

Tomas Iriarte

📍 Data Science & AI Engineering at Leniolabs_

Key Expertise

- Data Science
- Analytical Skills
- Communication
- Programming

Summary

Motivated by the potential of data to uncover valuable insights and always eager to tackle new challenges and projects. I see data as a valuable resource that can enhance our understanding and drive innovation. My approach is continuously shaped by a commitment to learning and applying the latest techniques in the field.

Python

NumPy, Pandas, Scikit-learn, Tensorflow,

Amazon Web Services

Sagemaker, S3, EC2, Batch

Jupyter, OpenCV, Matplotlib,
Seaborn, Plotly

Databases

SQL, MongoDB

BI

PowerBI, Tableau

Selected Work History

Aug 2023 - Present

Data Scientist and AI engineer @ Leniolabs_

- Large Language Models (LLMs) and prompt engineering: GPT, Claude, Mixtral, Llama, codeLlama.
- Image generation: Stable Diffusion, Dall-E
- Developed internal tools and client-specific applications to support business operations and customer interactions.
- Engineered chatbots and virtual assistants using LLMs, incorporating Langchain agents in some cases, to facilitate communication and service while addressing memory issues when surpassing the model's context limits..
- Implemented Retriever-Augmented Generation (RAG) to integrate relevant documents into the knowledge base of assistants, improving response accuracy. (Pinecone, LlamaIndex)
- Developed code for integrating image generation in video game asset production.
- Created a storyteller agent for a children's app that generates bedtime stories based on user-input moral values and produces a corresponding image for the story cover.
- Worked on the detection and classification of Personally Identifiable Information (PII) to ensure data privacy and compliance.

- Built a customer service agent that uses machine learning to assist in resolving customer inquiries effectively.
- Developed a voice detection and interaction tool as a widget to facilitate form completion on webpages, enhancing user accessibility and efficiency.
- Developed realistic deepfakes to raise awareness among a client's workers, emphasizing the importance of recognizing and understanding synthetic media. Wrote an article about it.

Sept 2022 - Aug 2023

Data Scientist @ Unumbio, Argentina

- Drive the interaction between the technical and business areas of the

company, to ensure active cooperation in identifying analytical needs,

and develop insights into different aspects of the business.

- Developed and trained several NLP NER models by fine tuning spaCy.
- Build various computer vision and OCR systems for specific types of

images, automating the text extraction process.

- Develop the company's custom BI platform connected to a MongoDB

data base for all users attending the specific needs of each role.

- Build a predictive model using various machine learning tools to prevent

the possibility of outdated sources or data and automate an email

notification alerting this behaviour.

- Develop code to automatically connect to APIs and download over 10M files.

Feb 2022 - Sept 2022

Data Scientist Intern NLP @ Orcawise, UK

- Developed and trained a custom NLP model for relation extraction by

fine tuning BERT

- Processed and tagged data to train the model
- Wrote and published articles in the company's web page: "Why shift from

a pretrained model to a custom model" and "How to create a custom

2019 - 2022

model for relationship extraction".

Scientific Research in Physics @ Universidad Nacional del Sur (UNS)

- Data handling, statistical analysis, finding patterns, literature review,

problem-solving, and communication with professors and a team.

2016 - 2021

Tutoring - Physics and Mathematics

- Communicate and present to a non-technical audience.
- Data driven analysis: assess what students already know and develop

strategies to attend specific difficulties.

Education

2016-2024

Universidad Nacional del Sur (UNS), Bahía Blanca, Argentina

Physics - Licenciate's Degree (MS level)

Thesis: Multilayers of opposite charge species on interfaces: An X-ray reflectivity study.

Courses

- **IBM Data Science Professional Certificate (44 weeks)**

  This Professional Certificate from IBM consists of 9 courses that include: open source tools and libraries, methodologies, Python, databases, SQL, data visualization, data analysis, and machine learning. IBM Capstone Project

- **Artificial Intelligence Multidisciplinary Training Program - Universidad de Buenos Aires (UBA)**
  - Data governance (120 hs)
  - AI governance (50 hs)
  - AI design, development and deployment (200 hs)