

Cristian Enrique Muñoz Villalobos

📍 Rio de Janeiro, Brazil ✉ cristrink@gmail.com ☎ 55 21 997262400 🔗 crismunoz.github.io
in crismunozv 🌐 crismunoz

Education

Pontifical Catholic University of Rio de Janeiro (PUC-Rio) Ph.D. in Electrical Engineering	August 2015 to August 2019
Pontifical Catholic University of Rio de Janeiro (PUC-Rio) M.Sc. in Electrical Engineering	March 2012 to March 2014
National University of Engineering (UNI) B.Sc. in Mechatronics Engineering	June 2005 to July 2010

Experience

Senior Research - Machine Learning Engineer <i>Holistic AI</i>	<i>London, UK</i> <i>May 2002 – present</i>
<ul style="list-style-type: none">◦ Developed benchmarks for evaluating Large Language Models (LLMs) across key technical risks, including efficacy, robustness, and privacy, deploying these solutions on Azure Platform and Google Cloud Platform (GCP).◦ Engineered a risk mitigation tool for AI systems, addressing issues related to model safety, security, and performance, using Jax/Optax/Flax.◦ Led the development of an AI trustworthiness toolkit, defining Responsible AI metrics and integrating advanced bias mitigation techniques for bias, fairness, and explainability.	
Data Science Advisor <i>Applied Computational Intelligence Laboratory - ICA</i>	<i>Rio de Janeiro, Brazil</i> <i>May 2022 – present</i>
<ul style="list-style-type: none">◦ Provided expertise in the design and construction of semantic search systems based on knowledge graphs, tables, and raw text, using Python.◦ Advised on the architecture of data provisioning systems, utilizing REST APIs and gRPC for anchoring data access.	
Senior Data Scientist <i>Applied Computational Intelligence Laboratory - ICA</i>	<i>Rio de Janeiro, Brazil</i> <i>March 2018 – April 2022</i>
<ul style="list-style-type: none">◦ Led a large-scale R&D project with a 20-person team over two years, delivering innovative AI solutions that enhanced system trustworthiness and generated measurable impact.◦ Developed transformer-based AI models, scaling training to millions of simulations across 100 GPUs, using Torch and Tensorflow.	
Researcher <i>Applied Computational Intelligence Laboratory - ICA</i>	<i>Rio de Janeiro, Brazil</i> <i>March 2012 – March 2018</i>
<ul style="list-style-type: none">◦ Participated in four R&D projects, developing neural networks for client non-compliance classification and optimization algorithms using C#, C++, and Bash on Linux systems.◦ Configured and optimized compute clusters using PBS Torque and SLURM for efficient resource management.	

Consulting Experience

Consultant - Natural Language Processing (NLP) <i>International Labour Organization (ILO)</i>	<i>Lima, Peru</i> <i>March 2021 – September 2021</i>
<ul style="list-style-type: none">◦ Implemented NLP solutions to extract and classify corporate documents, improving data analysis efficiency with Python.◦ Built an Information Extraction System using machine learning to automate the organization of news data.	

Consultant - AI Systems
Holistic AI

London, UK
May 2020 – December 2020

- Developed **synthetic data generation systems** with **deep learning**, enhancing model training using **TensorFlow**.

Teaching Experience

Assistant Professor - NLP and Data Science
Pontifical Catholic University of Peru (PUCP)

Lima, Peru
June 2022 – December 2022

- Taught **Natural Language Processing** using **Python**, focusing on real-world applications in public services and social sciences.


Assistant Professor - Data Science and AI
Pontifical Catholic University of Rio de Janeiro (PUC-Rio)

Rio de Janeiro, Brazil
March 2016 – April 2022

- Taught **Decision Support Systems** and **Deep Learning** with **Python**, focusing on practical applications in data science.

Projects

Holistic AI Open Source Library

[holisticai - github](#) 

- Spearheaded the development of an open-source toolkit focused on assessing bias, fairness, and explainability in AI systems.
- Led cross-functional collaboration to design and manage GitFlow pipelines and CI/CD workflows, ensuring seamless integration and automation.

Patents

Method for extracting and structuring information
Patent number BR 10 2021 023977-8

[BR 10 2021 023977-8](#) 

Publications

Navya Jain, Zekun Wu, **Cristian E. M. Villalobos**, Airlie Hilliard, Adriano Koshiyama, Emre Kazim, Philip Treleaven. From Text to Emoji: How PEFT-Driven Personality Manipulation Unleashes the Emoji Potential in LLMs, Sep 2024.

[NeurIPS 2024 Workshop on Behavioral Machine Learning](#) 

Mengfei Liang, Archish Arun, Zekun Wu, **Cristian E. M. Villalobos**, Jonathan Lutch, Emre Kazim, Adriano Koshiyama, Philip Treleaven. THaMES: An End-to-End Tool for Hallucination Mitigation and Evaluation in Large Language Models, Sep 2024.

[Workshop on Socially Responsible Language Modelling Research](#) 

Airlie Hilliard, **Cristian E. M. Villalobos**, Zekun Wu, and Adriano S. Koshiyama. Eliciting Big Five Personality Traits in Large Language Models: A Textual Analysis with Classifier-Driven Approach, Feb 2024.

[10.48550/arXiv.2402.08341](#) 

Cristian E. M. Villalobos, Kleyton da Costa, Bernardo Modenesi, and Adriano S. Koshiyama. Evaluating explainability for machine learning predictions using model-agnostic metrics, Jan 2024.

[10.48550/arXiv.2302.12094](#) 


Cristian E. M. Villalobos, Leonardo Forero, Harold De Mello, Cesar Valencia, and Alvaro Orjuela. Sentimental Analysis on Social Media Comments with Recurring Models and Pretrained Word Embeddings in Brazilian Portuguese, Jun 2023.


[10.1145/3582768](#) 


Cristian E. M. Villalobos, Sara Zannone, Umar Mohammed, and Adriano S. Koshiyama. Uncovering Bias in Face Generation Models, Feb 2023.

[10.48550/arXiv.2302.11562](#) 


Jose David Bermudez Castro, Ricardo Rei, Jose E. Ruiz, Pedro Achanccaray Diaz, Smith Arauco Canchumuni, **Cristian Muñoz Villalobos**, Felipe Borges Coelho, Leonardo Forero Mendoza, Marco Aurelio C. Pacheco. A free web service for fast COVID-19 classification of chest X-Ray images, Jul 2022.


[10.1007/978-3-031-10522-7_29](#) 

Potratz, Júlia, **Cristian E. M. Villalobos**, Smith WA Canchumuni and Marco Aurélio C. Pacheco. Deep learning for mapping rainwater drainage networks using Remote Sensing Data, Dec 2021.
[5774/4838](#) 

Jose D. Bermudez Castro, Smith WA Canchumuni, **Cristian E. M. Villalobos**, Fábio Corrêa Cordeiro, Antônio Marcelo Azevedo Alexandre, and Marco A. Cavalcanti Pacheco. Improvement Optical Character Recognition for Structured Documents using Generative Adversarial Networks, Dec 2021.
[10.1109/ICCSA54496.2021.00046](#) 


Fábio Corrêa Cordeiro, **Cristian E. M. Villalobos**. Petrolês - How to Build a Specialized Oil and Gas Corpus in Portuguese. Rio Oil and Gas Expo and Conference, Dec 2020.

Cristian E. M. Villalobos, Leonardo Mendoza and Ricardo Tanscheit. Construction of a Transformer-based model for Relation Extraction, Dec 2019
[CBIC2019/CBIC2019-117](#) 

Costa Leandro Santos DA, Blank Frances Fischberg Oliveira, Fernando Luiz Cyrino, **Cristian E. M. Villalobos**. Conditional Pricing Model with Heteroscedasticity: Evaluation of Brazilian Funds, Aug 2019.
[10.1590/S0034-759020190402](#) 

Arthur Silveira, **Cristian E. M. Villalobos**, Leonardo Mendoza. Severe Asthma Exacerbations Prediction Using Neural Networks, May 2019.
[10.1007/978-3-030-20257-6_10](#) 

Leonardo A. Forero Mendoza, **Cristian E. M. Villalobos**, Manoela Kohler, Evelyn Batista and Marco A. Pacheco. Analysis and Classification of Voice Pathologies using Glottal Signal Parameters with Recurrent Neural Networks and SVM, Jan 2019.
[10.5220/0007250700190028](#) 

Gabriel Lins Tenorio, **Cristian E. M. Villalobos**, Leonardo A. Forero Mendoza, Eduardo Costa da Silva and Wouter Caarls. Improving Transfer Learning Performance: an Application in the Classification of Remote Sensing Data, Jan 2019.
[10.5220/0007372201740183](#) 

Skills

Programming Languages: Python, C++, C# JavaScript

Machine Learning Frameworks: TensorFlow, PyTorch, Jax, Scikit-learn

Cloud Platforms: Google Cloud Platform (GCP), Azure

Containerization and Orchestration: Kubernetes, Docker, Singularity

CI/CD Tools: GitHub Actions, GitFlow

Specializations: AI Ethics, Explainability, Bias Mitigation, Large Language Models (LLMs)

Languages

Spanish (Native)

English (Fluent)

Portuguese (Fluent)