

Juan Bernal

Ingeniero de Sistemas e Industrial | Entusiasta de IA

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

PERFIL PROFESIONAL

Soy Ingeniero de Sistemas e Industrial, especializado en la arquitectura de sistemas distribuidos de alta escala y ecosistemas impulsados por IA. Al integrar principios de Ingeniería Industrial, abordo el desarrollo de software como un proceso riguroso de optimización, donde la eficiencia, la escalabilidad y la confiabilidad de largo plazo se diseñan desde la base.

EXPERIENCIA PROFESIONAL

Ingeniero de IA / Ingeniero Backend | Nebula Medical

Agosto 2025 – Febrero 2026

- Lideré la arquitectura backend de una plataforma médica impulsada por IA, usando LangChain para orquestación multi-modelo con Claude, GPT-5 y Gemini, reduciendo la latencia promedio de respuesta en 35% mediante enrutamiento inteligente y caché.
- Diseñé e implementé un sistema de memoria conversacional basado en RAG con búsqueda vectorial en Pinecone, habilitando contexto clínico persistente y reduciendo la tasa de alucinaciones de la IA, con mejora en la precisión de las respuestas diagnósticas.
- Construí una capa escalable de acceso a datos (DAL) con Prisma 7 y PostgreSQL, optimizando consultas complejas mediante indexación avanzada y estrategias de caché, con p99 < 500 ms y p95 < 200 ms bajo cargas clínicas concurrentes.
- Implementé cifrado de extremo a extremo y protocolos de seguridad alineados con HIPAA en los flujos de datos de pacientes, logrando cero hallazgos críticos en auditorías internas.
- Integré más de 3 APIs de interoperabilidad en salud (FHIR R4, SaludTools), facilitando el intercambio de datos entre sistemas de proveedores y reduciendo el esfuerzo de integración en 50%.
- Arquitecté y desplegué un sistema de autenticación multi-tenant con Clerk, reduciendo la fricción de incorporación para nuevas organizaciones de salud en 70%.
- Implementé un API Gateway con AWS API Gateway y funciones Lambda, reduciendo la latencia de respuesta de APIs en 80% y disminuyendo el tamaño del bundle de la aplicación.

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

Practicante - Desarrollador de Software I | Caseware

Junio 2024 – Julio 2025

- Desarrollé integraciones dentro del módulo de imports/bindings de la plataforma Caseware, una capa crítica de ingestión de datos usada por más de 475.000 profesionales en 130 países, garantizando mapeo de datos confiable y compatible con esquemas a escala empresarial.
- Desarrollé microservicios backend críticos en Java Spring Boot y Node.js, y desplegué infraestructura serverless en AWS con CDK, optimizando costos operativos. Llevé proyectos CDK desde desarrollo hasta producción.
- Lideré migraciones masivas de datos de más de 500.000 registros, asegurando atomicidad transaccional y confiabilidad. Mejoré la latencia de patrones de acceso a datos en 40%.
- Implementé componentes reactivos en Angular, reduciendo el tiempo de carga de la aplicación en 30% mediante lazy loading y optimización del bundle.
- Configuré despliegues orquestados en Kubernetes.

- Atendí incidentes de producción en turnos on-call con diagnóstico y resolución rápida para clientes.
- Lideré ceremonias Scrum para un equipo de 7 desarrolladores, asegurando entrega continua de valor para el equipo de imports.
- Trabajé con servicios de AWS como S3, DynamoDB, EC2, EKS y RDS.

Tecnologías: *Java, Spring Boot, AWS CDK, Angular, Kubernetes, Scrum, DynamoDB, S3, Jira, Agile, Copilot*

EDUCACIÓN

Ingeniería de Sistemas y Computación | *Universidad de los Andes* | Bogotá, Colombia | Enero 2021 – Junio 2025 | GPA: 4.21 / 5.0

Cursos relevantes: *Sistemas transaccionales, Arquitectura de software, Estructuras de datos y algoritmos, Inteligencia de negocios, Desarrollo de APIs, Programación orientada a objetos, Desarrollo web*

Ingeniería Industrial | *Universidad de los Andes* | Bogotá, Colombia | Enero 2022 – Junio 2025 | GPA: 4.21 / 5.0

Cursos relevantes: *Optimización avanzada, Metaheurísticas, Modelos probabilísticos, Simulación discreta, Probabilidad y estadística*

PROYECTOS DESTACADOS

Trove: Motor de búsqueda local open-source y servidor MCP [\[enlace\]](#) -- Motor de búsqueda local de alto rendimiento que indexa contenido de GitHub, Notion, Slack y archivos locales en una base vectorial unificada. Incluye una implementación de servidor MCP para optimizar flujos de trabajo de agentes de IA.

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

Aura AI: Plataforma inteligente de matching de talento con NLP, GNN y LLMs [\[enlace\]](#) -- Plataforma end-to-end de ML que transforma la adquisición de talento al reemplazar el matching por palabras clave con ajuste semántico. Incluye comprensión multi-modal de documentos, taxonomía de habilidades aprendida mediante Knowledge Graphs (Neo4j) y matching candidato-vacante con GNN.

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

Generador de CV [\[enlace\]](#) -- Librería modular en Python para generar CVs profesionales en formato DOCX y PDF.

Tecnologías: *Python, docx, reportlab*

Motor de predicción para apuestas deportivas - Pipeline ML con criterio de Kelly [\[enlace\]](#) -- Pipeline de ML que implementa variables temporales a partir de datos históricos de partidos, entrena y calibra modelos ensemble (Random Forest, XGBoost, LightGBM), y dimensiona apuestas mediante el criterio de Kelly.

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

Sistema WebSocket de alertas de mercado financiero [\[enlace\]](#) -- Sistema de alertas financieras en tiempo real basado en WebSocket. Implementa arquitectura orientada a eventos con Redis para mensajería y PostgreSQL para persistencia de datos.

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

Optimizador multiobjetivo de estaciones de carga para vehículos eléctricos -- Sistema de optimización de alto rendimiento desarrollado en Java y C++ para infraestructura de carga de vehículos eléctricos. Utiliza algoritmos multiobjetivo para minimizar costos de energía y maximizar niveles de servicio al cliente.

Tecnologías: *Java, C++, Algoritmos de optimización, Modelado de datos*

Plataforma de gestión empresarial (Hogar de Abuelos) [\[enlace\]](#) -- Plataforma integral de gestión para administración de instalaciones, orientada a optimizar operaciones, manejo de registros y asignación de recursos.

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

Orquestación de microservicios con Spring Boot y Docker -- Diseñé un sistema distribuido escalable con Java Spring Cloud (Eureka, Config, Gateway) dentro de un monorepo Nx. Implementé mensajería asíncrona con RabbitMQ y patrones de resiliencia con Resilience4j.

Tecnologías: *Spring Boot, Docker, AWS*

INVESTIGACIÓN

Modelado predictivo biológico (tesis de pregrado) [\[enlace\]](#) - 2025 - I

Desarrollé modelos predictivos para datos biológicos complejos usando técnicas de machine learning, aplicando preprocesamiento avanzado (PCA), balanceo de clases y deep learning (TensorFlow/Keras) para alcanzar 98% de precisión.

Analytics Forum 2025 - 2025 - I

Presenté los resultados de mi tesis de pregrado en el Analytics Forum 2025, exponiendo aplicaciones avanzadas de machine learning en bioinformática ante una audiencia de expertos de industria e investigadores.

Presentación en el grupo de investigación COPA - 2025 - I

Fui ponente en el Centro de Optimización y Probabilidad Aplicada (COPA), donde discutí la integración de redes neuronales y modelado probabilístico en sistemas biológicos.

HABILIDADES TÉCNICAS

Ciencia de datos: Python, pandas / NumPy, scikit-learn, Modelado estadístico, Ingeniería de características, Visualización de datos, Diseño experimental

Machine Learning e IA: Desarrollo con LLMs, Sistemas RAG, LangChain, Prompt Engineering, PyTorch, TensorFlow / Keras, XGBoost / LightGBM, Graph Neural Networks

Ingeniería de datos: Pipelines de datos, ETL / ELT, SQL, Modelado de datos, Apache Airflow, RabbitMQ, Flujos de datos orientados a eventos

Bases de datos: PostgreSQL, MySQL, MongoDB, DynamoDB, Redis, Pinecone, Neo4j, Qdrant

MLOps y Cloud: AWS, Docker, Kubernetes, GitHub Actions, APIs de serving de modelos, Monitoreo y observabilidad

Backend y APIs: FastAPI, Node.js / NestJS, Java / Spring Boot, APIs REST, GraphQL, Microservicios

Habilidades blandas y herramientas: Inglés, Comunicación técnica, Scrum, Design Thinking, Confluence y Jira

Idiomas: Inglés (C1 fluido), Español (nativo)

LOGROS Y RECONOCIMIENTOS

Finalista - Quala Business Challenge - 2025-1

Finalista en el Quala Business Challenge 2025-1, compitiendo frente a estudiantes destacados de negocios e ingeniería en una competencia basada en casos.