



Reproducible and Responsible Data Science Practices Training Course

2 – 6 May 2027
Johannesburg

Reproducible and Responsible Data Science Practices Training Course

Ref: 793_160281 **Date:** 2 – 6 May 2027 **Location:** Johannesburg **Fees:** 4500 Euro

Overview

Reproducible and Responsible Data Science Practices Training Course is a five-day professional course for data science practitioners, analytics leads, and model owners who leave with a Reproducible Data Science Control Pack. Participants connect decision context, versioned data and code, experiment tracking, validation, documentation, deployment gates, monitoring, risk, and review. The course addresses fragile analytical work by making results traceable, repeatable, testable, and governed. Agile Leaders Training Center develops responsible data science practice.

Who Should Attend

- Data science functions responsible for experiments, models, evidence, and technical decisions
- Analytics leadership functions responsible for quality, review, prioritization, and delivery
- Model ownership functions responsible for performance, risk, approval, and continued use
- Machine learning engineering functions responsible for pipelines, releases, monitoring, and rollback
- Data governance functions responsible for provenance, quality, privacy, access, and accountability
- Risk and assurance functions responsible for independent testing, oversight, and documentation

The course assumes participants can interpret datasets, code, model metrics, and analytical experiments, and leaves out introductory statistics, basic programming instruction, platform administration, and external certification preparation.

Departments and Industries

The course supports departments and industries that build, deploy, or oversee analytical models and data products.

- Financial services risk and decision analytics
- Retail demand and customer modeling
- Manufacturing quality and predictive maintenance
- Healthcare operations and outcome analytics
- Telecommunications network and customer intelligence
- Public-service policy and resource analytics

Learning Objectives

By the end of this course, participants will be able to:



- Apply reproducible experiment and asset controls
- Build traceable data, feature, code, and model records
- Evaluate models through testing, validation, and review
- Use risk, fairness, privacy, and security evidence
- Build deployment, monitoring, and incident controls
- Prioritize improvement, rollback, and retirement decisions

Course Agenda

Day 1: Frame Responsible Practice

- Decision Context and Intended Use Canvas
- NIST AI RMF Govern, Map, Measure, Manage Map
- Stakeholder, Impact, and Accountability Matrix
- Risk, Trustworthiness, and Control Register
- Data Science Lifecycle and Review Gate Model

Day 2: Make Experiments Reproducible

- Data Provenance and Version Record
- Code, Dependency, and Environment Manifest
- Experiment Objective, Baseline, and Metric Card
- Feature, Label, and Sampling Traceability Sheet
- Run Logging and Result Comparison Protocol

Day 3: Test and Document Models

- Training, Validation, and Test Separation Plan
- Performance, Robustness, and Error Segment Matrix
- Bias, Fairness, Privacy, and Security Review
- Independent TEVV Evidence Checklist
- Model Card and Limitation Statement

Day 4: Govern Deployment and Monitoring

- Training-Serving Consistency Assessment
- Release Approval and Deployment Gate Checklist
- Data Drift, Concept Drift, and Performance Monitor
- Incident, Escalation, and Human Oversight Playbook
- Rollback, Retraining, and Retirement Decision Tree



Day 5: Practice Data Science Controls

- Exercise: Map Context, Stakeholders, and Risks
- Exercise: Reproduce an Experiment from Versioned Assets
- Exercise: Evaluate Model Evidence and Limitations
- Exercise: Design Monitoring and Incident Responses
- Capstone: Reproducible Data Science Control Pack

Practical Exercises

The course uses suggested activities based on financial services, retail, manufacturing, healthcare, telecommunications, and public-service analytical scenarios.

- Suggested activity: define intended use, affected stakeholders, objectives, risk tolerances, oversight, and evidence requirements for a model.
- Suggested activity: reconstruct an experiment from versioned data, features, code, environment, parameters, metrics, and run records.
- Suggested activity: compare validation, robustness, error, fairness, privacy, security, and limitation evidence before a release decision.
- Suggested activity: assemble review gates, deployment evidence, monitors, incidents, rollback criteria, and ownership in the final control pack.

FAQs

Who suits responsible data science practices training, and what does it assume?

Data scientists, analytics leads, model owners, ML engineers, governance teams, and assurance functions suit the training; it assumes experience with data, code, model metrics, and experiments.

How does responsible data science training differ from an introductory data science course?

Responsible data science training governs reproducibility, validation, risk, releases, monitoring, and accountability, while introductory training teaches foundational analysis, programming, statistics, and model-building techniques.

What makes a data science experiment reproducible?

A reproducible experiment records the exact data, code, environment, features, labels, parameters, random states, dependencies, metrics, and execution steps needed to obtain and explain the result.



Which evidence supports a responsible model release?

A responsible release uses intended-use approval, representative data evidence, validation and robustness results, error analysis, risk and impact review, documented limitations, human oversight, monitoring, rollback criteria, and accountable owners.

How should data science teams monitor deployed models?

Teams should monitor data quality, drift, prediction behavior, performance, errors by segment, incidents, user feedback, operating context, overrides, and business outcomes against thresholds linked to investigation and action.

Conclusion

Participants take back a Reproducible Data Science Control Pack containing context, provenance, versions, experiments, validation, risk evidence, model documentation, release gates, monitoring, incidents, rollback, and ownership. It changes isolated model work into a traceable lifecycle. The pack supports repeatable results and reviewable decisions throughout model use.