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

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 <u>from scratch</u>.

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

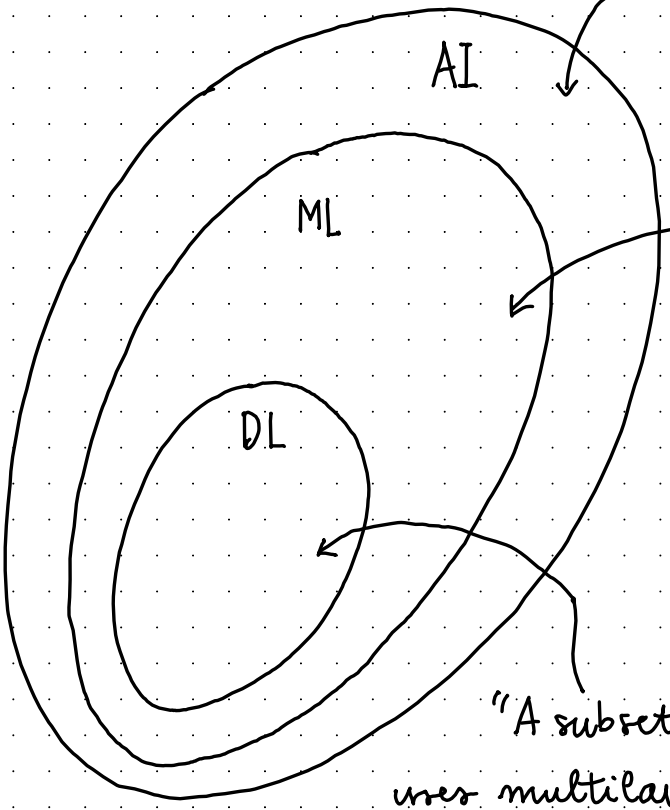
- Hands-on exercises in Python to build a DL library

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

ANOMALY DETECTION
Fraud detection
Security

COMPUTER VISION
Recognition
Autonomous systems

RECOMMENDATION SYSTEMS
Entertainment
Retail
e-commerce

REINFORCEMENT LEARNING
Gaming
Robot
Strategies

GENERATIVE
Design
Audio
Content
Synthetic data

NATURAL LANGUAGE PROCESSING
Assistants
Translation

A SHORT HISTORY RECAP

~1760

~1870

~1947

PRE-INDUSTRIAL SOCIETY

FIRST INDUSTRIAL REVOLUTION

Textile
Steam
Iron

SECOND INDUSTRIAL REVOLUTION

Combustion engine
Electrical power
Steel

THIRD INDUSTRIAL REVOLUTION

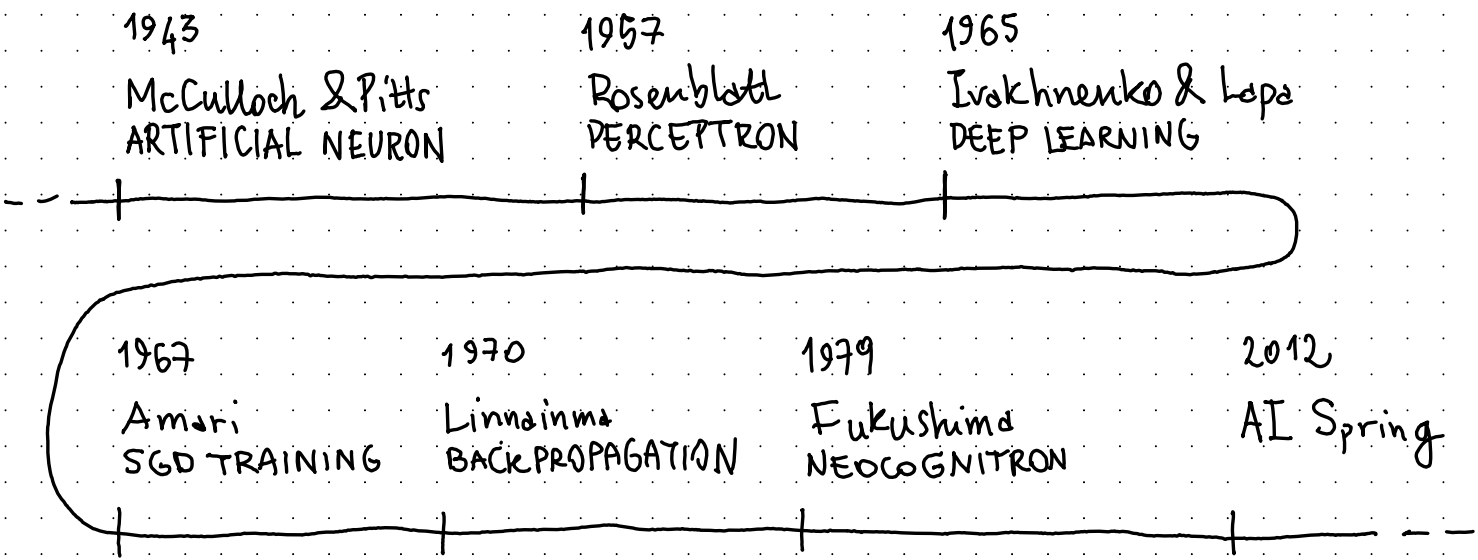
Transistor
Internet
Miniaturization

~2000

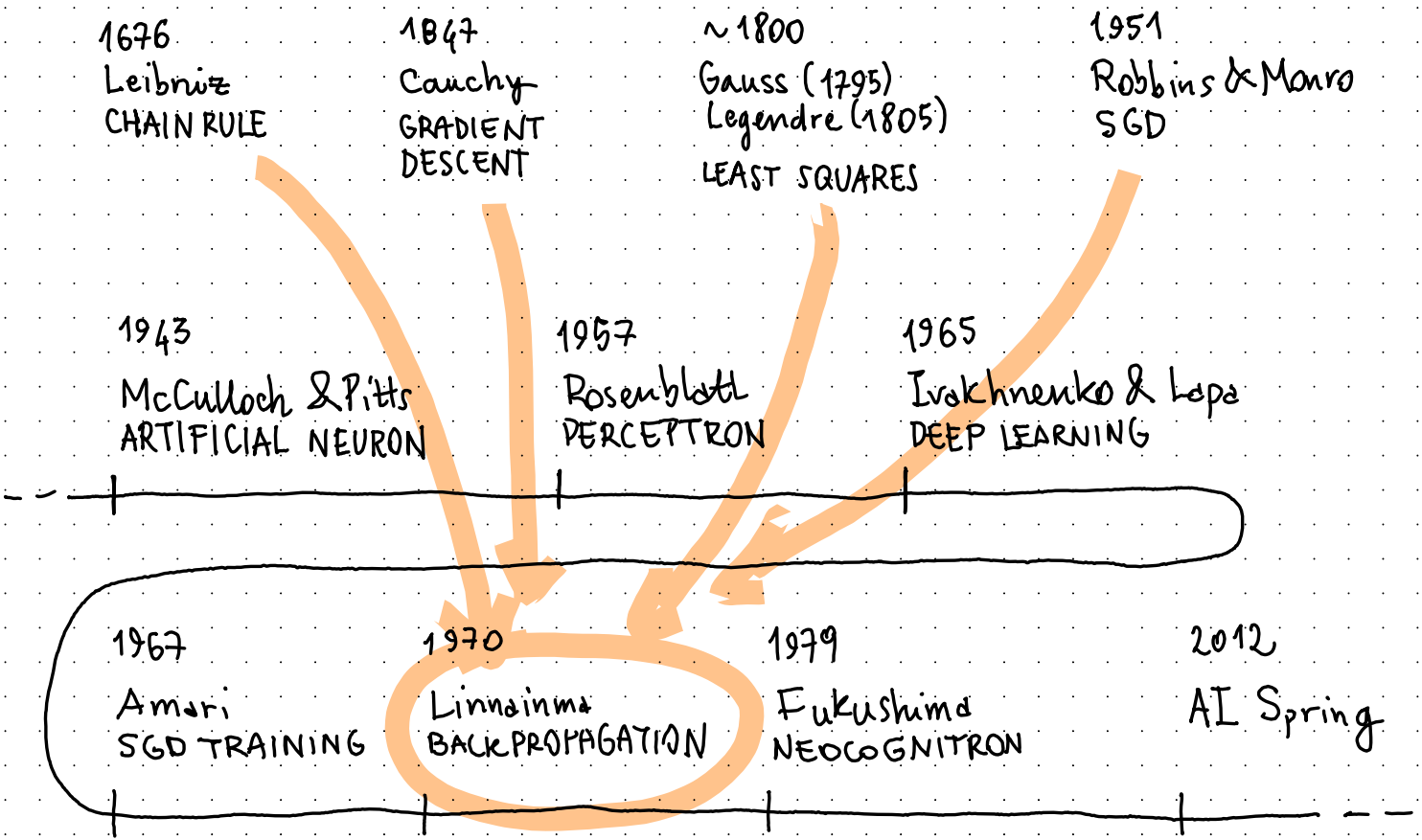
INDUSTRY 4.0

CPS
IoT
AI

A DEEP LEARNING TIMELINE



A DEEP LEARNING TIMELINE



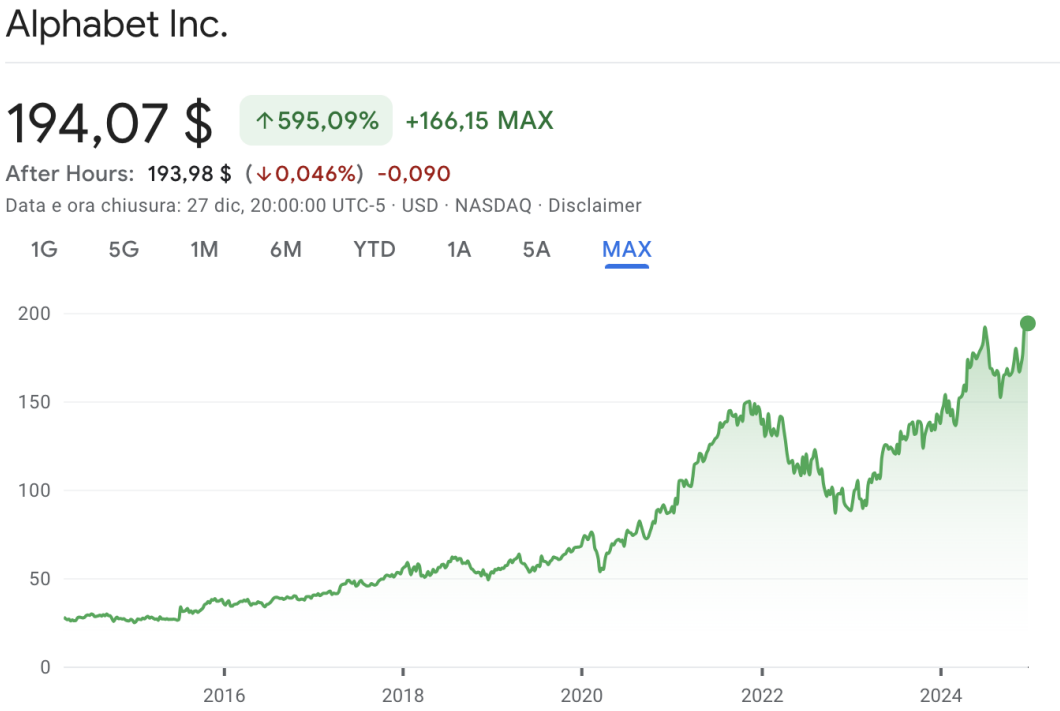
THE AI SPRING

Impact
- cultural
- philosophical
- religious
- economic

- 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



Source: Google Finance

IMPACT

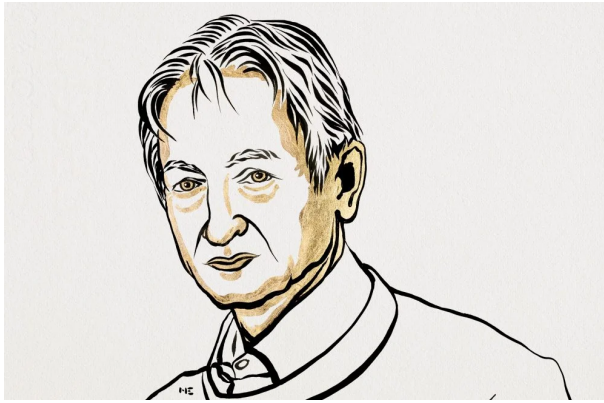
Source: Google Finance

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 ~ 2 years)

- Fastest supercomputer: 1.742 → 2.746 exaFLOPS (10^{18})

THE ZETTABYTE ERA

- In 2016: > 1 ZB of IP traffic
- In 2024: ~ 147 ZB

NEW ARCHITECTURES

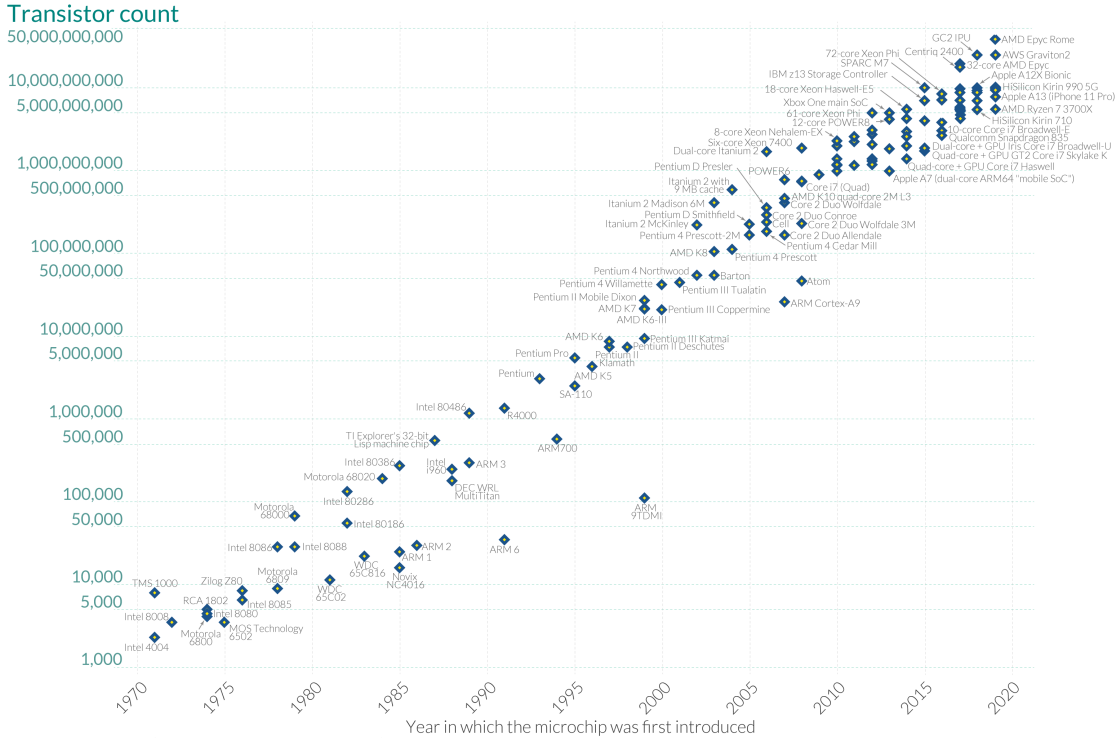
- Transformer

OK, BUT WHY TODAY?

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

Our World in Data

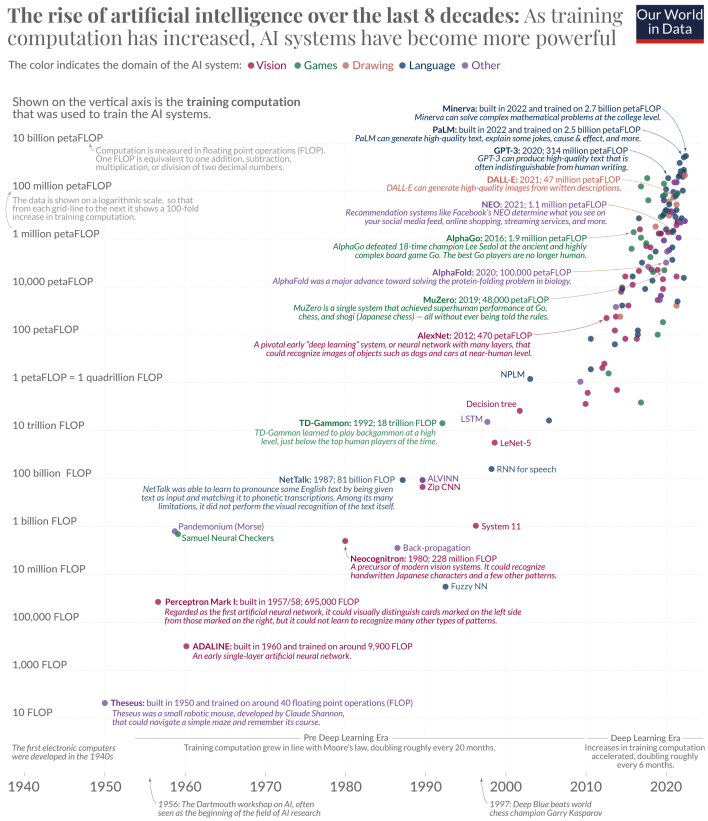
Moore's law describes the empirical regularity that the number of transistors on integrated circuits doubles approximately every two years. This advancement is important for other aspects of technological progress in computing – such as processing speed or the price of computers.



Data source: Wikipedia (wikipedia.org/wiki/Transistor_count)
OurWorldinData.org – Research and data to make progress against the world's largest problems.

Source: OurWorldInData.org

OK, BUT WHY TODAY?



Source: OurWorldInData.org

OK, BUT WHY TODAY?

Source: Google Finance

OK, BUT WHY TODAY?

Source: Google Finance

OK, BUT WHY TODAY?

COMPUTATIONAL POWER

- Moore's Law (Transistor count in microchip doubles every ~ 2 years)

- Fastest supercomputer: 1.742 → 2.746 exaFLOPS (10^{18})

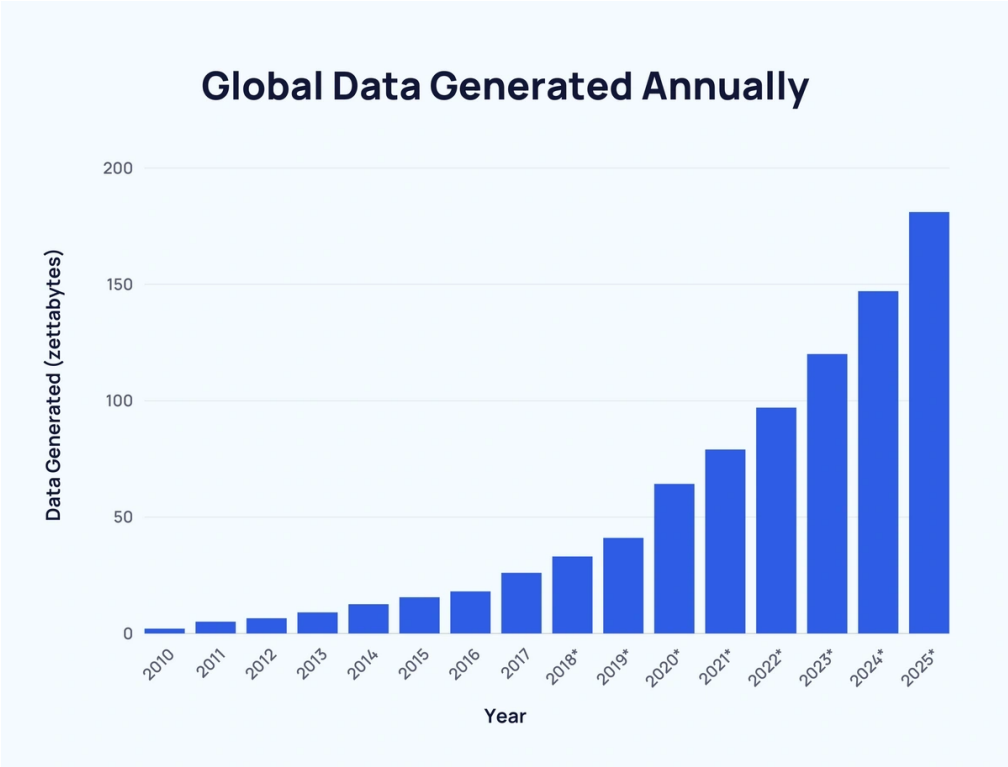
THE ZETTABYTE ERA

- In 2016: > 1 ZB of IP traffic
- In 2024: ~ 147 ZB

NEW ARCHITECTURES

- Transformer

OK, BUT WHY TODAY?



Source: Exploding Topics