

Anton Romeo

AI Full Stack Engineer

✉ antonr24710@gmail.com ☎ 12512500203 ⬇ Edmundston, NB, Canada

PROFILE

AI Full-Stack Engineer with 10 years of experience building production-grade web and mobile applications powered by modern AI systems. Specialized in LLM-driven features, RAG pipelines, and intelligent backend services, with a strong balance of web and mobile app development.

Experienced in designing and deploying scalable AI microservices, integrating LLMs into real user workflows such as chat assistants, document automation, and semantic search. Skilled in FastAPI, Node.js, cloud-native deployment on AWS, and delivering secure, low-latency systems that are built for real-world adoption. Known for writing clean, maintainable code and translating complex AI capabilities into reliable, user-focused products.

SKILLS

Languages: Python, Java, JavaScript, TypeScript, Go, SQL

AI & GenAI: LangChain, LangGraph, CrewAI, MCP, RAG, embeddings, semantic search, vector DBs (Pinecone, FAISS), prompt engineering, LLM routing (GPT-4, Claude, Mistral, Gemini), LangSmith, multi-agent workflows, speech I/O(Retell.ai), LiveKit, Deepgram, ElevenLabs

Backend: FastAPI, Flask, Spring Boot, Django, Node.js, Express, REST/GraphQL, Celery, Redis, WebSockets

Frontend: React, Next.js (App Router), TypeScript, Redux Toolkit, Zustand, responsive UIs

Mobile: Swift, Kotlin, Flutter, React Native

Data: PostgreSQL, MySQL, DynamoDB, Redshift, ETL pipelines, Supabase, real-time analytics

Cloud & DevOps: AWS (S3, Lambda, Kinesis, Step Functions, EventBridge, RDS, DynamoDB, EMR, IAM, ECS/EKS), GCP, Docker, Kubernetes, Terraform, CI/CD (GitHub Actions), Vercel, Railway

Practices: SOC 2 compliance, AES-256 encryption, TDD, Agile/Scrum, Jira, Git, system observability, cross-functional collaboration

PROFESSIONAL EXPERIENCE

AI Software Engineer, BeckTek

03/2022 – 05/2025 | Riverview, Canada

- Designed and delivered LLM-driven AI systems including conversational assistants, document Q&A, and intelligent data extraction pipelines, enabling users to query and automate workflows over large internal knowledge bases.
- Built and optimized retrieval-augmented generation (RAG) pipelines using embeddings and vector databases to produce accurate, context-aware responses from unstructured documents and customer data in production environments.
- Engineered full-stack development across web and mobile, building AI-enabled user interfaces with React/Next.js and integrating real-time AI features into mobile applications via REST and WebSocket APIs.
- Architected FastAPI-based AI microservices for inference, retrieval, and orchestration, containerized with Docker and deployed on AWS ECS to support scalable, low-latency workloads.
- Implemented cloud infrastructure, CI/CD pipelines, and monitoring using AWS services (ECS, S3, CloudWatch), improving deployment reliability, observability, and operational efficiency for AI services.

Full-Stack Engineer, Sedin Technologies

09/2020 – 02/2022 | Sunnyvale, CA

- Built and maintained full-stack web applications using Next.js, React, and TypeScript, delivering performant, SEO-friendly interfaces for enterprise and SaaS platforms.
- Introduced early AI-driven features such as intelligent search, text summarization, and rule-based recommendation logic to enhance user experience and data discoverability.
- Developed backend services with Flask and Django, designing RESTful APIs and business logic to support web and mobile clients at scale.
- Implemented secure authentication and authorization using JWT and role-based access control, ensuring safe access for multi-tenant applications.
- Deployed and operated applications on AWS, leveraging Docker for containerization and CI pipelines to support reliable staging and production releases.
- Collaborated closely with mobile teams to integrate backend APIs with applications, maintaining consistent API contracts and optimizing response times.

- Full Stack Developer, Aatio** 04/2018 – 08/2020 | Saint John, Canada
- Delivered responsive web applications using React, JavaScript, and REST APIs for business and internal tooling use cases.
 - Built backend services with Node.js and Django, integrating relational databases and third-party APIs.
 - Developed mobile-friendly web interfaces and supported early mobile app integrations with backend services.
 - Optimized database queries and backend logic, improving application performance and reducing response latency.
 - Worked closely with product and design teams to translate requirements into reliable, user-focused features.

- Software Intern, MNP Digital** 09/2016 – 03/2018 | Saint John, Canada
- Assisted in developing internal web tools and dashboards using JavaScript, Python, and SQL.
 - Supported backend development tasks, including API enhancements, database maintenance, and bug fixes.
 - Participated in requirements gathering, testing, and documentation as part of agile project teams.
 - Gained foundational experience in software engineering best practices, version control, and production support.

EDUCATION

Bachelor's Degree, University of New Brunswick 2012 – 2016