

LINKSFOUNDATION.COM

Applied Data Science Project

Lecture 1

Turin

September 28, 2021

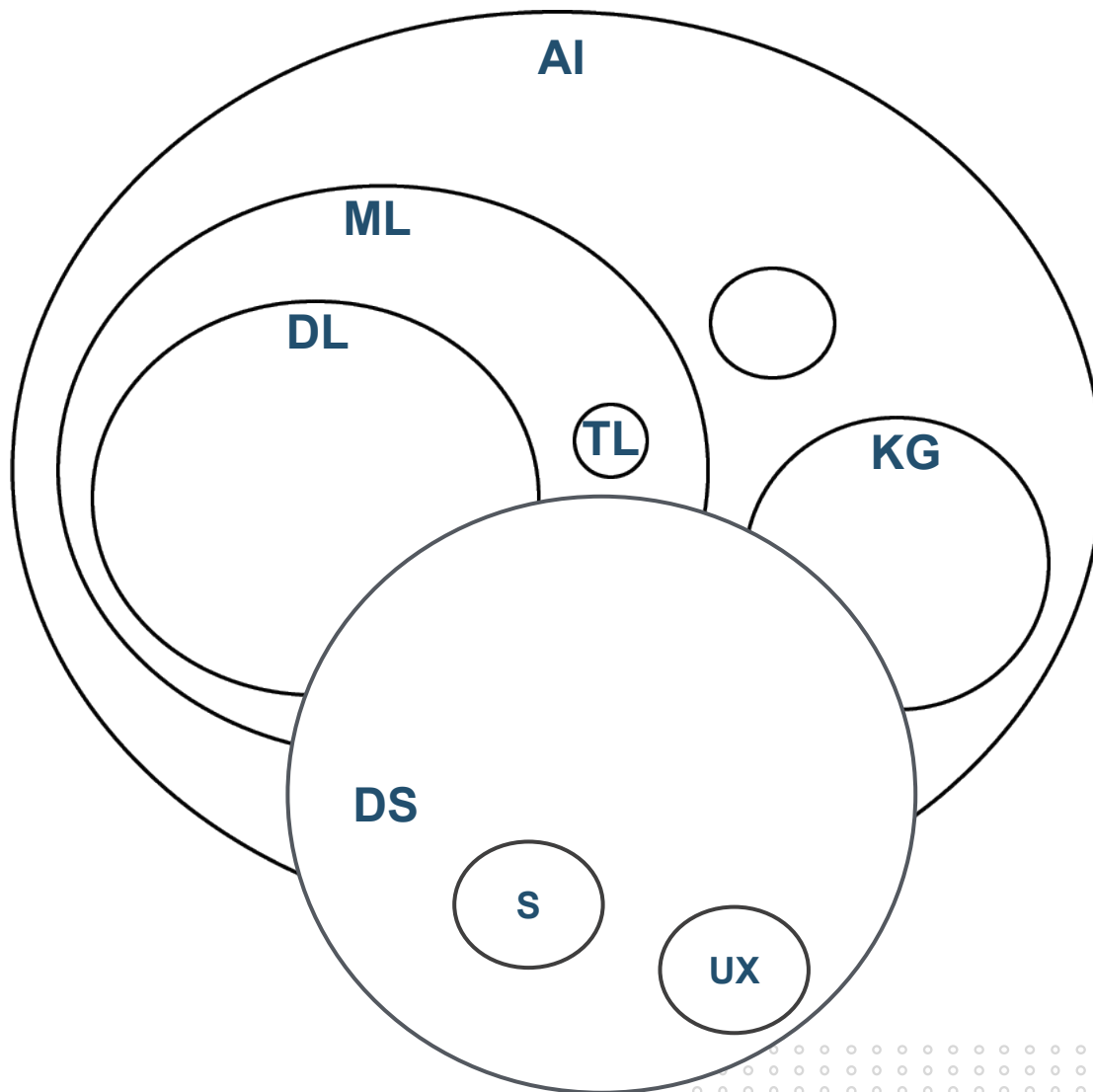


**Politecnico
di Torino**



e l i s
European Laboratory for Learning and Intelligent Systems

Interplay



AI: Artificial Intelligence

ML: Machine Learning

DL: Deep Learning

TL: Transfer Learning

KG: Knowledge Graph

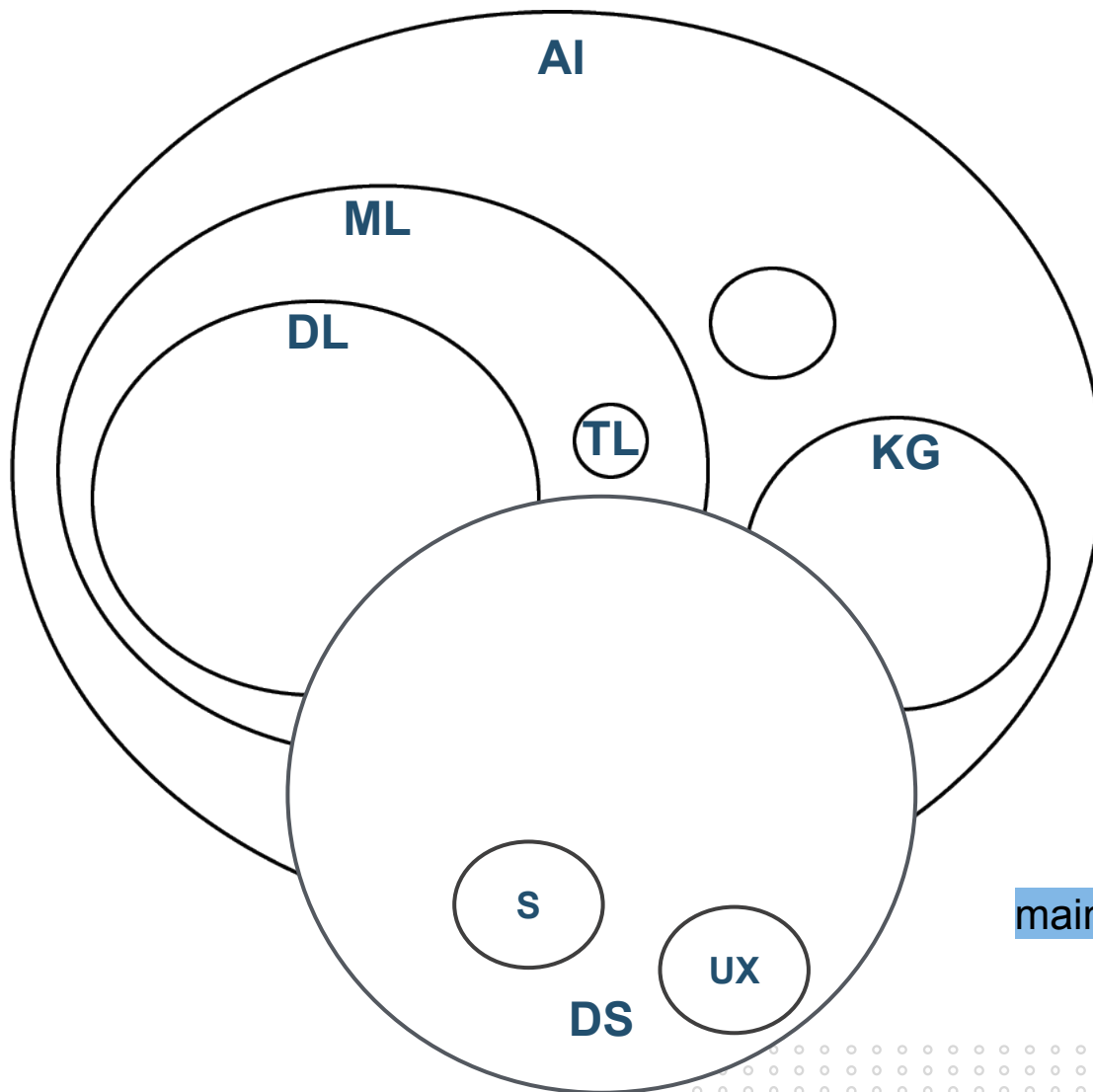
S: Statistics

UX: User eXperience

DS: Data Science

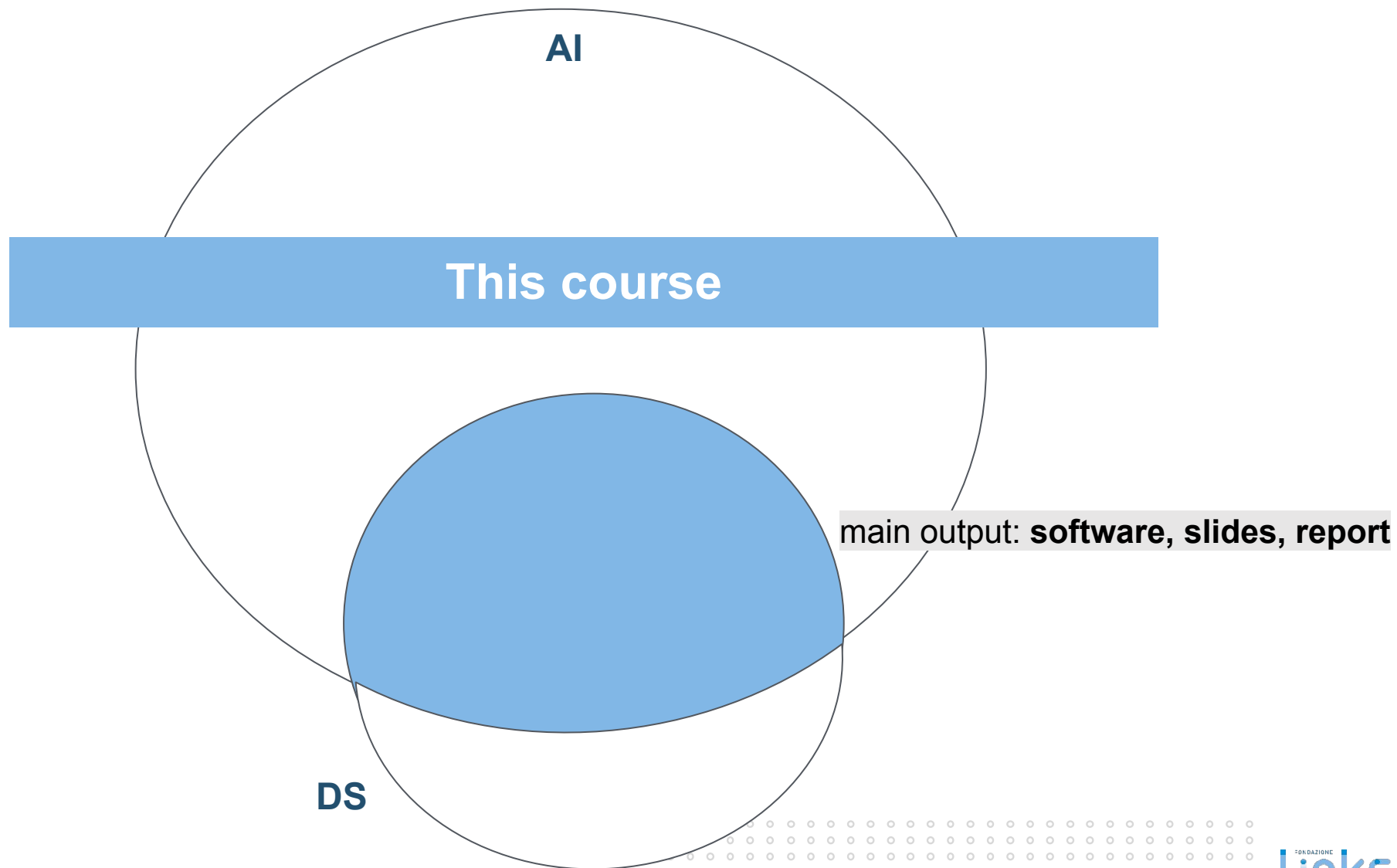
Interplay

main output:
software



main output: visuals, slides

Applied Data Science Project



Applied Data Science Project 101

Bear in mind: your ambition is to develop **machine intelligence** clearly **documented**

- 1 Start from a real world challenge and define the goal/value
- 2 Design a project to
 - manage the activity & monitor progress
 - select the right tools
 - develop a machine intelligence to answer the challenge and generate value
- 3 Quantify the benefits and impact of the project

R&I pathway

1

Start from a real world challenge and define the goal/value

2

Design a project to

- manage the activity & monitor progress
- select the right tools
- develop a machine intelligence to answer the challenge and generate value

3

Quantify the benefits and impact of the project

research and innovation pathway



Goal and value

Why are you conducting the project?

Where it can be used?

Who are the potential users?



Align project goals with the SDGs





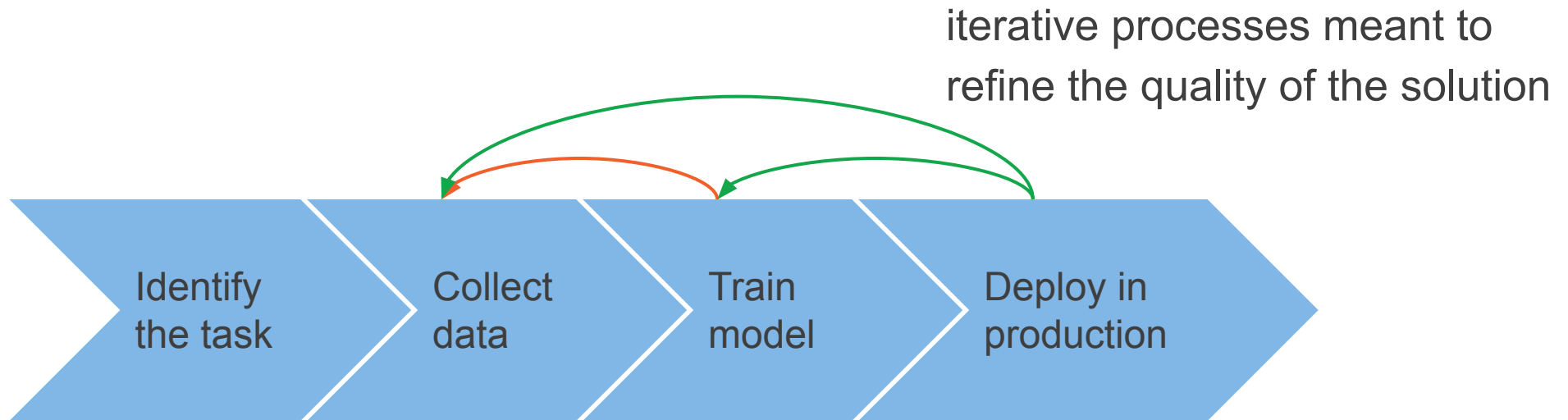
Scoping a project



How the machine intelligence generated in the project will look like?



Scoping a project



Management tools

“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

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:

$$\text{value} - \text{cost of fulfilment} > 0$$

Team



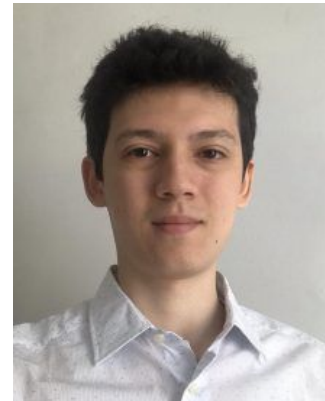
Giuseppe Rizzo
Prof, 30 ore



Marco Torchiano
Prof, 9 ore



Alberto Benincasa
TA, 21 ore



Luca Colomba
TA, 15 ore



Salvatore Greco
TA, 45 ore

Logistics



Topics

- Introduction
- Model and Data-centric projects
- Foundation models
- Sustainable development goals and data science examples
- Data Science project pillars:
 - Design
 - Develop
 - Manage
 - Communicate
- 10 tips
- Agile and Scrum methodologies
- Collaborative and interactive project design
- Collaborative and interactive project development
 - Google Colab setup and use
 - Github setup and use
- Collaborative project management
 - Asana setup and use
- How to preparing clear slides to present the project
- How to write a technical report
- Project champions
- Project proposals

Recommended Background

The only formal prerequisite for this course is an intro to **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**

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

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 guess 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>

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

Additional reading materials

- Data Science Applications Across The Industries: Applied Data Science, by Goran Dragosavac
- Data Science from Scratch, Joel Grus
- On the Opportunities and Risks of Foundation Models, by Bommasani et al.
- Machine Learning Yearning, by Andrew Ng
- Transforming our world: the 2030 Agenda for Sustainable Development, by United Nations
- Scrum Mastery: From Good To Great Servant-Leadership, by Geoff Watts
- Agile Project Management for Dummies, by Mark C. Layton

Discussions and reporting

Slack group adsp-polito.slack.com

Please all join the group:

<https://join.slack.com/t/adsp-polito/signup?x=x-p2410953592102-2417895063075-2410979403766>

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 (mandatory)

Laboratories for projects

A major part of the course is to conduct a real-world project, this accounts about $\frac{3}{4}$ of the whole course

There will be up to 5 projects: you will be asked to express your interest as team (each team is up to 3 students)

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 GSlides, Asana, Colab Notebook, GDoc that will include the analysis of your data, including code and visual results

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:

- 1 set of slides
- document
- software

3 intermediate checks, mid November, mid December, and mid of January: a 15-minute presentation for each group. The exact days will be announced with 1 week of notice

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 the external tutors who have followed the execution of the project

Your presentation will be listened by the instructors and external tutors will factor in to agree on your final grade

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:

$$0.8 * \text{grade team project} + 0.2 * \text{grade individual oral part}$$

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 team work (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 taken into account 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 diff of the delta for each output. If the diff is missing, there are no conditions to have the retake
- **be aware**: support from the project managers 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?

CONTACTS

Giuseppe Rizzo

Team Leader

p. +39 011 2276244

e. giuseppe.rizzo@linksfoundation.com



FONDAZIONE LINKS

Via Pier Carlo Boggio 61 | 10138 Torino

P. +39 011 22 76 150

LINKSFOUNDATION.COM