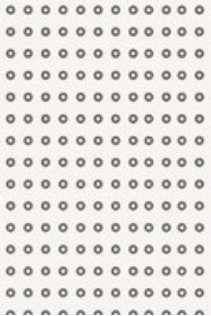
LINKSFOUNDATION.COM

FONDAZIONE links
PASSION FOR INNOVATION

Applied Data Science Project

L1 - Introduction

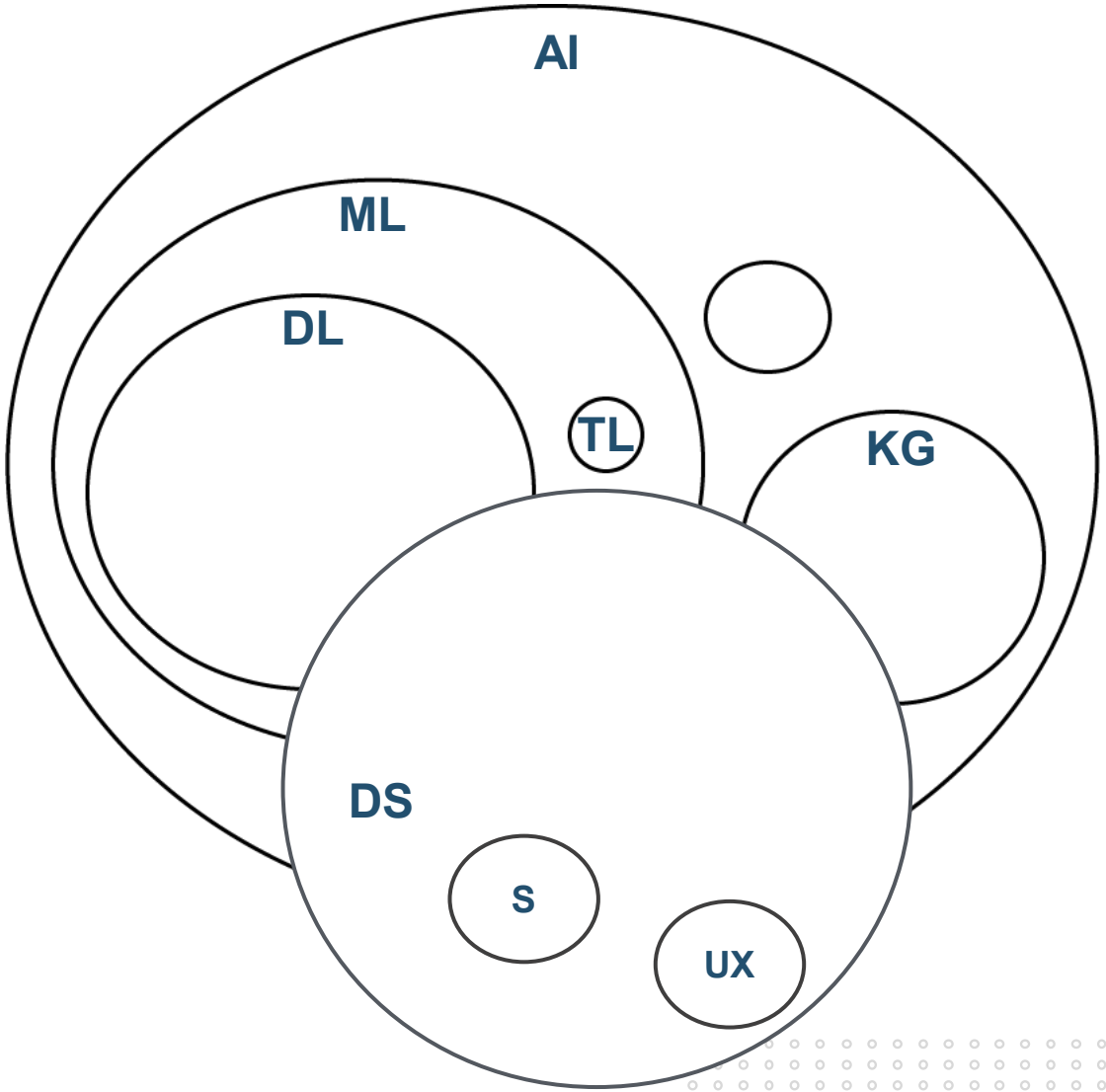
ellis
European Laboratory for Learning and Intelligent Systems



Course objective

Building an artificial intelligence solution with a data science project

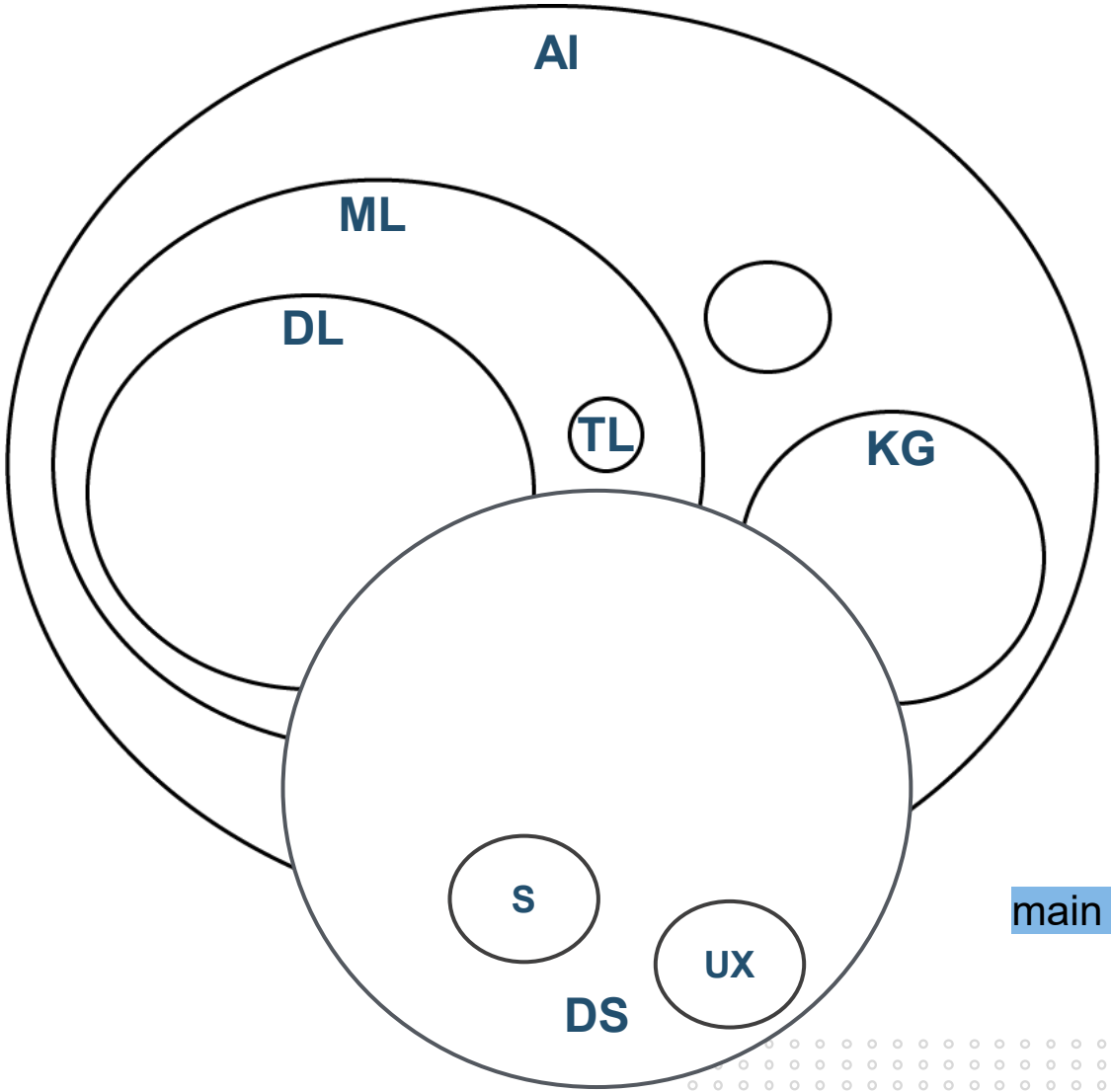
Interplay



AI: **A**rtificial **I**ntelligence
ML: **M**achine **L**earning
DL: **D**eep **L**earning
TL: **T**ransfer **L**earning
KG: **K**nowledge **G**raph
S: **S**tatistics
UX: **U**ser e**X**perience
DS: **D**ata **S**cience

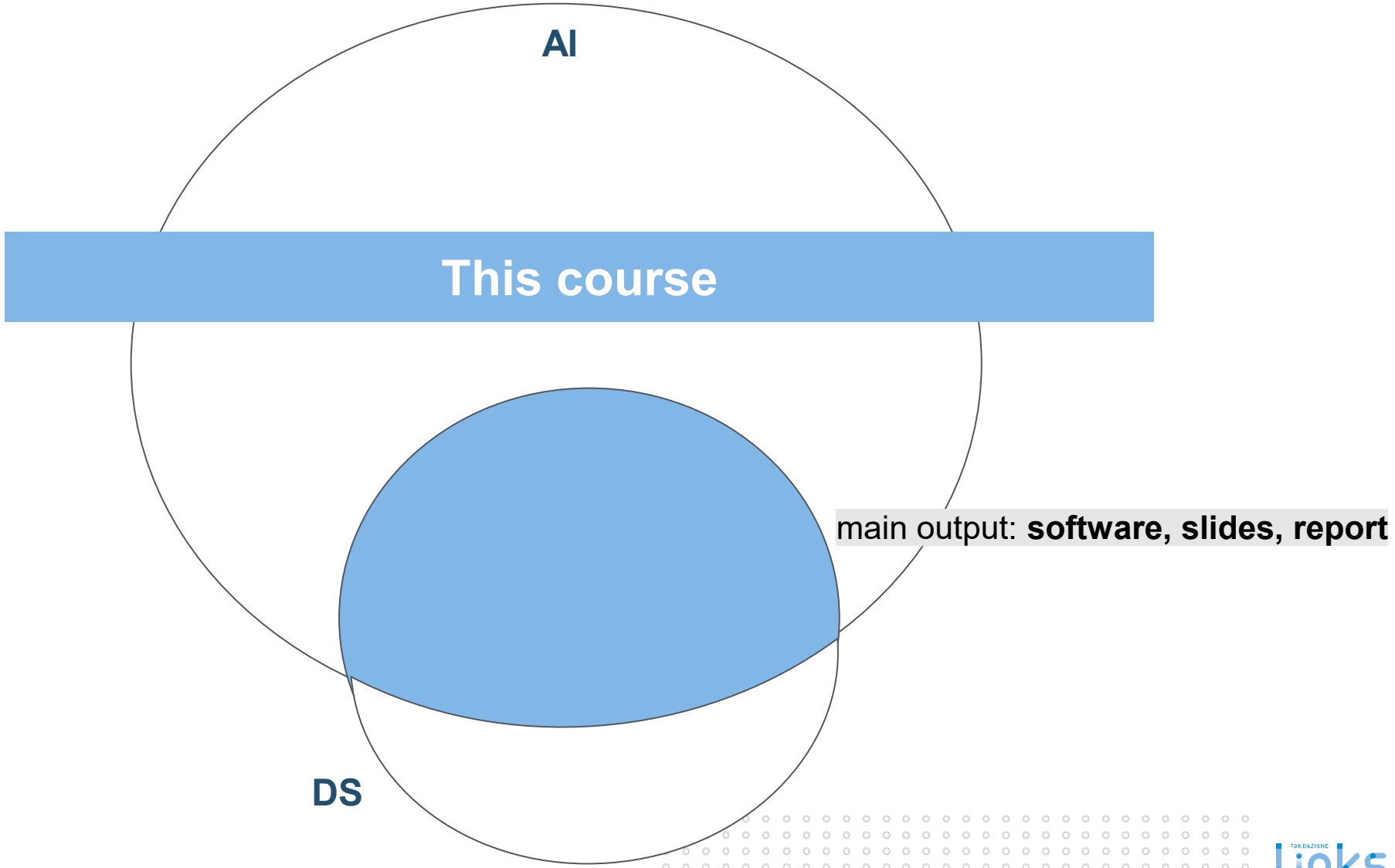
Interplay

main output: **software**



main output: **visuals, slides**

Applied Data Science Project



main output: **software, slides, report**

Applied Data Science Project 101

Your goal is to develop and document an artificial intelligence solution by

1. Starting from a real-world challenge and define project objectives
2. Design a project
   - that is centered to people
   - plan the work in advance, manage the activity & monitor progress
   - select the right tools (foundation models such as a LLM)
   - customize an artificial intelligence to answer the project challenges and generate value
3. Quantify the benefits and impact of the project

R&I pathway

Your goal is to develop and document an artificial intelligence solution by

1. Starting from a real-world ch lenge and define project objectives
2. Design a project
   - that is centered to peop
   - plan the work in advanc manage the activity & monitor progress
   - select the right tools (fo dation models such as a LLM)
   - customize an artificia gence to answer the project challenges and generate value
3. Quantify th research and innovation pathway

Define project objectives

1

Why are you conducting the project?

Where it can be used?

Align project objectives with the SDGs

Design a project

2

Who are the users?

How a user will utilize the solution and for which task?

How the solution will look like?

Plan the work

“Divide and rule”

Project activities are usually grouped in:
- macro: Work Package(s)
- micro: Task(s)

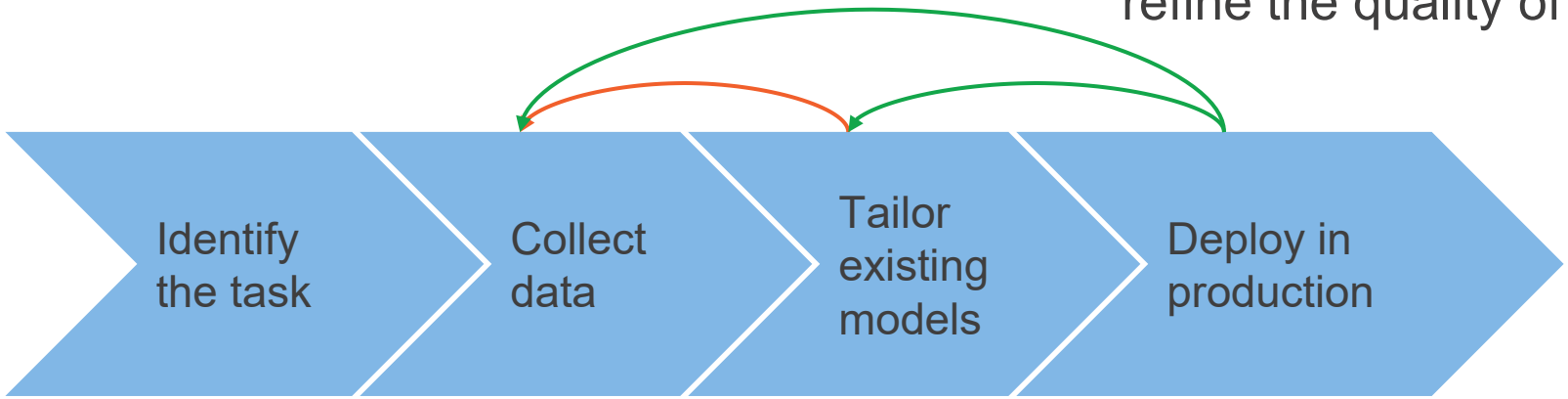
with a specific due date of completion

Projects have Milestones, aka dates by which it is expected a release

A project is under Risks of actions (for instance acquisition of data or recruitment of testers), the sooner the risks are identified, the clearer it is to address them

Design the solution

iterative processes meant to refine the quality of the solution



Impact

A project generates value meaning the quantifiable benefit that the project impacts to users

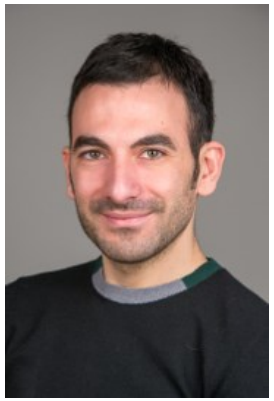
The quantification of the project impact to its value is studied utilizing the concept of Key Performance Indicator(s)

Value can be framed according to economy, society, environment

Recall that a successful project will have the chance to apply it more on other challenges **iff**:

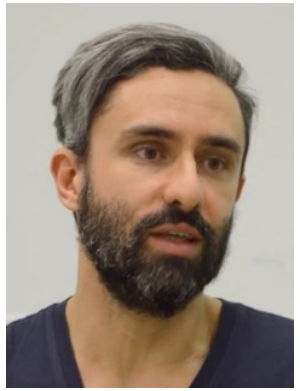
value - cost of fulfilment > 0

Team

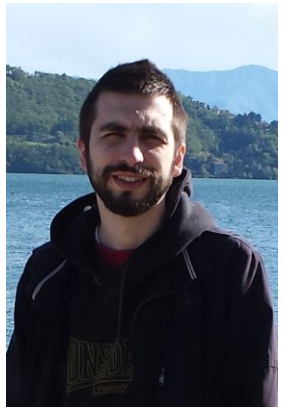


Giuseppe Rizzo
Prof, 37.5 ore

Antonella Frisiello
Prof, 9 ore



Giuseppe Tipaldo
Prof, 6h
TA, 12 ore



Alessandro Fiori
TA, 19.5 ore

Edoardo Arnaudo
TA, 21 ore

Bartolomeo Vacchetti
TA, 15 ore

Logistics

Topics

Lectures (30 hours)

- Building an artificial intelligence solution with a data science approach (1.5h)
- Data Science project pillars (1.5h)
  - Design: People and artificial intelligence solutions
  - Development: foundation models and domain-adaptation
  - Management: GANTT e work breakdown structure
  - Communication: paper, deliverable and slides
- Model and Data-centric projects (1.5h)
- Foundation models (1.5)
- Artificial intelligence ethics (1.5h)
- Impact of a project and SGDs (1.5h)
- Data Science project tools (18h)
  - Project design tools
    - Stakeholder maps and user personas
    - User journey
  - Project development tools
    - Functional requirements: From a user research to solution definition
    - Existing foundation models
    - Domain adaptation and downstream tasks
    - Version control and testing
  - Project management tools
    - GANTT
    - Work breakdown structure, work packages and tasks, milestones
  - Project communication tools
    - Project communication
    - Presentation
    - Paper
    - Deliverable
- Success stories of past projects (1.5h)
- Project proposals (1.5h)

Laboratory activities (90h)

- Generation of the solution

Lectures until Oct 9th

10/2/23 10:00	10/2/23 11:30	Aula ACSLab	1.5	1.5		Rizzo	L	L- Introduction
10/2/23 13:00	10/2/23 14:30	Aula 17	1.5	1.5		Rizzo	L	L- Project pillars
~~10/3/23 10:00~~	~~10/3/23 13:00~~	~~Aula ACSLab~~	~~3~~	~~3~~				
10/4/23 10:00	10/4/23 13:00	Aula ACSLab	3	3		Rizzo	L	L - Models & data-centric data science projects; L - Foundation models; L - Model adaptation and tuning
~~10/5/23 11:30~~	~~10/5/23 13:00~~	~~Aula ACSLab~~	~~1.5~~	~~1.5~~				
~~10/5/23 14:30~~	~~10/5/23 16:00~~	~~Aula 10I~~	~~1.5~~	~~1.5~~				
10/9/23 10:00	10/9/23 11:30	Aula ACSLab	1.5	1.5		Rizzo	L	L - SDGs
10/9/23 13:00	10/9/23 14:30	Aula 17	1.5	1.5		Rizzo	L	L - Functional requirements and diagrams

Recommended Background

Master technologies that are linked to AI in particular deep learning and machine learning

Formal prerequisites for this course are both software engineering and programming (if you have taken one at another university, this is fine)

We strongly recommend that students have experience with Python, have a background in probability and statistics, and linear algebra

If you don’t have background in these areas, please compensate in teaming up with mates who master those fields and ask for references to get up to speed (at least understand the terminology and what we are talking about)

General rule of thumb: If the project seems hard, but you have ideas about how to proceed, you probably have the right level of background; if the project seems hard and you have no idea how to proceed, this may be the wrong course (I bet this won’t happen)

Course material

All course material (slides, lecture videos, project descriptions) is available on the *Portale della Didattica* of *Politecnico di Torino* course webpage:

https://didattica.polito.it & https://adsp-polito.github.io

Slides & class videos - if recorded - will be uploaded after the lecture (best effort, up to 1 day)

Additional general reading materials

- Machine Learning Yearning, by Andrew Ng
- Data Science from Scratch, Joel Grus
- Harvard Business Review Project Management Handbook: How to Launch, Lead, and Sponsor Successful Projects, by Antonio Nieto-Rodriguez
- Oxford Guide to Effective Writing and Speaking: How to Communicate Clearly, by John Seely
- The Design of Everyday Things: Revised and Expanded Edition, by Donald Norman
- Noessel C. Designing Agentive Technology. AI That Works for People. Rosenfeld, 2013

Discussions and reporting

Slack group adsp-polito.slack.com

Please all join the group:
https://join.slack.com/t/adsp-polito/shared_invite/zt-249in44aa-mYNU6Rs6ej_wtHd044hx8g
(link expires 11/02/2023)

Walkthrough the workspace:

- #general channel to make questions about the course
- #feedback channel to give us tips and immediate feedback about lectures and projects
- 1 channel for each project managed by each team
  - make project specific questions to the TAs (optional according to needs)
  - provide a bi-weekly report about the project status (this is mandatory and contributes to the final grade)

Laboratories for projects

A major part of the course is to conduct a real-world project, this accounts about ¾ of the whole course

There will be up to 8 projects: you will be asked to express your interest as team.
1 team will be assigned to 1 project.
We will organize 8 teams with 3 aprox

The project will be conducted under the supervision of the whole teaching team and 1 representative for each project coming for the involved companies or research labs

Final report will be a latex document, slides, and a Colab Notebook

Project delivery

Projects are due at the end of the semester

The exact day will be communicated in the first half of the course

The delivery includes 3 outputs for each project:

- slide deck
- document
- software

3 intermediate checks, 15-minute presentation for each group:

- 08/11/2023
- 06/12/2023
- 17/01/2024

Project presentation

The best way to learn a subject is to teach it

Students will prepare a presentation to be given in front of all students of the course and (possibly) the external tutors who have followed the execution of the project

Grading

Team project (80%) + compulsory individual oral exam (20%)

The maximum grade for the team project is 32
The maximum grade for the oral part is 32

The final grade is given by weighted average of the two parts:

0.8 * grade team project + 0.2 * grade individual oral part

The exam is passed if the grade of the team project is greater than or equal to 18 and the grade of the oral part is greater than or equal to 18

What will be assessed?

Team project

- team project assessment is based on the performance and accuracy of the proposed solution, in terms of standard quality measures (e.g., prediction, accuracy) and completeness (i.e., in depth analysis of each phase of the designed process and motivation for selecting given techniques and algorithms). The clearness and completeness of the delivered reports will also be considered
- the grade is valid for the entire academic year

Individual oral part

- the assessment covers all the theoretical parts of the course. The score is based on the completeness and clarity of the answers
- the grade is valid for the session

Autograding

Reasoning on assessing a proper performance is the best way to be aware of the value of what has been done

Students will be asked to grade both the project teamwork (1 grade for the whole team, this means you should agree beforehand) and for your individual oral part

Along with the grade, we will expect to hear the reasons

Both grades will be considered as proposals and be considered for the final grade

Not happy with your final grade?

Project team

- you can further extend some parts that resulted be brittle of your assigned project
- in the final delivery (slides, document and software) we expect a delta for each output. If the delta is missing, there are no conditions to have the retake
- be aware: support from the project mentors and from the teaching assistants will become best effort, thus do not pretend any help

Individual oral part

- You can retake the oral part

Grades are kept within the academic year

Class and lab participation

Both participations are recommended but not mandatory

We believe they are an important moment for learning and exchanging ideas and clear doubts

These moments cannot be substituted by questions on slack or requests of meetings

Meetings

Maximize your participation during classes and labs to shed light on your doubts

External meetings (not part of the teaching duties) have an enormous costs to our teaching execution

Anyway, we offer you the chance to digging into the topics and clear doubts

A meeting is subject to our decision, and it will last no more than 30 minutes

In order to maximize the meeting, make sure you did your homework preparing a detailed agenda

Student well-being

Courses are stressful environments, we acknowledge this

In my experience, most student integrity violations are the product of these environments and decisions made out of desperation

Is there a need to get to that point?

Don’t sacrifice quality of life for this course, which remains a course where to learn and not to perish

Make time to sleep, eat well, exercise, socialize and have fun when working at the project, these are key for success

Thank you for your attention.

Questions?

FONDAZIONE links

PASSION FOR INNOVATION

CONTACTS

Giuseppe Rizzo

Program Manager (LINKS Foundation) and
Adjunct Professor (Politecnico di Torino)

giuseppe.rizzo@polito.it

FONDAZIONE LINKS
Via Pier Carlo Boggio 61 | 10138 Torino
P. +39 011 22 76 150
LINKSFOUNDATION.COM