

MLOps Engineer

Training structure



Context

How to become an AI Engineer

In the field of AI, many companies are experiencing **major difficulties** in **deploying and operating** AI models in production in an industrial manner. This is mainly due to a lack of **global maturity** on the subject and a still too **important siloing** of all the teams involved in building this new value centered on AI.

The objective of this course is therefore to reinforce the knowledge of tomorrow's engineers around the issues of **operationalization of AI** in production, with regard to the needs, constraints and demanding processes of today's industry.

The ambition of this course is to provide students with:



An awareness of **MLOps** and **DataOps** methodologies so that they can apply them on simple examples



A better knowledge of **cloud native** environments and technologies



A good visibility on the complete **life cycle of a Machine Learning** application, with CI/CD, technical monitoring and business monitoring



Some notions about sensitivity, security, encryption and data isolation

Context



• What will this course not be :

- Ai deep dive
- Optimisation
- Scientific course



• What will this course be :

- Awareness
- Architecture
- Good practices
- Industrialization

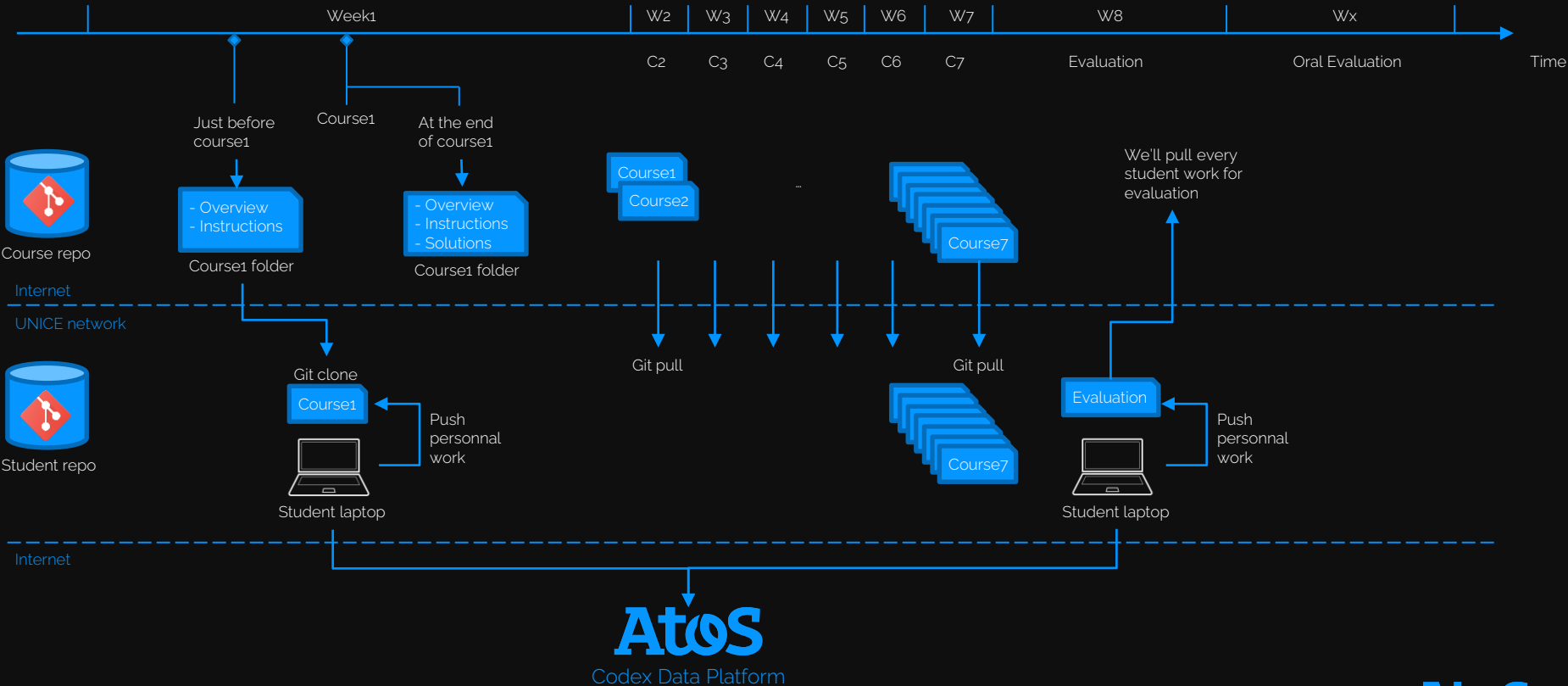
Planning

Ambitions

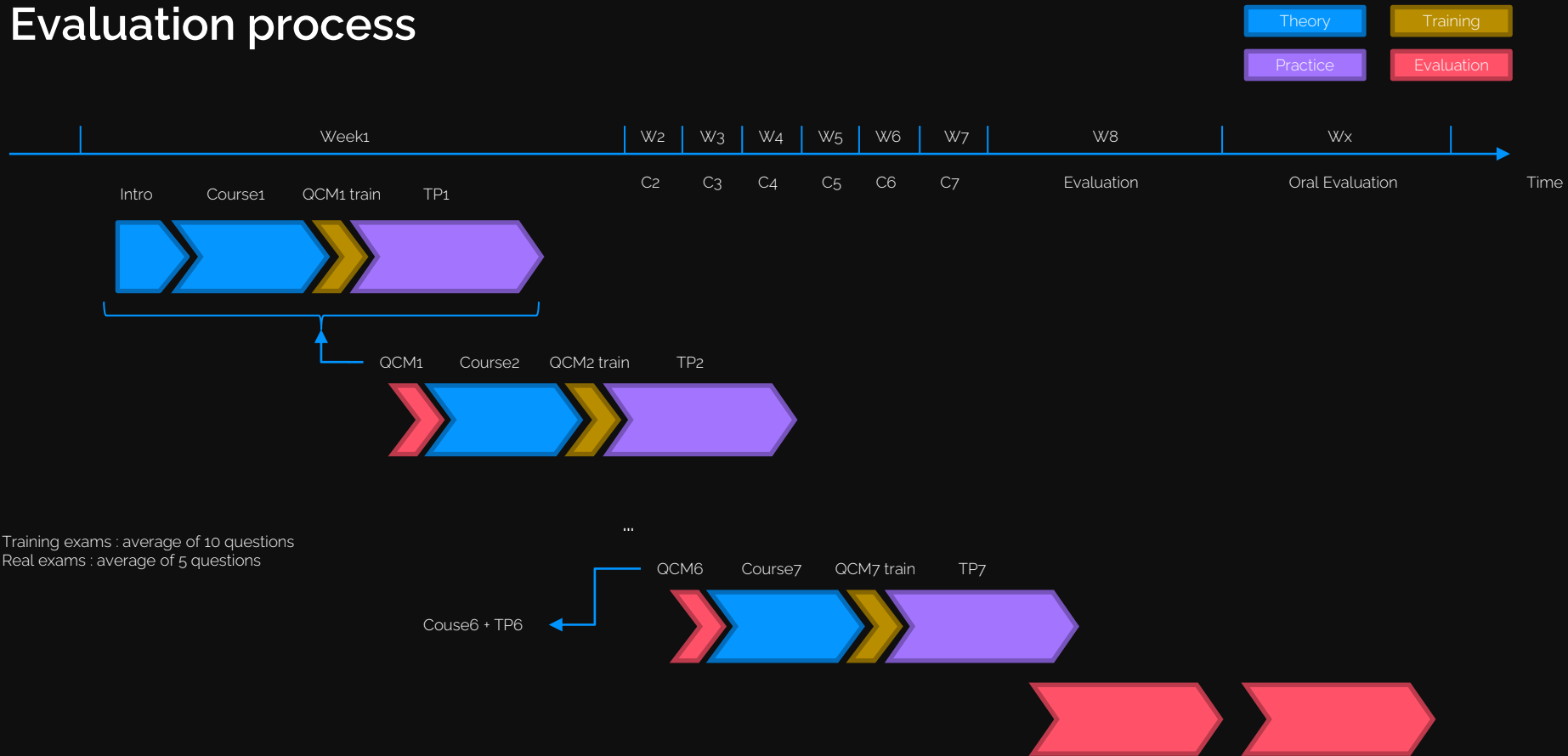
Weeks	Date	Title	Theory	Practice
1	06/12 13h30 – 15h 15h15 – 17h15	Data Platform Foundation	Introduction MLOps, DataOps, Cloud, Dataplatform/lakehouse/mesh	Interaction with object store Quick data analyse with python Communication with datawarehouse and data viz Use events
2	13/12 13h30 – 15h 15h15 – 17h45	Data Pipeline	Trigger, orchestration ETL, ELT, DAG Templating, Heterogeneous processing (tf, spark, ..) Tracking experiment AutoML	Kubeflow discovering KF component creation Light model training
3	20/12 13h30 – 15h 15h15 – 17h45	Data Pipeline 2	Streaming processing Serving complex graph : dataprep, inference, monitoring, explainability	Creation of an simple inference service And a complexe inference service Chain it inside event pipeline
4	10/01 13h30 – 15h 15h15 – 17h45	Data Governance ML Governance	Data gov : catalog, lineage ML gov : model registry, feature store	Interact with data catalog Use feature store in build mode and in inference mode
5	17/01 13h30 – 15h 15h15 – 17h45	CICD	CICD and Gitops mindset Standardization in ML (model formats)	CICD (git, gitflow, gitlab pipelines/jenkins?, argo/flux) Progressive deployment : Canary, Blue Green
6	24/01 13h30 – 15h 15h15 – 17h45	ML Monitoring	Model monitoring : data drift, model drift, outliers, .. Global monitoring : data (lineage, tracing), infrastructure	Evidently Prometheus, loki, fluentd, grafana
7	31/01 13h30 – 15h 15h15 – 17h45	Wrap up & Opening	Opening on api management Opening on fairness, biases	TBD
8	07/02 13h30 – 15h	Evaluation	Test – Build things	Test – Convince us

Training process

Gitflow



Evaluation process



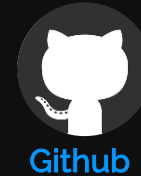
Course Asset Structuration

Between Moodle and Github



Moodle

- Slides as pdf
- 1 multiple-choice exam at the end of each theoretical course
- Final exam instructions



Github

For each practical session

- What should be done (notebooks with missing code)
- Interesting links, documentation
- Answers published after classe

<https://github.com/A709509/aiengineerPolytech>

Questions

Thank you!

Atos, the Atos logo, Atos|Syntel are registered trademarks of the Atos group.
June 2021. © 2021 Atos. Confidential information owned by Atos, to be used by the recipient only. This document, or any part of it, may not be reproduced, copied, circulated and/or distributed nor quoted without prior written approval from Atos.