Machine Learning: Il motore dell'Intelligenza Artificiale e dell'investigazione scientifica guidata dai Big Data.

Stefano Leonardi

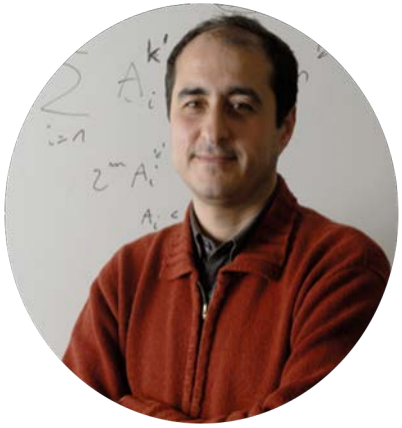
Dipartimento di Ingegneria informatica automatica e gestionale Antonio Ruberti

Sapienza Università di Roma

European Research Council

Established by the European Commission

Ricerca



- **Theory of Computing/Algorithmic Theory**
- **Algorithms and Data Science**

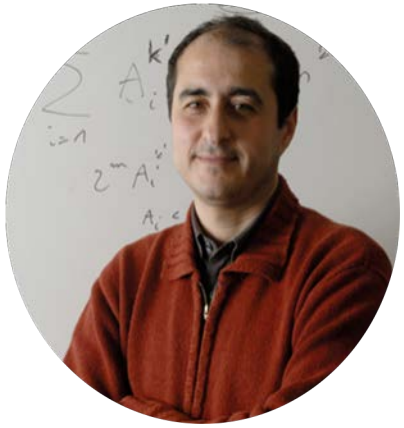
  Algorithms and Uncertainty

  Algorithmic Game Theory/Economics and Computation

- ERC Adv. Grant ’’ Algorithmic and Mechanism Design Research in Online MArkets’’ (AMDROMA) 2018-2023

- Supported by Google and Facebook since 2012

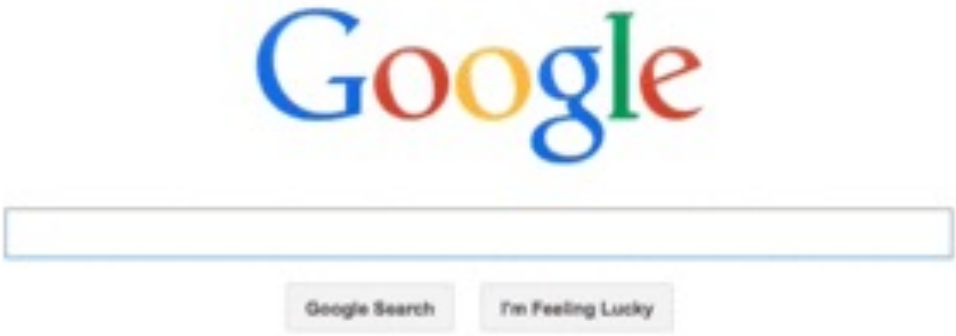
Attività Accademiche



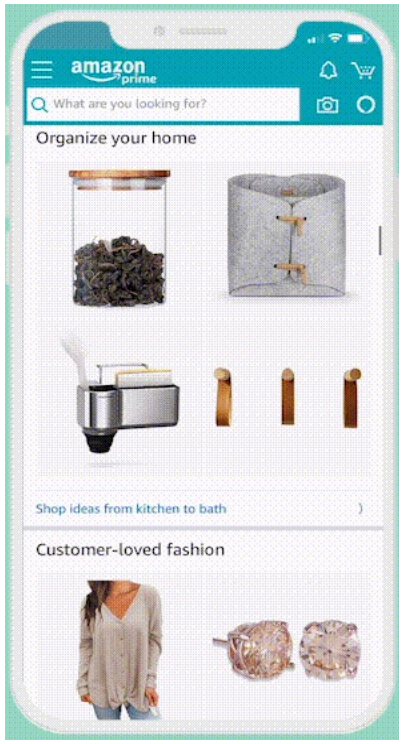
- Dottorato di Ricerca in Data Science [2018 -]
- Laurea Magistrale in Data Science [2014 -]
- Laurea Triennale in ''Scienze per l'Intelligenza Artificiale'' [2022 -]
- Sapienza School of Advanced Studies (SSAS) [2012-2018]
- Cultura e Creatività Digitale presso Fondazione ''I Lincei per la Scuola'' [2018 -]
- Commissione Scuola Accademia dei Lincei [2020 -]
- Conferenza Lincea ''Theory of Computing: A Multidisciplinary Perspective'' in occasion of the 54° ACM Symposium on Theory of Computing, Rome 20 – 24 June 2022.

Machine Learning: il motore dell’Intelligenza Artificiale

Ringraziamenti:
Fabrizio Silvestri (Sapienza)
Fabio Petroni (Facebook)
Tancredi Massimo Pentimalli (Berlin School of Integrative Oncology)
Stefano Giacu (Sapienza)



SAPIENZA
UNIVERSITÀ DI ROMA



SAPIENZA
UNIVERSITÀ DI ROMA

SAPIENZA
UNIVERSITÀ DI ROMA

SAPIENZA
UNIVERSITÀ DI ROMA

SAPIENZA
UNIVERSITÀ DI ROMA

SAPIENZA
UNIVERSITÀ DI ROMA

The importance of being on twitter

by Jerome K. Jerome
London, Summer 1897

It is a curious fact that the last remaining form of social life in which the people of London are still interested is Twitter. I was struck with this curious fact when I went on one of my periodical holidays to the sea-side, and found the whole place twittering like a starling cage. I called it an anomaly, and it is.

I spoke to the waiter, whose cottage, like all waiter's cottages, is full of antiquities and interesting relics of former centuries. I said to him, "My dear waiter, what does all this twittering mean?" And he replied, "Why, sir, of course it means Twitter." "Ah!" I said, "I know about that. But what is Twitter?"

"It is a system of short and pithy sentences strung together in groups, for the purpose of conveying useful information to the initiated, and entertainment and the exercise of wits to the initiated, and entertainment and the exercise of wits to the rest of us."

"Very interesting," I said. "But it is a name?"
"It has," he said. "It is called Twitter."
"Yes," I said, "I know that, but what is it?"
"It is a system of information," he said.
"Oh, yes," I replied. "But what is it?"

"Why, sir," he said, "you can go up to any of the gentlemen you see twittering in the street, and say to him, 'You are a fool,' or 'Your wife is an adulteress,' or 'You have stolen that hat,' and if he is a member of the initiated he will answer you in the same form and tell you that you are a hat, or that your eyes resemble the eyes of a duck, or that you have stepped out of your part in the last charade you acted in, or that you were for a short time a statistician in a Government Office, and he will go on to tell you the whole story of your life, in language so exceedingly small and pointed that even you will be glad you can't understand it.

SAPIENZA
UNIVERSITÀ DI ROMA

TEXT PROMPT an armchair in the shape of an avocado. . . .

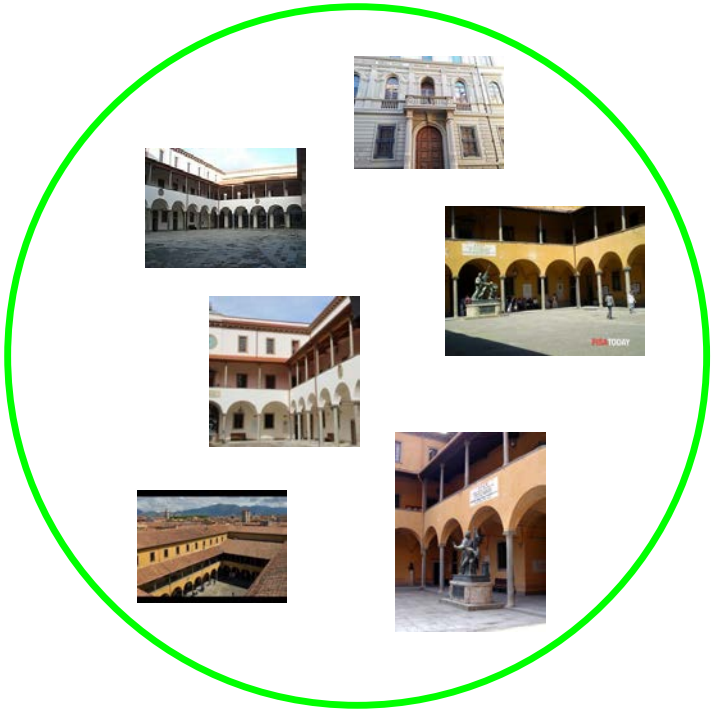
AI-GENERATED IMAGES

Edit prompt or view more images↓

SAPIENZA
UNIVERSITÀ DI ROMA

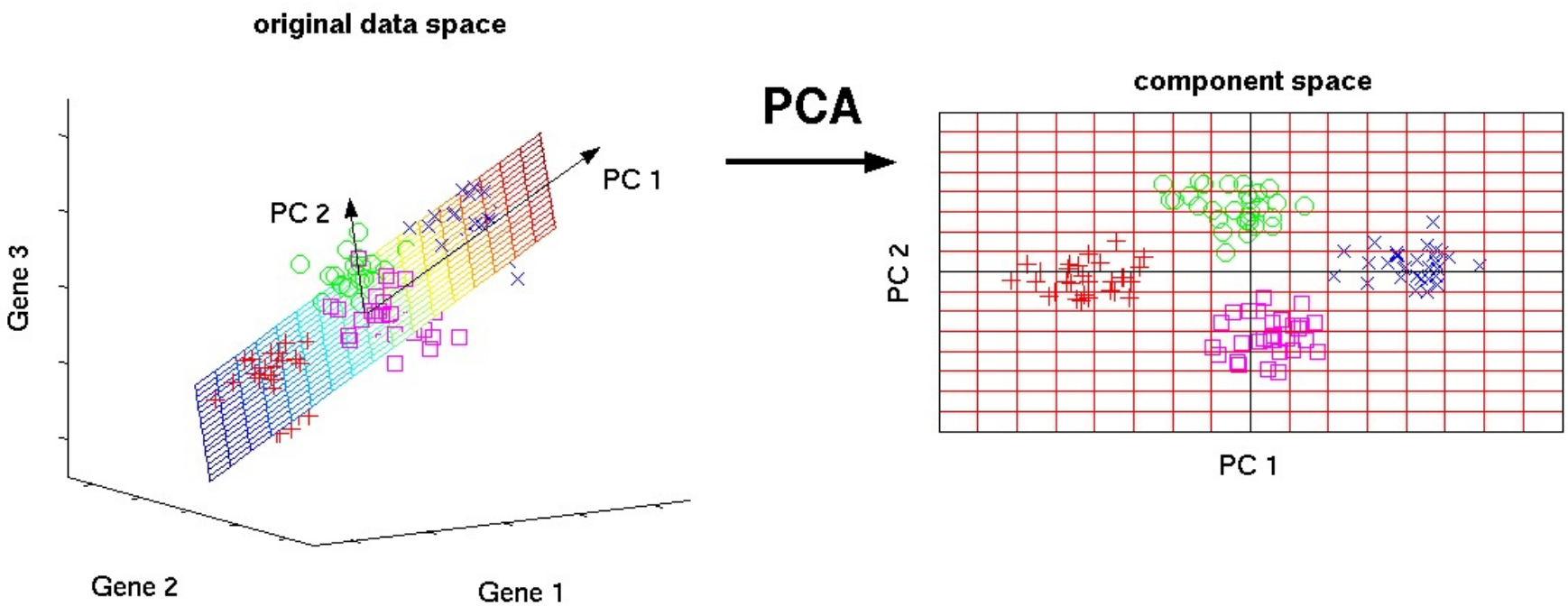
Cosa è il Machine Learning?

Unsupervised Machine Learning



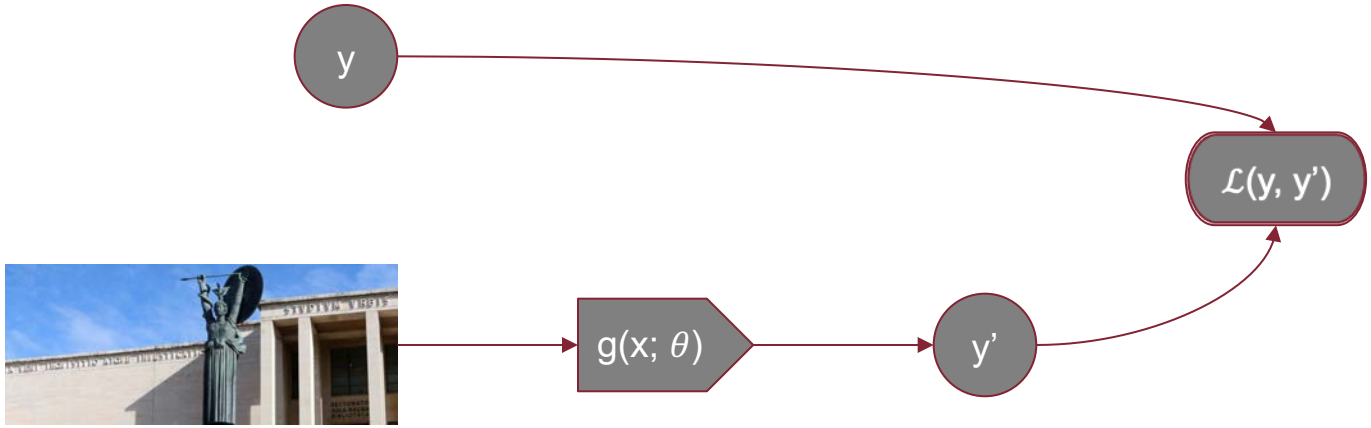
d( , )

What does (unsupervised) learning mean?

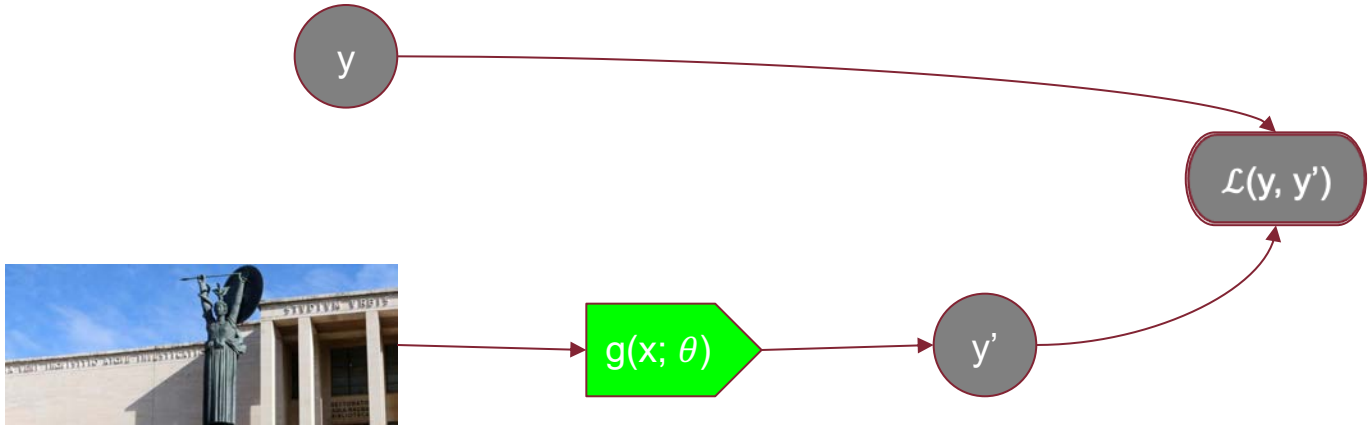


SAPIENZA
UNIVERSITÀ DI ROMA

Supervised Machine Learning



What does (supervised) learning mean?

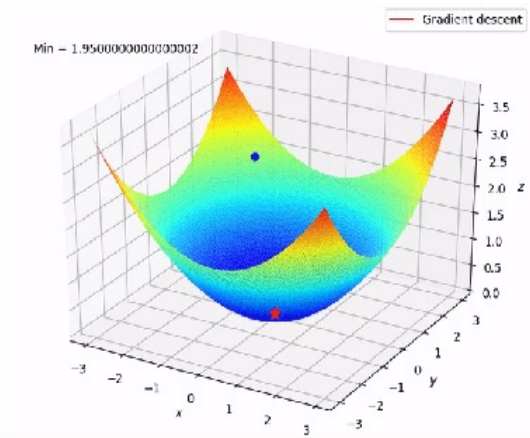


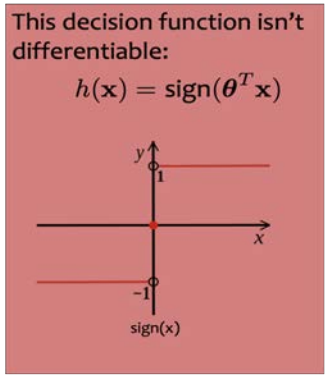
$\theta^t \leftarrow u(\theta^{t-1}, \nabla g)$

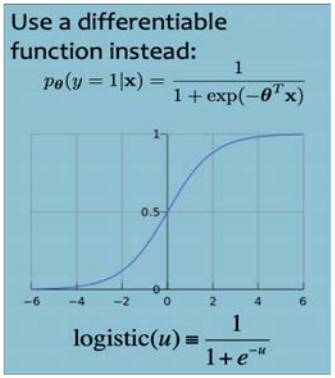
e.g.,
$\theta^t \leftarrow \theta^{t-1} - \eta \nabla g$

Optimization

- Common to all ML (and DL) techniques there is the optimization component
  - We want to find the parameters of a model that minimize/maximize some cost functions.
- An ML model is essentially a member of a family of functions that we call hypothesis
- A model depends generically on parameters $\boldsymbol{\theta} \rightarrow P(X;\boldsymbol{\theta})$
- How do you find the best model?
  - You pick the best parameters!

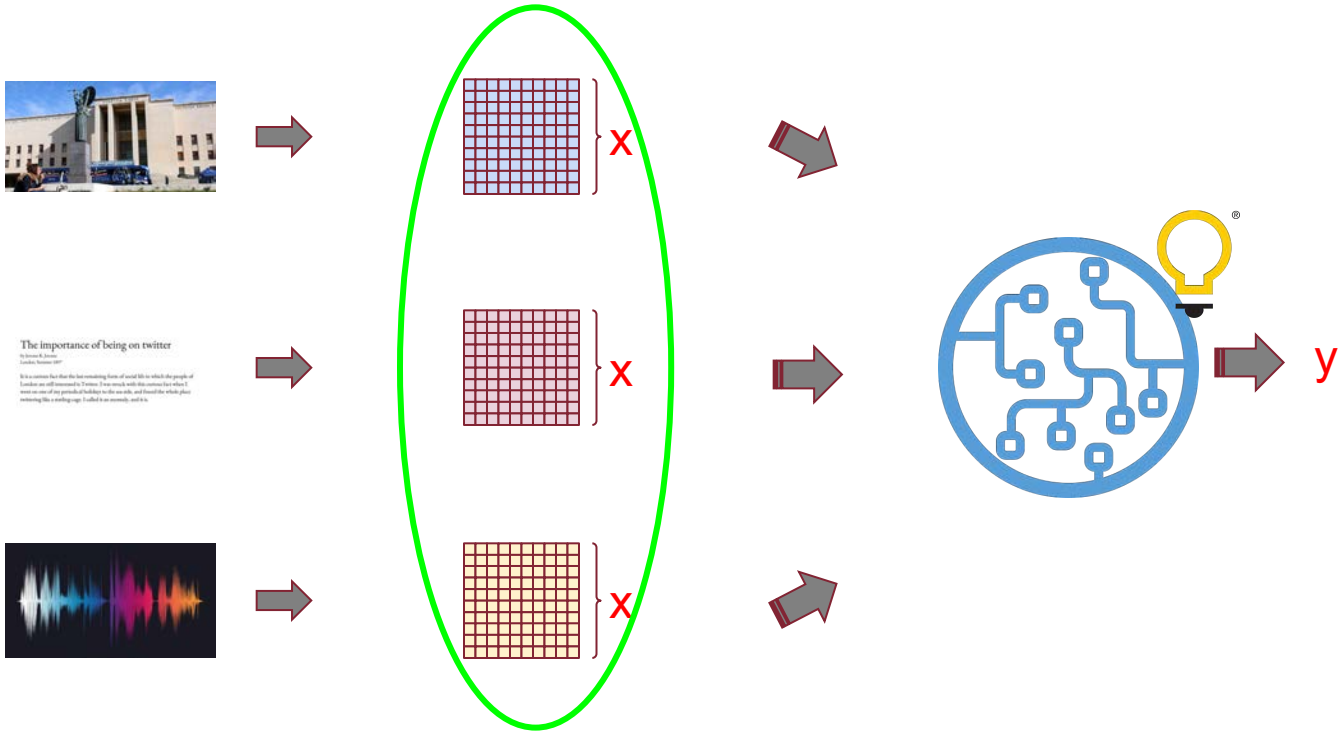






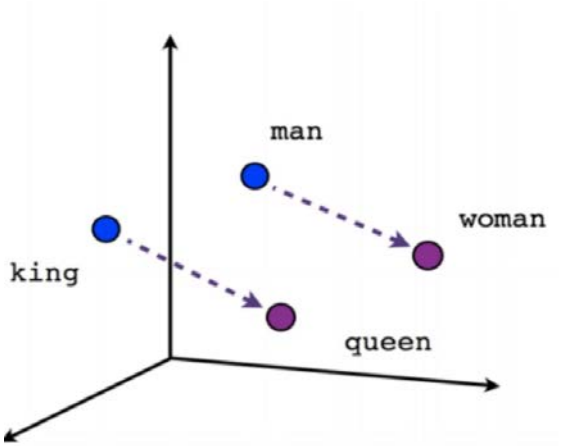
The central challenge in ML is that our algorithm must perform well on new, previously unseen inputs

Data Representation

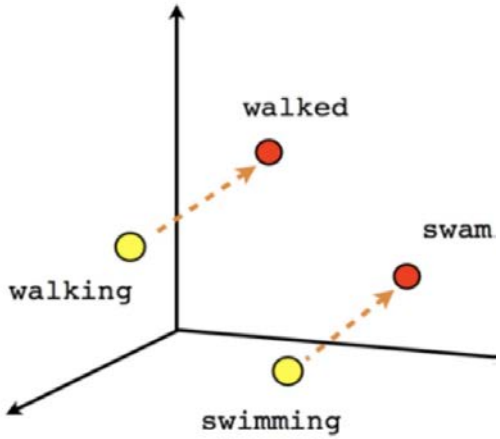


SAPIENZA
UNIVERSITÀ DI ROMA

Data Representation



Male-Female



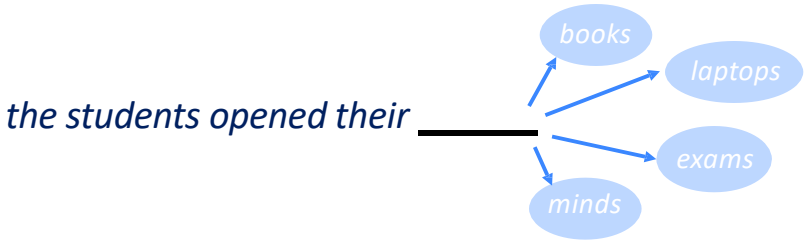
Verb tense

EMBEDDING

The big data revolution

SAPIENZA
UNIVERSITÀ DI ROMA

Language Modeling



- More formally: given a sequence of words $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots, \boldsymbol{x}^{(t)}$ compute the probability distribution of the next word $\boldsymbol{x}^{(t+1)}$

$$P(\boldsymbol{x}^{(t+1)} | \ \boldsymbol{x}^{(t)}, \ldots, \boldsymbol{x}^{(1)})$$

where $\boldsymbol{x}^{(t+1)}$ can be any word in the vocabulary $V = \{\boldsymbol{w}_1, ..., \boldsymbol{w}_{|V|}\}$

- A system that does this is called a **Language Model**.

A RNN Language Model

output distribution

$$\hat{\boldsymbol{y}}^{(t)} = \text{softmax}\left(\boldsymbol{U}\boldsymbol{h}^{(t)} + \boldsymbol{b}_2\right) \in \mathbb{R}^{|V|}$$

hidden states

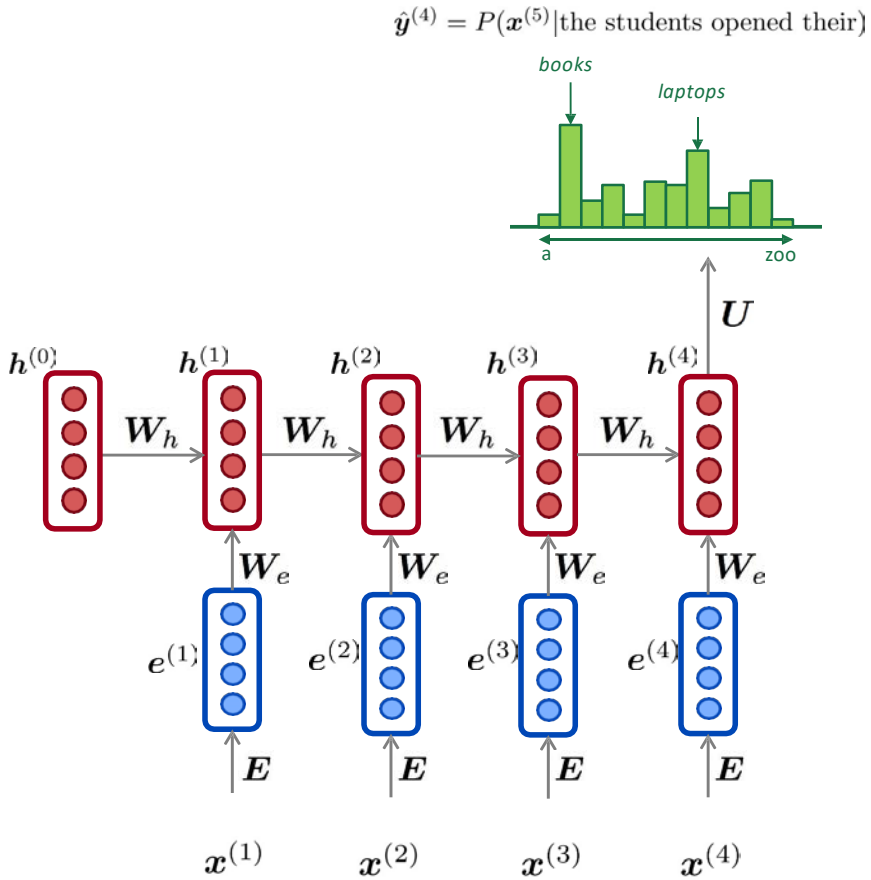
$$\boldsymbol{h}^{(t)} = \sigma\left(\boldsymbol{W}_h\boldsymbol{h}^{(t-1)} + \boldsymbol{W}_e\boldsymbol{e}^{(t)} + \boldsymbol{b}_1\right)$$

$$\boldsymbol{h}^{(0)}$$

word embeddings

$$\boldsymbol{e}^{(t)} = \boldsymbol{E}\boldsymbol{x}^{(t)}$$

$$\boldsymbol{x}^{(t)} \in \mathbb{R}^{|V|}$$



SAPIENZA
UNIVERSITÀ DI ROMA

Generating text with a RNN Language Model

- You can train a RNN-LM on any kind of text, then generate text in that style.
- RNN-LM trained on Obama speeches:

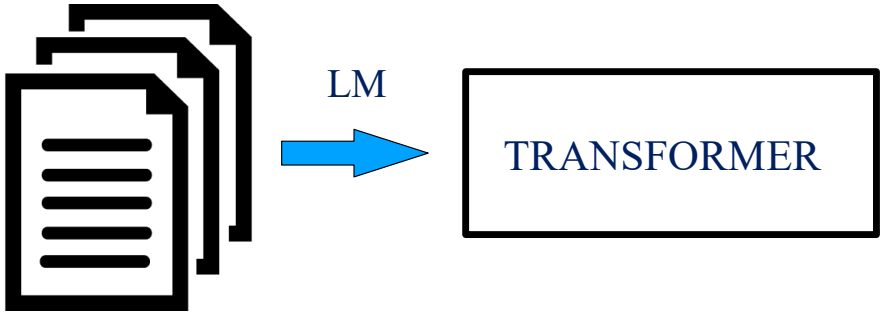
The United States will step up to the cost of a new challenges of the American people that will share the fact that we created the problem. They were attacked and so that they have to say that all the task of the final days of war that I will not be able to get this done.

https://medium.com/@samim/obama-rnn-machine-generated-political-speeches-c8abd18a2ea0

Transformers (e.g., BERT 2018)

1. pretrain

Huge corpora (B of words)

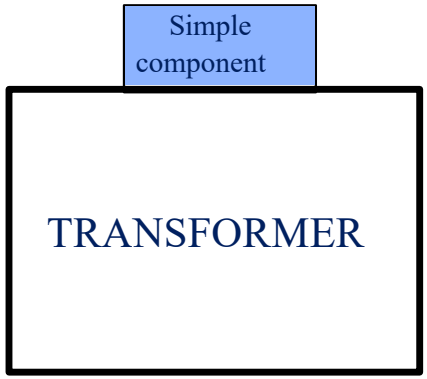


Unsupervised

This takes days/weeks on several GPUs / TPUs

Pre-trained models publicly available on the web!

2. fine-tune

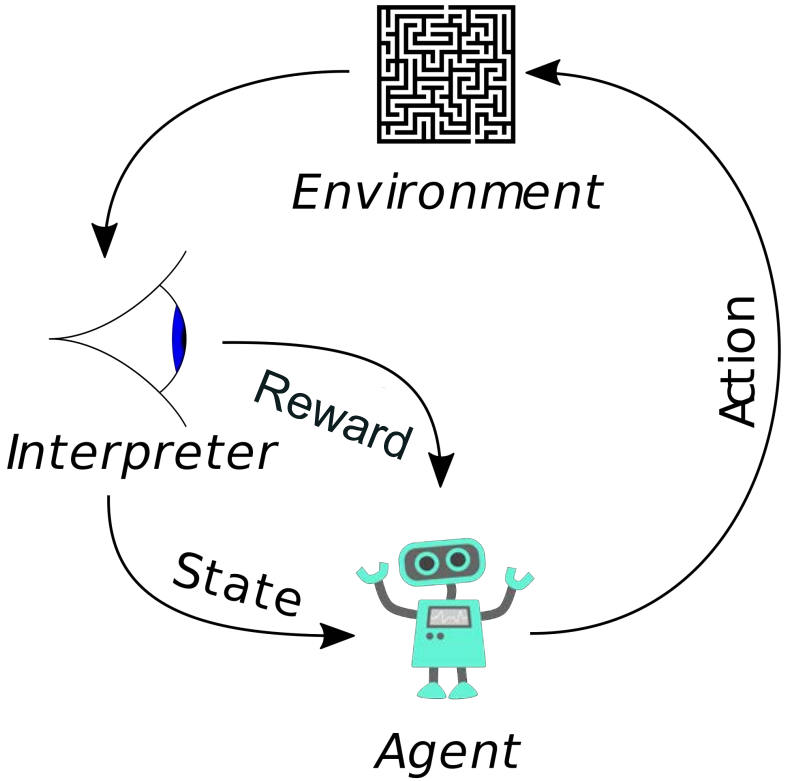


Supervised

Fine-tune for a specific task on as much labeled data as you have. Fast !

SAPIENZA
UNIVERSITÀ DI ROMA

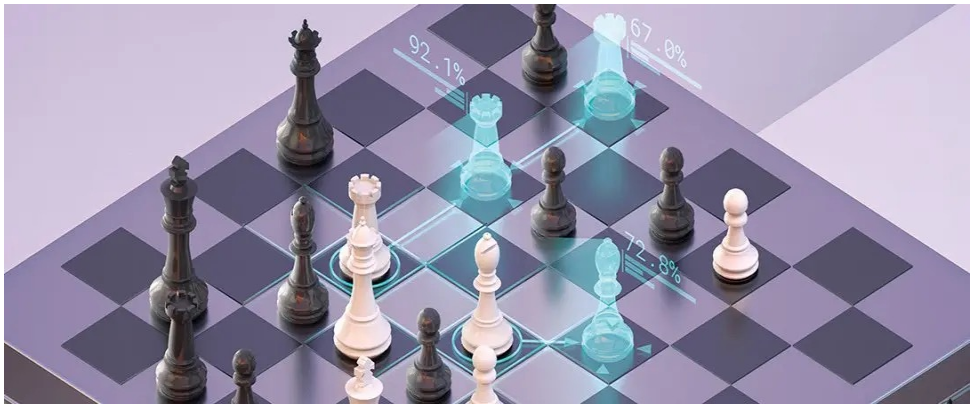
Reinforcement Learning



Credits Wikipedia

Alpha Zero (Deep Mind 2018)

- A single system that taught itself from scratch how to master the games of chess, shogi (Japanese chess), and Go, beating a world champion program in each case.
- To learn each game, an untrained neural network plays millions of games against itself via a process of trial and error called reinforcement learning.
- At first, it plays completely randomly, but over time the system learns from wins, losses, and draws to adjust the parameters of the neural network, making it more likely to choose advantageous moves in the future.



I can't disguise my satisfaction that it plays with a very dynamic style, much like my own!"

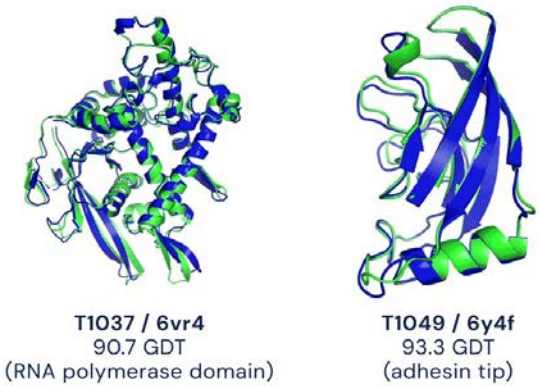
GARRY KASPAROV
FORMER WORLD CHESS CHAMPION

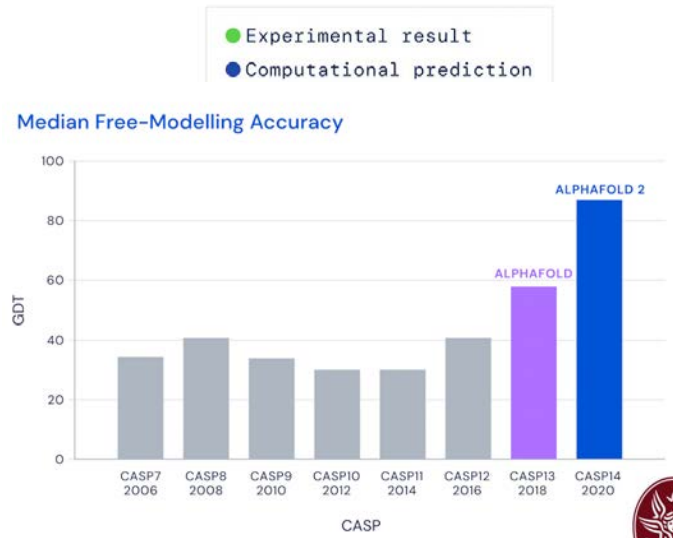
SAPIENZA
UNIVERSITÀ DI ROMA

Machine Learning e l’investigazione scientifica basata sui big data

Alpha Fold (Deep Mind 2018-2020)

- What a protein does largely depends on its unique 3D structure. Figuring out what shapes proteins fold into is known as the "protein folding problem", and has stood as a grand challenge in biology for the past 50 years.
- The latest version of AlphaFold, used at CASP14, is an attention-based neural network system, trained end-to-end, that attempts to interpret the structure of the graph between residues in close proximity.
- The system is trained on publicly available data consisting of ~170,000 protein structures from the protein data bank together with large databases containing protein sequences of unknown structure.





SAPIENZA
UNIVERSITÀ DI ROMA

Alpha Fold (Deep Mind 2018-2020)

This computational work represents a stunning advance on the protein-folding problem, a 50-year-old grand challenge in biology. It has occurred decades before many people in the field would have predicted. It will be exciting to see the many ways in which it will fundamentally change biological research.

PROFESSOR VENKI RAMAKRISHNAN
NOBEL LAUREATE AND PRESIDENT OF THE ROYAL SOCIETY

We have been stuck on this one problem – how do proteins fold up – for nearly 50 years. To see DeepMind produce a solution for this, having worked personally on this problem for so long and after so many stops and starts, wondering if we'd ever get there, is a very special moment.

PROFESSOR JOHN MOULT
CO-FOUNDER AND CHAIR OF CASP, UNIVERSITY OF MARYLAND

In partnership with EMBL-EBI, it has been launched the AlphaFold Protein Structure Database.

SAPIENZA
UNIVERSITÀ DI ROMA

ML and disease classification: brain tumors

Problem:

- How many subtypes of brain tumors exist?
- How to identify (i.e. classify and Dx) them?
- Tumors are rare and Dx is not standardized

Relevance:

- Completely changes prognosis and treatment
- A process combining an advanced imaging technology and artificial intelligence (AI) can accurately diagnose brain tumors in fewer than 3 minutes during surgery.

Avaliable data:

- Tumor methylation profiles
  (Epigenetic marks of active and inactive DNA)
  (What are the tumor cells doing? Which genes are active?)

https://www.nytimes.com/2020/01/06/health/artificial-intelligence-brain-cancer.html

CNNs and digital pathology: AI-assisted Dx

Problem:

- How to reliably identify skin cancer?
- Clinicians have different levels of experience

Relevance:

- Skin lesions extremely frequent (most common cancer!)
- Early detection dramatically changes treatment (i.e simple surgical excision)
- Low-cost, widespread access (e.g. smartphones!)

- CNN end-to-end training
  Google's Inception CNN
  130'000 images for training
  In: pixels and labels (biopsy results, 2'000 diseases)
  Out: Benign vs Malignant

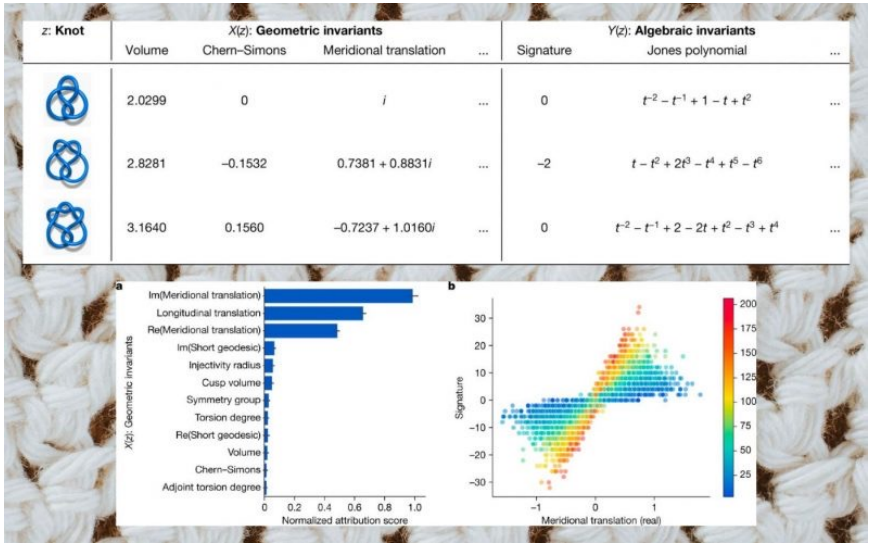
- Compared to 21 dermatologists, CNN outperformed humans!

https://mashable.com/article/ibm-research-melanoma-testing-with-phone-camera

Math Conjectures (Deep Mind 2021)

- Use machine learning to make significant new discoveries in pure mathematics
- Conjecture a new connection between the algebraic and geometric structure of knots, and a candidate algorithm predicted by the combinatorial invariance conjecture for symmetric groups
- A supervised learning model was able to detect the existence of a pattern between a large set of geometric invariants and the signature $\sigma(K)$ of a knot that was not previously known to be related to the hyperbolic geometry.

Nature **volume 600**, pages70–74 (2021)



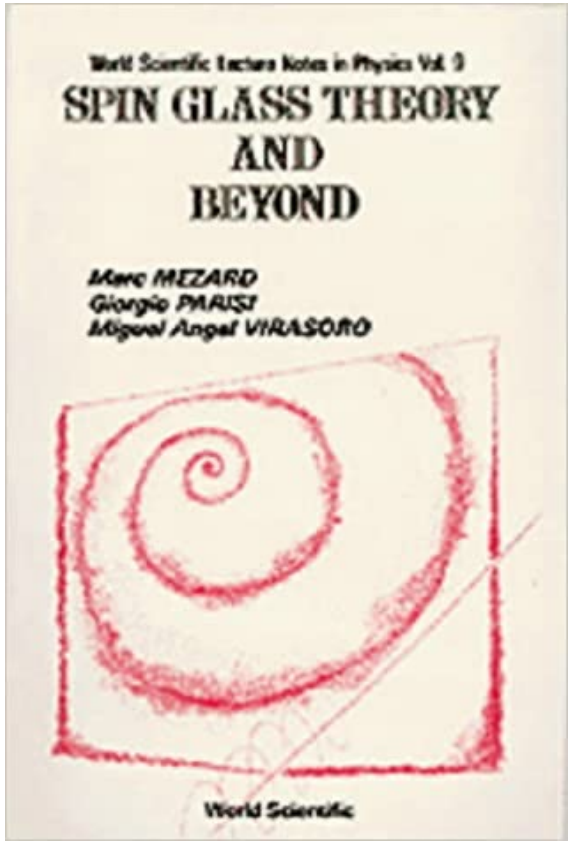
z: Knot	$X(z)$: Geometric invariants				$Y(z)$: Algebraic invariants		
	Volume	Chern–Simons	Meridional translation	...	Signature	Jones polynomial	...
	2.0299	0	i	...	0	$t^{-2} - t^{-1} + 1 - t + t^2$	...
	2.8281	−0.1532	$0.7381 + 0.8831i$	...	−2	$t - t^2 + 2t^3 - t^4 + t^5 - t^6$	...
	3.1640	0.1560	$-0.7237 + 1.0160i$	...	0	$t^{-2} - t^{-1} + 2 - 2t + t^2 - t^3 + t^4$	...

- Neural networks pre-trained on text and fine tuned on code perfectly solves university-level problems.
  Drori et al (PNAS 2022)

SAPIENZA
UNIVERSITÀ DI ROMA

Spin Glass Theory

- Finding the ground states of spin glasses is an essential computationally hard problem for the understanding of the nature of disordered magnetic and other physical systems. DIRAC Fan et al [2021] uses Deep Neural Networks trained purely on small-scale spin glass instances and then applied to arbitrarily large ones.
- Use methods developed in statistical physics of glassy systems to analyze numerically the training dynamics of deep neural networks. Baity-Jesi et al. [ICML 2018]

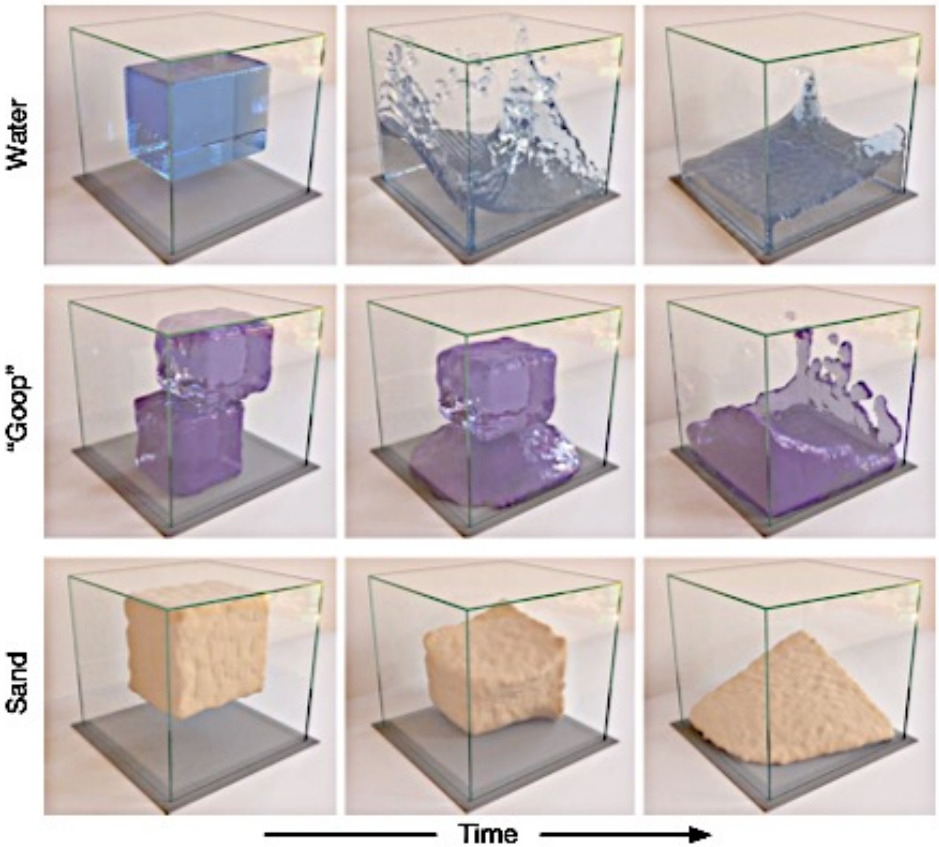


SAPIENZA
UNIVERSITÀ DI ROMA

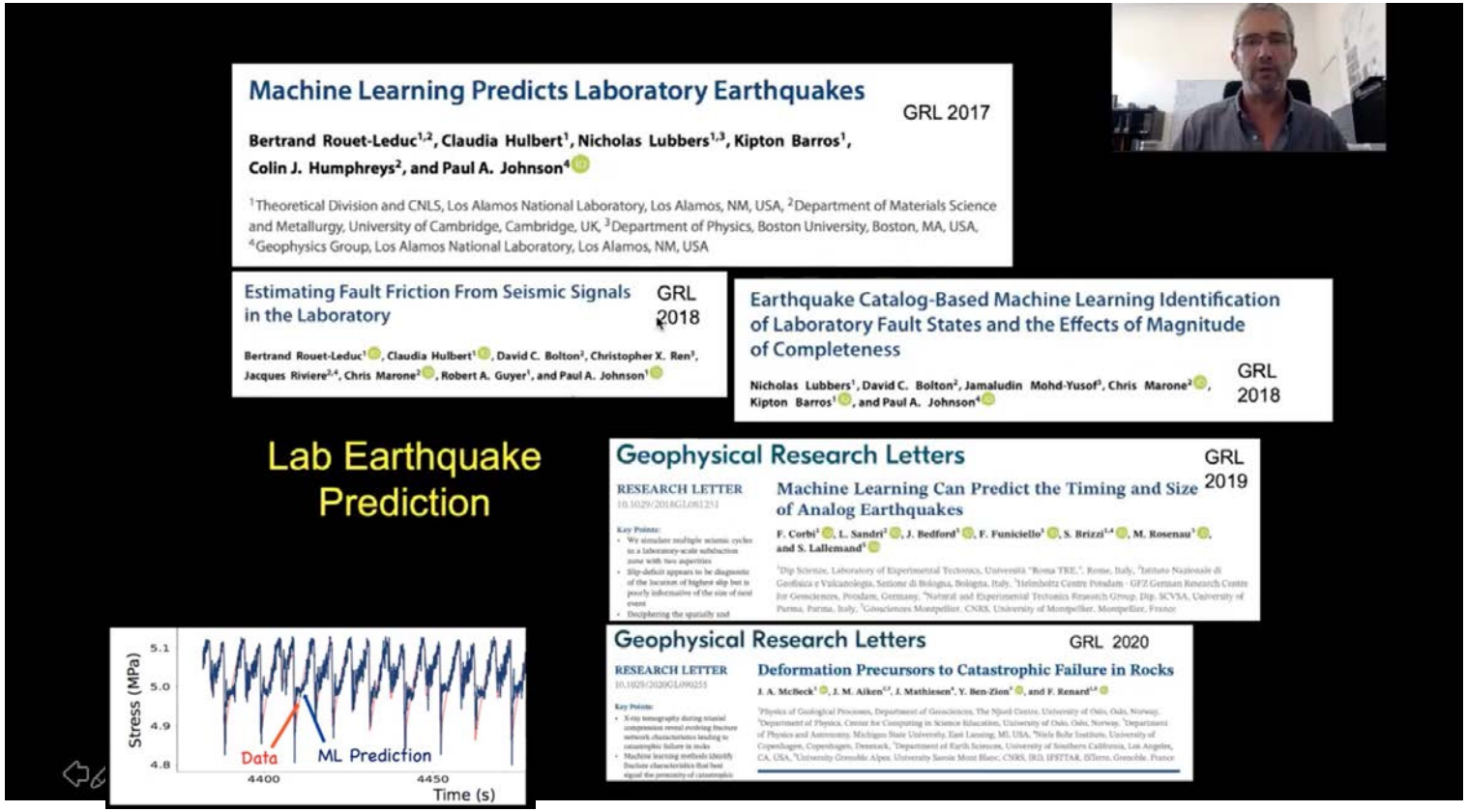
LEARNING TO SIMULATE COMPLEX PHYSICS WITH GRAPH NETWORKS

arXiv:2002.09405 [cs.LG]

- **Graph Network-based Simulators** that can learn to simulate challenging physical domains: fluids, rigid solids, deformable materials interacting with one another.
- State of a physical system represented with particles, expressed as nodes in a graph.
- Computes dynamics via learned message-passing.
- Shown that it can generalize from single-timestep predictions with thousands of particles during training, to different initial conditions, thousands of timesteps, and at least an order of magnitude more particles at test time.



Forecast Earthquakes on Tecnonic Faults



Precursors:
Consistent changes prior to failure (elastic wave speed, amplitude, fault dilation, fluid pressure/chemistry, etc.)

https://www.lanl.gov/discover/news-release-archive/2018/December/1217-machine-learning.php

SAPIENZA
UNIVERSITÀ DI ROMA

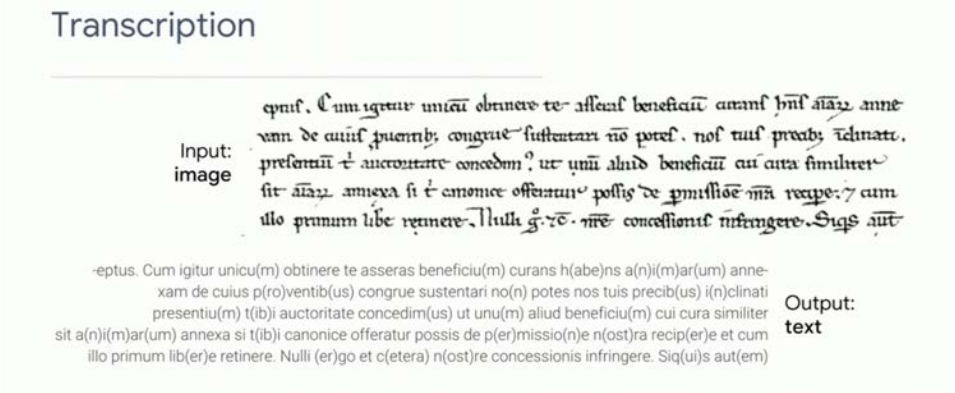
In Codice Ratio – ML based transcription of the 85 km shelves of the Vatican Secret Archive

Automatically transcribe the contents of the manuscripts. Follow a novel approach, based on character segmentation.

Govern imprecise character segmentation by considering that correct segments are those that give rise to a sequence of characters that more likely compose a Latin word.

Designed a principled solution that relies on **convolutional neural networks** and **statistical language models**.

High School students will be involved in collecting and labelling training data through the Program of digital of Fondazione "I Lincei per La Scuola" e Ministero dell'Istruzione



http://www.inf.uniroma3.it/db/icr/

SAPIENZA
UNIVERSITÀ DI ROMA

“I don’t know how long it’s going to be before the name of our field is changed from computer science to machine learning…”

Donald Knuth on Machine Learning and the Meaning of Life
(2021)