DANIEL LANDI

Engenheiro de IA e Infraestrutura | LLMs Locais · Cloud · Programação de Sistemas

São Paulo, Brasil | drlandi@gmail.com | +55 11 9 5927-0064 (WhatsApp) | drlandi.dev | github.com/drlandi | linkedin.com/in/daniel-landi-ai

RESUMO PROFISSIONAL

Engenheiro de IA e infraestrutura com mais de 25 anos em operações de telecomunicações — NOC no Brasil e projetos de processos operacionais para operadoras em seis outros países — agora construindo aplicações de IA e ferramentas de LLM local. Trabalhos recentes: um runtime de inferência em C sobre o llama.cpp para hardware de 8 GB, apenas com CPU, com um harness de avaliação em Python que mede o que modelos pequenos conseguem construir; um gerador de atividades com LLM que se autocorrige, construído em um hackathon de 48 horas; e uma arquitetura serverless na AWS com failover. AWS Certified Cloud Practitioner; concluindo o currículo de programação de sistemas da 42 São Paulo. Trago a disciplina operacional de nível carrier — SLAs, projeto pensado para falhas, documentação que outra pessoa consegue usar às 3 da manhã — para sistemas de IA que precisam continuar funcionando.

COMPETÊNCIAS

- **Engenharia de IA e LLMs:** inferência local com llama.cpp (API em C, GGUF, amostragem e decodificação); avaliação de modelos com juízes escritos por humanos (taxa de acerto, estratégias de retentativa); aplicações com LLM (pipelines de prompts, ciclos de autocrítica); Llama 3 via Groq; AWS Bedrock e Textract; CLI do IBM Bob 2.0
- **Programação e ferramentas:** C (Norma da 42), Python, Node.js/Express, Bash, Git/GitHub, Makefiles, Docker, gerenciamento manual de memória
- **Cloud e infraestrutura:** AWS: Lambda, API Gateway, DynamoDB, S3, CloudFront, WAF, SQS, SNS, CloudWatch, IAM; Terraform (auditoria de ambiente); Linux/Unix (Pop!_OS, Solaris, HP-UX)
- **Operações e processos:** ITIL, eTOM, gestão de SLA e uptime, operação de NOC, procedimentos operacionais e documentação técnica

PROJETOS DE IA E SOFTWARE

uLLAMA — Runtime de Inferência Local de LLMs em C e Harness de Avaliação | Projeto pessoal, Set 2026 – Atual

- Escrevi um runtime em C sobre a API do llama.cpp que carrega modelos GGUF e executa chat multi-turno com saída em streaming, amostragem configurável (top-k, top-p, temperatura) e medição de tokens por segundo; projetado para 8 GB de RAM só com CPU, escrito sob a Norma da 42.
- Construí um harness de avaliação em Python: um modelo pequeno recebe uma tarefa, um juiz escrito por humanos testa o código, e o harness repete e conta quantas vezes passa. Medi que detalhar a especificação no prompt elevou a taxa de acerto por tentativa de cerca de 3% para cerca de 32% em uma tarefa. Revisar o código aprovado revelou um vazamento de arquivo que o juiz não detectava, e o juiz foi reforçado: passar no juiz é só um limite inferior.
- Medi uma busca binária em arquivo de 100 milhões de registros (2,4 GB) gerada pelo harness: cerca de 60 µs a quente; uma versão em C com mmap chegou a cerca de 1,5 µs com o arquivo já mapeado, e o repositório documenta o que foi e o que não foi estabelecido. Código: github.com/drlandi/ullama

Gerador de Atividades Pedagógicas com Autocorreção por IA | Hackathon IBM Bob 2.0, Set 2026

- Construí e publiquei, durante um hackathon de 48 horas, uma aplicação web em Node.js/Express que gera atividades práticas e editáveis pelo professor, adaptadas aos recursos da sala de aula, com uma segunda passada de autocrítica que revisa e regenera a saída conforme critérios de qualidade.
- Integrei a CLI do Bob por processos filhos, resolvi travamentos no tratamento de streams, adicionei exportação em PDF e fluxogramas e containerizei com Docker para implantação no Railway. Validado com atividades que já havia criado no CIEBP. Código: github.com/drlandi/ibm-bob-activity-generator

Arquitetura Cloud de Resiliência Digital | Hackathon EDN "Future Builders" (Escola da Nuvem × AWS), Set 2026

- Projetei, construí e testei ao vivo uma arquitetura serverless resiliente a falhas: WAF e CloudFront com failover de dupla origem (S3 em sa-east-1 e us-east-1), API Gateway com integração direta com SQS, worker Lambda, DynamoDB, CloudWatch/SNS.

- Validei de ponta a ponta com envios reais e auditei o ambiente de um colega, provisionado com Terraform, contra os requisitos.

EXPERIÊNCIA PROFISSIONAL

Professor Especialista | CIEBP — Unidade Especializada, escola pública estadual de SP — Fev 2023 – Atual

- Crio e ministro séries de atividades que levam pensamento computacional (programação desplugada) e letramento em IA a salas de aula com pouco acesso a tecnologia; construo ferramentas com IA para gerar e checar a qualidade de atividades de sala de aula.

Consultor de Engenharia Elétrica | TAUPO – Consultoria Ambiental — Nov 2015 – Dez 2021

- Entreguei projetos elétricos prediais de baixa tensão e projetos de energia solar em áreas rurais do noroeste de Minas Gerais.

Coordenador de Operações de Manutenção | SEICOM (rede Vivo, Espírito Santo) — Nov 2014 – Mar 2015

- Coordenei a manutenção de RF e infraestrutura dos sites celulares da Vivo e gerenciei a equipe técnica e administrativa.

Educação Técnica e Coordenação de Cursos | UFES, Multivix, Master Grupo Educacional — 2010 – 2013

- Docente em termodinâmica, resistência dos materiais e mecânica dos sólidos (UFES, Multivix); Coordenador de Cursos Técnicos no Master Grupo Educacional (Eletrotécnica, Automação Industrial, Mecânica, Petróleo e Gás).

Especialista em Telecomunicações (Telecom Specialist) | Ericsson Telecomunicações Ltda. — Nov 2005 – Jul 2010

- Especialista internacional em projetos GSM e 3G (WCDMA) para operadoras em seis países fora do Brasil, projetando e implantando processos operacionais sob eTOM e ITIL.
- Zain (Nigéria): gerenciei o subprojeto de Operações de Campo e novos procedimentos operacionais sob eTOM. Hutchison 3G (Inglaterra, Austrália) e Etisalat (Egito): avaliação contratual e desenvolvimento de processos operacionais.

Carreira anterior em operações de rede — 1991 – 2005

- Engenheiro de NOC na Nextel (1998 – 2005) e na Telesp Celular (1993 – 1998): manutenção de centrais celulares (Nortel, NEC, Ericsson), sistemas de sinalização, mensagens e pré-pago, análise de desempenho, drive tests, procedimentos operacionais; administração de Unix (Solaris, HP-UX) e projeto de uma LAN em Windows NT 4.0.
- Técnico em telecomunicações na Bastec e Telepatch (1991 – 1993): PABX e links de micro-ondas de uma rede bancária.

FORMAÇÃO

- **42 São Paulo — Software Engineering Cadet** (2025 – atual): currículo intensivo, baseado em projetos e peer-to-peer, em C, algoritmos e programação de sistemas Unix. Projetos: libft, get_next_line, push_swap (em dupla), ft_printf.
- **Licenciatura em Matemática** (Curso de Formação Pedagógica em Matemática) — Universidade Pitágoras Unopar; concluído em dez/2020, grau conferido em mar/2021
- **Graduação em Engenharia Elétrica** — Faculdade de Engenharia de São Paulo (FESP), 2002
- **Técnico em Eletrônica** — Escola Técnica de Eletrônica Francisco Moreira da Costa, Santa Rita do Sapucaí, MG, 1990

CERTIFICAÇÕES E CURSOS

- **AWS Certified Cloud Practitioner** — Amazon Web Services (válida até jun/2028) — credly.com/badges/564e8c2f-bc11-46a9-b1fe-7b0aa87dc545
- AWS Certified Developer – Associate: em preparação
- AWS re/Start Graduate (mai/2025); AWS Educate: Introduction to Generative AI (mar/2025); Escola da Nuvem (EDN): AWS Fundamentos + IA
- Outros cursos: Google Cloud Skill Badges (Vertex AI, Gemini multimodal RAG, apps GenAI com Streamlit, 2024); ITIL (2007, 2008); eTOM Distilled (TM Forum); cursos Ericsson de WCDMA, GPRS e ATM (2006).

IDIOMAS

Português (nativo) · Inglês (fluente) · Espanhol (básico)