Rafael Silva de Souza

Computer Engineering Student · Machine Learning, HPC & Quantitative Finance

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

SUMMARY

Computer Engineering student at UFRGS researching hardware-aware neural networks (GPU and FPGA). Open to ML engineering, AI research, quantitative research and HPC roles.

EDUCATION

UFRGS, Federal University of Rio Grande do Sul: B.Eng. Computer Engineering Expected 2030

UFSC, Federal University of Santa Catarina: Electrical Engineering (prior studies)

RESEARCH & EXPERIENCE

Undergraduate Researcher, UFRGS (advisor: Prof. Mateus Grellert) Ongoing

- Study how restricting CNN kernels to Winograd-friendly 2×2/3×3 sizes affects accuracy, efficiency and INT8 (QAT) robustness (152 runs on PCAD): alternating 3×3/2×2 kernels cut MACs 24% with no measurable accuracy loss.
- With Thomas Schmitz: evaluated his Winograd FPGA accelerators on 16 networks and trained their 4-bit network (68.51% ImageNet top-1). Co-author: IEEE OJCAS (in prep.), ISCAS (under review).

Projeto VORIS, LabMetro UFSC 2025

- Built a laser-triangulation 3D scanner (camera, laser, encoder-driven mirror, Arduino): subpixel line extraction, geometric triangulation and a gradient-boosted depth model.

E3 UFSC, Shell Eco-marathon Team 2025

- Engineering work on the ultra-efficient prototype, including carbon-fiber composites; ran and organized the team's marketing; technical visit to Nanopoxy (epoxy resins for composites).

PROJECTS & COMPETITIONS

LUNAX Fire Calendar, NASA Space Apps Challenge 2026

- Harmonized MODIS and VIIRS (NASA FIRMS) fire detections from 2000 to 2026 into a monthly burning calendar for Brazil.
- One of two main developers (67 commits): next-season outlook, national pilot and field tools; benchmarked an ML challenger against the ratio baseline.

Conecta+, IBM SkillsBuild & Fundação Telefônica Vivo Challenge 2026

- Led the team of 3; owned the decision engine ranking 5,570 municipalities for connectivity expansion: priority scores, NPV/payback, portfolio optimization (PuLP/CBC, IBM CPLEX) and a 2021→2025 backtest.
- Integrated the FastAPI backend, checks of IBM Granite (watsonx) output against engine numbers, CI and the Docker deploy.

Quantitative Finance, Brazilian Equity Market (B3) Ongoing

- Point-in-time, survivorship-bias-free pipeline: B3 since 2000 incl. delisted companies, ~9,600 US tickers (SEC EDGAR, FRED); LightGBM, risk model and cost-aware optimizer in walk-forward backtests (in progress).

vLLM Parallelism Benchmark, UFRGS Performance Evaluation

- Tensor vs pipeline parallelism for Qwen2.5-7B on 1–4 GPUs (vLLM + Ray, PCAD): across nodes, tensor parallelism spent up to 97% of GPU time on communication; I led the GPU telemetry and statistics (team of 3).

PAM UFRGS, open-source contribution

- Pull request to a UFRGS enrollment app (Vue) adding an AI assistant, a graduation forecast, .ics schedule export and a Vitest test suite.

SKILLS

- **Programming & HPC:** Python, C, C++, CUDA, TypeScript, JavaScript · vLLM, Ray, GPU profiling, FPGA (Vivado)
- **ML & Data:** PyTorch, Quantization (QAT/INT8), Reinforcement Learning, LightGBM, Computer Vision, NLP
- **Software:** FastAPI, Vue, React, Node.js, Express, Firebase, Docker, Git, PuLP
- **Languages:** Portuguese (native) · English (fluent) · French (proficient) · Spanish (proficient)

ACTIVITIES

UFRGSMUN, Delegate of Israel at the UN Office of Counter-Terrorism (UNOCT) 2026

- English-language committee on terrorism in the Sahel and cyberterrorism: negotiated counter-terrorism and cybersecurity measures with other delegations; one of four delegates chosen by the press for a special interview before all participants.