Rafael Silva de Souza

Estudante de Engenharia de Computação · Machine Learning, HPC e Finanças Quantitativas

Porto Alegre, RS, Brasil · rafael.souza@inf.ufrgs.br · linkedin.com/in/rafael-silva-de-souza · github.com/RafaelSdeS · rafaelsdes.github.io

RESUMO

Estudante de Engenharia de Computação na UFRGS, com pesquisa em redes neurais conscientes de hardware (GPU e FPGA). Aberto a vagas em engenharia de ML, pesquisa em IA, pesquisa quantitativa e HPC.

FORMAÇÃO

UFRGS, Universidade Federal do Rio Grande do Sul: Engenharia de Computação Previsão: 2030

UFSC, Universidade Federal de Santa Catarina: Engenharia Elétrica (estudos anteriores)

PESQUISA E EXPERIÊNCIA

Iniciação Científica, UFRGS (orientador: Prof. Mateus Grellert) Em andamento

- Estudo o efeito de restringir kernels de CNN a 2×2/3×3 (adequados ao Winograd) na acurácia, eficiência e robustez INT8 (QAT), com 152 execuções no PCAD: alternar 3×3/2×2 cortou 24% dos MACs sem perda mensurável de acurácia.
- Com Thomas Schmitz: avaliei seus aceleradores Winograd em FPGA em 16 redes e treinei a rede de 4 bits (68,51% de top-1 no ImageNet). Coautor: IEEE OJCAS (em prep.), ISCAS (em avaliação).

Projeto VORIS, LabMetro UFSC 2025

- Construí um scanner 3D por triangulação a laser (câmera, laser, espelho com encoder, Arduino): extração subpixel da linha, triangulação geométrica e um modelo de profundidade com gradient boosting.

E3 UFSC, Equipe Shell Eco-marathon 2025

- Engenharia do protótipo ultraeficiente, incluindo compósitos de fibra de carbono; marketing da equipe e organização da área; visita técnica à Nanopoxy (resinas epóxi para compósitos).

PROJETOS E COMPETIÇÕES

LUNAX Fire Calendar, NASA Space Apps Challenge 2026

- Harmonizei detecções de fogo MODIS e VIIRS (NASA FIRMS) de 2000 a 2026 em um calendário mensal de queimadas para o Brasil.
- Um dos dois principais desenvolvedores (67 commits): previsão da próxima temporada, piloto nacional e ferramentas de campo; comparei um modelo de ML com o baseline de razão.

Conecta+, Desafio IBM SkillsBuild & Fundação Telefônica Vivo 2026

- Liderei a equipe de 3 e fui responsável pelo motor de decisão que prioriza 5.570 municípios para expansão de conectividade: scores, VPL/payback, otimização de portfólio (PuLP/CBC, IBM CPLEX) e backtest 2021→2025.
- Integrei o backend em FastAPI, a conferência dos textos do IBM Granite (watsonx) contra os números do motor, a CI e o deploy em Docker.

Finanças Quantitativas, mercado de ações brasileiro (B3) Em andamento

- Pipeline point-in-time sem viés de sobrevivência: B3 desde 2000 com deslistadas, ~9.600 tickers dos EUA (SEC EDGAR, FRED); LightGBM, modelo de risco e otimizador com custos em backtest walk-forward (em andamento).

vLLM Parallelism Benchmark, Avaliação de Desempenho (UFRGS)

- Tensor vs pipeline parallelism para o Qwen2.5-7B em 1 a 4 GPUs (vLLM + Ray, PCAD): entre nós, o tensor parallelism gastou até 97% do tempo de GPU em comunicação; liderei a telemetria e a estatística (equipe de 3).

PAM UFRGS, contribuição open source

- Pull request para um app de matrícula da UFRGS (Vue) com assistente de IA, previsão de formatura, exportação do horário em .ics e suíte de testes com Vitest.

HABILIDADES

- **Programação e HPC:** Python, C, C++, CUDA, TypeScript, JavaScript · vLLM, Ray, profiling de GPU, FPGA (Vivado)
- **ML e Dados:** PyTorch, Quantização (QAT/INT8), Aprendizado por Reforço, LightGBM, Visão Computacional, NLP
- **Software:** FastAPI, Vue, React, Node.js, Express, Firebase, Docker, Git, PuLP
- **Idiomas:** Português (nativo) · Inglês (fluente) · Francês (proficiente) · Espanhol (proficiente)

ATIVIDADES

UFRGSMUN, Delegado de Israel no Escritório da ONU de Combate ao Terrorismo (UNOCT) 2026

- Comitê em inglês sobre terrorismo no Sahel e ciberterrorismo: negociei medidas antiterror e de cibersegurança; um dos quatro delegados escolhidos pela imprensa para uma entrevista especial.