

01VIXSM

Deep Natural Language Processing

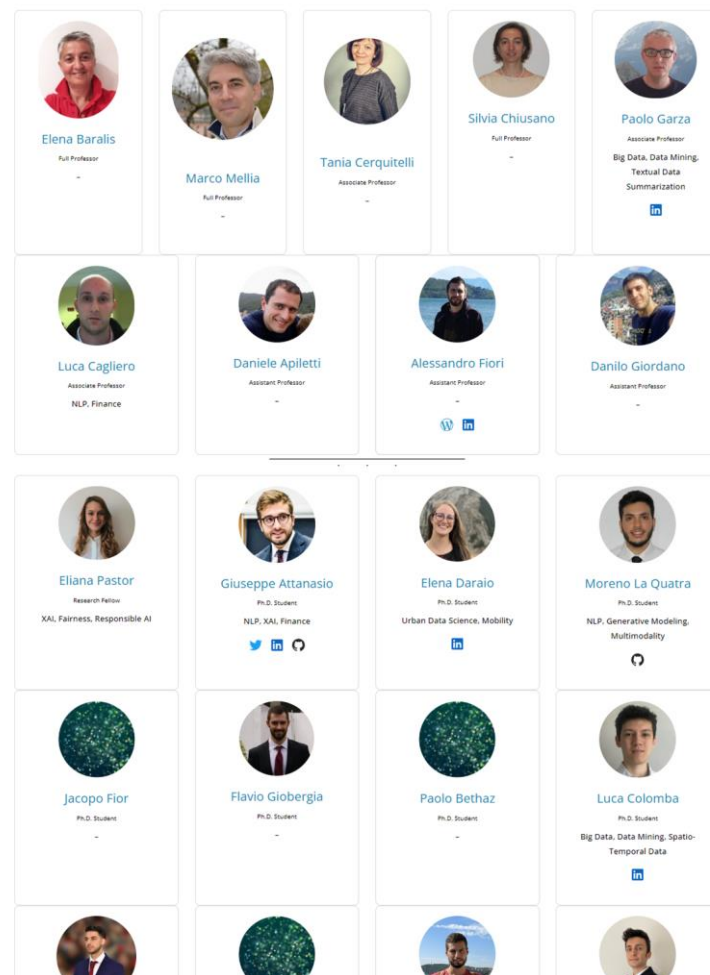
Prof. Luca Cagliero
Dipartimento di Automatica e Informatica
Politecnico di Torino



**Politecnico
di Torino**

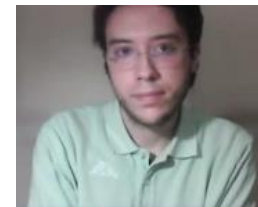
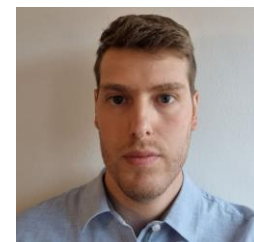
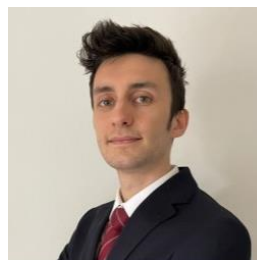
Who we are

- 4 Full Professors
- 2 Associate Professors
- 3 Researchers with Tenure Track
- 2 Researchers with Non-Tenure Track
- Around 20 PhD/Post Doc



NLP, Multimodal Learning and related areas

- Luca Cagliero, Associate Professor
- Jacopo Fior, Post Doc
 - Time series Analysis, Financial Data
- Lorenzo Vaiani, PhD Student
 - Multimodal Learning, NLP
- Giuseppe Gallipoli, PhD Student
 - NLP, Conversational Agents
- Davide Napolitano, PhD Student
 - Multimodal Learning, Explainable AI
- Irene Benedetto, PhD Student
 - Apprenticeship PhD with Jakala
 - Manager: Dr. Francesco Tarasconi
 - Legal AI
- Simone Papicchio, PhD Student
 - Joint program with EURECOM (Sophia Antipolis, France)
 - Co-supervised by Prof. Paolo Papotti
 - Table Representation Learning, Large Language Models



Current research interests - NLP

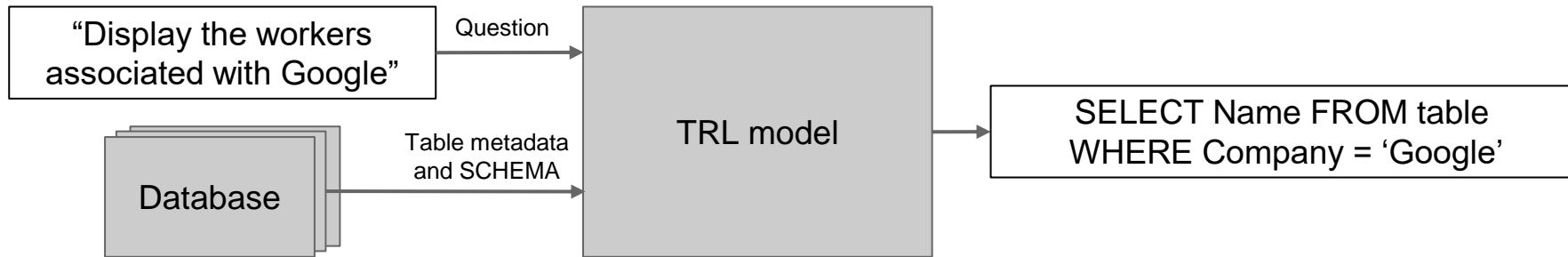
- Summarization
 - News, Scientific Articles, Patents, Social Posts, etc.
- Text Style Transfer
 - From informal to formal, from negative sentiment to positive
- Entity Recognition and Disambiguation
 - Legal Terms, Financial Parties, Figurative Language
- Large Language Models
 - Few-Shot Learning, NLU model evaluation

Current research interests – Multimodal Learning

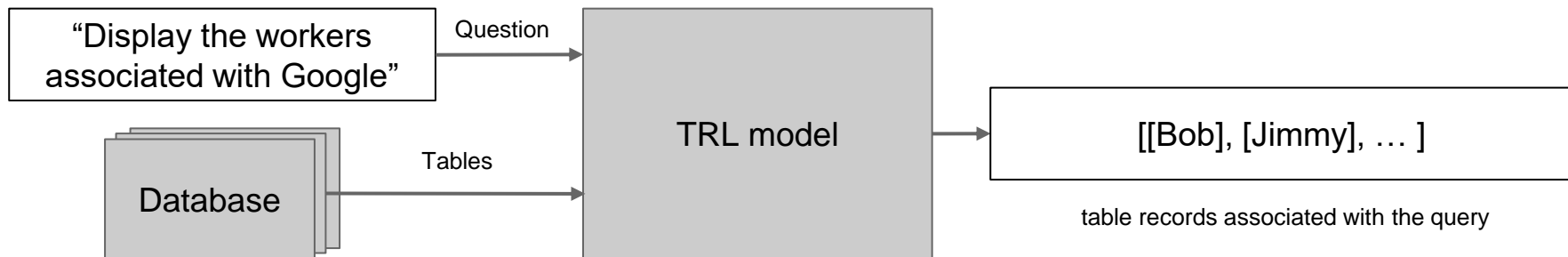
- Hate Speech Detection in Social Media Contents
- Multimodal Fake News detection
- Multimodal Machine Translation
- Multimodal Word Sense Disambiguation
- Emotion Recognition from Video Clips
- Automatic Slide Generation
- Audio Podcast and Video Meeting Summarization
- Explainable AI
- Modality-Agnostic Architectures

Current research interests – Table Representation Learning

- **Semantic Parsing**

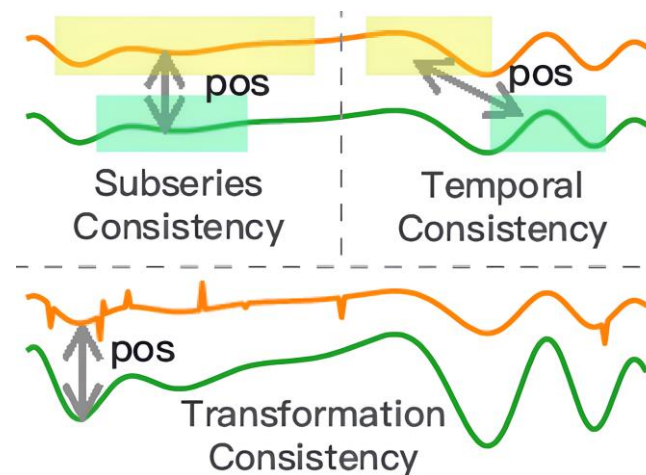


- **Question Answering**



Current research interests – Time Series Analysis

- Contrastive representation of time series
 - Augmentation techniques for self-supervised learning
 - Time series segmentation
 - Time series forecasting
 - Anomaly detection



Calendar

- 8 CFU -> 80 hours of frontal lecture/practice
 - 41 hours of frontal lecture
 - 18 hours of in-class practice
 - In presence, except for
 - Catch-up lessons
 - We will use BBB Virtual Classroom if necessary
 - 21 hours of laboratory practices (3 hours each)
 - All in presence

Calendar

- Planning

- 14 weeks
- 7 labs -> half of the weeks
- 59 hours of in-class lectures/practices
 - Ideally
 - 4.5 hours per week
 - In practice
 - 6 hours per week for the first half of the semester
 - 3 hours in the last weeks
- Why?
 - We need to introduce NLP fundamentals asap to start the project work earlier

Timetable

- Available on the Didactic Portal
- Lectures
 - Tuesday from 4pm to 7pm Room 29
 - Friday from 1pm to 4pm Room R1
 - Friday 20/10 NO LECTURE
 - (Further changes to be announced)
- Practices
 - Thursday from 4pm to 7pm at LADISPE
 - NO LAB IN THE FIRST WEEK

Videorecordings

- In-class lectures will be recorded using the Virtual Classroom (BBB)
- Video recordings will be made available to enrolled students only
 - The publication delay is unpredictable (typically, few days)
- Lab practices will not be recorded
 - We will share all the necessary materials (text, solutions, etc) through the Didactic Portal in advance

Videorecordings

- No interactions with remote students
 - We foster students' participation to in-presence lectures
 - Dual-channel communication (both remote and in-presence) degrades the quality of in-presence lectures
 - If you have specific questions, please send an email
- Students' microphones and cameras we will disabled

Prerequisites

- Fundamentals of Data Sciences, Machine Learning, and Deep Learning
- Basic knowledge of the Python language

Content outline

- Fundamentals of Natural Language Processing
- Main Deep Learning solutions for learning word, sentence, and contextualized embeddings
- Fundamentals of recommender systems
- Main NLP applications

Content outline

- Fundamentals of Natural Language Processing
- Main Deep Learning solutions for learning word, sentence, and contextualized embeddings
- Fundamentals of recommender systems
- Main NLP applications

Main topics

- Natural Language Processing fundamentals
 - text characteristics, text preparation, topic modelling, main NLP applications
- Vector representations of text
 - word embedding architectures and shallow sentence embedding architectures
- Contextualized embedding, the attention mechanism
- Entity Recognition, Intent Detection, Question Answering
- Text summarization
- Machine Translation
- Recommender Systems
- Application of NO SQL Databases for Information Retrieval: Elastic Search
- Text Categorization and Sentiment Analysis
- NLP pipeline design
 - requirement analysis, methodology design and implementation, empirical assessment, outcome presentation

Main topics

- More details available at

https://didattica.polito.it/pls/portal30/gap.pkg_guide.viewGap?p_cod_ins=01VIXSM&p_a_acc=2024&p_header=S&p_lang=EN

Laboratories

1. Text preprocessing and topic modelling
2. Word and sentence embeddings
3. Entity Recognition and intent detection
4. Information Retrieval and recommender systems
5. Summarization
6. Machine translation
7. AI Chatbots

LABS START IN THE 2ND WEEK

(schedule of the next labs to be announced)

Laboratories

- We use Google Colab
 - <https://colab.research.google.com/>
- You need a Google account
- Practice text and solutions will be shared through the didactic portal to all enrolled students

Teaching materials

- Handouts will be uploaded on the didactic portals
 - They will be made available throughout the course
- Additional materials
 - Scientific papers used to deepen the knowledge on specific topics covered by the slides
 - Not mandatory to pass the exam

Teaching materials

- Books / additional readings
 - Embeddings in Natural Language Processing: Theory and Advances in Vector Representations of Meaning. Mohammad Taher Pilehvar, Jose Camacho-Collados. Morgan & Claypool. ISBN: 9781636390215
 - Neural Network Methods for Natural Language Processing. Yoav Goldberg. Morgan & Claypool. ISBN: 9781627052955
 - Deep Learning in Natural Language Processing. Li Deng and Yang Liu Editors. Springer. ISBN: 9789811052088

Exam rules

- The exam consists of
 - a written test (max. 22 points)
 - Covering theoretical aspects introduced during the course
 - Closed and/or open-ended questions
 - A group project (max. 10 points)
 - A team project is assigned during the course
 - The project must be submitted through the Didactic Portal (section “Elaborati”)
 - The evaluation comprises an oral discussion
 - The final score is given by the sum of the points achieved in the written part and in the evaluation of the final report

Exam rules

- The written test is scheduled during the official exam session
 - All marks will be recorded by the end of the session
 - Exam bookings are mandatory
 - The exam is closed-books
- The group project will be assigned well before the end of the course
 - Teams consist of 3-4 students
 - The assignment of the specific topic is at the discretion of the teachers

Exam rules

- The project score is valid until the end of the academic year (until September 2024)
- When you pass the written test we automatically assign the final grade by summing the achieved project score (if any)
- If you want to improve the mark of the written test you can reject it and redo the test at the next available exam
- If you want to improve the score of the project you have to redo it from scratch
 - No supplementary oral part is allowed

Project rules

- Each team has to
 - Submit a project report (including references to project code, data, and models for reproducibility reasons)
 - Orally present the work (in group)
 - Whether using presentation slides or not
- To sum up the achieved project score within a exam session, you have to deliver the project report and code 1 week before the end of the session
 - We will revise report and code and set up an appointment for the discussion by the end of the session
- All grades of the written part (possibly including the project score) **MUST** be recorded within the official exam session

Project rules: examples

- I plan to give the written exam in the winter examination session (January-February 2024)

Tuesday, 19/12/2023

Deadline to complete the enrollment (candidates from other universities)

Wednesday, 27/12/2023 to Friday, 05/01/2024

Closure of the Office of the University Registrar and all Offices

(University closure)

Monday, 22/01/2024 to Saturday, 02/03/2024

WINTER EXAMINATION SESSION – a.y. 2023/24

Sunday, 03/03/2024

Credit Transfer - Master's degree programmes - second semester (recognition of credits earned at another university)

Deadline for submitting your Apply@polito application

Monday, 04/03/2024

Application deadline for admissions to Master's degree programmes in **second semester**

Monday, 04/03/2024

Application deadline for credit transfer for Master's Degree programmes - **second semester**

To get the extra points my team has to deliver the project at most 1 week before the end, i.e., by 24/2/2023

Project rules: examples

- I plan to give the written exam in the winter examination session (January-February 2024)

Tuesday, 19/12/2023

Deadline to complete the enrollment (candidates from other universities)

Wednesday, 27/12/2023 to Friday, 05/01/2024

Closure of the Office of the University Registrar and all Offices

Monday, 22/01/2024 to Saturday, 02/03/2024

WINTER EXAMINATION SESSION – a.y. 2023/24

Sunday, 03/03/2024

Credit Transfer - Master's degree programmes - second semester (recognition of credits earned at another university)

Deadline for submitting your Apply@polito application

Monday, 04/03/2024

Application deadline for admissions to Master's degree programmes in **second semester**

Monday, 04/03/2024

Application deadline for credit transfer for Master's Degree programmes - **second semester**

If you deliver the project on time, we will set up a 20/30min call/meeting to discuss about the project during the last week of the session, i.e., between 24/2 and 2/3

Consulting

- At the end of the lecture or by setting up an appointment
 - Send an email
- Practical questions about code and libraries
 - Mainly during the laboratory activities and in-class practices
- Theoretical/methodological questions
 - Mainly during frontal lectures

Consulting

- A couple of in-class lectures will be devoted to the help desk for the final project
 - Mainly for methodological questions and answers
 - No real-time code debugging!
- We recommend students to start the project work as soon as possible to take advantage of the help desk sessions

Acknowledgements and copyright license

- Copyright licence

- Attribution + Noncommercial + NoDerivatives



- Acknowledgements

- I would like to thank Dr. Moreno La Quatra, who collaborated to the writing and revision of the teaching content.

- Affiliation

- The author and his staff are currently members of the Database and Data Mining Group at Dipartimento di Automatica e Informatica (Politecnico di Torino) and of the SmartData interdepartmental centre
 - <https://dbdmg.polito.it>
 - <https://smartdata.polito.it>

Thank you!