

Production-Ready Machine Learning: Designing Scalable, Reliable, and Real-World AI Systems

18 – 22 January 2027
London

AGILE LEADERS
Training Center

AGILE LEADERS
Training Center

Production-Ready Machine Learning: Designing Scalable, Reliable, and Real-World AI Systems

Ref: 60_86603 **Date:** 18 – 22 January 2027 **Location:** London **Fees:** 5700 Euro

Course Overview:

In today's fast-paced digital landscape, deploying scalable and reliable machine learning systems is no longer optional — it is essential. Production-Ready Machine Learning: Designing Scalable, Reliable, and Real-World AI Systems is an intensive, practical training program grounded in the best practices from the authoritative book "Designing Machine Learning Systems." This course demystifies the challenges of transforming ML prototypes into robust, real-world AI systems. Participants will explore the entire lifecycle of production-ready ML — from system design and feature engineering techniques to ML model deployment, continuous training, model versioning, and monitoring.

Target Audience:

- Machine Learning Engineers
- AI System Architects
- Data Scientists
- DevOps Engineers
- Software Engineers in ML Ops
- AI/ML Product Managers
- Cloud Infrastructure Engineers

Targeted Organisational Departments:

- AI/ML Engineering
- Data Science & Advanced Analytics
- IT Operations & Infrastructure
- Digital Transformation
- Product Development & Innovation
- Quality Assurance & Monitoring
- Cloud & DevOps Teams

AGILE LEADERS
Training Center

Targeted Industries:

- Technology & SaaS
- Healthcare & Biotech
- Finance & FinTech
- E-commerce & Retail
- Telecommunications
- Manufacturing & IoT
- Automotive Self-driving Systems
- Logistics & Smart Supply Chain

Course Offerings:

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

- Design and implement scalable machine learning system architectures.
- Build production-ready ML pipelines and deploy models to cloud and edge environments.
- Apply data-centric AI principles to optimize feature engineering and data pipelines.
- Monitor, debug, and maintain ML systems using observability tools.
- Implement iterative ML development and continuous training practices.
- Manage model versioning and lifecycle with real-time deployment strategies.
- Ensure robust performance, fairness, and low-latency operation of AI systems in production.

Training Methodology:

This course integrates real-world machine learning case studies, interactive labs, and group-based projects that simulate production machine learning environments. Trainees will engage in iterative machine learning development cycles, explore debugging techniques for machine learning systems, and assess model performance using live monitoring methods. Each module blends conceptual discussions, hands-on exercises, and feedback-driven refinement of deployed artificial intelligence systems.

Course Toolbox:

- Course ebook & Slides
- Jupyter Notebooks with example ML pipelines
- Code templates for real-time ML systems
- Tools: MLflow, TensorFlow Serving, Streamlit, Airflow, Docker, Prometheus/Grafana
- Access to curated reading materials, case studies & GitHub repos
- Model evaluation checklists & deployment templates
- Monitoring dashboards for ML performance
- Troubleshooting & debugging flowcharts
- Production ML best practices cheat sheets

AGILE LEADERS
Training Center

Course Agenda:

Day 1: Foundations of Production-Ready ML Systems

- **Topic 1:** Introduction to Machine Learning Systems in Production
- **Topic 2:** Designing Reliable and Scalable ML Systems
- **Topic 3:** Differences Between Traditional Software and ML Engineering
- **Topic 4:** ML System Requirements: Reliability, Scalability, Maintainability, Adaptability
- **Topic 5:** Overview of Real-World ML Use Cases and Business Impact
- **Topic 6:** Introduction to Iterative ML Development and Deployment
- **Reflection & Review:** Assessing readiness for real-world ML system design

Day 2: Data-Centric AI and Feature Engineering

- **Topic 1:** The Critical Role of Data in ML System Performance
- **Topic 2:** Creating and Validating High-Quality Datasets for Production
- **Topic 3:** Feature Engineering Techniques and Data Preprocessing Best Practices
- **Topic 4:** Data Versioning and Validation in ML Pipelines
- **Topic 5:** Understanding Train-Serving Skew and Data Distribution Shifts
- **Topic 6:** Managing ML Data Infrastructure at Scale
- **Reflection & Review:** Data-centric challenges in scalable machine learning

Day 3: Model Development, Evaluation, and Deployment

- **Topic 1:** Building Robust ML Models for Real-World Applications
- **Topic 2:** Model Selection, Training Strategies, and Evaluation Metrics
- **Topic 3:** Deployment Strategies: Online vs Batch Prediction
- **Topic 4:** Infrastructure for ML Model Deployment and Integration
- **Topic 5:** Model Versioning Tools and Continuous Deployment Pipelines
- **Topic 6:** Debugging ML Systems and Handling Edge Cases
- **Reflection & Review:** Strengthening ML model deployment pipelines

AGILE LEADERS
Training Center

Day 4: Monitoring, Retraining, and Observability

- **Topic 1:** ML Model Monitoring in Production Environments
- **Topic 2:** Detecting and Responding to Concept Drift and Data Shifts
- **Topic 3:** Continual Learning and Retraining Cycles
- **Topic 4:** Observability Tools and Logging for ML Systems
- **Topic 5:** ML Reliability Engineering: Failures, Alerts, and Mitigations
- **Topic 6:** Real-Time ML Pipelines and Streaming Data Considerations
- **Reflection & Review:** ML lifecycle management and observability

Day 5: Scaling, Fairness, and Business Alignment

- **Topic 1:** Scaling AI Systems: From Prototypes to Global Infrastructure
- **Topic 2:** Ethical AI: Fairness, Bias, and Interpretability in Production
- **Topic 3:** Performance Optimization in Low-Latency AI Systems
- **Topic 4:** Business Metrics Alignment and Post-Deployment Analytics
- **Topic 5:** Case Studies of Real-World ML System Failures and Recoveries
- **Topic 6:** Best Practices in Production ML: End-to-End Workflows
- **Reflection & Review:** Final synthesis of scalable, production-ready ML systems

FAQ:

What specific qualifications or prerequisites are needed for participants before enrolling in the course?

Basic understanding of machine learning concepts and experience with Python programming is recommended. Prior experience with ML model development or deployment is helpful but not mandatory.

How long is each day's session, and is there a total number of hours required for the entire course?

Each day's session is generally structured to last around 4–5 hours, with breaks and interactive activities included. The total course duration spans one week, approximately 20–25 hours of instruction.

What’s the difference between deploying a model and making it production-ready?

Deploying a model means making it technically accessible. But making it production-ready involves designing scalable, low-latency pipelines, building monitoring and alerting systems, ensuring fairness, and preparing for continuous retraining, as emphasised in this course.

AGILE LEADERS
Training Center

How This Course is Different from Other Production ML Courses:

Unlike general-purpose ML bootcamps, Production-Ready Machine Learning is structured around real-world requirements for reliability, scalability, and adaptability, drawn directly from the acclaimed reference “Designing Machine Learning Systems.” It encompasses not only model development but also critical infrastructure design, continuous deployment, monitoring, and feedback loops. The curriculum is rich in use cases and practical challenges faced by companies like Netflix, Uber, and Google. Trainees gain hands-on experience with ML observability tools, iterative workflows, and scalable ML model deployment pipelines. Additionally, the course includes production ML best practices for debugging, data versioning, fairness checks, and retraining strategies — ensuring you are equipped for real-world success, not just academic exercises.