

# Héctor Javier Hortúa

✉ orjuelajavier@gmail.com

☎ +57-3102247678

🏠 Bogotá, Colombia

🌐 [JavierOrjuela](#)

in [hector-hortua-6292b41b3](#)

🇨🇴 **Citizenship:** Colombia

## Research interests

Applications of Machine(Deep) Learning in several scientific disciplines and real-world projects, along with the study of Trustworthy AI using Explainable artificial intelligence (XAI), uncertainty quantification, and generative models. Designing different end-to-end data science/ML projects starting from building an ML-focused strategy and progressing into model training, optimization, tracking experiments, and packaging code. Deployments via Google Cloud Platform and Git ecosystems. Building RESTful APIs and containerizing the application with Docker.

## Professional Experience

### Data Science Consultant, University el Bosque

Bogotá, Colombia

2024

**Description:** Solving scientific and practical questions with data-driven analysis, insight creation, and data science solutions. Responsible for aspects and implementation of medical solutions through data understanding, modeling, and deployment. Application of machine learning, natural language processing, and computer vision. The consultancy is provided to researchers enrolled in the *AI for cervical cancer screening* project and professors.

### Data Scientist at Quaesta AI

Cluj, Romania

2023-2024

**Description:** Doing research related to novel event cameras applied for optical flow, segmentation, pose & object detection, and anomaly detection. Providing inference descriptions for several methodologies and Deep Learning models based on client requirements.

**Products:** Construction software demo to highlight the abilities and performance that event cameras provide for standard cameras in autonomous driving vehicles and high-speed environment tasks. Writing scientific papers and attending conferences.

*Technologies: For the demo design*, technologies such as TensorFlow, ONNX, TFLite, and Torch-lighting were employed.

- Transformation and adaptation of spare dataset were made with spectral kernels using fully connected custom layers. Deep architectures were based on Convolutional and recurrent (LSTM) layers to build methodologies based on VAE, and temporal contexts like SegNet, PSPNet DeepLabV2, and YOLO.
- Methodologies based on Graph Convolutional Neural Networks deployed using PyTorch Geometric and TF-GNN.
- Experimentation was suitable thanks to MLFlow.
- Model optimization was applied at the end of the deployment model via pruning and Quantization.
- Bitbucket management for source control collaboration.

*Technologies: For Research and paper contributions*, PyTorch, Faiss and HuggingFace were used.

- Analyzing and compiling benchmark computer vision and anomaly detection methods that might be employed in event cameras to compare their high-level structures, strategies, and performance.
- AI Similarity Search (Faiss) library was adapted due to the ability to quickly search for billion-scale data sets similar to each other 8.5x faster than the previously reported state-of-the-art on the GPU.
- Manage Deep Learning experiments along with systematic tracking methods. Different methodologies were designed to compare with benchmarks.
- Uncertainty techniques developed in Bayesian-Torch were adapted to multiple models to determine reliability and interpretability.

### **Data Scientist at Dispell Magic Inc., Competitoor**

Dover, County of Kent, Delaware, USA

2021-2023

**Description:** Stays ahead of the Machine Learning needs by identifying opportunities for improvement modeling tools, processes, or infrastructure; and working on workflow from data access, processing and modeling, and modern software process to develop production.

**Products:** *Containerization* scripts for preprocessing data, training models, and deploying them via docker, to deploy future ML pipelines in cloud services.

*Leading data labeling team* to build, improve, and maintain models capable of recognizing products, increasing the detection performance by 92%.

*Improving Deep Learning model's performance* employed in Multimodal models that infer properties based on images and text.

*Building* robust unsupervised models to help our E-commerce services meet the customer's needs.

*Proposing* alternative technologies to store data objects and query them based on similarity in a faster way. Also, make experiments using tools such as Polar and Dask to manipulate big data tables, as an alternative to pandas.

*Recommending* the use of fine-tuning LLMs to retrieve information from the database faster and accurately to standard approaches. Also, I started to implement Ludwig in the DS team to deploy a low-code framework for building our custom LLMs.

*Technologies:* Docker, Flask, MLFlow, TensorFlow, Scikit-Learn, ElasticSearch, Kibana, MySQL, Weviate, Pandas, Polar, Dask, HuggingFace and Ludwig were used.

- Containerizing RESTful API's written in Flask, script for production, and pipelines written in MLFlow via docker for production and deployment into cloud platforms.
- First approach towards the use of Kubernetes for automating deployment, and management of containerized APIs.
- Designing Deep Learning models related to images, text, Multimodal models, object detection, and similarities with TensorFlow, Scikit-Learn, and PyCaret.
- Coding MLFlow pipelines for production.
- Designing and Computing model performance for generative models and autoML approaches for text applications.
- Leading team for labeling dataset used to improve Computer Vision approaches for the company.
- Exploring, analyzing, and preprocessing tabular MySQL data and non-structured information from elastic search server and Weviate.
- Proposing alternative technologies such as Weviate to store objects and query them based on similarity.
- Implementing Ludwig to deploy fine-tuning LLMs with our data, and combine it with HuggingFace to custom the model according to the company's requirement.

	PhD in Physics	04 2019
	<b>Universidad Nacional de Colombia</b>	Bogotá, Colombia
	MSc in Astrophysics	07, 2011
	<b>Universidad Nacional de Colombia</b>	Bogotá, Colombia
	BA in Physics	07, 2008
Honors, Grants & Scholarships	Helmholtz Information & Data Science Academy (HIDA) grant. Visiting Researcher at Deutsches Elektronen-Synchrotron (DESY), Germany	2023-2024
	Academic Research Grant, Google Cloud research.	2022-2024
	Unicore Stipendium der DAAD-Stiftung programme. Visiting Researcher at Max Planck Institute for Astrophysics, Germany.	2022
	Laureate Distinction: Doctoral Thesis. Universidad Nacional de Colombia.	2020
	Colciencias-Conv. 647, Doctorado Nacional Fellowship	2015
	Meritorious Distinction: Master Thesis. Universidad Nacional de Colombia.	2011
Internships	Romanian Institute of Science and Technology	2018-2019
	Implement supervised and unsupervised techniques in Machine and Deep Learning, and the use of TensorFlow and Sonnet for building an end-to-end pipeline for big astrophysical dataset.	
Research experience	<b>Long-term visiting researcher at Deutsches Elektronen-Synchrotron (DESY)</b>	
	Zeuthen, Germany	2023-2024
	Analysis of Bayesian Recurrent Neural Networks into Anomaly neural network detectors for observing transient phenomena in the next-generation ground-based instrument called Cherenkov Telescope Array Observatory (CTAO) for gamma-ray astronomy at very-high energies. Besides, novel recurrent Bayesian layers were built to measure model's reliability	
	<u>Technologies:</u> TensoFlow and TFProbabilities. Use of GitLab CI/CD to continuously build, test, deploy, and monitor iterative code changes deployment.	
	<b>Long-term visiting researcher at Max Planck Institute for Astrophysics</b>	
	Garching, Germany	2022
	Applications of Deep Learning generative models along with Graph Neural Networks to extract underlying spectral information (in the latent space) contained in Galactic all-sky data covering spectral bands from $\gamma$ -ray to radio waves.	

*Technologies:* TensorFlow and TFPProbabilities. Use of Graph Convolutional Neural Networks to deploy generative models based on Normalizing Flows. The infrastructure used was Q-Blocks cloud service to create optimized compute cluster. Kubernetes deployment was used also in the project.

### **Postdoctoral Networking in Artificial Intelligence**

DAAD, Germany

2021-2022

Initiative funded by the German Federal Ministry of Education and Research to scientific talents from all over the world a personalized gateway to the German AI research community.

### **Postdoctoral Research Scientist**

Romanian Institute of Science and Technology

2019-2020

Design and analysis of novel training algorithms for neural networks in deep learning, by applying notions of Riemannian optimization and differential geometry [RIST](#). Proposal methods for including uncertainties in Deep Learning for reliability and interpretability. Analysis of robustness via Bayesian Neural Nets and defense against adversarial attacks.

### **Argo project**

DeepRiemann - Riemannian Optimization Methods for Deep Learning project.  
POC 2014-2020

Argo is a library for deep learning algorithms based on TensorFlow and Sonnet. The library allows you to train different models (feed-forwards neural networks for regression and classification problems, autoencoders and variational autoencoders, Bayesian neural networks, Helmholtz machines, etc) by specifying their parameters as well as the network topologies in a configuration file. The models can then be trained in parallel in presence of multiple GPUs. The library is easy to expand for alternative models and training algorithms, as well as for different network topologies. [ARGO](#)

## Teaching experience

### **Teaching associate, Maestría en Estadística aplicada y Ciencia de Datos, Universidad el Bosque**

2022-2024

Deep Learning I-II, Machine Learning I-II

Relevant products: TF-Notebooks and GitHub-classroom repositories. Thesis supervisor and referee.

### **Teaching associate, Maestría en Ciencia de Datos, Escuela Colombiana de Ingeniería Julio Garavito**

2021-2022

Hello world!! con Tensorflow 2. Una Introducción a Deep Learning. Tópicos avanzados en Deep Learning.

Relevant products: TF-Notebooks and GitHub-classroom repositories. Thesis supervisor and referee.

**Teaching associate, Department of Basic Science and Technology, (UNAD)**

2021

Physics, Head and instructor in Data Science specialization program.

Relevant products: Constraining cosmological parameters from N-body simulations with BNNs, results accepted at NeurIPS, Bayesian Deep Learning, 2021

Relevant products: Property estimation method for Terpenes using Machine Learning, ESCAPE32, Toulouse, France, 2022

Relevant products: Un enfoque basado en Machine Learning para estimar el punto de fusión en compuestos químicos, 4ta semana de la Química, Colombia, 2021

**Teaching associate, Department of Basic Science, Universidad Los Libertadores**

2011-2014

Physics and Maths

Relevant products: First Astrostatistics School: Bayesian Methods in Cosmology, held in U. Los Libertadores Bogotá D.C., Colombia. [Proceedings](#)

Relevant products: IAU Symposium 306: Statistical Challenges in 21st Cosmology [Cambridge University Press](#)

**Projects**

**Google Cloud Research Credits Program: Robust Simulation-Based Inference in Cosmology with AI.**

Design new ML methods that allow Graph neural networks and Deep models to quantify the uncertainty of its predictions in the cosmological context. Bayesian inference via Deep Bayesian Nets.

**Machine Learning-based approaches for estimating melting point in chemical compounds**

Design new ML methods based on ensembles, graphs and BNNs, to determine properties estimation of chemical components.

*32nd European Symposium on Computer Aided Process Engineering, ESCAPE 2022, Elsevier, ISBN:9780323958790*

**Decomposition of the Galactic multi-frequency sky via Deep Generative Models**

Build Variational Autoencoders (VAEs) based graph network for spherical dataset for detecting the essential spectral information contained Galactic all-sky data covering all spectral bands.

**Reliable Uncertainties for Bayesian Neural Networks using Alpha divergences**

Design new metrics and methods to determine the accuracy in prediction's neural models. Generalized divergences were implemented for different dataset.

*Results presented and published at ICML, Uncertainty and Robustness in Deep Learning, 2020.*

### **Accelerating MCMC algorithms through Bayesian Deep Networks**

New approach to accelerate MCMC techniques by adding at the top of the model the Neural network. It allows to include reliable parameter space parameter and be able to speed up the convergence.

*Results presented and published at NeurIPS, Machine Learning and Physical Sciences, 2020.*

### **Constraining the reionization history using Bayesian normalizing flows**

Apply Normalized flow in computer vision tasks. Additionally, implement uncertainties in Object detection models to obtain robust vision models. It can be apply to style transfer, segmentation and resolution images.

*Results presented and published at Machine Learning Sci.Tech , 2020. and ICLR 2020 workshop Fundamental Artificial Intelligence in science*

*For publications click [here](#).*

## Conferences

**Debating the potential of Machine Learning in astronomical Surveys**, Paris, France, 12 2023

**IAIFI: NSF Institute for Artificial Intelligence and Fundamental Interactions**, virtual, 8 2023

**IAIFI: NSF Institute for Artificial Intelligence and Fundamental Interactions**, virtual, 8 2022

**ICTP: Latin American regional workshop on SciTinyML: Scientific Use of Machine Learning on Low-Power Devices**, virtual, 7 2022

**Conference on Neural Information Processing Systems, NeurIPS**, Contribution: "Constraining cosmological parameters from N-body simulations with Bayesian Neural Networks" virtual, 12 2021

**Conference on Neural Information Processing Systems, NeurIPS**, Contribution: "Constraining cosmological parameters from N-body simulations with Bayesian Neural Networks" virtual, 12 2021

**Conference on Neural Information Processing Systems, NeurIPS**, Contribution: "Accelerating MCMC algorithms through Bayesian Deep Networks" virtual, 12 2020

**ICTP-SAIFR Latin American Workshop on Observational Cosmology**, Contribution: "Constraining the Reionization History using Bayesian Normalizing Flows" virtual, 5 2020

**Joint Structures and Common Foundation of Statistical Physics, Information Geometry and Inference for Learning**, Contribution: "Calibrating BNNs with  $\alpha$ -divergences and Normalizing Flows", Les Houches, France, 07 2020

**Thirty-seventh International Conference on Machine Learning, ICML**,  
Contribution: "Reliable Uncertainties for Bayesian Neural Networks using  $\alpha$ -divergences", Virtual, 08 2020

**Eighth International Conference on Learning Representations, ICLR**,  
Contribution: "Parameters Estimation from the 21 cm signal using Variational Inference", Virtual, 05 2020

**Bayesian Deep Learning for Cosmology and Gravitational waves**, Contribution: "Constraining Cosmological Parameters from CMB maps using Neural Networks", Paris, France, 03 2020

**AICosmo2019: Artificial Intelligence Methods in Cosmology**, Contribution: "Estimation of Cosmological Parameters via ConvNets", Ascona, Switzerland, 06 2019

#### Certifications

**Google Cloud Computing Foundations Certificate, Google** 09,2023

**TensorFlow Developer Certificate, Google** 01 2021

**Artificial Intelligence Technology Certificate, Huawei** 02 2021

#### Certificated courses

**Machine Learning Engineering for Production, Coursera** 10 2022

**Advanced Machine Learning on Google Cloud, Coursera** 02 2021

**TensorFlow: Advanced Techniques, Coursera** 11 2020

**Machine Learning with TF on GCP, Coursera** 02 2021

**TensorFlow: Data and Deployment, Coursera** 11 2020

**TensorFlow 2 for Deep Learning, Coursera** 10 2020

**Advanced Data Science with IBM, Coursera** 06 2020

**Learn SQL Basics for Data Science, Coursera** 08 2020

#### Scientific courses

**Transylvanian Machine Learning Summer School (TMLSS)** 02 2018

**Microsoft Research Frontiers in Machine Learning** 07 2020

**Quantum 2020, The Institute of Physics and IOP Publishing** 10 2020

#### Skills

##### Programming

Proficient in: Mathematica, C++, Python-libraries, scikit-learn, CUDA, PyTorch, Tensorflow, Keras, TFProbability, TFX, Pandas ecosystem libraries, Dask, mySQL, GIT, bash, DBeaver, Docker, GCP, Elastic Search, MLflow, Kibana, FastAPI, HuggingFace, FlaskAPI, Orange Data-Mining

Familiar with: MPICH2, PyMC, Pystat, Apache Spark, MLlib, Sonnet, databricks, weaviate, PM2, Kubernetes, Polar

##### OS

Linux

##### Editor

LaTeX, Pycharm, Emacs, Visual Studio



**Languages**

Spanish (mother tongue), English (Professional working), German (beginner)