

# João Pereira

📍 Amsterdam, The Netherlands

🌐 [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 06-19792025

Data science lead with a proven track record of building and deploying impactful AI/ML systems at scale, serving 300M+ consumers. Experienced in taking innovative ideas from research to production. Passionate about building a culture of technical excellence where people feel empowered to be the best version of themselves.



## EXPERIENCE

**Data Science Lead – Discovery & Personalization** 2024 - Now

**Technical Lead & Senior Data Scientist – Forecasting** 2023

**Data Scientist – Forecasting** 2021 - 2022

adidas

- Spearheaded the end-to-end development of AI/ML solutions from idea to production across multiple product areas including discovery, personalization, and forecasting, enhancing experiences for 300M+ consumers globally throughout their journeys:
  - Discovery: Product tagging using fine-tuned deep learning models, improving discovery by 20% and saving 5k manual working hours/year. Plus automated review of product metadata using LLM-as-a-Judge to ensure high data quality.
  - Personalization: Recommendation engines that recommend content and offers 1:1 and in real-time, maximizing consumer engagement at scale. This capability includes GenAI-based tagging of content assets.
  - Forecasting: Pre-season and in-season demand forecasting using probabilistic deep learning models, driving +€70M net sales/year by increasing product availability while reducing excess stock. As a technical lead, I looked after the science, ML engineering, and architecture aspects of the products.
- Mentored a globally distributed team of Data Scientists and ML Engineers.
- Aligned cross-functionally between Digital Product, Tech, and Market teams to translate business opportunities into impactful and scalable ML solutions.
- Established ML Engineering and Operations best practices, reducing time-to-market and raising the technical standards (e.g., experiment tracking and evaluation, model training and inference, IaC).
- Partnered with the AWS ML Solutions Lab to develop an innovative graph neural network model for demand forecasting [↗](#).

### Data Scientist

Jheronimus Academy of Data Science 2019 – 2021

Delivered machine learning solutions for diverse partners:

- *Dutch Tax Authorities* – Developed an image recognition application powered by custom built deep learning models.
- *ASML* – Developed deep learning models to detect dust particles on silicon wafers.
- *Van Lanschot Kempen* – Applied community detection methods to identify suspicious cases of money laundering.
- *TE Connectivity* – Built a product recommendation engine.
- *Heijmans* – Developed a deep learning model to predict the occupancy in working spaces using sensor data.

## EDUCATION

**Engineering Doctorate (EngD) in Data Science** | Thesis Grade: 9.5/10  
Eindhoven University of Technology 2019 – 2021

**Master's & Bachelor's in Electrical & Computer Engineering** | Thesis Grade: 20/20  
Instituto Superior Técnico, University of Lisbon 2013 – 2018

## SELECTED PUBLICATIONS

390+ citations on Google Scholar [↗](#)

- Probabilistic Demand Forecasting With Graph Neural Networks, ECML-PKDD'23 [↗](#)
- Unsupervised Anomaly Detection in Energy Time Series Data Using VAEs With Attention, ICMLA'18 [↗](#)
- Learning Representations From Healthcare Time Series Data for Unsupervised Anomaly Detection, BigComp'19 [↗](#)
- Fine-tune MPT-7B on Amazon SageMaker, Medium [↗](#)
- Fast and Scalable Hyperparameter Tuning and Cross-validation in AWS SageMaker, Medium [↗](#)

## SKILLS

- **Programming Languages** – Python, SQL, Matlab, C.
- **Data Processing** – Spark, Kafka, Pandas, and Polars.
- **Cloud Platforms** – AWS (Bedrock, SageMaker, Lambda, S3, Glue, EMR), Databricks (Workflows, Model Serving, Unity Catalog).
- **AI/ML** – PyTorch, Transformers, MLFlow, TensorFlow.
- **Infrastructure as Code** – Terraform, Databricks Asset Bundles, AWS CDK.