



Fondamenti di Analisi dei Dati

from **data analysis** to **predictive techniques**

Prof. Antonino Furnari (antonino.furnari@unict.it)
Corso di Studi in Informatica
Dip. di Matematica e Informatica
Università di Catania



Università
di Catania

Introduction to the course

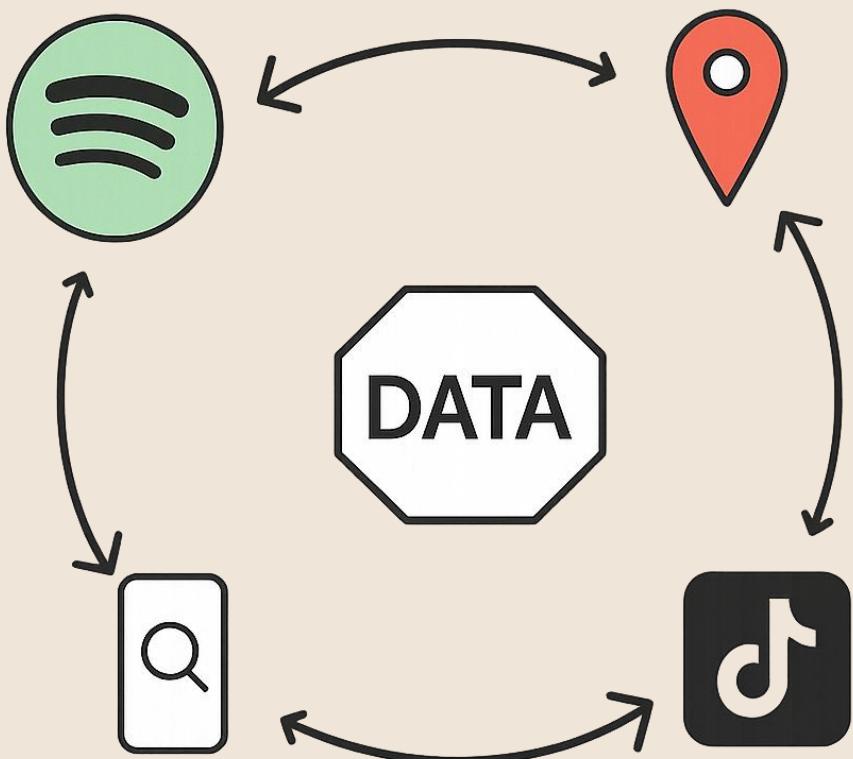
General Information

- **Course title:** Fondamenti di Analisi dei Dati
- **Teacher:** Antonino Furnari
- **CFU:** 9
- **Lectures:** Mondays and Wednesdays 8-11. Additional ~5 make-up classes on Tuesdays 14-17. Lectures follow a schedule shared on the Team.
- **Lecture rooms:** Room 23 (Tuesdays and Fridays) + **Room 22 (Thursdays)**
- **Teaching material:** <https://antoninofurnari.github.io/fadlecturenotes2526/>
- **Announcements and schedule:** MS Teams team (code i87g4nb)
- **Office Hours:** Mondays 11.30-13.30 – check for variations and book here:
<https://antoninofurnari.github.io/ricevimento/>
- **Syllabus:** <https://web.dmi.unict.it/corsi/l-31/insegnamenti?seuid=4EDF6456-342B-4D35-909F-33B4B835AB44>

Fondamenti di Analisi dei Dati

Fondamenti di Analisi dei Dati

Why Data Matters?



- Your Spotify playlist is curated by data;
- Your Google Maps route is optimized by data;
- Your TikTok feed is personalized by data;
- With every interaction (swipe, like, purchase, download, view, etc.), you contribute to a data science model.

We are Living in a Data Explosion



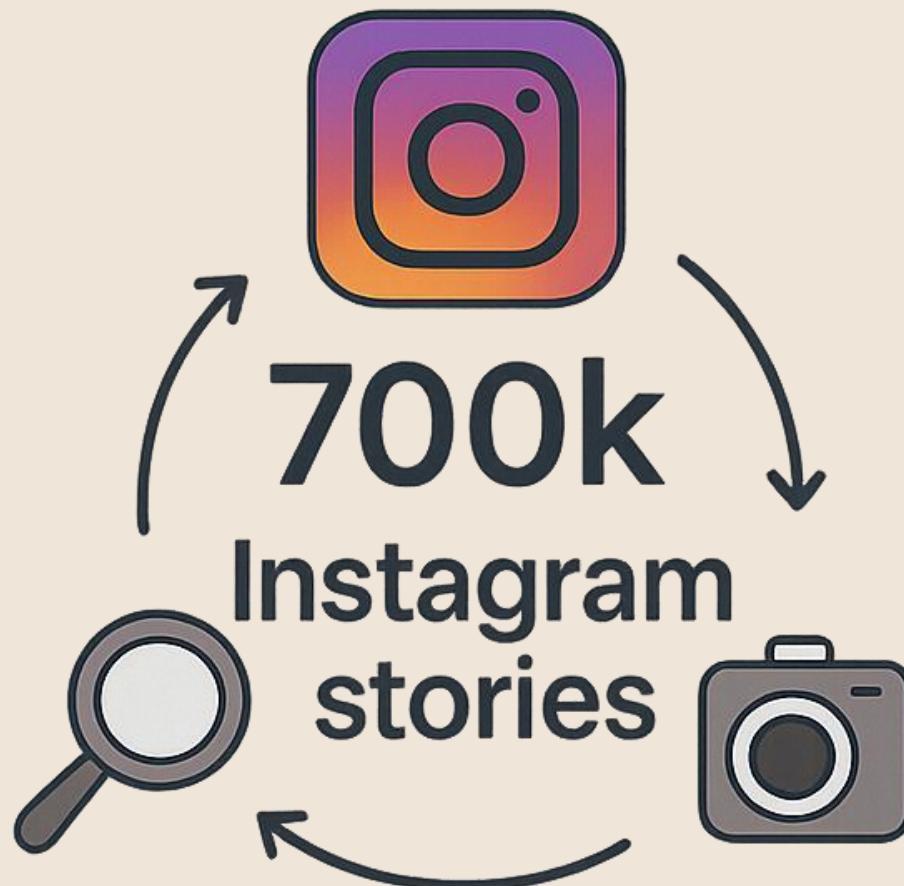
- In the early 1990s, the world's digital computing data—excluding consumer media like music CDs and analog formats—was estimated at just a few dozen gigabytes, roughly **the size of 50 CDs**.
- Today, we generate that much data **in seconds**.

We are Living in a Data Explosion

Every Minute:



7M
Google
searches

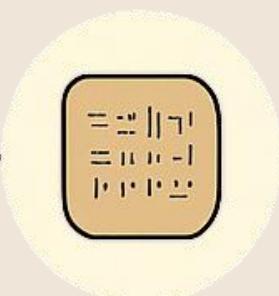


500k
hours of
YouTube
video

A (very) Brief History of Data

Ancient Civilizations

Sumerians used clay tablets to record grain inventories



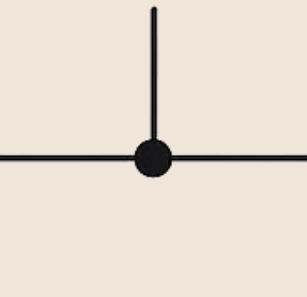
Middle Ages

Census-taking and tax records in medieval Europe



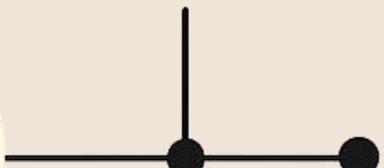
Enlightenment

Birth of statistics: probability theory, population studies



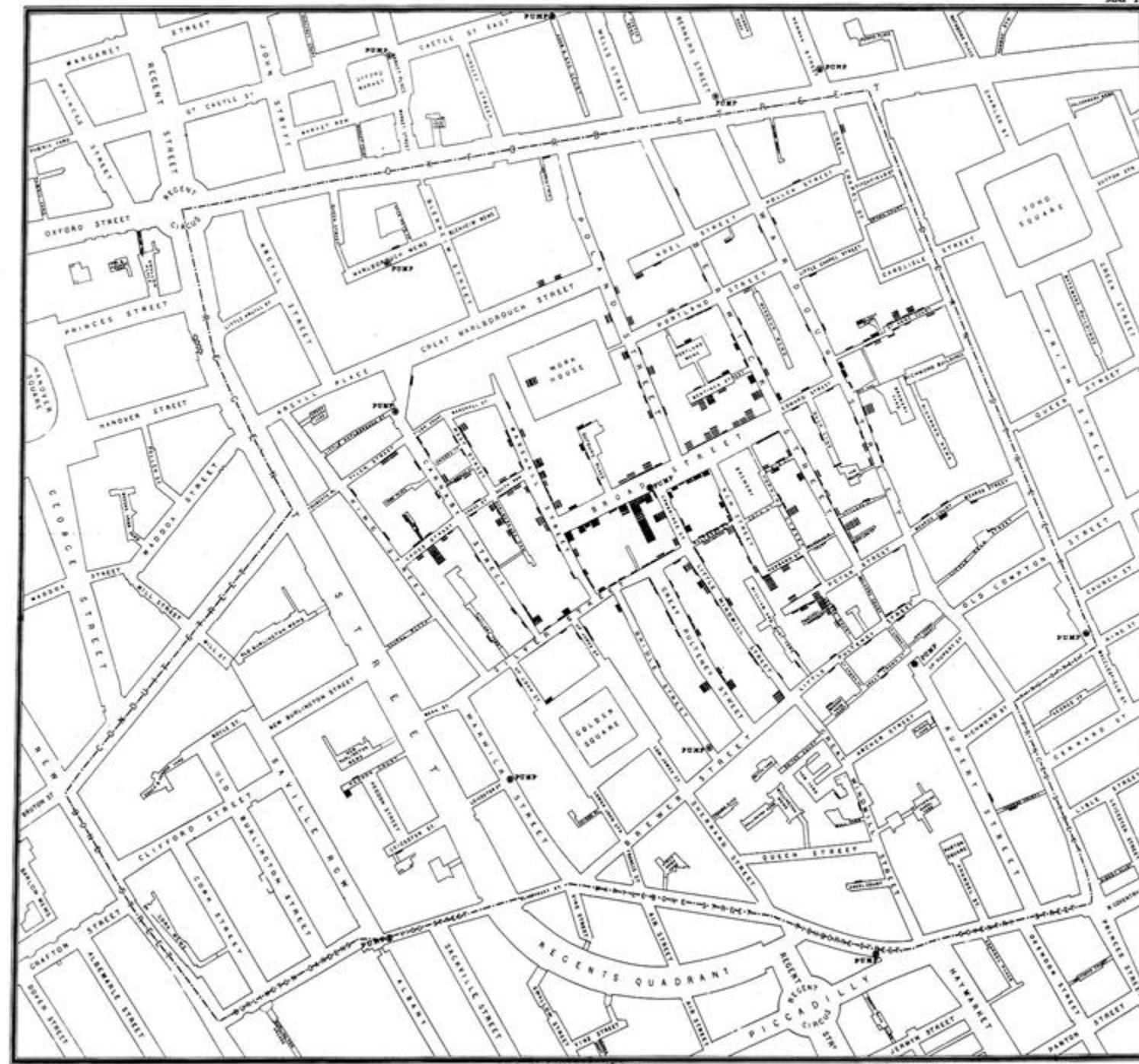
20th Century

IBM punch cards, government data systems



Jon Snow's Ghost Map (1954)

- Map of deaths during a cholera outbreak in London;
- Black lines represent deaths, localized in the map;
- Deaths concentrate in an area where citizens got water from a specific water pump;
- Jon Snow hypothesized that the cause was a contaminant and had the pump closed





What's Data Impact Today?

A few examples / case studies

Vaccines and Public Health



- Epidemiologists track infection rates to decide when and where vaccines are needed
- Vaccine effectiveness is measured using large-scale clinical trial data
- Public health agencies use data to allocate resources and prevent outbreaks

Polls and Public Opinion



- Pollsters use representative samples to estimate public opinion
- Margin of error and confidence intervals help interpret results
- Poll data influences campaign strategy, media coverage, and policy decisions

Elections and Turnout

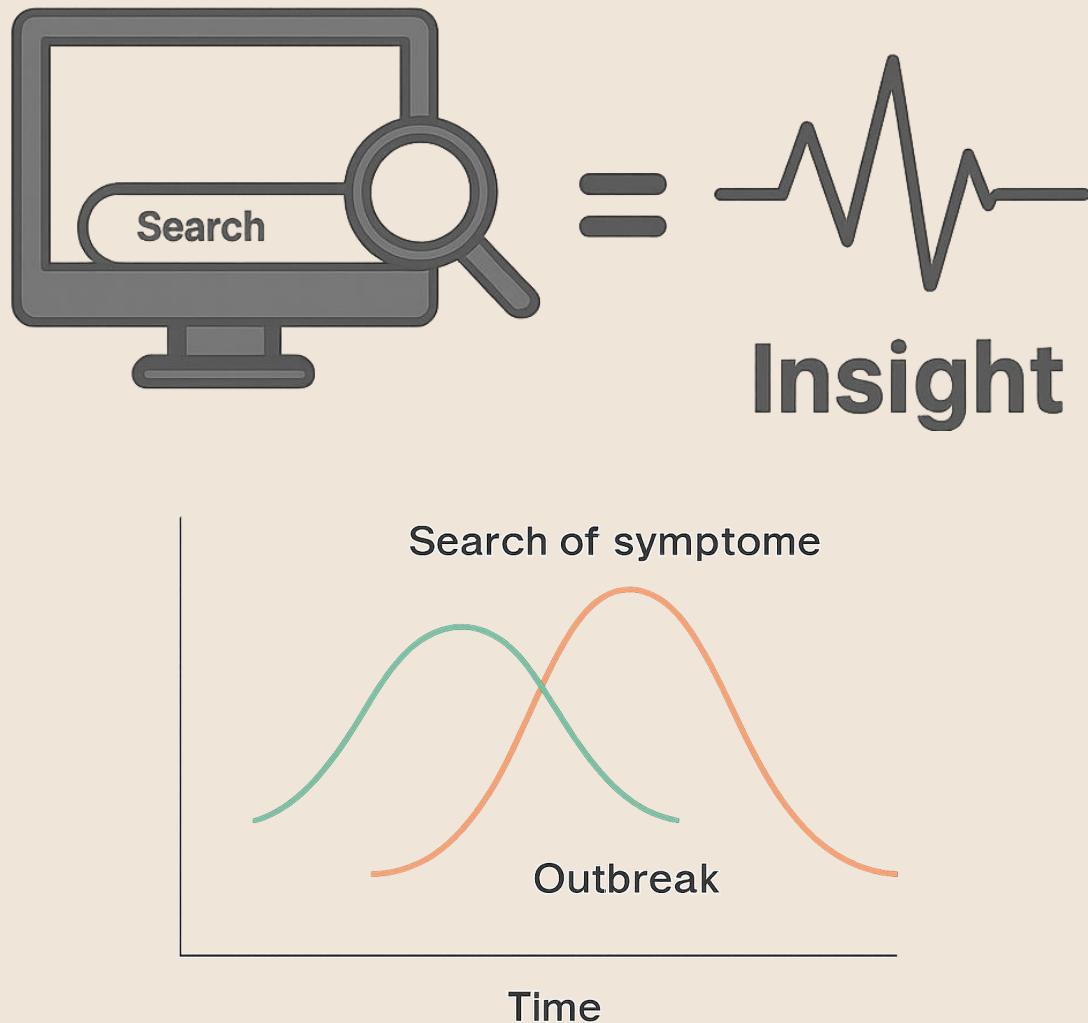


- Election officials use data to manage polling stations and prevent bottlenecks
- Geographic turnout data helps parties target outreach
- Historical voting patterns are used to forecast results

Beyond Analysis

Using Data for Automated Predictions and Decisions

Predicting Disease Outbreaks with Search Data



Search engines don't just reflect curiosity—they reveal emerging health trends.

- Symptom-related searches (e.g., “fever,” “cough,” “loss of smell”) often surge before official case reports.
- Google Trends and similar tools provide real-time, population-scale signals.
- Used in epidemiology to complement traditional surveillance systems.

Predicting Arterial Fibrillation



- ⌚ Apple Watch uses optical sensors and algorithms to detect irregular heart rhythms.
- ❤️ It can notify users of potential atrial fibrillation—often before symptoms appear.
- 🧠 FDA-cleared feature, backed by clinical studies.
- 🌐 Millions of users have received early warnings, prompting timely medical intervention.

Personalized Entertainment— Recommendations That Surprise



- Netflix's recommendation engine: trained on billions of viewing hours
- Netflix Prize (2006): crowdsourced 10 % accuracy boost in movie suggestions.
- TikTok's algorithm: learning your preferences in minutes
- Spotify Discover Weekly (2015): reached 40 M users in one year by clustering listening habits.

Smarter Journeys—Navigation & Smart Cities



- Waze community reroutes saved 10 min/trip on average; guided 500 K drivers around marathon closures.
- Google Maps ingests 20 PB of location data daily to avoid traffic.
- During an NYC marathon, Waze rerouted half a million drivers seamlessly.

Finance



- Fraud detection algorithms flagging suspicious transactions instantly
- Use statistical thresholds and anomaly detection—not always AI
- Robo-advisors managing investment portfolios based on data-driven strategies
- Offer low-cost, accessible financial planning—no human advisor needed

Climate & Environment



- Satellite data predicting deforestation (Global Forest Watch)
- AI models optimizing energy usage in smart cities
- NOAA runs calculations daily to forecast storms.
- NASA's earth satellites now produce more data than Apollo missions did in their entire run.

Our Everyday Smart Homes



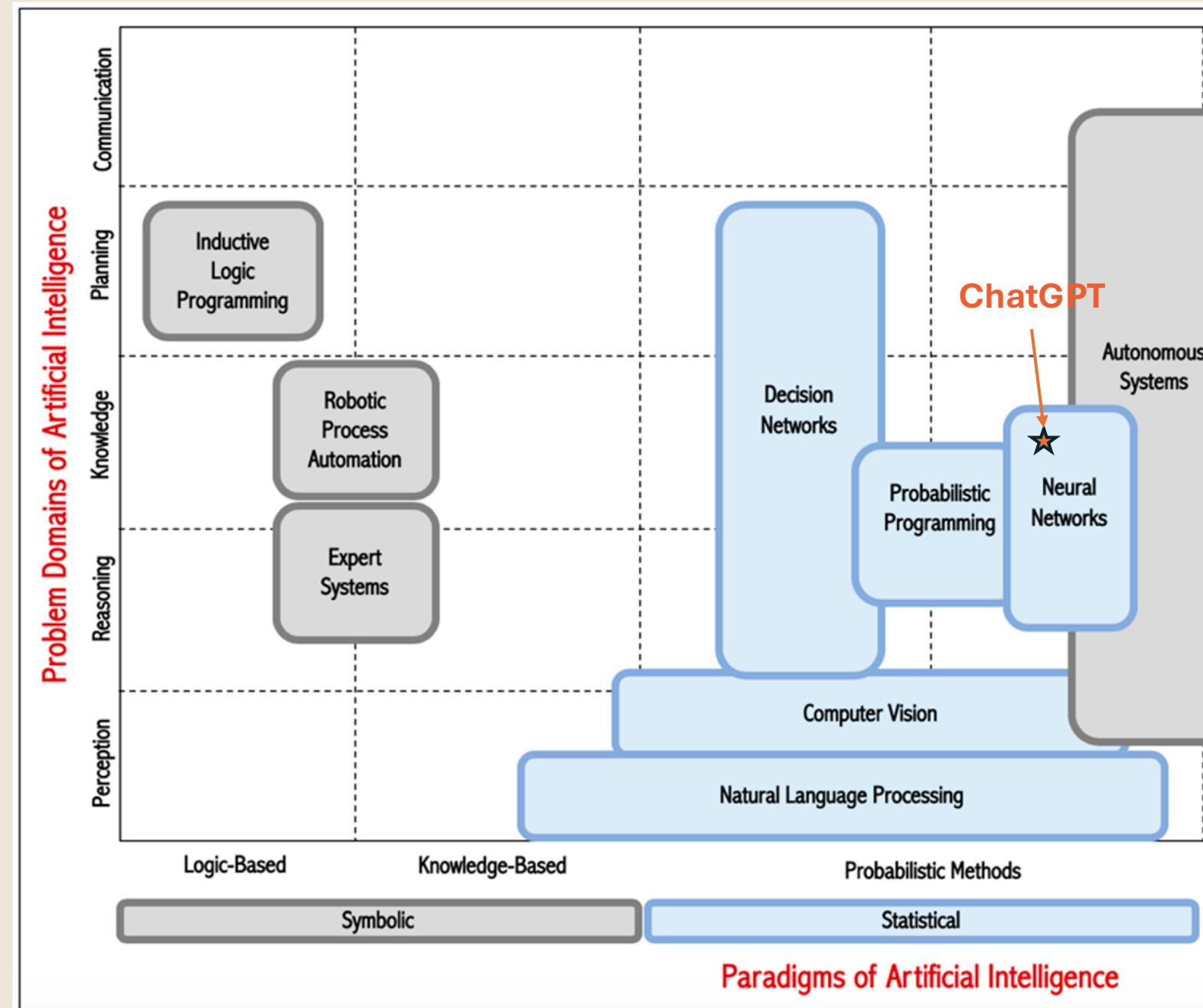
- Nest thermostat learns your routine—cuts heating bills by 10–12 %
- Smart fridges track groceries and suggest recipes from items about to expire
- Your home is collecting data to make you more comfortable—often before you know you're cold”

From Data to Intelligence

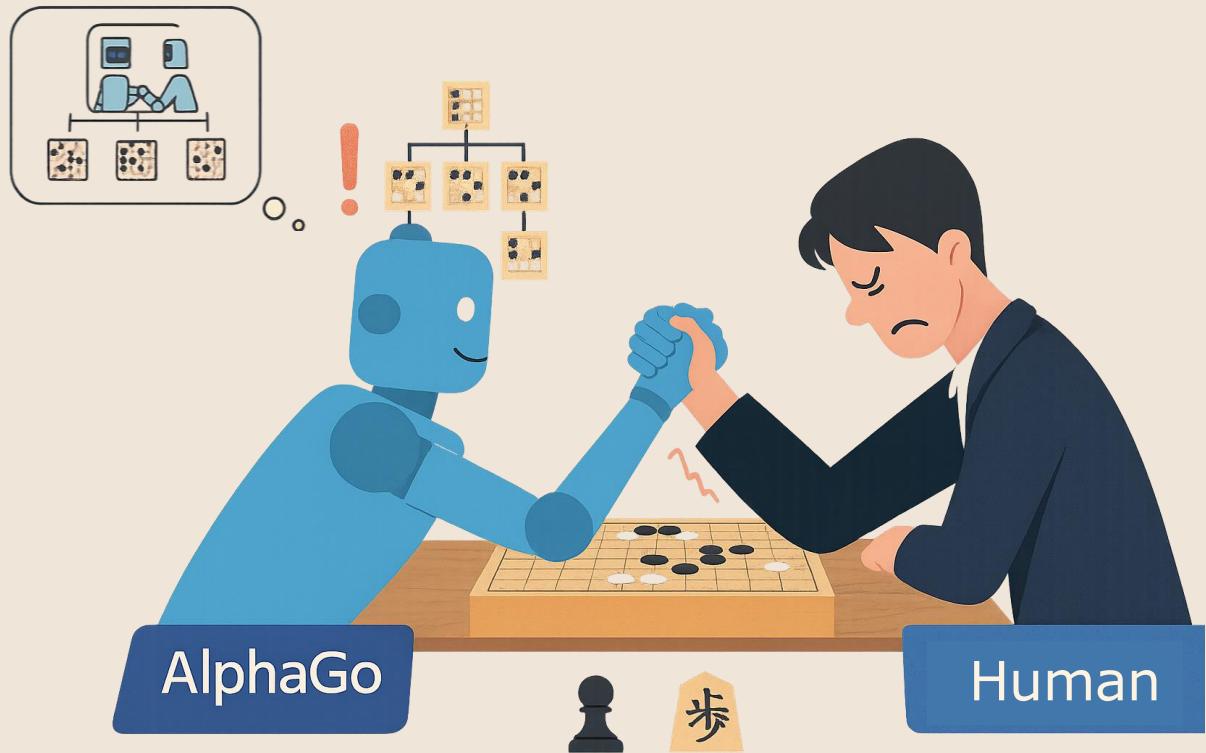
How Data is Fueling the AI Revolution

Data Powers AI

- The AI revolution lives in the realm of «statistical» AI;
- Modern AI is fueled by data, which is used to train, refine, and validate AI;
- If AI is the brain, data is the experience.

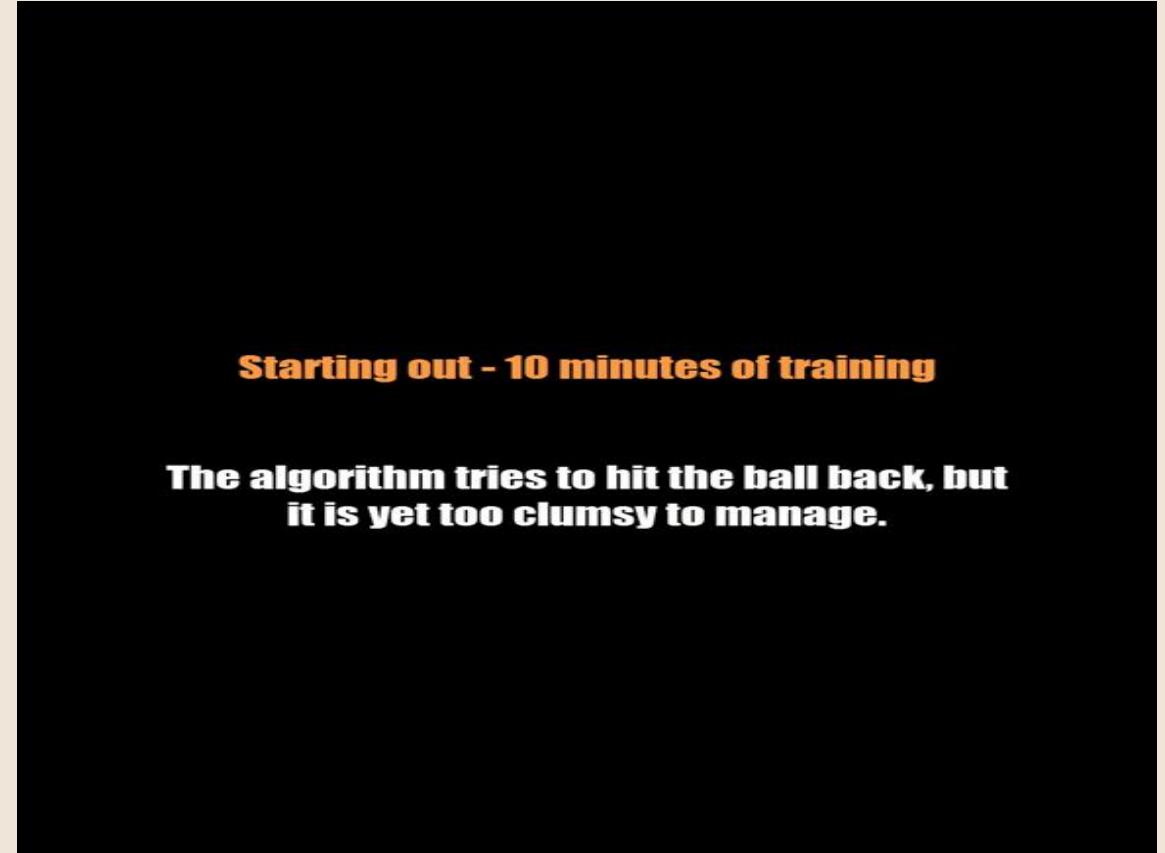


Beating the Best—Games & Self-Learning AI



- AlphaGo vs. Lee Sedol (2016): AI won 4–1 using deep learning and Monte Carlo tree search.
- Move 37 in Game 2 of AlphaGo stunned experts—an unorthodox strategy never seen before.
- AlphaZero (2017): mastered chess, shogi, Go from scratch—no human data.

Beating the Best—Games & Self-Learning AI



AlphaFold – Revolutionizing Protein Folding



- Proteins fold into unique 3D shapes essential for function—experimental methods once took years and cost ~\$120 k per structure
- AlphaFold was trained on millions of amino-acid sequences paired with thousands of experimentally solved structures from the Protein Data Bank
- Predicts structures in minutes with near-experimental accuracy
- Over 200 million structures freely available via the AlphaFold Protein Structure Database, enabling breakthroughs in biology, drug discovery, agriculture, and environmental science
- Won the Nobel Prize!

AlphaFold – Revolutionizing Protein Folding



<https://www.youtube.com/watch?v=Gk-PyJSNlcl>

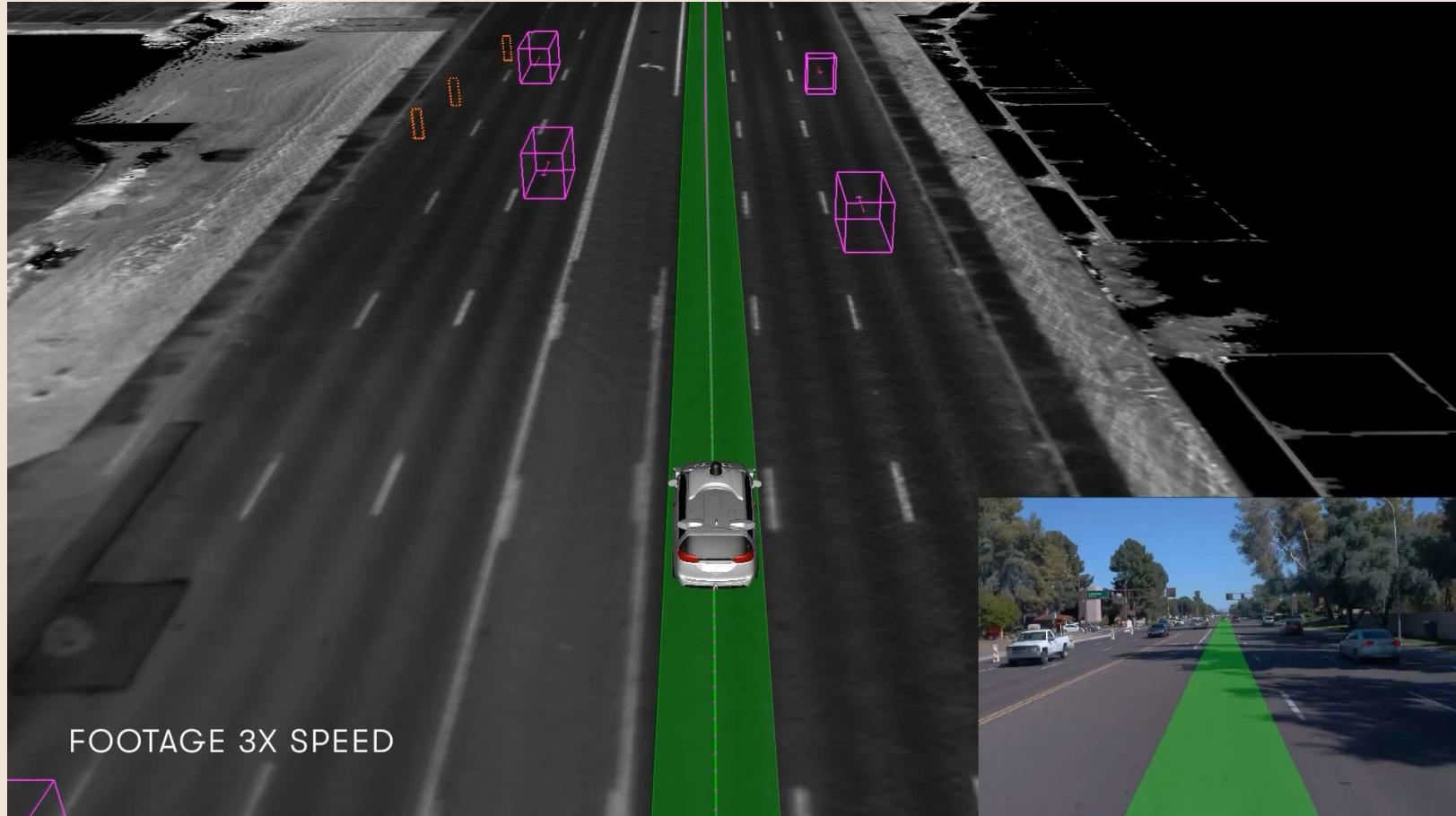
Self-Driving Cars



- Self-driving cars rely on vast sensor data—lidar, radar, cameras, GPS—to perceive their environment and predict movement.
- Algorithms trained on large datasets, including synthetic and simulated scenarios, learn to map sensory input to driving actions.
- Reinforcement learning is often used in simulation to teach agents how to drive by rewarding safe, goal-directed behavior.

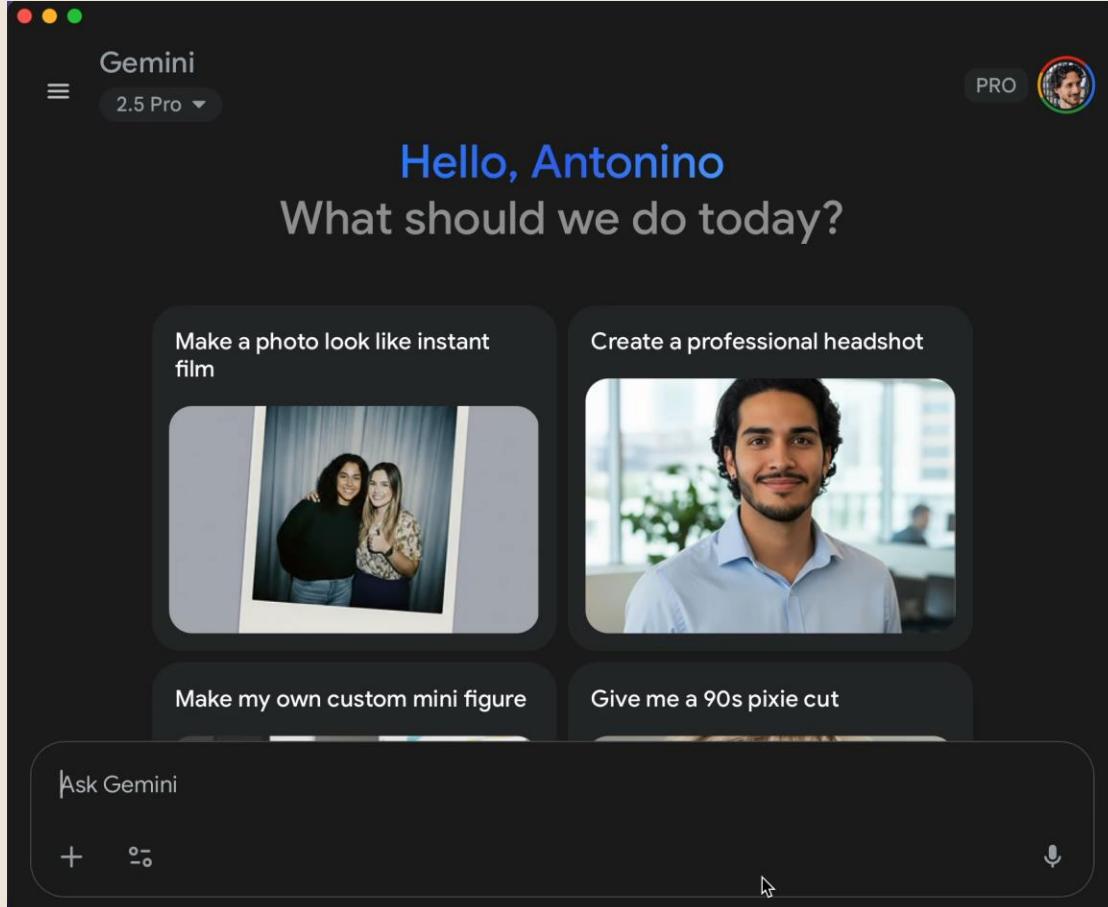


Self-Driving Cars



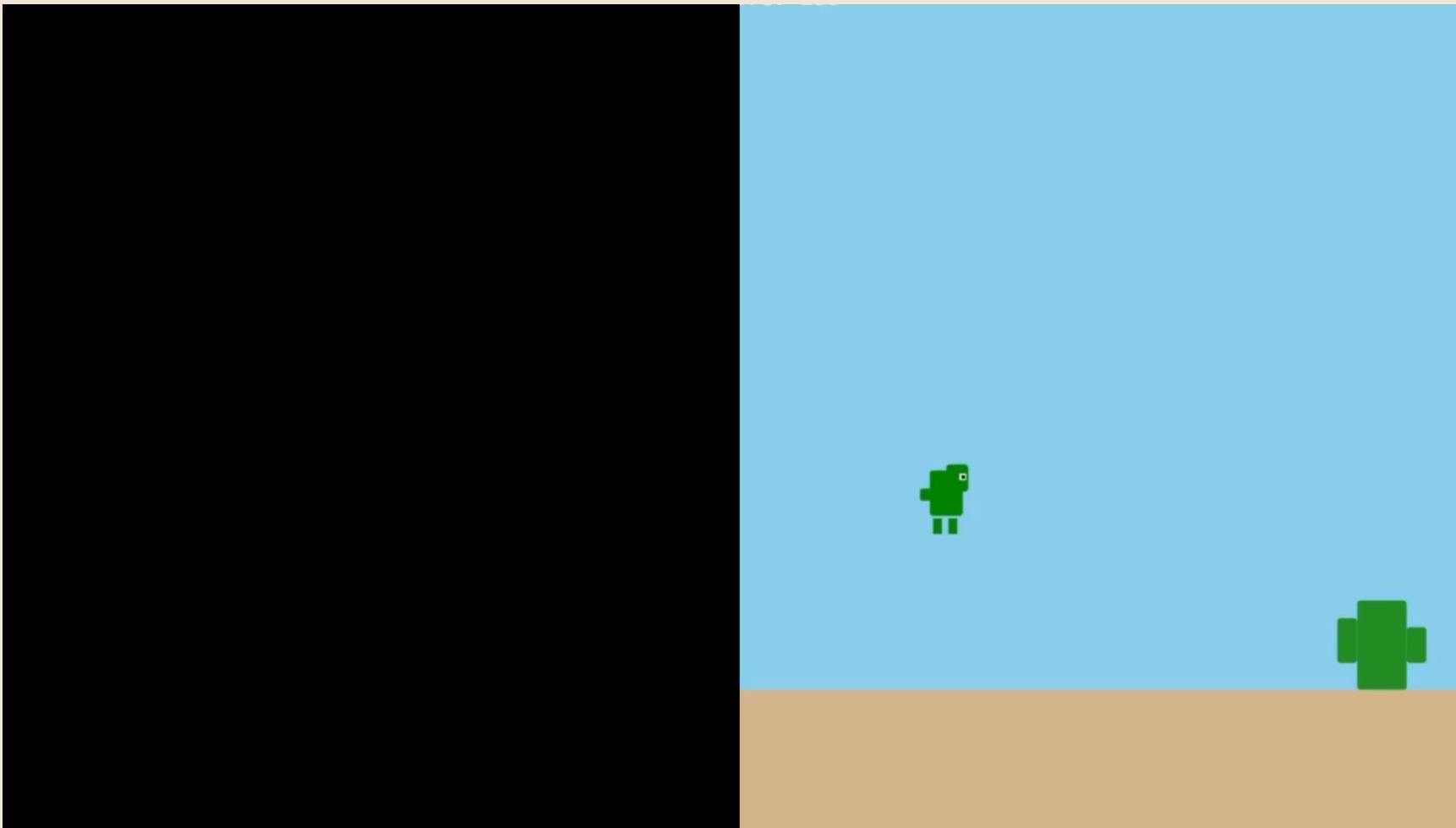
<https://www.youtube.com/watch?v=OopTOjnD3qY>

(Multimodal) Large Language Models, including ChatGPT



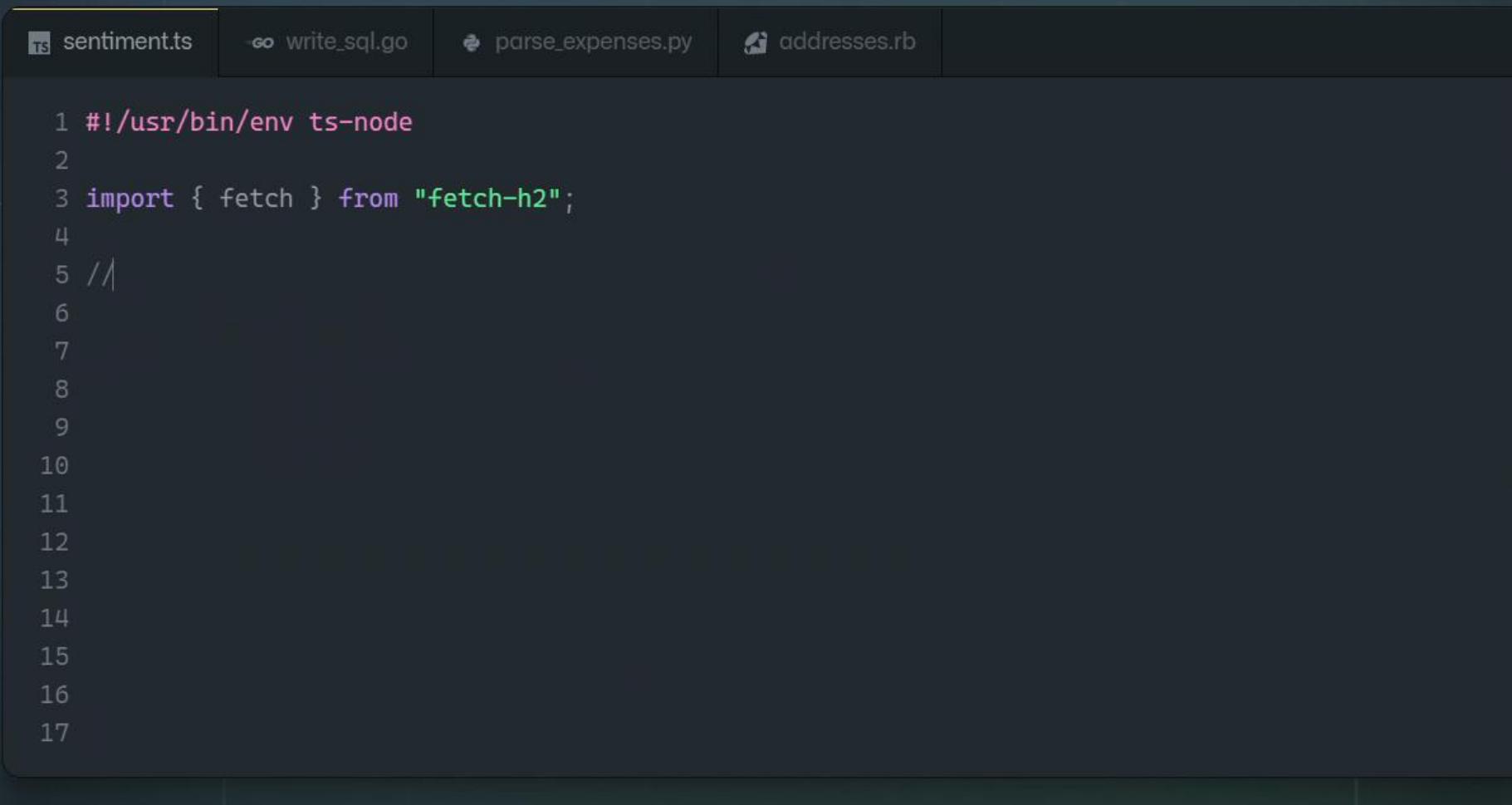
- Trained on massive datasets of text
- Predict the next word
- Learn patterns, grammar, facts, and reasoning from human language
- Can process text, images, audio, and even video
- Understand and generate across formats

Generating code based on a description



<https://www.youtube.com/watch?v=RLCBSpgos6s>

Github Copilot



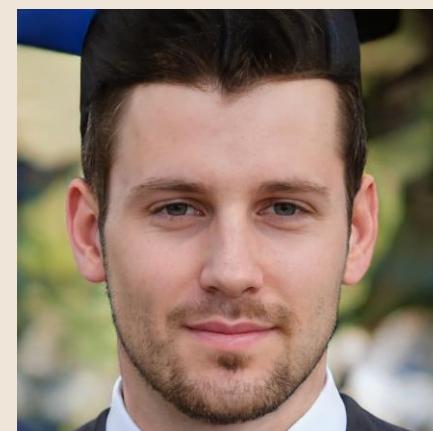
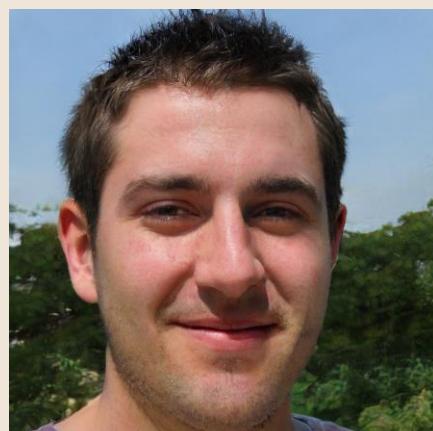
A screenshot of a code editor interface. At the top, there is a tab bar with four tabs: 'sentiment.ts' (which is currently active), 'write_sql.go', 'parse_expenses.py', and 'addresses.rb'. The main workspace shows a single line of code:

```
1 #!/usr/bin/env ts-node
```

The code is numbered from 1 to 17. Lines 1 through 16 are empty, while line 17 contains the code shown above. The background of the editor is dark.

<https://copilot.github.com/>

Impact - Generating Faces



DeepFake



https://www.youtube.com/watch?v=z7e08rWpGHY&ab_channel=VFXChrisUme

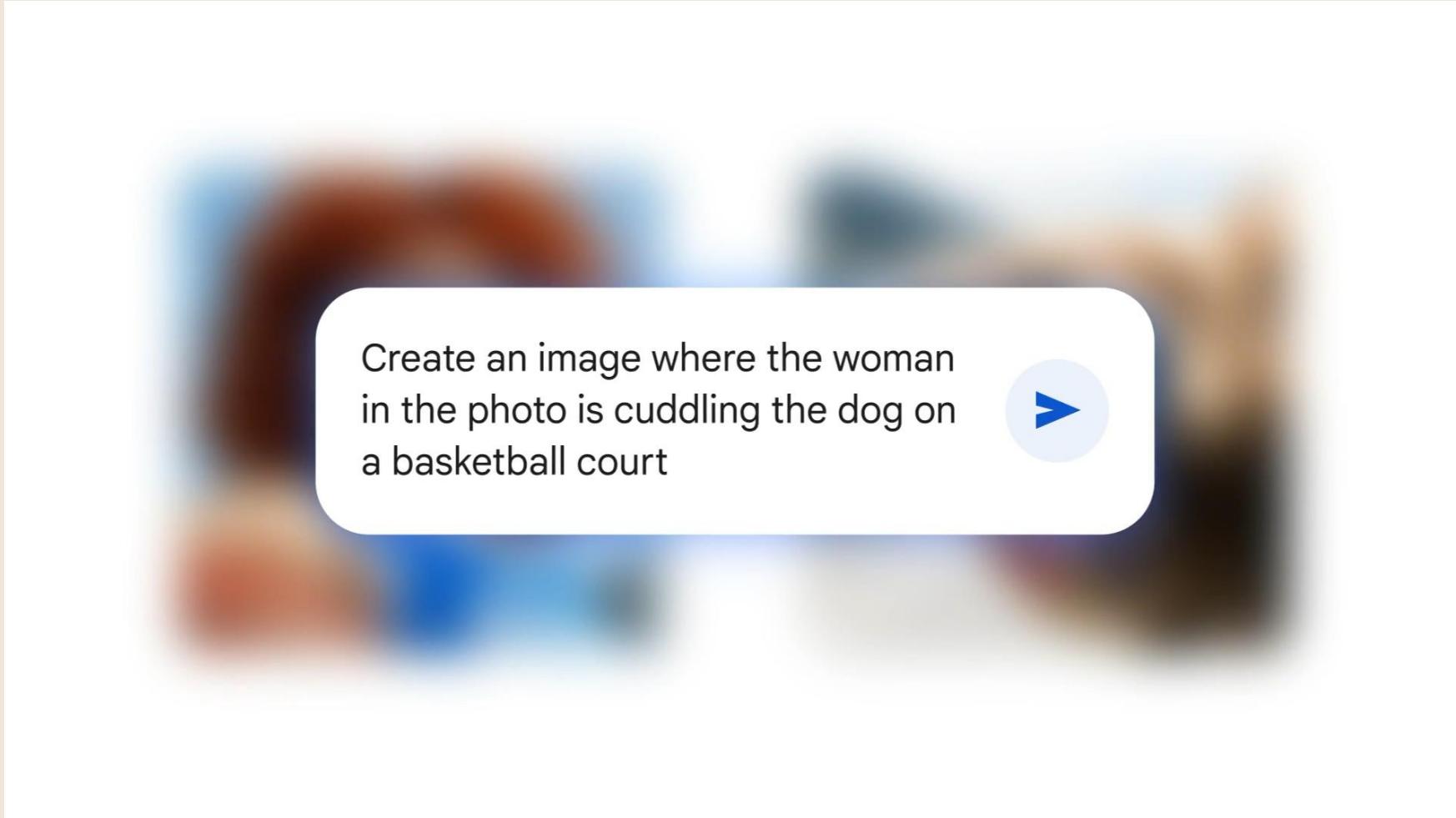
Image Generation – Imagen (Google Deepmind)



Create a cinematic, photorealistic medium shot capturing the nostalgic warmth of a late 90s indie film. The focus is a young woman with brightly dyed pink hair (slightly faded) and freckled skin, looking directly and intently into the camera lens with a hopeful yet slightly uncertain smile. She wears an oversized, vintage band t-shirt (slightly worn) over a long-sleeved striped top and simple silver stud earrings. The lighting is soft, golden hour sunlight streaming through a slightly dusty window, creating lens flare and illuminating dust motes in the air. The background shows a blurred, cluttered bedroom with posters on the wall and fairy lights, rendered with a shallow depth of field. Natural film grain, a warm, slightly muted color palette, and sharp focus on her expressive eyes enhance the intimate, authentic feel.

<https://deepmind.google/models/imagen/>

Image Editing – Nanobanana (Google)



<https://gemini.google/overview/image-generation/>

Video Generation – Veo3



A medium shot frames an old sailor, his knitted blue sailor hat casting a shadow over his eyes, a thick grey beard obscuring his chin. He holds his pipe in one hand, gesturing with it towards the churning, grey sea beyond the ship's railing. "This ocean, it's a force, a wild, untamed might. And she commands your awe, with every breaking light"

Ethics, Data Science, and Artificial Intelligence

- The great power arising from Artificial Intelligence and Data Science brings several ethics concerns;
- Like everything so powerful, these tools can be used in unethical ways;
- We won't delve into this discussion, but data scientists need to be aware of ethics;
- The one on the right is a good summer read on the topic.



See “The great hack” documentary on Netflix

Ethics & Responsibility— Cautionary Tale

- Cambridge Analytica claimed to use data to change audience behavior;
- Through the analysis of such data, they knew what kind of message every person was susceptible to;
- They could hence create messages influence people's opinion and change the way they voted;
- In 2018 Cambridge Analytica has been accused to have used illegal ways to harvest personal data from Facebook;
- They have been accused to use this data in Trump's presidential campaign and Brexit campaign.



How to Get There?



- In order to get there, we need to learn the **language of data**
- Data is never «given» and we need to learn how to transform into actionable feedback and models
- To do so, we need to understand both **fundamental concepts** and **advanced AI algorithms**
- Put simply, a **data scientist** is a professional who is **proficient in the language of data**

The Sexiest Job of the 21st Century



- Coined by Harvard Business Review in 2012
- Fusion of statistics, programming & domain expertise
- Powers decisions in healthcare, finance, marketing & AI
- Projected >30% job growth; median salaries > \$100 K
- High autonomy, cross-disciplinary teamwork & real-world impact

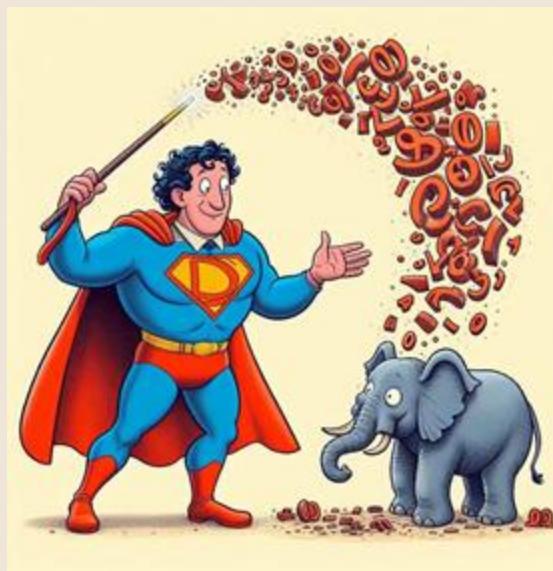
What's in a Data Scientist's Toolbox?



examining data



cleaning data

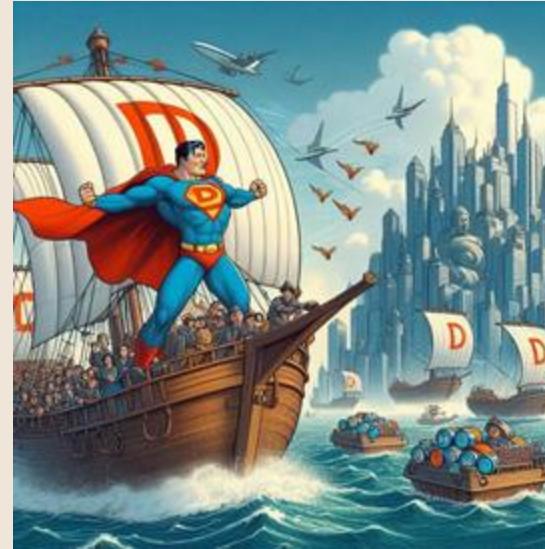


transforming data

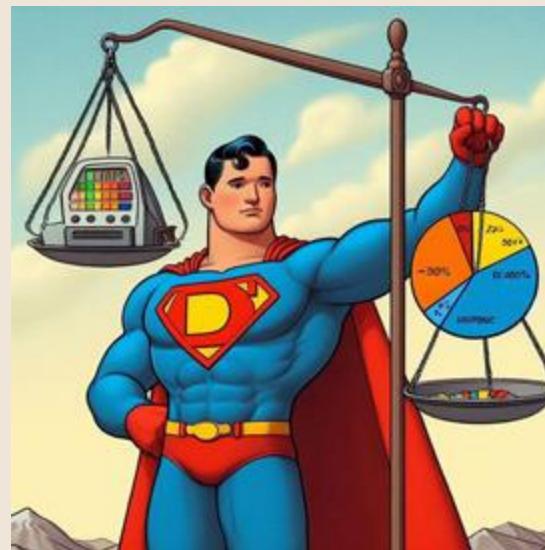


modeling data

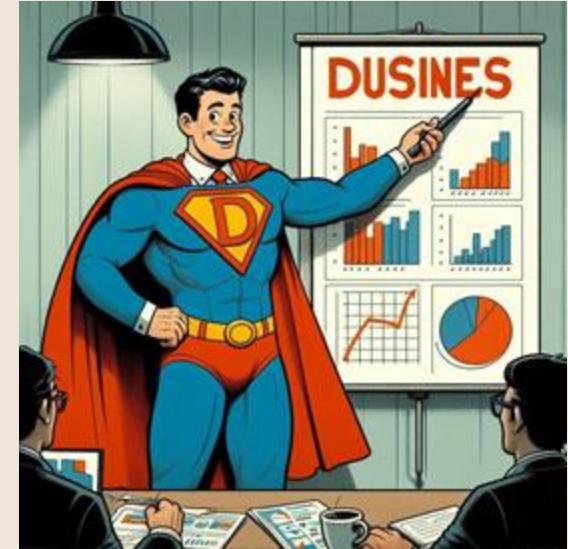
The Data Scientist's Superpowers



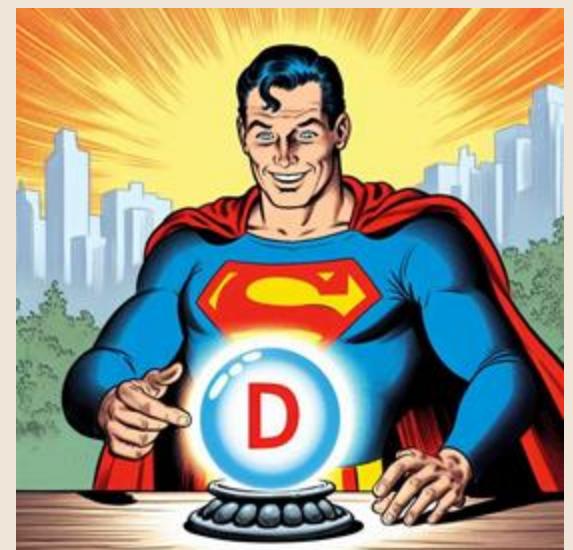
discover information



make decisions



draw conclusions



build predictive models

Art or Science? (Maybe Both)



More than a **technique** or an **algorithm**, data analysis is a **process** requiring a series of steps, not always in the same order, to answer some data **analysis questions**.

It requires the knowledge of **theory behind algorithms**, some **programming skills**, and a certain degree of **intuition**, that is **developed through practice**.

Course Organization

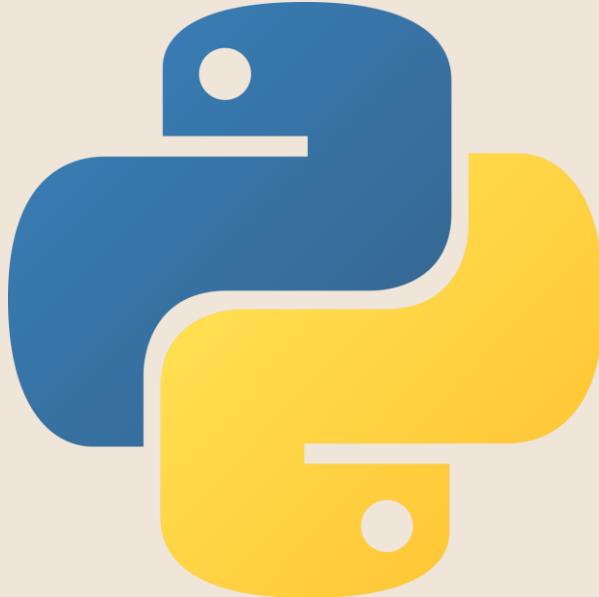
Two modules:

- **Theory** (6 CFU)
- **Laboratory** (3 CFU)

One kind of lecture:

Lectures will encompass both theoretical and practical aspects. In practice, each lecture will «switch» from a digital board to the Python interpreter to keep the course practical and interactive.

Practice in Python

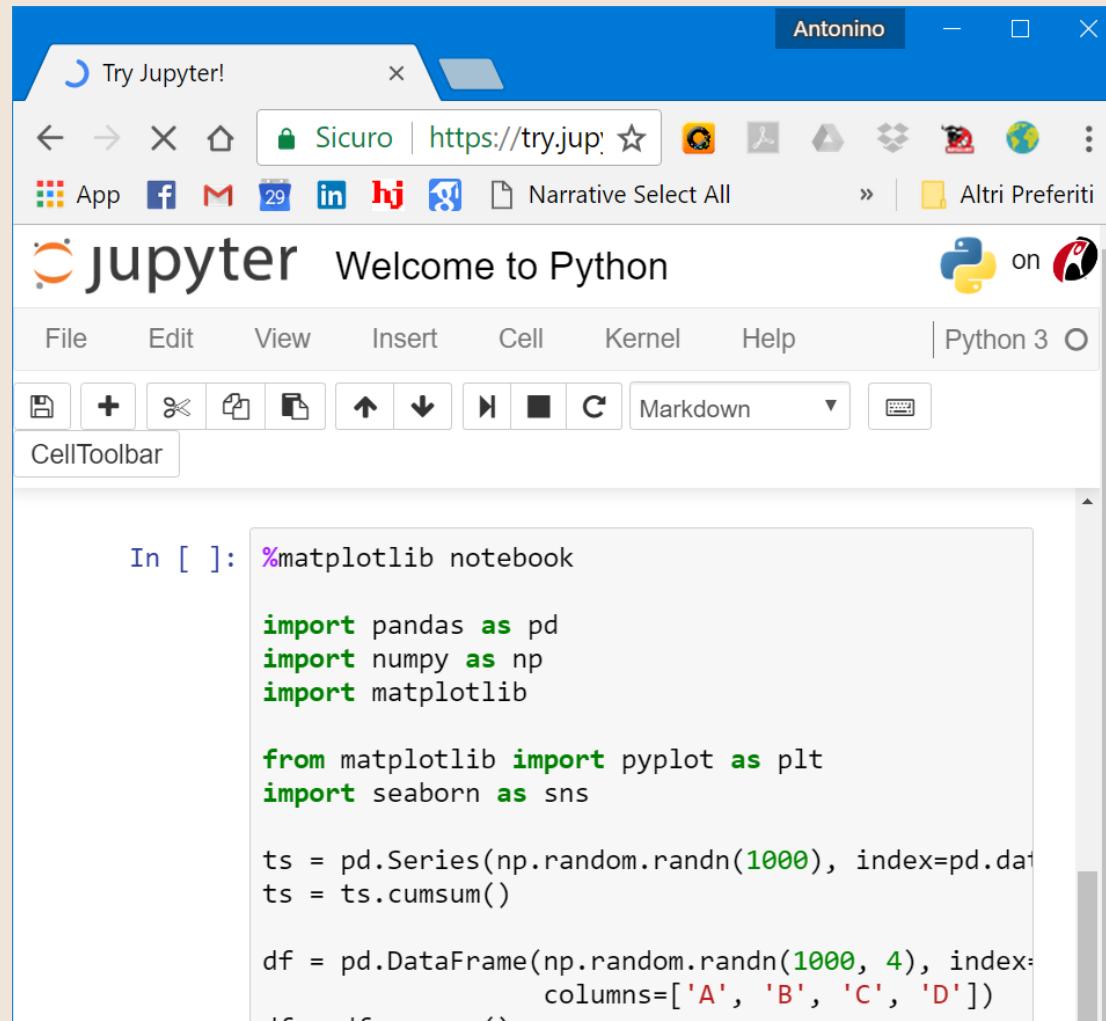


SciPy.org Sponsored By ENTHOUGHT

NumPy Base N-dimensional array package	SciPy library Fundamental library for scientific computing	Matplotlib Comprehensive 2D Plotting
IP[y]: IPython Enhanced Interactive Console	Sympy Symbolic mathematics	pandas Data structures & analysis

We will use Python, as a modern and flexible language for data science, together with a number of standard packages. We will use the Anaconda distribution.

Jupyter Notebooks



The screenshot shows a Jupyter Notebook window titled "Antonino". The browser tab is "Try Jupyter!". The notebook interface includes a toolbar with file operations like Open, Save, and New, along with cell selection and execution controls. Below the toolbar, the title bar says "jupyter Welcome to Python" and "Python 3". The main area contains a code cell labeled "In []:" with the following Python code:

```
%matplotlib notebook

import pandas as pd
import numpy as np
import matplotlib

from matplotlib import pyplot as plt
import seaborn as sns

ts = pd.Series(np.random.randn(1000), index=pd.date_range('1/1/2013', periods=1000))
ts = ts.cumsum()

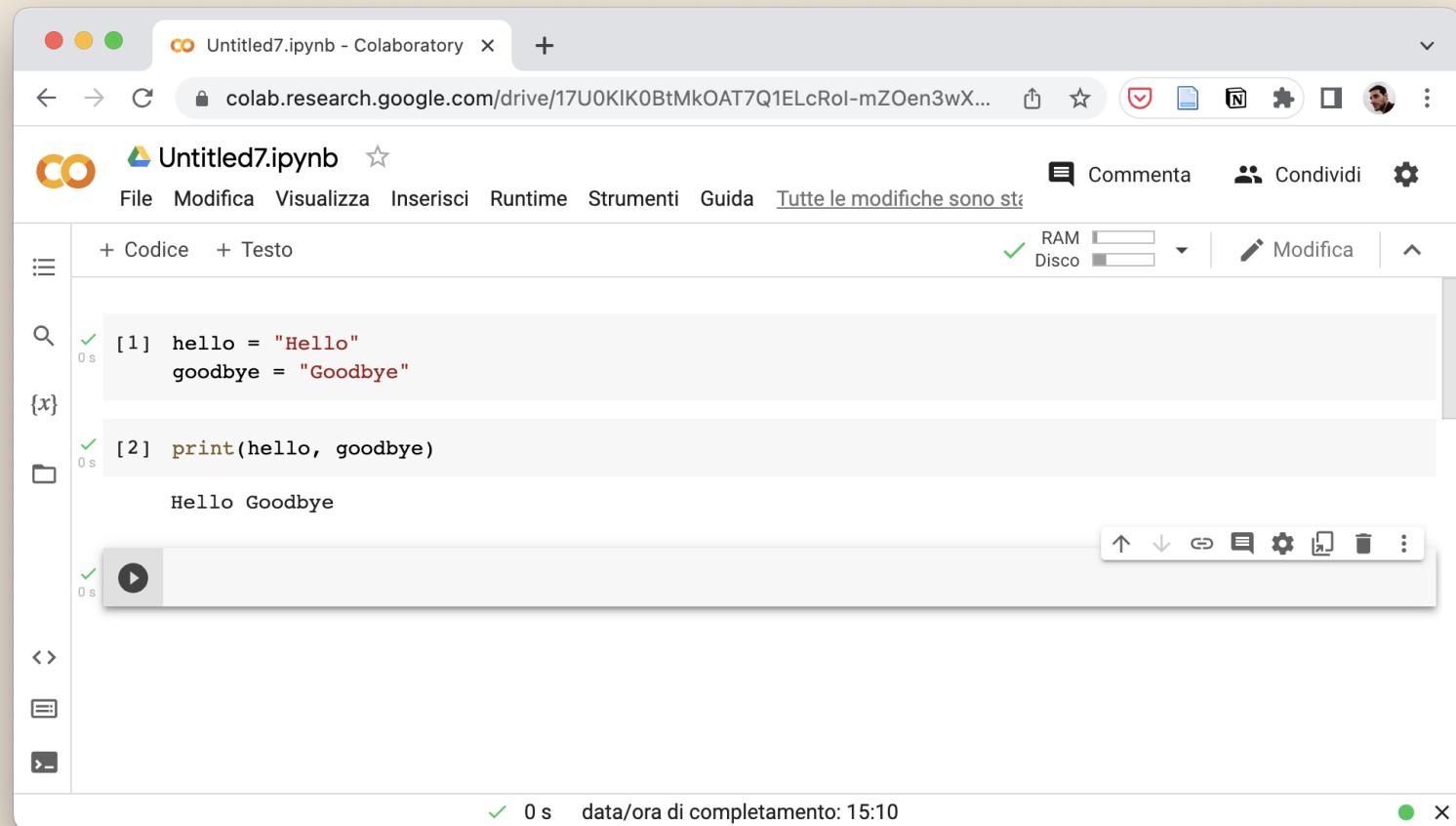
df = pd.DataFrame(np.random.randn(1000, 4), index=ts.index,
                  columns=['A', 'B', 'C', 'D'])
```

We will make extensive use of Jupyter Notebooks as they allow the creation (via a web interface) of “notebooks” containing:

- Formatted text
- Executable code
- The results (text, images) of computations

This is a very powerful tool and enables the generation of full-fledged reports of data analyses.

Google Colab



Google Colab offers a free-to-use interface to notebooks with limited computation on the cloud, which is sufficient for this course.

<https://colab.research.google.com>

Kaggle

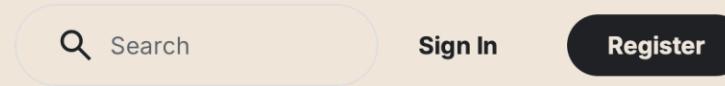
≡ kaggle

Level up with the largest AI & ML community

Join over 26M+ machine learners to share, stress test, and stay up-to-date on all the latest ML techniques and technologies. Discover a huge repository of community-published models, data & code for your next project.

 Register with Google

Register with Email



- A global platform for data science and machine learning
- Hosts competitions, datasets, and collaborative projects
- Offers free access to notebooks, GPUs, and tutorials
- Great for learning, experimenting, and showcasing your skills
- Community-driven: connect with experts, share code, and get feedback

<https://www.kaggle.com>

Python Setup Guide

The screenshot shows a web browser window with the URL antoninofurnari.github.io/fadlecturenotes2526/laboratories/01_setup.html. The page content is as follows:

Lecture Notes on Fundamentals of Data Analysis

Setup and Python Crash Course

- Introduction to the Labs and Work Environment Setup
- Python for Data Science Crash Course

Analyze the content of workspace variables.

We can launch the interactive shell with the `ipython` command:

```
(base) furnari@macbook ~ % ipython
Python 3.9.12 (main, Apr  5 2022, 01:53:17)
Type 'copyright', 'credits' or 'license' for more information
IPython 8.2.0 -- An enhanced Interactive Python. Type '?' for help.

In [1]: hello = "Goodbye"
In [2]: world = "Cruel World"
In [3]: print(hello, world)
Goodbye Cruel World

In [4]:
```

IDE

An IDE generally integrates an ipython shell and various debugging tools. It is an excellent

Contents

- Installing Python
- Programming in Python
 - Python Interpreter
 - ipython Interactive Shell**
- IDE
 - Jupyter Notebook
 - IDE vs Notebook
 - Google Colab

https://antoninofurnari.github.io/fadlecturenotes2526/laboratories/01_setup.html

Flipped Classroom: Python Crash Course

The screenshot shows a web browser window with the following details:

- Title Bar:** antoninofurnari.github.io/fadlecturenotes2526/laboratories/02_in
- Toolbar:** Back, Forward, Stop, Refresh, Home, etc.
- Header:** ExamBox, DropTheMark, Pattern Recognition AE, PAMI AE
- Left Sidebar:**
 - Lecture Notes on Fundamentals of Data Analysis
 - Setup and Python Crash Course
 - Introduction to the Labs and Work Environment Setup
 - Python for Data Science Crash Course
- Main Content Area:**
 - Section 1:** Questa stringa è formattata. Posso inserire numeri, ad esempio 3.000002
It is possible to specify the type of each argument using the colon:

```
print("This {:s} is formatted. I can insert numbers, for example {:.2f}"\
      .format("string",3.00002)) #Positional parameters, without specifying
```
 - Section 2:** Questa stringa è formattata. Posso inserire numeri, ad esempio 3.00
It's also possible to assign names to arguments so you can call them out of order:

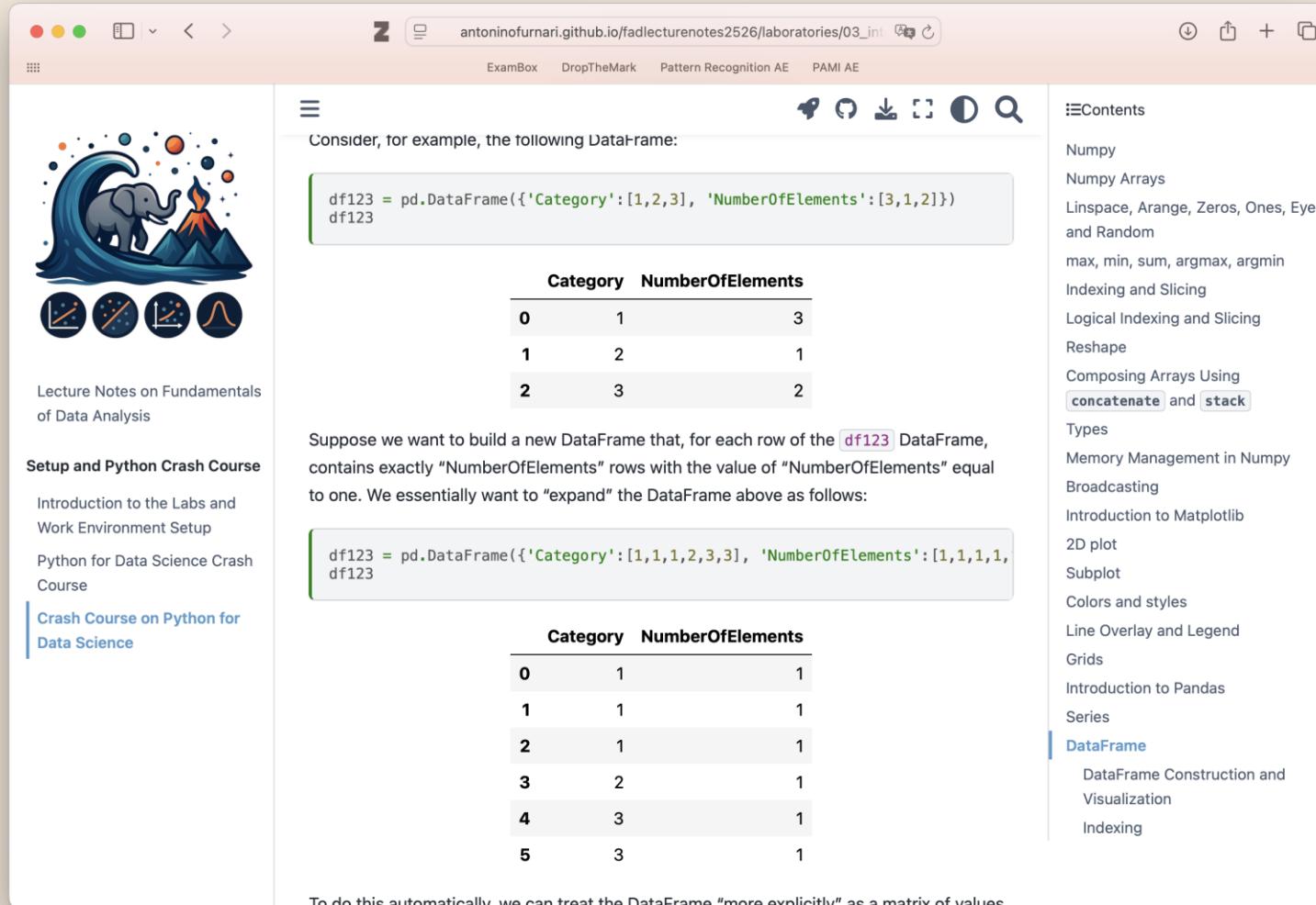
```
print("This {str:s} is formatted. I can insert numbers, for example {num:0.\
      .format(num=3.00002, str="string"))
```
 - Question 5:** Given the variables:

```
a = "hello"
b = "world"
c = 2.0```
Use string formatting to print the string `hello 2 times world`.
```
- Right Sidebar:** Contents menu with links to various Python topics.

- Crash Course on Python;
- Most of you probably don't need it!
- If you know C, you will be get up and running in little time;
- Start looking into it before the next lecture, but we'll recall important concepts as we go.

https://antoninofurnari.github.io/fadlecturenotes2526/laboratories/02_python_crash_course.html

Python For Data Science Crash Course



The screenshot shows a web browser window with the URL antoninofurnari.github.io/fadlecturenotes2526/laboratories/03_in/. The page contains a sidebar with a logo of an elephant on a wave, navigation links for 'Lecture Notes on Fundamentals of Data Analysis', 'Setup and Python Crash Course', and 'Crash Course on Python for Data Science'. The main content area displays code snippets and tables. A code snippet creates a DataFrame with 3 rows and 2 columns:

```
df123 = pd.DataFrame({'Category':[1,2,3], 'NumberOfElements':[3,1,2]})  
df123
```

A table is shown:

	Category	NumberOfElements
0	1	3
1	2	1
2	3	2

Text explains that we want to build a new DataFrame where each row has exactly one 'NumberOfElements' value of 1. A second code snippet creates a DataFrame with 6 rows and 2 columns:

```
df123 = pd.DataFrame({'Category':[1,1,1,2,3,3], 'NumberOfElements':[1,1,1,1,1,1]})  
df123
```

A second table is shown:

	Category	NumberOfElements
0	1	1
1	1	1
2	1	1
3	2	1
4	3	1
5	3	1

To do this automatically, we can treat the DataFrame "more explicitly" as a matrix of values.

The right sidebar lists contents related to DataFrames:

- Contents
- Numpy
- Numpy Arrays
- Linspace, Arange, Zeros, Ones, Eye and Random
- max, min, sum, argmax, argmin
- Indexing and Slicing
- Logical Indexing and Slicing
- Reshape
- Composing Arrays Using `concatenate` and `stack`
- Types
- Memory Management in Numpy
- Broadcasting
- Introduction to Matplotlib
- 2D plot
- Subplot
- Colors and styles
- Line Overlay and Legend
- Grids
- Introduction to Pandas
- Series
- DataFrame
- DataFrame Construction and Visualization
- Indexing

- Crash Course on Python for data science;
- Many of you may need it!
- Also for this, start looking into it before the next lecture, but we'll recall important concepts as we go.

Syllabus

Three Main modules:

Data Analysis

Understanding your data



Description & Visualization



Correlation Analysis



Linear & Logistic
Regression



Statistical Tests

Predictive Techniques

Using data to make predictions



Classification & Regression



Overfitting & Regularization



Evaluation Measures



Model Selection

Data Representation

Revealing structure in the data



Data Representation &
Feature Extraction



Clustering &
Density Estimation



Dimensionality Reduction



Data Interpretation

Skills and Career Paths

Skills You Will Learn

Theoretical and practical data analysis skills

-  Python for Data Science
-  Data Acquisition & Management
-  Data Cleaning & Preprocessing
-  Exploratory Data Analysis
-  Data Visualization
-  Predictive Modeling

Career Paths

The course provides the foundations to become

-  Data Analyst
-  Business Intelligence (BI) Analyst
-  Data Scientist
-  Research Scientist
-  Analytics Consultant
-  Machine Learning Engineer



2025/2026

(triennale)



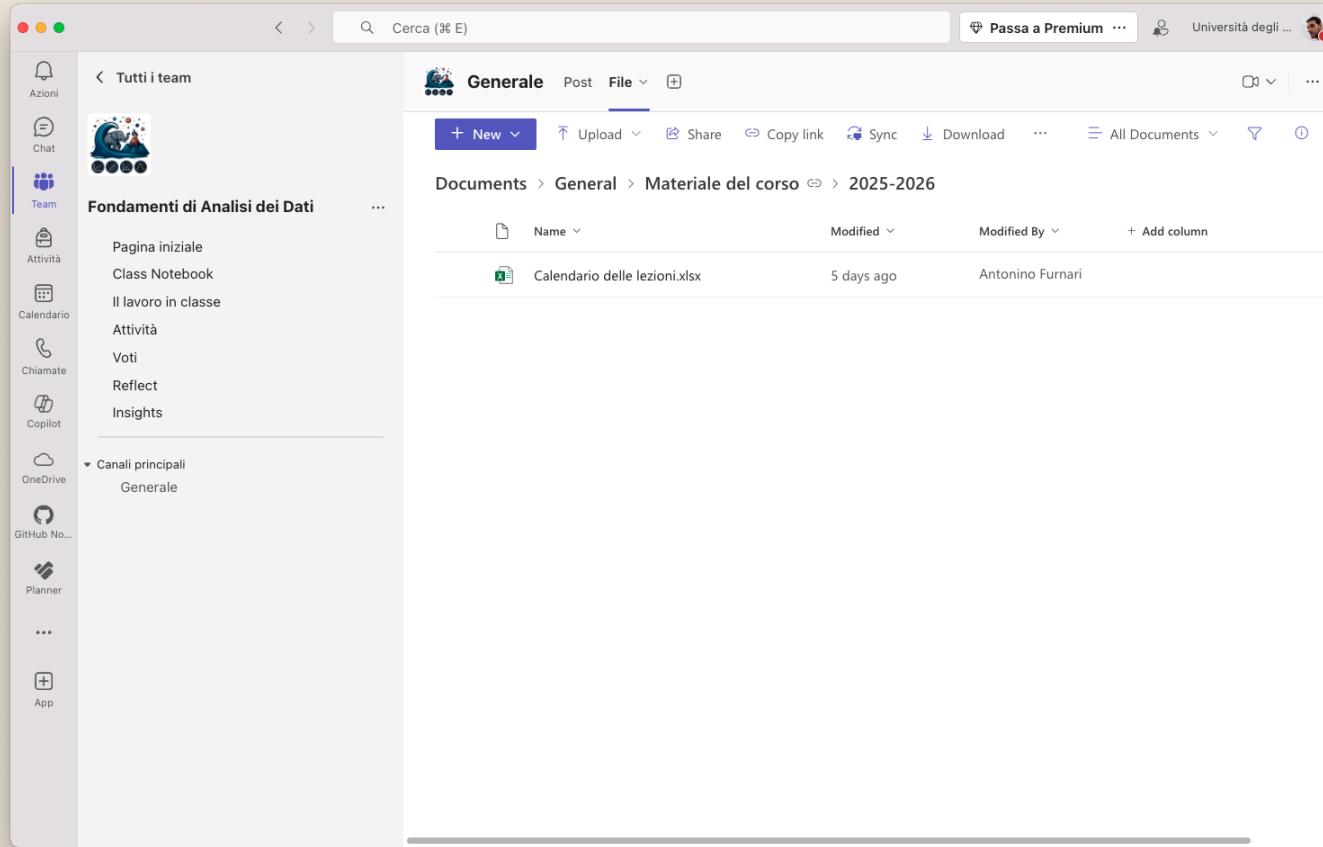
2023-2025

(magistrale)

Updated Syllabus

- Bilanciato il carico per la transizione
- Taglio maggiormente pratico con laboratorio e teoria affrontati in contemporanea
- Prove in itinere e modalità di esame seguono il taglio maggiormente pratico

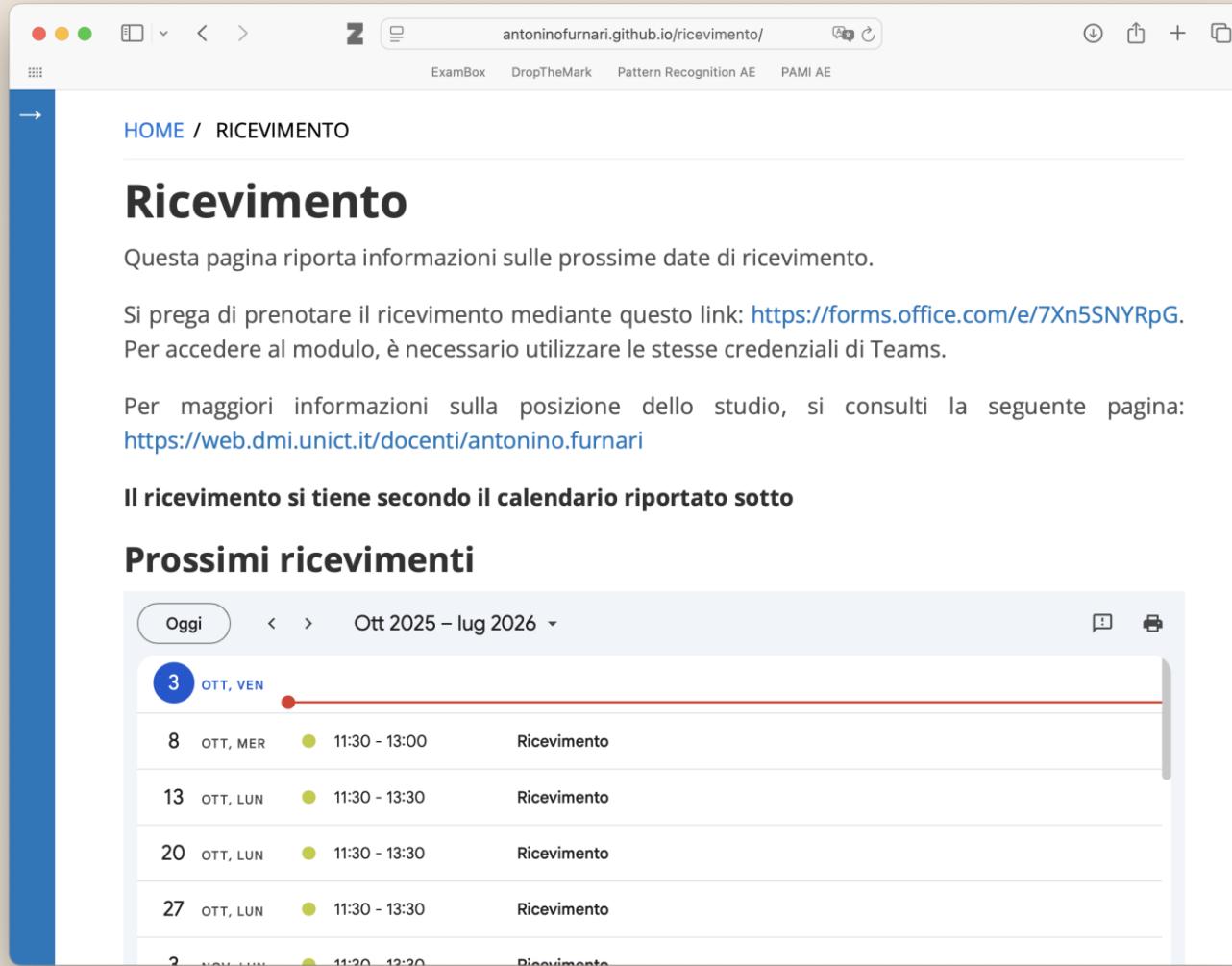
Communication



Announcements and pointers to teaching material will be given through Microsoft Teams.

Team code: **i87g4nb**

Office Hours – Online Calendar



The screenshot shows a web browser window with the URL antoninofurnari.github.io/ricevimento/. The page title is "Ricevimento". The content includes instructions for booking via a Microsoft Forms link (<https://forms.office.com/e/7Xn5SNYRpG>) and provides a link to the studio's location (<https://web.dmi.unict.it/docenti/antonino.furnari>). It also states that "Il ricevimento si tiene secondo il calendario riportato sotto". Below this, a "Prossimi ricevimenti" section displays a calendar for October 2025 to July 2026. The calendar shows weekly slots from Monday to Friday, with specific days highlighted in blue. A red dot on the 3rd of October indicates a reception. The following days show receptions on Monday, October 8th, 13th, 20th, and 27th.

Data	Giorno	Ora	Descrizione
3	OTT, VEN	11:30 - 13:00	Ricevimento
8	OTT, MER	11:30 - 13:00	Ricevimento
13	OTT, LUN	11:30 - 13:30	Ricevimento
20	OTT, LUN	11:30 - 13:30	Ricevimento
27	OTT, LUN	11:30 - 13:30	Ricevimento
3	NOV, LUN	11:30 - 13:30	Ricevimento

- Generally on Mondays 11.30 – 13.30
- Please book in advance, so I know I should wait for you
- Check the online page for variations: <https://antoninofurnari.github.io/ricevimento/>

Tutor for the Course



Rosario Forte

E-mail: rosario.forte@phd.unict.it

- Assistance in class when assigning and revising projects
- Coaching for projects on office hours (will write on the Team)

E-Mails

- Students can communicate with the teacher by sending an email to antonino.furnari@unict.it
- Some simple rules to follow when sending an email:
 - Use the studium address provided by the University. This is to minimize the likelihood of your email ending up in spam;
 - Always sign at the end of the email, so that I can know who is contacting me;
 - Always identify yourself in relation to the course (e.g. "I am a student of Fundamentals of Data Analysis 2025/2026");
 - Reread the email before sending it.

Teaching Material: Notes

Lecture Notes on Fundamental of Data Analysis

Theory

1. Introduction to Data Analysis
2. Main data analysis concepts
3. Misure di Frequenze e Rappresentazione Grafica dei Dati
4. Misure di Tendenza Centrale, Dispersione e Forma
5. Associazione tra Variabili
6. Probability for Data Manipulation
7. Common Probability Distributions

RSS = $(y_1 - \hat{\beta}_0 - \hat{\beta}_1 x_1)^2 + (y_2 - \hat{\beta}_0 - \hat{\beta}_1 x_2)^2 + \dots + (y_n - \hat{\beta}_0 - \hat{\beta}_1 x_n)^2$

This number will be the sum of the square values of the dashed segments in the plot below:

Linear Regression with Residual Line Segments

Intuitively, if we minimize these numbers, we will find the line which **best fits the data**.

We can obtain estimates for $\hat{\beta}_0$ and $\hat{\beta}_1$ by minimizing the RSS using an approach called **ordinary least squares**.

We can write the RSS as a function of the parameters to estimate:

$$RSS(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2$$

This is also called a **cost function** or **loss function**.

We aim to find:

antoninofurnari / fadlecturenotes

Code Issues Pull requests Actions Projects Security Insights Settings

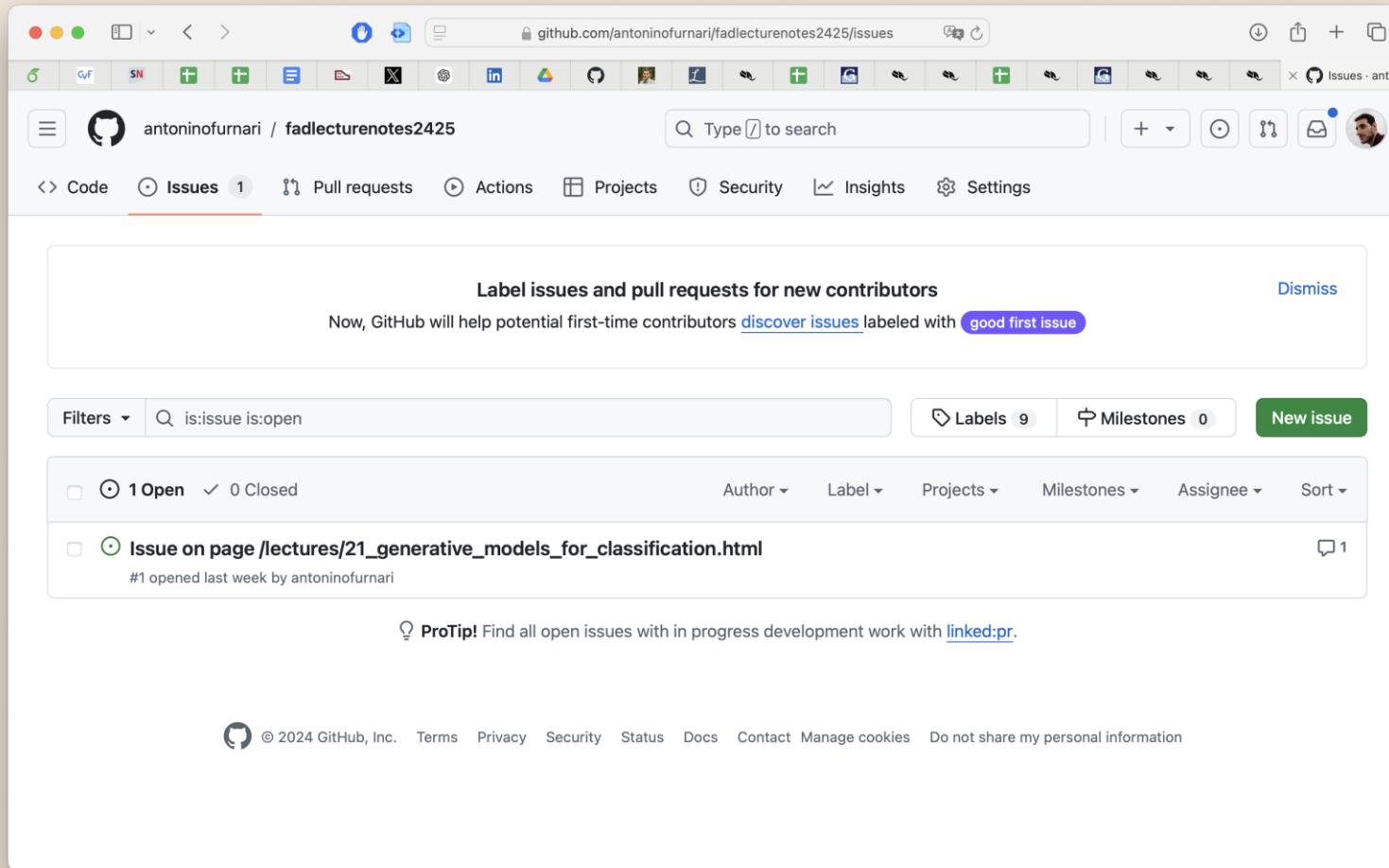
master / fadlecturenotes / lecturenotes /

antoninofurnari fix errors · aea1c9c · 5 months ago

Name	Last commit message	Last commit date
..		
_bibliography	rename book	9 months ago
_static/lecture_specific	fix mistakes	5 months ago
laboratories	fix errors	5 months ago
lectures	fix errors	5 months ago
slides	fix figures	8 months ago
_config.yml	refactoring	8 months ago
_toc.yml	add lab on classification	5 months ago
favicon.ico	updates	9 months ago
logo.png	updating many files	6 months ago

<https://antoninofurnari.github.io/fadlecturenotes2526/> OPEN SOURCE!

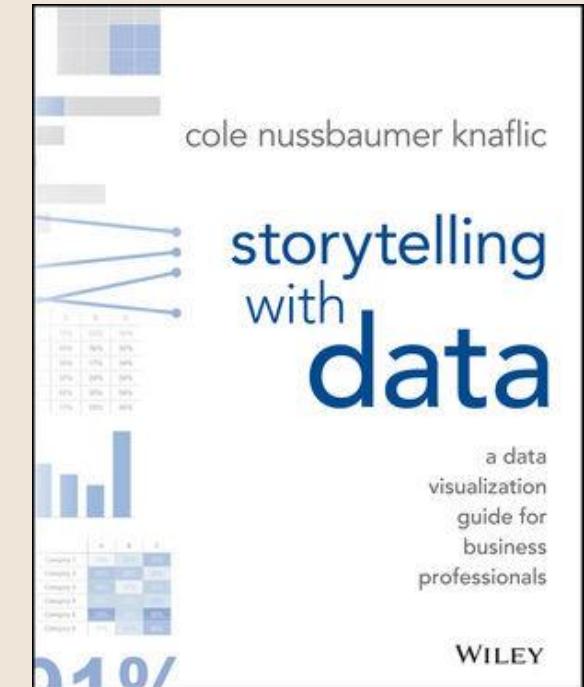
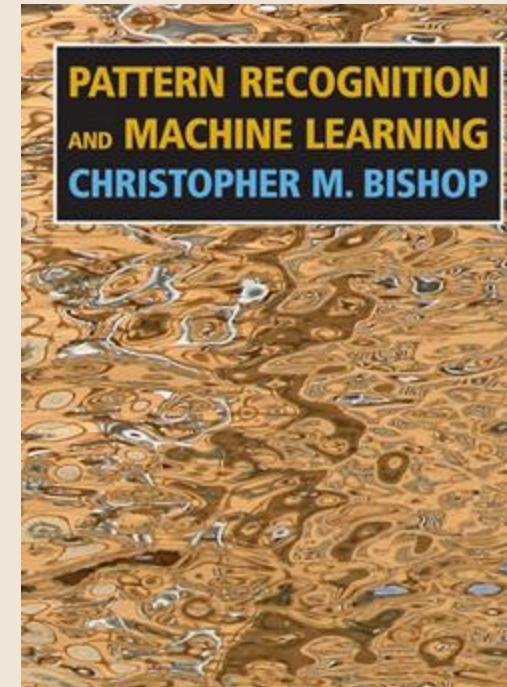
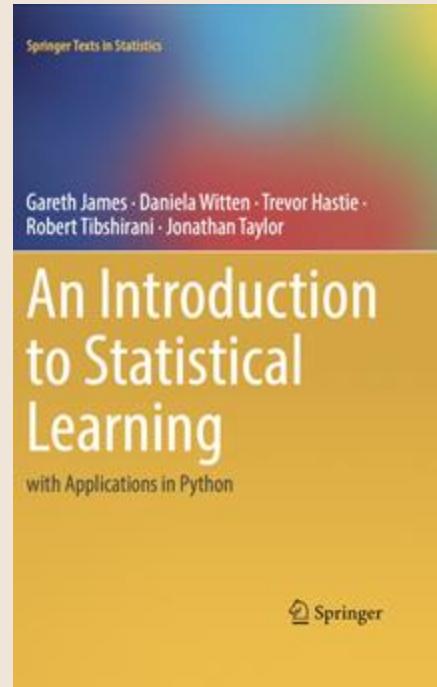
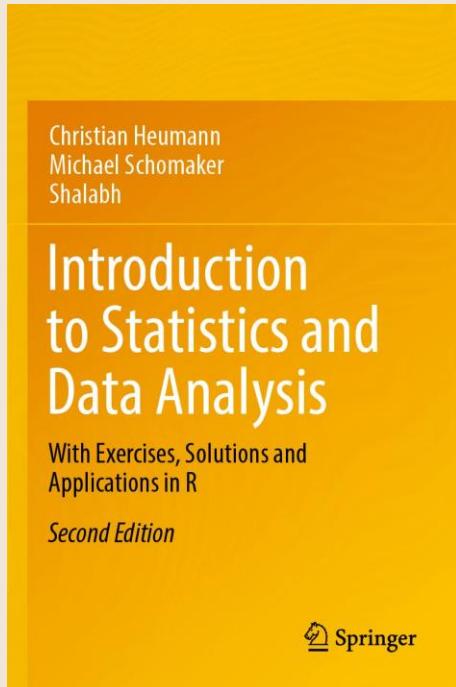
Teaching Material: Welcome to contribute!



If you want to notify a mistake, you can open an issue (or even better a pull request!)

<https://antoninofurnari.github.io/fadlecturenotes2526/> OPEN SOURCE!

Teaching Material: Pointers to Book Chapters



- Peck, Roxy, Chris Olsen, and Jay L. Devore. *Introduction to statistics and data analysis*. Cengage Learning, 2015.
- James, Gareth Michael. *An introduction to statistical learning: with applications in Python*, 2023. <https://www.statlearning.com>
- Bishop, Christopher M. "Machine Learning". *Machine learning*, 2006. <https://www.microsoft.com/en-us/research/publication/pattern-recognition-machine-learning/>
- Knaflic, Cole Nussbaumer. *Storytelling with data: A data visualization guide for business professionals*. John Wiley & Sons, 2025.

Esami: Due parti

Teoria (40% del voto finale)

- Un test scritto a scelta multipla per verificare la conoscenza dei concetti di base e della teoria
- Sebbene il corso sia progettato con un orientamento pratico, una buona data science richiede una solida conoscenza degli aspetti metodologici
- È previsto il rilascio di una piattaforma online per prepararsi a questo test
- La valutazione sarà espressa in trentesimi

Pratica (60% del voto finale)

- Un progetto assegnato dal docente, da svolgere in gruppi da 1 a 3 studenti
- Il progetto prevede l'analisi di un dataset, inclusa la progettazione e la verifica di algoritmi predittivi
- Il progetto viene presentato tramite una presentazione PowerPoint
- Il codice viene inviato al docente
- La valutazione è espressa in trentesimi

Il voto finale è una media pesata dei due voti.

Esami: Due Percorsi

Prove in itinere



Solo esame finale



In-Itinere



Teoria (40% del voto finale)

- Tre prove in itinere durante il corso, ciascuna relativa a uno dei tre moduli principali
- Ogni prova contiene 22 domande da completare in 45 minuti
- Ogni prova vale 11 punti
- Il totale è 33 punti, arrotondato a 30 (quindi gli studenti hanno 3 punti extra)

Pratica (60% del voto finale)

- Il progetto viene assegnato il giorno della prima prova in itinere
- Gli studenti lavorano in aula sotto la supervisione del docente e del tutor
- Il lavoro prosegue a casa
- La seconda e la terza parte del progetto vengono assegnate durante la seconda e la terza prova in itinere
- Il progetto viene presentato in un giorno precedente al primo appello d'esame (che verrà comunicato per tempo)

Solo Esame Finale



Teoria (40% del voto finale)

- Una prova scritta con 30 domande da completare in un'ora
- Ogni prova vale un totale di 30 punti

Pratica (60% del voto finale)

- Il progetto è assegnato agli studenti dal docente
- Gli studenti lavorano in autonomia e possono chiedere una revisione al docente
- Il progetto viene presentato e valutato il giorno dell'esame
- Il codice viene inviato al docente
- Il progetto vale 30 punti

Timeline e Dettagli

In itinere:

- È necessario completare teoria e pratica secondo la timeline prevista
- Gli studenti devono prenotarsi per il primo appello disponibile almeno 48 ore prima per la registrazione

Solo esame finale:

- Le due parti possono essere sostenute in **sessioni diverse**, ma dopo la prova scritta l'esame deve essere completato entro dicembre dello stesso anno solare
- Gli studenti devono prenotarsi per l'esame almeno 48 ore prima
- Se l'esame non viene completato entro la data del prossimo appello, verrà registrato come «ritirato» o «assente» a seconda del caso, e sarà necessaria una nuova prenotazione

Schedule of the Lectures (on Teams)

Numero	Giorno	Giorno della settimana	Aula	Orario	Argomenti della lezione
1	06/10/25	Lunedì	24	08:00 - 11:00	Introduzione al corso
2	08/10/25	Mercoledì	24	08:00 - 11:00	Introduzione a Python e lo stack per la data science. Concetti principali di data science, ottenere i dati
3	13/10/25	Lunedì	24	08:00 - 11:00	Descrivere e visualizzare i dati
4	14/10/25	Martedì	24	14:00 - 17:00	Probabilità per l'analisi dei dati
5	15/10/25	Mercoledì	24	08:00 - 11:00	Associazione tra variabili
6	27/10/25	Lunedì	24	08:00 - 11:00	Distribuzioni di dati
7	28/10/25	Martedì	24	14:00 - 17:00	Inferenza statistica per l'analisi dei dati



Fondamenti di Analisi dei Dati

9 CFU

from **data analysis** to **predictive techniques**

Prof. Antonino Furnari (antonino.furnari@unict.it)

Corso di Studi in Informatica

Dip. di Matematica e Informatica

Università di Catania



Università
di Catania

Data Analysis

Understanding your data

- Description & Visualization
- Correlation Analysis
- Linear & Logistic Regression
- Statistical Tests

Predictive Techniques

Using data to make predictions

- Classification & Regression
- Overfitting & Regularization
- Evaluation Measures
- Model Selection

Data Representation

Revealing structure in the data

- Data Representation & Feature Extraction
- Clustering & Density Estimation
- Dimensionality Reduction
- Data Interpretation

Skills You Will Learn

Theoretical and practical data analysis skills

- Python for Data Science
- Data Acquisition & Management
- Data Cleaning & Preprocessing
- Exploratory Data Analysis
- Data Visualization
- Predictive Modeling

Career Paths

The course provides the foundations to become

- Data Analyst
- Business Intelligence (BI) Analyst
- Data Scientist
- Research Scientist
- Analytics Consultant
- Machine Learning Engineer

Lectures
72 hours of frontal lectures covering formal concepts and data analysis examples in Python

Exams
Two paths: in itinere tests and project and standalone exam with written text and project

In itinere tests + **Guided group project** OR **Written exam** + **Individual project**

Theses & Internships

Info & Notes
open source on github

More info & Notes open source on GitHub



<https://antoninofurnari.github.io/fadlecturenotes/>

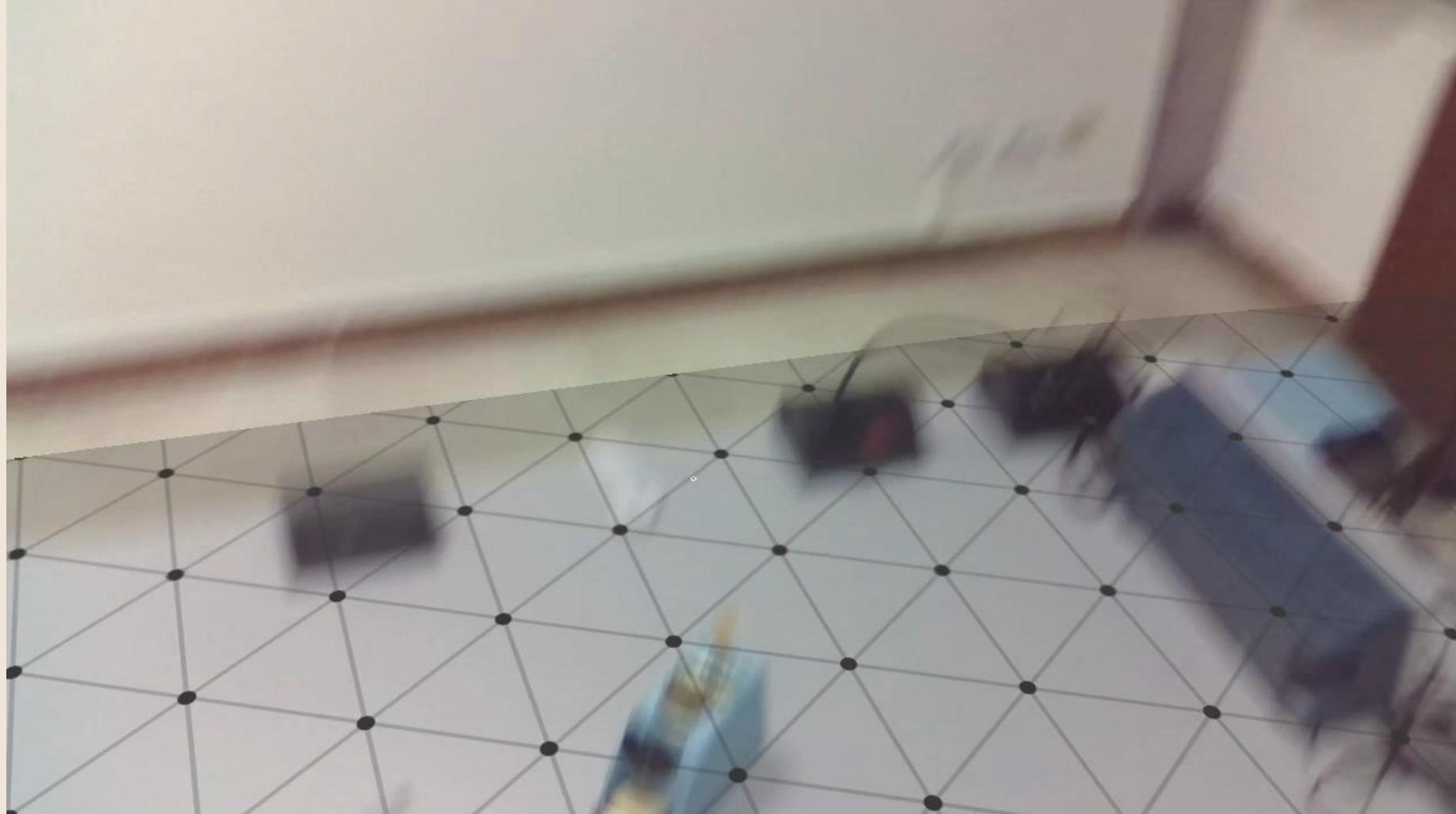


Opportunities for Internships and Theses on the Topics of the Course and Beyond

some examples of past projects in the following

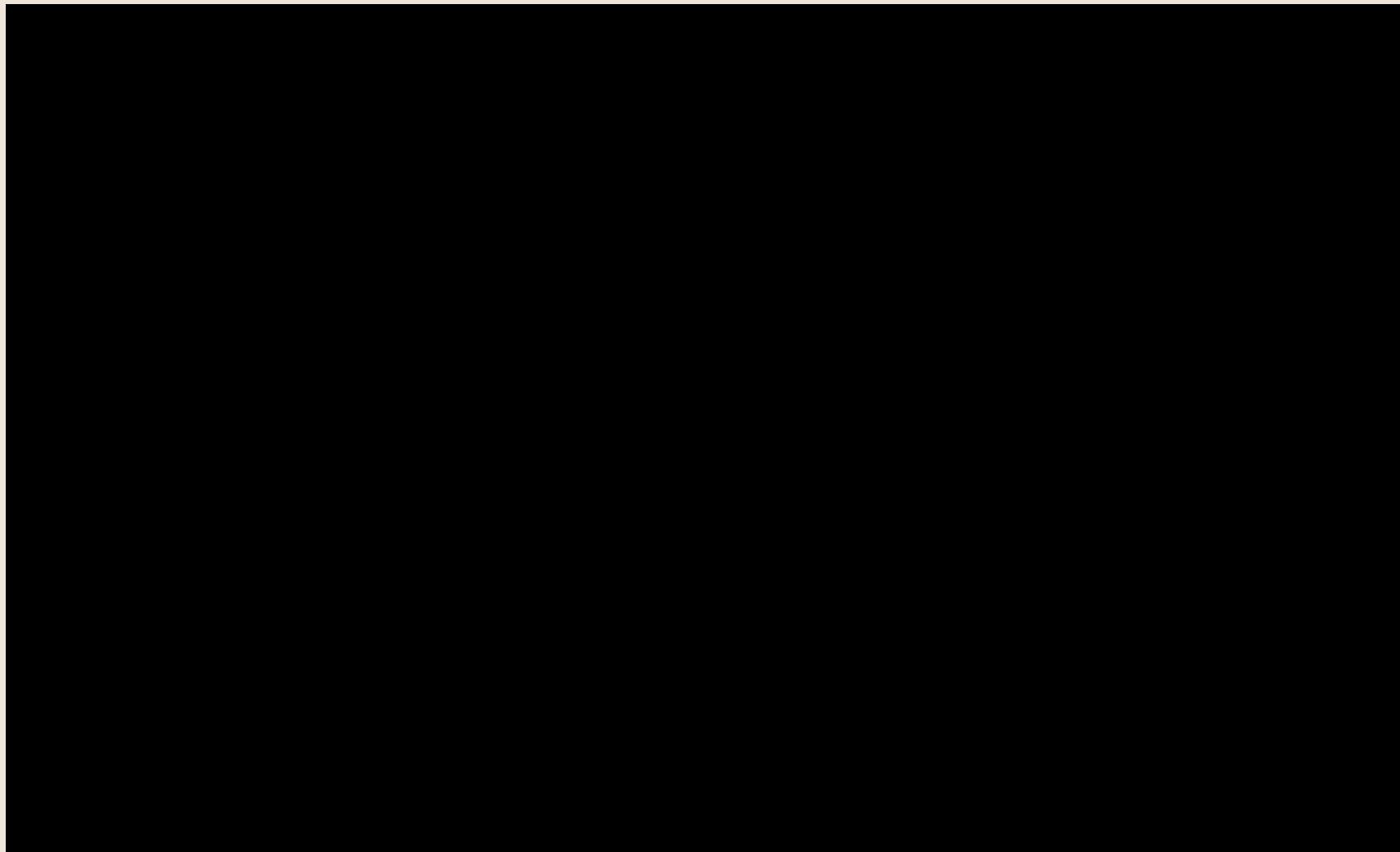
Riconoscimento di interazioni da dispositivi indossabili

VISAPP 2023



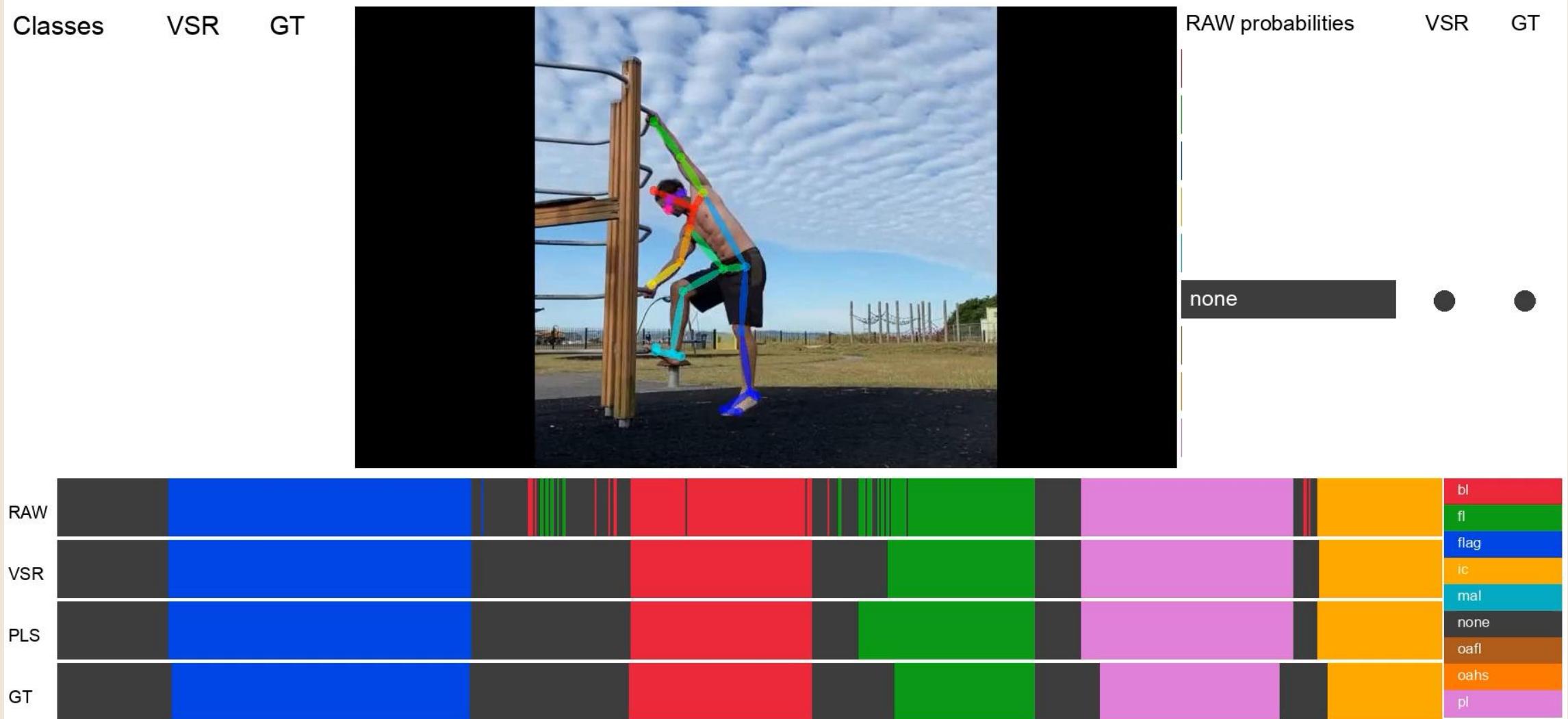
Alessandro Resta, 2021/2022

Chatbots (Multi-Agent RAG System)



Luca Strano, 2023/2024

Sports Video Analysis



Egocentric Streaming Object Memory



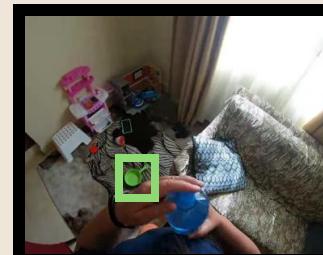
Egocentric Video



Query



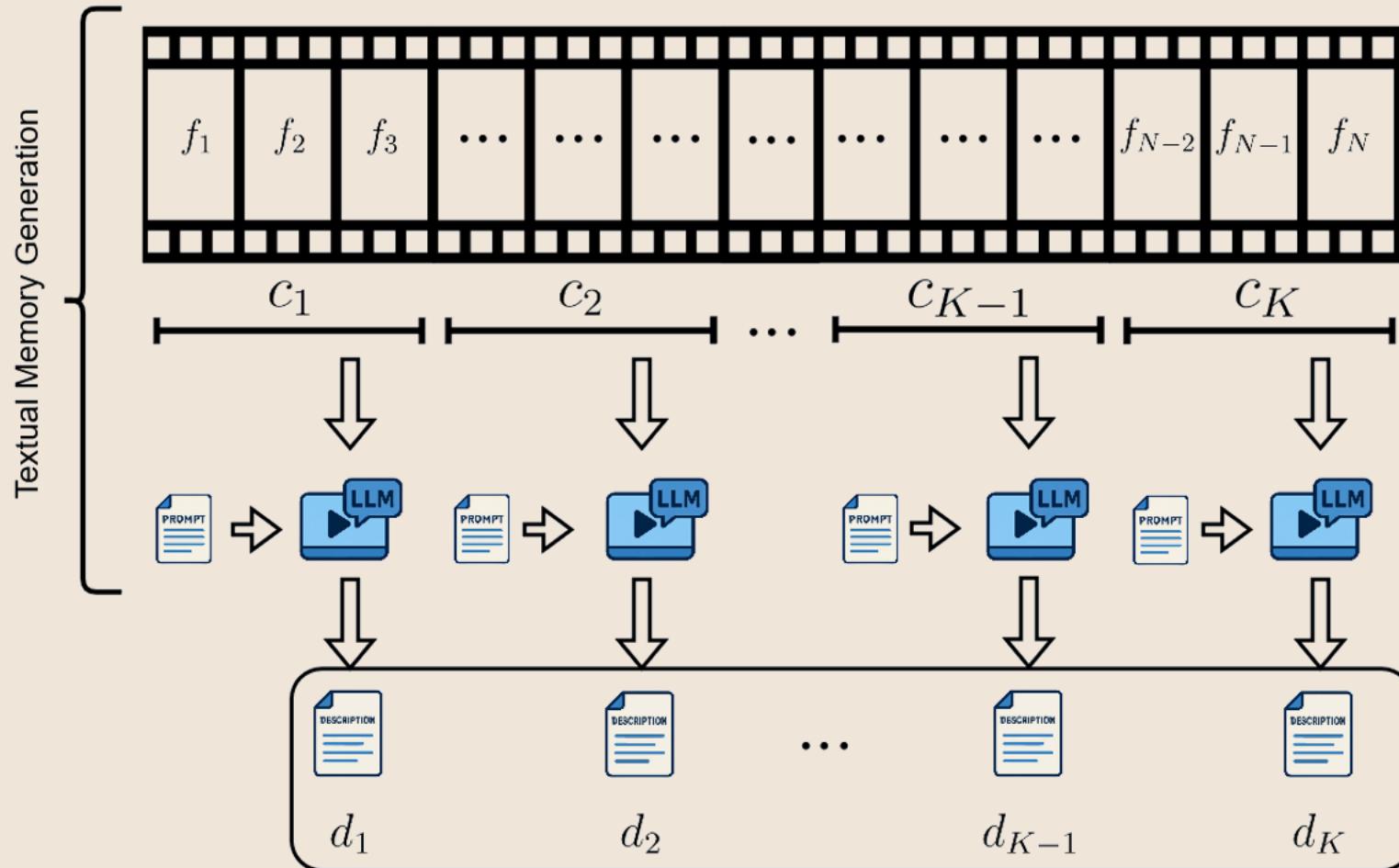
Model



Problems:

- The visual query is known **a priori**.
- Different queries on the same video lead to reprocess the sequence.
- Edge devices cannot keep in memory the full video until the query is issued.

LLM-Based Online Episodic Memory Retrieval



--- s seconds clip

Video LLM Descriptor

Prompt

Description

Question

LLM Reasoner

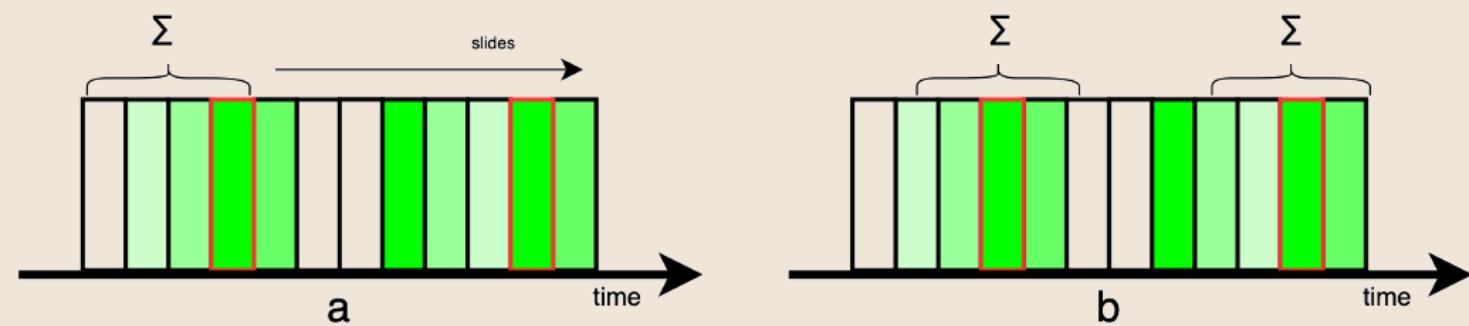
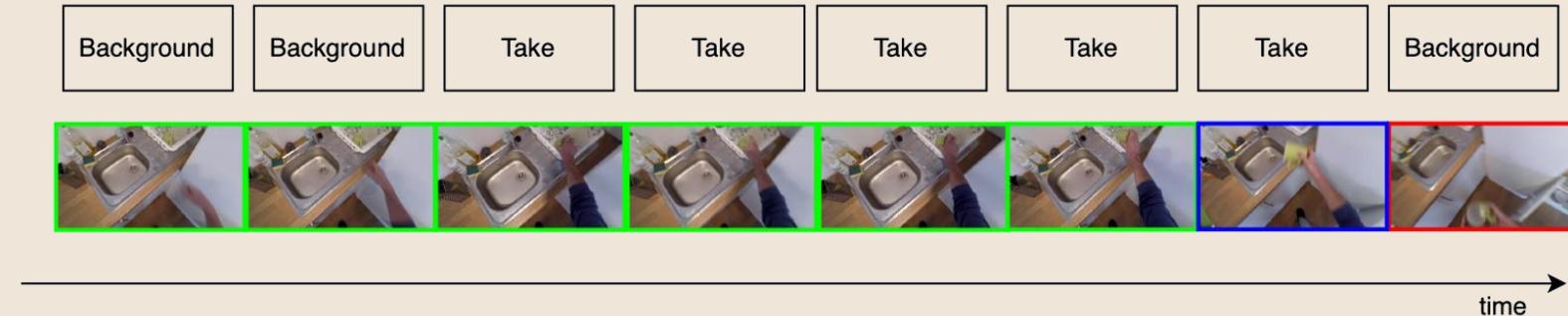
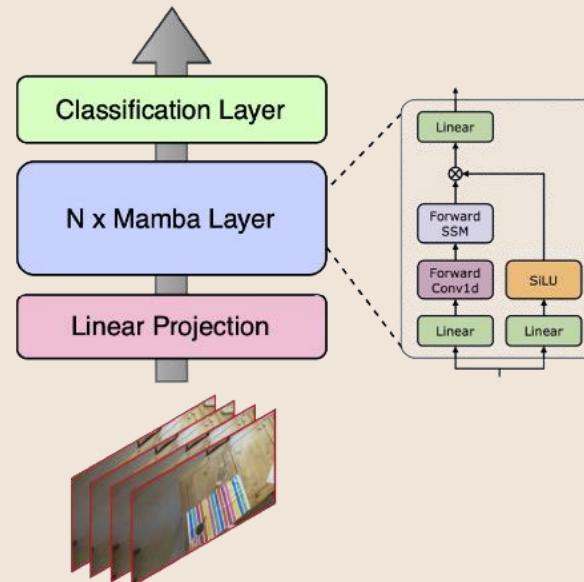
Answer

MAMBA-OTR for Take/Release Recognition



Take

Release



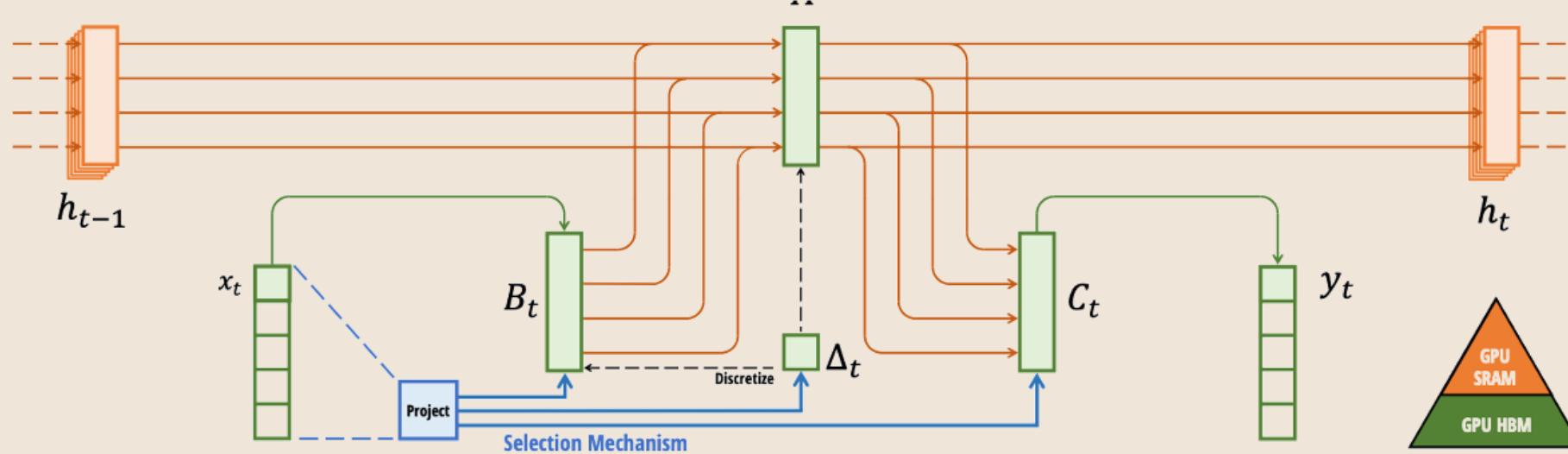
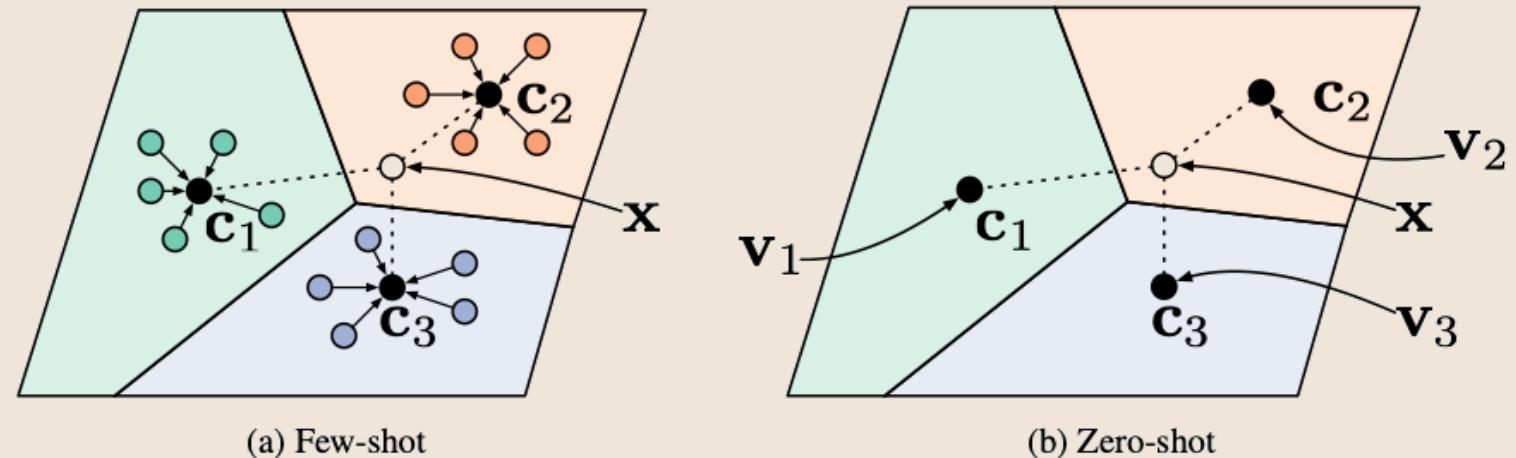
Model	Video	Frame	mp-mAP↑	
			Sliding↑	Streaming↑
Mamba-OTR	0.14s	8ns	45.48	43.35
Transformer	0.14s	8ns	38.96	0.04

Thesis Proposals

Few-Shot Action Recognition from Wearable Devices

L'obiettivo della tesi è sviluppare modelli di action recognition da wearable devices che possano essere allenati con pochi dati di training.

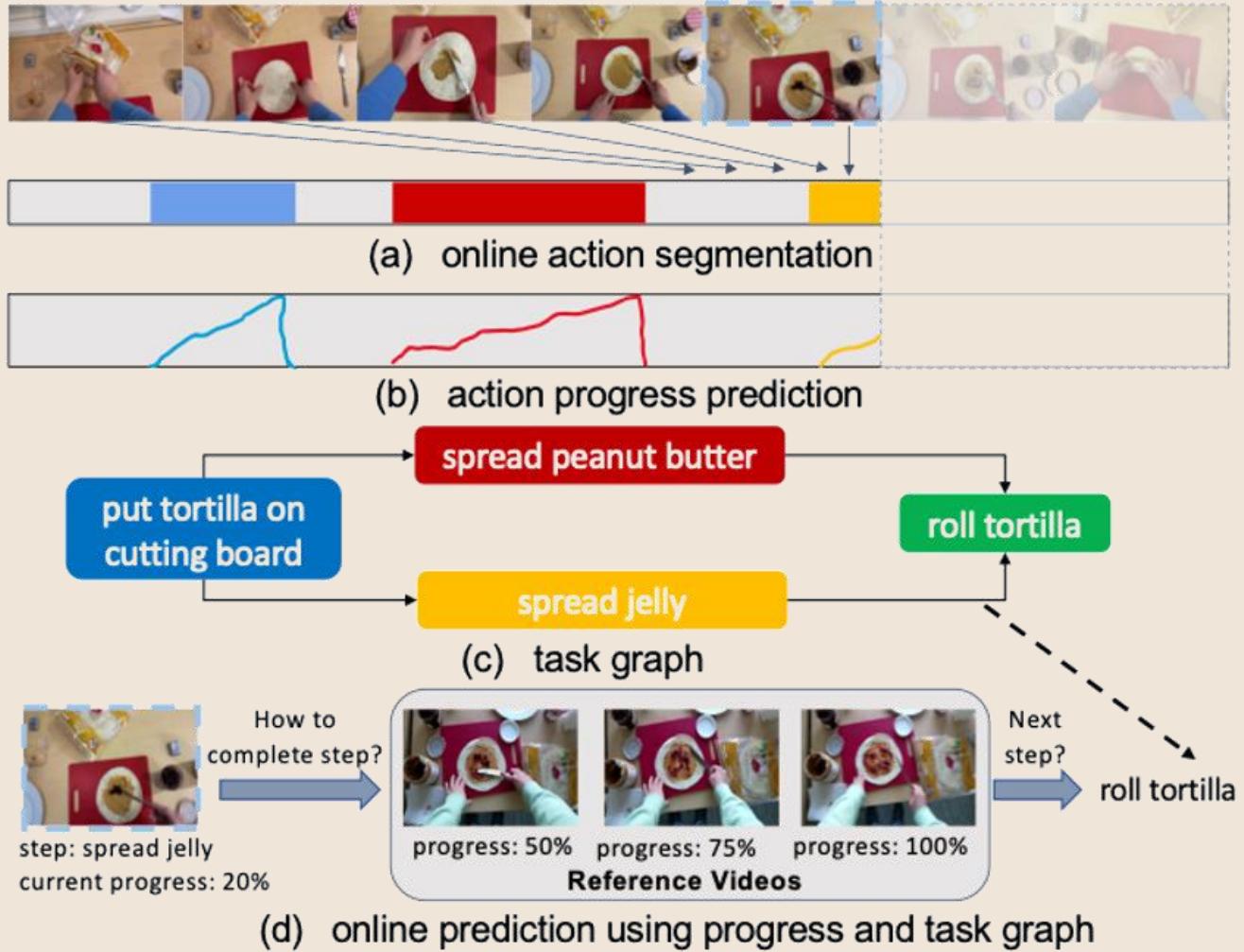
<https://antoninofurnari.github.io/proposte-tesi/>



Action Verification from Procedural Video

L'obiettivo della tesi è sviluppare algoritmi capaci di verificare le azioni effettuati dall'utente da video procedurale acquisito da dispositivi indossabili.

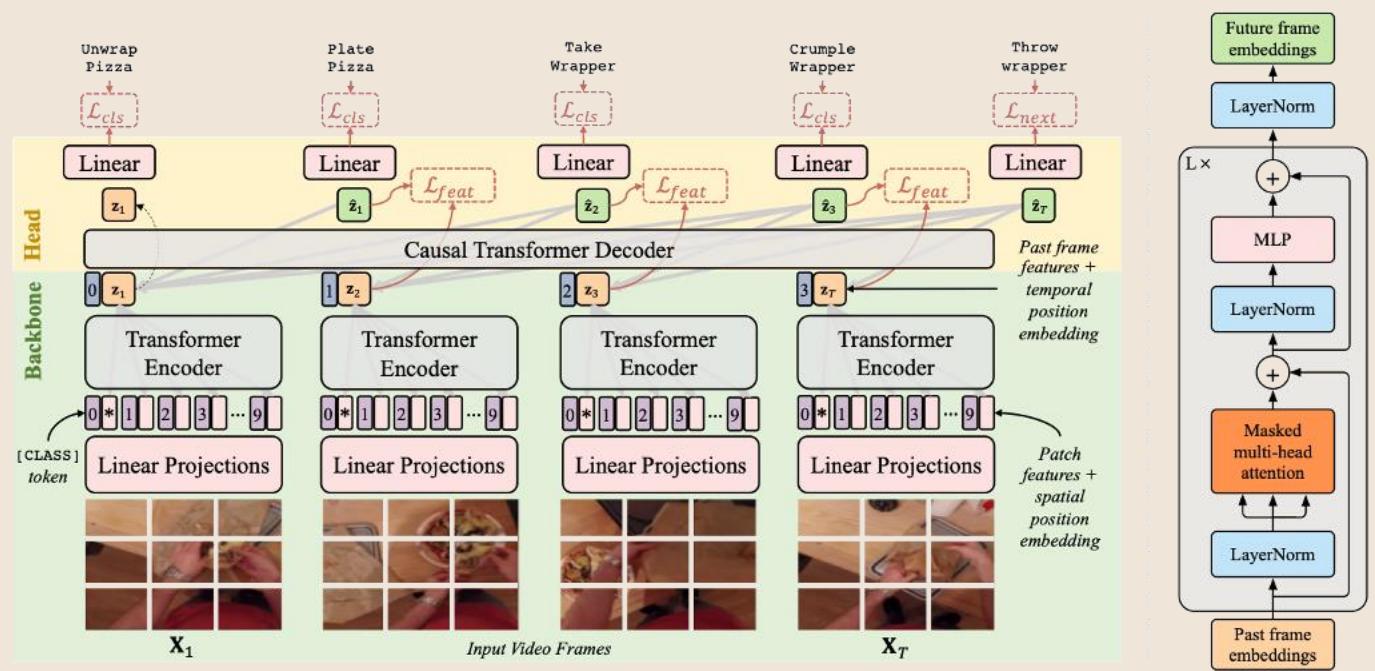
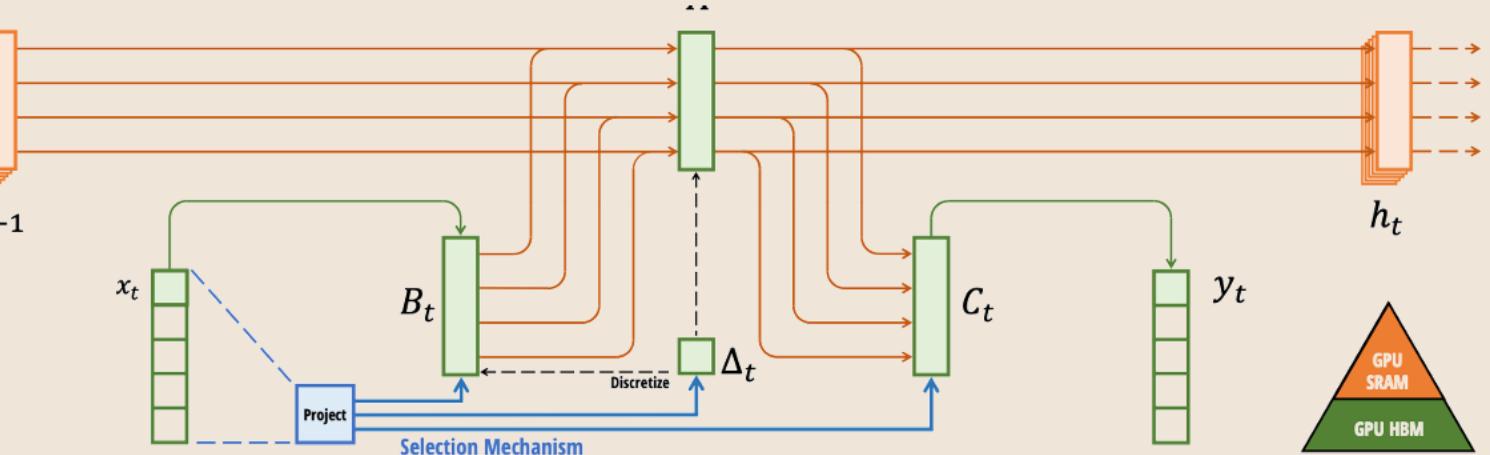
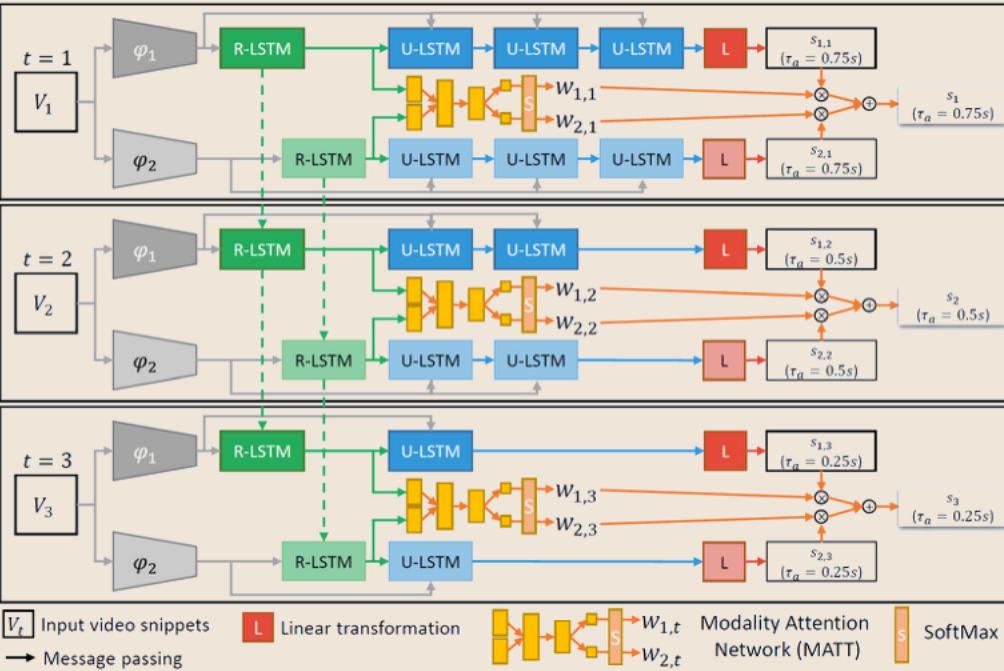
<https://antoninofurnari.github.io/proposte-tesi/>



Egocentric Action Anticipation con Architetture Mamba

L'obiettivo della tesi è quello di sviluppare modelli di egocentric action anticipation facendo uso di modelli di tipo Mamba e optionalmente di Transformer.

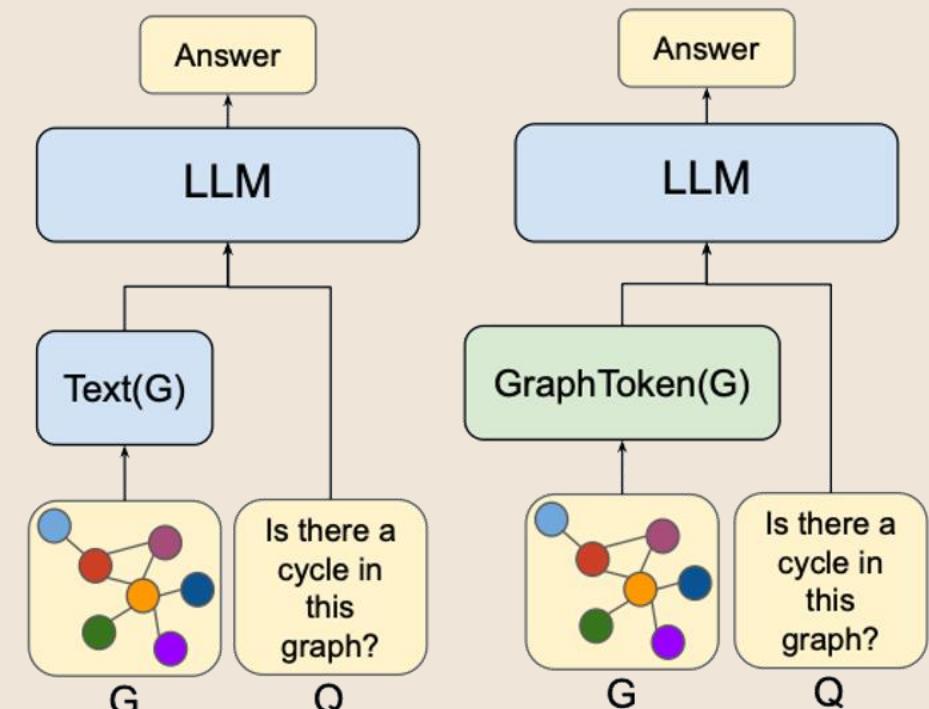
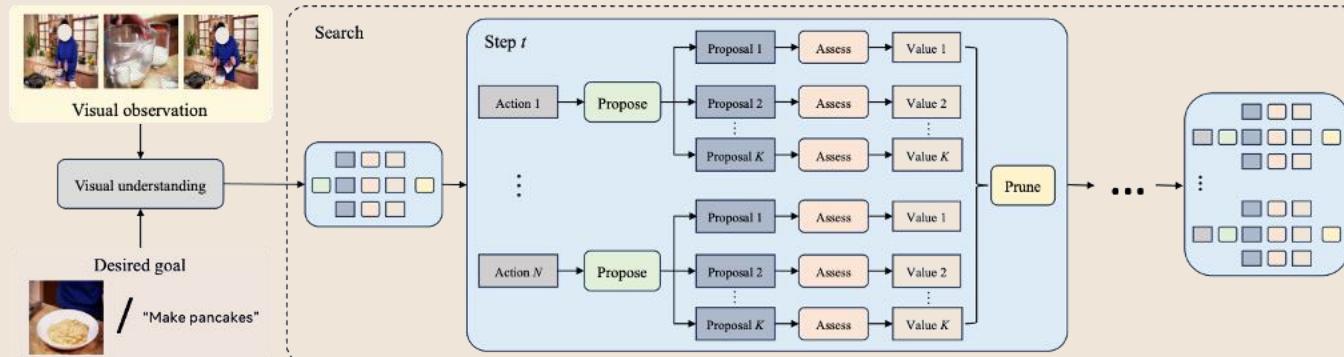
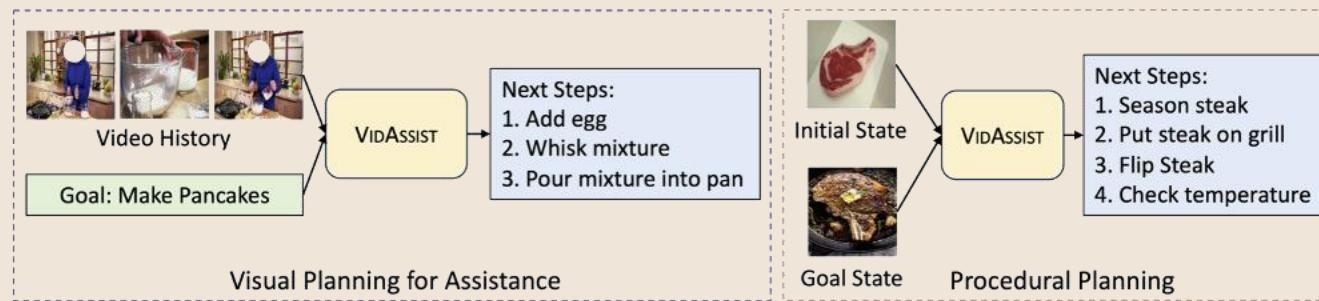
<https://antoninofurnari.github.io/proposte-tesi/>



Planning da video mediante rappresentazioni a grafo e Large Language Models

Procedure planning da video mediante rappresentazioni a grafo e Large Language Models

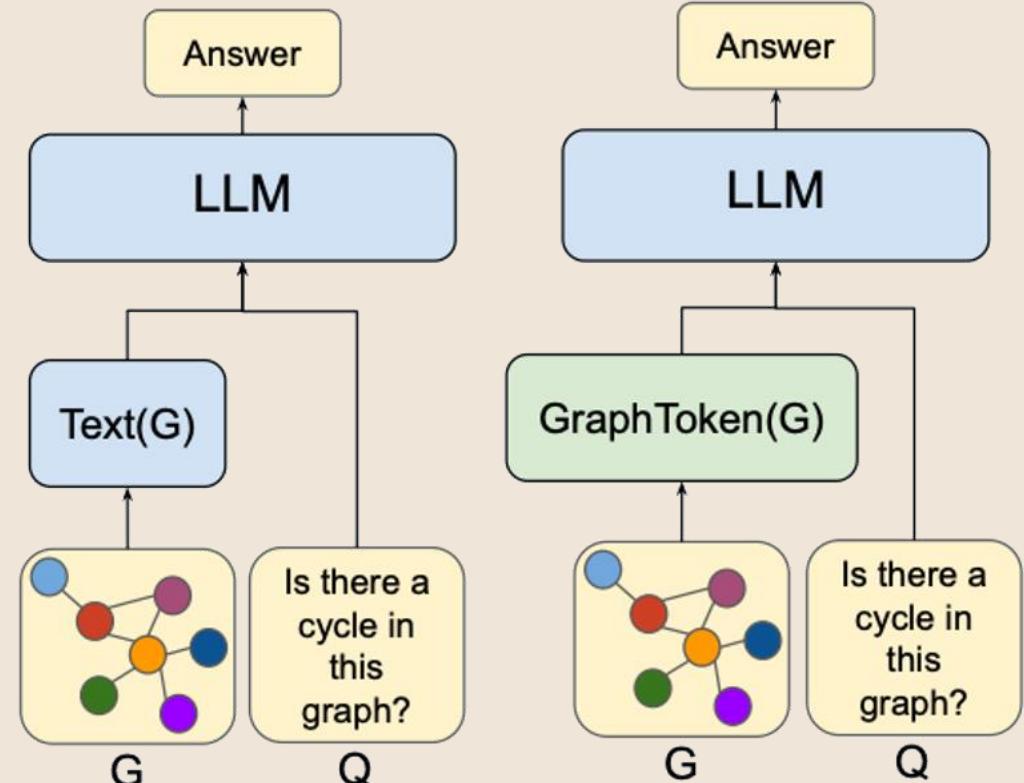
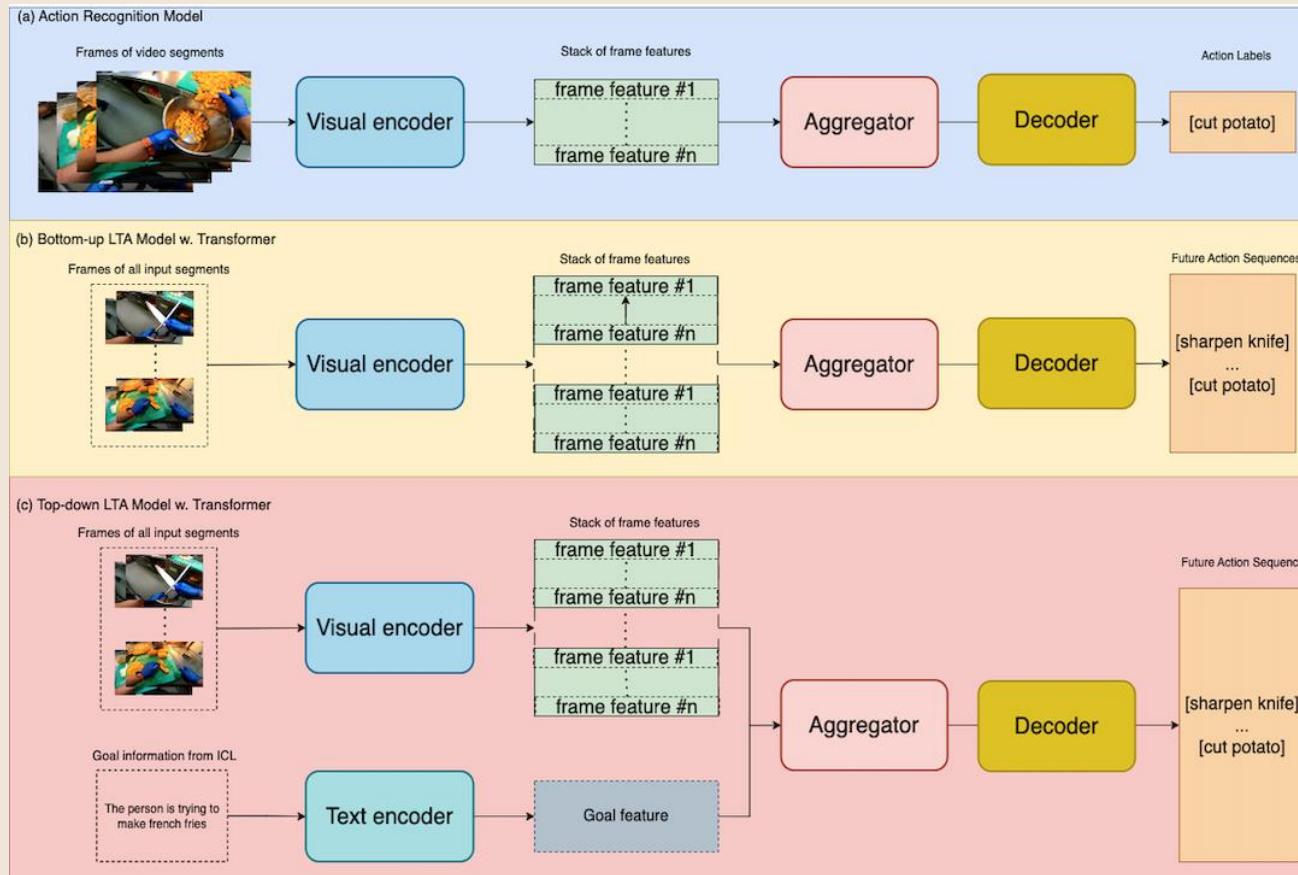
<https://antoninofurnari.github.io/proposte-tesi/>



Action anticipation da video mediante rappresentazioni a grafo e Large Language Models

Lo scopo della tesi è quello di integrare la conoscenza fornita da un grafo all'interno di un modello LLM per la predizione di azioni future.

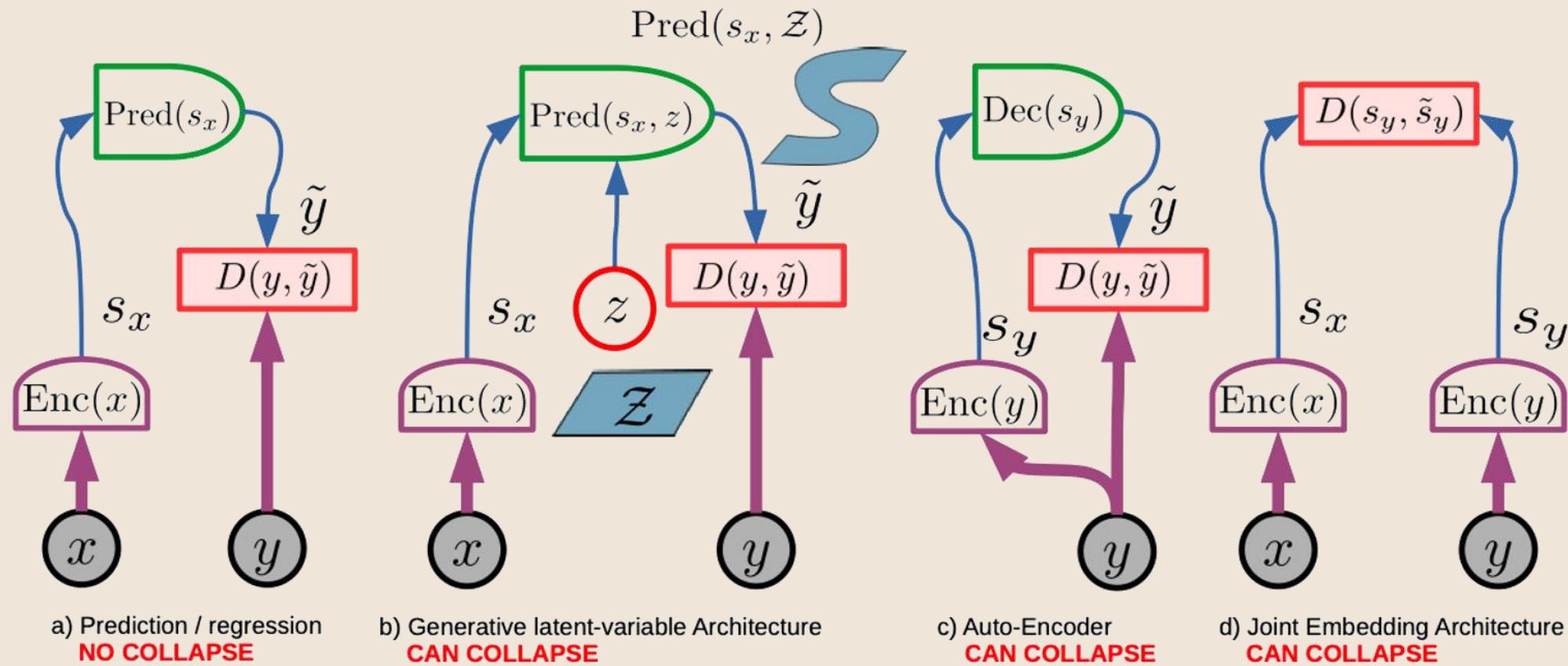
<https://antoninofurnari.github.io/proposte-tesi/>



Egocentric Action Anticipation con Architetture JEPA

L'obiettivo della tesi è quello di sviluppare modelli di egocentric action anticipation facendo uso di metodi di representation learning basati sul paradigma JEPA.

<https://antoninofurnari.github.io/proposte-tesi/>





Fondamenti di Analisi dei Dati

from **data analysis** to **predictive techniques**

Prof. Antonino Furnari (antonino.furnari@unict.it)
Corso di Studi in Informatica
Dip. di Matematica e Informatica
Università di Catania



Università
di Catania

The End