

FEDERICO FIORIO

Data & Machine Learning Engineer

Milan, Italy

+39 342 596 4621 | fioriofederico99@gmail.com

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

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** in **Databricks**, implementing **Medallion Architecture** for a fashion **luxury brand**. Replaced inefficient **JOIN-heavy historical tracking** with **SCD Type 2**, reducing query complexity and processing time by 80%, improving scalability, performance, and maintainability
- **Optimized Data Orchestration:** Migrated **Airflow DAGs** from legacy composer to **GCP Composer 2.x**, boosting **scheduling efficiency by 10x** for a global manufacturing client
- **AI-Driven Document Retrieval:** Developed **RAG** pipelines for a leading bank, leveraging **vectorDB** and **OpenAI** embeddings, and custom document techniques
- **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 ETL optimization:** Developed ETL workflows with **IBM DataStage** and **SQL** for banking data warehouses, improving legacy **SQL**
- **Data Governance & Compliance:** Enforced **data governance** policies in cross-functional teams

Software Engineer Intern | Dilium, 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
 - **BSc in Computer Science | University of Milan**
-

CERTIFICATIONS

- [Google Cloud Professional Machine Learning Engineer](#)
 - [AWS Machine learning Specialty](#)
 - [Databricks Certified Data Engineer Associate](#)
-

PROJECTS

- **BoardGames RAG** | [Repo](#)
 - **AlphaEvolve Mini – A simple copy of AlphaEvolve** | [Repo](#)
 - **Thesis: Membership inference attacks on text-to-image models** | [Repo](#)
 - **Competitive programming** | [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), Amazon Web Services (S3), Docker
- **Data Engineering:** Databricks, Apache Spark, SQL, NoSQL (mongoDb), RDBMS (postgre, SQL server), IBM DataStage, Airflow
- **Machine Learning & AI:** Scikit-learn, Vector Databases (VectorDB, Qdrant), Langchain, OpenAI API
- **Agile & Scrum:** Jira
- **Others:** Streamlit, Unity