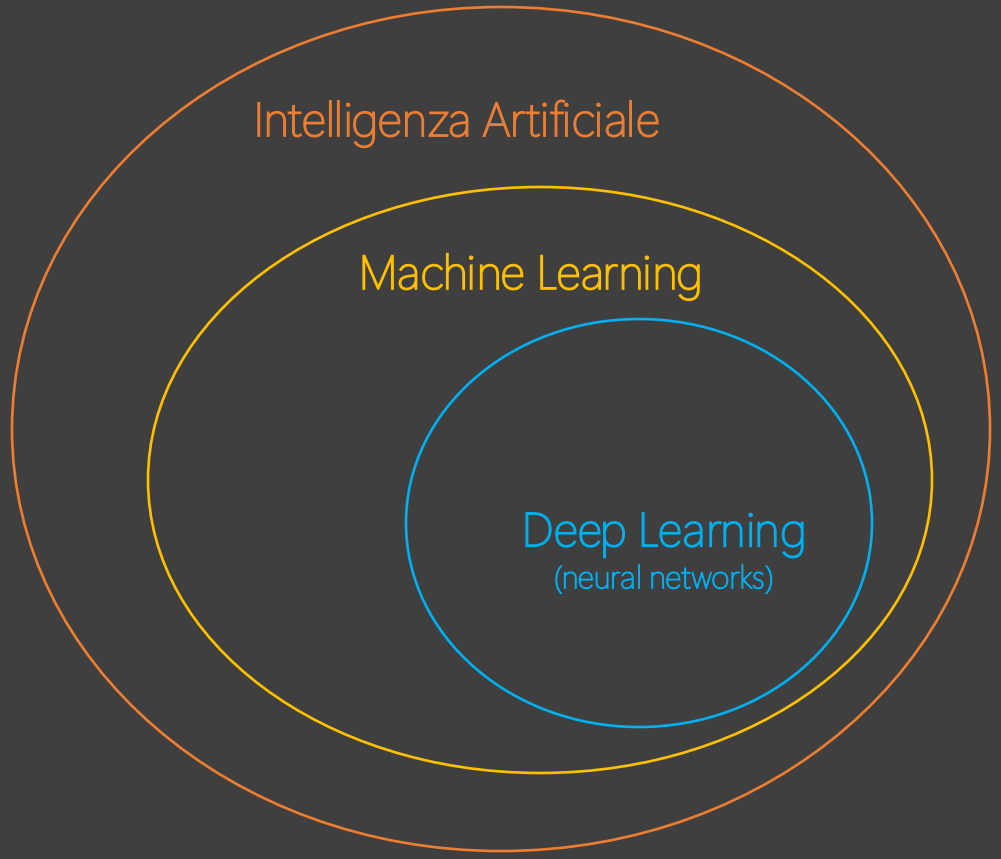
Artificiale o Intelligente?

Contesto


Intelligenza Artificiale
Machine Learning
Deep Learning
(neural networks)

What is Machine Learning?

ML is the study of **ALGORITHMIC** prediction

Machines that **LEARN** how to **PREDICT**

Algo1: LEARNER Algo2: PREDICTOR

Algorithm: is a finite sequence of rigorous instructions, typically used to

SOLVE A CLASS of specific problems

ML - definitions

We want the machine LEARN a PREDICTOR (the learner fixes parameters in the predictor)

Domain set:

$\mathcal{X}$ Usually, you have vectors $\underline{x}$ of features

Label set:

$\mathcal{Y}$ for example $[0,1]$

Training set:

$S = ((x_1, y_1), \ldots, (x_N, y_N))$
Sequence of pairs $\mathcal{X} \times \mathcal{Y}$

What the learner can use to determine the values of the predictor's parameters

Learner output:

Is the predictor: a function $h: \mathcal{X} \to \mathcal{Y}$

Data-generation model

What are the data we use?

1. Instances from the training dataset are generated by some probability distribution

   $\mathcal{D}$ is a probability distribution over the domain set $\mathcal{X}$

   The learner knows nothing about the distribution $\mathcal{D}$

2. There is a "correct" labeling function $f: \mathcal{X} \to \mathcal{Y}$

   $$y_i = f(x_i) \; \forall i$$

   This function is what the learner must learn.

The training dataset is the outcome of:
- Use $\mathcal{D}$ to extract x_i from $\mathcal{X}$
- Use f to compute y_i given x_i

Recap:

- We have a training set S sampled from an UNKNOWN distribution $\mathcal{D}$ and labeled by some target function f
- The learner output is a predictor

$$h_S: \mathcal{X} \to \mathcal{Y}$$

The predictor depend on S

GOAL: find h_S that minimizes the error with respect to the UNKNOWN $\mathcal{D}$ and f

Empirical Risk Minimization & loss function

$L_{\mathcal{D},f}(h)$, the true error, is not directly available because we do not know $\mathcal{D}$ and f

TRAINING ERROR: error of the predictor on the training samples

$$L_S(h) \stackrel{\text{def}}{=} \frac{|\{i \in [N] : h(x_i) \neq y_i\}|}{N} \qquad [N] = \{1, \dots, N\}$$

EMPIRICAL RISK/ERROR

A loss function takes two inputs, $\hat{y} = h(x)$ and the target values y, and returns a **real number** $loss(\hat{y}, y)$ that we interpret as a quantified error for predicting $\hat{y}$ when the target is y .

The risk associated with h is defined as

$$L(h) = \mathbb{E}(loss(h(X), Y)$$

$$L_S(h) = \frac{1}{N} \sum_{i=1}^{N} loss(h(x_i), y_i)$$

Deep Learning

How do we choose $h(x)$?

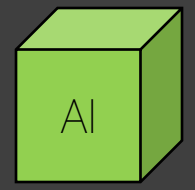
Many different algorithms available to approximate the target function

BUT

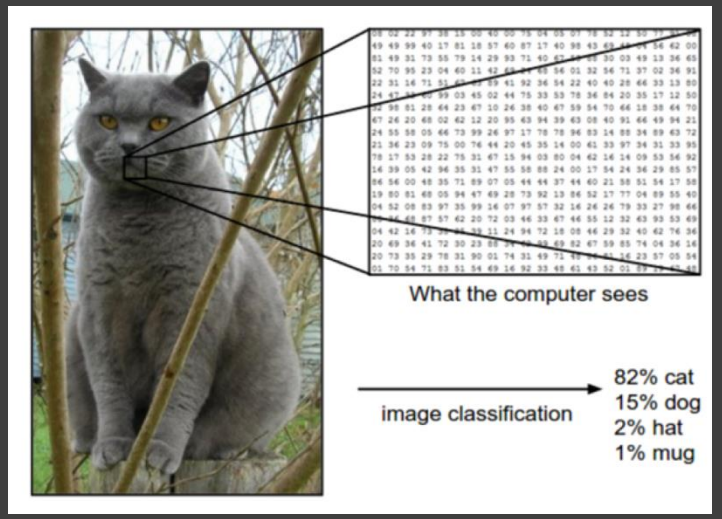
Today most of the people use:

Deep Neural Networks

The core of a neural network



$$\sigma\left(\begin{bmatrix} w_{11} & w_{12} & w_{13} & w_{14} \\ w_{21} & w_{22} & w_{23} & w_{24} \\ w_{31} & w_{32} & w_{33} & w_{34} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}\right) = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$



Kevin P. Murphy,
Probabilistic Machine Learning, An Introduction

$$y_j = \sigma(x_i w_{ij} + b_j)$$

Deep Learning: types of network

- Fully connected → Tabular data
- Convolutional → Images
- Recurrent → Sequence
- Transformer → Sequence
- [Diffusion] → Images

Large Language models [ChatGPT]

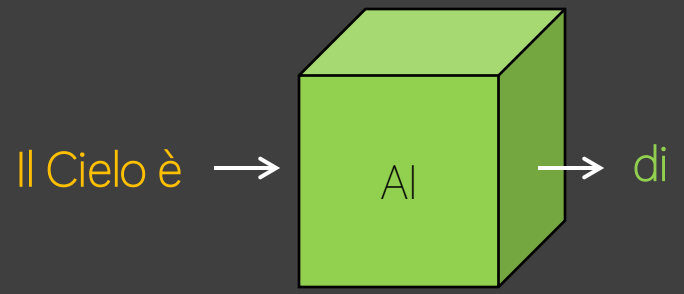
Images generative models [Dall-E]

Large Language Models

Intelligenza Artificiale:

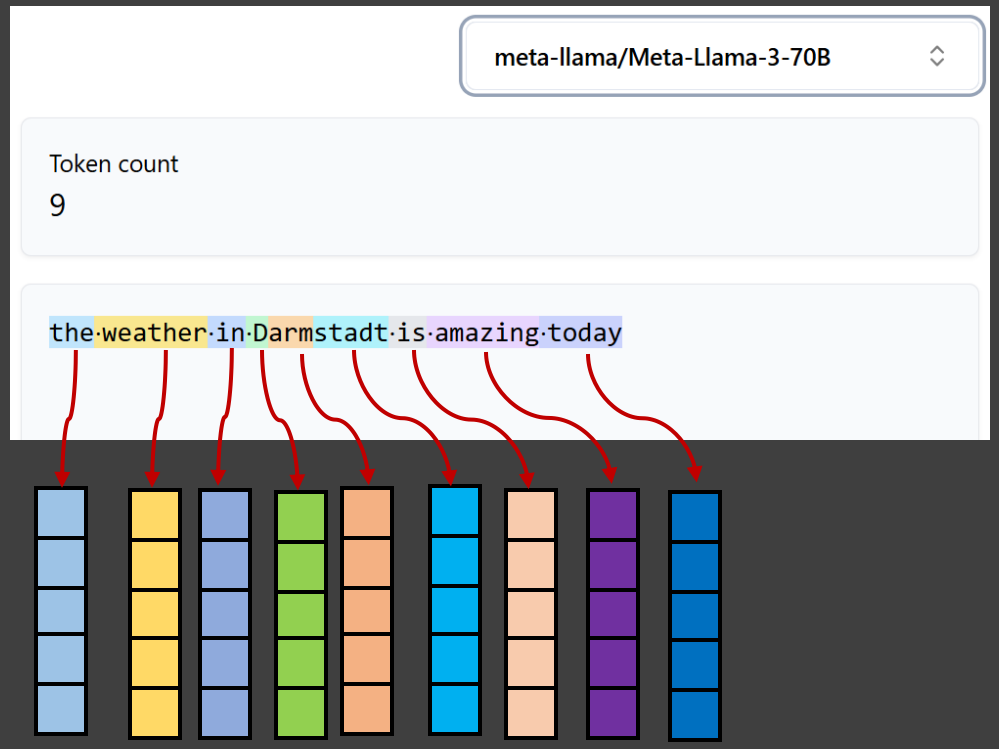
Teoria delle probabilità sotto steroidi

1. Cielo – Wikipedia
2. Rino Gaetano - Ma il cielo è sempre più blu (Official Video)
3. Ma il cielo è sempre più blu (Extended Version) – YouTube
4. Bungaro e Fiorella Mannoia - Il Cielo è di Tutti ... – YouTube
5. Perché il cielo è blu?
6. Perché il cielo è blu? La spiegazione semplice
7. Scienza Per Tutti - 0264. Perchè il cielo è azzurro? – INFN
8. Perchè il cielo è blu – base
9. Il cielo è di tutti. Ediz. a colori: Il cielo e di tutti - Amazon.it
10. Il cielo è di tutti
11. Il cielo è di tutti - Gianni Rodari
12. Il cielo non è blu, lo dice la scienza
13. Il cielo è dei leggeri - Matteo Munaretto
14. Il cielo è di tutti, la terra è di tutti - Mirca Benetton | Ed. ETS
15. Il cielo è dipinto di stelle
16. Il Cielo è di Tutti! - Video – RaiPlay
17. Perché il Cielo è Azzurro? Come spiegarlo ai bambini
18. Il cielo è dei violenti di Flannery O'Connor
19. Il cielo è di tutti - Rodari/Costa | Emme Edizioni
20. Perchè il cielo è blu?



Tokens and Embeddings

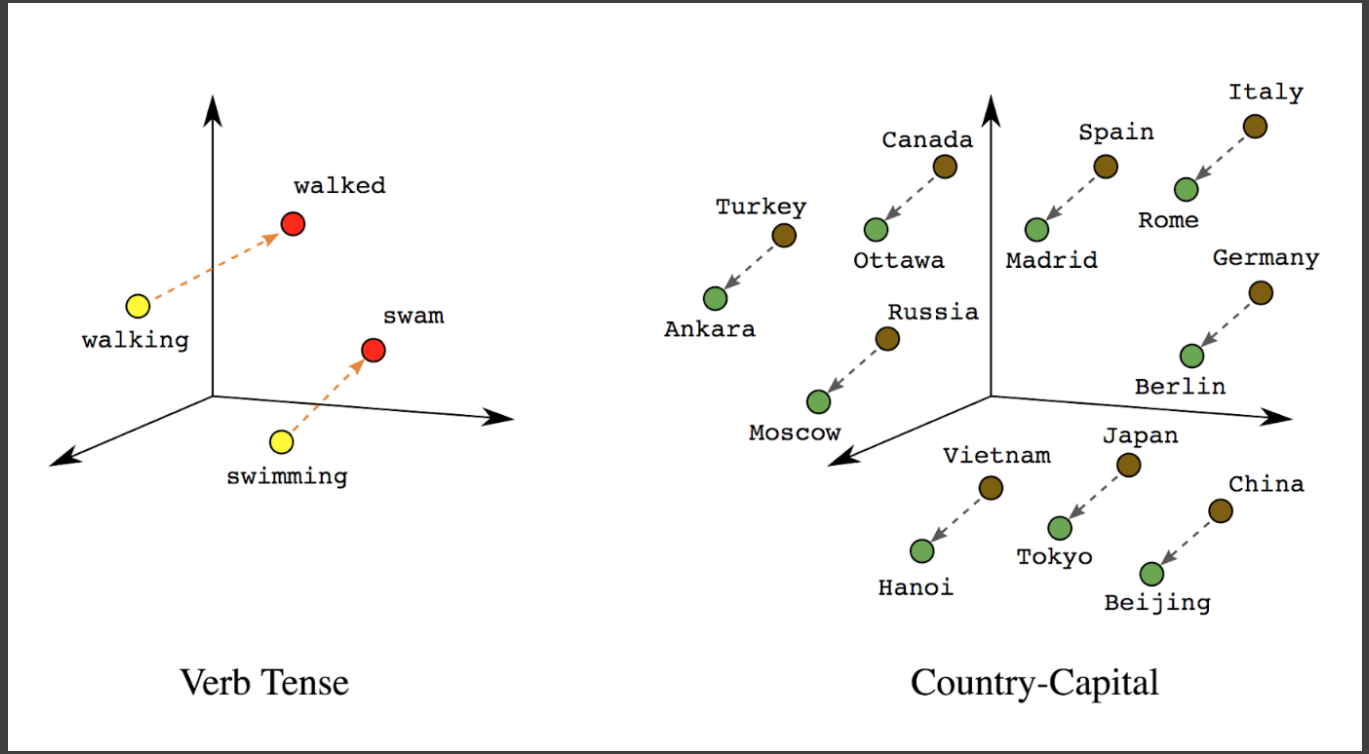
Computer cannot understand natural language, words become vectors.



https://tiktokenizer.vercel.app

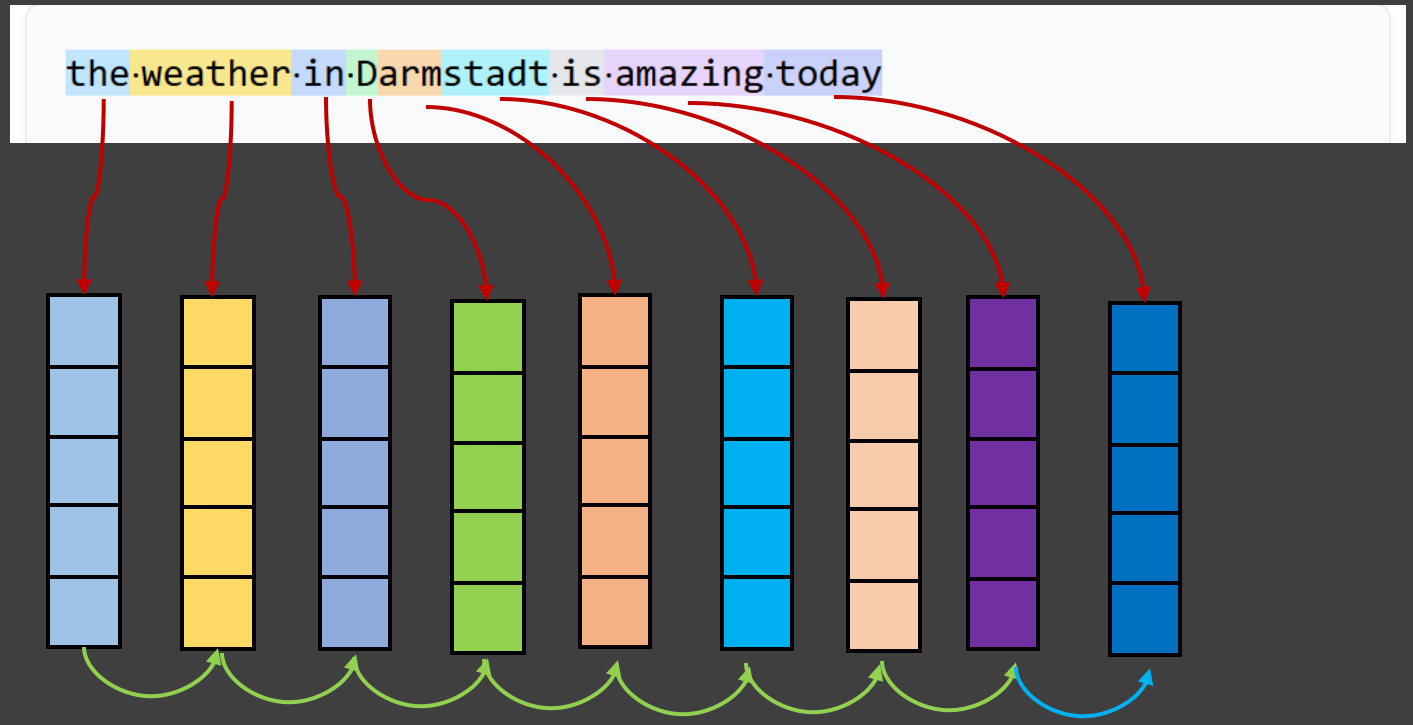
Vector embedding

Can capture semantic meaning and relationships.



Source: Google for Developers

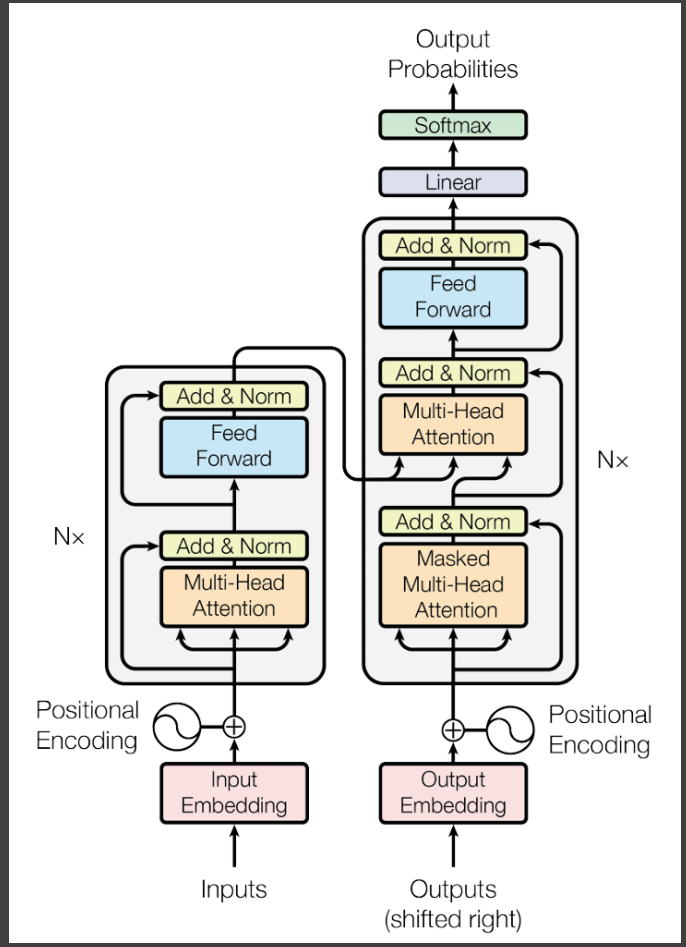
Tokens and Embeddings



Tokens are processed autoregressively

Attention mechanism

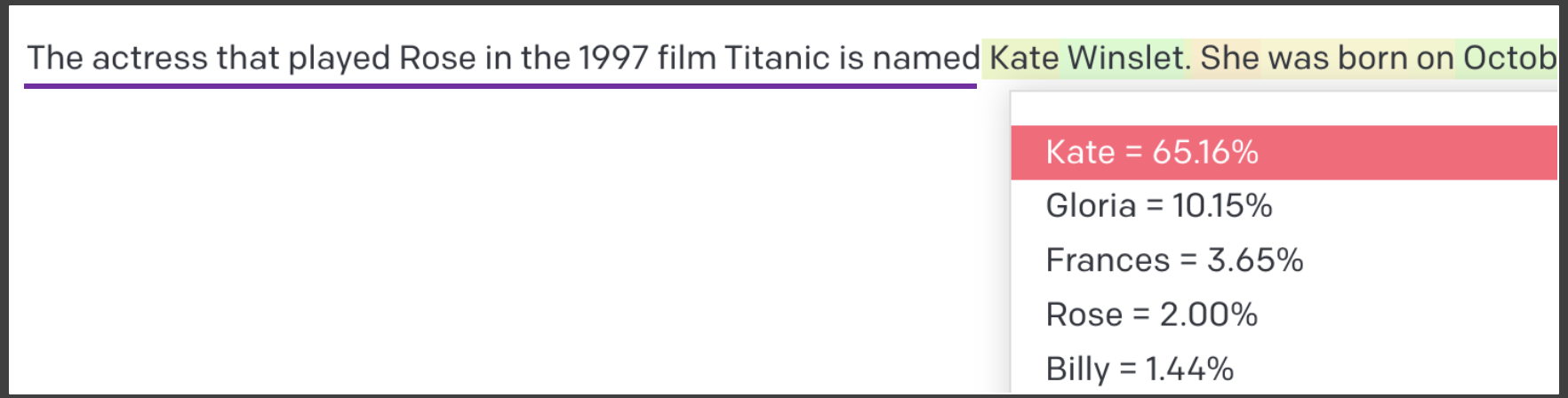
- Transformer is the go-to architecture for deep learning models, powering text-generative models like OpenAI's GPT, Meta's Llama (open), and Google's Gemini.
- Main innovation is the self-attention mechanism, which allows them to process entire sequences and capture long-range dependencies.
- Latest models can have more than 500 billion parameters, which are weights learned by the model during training.
- LLMs demand enormous data and compute power, balancing accuracy, cost, and sustainability.



Source: Attention Is All You Need

What is an LLM

Auto-regressive system* that, given the user prompt, will complete it with the most likely words.



Different from humans (typically) thinking first about what they want to say and then how.

*the output variable depends on its previous values.

Due questioni:

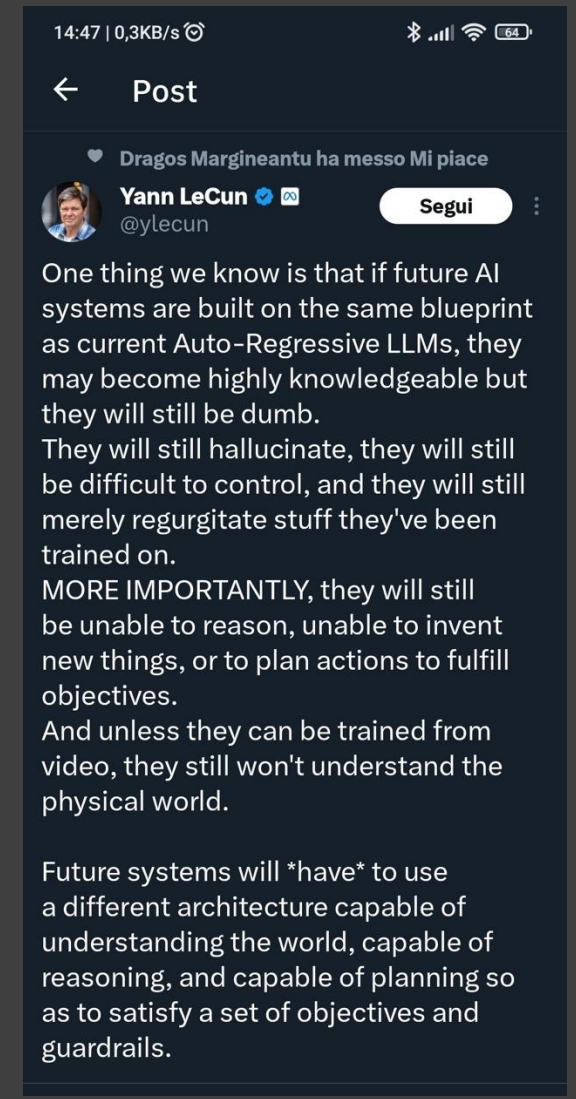
1. Quali sono i rischi legati allo sviluppo dei sistemi basati su Intelligenza Artificiale?
2. l'Intelligenza Artificiale è intelligente?

2018 Turing Award winners

Managing AI Risks in an Era of Rapid Progress

Yoshua Bengio — Mila - Quebec AI Institute, Université de Montréal, Canada CIFAR AI Chair
Geoffrey Hinton — University of Toronto, Vector Institute
Andrew Yao — Tsinghua University

Yoshua Bengio, Geoffrey Hinton, Andrew Yao, Dawn Song, Pieter Abbeel, Yuval Noah Harari, Ya-Qin Zhang, Lan Xue, Shai Shalev-Shwartz, Gillian Hadfield, Jeff Clune, Tegan Maharaj, Frank Hutter, Atılım Güneş Baydin, Sheila McIlraith, Qiqi Gao, Ashwin Acharya, David Krueger, Anca Dragan, Philip Torr, Stuart Russell, Daniel Kahneman, Jan Brauner, Sören Mindermann, arXiv:2310.17688, 26 Oct 2023



Twitter, 28/10/2023

Quali sono i rischi legati allo sviluppo dei sistemi basati su Intelligenza Artificiale?

Societal-scale risks

AI systems could rapidly come to outperform humans in an increasing number of tasks. If such systems are not carefully designed and deployed, they pose a range of societal-scale risks. They threaten to amplify social injustice, erode social stability, and weaken our shared understanding of reality that is foundational to society. They could also enable large-scale criminal or terrorist activities. Especially in the hands of a few powerful actors, AI could cement or exacerbate global inequities, or facilitate automated warfare, customized mass manipulation, and pervasive surveillance[12,13].

Many of these risks could soon be amplified, and new risks created, as companies are developing *autonomous* AI: systems that can plan, act in the world, and pursue goals. While current AI systems have limited autonomy, work is underway to change this[14]. For example, the non-autonomous GPT-4 model was quickly adapted to browse the web[15], design and execute chemistry experiments[16], and utilize software tools[17] including other AI models[18].

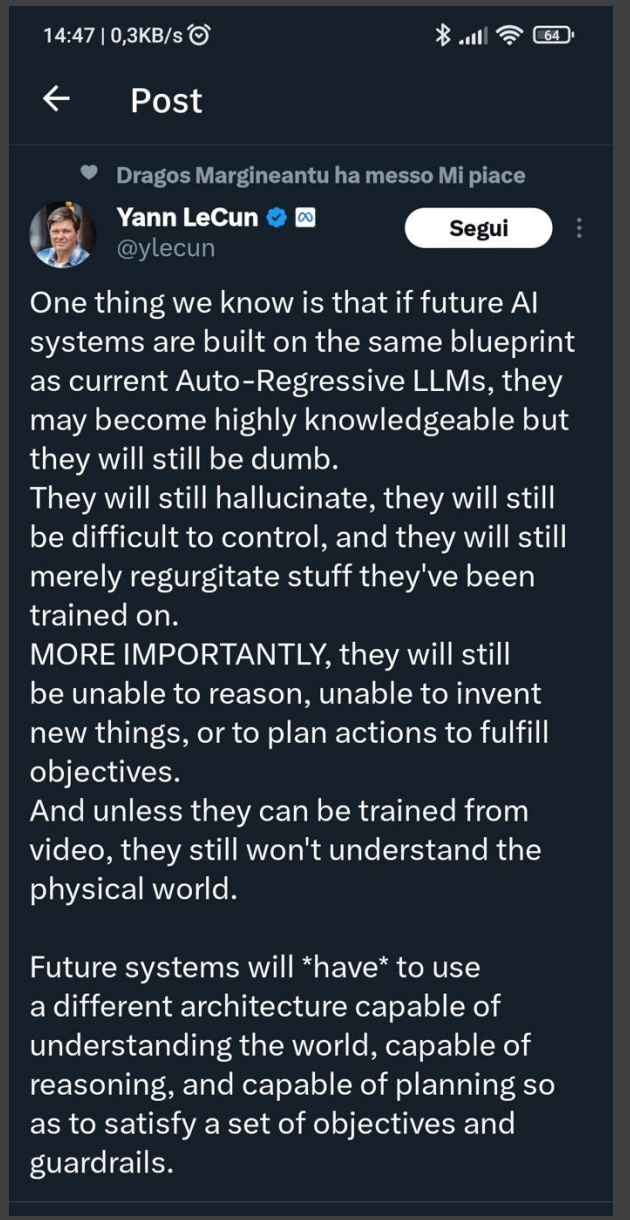
Quali sono i rischi legati allo sviluppo dei sistemi basati su Intelligenza Artificiale?

LLAMA 2: Open Foundation and Fine-Tuned Chat Models

Contents

https://arxiv.org/abs/2307.09288

l'Intelligenza Artificiale è intelligente?



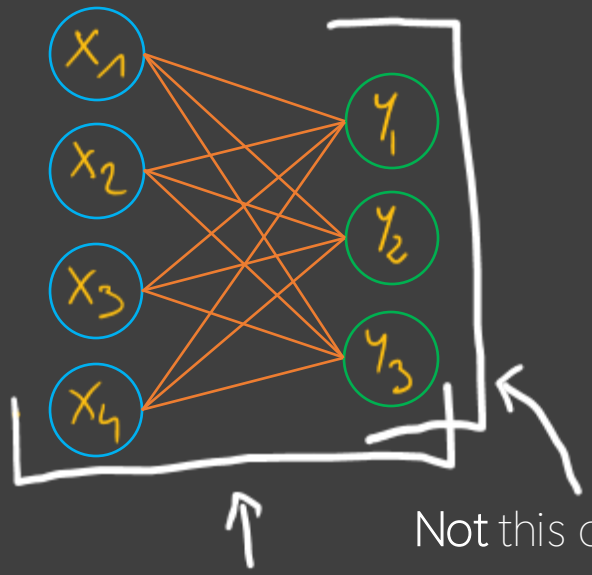
Twitter, 28/10/2023

l'Intelligenza Artificiale è intelligente?

Un inizio di risposta è «la domanda»
l'Intelligenza Artificiale non si pone domande

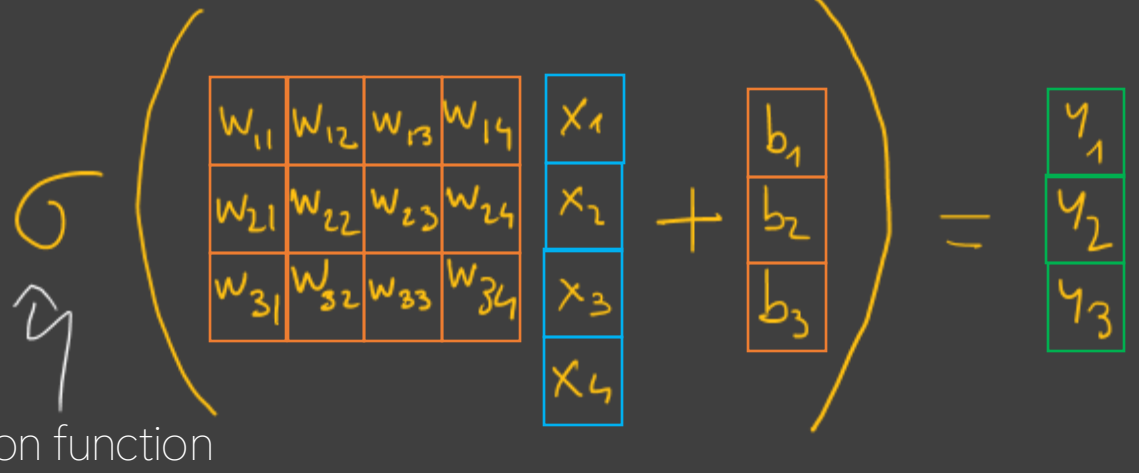
Backup

Neural Network: the layer



Not this one

The layer for me is this one

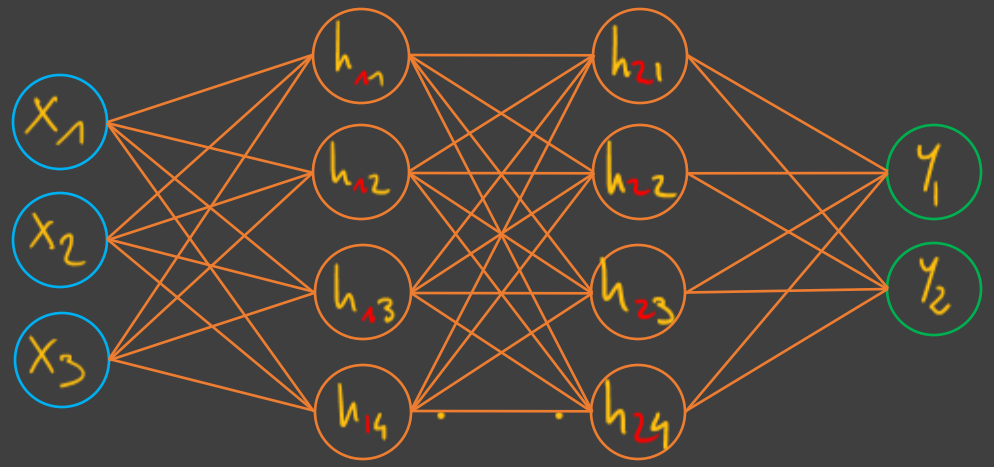


Activation function

$$y_j = \sigma(x_i w_{ij} + b_j)$$

Linear combination +
Non linear activation function

Neural Network with hidden layers

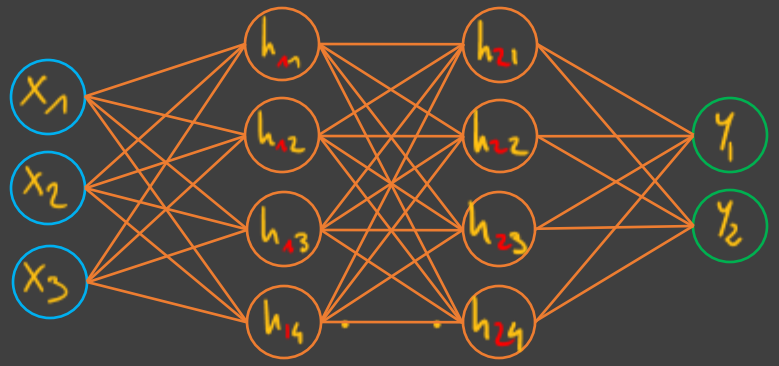


$$h_{1i} = \sigma(x_i w^1_{ij} + b^1_j)$$
$$h_{2i} = \sigma(h_{1i} w^2_{ij} + b^2_j)$$
$$y_j = \sigma(h_{2i} w^3_{ij} + b^3_j)$$

E.g.: Sigmoid activation function

$$\sigma = \frac{1}{1 + e^{-x}}$$

Empirical risk minimization

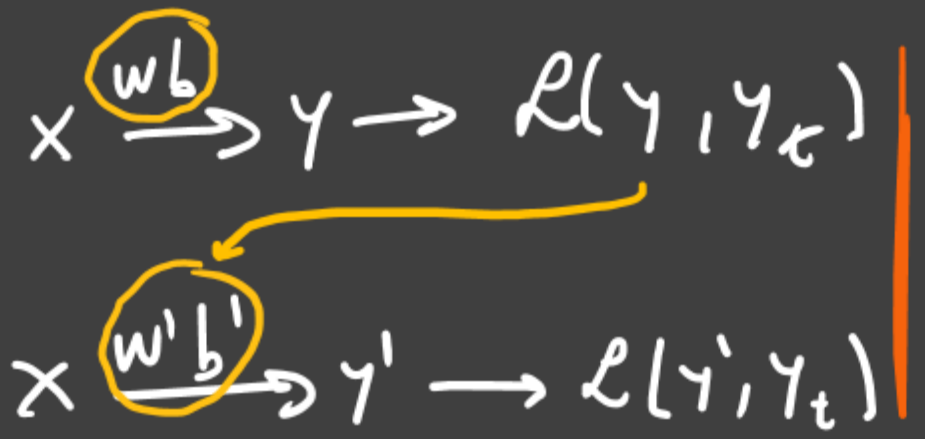


$$h_{1i} = \sigma(x_i w^1_{ij} + b^1_j)$$
$$h_{2i} = \sigma(h_{1i} w^2_{ij} + b^2_j)$$
$$y_j = \sigma(h_{2i} w^3_{ij} + b^3_j)$$

$$\mathcal{L}_{\mathcal{D}\ell}(h)$$

$$x \xrightarrow{wb} y \rightarrow \ell(y, y_t)$$
$$x \xrightarrow{w'b'} y' \rightarrow \ell(y', y_t)$$

Empirical risk minimization



Gradient Descent

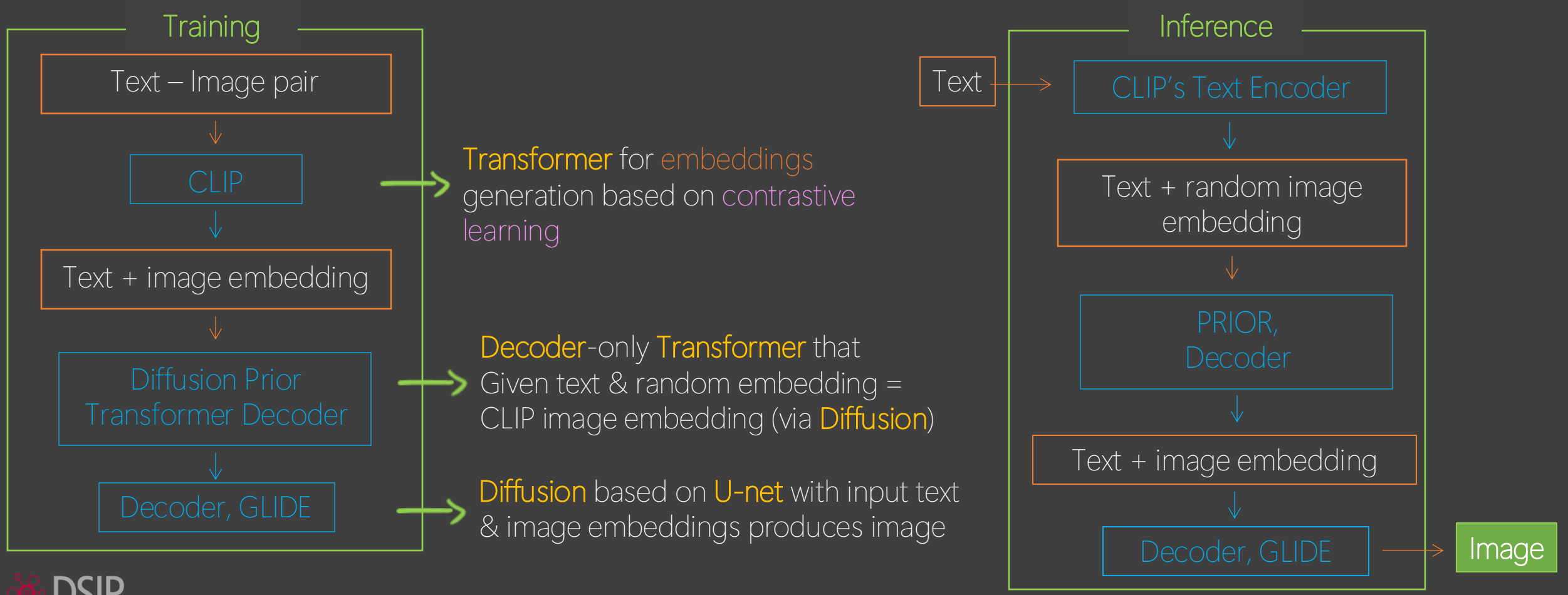
$$w' = w - \eta \nabla_w \mathcal{L}(w)$$

Learning rate

Need to compute the gradient wrt each weight

Deep Learning example: DALL-E

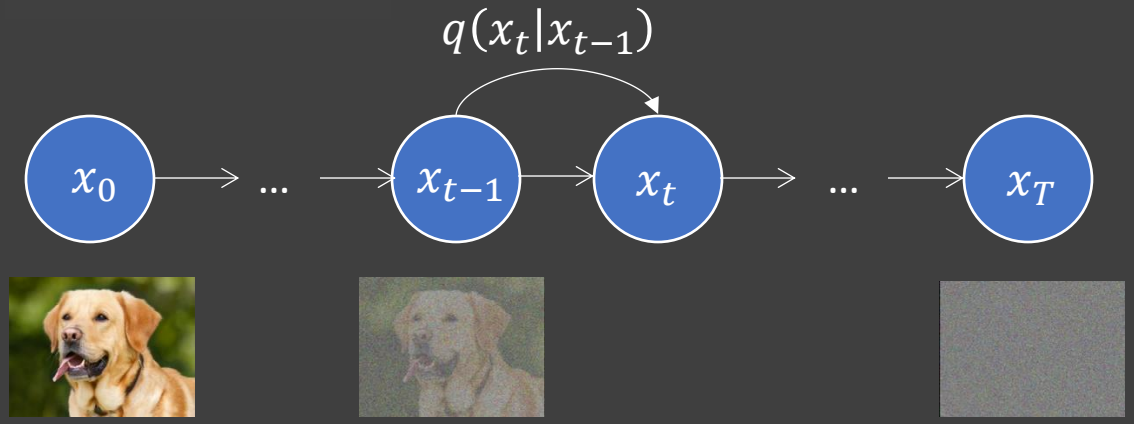
Text-to-image generator → Multimodal



https://medium.com/data-science-in-your-pocket/how-does-dall-e-the-text-to-image-generator-work-c2d9f4a0f26c

Diffusion model

Forward diffusion



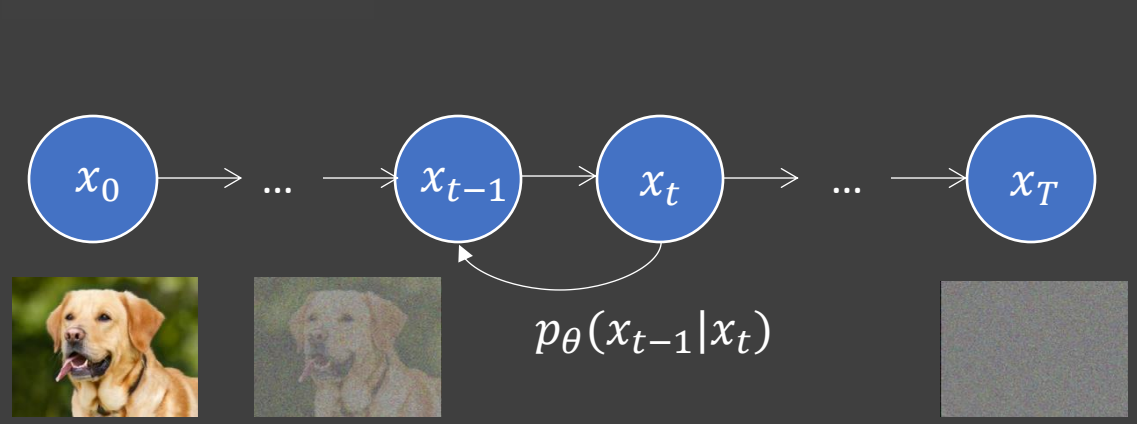
$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \mu_t = \sqrt{1-\beta_t}x_{t-1}, \Sigma_t = \beta_t)$$

Reparameterization trick

$$x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon_0 \qquad \alpha_t = 1-\beta_t \quad \bar{\alpha}_t = \prod_{s=0}^{t}\alpha_s$$

$$\epsilon_t \sim \mathcal{N}(0,1)$$

Reverse diffusion



$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t,t), \Sigma_\theta(x_t,t))$$

Neural Network

This works because the process is Gaussian