Università di Catania

Prospettive Informatiche

From Deep Learning to ChatGPT: l'evoluzione dell'Intelligenza Artificiale, dai modelli profondi al linguaggio naturale

2 Marzo 2026

Introduzione al Deep Learning

Who Am I?

Prof. Antonino Furnari

Contacts:

- http://antoninofurnari.github.io
- antonino.furnari@unict.it

Courses:

- Fondamenti di Analisi dei Dati (L-31)
- Deep Learning - Advanced Models and methods (LM-18)

Member: Image Processing Laboratory (IPLAB)

Research Focus: Embodied AI, Egocentric Vision, Video Understanding

When Computer Science "Hacked" Physics & Chemistry

A Turning Point in Science

- **Physics Nobel:** Hopfield & Hinton for Artificial Neural Networks
- **Chemistry Nobel:** David Baker, Demis Hassabis, and John Jumper (DeepMind) for AlphaFold

Why This Matters

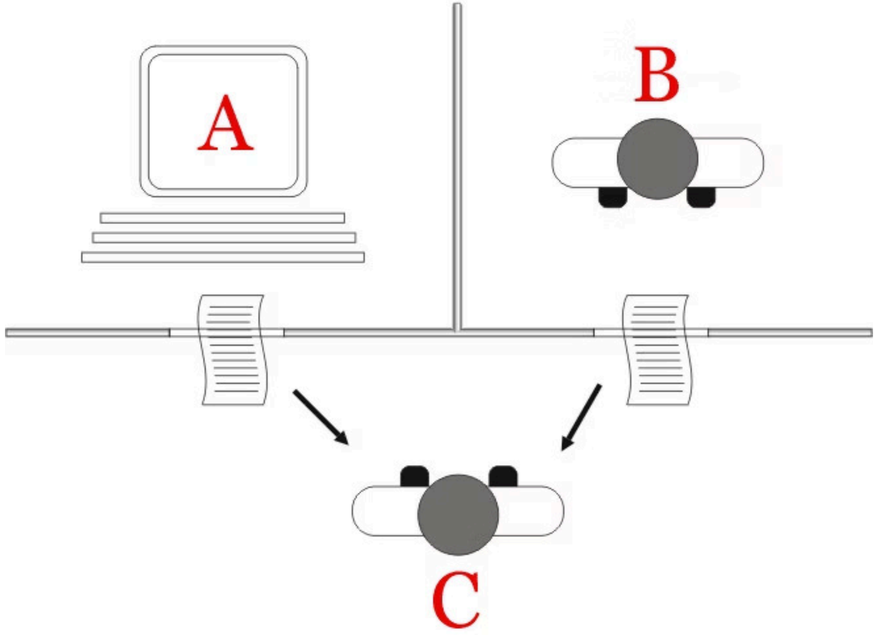
AI has graduated from "Computer Science" to "**Natural Science**"

> We are no longer just writing software—we are modeling the physical world through learned representations

"Can Machines Think?"

Alan Turing, 1950 — The "Imitation Game"

> If a machine behaves intelligently, does it matter if it has a soul?



https://en.wikipedia.org/wiki/Turing_test#/media/File:Turing_test_diagram.png

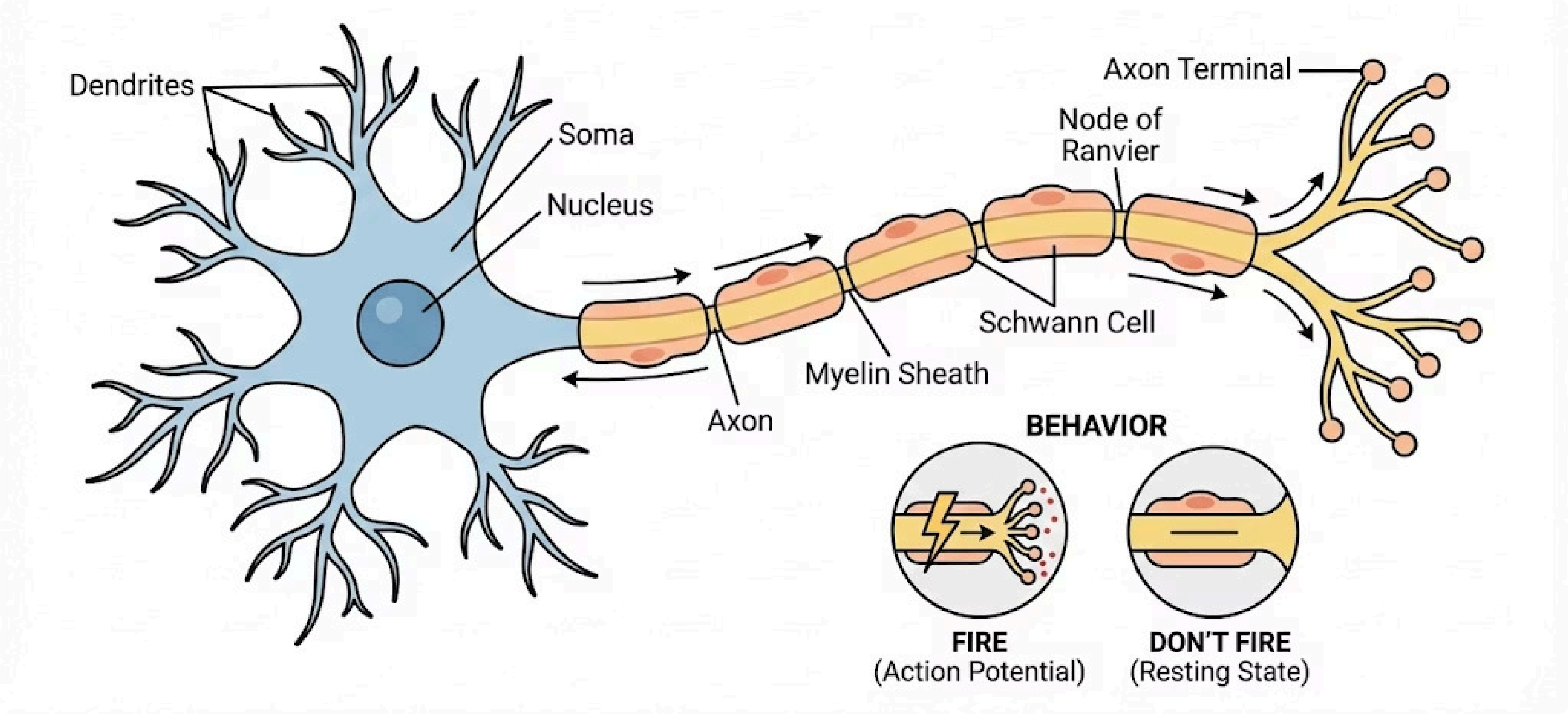
The Biological Blueprint

Long before computers, nature engineered the ultimate intelligence: the human brain.

Its intricate architecture, featuring billions of interconnected neurons, has served as the foundational blueprint for artificial intelligence since its inception. The brain's ability to learn, adapt, and process information in parallel inspired the very concept of neural networks.

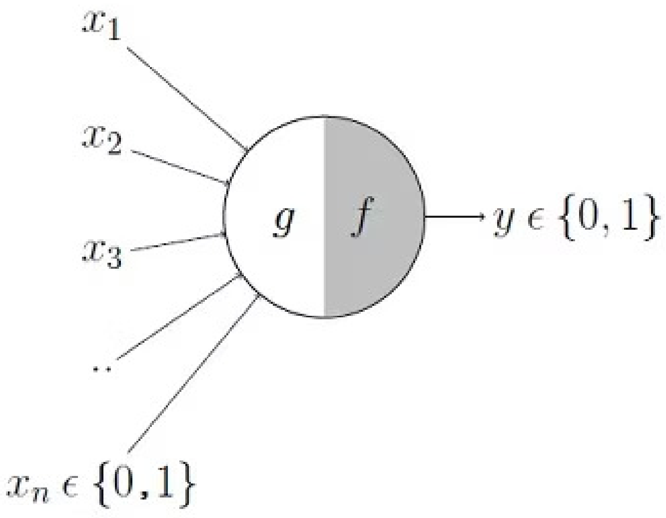
> Understanding the brain's mechanisms provides a way to unlocking AI capabilities.

The Biological Neuron



The McCulloch-Pitts Neuron (1943)

Inspired by the biological neuron, Warren McCulloch and Walter Pitts developed the first simplified mathematical model of a neural network in 1943. This binary threshold unit formalized how neurons could perform basic logical operations.



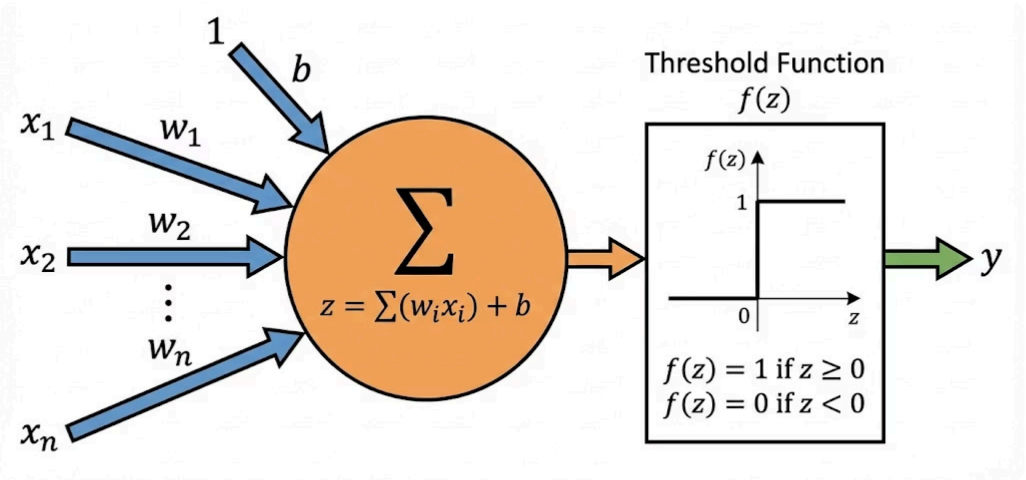
This foundational model demonstrated that neural networks could, in principle, compute any computable function, laying the theoretical groundwork for artificial intelligence.

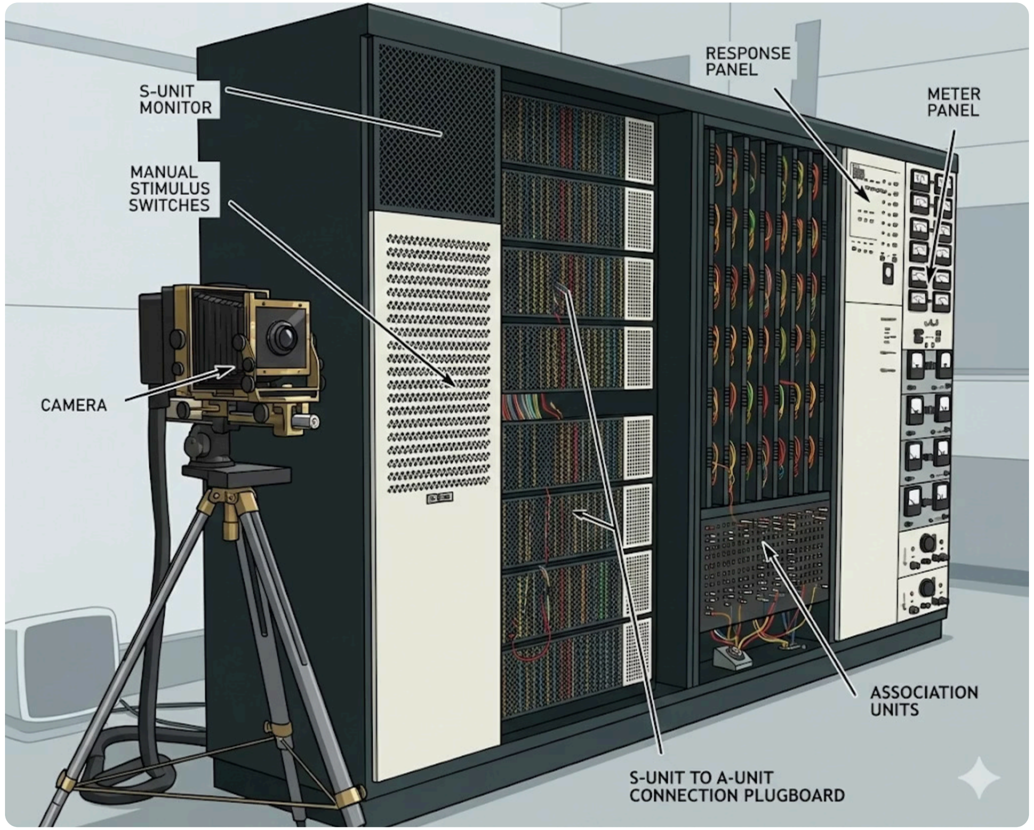
[McCulloch-Pitts Neuron — Mankind's First Mathematical Model Of A Biological Neuron]

The Perceptron — "Let's Build It" (1958)

Frank Rosenblatt tried to mimic biology with hardware.

A closet of wires and potentiometers — not code. A single-layer attempt that learned simple patterns but lacked the hierarchy the brain uses.





Perceptrons in action

YouTube

Perceptron Research from the 50's & ...

Short clip about perceptron research done in the 1950's and 1960's.

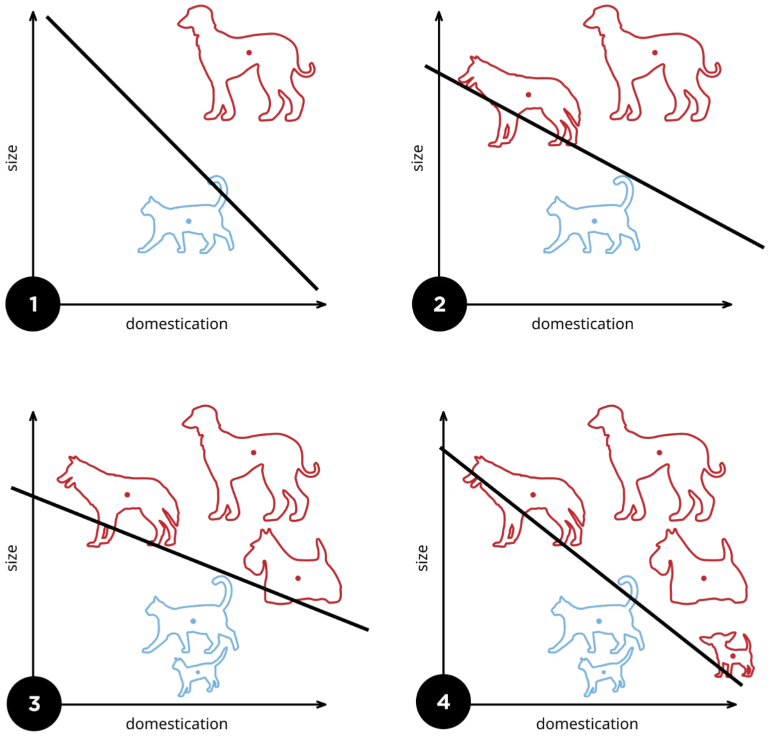
Perceptrons had a learning algorithm based on the comparison between the **output** of the perceptron and its **expected output**.

This happens by providing the perceptron with **labeled examples**, i.e., inputs and associated outputs.

As initial perceptrons were built in hardware, these inputs and outputs could be provided with **a sensing stack** to "perceive" the input visually and **mechanical switches** provide the actual output.

The machine is presented many examples until it **learns** to recognise objects by itself.

The Perceptron - Linear Classification Problems



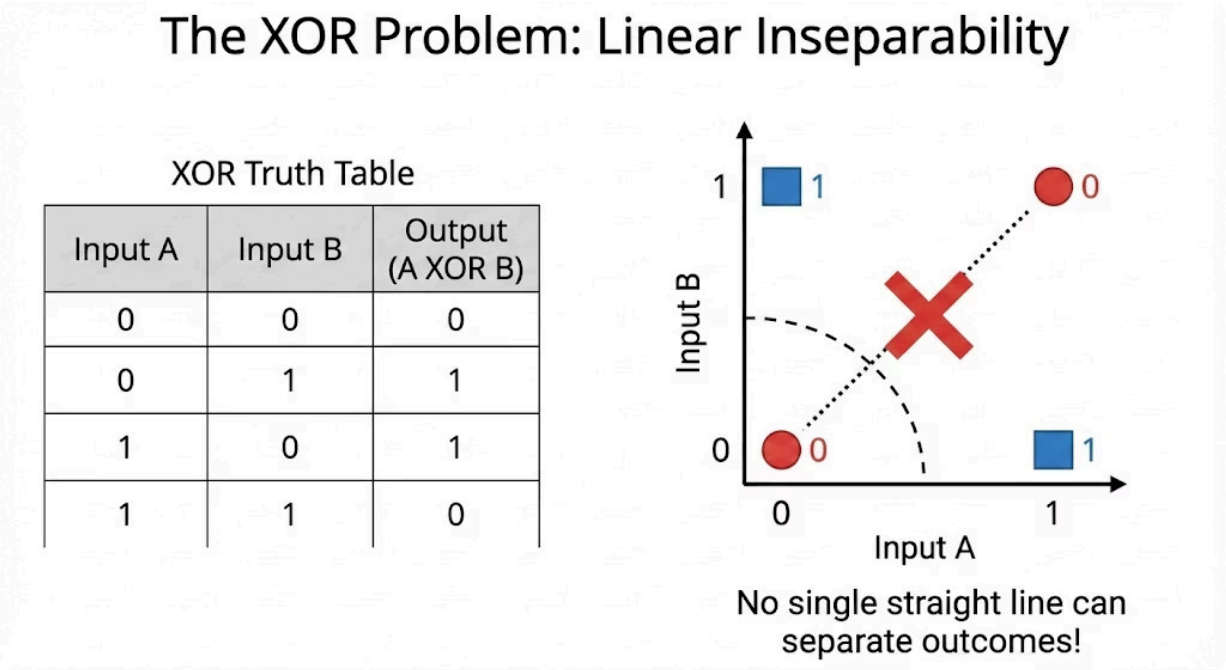
In practice, the perceptron models a very simple function. If the inputs are only two, the decision rule (fire / don't fire) becomes:

$$ax + by + x \geq 0$$

Based on the values of a and b, this finds a line and makes decisions based on the line, as shown in the illustration on the light, which shows **how the perceptron adjusts its weights** to identify the line **separating two classes of objects** based on input values.

Image Source

The Cold Shower — XOR & the First AI Winter (1969)



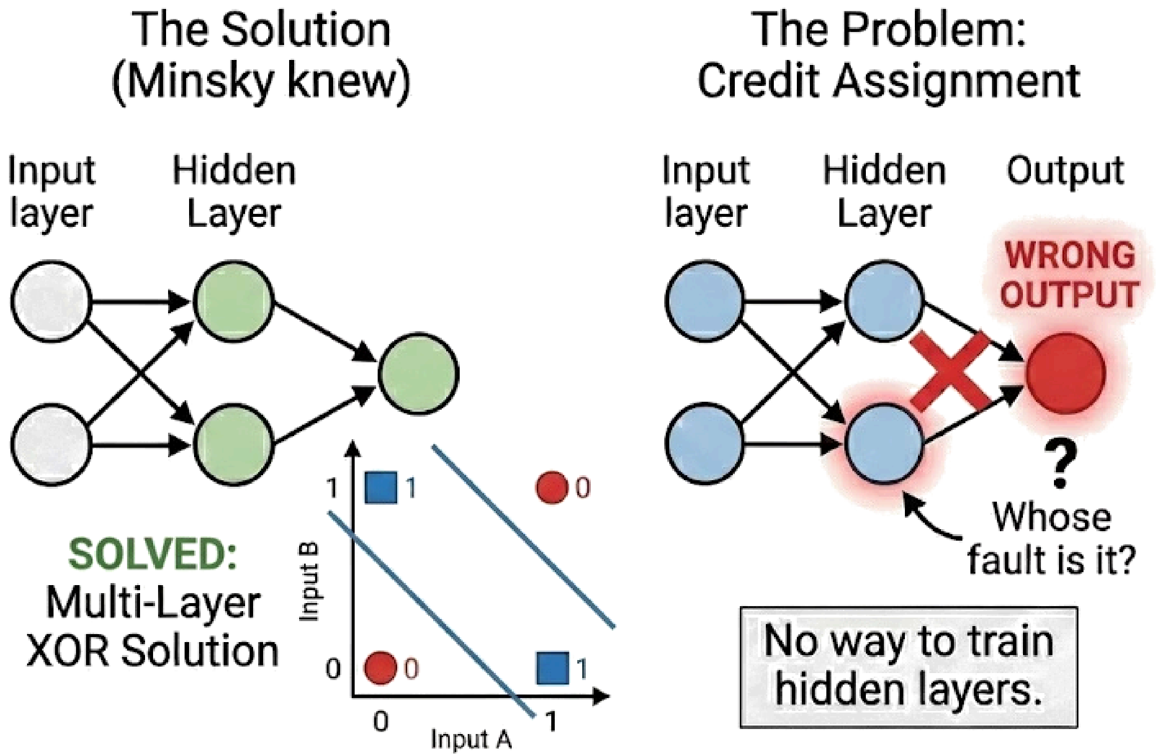
Minsky & Papert (MIT)

Proved a single layer **cannot solve XOR** — non-linear problems.

You can't separate a checkerboard with one straight line.

> Result: Funding dried up. The "AI Winter" began.

The "Yes, But..." — Multi-Layer Perceptrons



The Problem

No way to train hidden layers

Credit Assignment

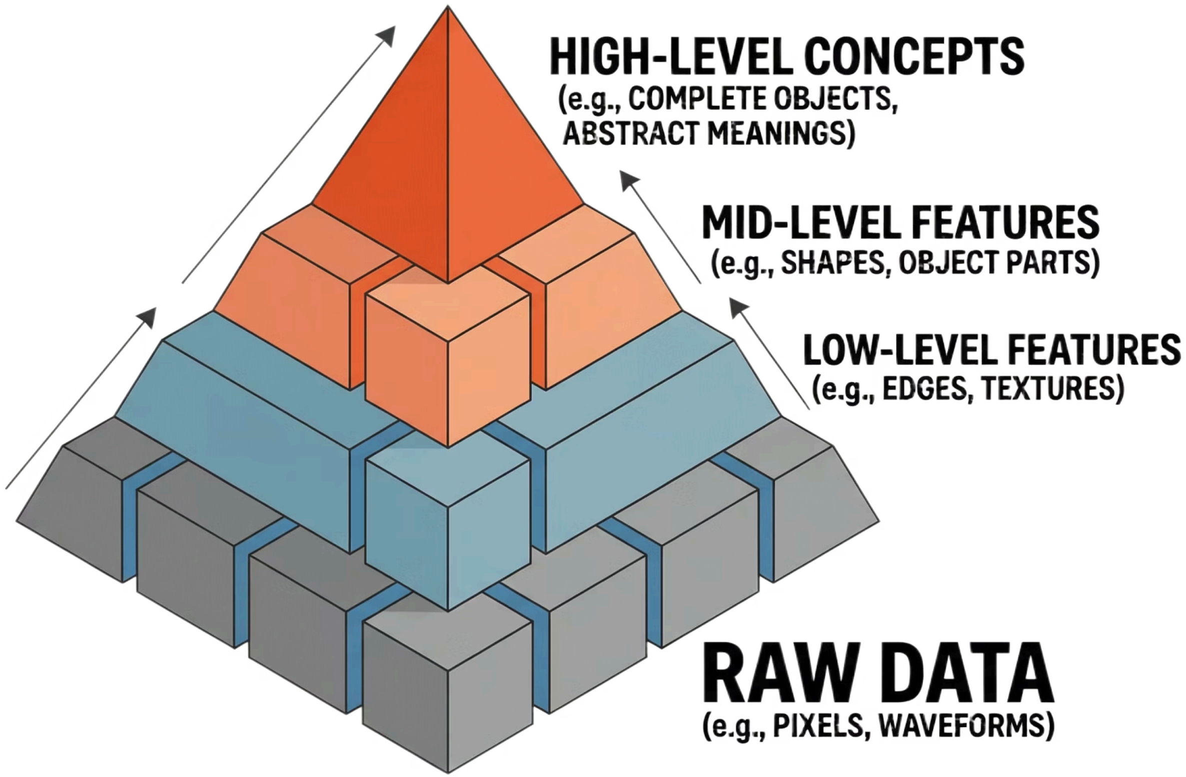
If the output is wrong — whose fault is it?

The Result

MLPs remained theoretical curiosities

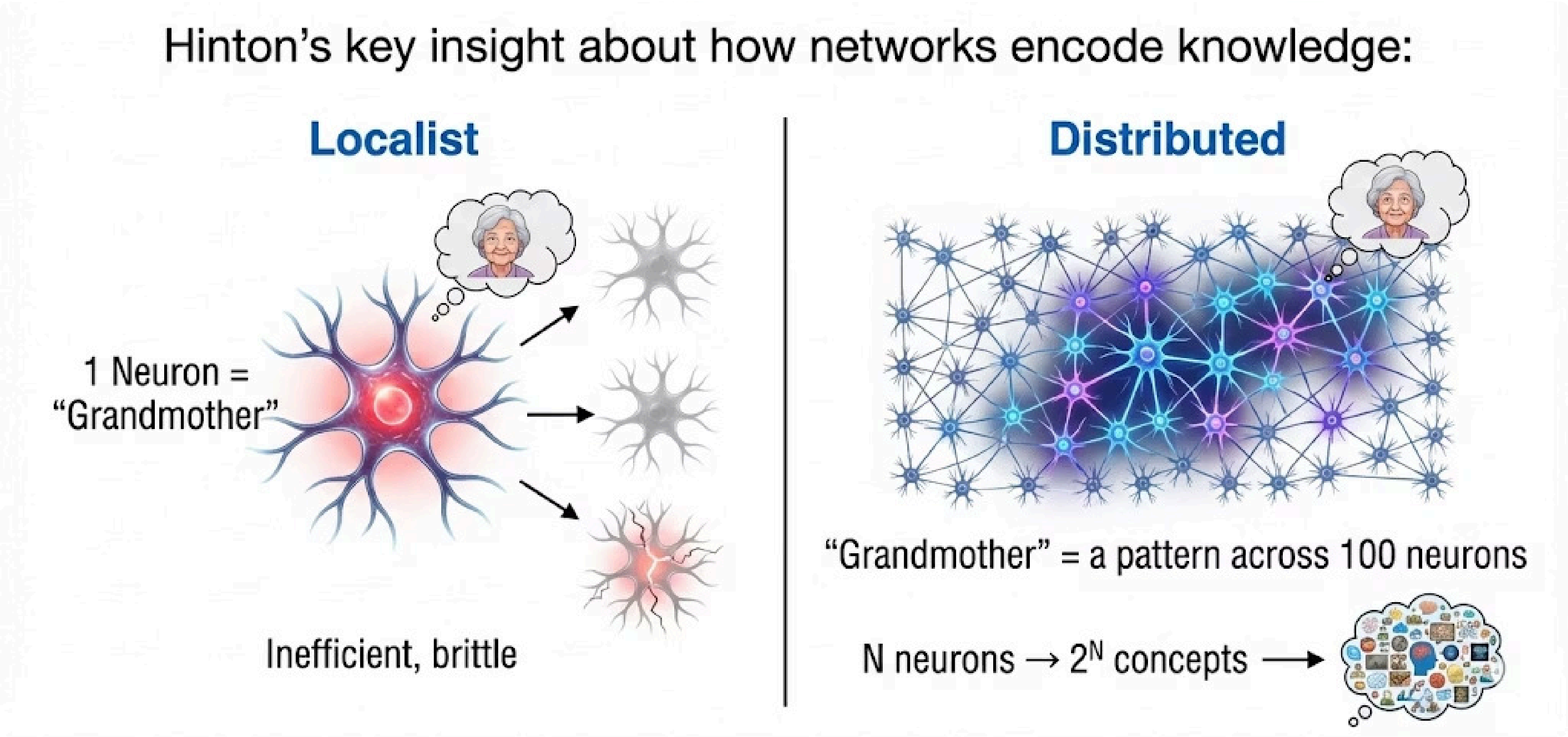
> We'll have to wait a bit for a suitable solution to this problem

Enter Deep Learning



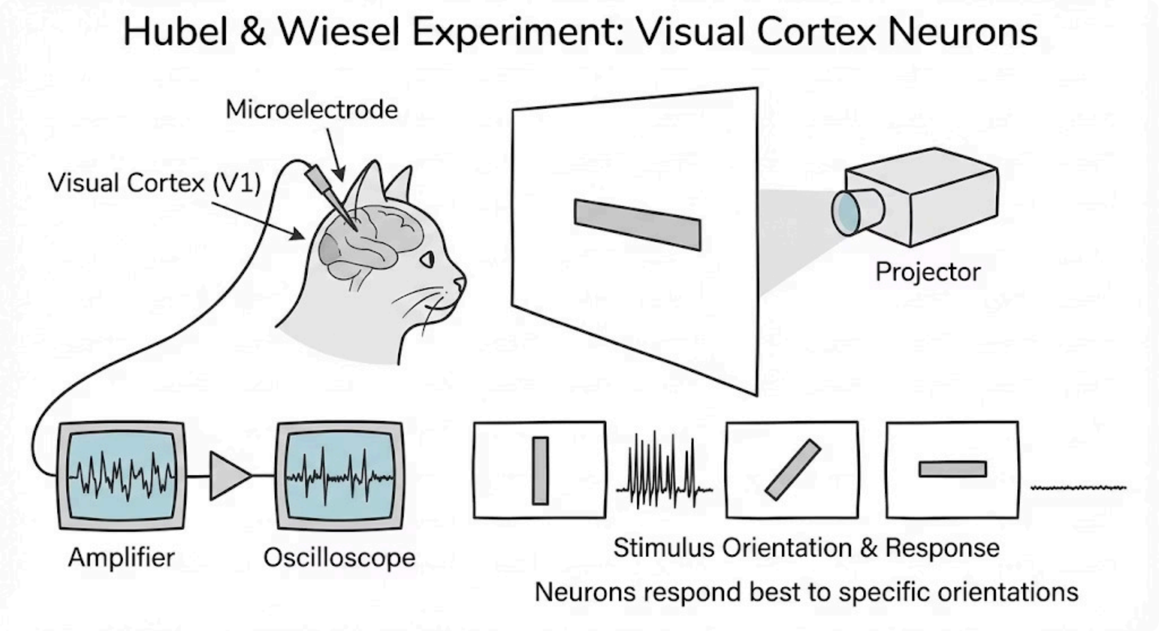
- Early models like the Perceptron **struggled with complex, non-linear problems**.
- Learning is "**shallow**", as we wish to learn a function of the input in just one pass (1 neuron or 1 layer of neurons).
- **Deep Learning** proposes a hierarchical approach, inspired by the human brain's visual cortex, where neural networks build sophisticated representations of the input, moving from simple elements to complex concepts.
- The input is raw data and each layer of the hierarchy is tuned to understand **increasingly abstract concepts**.

Distributed Representations (1986)



Hubel & Wiesel (1959–1962)

Layers of neurons



Hubel & Wiesel's Discovery

The brain is **hierarchical**, not a blank slate.

V1 — Simple Cells

Detects edges

V2 — Complex Cells

Detects motion & shapes

Intelligence

Layers of increasing abstraction

Fukushima's Neocognitron (1980)

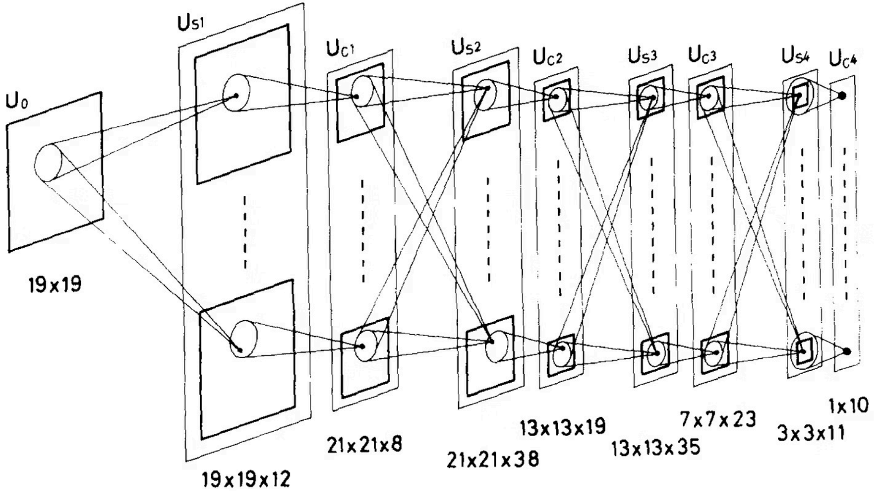


FIGURE 2. Hierarchical network structure of the neocognitron. The numerals at the bottom of the figure show the total numbers of S- and C-cells in individual layers of the network which are used for the handwritten numeral recognition system discussed in Section 4.

Mimicking Hierarchy

Following what was known about the brain, Fukushima invented the Neocognitron, which arranged simple neurons in a hierarchical structure and used convolutions.

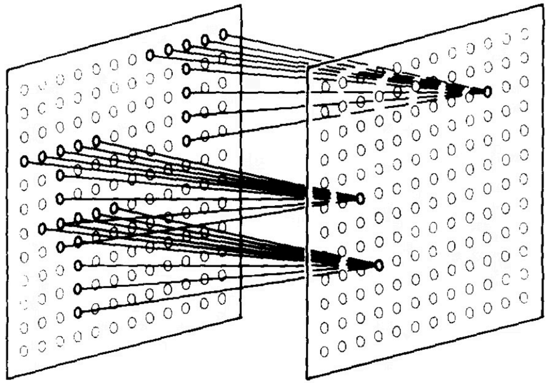


FIGURE 3. Illustration showing the spatial arrangement of the connections converging to single cells of a cell-plane.

Neocognitron in Action

YouTube

Neocognitron Movie – Part #2

This movie, produced in 1986, introduces the Neocognitron, which is a neural network...

The Neocognitron was able to recognize handwritten digits by using a layered architecture made of specialized neurons.

The Neocognitron used a **self-organizing** unsupervised training algorithm followed by user-assisted supervised components.

However, the algorithm was still very basic and a global training algorithm was missing, with the model **trained in discrete stages**, which prevented from reaching **a global optimum**.

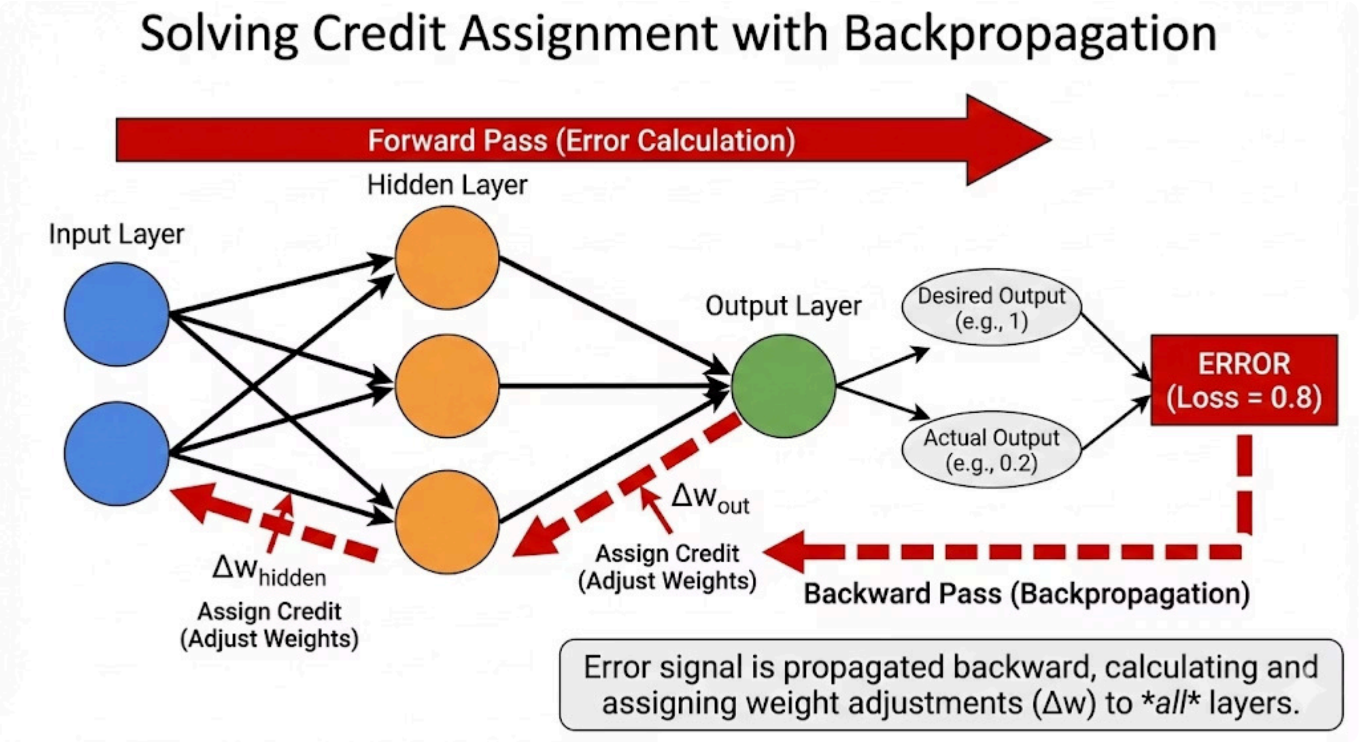
Backpropagation (1986)

Rumelhart, Hinton & Williams

Not magic — the **Chain Rule** (Leibniz) applied to errors.

$$\frac{d}{dx}f(g(x)) = f'(g(x)) \cdot g'(x)$$

Calculate exactly how much each neuron contributed to the error. Adjust it. Credit Assignment problem: solved.

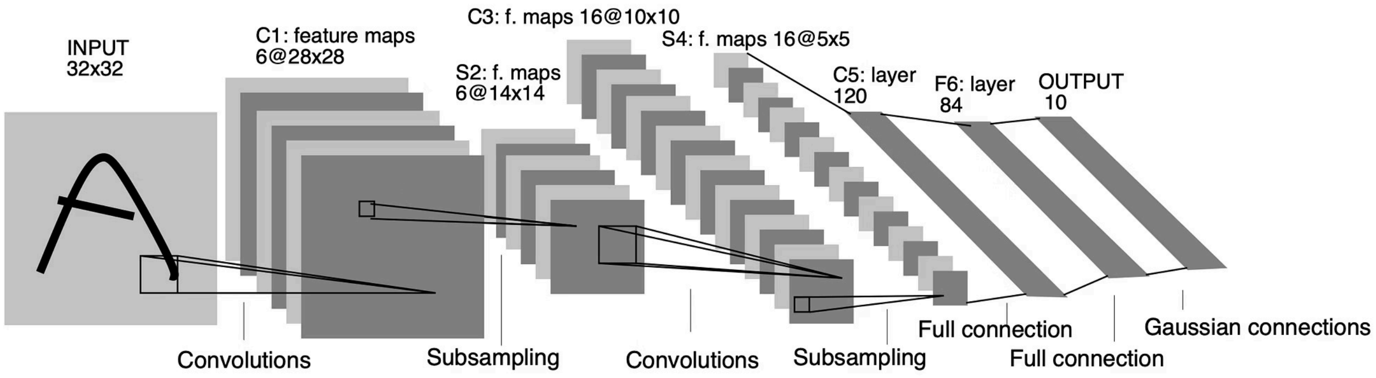


LeCun's LeNet (1998)

Engineering Hierarchical Processing

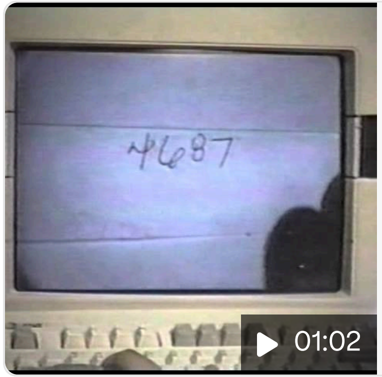
Combined **Convolution** (Structure) + **Backpropagation** (Learning).

Real-world proof: **reading zip codes**.



1	2	3
Fukushima's Architecture	**Hinton's Backprop**	**LeNet**

LeNet in Action



YouTube

Convolutional Network Demo f...

This is a demo of "LeNet 1", the first convolutional network that could...

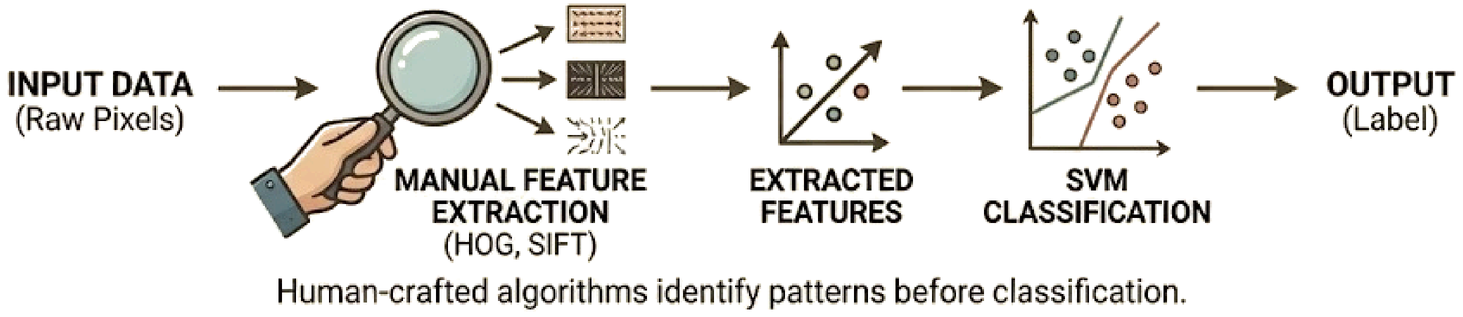
LeNet was able to recognise handwritten and machine-written letters.

This was integrated in automated check reading system in an ATM.

One of the first examples of commercial success for a Deep Learning application.

The SVM Era — The "Adult in the Room" (2000s)

Support Vector Machines dominated machine learning from the mid-90s to the early 2010s thanks to their mathematical properties.



> The only learnable component is the SVM classifier, hence leading to a form of "shallow learning", as in perceptrons, but now the input is not raw data.

Why SVMs Won

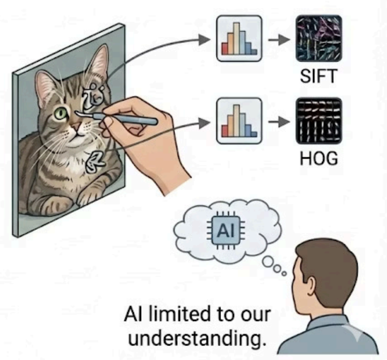
Mathematically elegant with a guaranteed global minimum — same result every re-train. No black box, fully interpretable.

They were extended to non-linear classification with the Kernel trick.

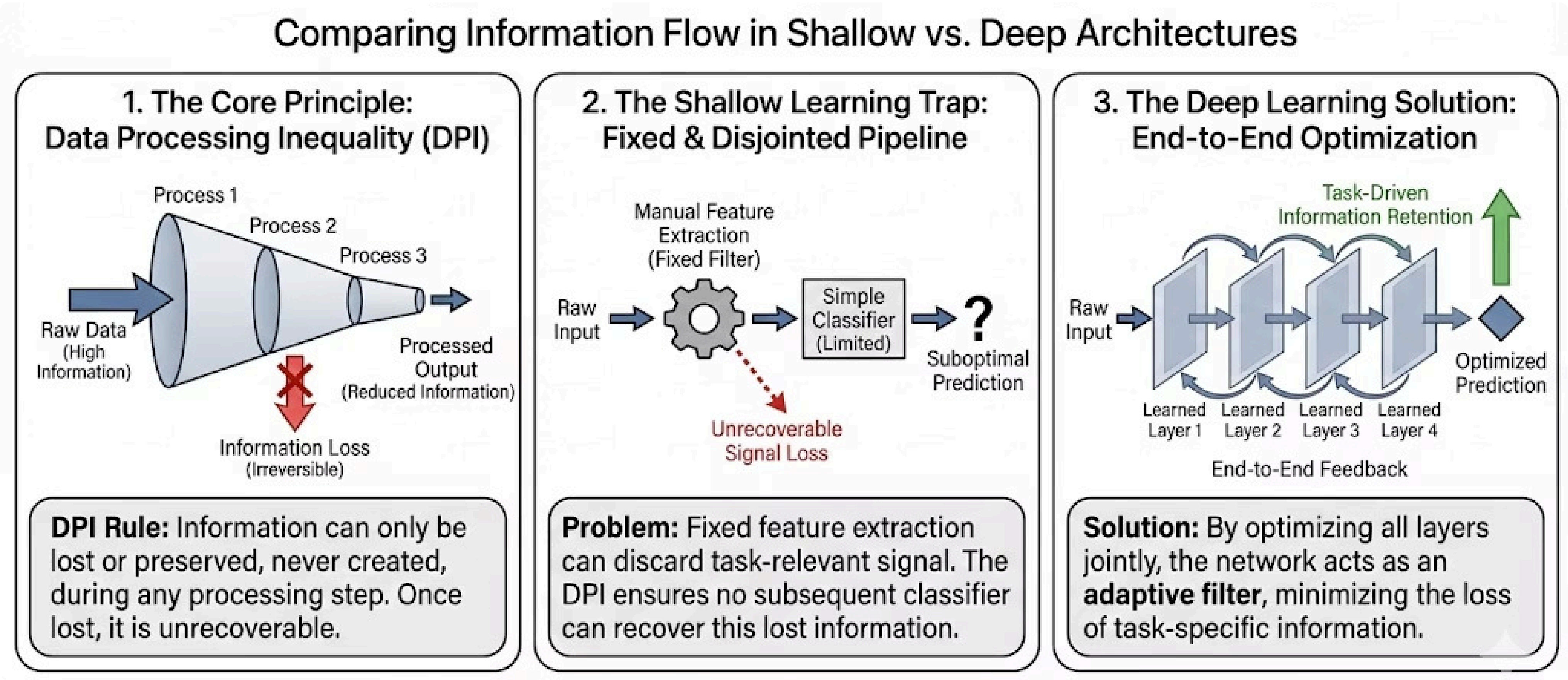
The Limitation

Required hand-crafted features (SIFT, HOG). AI was bottlenecked by human understanding of the data.

In contrast, Neural Network could **in principle** process raw data and find the best features for the problem at hand.



Shallow vs Deep Learning



ImageNet — The Fuel (2009)

Fei-Fei Li (Stanford)

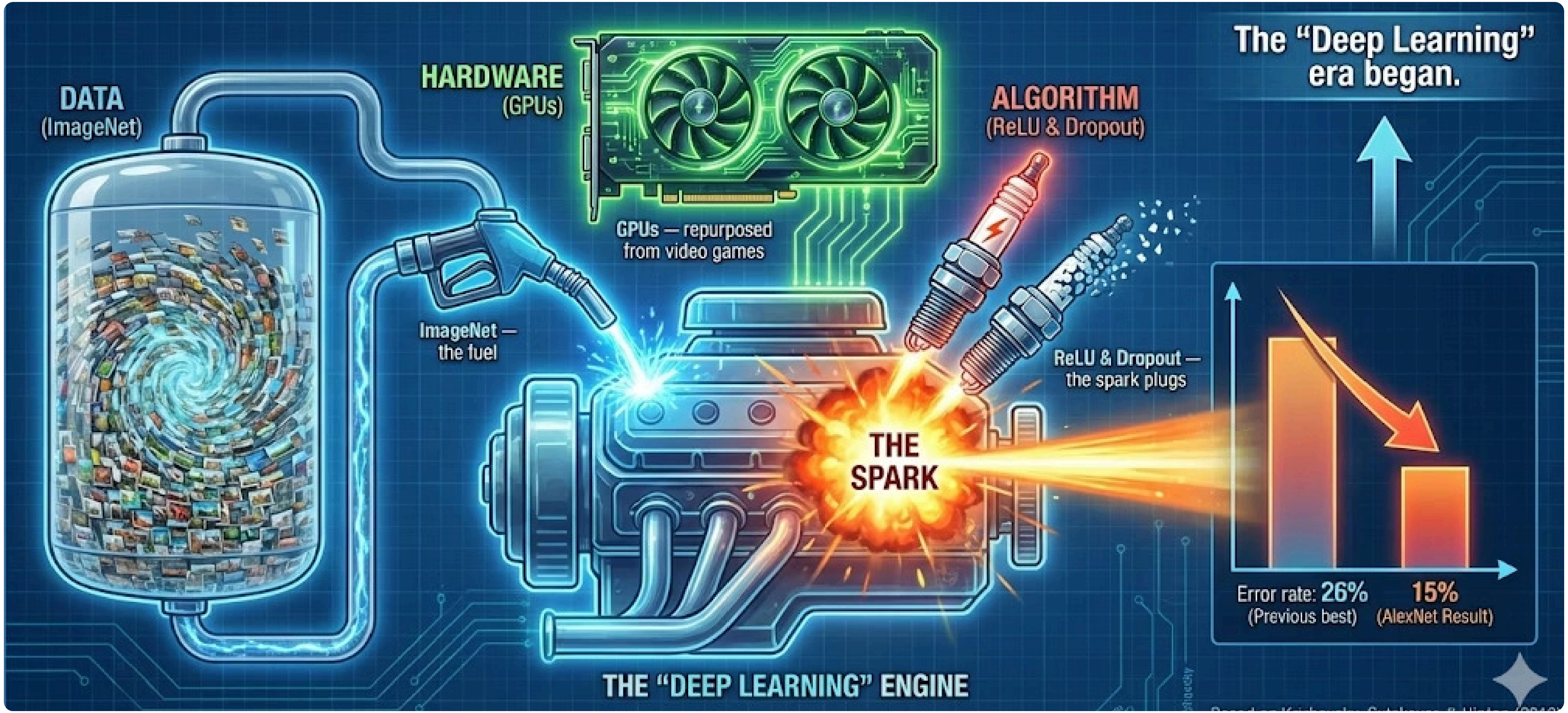
Algorithms weren't the bottleneck — **data** was.

14M	21K
Labeled Images	Categories

ImageNet was associated with the **ILSVRC (ImageNet Large Scale Visual Recognition Challenge)**, an annual competition launched in 2010 to benchmark progress in image recognition and object detection.

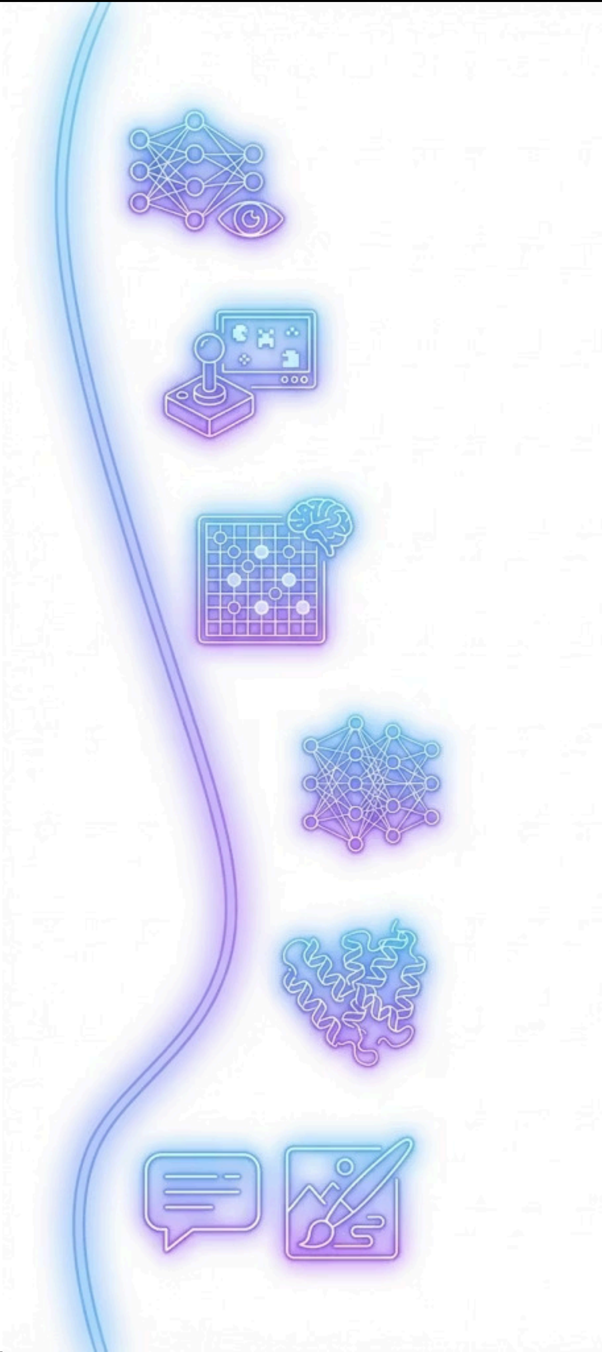
This challenge became the key metric for **measuring AI progress in computer vision**, and winning it became the holy grail for research teams worldwide.

AlexNet — The Spark (2012)

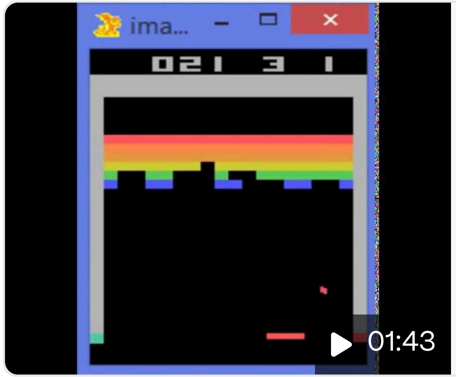


The Explosion — A Decade of Breakthroughs

1. **2012 — AlexNet**
   The Big Bang of Computer Vision
2. **2013 — DeepMind Atari**
   AI learns games from raw pixels
3. **2016 — AlphaGo**
   Search space larger than the universe
4. **2017 — Transformers**
   "Attention Is All You Need"
5. **2020 — AlphaFold 2**
   Solves protein folding
6. **2022 — GenAI Boom**
   ChatGPT & Midjourney — AI creates



DeepMind AI Learns Atari (2013)



YouTube

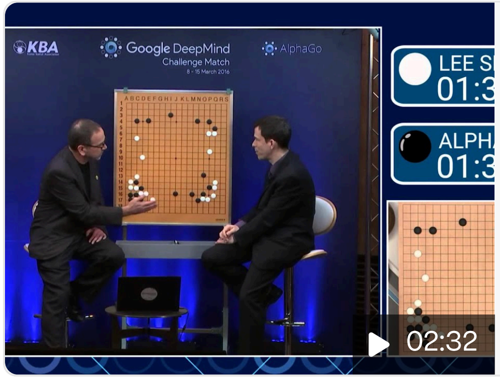
Google DeepMind's Deep Q-learning pl...

Google DeepMind created an artificial intelligence program using deep...

In **2013**, DeepMind achieved a landmark breakthrough by training an AI to play classic Atari 2600 games like Breakout and Space Invaders.

- The AI learned directly from **raw pixel data**, just like a human player would perceive the game.
- It achieved **superhuman performance** in many games, demonstrating the power of combining reinforcement learning with deep neural networks.

AlphaGo — "Move 37" (2016)



YouTube

Move 37!! Lee Sedol vs AlphaGo Match 2

Beautiful move 37 from AlphaGo vs Lee Sedol in Match 2. The move forces Sedol out of his chair,...

The Moment AI Showed Creativity

A move no human had ever played.
Professionals called it *beautiful*.

Search space: 10^{170} legal board positions — larger than atoms in the universe.

Generating Realistic Faces — The Uncanny Valley (~2018)

Using advanced generative adversarial networks (GANs), AI can create hyper-realistic human faces that do not belong to any real person. Websites like thispersondoesnotexist.com demonstrate this capability, producing endless unique portraits with astonishing detail.




This technological leap, largely driven by architectures like StyleGAN, highlights both the immense creative potential of AI and the growing challenges in distinguishing synthetic media from reality, impacting areas from digital art to combating misinformation.

Deepfakes — The Rise of Synthetic Media (2017-)

Building on the ability to generate hyper-realistic faces and voices, generative AI has advanced to create highly convincing fake videos and audio, commonly known as deepfakes. These synthetic media pieces can manipulate speech, expressions, and actions, making it appear as if someone said or did something they never did.

YouTube

How to spot a deepfake – BBC News

Deepfake videos are becoming ever more realistic. Can you spot one?

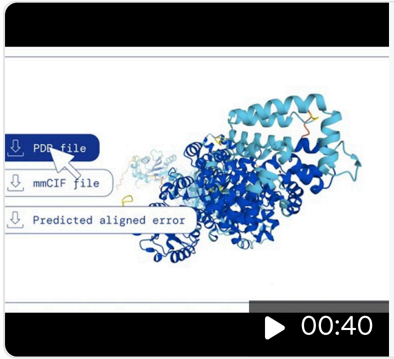
03:23

First appearing around **2017**, deepfake technology has rapidly evolved, posing significant challenges across various domains:

- **Misinformation:** Spreading false narratives and propaganda with fabricated evidence.
- **Reputation Damage:** Creating damaging content involving public figures or private citizens.
- **Security Risks:** Exploiting biometric authentication systems and enabling sophisticated scams.

The increasing realism of deepfakes necessitates enhanced media literacy and detection tools, as showcased in the video.

AlphaFold — Solving the Protein Folding Problem (2020)



YouTube

AlphaFold Protein Structure Datab...

With EMBL-EBI, we're proud to launch the AlphaFold Protein Structure Database,...

DeepMind's AlphaFold achieved a monumental breakthrough in **2020**, accurately predicting the 3D structure of proteins from their amino acid sequences.

> This was a 50-year-old "grand challenge" in biology, considered one of the most significant unsolved problems in understanding life itself.

By overcoming this hurdle, AlphaFold significantly accelerates drug discovery, vaccine development, and fundamental biological research, opening vast new avenues for medical science and biotechnology.

Embodied AI — Robots That Learn

The Paradigm Shift

1. **Old: Hard-coded Physics**
2. **New: Learned Visuomotor Policies**

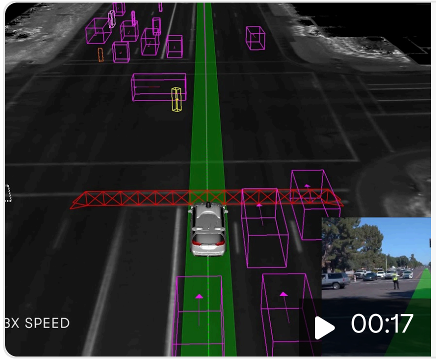
Robots that learn by **trying**, not by programming.

YouTube

RMA: Rapid Motor Adaptation for Legged ...

Project Page: https://ashish-kmr.github.io/rma-legged-robots/ Paper:...

Self-Driving Cars — Navigating the Real World



YouTube

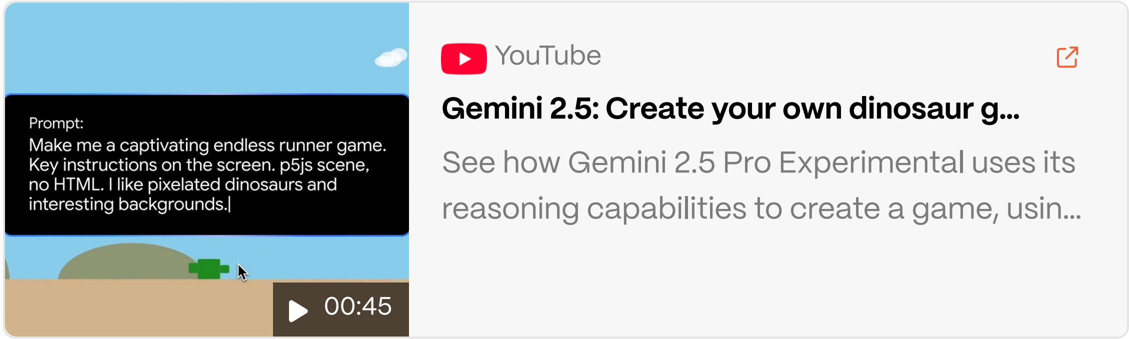
Waymo's autonomous driving technolo...

This video shows Waymo's autonomous driving technology approaching a traffic ligh...

Autonomous vehicles exemplify embodied AI, learning to perceive, predict, and plan in dynamic and unpredictable physical environments.

- They integrate advanced computer vision, sensor fusion (LiDAR, radar, cameras), and complex decision-making algorithms.
- Unlike controlled simulations or games, self-driving cars confront the full complexity of real-world physics, human behavior, and unforeseen obstacles.
- This field pushes AI to its limits, combining perception with real-time physical interaction and safety-critical decisions on the road.

Generating Apps & Websites — From Prompt to Product



Generative AI is rapidly evolving beyond creating images and text, now directly impacting software development.

- AI models can interpret natural language descriptions and translate them into functional code, entire websites, or even mobile applications.
- This capability drastically accelerates the prototyping and development process, enabling users to go from an idea to a working product in minutes or hours.
- It democratizes access to software creation, allowing individuals without extensive coding expertise to build digital tools and experiences.

Generating Art & Images — From Text to Pixels

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.

[Nano Banana Pro - Gemini AI image generator & photo editor]

Generating Video — The Next Frontier in GenAI

01:00

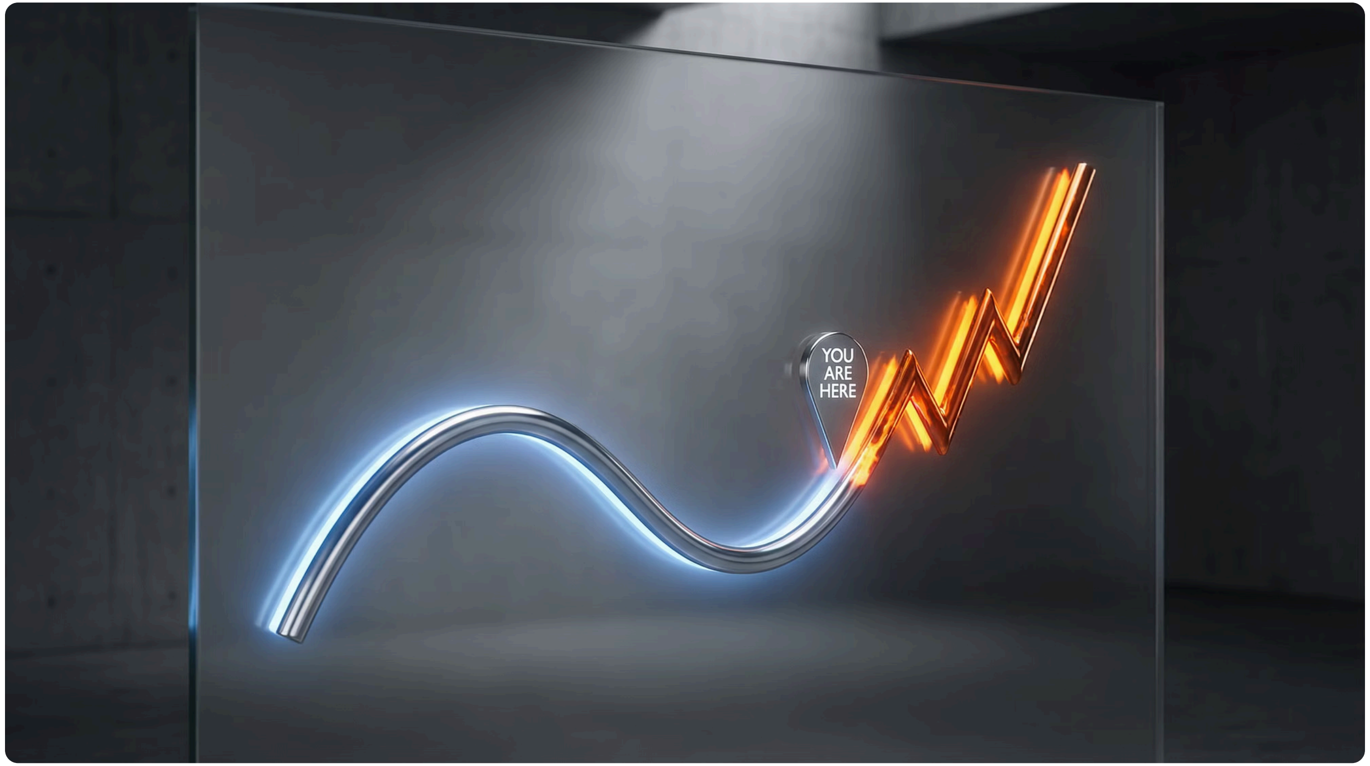
YouTube

Veo: Google DeepMind's new generative ...

Introducing Veo, our latest text-to-video generative AI model. Veo generates high-...

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"

The "Hockey Stick" Curve



50 Years of Slow Progress

10 Years of Vertical Ascent

You are entering the field HERE.

Builders vs. Users

Coding

Being automated by AI

AI Architecture

Designing loss functions, curating datasets, building systems

Our Courses

We teach you to build the engine — not call an API

MODULE 1

The Concrete — Building the Foundation

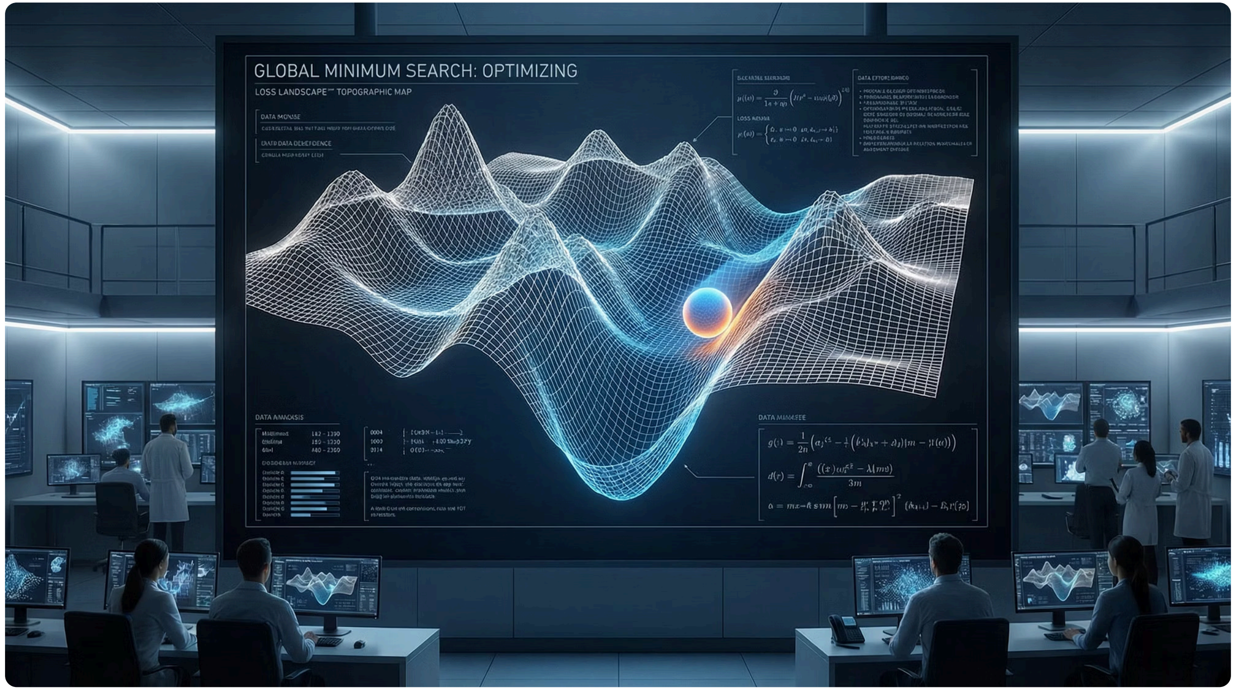
From biological inspiration to mathematical rigor. The essential building blocks of every modern AI system.

Deep Learning
Modulo Core Models and Methods

Docenti: Prof. Giovanni Maria Farinella & Prof. Francesco Ragusa

Syllabus: https://web.dmi.unict.it/corsi/lm-18/insegnamenti?seuid=414654A5-811B-4AAB-9031-836CC788F119

The Bricks — MLPs & Optimization



Core Concepts

01

Loss Landscapes

Navigating mountains of error

02

SGD

Stochastic Gradient Descent

03

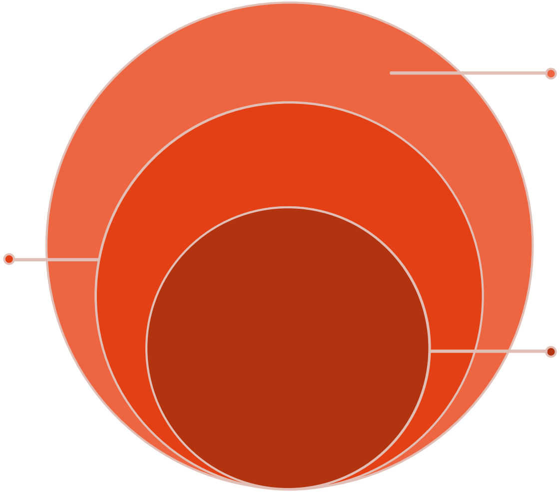
Differentiable Programming

Everything is a gradient

The Eyes — Convolutional Neural Networks

Convolutions & Inductive Bias

Teaching networks **where** to look and **how** to see.



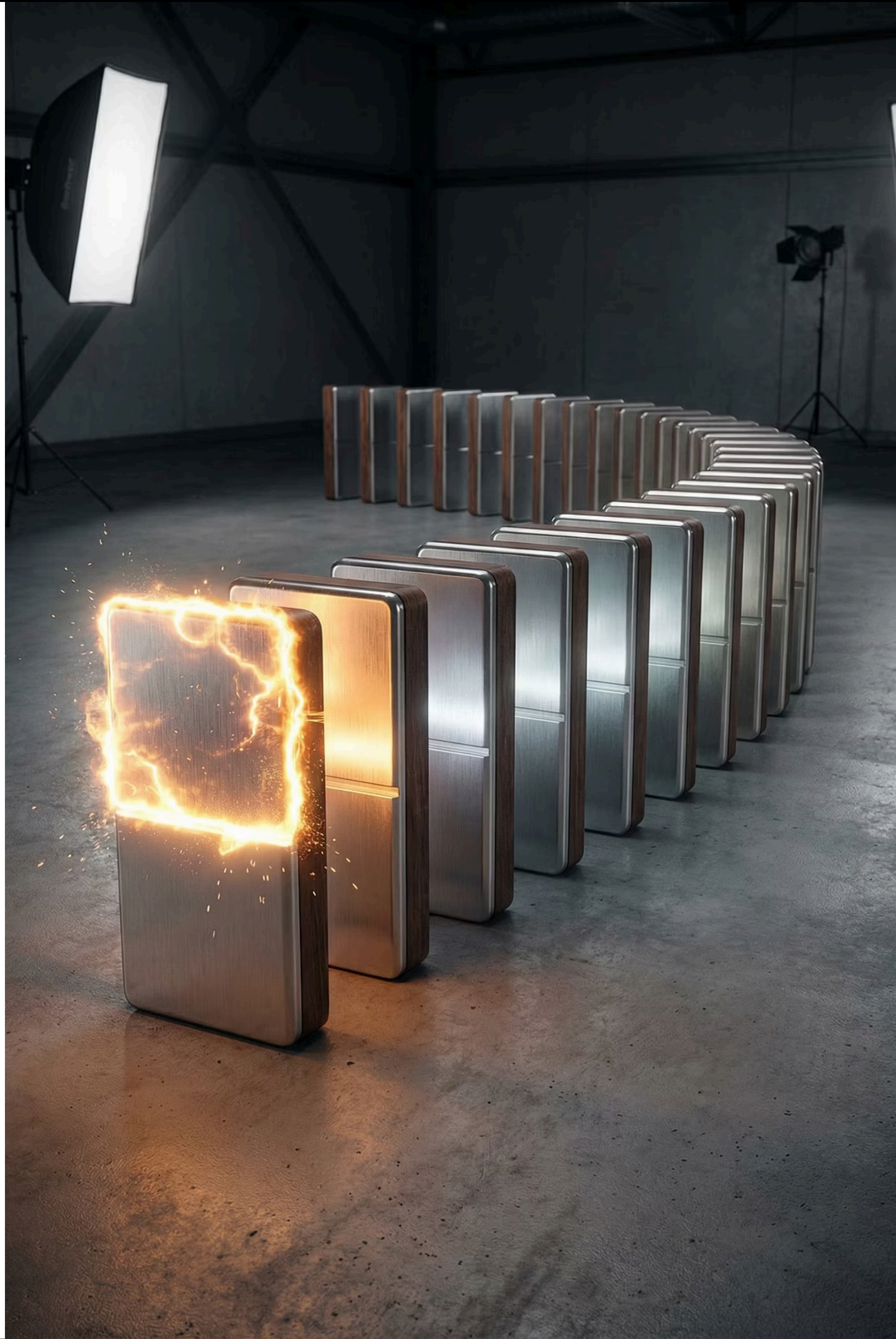
The Memory — Recurrent Neural Networks

The Logic

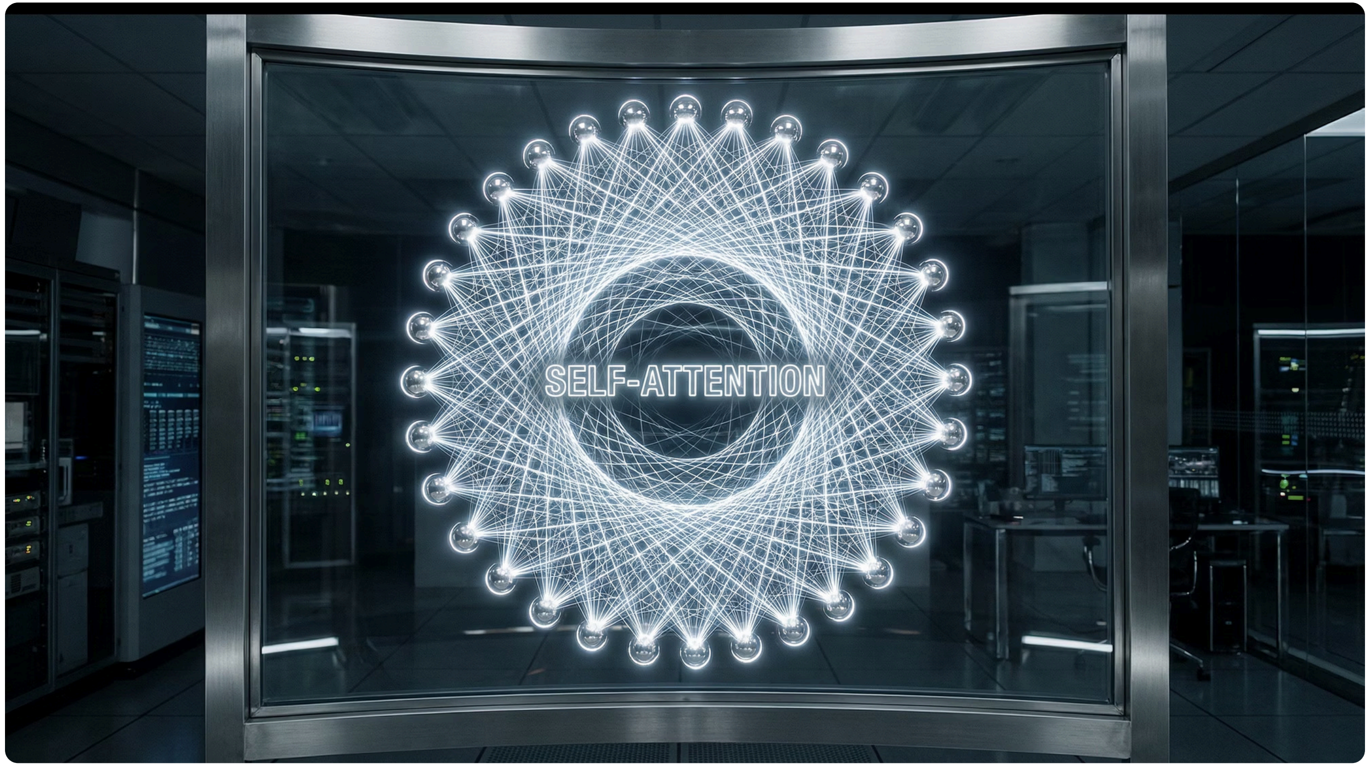
Handling sequences — understanding time

The Limitation

Vanishing Gradient — forgetting the past



The Revolution — Transformers



Parallel Processing + Attention

Every token sees every other token — simultaneously.

The engine behind ChatGPT, GPT-4, and the entire LLM revolution.

Unict DEEP LEARNING
ADVANCED MODELS AND METHODS

MODULE 2

Advanced Models — Real-World Messy Problems

Beyond the fundamentals. Where theory meets the chaos of reality.

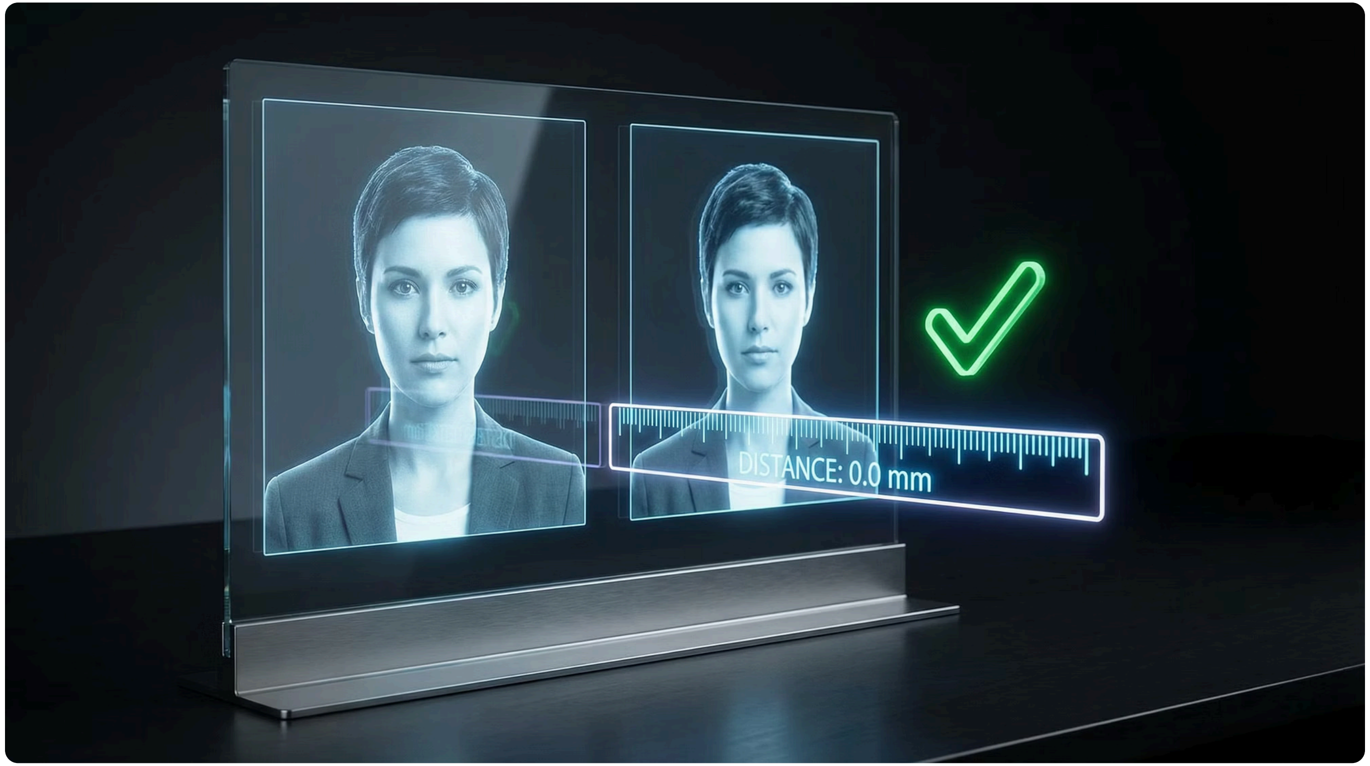
Deep Learning
Modulo Advanced Models and Methods

Docente: Prof. Antonino Furnari

Syllabus: https://web.dmi.unict.it/corsi/lm-18/insegnamenti?seuid=3540D939-DA16-4C1D-983C-E6B85C403F2F

Website: https://antoninofurnari.github.io/deeplearning/

Metric Learning — FaceID



Learning Distances, Not Classes

Don't ask "Who is this?"

Ask "**How similar are these two?**"

> Same person → distance ≈ 0 ✅
> Different person → distance >> 0 ❌

Domain Adaptation — Sim-to-Real

Learn in GTA V. Drive in reality.

Transfer knowledge from simulated environments to the real world — bridging the domain gap.

Knowledge Distillation — Mobile AI

Teacher → Student Compression

1

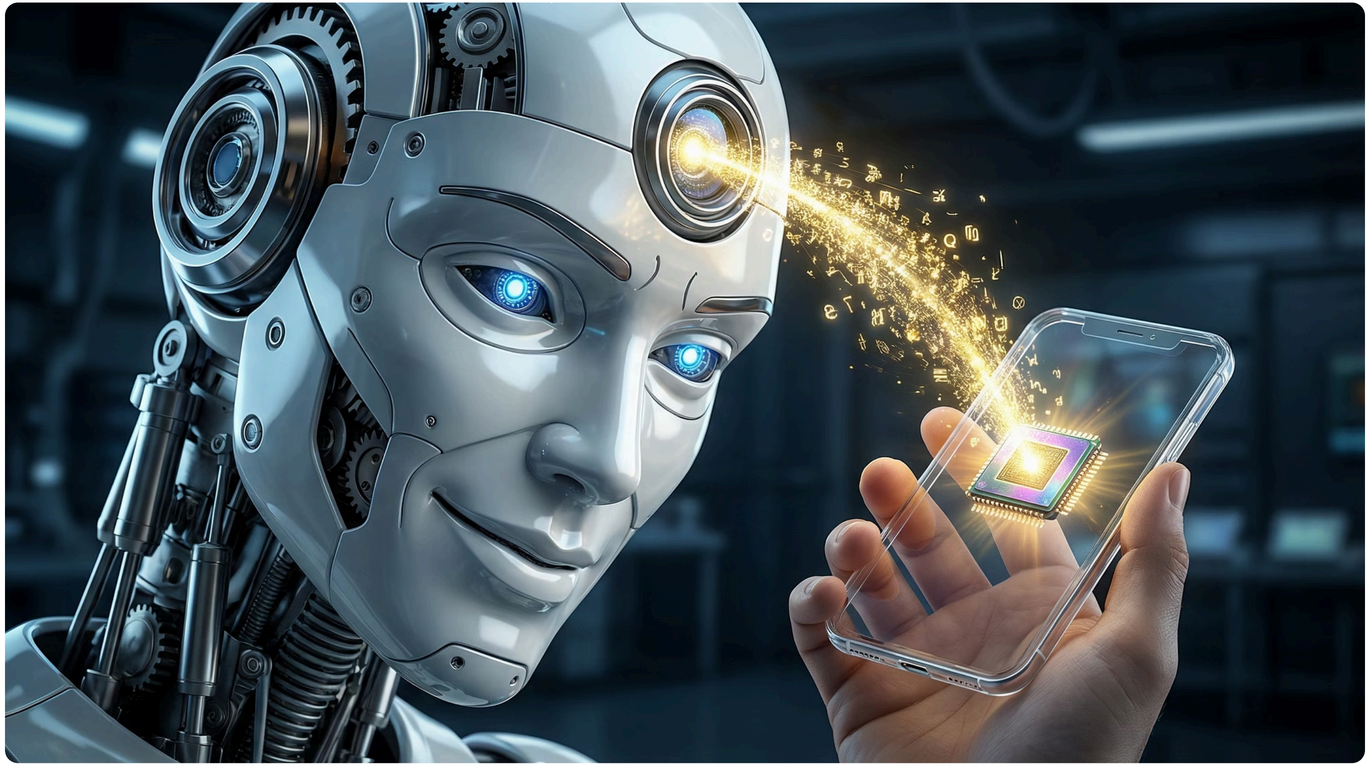
Massive Teacher Model

Billions of parameters

2

Tiny Student Model

Runs on your phone

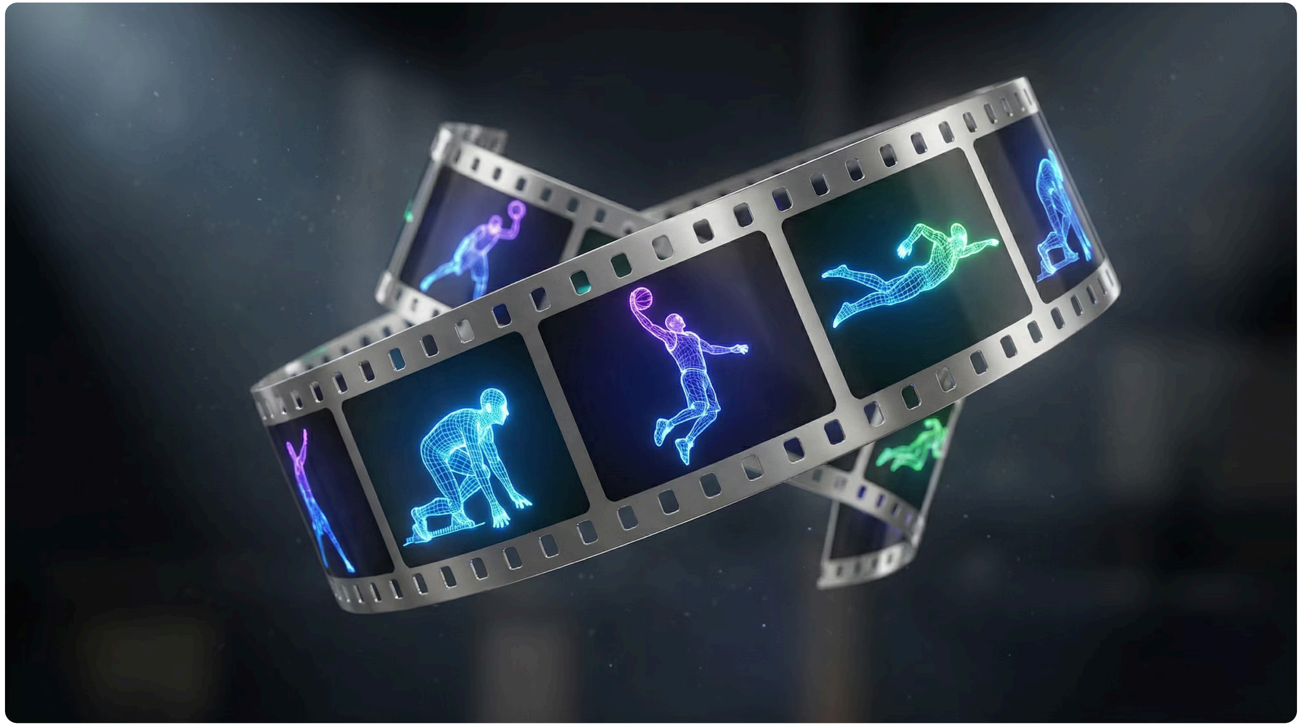


Self-Supervised Learning

No labels. No humans. Just **puzzles and invariance**.

The network learns structure from unlabeled data — the most abundant resource on Earth.

Video Understanding — Spatiotemporal Networks



Understanding Time

Images capture *what*.

Video captures **what happens next**.

Spatiotemporal networks fuse spatial features with temporal dynamics.

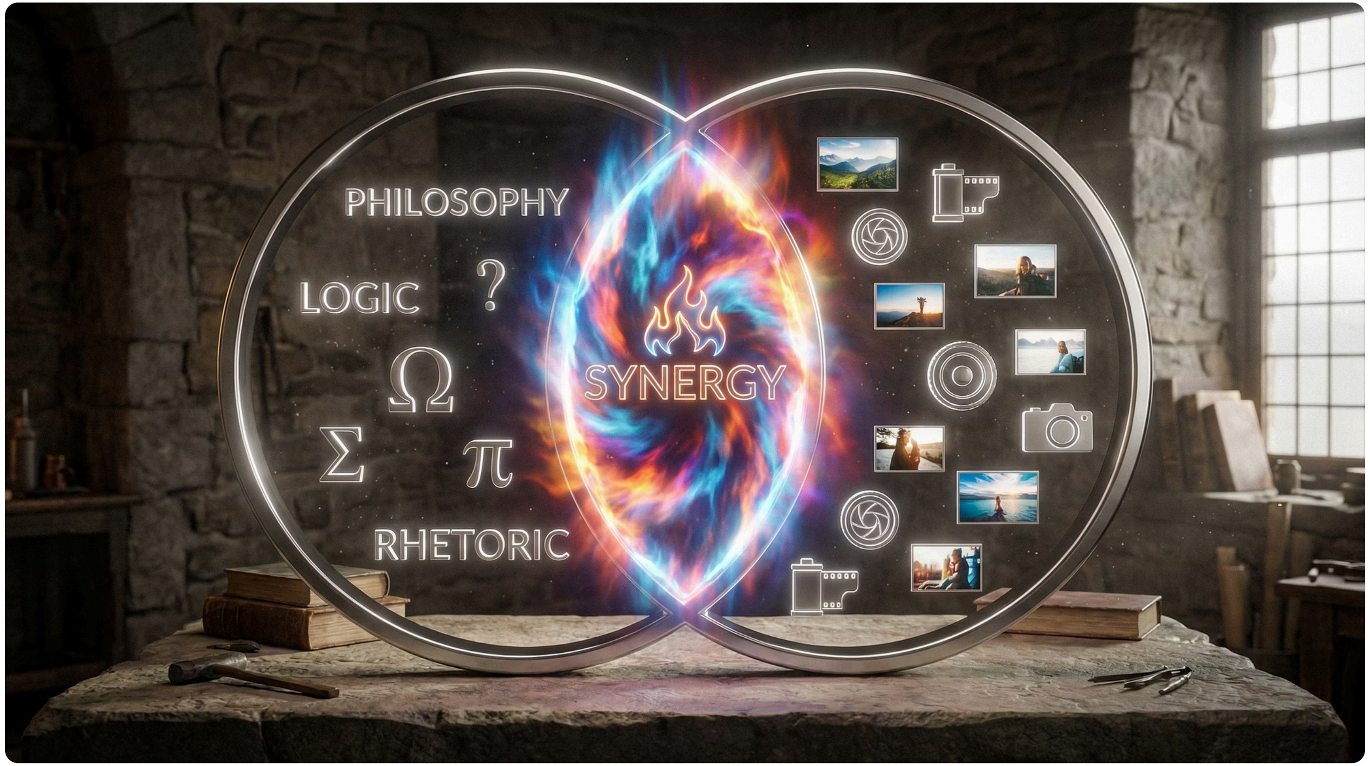
Multimodal Learning — Vision + Language

CLIP — Bridging Two Worlds

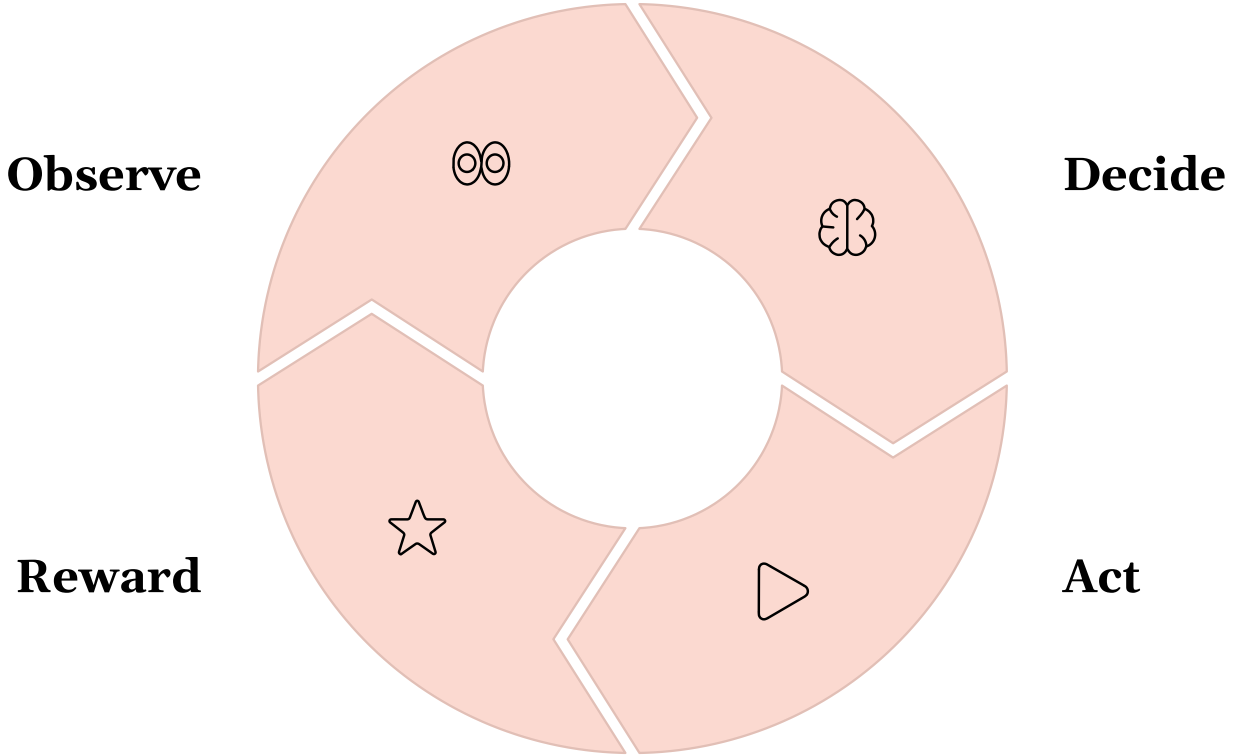
Map images and text into a **shared space**.

A photo of a dog and the phrase "a photo of a dog" land at the same point.

> This is the foundation of text-to-image generation and visual search.



Reinforcement Learning — Agents & Rewards



The Frontier — Beyond Transformers

Mamba & Large Multimodal Models

State Space Models — linear scaling replaces quadratic attention.

LMMs combine vision, language, audio, and action in a single model.

The next chapter is being written *right now.*

You may also like ...

Other LM-18 courses related to the Deep Learning journey:

- **Computer Vision**. Prof. Sebastiano Battiato, Prof. Francesco Guarnera
- **Multimedia**. Prof. Dario Allegra
- **Artificial Intelligence for Language Processing**
- **Generative Artificial Intelligence**

See all the details here: https://web.dmi.unict.it/it/corsi/lm-18/artificial-intelligence-machine-learning

Your Roadmap Starts Now

Vision & Sequence

CNNs, RNNs, and Transformers for perception.

Advanced Topics

Metric learning, self-supervised and multimodal methods.

Foundations

MLPs and optimization basics for model building.

The Frontier

Mamba, LMMs, and directions for your research.

Join LM-18 - Artificial Intelligence and Machine Learning. Build the future.

Thank You

Scan the QR code to access the full presentation slides and resources.

https://antoninofurnari.github.io/deeplearning/slides/02-03-26-intro-slides-prospettive-informatiche.pdf