COURSE SYLLABUS

AI520 • Responsible ML Systems

School of Computational Arts & Sciences • Graduate • 6 ECTS

Overview

Design and operate responsible ML services: evaluation protocols, bias and slice checks, monitoring/alerts, rollback strategies, and incident-style postmortems. Emphasis on documentation, governance, and shipping models that remain reliable under drift.

LOGISTICS

Credits: 6 ECTS
Level: Graduate
School: School of Computational Arts & Sciences
Prerequisites: Programming experience, Basic probability and statistics
Tags: ai, ethics, mlops
Meeting time: Weekly seminar + applied MLOps lab
Instruction mode: Case-based: incidents, audits, and rollback drills

LEARNING OUTCOMES

You will be able to:

- Design evaluation protocols for deployed ML systems
- Implement monitoring and incident response playbooks
- Document model behavior with audit-friendly artifacts
- Design an evaluation protocol that includes bias checks and safety constraints
- Implement monitoring signals and rollback criteria for ML services
- Write an audit-friendly model card and incident postmortem

ASSESSMENT

Components

- Evaluation memo: 25%
- System labs: 35%
- Incident postmortem simulation: 15%
- Final project (deployment plan + review): 25%

Assessment rewards operational rigor: clear assumptions, measurable constraints, and documentation that an external reviewer could follow. Projects are evaluated on system design, monitoring plan, and the quality of your risk analysis.

WEEKLY PLAN

Schedule

Week 1: Failure modes that matter

- What counts as harm
- Stakeholders
- Boundaries and constraints

Week 2: Baselines and evaluation

- Datasets and drift
- Metrics vs. values
- Reproducible protocols

Week 3: Monitoring and logging

- Telemetry
- Alert fatigue
- SLOs for ML

Week 4: Governance

- Documentation
- Review workflows
- Change control

Week 5: Incident response

- Runbooks
- Rollback
- Retirement plans

Extended outline

- Threat models for ML: failure modes and what ’safe’ means
- Dataset governance: provenance, consent, and minimization
- Bias evaluation: slices, parity, and trade-offs
- Monitoring: drift, data quality, and performance regression
- Rollback plans: thresholds, feature flags, and incident drills
- Capstone: ship a small ML service with an evaluation + governance pack

POLICIES & RESOURCES

- Ethics: document any sensitive features and mitigation decisions.
- Operational honesty: report failures; postmortems are graded positively when rigorous.
- Security: do not use real credentials or production keys in assignments.

Suggested resources
- Model card template (assumptions, data, metrics, limitations)
- Incident postmortem template (timeline, root cause, actions)
- Monitoring checklist (data quality, drift, alerting, rollback)