

Juan Andres Bernal Gil

M.Sc. Systems and Computing Engineering | Backend & AI Engineer | Technical Leadership

juan.bernal.2004.gil@gmail.com | Bogotá, Colombia

<https://linkedin.com/in/juan-andres-bernal/> | <https://github.com/JuanBernal13>

PROFESSIONAL SUMMARY

Backend and AI Engineer with experience owning backend architecture, designing distributed systems, and delivering cloud-native services in regulated enterprise and healthcare environments. Combines Java/Spring Boot, Node.js, AWS, Kubernetes, data platforms, and LLM/RAG systems with technical leadership across architecture, integrations, migrations, and production operations.

PROFESSIONAL EXPERIENCE

Scotiabank

February 2026 – Present · 7 mos

Software Engineer Associate

- Develop backend services and integrations for banking technology teams, supporting financial operations with reliable and maintainable software. Work with Java, Spring Boot, AWS, CI/CD pipelines, and microservice-based architectures within a regulated enterprise environment. Collaborate with cross-functional teams to translate business requirements into scalable technical solutions.

Tech Stack: Java, Spring Boot, AWS, Microservices, CI/CD, Jenkins, ArgoCD, Jira

Nebula Medical

August 2025 – February 2026 · 6 mos

AI Engineer / Backend Engineer

- At Nebula Medical, an AI healthcare startup valued at USD 800,000, led backend architecture of an AI-driven medical platform using LangChain for multi-model orchestration across Claude, GPT-5, and Gemini, reducing average AI response latency by 35% through intelligent model routing and caching. Designed and implemented a RAG-based conversational memory system with vector search in Pinecone, enabling persistent clinical chat context, reducing AI hallucination rate, and improving diagnostic response accuracy. Engineered a scalable Data Access Layer (DAL) with Prisma 7 and PostgreSQL, optimizing complex queries through advanced indexing and caching strategies, sustaining p99 < 500ms and p95 < 200ms under concurrent clinical workloads.
- Implemented end-to-end encryption and HIPAA-compliant security protocols across all patient data flows, achieving no critical findings in our internal audits. Integrated 3+ healthcare interoperability APIs (FHIR R4, SaludTools) enabling seamless data exchange across provider systems and reducing integration overhead by 50%.
- Architected and deployed a multi-tenant authentication system using Clerk, reducing onboarding friction for new healthcare organizations by 70%. Implemented an API Gateway with AWS API Gateway and Lambda functions, reducing API response latency by 80% and decreasing application bundle size.

Tech Stack: Next.js, LangChain, RAG, Prisma, PostgreSQL, HIPAA, React, AWS Lambda, Deepgram, FHIR, Pinecone, Confluence, Jira, Clerk

Caseware

June 2024 – July 2025 · 1 yr 1 mo

Intern - Software Developer I

- Engineered integrations within the imports/bindings module of the Caseware platform, a mission-critical data ingestion layer used by 475,000+ professionals across 130 countries, ensuring reliable, schema-compliant data mapping at enterprise scale. Developed critical backend microservices in Java Spring Boot and Node.js and deployed serverless infrastructure on AWS through CDK, optimizing operational costs. Launched CDK projects from development to production environment. Led massive data migrations (500k+ records) ensuring transactional atomicity and reliability. Improved data access patterns latency time by 40%.
- Implemented reactive components in Angular, reducing application load time by 30% through lazy loading and bundle optimization techniques. Configured orchestrated deployments in Kubernetes. Responded to on-call production incidents with rapid triage and resolution for clients.
- Led Scrum ceremonies for a 7-developer team, ensuring continuous value delivery for the imports team. Worked with AWS services such as S3, DynamoDB, EC2, EKS, RDS.

Tech Stack: Java, Spring Boot, AWS CDK, Angular, Kubernetes, Scrum, DynamoDB, S3, Jira, Agile, Copilot

EDUCATION

Universidad de los Andes

M.Sc. Systems and Computing Engineering

Bogotá, Colombia

2026 – Current

Universidad de los Andes

B.S. Systems and Computing Engineering

Bogotá, Colombia

Jan 2021 – June 2025

GPA: 4.21 / 5.0

Universidad de los Andes

B.S. Industrial Engineering

Bogotá, Colombia

Jan 2022 – June 2025

GPA: 4.21 / 5.0

PROJECTS

Trove: Local Open-Source Search Engine & MCP Server

A high-performance local search engine that indexes content across GitHub, Notion, Slack, and local files into a unified vector database. Includes an MCP server implementation that optimizes AI agent workflows. Technologies: TypeScript, Node.js, SQLite, Electron, React, MCP, Transformers.js, Ollama. <https://github.com/antoinebecker10-afk/Trove>

Aura AI: Intelligent Talent Matching Platform with NLP, GNN & LLMs

An end-to-end ML platform that transforms talent acquisition by replacing keyword matching with semantic fit. Features multi-modal document understanding, learned skill taxonomy via Knowledge Graphs (Neo4j), and contrastive candidate-job matching using GNN. Technologies: FastAPI, Next.js, PyTorch (GNN), Neo4j, Qdrant, PostgreSQL, MongoDB, Redis/Celery, LangChain. <https://github.com/JuanBernal13/aura-ai>

CV Generator

A modular Python library for generating professional CVs in DOCX and PDF format. Technologies: Python, docx, reportlab. https://github.com/JuanBernal13/cv_generator

Sports Betting Prediction Engine - ML Pipeline with Kelly Criterion

ML pipeline that implements temporal features from historical match data, trains and calibrates ensemble models (Random Forest, XGBoost, LightGBM), and sizes bets using the Kelly Criterion. Technologies: Python, scikit-learn, XGBoost, LightGBM, pandas, NumPy, matplotlib, seaborn. <https://github.com/JuanBernal13/bet-lck-2025>

Financial Market Alert WebSocket System

Real-time financial market alert system using WebSocket technology. Implements event-driven architecture with Redis for message brokering and PostgreSQL for data persistence. Technologies: TypeScript, NestJS, Redis, PostgreSQL, WebSocket, Docker. <https://github.com/JuanBernal13/web-sockets-finance-yahoo>

EV Charging Station Multi-Objective Optimizer

Developed a high-performance optimization system in Java and C++ for Electric Vehicle charging infrastructure. The system leverages multi-objective algorithms to minimize energy costs while maximizing customer service levels. Technologies: Java, C++, Optimization Algorithms, Data Modeling.

Enterprise Management Platform (Hogar de Abuelos)

A comprehensive management platform for facility administration, streamlining operations, records management, and resource allocation. Technologies: Python, Django, SQLite, HTML/CSS. <https://github.com/je-lopezul/hogarAbuelos>

Microservice orchestration using Spring Boot and Docker

Architected a scalable distributed system using Java Spring Cloud (Eureka, Config, Gateway) within an Nx monorepo. Implemented asynchronous messaging via RabbitMQ and resilience patterns with Resilience4j. Technologies: Spring Boot, Docker, AWS.

RESEARCH

Biological Predictive Modeling (B.S. Thesis)

2025 - I

Developed predictive models for complex biological data using machine learning techniques, applying advanced preprocessing (PCA), class balancing, and Deep Learning (TensorFlow/Keras) to achieve 98% accuracy.

<https://hdl.handle.net/1992/76592>

Analytics Forum 2025

2025 - I

Showcased undergraduate thesis findings at the 2025 Analytics Forum, presenting advanced machine learning applications in bioinformatics to an audience of industry experts and researchers.

COPA Research Group Presentation

2025 - I

Speaker at the Center for Optimization and Applied Probability (COPA) to discuss the integration of neural networks and probabilistic modeling in biological systems.

SKILLS

Data Science: Python, pandas / NumPy, scikit-learn, Statistical Modeling, Feature Engineering, Data Visualization, Experiment Design

Machine Learning & AI: LLM Development, RAG Systems, LangChain, Prompt Engineering, PyTorch, TensorFlow / Keras, XGBoost / LightGBM, Graph Neural Networks

Data Engineering: Data Pipelines, ETL / ELT, SQL, Data Modeling, Apache Airflow, RabbitMQ, Event-Driven Data Flows

Databases: PostgreSQL, MySQL, MongoDB, DynamoDB, Redis, Pinecone, Neo4j, Qdrant

MLOps & Cloud: AWS, Docker, Kubernetes, GitHub Actions, Model Serving APIs, Monitoring & Observability

Backend & APIs: FastAPI, Node.js / NestJS, Java / Spring Boot, REST APIs, GraphQL, Microservices

Soft Skills & Tools: English, Technical Communication, Scrum, Design Thinking, Confluence & Jira

ACHIEVEMENTS & AWARDS

Finalista - Quala Business Challenge

2025-1

Finalist in the Quala Business Challenge 2025-1, competing against top business and engineering students in a case-based competition.

ICFES Saber 11 Score

2020

Achieved a score of 466/500 on Colombia's national high school exit exam.

LANGUAGES

English (Fluent C1), Spanish (Native)