David Katz

david.katz430@gmail.com | +1 289-380-0100 | [LinkedIn](#) | [dkatz.ca](#) | [Github](#)

Work Authorization: Canadian citizen — TN visa eligible (no employer sponsorship required for U.S. roles)

Summary

Senior Software Engineer with 4+ years of experience in architecting and deploying production-grade AI systems and full-stack applications. Demonstrated success in launching LLM platforms, optimizing ML workflows in cloud and edge environments, and securing multi-cloud infrastructure at scale. Skilled in building cloud-native apps, leading AI/ML integration, and improving efficiency across teams with cutting-edge technologies.

Technical Skills

- **Languages**: Python, Java, Go, TypeScript, JavaScript, SQL, C#, Bash/Shell, Ruby, Scala, Elixir
- **Frameworks**: FastAPI, Spring Boot, Flask, Django, Node.js, .NET, Ruby on Rails, SQLModel
- **Frontend**: React.js, Vue.js, Angular.js, HTML, Tailwind CSS, Vite, Cypress, jQuery, Bootstrap
- **Cloud**: Azure, AWS, GCP, Serverless, Kubernetes (AKS, EKS), Databricks, Supabase, Vercel
- **Data & ML**: Apache Spark, Hadoop, Azure ML, TensorFlow, PyTorch, scikit-learn, Pandas, NLP, OpenCV
- **DevOps**: Docker, Terraform, Helm, Jenkins, GitHub Actions, CI/CD, Kafka, Grafana, Prometheus, Splunk
- **Databases**: PostgreSQL, SQL Server, MySQL, Redis, MongoDB, Snowflake, CockroachDB
- **Security**: IAM, HashiCorp Vault, OAuth 2.0, Azure Defender, AWS Security Hub, Wiz, Palo Alto, Cloudflare

Professional Experience

CIBC — *Toronto, Canada*
Senior Software Engineer — *Nov 2024 – Present*

- Deployed a scalable **LLM assistant** for **48,000+ employees**, supporting secure workflows across the enterprise.
- Integrated OneDrive file upload, chat history, and CIBC-specific lexicon recognition, boosting adoption **3x** across teams.
- Designed and implemented a **Retrieval-Augmented Generation (RAG)** system for teams with large document repositories.
- Built a **Teams Bot AI Agent** for meeting transcription and summarization using **Azure AI NLP**, reducing manual effort by **95%**.
- Mentored junior engineers on **LLM integration** and scalable backend architecture, accelerating delivery and onboarding speed.
- **Stack:** Python, FastAPI, Alembic, SQL Server, Redis, Docker, Azure AKS, OpenAI API, Vue.js

AI Spectral — *Toronto, Canada*
Senior Software Engineer — *Jul 2024 – Dec 2024*

- Led a team of 5 to deploy of **40+ ML models** on **500+ edge medical devices**, while maintaining **99.9% SLA uptime**.
- Delivered a real-time analytics API using **FastAPI + Grafana**, cutting North American clinical trial feedback cycles by **40%**.
- Engineered a **low-latency scheduler** in Go using Unix sockets + dynamic Docker orchestration, reducing device lag by **60%**.
- Ran experiments on **Azure ML** using **Python, scikit-learn, PyTorch, Pandas**, and **NumPy**, ensuring high model performance.
- **Stack:** Python, FastAPI, Go, Grafana, Loki, Docker, Azure ML, PyTorch, Pandas, scikit-learn, NumPy

City National Bank / RBC — *Los Angeles, CA / Toronto, Canada*
Senior Software Engineer — *May 2023 – May 2024*

- Built a **Cloud Security Exemptions Platform** to track **20K+ exceptions** across Azure, AWS, GCP, and HashiCorp Sentinel.
- Delivered a cloud-agnostic solution with **Azure SDK, Boto3, and Open Policy Agent**, automating compliance for $2T+ assets.
- Developed an **AI change monitor** for cloud security (Azure, AWS, GCP) blog updates, saving **120+ hours/year per architect**.
- Mentored an intern in **REST API development** and **cloud security design**, resulting in independent production feature delivery.
- **Stack:** Python, FastAPI, Azure AKS, Azure SDK, AWS Boto3, Kubernetes, SQL, HashiCorp Sentinel, GCP, OpenAI API

FINTRAC — *Toronto, Canada*
Software Engineer — *Jan 2023 – May 2023*

- Delivered a bilingual (EN/FR) **financial crime reporting API** (Vue + Spring Boot) in a **Top Secret Clearance** federal setting.
- Architected **Spark-based ETL pipelines** to transform high-volume financial data for audit/compliance readiness.
- **Stack:** Java, Spring Boot, Apache Spark, Hadoop, SQL Server, Vue.js

Just Web Agency — *Toronto, Canada*
Software Engineer (Entry Level) — *Sep 2021 – Feb 2023*

- Built a B2B lead-gen platform (React + FastAPI), automating scraping, validation, scoring, **increasing qualified leads by 200%**.
- Reduced hosting costs **45%** by implementing Kubernetes autoscaling on resource-limited DigitalOcean environments.
- Developed an AI-powered image search tool using **AWS Rekognition**, serving **3K+ monthly users**.
- **Stack:** React.js, Python, FastAPI, PostgreSQL, Docker, Kubernetes, AWS Rekognition, DigitalOcean

Education

Carleton University — *Ottawa, Canada*
Bachelor of Computer Science (Honours), Minor in Business — *2019 – 2022*

Graduated with Distinction (**3.8/4.0** GPA), Dean's List (All Years), A.D. Dunton Scholarship Recipient

Certifications

[Azure Fundamentals (2023)](#) | [Azure AI Fundamentals (2023)](#) | [AWS CCP (2024)](#) | [GCP Cloud Digital Leader (2024)](#)