

# João Pereira

Data Science Lead

🌐 [www.joaopereira.ai](http://www.joaopereira.ai) | [in jpcpereira](https://www.linkedin.com/in/jpcpereira)

✉ [joao.pereira.abt@gmail.com](mailto:joao.pereira.abt@gmail.com) | 📞 +31 619792025

Enjoy creating scalable machine learning solutions in the cloud. Love building a culture of innovation, technical excellence, and where people feel empowered to be the best version of themselves. Passionate about transforming cutting-edge AI technologies into impactful products that drive business growth.

## Experience

adidas

Amsterdam, The Netherlands

**Data Science Lead, Content Intelligence**

2024 - Now

- Leading GenAI products deployed on Databricks and AWS Bedrock that enable better product discovery and deliver personalized experiences to millions of users across adidas.com and App.
- Providing guidance and mentoring to Data Scientists and supporting global hiring efforts.

**Technical Lead, Demand Forecasting**

2023

- Led development of ML-based demand forecasting products deployed on AWS that drive +70M net sales by optimizing buying, planning, and trading processes globally.
- Provided technical leadership to a team of 5-7 Data Scientists and ML Engineers.
- Led innovation project on graph neural networks for demand forecasting [↗](#)

**Data Scientist, Demand Forecasting**

2021 - 2022

- Built demand forecasting models (PyTorch) and deployed them to production on AWS SageMaker.
- Implemented software engineering best practices, elevating technical standards and maturity.

Belastingdienst

Utrecht, The Netherlands

**Data Scientist** (via JADS)

2020

- Developed custom object detection and image classification models (TensorFlow).
- Created an application that enabled non-technical experts to use the ML models.

ASML

Veldhoven, The Netherlands

**Data Scientist** (via JADS)

2019

- R&D Department: Developed computer vision models to detect dust particles on silicon wafers.
- Supply Chain Department: Applied process mining to identify bottlenecks in the supply chain.

Van Lanschot Kempen

Den Bosch, The Netherlands

**Data Scientist** (via JADS)

2019

- Client Monitoring Department: Applied community detection to identify money laundering cases.

Institute for Systems and Robotics

Lisbon, Portugal

**Student Researcher**

2018

- Developed state-of-the-art, unsupervised deep learning models for time-series anomaly detection [↗](#)

## Education

Eindhoven University of Technology, JADS

Eindhoven, The Netherlands

**Engineering Doctorate (EngD), Data Science**

2019 - 2021

- Thesis topic: Large-scale object detection and image analysis. | Grade: 9.5/10
- Worked on projects for ASML, Van Lanschot, TE Connectivity, Heijmans, Gemeente Den Bosch.

Instituto Superior Técnico

Lisbon, Portugal

**Masters and Bachelors, Electrical and Computer Engineering**

2013 - 2018

- Thesis topic: Anomaly detection in time series data using deep learning [↗](#) | Grade: 20/20

Université Catholique de Louvain

Louvain-la-Neuve, Belgium

**Exchange Student**

2016

## Talks

- ECML-PKDD'23, Probabilistic Demand Forecasting With Graph Neural Networks [↗](#)
- AWS Build On Generative AI, Hyperparameter Tuning on SageMaker [↗](#)
- adidas, Fueling the adidas Digital Business Through Data Science and AI
- Tech Summit, adiFormer: A Large Language Model for Generating Product Copies

## Blog Posts

- Fast and Scalable Hyperparameter Tuning and Cross-validation in AWS SageMaker [↗](#)
- Fine-tune MPT-7B on Amazon SageMaker [↗](#)

## Selected Publications (300+ citations, see more on Google Scholar [↗](#))

- N. Kozodoi, L. Zinovyeva, S. Valentin, J. Pereira, R. Agundez, Probabilistic Demand Forecasting With Graph Neural Networks [↗](#), Turin, Italy, ECML-PKDD'23 (ML4ITS).
- J. Pereira, M. Silveira, Unsupervised Anomaly Detection in Energy Time Series Data Using Variational Recurrent Autoencoders With Attention [↗](#), Orlando, USA, ICMLA'18.
- J. Pereira, M. Silveira, Learning Representations From Healthcare Time Series Data for Unsupervised Anomaly Detection [↗](#), Kyoto, Japan, BigComp'19.

## Skills

- Cloud & Infrastructure: AWS (SageMaker, Bedrock, Lambda, S3), Databricks, Terraform for IaC.
- Data Processing: Apache Spark, Pandas, and Polars.
- Programming Languages: Python, SQL, Matlab, C.
- CI/CD: Jenkins and Github Actions.

## Trainings

- AWS LLMOps, Herzogenaurach, Germany . . . . . Nov 2023
- Manager Development Experience, adidas [↗](#) . . . . . Jun - Dec 2022
- AWS Distributed Training and MLOps, London, United Kingdom . . . . . Sep 2022
- AWS SageMaker Inference, Virtual . . . . . Sep 2022