Introduction to the course

2023-2024

Explainable and Trustworthy AI
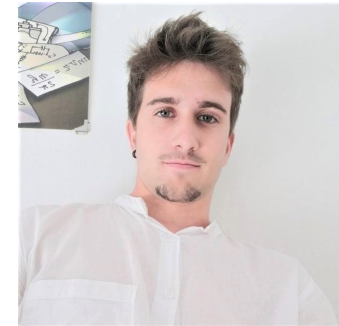
Eliana Pastor

Teaching staff

Eliana Pastor

Elena Baralis

Gabriele Ciravegna

Salvatore Greco

Eleonora Poeta

Course topics in a glance

- Trustworthy AI: Notions and Requirements
- Taxonomy of Explainable AI (XAI)
- Explainable by design models
- Post-modelling approaches for XAI
- Global XAI approach
- Local XAI approaches
- Concept-based XAI
- Model robustness
- Fairness in AI

Material

Announcements on the official «teaching portal» private page https://didattica.polito.it/ or through institutional email

Explainable and trustworthy AI

01HFOOV, 01HFOSM

A.A. 2023/24

Slides, texts of the practices, and all other materials are available on the public page of the course: https://dbdmg.polito.it/dbdmg_web/2024/explainable-and-trustworthy-ai-2023-2024/

📁 Courses

Explainable and trustworthy AI (2023/2024)

General Information

SSD: ING-INF/05

CFU: 6

Lecturer: Eliana Pastor

Teaching Staff: Elena Baralis, Gabriele Ciravegna, Salvatore Greco, Eleonora Poeta

Course structure

The course includes

- Lectures

- Hands-on and exercises

- Laboratories
    - Experimental activities
    - Practical analysis and evaluations of the methods covered in the course

- Lab practises will start from the third week

Schedule

- **Monday**

  8:30-10:00

  classroom 12I

- **Thursday**

  8:30-10:00

  classroom 7D

- **Friday**

  8:30-10:00

  classroom 11I

No fixed slot distinction between Lecture and Lab Practise

Exam

- Written exam
  - Main definitions and concepts of Explainable and Trustworthy AI
  - Explanation techniques and their main characteristics
  - Main libraries implementing explanation methods

- Group project (3-4 students)
  - Implement and evaluate a complete data science pipeline and its explanation
  - Design and evaluate explanation methods
  - Presentation in written form

The detailed exam rules will be soon published in the web page of the course