checks and error analysis on data batches before training to keep labelling consistent and data clean.

AI Evaluation Engineer | Cognyzer *Feb 2026 – Jul 2026*

- Engineered reinforcement learning (RL) environments and Terminal-Bench-style tasks to stress-test frontier LLMs and agents, with deterministic setup, programmatic verifiers and reproducible scoring.
- Delivered **76 accepted, review-approved tasks in a single month**, authoring multi-step command-line and agentic tasks that met strict quality and difficulty standards.
- Benchmarked tasks against frontier models, primarily Claude Opus 5 and Claude Fable 5.1; analysed failure modes and wrote structured feedback to calibrate difficulty and surface capability gaps.

- Designed grading rubrics and acceptance criteria that made model scoring consistent and objective across reviewers.
- Leveraged Snorkel-based workflows to create, label and validate datasets against quality guidelines.

Co-Founder & Full-Stack Engineer | Sarveio *Jan 2026 – Present*

- Co-founded and engineered Sarveio, a multi-tenant restaurant operating system (QR guest ordering, POS, KDS, inventory, CRM & analytics) on a MERN-based stack (React, TanStack Start, Express, MongoDB Atlas), taking it from architecture to $2,000+ in generated revenue.
- Designed a multi-tenant data layer across 23 MongoDB collections with strict tenant isolation, atomic order-counter increments, and idempotency-key–protected order placement to eliminate duplicate charges under concurrent load.
- Engineered a layered CDN and caching strategy — Cloudflare edge caching, ETag revalidation, stale-while-revalidate headers, and in-memory request coalescing — serving public menu pages with fast, cookie-less, cacheable responses.
- Built real-time order and kitchen-display sync with Pusher Channels and Socket.IO, backed by JWT authentication (httpOnly cookies), 5-role RBAC, rate limiting, and CSRF protection to secure a production multi-tenant SaaS.
- Automated WhatsApp order confirmations and marketing messaging via a Puppeteer-driven browser automation pipeline, and shipped WebXR/AR menu-item previews using Three.js and ARCore/ARKit.

PROJECTS

Vernier | AI Benchmarking & Evaluation Platform (Founder) Live site

Frontend live: Next.js 16, React 19, TypeScript, Tailwind CSS v4, Framer Motion, Vercel. Backend architecture designed: TypeScript API, PostgreSQL, Redis, S3/R2, durable workflows, sandboxed workers, Python SDK

- Founded Vernier, a Benchmark-as-a-Service (BaaS) platform for on-demand benchmark environments, tasks and evaluation infrastructure for frontier models and AI agents; applying hands-on RL-environment, rubric and task-validation experience to the cost and fragmentation of building reproducible benchmarks.
- Shipped the live product site and an interactive evaluation dashboard UI with Next.js 16, React 19 and Tailwind v4, including scroll-linked Framer Motion walkthroughs, an accessible focus-trapped dialog, reduced-motion support, and GitHub-connected Vercel deploys.
- Architected the backend: a multi-tenant PostgreSQL schema (orgs, API keys, immutable benchmark revisions, runs, trials, artifacts, usage events), idempotent run creation, Redis rate limiting and S3/R2 trace storage with signed URLs.
- Designed durable evaluation workflows (admit, boot, execute, collect, grade, finalize) with fair-share scheduling, per-plan concurrency caps, isolated sandbox workers (browser and shell pools) and deterministic, replayable graders.
- Defined the public API and Python SDK contract (`Benchmark(...).evaluate(agent, parallel=256)`), plus Stripe usage metering, OpenTelemetry observability and a CI plan with Playwright.

Portfolio World | Interactive 3D Voxel Portfolio Game Live site

TypeScript, Three.js, WebGL2, GLSL, Web Workers, Web Audio API, Vite, Vitest, Playwright, GitHub Actions

- Built a playable Minecraft-style 3D portfolio with a custom engine (fixed-step physics, camera rig, collision) and a Node.js pipeline converting Minecraft Anvil/NBT map data into a compact binary voxel format, meshed in Web Workers with greedy meshing and baked ambient occlusion.
- Held 120 fps on desktop within hard budgets (max 250 draw calls / 2M triangles) using culling, quality tiers, half-float geometry (about 34% less GPU memory) and zero-allocation render loops.
- Wrote custom GLSL shaders and a post-processing stack (SSAO, bloom, god rays, depth of field) and generated all audio procedurally with the Web Audio API.
- Ensured accessibility (keyboard-only, reduced motion, touch controls, pre-rendered no-WebGL fallback) and quality via 97 unit tests, 7 Playwright E2E tests and CI.

Cura | AI health companion app

Flutter, Dart, Firebase, Riverpod, Claude API

- Built a cross-platform health app with AI-powered medical report interpretation, symptom Q&A, diet and exercise guidance, appointment management and family connect.
- Designed a clean mobile UI/UX and AI assistant flows backed by a secured API proxy to keep keys off the client.

ACHIEVEMENTS

- Co-founded and scaled **Sarveio**, a multi-tenant restaurant SaaS platform, from architecture to $2,000+ in generated revenue.
- Independently designed, built, and shipped 6+ production web and mobile applications spanning e-commerce, AI/automation, 3D/WebGL, and health-tech domains — across full-stack, mobile, and AI/ML engineering.
- Participated in 5+ hackathons, reaching the finalist round in several and winning the Algoverse Hackathon.