

# MATHEMATICAL METHODS IN DEEP LEARNING

SCUDO Course - Politecnico di Bari  
year 2025/2026

# INFO

---

Teacher: Gianluca Orlando

e-mail: gianluca.orlando@poliba.it

Office: Mathematics building of DMMM

Microsoft Teams channel: [tinyurl.com/scudo-mmindl](https://tinyurl.com/scudo-mmindl)

Duration of the course: 20 hours (2 CFUs)

Calendar: See Teams channel!

# THIS COURSE

---

**Objective**: Provide the mathematical tools needed to understand how (and why ???) neural networks work.

**What you should expect**:

- ✓ Point of view of a mathematician.
- ✓ Some proofs (don't get scared).
- ✓ Looking at some tools from different perspectives.
- ✓ We will build a neural network library from scratch.

**What you should not expect**:

- ✗ Point of view of a data scientist.
- ✗ See all possible tools in deep learning.
- ✗ Building a state-of-the-art GPT or similar "fireworks".

**Why**:

- Deep learning does not work "automagically".
- Need for AI education.

# PREREQUISITES

---

You are supposed to know something about this:

- Multivariate calculus (partial derivatives, chain rule, ODEs)
- Probability & statistics (random variables, expectation, variance, Gaussian, uniform, estimators)
- Linear algebra (matrices, eigenvalues/eigenvectors)
- Basic programming concepts (if/then/else, for loops, functions)

You are not supposed to know:

- Neural networks
- Optimization methods
- Python

# TOPICS

---

- General introduction to machine learning
- Linear regression, logistic regression
- Maximum likelihood estimators, cross-entropy
- Sigmoid layers
- Neural networks
- Optimization methods: gradient descent, stochastic gradient descent, momentum
- Universal approximation Theorem
- ? Regularization in neural networks ( $L^2$ , dropout)
- ? Convolutional neural networks
- ? Superposition of features
- Hands-on exercises in Python to build a DL library

# EXAM

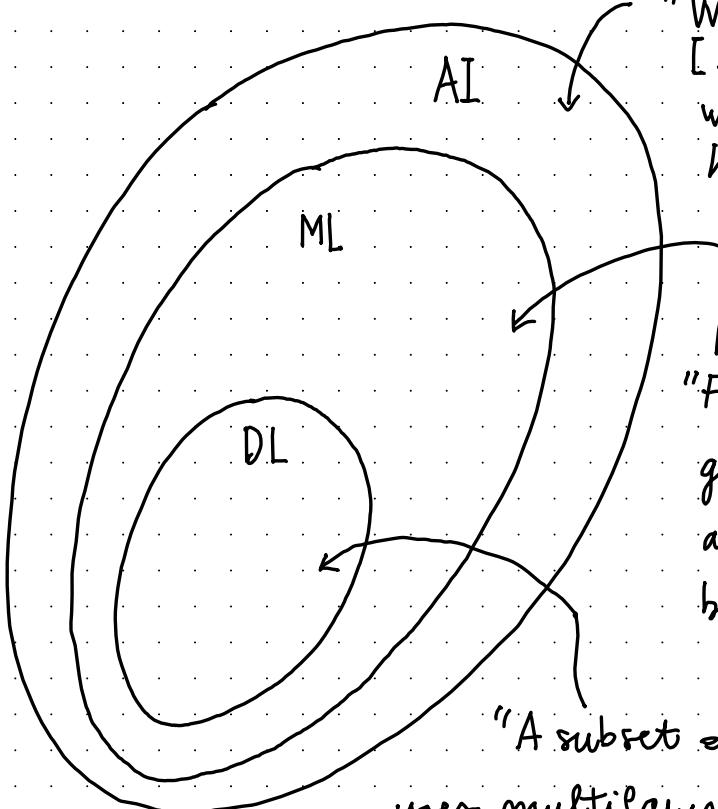
---

## Possibilities:

- Oral exam on the topics covered in the course (theoretical and exercises)
- A seminar on a theoretical topic not covered in the course (agreed upon)
- Application of the tools developed in the course to a case - study agreed - upon
- Expansion of the library with tools agreed - upon
- ( Required knowledge of LaTeX ) : writing notes for this course ( Chapters agreed - upon )

# DEEP LEARNING IN AI

---



"We call ourselves *Homo sapiens*, [...] For thousands of years, we have tried to understand how we think [...]. The field of AI attempts to build intelligent agents"

— S. Russell, P. Norvig

"Field of study that gives computers the ability to learn without being explicitly programmed"

— A. Samuel, 1959

"A subset of machine learning that uses multilayered neural networks to simulate the complex decision-making power of the human brain" — IBM

# SOME DEEP LEARNING USE CASES

---

## PREDICTIONS

Healthcare

Finance

## RECOMMENDATION SYSTEMS

Entertainment

Retail

e-commerce

## ANOMALY DETECTION

Fraud detection

Security

## COMPUTER VISION

Recognition

Autonomous systems

## REINFORCEMENT LEARNING

Gaming

Robot

Strategies

## GENERATIVE

Design

Audio

Content

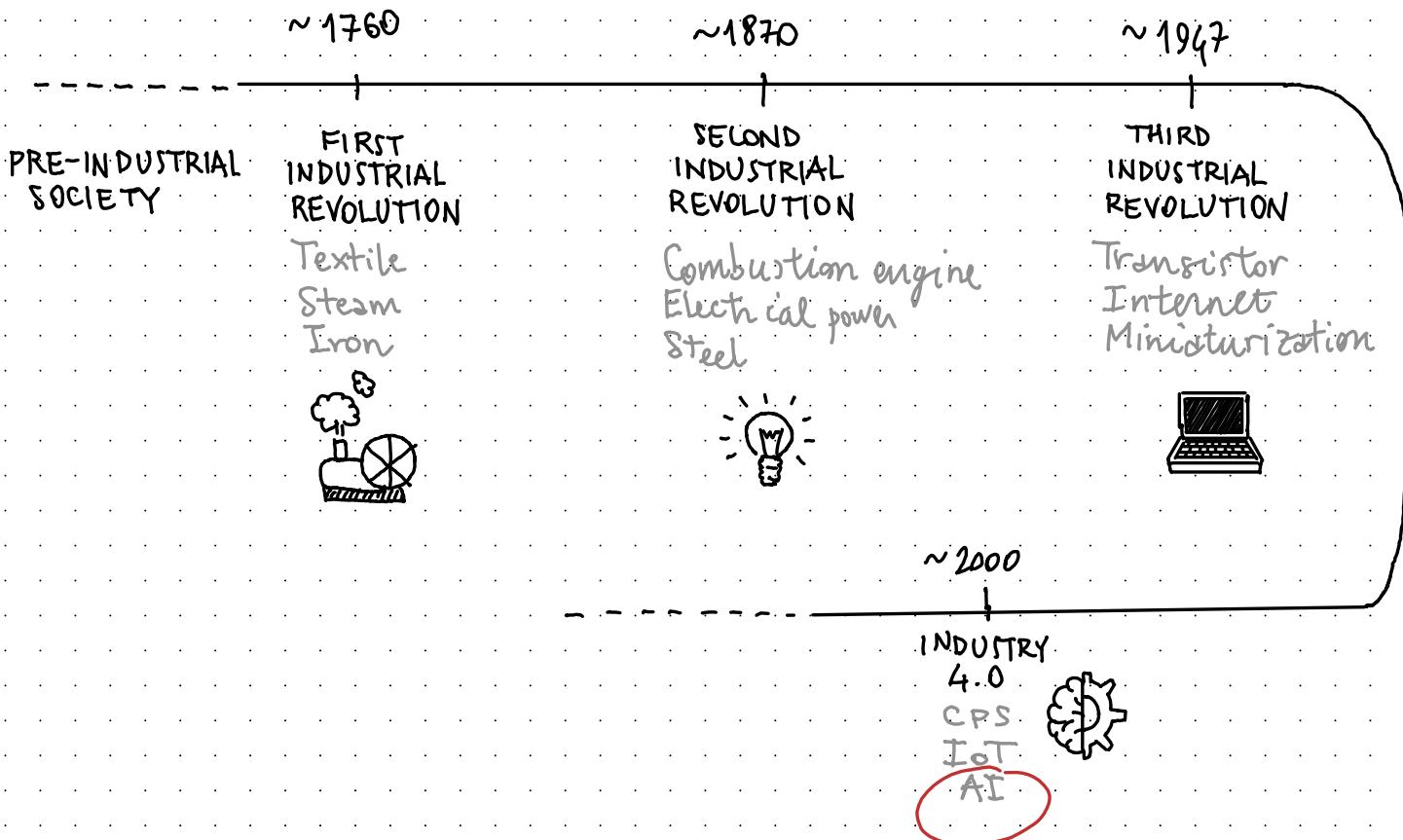
Synthetic data

## NATURAL LANGUAGE PROCESSING

Assistants

Translation

# A SHORT HISTORY RECAP



# A DEEP LEARNING TIMELINE

---

1943

McCulloch & Pitts  
ARTIFICIAL NEURON

1957

Rosenblatt  
PERCEPTRON

1965

Ivakhnenko & Lapa  
DEEP LEARNING

1967

Amari  
SGD TRAINING

1970

Linnainmaa  
BACKPROPAGATION

1979

Fukushima  
NEOCOGNITRON

2012

AI Spring

# A DEEP LEARNING TIMELINE

1676

Leibniz

CHAIN RULE

1847

Cauchy  
GRADIENT  
DESCENT

~1800

Gauss (1795)  
Legendre (1805)  
LEAST SQUARES

1951

Robbins & Monro  
SGD

1943

McCulloch & Pitts  
ARTIFICIAL NEURON

1957

Rosenblatt  
PERCEPTRON

1965

Ivakhnenko & Lapa  
DEEP LEARNING

1967

Amari  
SGD TRAINING

1970

Linnainmaa  
BACKPROPAGATION

1979

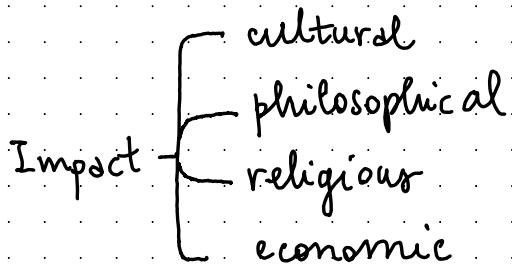
Fukushima  
NEOCOGNITRON

2012

AI Spring

# THE AI SPRING

---



- 2012 AlexNet (by A. Krizhevsky, I. Sutskever, G. Hinton)  
obtains a 15.3% error rate in ImageNet competition (second classified: > 26.1% !)
- 2016 AlphaGo by DeepMind beats Lee Sedol at Go (4-1)
- 2018 - 2024 AlphaFold by DeepMind performs unprecedented protein structure prediction
- 2022 ChatGPT is launched by OpenAI.

# IMPACT

---

Alphabet Inc.

194,07 \$ ↑595,09% +166,15 MAX

After Hours: 193,98 \$ (-0,046%) -0,090

Data e ora chiusura: 27 dic, 20:00:00 UTC-5 · USD · NASDAQ · Disclaimer

1G 5G 1M 6M YTD 1A 5A MAX



Source: Google Finance

# IMPACT

Alphabet Inc.

328,92 \$ ↑ 12.037,27% +326,21 MAX

26 gen, 09:38:56 UTC-5 · USD · NASDAQ · Disclaimer

1G 5G 1M 6M YTD 1A 5A MAX



[Confronta con](#)

Source: Google Finance

# THE NOBEL PRIZE IN PHYSICS 2024

---

"for foundational discoveries and inventions that enable machine learning with artificial neural networks"

John J. Hopfield



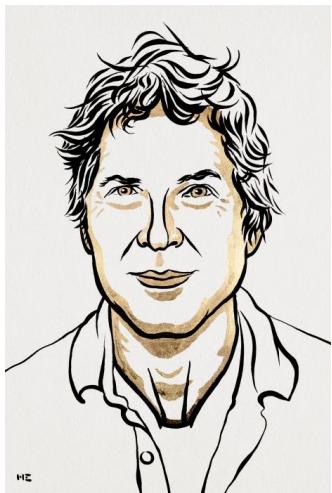
Geoffrey Hinton



# THE NOBEL PRIZE IN CHEMISTRY 2024

---

David Baker



Demis Hassabis



John Jumper



"for protein structure prediction"

# OK, BUT WHY TODAY?

---

## COMPUTATIONAL POWER

- Moore's Law (Transistor count in microchip doubles every  $\sim$  2 years)
- Fastest supercomputer :  $1.742 \rightarrow 2.746$  exaFLOPS  
 $\underbrace{\text{exa}}_{10^{18}}$

## THE ZETTABYTE ERA

- In 2016 :  $> 1\text{ZB}$  of IP traffic
- In 2024 :  $\sim 147\text{ZB}$

## NEW ARCHITECTURES

- Transformer

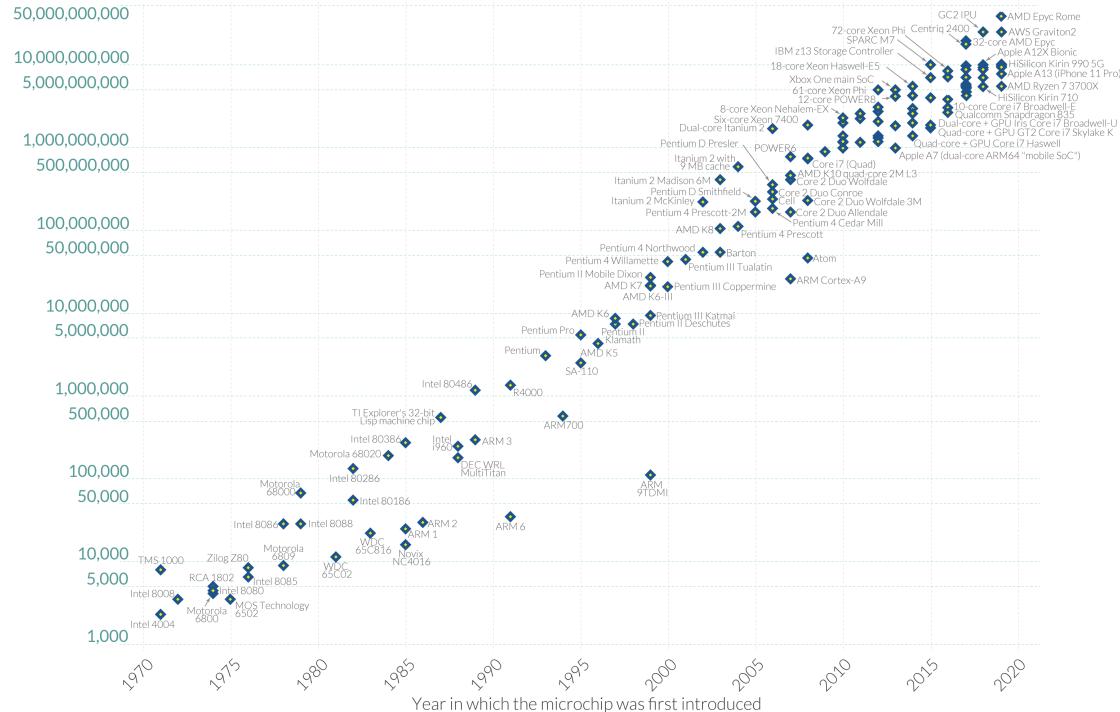
# OK, BUT WHY TODAY?

Moore's Law: The number of transistors on microchips has doubled every two years

Moore's law describes the empirical regularity that the number of transistors on integrated circuits doubles approximately every two years. This advancement is important for other aspects of technological progress in computing – such as processing speed or the price of computers.

Our World  
in Data

Transistor count



# OK, BUT WHY TODAY?

The rise of artificial intelligence over the last 8 decades: As training computation has increased, AI systems have become more powerful



The color indicates the domain of the AI system: • Vision • Games • Drawing • Language • Other

Shown on the vertical axis is the training computation that was used to train the AI systems.

10 billion petaFLOP  
Computation is measured in floating point operations (FLOP). One FLOP is equivalent to one addition, subtraction, multiplication, or division of two decimal numbers.

100 million petaFLOP  
The data is shown on a logarithmic scale, so that from each grid-line to the next it shows a 100-fold increase in training computation.

1 million petaFLOP

10,000 petaFLOP

100 petaFLOP

1 petaFLOP = 1 quadrillion FLOP

10 trillion FLOP

100 billion FLOP

1 billion FLOP

10 million FLOP

100,000 FLOP

1,000 FLOP

10 FLOP

• **Perceptron Mark I:** built in 1957/58; 695,000 FLOP

Regarded as the first artificial neural network, it could visually distinguish cards marked on the left side from those marked on the right, but it could not learn to recognize many other types of patterns.

• **ADALINE:** built in 1960 and trained on around 9,900 FLOP

An early single-layer artificial neural network.

The first electronic computers were developed in the 1940s

1956: The Dartmouth workshop on AI, often seen as the beginning of the field of AI research

The data on training computation is taken from Sevilla et al. (2022) - Parameter, Compute, and Data Trends in Machine Learning. It is estimated by the authors and comes with some uncertainty. The authors expect the estimates to be correct within a factor of two.

OurWorldInData.org - Research and data to make progress against the world's largest problems.

Pre Deep Learning Era

Deep Learning Era

1997: Deep Blue beats world chess champion Garry Kasparov

Increases in training computation accelerated, doubling roughly every 6 months.

Licensed under CC-BY by the authors

Charlie Giattino, Edouard Mathieu, and Max Roser

Source: OurWorldInData.org

# OK, BUT WHY TODAY?

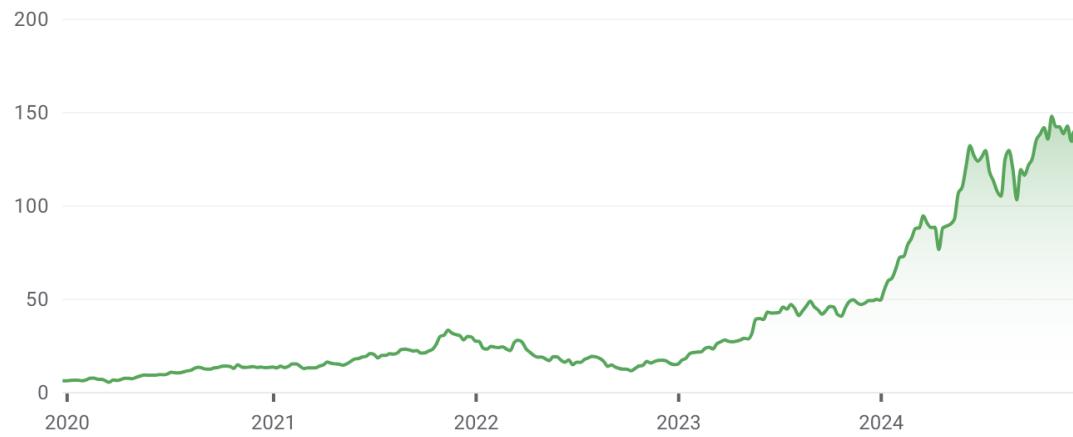
## NVIDIA

137,09 \$ ↑ 2.215,71% +131,17 5A

After Hours: 136,62 \$ (↓ 0,34%) -0,47

Data e ora chiusura: 27 dic, 19:59:59 UTC-5 · USD · NASDAQ · Disclaimer

1G 5G 1M 6M YTD 1A 5A MAX



Source: Google Finance

OK, BUT WHY TODAY?

NVIDIA

189,00 \$ ↑472.387,50% +188,96 MAX

26 gen, 09:42:44 UTC-5 · USD · NASDAQ · Disclaimer

1G

5G

1M

6M

YTD

1A

5A

MAX

250

200

150

100

50

0

2000

2005

2010

2015

2020

2025



Source: Google Finance

# OK, BUT WHY TODAY?

---

## COMPUTATIONAL POWER

- Moore's Law (Transistor count in microchip doubles every  $\sim$  2 years)
- Fastest supercomputer :  $1.742 \rightarrow 2.746$  exaFLOPS  
 $\underbrace{\text{exa}}_{10^{18}}$

## THE ZETTABYTE ERA

- In 2016 :  $> 1\text{ZB}$  of IP traffic
- In 2024 :  $\sim 147\text{ZB}$

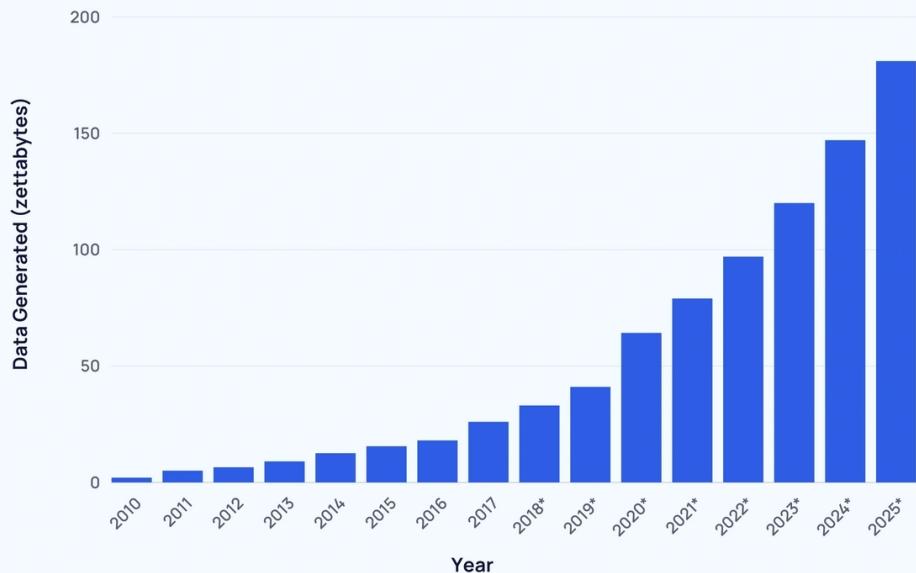
## NEW ARCHITECTURES

- Transformer

OK, BUT WHY TODAY?

---

## Global Data Generated Annually



Source: Exploding Topics