Natural Language Processing - 2nd Semester (2024-2025)
1038141

# 1.1 - Course Introduction

SAPIENZA
UNIVERSITÀ DI ROMA

**Prof. Stefano Faralli**
faralli@di.uniroma1.it

**Prof. Iacopo Masi**
masi@di.uniroma1.it

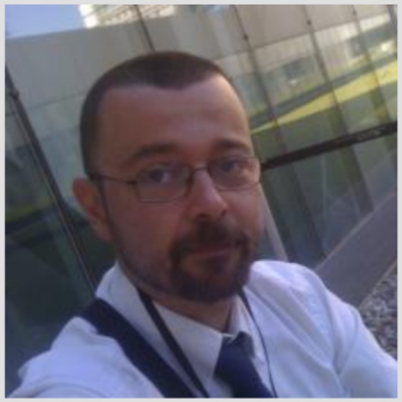
SAPIENZA
UNIVERSITÀ DI ROMA


SAPIENZA
UNIVERSITÀ DI ROMA

# 1.1 - Course Introduction

- Instructors
- Syllabus
- Course structure
- Course logistics
- Exams
- Books
- Q&A

# Instructors



## STEFANO FARALLI

**E-mail**
stefano.faralli@uniroma1.it

**Department**
INFORMATICA

**SSD**
INF/01

Profilo Research - Pubblicazioni IRIS


## IACOPO MASI

**E-mail**
iacopo.masi@uniroma1.it

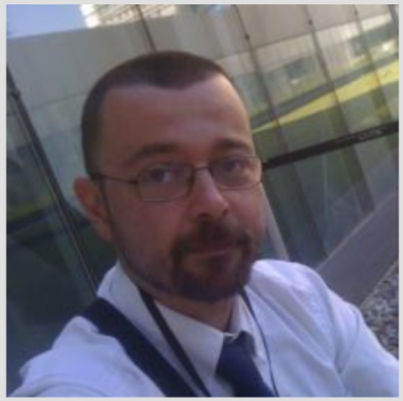
**Department**
INFORMATICA

**SSD**
INF/01

Profilo Research - Pubblicazioni IRIS

**Professor website**
AI & ML - Unit II AA21-22
Forum di Fondamenti di Programmazione
Fondamenti di Programmazione AA21-22
Fondamenti di Programmazione AA22-23
Sito Web Docente

# Teachers: Prof. Stefano Faralli




STEFANO FARALLI

**E-mail**
stefano.faralli@uniroma1.it

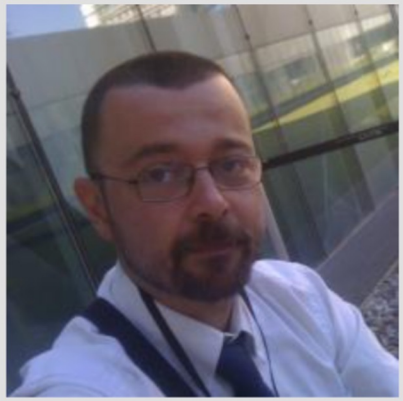
**Department**
INFORMATICA

**SSD**
INF/01

Profilo Research - Pubblicazioni IRIS


Dr. Stefano Faralli is an Associate Professor at Sapienza University of Rome. He is a co-head of the **Intelligent Information Mining (IIM)** group (http://iim.di.uniroma1.it/). He has a PhD in Computer Science and defended a thesis on **ontology learning from scratch** under the supervision of Prof. Roberto Navigli and Prof. Paola Velardi. He was a post-doc at the **Data and Web Science Group, University of Mannheim (Germany)** in the NLP group led by Prof. Simone Paolo Ponzetto, working in the "**JOIN-T: Joining Ontologies and Semantics Induced from Text**" project founded by Deutsche Forschungsgemeinschaft (German Research Foundation).

# Teachers: Prof. Stefano Faralli




STEFANO FARALLI

**E-mail**
stefano.faralli@uniroma1.it

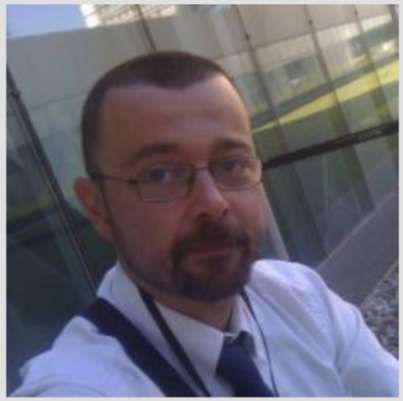
**Department**
INFORMATICA

**SSD**
INF/01

Profilo Research - Pubblicazioni IRIS


He also co-organized important workshops, including the SemEval 2015 task 17 "**Taxonomy extraction evaluation**" and the series of four international workshops "**BIAS in Search and Recommendation Algorithms**". He worked and published with around 70 researchers worldwide in more than 80 papers, with a particular focus on NLP-based knowledge acquisition/induction and representation applications, attracting 1500+ citations. He is also the co-author of the book "**Mining User Interests from Social Media**" (Zarrinkalam et. al., 2019) where he contributed to the chapters discussing the advancements of **NLP-based semantic-enabled user interest profiling methodologies** and the related **bias** and **fairness issues**.

# Teachers: Prof. Stefano Faralli




STEFANO FARALLI

**E-mail**
stefano.faralli@uniroma1.it

**Department**
INFORMATICA

**SSD**
INF/01

Profilo Research - Pubblicazioni IRIS


Stefano Faralli has experience in **lexical-semantic knowledge graphs induction**, **representation**, and **quality assessment**. Specifically, on the **large-scale induction of lexical-semantic knowledge** (Distante et al. JCognitiveComputation 2020; Faralli et al. CLiC-IT 2019; Faralli et al JWebSemantics 2017; Ristoski et al. WI 2017: Panchenko et al. SemEval, 2016; Faralli et al. ISWC 2016a; Velardi et. al., JComputationalLinguistics 2013; Navigli et al. IJCAI 2011), **entity linking** and **knowledge graph interlinking** (Faralli et al. ISWC 2021; Di Tommaso et al. ISWC 2018), **pruning of noisy knowledge graphs** (Faralli et al. JKnowledgeBasedSystems 2018; Faralli et al. IJCAI 2018; Faralli et al. EACL 2017) and in **unsupervised** or **supervised word sense disambiguation/Induction algorithms** (Faggiani et al. SWODSCH 2022; Biemann et al. JNaturalLanguageEngineering 2018; Panchenko et al. EMNLP 207; Panchenko et al. EACL 2017; Faralli et al. ISWC 2016b) and **evaluation** (Bordea et al. LREC 2020; Bordea et al. SemEval 2015; Velardi et al. LREC 2012).

# Teachers: Prof. Iacopo Masi


IACOPO MASI

**E-mail**
iacopo.masi@uniroma1.it

**Department**
INFORMATICA

**SSD**
INF/01

Profilo Research - Pubblicazioni IRIS

**Professor website**
AI & ML - Unit II AA21-22
Forum di Fondamenti di Programmazione
Fondamenti di Programmazione AA21-22
Fondamenti di Programmazione AA22-23
Sito Web Docente

Dr. Iacopo Masi is **Associate Professor** in the Computer Science Department at Sapienza, University of Rome. Till August 2022, I was also **Adjunct Research Assistant Professor** in the Computer Science Department at the **University of Southern California** (**USC**). Previously Dr. Masi was Research Assistant Professor and Research Computer Scientist at the USC Information Sciences Institute (ISI). Dr. Masi earned his Ph.D. degree in Computer Engineering from the University of Firenze, Italy. Immediately after, he moved to California and joined USC, where he was a postdoctoral scholar. Dr. Masi has been **Area-Chair of several conferences in computer vision** (WACVs, ICCV 21, ECCV 22) and currently serves as Associate Editor for The Visual Computer - International Journal of Computer Graphics. He organized an International Workshop on Human Identification at ICCV 17 and was Workshop Chair at SIBGRAPI 18. Dr. Masi was awarded the **prestigious Rita Levi Montalcini** award by the Italian government in 2018. Dr. Masi s main research interest lies in solving the computer vision problem. His background covers topics such as tracking, person re-identification, 2D/3D face recognition, and modeling, adversarial robustness, and facial manipulation detection.

# Teachers: Prof. Iacopo Masi

I do research in AI

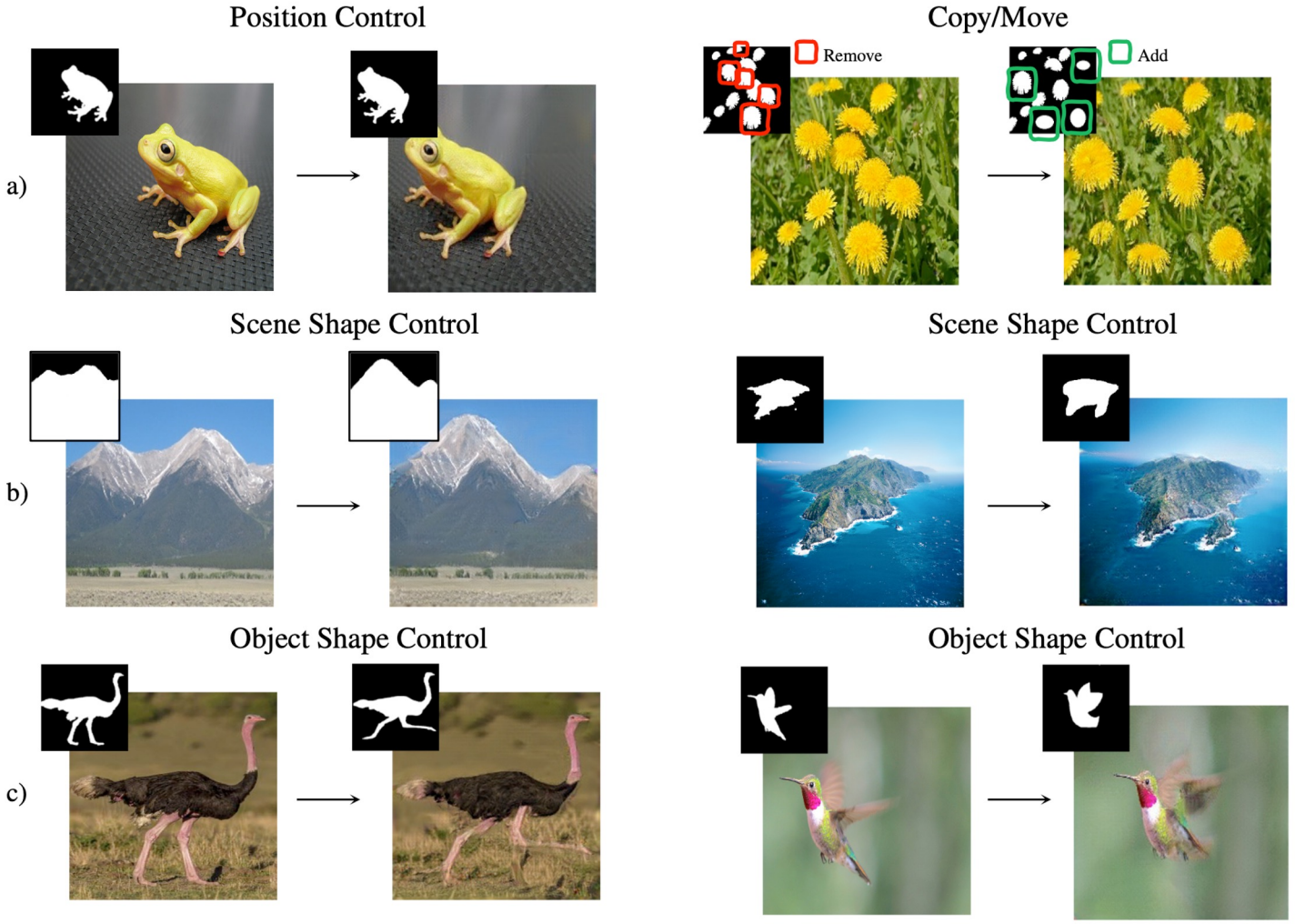
Biometrics, Face Recognition, Adversarial Robustness, Generative Models

Mainly visual domain (images) but I am broadening my range (e.g. NLP!) 😛



# Teachers: Prof. Iacopo Masi

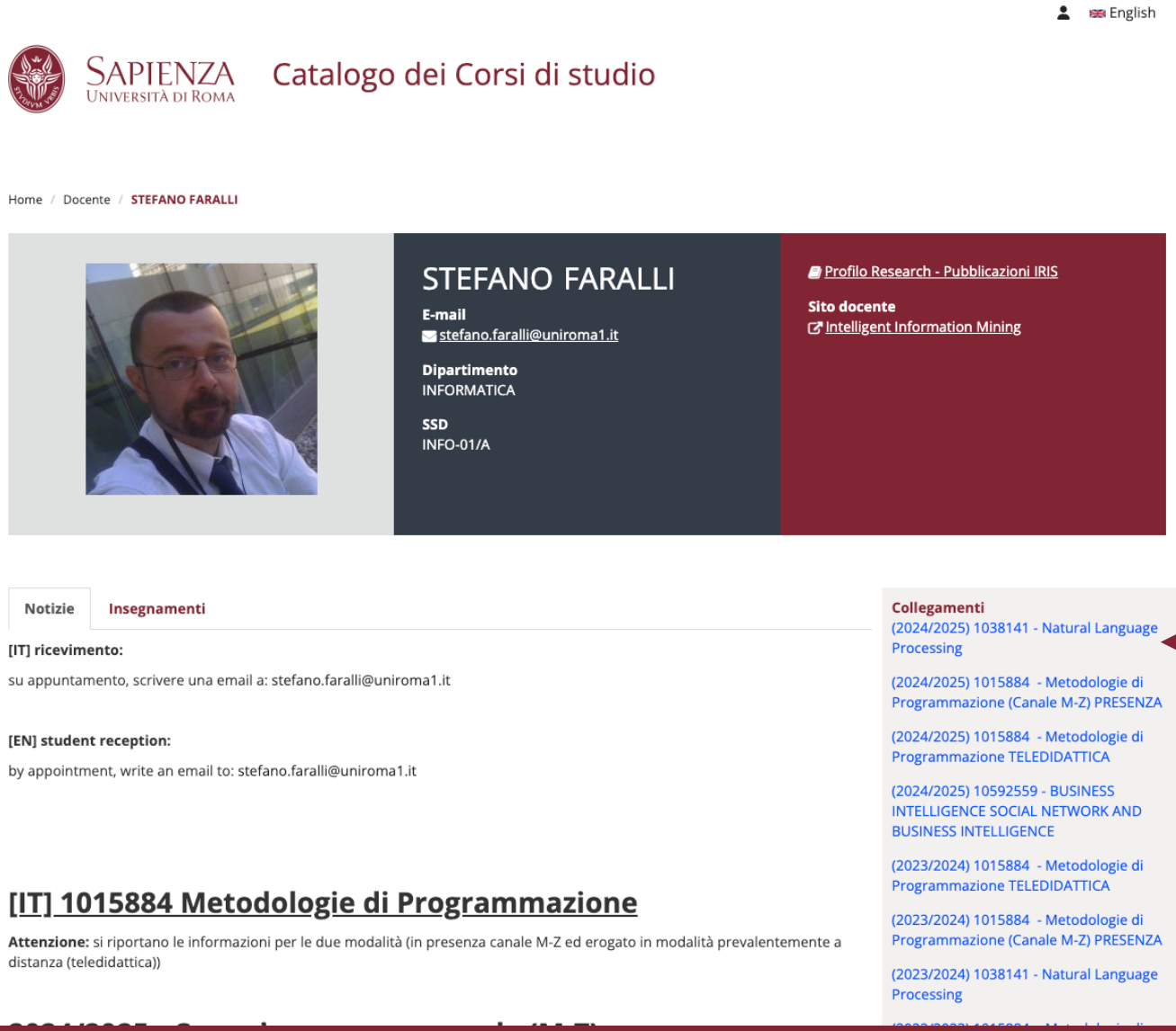
Image Synthesis by Inverting a Quasi-Robust Classifier in AAAI-23



# Course Logistics

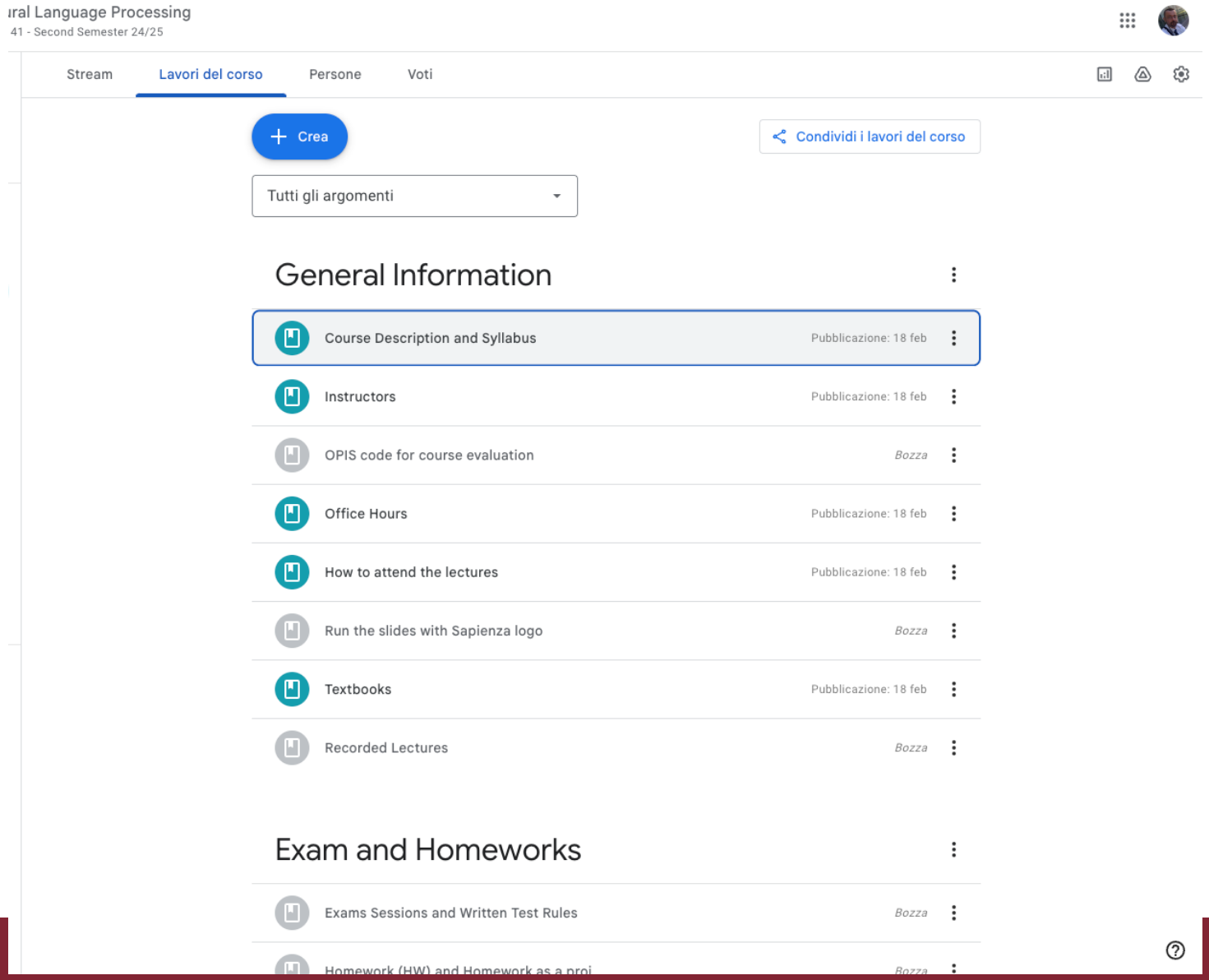
- course Web page "classroom":
  https://classroom.google.com/c/MjMxMjU0OTczOTFa?cjc=7b5izta

classroom
7b5izta



# NLP Classroom Code: mark it down!

7b5izta



# Syllabus [tentative]

classroom
7b5izta

- Introduction to Natural Language Processing
- N-gram models; smoothing; interpolation; backoff
- Part-of-speech assignment (including multilingual POS tagging)
- Syntactic analysis: statistical and neural techniques
- Computational semantics and lexical semantics
- Computational Lexicon: WordNet- Multilingual semantic networks: BabelNet- Word Sense Disambiguation; Entity Linking
- Multilingualism in Natural Language Processing- Neural networks, embodiments of words and senses, and deep learning
- Labeling of neural semantic roles and semantic analysis
- Statistical machine translation
- Natural Language Generation and Question Answering
- Neural language models (Transformers, BERT, XLM-RoBERTa, GPT)
- Neural machine translation
- Advanced multimodal NLP applications (text to images, text to video, diffusion model, Dall-E)

# Course Structure

classroom
7b5izta

## Part I (Prof. Stefano Faralli):

NLP introduction and downstream tasks

- Introduction to NLP, Regular Expressions, Finite State Automata and REs
- Words, Corpora and Text Normalization
- Spelling Correction and Minimum Edit Distance
- Language models, Part-of-speech-tagging
- Syntax, Semantics, Vector semantics (sparse), NLP tasks

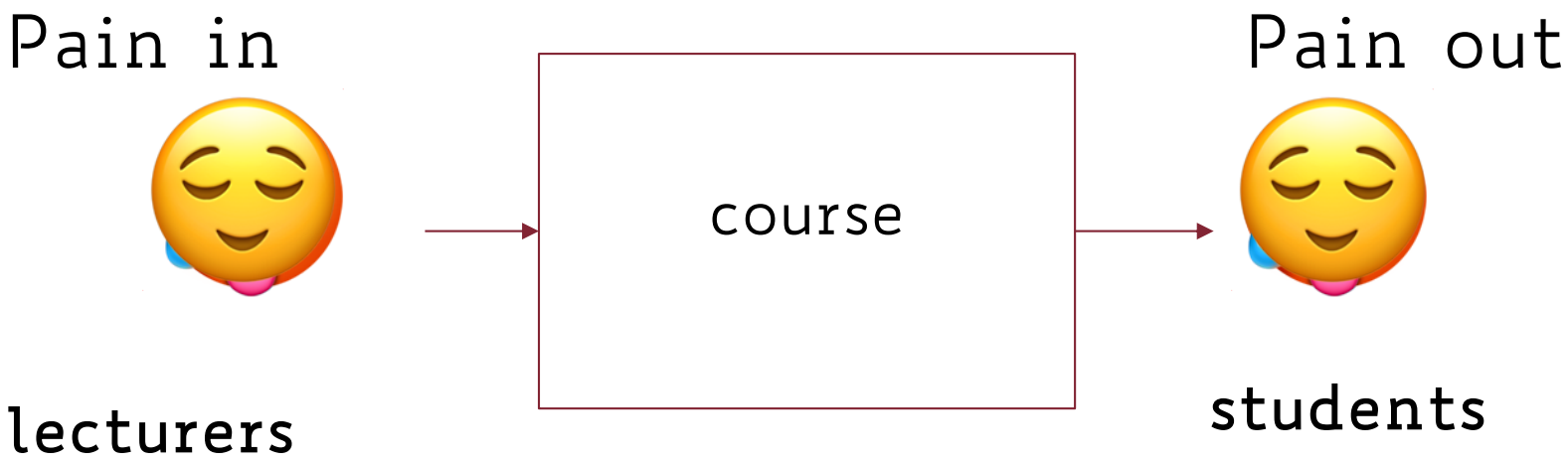
## Part II (Prof. Iacopo Masi):

Neural and multimodal NLP

- Latent Semantic Analysis and word2vec [hierarchical softmax & neg. sampling]
- Scaling word2vec, Sentiment Analysis, Language Model w/ Neural Net
- Sequence modeling w/ Deep Learning: LM /w RNN, POS, Image Captioning
- from LSTM to Transformers
- Neural Machine Translation, Encoder/Decoder, Beam Search
- Contextual Embedding: BERT, GPT, Transfer Learning
- Multimodal NLP: Diffusion models (images), NLP as supervision for Vision (CLIP)
- text2image application (Dall-E 2): based on diffusion and CLIP

# Bear with us while we [you] teach [learn]

classroom
7b5izta




- Let us live this course in an harmonious way and enjoy together the journey
- We have made our best to let you enjoy and learn the NLP world
- We hope we have made it reasonable in terms of exam/effort yet still enjoyable and fun

# Fair enough

classroom
7b5izta

6 CFU + Grade + precious NLP knowledge

Content + exam load

# Recorded Lectures

classroom
7b5izta

Lectures **will be recorded** so that students can integrate the classes "offline" in case they cannot attend in person a few times.

Recording quality is **"best effort": the screen will be recorded with the voice coming from a laptop microphone.** This is the best that we can do. In-person attendance is still strongly recommended.

We will upload the recordings **every week after two lectures** are given.

# Exam modality 🧐

classroom
7b5izta

The final grade will depend on two outcomes:

1. **Written test** — open and closed questions or exercise seen in class
2. **Homework (HW)** or extending HW as a project (**HWp**)

The final grade formula is:

```
ROUND[ 0.6 * written_test + 0.4* (HW or HWp) ]
```

An outcome is passed if grade >= **18**. If not passed then we cannot record your grade.
So to record your grade: written test >=**18** and (HW or HWp) >=**18.**

The HW evaluation is as follows: **at best**, if you submit the HW only you may get **28** points.

The HWp evaluation is as follows: **at best**, if you submit the HWp you may get **32** points.

More info at this google doc

# Course Evaluation

classroom
7b5izta

Let's review together the section of the Google classroom dedicated to the description of the evaluation method.

# Textbooks

classroom
7b5izta

- Jurafsky and Martin. Speech and Language Processing, Prentice Hall, third edition, in preparation.
  https://web.stanford.edu/~jurafsky/slp3/
- Jacob Eisenstein. Introduction to Natural Language Processing, MIT Press, 2019.
  https://cseweb.ucsd.edu/~nnakashole/teaching/eisenstein-nov18.pdf
- Yoav Goldberg. Neural Network Methods for Natural Language Processing, Morgan & Claypool, 2017.
  https://github.com/Michael2Tang/ML_Doc/blob/master/Neural%20Network%20Methods%20in%20Natural%20Language%20Processing-Morgan%20%26%20Claypool%20Publishers%20(2017)%20-%20Yoav%20Goldberg%2C%20Graeme%20Hirst.pdf

the provided links point to free editions.

classroom
7b5izta

# Q&A

classroom
7b5izta

SAPIENZA
UNIVERSITÀ DI ROMA

# Credits

Highly readable font Biancoenero© by biancoenero edizioni srl, designed by Umberto Mischi, with the consultancy of Alessandra Finzi, Daniele Zanoni and Luciano Perondi (Patent no. RM2011O000128).

Available free of charge for all institutions and individuals who use it for non-commercial purposes.