Francisco Fernando Roberto Pereira, PhD

+5583999504777 francisco.fernando@ufpe.br https://www.linkedin.com/in/engfranciscofernando/

EDUCATION

09/2023 - 06/2024
Rio de Janeiro, Brazil

Postgraduate Diploma in Data Science and Analytics
Pontifical Catholic University of Rio de Janeiro
- Coursework: Python, Exploratory Data Analysis, Machine Learning, Database, SQL, and Fundamentals of Project Management

02/2015 - 04/2019
Cambridge, United Kingdom

PhD in Materials Science and Metallurgy
University of Cambridge
- Main activities: Research and Design of Experiments, Data Processing, Computational Modelling, and Implementation of Optimisation Algorithms

01/2013 - 12/2014
Campina Grande, Brazil

Master of Science in Mechanical Engineering
Federal University of Campina Grande
- Main activities: Research, Design of Experiments, Data Processing, and Advanced Statistical Analysis

05/2007 - 12/2012
Campina Grande, Brazil

Bachelor of Science in Mechanical Engineering
Federal University of Campina Grande
- Coursework: Differential Calculus, Physics, Statistics, and Principles of Design

EXPERIENCE

07/2025 - Present
Remote (Brazil and US)

AI Researcher
Progress Rail, a Caterpillar Company
- Design and develop context-aware AI chatbots, leveraging frameworks such as LangChain, LangGraph, and others.
- Integrate LLMs and agentic frameworks into scalable products to improve functionality and user experience.
- Implement RAG pipelines and manage vector databases for knowledge retrieval.
- Prototype and deliver proofs of concept (PoC) to demonstrate AI-driven solutions for product innovation.
- Design and facilitate AI workshops to train and upskill teams on cutting-edge technologies.
- Collaborate with teams of scientists and quality analysts on cross-functional projects.

12/2020 - 06/2025
Recife, Brazil

AI Researcher, Data Scientist and Lecturer
Federal University of Pernambuco (UFPE), Department of Engineering
- Design and conduct research in AI applications targeting industry problems
- Design and structure course content for Data Science and Engineering, integrating foundational topics such as differential calculus, linear algebra and machine learning
- Conduct lectures, practical sessions, and workshops on emphasizing hands-on applications and industry-relevant skills on topics such as classic machine learning and deep learning
- Develop interactive learning materials, including lecture notes, project guides, coding exercises, and case studies to connect theory with practice
- Write technical documents to communicate project findings to academic, industrial, and general audiences
- Supervise undergraduate projects, guiding students from problem definition and data collection to analysis, modelling, and implementation

09/2023 - 06/2024
Rio de Janeiro, Brazil

Junior Data Scientist
Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Department of Computing
- Clean and prepare data to identify patterns and generate insights
- Create interactive visualizations to communicate results, following best practices in design and perception
- Develop predictive machine learning models for classification and regression, ensuring model optimisation and performance evaluation
- Perform hyperparameter optimisation and implement ensemble methods to enhance model performance
- Implement and manage data pipelines, build data lakes and data warehouses, and use SQL for querying and analysing large datasets
- Conduct multiple projects on real-world datasets from a wind farm

01/2020 - 12/2020
São Paulo, Brazil

Postdoctoral Fellowship
University of São Paulo (USP), Department of Engineering
- Plan and conduct controlled experiments to validate hypotheses and develop technical solutions
- Utilise specialised computational simulation software to predict outcomes and optimise project performance
- Collect, process, and analyse data, ensuring precision and reliability in result interpretation
- Create detailed technical drawings using CAD tools, adhering to technical standards to support projects and experiments
- Evaluate input and output variables to optimise processes and enhance project outcomes
- Prepare technical and scientific reports covering all project stages, from planning to final presentation

EXPERIENCE

08/2011 - 12/2011

Lexington, KY, USA

Exchange and Internship Program

University of Kentucky, Department of Agricultural and Biosystems Engineering

- Participate in an intensive English language programme at the Language Center, focusing on technical and academic communication skills
- Engage in classes and academic activities with graduate students
- Conduct an internship on the application of cameras for the automation of agricultural equipment

KEY PROJECTS

AI-Powered Web Platform for Exploring High School National Exam Data

10/2024 - Present

This project uses open data from the Brazilian High School National Exam to build a free AI-powered platform for exploring educational trends, regional disparities, and socioeconomic factors through natural language queries. Key objectives:

- Clean and preprocess over 30 GB of raw data from open-source government database
- Develop and refine socioeconomic indicators, attendance rates, and subject-specific scores (Mathematics, English, and Writing) along with other critical variables to enhance data analysis
- Implement LLM API integration to allow users to explore insights through natural language queries
- Fine-tune API temperature settings and apply prompt filtering to enhance accuracy, relevance, and reliability of responses
- Build a user-friendly, interactive web platform, freely accessible to the public
- Skills: Data Preprocessing and Cleaning, LLM APIs, Web Development, Backend Processing, and Cloud Deployment

Wind Energy Data Management

04/2024 - 06/2024

https://github.com/proffranciscofernando/DataScience-PUCRio-MVP3

Key results:

- Optimised turbine positioning based on peak production (noon), lowest output (10 PM), and wind directions (170-220 degrees)
- Found a 0.94 correlation between wind speed and energy output, improving forecast accuracy
- Detected a small discrepancy between actual and theoretical energy output, indicating near-optimal performance
- Streamlined data management using delta lake with ACID transactions
- Skills: Data Pipeline Development and Metadata Management

Wind Farm Power Prediction

01/2024 - 03/2024

https://github.com/proffranciscofernando/DataScience-PUCRio-MVP2

Key results:

- Conducted representative train-test splits for diverse data coverage
- Standardised features to enhance model performance
- Optimised power output predictions using XGBRegressor, achieving top R^2 and RMSE metric
- Applied hyperparameter tuning and cross-validation to ensure model generalization
- Delivered accurate power output predictions, closely matching theoretical values
- Skills: Machine Learning, Regression Modelling, and Ensemble Methods

Preparation and Exploration of Wind Farm Data

09/2023 - 12/2023

https://github.com/proffranciscofernando/DataScience-PUCRio-MVP1

Key results:

- Managed outliers from failures and low-performance periods in the wind turbine.
- Planned and conducted cleaning, data preprocessing, and EDA to identify patterns, trends, and anomalies in wind farm operations.
- Skills: Exploratory Data Analysis, Data Preprocessing, Cleaning, and Visualisation.

KEY PUBLICATIONS

Extraction of Superelastic Parameter Values from Instrumented Indentation Data

University of Cambridge

Roberto Pereira, FF

https://doi.org/10.17863/CAM.37445

PhD Thesis

Extraction of Superelasticity Parameter Values from Instrumented Indentation via Iterative FEM Modelling

Roberto Pereira, FF, Campbell, JE, Dean, J, Clyne, TW

https://doi.org/10.1016/j.mechmat.2019.04.007

Key results:

- Iterative modelling combined with grid search across the material property parameter space
- Automation of error minimisation with systematic algorithm
- Extraction of material properties and validation against experimental results for accuracy

TRAINING / COURSES

General Course on Intellectual Property
World Intellectual Property Organization

Basics of Patent Drafting
World Intellectual Property Organization

Advanced Patent Information Search
World Intellectual Property Organization

LANGUAGES

English
Native

Spanish
Advanced

Portuguese
Native

SKILLS

Python | **PyCharm** | **Databricks** | **SQL** | **LLM APIs** | **GitHub**

AWARDS

CISER Award of Technological Innovation, 2009.

CISER Award of Technological Innovation, 2016.

REFERENCES

Professor Emeritus Bill Clyne (Trevor William Clyne)

PhD supervisor
Email: twc10@cam.ac.uk
Department of Materials Science and Metallurgy - University of Cambridge
27 Charles Babbage Rd, Cambridge CB3 0FS, United Kingdom

Dr James Dean - CEO at Plastometrex

Research collaborator
Email: j.dean@plastometrex.com
204 Cambridge Science Park Milton Rd, Milton, Cambridge CB4 0GZ, United Kingdom