

SIMON BOUCHARD

Montreal, QC • 438-392-6707 • simon.bouchard31@gmail.com • simonbouchard.space • linkedin.com/in/simon-bouchard-54580b339

PROFESSIONAL SUMMARY

Machine learning developer with production experience building recommendation systems and agentic AI applications. Built real-time recommendation engine with dual-factor architecture (ALS collaborative filtering and attention-pooled embeddings) achieving sub-50ms latency, and multi-agent conversational AI system with intent routing and a multi-stage recommendation pipeline. AWS Certified Machine Learning - Specialty and Generative AI Developer - Professional. Self-taught developer with strong foundation in backend development, cloud infrastructure, and applied AI. Business administration background (BBA, HEC Montreal) enables effective bridge between technical implementation and business requirements. Seeking ML role to apply demonstrated capabilities in production environment.

CERTIFICATIONS

AWS Certified Generative AI - Professional (AIP-C01) | Early Adopter (First 5000) | Jan 2026

AWS Certified Machine Learning - Specialty (MLS-C01) | Score: 890/1000 | Jan 2026

TECHNICAL SKILLS

Machine Learning & AI: PyTorch, scikit-learn, FAISS, LangGraph, Sentence-Transformers, collaborative filtering, neural networks, attention mechanisms, RAG systems, agentic frameworks, model deployment, real-time inference

Cloud & Infrastructure: AWS, Docker, Nginx, CI/CD (GitHub Actions), Linux/Bash, Redis

Backend Development: Python, FastAPI, Git, RESTful APIs, microservices architecture, unit testing

Data Engineering: SQL (MySQL), NoSQL (MongoDB), ETL/ELT pipelines, Kafka, Spark, pandas, NumPy

Observability: Prometheus, Grafana, Jaeger, OpenTelemetry

Languages: French (native), English (fluent)

TECHNICAL PROJECTS

Production ML Recommendation System recsys.simonbouchard.space

- Engineered real-time recommendation API achieving sub-50ms latency for personalized book recommendations, deployed with a model server architecture across dedicated containers, enabling independent scaling and deployment of each model component
- Architected dual-factor recommendation system: ALS collaborative filtering for warm recommendations and behavioral similarity, attention-pooled subject embeddings trained with dual regression and contrastive loss for cold-start recommendations and content similarity, with hybrid similarity combining both approaches
- Built automated MLOps pipeline with recall@30 quality gate blocking artifact promotion on regression, versioned artifact store, failure notifications, and pointer-file-based active version tracking for zero-downtime hot-reload across 5 model servers
- Designed and managed MySQL database with automated data ingestion pipelines and schema optimization for ML workloads
- Integrated Meilisearch for full-text search across 250,000+ books with millisecond-range response times
- Deployed on cloud infrastructure with CI/CD automation, SSL security, monitoring, and firewall hardening (UFW, Fail2ban)

Multi-Agent Conversational Recommendation System recsys.simonbouchard.space/chat

- Developed production multi-agent system using LangGraph with intent-routing across four specialized agents (recommendation, docs, web, response); the recommendation agent runs a multi-stage pipeline (planning → retrieval → selection → response) integrating recommendation APIs, semantic search, user personalization, and external knowledge sources with streaming responses
- Implemented retrieval-augmented generation with custom in-memory vector store using FAISS HNSW index for semantic search across book catalog, grounding chatbot responses to reduce hallucinations
- Designed event-driven embedding generation pipeline using Kafka for message queuing and Spark for batch processing of LLM-generated item descriptions
- Implemented comprehensive test and evaluation suites ensuring system reliability and quality and enabling rapid iteration

WORK EXPERIENCE

Grocery Clerk | Metro Renaud Poissant, Montreal | Summer 2022

Co-owner | Gazonnet Royal (Lawn Care Business) | 2016-2021

Managed operations including task allocation, equipment management, marketing, and customer service

EDUCATION

Bachelor of Business Administration (BBA), Bilingual Program | 2021-2025

HEC Montreal - Information Technology Specialization

PROFESSIONAL DEVELOPMENT

Advanced Machine Learning: MIT Introduction to Deep Learning, fast.ai Practical Deep Learning for Coders (Parts 1 & 2)

AI & Generative AI: DeepLearning.AI - AI for Everyone, Generative AI for Everyone, Agentic AI

Software Engineering: MIT The Missing Semester of Your CS Education

Backend Development: FreeCodeCamp (Backend APIs, Quality Assurance, Information Security, Scientific Computing, Data Analysis, Machine Learning with Python)