# João Rietra

**Researcher** · AI model reliability: calibration, OOD generalisation and explanation faithfulness

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) · Lattes: *(to fill)* · ORCID: *(to fill)*

## SUMMARY

Researcher at the intersection of machine learning and healthcare, with a medical degree and a Professional Master's in Software Engineering (CESAR School). My research question is a single one, attacked from two sides: **when is a model reliable enough to support a decision?** Accuracy does not answer it — I investigate the calibration of declared confidence, generalisation outside the training distribution, and whether the explanation produced actually corresponds to the decision the model made. I work with controlled experimental protocol, multiple seeds, statistical testing and end-to-end reproducibility, and I **report negative results** when they appear — which is often, and which is the part usually missing from applied literature.

## EDUCATION

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

*The first professional Master's in Software Engineering accredited by CAPES (Brazil's federal graduate-programme evaluation body) in the country; highest CAPES rating among professional programmes in Computing.*

- **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 · Aug 2015 – Jul 2016 · Brazilian federal merit scholarship
- **Computer Science** — CIn / UFPE · Jan 2014 – discontinued

## WORK IN PROGRESS

**Master's dissertation — CESAR School (2026).** Two connected papers.

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

**Design.** Controlled comparison of 4 backbone origins (random initialisation, ImageNet, forensic intermediate fine-tuning and a variant) across **5 heterogeneous MedMNIST domains**, under an identical training and evaluation protocol.

**Metrics.** AUC, sensitivity and specificity, **calibration (Expected Calibration Error)** and **occlusion-based explanation faithfulness** — deliberately beyond what applied literature usually reports.

**Results.** Forensic intermediate fine-tuning consistently outperformed the other origins: **AUC 0.842 vs 0.796** for ImageNet, with far better calibration — **ECE 0.045 vs 0.219**. The calibration gap is larger in relative magnitude than the AUC gap, which supports the central claim: **the choice of pretraining affects the reliability of declared confidence more than it affects raw discriminative power.**

**Negative result.** Grad-CAM produced spatially concentrated, visually convincing maps with **zero faithfulness to the model's actual decision** under occlusion testing. Reported explicitly: the visual plausibility of a saliency map is not evidence that it explains anything.

### 2. Manipulated-image detection with lightweight convolutional networks

**Design.** Empirical and adversarial evaluation of in-house detectors (EfficientNet-B0) against external ones (ViT, Swin) and ensembles, across **3 distinct distributions**, with emphasis on out-of-distribution generalisation.

**Findings.** (i) No single detector generalises across manipulation types — in-distribution performance does not predict out-of-distribution performance. (ii) "Lightweight in parameters" does not imply "lightweight in practice": the relevant cost metric is

measured latency, not parameter count. (iii) Proposed a **domain-signal *gated* fusion architecture**, where routing between experts is conditioned on an estimate of the input domain.

### Experimental infrastructure and reproducibility

MLflow on a SQLite backend: per-epoch metrics, an evaluation namespace comparable across runs, confusion matrices as artefacts, **inference logging with content-addressed deduplication (sha256)** and a statistical aggregation layer (mean, confidence interval, Wilcoxon test) over **multiple seeds**. Every reported result is reproducible from training through evaluation. Public code at github.com/jhlr/pedrita.

---

## RESEARCH AND DEVELOPMENT EXPERIENCE

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

Manipulated-media detection. Computer vision pipeline with Grad-CAM, systematic false-positive analysis and OOD evaluation; composition of model evidence with a natural-language explanation from a generative model. Distributed large-scale processing.

### Technical lead (data and computer vision) — MAPEN · 2026 – present

Electricity-grid loss detection from the divergence between distributed and consumed energy. Built the data layer (hourly series from Brazil's national grid operator 2000→2026, consumption per state, IBGE geographic boundaries in DuckDB) and the meter-OCR subproject — TrOCR-small fine-tuned on UFPR-AMR, benchmarked against PP-OCRv6-tiny, with a CLI pipeline and MLflow tracking. **Provenance documented for every dataset** (source, author, licence, granularity), a requirement for regulatory use of the results.

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

AI-assisted people development platform: LLM-backed conversational agent, plan generation decomposed into micro-tasks, and RAG grounding for recommendations. Fairness evaluation of the recommendation pipeline, reported in a versioned document.

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

Decision-support system over a rule tree derived from published consensus; domain modelling and rule validation using my own clinical background.

### Physician — Paediatric Emergency and Primary Care · 2022 – 2025

Clinical practice that underpins the choice of clinically meaningful metrics and validation with domain experts.

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

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

### Teaching Assistant — CESAR School · 2025 – 2026

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

---

## DISTINCTIONS

**First student ever selected with two simultaneous projects in CESAR School's School Innovation programme** (2026.2 cohort): LetzGrow and MAPEN. A pre-incubation track with competitive selection, faculty supervision and artefact-based assessment — eight teams selected in the cohort, two of them mine. One of the projects (LetzGrow) was selected as a two-person team, a rare format in the programme. Third participation overall (the first, Veritas, in 2025).

---

## RESEARCH SOFTWARE

**pedrita** · github.com/jhlr/pedrita — dissertation code: multi-domain transfer learning, evaluation (AUC, ECE, occlusion), MLflow infrastructure, FastAPI backend serving the model.

**EfficientNeXt (mini-llm)** · github.com/jhlr/MiniLM — autoregressive LLM built from scratch: a causal *hourglass* architecture translating EfficientNet-B0 into language, with pretrained-skeleton transfer and a causality test as the acceptance criterion.

**btc-forecast** · github.com/jhlr/BTC-Forecast — comparative evaluation of zero-shot time-series foundation models (Chronos, TimesFM, Moirai, Lag-Llama, TTM) with fine-tuning and ensembling. Honest conclusion reported: short-horizon directional forecasting is indistinguishable from a random walk.

**Minnow** · github.com/jhlr/minnow — AI agent executing entirely on the client (mobile browser), with optional on-device inference via WebAssembly. Published measurement of per-turn latency (420 s → 176 s with cross-origin isolation) and of the limit found: models small enough to fit on the device report tool actions they never executed, which is why they were restricted to conversation.

**ragcode** · github.com/jhlr/ragcode — Model Context Protocol server exposing a local LLM as tools, with a semantic code index updated incrementally per commit.

---

## RESEARCH SKILLS

**Methodology:** controlled experimental design · multiple seeds and statistical aggregation · confidence intervals and non-parametric tests (Wilcoxon) · adversarial and OOD evaluation · reproducibility and experiment versioning · negative-result reporting

**Machine learning:** PyTorch · Hugging Face Transformers · scikit-learn · OpenCV · transfer learning and fine-tuning · calibration (ECE) · explainability (Grad-CAM, occlusion) · OCR · time series and foundation models · LLMs, RAG and agentic systems

**Infrastructure:** MLflow · Docker · AWS · DuckDB · PostgreSQL · Git · Linux · Python, R, SQL, TypeScript, C, Go

**Domain:** medical imaging · clinically meaningful metrics · expert-in-the-loop validation · sensitive data under LGPD (GDPR-equivalent) · Brazilian public datasets (IBGE, national grid operator, education census)

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