Master AI per il Digital Marketing 2023

AI ESSENTIALS

Diego Vendramini

STUDIOSAMO®

Chi sono

Direttore Operativo Master Executive
MIB Trieste School of Management

Valutatore di processi formativi, programmi e competenze professionali
APAFORM - ASFOR

Componente del Comitato Scientifico

Generativa
l'AI
umani
Le
AI
Rete
Bias
Gemini
LLM
Bert
Agente
neurale
Gpt
Token
Machine
Learning
apprendono
devono
esseri
Tutte
pensano
cervelli
più
degli
è
aziende
diventando
intelligenti
per
avere
rimanere
stanno
macchine
come
competitive
creativa

Cosa fa male all’ AI

CORRIERE DELLA SERA
CORRIERE DEL MEZZOGIORNO
ABBONATI A 6€/ANNO
Accedi

IN EVIDENZA

5 Marzo 2025

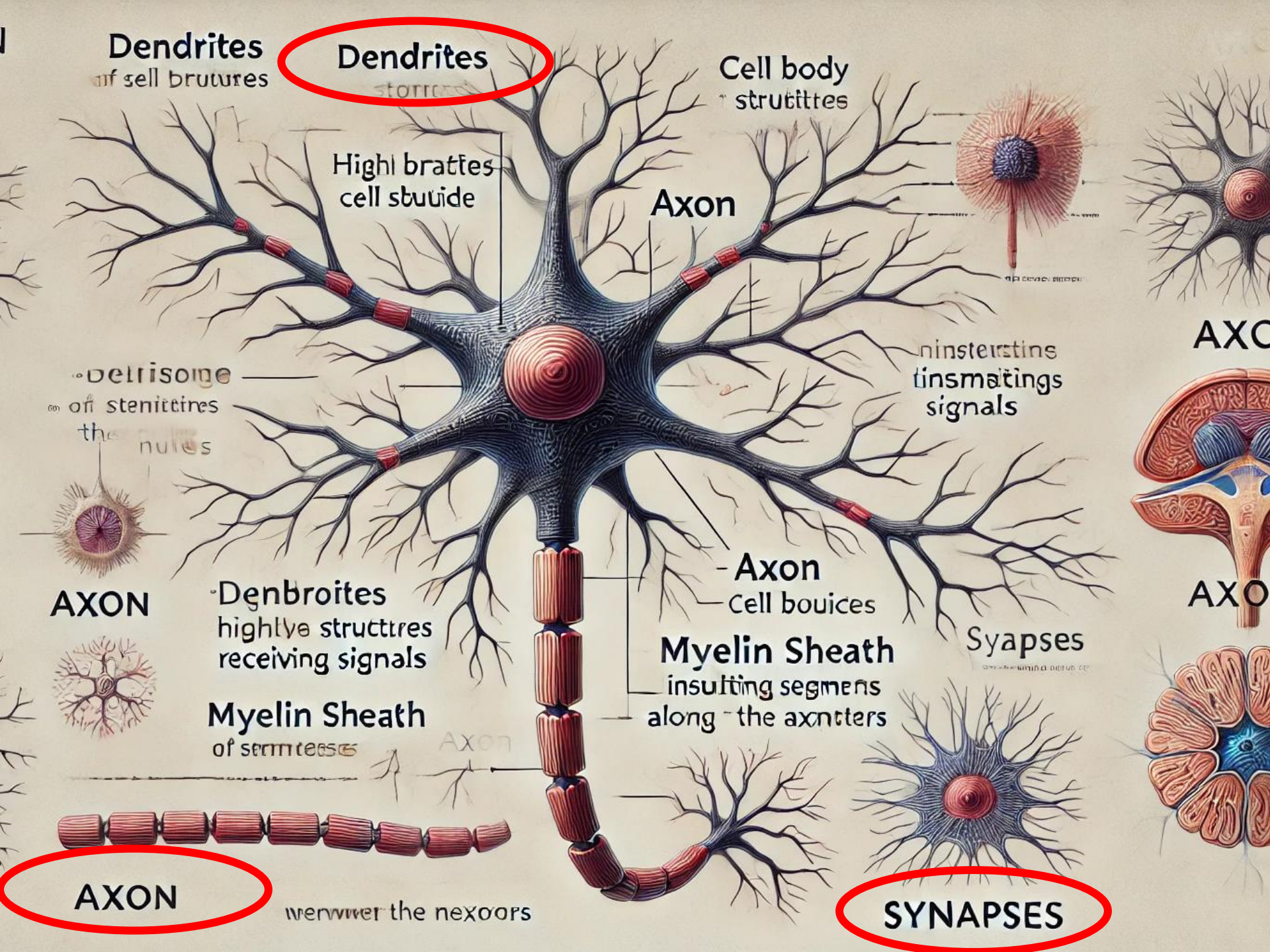
Lecce, i numeri del Lotto con l'Intelligenza artificiale: tre studenti universitari ci provano e vincono fino a 43mila euro

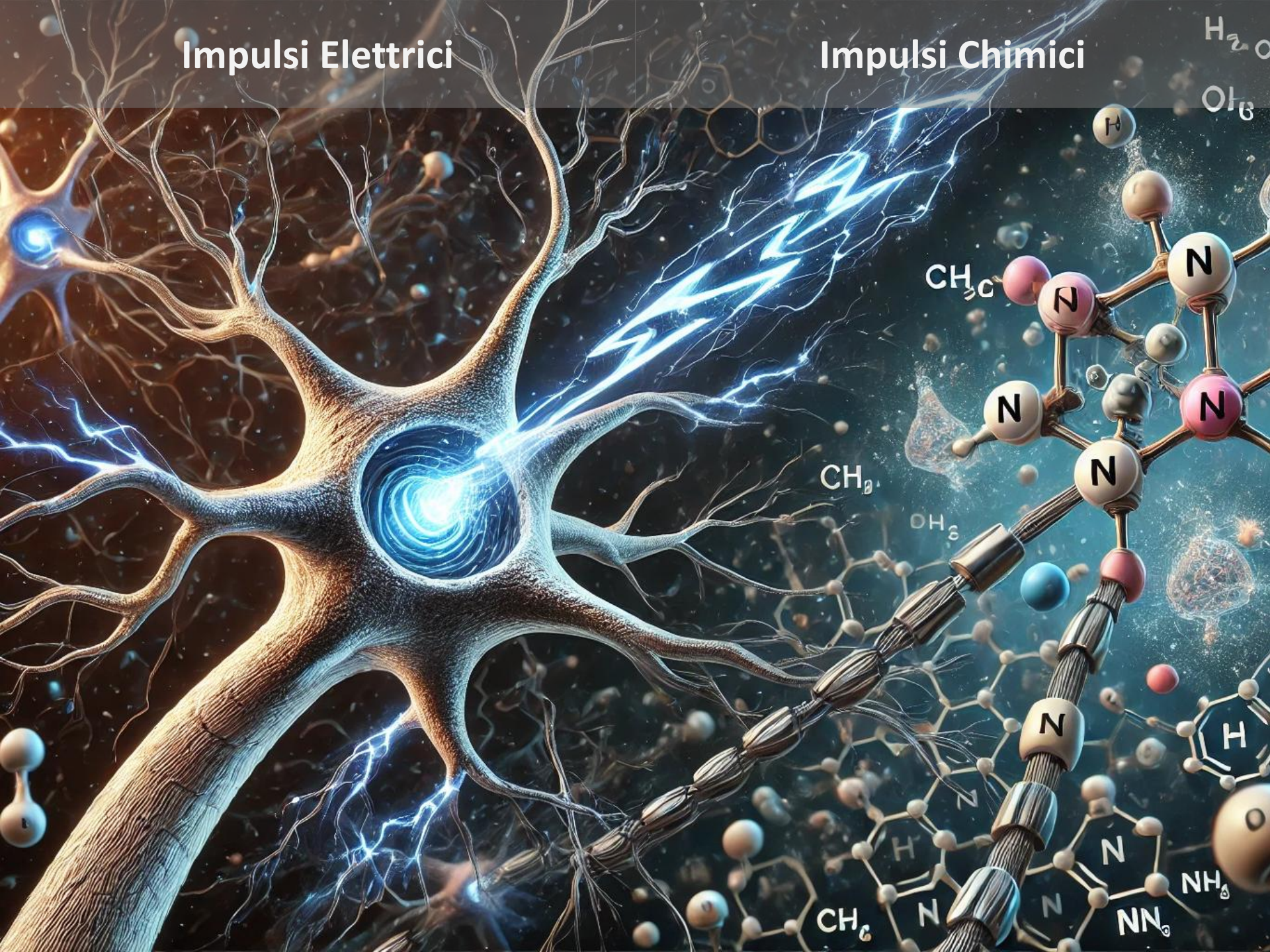
di Claudio Tadicini

Gli universitari, tutti provenienti dal Dipartimento di Matematica dell'ateneo leccese, sono infatti riusciti a sviluppare un algoritmo di machine learning, che ha esaminato i dati delle estrazioni del Lotto degli ultimi due anni

La ricevitoria dove è stata messa a segno la vincita

Ridurre e Minimizzare


Dendrites
Dendrites
Cell body
Axon
Axon
Cell bouices
Myelin Sheath
insulting segmens
Syapses
AXON
AXON
SYNAPSES


Impulsi Elettrici
Impulsi Chimici

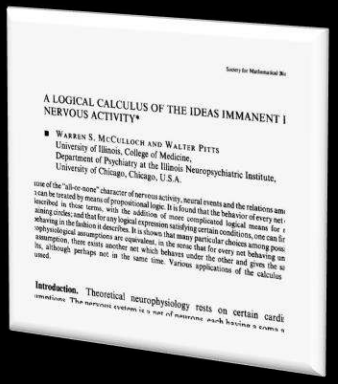
STORIA DELL' AI

Una storia di uomini, tecnologie, inverni e paper

1943

Modello di Neurone Artificiale di McCulloch e Pitts

“A Logical Calculus of the Ideas Immanent in Nervous Activity”, primo modello matematico di un neurone artificiale.


A LOGICAL CALCULUS OF THE IDEAS IMMANENT I
NERVOUS ACTIVITY*

WARREN S. MCCULLOCH AND WALTER PITTS
University of Illinois, College of Medicine,
Department of Psychiatry at the Illinois Neuropsychiatric Institute,
University of Chicago, Chicago, U.S.A.

Introduction. Theoretical neurophysiology rests on certain cardi

Image souce: www.semanticscholar.org
www. archive.org

1956

Conferenza di Dartmouth

John McCarthy, M.Minsky, N.Rochester , C. Shannon

nasce ufficialmente il termine Artificial Intelligence

Image Souce: www.ihal.it
www.jmc.stanford.edu

1957

Perceptron di Rosenblatt

Frank Rosenblatt sviluppa il Perceptron, il primo algoritmo di apprendimento supervisionato ispirato ai neuroni biologici, che costituisce la base per lo sviluppo delle reti neurali.

Image Souce: www. news.cornell.edu

1966

Eliza

Joseph Weizenbaum crea ELIZA, primo chatbot per simulare una conversazione in linguaggio naturale con tecniche di pattern matching.

Image Souce: www. Hdblog.it

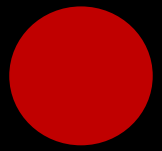
1969 - 1974

1° INVERNO DELL' AI

1966: la traduzione automatica russo-> inglese fallisce

1969: Marvin Minsky vs. Perceptron

1974: DARPA taglia i fondi per lo Speech Understanding Research

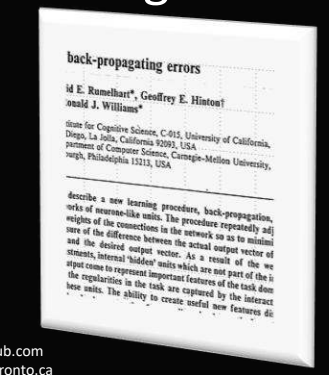


1986

Backpropagation

“Learning representations by back-propagating errors” (Rumelhart, Hinton e Williams)

Riscoperta e applicazione efficace dell’algoritmo di backpropagation per l’addestramento delle reti neurali multilivello: basi per il deep learning


back-propagating errors

id E. Rumelhart*, Geoffrey E. Hinton†
onald J. Williams*

itute for Cognitive Science, C-015, University of California,
Diego, La Jolla, California 92093, USA
partment of Computer Science, Carnegie-Mellon University,
burgh, Philadelphia 15213, USA

describe a new learning procedure, back-propagation,
orks of neurone-like units. The procedure repeatedly adj
eights of the connections in the network so as to minimi
sure of the difference between the actual output vector of
and the desired output vector. As a result of the we
ntments, internal 'hidden' units which are not part of the i
atput come to represent important features of the task dom
the regularities in the task are captured by the interact
hese units. The ability to create useful new features di

Image Souce: : www.github.com
www.cs.utoronto.ca

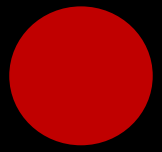
1989 - 1996

2° INVERNO DELL' AI

Falliscono i sistemi Esperti di Xcon

Pochi dati disponibili - Promesse disattese

Nuovi attori: Database relazionali, Sistemi ERP, World Wide Web



1997

Deep Blue

Deep Blue (e Feng-hsiung Hsu) battono il campione di scacchi Garry Kasparov

NYT 1989: «Non riesco a immaginarmi cosa vorrebbe dire vivere sapendo che un computer è più forte della mente umana, dovevo sfidare Deep Thought per proteggere la razza umana»

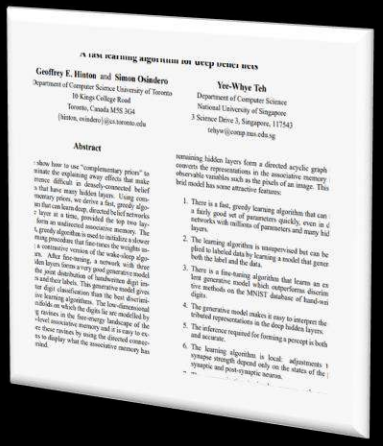
Image Souce: :www.il post.it

2006

Deep Learning

Geoffrey Hinton introduce il concetto di deep learning.

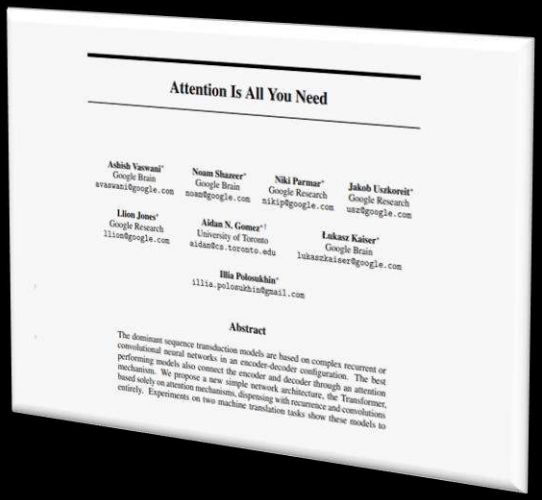
L’ AI cambia con grandi dataset e hardware più potente.



2017

Attention is all you need

Nascita del Transformer



Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com

Noam Shazeer* Google Brain noam@google.com

Niki Parmar* Google Research nikip@google.com

Jakob Uszkoreit* Google Research usz@google.com

Llion Jones* Google Research llion@google.com

Aidan N. Gomez*† University of Toronto aidan@cs.toronto.edu

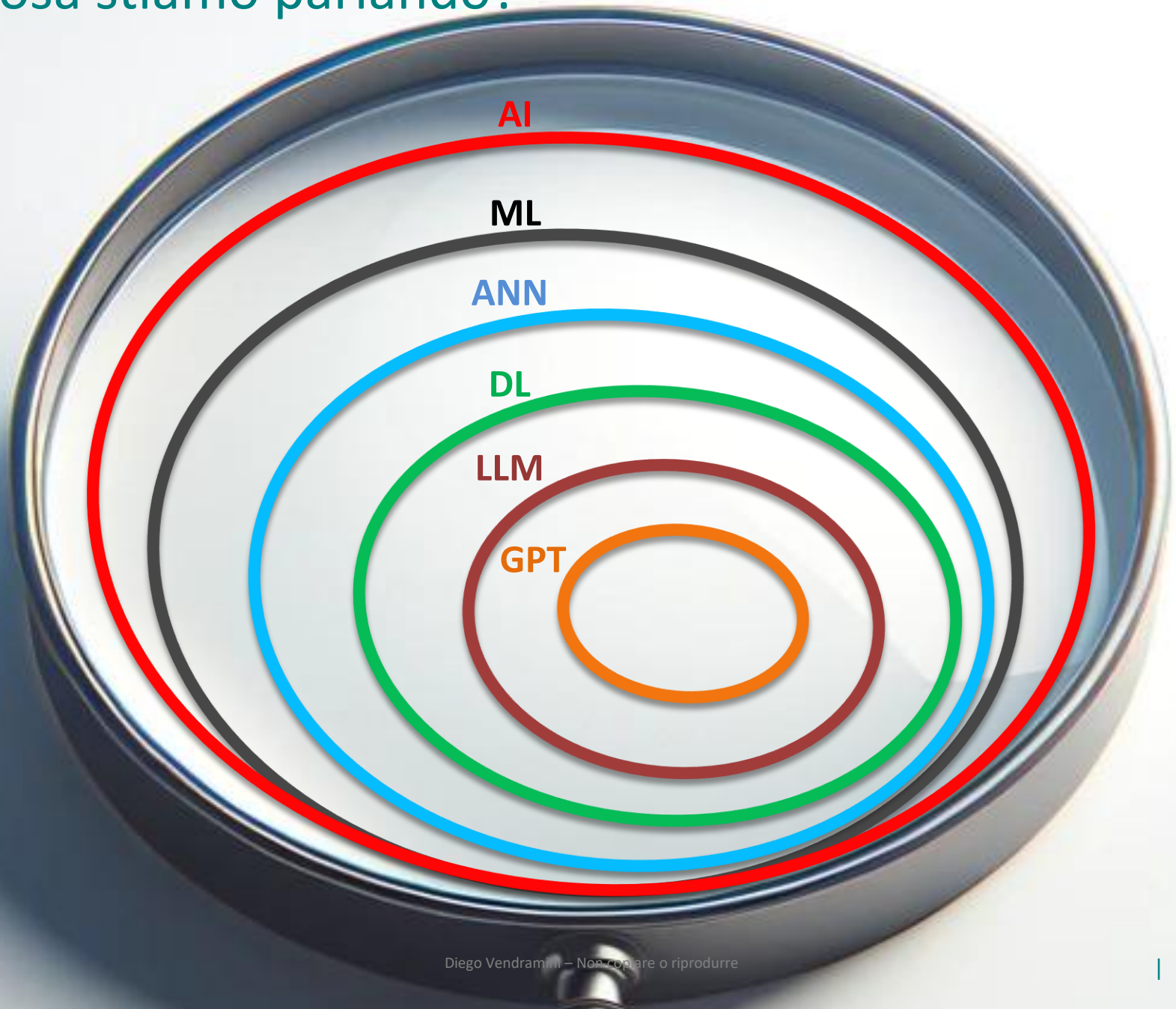
Łukasz Kaiser* Google Brain lukaszkaiser@google.com

Illia Polosukhin* illia.polosukhin@gmail.com

Abstract

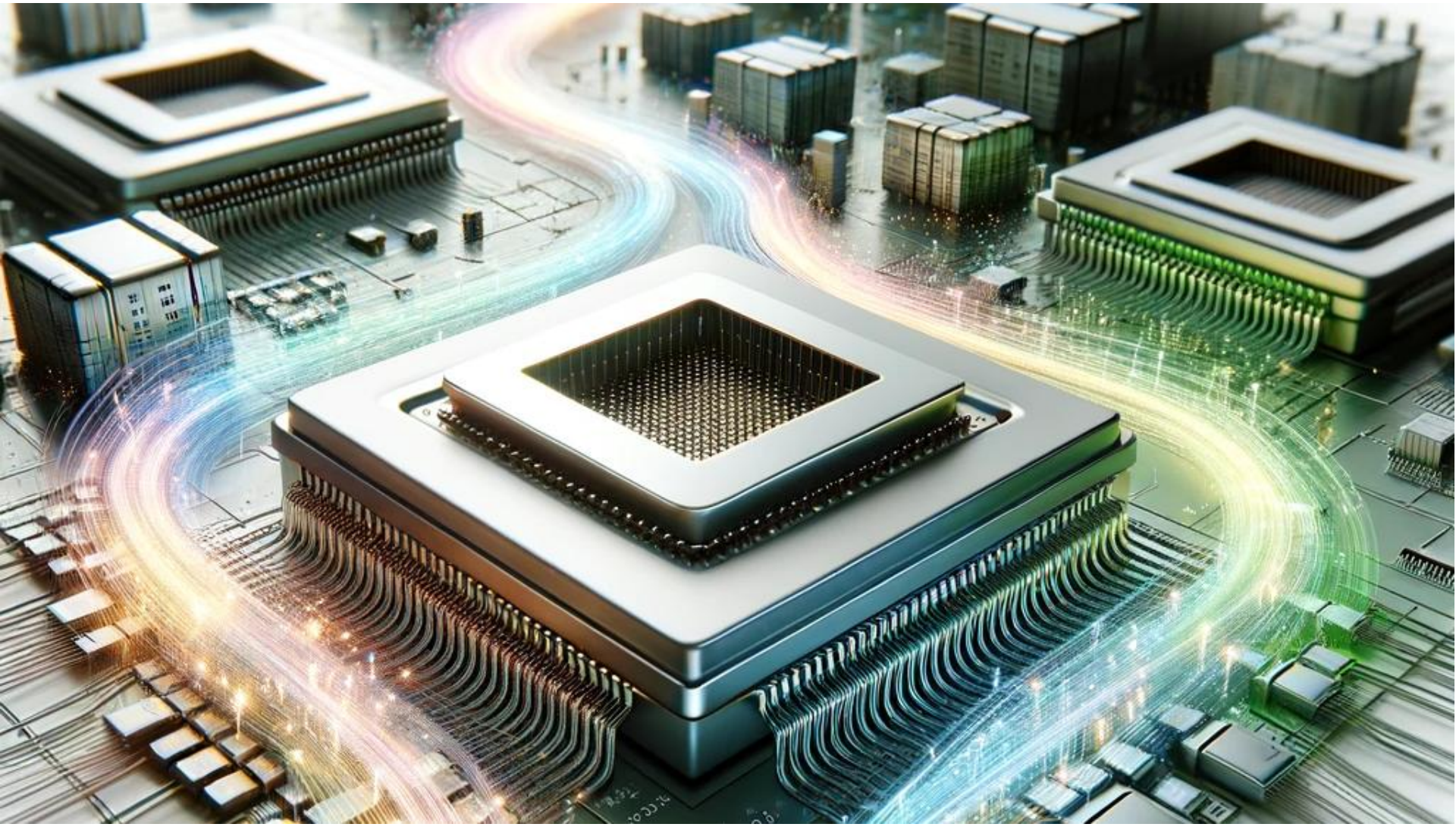
The dominant sequence transduction models are based on complex recurrent or convolutional neural networks in an encoder-decoder configuration. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to

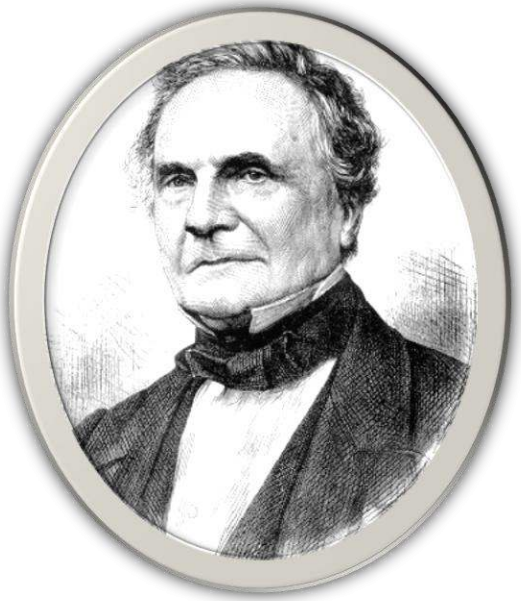
Di cosa stiamo parlando?



Diego Vendramini – Non copiare o riprodurre

Perché ne parliamo oggi?



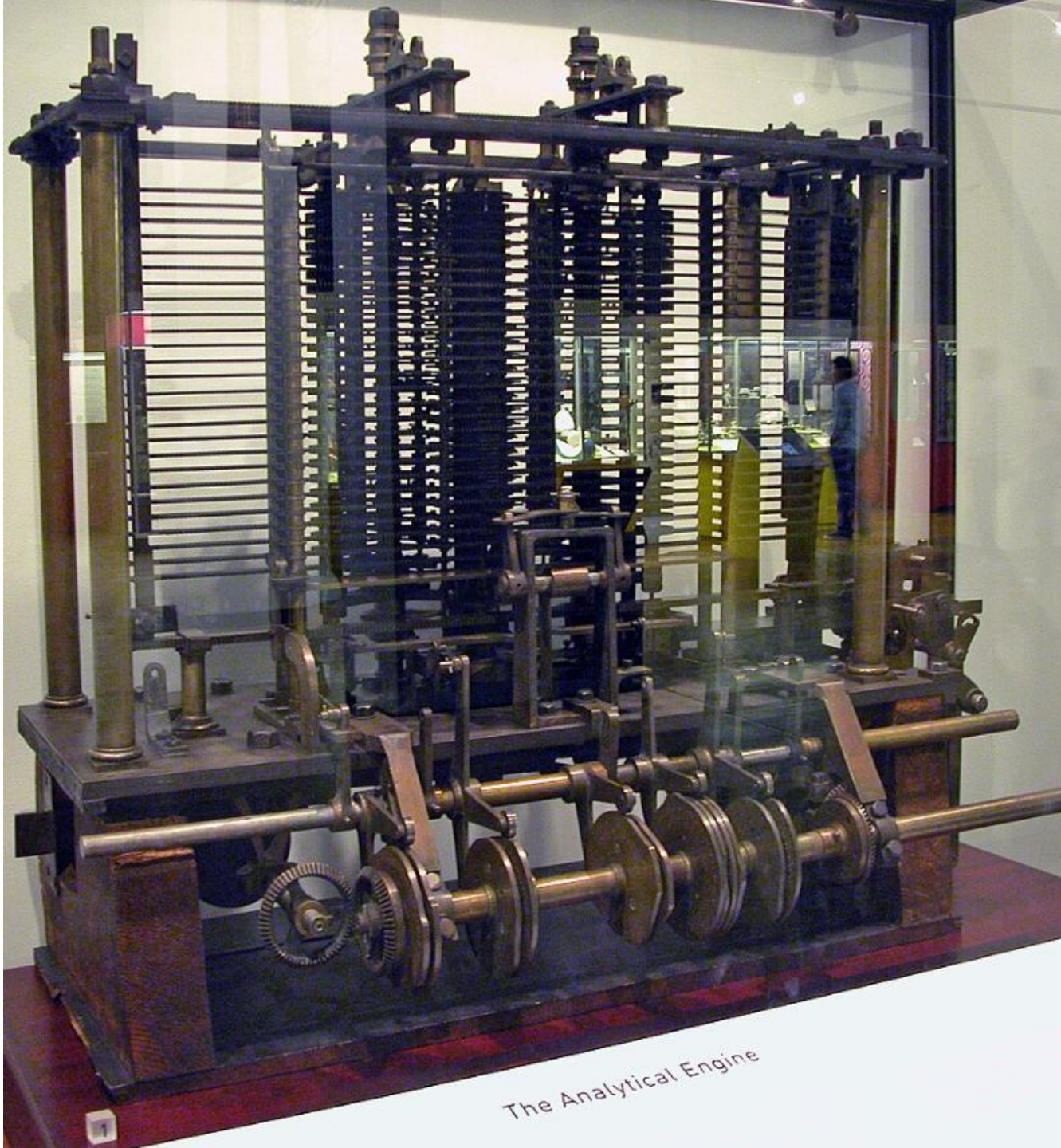


Charles Babbage

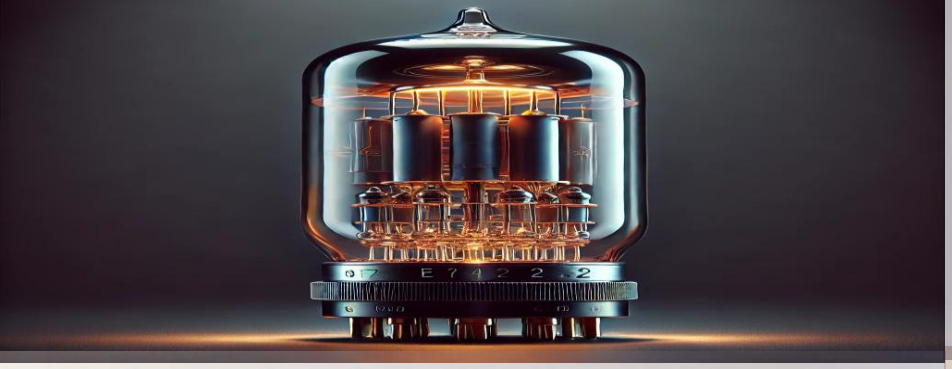
-

The Analytical Engine

1837



Diego Ventrammi – Non copiare o riprodurre

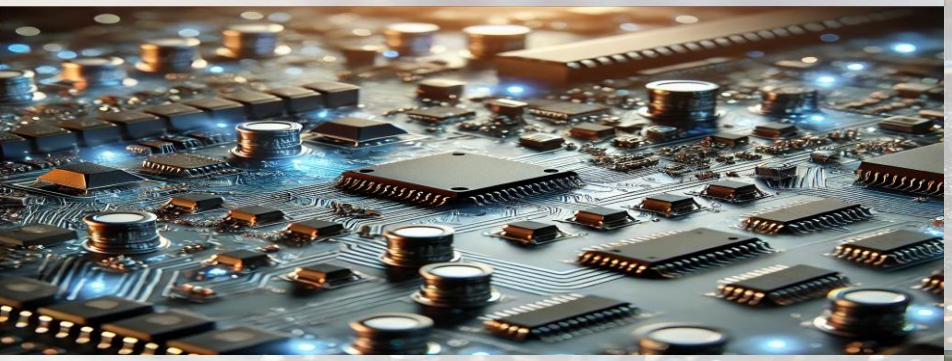


Prima Generazione (1940-1956)

Valvole termoioniche (tubi a vuoto).
Grandi dimensioni, elevato consumo energetico e necessità di raffreddamento.

Seconda Generazione (1956-1963)

Transistor.
Riduzione delle dimensioni, maggiore affidabilità e minore consumo energetico.

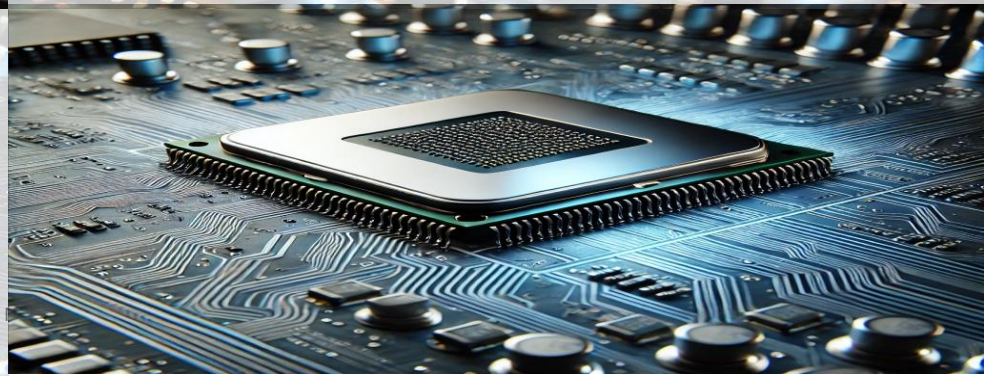


Terza Generazione (1964-1971)

Circuiti integrati (IC)
Miniaturizzaizone, aumento della potenza di calcolo e riduzione dei costi.

Quarta Generazione (1971-2010)

Microprocessori: migliaia di transistor su un singolo chip. PC diventano accessibili al grande pubblico.



CPU

GPU

NPU

TPU

Quanti calcoli fa un GPU?

Diego Vendramini – Non copiare o riprodurre

Quanti calcoli fa un GPU?

100 Milioni di calcoli / sec ?

Mario 64 ®
1996

100 Miliardi di calcoli / sec ?

Minecraft ®
2011

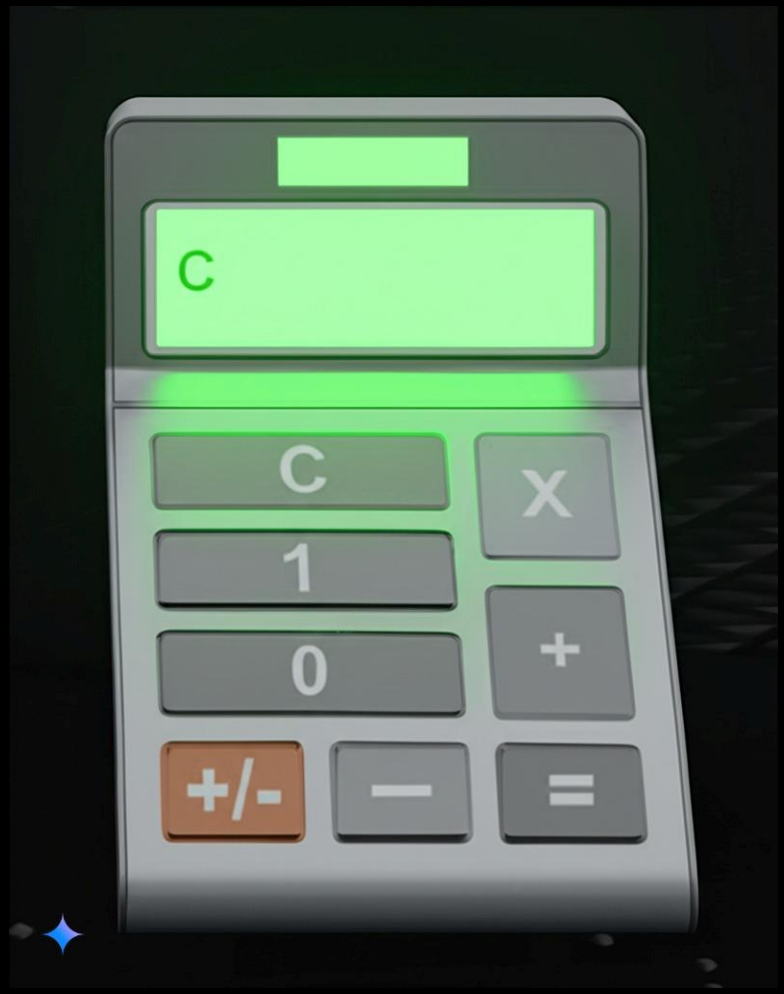
Quanti calcoli fa un GPU?

Cyberpunk 2077®

36 trillioni di calcoli al secondo

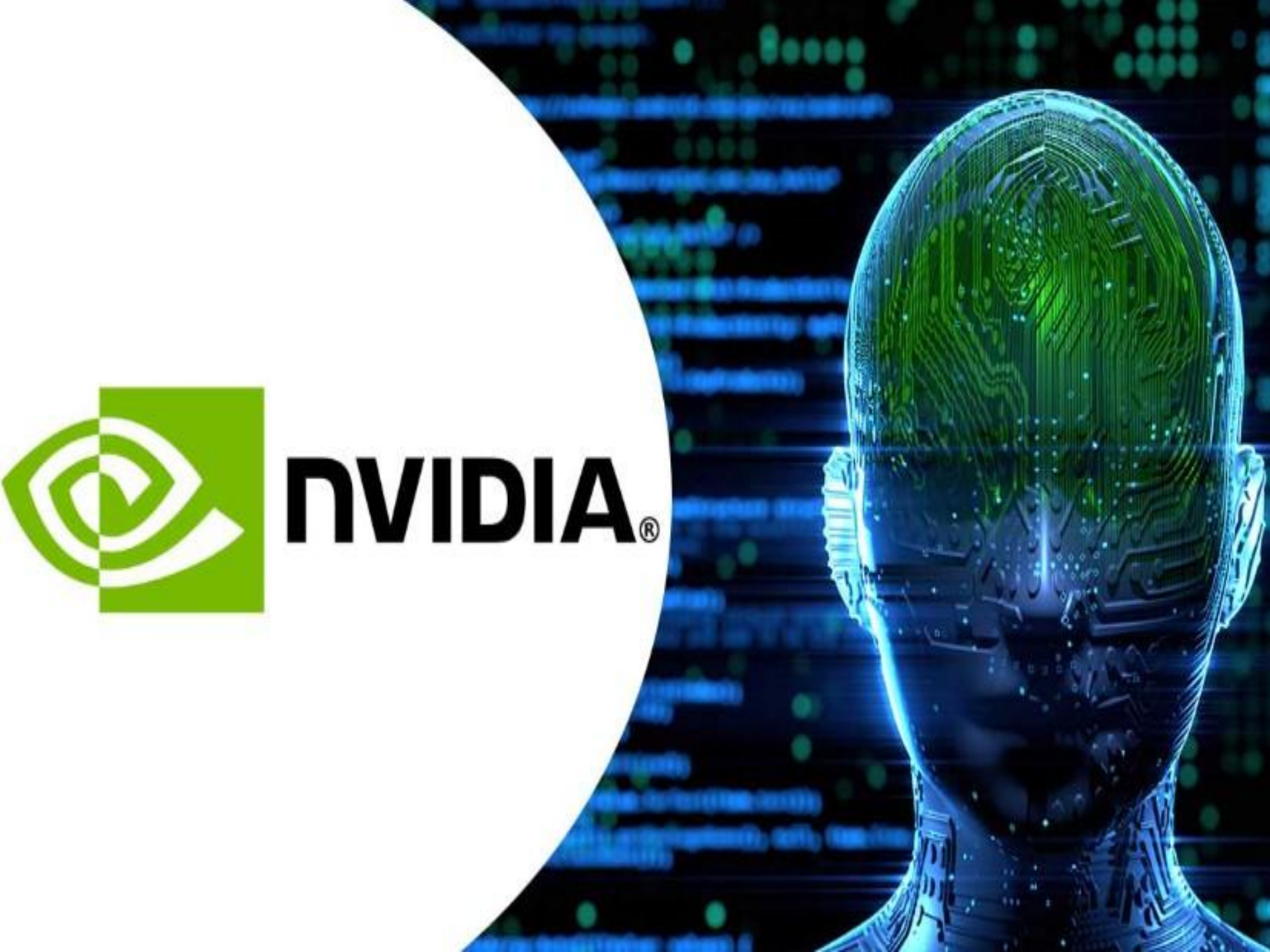
Diego Vendramini – Non copiare o riprodurre

Quanti calcoli fa un GPU?



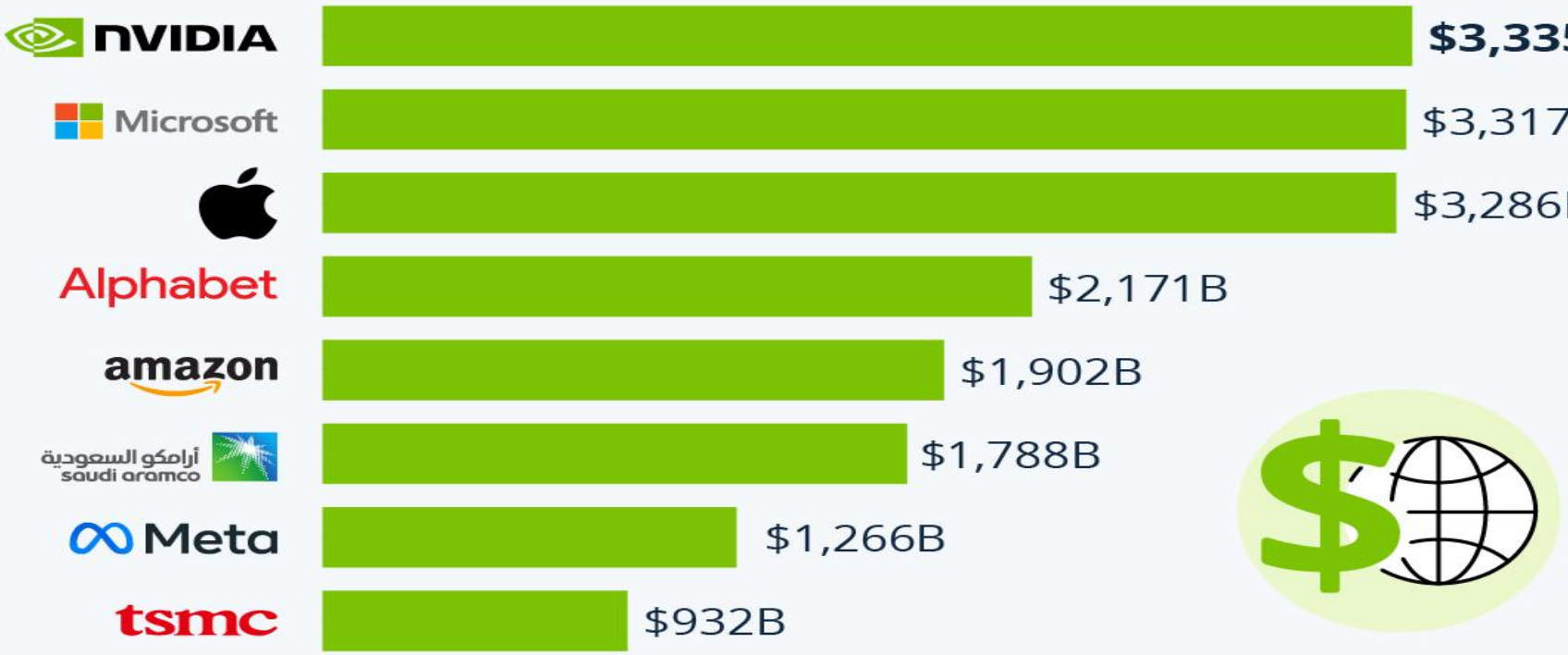
Tensor cores

Diego Vendramini – Non copiare o riprodurre


NVIDIA®

AI Boom: Nvidia Becomes Most Valuable Company in the World

Market capitalization of the highest-valued public companies in the world*



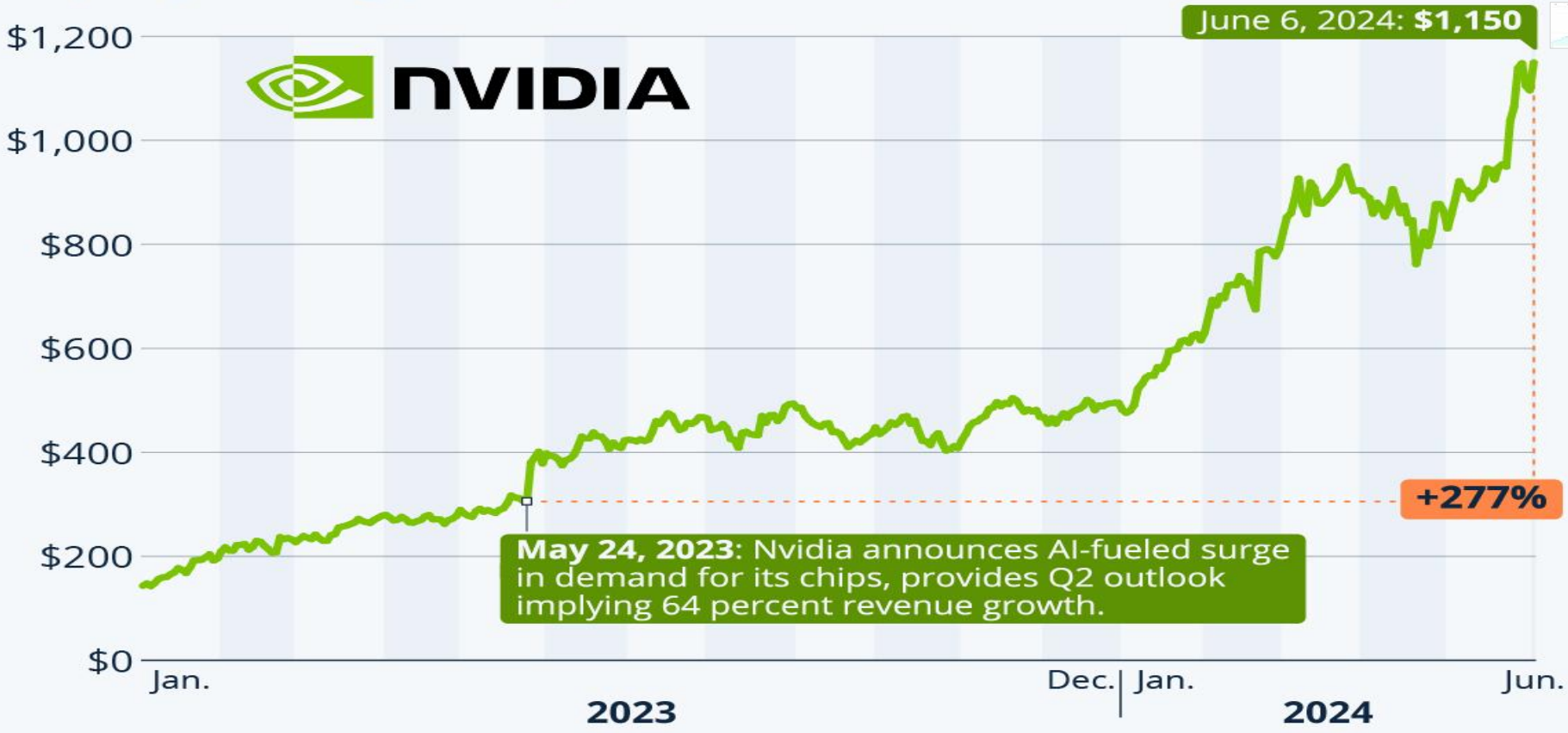
* As of market close on June 19, 2024

Source: Companiesmarketcap.com

statista

Nvidia's AI-Fueled Stock Market Rally

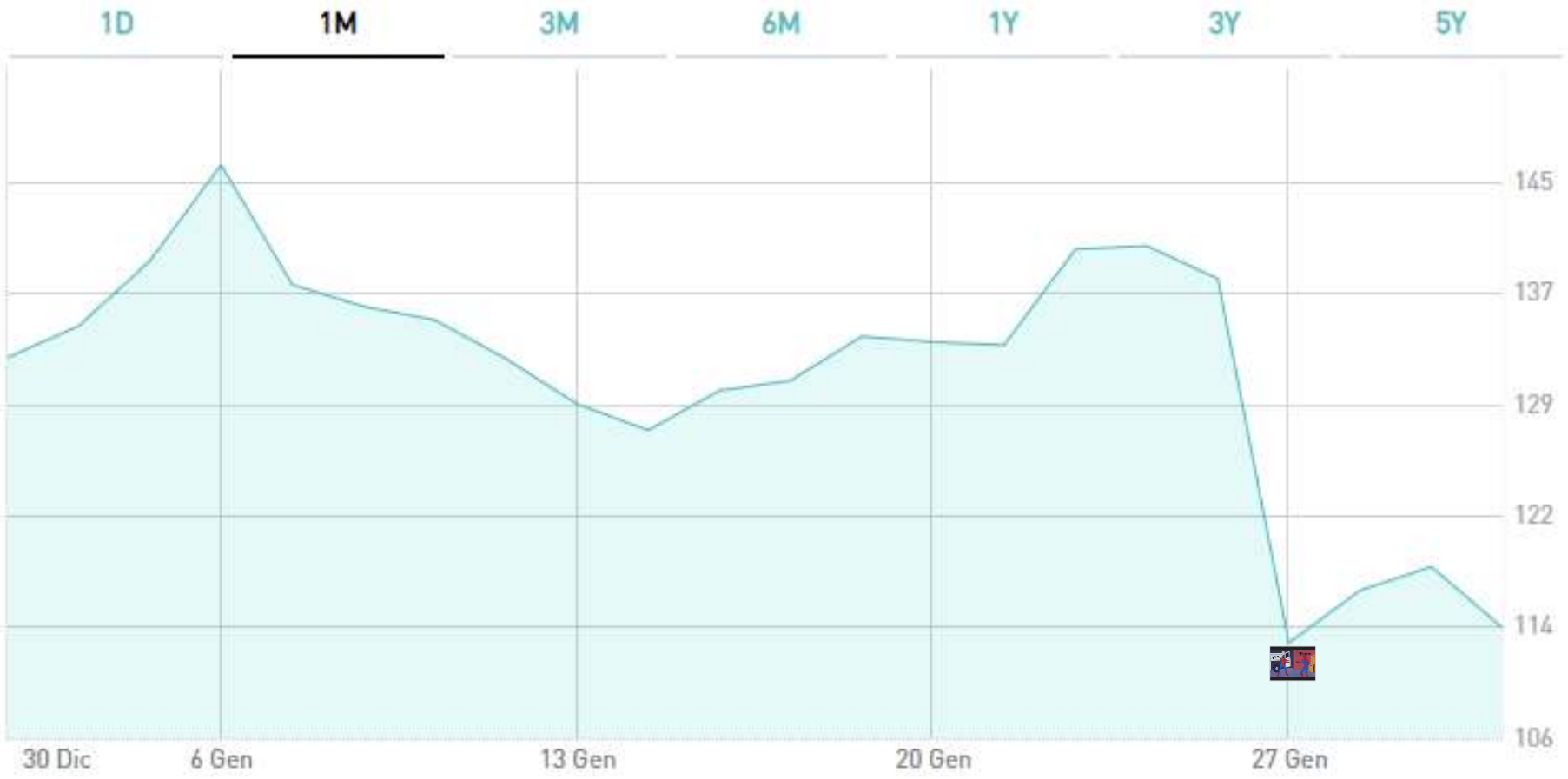
Nvidia's stock price since January 2, 2023
(daily closing prices)



Source: Yahoo Finance

Diego Vendramini – Non copiare o riprodurre

3M
6M
1Y
3Y
5Y
14
11
86,
57,
28,
0,0
n '23
Gen '24
Gen '25


1D
1M
3M
6M
1Y
3Y
5Y
145
137
129
122
114
106
30 Dic
6 Gen
13 Gen
20 Gen
27 Gen

NYPD
deepseek

LEONARDO
EuroHPC
CINECA
Funded by
the European Union
LEONARDO

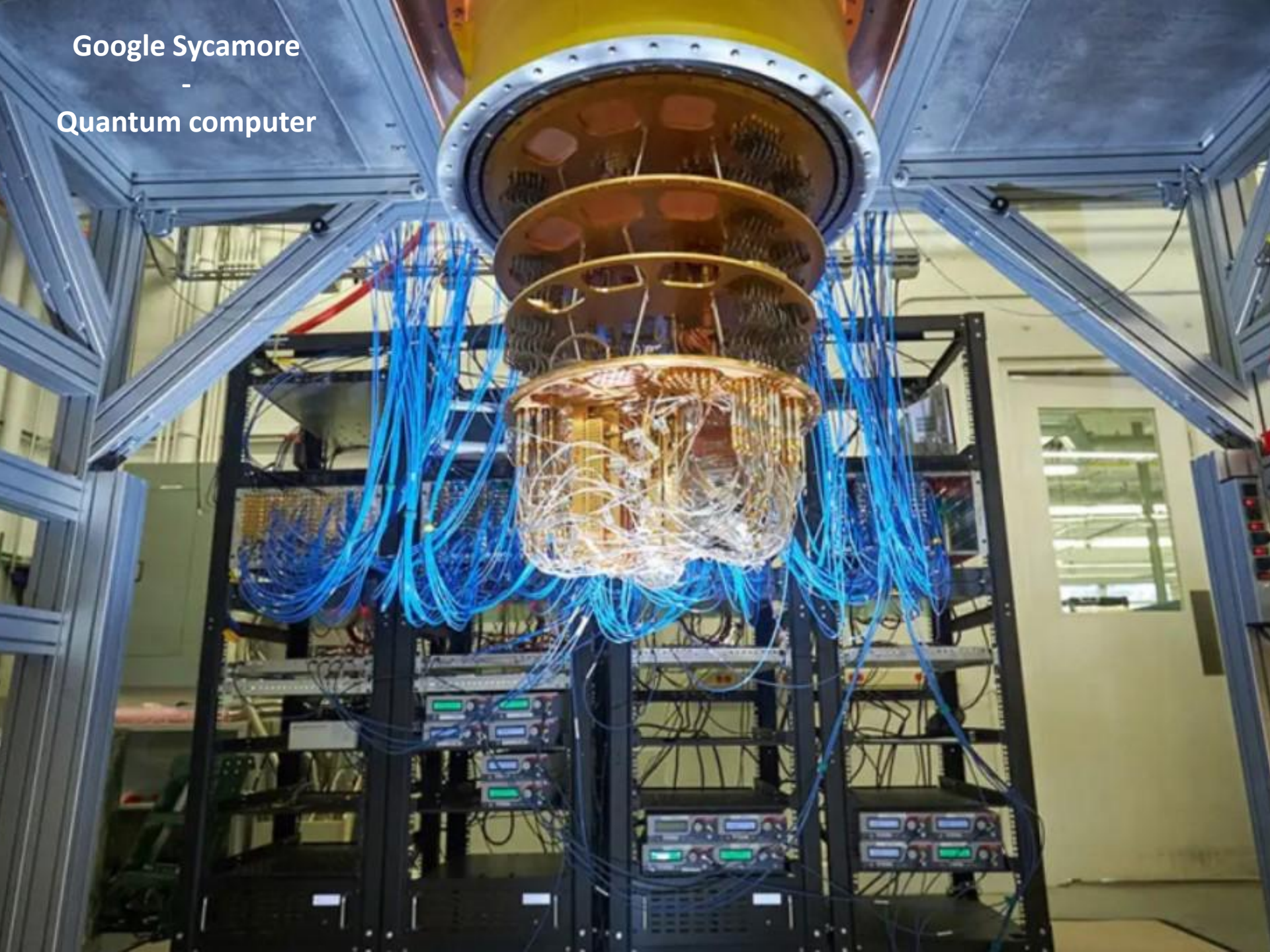
Computer Exascale - Classifica

Prime 10 posizioni della 60ª TOP500 (novembre 2022)[3]

Posizione (variazione)	Rmax Rpeak (PetaFLOPS)	Nome	Modello	N. core CPU	N. core acceleratore (es. GPU)	Interconnessione	Produttore	Località e nazione	Anno	Sistema operativo
1 –	1.102,00 1.685,65	Frontier	HPE Cray EX235a	591.872 (9.248 × 64-core Optimized 3rd Generation EPYC 64C @2,0 GHz)	36.992 × 220 AMD Instinct MI250X	Slingshot-11	HPE	Oak Ridge National Laboratory Stati Uniti	2022	Linux (HPE Cray OS)
2 –	442,010 537,212	Fugaku	Supercomputer Fugaku	7.630.848 (158.976 × 48-core Fujitsu A64FX @2,2 GHz)	0	Tofu interconnect D	Fujitsu	RIKEN Center for Computational Science Giappone	2020	Linux (RHEL)
3 –	309,10 428,70	LUMI	HPE Cray EX235a	150.528 (2.352 × 64-core Optimized 3rd Generation EPYC 64C @2,0 GHz)	9.408 × 220 AMD Instinct MI250X	Slingshot-11	HPE	EuroHPC JU Finlandia (Kajaani, Finlandia)	2022	Linux (HPE Cray OS)
4 NEW	174,70 255,75	Leonardo	BullSequana XH2000	110.592 (3.456 × 32-core Xeon Platinum 8358 @2,6 GHz)	13.824 × 108 Nvidia Ampere A100	Nvidia HDR100 Infiniband	Atos	EuroHPC JU Italia (Bologna, Italia)	2022	Linux
5 ▼ (4)	148.600 200.795	Summit	IBM Power System AC922	202.752 (9.216 × 22-core IBM POWER9 @3,07 GHz)	27.648 × 80 Nvidia Tesla V100	InfiniBand EDR	IBM	Oak Ridge National Laboratory Stati Uniti	2018	Linux (RHEL 7.4)
6 ▼ (5)	94,640 125,712	Sierra	IBM Power System S922LC	190.080 (8.640 × 22-core IBM POWER9 @3,1 GHz)	17.280 × 80 Nvidia Tesla V100	InfiniBand EDR	IBM	Lawrence Livermore National Laboratory Stati Uniti	2018	Linux (RHEL)
7 ▼ (6)	93,015 125,436	Sunway TaihuLight	Sunway MPP	10.649.600 (40.960 × 260-core Sunway SW26010 @1,45 GHz)	0	Sunway[4]	NRCPC	National Supercomputing Center in Wuxi Cina[4]	2016	Linux (RaiseOS 2.0.5)
8 ▼ (7)	70,87 93,75	Perlmutter	HPE Cray EX235n	? × ?-core AMD Epyc 7763 64-core @2,45 GHz	? × 108 Nvidia Ampere A100	Slingshot-10	HPE	NERSC Stati Uniti	2021	Linux (HPE Cray OS)
9 ▼ (8)	63,460 79,215	Selene	Nvidia	71.680 (1.120 × 64-core AMD Epyc 7742 @2,25 GHz)	4.480 × 108 Nvidia Ampere A100	Mellanox HDR Infiniband	Nvidia	Nvidia Stati Uniti	2020	Linux (Ubuntu 20.04.1)
10 ▼ (9)	61,445 100,679	Tianhe-2A	TH-IVB-FEP	427.008 (35.584 × 12-core Intel Xeon E5–2692 v2 @2,2 GHz)	35.584 × Matrix-2000[5] 128-core	TH Express-2	NUDT	National Supercomputer Center in Guangzhou Cina	2018[6]	Linux (Kylin)

https://it.wikipedia.org/wiki/TOP500

n°	Nazione	Numero di supercomputer nella TOP500	Rmax Rpeak (GFLOPS)	Core
1	Cina	162	514.491.614 1.132.071.398	28.865.036
2	Stati Uniti	127	2.122.791.370 3.216.124.169	27.858.532
3	Germania	34	219.253.860 331.231.244	4.030.068
4	Giappone	31	624.251.300 815.667.288	11.948.484
5	Francia	24	174.854.530 251.166.773	4.052.696
6	Regno Unito	15	64.078.644 92.563.552	1.971.888
7	Canada	10	41.208.360 71.911.529	845.984
8	Corea del Sud	8	88.682.560 128.264.430	1.551.012
9	Paesi Bassi	8	33.959.120 56.740.440	547.728
10	Brasile	8	46.729.150 88.175.939	775.232


Google Sycamore
-
Quantum computer

I.O.T.

Diego Vendramin – Non copiare o riprodurre

Perché ne parliamo oggi?

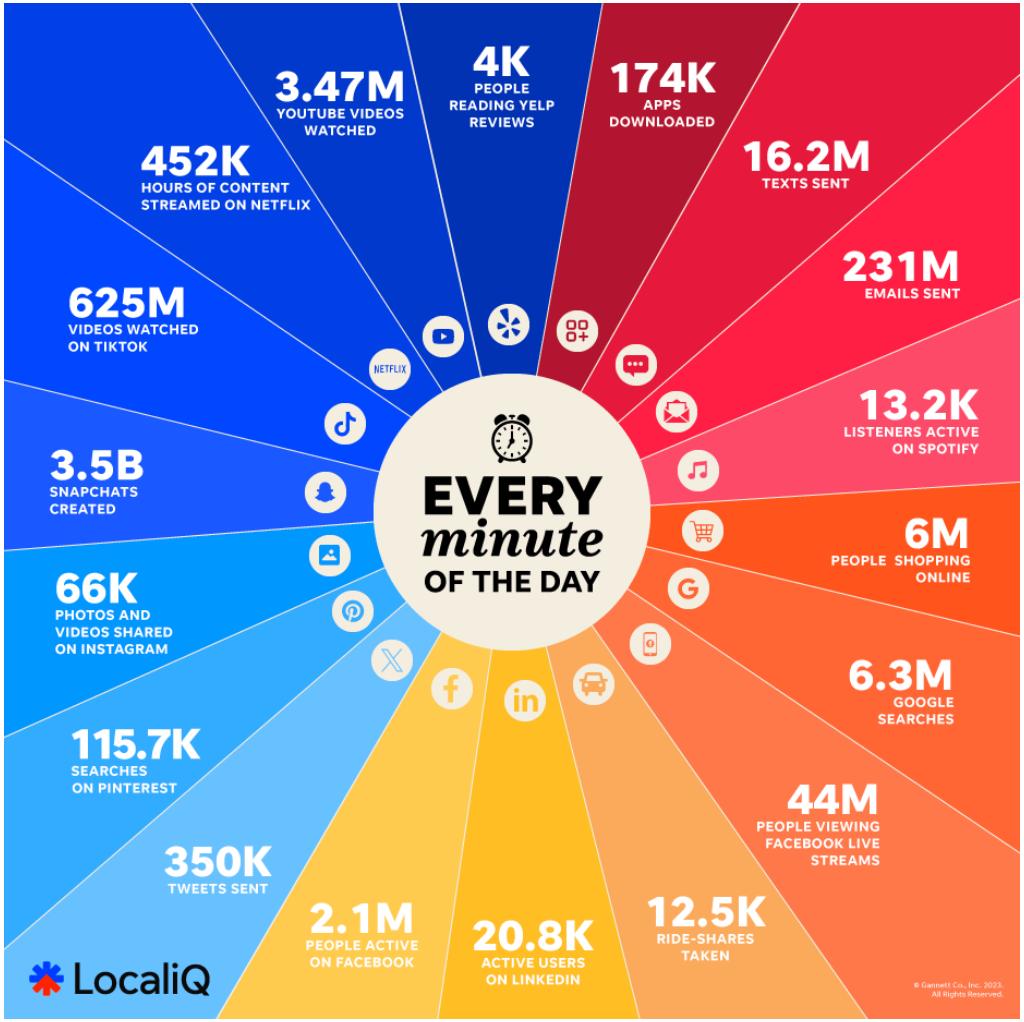
IOT Devices

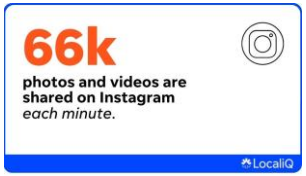
Region	2023	2030
Greater China	5.04 billion	8.57 billion
Europe	3.48 billion	6.56 billion
North America	3.22 billion	6.24 billion
Japan	0.93 billion	1.77 billion
Latin America	0.64 billion	1.54 billion
South Korea	0.38 billion	0.62 billion
South East Asia	0.35 billion	1.04 billion
Russia and Central Asia	0.29 billion	0.66 billion
Middle East and North Africa	0.28 billion	0.73 billion
India and South Asia	0.28 billion	1.03 billion
Sub-Saharan Africa	0.1 billion	0.26 billion
TOT	15 B	30 B

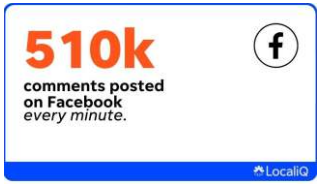
https://explodingtopics.com/blog/number-of-iot-devices

BIG DATA

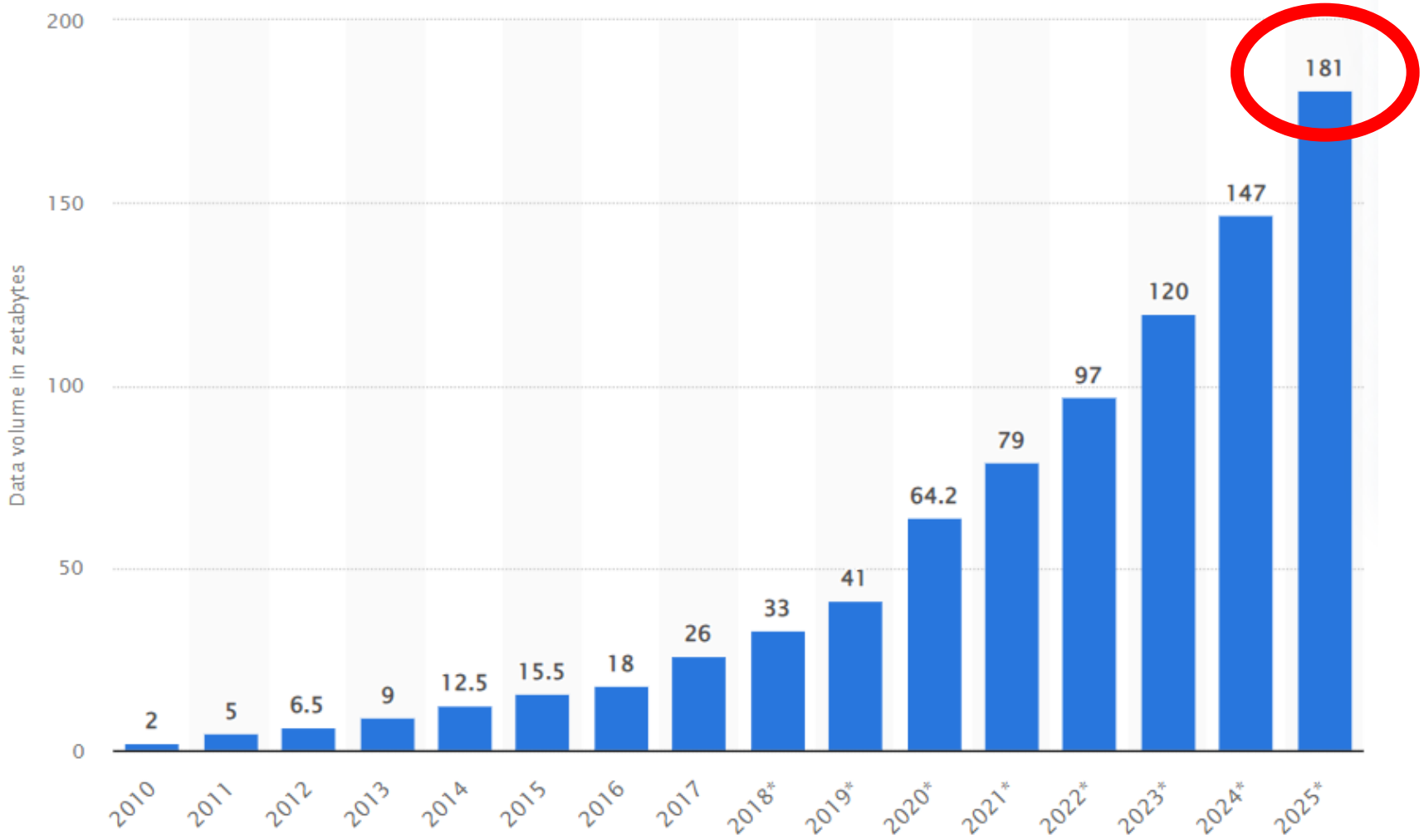
I Big Data sono davvero BIG







I Big Data sono davvero BIG



Fonte: https://www.statista.com/statistics/871513/worldwide-data-created/

Diego Vendramini – Non copiare o riprodurre

I Big Data sono davvero BIG

Numerico	Italia	US	Simbolo	Prefisso
10^6	Milione	Million	Mb	Megabyte
10^9	Miliardo	Billion	Gb	Gigabyte
10^12	Bilione	Trillion	Tb	Terabyte
10^15	Biliardo	Quadrillion	Pb	Petabyte
10^18	Trillione	Quintillion	Eb	Exabyte
10^21	Triliardo	Sixtillion	Zb	Zettabyte

181 Zettabyte

181.000.000.000.000.000.000.000

bytes

Diego Vendramini – Non copiare o riprodurre

I Big Data sono davvero BIG

1 Gb

181.000.000.000.000

Bilioni / Trillions

1,44 Mb

125.700.000.000.000.000

Biliardi / Quadrillions

Diego Vendramini – Non copiare o riprodurre

N° floppy	125.700.000.000.000.000 (biliardi)
Spessore	2 mm
mm	251.400.000.000.000.000 (biliardi)
Km	251.400.000.000 (miliardi)

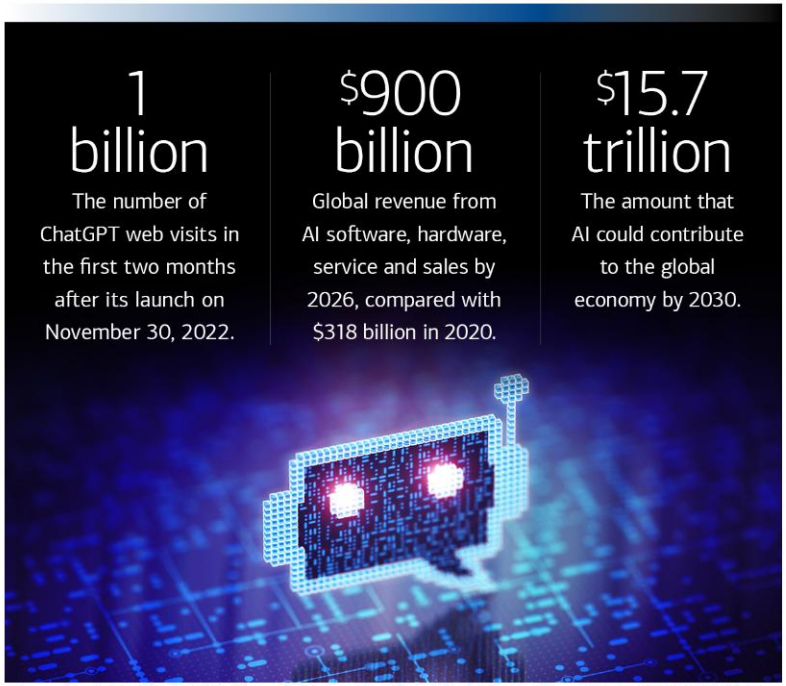
U.A. in Km
149.597.870
(milioni)
1.680
volte la distanza
Terra - Sole

31.888 anni di volo
continuo, a 900km/,
senza soste, attorno
alla Terra

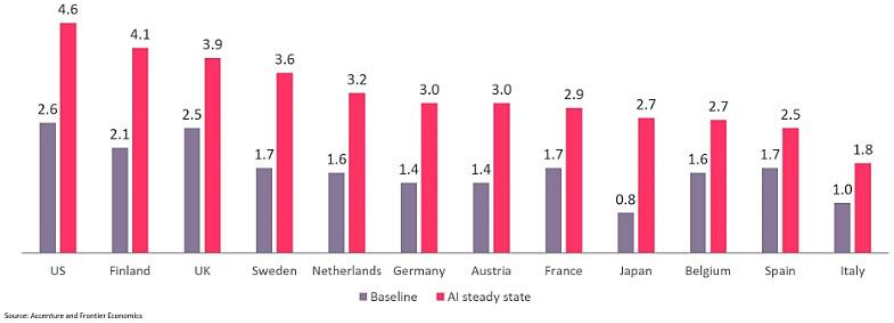
Big Data, Big Money

BANK OF AMERICA Who we serve Solutions Research & insights

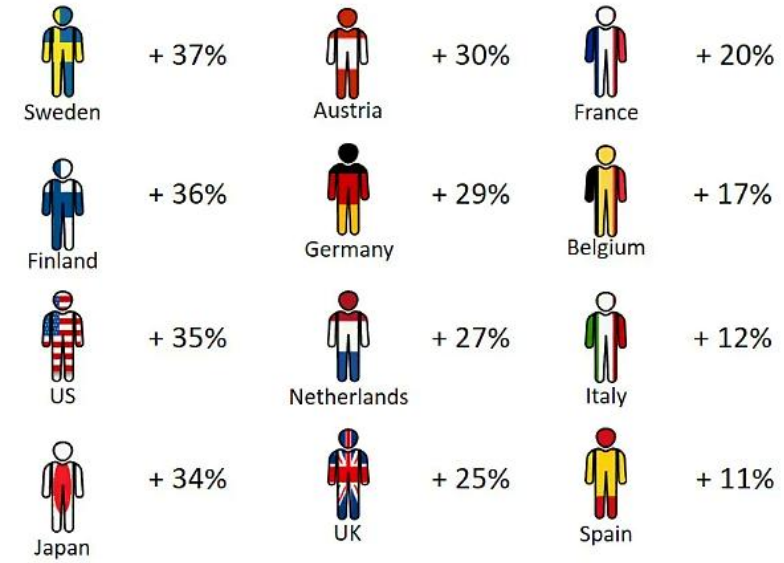
Businesses & Institutions
companies, he adds.



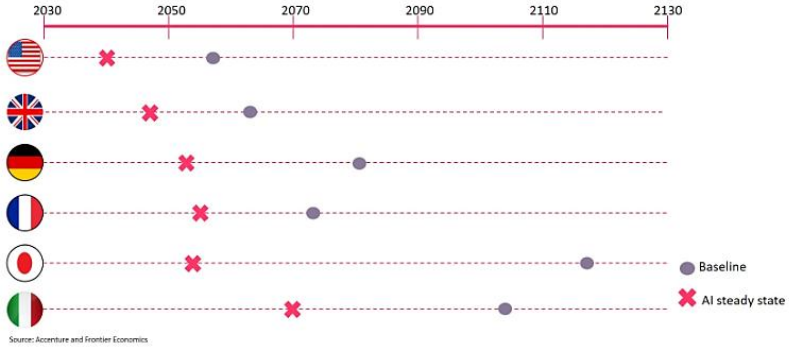
Sources, from left to right: BofA Global Research, "Me, Myself and AI—Artificial Intelligence Primer," February 28, 2023; BofA Global



Source: Accenture and Frontier Economics



Aumento percentuale della produttività del lavoro con l'intelligenza artificiale, rispetto ai livelli di produttività di base previsti entro il 2035



Source: Accenture and Frontier Economics

Attori in gioco: Istituzioni

I costruttori di modelli
Sviluppano e addestrano grandi modelli di intelligenza artificiale, i cosiddetti LLM (Large Language Model) o modelli di deep learning di varia natura.
Spendono enormi risorse in potenza di calcolo e in dati, spesso più di quanto riescano a monetizzare nell'immediato.

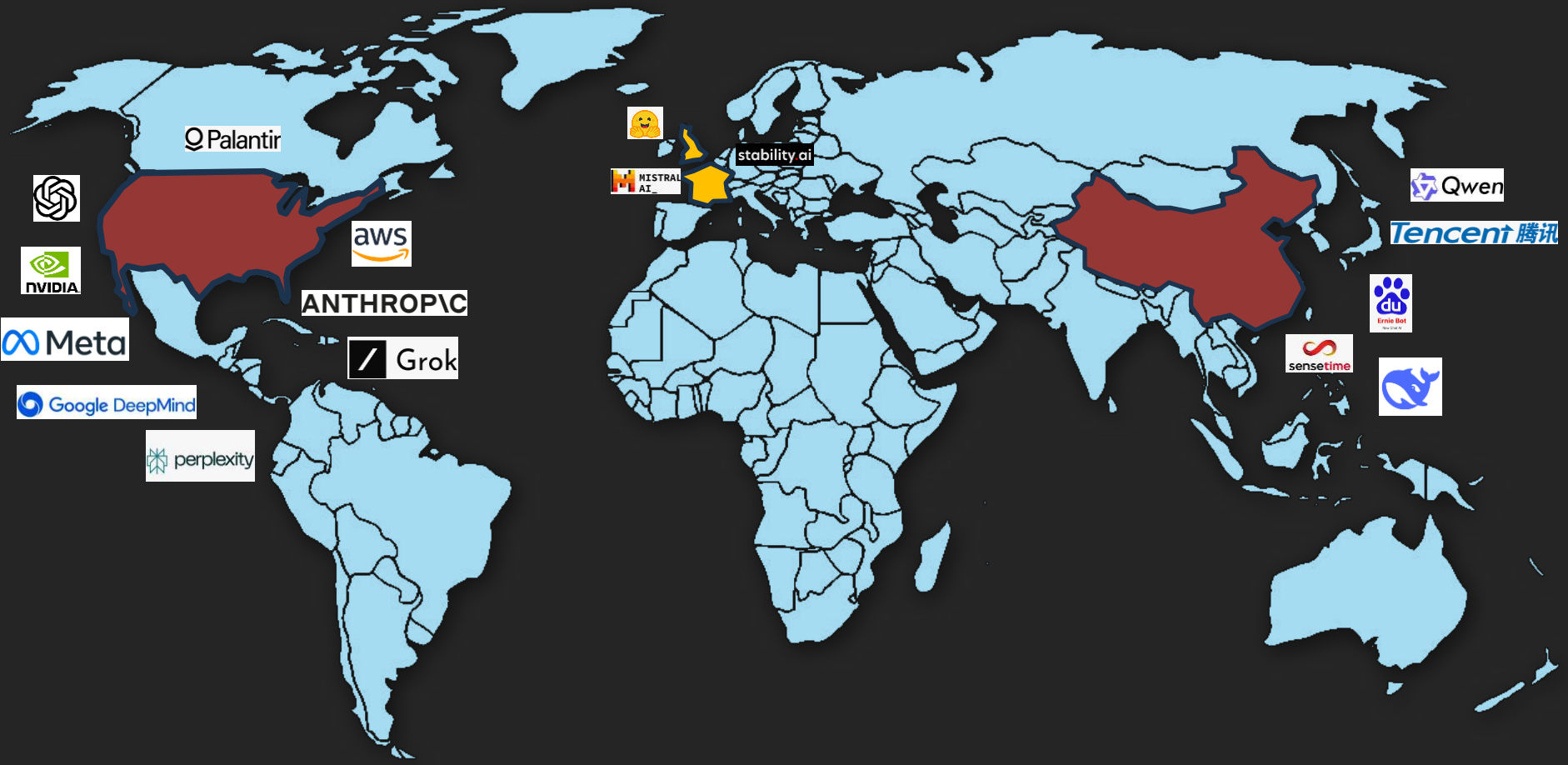
I provider di infrastrutture
Offrono la potenza di calcolo (GPU, CPU specializzate, TPU, ecc.), lo storage e i servizi cloud indispensabili ai "costruttori di modelli" per addestrare e distribuire le soluzioni di AI.
Vendono servizi e hardware di cui l'AI ha "disperatamente" bisogno.

I data owners
Tutti i detentori di dati (aziende, enti pubblici, social network, piattaforme online, società civile), che forniscono la materia prima per l'addestramento dei modelli AI.

Source: Accenture and Frontier Economics

Diego Vendramini – Non copiare o riprodurre

Attori in gioco: Paesi ed Istituzioni



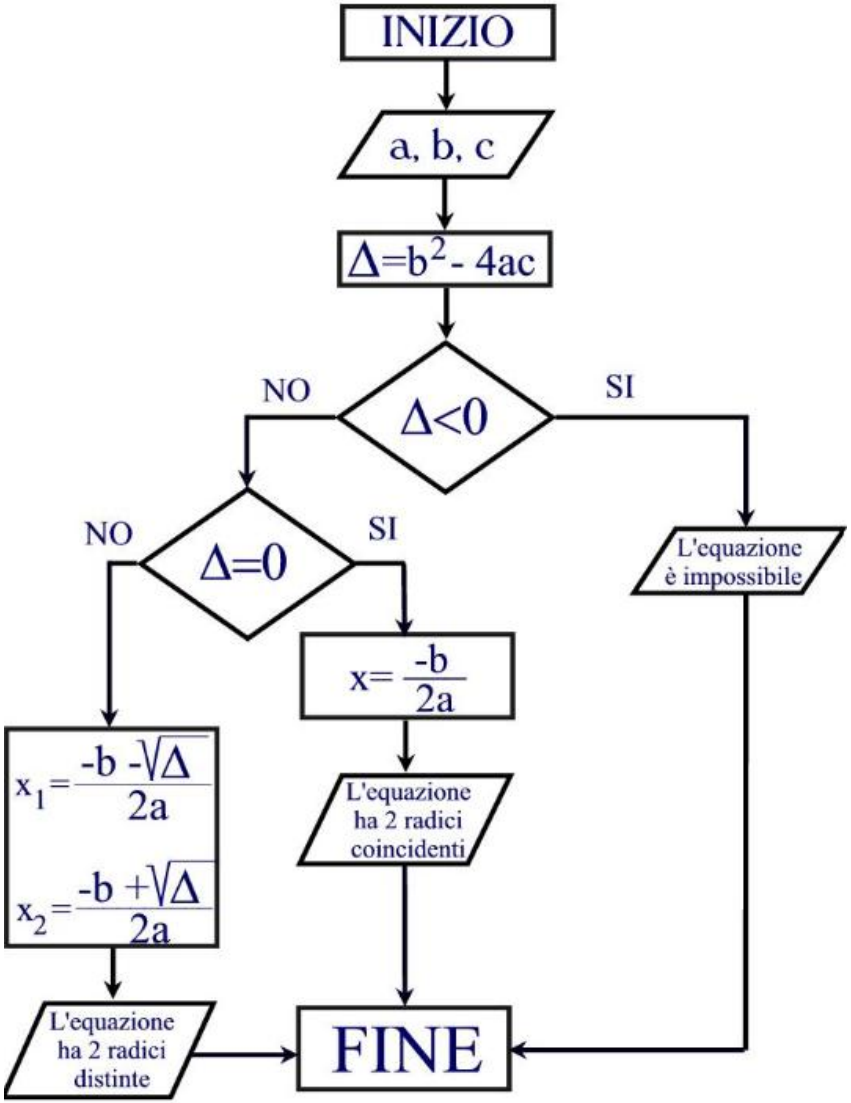
Diego Vendramini – Non copiare o riprodurre

Di cosa stiamo parlando? AI

Area multidisciplinare della scienza che si concentra sullo sviluppo di macchine e **algoritmi** in grado di agire comportamenti intelligenti.

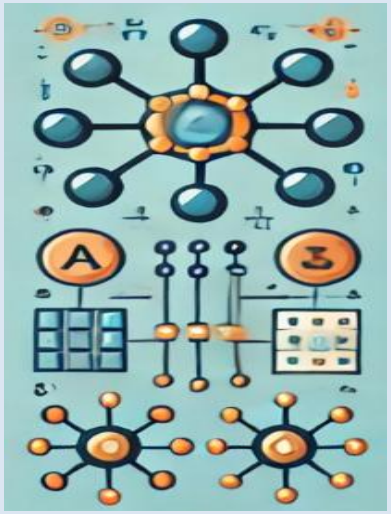
Algoritmi

Sequenze di istruzioni logiche e matematiche per risolvere un problema o eseguire un' operazione imitando il ragionamento umano e l'apprendimento, che consente alle macchine di prendere decisioni intelligenti o di apprendere da **grandi quantità di dati.**



AI – DIMENSIONI DI CLASSIFICAZIONE

1) Secondo le modalità con cui rappresenta e manipola la conoscenza



IA Statistica (Connessionista o basata su Dati)

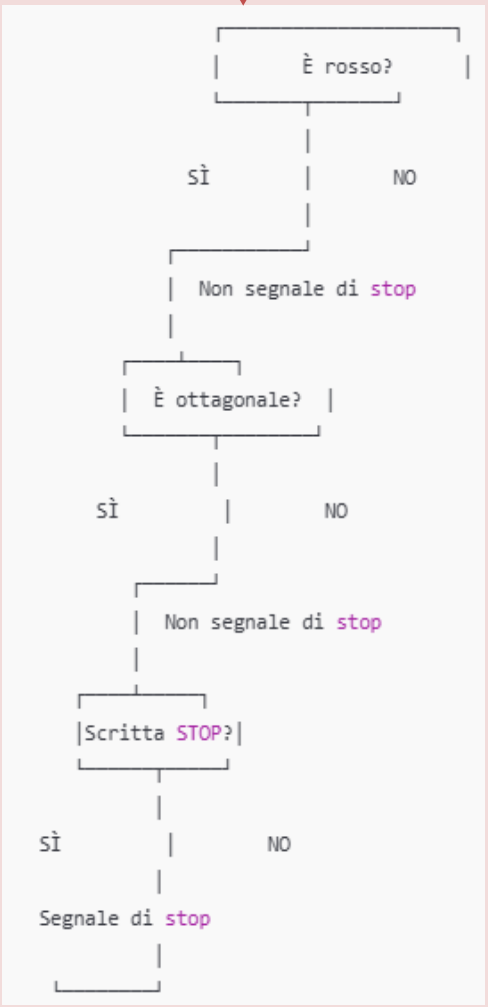
Fondata su **modelli matematici**, **statistici** (es. regressioni, Random Forest, Support Vector Machines) e le **reti neurali** acquisiscono la capacità di predire o classificare analizzando esempi e ottimizzando funzioni obiettivo. "Apprende" da pattern dai dati. La conoscenza non è codificata in forma di regole esplicite, ma "emerge" dall'addestramento su dataset.

AI Logico-Simbolica (Basata su regole)

La conoscenza è rappresentata attraverso simboli, regole logiche, ontologie e strutture formali. Inferisce per dedurre a partire da premesse note e regole o processi formalizzati.
La conoscenza è codificata in forma di regole "*if-then*", alberi decisionali, grafi della conoscenza o logiche descrittive.

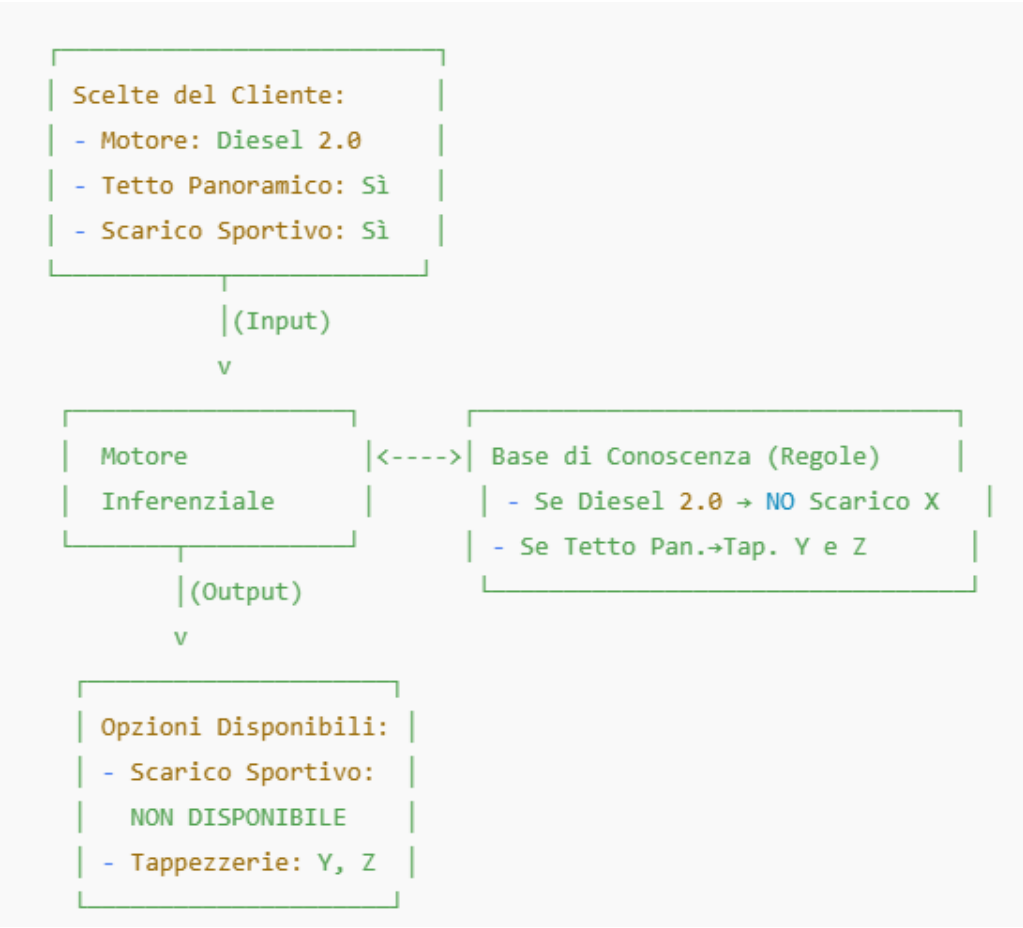
2 Approcci

AI Logico simbolica



AI Statistica

Esempio di AI Logico-Simbolica (Basata su regole): Configuratore online



Esempi di AI Logico-Simbolica (Basata su regole): Ausilio Diagnostico

Input:

Sintomi (febbre, tosse, livello di ossigeno, esiti di esami del sangue), condizioni pregresse del paziente, età, stili di vita.

Elaborazione:

Il motore inferenziale applica regole

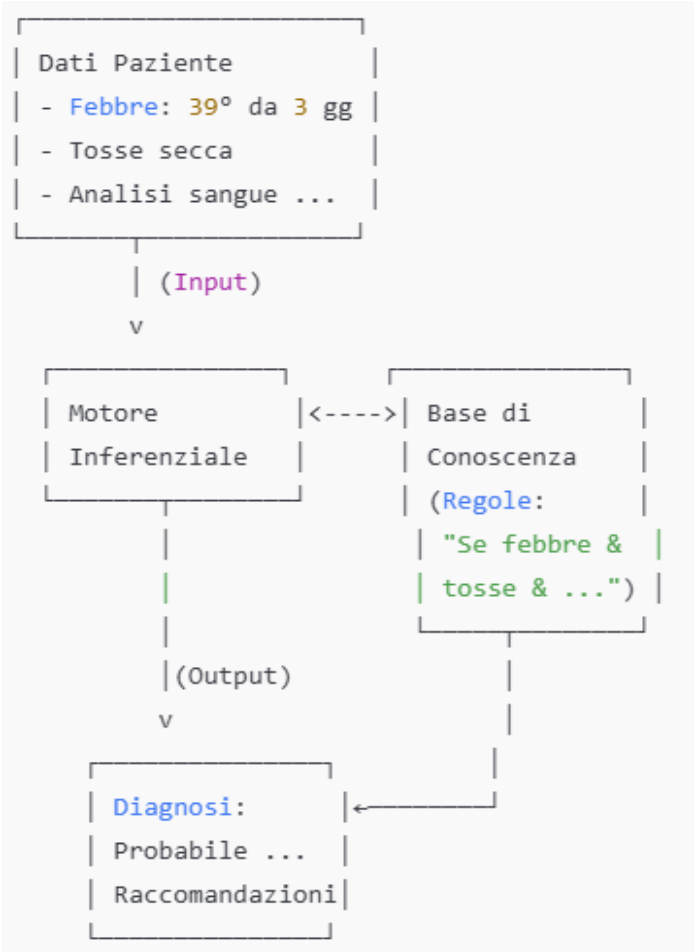
"Se il paziente ha febbre alta da più di 3 giorni e ha tosse secca e alterazione di determinate cellule nel sangue

= >

considerare l'ipotesi di polmonite

Output:

diagnosi probabile o raccomandazioni su esami diagnostici aggiuntivi.



AI – DIMENSIONI DI CLASSIFICAZIONE

2) Secondo lo scopo che si danno

DISCRIMINATIVE A.I.

Scopo: Si concentra sul **distinguere** tra diverse classi o categorie nei dati. Modella il confine decisionale tra le classi.

Funzione: Dato un input di dati, predice un'etichetta o una classe. Ad esempio, nella rilevazione dello spam, i modelli discriminativi **classificano** le email come "spam" o "non spam".

Esempi di Algoritmi: Regressione, Support Vector Machines (SVM), Alberi Decisionali, Random Forests, Reti Neurali (per compiti di classificazione).

Chiave: Si focalizza sul trovare il confine tra le classi di dati per fare **previsioni** accurate. Non tenta di modellare la distribuzione dei dati, ma solo la relazione tra input e output.

GENERATIVE A.I.

Scopo: Può **generare nuovi dati che** sembrano provenire dalla stessa distribuzione dei dati di addestramento.

Funzione: Dato un input, può generare dati simili a quelli del set di addestramento. Ad esempio, può creare **immagini, testi o suoni** simili a quelli su cui è stato addestrato.

Esempi di Algoritmi: ANN,Variational Autoencoders (VAE), Generative Adversarial Networks (GAN) , CCN, Modelli di Markov, Boltzmann Machines

Chiave: Tenta di modellare la distribuzione dei dati e può **generare nuovi campioni** che appartengono a tale distribuzione. È utile per la **generazione di contenuti, il completamento di dati mancanti e la simulazione.**

Alberi decisionali

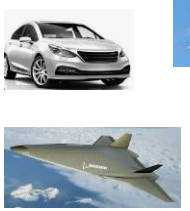
Clusterizzazione

Immagine	Forma	Lunghezza (m)	Larghezza (m)	Altezza (m)	CLASSE
	Struttura compatta con profilo aerodinamico	3,5 - 5,0	1,5 - 2,0	1,4 - 1,6	Auto
	Corpo allungato con due ruote	1,8 - 2,2	0,7 - 0,9	1,0 - 1,3	Moto
	Telaio sottile con due ruote	1,7 - 2,0	0,5 - 0,6	1,0 - 1,2	Bicicletta
	Rettangolare e allungato, finestrini su fiancate	8,0 - 12,0	2,5 - 3,0	3,0 - 3,5	Autobus
	Fusoliera allungata con ali laterali	30,0 - 70,0	28,00 - 60,00	10,00 - 15,00	Aereo
	Scafo lungo e ampio con ponte superiore	50,0 - 400,00	10,00 - 60,00	15,00 - 70,00	?

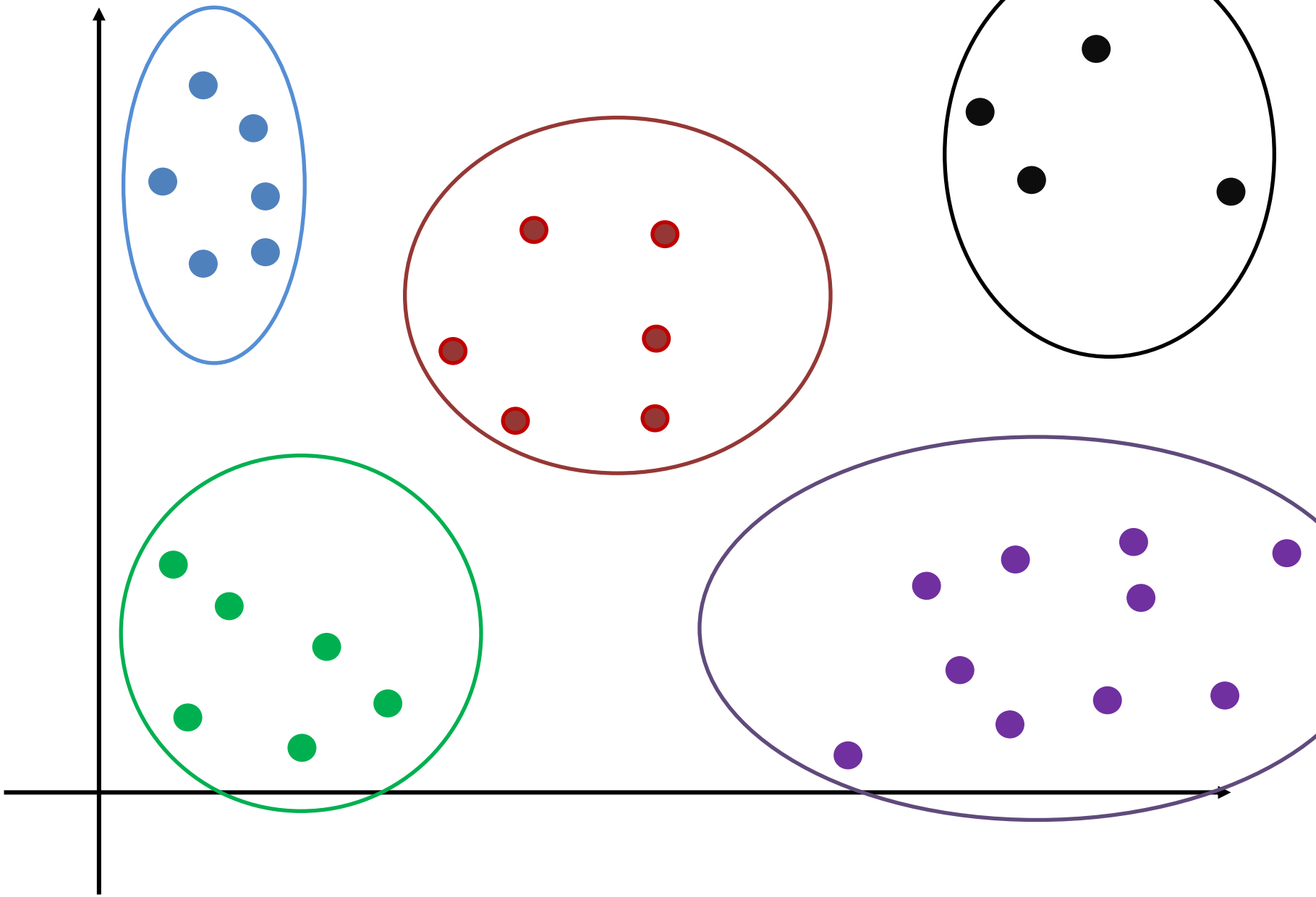
Regressione

Immagine	Forma	Lunghezza (m)	Larghezza (m)	Altezza (m)	PESO (t)
	Struttura compatta con profilo aerodinamico	3,5 - 5,0	1,5 - 2,0	1,4 - 1,6	1 – 2 t
	Corpo allungato con due ruote	1,8 - 2,2	0,7 - 0,9	1,0 - 1,3	0,1 – 0,2 t
	Telaio sottile con due ruote	1,7 - 2,0	0,5 - 0,6	1,0 - 1,2	0,01 – 0,02 t
	Rettangolare e allungato, finestrini su fiancate	8,0 - 12,0	2,5 - 3,0	3,0 - 3,5	10 – 18 t
	Fusoliera allungata con ali laterali	30,0 - 70,0	28,00 - 60,00	10,00 - 15,00	22 – 250 t
	Scafo lungo e ampio con ponte superiore	50,0 - 400,00	10,00 - 60,00	15,00 - 70,00	? t

BIGBUS VIENNA



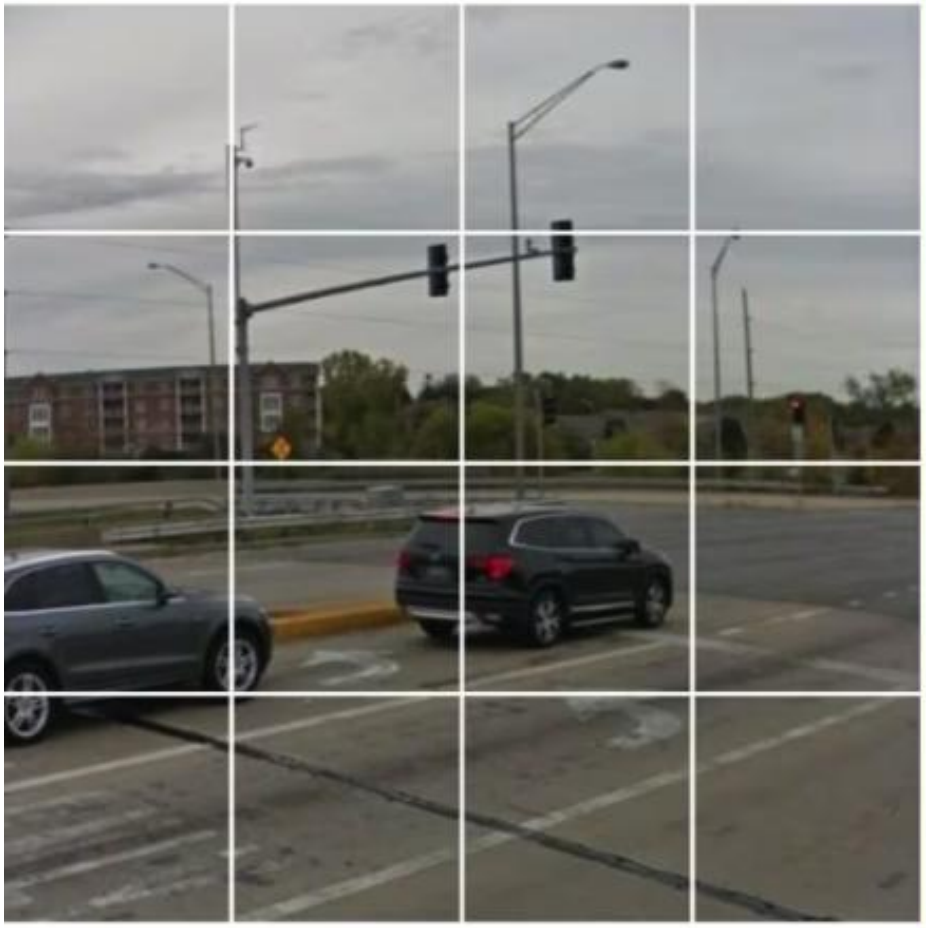
MSC



Select all squares with

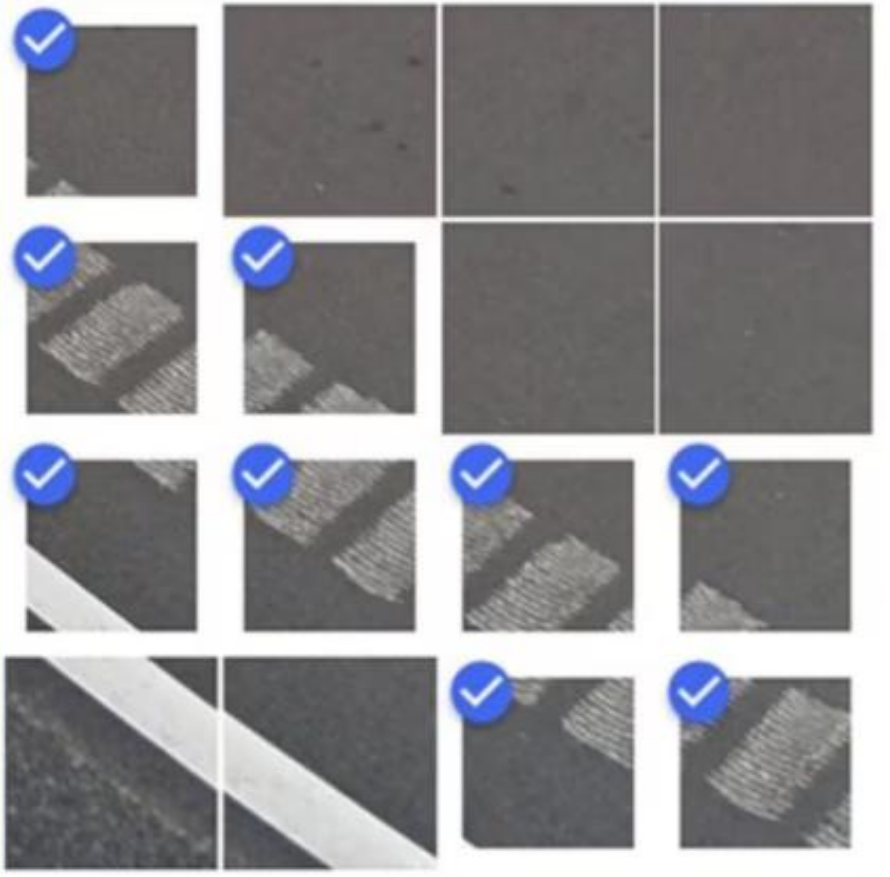
traffic lights

If there are none, click skip



Select all squares with

crosswalks



I Big Data sono complessi

- VOLUME
- VELOCITA’
- VARIETA’
- VERIDICITA’
- VALORE

STRUTTURATI
Tabelle, database, fogli di calcolo…

NON STRUTTURATI
testi, immagini, audio, video, post social media, e-mail…

Diego Vendramini – Non copiare o riprodurre

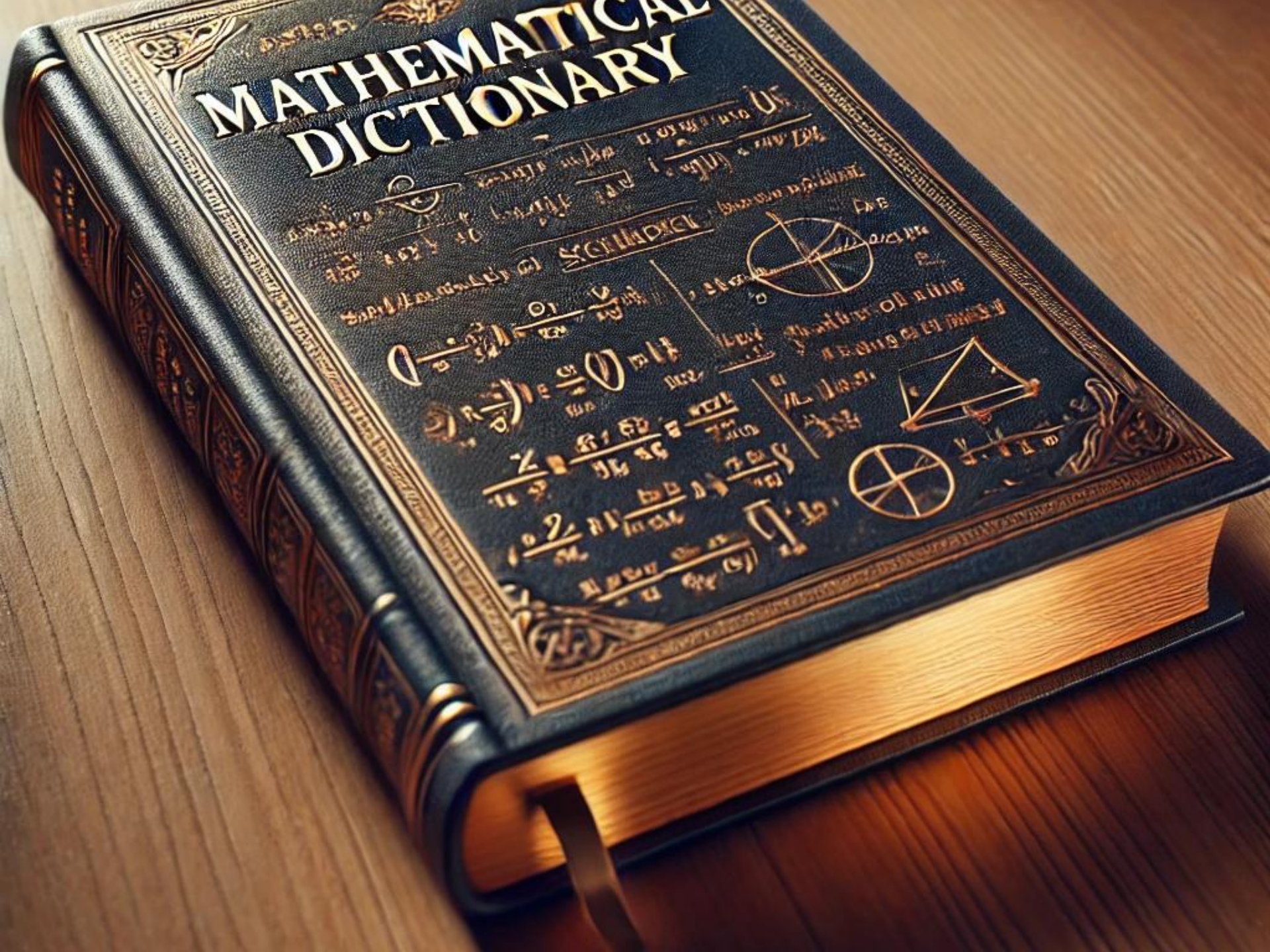
Concetti di base: Tipi di apprendimento

SUPERVISIONATO: algoritmo viene addestrato su un set di dati etichettati (set di addestramento) per imparare a fare previsioni o prendere decisioni. Congruità tra input ed output.

NON SUPERVISIONATO: l'algoritmo analizza i dati senza la presenza di etichette o target di output predefiniti.

CON RINFORZO: l'agente riceve feedback positivi o negativi, sotto forma di ricompense o penalità, in base alle azioni intraprese.

Diego Vendramini – Non copiare o riprodurre


MATHEMATICAL
DICTIONARY

Dizionario Minimo: Vettore

Un vettore può essere rappresentato come un **array di numeri**:

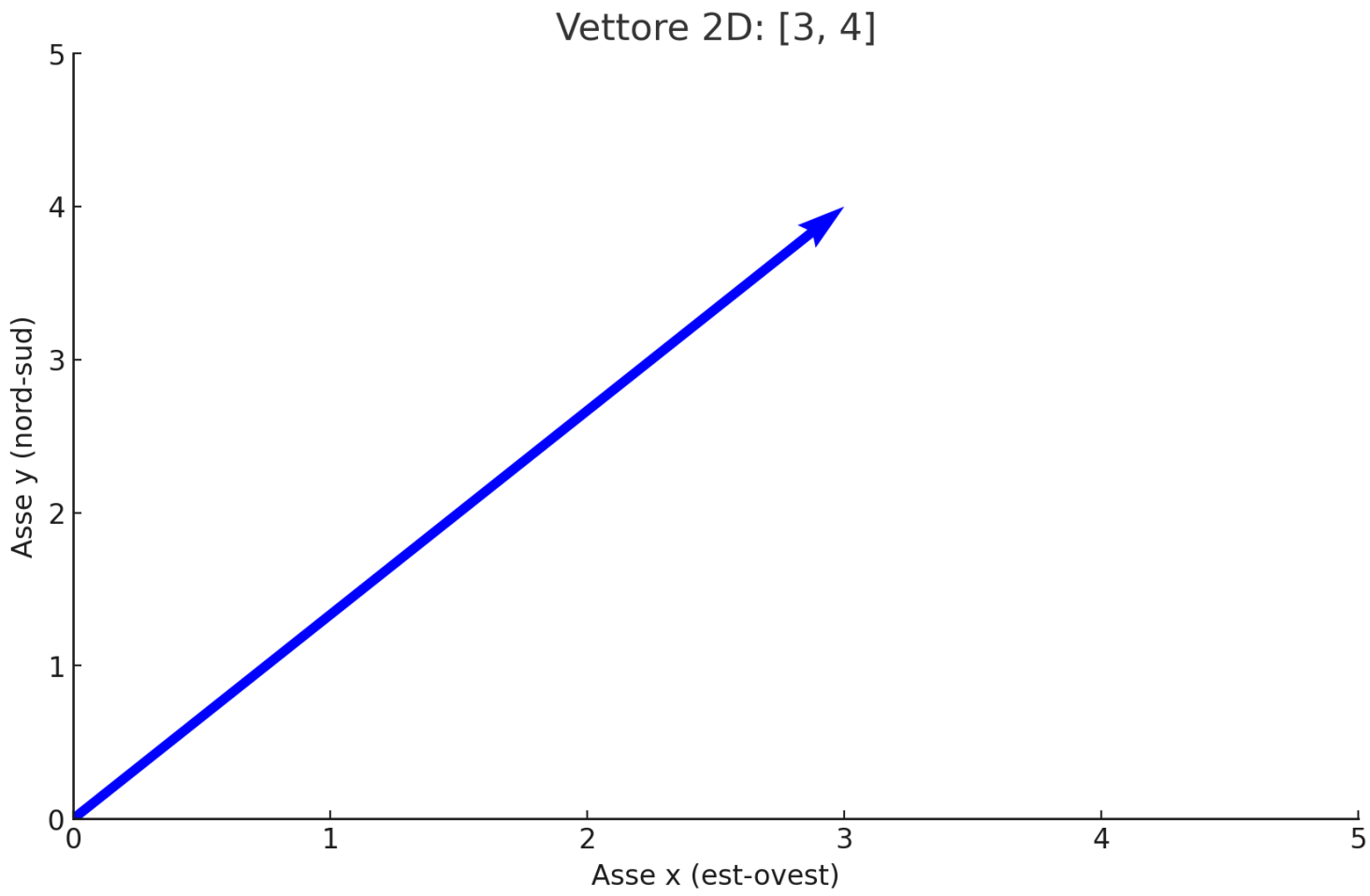
$$\mathbf{v} = [v_1, v_2, \ldots, v_n]$$

- Ogni numero (componente) indica **quanto si sposta** in una determinata direzione.
- Ad esempio, in uno spazio 2D

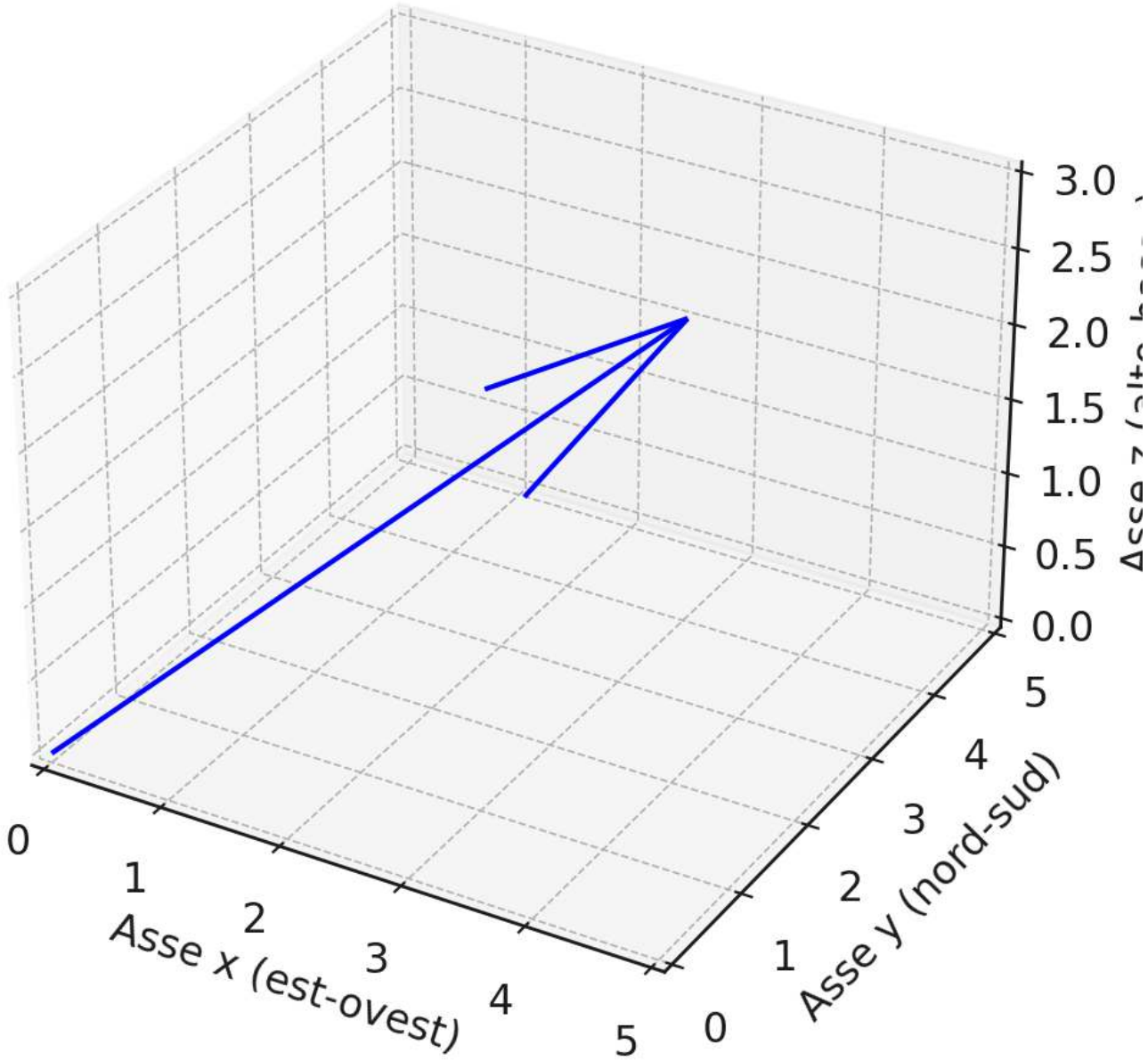
$$\mathbf{v} = [3, 4]$$

- rappresenta un movimento di 3 unità in x e 4 unità in y.

Diego Vendramini – Non copiare o riprodurre



Diego Vendramini – Non copiare o riprodurre


Vettore 3D: [3, 4, 2]
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
1
2
3
4
5
Asse x (est-ovest)
0
1
2
3
4
5
Asse y (nord-sud)

Matrici

Una **matrice** è una tabella rettangolare di numeri (o simboli) disposti in **righe** e **colonne**. È una struttura matematica utilizzata per rappresentare e manipolare dati o relazioni.

Una matrice di dimensione m×n

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

- m: Numero di righe.
- n: Numero di colonne.
- a_{ij}: L'elemento della matrice nella riga i e nella colonna j.

Diego Vendramini – Non copiare o riprodurre

Immagini

Un'immagine a colori è rappresentata come un **array tridimensionale** in cui:

La prima dimensione rappresenta l'**altezza** (numero di righe di pixel).

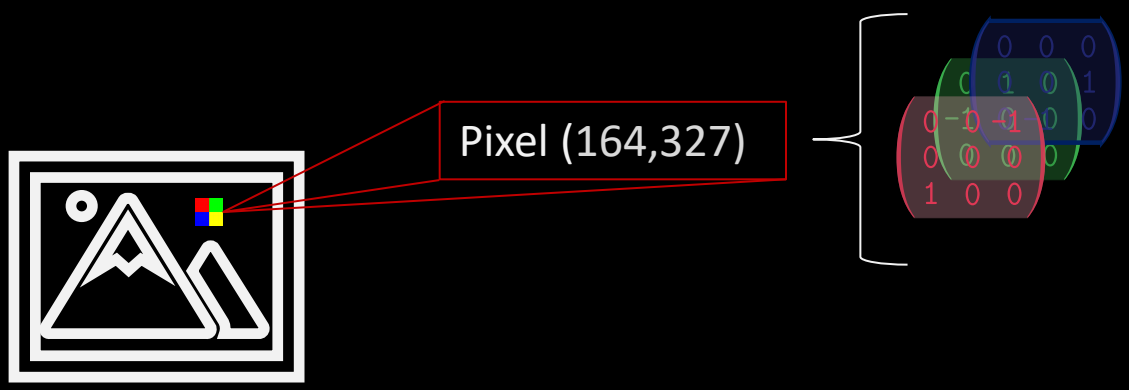
La seconda dimensione rappresenta la **larghezza** (numero di colonne di pixel).

La terza dimensione contiene i **canali di colore** (tipicamente 3: Rosso Verde Blu - RGB).

REDUCED SPEED
30
AHEAD

📐 Altezza (righe): 164
📏 Larghezza (colonne): 327
🌈 Colori RGB: [🔴 Rosso 124, 🟢 Verde 201, 🔵 Blu 179]

(Y,X)=(164,327)
Pixel (164,327)=[124,201,179]



💻 Tensore : (x, y, 3)

T[164,327] → [124,201,179]

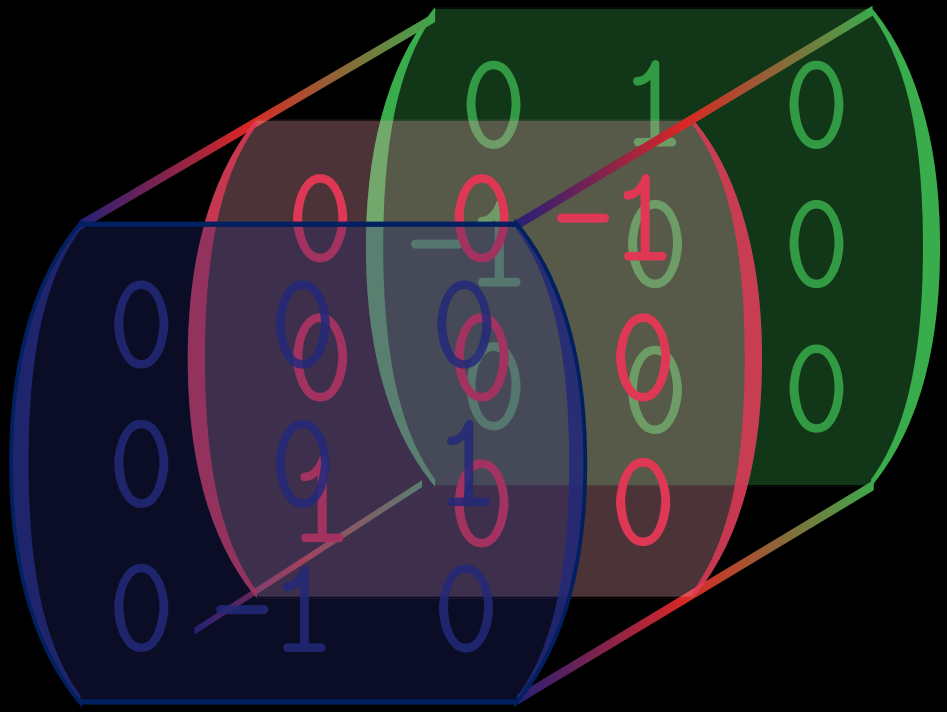
Tensore

Un tensore è un “array multidimensionale” di numeri, che serve per rappresentare e manipolare dati e parametri nei modelli di machine learning applicando operazioni vettoriali e matriciali.

- un **numero** singolo (anche detto scalare) -> è un tensore di ordine 0
- un **vettore** di numeri (ad es. [2, 3, 7]) -> tensore di ordine 1
- una **matrice** $\begin{pmatrix} 1 & 4 & 5 \\ 2 & 7 & 9 \end{pmatrix}$ ->tensore di ordine 2

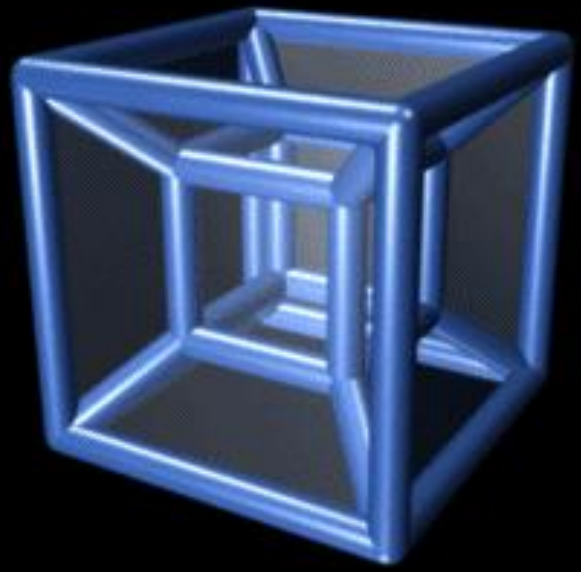
Tensori

tensore di ordine 3
ϵ_{ijk}



https://it.wikipedia.org/wiki/Tensore#/media/File:Epsilontensor.svg

tensore di ordine 4:
ipercubo (tesseratto)



https://en.wikipedia.org/wiki/File:8-cell-simple.gif

Nel deep learning tutto è rappresentato come tensori.

I tensori gestiscono dati di qualsiasi forma e dimensione:

- testo (vettori di embedding)
- Immagini
- segnali audio
- parametri di rete neurale

Servono a compiere operazioni vettoriali (somma, prodotto, trasformazioni lineari) in modo molto veloce, soprattutto sfruttando GPU e TPU

L'intero framework dell'apprendimento automatico si basa su queste operazioni tensoriali, che rappresentano retropropagazione, convoluzioni, moltiplicazioni di matrici

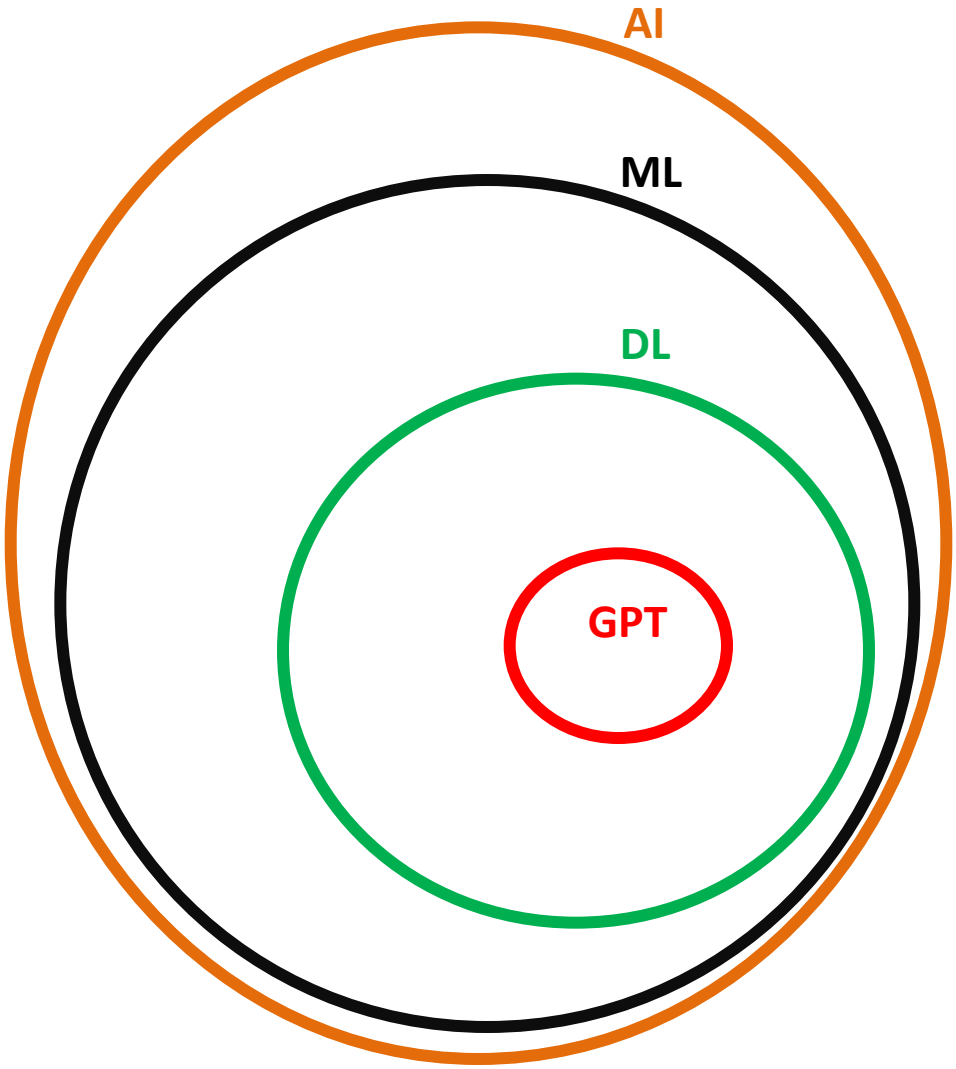
Le immagini, ad esempio, sono tensori 3D (altezza, larghezza, canali RGB)

Un' immagine a colori 256×256 pixel con 3 canali (**RGB**) è un tensore con shape (256, 256, 3).
Un "batch" di 32 immagini è un tensore 4D con shape (32, 256, 256, 3)

MACHINE LEARNING Vs DEEP LEARNING

MACHINE LEARNING:
Le macchine apprendono con dati spesso elaborati a priori (normalizzazioni, etichette, struttura)
La qualità dell'output è fortemente condizionata dall'attività di input.

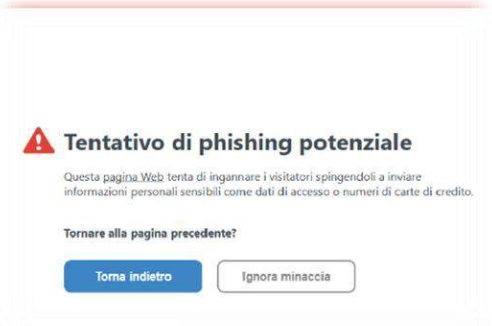
DEEP LEARNING:
E' una parte del machine learning noto per la sua capacità di apprendere automaticamente caratteristiche rilevanti dai dati, rispetto ad un modello di addestramento, eliminando così la necessità di una pre-elaborazione manuale. Usa spesso le **Reti Neurali Artificiali (ANN)**

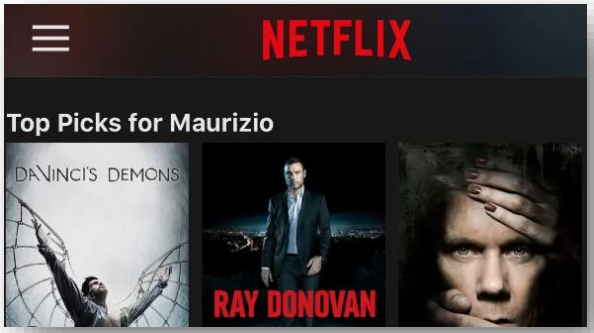


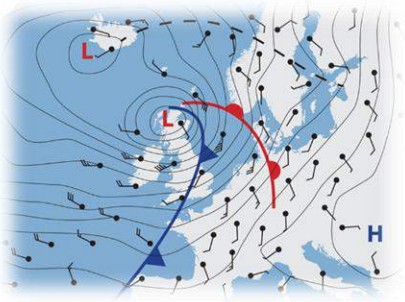
Diego Vendramini – Non copiare o riprodurre

Applicazioni?

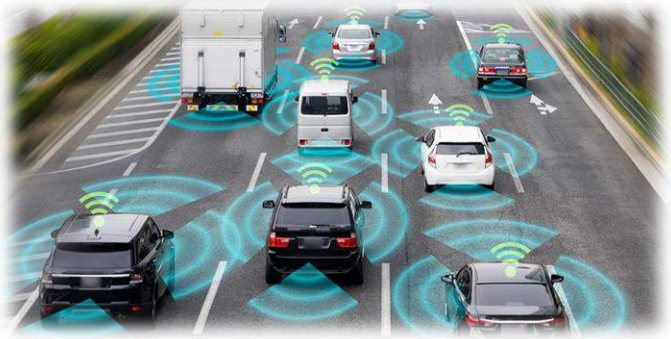
ML







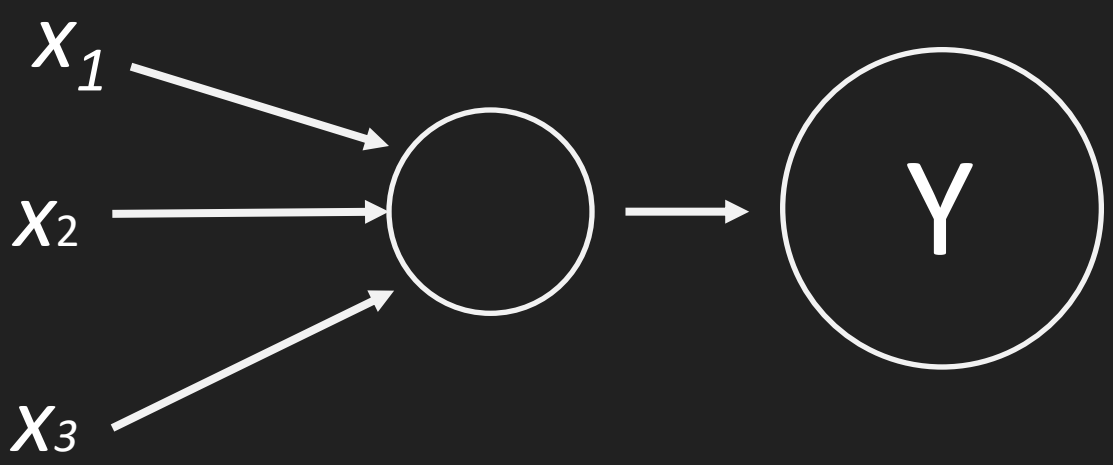
DL



Rete neurale - Funzionamento

1943 - Warren McCulloch e Walter Pitts

INPUT NEURONE OUTPUT



$$y = f\left(\sum_{i=1}^{n} x_i\right)$$

Output 1 = il neurone si attiva

Output 0 = il neurone non si attiva

Diego Vendramini – Non copiare o riprodurre

Rete neurale - Funzionamento

1943 - Warren McCulloch e Walter Pitts

La funzione di attivazione (funzione step) è così definita:

$$f(x) = \begin{cases} 1 & \text{se } \sum_{i=1}^{n} x_i \geq \theta \\ 0 & \text{altrimenti} \end{cases}$$

Output 1 = il neurone si attiva

Output 0 = il neurone non si attiva

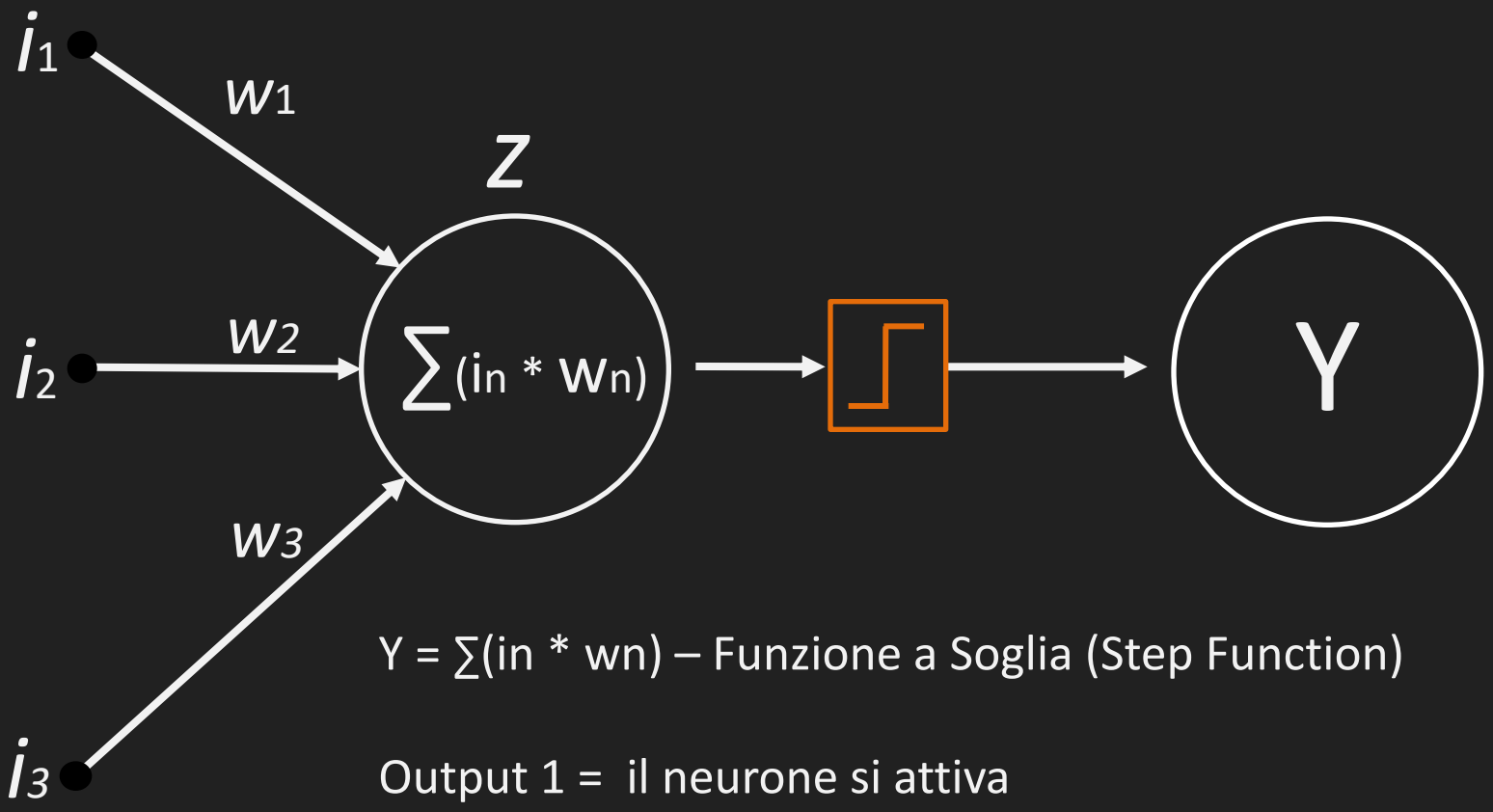
Diego Vendramini – Non copiare o riprodurre

Rete neurale – Funzionamento

Rosenblatt 1957

$$y = f\left(\sum_{i=1}^{n} w_i x_i + b\right)$$

INPUT LIVELLO 1 LIVELLO OUTPUT



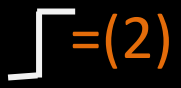
Y = ∑(in * wn) – Funzione a Soglia (Step Function)

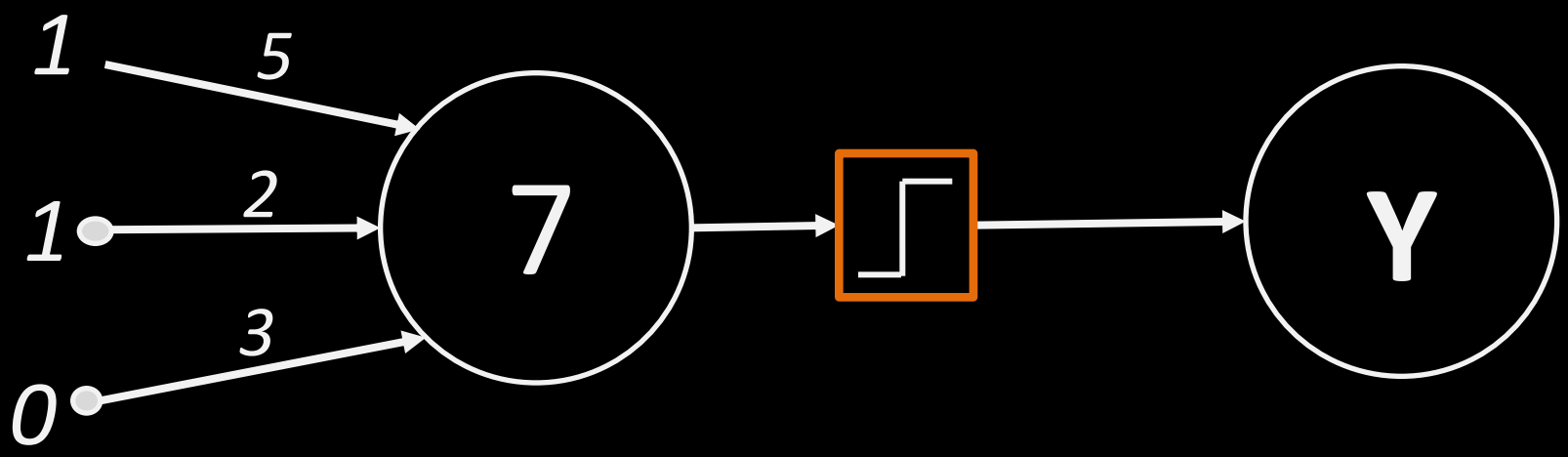
Output 1 = il neurone si attiva
Output 0 = il neurone non si attiva

Esempio di attivazione del neurone

- prezzo <15 k euro
- colore rosso
- consumo < 22 km/l

i1= prezzo (0;1) w1= 5
i2= colore (0;1) w2= 2
i3=consumo (0;1) w3= 3

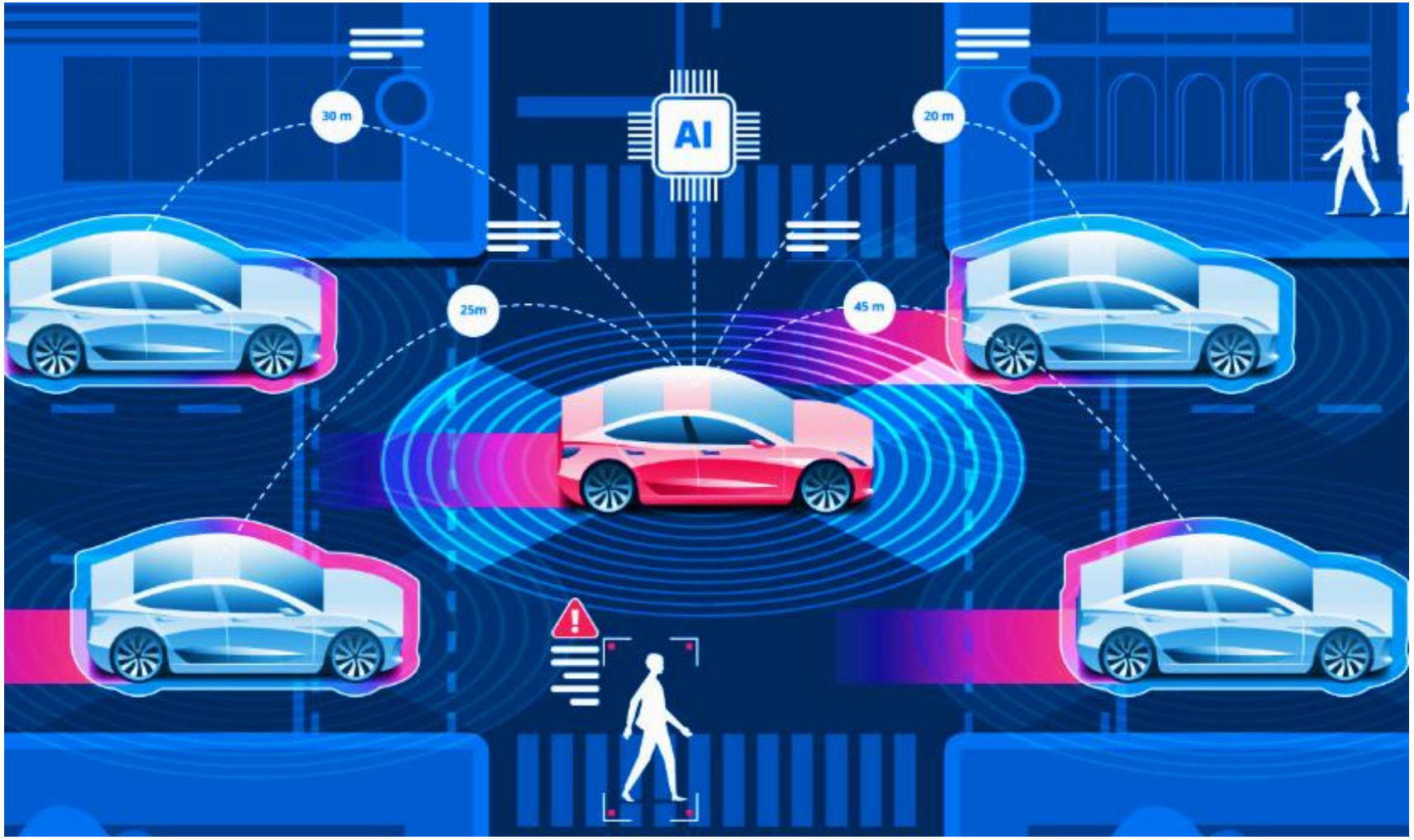




$\sum(in * wn)) = z$ z - Soglia

Output 1 Se z > Soglia => Neurone Attivo
Output 0 Se z < Soglia => Neurone Inattivo

I limiti dei primi modelli



Diego Vendramini – Non copiare o riprodurre

I limiti del modello di McCulloch e Pitts

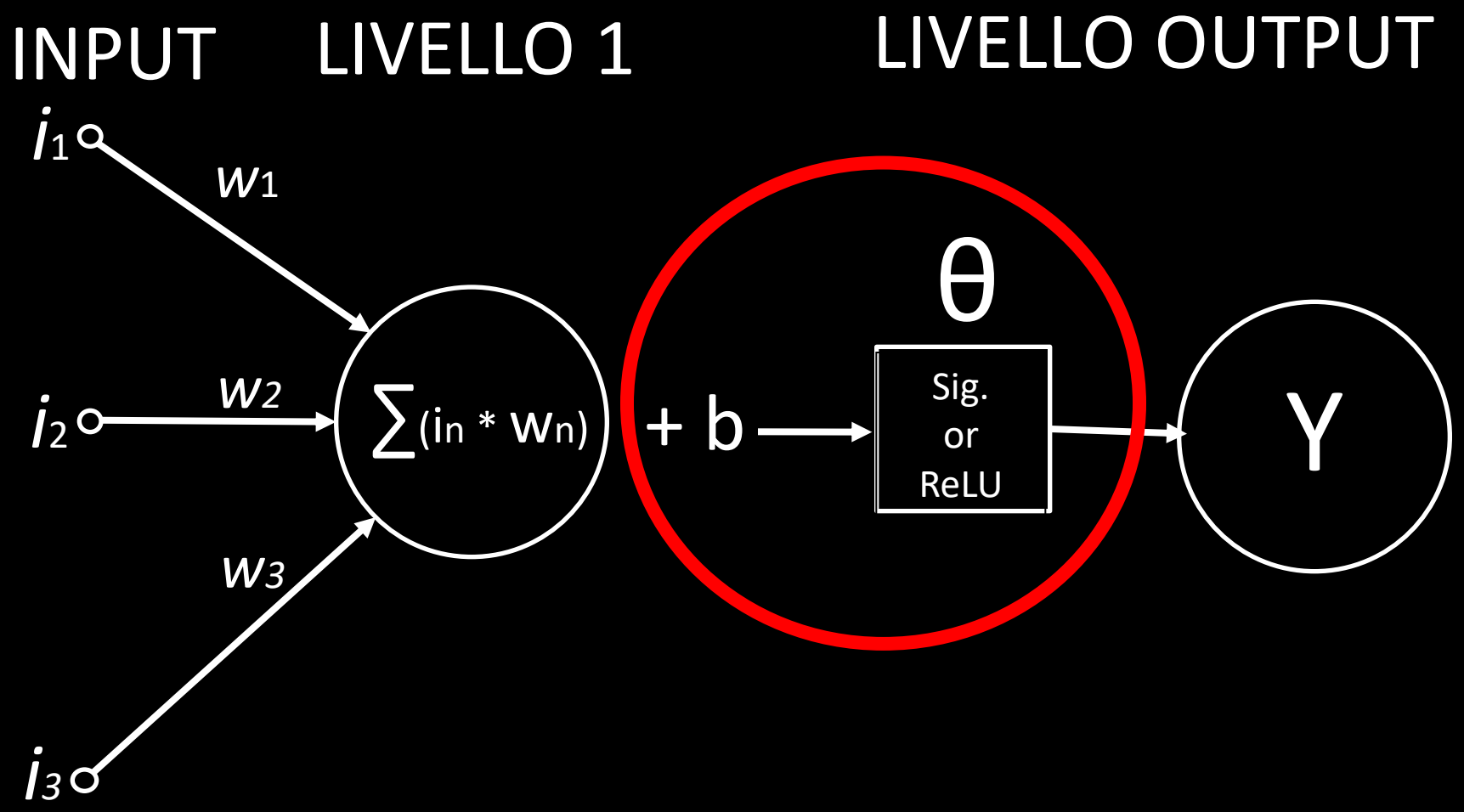
Binarietà

Soglia fissa

Mancanza di apprendimento

Diego Vendramini – Non copiare o riprodurre

Rete neurale - Funzionamento

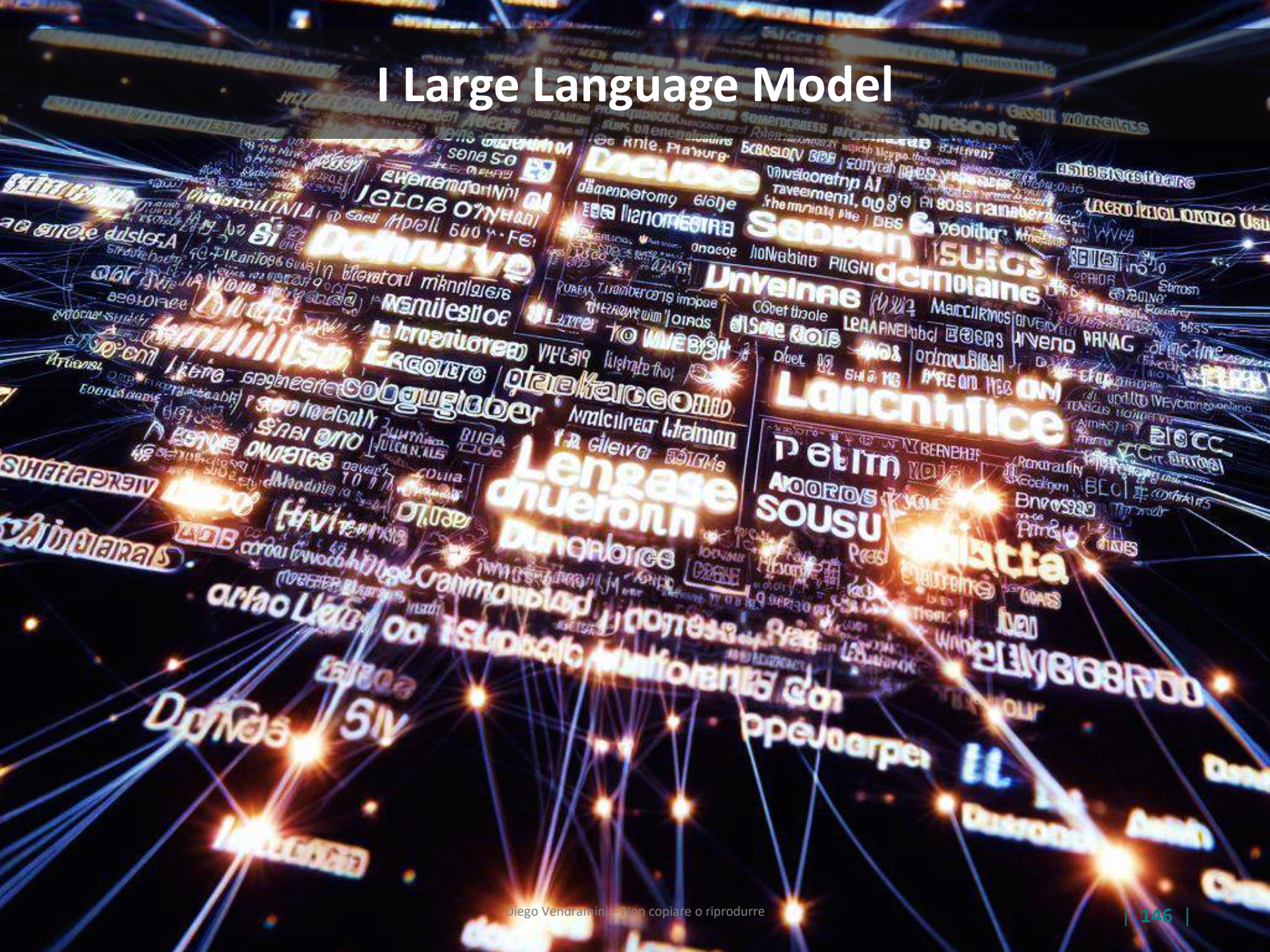


Bias

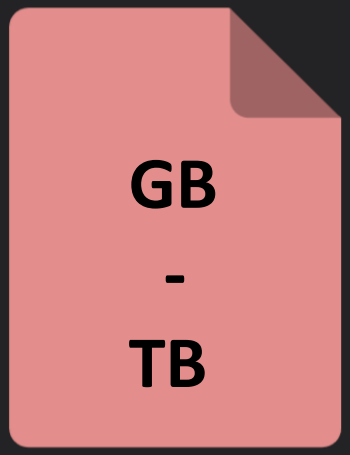
I bias servono a salvare i neuroni, garantendo un’attivazione anche in caso di valori elaborati molto bassi.

Diego Vendramini – Non copiare o riprodurre

I Large Language Model



Base Model

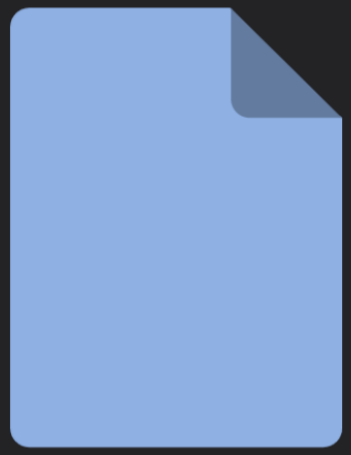


File 1:

Parametri

2 byte x Parametro

(float16)



File 2:

Runner File

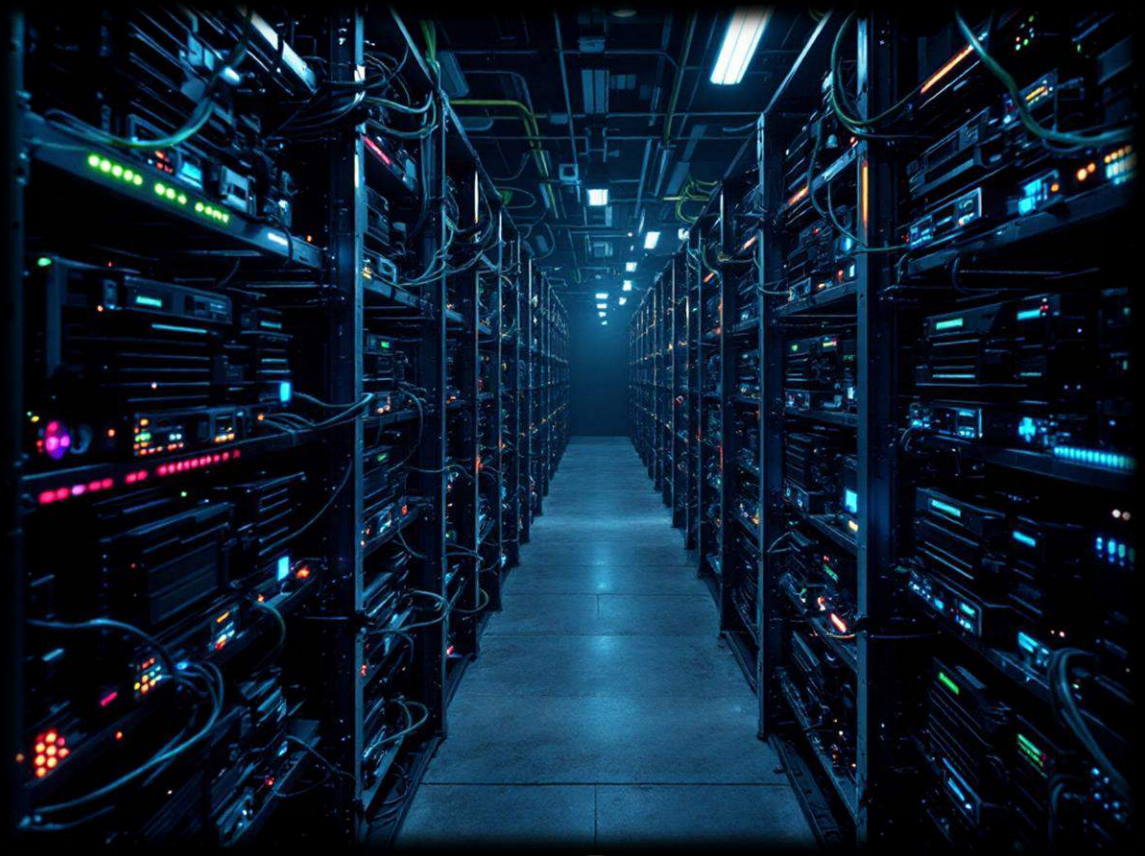
(C, C+, Py…)

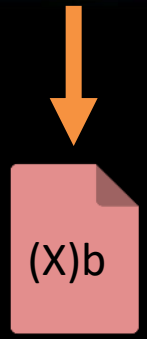
Diego Vendramini – Non copiare o riprodurre

HALF-PRECISION FLOATING-POINT FORMAT
16 BIT

Segno	Esponente	Mantissa
1	5	10
Questo bit indica se il numero è positivo o negativo. **0** = positivo **1** = negativo	Parte esponenziale del numero: la potenza di 2 che lo scala Formula esponente effettivo = valore binario dei 5 bit - 15 (bias)	Parte frazionaria del numero

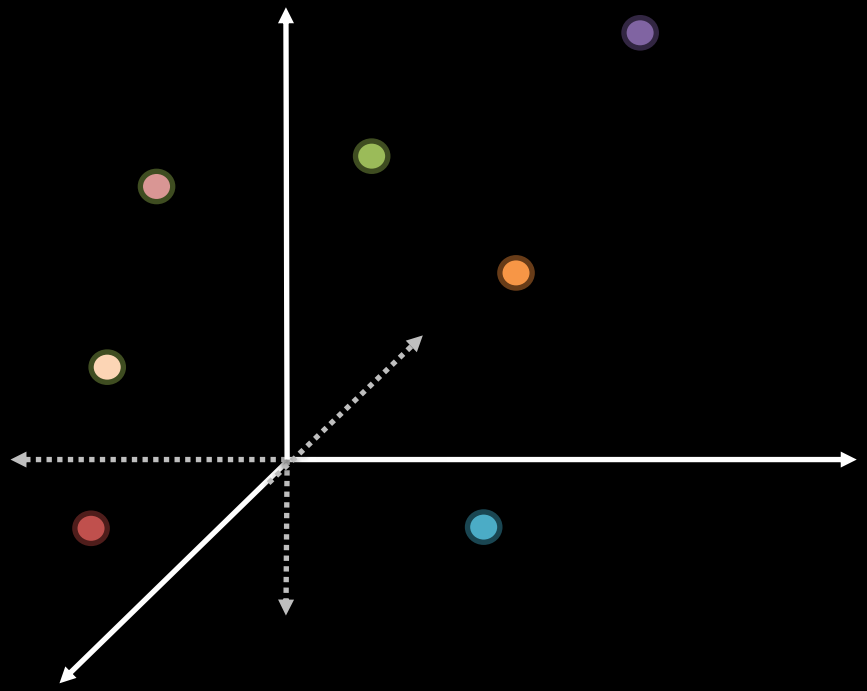
Parametri





L'energia e la tecnologia sono pilastri del progresso umano, con la rivoluzione industriale che ha segnato l'inizio della loro simbiosi. Oggi, la transizione verso fonti rinnovabili è una sfida geopolitica cruciale, con Germania e Cina in prima linea. L'Internet delle Cose (IoT) e l'intelligenza artificiale (IA) ottimizzano l'uso delle risorse energetiche, ma sollevano questioni di sicurezza informatica. La geopolitica dell'energia è influenzata dalle innovazioni tecnologiche, come il fracking, che ha ridisegnato gli equilibri di potere globali. La dipendenza dalle fonti fossili rimane urgente, con la tecnologia che offre soluzioni promettenti per un'economia a basse emissioni di carbonio. La transizione richiede impegno politico e sociale, con accordi internazionali come l'Accordo di Parigi che giocano un ruolo chiave. In conclusione, il legame tra energia e tecnologia è complesso e multidimensionale, determinando il futuro sostenibile del nostro pianeta.

L'energia e la tecnologia sono pilastri del progresso umano, con la rivoluzione industriale che ha segnato l'inizio della loro simbiosi. Oggi, la transizione verso fonti rinnovabili è una sfida geopolitica cruciale, con Germania e Cina in prima linea. L'Internet delle Cose (IoT) e l'intelligenza artificiale (IA) ottimizzano l'uso delle risorse energetiche, ma sollevano questioni di sicurezza informatica. La geopolitica dell'energia è influenzata dalle innovazioni tecnologiche, come il fracking, che ha ridisegnato gli equilibri di potere globali. La dipendenza dalle fonti fossili rimane urgente, con la tecnologia che offre soluzioni promettenti per un'economia a basse emissioni di carbonio. La transizione richiede impegno politico e sociale, con accordi internazionali come l'Accordo di Parigi che giocano un ruolo chiave. In conclusione, il legame tra energia e tecnologia è complesso e multidimensionale, determinando il futuro sostenibile del nostro pianeta.



Si genera un nuovo Output

Criteri di Valutazione delle Risposte AI

1. Utilità

- Linguaggio semplice, conciso e pertinente.
- Tiene conto del contesto e dell'intenzione reale dell'utente.
- In caso di istruzioni confuse, richiede chiarimenti.

2. Veridicità

- Presenta esclusivamente informazioni corrette e affidabili.
- Non inventa fatti o dettagli inesistenti.
- Evita informazioni fuorvianti o ambigue.
- In caso di domande basate su premesse errate, chiarisce l'errore invece di fornire risposte ambigue.

3. Non Nocumento (Innocuità)

- Non produce danni fisici, psicologici, sociali o ambientali.
- Rispetta individui e gruppi, evitando linguaggi offensivi o discriminatori.
- Non suggerisce comportamenti pericolosi o illegali.
- Evita contenuti violenti o sessuali, a meno che non esplicitamente richiesti.

Linee guida nella gestione del trade-off tra criteri

- **Priorità assoluta a Veridicità e Non Nocumento:**
  Nella maggior parte dei casi, essere veritieri e innocui è più importante dell'essere utili.
- **Eccezione – Maggiore utilità:**
  Se un output è molto più utile e solo leggermente meno veritiero o innocuo, e se il compito non è ad alto rischio (es.: richieste finanziarie, consigli medici o legali, ecc.), è possibile privilegiare la maggiore utilità.
- **Gestione dei casi limite:**
  In situazioni dubbie, scegliere l'output che vorresti ricevere personalmente da un assistente che cerca sinceramente di aiutarti.
- **Regola fondamentale finale:**
  La decisione finale deve essere presa applicando il tuo miglior giudizio, considerando sempre le possibili conseguenze nel mondo reale.

Magical thinking is not a religion. It's a different kind of faith -- a faith in the power of positive thoughts and feelings. Yet as unscientific as magical thinking sounds, Pronin said studies have shown there are times when it seems to have a real effect: "There was a study where people in their mid-20s were measured in terms of their optimism," she said, "and then, 50 years later, those who were more optimistic, were actually more likely to still be alive. So it's not always magical to believe that your positive thoughts are having a positive effect."

Magical thinking starts in childhood. At the University of Texas, Professor Jacqui Woolley has examined how children who know the difference between what's real and what's not believe that wishing can cause a penny to appear in what has just been shown to be an empty box. "We find that, by about the age of 4, most of the kids we test seem to really believe that wishing works," said Woolley. "So that would be an example of magical thinking. Nick Barber spent his childhood wishing for riches, focusing on something his father gave him. "I was about 8 or 9 when my dad came home and gave me a fake million dollar bill," he said. "And that became something that represented my goals. I've hung on to it every since.« Barber would even sleep with the bill under his mattress. Now, at 27, he runs a multimillion dollar real estate company in Dallas called UMoveFree, and he still has that million dollar bill in his wallet. How did that bill help him get to this point? "It allowed me to believe in myself at a very young age," Barber said, "not to pay attention to those that said ActiveRain University (ARU) provides free on-line training. We coach, consult and support real estate professionals about real estate trends, technology and social media. A very high achiever who settles for nothing less than the best, Fred has excelled in selling all types of real estate throughout Victoria, from bare land and acreages to new and existing homes, and prime luxury and waterfront properties.He also has expertise locating and selling land for condominium projects in which he's involved from start to finish, as well as new and existing apartment buildings in Victoria, Oak Bay, James Bay, Fairfield, Rockland, Saanich and the West This native Victorian, a rare breed in his field, possesses an unparalleled understanding of his market, which he considers "Canada's Best Kept Secret."Fred maintains a wealth of knowledge, resources and connections that help streamline deals and ensure that the job always gets done right.

Whether you are looking to buy or sell a home in the Greater Victoria area, with Fred on your side you'll always enjoy a win-win situation thanks to his eye for value, thorough service, proven marketing, and no-nonsense negotiating.

It's no wonder that Fred is rewarded with 100% referrals - because he gives back 100% to pleasing all his clients."As my client you'll receive my undivided attention when it comes to providing the level of service you deserve, and tending to each detail so your transaction closes smoothly."

In the course of his career, Fred has served on many committees on the Victoria Real Estate Board. Fred sat on the MLS Committee's sub-committee to develop a NEW Live Open House website for consumers in Victoria and beyond.Fred is also a founding member of the original IC&I Division of the Victoria Real Estate Board, now known as the Commercial Division.

Yes, experience counts, so call the professional who puts his experience to work for you -

Call Fred CarverDirect 250-598-2963. Toll-Free 1-800-663-2121.

Fred is a high achiever and settles for nothing less than selling your home fast and for the right price and believes in a win-win relationship!

Fred is also a past member of the Oak Bay Lions Club and has held many positions, including President.

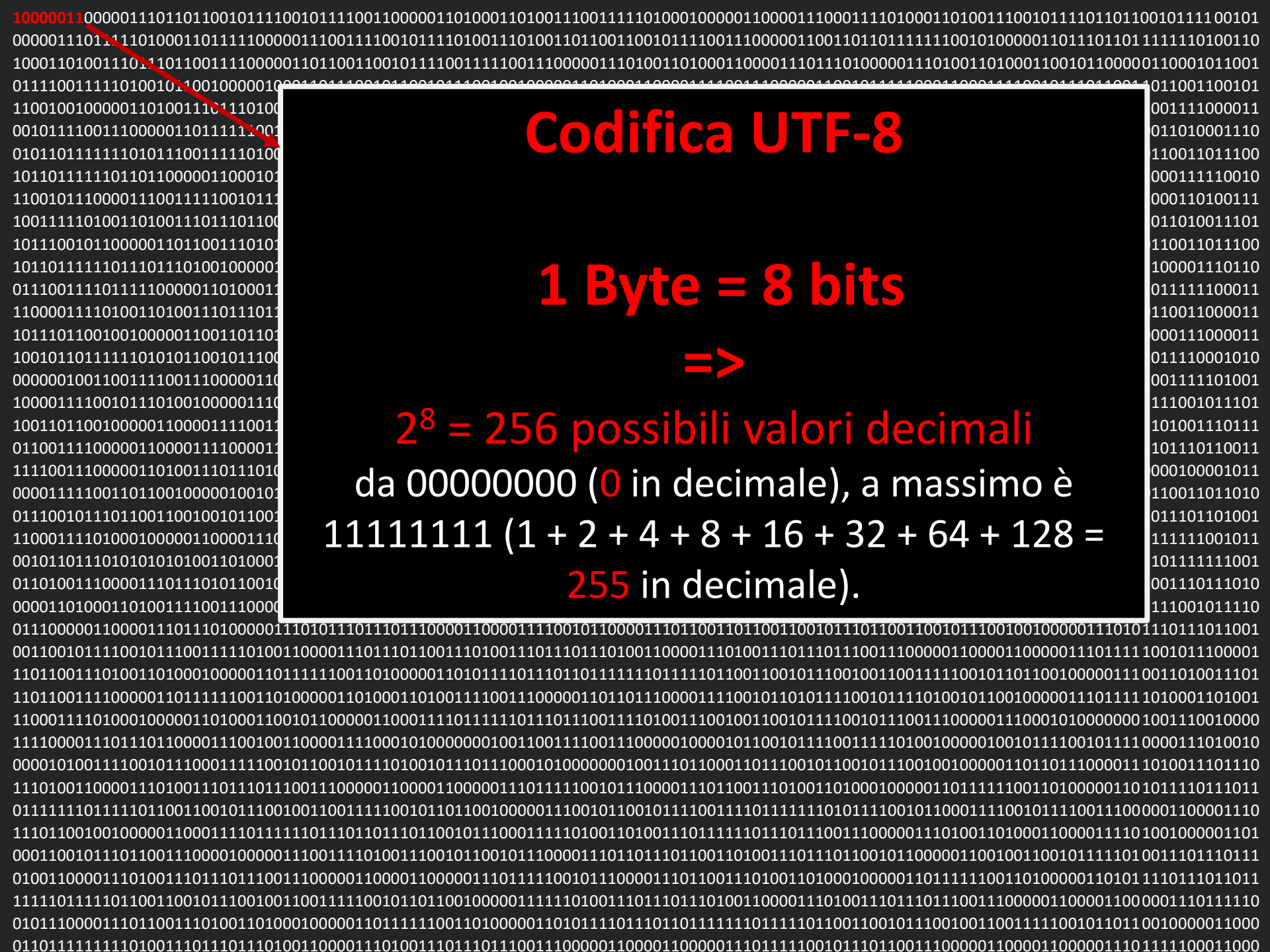
If you're serious about selling your Victoria home or property , meet with Fred. He is a true real estate professional interested in getting your property sold fast and for the most amount of money! Fred has been an award winner many times and uses his vast experience to consult his clients on buying and selling real estate in Victoria. After all these years, Fred does what he has done most of his life...sell real estate successfully, honestly and effectively.

In this rapidly changing world of computers and technology, we have found that positive thinking continuous education, teamwork and adaptability are the keys to providing excellent customer service and success in this decade.

Our vision is to be the Real Estate company of choice for a new generation of professional real estate agents and owners.

Our belief is real estate buying and selling to the public is best served when the real estate industry is viewed as a local service business driven by individual real estate agents...such as a Victoria Realtor like Fred.....

100% Referral... means,A 100% of my time and energies to servicing my clients.While most agents are spending the vast majority of time, energy and money seeking new clients and promoting themselves to the general public, I approach my business differently. You will receive my

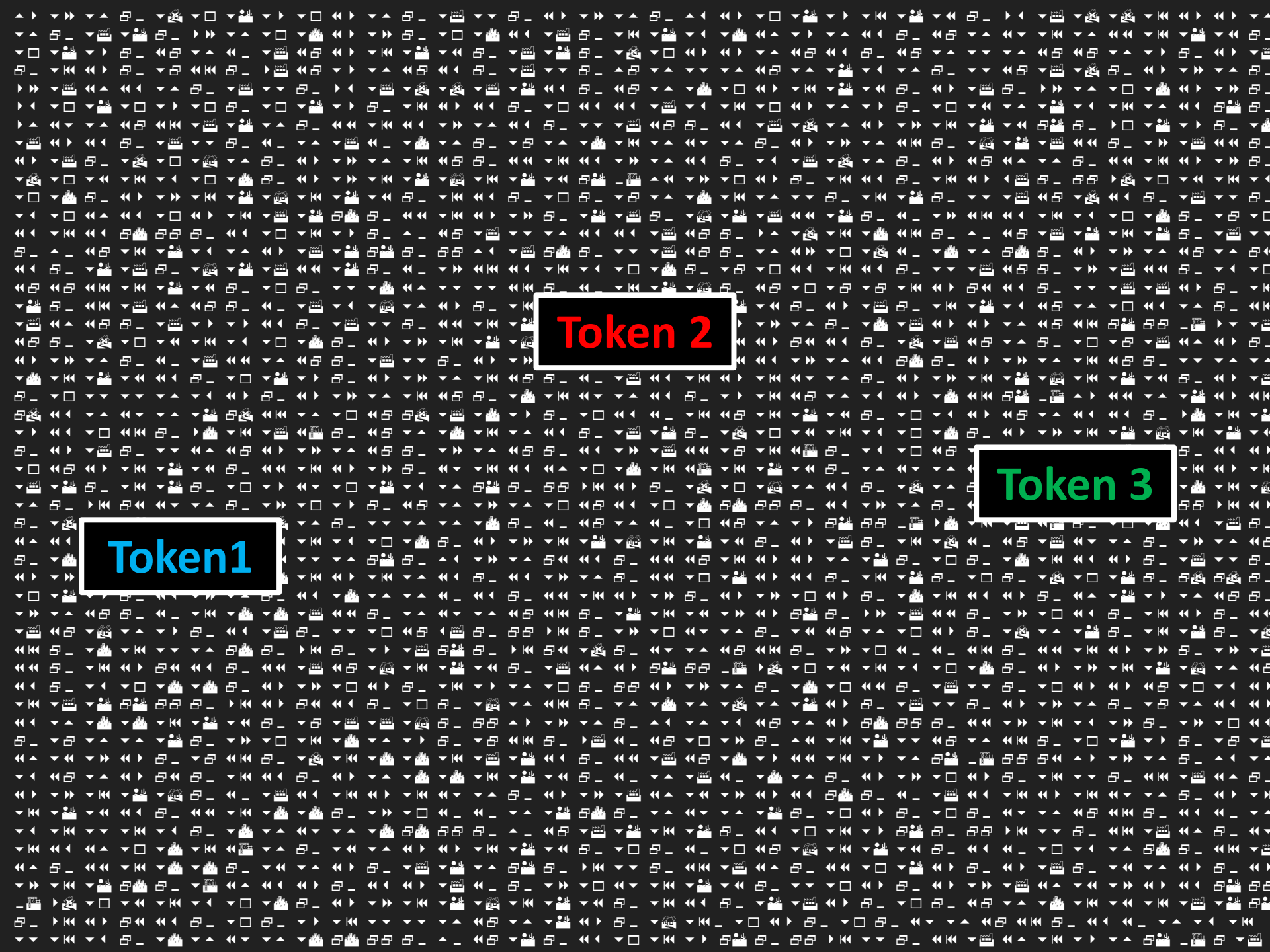

Codifica UTF-8
1 Byte = 8 bits
=>
2^8 = 256 possibili valori decimali
da 00000000 (0 in decimale), a massimo è 11111111 (1 + 2 + 4 + 8 + 16 + 32 + 64 + 128 = 255 in decimale).

108 100 32 97 98 111 117 116 32 97 114 74 101 110 103 111 39 116 32 99 97 114 101 32 67 104 108 101 97 114 84 101 32 97 110 117 116 105 99 32 83 97 116 117 114 100 97 121 44 32 98 117 116 32 110 101 119 32 103 101 110 101 114 97 116 105 111 110 32 105 110 32 116 104 101 32 65 110 100 32 84 104 101 32 46 92 110 83 84 69 70 65 78 79 44 32 83 84 69 70 65 78 79 79 58 32 58 58 104 101 114 111 32 124 83 112 111 105 108 101 100 32 70 119 105 44 32 83 112 111 105 108 101 100 32 101 110 100 32 97 108 116 32 99 111 109 109 101 110 32 121 44 32 121 101 115 32 105 110 100 105 101 100 44 32 100 105 100 32 121 111 117 32 114 101 97 108 108 121 32 100 111 32 116 104 105 115 32 121 111 117 32 103 117 101 115 115 44 32 34 121 101 97 104 44 32 121 101 115 34 32 97 110 100 32 34 104 97 104 34 32 97 108 108 32 104 111 112 32 111 110 32 109 121 32 97 108 97 114 109 32 97 108 32 109 121 32 102 101 101 108 105 110 103 32 99 104 101 99 107 32 116 104 101 32 100 111 111 114 32 116 111 32 109 121 32 104 111 117 115 101 32 115 104 111 117 116 105 110 103 32 97 110 100 32 115 97 121 105 110 103 32 121 111 117 32 119 101 114 101 32 119 101 110 116 32 116 104 101 32 119 114 111 110 103 32 119 97 121 32 97 110 100 32 110 111 116 32 116 111 32 109 101 110 116 105 111 110 32 116 104 101 32 102 97 99 116 32 116 104 97 116 32 105 32 119 97 115 32 97 99 116 117 97 108 108 121 32 115 99 114 101 97 109 105 110 103 32 108 111 108 32 121 111 117 32 121 101 115 32 105 116 32 119 97 115 32 97 99 116 117 97 108 108 121 32 115 99 114 101 97 109 105 110 103 32 100 105 100 32 121 111 117 32 105 110 32 116 104 101 32 98 97 99 107 103 114 111 117 110 100 46 32 70 119 105 44 32 83 112 111 105 108 101 100 32 101 110 100 32 97 108 116 32 248 101 32 36 6 87 116 111 84 144 42 17 240 58 110 161 90 137 69 37 234 97 116 237 148 2 98 185 102 56 119 117 72 111 253 87 145 43 108 161 12 97 98 116 32 101 105 105 172 32 237 148 110 23 86 5 32 87 230 73 173 95 114 110 101 233 67 196 107 32 121 32 119 120 118 103 198 116 9 29 98 244 110 111 108 101 15 247 4 32 97 187 103 100 100 32 233 44 141 160 229 113 119 39 110 74 88 97 101 117 101 198 110 100 32 105 78 34 215 44 3 137 84 155 51 97 101 50 111 108 124 198 103 32 110 32 116 110 32 227 203 203 104 101 141 135 139 105 83 105 32 111 114 101 116 51 90 184 101 247 97 158 99 105 97 114 32 99 212 46 111 114 103 110 45 99 32 49 75 114 117 55 83 114 125 101 211 32 116 196 104 115 70 110 97 101 116 32 53 117 154 17 104 255 206 172 101 99 99 111 109 109 101 110 32 121 44 32 121 101 115 32 105 110 100 105 101 100 44 32 100 105 100 32 121 111 117 32 114 101 97 108 108 121 32 100 111 32 116 104 105 115 32 121 111 117 32 103 32 70 119 105 44 32 83 112 111 105 108 101 100 32 101 110 100 32 97 108 116 32 99 111 109 109 101 110 32 121 44 32 121 101 115 32 105 110 100 105 101 100 44 32 100 105 100 32 121 111 117 32 114 101 97 108 108 121 32 100 111 32 116 104 105 115 32 121 111 117 32 103 124 86 105 101 119 105 110 103 32 83 105 110 103 101 32 87 111 101 32 111 110 32 116 104 101 32 87 101 98 32 105 110 32 116 104 101 32 119 111 114 108 100 32 97 98 111 117 116 32 97 114 74 101 110 103 111 39 116 32 99 97 114 101 32 67 121 32 100 111 32 116 104 105 115 32 121 111 117 32 103 117 101 115 115 44 32 34 121 101 97 104 44 32 121 101 115 34 32 97 110 100 32 34 104 97 104 34 32 97 108 108 32 104 111 112 32 111 110 32 109 121 32 97 108 97 114 109 32 97 108 32 109 121 32 102 101 101 108 105 110 103 32 99 104 101 99 107 32 116 104 101 32 100 111 111 114 32 116 111 32 109 121 32 104 111 117 115 101 32 115 104 111 117 116 105 110 103 32 97 110 100 32 115 97 121 105 110 103 32 121 111 117 32 119 101 114 101 32 119 101 110 116 32 116 104 101 32 119 114 111 110 103 32 119 97 121 32 97 110 100 32 110 111 116 32 116 111 32 109 101 110 116 105 111 110 32 116 104 101 32 102 97 99 116 32 116 104 97 116 32 105 32 119 97 115 32 97 99 116 117 97 108 108 121 32 115 99 114 101 97 109 105 110 103 32 108 111 108 32 121 111 117 32 121 101 115 32 105 116 32 119 97 115 32 97 99 116 117 97 108 108 121 32 115 99 114 101 97 109 105 110 103 32 100 105 100 32 121 111 117 32 105 110 32 116 104 101 32 98 97 99 107 103 114 111 117 110 100 46 32 70 119 105 44 32 83 112 111 105 108 101 100 32 101 110 100 32 97 108 116 32 99 111 109 109 101 110 32 121 44 32 121 101 115 32 105 110 100 105 101 100 44 32 100 105 100 32 121 111 117 32 114 101 97 108 108 121 32 100 111 32 116 104 105 115 32 121 111 117 32 103 32 70 119 105 44 32 83 112 111 105 108 101 100 32 101 110 100 32 97 108 116 32 99 111 109 109 101 110 32 121 44 32 121 101 115 32 105 110 100 105 101 100 44 32 100 105 100 32 121 111 117 32 114 101 97 108 108 121 32 100 111 32 116 104 105 115 32 121 111 117 32 103 117 101 115 115 44 32 34 121 101 97 104 44 32 121 101 115 34 32 97 110 100 32 34 104 97 104 34 32 97 108 108 32 104 111 112 32 111 110 32 109 121 32 97 108 97 114 109 32 97 108 32 109 121 32 102 101 101 108 105 110 103 32 99 104 101 99 107 32 116 104 101 32 100 111 111 114 32 116 111 32 109 121 32 104 111 117 115 101 32 115 104 111 117 116 105 110 103 32 97 110 100 32 115 97 121 105 110 103 32 121 111 117 32 119 101 114 101 32 119 101 110 116 32 116 104 101 32 119 114 111 110 103 32 119 97 121 32 97 110 100 32 110 111 116 32 116 111 32 109 101 110 116 105 111 110 32 116 104 101 32 102 97 99 116 32 116 104 97 116 32 105 32 119 97 115 32 97 99 116 117 97 108 108 121 32 115 99 114 101 97 109 105 110 103 32 108 111 108 32 121 111 117 32 121 101 115 32 105 116 32 119 97 115 32 97 99 116 117 97 108 108 121 32 115 99 114 101 97 109 105 110 103 32 100 105 100 32 121 111 117 32 105 110 32 116 104 101 32 98 97 99 107 103 114 111 117 110 100 46 32 70 119 105 44 32 83 112 111 105 108 101 100 32 101 110 100 32 97 108 116 32 99 111 109 109 101 110 32 121 44 32 121 101 115 32 105 110 100 105 101 100 44 32 100 105 100 32 121 111 117 32 114 101 97 108 108 121 32 100 111 32 116 104 105 115 32 121 111 117 32 103 32 70 119 105 44 32 83 112 111 105 108 101 100 32 101 110 100 32 97 108 116 32 99 111 109 109 101 110 32 121 44 32 121 101 115 32 105 110 100 105 101 100 44 32 100 105

108 100 32 97 98 111 117 116 32 97 114 74 101 110 103 111 39 116 32 99 97 114 101 32 67 104 108 101 97 114 84 101 32 97 110 117 116 105 99 32 83 97
116 117 114 100 97 121 44 32 98 117 116 32 110 101 119 32 103 101 110 101 114 97 116 105 111 110 32 105 110 32 116 104 101 32 65 110 100 32 84 104
101 32 46 92 110 83 84 69 70 65 78 79 44 32 83 84 69 70 65 78 79 79 58 32 58 58 104 101 114 111 32 124 83 112 111 105 108 101 100 32 70 119 105 44 32
83 112 111 105 108 101 100 32 101 110 100 32 97 108 116 32 99 111 109 109 101 110 32 121 44 32 121 101 115 32 105 110 100 105 101 100 44 32 100 105
100 32 121 111 117 32 114 101 97 108 108 121 32 100 111 32 116 104 105 115 32 121 111 117 32 103 117 101 115 115 44 32 34 121 101 97 104 44 32 121
101 115 34 32 97 110 100 32 34 104 97 104 34 32 97 108 108 32 104 111 112 32 111 110 32 109 121 32 97 108 97 114 109 32 97 108 32 109 121 32 102 101
101 108 105 110 103 32 99 104 101 99 107 32 116 104 101 32 100 111 111 114 32 116 111 32 109 121 32 104 111 117 115 101 32 115 104 111 117 116 105
110 103 32 97 110 100 32 115 97 121 105 110 103 32 121 111 117 32 119 101 114 101 32 119 101 110 116 32 116 104 101 32 119 114 111 110 103 32 119 97
121 32 97 110 100 32 110 111 116 32 116 111 32 109 101 110 116 105 111 110 32 116 104 101 32 102 97 99 116 32 116 104 97 116 32 105 32 119 97 115 32
97 99 116 117 97 108 108 121 32 115 99 114 101 97 109 105 110 103 32 108 111 108 32 121 111 117 32 121 101 115 32 105 116 32 119 97 115 32 97 99 116
117 97 108 108 121 32 115 99 114 101 97 109 105 110 103 32 100 105 100 32 121 111 117 32 105 110 32 116 104 101 32 98 97 99 107 103 114 111 117 110
100 46 32 70 119 105 44 32 83 112 111 105 108 101 100 32 101 110 100 32 97 108 116 32 248 101 32 36 6 87 116 111 84 144 42 17 240 58 110 161 90 137
69 37 234 97 116 237 148 2 98 185 102 56 119 117 72 111 253 87 145 43 108 161 12 97 98 116 32 101 105 105 172 32 237 148 110 23 86 5 32 87 230 73 173
95 114 110 101 233 67 196 107 32 121 32 119 120 118 103 198 116 9 29 98 244 110 111 108 101 15 247 4 32 97 187 103 100 100 32 233 44 141 160 229 113
119 39 110 74 88 97 101 117 101 198 110 100 32 105 78 34 215 44 3 137 84 155 51 97 101 50 111 108 124 198 103 32 110 32 116 110 32 227 203 203 104
101 141 135 139 105 83 105 32 111 114 101 116 51 90 184 101 247 97 158 99 105 97 114 32 99 212 46 111 114 103 110 45 99 32 49 75 114 117 55 83 114
125 101 211 32 116 196 104 115 70 110 97 101 116 32 53 117 154 17 104 255 206 172 101 99 99 111 109 109 101 110 32 121 44 32 121 101 115 32 105 110
100 105 101 100 44 32 100 105 100 32 121 111 117 32 114 101 97 108 108 121 32 100 111 32 116 104 105 115 32 121 111 117 32 103 32 70 119 105 44 32
83 112 111 105 108 101 100 32 101 110 100 32 97 108 116 32 99 111 109 109 101 110 32 121 44 32 121 101 115 32 105 110 100 105 101 100 44 32 100 105
100 32 121 111 117 32 114 101 97 108 108 121 32 100 111 32 116 104 105 115 32 121 111 117 32 103 124 86 105 101 119 105 110 103 32 83 105 110 103
101 32 87 111 101 32 111 110 32 116 104 101 32 87 101 98 32 105 110 32 116 104 101 32 119 111 114 108 100 32 97 98 111 117 116 32 97 114 74 101 110
103 111 39 116 32 99 97 114 101 32 67 121 32 100 111 32 116 104 105 115 32 121 111 117 32 103 117 101 115 115 44 32 34 121 101 97 104 44 32 121 101
115 34 32 97 110 100 32 34 104 97 104 34 32 97 108 108 32 104 111 112 32 111 110 32 109 121 32 97 108 97 114 109 32 97 108 32 109 121 32 102 101 101
108 105 110 103 32 99 104 101 99 107 32 116 104 101 32 100 111 111 114 32 116 111 32 109 121 32 104 111 117 115 101 32 115 104 111 117 116 105 110
103 32 97 110 100 32 115 97 121 105 110 103 32 121 111 117 32 119 101 114 101 32 119 101 110 116 32 116 104 101 32 119 114 111 110 103 32 119 97 121
32 97 110 100 32 110 111 116 32 116 111 32 109 101 110 116 105 111 110 32 116 104 101 32 102 97 99 116 32 116 104 97 116 32 105 32 119 97 115 32 97
99 116 117 97 108 108 121 32 115 99 114 101 97 109 105 110 103 32 108 111 108 32 121 111 117 32 97 108 116 32 99 111 109 109 101 110 32 121 44 32
121 101 115 32 105 110 100 105 101 100 44 32 100 105 100 32 121 111 117 32 114 101 97 108 108 121 32 100 111 32 116 104 105 115 32 121 111 117 32
103 117 101 115 115 44 32 34 121 101 97 104 44 32 121 101 115 34 32 97 110 100 32 34 104 97 104 34 32 97 108 108 32 104 111 112 32 111 110 32 109 121
32 97 108 97 114 109 32 97 108 32 109 121 32 102 101 101 108 105 110 103 32 99 104 101 99 107 32 116 104 101 32 100 111 111 114 32 116 111 32 109
121 32 104 111 117 115 101 32 115 104 111 117 116 105 110 103 32 97 110 100 32 115 97 121 105 110 103 32 121 111 117 32 119 101 114 101 32 119 101
110 116 32 116 104 101 32 119 114 111 110 103 32 119 97 121 32 97 110 100 32 110 111 116 32 116 111 32 109 101 110 116 105 111 110 32 116 104 101 32
102 97 99 116 32 116 104 97 116 32 105 32 119 97 115 32 97 99 116 117 97 108 108 121 32 115 99 114 101 97 109 105 110 103 32 108 111 108 32 121 111
117 32 121 101 115 32 111 114 32 116 111 32 109 121 32 104 111 117 115 101 32 115 104 111 117 116 105 110 103 32 97 110 100 32 115 97 121 105 110
103 32 121 111 117 32 119 101 114 101 32 119 101 110 116 32 116 104 101 32 119 114 111 110 103 32 119 97 121 32 97 110 100 32 110 111 116 32 116 111
32 109 101 110 116 105 111 110 32 116 104 101 32 102 97 99 116 32 116 104 97 116 32 105 32 119 97 115 32 97 99 116 117 97 108 108 121 32 115 99 114
101 97 109 105 110 103 32 108 111 108 32 121 111 117 32 121 101 115 32 105 116 32 119 97 115 32 97 99 116 117 97 108 108 121 32 115 99 114 101 97
109 105 110 103 32 100 105 100 32 121 111 117 32 105 110 32 116 104 101 32 98 97 99 107 103 114 111 117 110 100 46 32 70 119 105 44 32 83 112 111
105 108 101 100 32 101 110 100 32 97 108 116 32 99 111 109 109 101 110 32 121 44 32 121 101 115 32 105 110 100 105 101 100 44 32 100 105 100 32 121
111 117 32 114 101 97 108 108 121 32 100 111 32 116 104 105 115 32 121 111 117 32 103 32 70 119 105 44 32 83 112 111 105 108 101 100 32 101 110 100
32 97 108 116 32 99 111 109 109 101 110 32 121 44 32 32 121 101 115 32 105 116 32 119 97 115 32 97 99
116 117 97 108 108 121 32 115 99 114 101 97 109 105 110 103 32 100 105 100 32 121 111 117 32 105 110 32 116 104 101 32 98 97 99 107 103 114 111 117

108 100 32 97 98 111 117 116 32 97 114 74 101 110 103 111 39 116 32 99 97 114 101 32 67 104 108 101 97 114 84 101 32 97 110 117 116 105 99 32 83 97
116 117 114 100 97 121 44 32 98 117 116 32 110 101 119 32 103 101 110 101 114 97 116 105 111 110 32 105 110 32 116 104 101 32 65 110 100 32 84 104
101 32 46 92 110 83 84 69 70 65 78 79 44 32 83 84 69 70 65 78 79 79 58 32 58 58 104 101 114 111 32 124 83 112 111 105 108 101 100 32 70 119 105 44 32
83 112 111 105 108 101 100 32 101 110 100 32 97 108 116 32 99 111 109 109 101 110 32 121 44 32 121 101 115 32 105 110 100 105 101 100 44 32 100 105
100 32 121 111 117 32 114 101 97 108 108 121 32 100 111 32 116 104 105 115 32 121 111 117 32 103 117 101 115 115 44 32 34 121 101 97 104 44 32 121
101 115 34 32 97 110 100 32 34 104 97 104 34 32 97 108 108 32 104 111 112 32 111 110 32 109 121 32 97 108 97 114 109 32 97 108 32 109 121 32 102 101
101 108 105 110 103 32 99 104 101 99 107 32 116 104 101 32 100 111 111 114 32 116 111 32 109 121 32 104 111 117 115 101 32 115 104 111 117 116 105
110 103 32 97 110 100 32 115 97 121 105 110 103 32 121 111 117 32 119 101 114 101 32 119 101 110 116 32 116 104 101 32 119 114 111 110 103 32 119 97
121 32 97 110 100 32 110 111 116 32 116 111 32 109 101 110 116 105 111 110 32 116 104 101 32 102 97 99 116 32 116 104 97 116 32 105 32 119 97 115 32
97 99 116 117 97 108 108 121 32 115 99 114 101 97 109 105 110 103 32 108 111 108 32 121 111 117 32 121 101 115 32 105 116 32 119 97 115 32 97 99 116
117 97 108 108 121 32 115 99 114 101 97 109 105 110 103 32 100 105 100 32 121 111 117 32 105 110 32 116 104 101 32 98 97 99 107 103 114 111 117 110
100 46 32 70 119 105 44 32 83 112 111 105 108 101 100 32 101 110 100 32 97 108 116 32 248 101 32 36 6 87 116 111 84 144 42 17 240 58 110 161 90 137
69 37 234 97 116 237 148 2 98 185 102 56 119 117 72 111 253 87 145 43 108 161 12 97 98 116 32 101 105 105 172 32 237 148 110 23 86 5 32 87 230 73 173
95 114 110 101 233 67 196 107 32 121 32 119 120 118 103 198 116 9 29 98 244 110 111 108 101 15 247 4 32 97 187 103 100 100 32 233 44 141 160 229 113
119 39 110 74 88 97 101 117 101 198 110 100 32 105 78 34 215 44 3 137 84 155 51 97 101 50 111 108 124 198 103 32 110 32 116 110 32 227 203 203 104
101 141 135 139 105 83 105 32 111 114 101 116 51 90 184 101 247 97 158 99 105 97 114 32 99 212 46 111 114 103 110 45 99 32 49 75 114 117 55 83 114
125 101 211 32 116 196 104 115 70 110 97 101 116 32 53 117 154 17 104 255 206 172 101 99 99 111 109 109 101 110 32 121 44 32 121 101 115 32 105 110
100 105 101 100 44 32 100 105 100 32 121 111 117 32 114 101 97 108 108 121 32 100 111 32 116 104 105 115 32 121 111 117 32 103 32 70 119 105 44 32
83 112 111 105 108 101 100 32 101 110 100 32 97 108 116 32 99 111 109 109 101 110 32 121 44 32 121 101 115 32 105 110 100 105 101 100 44 32 100 105
100 32 121 111 117 32 114 101 97 108 108 121 32 100 111 32 116 104 105 115 32 121 111 117 32 103 124 86 105 101 119 105 110 103 32 83 105 110 103
101 32 87 111 101 32 111 110 32 116 104 101 32 87 101 98 32 105 110 32 116 104 101 32 119 111 114 108 100 32 97 98 111 117 116 32 97 114 74 101 110
103 111 39 116 32 99 97 114 101 32 67 121 32 100 111 32 116 104 105 115 32 121 111 117 32 103 117 101 115 115 44 32 34 121 101 97 104 44 32 121 101
115 34 32 97 110 100 32 34 104 97 104 34 32 97 108 108 32 104 111 112 32 111 110 32 109 121 32 97 108 97 114 109 32 97 108 32 109 121 32 102 101 101
108 105 110 103 32 99 104 101 99 107 32 116 104 101 32 100 111 111 114 32 116 111 32 109 121 32 104 111 117 115 101 32 115 104 111 117 116 105 110
103 32 97 110 100 32 115 97 121 105 110 103 32 121 111 117 32 119 101 114 101 32 119 101 110 116 32 116 104 101 32 119 114 111 110 103 32 119 97 121
32 97 110 100 32 110 111 116 32 116 111 32 109 101 110 116 105 111 110 32 116 104 101 32 102 97 99 116 32 116 104 97 116 32 105 32 119 97 115 32 97
99 116 117 97 108 108 121 32 115 99 114 101 97 109 105 110 103 32 108 111 108 32 121 111 117 32 97 108 116 32 99 111 109 109 101 110 32 121 44 32
121 101 115 32 105 110 100 105 101 100 44 32 100 105 100 32 121 111 117 32 114 101 97 108 108 121 32 100 111 32 116 104 105 115 32 121 111 117 32
103 117 101 115 115 44 32 34 121 101 97 104 44 32 121 101 115 34 32 97 110 100 32 34 104 97 104 34 32 97 108 108 32 104 111 112 32 111 110 32 109 121
32 97 108 97 114 109 32 97 108 32 109 121 32 102 101 101 108 105 110 103 32 99 104 101 99 107 32 116 104 101 32 100 111 111 114 32 116 111 32 109
121 32 104 111 117 115 101 32 115 104 111 117 116 105 110 103 32 97 110 100 32 115 97 121 105 110 103 32 121 111 117 32 119 101 114 101 32 119 101
110 116 32 116 104 101 32 119 114 111 110 103 32 119 97 121 32 97 110 100 32 110 111 116 32 116 111 32 109 101 110 116 105 111 110 32 116 104 101 32
102 97 99 116 32 116 104 97 116 32 105 32 119 97 115 32 97 99 116 117 97 108 108 121 32 115 99 114 101 97 109 105 110 103 32 108 111 108 32 121 111
117 32 121 101 115 32 111 114 32 116 111 32 109 121 32 104 111 117 115 101 32 115 104 111 117 116 105 110 103 32 97 110 100 32 115 97 121 105 110
103 32 121 111 117 32 119 101 114 101 32 119 101 110 116 32 116 104 101 32 119 114 111 110 103 32 119 97 121 32 97 110 100 32 110 111 116 32 116 111
32 109 101 110 116 105 111 110 32 116 104 101 32 102 97 99 116 32 116 104 97 116 32 105 32 119 97 115 32 97 99 116 117 97 108 108 121 32 115 99 114
101 97 109 105 110 103 32 108 111 108 32 121 111 117 32 121 101 115 32 105 116 32 119 97 115 32 97 99 116 117 97 108 108 121 32 115 99 114 101 97
109 105 110 103 32 100 105 100 32 121 111 117 32 105 110 32 116 104 101 32 98 97 99 107 103 114 111 117 110 100 46 32 70 119 105 44 32 83 112 111
105 108 101 100 32 101 110 100 32 97 108 116 32 99 111 109 109 101 110 32 121 44 32 121 101 115 32 105 110 100 105 101 100 44 32 100 105 100 32 121
111 117 32 114 101 97 108 108 121 32 100 111 32 116 104 105 115 32 121 111 117 32 103 32 70 119 105 44 32 83 112 111 105 108 101 100 32 101 110 100
32 97 108 116 32 99 111 109 109 101 110 32 121 44 32 121 101 115 32 105 110 100 105 101 100 44 32 32 121 101 115 32 105 116 32 119 97 115 32 97 99
116 117 97 108 108 121 32 115 99 114 101 97 109 105 110 103 32 100 105 100 32 121 111 117 32 105 110 32 116 104 101 32 98 97 99 107 103 114 111 117

108 100 32 97 98 111 117 116 32 97 114 74 101 110 103 111 39 116 32 99 97 114 101 32 67 104 108 101 97 114 84 101 32 97 110 117 116 105 99 32 83 97
116 117 114 100 97 121 44 32 98 117 116 32 110 101 119 32 103 101 110 101 114 97 116 105 111 110 32 105 110 32 116 104 101 32 65 110 100 32 84 104
101 32 46 92 110 83 84 69 70 65 78 79 44 32 83 84 69 70 65 78 79 79 58 32 58 58 104 101 114 111 32 124 83 112 111 105 108 101 100 32 70 119 105 44 32
83 112 111 105 108 101 100 32 101 110 100 32 97 108 116 32 99 111 109 109 101 110 32 121 44 32 121 101 115 32 105 110 100 105 101 100 44 32 100 105
100 32 121 111 117 32 114 101 97 108 108 121 32 100 111 32 116 104 105 115 32 121 111 117 32 103 117 101 115 115 44 32 34 121 101 97 104 44 32 121
101 115 34 32 97 110 100 32 34 104 97 104 34 32 97 108 108 32 104 111 112 32 111 110 32 109 121 32 97 108 97 114 109 32 97 108 32 109 121 32 102 101
101 108 105 110 103 32 99 104 101 99 107 32 116 104 101 32 100 111 111 114 32 116 111 32 109 121 32 104 111 117 115 101 32 115 104 111 117 116 105
110 103 32 97 110 100 32 115 97 121 105 110 103 32 121 111 117 32 119 101 110 116 32 116 104 101 32 119 114 111 110 103 32 119 97
121 32 97 110 100 32 110 111 116 32 116 111 32 109 101 6 104 101 32 102 97 99 116 32 116 104 97 116 32 105 32 119 97 115 32
97 99 116 117 97 108 108 121 32 115 99 114 101 97 109 10 32 121 111 117 32 121 101 115 32 105 116 32 119 97 115 32 97 99 116
117 97 108 108 121 32 115 99 114 101 97 109 105 110 103 1 117 32 105 110 32 116 104 101 32 98 97 99 107 103 114 111 117 110
100 46 32 70 119 105 44 32 83 112 111 111 108 101 100 32 101 110 100 32 97 108 116 32 248 101 32 36 6 87 116 111 84 144 42 17 240 58 110 161 90 137
69 37 234 97 116 237 148 2 98 185 102 56 119 117 72 111 253 87 145 43 108 161 12 97 98 116 32 101 105 105 172 32 237 148 110 23 86 5 32 87 230 73 173
95 114 110 101 233 67 196 107 32 121 32 119 120 118 103 198 116 9 29 98 244 110 111 108 101 15 247 4 32 97 187 103 100 100 32 233 44 141 160 229 113
119 98 110 100 32 105 78 34 215 44 3 137 84 155 51 97 101 50 111 108 124 198 03 203 104
101 1 114 101 116 51 90 184 101 247 97 158 99 105 97 114 32 99 212 46 111 114 7 55 83 114
125 0 110 97 101 116 32 53 117 154 17 104 255 206 172 101 99 99 111 109 109 101 32 105 110
100 0 32 121 111 117 32 114 101 97 108 108 121 32 100 111 32 116 104 105 115 3 9 105 44 32
83 112 111 105 108 101 100 32 101 110 100 32 97 108 116 32 99 111 109 109 101 110 32 121 44 32 121 101 115 32 105 110 100 105 101 100 44 32 100 105
100 32 121 111 117 32 114 101 97 108 108 121 32 100 111 32 116 104 105 115 32 121 111 117 32 103 124 86 105 101 119 105 110 103 32 83 105 110 103
101 32 87 111 101 32 111 110 32 116 104 101 32 87 101 98 32 105 110 32 116 104 101 32 119 111 114 108 100 32 97 98 111 117 116 32 97 114 74 101 110
103 111 39 116 32 99 97 114 101 32 67 121 32 100 111 32 116 104 105 115 32 121 111 117 32 103 117 101 115 115 44 32 34 121 101 97 104 44 32 121 101
115 34 32 97 110 100 32 34 104 97 104 34 32 97 108 108 32 104 111 112 32 111 110 32 109 121 32 97 108 97 114 109 32 97 108 32 109 121 32 102 101 101
108 105 110 103 32 99 104 101 99 111 111 32 109 121 32 104 111 117 115 101 32 115 104 111 117 116 105 110
103 32 97 110 100 32 115 97 121 2 119 101 110 116 32 116 104 101 32 119 114 111 110 103 32 119 97 121
32 97 110 100 32 110 111 116 32 116 111 32 109 101 110 116 105 111 110 32 116 104 101 32 102 97 99 116 32 116 104 97 116 32 105 32 119 97 115 32 97
99 116 117 97 108 108 121 32 115 99 114 101 97 109 105 110 103 32 108 111 108 32 121 44 32
121 101 115 32 1 100 44 32 100 105 100 32 121 111 117 32 114 101 97 108 108 121 32 100 111 32 116 104 105 115 32 121 111 117 32
103 117 101 115 01 97 104 44 32 121 101 115 34 32 97 110 100 32 34 104 97 104 34 32 97 108 108 32 104 111 112 32 111 110 32 109 121
32 97 108 97 114 109 32 97 108 32 109 121 32 102 101 101 108 105 110 103 32 99 104 101 99 107 32 116 104 101 32 100 111 111 114 32 116 111 32 109
121 32 104 111 117 115 101 32 115 104 111 117 116 105 110 103 32 97 110 100 32 115 97 121 32 119 101 114 101 32 119 114 111 110 103 32 119 101
110 116 32 116 104 101 32 119 114 111 110 103 32 119 97 121 32 97 110 100 32 110 111 116 32 116 111 32 109 101 110 116 105 111 110 32 116 104 101
102 97 99 116 32 116 104 97 116 32 105 32 119 97 115 32 97 99 116 117 97 108 108 121 32 115 99 114 101 97 109 105 110 103 32 108 111 108 32 121 111
117 32 121 101 115 32 111 114 32 116 111 32 109 121 32 104 111 117 115 101 32 115 104 111 117 116 105 110 103 32 97 110 100 32 115 97 121 105 110
103 32 121 111 117 32 119 101 114 101 119 101 110 116 32 116 104 101 32 119 114 111 110 103 32 119 97 121 32 97 110 100 32 110 111 116 32 116 111 111
32 109 101 110 116 105 111 110 32 116 104 101 102 97 99 116 32 116 104 97 116 32 6 117 97 108 108 121 32 115 99 114 101 97 109 105 110 103 32 108
101 97 109 105 110 103 32 108 111 108 32 121 111 117 32 121 101 115 32 105 116 32 119 97 115 32 97 99 116 117 97 108 108 121 32 115 99 114 101 97
109 105 110 103 32 100 105 100 32 121 111 117 32 105 110 32 116 104 101 32 98 97 99 107 103 114 111 117 110 46 32 70 119 105 44 32 83 112 111
105 108 101 100 32 101 110 100 32 97 108 116 32 99 111 109 109 101 110 32 121 44 32 121 101 115 32 105 110 100 105 101 100 44 32 100 105 100 32 121
111 117 32 114 101 97 108 108 121 32 100 111 32 116 104 105 115 32 121 111 117 32 103 32 70 119 105 44 32 83 112 111 108 101 100 32 101 110 100
32 97 108 116 32 99 111 109 109 101 110 32 121 44 32 121 101 115 32 105 110 100 105 101 100 44 32 32 121 101 115 32 105 116 32 119 97 115 32 97 99
116 117 97 108 108 121 32 115 99 114 101 97 109 105 110 103 32 100 105 100 32 121 111 117 32 114 101 97 108 108 121 32 100 111 32 116 104 105 115

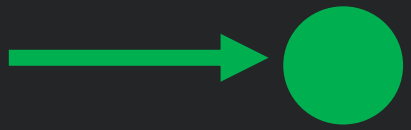

Token 2
Token 3
Token1

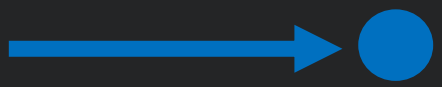
Magical thinking is not a religion. It's a different kind of faith -- a faith in the power of positive thoughts and feelings. Yet as unscientific as magical thinking sounds, Pronin said studies have shown there are times when it seems to have a real effect: "There was a study where people in their mid-20s were measured in terms of their optimism," she said, "and then, 50 years later, those who were more optimistic, were actually more likely to still be alive. So it's not always magical to believe that your positive thoughts are having a positive effect."

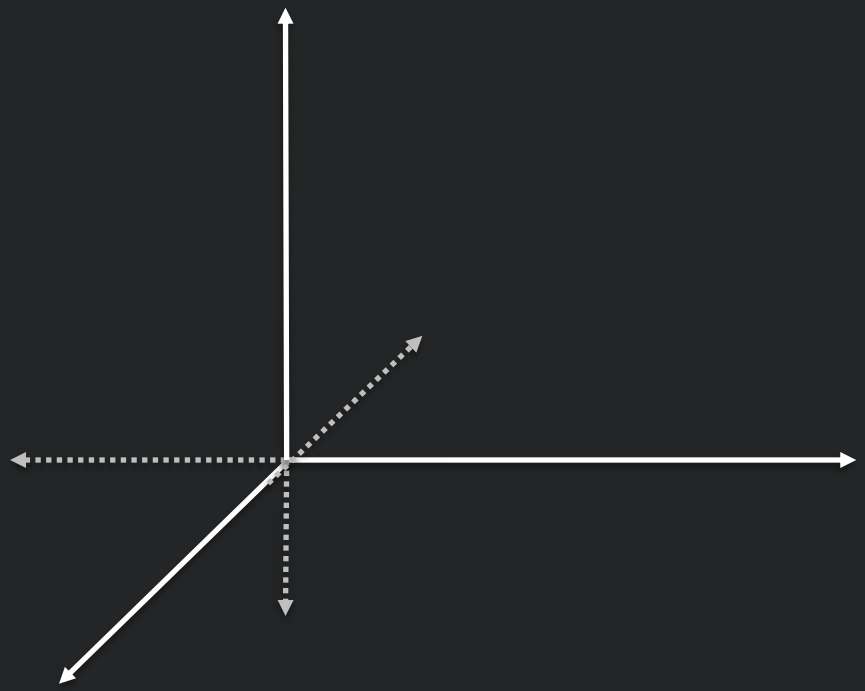
Magical thinking starts in childhood. At the University of Texas, Professor Jacqui Woolley has examined how children who know the difference between what's real and what's not believe that wishing can cause a penny to appear in what has just been shown to be an empty box. "We find that, by about the age of 4, most of the kids we test seem to really believe that wishing works," said Woolley. "So that would be an example of magical thinking. Nick Barber spent his childhood wishing for riches, focusing on something his father gave him. "I was about 8 or 9 when my dad came home and gave me a fake million dollar bill," he said. "And that became something that represented my goals. I've hung on to it every since.« Barber would even sleep with the bill under his mattress. Now, at 27, he runs a multimillion dollar real estate company in Dallas called UMoveFree, and he still has that million dollar bill in his wallet. How did that bill help him get to this point? "It allowed me to believe in myself at a very young age," Barber said, "not to pay attention to those that said ActiveRain University (ARU) provides free online training. We coach, consult and support real estate professionals about real estate trends, technology and social media. A very high achiever who settles for nothing less than the best, Fred has excelled in selling all types of real estate throughout Victoria, from bare land and acreages to

34015, 950, 7422, 374, 539, 264, 13901, 13, 1102, 596, 264, 2204, 3169, 315, 10082, 1198, 264, 10082, 304, 279, 2410, 315, 6928, 11555, 323, 16024, 13, 14968, 439, 653, 85930, 439, 24632, 7422, 10578, 11, 88015, 258, 1071, 7978, 617, 6982, 1070, 527, 3115, 994, 433, 5084, 311, 617, 264, 1972, 2515, 25, 330, 3947, 574, 264, 4007, 1405, 1274, 304, 872, 5209, 12, 508, 82, 1051, 17303, 304, 3878, 315, 872, 54508, 1359, 1364, 1071, 11, 330, 438, 1243, 11, 220, 1135, 1667, 3010, 11, 1884, 889, 1051, 810, 37036, 11, 1051, 3604, 810, 4461, 311, 2103, 387, 13989, 13, 2100, 433, 596, 539, 2744, 24632, 311, 4510, 430, 701, 6928, 11555, 527, 3515, 264, 6928, 2515, 10246, 34015, 950, 7422, 8638, 304, 20587, 13, 2468, 279, 3907, [redacted] 7391, 47400, 3258, 706, 2507[redacted] 6811, 1990, 1148, 596, 1972, [redacted], 46227, 649, 5353, 264, 501[redacted] 1120, 1027, 6982, 311, 387, [redacted] 1505, 430, 11, 555, 922, 2[redacted], 315, 279, 6980, 584, 1296[redacted] 6227, 4375, 1359, 1071, 4740[redacted] 53, 387, 459, 3187, 315, 246[redacted], 813, 20587, 46227, 369, 70[redacted] 7126, 6688, 1461, 13, 330, [redacted], 24, 994, 856, 18233, 3782, 2162, 323, 6688, 757, 264, 12700, 3610, 18160, 4121, 1359, 568, 1071, 13, 330, 3112, 430, 6244, 2555, 430, 15609, 856, 9021, 13, 358, 3077, 18799, 389, 311, 433, 1475, 2533, 13, 24633, 69235, 1053, 1524, 6212, 449, 279, 4121, 1234, 813, 33326, 13, 4800, 11, 520, 220, 1544, 11, 568, 8640, 264, 80149, 15068, 18160, 1972, 12675, 2883, 304, 19051, 2663, 549, 10061, 11180, 11, 323, 568, 2103, 706, 430, 3610, 18160, 4121, 304, 813, 15435, 13, 2650, 1550, 430, 4121, 1520, 1461, 636, 311, 420, 1486, 30, 330, 2181, 5535, 757, 311, 4510, 304, 7182, 520, 264, 1633, 3995, 4325, 1359, 69235, 1071, 11, 330, 1962, 311, 2343, 6666, 311, 1884, 430, 1071, 10106, 60139, 3907, 320, 946, 52, 8, 5825, 1949,

ID







L'energia e la tecnologia sono pilastri del progresso umano, con la rivoluzione industriale che ha segnato l'inizio della loro simbiosi. Oggi, la transizione verso fonti rinnovabili è una sfida geopolitica cruciale, con Germania e Cina in prima linea. L'Internet delle Cose (IoT) e l'intelligenza artificiale (IA) ottimizzano l'uso delle risorse energetiche, ma sollevano questioni di sicurezza informatica. La geopolitica dell'energia è influenzata dalle innovazioni tecnologiche, come il fracking, che ha ridisegnato gli equilibri di potere globali. La dipendenza dalle fonti fossili rimane urgente, con la tecnologia che offre soluzioni promettenti per un'economia a basse emissioni di carbonio. La transizione richiede impegno politico e sociale, con accordi internazionali come l'Accordo di Parigi che giocano un ruolo chiave. In conclusione, il legame tra energia e tecnologia è complesso e multidimensionale, determinando il futuro sostenibile del nostro pianeta.

Post Training

Supervised Fine-Tuning (SFT)

Speciali annotatori (**Human Labellers**) creano **esempi ideali di risposte**, mostrando chiaramente al modello come dovrebbe comportarsi idealmente di fronte a determinati input o richieste.

Addestramento del Reward Model (RM)

Attraverso la scelta di risposte corrette vien addestrato un nuovo modello (detto Reward Model), il cui compito è imparare a prevedere automaticamente quale risposta preferirebbe un essere umano tra due o più alternative per rendere le risposte più umane

Reinforcement Learning tramite Proximal Policy Optimization (PPO)

Il Reward Model fornisce al modello principale un "premio" (reward) numerico per ogni risposta che il modello produce valorizzando risposte che si avvicinano maggiormente alle preferenze umane (cioè quelle preferite dal Reward Model)

Funzione Loss del Reward Model

$$\text{loss}(\theta) = -\frac{1}{\binom{K}{2}} \mathbb{E}_{(x,y_w,y_l)\sim D} \left[\log\left(\sigma\left(r_\theta(x, y_w) - r_\theta(x, y_l)\right)\right)\right]$$

log σ = REGRESSIONE LOGISTICA BINARIA

$$\text{logit}(p) = \log \frac{p}{1-p}$$

odds ratio
(p. di successo / p. insuccesso)

$$\sigma(z) = \frac{1}{1+e^{-z}}$$

Sigmoide con valore da 0 ad 1

valore scalare restituito dal Reward Model con parametri theta per il prompt x e la risposta y.

risposata winner e risposta loser

E: Expected Value
D: dataset contenente le preferenze umane raccolte (cioè coppie di risposte, delle quali una preferita rispetto all'altra).
K: numero totale di risposte diverse generate dal modello per ogni prompt

Metriche

1. Accuratezza (Accuracy)
- **Definizione: Percentuale di predizioni corrette sul totale delle predizioni effettuate.**
- **Applicazione: Utilizzata nei problemi di classificazione con classi bilanciate.**

2. Precisione (Precision)
- **Definizione: Rapporto tra il numero di veri positivi e il numero totale di elementi che il modello ha predetto come positivi.**
- **Applicazione: Importante quando il costo di un falso positivo è alto.**

3. Richiamo (Recall)
- **Definizione: Rapporto tra il numero di veri positivi e il numero totale di elementi che avrebbero dovuto essere predetti come positivi.**
- **Applicazione: Cruciale nei casi in cui è importante catturare tutti i positivi, come nelle diagnosi mediche.**

4. F1-score
- **Definizione: Media armonica di precisione e richiamo, utile quando c'è uno squilibrio tra classi.**
- **Applicazione: Fornisce una misura bilanciata delle performance del modello.**

5. Area sotto la curva ROC (AUC-ROC)
- **Definizione: Rappresenta la capacità del modello di distinguere tra classi positive e negative.**
- **Applicazione: Utilizzata per valutare modelli di classificazione binaria, specialmente con dataset sbilanciati.**

Cosa evitare durante l'addestramento

Overfitting: un modello che impara troppo accuratamente i dati di training tende a performare male su dati nuovi. Strategie come la **cross-validation**, il **dropout** o la **regularizzazione** sono essenziali.

Il data leakage: quando informazioni dai dati di test filtrano accidentalmente nel set di training, si ottiene un modello apparentemente molto accurato ma poco utile nella realtà.

Instabilità del gradiente: gradienti troppo instabili possono compromettere la convergenza del modello. L'impostazione corretta del **learning rate** aiuta a prevenire questo problema.

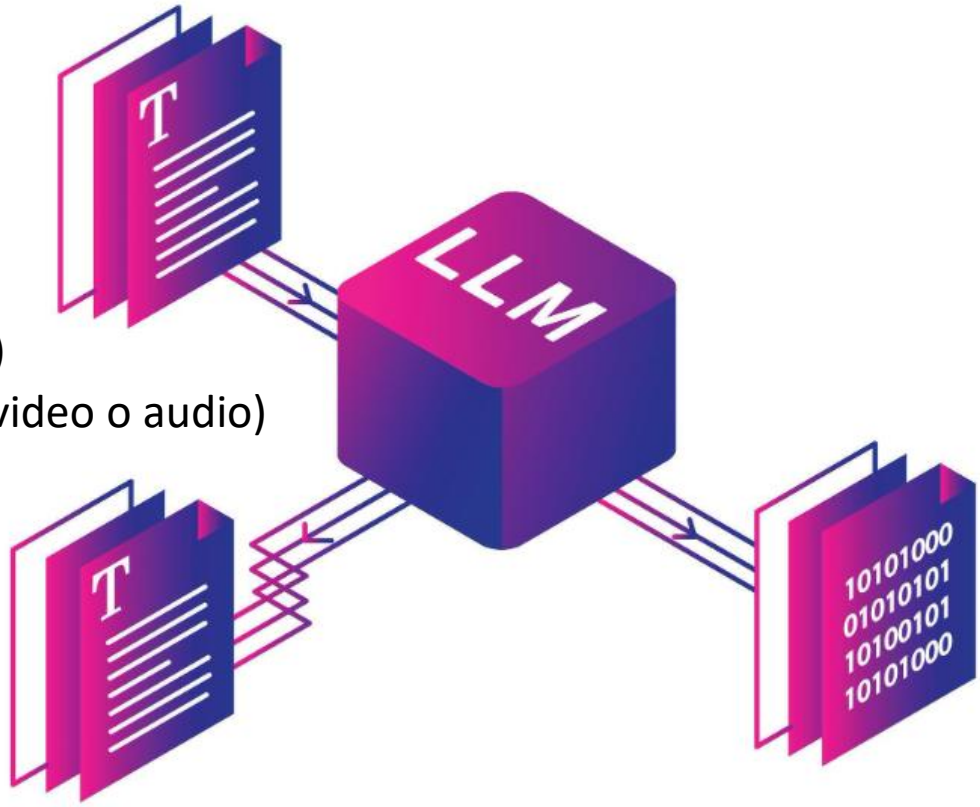
Bias e distorsioni: se il dataset di addestramento non è rappresentativo di tutti i contesti d'uso, il modello può esibire pregiudizi.

La scarsa diversità dei dati: se il dataset è troppo uniforme, il modello fatica ad apprendere pattern più generali e diventa fragile in contesti diversi.

Il Ruolo degli LLM nella Generative AI

Un **Large Language Model** è un modello di apprendimento automatico addestrato su enormi quantità di dati testuali, capace di riconoscere il contesto in input testuali e fornire risposte pertinenti

- **Generazione di Contenuti Testuali**
- **Interazioni Umano-Macchina** (coding)
- **Supporto per Altri Media** (immagini, video o audio)
- **Automazione della Creatività**

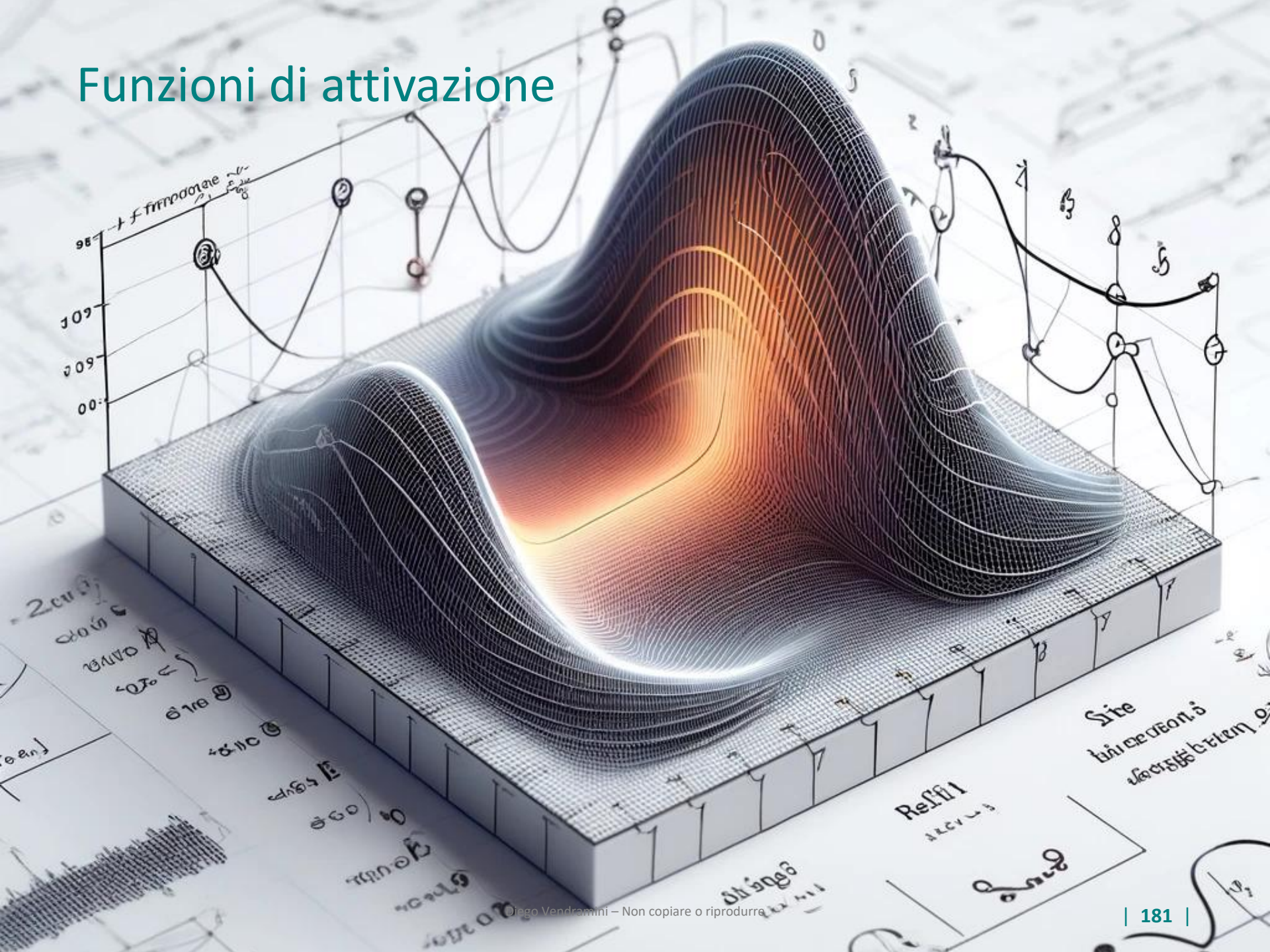


Diego Vendramini – Non copiare o riprodurre

Quanti ne esistono?

LLM Name	Developer	Release Date	Access	Parameters
GPT-4.5	OpenAI	Feb 27, 2025	API	Unknown
Claude 3.7 Sonnet	Anthropic	Feb 24, 2025	API	Unknown (est. 200B+)
Grok-3	xAI	Feb 17, 2025	API	Unknown
Gemini 2.0 Flash-Lite	Google DeepMind	Feb 5, 2025	API	Unknown
Gemini 2.0 Pro	Google DeepMind	Feb 5, 2025	API	Unknown
GPT-o3-mini	OpenAI	Jan 31, 2025	API	Unknown
Qwen 2.5-Max	Alibaba	Jan 29, 2025	API	Unknown
DeepSeek R1	DeepSeek	Jan 20, 2025	API, Open Source	671B (37B active)
DeepSeek-V3	DeepSeek	Dec 26, 2024	API, Open Source	671B (37B active)
Gemini 2.0 Flash	Google DeepMind	Dec 11, 2024	API	Unknown
Sora	OpenAI	Dec 9, 2024	API	Unknown
Nova	Amazon	Dec 3, 2024	API	Unknown
Claude 3.5 Sonnet (New)	Anthropic	Oct 22, 2024	API	Unknown
GPT-o1	OpenAI	Sept 12, 2024	API	Unknown (o1-mini est. ~100B)
DeepSeek-V2.5	DeepSeek	Sept 5, 2024	API, Open Source	Unknown
Grok-2	xAI	Aug 13, 2024	API	Unknown
Mistral Large 2	Mistral AI	July 24, 2024	API	123B
Llama 3.1	Meta AI	July 23, 2024	Open Source	405B
GPT-4o mini	OpenAI	July 18, 2024	API	~8B (est.)
Nemotron-4	Nvidia	July 14, 2024	Open Source	340B
Claude 3.5 Sonnet	Anthropic	June 20, 2024	API	~175-200B (est.)
GPT-4o	OpenAI	May 13, 2024	API	~1.8T (est.)
DeepSeek-V2	DeepSeek	May 6, 2024	API, Open Source	Unknown
Phi-3	Microsoft	April 23, 2024	API, Open Source	Mini 3B, Small 7B, Medium 14B
Mixtral 8x22B	Mistral AI	April 10, 2024	Open Source	141B (39B active)
Jamba	AI21 Labs	Mar 29, 2024	Open Source	52B (12B active)
DBRX	Databricks' Mosaic ML	Mar 27, 2024	Open Source	132B
Command R	Cohere	Mar 11, 2024	API, Open Source	35B
Inflection-2.5	Inflection AI	Mar 7, 2024	Proprietary	Unknown (predecessor ~400B)
Gemma	Google DeepMind	Feb 21, 2024	API, Open Source	2B, 7B
Gemini 1.5	Google DeepMind	Feb 15, 2024	API	~1.5T Pro, ~8B Flash (est.)
Stable LM 2	Stability AI	Jan 19, 2024	Open Source	1.6B, 12B
Grok-1	xAI	Nov 4, 2023	API, Open Source	314 billion
Mistral 7B	Mistral AI	Sept 27, 2023	Open Source	7.3 billion
Falcon 180B	Technology Innovation Institute	Sept 6, 2023	Open Source	180 billion
XGen-7B	Salesforce	July 3, 2023	Open Source	7 billion
PaLM 2	Google	May 10, 2023	API	340 billion
Alpaca 7B	Stanford CRFM	Mar 13, 2023	Open Source	7 billion
Pythia	EleutherAI	Mar 13, 2023	Open Source	70 million to 12 billion

Funzioni di attivazione



Diego Vendramini – Non copiare o riprodurre

In una rete neurale, dopo ogni trasformazione lineare (es. una moltiplicazione di matrici o una somma pesata), si applica una **funzione non lineare** chiamata **funzione di attivazione**.
Il suo compito è:
•Introdurre **non linearità**, altrimenti la rete si ridurrebbe a una grande trasformazione lineare;
•Permettere al modello di apprendere **relazioni complesse e non banali** nei dati.

.

Vanishing Gradient (gradiente che scompare)

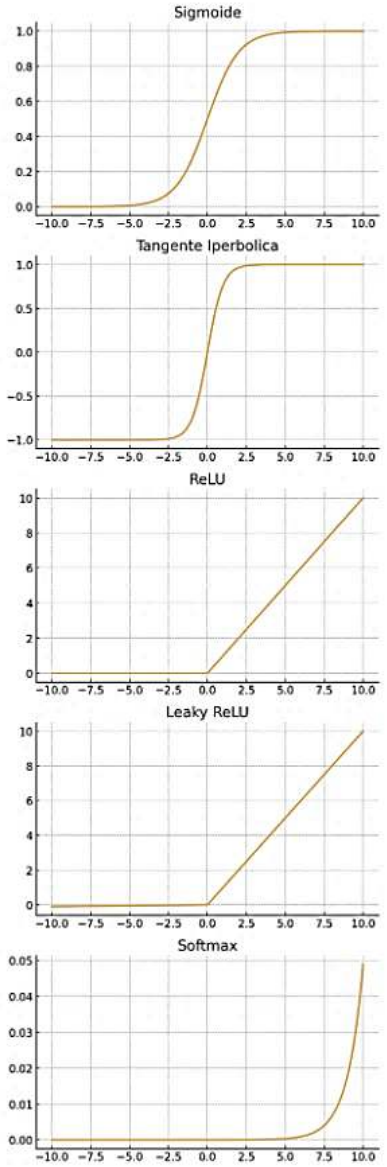
✅ **Cos’è:**
Durante la **backpropagation**, i gradienti (cioè le “correzioni” ai pesi) diventano **sempre più piccoli** man mano che si risale ai primi strati della rete.

🎯 **Perché accade:**
Succede soprattutto con **funzioni di attivazione saturanti** come la **sigmoide** o la **tanh**, la cui derivata:
•è **piccola** (vicina a 0) quando l’input è molto positivo o negativo;
•quindi, il gradiente **si riduce moltiplicazione dopo moltiplicazione**.

🧨 **Effetto:**
•I **primi layer della rete non si aggiornano**.
•Il modello **non impara** rappresentazioni utili nei livelli iniziali.
•La rete sembra "bloccata" o "superficiale".

Diego Vendramini – Non copiare o riprodurre

Nome	Nome Completo	Formula	Intervallo	Pro	Contro
Sigmoide	Sigmoide	$\sigma(x) = \frac{1}{1+e^{-x}}$	0 to 1	Buona per modelli probabilistici, semplice.	Problema del gradiente che scompare, output non centrato su zero.
Tanh	Tangente Iperbolica	$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$	-1 to 1	Output centrato su zero, buona per modelli che necessitano di valori negativi.	Problema del gradiente che scompare.
ReLU	Rectified Linear Unit	$\text{ReLU}(x) = \max(0, x)$	0 to ∞	Efficiente, risolve il problema del gradiente che scompare.	Problema dei neuroni morti.
Leaky ReLU	Leaky Rectified Linear Unit	$\text{Leaky ReLU}(x) = \max(0.01x, x)$	[-∞, 0) to ∞	Mitiga il problema dei neuroni morti, mantiene la semplicità computazionale della ReLU.	Richiede l'impostazione di iperparametri aggiuntivi.
Softmax	Softmax	$\text{Softmax}(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}}$	0 to 1	Utile per classificazione multiclasse, output normalizzati.	Non utilizzata negli strati intermedi della rete.



Diego Vendramini – Non copiare o riprodurre

E quindi Chatgpt cosa usa?

GELU (Gaussian Error Linear Unit)

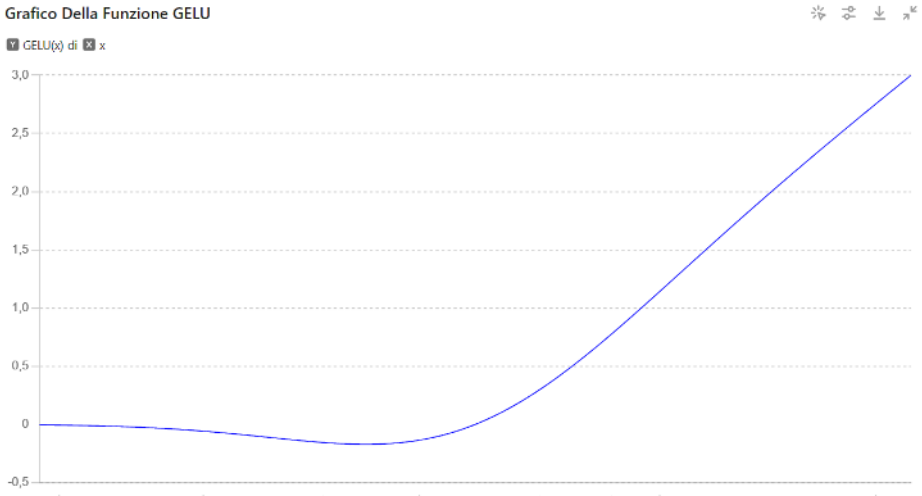
GELU: La funzione GELU è definita come:

$$\text{GELU}(x) = x \cdot \Phi(x)$$

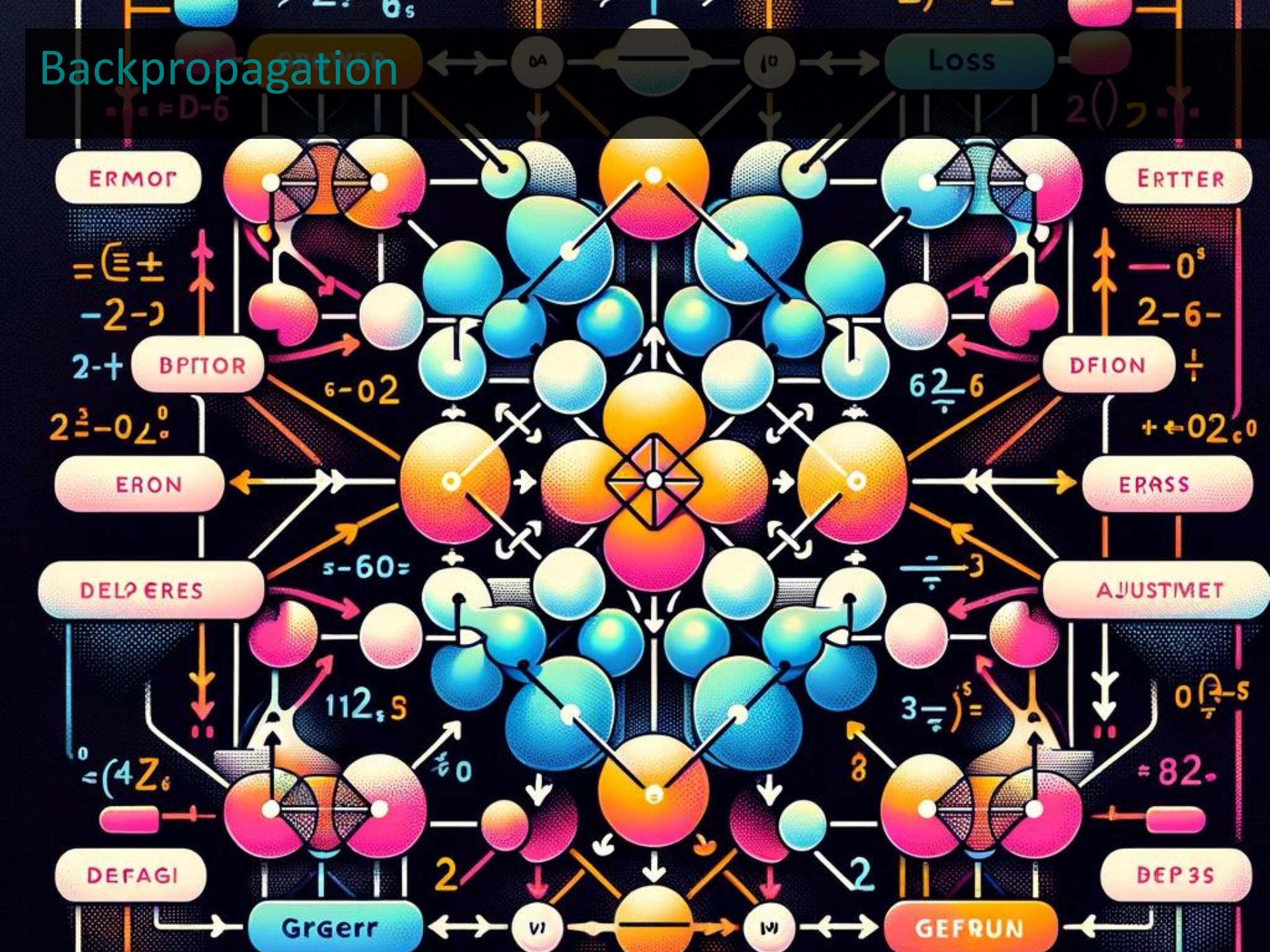
dove $\Phi(x)$ è la funzione cumulativa della distribuzione normale standard.

Una forma approssimativa comunemente usata è:

$$\text{GELU}(x) = 0.5 \cdot x \cdot \left(1 + \tanh\left(\sqrt{\frac{2}{\pi}}(x + 0.044715x^3)\right)\right)$$



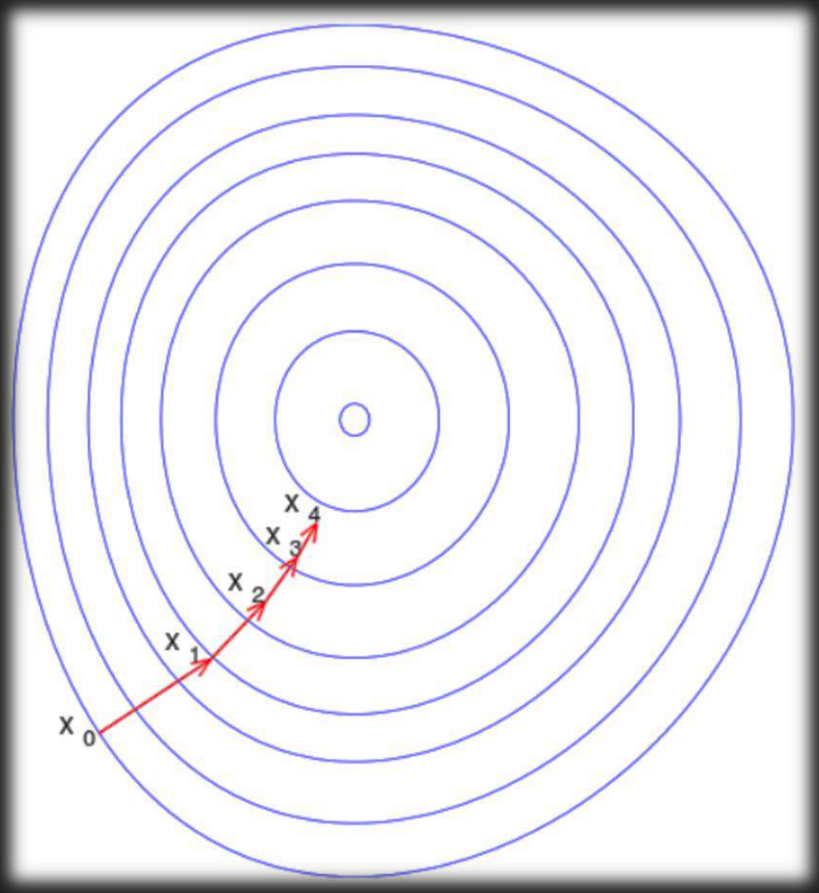
Backpropagation



Backpropagation

Algoritmo per calcolare i gradienti in una rete neurale

L’ algoritmo confronta il valore in uscita del sistema con il prodotto con quello atteso, calcola l’errore e modifica i pesi sinaptici della rete neurale e i bias (riducendoli quanto più possibile), facendo convergere progressivamente il set dei valori di uscita verso quelli desiderati.



Fonte: Wikipedia

Rete neurale multilivello - Funzionamento

FASI

Input di dati: gli input passano ai neuroni nello livello successivo.

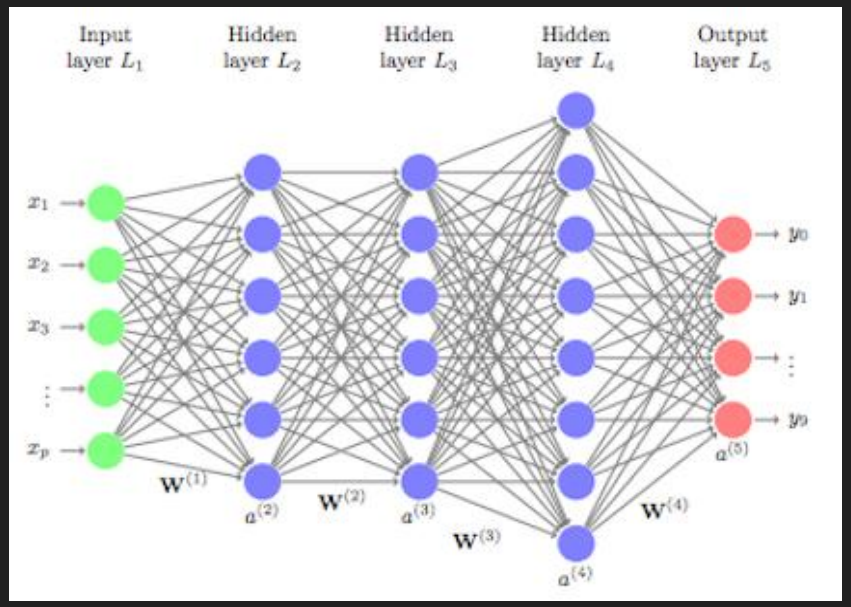
Elaborazione dei dati: ogni neurone nello strato nascosto calcola un valore ponderato **(casuale)** degli input, somma questi valori e applica una **funzione di attivazione**. Il risultato viene poi passato ai neuroni del prossimo livello.

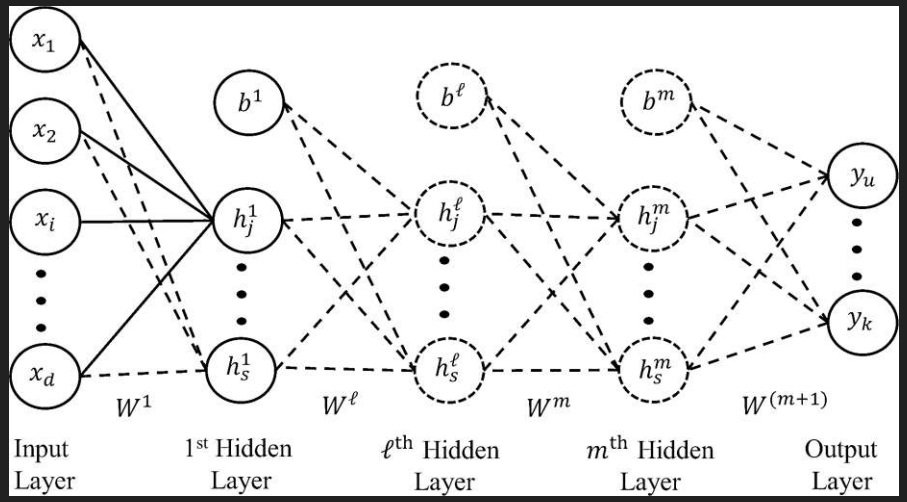
Output dei dati: livello dopo livello si produce il risultato finale.

Apprendimento: Confronta l'output prodotto con l'output atteso, (funzione di costo). La differenza tra i due è l'errore.

Backpropagation: I pesi delle connessioni tra i neuroni e i bias vengono aggiornati in modo iterativo per ridurre l'errore. L'obiettivo dell'apprendimento è minimizzare questo errore.

Iterazione: Il processo di input, elaborazione, output e apprendimento viene ripetuto per molti esempi di dati





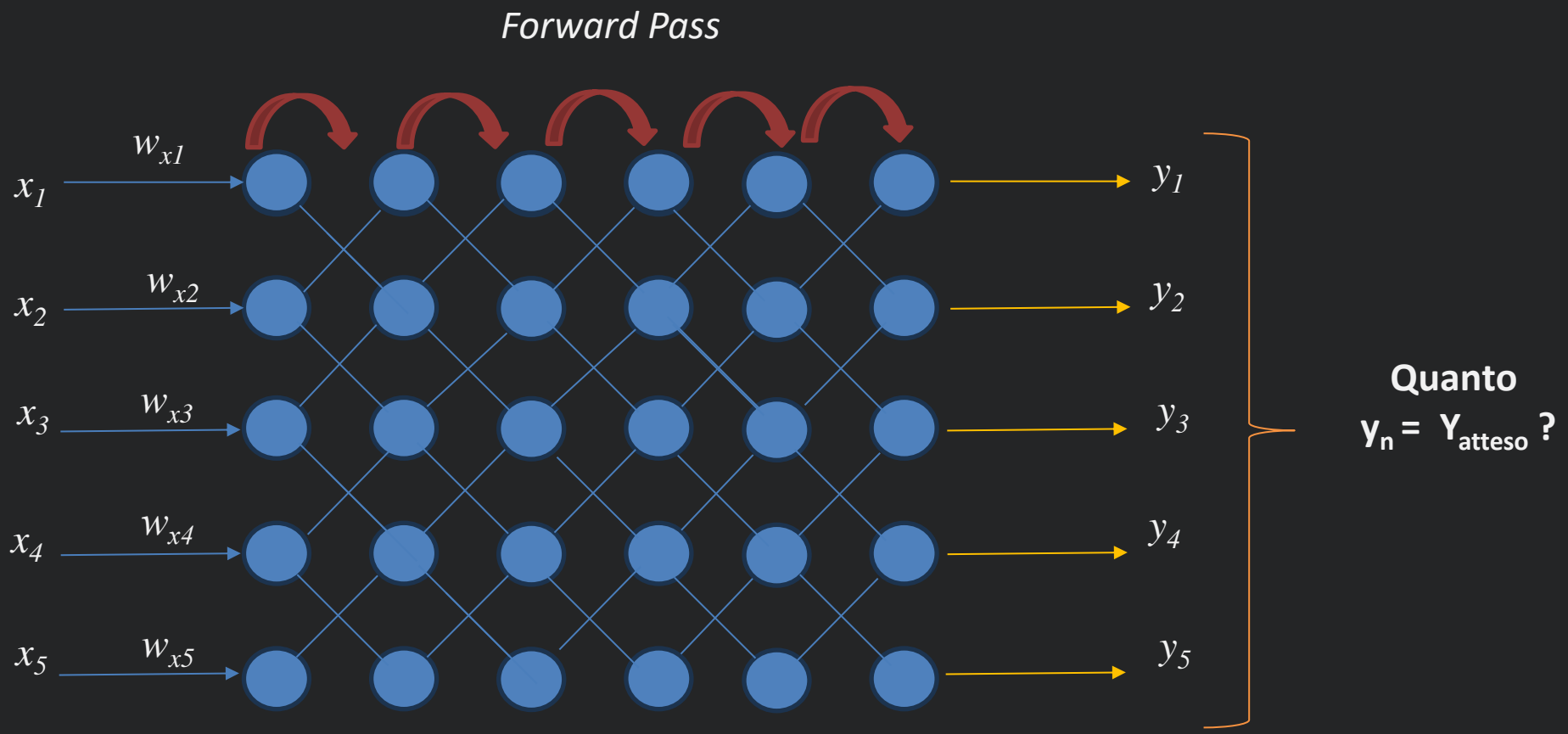
Fonte immagini: Wikipedia

Backpropagation

Nella fase di training abbiamo insegnato alla modello che ad un input X_1 corrisponde un determinato Y corretto

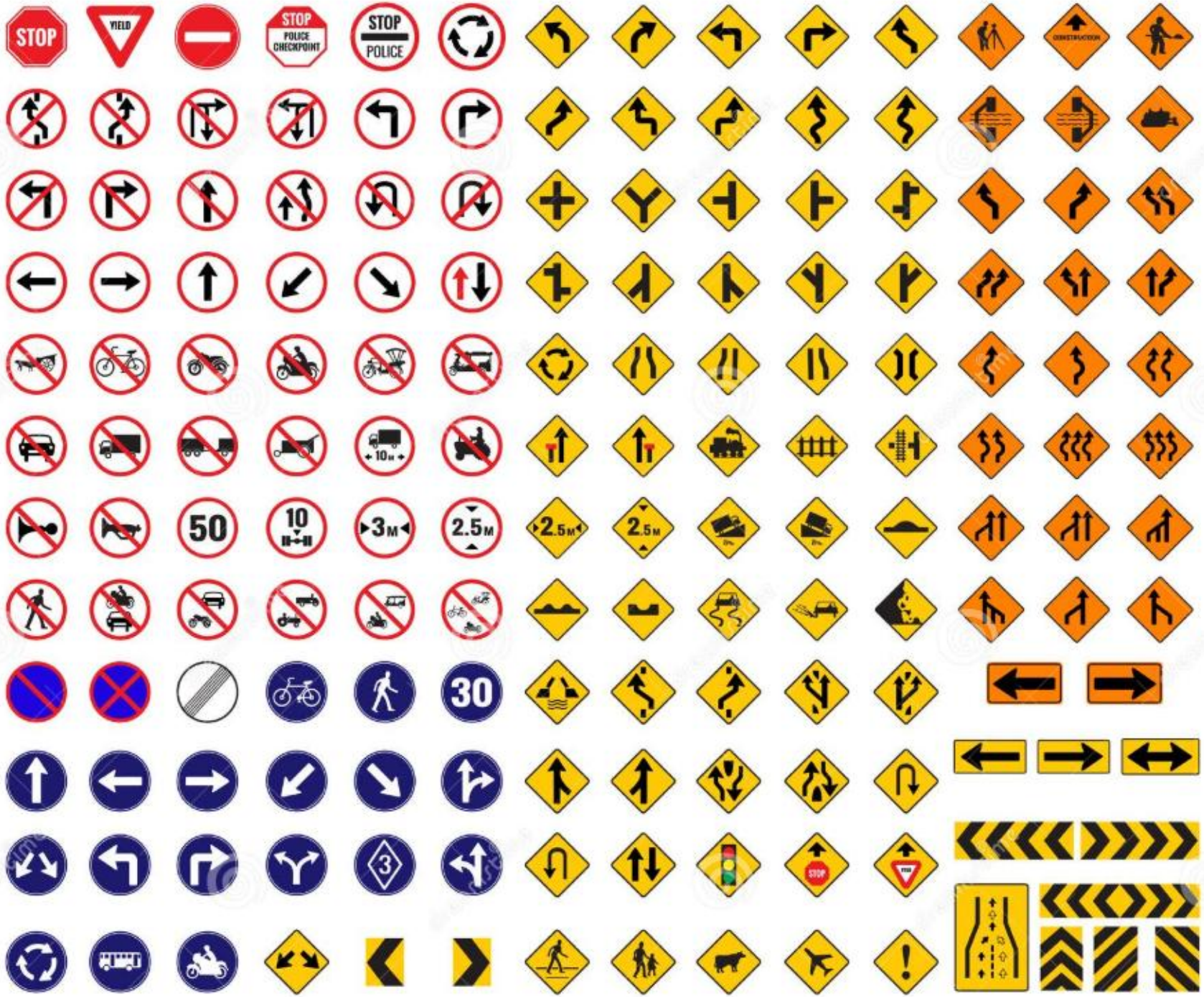
Fonte immagine: https://matthewmazur

Funzioni di Backpropagation



Fonte immagine: https://matthewmazur

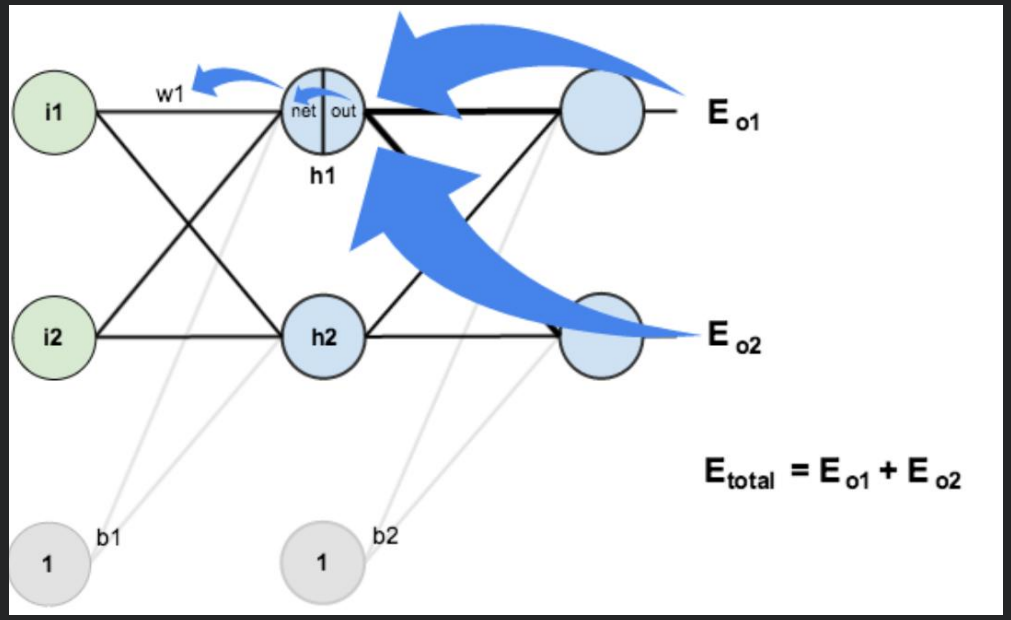
Diego Vendramini – Non copiare o riprodurre


STOP
YIELD
STOP
POLICE CHECKPOINT
STOP
POLICE
50
10
3M
2.5M
2.5M
2.5M
10M
30

Backpropagation

Algoritmo di apprendimento delle reti neurali che

1. confronta il valore in uscita del sistema con il prodotto con quello atteso
2. calcola l'errore con una **funzione di costo**
3. modifica i pesi sinaptici della rete neurale e il bias, facendo convergere progressivamente il set dei valori di uscita verso quelli desiderati minimizzando l'errore



Fonte immagine: https://matthewmazur

Backpropagation - Processo

1 - Lavoriamo su vettori
Ogni neurone (o nodo) nella rete ha pesi, che vengono rappresentati come un vettore di pesi.
Ad esempio, se ci sono 3 pesi, il vettore sarà: $\mathbf{w} = [w_1, w_2, w_3]$

2- Calcoliamo il gradiente :
un vettore che indica quanto varia l'errore se modifichiamo ciascun peso.
Si calcola tramite le **derivate parziali** della funzione errore rispetto a ciascun peso:

$$\nabla E(\mathbf{w}) = \begin{bmatrix} \frac{\partial E}{\partial w_1} \\ \frac{\partial E}{\partial w_2} \\ \frac{\partial E}{\partial w_3} \end{bmatrix}$$

3 - Otteniamo nuovi vettori : il gradiente discendente aggiorna i pesi

$$\mathbf{w}^{(k+1)} = \mathbf{w}^{(k)} - \eta \nabla E(\mathbf{w}^{(k)})$$

4 - Ogni nuovo vettore w(k+1) rappresenta una **nuova configurazione** dei pesi, più vicina a quella che minimizza l'errore.

Fonte immagine: https://matthewmazur

Backpropagation - Processo

5 - La funzione errore viene ricalcolata
Dopo ogni aggiornamento dei pesi, la rete ricalcola l'**errore** usando la nuova configurazione dei pesi

$$E(\mathbf{w}^{(k+1)}) = \text{Funzione Errore}$$

Questo permette di verificare se il processo di apprendimento sta migliorando la rete.

6 - Il processo si ripete iterativamente
Si continua a calcolare il gradiente e ad aggiornare i pesi finché l'errore non raggiunge un valore sufficientemente basso o fino a un numero massimo di iterazioni.

Fonte immagine: https://matthewmazur

Diego Vendramini – Non copiare o riprodurre

Gradiente discendente

(*Nabla*)

$$w^{(k+1)} = w^{(k)} - \eta \nabla E(w^{(k)})$$

(*Eta*)

$w^{(k)}$ = Vettore dei pesi al passo k.

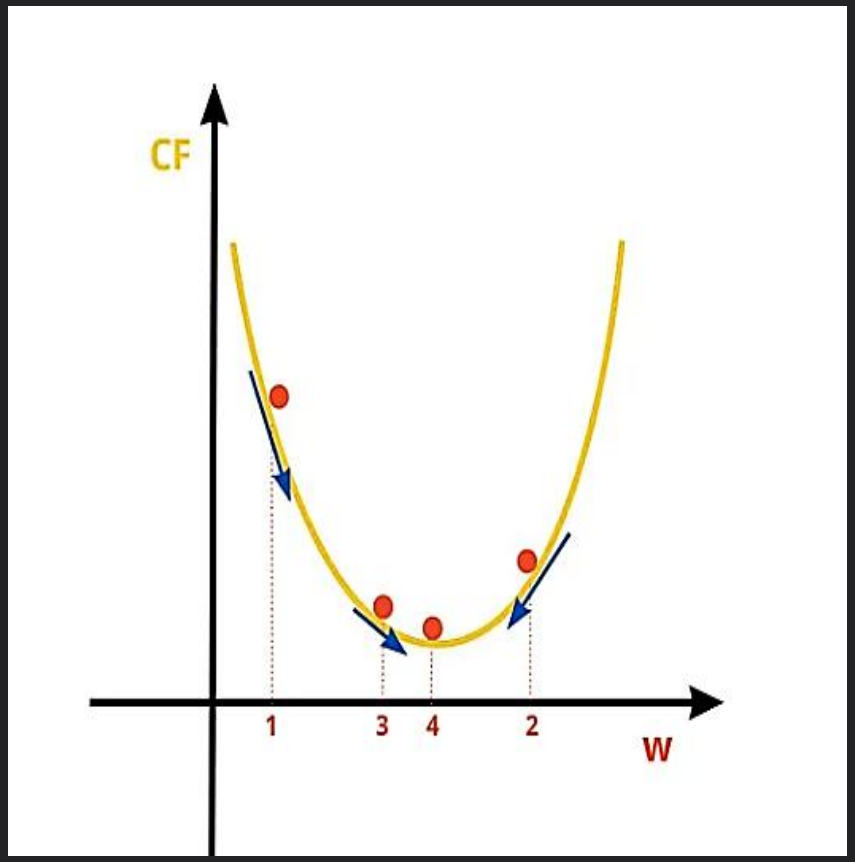
η = Learning rate, un parametro positivo che regola la dimensione del passo nella direzione del gradiente.

$\nabla E(w^{(k)})$ = Gradiente della funzione obiettivo E valutato nel punto $w^{(k)}$.

Fonte: Wikipedia

Funzione di Errore	Formula	Tipologia	Caratteristiche
Mean Squared Error (MSE)	$MSE = \frac{1}{m} \sum_{i=1}^{m} (y_{pred}^{(i)} - y_{true}^{(i)})^2$	Regressione	Molto sensibile agli outlier.
Binary Cross-Entropy	$-\sum_{i=1}^{m} \left[y_{true}^{(i)} \log(y_{pred}^{(i)}) + (1 - y_{true}^{(i)}) \log(1 - y_{pred}^{(i)}) \right]$	Classificazione binaria	Molto efficace per compiti di classificazione binaria.
Categorical Cross-Entropy	$-\sum_{i=1}^{m} \sum_{c=1}^{C} y_{true}^{(i,c)} \log(y_{pred}^{(i,c)})$	Classificazione multi-classe	Ideale per classificazione multi-classe, usata tipicamente con Softmax.
KL Divergence (Divergenza KL)	$D_{KL}(P \parallel Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}$	Misura di distanza tra distribuzioni	Misura la differenza tra due distribuzioni di probabilità (P rispetto a Q). Non è simmetrica.
Cross-Entropy Loss (tipica per LLM)	$-\sum_{i=1}^{m} \sum_{c=1}^{C} y_{true}^{(i,c)} \log(y_{pred}^{(i,c)})$ (con Softmax)	Previsione token nei modelli LLM	Gestisce previsioni di token basate su distribuzioni di probabilità.

Gradiente Discendente



∇f

Vettore delle derivate parziali di una funzione con più variabili f(x, y, z) rispetto ad un punto.

Ci dà direzione e pendenza della funzione di costo.

Se seguiamo la direzione del gradiente, aumentiamo il valore della funzione; **se andiamo nella direzione opposta, diminuiamo il valore della funzione.**

Fonte:https://medium.com

Limitazioni del Gradient Descent e alternative

Scelta del tasso di apprendimento:

Se troppo alto → oscillazioni.
Se troppo basso → convergenza lenta.

Diego Vendramini – Non copiare o riprodurre

RMSprop (Root Mean Square Propagation)

Introduce una **media mobile esponenziale** dei quadrati dei gradienti per stabilizzare il processo di apprendimento in caso di gradienti molto variabili.

I pesi con gradienti grandi vengono aggiornati con un **passo più piccolo**.
I pesi con gradienti piccoli vengono aggiornati con un **passo più grande**.

Questo consente una discesa più stabile e riduce le oscillazioni nei gradienti.

Diego Vendramini – Non copiare o riprodurre

Momentum

Si ispira alla **fisica**: il peso del modello si aggiorna simulando il movimento di una palla che rotola lungo una superficie irregolare.

Momentum: $$v_t = \beta v_{t-1} + (1 - \beta)\nabla E(w_t)$$

Aggiornamento dei pesi: $$w_{t+1} = w_t - \eta v_t$$

w_t → Vettore dei pesi al tempo t.

v_t → Velocità accumulata (il "momentum").

β → Fattore di decadimento del Momentum (tipicamente tra 0.8 e 0.99).

η → Learning rate (passo di apprendimento).

$\nabla E(w_t)$ → Gradiente della funzione errore rispetto ai pesi al tempo t.

Diego Vendramini – Non copiare o riprodurre

ADAM (Adaptive Moment Estimation)

Adam è un ottimizzatore avanzato che combina le migliori caratteristiche di:

✅ **Momentum** → Per accelerare la convergenza e stabilizzare i pesi.

✅ **RMSprop** → Per adattare dinamicamente il learning rate e gestire gradienti instabili.

Primo momento (mt) → Simile al Momentum; rappresenta una media mobile esponenziale del gradiente (direzione preferita).

Secondo momento (vt) → Simile a RMSprop; rappresenta una media mobile esponenziale dei **quadrati** del gradiente (stima della variabilità).

Diego Vendramini – Non copiare o riprodurre

ADAM (Adaptive Moment Estimation)

Step 1: Calcolo del primo momento (la "velocità")

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1)\nabla E(w_t)$$

- m_t = Primo momento (velocità del gradiente)
- β_1 = Parametro che determina quanto contano i gradienti passati (tipico valore: 0.9)
- $\nabla E(w_t)$ = Gradiente corrente

Step 2: Calcolo del secondo momento (la "varianza")

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2)(\nabla E(w_t))^2$$

- v_t = Secondo momento (stima della variabilità del gradiente)
- β_2 = Parametro che controlla quanto contano i gradienti passati nel calcolo della varianza (tipico valore: 0.999)

Diego Vendramini – Non copiare o riprodurre

ADAM (Adaptive Moment Estimation)

🔷 Step 3: Correzione del bias

Poiché m_t e v_t iniziano entrambi vicini a zero, viene applicata una correzione del bias per renderli più accurati nelle prime iterazioni:

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \qquad \hat{v}_t = \frac{v_t}{1 - \beta_2^t}$$

✅ Questo passaggio garantisce che il processo parta in modo stabile anche quando i gradienti iniziali sono piccoli.

Diego Vendramini – Non copiare o riprodurre

ADAM (Adaptive Moment Estimation)

🔷 Step 4: Aggiornamento dei pesi

Infine, Adam aggiorna i pesi secondo questa regola:

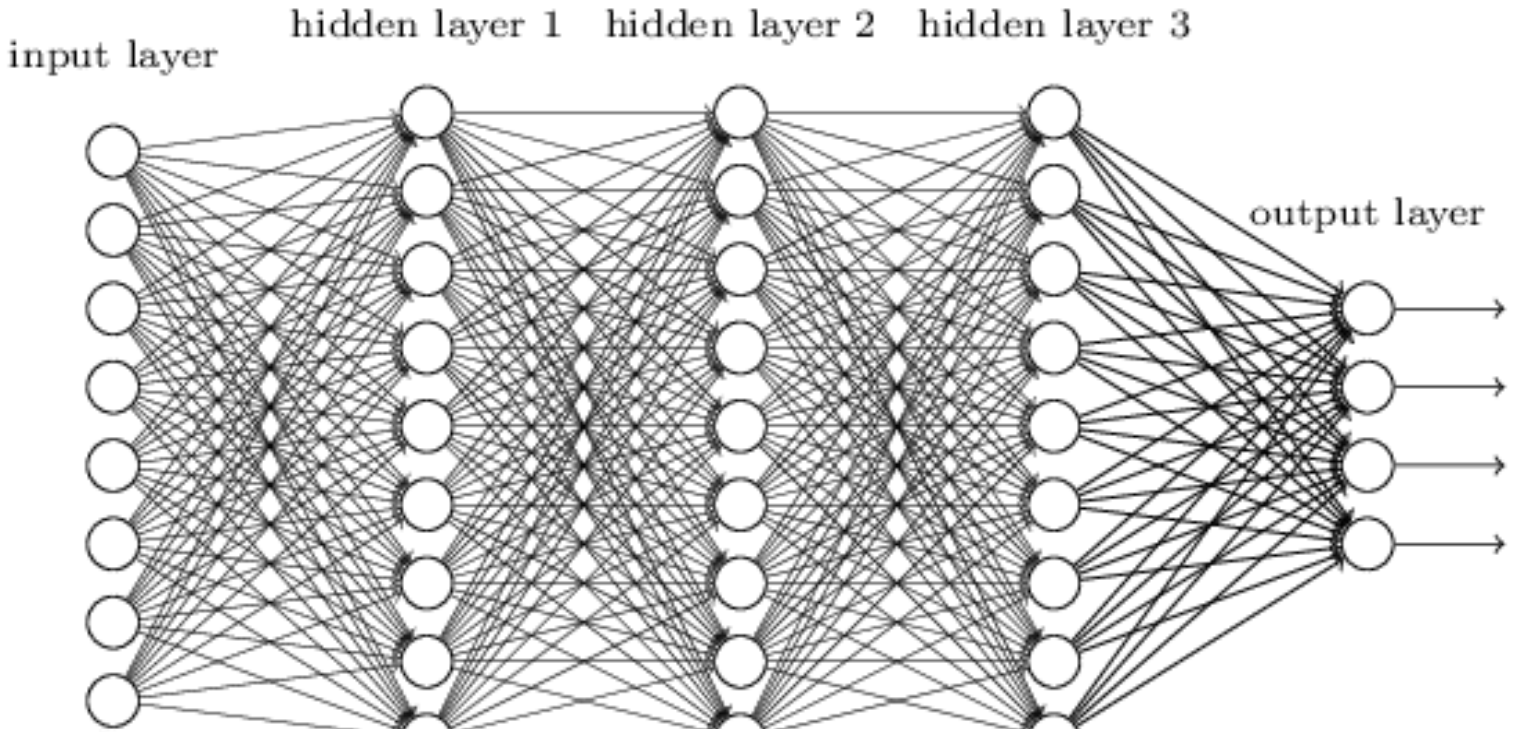
$$w_{t+1} = w_t - \frac{\eta \hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon}$$

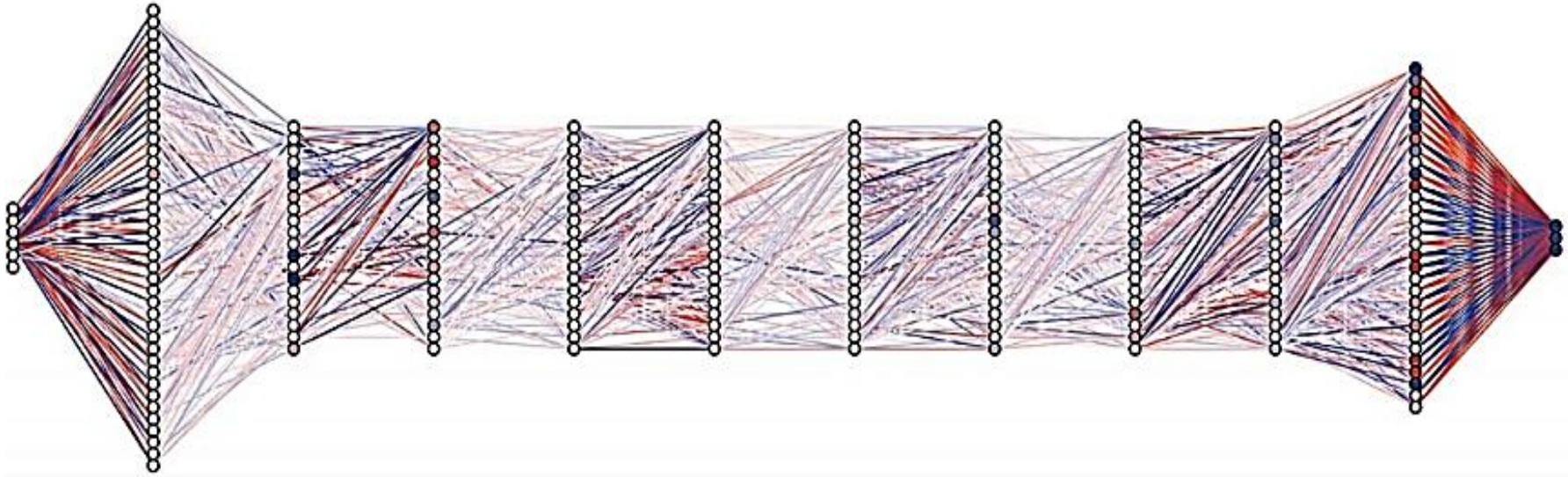
- η = Learning rate (tipico valore: 0.001)
- ϵ = Valore piccolo (es. 10^{-8}) per evitare divisioni per zero

✅ Questo passaggio consente di regolare il passo di aggiornamento in modo flessibile:

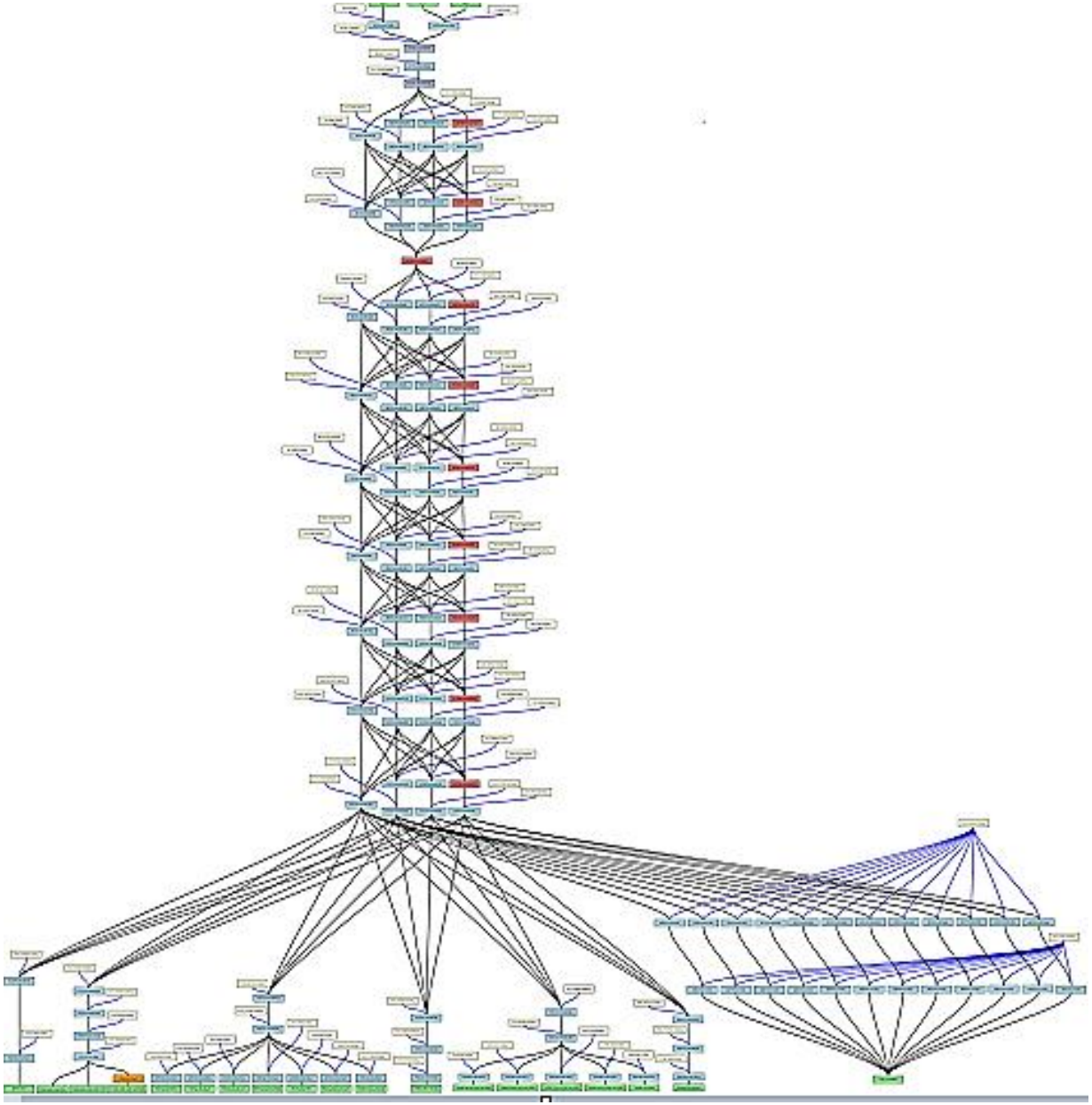
- Per gradienti **grandi**, il denominatore (radice di v_t) è elevato → **passi piccoli**.
- Per gradienti **piccoli**, il denominatore è basso → **passi grandi**.

Diego Vendramini – Non copiare o riprodurre





Fonte immagini www. teslari.it



Fonte Immagini www. teslari.it

Caratteristica	Gradient Descent	Momentum	RMSprop	Adam
Aggiornamento basato su gradiente corrente	✅ Sì	✅ Sì (con memoria del gradiente precedente)	✅ Sì (con media mobile esponenziale dei gradienti)	✅ Sì (con memoria e correzione del bias)
Memorizza velocità del gradiente (Momentum)	❌ No	✅ Sì	❌ No	✅ Sì
Adatta dinamicamente il learning rate	❌ No	❌ No	✅ Sì	✅ Sì
Correzione del bias iniziale	❌ No	❌ No	❌ No	✅ Sì
Gestione dei gradienti rumorosi	❌ Debole	✅ Buona	✅ Ottima	✅ Eccellente
Gestione di gradienti molto piccoli (sparse gradients)	❌ Debole	✅ Moderata	✅ Buona	✅ Eccellente
Velocità di convergenza	🟠 Lenta	🟠 Moderata	✅ Veloce	✅ Molto veloce
Robusto per funzioni di perdita con gradienti oscillanti	❌ No	✅ Sì	✅ Sì	✅ Sì
Adatto a reti neurali profonde	❌ Difficile da gestire	🟠 Discreto	✅ Buono	✅ Ottimo
Ideale per reti ricorrenti (RNN/LSTM)	❌ No	🟠 Moderato	✅ Ottimo	✅ Eccellente
Facile da implementare	✅ Molto semplice	🟠 Moderata	🟠 Moderata	🔴 Più complesso

I token

Chat GPT - Token

I token

Sono singole unità di un testo o di un linguaggio.

I token vengono utilizzati nelle reti neurali per rappresentare le unità di **input** e di **output** durante il processo.

LLM: la dimensione linguistica

Linguistica e LLM

PAROLA

Cosa è una parola?
"*segmento della catena parlata e del testo scritto, tale che non sia interrompibile da altri elementi, che sia mobile, che possa comparire da solo e che abbia significato*" (fonte: Beccaria, 1996)

Quante tipologie linguistiche esistono?

Tipologia di Lingua	Esempi di Lingue (diffuse)	Caratteristiche
ANALITICHE	Inglese, Francese, Italiano (parzialmente)	Uso minimo della morfologia, parole autonome e struttura chiara
AGGLUTINANTI	Turco, Giapponese, Coreano, Ungherese	Parole complesse con morfemi distinti e aggiunti alla radice
FLESSIVE	Latino, Greco antico, Russo, Arabo, Tedesco, Spagnolo	Morfemi che combinano più informazioni grammaticali (caso, numero, ecc.)
INCORPORANTI (POLISINTETICHE)	Inuit	Parole molto lunghe che includono più informazioni grammaticali (soggetto, oggetto, verbo)
ISOLANTI	Cinese mandarino, Vietnamita, Tailandese	Parole invariabili e assenza di morfologia flessiva
FUSIONALI	Russo, Spagnolo, Greco antico	Morfemi che combinano molte informazioni grammaticali fuse in una radice
SINTETICHE	Giapponese, Coreano	Parole composte da più morfemi; alta complessità grammaticale

Diego Vendramini – Non copiare o riprodurre

Quante tipologie linguistiche esistono?

Tipologia di Lingua	Efficacia della Tokenizzazione	Motivo principale
ANALITICHE	Alta	Parole separate e struttura sintattica ben definita
AGGLUTINANTI	Moderata	Ogni morfema ha una funzione chiara e distinta
FLESSIVE	Moderata/Bassa	Morfemi polivalenti e radici modificabili
INCORPORANTI (POLISINTETICHE)	Bassa	Parole complesse che rappresentano frasi intere
ISOLANTI	Alta	Parole autonome e distinte; relazioni grammaticali espresse tramite l'ordine delle parole
FUSIONALI	Moderata/Bassa	Fusione di molte informazioni in un unico morfema
SINTETICHE	Variabile	Dipende dal predominio di caratteristiche agglutinanti o flessive

Diego Vendramini – Non copiare o riprodurre

Teorie

"**Grammatica delle dipendenze " :**

Si concentra sulle relazioni tra le parole, enfatizzando come ogni parola dipenda da un'altra in una frase.

Il gatto MANGIA il topo è diverso da il topo MANGIA il gatto

"Semantica Distribuzionale":
il significato delle parole può essere compreso analizzando il contesto in cui appaiono (Word2Vec)

Es: lingua flessiva - latino
"Marcus puellam amat." (Marco ama la ragazza)

"Puellam Marcus amat."
(Marco ama la ragazza)
maggiore enfasi sulla ragazza)

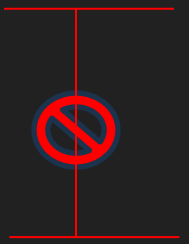
" Amat Marcus puellam."
(Ama Marco la ragazza)
Maggior importanza sull'azione)

Semantica Distributiva

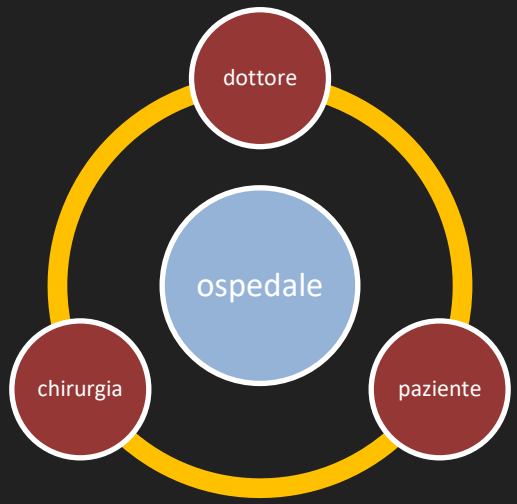
Il gatto dorme sul divano

Il felino riposa sulla poltrona

Il computer era posato sul banco.

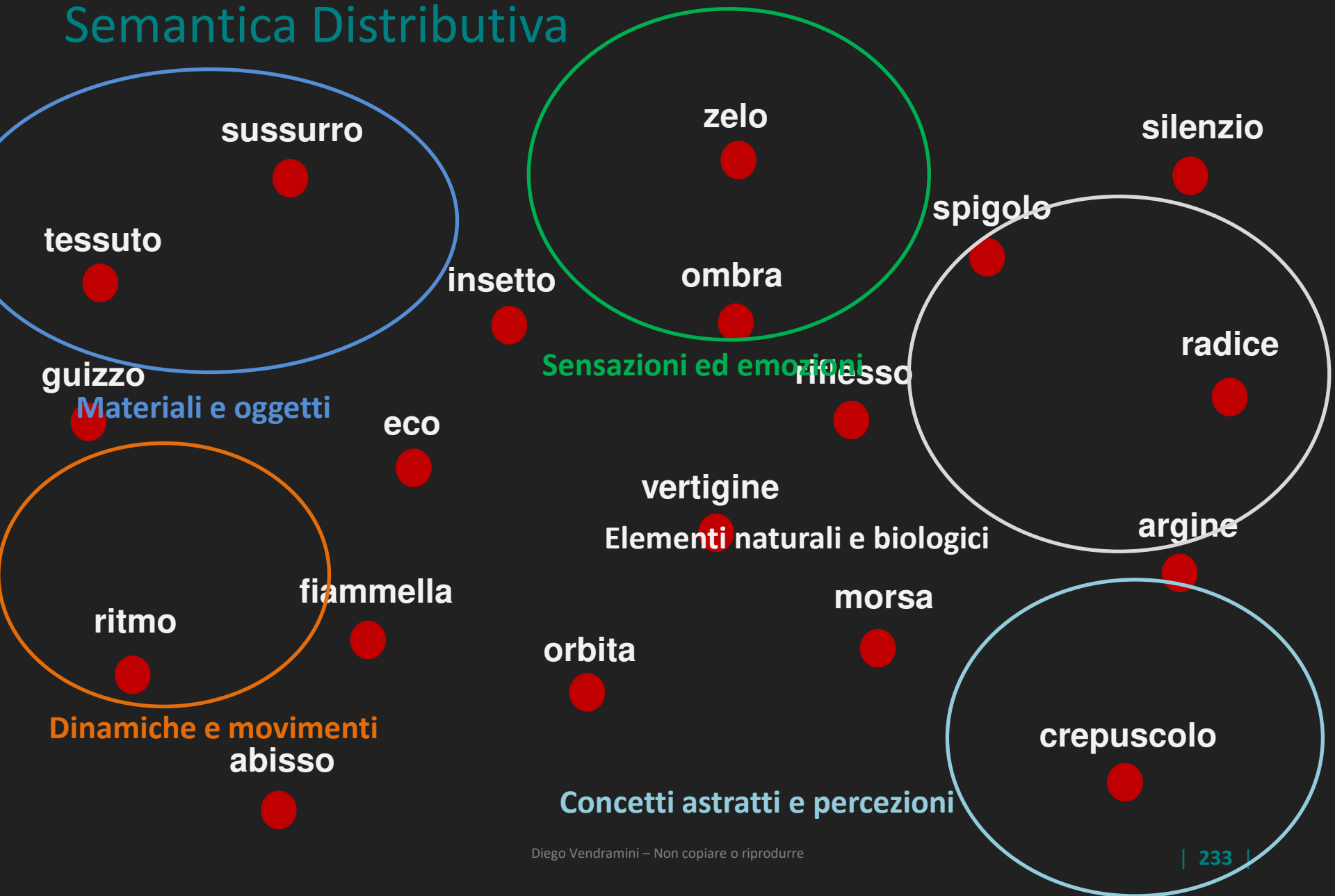


I moduli sono installati nel banco della memoria.



Diego Vendramini – Non copiare o riprodurre

Semantica Distributiva



Dimensioni del linguaggio

- **Sintassi**

  Struttura delle frasi e regole grammaticali.

- **Semantica**

  Significato delle parole nelle frasi. Relazioni di significato tra le parole

- **Morfologia**

  Struttura interna delle parole (radici, prefissi, suffissi).

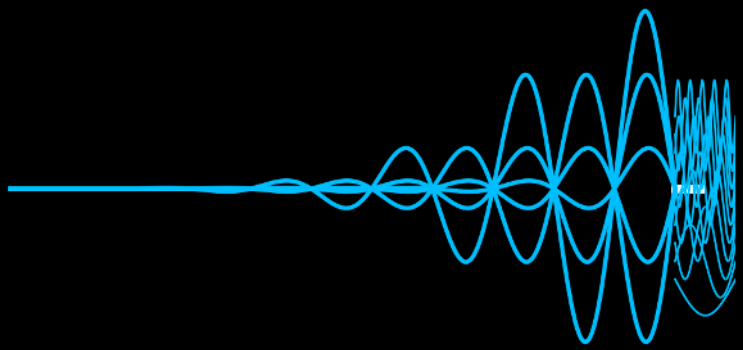
- **Fonologia/Prosodia**

  Aspetti sonori del linguaggio. Aspetti ritmici, melodici e intonativi del linguaggio parlato.

  Include elementi come l'intonazione, l'accento, il ritmo e la pausa, che aiutano a trasmettere emozioni, intenzioni e significati aggiuntivi oltre al testo stesso.

Diego Vendramini – Non copiare o riprodurre

Cosa succede quando parliamo con l'AI



Analisi del segnale audio: Il segnale vocale viene scomposto in piccole finestre temporali

Trasformata di Fourier: Per ogni finestra, si calcola la trasformata di Fourier per ottenere lo spettro di frequenza.

Filtri Mel: Si applicano filtri basati sulla scala Mel, che simula la percezione umana delle frequenze, passando da frequenza in Hz a scala Mel

Cepstrum: per separare le componenti periodiche di un segnale (come le vibrazioni delle corde vocali) dalle sue caratteristiche spettrali più lente (legate alla forma del tratto vocale)

Diego Vendramini – Non copiare o riprodurre

ITALIANO e Fonologia trasparente

L'italiano ha una forte corrispondenza tra grafia e fonetica (cioè tra scrittura e pronuncia). Rispetto a lingue come l'inglese o il francese, che hanno molte eccezioni e regole di pronuncia complesse, in italiano la relazione tra la forma scritta e quella parlata è più diretta.

Ciò semplifica sia i compiti di riconoscimento vocale che di elaborazione testuale.

Diego Vendramini – Non copiare o riprodurre

Chat GPT

Esempio:

Token: " *Come vado in Cina?*"

In questo esempio, i token sono 5: **Come** **vado** **in** **Cina** **?**

Ogni valore, nel vettore, rappresenta una peso associato al relativo token. Ad esempio un tono interrogativo, un riferimento geografico, una collocazione tipica in un periodo ecc…

La Conversione in Vettori Numerici

Un modello di intelligenza artificiale non può comprendere parole direttamente come esseri umani.

Il testo deve essere convertito in una rappresentazione numerica.

Questo avviene attraverso un processo noto come *embedding*, che associa a ogni "token" un vettore numerico di dimensioni elevate (*high-dimensional vector*).

Cos'è un Vettore di Alto Dimensionamento?
Un vettore è essenzialmente una lista di numeri.
Esempio: vettore a 3 dimensioni = [1.5, −2.3, 0.8].
I modelli linguistici avanzati le dimensioni sono centinaia o migliaia(512, 768, 1024, ecc.)

PERCHE?

Per "catturare" le sottili relazioni tra parole o frammenti di parole e collegare *caratteristiche semantiche* analoghe.

Come Funziona l’Embedding?

E una tecnica per rappresentare il significato delle parole come vettori.

L'addestramento serve per insegnare al modello relazioni semantiche tra parole.

Parole simili nel significato avranno vettori simili ed essere rappresentate geometricamente attravaerso la vicinanza o la direzione di questi vettori.

ESEMPI:

- vettori per "gatto" e "felino" saranno vicini tra loro in questo spazio ad alta dimensione
- sinonimie
- analogia (es. "re" è a "regina" come "uomo" è a "donna")

Tokenizzazione della frase
Prima di tutto, la frase "avere un diavolo per capello"
Per semplicità di analisi usiamo un modello di tokenizzazione basato su singole parole e ipotizzimao l'uso di sole 4 dimensioni

["avere", "un", "diavolo", "per", "capello"]

Matrice di Embedding con indice posizionale

Token	Indice Posizionale	Embedding (vettore)
avere	1	[0.2, 0.1, 0.4, -0.3]
un	2	[0.3, 0.5, 0.2, 0.1]
diavolo	3	[0.6, -0.2, 0.8, 0.4]
per	4	[0.1, 0.3, -0.4, 0.5]
capello	5	[0.4, -0.1, 0.2, 0.3]

$$\text{"avere un diavolo per capello"} \rightarrow \begin{bmatrix} 0.2 & 0.1 & 0.4 & -0.3 \\ 0.3 & 0.5 & 0.2 & 0.1 \\ 0.6 & -0.2 & 0.8 & 0.4 \\ 0.1 & 0.3 & -0.4 & 0.5 \\ 0.4 & -0.1 & 0.2 & 0.3 \end{bmatrix}$$

Large Language Models – Chat GPT

Avere un diavolo per …

amico

capello

coperchio

CHAT GPT

Il più famoso LLM

Generative

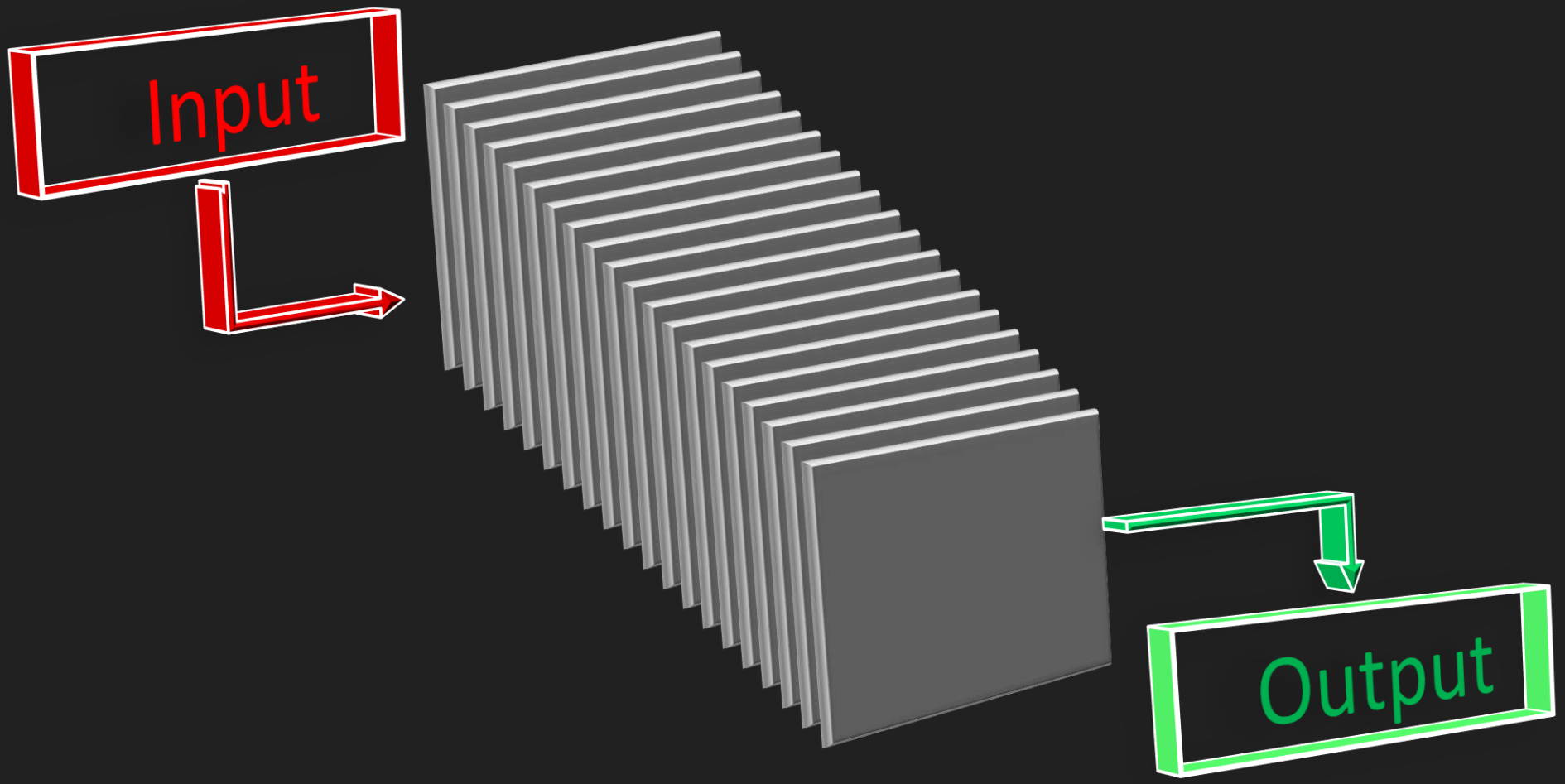
Pretrained

Transformer

LLM Formula

P (*capello / avere, un, diavolo, per*)

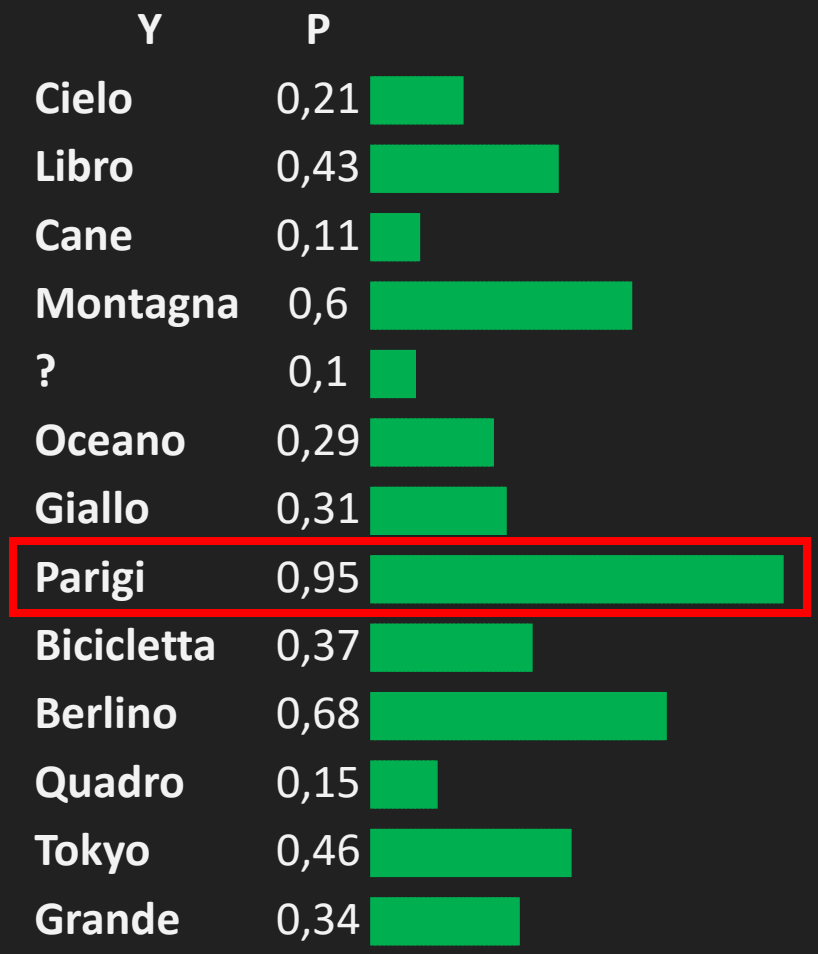
Diego Vendramini – Non copiare o riprodurre


Input
Output

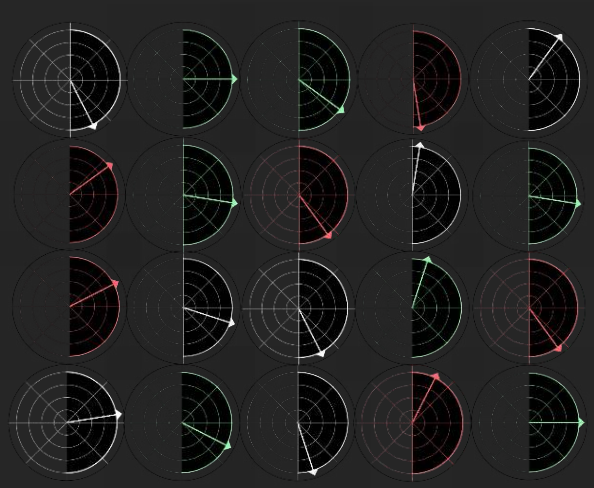
Qual è la città più grande della Francia?

La città più grande della Francia, nonché capitale dello Stato, è ->

La città più grande della Francia, nonché capitale dello Stato,
è_____

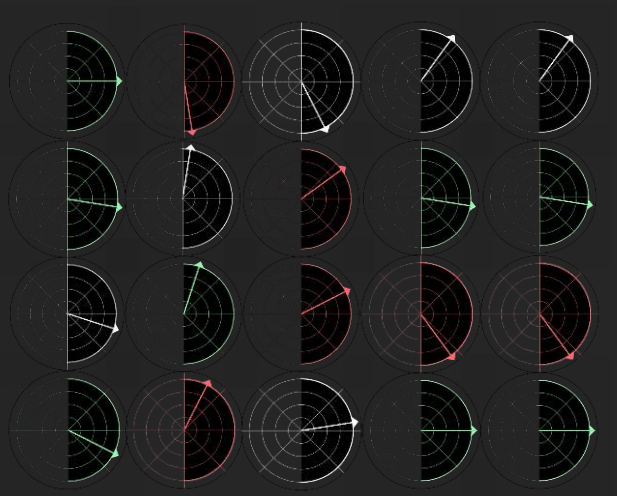


La città più grande della Francia, nonché capitale dello Stato, è_____



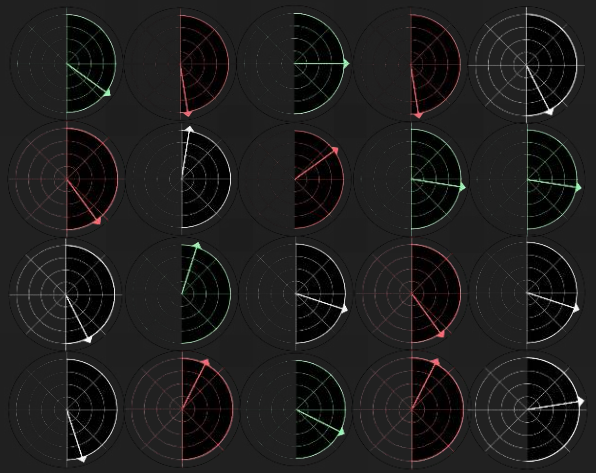
Y	P
Cielo	0,07
Libro	0,48
Cane	0,55
Montagna	0,21
?	0,36
Oceano	0,78
Giallo	0,53
Parigi	0,25
Bicicletta	0,01
Berlino	0,68
Quadro	0,5
Tokyo	0,72
Grande	0,64

La città più grande della Francia, nonché capitale dello Stato, è_____



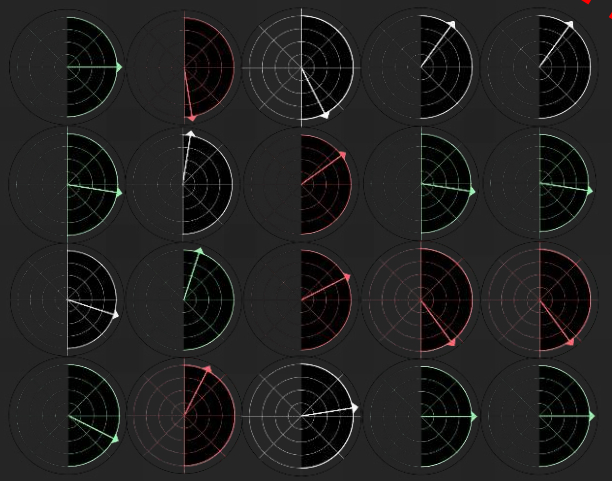
Y	P
Cielo	0,46
Libro	0,43
Cane	0,11
Montagna	0,81
?	0,84
Oceano	0,49
Giallo	0,29
Parigi	0,83
Bicicletta	0,79
Berlino	0,08
Quadro	0,44
Tokyo	0,82
Grande	0,08

La città più grande della Francia, nonché capitale dello Stato, è_____



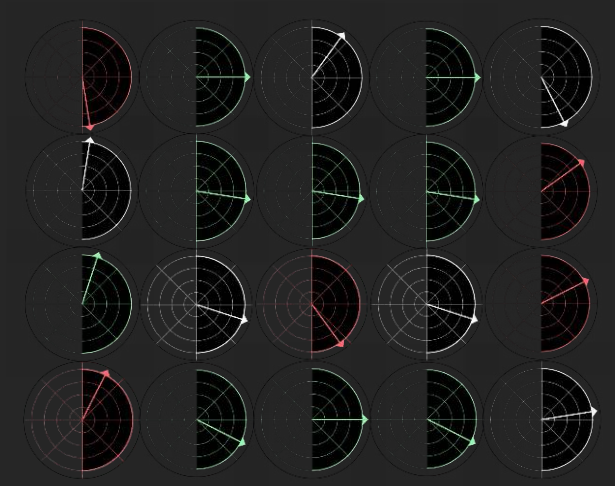
Y	P
Cielo	0,85
Libro	0,4
Cane	0,82
Montagna	0,41
?	0,08
Oceano	0,65
Giallo	0,46
Parigi	0,82
Bicicletta	0,34
Berlino	0,13
Quadro	0,58
Tokyo	0,76
Grande	0,91

La città più grande della Francia, nonché capitale dello Stato, è___ ***Parigi***



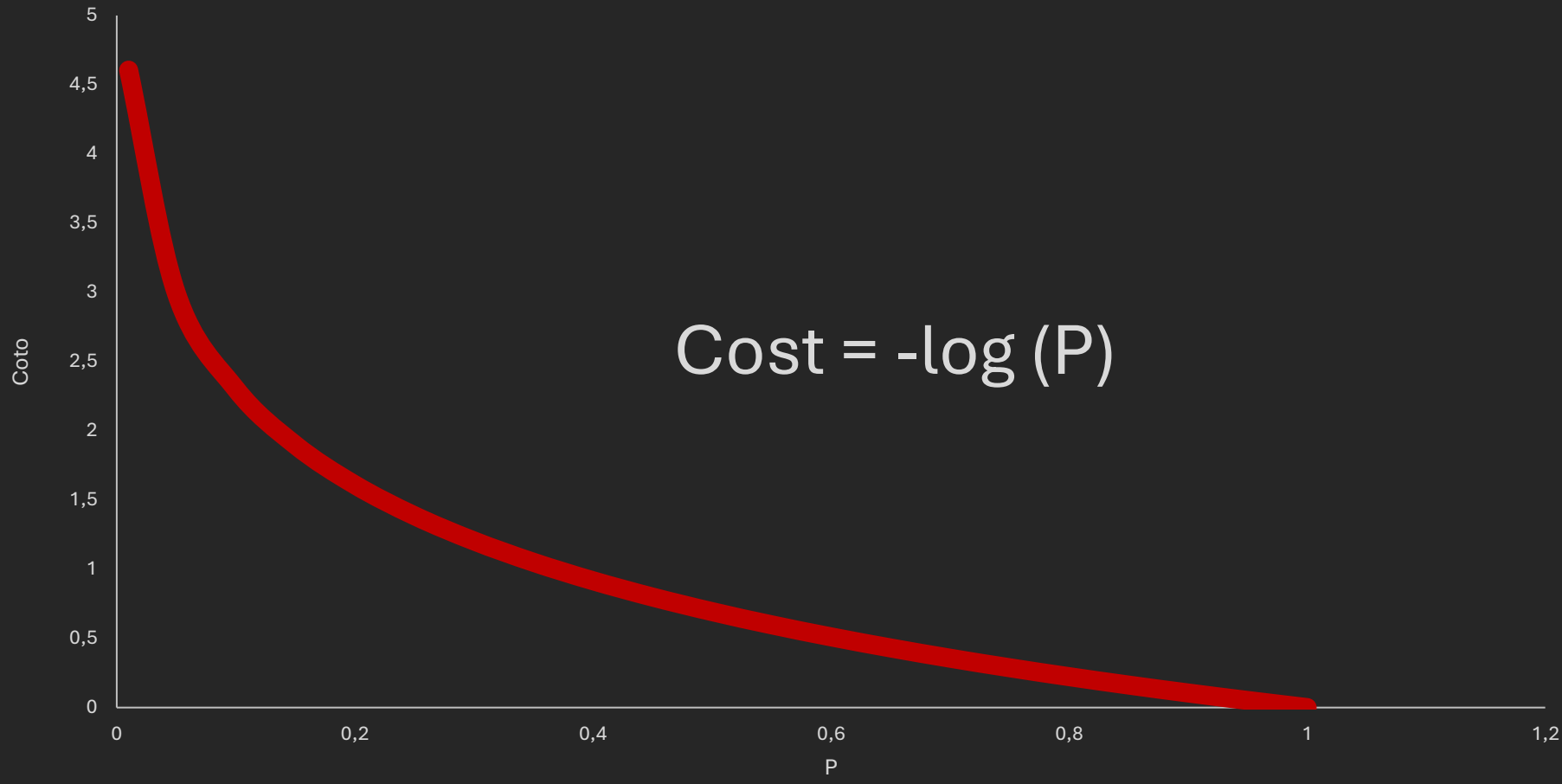
Y	P
Cielo	0,21
Libro	0,43
Cane	0,11
Montagna	0,6
?	0,1
Oceano	0,29
Giallo	0,31
Parigi	0,95
Bicicletta	0,37
Berlino	0,68
Quadro	0,15
Tokyo	0,46
Grande	0,34

La città più grande della Francia, nonché capitale dello Stato, è_____



Y	P
Cielo	0,21
Libro	0,43
Cane	0,11
Montagna	0,6
?	0,1
Oceano	0,29
Giallo	0,31
Parigi	0,95
Bicicletta	0,37
Berlino	0,68
Quadro	0,15
Tokyo	0,46
Grande	0,34

Funizone di Perdita



Chat GPT - Generative Pre-trained Transformer

Tipo di rete neurale	Caratteristiche
Perceptron a strato singolo	Il tipo più semplice di rete neurale, incapace di apprendere modelli complessi.
Perceptron multistrato (MLP)	Ha uno o più strati nascosti tra input e output, in grado di apprendere modelli non lineari.
Rete neurale convoluzionale (CNN)	Specializzata nell'elaborazione di immagini, utilizza la convoluzione per identificare le caratteristiche.
Rete neurale ricorrente (RNN)	Progettata per l'elaborazione di sequenze di dati, mantiene informazioni da input precedenti.
Long Short-Term Memory (LSTM)	Un tipo speciale di RNN, progettata per apprendere dipendenze a lungo termine nelle sequenze.
Gated Recurrent Unit (GRU)	Simile all'LSTM, ma con una struttura più semplice, progettata anch'essa per dipendenze a lungo termine.
Autoencoder	Utilizzata per l'apprendimento non supervisionato, codifica l'input in una rappresentazione compressa e poi lo decodifica.
Transformer	**Usa un meccanismo di "attenzione" per pesare l'importanza relativa di ogni elemento in una sequenza, efficace per il LLM.** **Attenzione maggiore alle dipendenze a lungo termine = meno Dimenticanza.**
Generative Adversarial Network (GAN)	Composta da due reti che lavorano in competizione, una genera nuovi dati, l'altra cerca di distinguere i dati generati dai dati reali.

Multihead Self - Attention

Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com
Noam Shazeer* Google Brain noam@google.com
Niki Parmar* Google Research nikip@google.com
Jakob Uszkoreit* Google Research usz@google.com
Llion Jones* Google Research llion@google.com
Aidan N. Gomez*† University of Toronto aidan@cs.toronto.edu
Łukasz Kaiser* Google Brain lukaszkaiser@google.com
Illia Polosukhin* illia.polosukhin@gmail.com

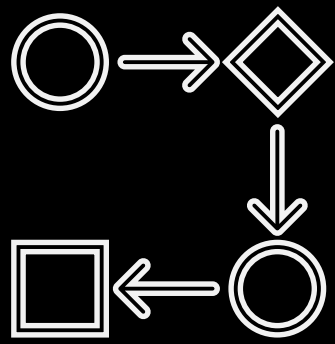
.03762v1 [cs.CL] 12 Jun 2017

Abstract

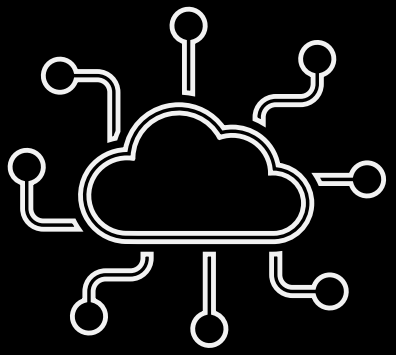
The dominant sequence transduction models are based on complex recurrent or convolutional neural networks in an encoder-decoder configuration. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles by over 2 BLEU. On the WMT 2014 English-to-French translation task,

Funzionamento

Il Trasformer impara a mescolare i contesti delle parole (dei token) in un modo che lo rende davvero bravo a indovinare la parola successiva.



Gli LLM sono addestrati su enormi quantità di informazioni raccolte da internet (libri, blog, siti di notizie, articoli di Wikipedia, discussioni su Reddit, social media).



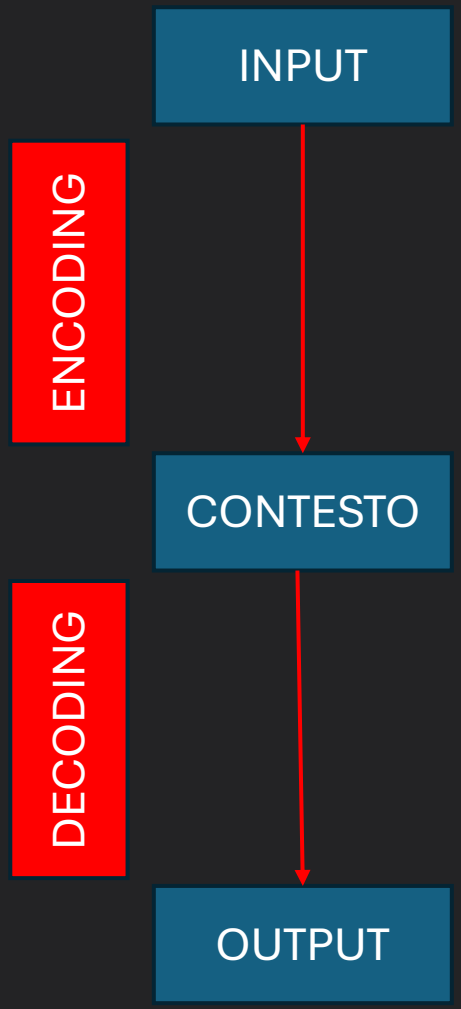
Diego Vendramini – Non copiare o riprodurre

ENCODER

Elabora la sequenza di input e la comprime in una rappresentazione di lunghezza fissa nota come vettore di contesto.

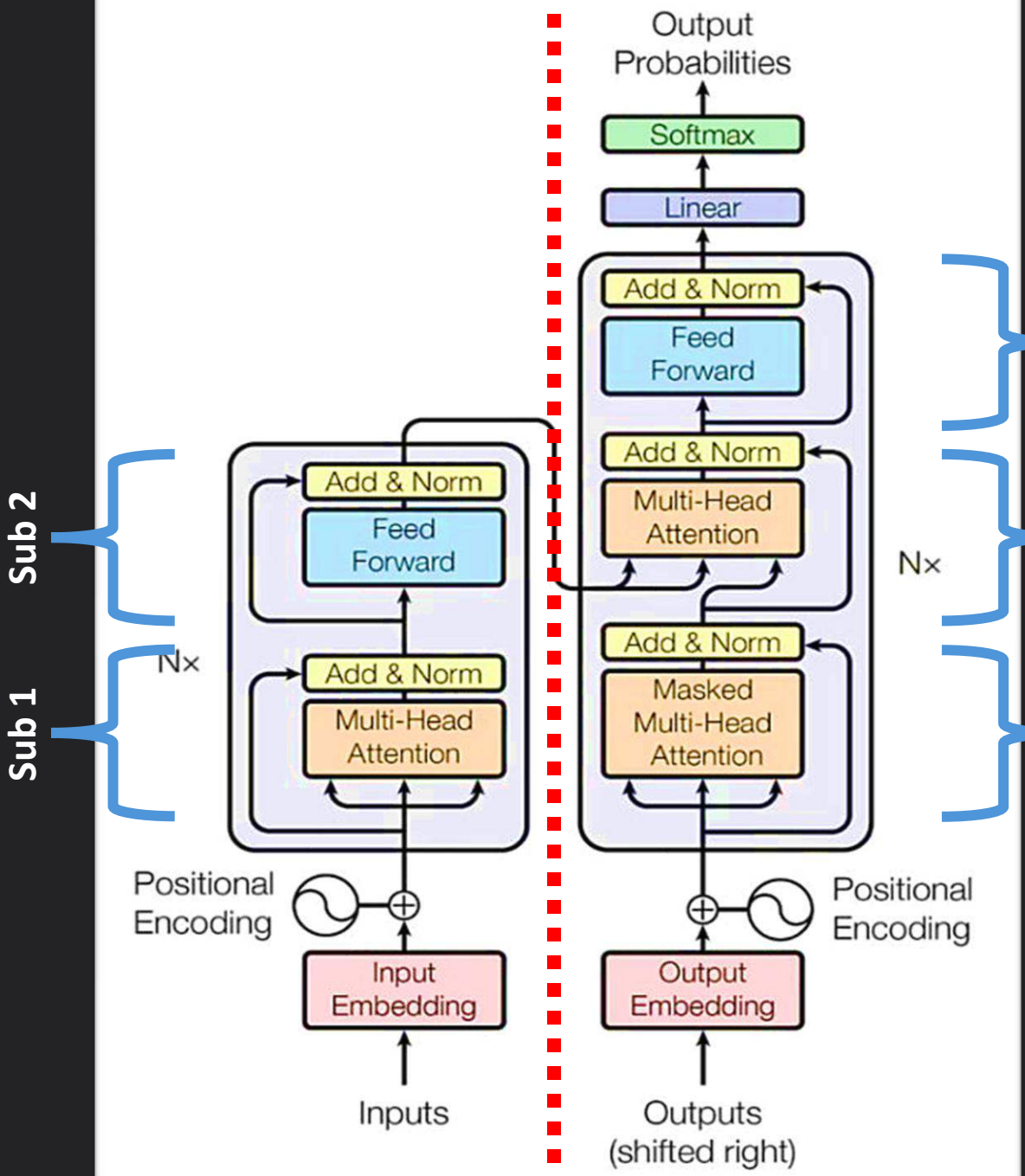
DECODER

Prende questo vettore di contesto come input e genera la sequenza di output .
Il vettore di contesto funge da riassunto delle informazioni essenziali contenute nell'input .



ENCODER

DECODER



Come funziona?

1. Embedding e Positional Encoding

Gli input, tipicamente token, vengono convertiti in vettori di embedding. A questi embedding viene aggiunta un'informazione di posizionamento (Positional Encoding) per mantenere l'ordine delle parole.

2. Calcolo di Q, K, V

Ogni embedding arricchito viene trasformato in tre nuovi vettori: Query (Q), Key (K) e Value (V) attraverso delle trasformazioni lineari. Queste proiezioni servono a calcolare quanto ogni parola deve **prestare attenzione** alle altre parole nella sequenza.

3. Meccanismo di Attenzione + Altri Passaggi

Il prodotto scalare tra **Q**uery e **K**ey determina i punteggi di attenzione, che vengono poi scalati e normalizzati tramite una funzione Softmax per ottenere i pesi di attenzione. Questi pesi vengono applicati ai **V**alue per ottenere una combinazione pesata che rappresenta la probabilità della relazione.

4. Feed-Forward e Output Finale

L'output del meccanismo di attenzione passa attraverso una rete ed è ripetuto per diversi strati del Transformer.

Alla fine, l'output viene utilizzato per fare previsioni, ad esempio, la prossima parola /token in una sequenza.

Diego Vendramini – Non copiare o riprodurre

cosa completa la frase avere un diavolo per...?

La frase completa è **"avere un diavolo per capello"**.

È un modo di dire italiano che significa essere estremamente arrabbiati, irritati o nervosi. La metafora suggerisce l'immagine di una persona talmente infuriata che ogni suo capello sembra essere animato, come se avesse un diavolo sopra.

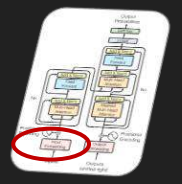
Esempio:

- "Non disturbare Marco, oggi ha un diavolo per capello dopo quella riunione disastrosa."

Tokenizzazione

La frase viene segmentata in una sequenza di token.

Token 1: "avere"
Token 2: "un"
Token 3: "diavolo"
Token 4: "per"
Token 5: "capello"



Lookup nella Matrice di Embedding e Positional Encoding

Ad ogni token viene associato un vettore continuo tramite una matrice degli embedding *e*

(dimensione dmodel=4.)

avere

$$e_1 = \begin{bmatrix} 0.20 \\ -0.10 \\ 0.50 \\ 0.30 \end{bmatrix}$$

un

$$e_2 = \begin{bmatrix} 0.00 \\ 0.10 \\ -0.20 \\ 0.40 \end{bmatrix}$$

diavolo

$$e_3 = \begin{bmatrix} 0.30 \\ 0.60 \\ -0.10 \\ 0.20 \end{bmatrix}$$

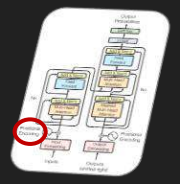
per

$$e_4 = \begin{bmatrix} -0.30 \\ 0.20 \\ 0.40 \\ -0.10 \end{bmatrix}$$

capello

$$e_5 = \begin{bmatrix} 0.10 \\ -0.20 \\ 0.30 \\ 0.50 \end{bmatrix}$$

Calcolo del Positional Encoding (PE)



Per ogni token si calcola un vettore di positional encoding per includere l'informazione sulla posizione.

- Nei Transformer l'input è processato in parallelo -> l'ordine dei token non è intrinsecamente mantenuto.
- L'ordine delle parole è però cruciale per comprendere il significato di una frase
- Il PE aggiunge informazioni sulla posizione di ciascun token nella sequenza aggiungendo un vettore di posizione all'embedding di ciascun token.
- Il vettore di posizione è calcolato utilizzando funzioni sinusoidali, che permettono di codificare la posizione in modo continuo e differenziabile ideale per la backpropagation durante l'ottimizzazione.

$$PE(pos, 2i) = \sin\left(\frac{pos}{10000^{2i/d}}\right)$$

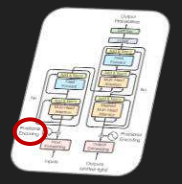
$$PE(pos, 2i+1) = \cos\left(\frac{pos}{10000^{2i/d}}\right)$$

pos è la posizione del token nella sequenza.

i è l'indice della dimensione dell'embedding.

d è la dimensione totale dell'embedding.

i scorre da 0 a $\frac{d}{2} - 1$



Con $d_{\text{model}} = 4$ e usando un'approssimazione, poniamo per semplicità:

- **Per $i = 0$:**
  $$\text{PE}(pos, 0) = \sin(pos)$$
  $$\text{PE}(pos, 1) = \cos(pos)$$
- **Per $i = 1$:**
  $$\text{PE}(pos, 2) = \sin\left(\frac{pos}{100}\right)$$
  $$\text{PE}(pos, 3) = \cos\left(\frac{pos}{100}\right)$$

Calcoliamo per ogni posizione (usando posizioni 1, 2, ..., 5):

Posizione 1:

$$\text{PE}(1) \approx \begin{bmatrix} \sin(1) \\ \cos(1) \\ \sin(0.01) \\ \cos(0.01) \end{bmatrix} \approx \begin{bmatrix} 0.8415 \\ 0.5403 \\ 0.0100 \\ 0.99995 \end{bmatrix}$$

Posizione 2:

$$\text{PE}(2) \approx \begin{bmatrix} \sin(2) \\ \cos(2) \\ \sin(0.02) \\ \cos(0.02) \end{bmatrix} \approx \begin{bmatrix} 0.9093 \\ -0.4161 \\ 0.0200 \\ 0.99980 \end{bmatrix}$$

Posizione 3:

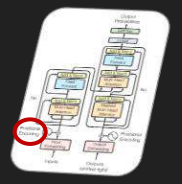
$$\text{PE}(3) \approx \begin{bmatrix} \sin(3) \\ \cos(3) \\ \sin(0.03) \\ \cos(0.03) \end{bmatrix} \approx \begin{bmatrix} 0.1411 \\ -0.9900 \\ 0.0300 \\ 0.99955 \end{bmatrix}$$

Posizione 4:

$$\text{PE}(4) \approx \begin{bmatrix} \sin(4) \\ \cos(4) \\ \sin(0.04) \\ \cos(0.04) \end{bmatrix} \approx \begin{bmatrix} -0.7568 \\ -0.6536 \\ 0.0400 \\ 0.99920 \end{bmatrix}$$

Posizione 5:

$$\text{PE}(5) \approx \begin{bmatrix} \sin(5) \\ \cos(5) \\ \sin(0.05) \\ \cos(0.05) \end{bmatrix} \approx \begin{bmatrix} -0.9589 \\ 0.2837 \\ 0.0500 \\ 0.99875 \end{bmatrix}$$



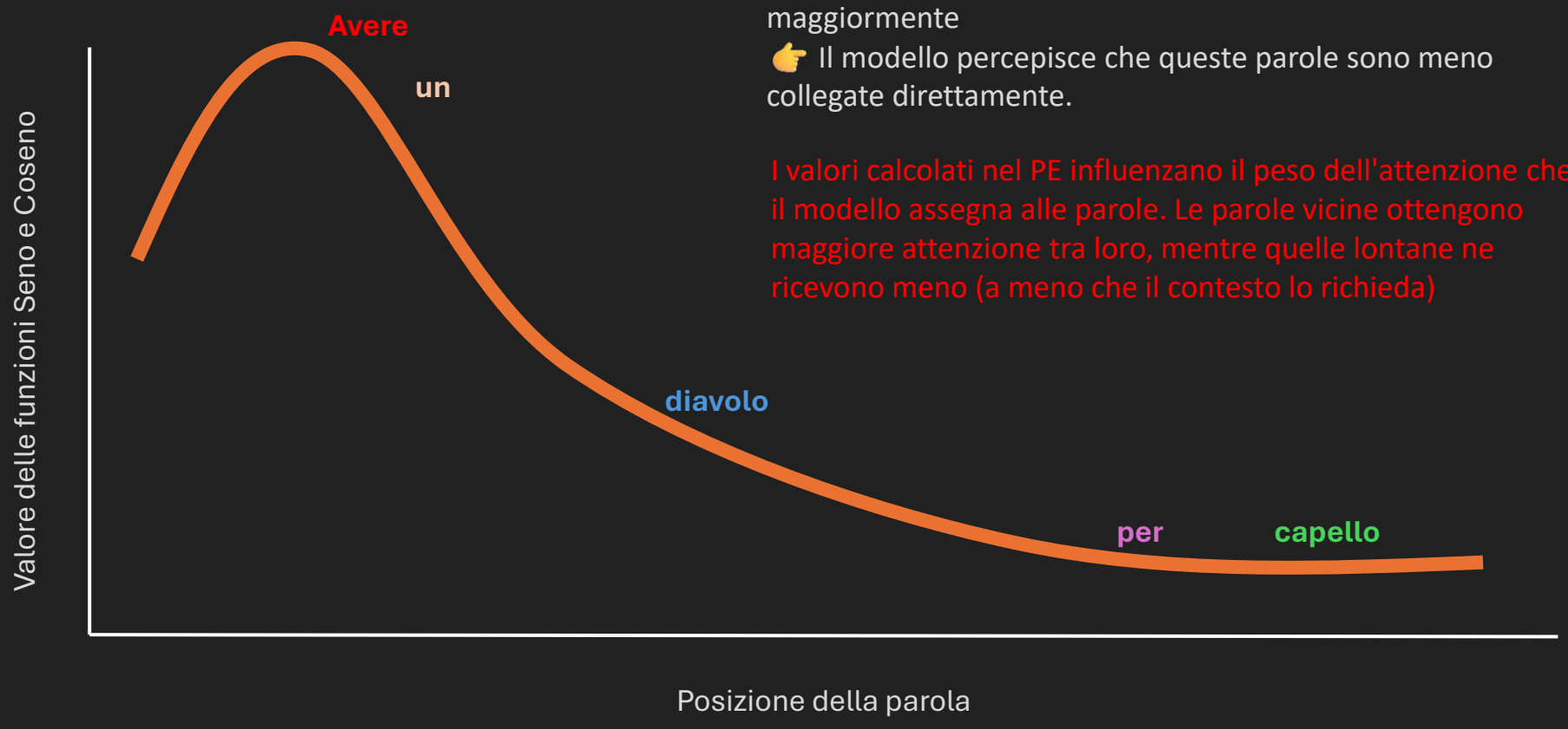
avere $\longrightarrow X_0^{(1)} = e_1 + \mathrm{PE}(1) = \begin{bmatrix} 0.20 \\ -0.10 \\ 0.50 \\ 0.30 \end{bmatrix} + \begin{bmatrix} 0.8415 \\ 0.5403 \\ 0.0100 \\ 0.99995 \end{bmatrix} \approx \begin{bmatrix} 1.0415 \\ 0.4403 \\ 0.5100 \\ 1.29995 \end{bmatrix}$

un $\longrightarrow X_0^{(2)} = e_2 + \mathrm{PE}(2) = \begin{bmatrix} 0.00 \\ 0.10 \\ -0.20 \\ 0.40 \end{bmatrix} + \begin{bmatrix} 0.9093 \\ -0.4161 \\ 0.0200 \\ 0.99980 \end{bmatrix} \approx \begin{bmatrix} 0.9093 \\ -0.3161 \\ -0.1800 \\ 1.39980 \end{bmatrix}$

diavolo $\longrightarrow X_0^{(4)} = e_4 + \mathrm{PE}(4) = \begin{bmatrix} -0.30 \\ 0.20 \\ 0.40 \\ -0.10 \end{bmatrix} + \begin{bmatrix} -0.7568 \\ -0.6536 \\ 0.0400 \\ 0.99920 \end{bmatrix} \approx \begin{bmatrix} -1.0568 \\ -0.4536 \\ 0.4400 \\ 0.8992 \end{bmatrix}$

per $\longrightarrow X_0^{(3)} = e_3 + \mathrm{PE}(3) = \begin{bmatrix} 0.30 \\ 0.60 \\ -0.10 \\ 0.20 \end{bmatrix} + \begin{bmatrix} 0.1411 \\ -0.9900 \\ 0.0300 \\ 0.99955 \end{bmatrix} \approx \begin{bmatrix} 0.4411 \\ -0.3900 \\ -0.0700 \\ 1.19955 \end{bmatrix}$

capello $\longrightarrow X_0^{(5)} = e_5 + \mathrm{PE}(5) = \begin{bmatrix} 0.10 \\ -0.20 \\ 0.30 \\ 0.50 \end{bmatrix} + \begin{bmatrix} -0.9589 \\ 0.2837 \\ 0.0500 \\ 0.99875 \end{bmatrix} \approx \begin{bmatrix} -0.8589 \\ 0.0837 \\ 0.3500 \\ 1.49875 \end{bmatrix}$



Parole vicine (es. "avere" e "un"):
Le loro rappresentazioni (sin e cos) sono molto simili. La distanza tra i valori è minima.
👉 Il modello capisce che queste parole sono strettamente collegate.

Parole lontane (es. "avere" e "cappello"):
Le loro rappresentazioni (sin e cos) differiscono maggiormente
👉 Il modello percepisce che queste parole sono meno collegate direttamente.

I valori calcolati nel PE influenzano il peso dell'attenzione che il modello assegna alle parole. Le parole vicine ottengono maggiore attenzione tra loro, mentre quelle lontane ne ricevono meno (a meno che il contesto lo richieda)

Self Attention

Proviamo a spiegarlo…

La self attention “spiegata semplice”

Marco va in biblioteca a cercare un libro sull’intelligenza artificiale: **Query (K)**

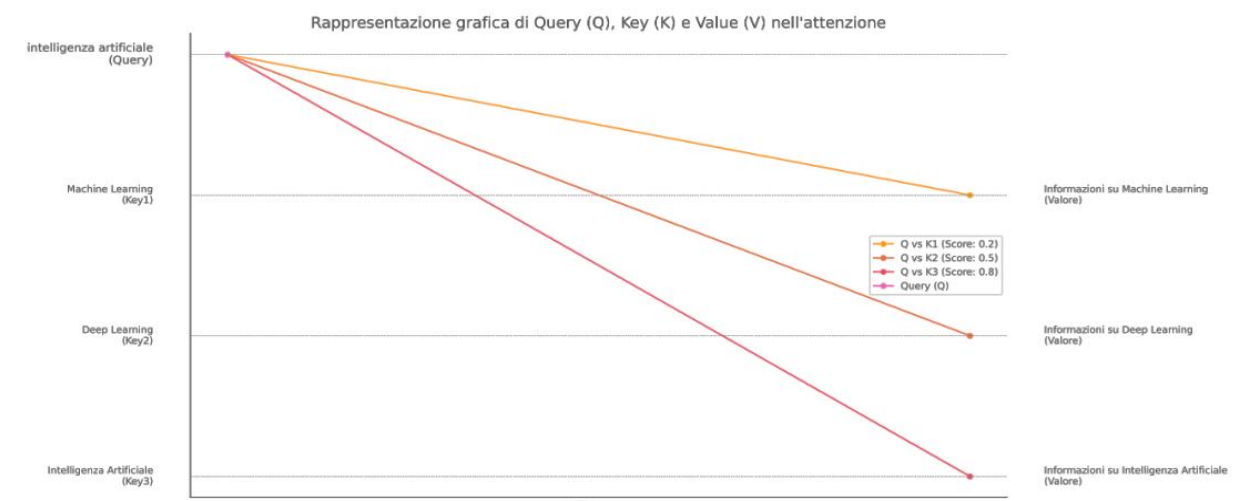
Nel contesto dei Transformer, Query rappresenta quello che stiamo cercando (**token che compongono le parole**)

Nella biblioteca, ogni libro ha un titolo e delle parole chiave che descrivono il suo contenuto. La Bibliotecaria fa una ricerca per i libri più corretti attraverso le parole chiave: **Key (K)**

Nel modello, il Key rappresenta l'indice o l'identificatore di contenuti importanti (**valore numerico da ponderare**

Ogni libro nella biblioteca contiene le informazioni vere e proprie, cioè il contenuto completo del libro: **Value (V)** Nel modello, il Value rappresenta l'informazione o il contenuto che si desidera recuperare (**parola corretta**).

Relazioni Q K e V



Query (Q) "intelligenza artificiale" è rappresentato sulla sinistra come la richiesta iniziale.

Key (K): Sono le parole chiave associate a ciascun libro (K1, K2, K3)

K1: “Applicazioni di Machine Learning e algoritmi decisionali"

K2: “Reti neurali e Deep Learning –Principi di matematica applicata"

K3: “Fondamenti di Intelligenza Artificiale"

Punteggi di Attenzione

Q vs K1: 0.2 Basso

Q vs K2: 0.5 Medio

Q vs K3: 0.8 Alto – Altamente rilevante per la funzione di attenzione

Value (V) Sono le informazioni contenute in ciascun libro:

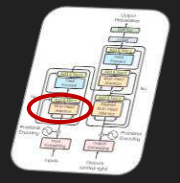
V1: "Informazioni su Machine Learning"

V2: "Informazioni su Deep Learning"

V3: "Informazioni su Intelligenza Artificiale"

Diego Vendramini – Non copiare o riprodurre

Multi Head Attention



L’input $X0^{(n)}$ viene elaborato da più strati del Transformer.

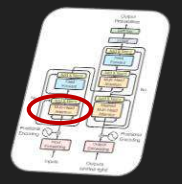
Per ogni token, si calcolano:

$$Q = X_0 W^Q, \quad K = X_0 W^K, \quad V = X_0 W^V$$

Qw, Qk, Qv sono matrici di dimensione 4x4 (nel nostro esempio) che possiedono valori appresi in fase di addestramento.

Supponiamo di avere le seguenti matrici di proiezione (valori ipotetici):

$$W_Q = \begin{bmatrix} 0.5 & 0.1 & 0.3 & -0.2 \\ 0.2 & 0.4 & -0.1 & 0.3 \\ -0.3 & 0.6 & 0.2 & 0.1 \\ 0.0 & -0.2 & 0.5 & 0.4 \end{bmatrix} \quad W_K = \begin{bmatrix} 0.4 & -0.1 & 0.2 & 0.0 \\ 0.3 & 0.5 & 0.1 & -0.2 \\ -0.2 & 0.4 & 0.3 & 0.1 \\ 0.1 & 0.0 & -0.3 & 0.6 \end{bmatrix} \quad W_V = \begin{bmatrix} 0.3 & 0.2 & -0.1 & 0.4 \\ -0.2 & 0.5 & 0.4 & 0.1 \\ 0.1 & -0.3 & 0.6 & 0.2 \\ 0.4 & 0.0 & 0.2 & -0.1 \end{bmatrix}$$



Avere

$$Q = X_0 W_Q$$

X_0 W_Q

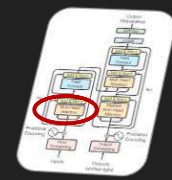
$$\begin{bmatrix} 1.0415 \\ 0.4403 \\ 0.5100 \\ 1.29995 \end{bmatrix} \times \begin{bmatrix} 0.5 & 0.1 & 0.3 & -0.2 \\ 0.2 & 0.4 & -0.1 & 0.3 \\ -0.3 & 0.6 & 0.2 & 0.1 \\ 0.0 & -0.2 & 0.5 & 0.4 \end{bmatrix}$$

$$Q_1^{(1)} = 1.0415 \times 0.5 + 0.4403 \times 0.2 + 0.5100 \times (-0.3) + 1.29995 \times 0.0.$$
$$Q_2^{(1)} = 1.0415 \times 0.1 + 0.4403 \times 0.4 + 0.5100 \times 0.6 + 1.29995 \times (-0.2).$$
$$Q_3^{(1)} = 1.0415 \times 0.3 + 0.4403 \times (-0.1) + 0.5100 \times 0.2 + 1.29995 \times 0.5.$$
$$Q_4^{(1)} = 1.0415 \times (-0.2) + 0.4403 \times 0.3 + 0.5100 \times 0.1 + 1.29995 \times 0.4.$$

$$Q^{(1)} \approx \begin{bmatrix} 0.45581 \\ 0.32628 \\ 1.02040 \\ 0.49577 \end{bmatrix}$$

Analogamente, si calcolano $K^{(1)} = X_0^{(1)} W_K$ e $V^{(1)} = X_0^{(1)} W_V$ usando le rispettive matrici.

Applicazione del Meccanismo di Attenzione



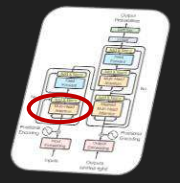
$$\text{Attention}(Q, K, V) = \underbrace{\overset{3}{\boxed{\text{softmax}}}\left(\frac{\overset{1}{\boxed{Q K^T}}}{\boxed{\sqrt{d_k}}}\right)}_{\text{pesi di attenzione}} \overset{4}{\boxed{\cdot \; V}}$$

2

Serve a determinare quanto ogni elemento di una sequenza di input sia rilevante per ogni altro elemento.

Permette al modello di pesare l'importanza di ciascun token in una sequenza rispetto agli altri, catturando dipendenze a lungo raggio e migliorando la capacità di comprendere il contesto.

$\sqrt{d_k}$



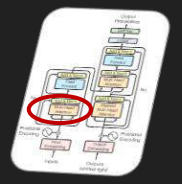
$\mathbf{d_k}$ è la dimensione dei vettori $\mathbf{K}$ e $\mathbf{Q}$

Perché dividere per $\sqrt{d_k}$

- Se $\mathbf{Q}$ e $\mathbf{K}$ hanno dimensione d_k, il loro prodotto scalare tende a crescere con d_k.
- Senza la divisione, i valori passati alla softmax sarebbero molto grandi e potrebbero **saturarla** (portando a gradienti molto piccoli).

> La softmax è sensibile a valori numerici grandi; se i punteggi non vengono "calmierati", la softmax può assegnare pesi quasi 1 a un singolo token e quasi 0 a tutti gli altri, perdendo sfumature.
>
> Dividendo per $\sqrt{d_k}$, si mantengono i punteggi in un intervallo più moderato, rendendo l'allenamento **più stabile**.

Softmax



La **Softmax** è una funzione che prende valori reali (punteggi dello Scaled Dot Product), siano essi positivi, negativi, grandi o piccoli, e li trasforma in una distribuzione di probabilità.

Serve a dare risposta alla domanda: “Quanto peso deve avere questo valore ”?

Rende più agevole imparare quali “punteggi” devono essere enfatizzati e quali ignorati.

La formula della softmax per un valore α_i corrispondente al punteggio z_i è:

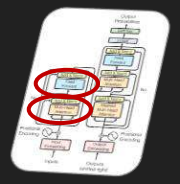
$$\alpha_i = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}}$$

e : numero di Nepero, Eulero, base della funzione esponenziale e^x, base del naturale $\log_e(x)$, limite della successione $(1+1/n)^n$

=

2,71828182845904523536

ADD & NORM



Migliorare il flusso dei gradienti attraverso la rete evitando il problema della scomparsa del gradiente.

Stabilizzare l'addestramento attraverso la normalizzazione.

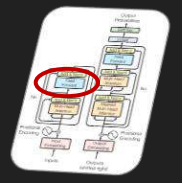
Facilitare l'ottimizzazione consentendo un apprendimento più stabile e veloce.

Questo modulo è composto da due operazioni principali:

Residual Connection (Add)

Layer Normalization (Norm)

1. Residual Connection (Add)



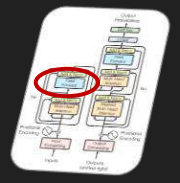
$$y = x + F(x)$$

x è l'input della trasformazione (ad esempio, l'output del livello precedente).

$F(x)$ è il risultato della trasformazione (Multi-Head Attention o Feed-Forward).

L'operazione $x + F(x)$ è un'**addizione elemento per elemento** (element-wise sum).

2. Layer Normalization (Norm)



$$\hat{y}_i = \frac{y_i - \mu}{\sigma} \cdot \gamma + \beta$$

y_i = Il **valore originale** dell'i-esima componente del vettore di input.

$\hat{y}_i$ = Il **valore normalizzato** (output della LayerNorm) per quella componente.

μ = La **media** dei valori nel vettore di input.

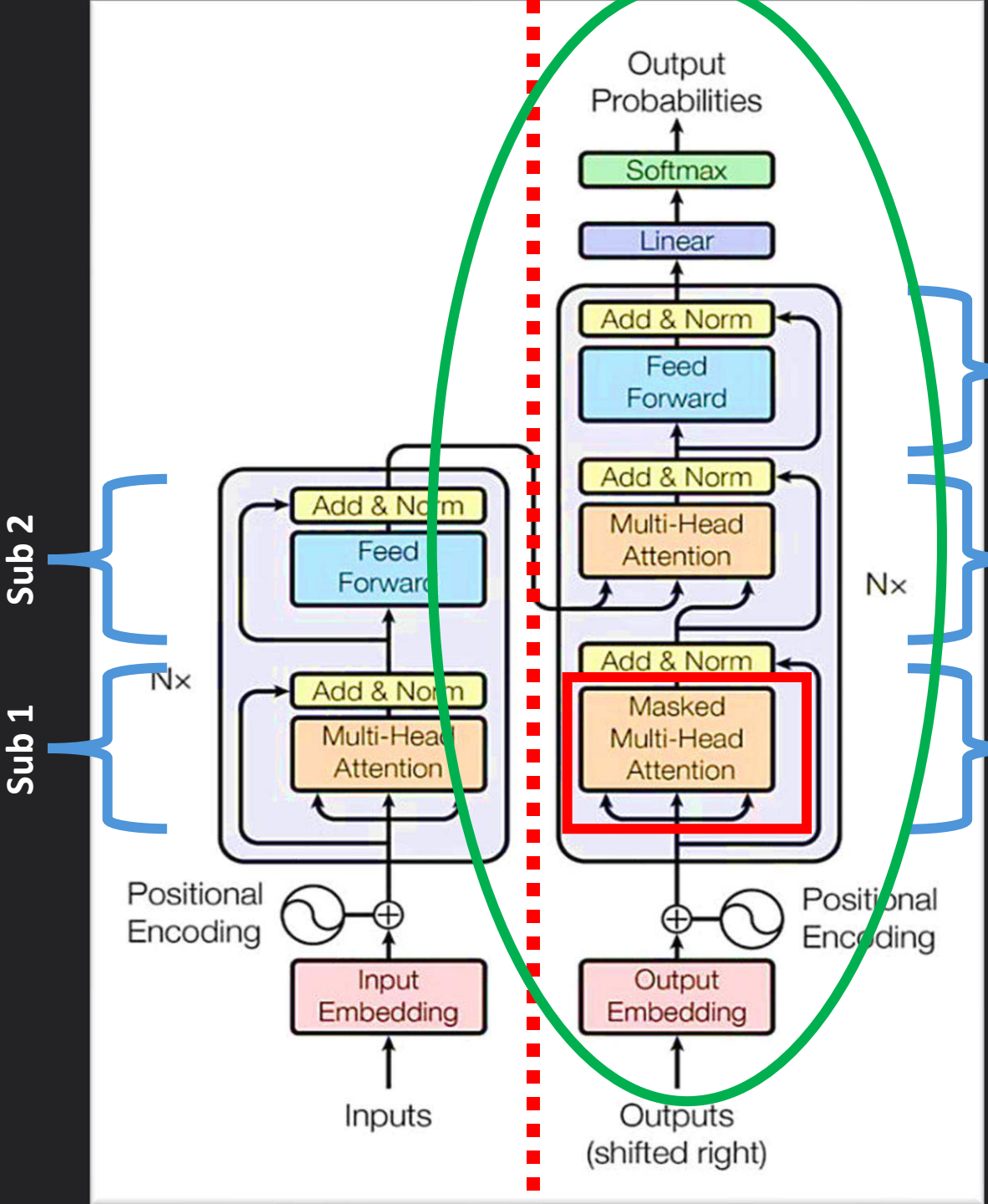
σ = La **deviazione standard** dei valori nel vettore di input.

γ = Un parametro che scala (allarga o restringe) la distribuzione normalizzata.

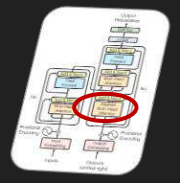
β = Un parametro che sposta (trasla) la distribuzione normalizzata.

ENCODER

DECODER



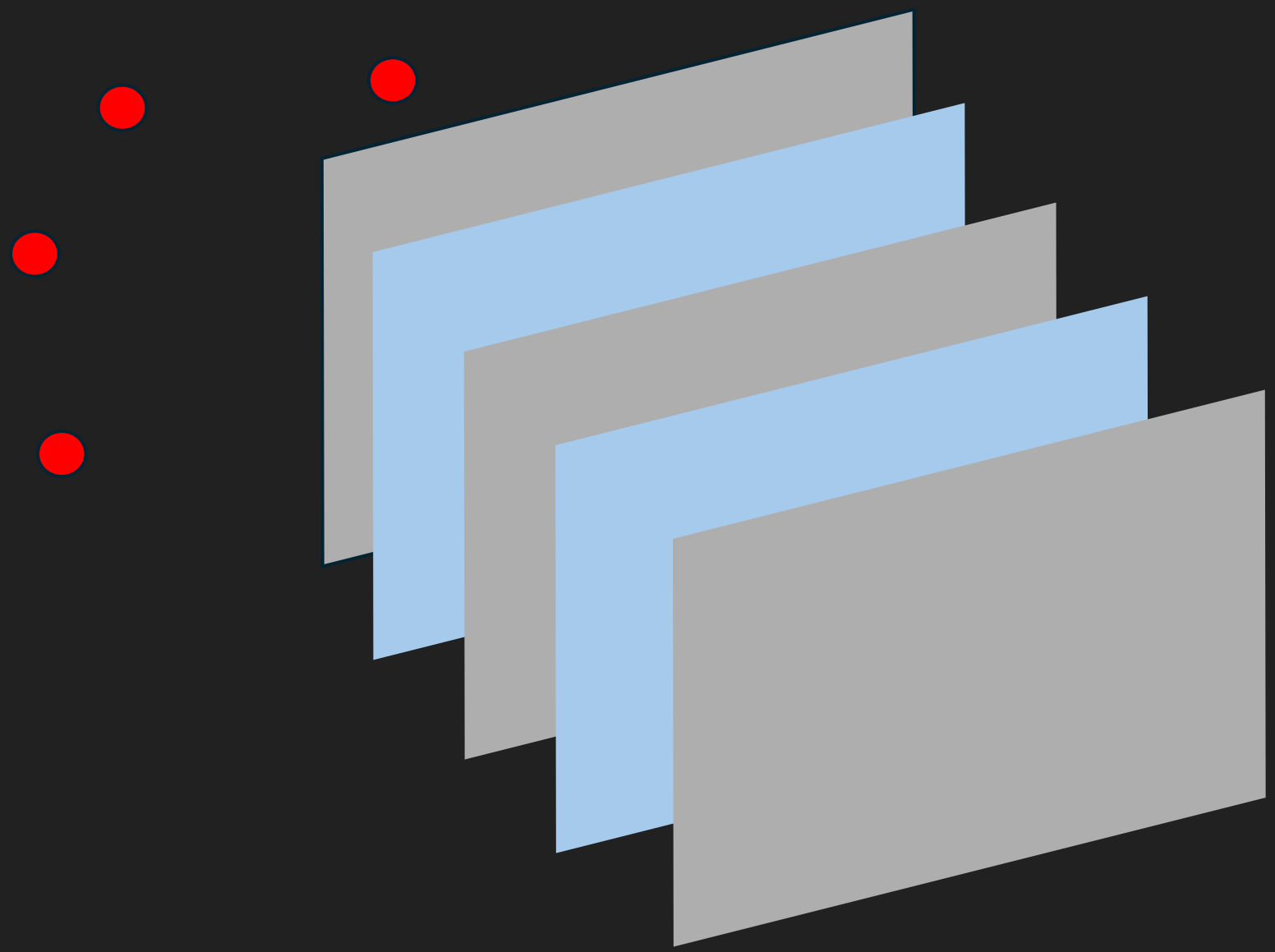
Masked Self - Attention



A differenza della self-attention nell'encoder, la masked self-attention impedisce al decoder di considerare i token futuri nella sequenza di output durante l'addestramento.

Questo viene fatto applicando una maschera che assicura che le previsioni per una data posizione dipendano solo dai token generati in precedenza.
Questo meccanismo è fondamentale per mantenere la proprietà autoregressiva del decoder, che imita il processo di generazione del testo parola per parola .

Token di Inizio: Il processo di decodifica inizia con un token speciale, spesso chiamato <START> o <SOS> (Start Of Sequence). Questo token viene fornito come input iniziale al decoder.



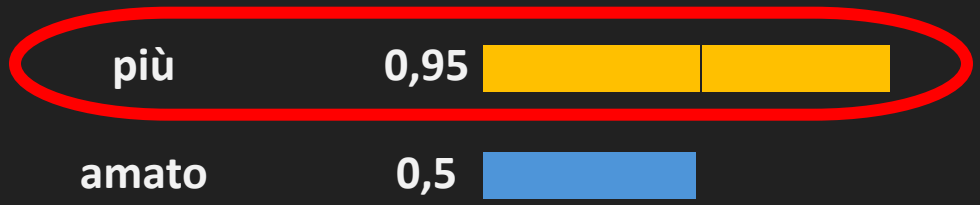
Ragionevole

ragionevole /radʒo'nevole/ agg. [rifacimento, secondo *ragione*, col suff. *-evole*, del lat. *rationabĭlis*]. - **1. a.** [che possiede la facoltà di ragionare: *l'uomo è un animale r.*] ≈ intelligente, pensante, raziocinante, razionale. ↔ irragionevole, irrazionale. **b.** [di persona, che si lascia guidare dalla ragione: *è stato molto r. nelle sue richieste*] ≈ coscienzioso, equilibrato, equo, misurato. ↔ esagerato, iniquo. **2.** (*estens.*) **a.** [di pensiero, discorso, comportamento, ecc., conforme alla ragione o al buon senso: *proposta r.*] ≈ assennato, equilibrato, saggio, sensato. ↔ dissennato, incosciente, insensato, irragionevole, irresponsabile, sconsiderato, (*lett.*) stolido, stolto. **b. [che si basa su giuste ragioni: *dubbi r.*] ≈ fondato, giustificato, lecito, legittimo, motivato, plausibile. c. [di giusta quantità o misura: *prezzi r.*] ≈ adeguato, congruo, conveniente, corretto, giusto, legittimo, onesto.**

Qual è la città più grande della Francia?

La città ->

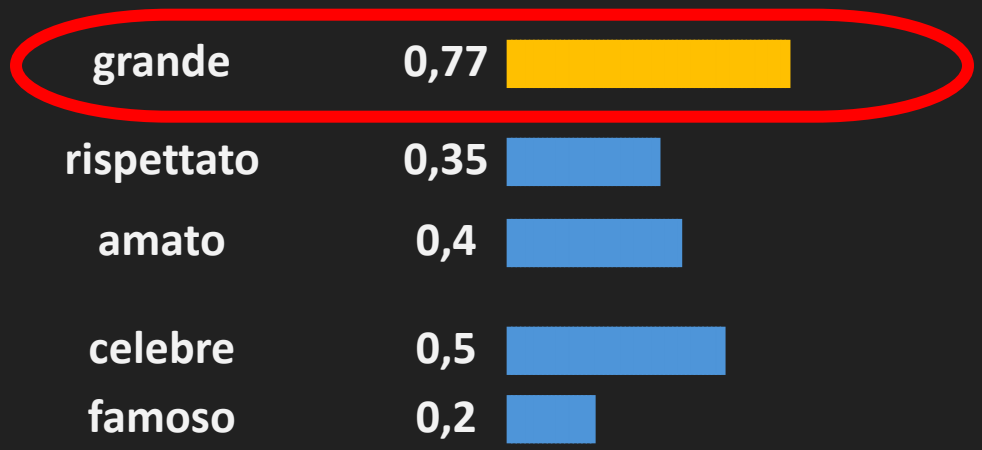
X	Y	P
La città	migliore	0,1
	famoso	0,25
	conosciuto	0,3
	rispettato	0,45
	più	0,95
	amato	0,5



Qual è la città più grande della Francia?

La città più ->

X	Y	P
più	importante	0,15
	grande	0,77
	rispettato	0,35
	amato	0,4
	celebre	0,5
	famoso	0,2



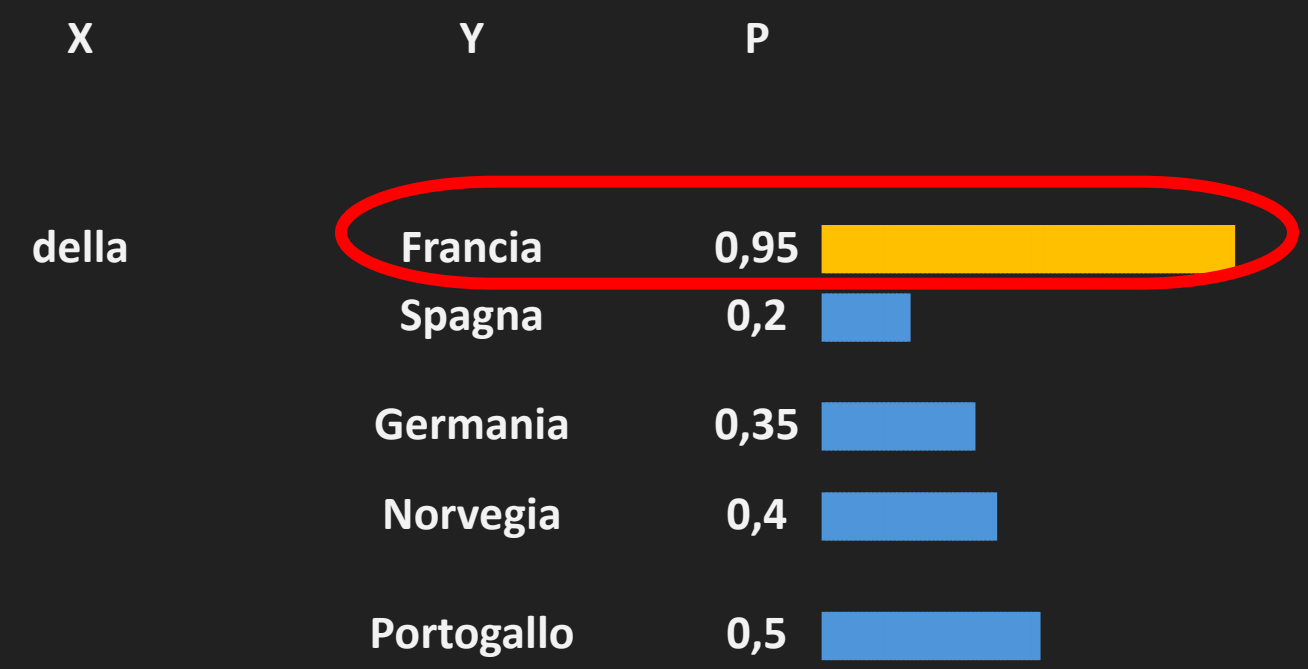
Qual è la città più grande della Francia?

La città più grande ->

X	Y	P
grande	in	0,1
	sul	0,25
	sotto	0,3
	vicino	0,45
	oltre	0,5
	della	0,95

Qual è la città più grande della Francia?

La città più grande della ->



Qual è la città più grande della Francia?

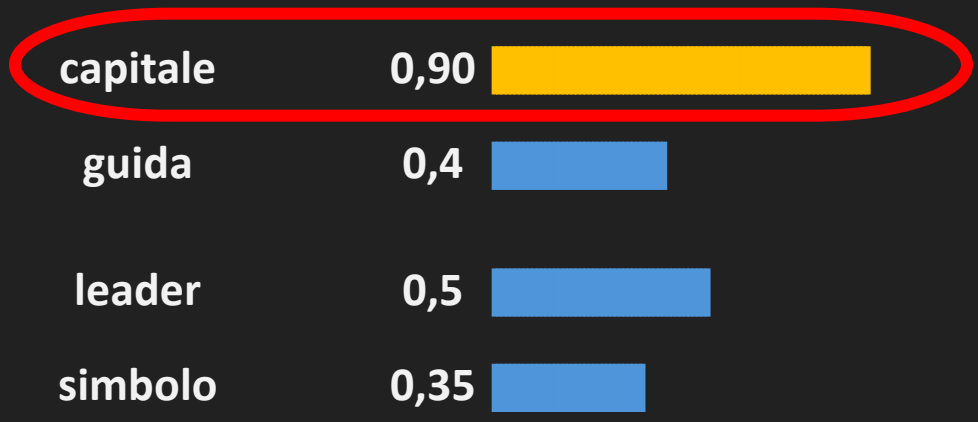
La città più grande della Francia, ->

X	*Y*	*P*
Francia,	oltre	0,1
	anche	0,25
	pure	0,3
	perfino	0,45
	addirittura	0,5
	nonché	0,75

Qual è la città più grande della Francia?

La città più grande della Francia, nonché ->

X	Y	P
nonché	centro	0,15
	cuore	0,2
	capitale	0,90
	guida	0,4
	leader	0,5
	simbolo	0,35



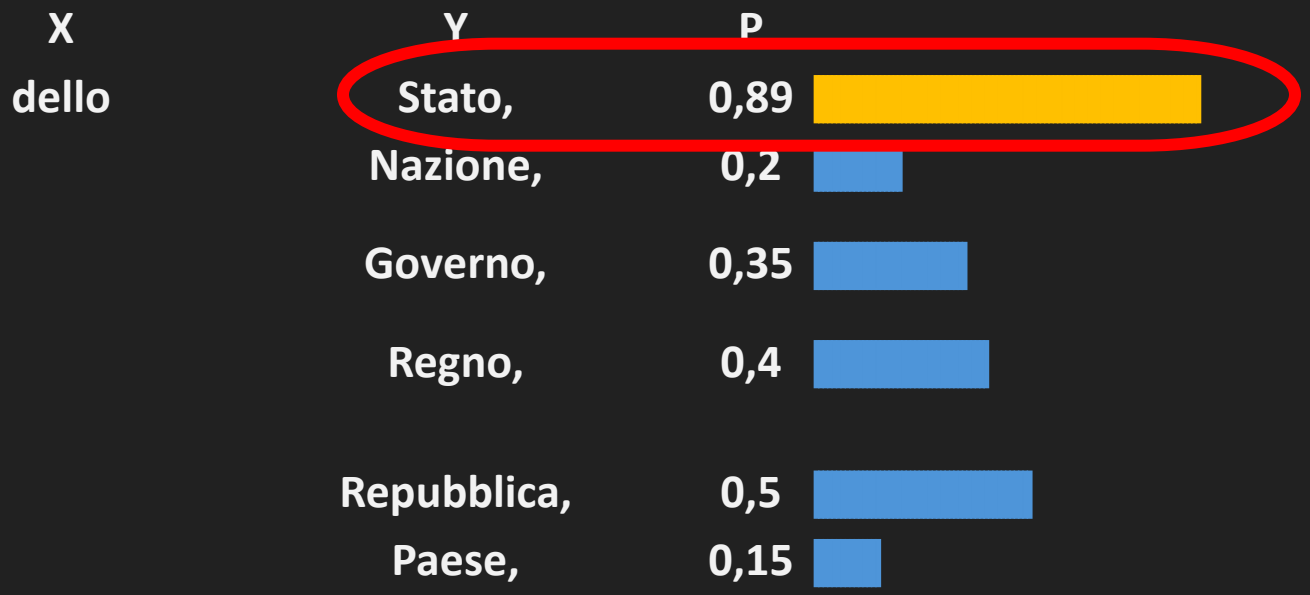
Qual è la città più grande della Francia?

La città più grande della Francia, nonché capitale ->

X	Y	P
capitale	sul	0,1
	nello	0,25
	vicino	0,3
	oltre	0,45
	verso	0,5
	dello	0,78

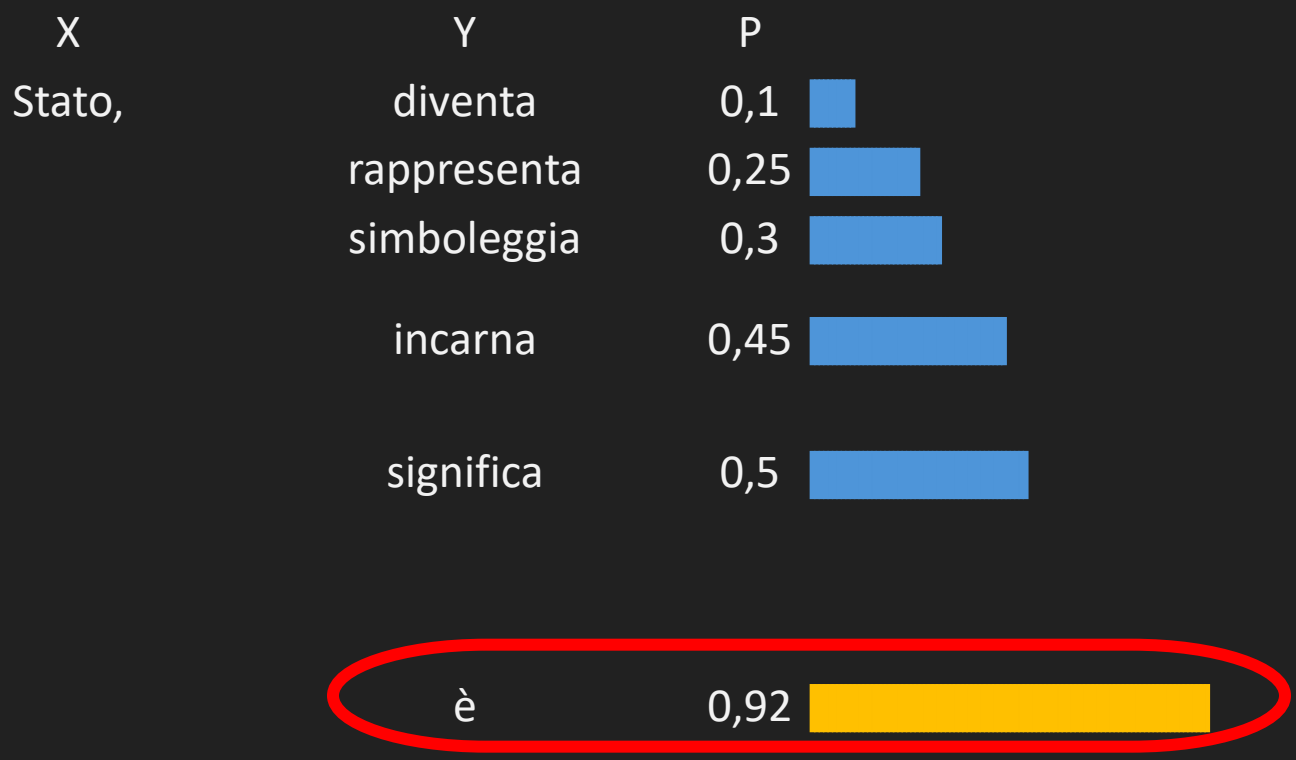
Qual è la città più grande della Francia?

La città più grande della Francia, nonché capitale dello ->



Qual è la città più grande della Francia?

La città più grande della Francia, nonché capitale dello Stato, ->



I Limiti degli LLM

I Limiti degli LLM

- **SONO ADDESTRATI SU INTERNET**
- **POSSONO COMMETTERE ERRORI => NECESSARIO VERIFICARE I RISULTATI**
- **QUALITÀ DELL'INPUT = QUALITÀ DELL'OUTPUT**
- **NON HANNO "VALORI", ETICA O PRINCIPI MORALI CHE LI ISPIRINO**
- **NON SANNO COSA SIA VERO, FALSO, CORRETTO O ERRATO**
- **SONO AUTO-REGRESSIVI (GLI ERRORI SI ACCUMULANO)**
- **NON POSSIEDONO COMPRENSIONE O CONSAPEVOLEZZA**

Diego Vendramini – Non copiare o riprodurre

Stochastic Parrots*

GLI LLM non sono collegati un insieme limitato fisso e chiuso di compiti particolari

Però….

Sono **incapaci di generalizzare** ciò che hanno imparato riportandolo in contesti qualitativamente molto diversi.

Sono **incapaci di interpretare in modo flessibile** i loro compiti in un contesto più ampio e la loro relazione con esso.

* Bender, Emily M.; Gebru, Timnit; McMillan-Major, Angelina; Shmitchell, Shmargaret (2021-03-01). "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? ". *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. FAccT '21. New York, NY, USA: Association for Computing Machinery. pp. 610–623.

Il gioco si sta rompendo?

«*Il giornale bagnato caduto dal tavolo è il mio giornale preferito. Ma ora che il mio Giornale preferito ha licenziato il direttore, forse non mi piacerà più leggerlo».*

Posso sostituire nella seconda frase "il mio giornale preferito" con "il giornale bagnato che è caduto dal tavolo"?

Chat GPT 4° - 23 maggio 2024

Sì, puoi sostituire "il mio Giornale preferito" con "il giornale bagnato che è caduto dal tavolo" nella seconda frase.

La frase risulterà così:

"*Ma ora che il giornale bagnato che è caduto dal tavolo ha licenziato il direttore, forse non mi piacerà più leggerlo."*

Questa modifica mantiene il significato originale della frase e rende chiaro il riferimento al giornale.

Il gioco si è rotto

Chat GPT o3 – 21 marzo 2025

Il giornale bagnato caduto dal tavolo è il mio giornale preferito. Ma ora che il mio Giornale preferito ha licenziato il direttore, forse non mi piacerà più leggerlo».

Posso sostituire nella seconda frase "il mio giornale preferito" con "il giornale bagnato che è caduto dal tavolo"?

La sostituzione proposta è grammaticalmente possibile ma logicamente inappropriata, in quanto mescola due piani di significato incompatibili (letterale e figurato).

La frase originale fa chiaramente riferimento al giornale in senso figurato (testata editoriale, azienda, redazione), non all'oggetto fisico caduto dal tavolo. Sostituire «il mio giornale preferito» con «il giornale bagnato che è caduto dal tavolo» rompe questa distinzione tra significato figurato e letterale, rendendo la frase illogica o paradossale.

Ma allora come fa ChatGPT a fare i calcoli?

Un po’ di Passaggi nel calcolo:

1. INTERPRETAZIONE DELLA RICHIESTA

Il modello interpreta la richiesta di calcolo.

Esempio: 'Quanto fa 5+3?' -> Comprende che desideri una somma.

2. GENERAZIONE DELLA RISPOSTA

Utilizza parametri pre-addestrati per generare una risposta

Potrebbe richiamare regole matematiche o logiche apprese durante l'addestramento

Richiama risultati noti per calcoli semplici.

E’ Possibile commetta errori nei calcoli complessi.

3. ESECUZIONE DI CALCOLI

Se necessario, passa la richiesta a un **motore di calcolo esterno** (Wolfram Alpha; Python con librerie scientifiche come NumPy; MATLAB; R; Google Calculator….

4. RESTITUZIONE DELLA RISPOSTA

Fornisce direttamente il risultato per calcoli semplici.

Spiega il processo per calcoli complessi.

Fornisce codice

Diego Vendramini – Non copiare o riprodurre

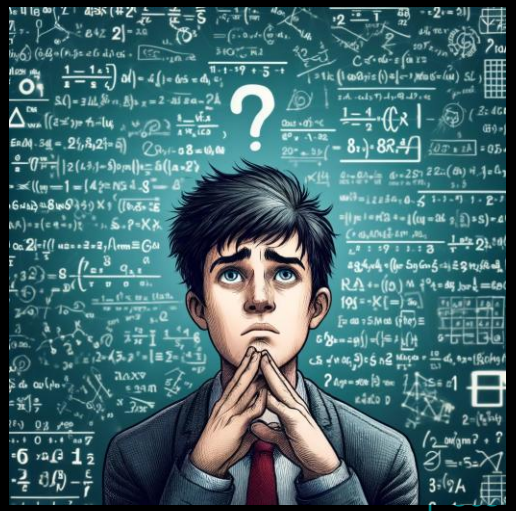
Rischi e problemi

- **Ambiguità** della Domanda
- **Parole Mancanti** o malintese
- **Assunzioni sbagliate** sul contesto o sulla necessità del calcolo (semplificando eccessivamente o complicando inutilmente la richiesta)
- **Errori di Parsing**: la conversione della domanda dell'utente in un formato comprensibile per il motore di calcolo potrebbe contenere errori di parsing.
- **Limiti** del Motore di Calcolo Esterno
- **Riformulazione dei Risultati**: Anche se il motore di calcolo esterno fornisce una risposta corretta, ChatGPT deve interpretare e riformulare questo risultato in un linguaggio naturale. Durante questo processo, potrebbero verificarsi errori di interpretazione o presentazione.

Esempio Pratico:

"Qual è la radice quadrata di 123 e 456" ?

$\sqrt{123 + 456}$ $\sqrt{123} + 456$



E quando Generiamo un’immagine, che succede?

Tipi di rete coinvolte

CNN (Convolutional Neural Networks): Le CNN sono reti neurali progettate per elaborare dati visivi. Utilizzano strati convoluzionali per rilevare caratteristiche come bordi, texture e forme. Ogni strato convoluzionale applica filtri per estrarre queste caratteristiche, mentre i livelli di pooling riducono la dimensionalità, mantenendo informazioni essenziali. Questo le rende ideali per compiti come il riconoscimento di oggetti e la classificazione di immagini.

GAN (Generative Adversarial Networks): Le GAN consistono in due reti neurali: un generatore e un discriminatore. Il generatore crea immagini false, mentre il discriminatore cerca di distinguere tra immagini reali e generate. Questo processo competitivo migliora la qualità delle immagini generate.

VAE (Variational Autoencoders): I VAE sono modelli generativi che apprendono una rappresentazione compatta dei dati (codifica) e poi li ricostruiscono (decodifica). Sono utili per generare nuove immagini simili a quelle su cui sono stati addestrati.

Diego Vendramini – Non copiare o riprodurre

Un po’ di passaggi:

Input: *" un gatto grigio vestito da astronauta nello spazio "*

['un', ' ', 'gatto', ' ', 'grigio', ' ', 'vestito', ' ', 'da', ' ', 'astronauta', ' ', 'nello', ' ', 'spazio']

1. *Tokenizzazione : La frase viene divisa in token.*
2. *Encoding: I token vengono codificati in vettori numerici.*
3. *Positional Encoding: Viene aggiunto un PE per considerare l'ordine dei token.*
4. MultiHead Attention nel Transformer
5. **Spazio Latente**: Il modello genera una rappresentazione nello spazio latente.
6. **Decodifica**: La rappresentazione latente viene convertita in un'immagine.
7. **Post-Processing**: L'immagine viene migliorata.
8. **Output**: L'immagine finale viene generata.

Step 5: Passaggio nel modello Text to Image

Dopo aver compreso il significato del testo attraverso i vari passaggi, il modello deve generare un'immagine.

L'immagine da generare viene trattata come una **sequenza di pixel** o **patch** (piccole aree dell'immagine, come blocchi di 16x16 pixel):

Ogni patch diventa un "token visivo".
Similmente al testo, ogni patch viene rappresentata come un vettore numerico (embedding del patch).

Il vettore di input viene quindi passato attraverso un modello di Text-to-Image (ad esempio, DALL-E, Stable Diffusion, ecc.) nello spazio latente.

Il modello converte l'embedding della frase in una rappresentazione visiva

Diego Vendramini – Non copiare o riprodurre

Step 6: Spazio Latente

I nuovi vettori entrano nello Spazio latente

Lo **spazio latente** è una rappresentazione astratta, multidimensionale e compatta di dati complessi (ad esempio immagini, testi, o suoni) che cattura le loro caratteristiche principali in un formato numerico ridotto.
È uno spazio in cui ogni punto (o vettore) corrisponde a un'istanza compressa e "codificata" dei dati, preservandone i tratti essenziali.
In altre parole, lo spazio latente è una sorta di "lingua nascosta" in cui il modello rappresenta informazioni complesse per elaborarle in modo efficiente e strutturato.

Il vettore nello spazio latente è una rappresentazione **non direttamente interpretabile.** È una combinazione di caratteristiche astratte apprese dal modello

Diego Vendramini – Non copiare o riprodurre

Utilità dello Spazio latente

Un'immagine RGB utilizza **3 canali** (Rosso, Verde e Blu) per rappresentare le informazioni visive.

Lo spazio latente, invece, non rappresenta direttamente i colori, ma una **codifica astratta** delle caratteristiche significative (texture, bordi, profondità, ecc.)

L'aumento del numero di canali nello spazio latente (ad esempio da 3 a 4) fornisce **più dimensioni per rappresentare pattern complessi** con meno pixel. Questo aiuta la rete neurale a catturare meglio i dettagli strutturali, la posizione spaziale e le caratteristiche astratte.

Esempio: Se la frase è "Un gatto grigio vestito da astronauta nello spazio", il modello cercherà di generare una rappresentazione latente che includa le seguenti dimensioni:

- Canale 1: Forma degli oggetti (es. il contorno del gatto o del casco).
- Canale 2: Texture superficiale (es. pelo del gatto, materiale del casco).
- Canale 3: Informazioni spaziali o disposizioni degli oggetti (es. stelle nello sfondo).
- Canale 4: Colori o illuminazione (es. sfondo nero con luci stellari).

Diego Vendramini – Non copiare o riprodurre

Tecniche di Generazione nello Spazio Latente: il Modello di Diffusione

Uno dei metodi più popolari per generare immagini nello spazio latente è l'uso di Diffusion Models.
Questi modelli aggiungono gradualmente rumore a un'immagine iniziale (che potrebbe essere completamente casuale) e poi rimuovono il rumore in modo iterativo per generare l'immagine finale.

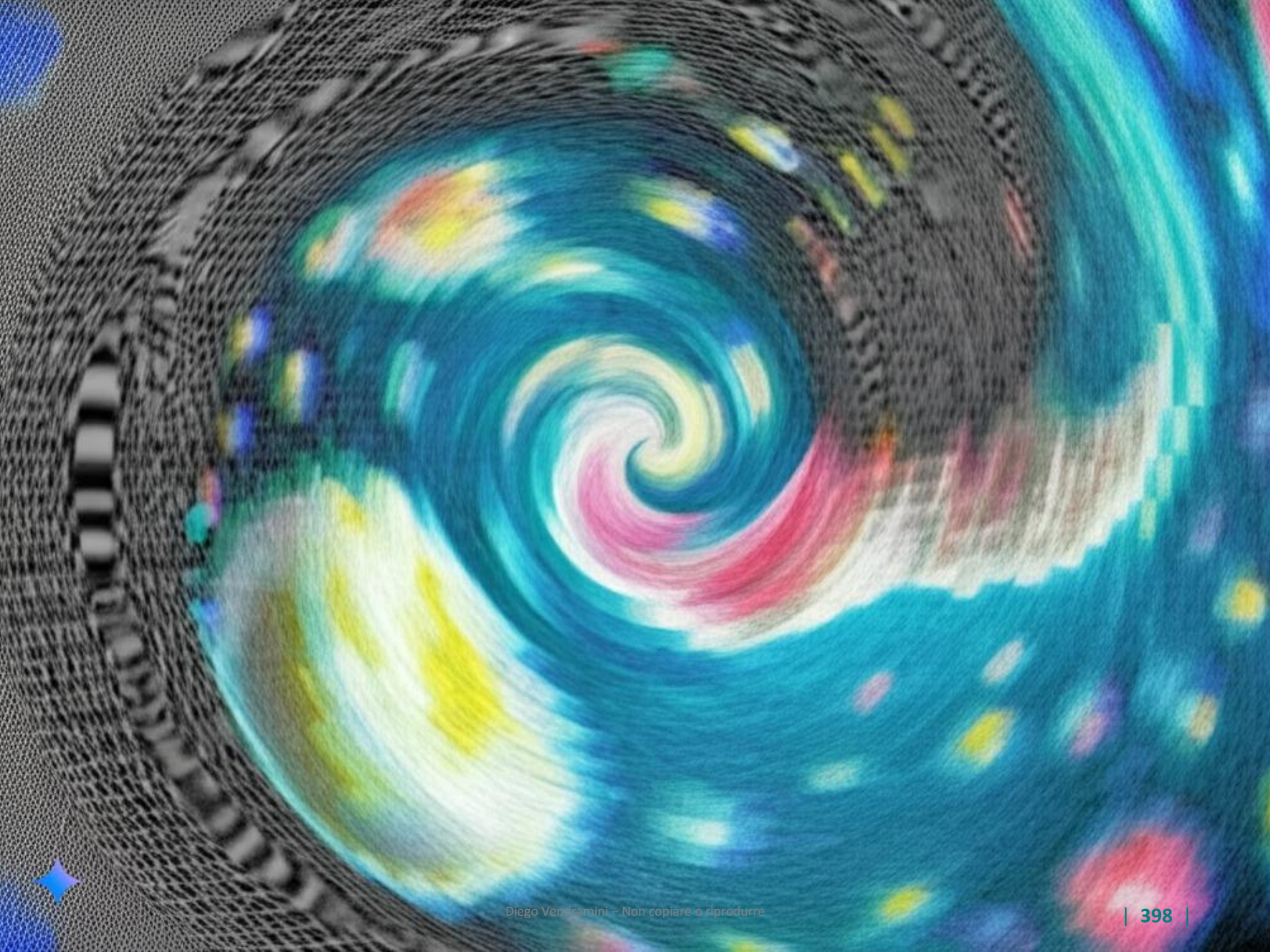
Generative Adversarial Networks (GANs):

Il generatore crea immagini nello spazio latente e un discriminatore valuta la qualità delle immagini generate.

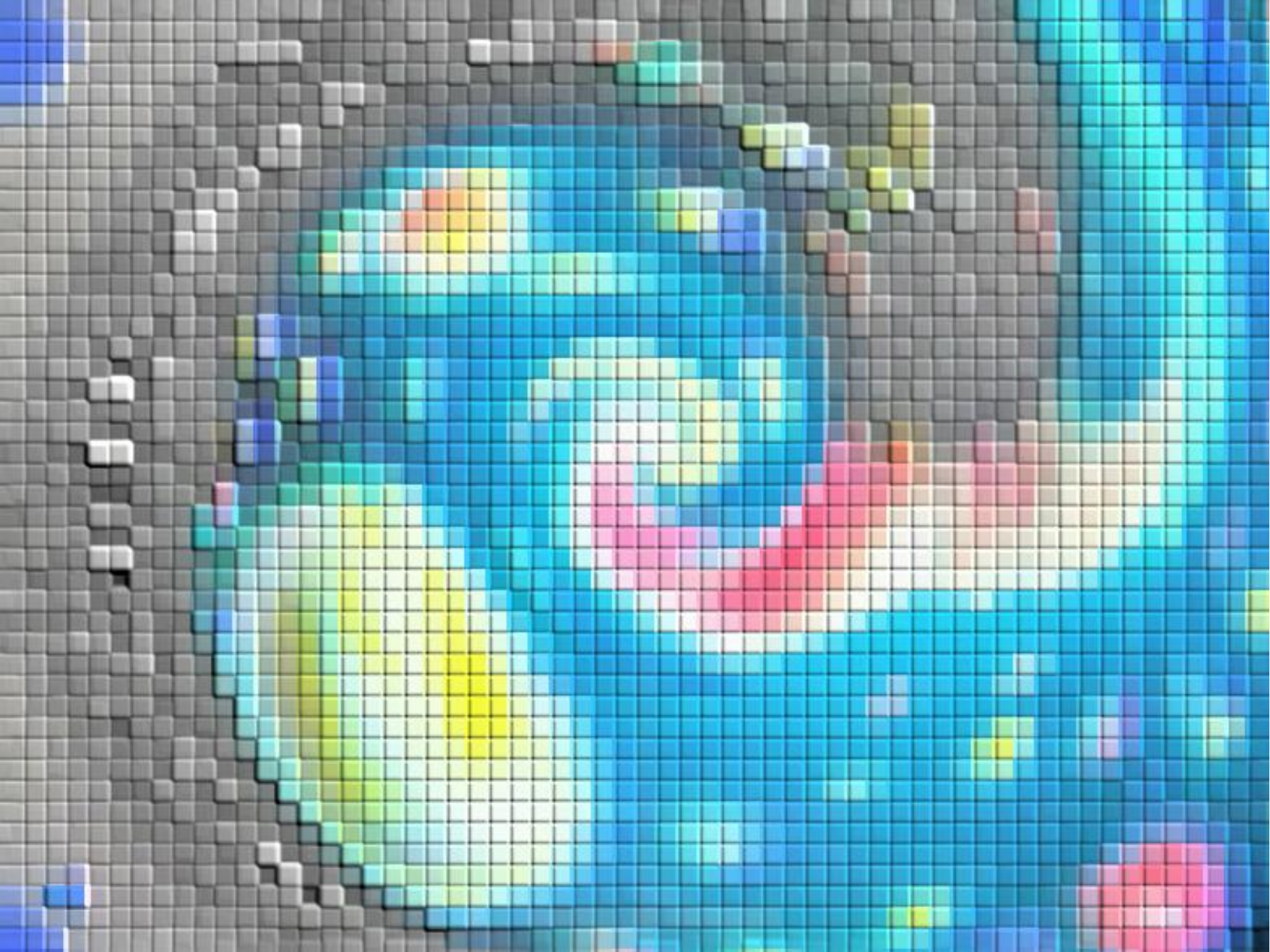
Il generatore cerca di ingannare il discriminatore generando immagini sempre più realistiche.

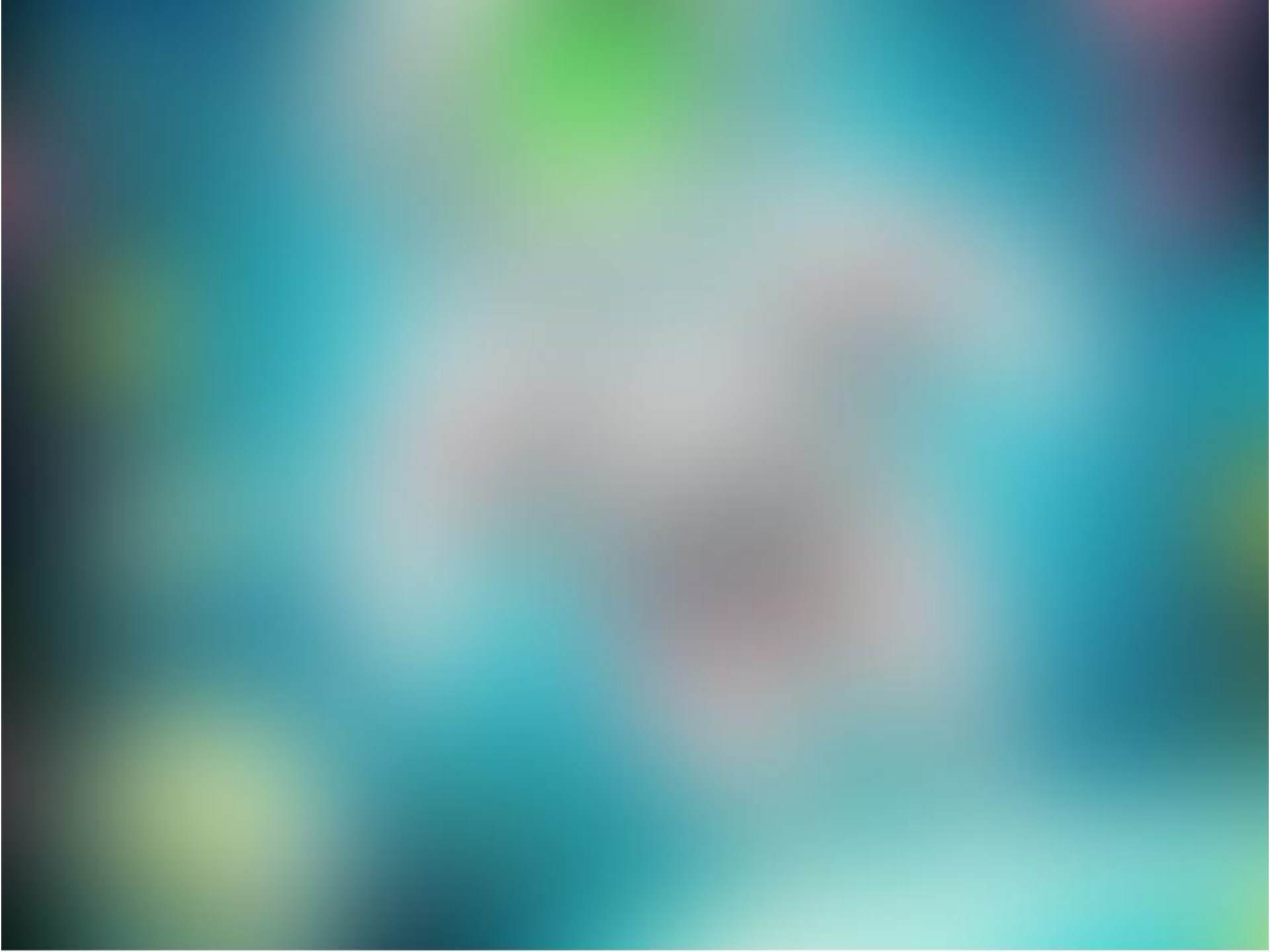
Diego Vendramini – Non copiare o riprodurre

Diego Vendramini – Non copiare o riprodurre



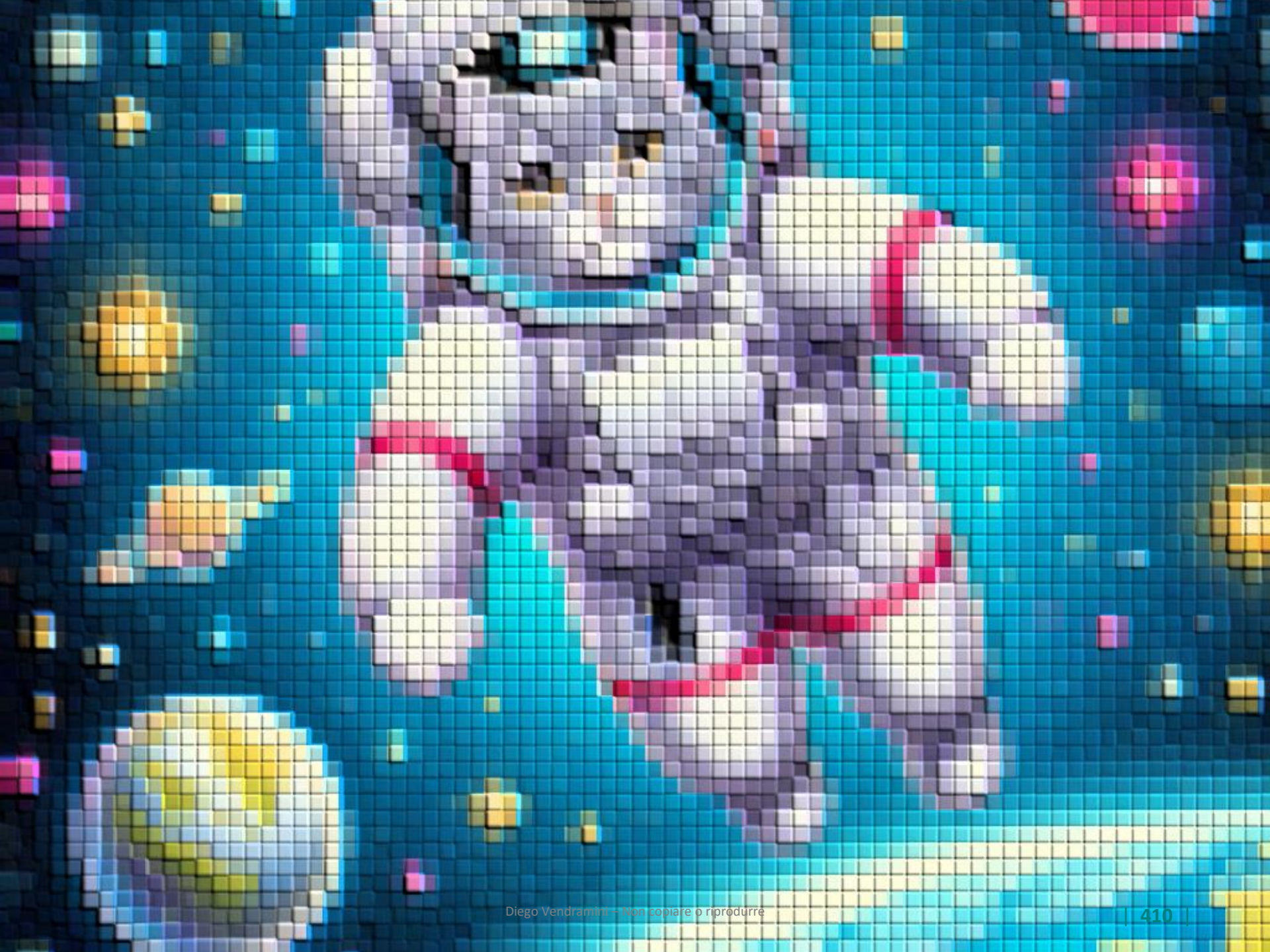
Diego Vendramini – Non copiare o riprodurre





Diego Vendramini – Non copiare o riprodurre

Diego Vendramini – Non copiare o riprodurre



Diego Vendramini – Non copiare o riprodurre

Diego Vendramini – Non copiare o riprodurre

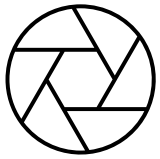
AGI- Artificial General Intelligence:

Diego Vendramin – Non copiare o riprodurre

Proprietà e caratteristiche di un AGI:

E’ un’ intelligenza che non è legata a un insieme di compiti altamente specifici

Mostra lo stesso tipo di intelligenza generale approssimativa degli esseri umani



Generalizza ciò che ha imparato, anche in contesti qualitativamente molto diversi dai precedenti

Adotta una visione ampia e interpreta flessibilmente i suoi compiti nel contesto del mondo in generale e della sua relazione con esso

Diego Vendramini – Non copiare o riprodurre

Ragionamento

Creatività

Associazione

Generalizzazione

Pianificazione

Riconoscimento di schemi

Risoluzione di problemi

Memorizzazione

Definizione di obiettivi

Apprendimento

Ottimizzazione

Auto-conservazione

Elaborazione sensoriale

Elaborazione del linguaggio

Classificazione

Induzione e Deduzione

Diego Vendramini – Non copiare o riprodurre

AI ACT

Obiettivi principali

1.Sicurezza: Assicurare che i sistemi di AI non comportino rischi inaccettabili per la sicurezza o la salute delle persone.

2.Protezione dei diritti fondamentali: Salvaguardare i diritti e le libertà fondamentali, come la privacy e la non discriminazione.

3.Promozione dell'innovazione: Creare un quadro normativo che non ostacoli lo sviluppo tecnologico, ma lo guidi in modo responsabile.

Diego Vendramini – Non copiare o riprodurre

AI ACT

Principali aspetti regolamentati

•**Classificazione dei rischi**: I sistemi di AI sono suddivisi in categorie basate sul livello di rischio: inaccettabile, alto, limitato e minimo. I sistemi ad alto rischio, come quelli usati in ambito sanitario o per il reclutamento, sono soggetti a requisiti più stringenti.

•**Proibizioni**: Alcuni usi dell'AI sono vietati, come il social scoring o la manipolazione comportamentale che possa causare danni.

•**Trasparenza**: Per i sistemi di AI con cui gli utenti interagiscono direttamente, è richiesto che ci sia trasparenza sull'uso dell'AI.

Diego Vendramini – Non copiare o riprodurre

AI ACT

Principali aspetti regolamentati

•**Classificazione dei rischi**: I sistemi di AI sono suddivisi in categorie basate sul livello di rischio: inaccettabile, alto, limitato e minimo. I sistemi ad alto rischio, come quelli usati in ambito sanitario o per il reclutamento, sono soggetti a requisiti più stringenti.

•**Proibizioni**: Alcuni usi dell'AI sono vietati, come il social scoring o la manipolazione comportamentale che possa causare danni.

•**Trasparenza**: Per i sistemi di AI con cui gli utenti interagiscono direttamente, è richiesto che ci sia trasparenza sull'uso dell'AI.

Diego Vendramini – Non copiare o riprodurre

AI ACT

Cosa cerca di evitare l'AI Act

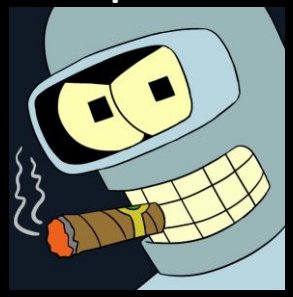
•**Rischi elevati e inaccettabili**: L'AI Act mira a prevenire l'uso di sistemi di AI che presentano rischi elevati per la sicurezza pubblica o che violano i diritti fondamentali.

•**Discriminazione e bias**: Si cerca di ridurre l'impatto di bias nei sistemi di AI, garantendo che non perpetuino discriminazioni ingiuste.

•**Mancanza di trasparenza**: Impone requisiti di trasparenza per garantire che gli utenti siano consapevoli quando interagiscono con un sistema di AI.

Diego Vendramini – Non copiare o riprodurre

N.AI vs G.AI vs S.AI

Tipo di Intelligenza Artificiale	Caratteristiche	Pro	Contro e Rischi per l'Umanità
Narrow AI	Progettata per compiti specifici, eccelle in un campo ristretto	Efficiente e precisa nei compiti specifici. Può aumentare la produttività e migliorare la qualità della vita	Limitata, non può adattarsi a nuovi compiti senza essere riprogrammata, rischio di **disoccupazione in settori automatizzati**
General AI	Capacità cognitive umane, apprendimento e adattamento a diversi compiti	Versatile e flessibile, potrebbe rivoluzionare molteplici settori inclusi i modelli economici e sociali	Sfide tecniche complesse da risolvere, **rischi etici** e di sicurezza, inclusa la **perdita di controllo**
Super AI 	Supera l'intelligenza umana in ogni campo, superiore in creatività, risoluzione dei problemi e decisioni strategiche	Potrebbe risolvere problemi complessi che gli esseri umani non possono affrontare, migliorando significativamente il progresso scientifico e tecnologico	**Rischi esistenziali**, potrebbe decidere di agire contro gli interessi umani, difficile da controllare e prevedere

Diego Vendramini – Non copiare o riprodurre

Grazie!

STUDIOSAMO®