



Industrial Machine Learning for Enterprises

Deliverable D4.3

Training and Education Material: Synopses



This document by the IML4E project (IML4E – 20219) is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0).

Project title:	IML4E
Project number:	20219
Call identifier:	ITEA AI 2020
Challenge:	Safety & Security

Work package:	WP4
Deliverable number:	D4.3
Nature of deliverable:	Report
Dissemination level:	PU
Internal version number:	1.0
Contractual delivery date:	2024-05-31
Actual delivery date:	2024-07-17
Responsible partner:	SILO AI

Contributors

Editor(s)	Harry Souris (Silo AI)
Contributor(s)	Jürgen Großmann, Dorian Knoblauch (Fraunhofer FOKUS), Johan Himberg (Reaktor), Mikko Raatikainen (Helsinki University)
Quality assesor(s)	Ihasalo Heikki (Granlund), Abhishek Shresta (Fraunhofer FOKUS)

Version history

Version	Date	Description
1.0	24-07-17	Version for publication

Abstract

This document introduces the collection of training materials created by various IML4E participants. The material is about ML and MLOps practices and regulation from various viewpoints. The document describes each separate training material by its intended target audience, summary of contents, expected learning outcome, and gives link or other reference to the material itself.

Keywords

MLOps, MLOps framework Governance, Audit, Monitoring, Deployment, ML/AI Architecture, MLOPs technology stack

Executive Summary

This document introduces a syllabus summarizing the efforts and the deliverables in the work package four (WP4) of the IML4E project. WP4 is primary focused on the methodologies and technologies utilised in the operations of ML systems and in this document a reader will be in position to find related use cases, educational material and references to an open-source technology stack for executing AI and ML services in production.

Table of contents

1 INTRODUCTION 6

1.1 ROLE OF THIS DOCUMENT 6

1.2 INTENDED AUDIENCE..... 6

1.3 APPLICABLE DOCUMENTS..... 6

2 IML4E EDUCATION MATERIAL 7

2.1 IML4E MLOPS OVERVIEW: TEAM TRAINING CASES 7

2.1.1 Short summary 7

2.1.2 Target audience and prerequisites 7

2.1.3 Learning outcomes 7

2.1.4 Content 7

2.1.5 Learning materials 7

2.2 IML4E MLOPS OVERVIEW: DESIGNER’S VIEWPOINT TO LLM SOLUTIONS..... 8

2.2.1 Short summary 8

2.2.2 Target audience and prerequisites 8

2.2.3 Learning outcomes 8

2.2.4 Content 8

2.2.5 Learning materials 8

2.3 IML4E’S OSS PLATFORM 9

2.3.1 Short Summary 9

2.3.2 Target Audience..... 9

2.3.3 Prerequisites..... 9

2.3.4 Learning Outcomes..... 9

2.3.5 Learning Material..... 9

2.4 IML4E’S OSS PLATFORMS REFERENCE USE CASES 10

2.4.1 Short summary 10

2.4.2 Prerequisites..... 10

2.4.3 Learning outcomes 10

2.4.4 Content 10

2.4.5 Additional info 10

2.5 IMPLEMENTING AND MEASURING COMPLIANCE WITH THE EU AI ACT IN AI SYSTEMS..... 11

2.5.1 Short summary 11

2.5.2 Prerequisites..... 11

2.5.3 Learning outcomes 11

2.5.4 Content 11

2.5.5 Learning materials 11

2.5.6 Literature 11

2.5.7 Additional info 12

2.6 ONE-DAY TRAINING SYLLABUS: INTRODUCTION TO MLOPS WITH OSS MLOPS PLATFORM 12

2.6.1 Short summary 12

2.6.2 Prerequisites..... 12

2.6.3 Learning outcomes 12

2.6.4 Content 12

2.6.5 Learning materials 12

2.6.6 Literature 13

2.6.7 Additional info 13

3 SUMMARY..... 14

1 Introduction

1.1 Role of this Document

The document facilitates as a syllabus on the introduced deliverables by IML4E partners working on work package four (WP4). The deliverables focus on MLOps technologies and practices aiming to enable Enterprises to integrated ML services to their operations. In summary a collection of use cases is presented and an IML4E developed open source MLOps platform is referenced and introduced together with several educational materials developed by partners.

1.2 Intended Audience

The intended audience of the present document is composed primarily of the IML4E consortium for the purpose of capturing the baseline of the project and the material introduced in work package four.

1.3 Applicable Documents

Reference	Referred document
[FPP]	IML4E – Full Project Proposal 20219
[PCA]	IML4E Project Consortium Agreement

Table 1 - Contractual documents

2 IML4E Education material

2.1 IML4E MLOps overview: Team training cases

2.1.1 Short summary

Reality based, integrative case exercises for teamwork in the practice of designing, deploying, and maintaining and ML solutions.

2.1.2 Target audience and prerequisites

A design and delivery team (or trainees / students) of machine learning specialists, developers, and project managers / owners. Target audience are professionals. They are expected to have at least basic knowledge of their respective fields. See for example: “Engineering of Machine Learning Systems” (University of Helsinki course).

2.1.3 Learning outcomes

Solution patterns for Ops situations on a team level

2.1.4 Content

List of topics

1. Training cases themselves
2. Commentary

2.1.5 Learning materials

- <https://github.com/reaktor/iml4e/>
- Relation to other learning material: see “Engineering of Machine Learning Systems” (University of Helsinki course)

2.2 IML4E MLOps overview: Designer's viewpoint to LLM solutions

2.2.1 Short summary

The material contains observations and guidelines about designers work in context of LLM solutions

2.2.2 Target audience and prerequisites

Service and Solutions Designers; Project/product owners

2.2.3 Learning outcomes

Good service and interaction design patterns when creating AI solutions: especially LLM solutions

2.2.4 Content

List of topics

- Design principles for natural language user interfaces
- Designing disruptive GenAI services

2.2.5 Learning materials

- <https://github.com/reaktor/iml4e/>

2.3 IML4E's OSS platform

2.3.1 Short Summary

An end-to-end machine learning platform enabling practitioners to:

1. Collect and analyze data
2. Process data and train ML models
3. Record experiments results and machine learning metadata
4. Register, deploy and monitor the operations of ML models

2.3.2 Target Audience

Engineers, scientists, solution designers aiming to develop and operate AI solutions at scale

2.3.3 Prerequisites

A solid background in software development, machine learning and data engineering is needed for understanding the capabilities of the platform

2.3.4 Learning Outcomes

Users of the platform would be in position to learn how to:

- develop end to end machine learning pipelines
- structure them and utilize the right tools
- monitor them and track their execution

2.3.5 Learning Material

Github link to the open-source OSS MLOps platform:

- [GitHub - OSS-MLOPS-PLATFORM/oss-mlops-platform: OSS MLOps Platform](https://github.com/itea4/oss-mlops-platform)

Practical tutorials:

- https://gitlab.fokus.fraunhofer.de/iml4e/iml4e_oss_exp_platform/-/tree/main/tutorials
- <https://itea4.org/project/workpackage/document/download/8285/IML4E-D4.1-V1.0-Requirements%20for%20the%20IML4E%20online%20experimentation%20and%20training%20platform.pdf>
- <https://itea4.org/project/workpackage/document/download/8607/IML4E-D4.2-Initial%20MLOps%20methodology%20and%20the%20architecture%20of%20the%20IML4E%20framework.pdf>

2.4 IML4E's OSS platforms reference use cases

Engineering of Machine Learning Systems (5 cr University of Helsinki course) Permanent link:

<https://studies.helsinki.fi/courses/course-unit/otm-995b4379-d234-4c64-bcf7-9981fc39ddfd/DATA11008>

2.4.1 Short summary

Learn the state-of-the-art challenges, practices and tools for engineering of ML systems.

2.4.2 Prerequisites

Basics of machine learning, basic programming skills

Related courses: <https://big-data-platforms-21.mooc.fi/> and [https://studies.helsinki.fi/courses/cu/hy-CU-118843631-2021-08-01/DATA11003/Distributed Data Infrastructures](https://studies.helsinki.fi/courses/cu/hy-CU-118843631-2021-08-01/DATA11003/Distributed%20Data%20Infrastructures)

2.4.3 Learning outcomes

- Compare the difference between developing ML systems with classical software.
- Outline ML workflow and its relationship to real-world use cases.
- Explain the anatomy of MLOps pipeline and the role of its elements.
- Apply MLOps pipeline for an ML model-based system.

2.4.4 Content

- Why study MLOps?
- MLOps pipelines, iterative workflow
- Data management (validation, feature stores, ...)
- (Re)Training
- Testing and validation
- Deploying
- Serving
- Monitoring

2.4.5 Additional info

Course Completion Methods

- Lectures
- Prior readings (incl. videos) + discussions
- Weekly exercises / Project
- Experiment with ML pipeline
- Learning diary: what is different in ML from classical
- No exam

2.5 Implementing and Measuring Compliance with the EU AI Act in AI Systems

2.5.1 Short summary

A one-day intensive course on implementing and measuring compliance with the EU AI Act in AI systems, focusing on practical strategies, automation techniques, and continuous improvement in adhering to regulatory and ethical standards.

2.5.2 Prerequisites

- Basic understanding of AI and machine learning principles.
- Awareness of general regulatory and compliance issues in technology.

2.5.3 Learning outcomes

- Understand the essentials of the EU AI Act and its implications for AI systems.
- Learn to operationalize and measure compliance with the Act technically.
- Recognize the importance of automated compliance monitoring in dynamic AI environments.
- Gain knowledge on continuous improvement strategies for AI compliance.
- Explore documentation and reporting roles in automated compliance systems.

2.5.4 Content

- Understanding the EU AI Act (30 minutes): Overview of the Act's structure, technical compliance requirements, and the role of automated systems.
- Technical Foundations of AI Compliance (30 minutes): Key concepts in AI trustworthiness including data governance, documentation, transparency, human oversight, and security.
- Automation Opportunities for Risk Management in AI (1 hour): Techniques for risk identification, assessment, and management in AI systems, focusing on automated solutions.
- Operationalizing Compliance through Automation (1 hour): Steps to translate regulatory requirements into automated checks and balances.
- Continuous Compliance Monitoring (1 hour): Using automated tools for ongoing monitoring and adjustments in AI systems.
- Documentation and Reporting in Automated AI Systems (1 hour): Best practices for documenting compliance processes and reporting.

2.5.5 Learning materials

- Comprehensive course materials on the EU AI Act.
- Case studies on automated compliance and monitoring.
- Access to simulations and automated tool demos.

2.5.6 Literature

Selection of case studies on various AI applications and their compliance needs under the EU AI Act, including object detection systems, fine-tuned LLMs for corporate chatbots, and IoT sensor-based prediction systems.

2.5.7 Additional info

- Target Audience: AI Developers, Compliance Officers, Risk Management Professionals, AI Policy Advisors, IT and AI Consultants.
- Evaluation and Feedback: Interactive quiz and participant feedback forms for course enhancement.
- Conclusion: Recap of key concepts and encouragement for application in respective fields.
- Resources: [ML Training / EU AI Act Course · GitLab \(fraunhofer.de\)](#)

2.6 One-Day Training Syllabus: Introduction to MLOps with OSS MLOps Platform

2.6.1 Short summary

A 5-hour training course aimed at Machine Learning Engineers and Data Scientists, focusing on the practical understanding of MLOps processes using the OSS MLOps platform, covering its application in real-world scenarios from experiments to pipelines.

2.6.2 Prerequisites

- Basic understanding of Machine Learning concepts.
- Familiarity with cloud computing and containerization, specifically Kubernetes (K8s).

2.6.3 Learning outcomes

- Comprehend the principles and processes of MLOps.
- Learn prerequisites for implementing MLOps, including Kubernetes (K8s).
- Gain hands-on experience in transitioning from ML experiments to scalable pipelines.
- Understand the OSS MLOps platform and its real-world applications.

2.6.4 Content

- Introduction to MLOps (1 hour): Understanding MLOps, OSS MLOps platform, features like automation, versioning, monitoring, scalability, and development environment isolation.
- MLOps Process and Principles (1 hour): Lifecycle of ML models in MLOps, key components like data preparation, model training, deployment, monitoring, and maintenance, service-based approach.
- Prerequisites for MLOps (1 hour): Introduction to Kubernetes (K8s) and its role in MLOps, other necessary technologies and tools.
- From Experiments to Pipelines (1 hour): Transitioning from ML experiments to scalable pipelines, building and deploying models using OSS MLOps platform, practical examples and case studies.
- Hands-On Workshop (1 hour): Setting up a basic MLOps pipeline with the OSS MLOps platform, deploying and managing a sample ML model.

2.6.5 Learning materials

- Access to the OSS MLOps platform.
- Documentation and case studies for reference.
- Tools for hands-on exercises including cloud access, IDEs, and relevant software.

2.6.6 Literature

- Documentation on MLOps processes and principles, case studies illustrating the application of the OSS MLOps platform in real-world scenarios.

2.6.7 Additional info

- Target Audience: Machine Learning Engineers, Data Scientists, and those interested in MLOps.
- Evaluation and Feedback: Interactive Q&A sessions, practical exercise review, and feedback forms for course improvement.
- Conclusion: Recap of key learnings, future trends in MLOps, and additional resources for continued learning.
- Resources: ML Training / Mlops Course · GitLab (fraunhofer.de)

3 Summary

Training in MLOps is essential for integrating machine learning systems into enterprise environments effectively. It ensures technical competency across data collection, model training, deployment, and monitoring, fostering collaboration among diverse teams. This document outlines the training offering developed in the IML4E project tackling MLOps in general, the IML4E OSS platform, University education and the governance challenges related to the European AI Act.