Aviral Garg

Vancouver, British Columbia, Canada · avral.garg@icloud.com · in/aviralgarg · aviralgarg.com/

SUMMARY

A highly creative Full Stack Software Engineer with over 5 years of experience specializing in the design, development, & maintenance of scalable, cloud-based applications that leverage ML & Generative AI. Proven expertise in building end-to-end, AI-driven solutions & automation pipelines on AWS that accelerate product development & enable rapid experimentation.

SKILLS

- **AI & Data Visualization:** Generative AI (OpenAI, Anthropic, Gemini, Ollama), Agentic Orchestration & Frameworks (LangChain, CrewAI, Google Agent Development Kit, Autogen), RAG, NLP, LLMs, Model Deployment (SageMaker Endpoints, Batch Transform), Intent Recognition, Query Parsing; **Python** (Flask, FastAPI), **Visualization** (Plotly, D3.js, Matplotlib)
- **Cloud Computing: AWS** (Lambda, SageMaker, API Gateway, DynamoDB, Cognito, AWS CDK & CloudFormation (IaC), SAM, SSM, Step Functions, S3, EC2, ECS, CloudWatch, X-Ray, SNS, SQS, Glue, Athena), Azure, GCP, Firebase

EXPERIENCE

Software Engineer

Amazon, Beauty Tech Team February 2021 – March 2025, Vancouver, BC

- Led a team of 8 engineers, designers, & data scientists to deploy serverless data pipelines with 50% increased efficiency.
- Leveraged AWS Step Functions, Lambda, DynamoDB, SageMaker, & Athena; automatically scaled for growth, processed 40TB/day of product catalog data, & served 40M+ monthly customers across 15+ countries.
- Built 4-staged AWS CDK CI/CD pipelines using CloudWatch & X-Ray for granular system monitoring & bottleneck tracing. Received org-wide accolades for robust documentation & operational readiness for every step of the SDLC.
- Built internal-documentation-powered AI agents to provide instant answers to CI/CD admin questions, identify documentation gaps, & flag outdated content. Improved question resolution time by 82%, with projected savings of 240+ dev-hrs/month from integration with the internal admin ticketing system.
- Led an organization-wide documentation workshop, establishing a systematic platform for subject-matter-expert reference & reaching 100% completion of product & CI/CD documentation for the Beauty Tech organization.
- Spearheaded CI/CD-driven store-page-builder project for Amazon clients using React & Python.
- Integrated external microservices & A/B testing framework, increasing client customization capabilities by 300%.
- Automated monthly business reports for the recommendations team, reducing manual reporting time by 90%.
- Led a team of 4 to win first place at the Amazon 2022 Hackathon with an Augmented Reality iOS makeup-tutorial app.

Software Engineer

Amazon, Prime Pantry Team June 2020 – February 2021, Vancouver, BC

- Architected robust Kotlin applications achieving 100% test coverage for Prime Pantry API development.
- Optimized API performance, reducing page-load impact by 37% & increasing add-to-cart conversion rates by 2.5%, which directly contributed to over $3 million monthly revenue for the Prime Pantry platform.
- Cut setup time by 75% & saved 3,528 dev-hrs/year by automating Prime Pantry API developer onboarding & environment setup while eliminating setup-related errors.
- Built automated CI/CD pipelines in AWS CDK, saving 12+ dev-hrs/week & cutting deployment errors by 95%.
- Delivered $30K in annual cost savings through systematic AWS CI/CD optimization & usage policy implementation.
- Led organization-wide CodeGuru Profiler adoption across 7 projects, reducing on-call incidents & maintenance hours.

AI Software Engineer

T4G Limited October 2018 – December 2019, Vancouver, BC

- Architected NLP chatbots with full CI/CD, leveraging Azure DevOps, using Node.js & SQL, securing $100K+ contracts.
- Demonstrated AI capabilities to enterprise clients, establishing the company's reputation in the AI/ML space.
- Developed mission-critical C# desktop applications & SSIS data integration packages serving 50+ million users globally.
- Optimized promo-code generation, achieving 40% efficiency improvement & enabling 4 additional high-value contracts.
- Automated extraction of 10 years of Git version history from 15 applications, saving 7 dev-hrs/week with Python scripts.

Software Engineer

Texavie Technologies Inc.   **December 2017 – May 2018, Vancouver, BC**
 · Engineered cross-platform mobile, desktop, & mixed-reality headset proof-of-concepts in Unity3D.
 · Demonstrated wearable prototypes, securing investor interest & validating product-market fit for AR/VR technologies.
 · Developed critical prototype debugging tools reducing calibration time by 60% & enabling on-time product launch.
 · Built cross-platform game development pipelines for Windows, macOS, & Linux, doubling the target customer base.

PROJECTS (2025)

FoodFlow Intelligence Platform – Multi-Agent Food Delivery Analytics
 · Production ecosystem integrating LangChain, CrewAI, Google Agent Development Kit, Autogen for food delivery analytics.
 · Conversational AI enabling natural English queries with automated Plotly/D3.js/Matplotlib visualizations.
 · Streaming analytics engine with SQL data warehouse/lake integration and external API connections.
 · Communication framework with message passing, load balancing, auto-scaling for optimal performance.
 · Kubernetes deployment with CI/CD pipeline, monitoring, security for enterprise operations.
 · Prompt optimization laboratory with A/B testing, intent classification, route optimization algorithms.

Bob – AI-Powered Development Assistant
 · Built a multi-threaded personal voice assistant with 8 operational modes for hands-free development & accessibility.
 · Reduced Agentic AI platform costs by 40x through extensive prompt engineering & optimization techniques.
 · Implemented zero speech data loss during processing with real-time speech-to-tool-calls conversion.
 · Integrated OpenAI Whisper for local ML-based speech recognition with real-time audio processing pipeline.

Live Transcriber (7.8K downloads) – Real-Time Speech Processing
pypi.org/project/livetranscriber
 · Deployed a multi-threaded open-source Python package with real-time speech transcription & zero data loss.
 · Architected a single-file WebSocket wrapper around Deepgram API with async/await supporting sync & async callbacks.
 · Implemented comprehensive resource management with automatic cleanup, graceful shutdown, & error handling.
 · Created a seamless PyPI distribution with proper dependency management, enabling pip installation & integration.

Pre-bunker health communications system – Health Misinformation Prevention
github.com/gaviral/pre_bunker_health_communications_system
 · Built AI agents for health misinformation prevention, simulating public reactions & generating evidence-backed prebunks.
 · Created a comprehensive prototype with implementation across 19 versions achieving 65-80% risk reduction.
 · Built a 5-stage pipeline: Claim Extraction, Risk Assessment, Audience Simulation, Evidence Validation, Countermeasures.
 · Developed 12 specialized agent personas for comprehensive audience simulation with multi-source evidence validation.
 · Created a FastAPI web interface with async processing & real-time risk assessment.
 · Developed a novel Misinterpretability@k metric for quantifying health communication risk.

Resume Coach – Full-Stack AI Application
coach.aviralgarg.com
 · Built full-stack GPT-powered AI app with LangChain on AWS Lambda, API Gateway, DynamoDB, S3 & CloudFront.
 · Created a React 19 + TypeScript frontend with a CI/CD pipeline using GitHub Actions, deploying to production.
 · Implemented session persistence with DynamoDB TTL & contextual follow-up chat system.

EDUCATION

Machine Learning Program
Interview Kickstart · 2025
 · *Coursework: EDA & Feature Engineering, Regression & Classification, Model Evaluation, Model Fine-Tuning, Supervised, Unsupervised & Reinforcement Learning, Neural Networks, NLP, Agents Frameworks.*

B.A.Sc., Computer Engineering (Software Engineering Major)
University of British Columbia · Vancouver, BC · 2019