João Rietra

Pesquisador · confiabilidade de modelos de IA: calibração, generalização OOD e fidelidade de explicação

Recife, PE · (81) 99243-3485

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

CRM-PE 32980 · Lattes: *(preencher)* · ORCID: *(preencher)*

RESUMO

Pesquisador na interseção de aprendizado de máquina e saúde, com formação em Medicina e mestrado profissional em Engenharia de Software (CESAR School). Minha pergunta de pesquisa é uma só, atacada por dois lados: **quando um modelo é confiável o bastante para sustentar uma decisão?** Não basta acurácia — investigo calibração da confiança declarada, generalização fora da distribuição de treino e se a explicação produzida corresponde de fato à decisão tomada pelo modelo. Trabalho com protocolo experimental controlado, múltiplas seeds, testes estatísticos e reprodutibilidade ponta a ponta, e **reporto resultado negativo** quando ele aparece — o que aparece com frequência e é a parte que costuma faltar na literatura aplicada.

FORMAÇÃO ACADÊMICA

- **Mestrado Profissional em Engenharia de Software** — CESAR School, Recife · Out/2025 – em andamento

Primeiro mestrado profissional em Engenharia de Software reconhecido pela CAPES no país; maior nota CAPES entre os profissionais da área de Computação.

- **Tecnólogo em Ciência de Dados** — CESAR School, Recife · Jan/2025 – em andamento
- **Medicina** — FPS / IMIP, Recife · Concluído em 2022
- **Ciências sem Fronteiras (Neuroscience)** — University of Alberta, Canadá · Ago/2015 – Jul/2016
- **Ciência da Computação** — CIn / UFPE · Jan/2014 – interrompido

PRODUÇÃO EM ANDAMENTO

Dissertação de Mestrado — CESAR School (2026). Dois artigos conectados.

1. Reuso de treino para triagem binária em imagem médica multidomínio

Desenho. Comparação controlada de 4 origens de backbone (inicialização aleatória, ImageNet, fine-tuning intermediário forense e variante) sobre **5 domínios heterogêneos do MedMNIST**, sob protocolo idêntico de treino e avaliação.

Métricas. AUC, sensibilidade e especificidade, **calibração (Expected Calibration Error)** e **fidelidade de explicação por oclusão** — deliberadamente além do que a literatura aplicada costuma reportar.

Resultados. O fine-tuning intermediário forense superou consistentemente as demais origens: **AUC 0,842 vs 0,796** do ImageNet, com calibração muito superior — **ECE 0,045 vs 0,219**. A diferença de calibração é maior em magnitude relativa que a de AUC, o que sustenta a tese central: **a escolha de pré-treino afeta mais a confiabilidade da confiança declarada do que o poder discriminativo bruto.**

Resultado negativo. O Grad-CAM produziu mapas espacialmente concentrados e visualmente convincentes, com **fidelidade nula à decisão real do modelo** sob teste de oclusão. Reportado explicitamente: a plausibilidade visual de um mapa de saliência não é evidência de que ele explica algo.

2. Detecção de imagens manipuladas com redes convolucionais leves

Desenho. Avaliação empírica e adversarial de detectores próprios (EfficientNet-B0) contra externos (ViT, Swin) e ensembles, em **3 distribuições distintas**, com ênfase em generalização fora-de-distribuição.

Achados. (i) Nenhum detector único generaliza entre tipos de manipulação — o desempenho intra-distribuição não prediz o extra-distribuição. (ii) "Leve em parâmetros" não implica "leve na prática": a métrica de custo relevante é latência medida, não contagem de parâmetros. (iii) Proposta de arquitetura de **fusão *gated* por sinal de domínio**, em que o roteamento entre especialistas é condicionado a uma estimativa do domínio de entrada.

Infraestrutura experimental e reprodutibilidade

MLflow com backend SQLite: métricas por época, namespace de avaliação comparável entre execuções, matrizes de confusão como artefato, **logging de inferência com deduplicação content-addressed (sha256)** e camada de agregação estatística (média, intervalo de confiança, teste de Wilcoxon) sobre **múltiplas seeds**. Todo resultado reportado é reprodutível do treino à avaliação. Código público em github.com/jhlr/pedrita.

EXPERIÊNCIA EM PESQUISA E DESENVOLVIMENTO

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

Detecção de mídia manipulada. Pipeline de visão computacional com Grad-CAM, análise sistemática de falsos positivos e avaliação OOD; composição de evidência do modelo com explicação em linguagem natural por modelo generativo. Processamento distribuído em larga escala.

Lead técnico (dados e visão computacional) — MAPEN · 2026 – atual

Detecção de perdas em rede elétrica a partir da divergência entre energia distribuída e consumida. Construção da camada de dados (séries horárias do ONS 2000→2026, consumo por UF, malha do IBGE em DuckDB) e do subprojeto de OCR de medidores — fine-tuning de TrOCR-small sobre UFPR-AMR, comparado a PP-OCRv6-tiny, com pipeline em CLI e tracking em MLflow. **Procedência de cada dataset documentada** (fonte, autor, licença, granularidade), requisito para uso regulatório do resultado.

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

Plataforma de desenvolvimento de pessoas assistida por IA: agente conversacional sobre LLM, geração de plano com decomposição em micro-tarefas e RAG para ancoragem das recomendações. Avaliação de fairness do pipeline de recomendação, com relatório versionado.

Engenheiro de Software (SMO + QA) — GasoGraph (Mestrado / CESAR School) · 2025 – 2026

Sistema de apoio à decisão sobre árvore de regras derivada de consensos publicados; modelagem do domínio e validação das regras com background clínico próprio.

Médico — Urgência Pediátrica e Atenção Primária · 2022 – 2025

Prática clínica que fundamenta a escolha de métricas clinicamente relevantes e a validação com especialistas.

Pesquisador — Neurociência Computacional, University of Alberta · 2015 – 2016

Intercâmbio Ciências sem Fronteiras: modelagem de fluxos neurais e análise de datasets de alta dimensionalidade, no cruzamento entre Computação, Neurociência e Genética. Bioinformática: identificação de indels em arquivos BAM (indelMINER).

Monitor — CESAR School · 2025 – 2026

Programação (C), Estatística (R), Cálculo I e Programação Competitiva.

DISTINÇÕES

Primeiro aluno classificado com dois projetos simultâneos no School Innovation da CESAR School (edição 2026.2): LetzGrow e MAPEN. Trilha de pré-incubação com seleção competitiva, orientação docente e avaliação por artefatos — oito grupos classificados na edição, dois deles meus. Um dos projetos (LetzGrow) foi classificado em dupla, formato raro no programa. Terceira participação (a primeira, Veritas, em 2025).

SOFTWARE DE PESQUISA

pedrita · github.com/jhlr/pedrita — código da dissertação: transfer learning multidomínio, avaliação (AUC, ECE, oclusão), infraestrutura MLflow, backend FastAPI servindo o modelo.

EfficientNeXt (mini-llm) · github.com/jhlr/MiniLM — LLM autoregressivo construído do zero: arquitetura *hourglass* causal traduzindo a EfficientNet-B0 para linguagem, com transferência de esqueleto pré-treinado e teste de causalidade como critério de aceite.

btc-forecast · github.com/jhlr/BTC-Forecast — avaliação comparativa de foundation models de série temporal zero-shot (Chronos, TimesFM, Moirai, Lag-Llama, TTM) com fine-tuning e ensemble. Conclusão honesta reportada: previsão direcional de curto prazo é indistinguível de passeio aleatório.

Minnow · github.com/jhlr/minnow — agente de IA executando integralmente no cliente (navegador móvel), com inferência opcional on-device via WebAssembly. Medição publicada de latência por turn (420 s → 176 s com cross-origin isolation) e do limite encontrado: modelos que cabem no dispositivo relatam ações de ferramenta que não executaram, o que motivou restringi-los a conversação.

ragcode · github.com/jhlr/ragcode — servidor Model Context Protocol expondo LLM local como ferramentas, com índice semântico de código incremental por commit.

COMPETÊNCIAS DE PESQUISA

Metodologia: desenho experimental controlado · múltiplas seeds e agregação estatística · intervalos de confiança e testes não-paramétricos (Wilcoxon) · avaliação adversarial e OOD · reprodutibilidade e versionamento de experimentos · reporte de resultado negativo

Aprendizado de máquina: PyTorch · Hugging Face Transformers · scikit-learn · OpenCV · transfer learning e fine-tuning · calibração (ECE) · explicabilidade (Grad-CAM, oclusão) · OCR · séries temporais e foundation models · LLMs, RAG e sistemas agênticos

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

Domínio: imagem médica · métricas clinicamente relevantes · validação com especialistas · dados sensíveis e LGPD · dados públicos brasileiros (IBGE, ONS, INEP)

Idiomas: Português (nativo) · Inglês (nativo) · Espanhol (básico) · Francês (básico)