

Gonzalo Romero

Engenheiro de IA | Engenheiro de Software | Pentester

Argentina (remoto) | deeprat.tec@gmail.com | linkedin.com/in/gonzalo-romero-b9b5b4355 | github.com/glromero | gonzalaromeroai.com

RESUMO PROFISSIONAL

Engenheiro de IA e de software com mais de 15 anos de experiência em programação e 6 anos em segurança ofensiva. Desenvolvo sistemas baseados em LLM que precisam funcionar bem em produção, não só em demonstrações: pipelines multiagente com contratos tipados e etapas de verificação, sistemas de recuperação com ranqueamento híbrido e plataformas back-end com um modelo de segurança rigoroso. Sou o único autor do Cortex, uma plataforma de operações para operadoras de telecom (BSS/OSS/CRM/ERP) que mantém latência p95 abaixo de 100 ms com 4 vCPUs. Também criei o Clinical Deep Research, um motor aberto baseado em LLM que vincula cada afirmação clínica ao trecho exato da fonte. Certificado pela IBM em AI Engineering e em Generative AI Engineering. Python é minha linguagem principal (9 anos), junto com TypeScript e Go.

COMPETÊNCIAS TÉCNICAS

- **IA / Engenharia de LLM:** LLMs, RAG, sistemas multiagente, LangGraph, LangChain, recuperação híbrida (BM25 + embeddings densos + reranking), Qdrant, engenharia de prompts, avaliação de LLMs, fine-tuning (LoRA/PEFT), Hugging Face, roteamento entre vários provedores de LLM, Model Context Protocol (MCP), segurança de LLMs (prompt injection)
- **Machine Learning / Deep Learning:** PyTorch, TensorFlow, Keras, scikit-learn, transformers, NLP, visão computacional (YOLOv8), MLOps
- **Back-end e dados:** Python, FastAPI, Pydantic, PostgreSQL 16, APIs REST, OpenAPI, webhooks (Standard Webhooks), workers orientados a eventos / padrão outbox, SQL
- **Front-end:** TypeScript, React, Reflex, Tailwind CSS
- **Cloud e DevOps:** Docker, Kubernetes, Helm, AWS, Google Cloud, Azure, infraestrutura como código, CI/CD com GitHub Actions, Nix, uv
- **Segurança:** testes de invasão, segurança de aplicações web e APIs, OWASP ZAP, bug bounty (HackerOne, Bugcrowd), modelagem de ameaças, sessões vinculadas a chave (DPoP), assinaturas Ed25519, autorização baseada em capacidades, segurança da cadeia de suprimentos (CodeQL, gitleaks, OpenSSF Scorecard)
- **Qualidade de código:** pytest, 100% de cobertura de branches, mypy (modo estrito), Ruff, registros de decisões de arquitetura (ADRs)
- **Linguagens de programação:** Python, TypeScript, Go, SQL, C++, C#, Ruby, Lua

EXPERIÊNCIA PROFISSIONAL

Fundador e Engenheiro Principal | Blue Rings Labs | Remoto, Argentina | 2026 - atual

Laboratório independente de engenharia, pesquisa e educação. Oferece desenvolvimento de sistemas de IA, software sob medida, auditorias de segurança e treinamento técnico.

- Projetei e desenvolvi como único autor o **Cortex 2.0**, uma plataforma proprietária para operadoras de telecomunicações que reúne BSS, OSS, CRM e ERP em um único modelo de dados. Cobre tarifação, saldos pré-pagos, faturamento e emissão fiscal eletrônica, portabilidade numérica, eSIM, contabilidade geral e folha de pagamento, com Python, PostgreSQL 16, React e TypeScript.
- Levei a latência p95 para **menos de 100 ms em todas as 67 leituras do console**, com três tenants simultâneos em uma única máquina de 4 vCPUs. Consultas por tenant custam **4 buffers em um tenant de 900.000 clientes**, o mesmo que em um de 5.000.
- Implementei autorização de crédito pré-pago em tempo real com **0 erros e nenhum crédito concedido em duplicidade** em 160 autorizações simultâneas vindas de 8 switches, e **0 violações em 65 testes de injeção de falhas** (transações interrompidas e novas tentativas). A verificação e o débito rodam em uma única instrução atômica.
- Atingi no processamento de fechamento de mês uma vazão de **10.649 registros/s na ingestão e 2.965 registros/s na mediação** (128.000 registros em 63 s). Reconstruí a folha de pagamento para fazer **3.750 vezes menos leituras no banco de dados** (8 leituras para 2.000 funcionários).

- Projetei o modelo de segurança do Cortex:
 - motor de autorização com 198 capacidades classificadas por nível de risco
 - sessões vinculadas a chave (DPoP), com elevação por chave de hardware
 - trilha de auditoria encadeada por hash e assinada com Ed25519, em um banco de dados separado
 - isolamento de tenants aplicado na camada de persistência
 - lista de permissões para o tráfego de saída
- Publiquei os SDKs de webhooks do Cortex em Python, TypeScript e Go, sem dependências externas, junto com a especificação OpenAPI. Entreguei kits de implantação para AWS, Google Cloud, Azure, Kubernetes (Helm) e bare metal.
- Criei o **Clinical Deep Research (CDR)**, um motor de evidências de código aberto (Apache 2.0) que automatiza o trabalho de revisões sistemáticas. É um **pipeline de 13 nós em LangGraph** com contratos Pydantic entre os nós, back-end em FastAPI e suporte a 8 provedores de LLM.
- Implementei no CDR a extração PICO, a busca no PubMed e no ClinicalTrials.gov com ranqueamento híbrido (BM25 + denso + rerank), a triagem com LLM no estilo PRISMA e a avaliação de risco de viés com RoB 2 / ROBINS-I. Adicionei um agente "cético" adversarial e uma etapa de verificação que marca como não publicável qualquer relatório com afirmações sem rastreabilidade.
- Desenvolvi o **awesome_cybersecurity**, uma base de conhecimento gerida por CI com **490 recursos em 18 categorias** alinhadas ao CyBOK e ao NICE. O pipeline de classificação com LLM alterna entre modelos Gemini, Gemma e Groq, e o repositório mantém 100% de cobertura de branches, tipagem estrita com mypy e CI fixado por SHA.
- Escrevi o **No Shortcuts**, um curso gratuito de engenharia de IA com **173 aulas e 166 laboratórios executáveis** que o CI roda a cada alteração. Cobre matemática, engenharia de dados, ML, deep learning, LLMs, MLOps e segurança em IA.

Pentester e Pesquisador de Segurança Independente | Autônomo | Remoto | 2020 – atual

- Comecei em programas públicos de bug bounty no HackerOne e no Bugcrowd e passei para projetos privados de testes de invasão, que se tornaram minha principal fonte de renda profissional.
- Avaliei a segurança de aplicações web e APIs para clientes. Essas descobertas hoje orientam uma abordagem de "quebrar para reconstruir" no projeto de arquiteturas back-end seguras.
- Desenvolvi uma camada de orquestração agêntica baseada em LLM sobre o OWASP ZAP que automatiza a descoberta e a triagem de vulnerabilidades, e a usei para encontrar vulnerabilidades reais em projetos autorizados.

Engenheiro de IA, Projetos Independentes | Autônomo | Remoto | 2021 – atual

- **MedX**: assistente médico de IA de código aberto com roteamento entre 8 modelos, base de conhecimento RAG com mais de 1.000 doenças e 500 medicamentos no Qdrant, triagem de emergências a partir de mais de 50 padrões e análise de imagens médicas (Python, Reflex, CI).
- **TrueEye**: sistema de letramento midiático que analisa notícias para identificar viés narrativo, público-alvo e intenção manipuladora (FastAPI, LangFlow, Nix, Tailwind CSS).
- **Building Intelligence**: programa aberto e bilíngue (EN/ES) de formação em IA, com 15 módulos e mais de 70 aulas em notebooks Jupyter, pensado para o plano gratuito do Google Colab.

FORMAÇÃO ACADÊMICA

Engenharia de Sistemas de Informação (em andamento) | Universidad Tecnológica Nacional (UTN), Argentina

CERTIFICAÇÕES

- **IBM AI Engineering Professional Certificate** (13 cursos) | IBM / Coursera | fev. 2025
- **IBM Generative AI Engineering with LLMs Specialization** (7 cursos) | IBM / Coursera | fev. 2025
- **CS50's Introduction to Artificial Intelligence with Python** (12 projetos) | 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** (com distinção) | IBM / Coursera | jul. 2024
- **Data Science Math Skills** | Duke University / Coursera | nov. 2024

IDIOMAS

Espanhol: nativo | **Inglês**: avançado, C1 (EF SET 66/100, mar. 2026)

INFORMAÇÕES ADICIONAIS

- Comecei a programar aos 15 anos criando mods em Lua para *S.T.A.L.K.E.R.* e colaborando na tradução comunitária para o espanhol de *Call of Pripyat*. Isso me levou ao C++ e depois a C#, Ruby e Python.
- Fundador da **Puppizz Games**, um estúdio indie que hoje desenvolve *Life Chronicles*.