

# FEDERICO FIORIO

## Data & Machine Learning Engineer

+39 342 596 4621 | [fioriofederico99@gmail.com](mailto:fioriofederico99@gmail.com)

[GitHub](#) | [LinkedIn](#) | [Portfolio](#)

Results-driven **Data & Machine Learning Engineer** with expertise in **cloud-based data engineering, AI, and big data**. Proven track record in designing and optimizing **high-performance data architectures** at scale. Passionate about **building resilient, scalable data solutions and deploying ML models** to drive business impact.

## EXPERIENCE

### Data & Machine Learning Engineer | Data Reply, Milan, Italy

May 2024 – Present

- **Enterprise-Scale Data Pipelines & Historical Data Optimization:** Designed **high-volume** data ingestion pipelines (**Databricks, Medallion Architecture**) for a luxury brand; optimized historical tracking with **SCD Type 2**, reducing query time by 80%.
- **Optimized Data Orchestration:** Migrated **Airflow** DAGs to **GCP Composer 2.x**, boosting scheduling efficiency by **10x**.
- **Sanctioned Shop Blocker:** Optimized sanctions-screening **lakeflow job** for customer onboarding; Optimized sanctions-screening process, significantly improving performance and reducing data size to a quarter.
- **CI/CD:** Established CI/CD pipelines in Azure DevOps and GitHub Actions to automate code testing with pytest.
- **RAG-powered CloudFormation Template Generator:** Developed a UI-driven tool to generate **AWS CloudFormation** templates from natural language. The engine retrieves AWS formatting documentation to guarantee a well-written YAML output

### Data Engineer | Management Solutions, Milan, Italy

October 2023 - April 2024

- **IBM DataStage refactoring:** Optimized IBM DataStage ETL workflows and **SQL** for banking data warehouses.
- **Data Governance & Compliance:** Enforced **data governance** policies in cross-functional teams

### Software Engineer Intern | Diliium, Milan, Italy

February 2021 - May 2021

- **GANs Research:** Conducted research on GANs for deepfake generation, emphasizing ethical considerations
- **Training & Awareness:** Led training sessions for employees on the emerging risks of these technologies, especially audio forgery

## EDUCATION

- **MSc in Computer Science | University of Milan, Erasmus+ (University of Copenhagen)** Grade: 108 / 110  
[Thesis:](#) Text-to-image diffusion models attacks in scarce resource environments and privacy issues
- **BSc in Computer Science | University of Milan**

## CERTIFICATIONS

- **All certifications:** Databricks Certified Data Engineer Professional, Databricks Certified Associate Developer for Apache Spark, Databricks Certified Data Engineer Associate, Neo4j Certified Professional, AWS Certified Machine Learning – Specialty, Professional Machine Learning Engineer Certification

## PROJECTS

- **World's biggest space hackathon winner 2025, Zurich local competition** – Team Leader and developer | [Project](#)
- **Personalized BoardGames RAG** – Enables interaction with game rules via natural language interface | [Repo](#)
- **AlphaEvolve Mini** – Developed a simplified version of AlphaEvolve to enhance understanding of evolutionary algorithms | [Repo](#)
- **QueryScorer** – Implemented CI pipeline to ensure only optimized SQL queries are committed | [Repo](#)
- **Data Structures & Algorithms** | [training & contests](#)

## SKILLS

- **Programming & Versioning:** Python, Java, C, SQL, Pyspark, GitHub, Bitbucket
- **Cloud & Deployment:** Google Cloud (Composer, BigQuery, IAM, Airflow, GCS), Microsoft Azure (Databricks, Azure Blobs, DevOps), Amazon Web Services (S3, Lambda, Step Functions, DynamoDB)
- **Data Engineering:** Databricks, Apache Spark, SQL, NoSQL (MongoDB), RDBMS (PostgreSQL, SQL server), IBM DataStage, Airflow
- **Operations:** Azure DevOps, Github Actions, Terraform, Docker
- **Machine Learning & AI:** Scikit-learn, Pytorch, Vector Databases (Qdrant), Langchain, OpenAI API, Streamlit, MLflow
- **Agile & Scrum:** Jira
- **Agents Automation:** [n8n](#)