

Reuben Roy

Software Engineer | AI Agents, Backend Systems & Native Apps

explosion.fun | github.com/reuben-roy | linkedin.com/in/reuben-roy

Profile

Software engineer building reliable AI workflows, native applications, and backend services. Combines Java/Spring Boot and C#/.NET experience with local inference, retrieval, browser automation, and explicit verification. Recent projects span macOS automation, job discovery infrastructure, remote AI assistance, and voice-driven productivity.

Experience

AI Engineer (Capstone) | Aligned Rewards

- Developed RAG features with LangChain, FastAPI, OpenAI embeddings, and Pinecone; implemented routing and fallback between OpenRouter and local models through Ollama.
- Built interfaces for model steering and AI workflow generation; documented integration decisions and failure behavior.

Graduate Teaching Assistant / AI Engineer | Arizona State University

- Built an AI auto-grader integrated with Canvas, structured rubrics, and PDF lab-result processing; supported self-hosted grading to keep coursework on university infrastructure.
- Reviewed software engineering submissions and explained design and testing feedback.

Software Engineer | RCKR Software - Trubridge / CPSI

- Developed C#/.NET backend services and SQL ETL pipelines for hospital data migration and FHIR interoperability.
- Optimized PostgreSQL access with Entity Framework; used Redis and asynchronous processing for large record-exchange workloads, with OAuth/JWT service authentication.

Full-Stack Developer | Tata Consultancy Services - IKEA

- Built Java/Spring Boot microservices on GCP for quotation generation, revisions, approvals, and order-status notifications.
- Optimized API latency with caching and SQL indexes; built concurrent workers and scheduled jobs, and contributed React checkout and 3D room-planner features.
- Maintained automated unit testing with JUnit and Mockito in CI/CD pipelines.

Technical Skills

- Languages: Python, Java, C#, TypeScript/JavaScript, Swift, Go, Kotlin, SQL.
- AI: Ollama, RAG, embeddings, structured output, tool calling, inference-provider fallback, Playwright/Patchright, LiveKit Agents, local speech processing.
- Applications: Spring Boot, .NET, FastAPI, Fastify, Hono, React, Next.js, React Native, SwiftUI, Electron, D3.js, Three.js.
- Data and infrastructure: PostgreSQL, SQLite, Redis, Flyway, Auth0/OIDC, Docker, AWS, GCP, Azure, Vercel, GitHub Actions.

Education

- Arizona State University | MS in Software Engineering | 2024-2026
- National Institute of Technology, Calicut | B.Tech | 2017-2021

Selected Projects

Job Applicator | Native macOS app - development preview

- Building a SwiftUI app with a bundled Python engine for queued Greenhouse and Ashby applications. Profile-first answers, reusable responses, per-field submission gates, and initial human review control automation.
- Integrates Keychain-backed inference credentials, Gmail security-code handling, and Auth0 PKCE sign-in. Cloud integrations require deployment configuration and validation.

Job Applicator API & Job Board Collector | Backend and discovery

- Implemented a Java 21 / Spring Boot 3 service with PostgreSQL account data, OIDC, Flyway, transactional outbox, telemetry, and a Next.js admin dashboard.
- Built a Python collector using shared Greenhouse/Ashby adapters, TOML source configuration, a SQLite publishing journal, bounded retries, and snapshot checks.
- Feed synchronization uses cursors, ETags, compressed page caching, and commit-ordered changes; hosted inference adds key-pool failover and per-user limits.

Auto-Apply | Local-first application automation

- Separated LLM question answering from deterministic browser filling and submission decisions. Added semantic answer recall, confidence gates, and post-submit verification.
- Built durable queues, bounded retries, audit logs, and human review paths; resume tailoring selects verified content and checks the generated PDF text layer.
- Technical write-up: explosion.fun/projects/auto-apply

Hired Gun | Remote AI task execution

- Built a one-use-code workflow that starts a macOS agent for remote assistance, with live operator visibility and structured completion reports.
- Kept model keys server-side behind per-session tokens; built recipient/operator CLIs and a portable agent plugin. TypeScript, Hono, PostgreSQL/Neon, PGlite.

NotAnother | Desktop accountability companion

- Built an Electron macOS overlay and menu-bar task interface with Markdown storage, planning tools, timers, and proactive nudges. Basic local intent handling works without a model key.

Local Voice Tools | CheckLister & Window for macOS

- CheckLister prototype combines LiveKit Agents, Ollama, faster-whisper, SQLite, and scheduled check-ins, with task estimates calibrated against completion time.
- Window for macOS combines SwiftUI, local speech synthesis, Parakeet transcription, and a Python/Ollama quiz evaluator.

Mobile, Browser & Data Projects

- Side-Track: React Native iOS weight-training app with fatigue tracking and Apple Health integration. Window Extension: calendar-aware focus blocking and idea capture with a React/TypeScript extension and Fastify backend.
- explosion.fun: Next.js/headless WordPress publishing and data lab. Built D3/Three.js visualizations, self-actualization scoring, and a coding-harness star-growth comparison with explicit data provenance.
- K2 Horizon experiment: Python CLI comparing inference latency and reasoning-effort settings through IFM's OpenAI-compatible API.

Full project details and additional work: explosion.fun/cv