Rafael Silva de Souza

Estudiante de Ingeniería de Computación · Machine Learning, HPC y Finanzas Cuantitativas

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

RESUMEN

Estudiante de Ingeniería de Computación en la UFRGS; investigo redes neuronales conscientes del hardware (GPU y FPGA). Abierto a puestos de ingeniería de ML, investigación en IA, investigación cuantitativa y HPC.

FORMACIÓN

UFRGS, Universidad Federal de Rio Grande do Sul: Ingeniería de Computación Previsto 2030

UFSC, Universidad Federal de Santa Catarina: Ingeniería Eléctrica (estudios previos)

INVESTIGACIÓN Y EXPERIENCIA

Investigador de grado, UFRGS (director: Prof. Mateus Grellert) En curso

- Estudio cómo limitar los kernels de CNN a 2×2/3×3, aptos para Winograd, afecta precisión, eficiencia y robustez INT8 (QAT) (152 ejecuciones en PCAD): alternar 3×3/2×2 reduce los MAC un 24 % sin pérdida de precisión medible.
- Con Thomas Schmitz: evalué sus aceleradores Winograd en FPGA en 16 redes y entrené su red de 4 bits (68,51 % top-1 en ImageNet). Coautor: IEEE OJCAS (en prep.), ISCAS (en revisión).

Projeto VORIS, LABMETRO UFSC (con el Prof. Diego Fabre) 2025

- Construí un escáner 3D por triangulación láser (cámara, láser, espejo con encoder, Arduino): extracción subpíxel de la línea, triangulación geométrica y un modelo de profundidad con gradient boosting.

E3 UFSC, equipo de la Shell Eco-marathon 2025

- Ingeniería del prototipo ultraeficiente, incluidos compuestos de fibra de carbono; llevé y organicé el marketing del equipo; visita técnica a Nanopoxy (resinas epoxi para compuestos).

PROYECTOS Y COMPETICIONES

LUNAX Fire Calendar, NASA Space Apps Challenge 2026

- Unifiqué las detecciones MODIS y VIIRS (NASA FIRMS) de 2000–2026 en un calendario mensual de quemas de Brasil.
- Uno de los dos desarrolladores principales (67 commits): perspectiva de la próxima temporada, piloto nacional y herramientas de campo; comparé un retador de ML con la línea base de proporciones.

Conecta+, desafío IBM SkillsBuild & Fundação Telefônica Vivo 2026

- Lideré el equipo de 3; responsable del motor de decisión que clasifica 5570 municipios para expandir la conectividad: puntajes de prioridad, VPN/payback, optimización de cartera (PuLP/CBC, IBM CPLEX) y un backtest 2021→2025.
- Integré backend FastAPI, CI, despliegue Docker y la verificación de IBM Granite (watsonx) contra los números del motor.

Finanzas cuantitativas, mercado de acciones brasileño (B3) En curso

- Pipeline point-in-time sin sesgo de supervivencia: B3 desde 2000 con empresas deslistadas, ~9600 tickers de EE. UU. (SEC EDGAR, FRED); LightGBM, modelo de riesgo y optimizador con costos en backtests walk-forward (en curso).

vLLM Parallelism Benchmark, Evaluación de Desempeño (UFRGS)

- Paralelismo de tensor vs. de pipeline para Qwen2.5-7B en 1–4 GPU (vLLM + Ray, PCAD): entre nodos, el paralelismo de tensor gastó hasta el 97 % del tiempo de GPU en comunicación; lideré telemetría de GPU y estadística (equipo de 3).

PAM UFRGS, contribución open source

- Pull request a una app de matrícula de la UFRGS (Vue) que añade un asistente con IA, un pronóstico de graduación, exportación del horario a .ics y pruebas con Vitest.

HABILIDADES

- **Programación y HPC:** Python, C, C++, CUDA, TypeScript, JavaScript · vLLM, Ray, profiling de GPU, FPGA (Vivado)
- **ML y datos:** PyTorch, cuantización (QAT/INT8), aprendizaje por refuerzo, LightGBM, visión por computadora, NLP
- **Software:** FastAPI, Vue, React, Node.js, Express, Firebase, Docker, Git, PuLP
- **Idiomas:** portugués (nativo) · inglés (fluido) · francés (avanzado) · español (avanzado)

ACTIVIDADES

UFRGSMUN, delegado de Israel en la Oficina de Lucha contra el Terrorismo de la ONU (UNOCT) 2026

- Comité en inglés sobre el terrorismo en el Sahel y el ciberterrorismo: negocié medidas antiterroristas y de ciberseguridad con otras delegaciones; uno de los cuatro delegados elegidos por la prensa para una entrevista especial ante todos los participantes.