

João Rietra

Physician and AI Engineer · AI applied to healthcare, clinical validation and sensitive data

Recife, Brazil · +55 81 99243-3485

joaorietra@gmail.com · jhlr.github.io · github.com/jhlr · linkedin.com/in/joaorietra

Licensed physician, CRM-PE 32980 (Brazilian medical board) · MD, FPS/IMIP (2022)

SUMMARY

Physician with three years of clinical practice — paediatric emergency and primary care — who moved into AI engineering without giving up the clinical side. I now build and evaluate machine learning models and generative AI applications, with a Professional Master's in Software Engineering (CESAR School) whose dissertation is precisely about **model reliability in multi-domain medical imaging**: calibration, out-of-distribution generalisation and explanation faithfulness. I am the rare interface between the two worlds — I model the domain, talk to the clinician in their own language, know what a metric means at the bedside, and treat sensitive data with the care it requires.

What separates me from an ML engineer with a good CV: I know why sensitivity matters more than accuracy in a triage setting, why a poorly calibrated model is dangerous even at high AUC, and why a pretty saliency map is evidence of nothing. That is not learned vocabulary — it is the yardstick I was trained to decide by.

EXPERIENCE

Physician — Paediatric Emergency and Primary Care · 2022 – 2025

- Diagnostic decisions under high demand in a paediatric emergency department; longitudinal care and care coordination in primary care.
- Practice in triage, risk stratification and communication with families — the same structure of decision-making under uncertainty that I now model in software.
- Critical appraisal of literature and clinical consensus as the basis for management, now reused when validating decision-support systems.

AI Engineer — LetzGrow (Grow Group) · 2026 – present

AI-assisted people development platform.

- **LLM-backed conversational agent** that drives the journey, interprets open-ended answers and generates personalised recommendations; development-plan generation decomposed into adaptive micro-tasks.
- **RAG** to ground recommendations in the user's profile and in the domain, reducing hallucination — the same *groundedness* problem any clinical LLM application faces.
- **Sensitive data by design**: conditional access, audit trail and compliance with Brazil's LGPD (the local equivalent of the GDPR); fairness evaluation of the recommendation pipeline, reported in a versioned document.
- Backend on NestJS + Prisma over PostgreSQL, LLM via Groq (Llama 3.3 70B), local embeddings with Ollama, AWS infrastructure via CDK.

Software Engineer (SMO + QA) — GasoGraph (Master's project, CESAR School) · 2025 – 2026

Clinical decision-support system over a rule tree derived from published consensus.

- **Translating published consensus into an auditable decision pipeline** — the part that usually stalls this kind of project, and that I can resolve because I sit on both sides of it.
- Strict input validation, inconsistency handling and test coverage on the critical rules.
- Domain modelling and rule validation from my own background, drastically shortening the iteration cycle with clinical experts.

ML Engineer — Veritas (School Innovation, CESAR School) · 2025 – 2026

Manipulated-media detection with deep learning.

- Computer vision pipeline with Grad-CAM, systematic false-positive analysis and out-of-distribution generalisation testing — the same protocol I apply to medical imaging.

- Composition of model evidence (forensic score + heatmap) with a natural-language explanation from a generative model.

Researcher — Computational Neuroscience, University of Alberta · 2015 – 2016

- Science Without Borders exchange (Brazilian federal merit scholarship): modelling of neural flows and analysis of high-dimensional datasets, at the intersection of computing, neuroscience and genetics.
- Applied bioinformatics: indel identification in BAM files (indelMINER).

Teaching Assistant — CESAR School · 2025 – 2026

- Programming (C), Statistics (R), Calculus I and Competitive Programming.

RESEARCH — MODEL RELIABILITY IN MEDICAL IMAGING

Master's dissertation, CESAR School (2026). Two connected papers on **when a model is reliable enough to come anywhere near a clinical decision**.

1. Training reuse for binary triage in multi-domain medical imaging.

Controlled comparison of 4 backbone origins across 5 heterogeneous domains (MedMNIST), evaluating AUC, sensitivity/specificity, **calibration (ECE)** and **occlusion-based explanation faithfulness**. Forensic intermediate fine-tuning won consistently (AUC **0.842** vs 0.796 for ImageNet) with far better calibration (ECE **0.045** vs 0.219) — meaning that beyond being more often right, the confidence the model declared matched reality, which is what makes a threshold safe to use.

Negative result reported: Grad-CAM produced spatially concentrated, convincing maps with **zero faithfulness to the model's actual decision**. A beautiful explanation is not an explanation.

2. Manipulated-image detection with lightweight convolutional networks.

Empirical and adversarial evaluation of in-house detectors against external ones (ViT, Swin) and ensembles, across 3 distributions, focused on OOD. No single detector generalises across manipulation types; "lightweight in parameters" ≠ "lightweight in practice" — the cost metric that matters is measured latency. Proposed a *gated* fusion architecture routed by a domain signal.

Experimental infrastructure. MLflow on a SQLite backend: per-epoch metrics, confusion matrices as artefacts, inference logging with content-addressed deduplication (sha256) and statistical aggregation (mean, confidence interval, Wilcoxon) over multiple seeds. End-to-end reproducibility from training to evaluation.

PROJECTS

pedrita — dissertation code · github.com/jhlr/pedrita

Multi-domain transfer learning over MedMNIST, evaluation with AUC/ECE/occlusion, MLflow infrastructure and a FastAPI backend serving the model.

GlucoTrack · github.com/jhlr/GlucoTrack

Diabetes management app: blood glucose and calorie logging.

indelMINER · github.com/jhlr/indelMINER

Bioinformatics: indel identification in BAM files.

klone-persona-rag — RAG with a vector store (Sentence-Transformers + ChromaDB), similarity retrieval with recency prioritisation. A *groundedness* foundation for LLM applications.

Minnow · github.com/jhlr/minnow

Coding agent that runs entirely inside a phone's browser — no account, no backend, with the workspace never leaving the device. Relevant to healthcare for its architectural argument: privacy obtained from where the data runs, not from a contractual promise.

EDUCATION

- **Professional Master's in Software Engineering** (industry-oriented track) — CESAR School, Recife · Oct 2025 – ongoing

- **Technologist Degree in Data Science** (3-year applied undergraduate degree) — CESAR School, Recife · Jan 2025 – ongoing
- **Doctor of Medicine (MD)** — FPS / IMIP, Recife · completed 2022 · six-year Brazilian medical degree, entered directly from secondary school
- **Science Without Borders exchange (Neuroscience)** — University of Alberta, Canada · 2015 – 2016 · Brazilian federal merit scholarship
- **Computer Science** — CIn / UFPE · Jan 2014 – discontinued

Recognition: first student ever selected with two simultaneous projects in CESAR School's School Innovation programme (2026.2) — a pre-incubation track with competitive selection and faculty supervision. Of the eight teams selected in the cohort, two are mine.

SKILLS

Healthcare and AI: clinical domain modelling · expert-in-the-loop validation · translating consensus into decision rules · medical imaging (MedMNIST) · clinically meaningful metrics (sensitivity, specificity, calibration) · sensitive data under LGPD (GDPR-equivalent) · critical appraisal of literature

Machine learning: PyTorch · scikit-learn · Hugging Face Transformers · OpenCV · transfer learning and fine-tuning · calibration (ECE) · OOD evaluation · explainability (Grad-CAM, occlusion) · inferential statistics (CI, Wilcoxon, multiple seeds)

Generative AI: LLMs (OpenAI, Anthropic, Groq, Ollama) · RAG and vector stores · agentic flows · groundedness and hallucination evaluation

Engineering: Python (FastAPI) · TypeScript (NestJS, React) · PostgreSQL · MLflow · Docker · AWS · Git · testing and CI · R, SQL

Languages: Portuguese (native) · English (native) · Spanish (basic) · French (basic)