

Karan Rai

Software Engineer

kr3287@drexel.edu

484-521-6836

github.com/Krz-rai

linkedin.com/in/krz-rai

karanrai.live

EDUCATION

Drexel University - Bachelor of Science, Computer Science

September 2023 - Expected May 2028

Philadelphia, PA

Relevant Coursework: Data Structures, Algorithms, Object-Oriented Programming, Web Development

WORK EXPERIENCE

Incyte - Advanced Data Analytics Co-op (Applied AI) | Azure, LangChain, SQL, Power BI

April 2026 - September 2026

Chadds Ford, PA

- Designed a text-to-SQL service where the agent emits a JSON query request that is schema-validated and compiled by code into parameterized, read-only SQL, so model-written SQL never reaches the database
- Overhauled the document ingestion pipeline with parallel vision-model page parsing (20 at a time), Markdown normalization, and structure-aware chunking into Azure vector storage, raising retrieval accuracy from 68% to 95%+
- Shipped an internal search service for the Commercial team over 500+ documents, emails, and meeting recordings plus live SQL and Power BI data, cutting lookups from days to seconds

Looma - Founding AI Engineer | Next.js, TypeScript, Python, Computer Vision, Convex

September 2025 - April 2026

Philadelphia, PA

- Built the web and iOS hiring and assessment platform the world's largest school (City Montessori School) uses to evaluate 5,000+ candidates, cutting a 3–5 week paper process to under an hour
- Scaled the live exam platform to 1,200+ concurrent test-takers with zero incidents, keeping it reliable through crashes and unstable school Wi-Fi
- Enforced 100% statement, branch, function, and line coverage as a CI gate across 690+ test files (Vitest, Playwright E2E), so no change merges untested
- Engineered sandboxed code execution and auto-grading for 9 types of hands-on exams, from coding to spreadsheets, running candidate code safely on isolated, short-lived VMs
- Created a public REST API and HMAC-signed webhooks with idempotent request handling for two-way sync with the school's HR system, ingesting candidates and pushing results back with zero manual data entry
- Developed a multi-model LLM grading pipeline for handwritten, typed, and diagram answers in 5 languages, saving 3,500+ teacher-hours per hiring round

Compelo - Founding Full-Stack Engineer | Next.js, Express.js, C++, AWS, MongoDB

April 2025 - September 2025

Philadelphia, PA

- Delivered a 26-endpoint REST API as a Dockerized microservice on AWS with HMAC-signed webhooks, scoped API keys, and rate limiting, letting third-party hiring tools securely access candidate data
- Wrote 2,000+ unit and end-to-end tests (Vitest, Playwright) with CI gating via GitHub Actions on every push, keeping 90%+ coverage on core modules and catching regressions before deployment
- Implemented an ATS-style CRM with kanban pipeline, SLA alerting, RBAC, and bulk import, enabling recruiters to manage 1,000+ candidates across 5-stage hiring lifecycles
- Architected a multi-agent hiring platform with 7 specialized services for candidate discovery, scoring, and retrieval over resume and interview data, cutting a week-long evaluation cycle to under 5 minutes
- Devised Dynamic Reference Bucketing, a score-calibration method that fixes model scores bunching in a narrow band by ranking candidates via pairwise comparisons and Bradley-Terry modeling; matched recruiters on 22 of 23 decisions

SKILLS

- **Languages:** Java, C++, C#, Python, TypeScript/JavaScript, Swift, Kotlin, Go, SQL, PHP, Ruby, HTML/CSS
- **Frameworks:** React, Next.js, React Native, Node.js, Express.js, Spring Boot, .NET, SwiftUI, REST, GraphQL
- **Databases:** PostgreSQL, MySQL, SQL Server, MongoDB, DynamoDB, Redis, Convex, Kafka
- **Cloud/DevOps:** AWS (EC2/ECS/Lambda/S3), Azure, GCP, Docker, Kubernetes, Linux, Git, GitHub Actions, CI/CD
- **Engineering:** OOP/Design Patterns, Data Structures and Algorithms, Multithreading, Distributed Systems, Embedded
- **Testing/Workflow:** JUnit, Vitest, Jest, Playwright, TDD, Agile/Scrum, Code Review, Claude Code, Cursor, Copilot