

# Gonzalo Romero

## AI Engineer | Software Engineer | Penetration Tester

Argentina (Remote) | [deeprat.tec@gmail.com](mailto:deeprat.tec@gmail.com) | [linkedin.com/in/gonzalo-romero-b9b5b4355](https://linkedin.com/in/gonzalo-romero-b9b5b4355) | [github.com/glromero](https://github.com/glromero) | [gonzalaromeroai.com](https://gonzalaromeroai.com)

### PROFESSIONAL SUMMARY

---

AI and software engineer with 15+ years of programming experience and 6 years of offensive security work. Builds LLM systems that have to be correct, not just impressive: multi-agent pipelines with typed contracts and verification gates, retrieval systems with hybrid ranking, and backend platforms with a strict security model. Sole author of Cortex, a telecom operations platform (BSS/OSS/CRM/ERP) that keeps p95 latency under 100 ms on 4 vCPUs. Also built Clinical Deep Research, an open LLM engine that traces every clinical claim back to its source passage. Holds IBM AI Engineering and Generative AI Engineering certifications. Python is the main language (9 years), alongside TypeScript and Go.

### TECHNICAL SKILLS

---

- **AI / LLM Engineering:** LLMs, RAG, multi-agent systems, LangGraph, LangChain, hybrid retrieval (BM25 + dense embeddings + reranking), Qdrant, prompt engineering, LLM evaluation, fine-tuning (LoRA/PEFT), Hugging Face, multi-provider LLM routing, Model Context Protocol (MCP), LLM security (prompt injection)
- **Machine Learning / Deep Learning:** PyTorch, TensorFlow, Keras, scikit-learn, transformers, NLP, computer vision (YOLOv8), MLOps
- **Backend & Data:** Python, FastAPI, Pydantic, PostgreSQL 16, REST APIs, OpenAPI, webhooks (Standard Webhooks), event-driven workers / outbox pattern, SQL
- **Frontend:** TypeScript, React, Reflex, Tailwind CSS
- **Cloud & DevOps:** Docker, Kubernetes, Helm, AWS, Google Cloud, Azure, infrastructure as code, GitHub Actions CI/CD, Nix, uv
- **Security:** penetration testing, web application and API security, OWASP ZAP, bug bounty (HackerOne, Bugcrowd), threat modeling, DPoP / key-bound sessions, Ed25519 signing, capability-based authorization, supply-chain security (CodeQL, gitleaks, OpenSSF Scorecard)
- **Code Quality:** pytest, 100% branch coverage, mypy (strict), Ruff, Architecture Decision Records (ADRs)
- **Programming Languages:** Python, TypeScript, Go, SQL, C++, C#, Ruby, Lua

### PROFESSIONAL EXPERIENCE

---

#### Founder & Principal Engineer | Blue Rings Labs | Remote, Argentina | 2026 – Present

*Independent engineering, research and education lab offering AI systems development, custom software, security audits and technical training.*

- Designed and wrote **Cortex 2.0** as sole author, a proprietary operations platform for telecom operators that combines BSS, OSS, CRM and ERP in one data model. It covers rating, prepaid balances, invoicing and fiscal filing, number portability, eSIM, general ledger and payroll, built on Python, PostgreSQL 16, React and TypeScript.
- Brought p95 latency **under 100 ms on all 67 console reads** across three concurrent tenants on a single 4-vCPU machine. Tenant-scoped queries cost **4 buffers in a 900,000-customer tenant**, the same as in a 5,000-customer one.
- Built real-time prepaid credit authorization with **0 double-grants and 0 errors** across 160 concurrent authorizations from 8 switches, and **0 violations in 65 kill/retry fault-injection tests**. The check and the deduction run as a single atomic statement.
- Raised month-end pipeline throughput to **10,649 records/s ingest and 2,965 records/s mediation** (128,000 records in 63 s). Rebuilt payroll to need **3,750× fewer database reads** (8 reads for 2,000 employees).
- Engineered the Cortex security model:
  - 198-capability authorization engine with risk tiers
  - DPoP key-bound sessions with hardware-key step-up
  - hash-chained audit trail signed with Ed25519 on a separate database
  - tenant isolation enforced in the persistence layer
  - egress allow-listing
- Published Cortex's dependency-free webhook SDKs in Python, TypeScript and Go, plus the OpenAPI spec. Delivered deployment kits for AWS, Google Cloud, Azure, Kubernetes (Helm) and bare metal.
- Created **Clinical Deep Research (CDR)**, an open-source (Apache 2.0) evidence engine that automates systematic-review work. It is a **13-node LangGraph pipeline** with Pydantic contracts between nodes, a FastAPI backend and support for 8 LLM providers.

- Implemented CDR's PICO extraction, PubMed and ClinicalTrials.gov retrieval with hybrid ranking (BM25 + dense + rerank), PRISMA-style LLM screening and RoB 2 / ROBINS-I risk-of-bias assessment. Added an adversarial "skeptical" agent and a verification gate that marks any report with untraceable claims as unpublishable.
- Built **awesome\_cybersecurity**, a CI-driven knowledge base of **490 resources in 18 categories** aligned to CyBOK and NICE. Its LLM classification pipeline rotates across Gemini, Gemma and Groq models, and the repository keeps 100% branch coverage, strict mypy typing and SHA-pinned CI.
- Authored **No Shortcuts**, a free AI engineering course of **173 lessons and 166 runnable labs** that CI executes on every change. It covers math, data engineering, ML, deep learning, LLMs, MLOps and AI security.

### Independent Penetration Tester & Security Researcher | Self-employed | Remote | 2020 – Present

- Started in public bug bounty programs on HackerOne and Bugcrowd and moved into private penetration-testing engagements, which became the main source of professional income.
- Tested web applications and APIs for clients. The findings now shape a "break it, then rebuild it" approach to secure backend architecture.
- Built an agentic LLM orchestration layer on top of OWASP ZAP that automates vulnerability discovery and triage, and used it to find real vulnerabilities in authorized engagements.

### AI Engineer, Independent Projects | Self-employed | Remote | 2021 – Present

- **MedX**: open-source medical AI assistant with multi-model routing across 8 LLMs, a RAG knowledge base of 1,000+ conditions and 500+ medications on Qdrant, emergency triage from 50+ patterns, and medical image analysis (Python, Reflex, CI).
- **TrueEye**: media-literacy system that scores news articles for narrative bias, target audience and manipulative intent (FastAPI, LangFlow, Nix, Tailwind CSS).
- **Building Intelligence**: open, bilingual (EN/ES) AI curriculum of 15 modules and 70+ lessons in Jupyter notebooks, designed for Google Colab's free tier.

## EDUCATION

---

**Systems Engineering (in progress)** | Universidad Tecnológica Nacional (UTN), Argentina

## CERTIFICATIONS

---

- **IBM AI Engineering Professional Certificate** (13 courses) | IBM / Coursera | Feb 2025
- **IBM Generative AI Engineering with LLMs Specialization** (7 courses) | IBM / Coursera | Feb 2025
- **CS50's Introduction to Artificial Intelligence with Python** (12 projects) | Harvard University | 2026
- **Anthropic Academy**: Claude Code in Action, Claude Code 101, Claude Platform 101, Claude 101, Model Context Protocol: Advanced Topics, AI Fluency: Framework & Foundations | Anthropic | 2026
- **Machine Learning with Python** (with honors) | IBM / Coursera | Jul 2024
- **Data Science Math Skills** | Duke University / Coursera | Nov 2024

## LANGUAGES

---

**Spanish**: Native | **English**: C1 Advanced, EF SET 66/100 (Mar 2026)

## ADDITIONAL

---

- Started programming at 15, writing Lua mods for *S.T.A.L.K.E.R.* and contributing to the community Spanish translation of *Call of Pripyat*. That work led to C++, then C#, Ruby and Python.
- Founder of **Puppizz Games**, an indie studio currently developing *Life Chronicles*.