

# Juan Bernal

Software & Industrial Engineer | AI Enthusiast

[juan.bernal.2004.gil@gmail.com](mailto:juan.bernal.2004.gil@gmail.com) | <https://linkedin.com/in/juan-andres-bernal/> | <https://github.com/JuanBernal13> |  
Bogotá, Colombia

## PROFESSIONAL SUMMARY

---

I am a Software and Industrial Engineer specializing in the architecture of high-scale distributed systems and AI-driven ecosystems. By integrating Industrial Engineering principles, I treat software development as a rigorous optimization process where efficiency, scalability, and long-term reliability are engineered into the foundation.

## PROFESSIONAL EXPERIENCE

---

### AI Engineer / Backend Engineer | Nebula Medical

*August 2025 – February 2026*

- 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 (Pinecone), enabling persistent clinical chat context and 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 API Gateway using AWS API Gateway and Lambda functions, reducing API response latency by 80% and reducing application bundle size.

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

### Intern - Software Developer I | Caseware

*June 2024 – July 2025*

- 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

---

**B.S. Systems and Computing Engineering** | *Universidad de los Andes* | Bogotá, Colombia | Jan 2021 – June 2025 | GPA: 4.21 / 5.0

**Relevant Coursework:** *Transactional System, Software architecture, Data structures and algorithms, Business Intelligence, API development, Object-oriented programming, Web development*

**B.S. Industrial Engineering** | *Universidad de los Andes* | Bogotá, Colombia | Jan 2022 – June 2025 | GPA: 4.21 / 5.0

**Relevant Coursework:** *Advanced Optimization, Metaheuristics, Probabilistic models, Discrete simulation, Probabilities and statistics*

## KEY PROJECTS

---

**Trove: Local Open-Source Search Engine & MCP Server** [\[link\]](#) -- 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.

*Tech: TypeScript, Node.js, SQLite, Electron, React, MCP, Transformers.js, Ollama*

**Aura AI: Intelligent Talent Matching Platform with NLP, GNN & LLMs** [\[link\]](#) -- 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.

*Tech: FastAPI, Next.js, PyTorch (GNN), Neo4j, Qdrant, PostgreSQL, MongoDB, Redis/Celery, LangChain*

**CV Generator** [\[link\]](#) -- A modular Python library for generating professional CVs in DOCX and PDF format.

*Tech: Python, docx, reportlab*

**Sports Betting Prediction Engine - ML Pipeline with Kelly Criterion** [\[link\]](#) -- 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.

*Tech: Python, scikit-learn, XGBoost, LightGBM, pandas, NumPy, matplotlib, seaborn*

**Financial Market Alert WebSocket System** [\[link\]](#) -- Real-time financial market alert system using WebSocket technology. Implements event-driven architecture with Redis for message brokering and PostgreSQL for data persistence.

*Tech: TypeScript, NestJS, Redis, PostgreSQL, WebSocket, Docker*

**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.

*Tech: Java, C++, Optimization Algorithms, Data Modeling*

**Enterprise Management Platform (Hogar de Abuelos)** [\[link\]](#) -- A comprehensive management platform for facility administration, streamlining operations, records management, and resource allocation.

*Tech: Python, Django, SQLite, HTML/CSS*

**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.

*Tech: Spring Boot, Docker, AWS*

## RESEARCH

---

**Biological Predictive Modeling (B.S. Thesis)** [\[link\]](#) - 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.

**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.

## TECHNICAL EXPERTISE

---

**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

**Languages:** English (Fluent C1), Spanish (Native)

## 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.