AI and Machine Learning Data Management Training Course

19 - 23 Oct 2026

Cape town

4600 € (Per Person)

Ref: #DM8537_264370

Course Introduction / Overview:

In the fast-paced world of artificial intelligence and machine learning, the quality and management of data are not just important, they're everything. This training course is designed to give participants a thorough understanding of the entire data lifecycle for AI and ML applications, from collection and cleaning to governance and deployment. We will explore how proper data management can make or break an AI project and cover the specific techniques needed to prepare data for training models. We will discuss topics such as data labeling, feature engineering, and dealing with imbalanced datasets. The course also addresses the ethical and privacy issues associated with managing large datasets for AI. As Thomas H. Davenport and R. S. Dalle discuss in their book, "The AI Advantage: How to Put the Artificial Intelligence Revolution to Work," successful AI implementation depends on a strong data strategy. At BIG BEN Training Center, we understand that building effective AI systems starts with a solid foundation of well-managed data. This training course will give participants the practical skills to build, maintain, and optimize data pipelines that are crucial for developing high-performing AI and ML models, giving them a significant advantage in this competitive field.

Target Audience / This training course is suitable for:

- Data scientists and machine learning engineers.
- AI product managers and project managers.
- Data architects and data engineers.
- Business intelligence analysts.
- IT professionals involved in AI and ML projects.
- Researchers and academics in related fields.
- Chief Data Officers and Chief Technology Officers.

Target Sectors and Industries:

- Technology and software development.
- Financial services, including banking and investment.
- Healthcare and pharmaceuticals.
- E-commerce and retail.
- Telecommunications.
- Automotive industry.
- Government agencies and defense.

Target Organizations Departments:

- Data Science and Analytics.
- AI and Machine Learning Engineering.
- IT Infrastructure and Operations.
- Product Development.
- Research and Development.
- Business Intelligence.

Course Offerings:

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

- Understand the complete data lifecycle for AI and ML projects.
- Apply techniques for effective data collection, cleaning, and preparation.
- Master the process of data labeling, annotation, and feature engineering.
- Implement strategies for managing large-scale datasets.
- Address and solve issues with imbalanced and noisy data.
- Establish data governance and quality frameworks for AI data.
- Build robust data pipelines for machine learning models.
- Understand ethical considerations in AI data management.
- Measure and optimize data readiness for AI deployment.

Course Methodology:

This training course uses a hands-on and practical approach to make sure participants can immediately apply their new skills. We will use a blend of interactive lectures, group workshops, and real-world case studies from various industries. A key part of our approach is using problem-based learning, where participants work together to tackle complex data challenges that are common in AI and ML projects. Participants will get their hands dirty with data, practicing data cleaning, feature engineering, and data pipeline creation using example datasets. Our expert trainers will provide personalized guidance and constructive feedback, helping participants improve their skills and understanding. At BIG BEN Training Center, we know that a theoretical understanding is not enough. Our goal is to make sure every participant is confident in their ability to handle the complexities of data for AI and machine learning in a professional setting.

Course Agenda (Course Units):

Unit One: The Foundation of Data for AI.

- Understanding the data lifecycle for AI and machine learning.
- The critical role of data quality in model performance.
- Data types and their impact on AI systems.
- Strategies for data collection and integration.
- Exploring the difference between structured and unstructured data.
- Introduction to data pipelines and ETL processes.
- Understanding the importance of data lineage.

Unit Two: Data Preparation and Preprocessing.

- Techniques for data cleaning and handling missing values.
- Dealing with noisy data and outliers.
- Data normalization and standardization.
- Feature engineering and its impact on model accuracy.
- Data labeling and annotation best practices.
- Handling imbalanced datasets with oversampling and under sampling.
- Using data augmentation for improved model training.

Unit Three: Data Storage and Management for AI.

- Comparing data lakes, data warehouses, and databases.
- Choosing the right storage solution for AI workloads.
- Managing large-scale datasets efficiently.
- Implementing version control for datasets.
- Data security and privacy in AI data management.
- Introduction to MLOps and DataOps.
- Best practices for data governance in an AI context.

Unit Four: Data Governance and Ethics in AI.

- Establishing data governance frameworks for AI.
- Defining data ownership and stewardship.
- Ethical considerations in data collection and use.
- Addressing bias in data and its effect on models.
- Ensuring data compliance with regulations like GDPR.
- Data transparency and explain ability.
- Creating a framework for responsible AI.

Unit Five: Building and Deploying AI Data Pipelines.

- Designing a scalable data pipeline for machine learning.
- Using automated tools for data transformation.
- Monitoring and managing data quality in pipelines.
- Real-world examples of AI data pipelines.
- Continuous integration and continuous delivery for data.
- Final project: designing a data strategy for a new AI product.

FAQ:

Qualifications required for registering to this course?

There are no requirements.

How long is each daily session, and what is the total number of training hours for the course?

This training course spans five days, with daily sessions ranging between 4 to 5 hours, including breaks and interactive activities, bringing the total duration to 20 - 25 training hours.

Something to think about:

Considering the rapid evolution of AI and the increasing demand for high-quality data, how can organizations build data management practices that are not only efficient but also scalable and adaptable to future technological changes and new ethical challenges?

What unique qualities does this course offer compared to other courses?

This training course is unique because it specifically addresses the critical intersection of data management and artificial intelligence. While many courses focus on just one of these topics, this course provides a comprehensive view of the entire data lifecycle for AI and ML projects. We don't just teach you how to clean data; we show you why that data needs to be clean for a model to perform well. Our content is rich with real-world case studies and problem-based learning, which means you will work with the same kinds of data challenges that professionals face every day. The course also puts a strong emphasis on the often-overlooked aspects of data management, such as data governance, ethics, and building scalable data pipelines. This holistic approach makes sure that participants leave with more than just a list of technical skills. They will have a strategic understanding of how data powers AI and how to build a robust foundation for any AI initiative.