João Victor Assaoka Ribeiro

(31) 99700-0405 joao.assaoka@unifesp.br github/Assaoka linkedin/assaoka

EDUCAÇÃO

Doutorado Direto em Ciência da Computação *Mar 2024 - Dez 2030*
Universidade Federal de São Paulo (UNIFESP)

- **Aluno Regular de Doutorado Direto (Jan 2027 - Dez 2030):** Convidado a ingressar diretamente no doutorado em reconhecimento ao desempenho acadêmico e produção científica durante a graduação.
- **Aluno Especial da Pós-Graduação (Mar 2024 - Presente):** Cursou disciplinas de pós-graduação concomitantemente à graduação, obtendo conceito máximo (A) em todas: Análise de Algoritmos e Estruturas de Dados, Inteligência Artificial, Redes Neurais, Pesquisa Operacional e Seminários.

Visiting International Research Student (VIRS) *Maio 2026 - Ago 2026*
Dalhousie University Halifax, Canadá

- Pesquisador integrante do DNLP Group e do Malanis Lab, atuando em discussões e experimentos sobre o estado da arte em Processamento de Linguagem Natural (NLP) e Aprendizado de Máquina.

Bacharelado em Ciência da Computação (BCC) *Fev 2026 - Presente*
Universidade Federal de São Paulo (UNIFESP) GPA: **9,25/10**

- 3º Lugar no processo seletivo de progressão do bacharelado interdisciplinar.
- **Disciplinas Relevantes (BICT/BCC):** Cálculo, Álgebra Linear, Otimização Linear, Pesquisa Operacional, Resolução de Problemas via Modelagem Matemática, Tópicos em Economia e Negócios, Algoritmos e Estruturas de Dados, Programação Orientada a Objetos, Engenharia de Software, Bancos de Dados, Probabilidade e Estatística, Inferência e Análise de Regressão, Aprendizado de Máquina, Inteligência Artificial e Redes Neurais.

Bacharelado Interdisciplinar em Ciência e Tecnologia (BICT) *Mar 2023 - Fev 2026*
Universidade Federal de São Paulo (UNIFESP) GPA: **9,23/10**

- Certificado de Ênfase em Fundamentos de Ciência de Dados.
- **Monitoria Acadêmico:** Atuação como monitor das disciplinas de Lógica de Programação e Algoritmos e Estruturas de Dados.

EXPERIÊNCIA

Pesquisador em Processamento de Linguagem Natural *Nov 2024 - Presente*
Grupo de Pesquisa PARANA | FAPESP | Serasa Experian Remoto

- Elevou a correlação com julgamento humano em 113% (de 0,23 para 0,49) na avaliação sem referência de sumarização de notícias financeiras, propondo e implementando decomposição da tarefa via Question Answering com LLMs em Python e LangChain.
- Acelerou o ciclo de retreinamento multitarefa em 99,5% (de 1h17min para 25s) preservando 94,7% do Macro-F1, ao arquitetar uma estratégia de Model Merging de SLMs via Mergekit.
- Aumentou o Macro-F1 em 31,2% relativo na detecção multiclasse de toxicidade em cenários desbalanceados, estruturando pipelines sintéticos de data augmentation via LLMs.
- Construiu um dataset proprietário do domínio financeiro (+5.000 amostras), desenvolvendo a raspagem via Selenium, o desenvolvimento da plataforma web de anotação em Streamlit, a modelagem relacional em PostgreSQL (Neon/SQLAlchemy) e a análise exploratória com Pandas e Matplotlib.
- Arquitetou pipelines de fine-tuning (Full, LoRA, IA3, Soft Prompts via Hugging Face) e inferência (vLLM) com saídas estruturadas, orquestrando experimentos em nós multi-GPU NVIDIA em ambiente HPC Linux (Digital Research Alliance of Canada) via Slurm.

Supervisor / Gerente do Setor de Dados e IA *Dez 2024 - Jan 2026*
SciTec Jr. Remoto

- Fundou e estruturou o setor dedicado de Ciência de Dados e Inteligência Artificial da consultoria, liderando o spin-off da área de computação e instituindo novos critérios de recrutamento que viabilizaram a atração de talentos de múltiplos cursos de Engenharia.
- Gerenciou o ciclo de vida e a entrega técnica de 6 projetos, compreendendo 3 iniciativas de consultoria para clientes externos e 3 produtos internos de automação operacional com foco em escalabilidade de processos.
- Capacitou novos consultores em competências de Dados e IA, elaborando a trilha técnica e ministrando treinamentos em Python, análise exploratória (Pandas, Matplotlib), metodologia CRISP-DM, interfaces interativas com Streamlit e orquestração de LLMs com LangChain.

Engenheiro de IA Generativa *Fev 2025 - Jul 2025*
Ativamente - Enactus UNIFESP São José dos Campos Remoto

- Reduziu o tempo de geração de relatórios clínicos em mais de 90% (1 a 3 horas para 5 minutos) ao arquitetar uma solução integrada de dados com interface em Streamlit e pipeline de sumarização automatizada via LLM.
- Desenvolveu agente conversacional inteligente (Text-to-SQL / Chatbot) capaz de interpretar consultas em linguagem natural e extrair métricas analíticas sob demanda a partir de dados de oficinas cognitivas e prontuários de participantes idosos.
- Centralizou e padronizou o ecossistema de dados das oficinas, substituindo planilhas legadas fragmentadas por arquitetura relacional estruturada com formulários de entrada.
- Demonstração em Vídeo.

Pesquisador em Manutenção Prescritiva & XAI *Jul 2023 - Dez 2024*
Grupo de Pesquisa PARANA | FAPESP | Plataforma IAsmin | Siemens Energy Remoto

- Viabilizou economia de custos operacionais e prevenção proativa de falhas industriais ao prever sinais de degradação em séries temporais de sensores, desenvolvendo e comparando desde baselines lineares (Regressão Linear) até ensembles de gradiente (XGBoost) e arquiteturas profundas recorrentes (LSTM).
- mplementou framework multi-técnica de Inteligência Artificial Explicável (XAI) utilizando SHAP, gráficos ALE (Efeitos Locais Acumulados) e dependência parcial (PDP 1D/2D), diagnosticando interações causais complexas para validar o alinhamento das previsões ao conhecimento técnico e elevar a transparência na tomada de decisão.
- Co-criou a plataforma educacional IAsmin (Link de Acesso), democratizando metodologias práticas de XAI e manutenção preditiva da Indústria 4.0 em português através da publicação de 7 tutoriais práticos detalhando o funcionamento conceitual e a implementação de técnicas de interpretabilidade.

PUBLICAÇÕES

- **ASSAOKA, J.**; MILIOS, E.; KESELJ, V.; REQUENA, J. V. S. C.; MIRANDA, D. A.; BERTON, L. United Experts: Efficient Multi-Task Sentiment Analysis in Portuguese via Model Merging. Anais do XXIII Encontro Nacional de Inteligência Artificial e Computacional (ENIAC), SBC, p. 3035–3046, 2026.
- **ASSAOKA, J.**; CORREIA, T. P.; REQUENA, J. V. S. C.; BERTON, L. Evaluating Reference-Free Summarization Quality Metrics for Portuguese: A Study with Human Judgments in Financial News. Proceedings of the 17th International Conference on Computational Processing of Portuguese (PROPOR), ACL, p. 899–907, 2026.
- REQUENA, J. V. S. C.; **ASSAOKA, J.**; BERTON, L. Mitigating Imbalance in Portuguese Toxicity Classification via LLM Augmentation and Cross-Dataset Label Propagation. Anais do XXIII Encontro Nacional de Inteligência Artificial e Computacional (ENIAC), SBC, p. 3011–3022, 2026.
- REQUENA, J. V. S. C.; **ASSAOKA, J.**; BERTON, L. A Multimodal Framework for Financial Fake News Detection for Brazilian Portuguese. Proceedings of the 17th International Conference on Computational Processing of Portuguese (PROPOR), ACL, p. 234–239, 2026.
- MIRANDA, D. A.; **ASSAOKA, J.**; BERTON, L. Benchmarking Text Classifiers for Economic News in Brazilian Portuguese: In-Domain Quality and Out-of-Distribution Generalization. Anais do XXIII Encontro Nacional de Inteligência Artificial e Computacional (ENIAC), SBC, p. 981–991, 2026.
- **ASSAOKA, J.**; BERTON, L. Negotiating LLMs for Enhanced Hate Speech Classification and Interpretability. Anais do XXII Encontro Nacional de Inteligência Artificial e Computacional (ENIAC), SBC, p. 1716–1727, 2025.

PRÊMIOS E RECONHECIMENTOS

- **Prêmio Destaque na Graduação** – Universidade Federal de São Paulo (UNIFESP), 2025. Vencedor na categoria Produtos Técnicos ou Tecnológicos.
- **Prêmio Destaque na Graduação** – Universidade Federal de São Paulo (UNIFESP), 2026. Projeto selecionado na etapa local do campus na categoria Pesquisa Científica em Ciências Exatas para concorrer à premiação institucional global da UNIFESP.
- **Prêmio Voluntário Admirável** – Prefeitura Municipal de São José dos Campos (Programa Decolar), 2025. Concedido pela atuação no ensino de Lógica de Programação e Ciência de Dados a alunos de altas habilidades da rede pública.
- **Trainee Mais Engajado** – SciTec Jr. (ICT/UNIFESP), 2024.
- **Medalha de Bronze** – Olimpíada Brasileira de Astronomia e Astronáutica (OBA), 2022.