# Erik Jan de Vries

AI & MLOps Architect | Data / MLOps Engineer | Data Scientist

Analytics Product Development | Data & Analytics Strategy | Building Analytics Capabilities

## Key words

Results-oriented
Strategic thinker
Innovation mindset
Agile approach
Strong communication skills
Team player

MLOps
Generative AI
Responsible AI
Marketing Analytics

## Education

| | |
|---|---|
| PhD | Mathematical Physics, Heriot-Watt University Edinburgh, UK |
| MSc | Theoretical Physics, Leiden University (cum laude, GPA 8.5) |
| TVWO | Bilingual high school English / Dutch (cum laude, GPA 9) |

## Contact details

*Hidden on public CV*

linkedin.com/in/erikjandevries

## Award-winning AI & MLOps Consultant

With over 13 years of experience, I have helped many organisations optimise the use of their data assets. To this end, I have developed a framework integrating **business**, **organisation** and **technology**, enabling me to demonstrate in a practical and pragmatic way how to:

(1) embed data-driven decision-making into all types of business processes,
(2) build effective analytics products, focussing on business value, and
(3) create an environment for success with data and analytics.

I have developed many AI solutions and built the MLOps platforms to bring them to production, for numerous organisations across different industries. Through trainings on various topics, including data science, MLOps and Generative AI, I empower my clients to fully own, operate and maintain these solutions independently after my engagement.

As a Data & Analytics strategy consultant, I aim for long-term success by defining ambitions for the future, laying out a roadmap to achieve those goals, and implementing tools to monitor progress along the way, enhancing the data and analytics capabilities and maturity of organisations.

In my spare time, I love playing the piano. A highlight in my musical career was playing Gershwin's Rhapsody in Blue at the concert hall in Haarlem. In 2021, I became father of a beautiful little girl.

## Highlights

| | | |
|---|---|---|
| 2024 | Designed and built the MLOps platform (Azure ML) | Van Oord |
| 2023 | Led the Data & Analytics team | Triple A - Risk Finance |
| 2022 | Designed and built the MLOps platform (GCP) | Liberty Global |
| 2020 | Created the Next Best Action Marketing solution (Azure) which **won the Data Driven Marketing Award 2023** | Eneco |
| 2017 | Advanced analytics platform (Hadoop, Spark) | VodafoneZiggo |

## Work experience

2023 – 2024 **Freelance consultant AI & MLOps, architect, engineer** Tadata

2022 – 2023 **Practice lead Data & Analytics** Triple A - Risk Finance

Led the Data & Analytics practice, defining a strategy for growth and team development. Managed a team of data professionals, providing coaching and mentorship, and ensuring high-quality deliverables for clients. Led analytics and reporting initiatives, utilising advanced techniques to provide valuable insights and recommendations to clients. Collaborated with cross-functional teams and clients to identify opportunities for data-driven optimisation and innovation in their business processes.

2018 – 2022 **Lead data scientist, consultant** BigData Republic

Co-developed the BigData Republic methodology, acted as coach, mentor and trainer.

Advised clients as all-round expert: developing the analytics strategy, interim manager data science, member of a Digital Transformation steering committee, developing analytics skills and technologies, data science project lead, machine learning architect, analytics translator, commercial product development, coach, mentor, trainer.

2016 – 2017 **Chief data scientist, consultant** Future Facts

Responsible for the technological roadmap, methodology, coaching and supervising data scientists and data engineers, and knowledge sharing and skills development.

Advised clients on predictive analytics, developed machine learning models.

2015 **BI and data science consultant** i3, Tadata & CGI

2010 – 2014 **Business Intelligence and Data Analytics specialist** Kas Bank

2009 **Strategy consultant** Strategy Development Partners

## Projects

2023 – 2024 **MLOps platform** Van Oord

Extended the Data Platform with an MLOps capability using Azure ML. Designed and architected the solution, created the infrastructure and built the first ML pipeline to serve as template for following projects. Created a roadmap for further development.

2024 **Workshops: Generative AI with Microsoft Copilot** Van Oord

Conducted workshops on Generative AI using Microsoft Copilot, focusing on prompt engineering for Large Language Models (LLMs). Designed and delivered training sessions, demonstrated practical use cases, and facilitated hands-on exercises.

2024 **Retrieval-Augmented Generation (RAG)** Van Oord

Developed a proof of concept for a Retrieval-Augmented Generation (RAG) system using LLMs via the OpenAI API and a Qdrant vector database. Refined the Request for Proposal (RfP), defining project objectives, technical requirements, and evaluation criteria. Led the vendor selection process, evaluating proposals, conducting interviews, and facilitating discussions to identify the best-fit solution provider.

2023 **Bias mitigation with Adversarial Learning** Triple A - Risk Finance

Built a solution for bias mitigation with Adversarial Learning using both TensorFlow / Keras and PyTorch, developing our Responsible AI proposition.

2021 – 2022 **Operational Data Hub, Technology Deployment Analytics** Liberty Global

Advised the CTO and senior management on a migration to the cloud, developed the machine learning architecture using MLOps on Google Cloud Vertex AI, and trained the internal team in the new way of working while implementing and deploying the first project in production using fully automated CI/CD pipelines in the new architecture.

2020 – 2021 **Next Best Action Marketing** Eneco

**With this solution, Eneco won the Data Driven Marketing Award 2023**.

Designed the Next Best Action architecture and end-to-end solution, created a proof-of-concept and an implementation roadmap, and acted as technical lead of a cross-functional team developing and implementing the solution. The modular design allowed for an incremental and agile approach.

Stakeholder management in the Digital Transformation programme, aligning the NBA initiative with other projects, such as the implementation of a new CRM system.

The architectural board selected my solution for their target architecture, favouring it over popular commercial products such as Pega.

2019 – 2021 **Data science platform** Eneco

In collaboration with the internal lead engineer, created a DevOps data science platform on Microsoft Azure and Snowflake. Coached and led various data science teams.

2018 – 2019 **Customer analytics and finance analytics** KLM

Led and coached two data science teams for Customer and Finance (using SAFe Scrum).

In one project, I extracted topics and sentiment from free format text feedback from customers using BERT, to provide focus for customer service improvement projects.

2018 **Image segmentation** Nutreco

Set up a deep learning framework for the development of an image segmentation model of microscope photos of fodder (using TensorFlow / Keras).

2016 – 2017 **Advanced analytics platform, marketing analytics** Vodafone Ziggo

Co-designed and -developed the advanced analytics platform, using MapR Hadoop and Spark. Built data and ML pipelines using Python and R. Designed the analytics way-of-working for the new platform.

## Other experience

| | |
|---|---|
| 2023 | Presented at the VSAE Actuary Conference on ethical and explainable AI. |
| 2019 | Won 1st prize at KLM Data Science Community Demo Festival |
| 2016, 2017 | Taught guest lectures on Data Science at Fontys University of Applied Sciences. |
| 2012 – 2014 | KasXchange (Young Professionals Kas Bank): first treasurer, then president. |
| 2009 | Represented Heriot-Watt University at an interdisciplinary, national conference. |
| 2006, 2007 | Co-organised the Post Graduate Research Conference, Heriot-Watt University. |
| 2001 – 2002 | Board member Sempre Crescendo, a students' music society. |

## PUBLICATIONS

| | |
|---|---|
| 2024 | **Empowering Efficiency: Building a Self-Service Platform for Analytics**, Erik Jan de Vries<br>https://erikjandevries.medium.com/empowering-efficiency-building-a-self-service-platform-for-analytics-9a28349ea0c6 |
| 2024 | **The Power of MLOps: Benefits for Business Leaders and Managers**, Erik Jan de Vries<br>https://erikjandevries.medium.com/the-power-of-mlops-eccdef9ef5f6 |
| 2024 | **Defining Machine Learning and MLOps**, Erik Jan de Vries<br>https://erikjandevries.medium.com/defining-machine-learning-and-mlops-d2b0e8325959 |
| 2020 | **Data science is boring**, Erik Jan de Vries<br>https://medium.com/bigdatarepublic/data-science-is-boring-1756a7be1899 |
| 2012 | **Rhapsody in Blue (Gershwin)**, Erik Jan de Vries, de Amsterdamse Tramharmonie<br>http://www.youtube.com/watch?v=lJsnGprFRa8 |
| 2010 | **Supercharges, Quantum States and Angular Momentum for N=4 Supersymmetric Monopoles**, Erik Jan de Vries, Bernd J. Schroers (published in Nucl.Phys.B) |
| 2008 | **Supersymmetric Quantum Mechanics of Magnetic Monopoles: A Case Study**, Erik Jan de Vries, Bernd J. Schroers (published in Nucl.Phys.B) |

## TECH SKILLS

### Cloud Platforms

- Microsoft Azure
- Google Cloud Platform

### Software development and programming

- Source code version control: Git, GitHub, GitLab, Subversion (SVN)
- CI/CD pipelines: GitHub Actions, Azure DevOps, Google Cloud Build
- Python: numpy, pandas, scikit-learn, matplotlib, etc.
- R: dplyr, ggplot2, caret, data.table, shiny, etc.
- C#, VBA, DAX (Power BI), MDX (SSAS)
- Shell scripting: bash, zsh, PowerShell
- Jupyter Notebooks, Jupyter Lab

### Machine Learning and MLOps

- Classification, regression, recommender systems, gradient boosted trees: XGBoost, LightGBM
- Neural networks and deep learning: TensorFlow, Keras, PyTorch
- Image classification and segmentation, object detection: CNN, U-Net
- Time series analysis: Prophet, RNN, CNN
- Reinforcement learning: Deep Q-Learning
- MLOps platforms: Azure ML, Vertex AI, Kubeflow, Airflow, MLflow
- A/B testing, statistical analysis, fairness, feature importance: Shapley analysis (SHAP)
- Virtualisation / containerisation: Docker, Kubernetes (K8s), Hyper-V, VirtualBox

### Data Engineering and Data Analysis

- SQL and relational databases: Microsoft SQL Server, Oracle, Postgres/PostgreSQL, Google Cloud SQL
- ETL processes and data pipelines: SSIS, dbt, Apache Airflow
- Data warehouses: Data Vault, Dimensional modelling, Google BigQuery
- Data lakes: HDFS, Google Cloud Storage, Azure Data Lake
- Distributed computing: Apache Hadoop, HDFS, Hive, MapR, Spark
- Streaming data: Kafka, Google Cloud Pub/Sub
- Data quality: dbt, Great Expectations
- Data visualization: matplotlib, seaborn, plotly, ggplot2
- Business Intelligence: Power BI

### Generative AI

- OpenAI ChatGPT, Microsoft Copilot
- Prompt engineering
- Retrieval-Augmented Generation (RAG), Large Language Models (LLMs)
- Vector databases: Qdrant

### Workflow and project management

- Atlassian: Confluence, Jira, BitBucket
- Agile project management: Scrum, Scaled Agile Framework (SAFe), Kanban