

Reti neurali per la generazione di immagini



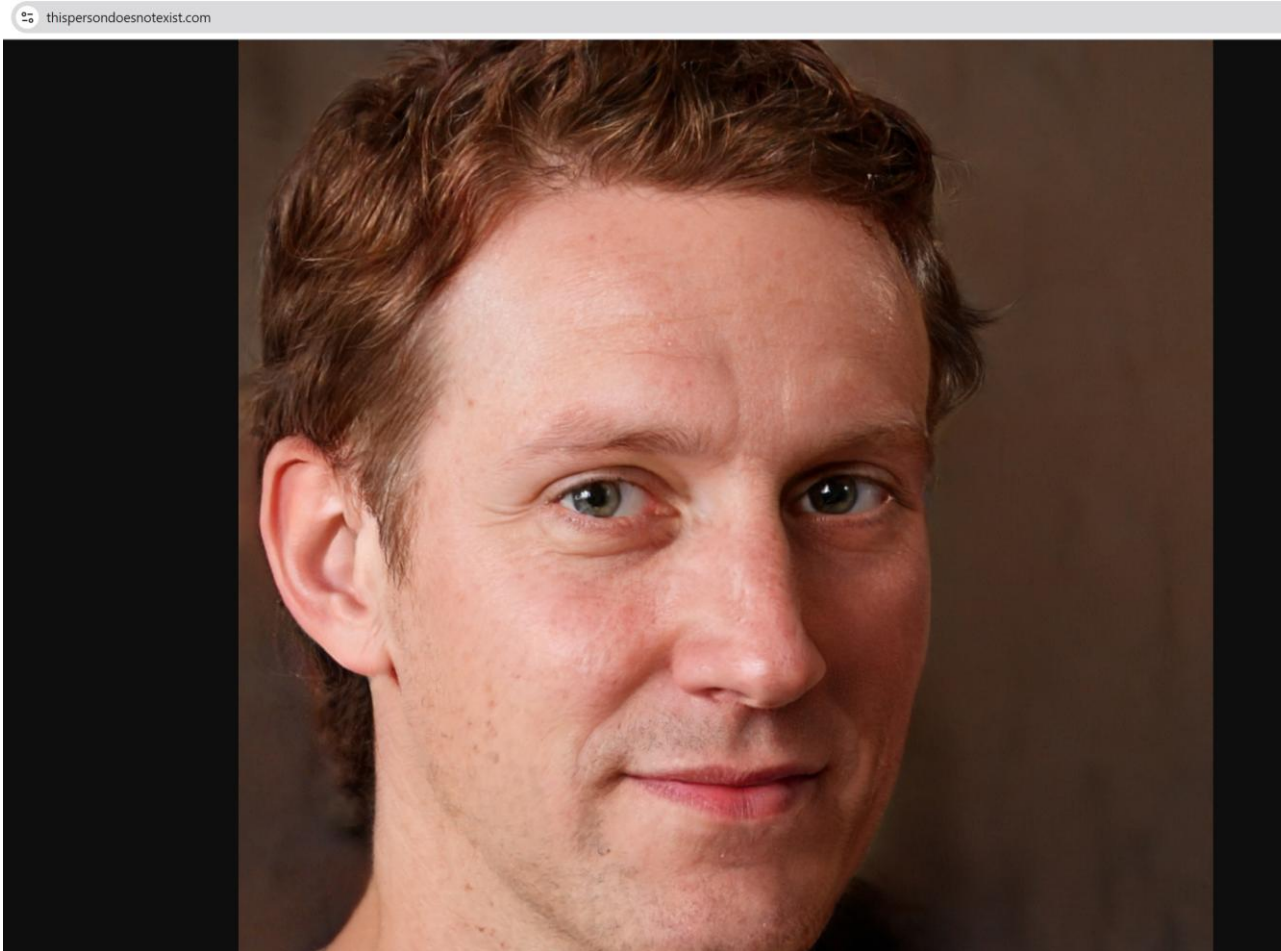
Prof.ssa Lucia Migliorelli

lmigliorelli@unite.it

Corso di Sistemi multimediali e web per il turismo

Dipartimento di Scienze Politiche, Università di Teramo

Generative AI (1)



GAN: imparano a generare immagini “dal nulla” partendo da un **rumore casuale**.

Possono essere condizionate con etichette (es. “cane vs gatto”), ma non con descrizioni in linguaggio naturale.

Sono state fondamentali per la prima generazione di immagini sintetiche realistiche (volti, paesaggi, artefatti deepfake).

Generative AI (2)



Modelli di diffusione (i più usati oggi):

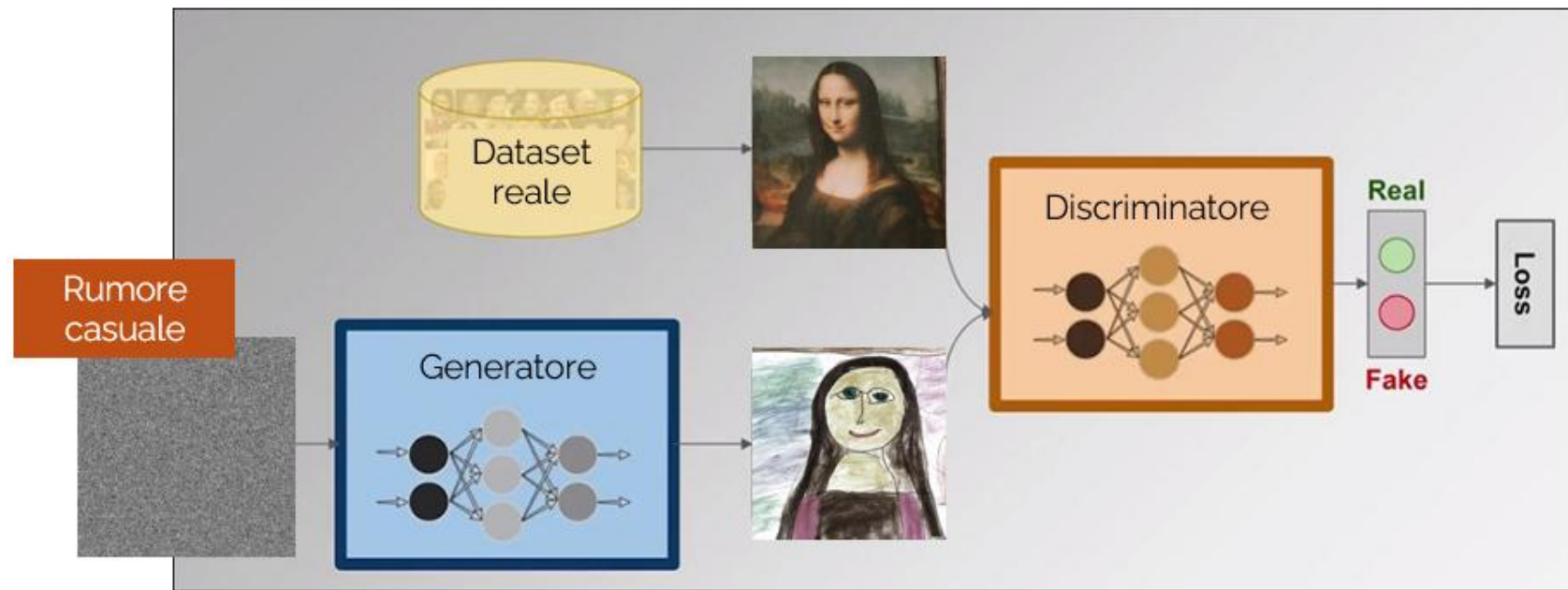
Lavorano a partire da un'immagine rumorosa e "la puliscono" passo dopo passo fino a formare l'immagine richiesta.

Qui l'input può essere un prompt testuale, una foto di partenza → molto più flessibili e controllabili.

insieme di modelli, che apprendono i pattern visivi da grandi dataset e li usano per creare immagini nuove a partire da istruzioni testuali o visive fornite dall'utente.

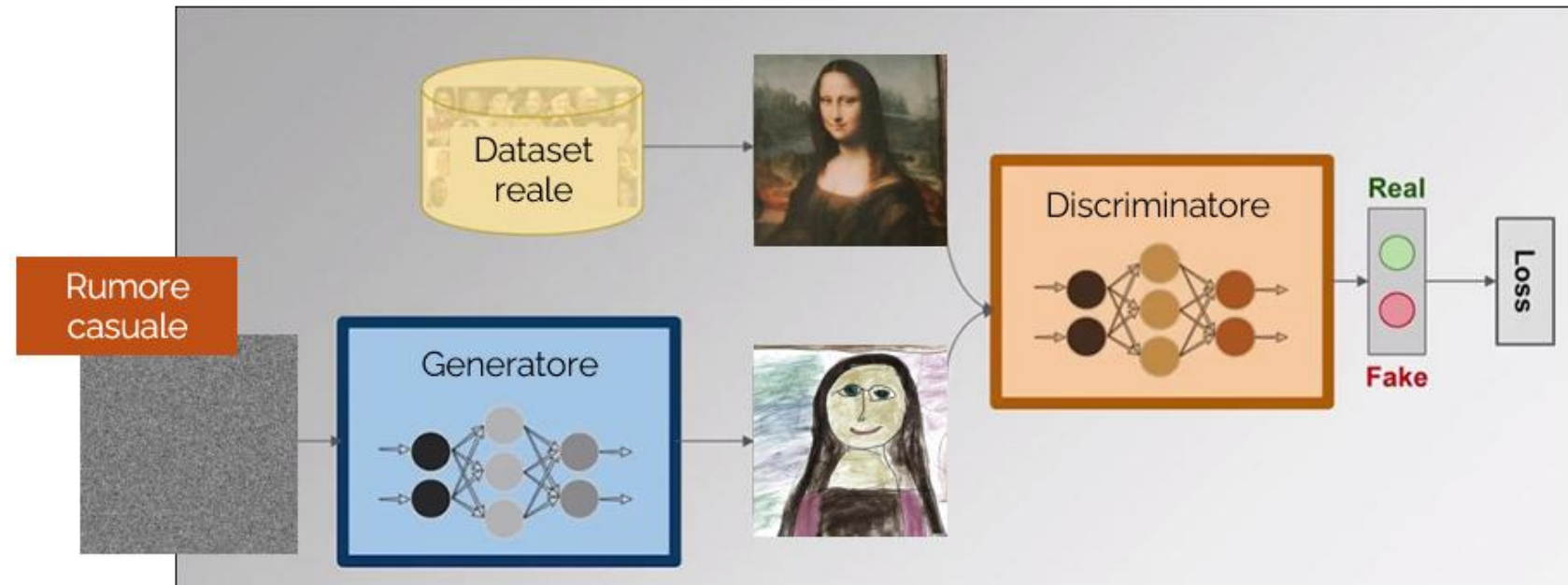
GAN (1)

Fondate sul paradigma di apprendimento senza supervisione (unsupervised learning)



GAN (2)

- Generatore (falsario): crea immagini nuove.
- Discriminatore (poliziotto): controlla se sono "vere" o "false".
- L'apprendimento nasce dal conflitto e dal miglioramento reciproco.



GAN (3)

- Ci sono due personaggi Il falsario 🖋️ → prova a disegnare la gioconda falsa.
 - Il poliziotto 🚔 → deve capire se la gioconda è vera o falsa
- All'inizio:
 - Il falsario disegna malissimo (le gioconde che disegna sembrano scarabocchi).
 - Il poliziotto è bravissimo a smascherare le gioconde false.
- Poi:
 - Ogni volta che il poliziotto smaschera una gioconda falsa, il falsario impara i suoi errori e prova a fare copie più credibili.
 - Ogni volta che il falsario riesce a ingannare il poliziotto, il poliziotto si allena a riconoscere meglio i dettagli veri.
- Con questo gioco a due:
 - il falsario diventa sempre più bravo a generare immagini realistiche (persone, paesaggi, oggetti).
 - il poliziotto diventa sempre più bravo a riconoscere il vero dal falso.
 - Alla fine, il falsario riesce a produrre immagini così convincenti che perfino il poliziotto non riesce più a distinguerle da quelle reali.

Modelli di diffusione (1)

Immaginiamo una goccia di inchiostro che cade in acqua



Modelli di diffusione (2)



Immaginiamo una goccia di inchiostro che cade in acqua

Stato iniziale: lontano dall'equilibrio

Quando la goccia di inchiostro entra nell'acqua, la concentrazione non è uniforme:

vicino alla goccia c'è alta concentrazione di molecole di colorante, lontano dalla goccia c'è solo acqua (concentrazione nulla).

Questo è un sistema fuori dall'equilibrio: esiste un forte gradiente di concentrazione (cioè una differenza di densità di particelle da una zona all'altra).

Modelli di diffusione (3)



Immaginiamo una goccia di inchiostro che cade in acqua

Evoluzione: processo irreversibile

Le molecole di inchiostro cominciano a diffondersi spontaneamente nell'acqua.

Il meccanismo fisico è governato dal moto browniano: le molecole urtano tra loro e con l'acqua, spostandosi in modo casuale.

A livello macroscopico, però, osserviamo un fenomeno ordinato: l'inchiostro tende a spostarsi dalle zone di maggiore concentrazione a quelle di minore concentrazione.

Questo processo aumenta l'entropia (il grado di disordine o di mescolamento).

Modelli di diffusione (4)



Immaginiamo una goccia di inchiostro che cade in acqua

Evoluzione: processo irreversibile

Le molecole di inchiostro cominciano a diffondersi spontaneamente nell'acqua.

Il meccanismo fisico è governato dal moto browniano: le molecole urtano tra loro e con l'acqua, spostandosi in modo casuale.

A livello macroscopico, però, osserviamo un fenomeno ordinato: l'inchiostro tende a spostarsi dalle zone di maggiore concentrazione a quelle di minore concentrazione.

Questo processo aumenta l'entropia (il grado di disordine o di mescolamento).

Perché non vediamo l'inverso?

Modelli di diffusione (5)



La termodinamica del non equilibrio spiega che:

i processi spontanei sono irreversibili:

l'inchiostro si diffonde, ma non torna mai spontaneamente a ricompattarsi in una goccia.

L'unico stato stabile a lungo termine è quello di equilibrio, dove la concentrazione è uniforme in tutto il bicchiere.

Se mai vedessimo il contrario (l'inchiostro che si ricompatta da solo), significherebbe che l'entropia diminuisce in modo spontaneo — ma questo è proibito dalla seconda legge della termodinamica.

Modelli di diffusione (6)

La goccia di inchiostro che si diffonde è un esempio di come un sistema passi da uno stato di **ordine locale** (goccia compatta) a uno stato di **disordine uniforme** (acqua colorata), aumentando l'entropia.

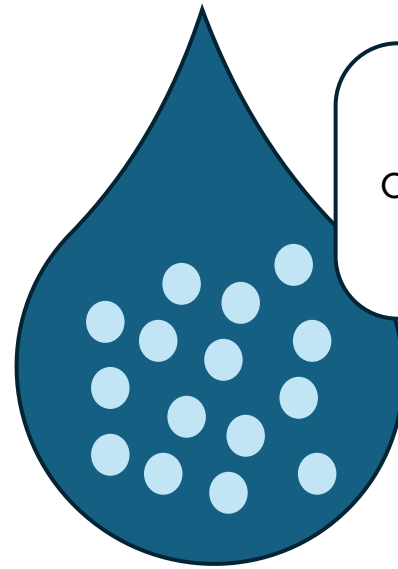
La **termodinamica del non equilibrio** ci permette di descrivere matematicamente questo cammino verso l'equilibrio che è:

Irreversibile → non assistiamo mai al processo inverso ovvero dall'inchiostro diffuso si torni alla singola goccia)



Modelli di diffusione (7)

Perché non esiste il processo inverso?



La goccia è un insieme di particelle con dei legami non così forti: possono separarsi facilmente

Modelli di diffusione (8)

Perché non esiste il processo inverso?

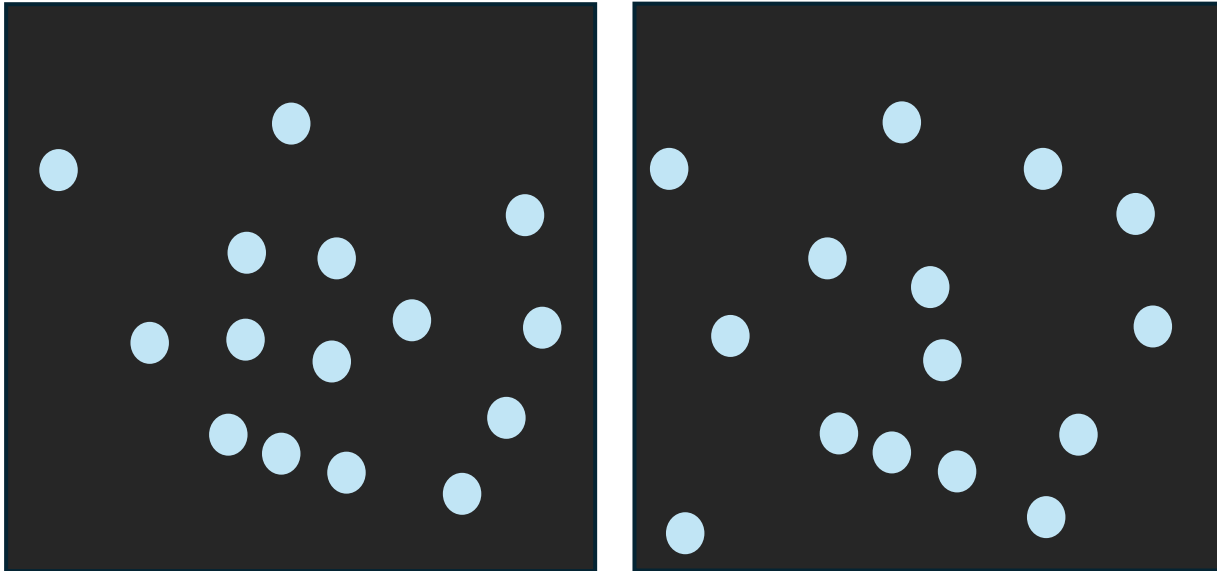


Nell'acqua le particelle iniziano a muoversi a caso

Si chiama **camminata casuale**: ogni particella nell'acqua si sposta in una direzione diversa senza un ordine preciso

Modelli di diffusione (9)

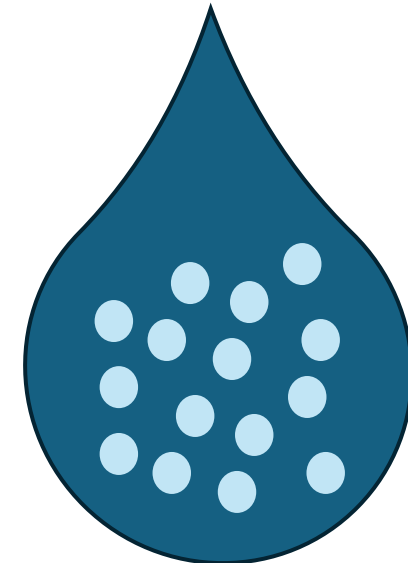
Perché non esiste il processo inverso?



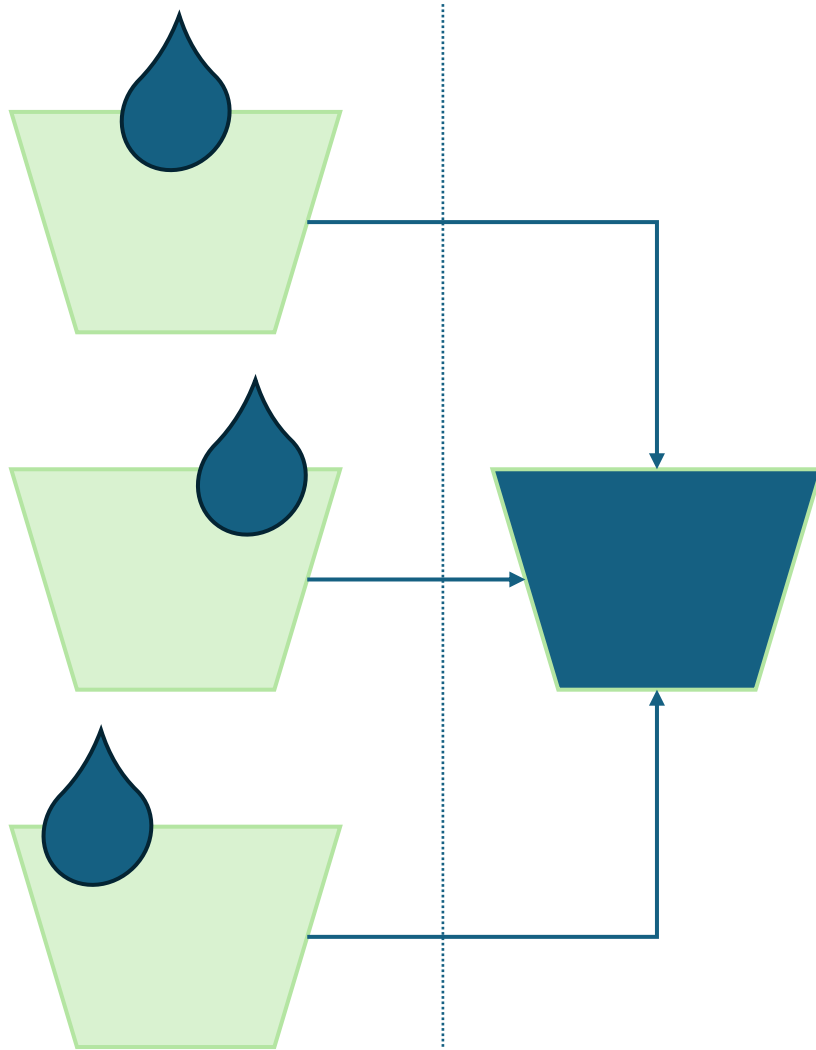
Se tutte le particelle restassero compatte in una goccia, le configurazioni possibili sarebbero poche.

Se invece si spargono in tutto il bicchiere, le combinazioni possibili diventano enormi.

Per questo motivo, è molto più probabile che le particelle si diffondano piuttosto che rimanere unite: la natura sceglie quasi sempre la situazione con **più possibilità** (cioè con più **disordine**, più entropia).

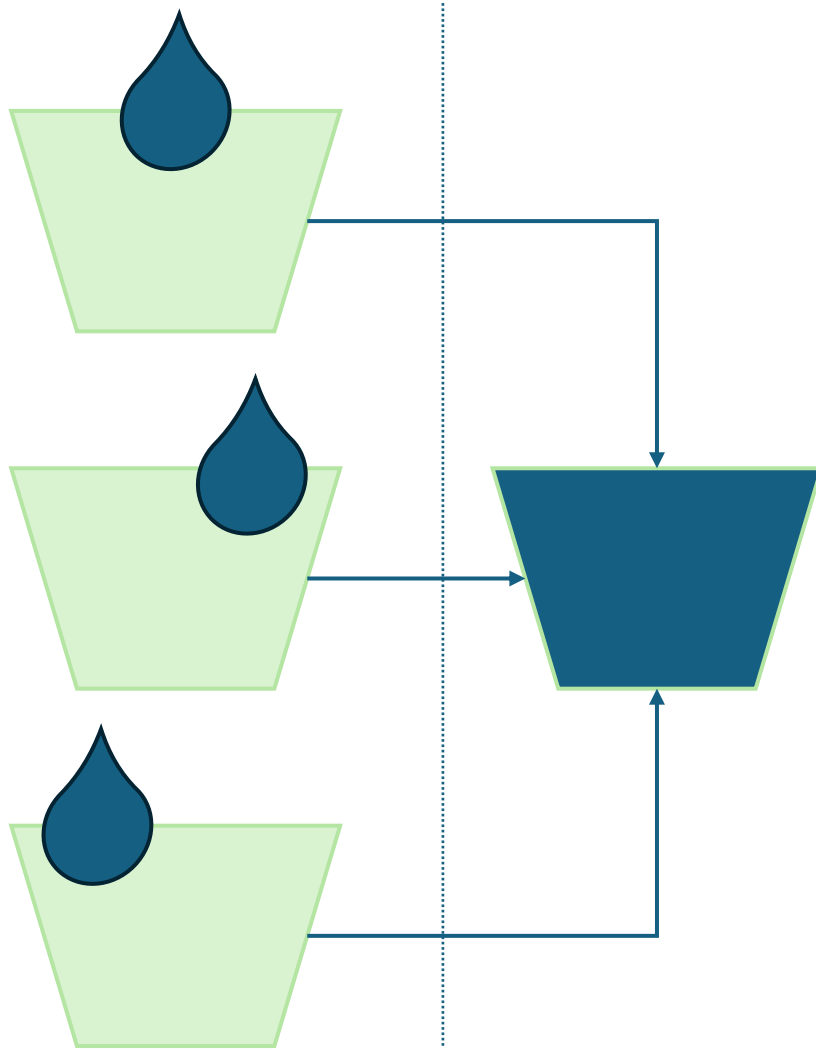


Modelli di diffusione (10)



Quando si butta una goccia di inchiostro nell'acqua non importa il punto preciso in cui cade la goccia: il risultato finale è sempre lo stesso ovvero l'inchiostro si mescola in acqua

Modelli di diffusione (11)

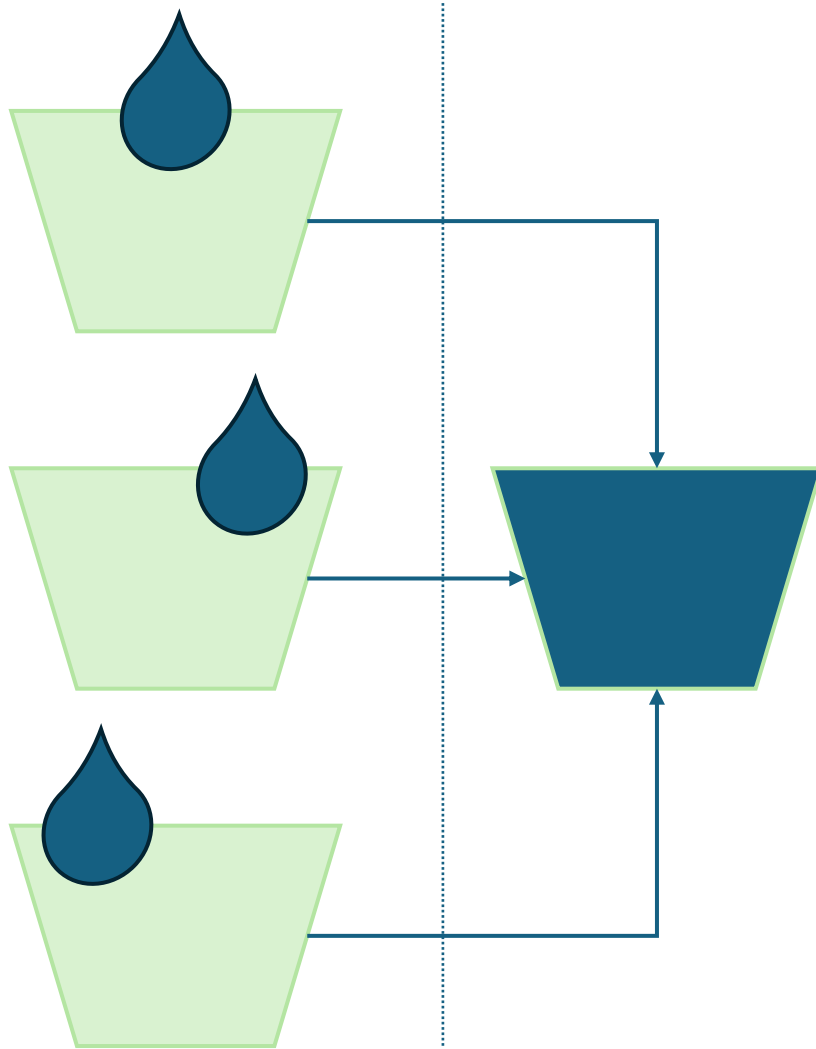


Quando si butta una goccia di inchiostro nell'acqua non importa il punto preciso in cui cade la goccia: il risultato finale è sempre lo stesso ovvero l'inchiostro si mescola in acqua

Trattasi di probabilità:

- È "statisticamente" quasi certo che l'inchiostro si diffonda.
 - Il contrario (che le particelle tornino tutte insieme formando di nuovo una goccia) non è impossibile in senso assoluto, ma è talmente improbabile che nella realtà non lo vedremo mai

Modelli di diffusione (11)



Quando si butta una goccia di inchiostro nell'acqua non importa il punto preciso in cui cade la goccia: il risultato finale è sempre lo stesso: avere l'inchiostro si mescola in acqua

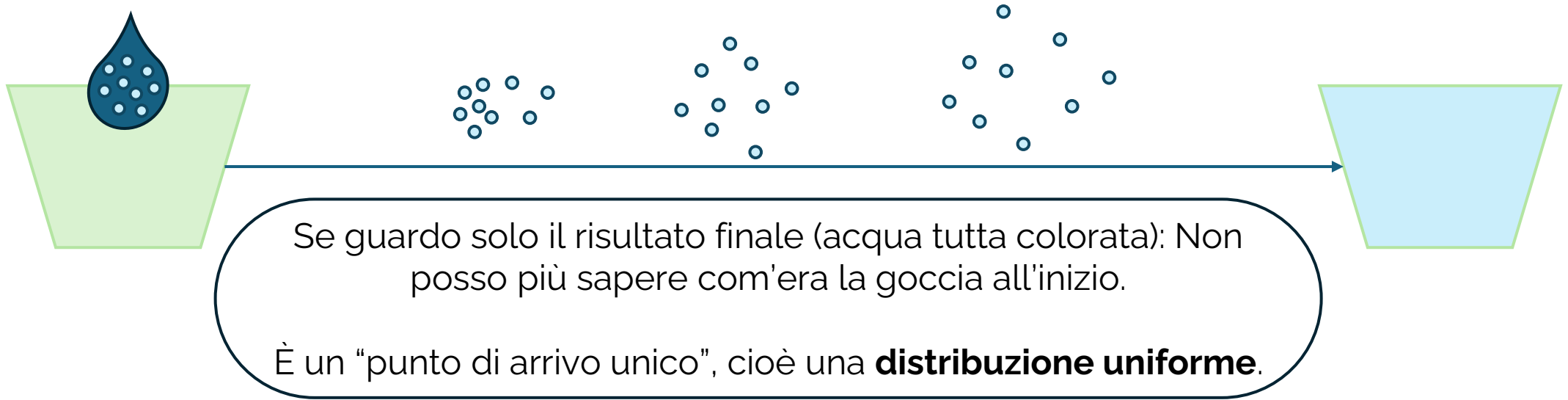
Proprio perché tanti stati iniziali portano allo stesso stato finale, guardando l'acqua colorata non possiamo più sapere com'era la goccia all'inizio.



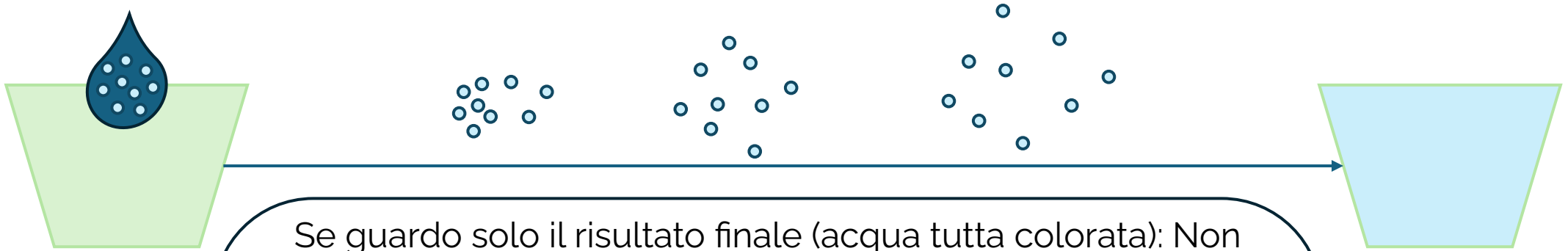
Questo è il senso di **irreversibilità**: si può andare solo in avanti (diffusione), non indietro (ricompattamento).

Anche se all'inizio la goccia cade in posizioni o forme diverse, dopo un po' il bicchiere avrà sempre lo stesso aspetto: acqua colorata in modo uniforme.

Intuizione dietro ai modelli diffusivi (1)



Intuizione dietro ai modelli diffusivi (2)



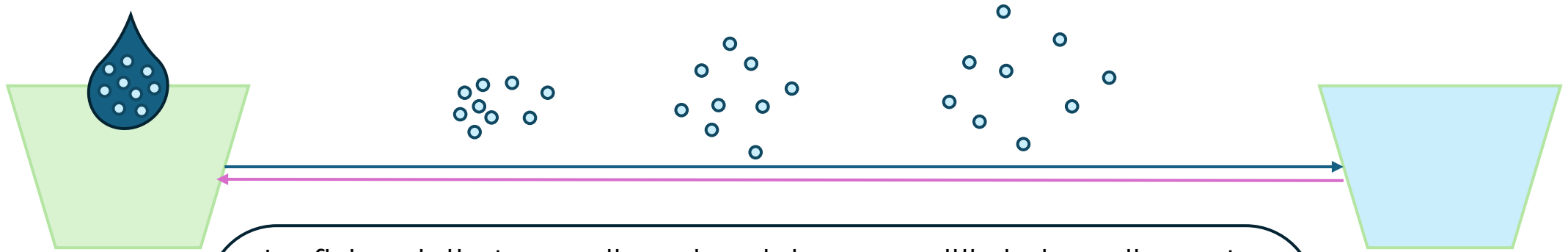
Se guardo solo il risultato finale (acqua tutta colorata): Non posso più sapere com'era la goccia all'inizio.

È un “punto di arrivo unico”, cioè una **distribuzione uniforme**.

Se invece osservo una configurazione particolare (ad esempio la goccia caduta proprio al centro): Posso descrivere passo per passo come le particelle si muovono e si diffondono.

Posso farlo sia in avanti (dal momento in cui cade fino a che si diffonde) sia “all'indietro” (ricostruendo il percorso della diffusione fino alla goccia iniziale).

Intuizione dietro ai modelli diffusivi (3)



La fisica della termodinamica del non equilibrio ha sviluppato strumenti matematici che permettono di simulare questo processo, cioè di passare da una distribuzione iniziale complicata (la goccia compatta) a quella finale semplice (uniforme).

E viceversa: ricostruire a ritroso come si è arrivati lì.

Diffusion
models

- L'idea è: parto da rumore (equivalente alla distribuzione finale uniforme → molto semplice) e, passo dopo passo, ricostruisco a ritroso un'immagine ordinata (equivalente alla goccia iniziale → distribuzione molto complessa).

Il metodo di diffusione (1)

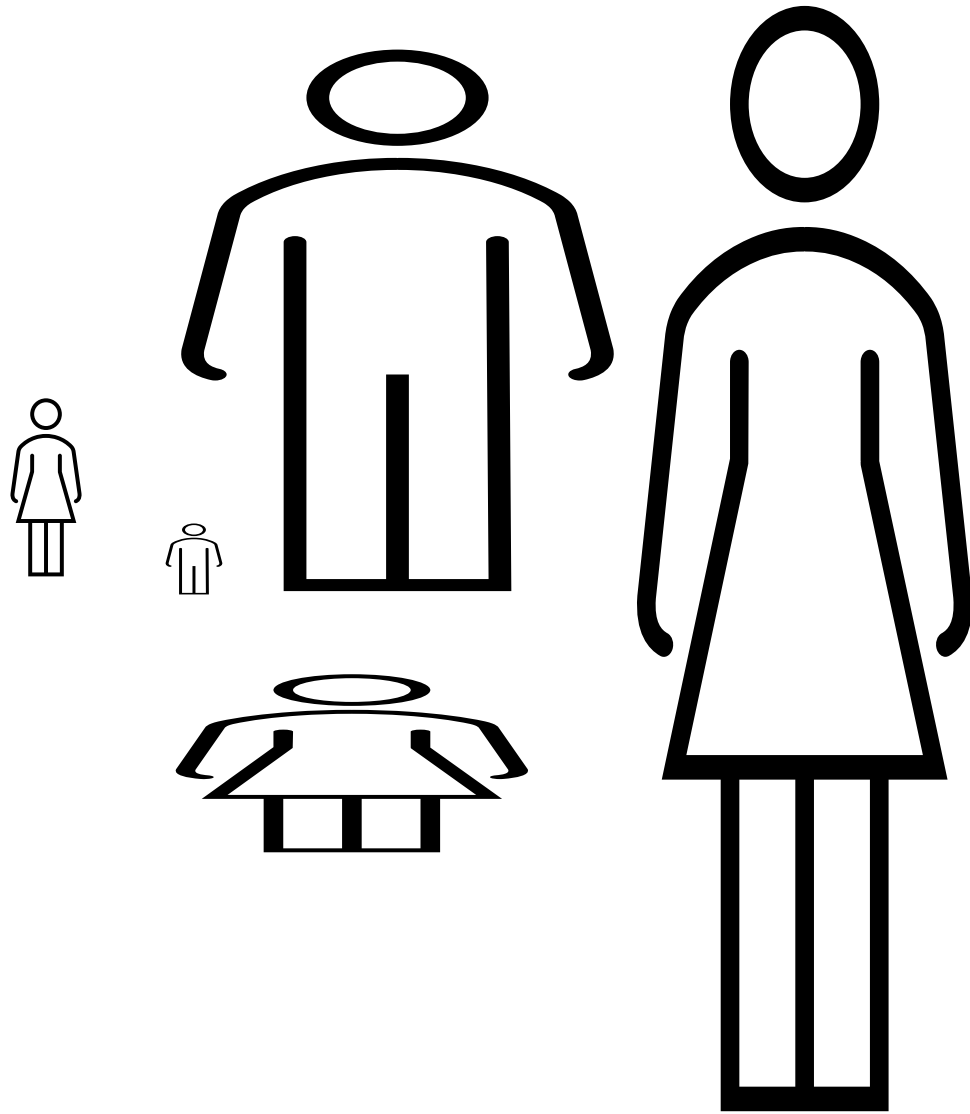
Vogliamo creare con una rete generativa
l'immagine di un essere umano

Il metodo di diffusione (2)

Vogliamo creare con una rete generativa l'immagine di un essere umano

Per riuscirci la rete deve sapere come «sono fatti» gli esseri umani (altezza, peso, proporzioni, ...)

Il metodo di diffusione (3)

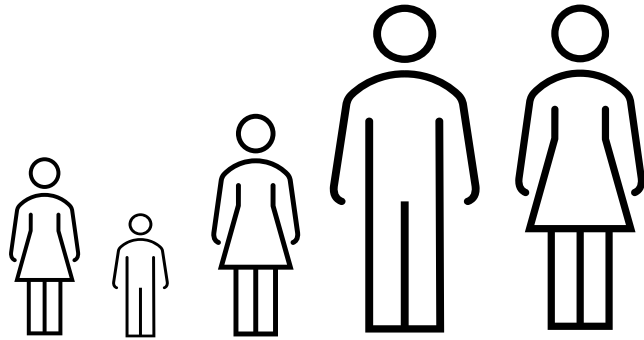


Vogliamo creare con una rete generativa l'immagine di un essere umano

Per riuscirci la rete deve sapere come «sono fatti» gli esseri umani (altezza, peso, proporzioni, ...)

Alcune immagini possono sembrarci sbagliate: persone troppo alte e strette, o troppo larghe e basse → proporzioni non realistiche.

Il metodo di diffusione (4)



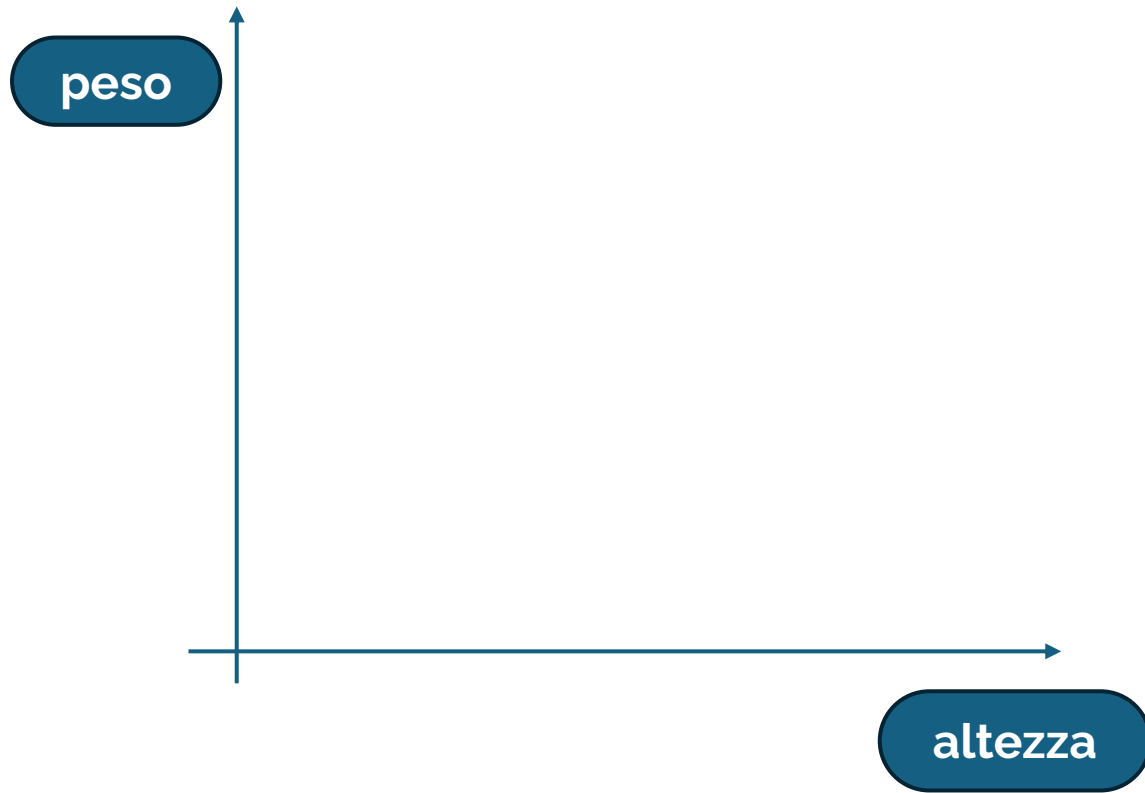
Vogliamo creare con una rete generativa l'immagine di un essere umano

Per riuscirci la rete deve sapere come «sono fatti» gli esseri umani (altezza, peso, proporzioni, ...)

Alcune immagini possono sembrarci sbagliate: persone troppo alte e strette, o troppo larghe e basse → proporzioni non realistiche.

Altre invece ci sembrano corrette, perché rispettano i rapporti normali di un corpo umano.

Il metodo di diffusione (5)



Vogliamo creare con una rete generativa l'immagine di un essere umano

Per riuscirci la rete deve sapere come «sono fatti» gli esseri umani (altezza, peso, proporzioni, ...)

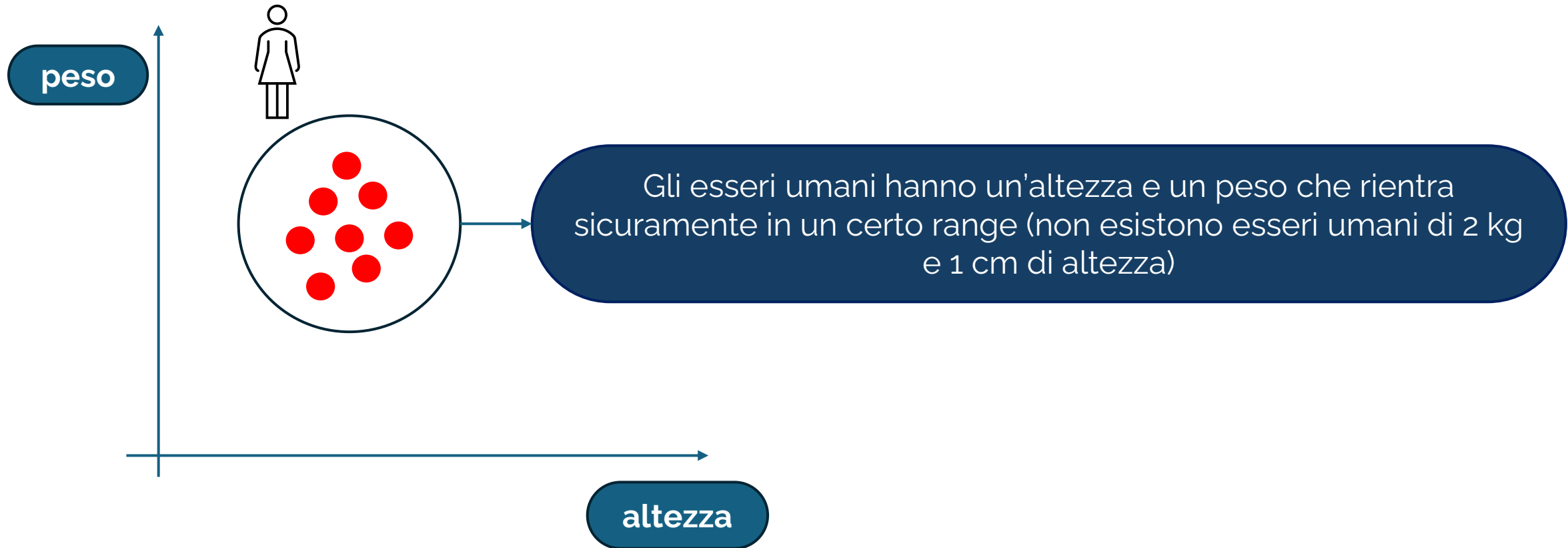
Alcune immagini possono sembrarci sbagliate: persone troppo alte e strette, o troppo larghe e basse → proporzioni non realistiche.

Altre invece ci sembrano corrette, perché rispettano i rapporti normali di un corpo umano.

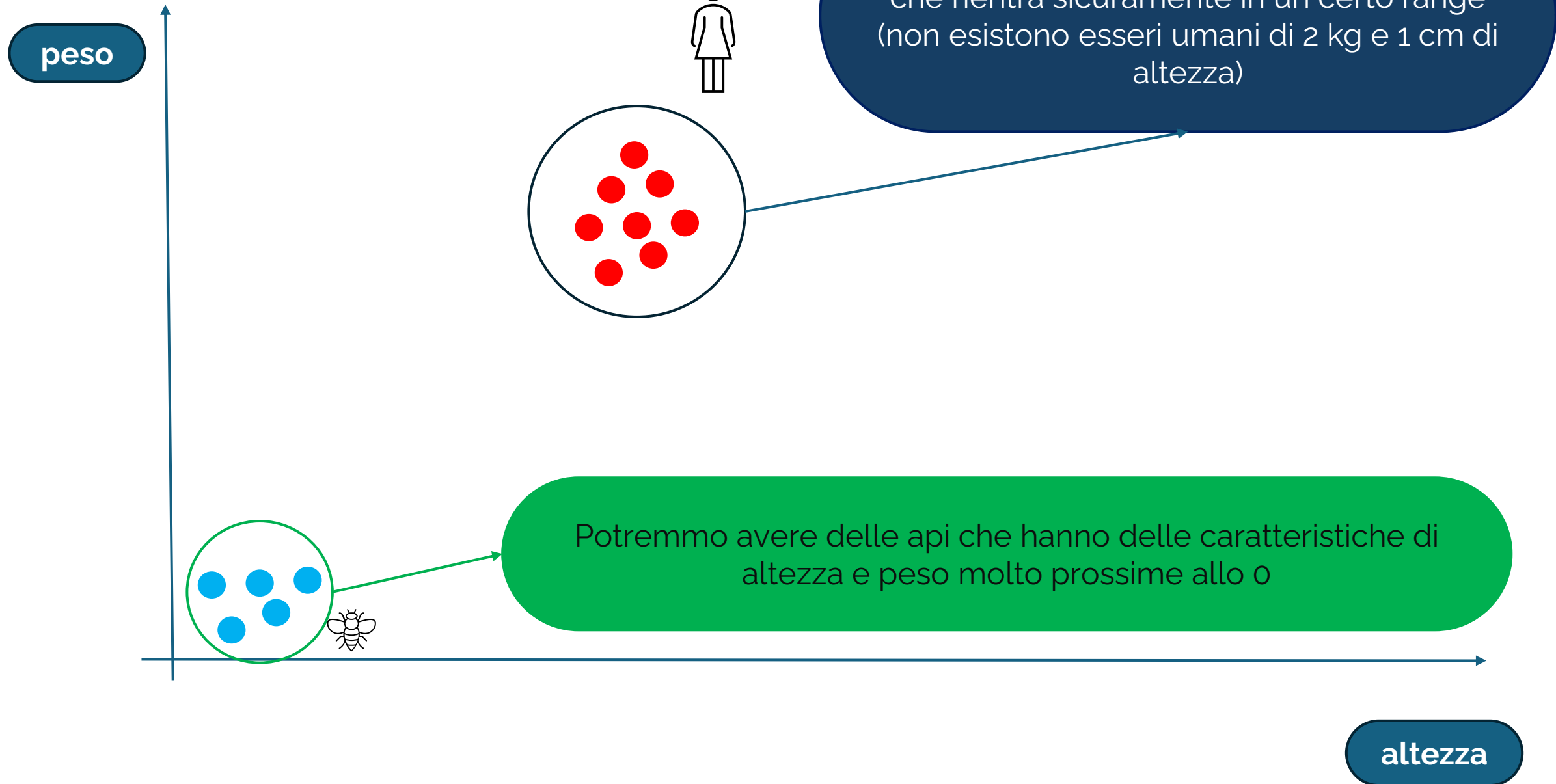
Possiamo immaginare che ogni essere umano sia rappresentato da un punto in un grafico: in questo spazio ogni combinazione plausibile di altezza e peso corrisponde a un essere umano realistico.

Il metodo di diffusione (6)

Possiamo immaginare che ogni essere umano sia rappresentato da un punto in un grafico: in questo spazio ogni combinazione plausibile di altezza e peso corrisponde a un essere umano realistico.



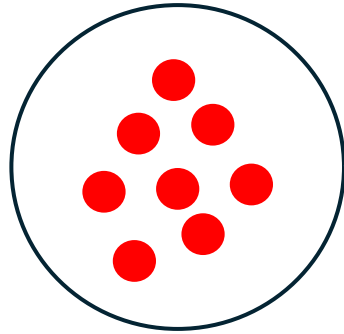
Il metodo di diffusione (7)



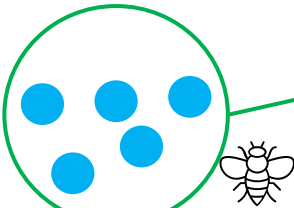
Il metodo di diffusione (8)

peso

Caratteristiche
come peso ed
altezza
permettono di
distinguere api da
persone



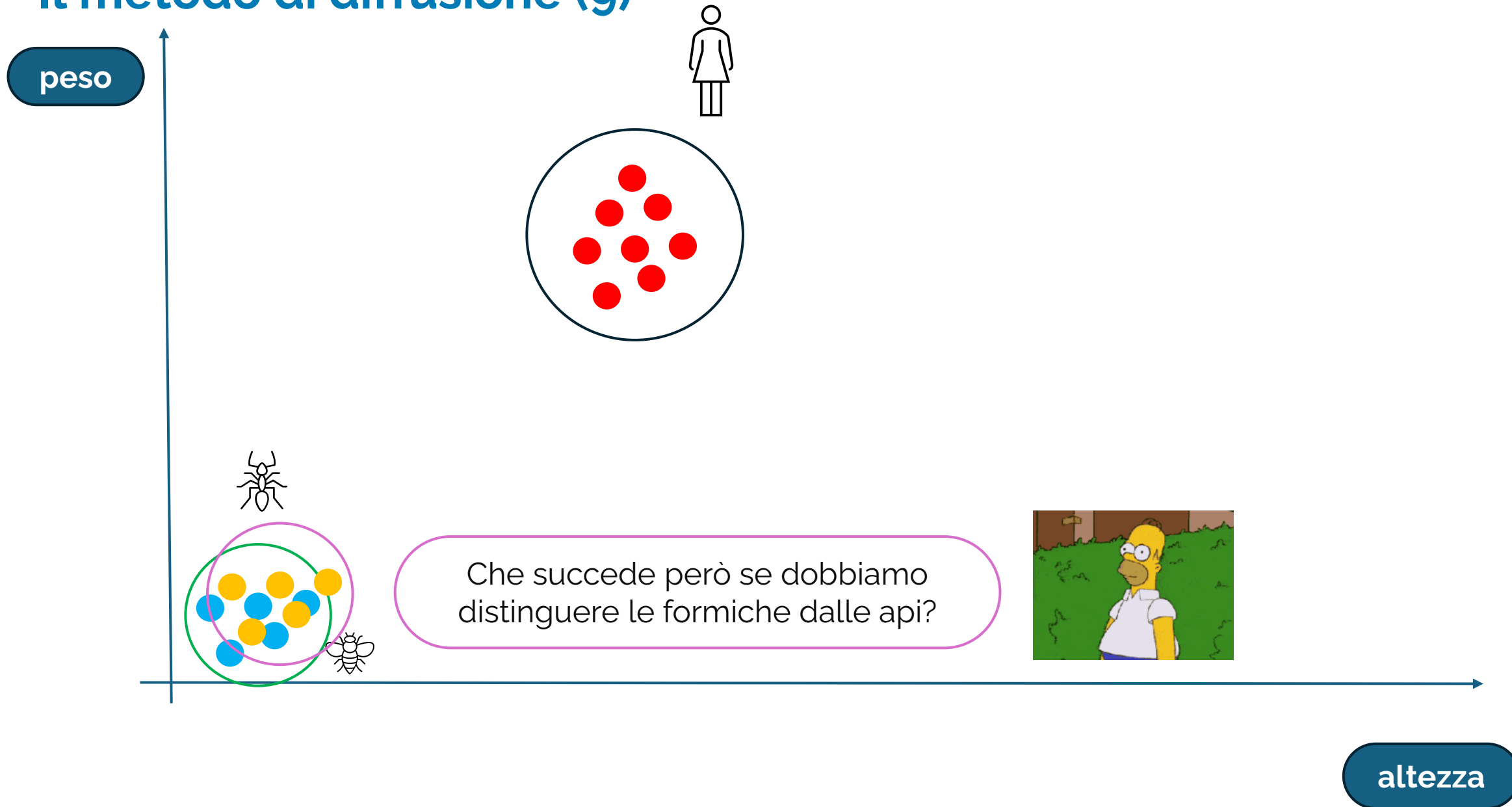
Gli esseri umani hanno un'altezza e un peso
che rientra sicuramente in un certo range
(non esistono esseri umani di 2 kg e 1 cm di
altezza)



Potremmo avere delle api che hanno delle caratteristiche di
altezza e peso molto prossime allo 0

altezza

Il metodo di diffusione (g)

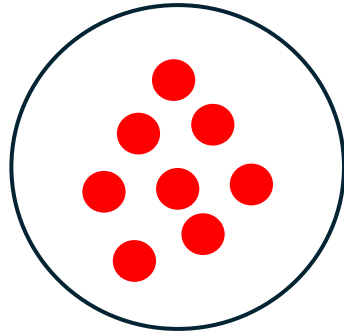


Il metodo di diffusione (10)



Il metodo di diffusione (11)

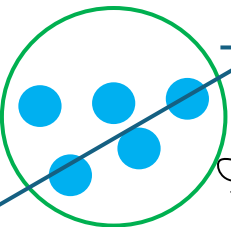
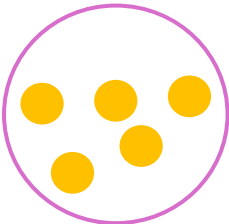
peso



Che succede però se dobbiamo distinguere le lumache?



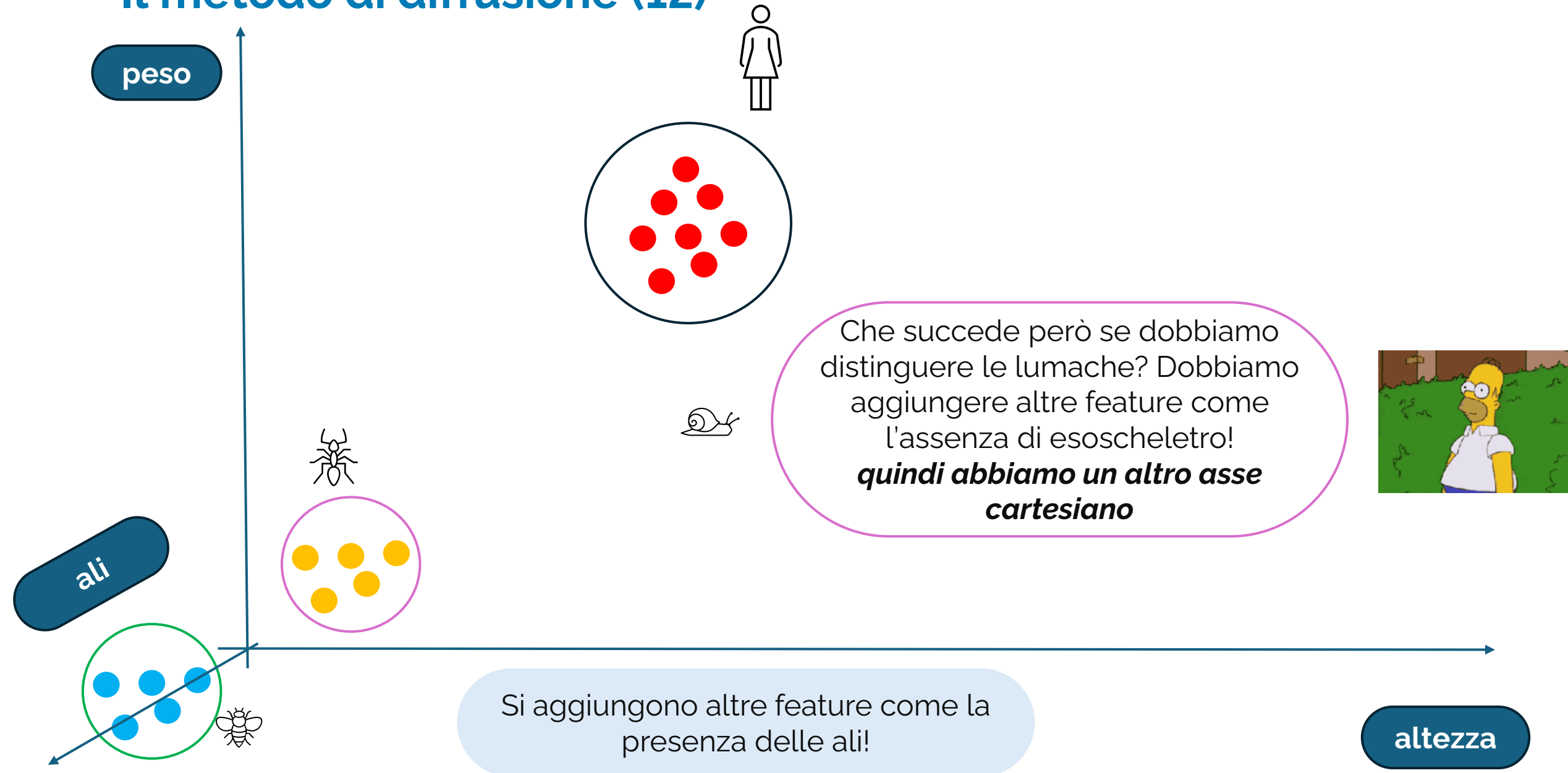
ali



Si aggiungono altre feature come la presenza delle ali!

altezza

Il metodo di diffusione (12)



Diffusione e distribuzione di probabilità (1)

peso



Lo spazio delle feature è
MULTIDIMENSIONALE.

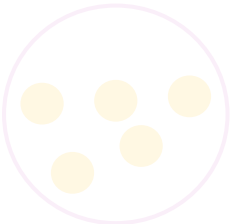
Sia che si abbiano reti di classificazione sia
che si lavori con reti per la generazione di
immagini la sfida dell'addestramento sta
proprio nella necessità di

**muoversi in uno spazio
multidimensionale**

se dobbiamo
? Dobbiamo
ure come
eletro!
tro asse

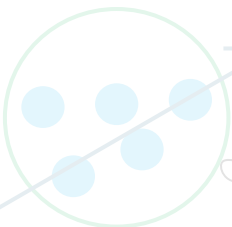


ali

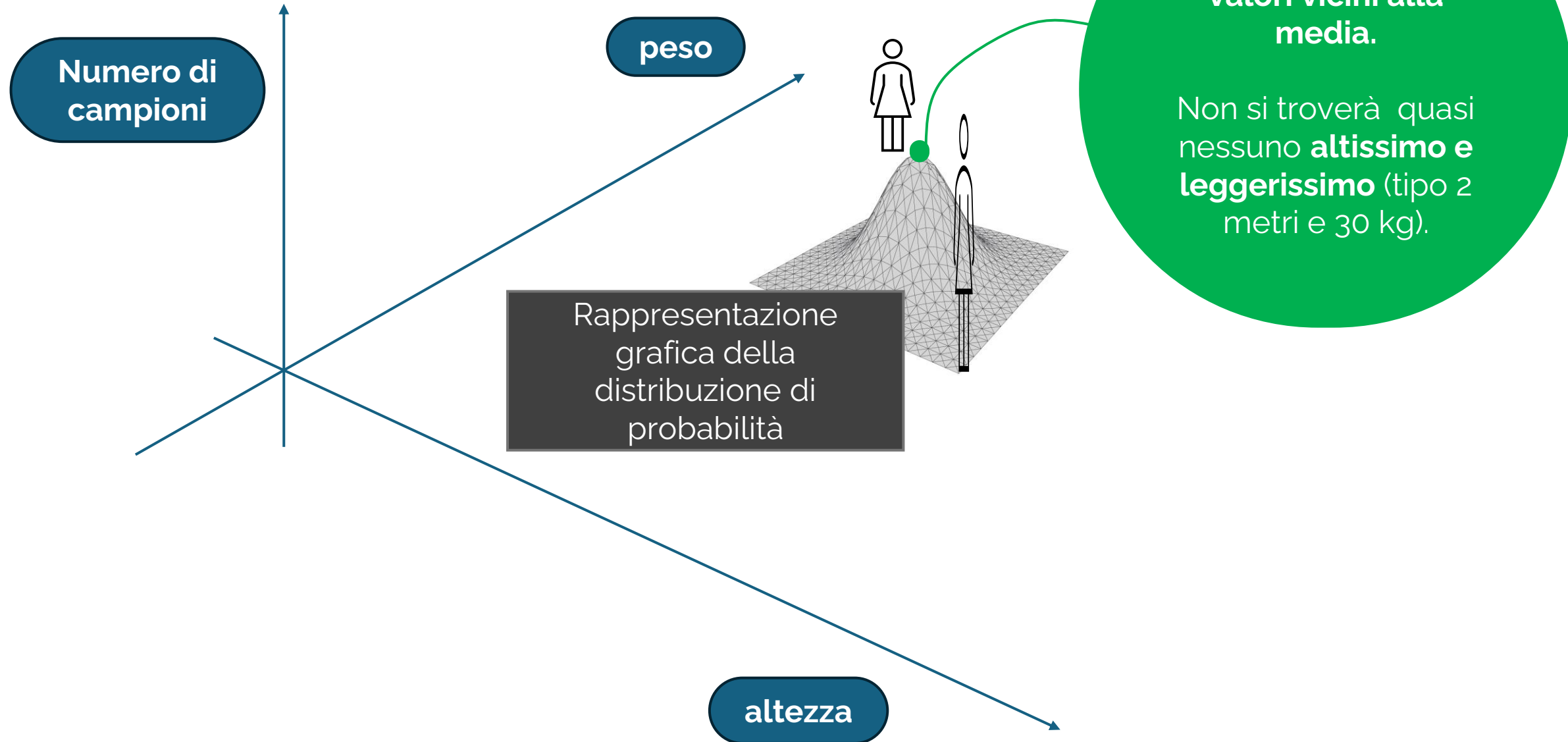


Si aggiungono altre feature come la
presenza delle ali!

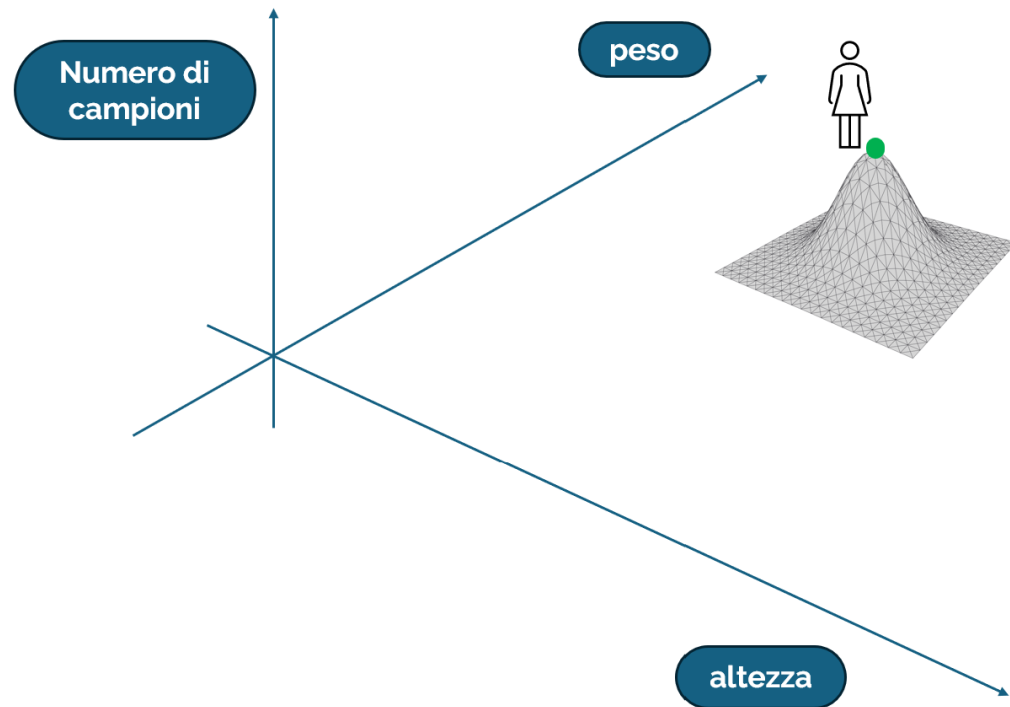
altezza



Diffusione e distribuzione di probabilità (2)



Diffusione e distribuzione di probabilità (3)



Un'**intelligenza artificiale generativa** non può inventare immagini "a caso":

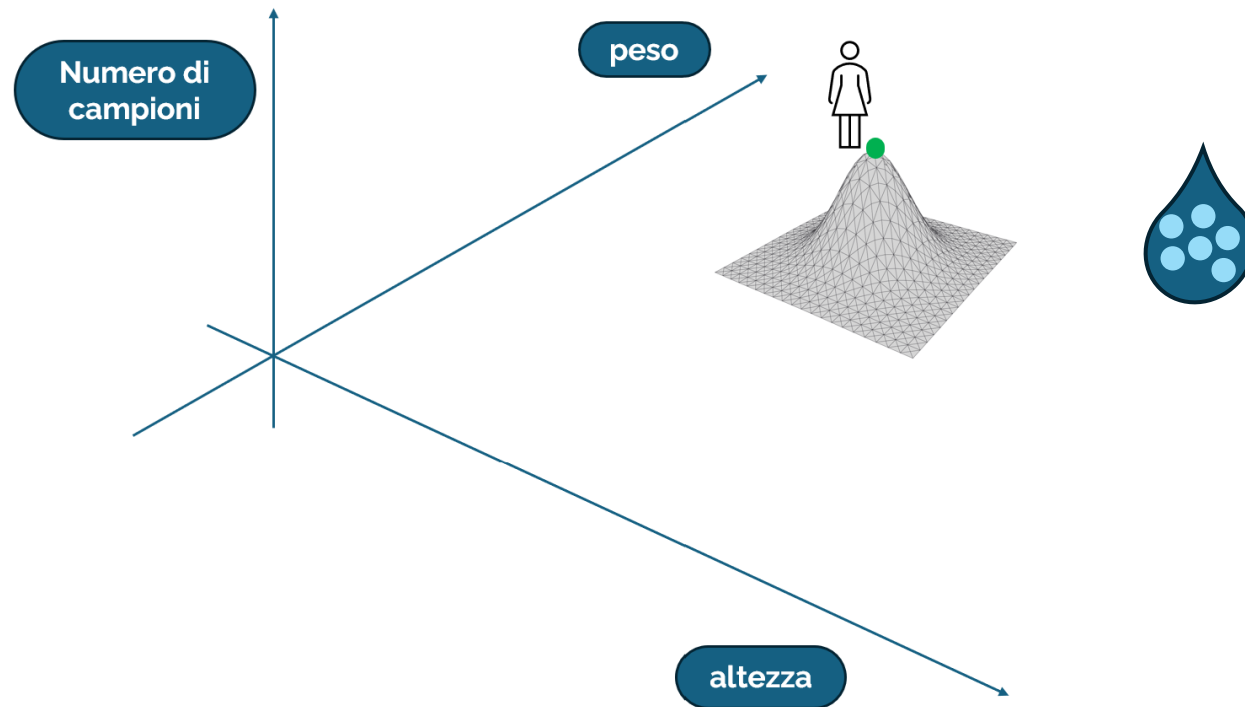
Se deve creare un nuovo essere umano, deve sapere **quali combinazioni di altezza, peso e altre caratteristiche** sono **più probabili**.

Per questo, deve imparare la **distribuzione di probabilità** dei dati reali.

Se l'IA genera un nuovo campione vicino al picco, sarà **riconoscibile come umano realistico**.

Se invece genera campioni troppo lontani (nelle code o fuori distribuzione), rischia di creare immagini strane e non credibili.

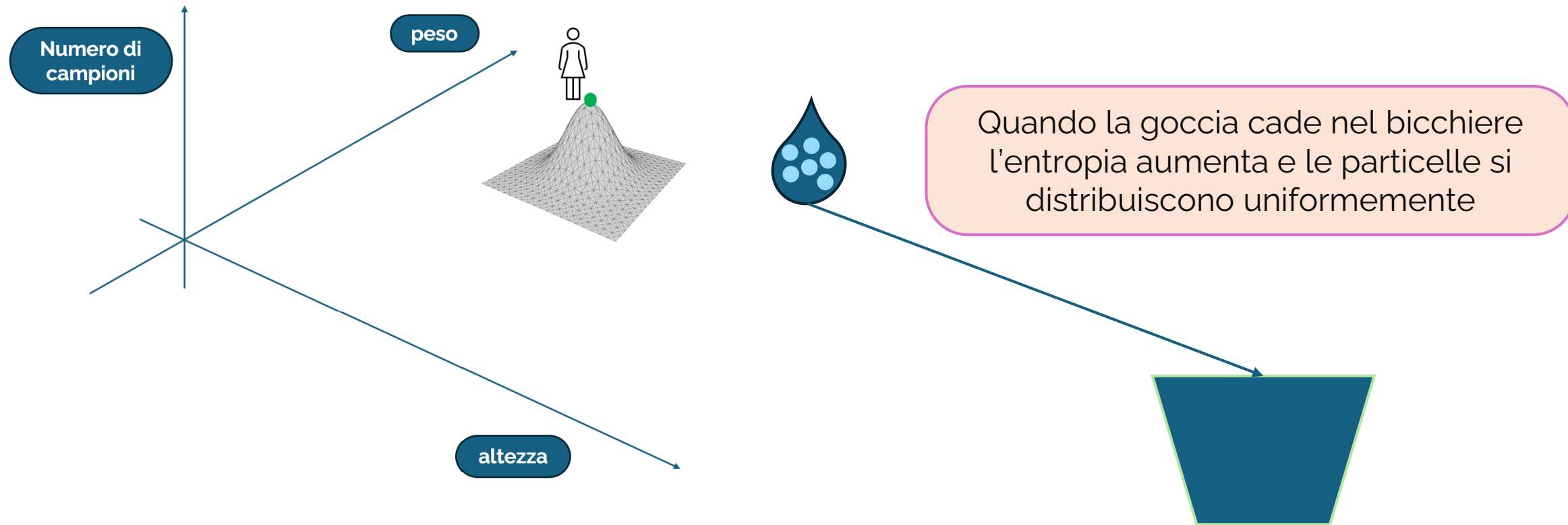
Modelli diffusivi e termodinamica di non equilibrio (1)



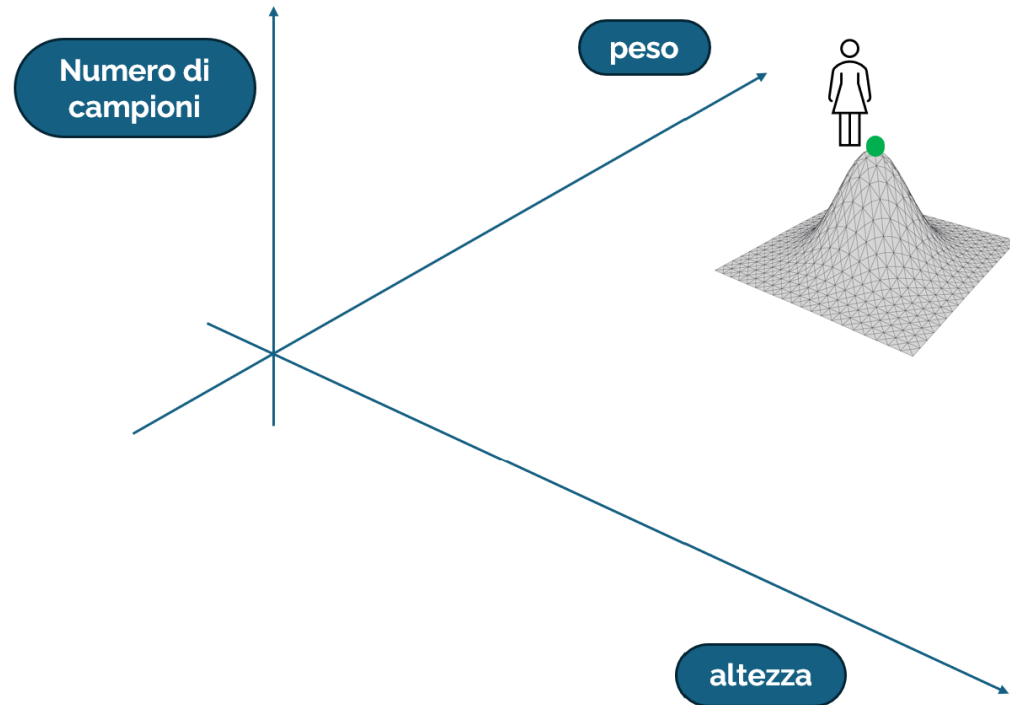
Possiamo assimilare la distribuzione di probabilità associata alla «classe» persona con la distribuzione di probabilità delle particelle nella goccia

(entrambe sono accomunate da bassa entropia e molta struttura)

Modelli diffusivi e termodinamica di non equilibrio (2)



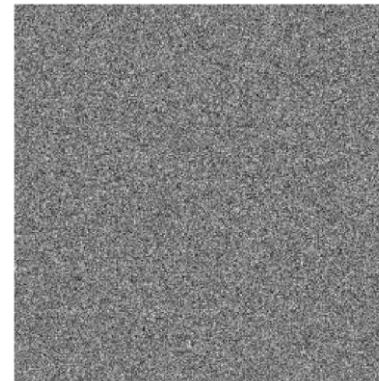
Modelli diffusivi e termodinamica di non equilibrio (3)



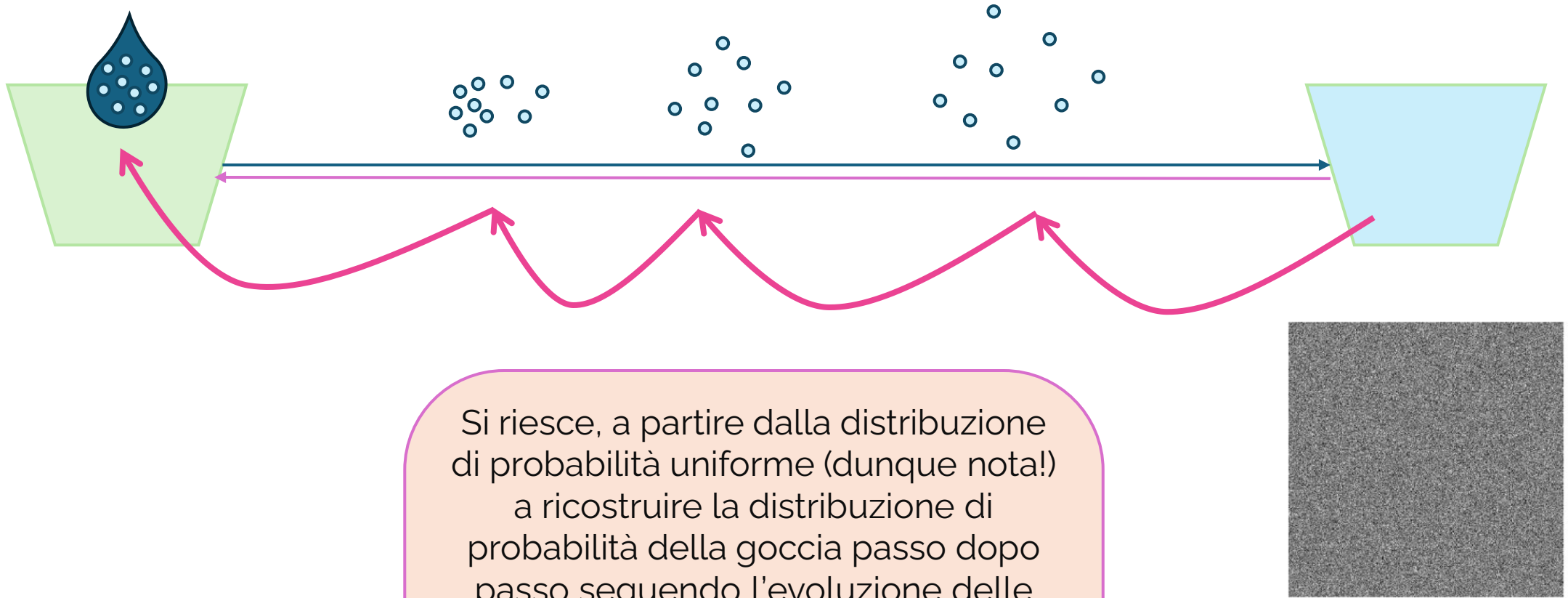
Se dovessimo fare un parallelismo la distribuzione uniforme rappresenta rumore bianco, assenza di immagine, pixel con livelli di grigio casuale



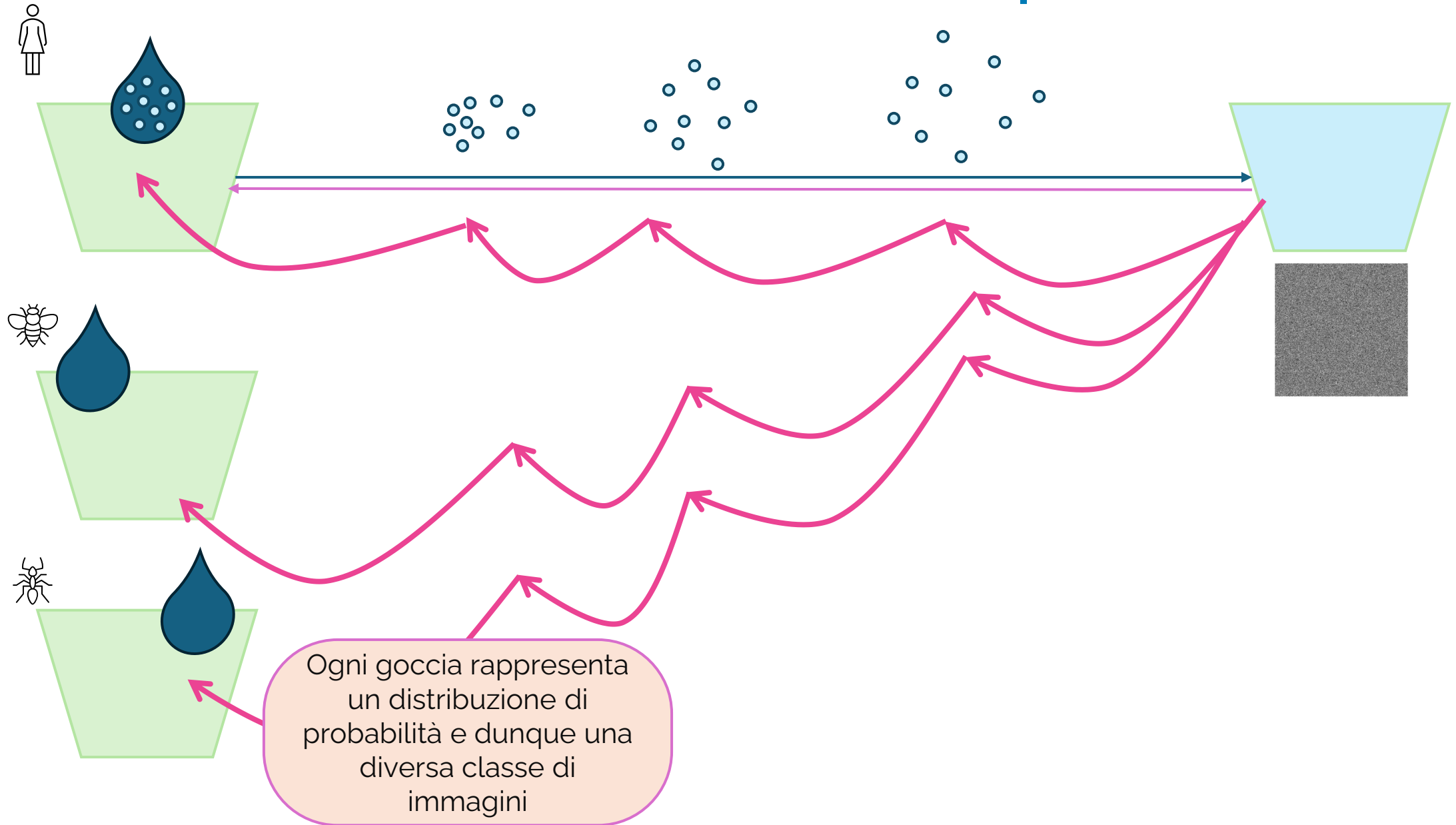
Quando la goccia cade nel bicchiere l'entropia aumenta e le particelle si distribuiscono uniformemente



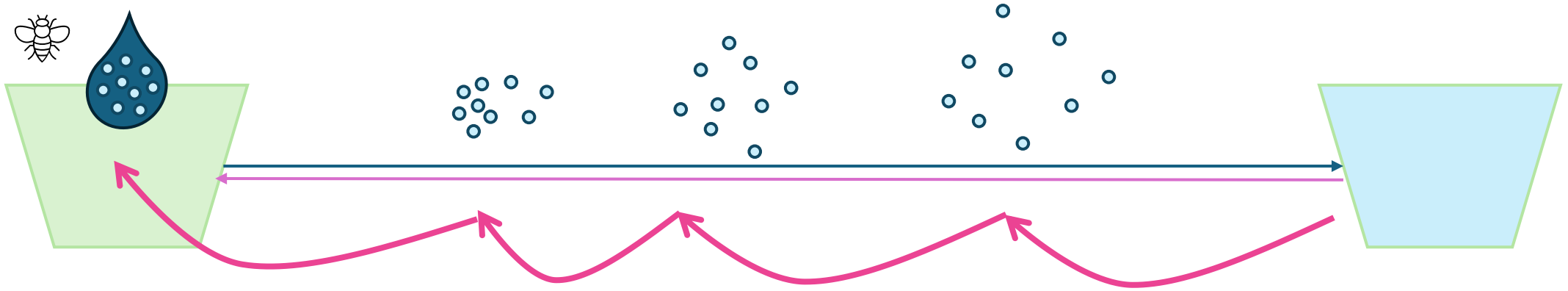
Modelli diffusivi e termodinamica di non equilibrio (4)



Modelli diffusivi e termodinamica di non equilibrio (4)



Forward pass & reverse diffusion



Deep Unsupervised Learning using Nonequilibrium Thermodynamics

Jascha Sohl-Dickstein
Stanford University

Eric A. Weiss
University of California, Berkeley

Niru Maheswaranathan
Stanford University

Surya Ganguli
Stanford University

Abstract

A central problem in machine learning involves modeling complex data-sets using highly flexible families of probability distributions in which learning, sampling, inference, and evaluation are still analytically or computationally tractable. Here, we develop an approach that simultaneously achieves both flexibility and tractability. The essential idea, inspired by non-equilibrium statistical physics, is to systematically and slowly destroy structure in a data distribution through an iterative forward diffusion process. We then learn a reverse diffusion process that restores structure in data, yielding a highly flexible and tractable generative model of the data. This approach allows us to rapidly learn, sample from, and evaluate probabilities in deep generative models with thousands of layers or time steps, as well as to compute conditional and posterior probabilities under the learned model. We addi-

- «Distruggere» gradualmente e sistematicamente la struttura di una distribuzione di dati attraverso un processo di diffusione in avanti, eseguito in modo iterativo.
- Successivamente, si apprende un processo di diffusione inverso, che consente di ricostruire la struttura originale dei dati.

Forward pass – modelli diffusivi (1)

- **Forward pass)** «Distruggere» gradualmente e sistematicamente la struttura di una distribuzione di dati attraverso un processo di diffusione in avanti, eseguito in modo iterativo.

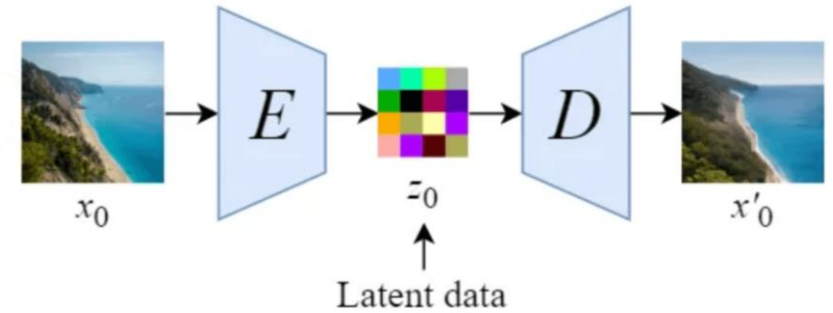
$t=1$



+



=



Forward pass – modelli diffusivi (2)

- **Forward pass)** «Distruggere» gradualmente e sistematicamente la struttura di una distribuzione di dati attraverso un processo di diffusione in avanti, eseguito in modo iterativo.

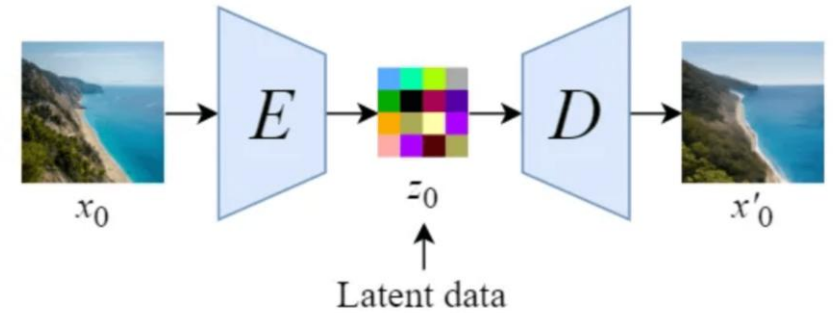
$t=2$



+

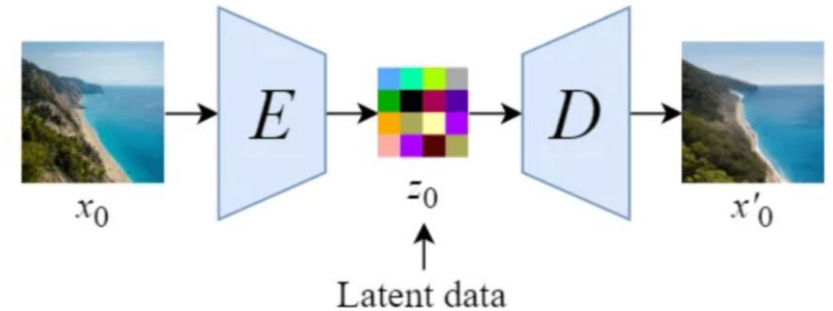


=



Forward pass – modelli diffusivi (3)

- **Forward pass)** «Distruggere» gradualmente e sistematicamente la struttura di una distribuzione di dati attraverso un processo di diffusione in avanti, eseguito in modo iterativo.



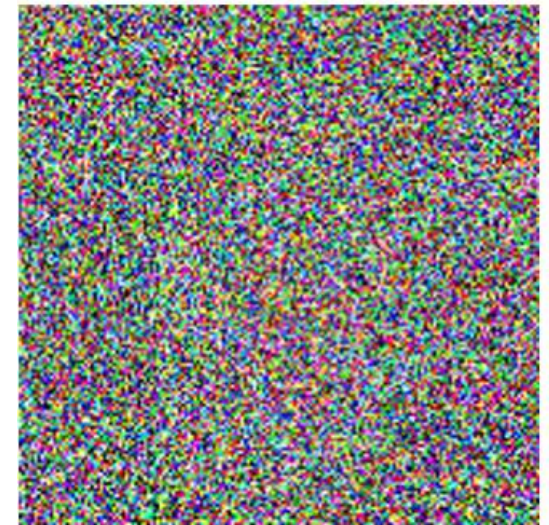
t=1000



+



=

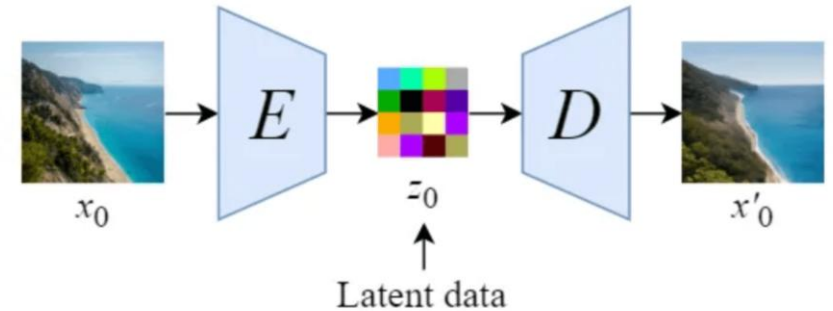


XT) immagine al
timestamp t=1000

<https://erdem.pl/2023/11/step-by-step-visual-introduction-to-diffusion-models#forward-diffusion-diagram>

Forward pass – modelli diffusivi (4)

- **Forward pass)** «Distruggere» gradualmente e sistematicamente la struttura di una distribuzione di dati attraverso un processo di diffusione in avanti, eseguito in modo iterativo.



Si aggiunge **rumore** passo dopo passo all'immagine fino a distruggere completamente la struttura visiva.

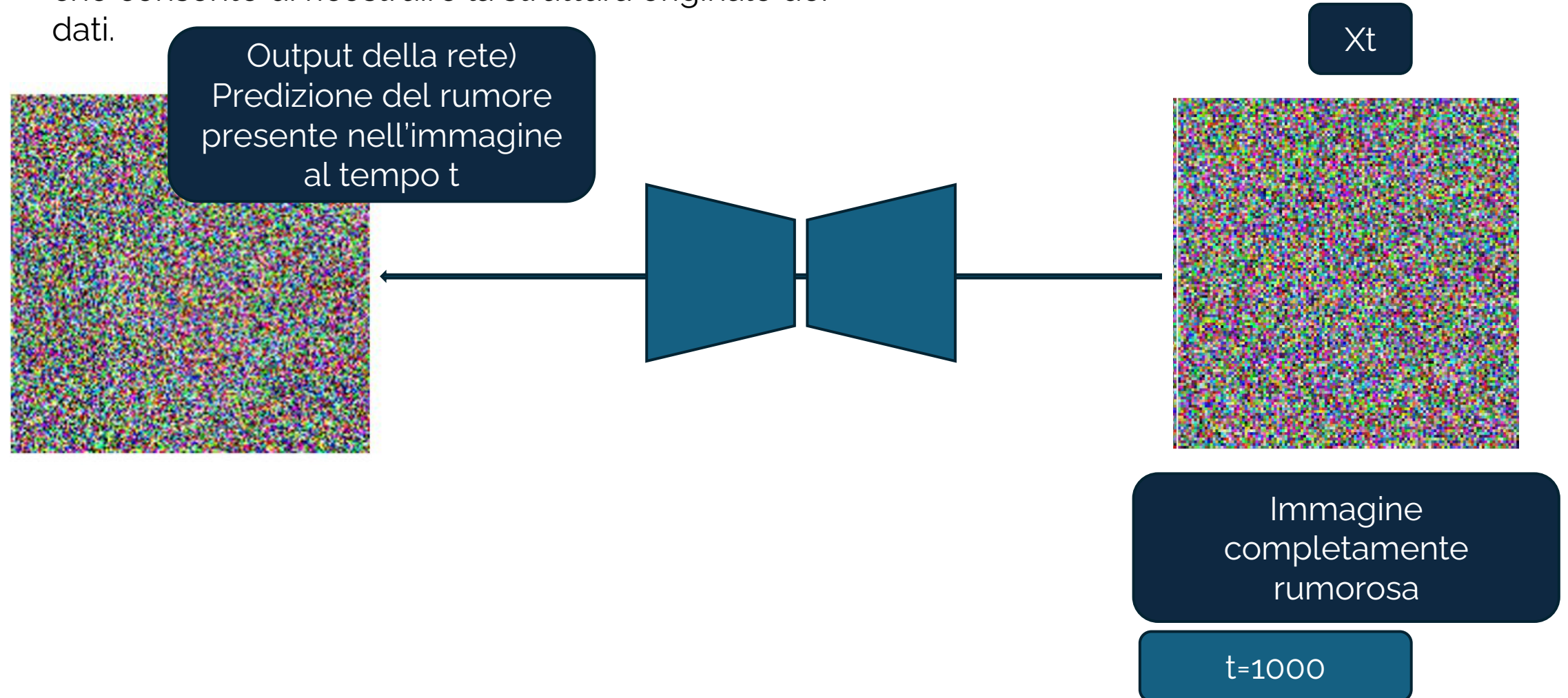
Dopo molti passi, l'immagine diventa puro rumore.

Tutto questo è conosciuto, non appreso, e infatti viene spesso chiamato **“diffusione diretta”** o **“forward noising process”**

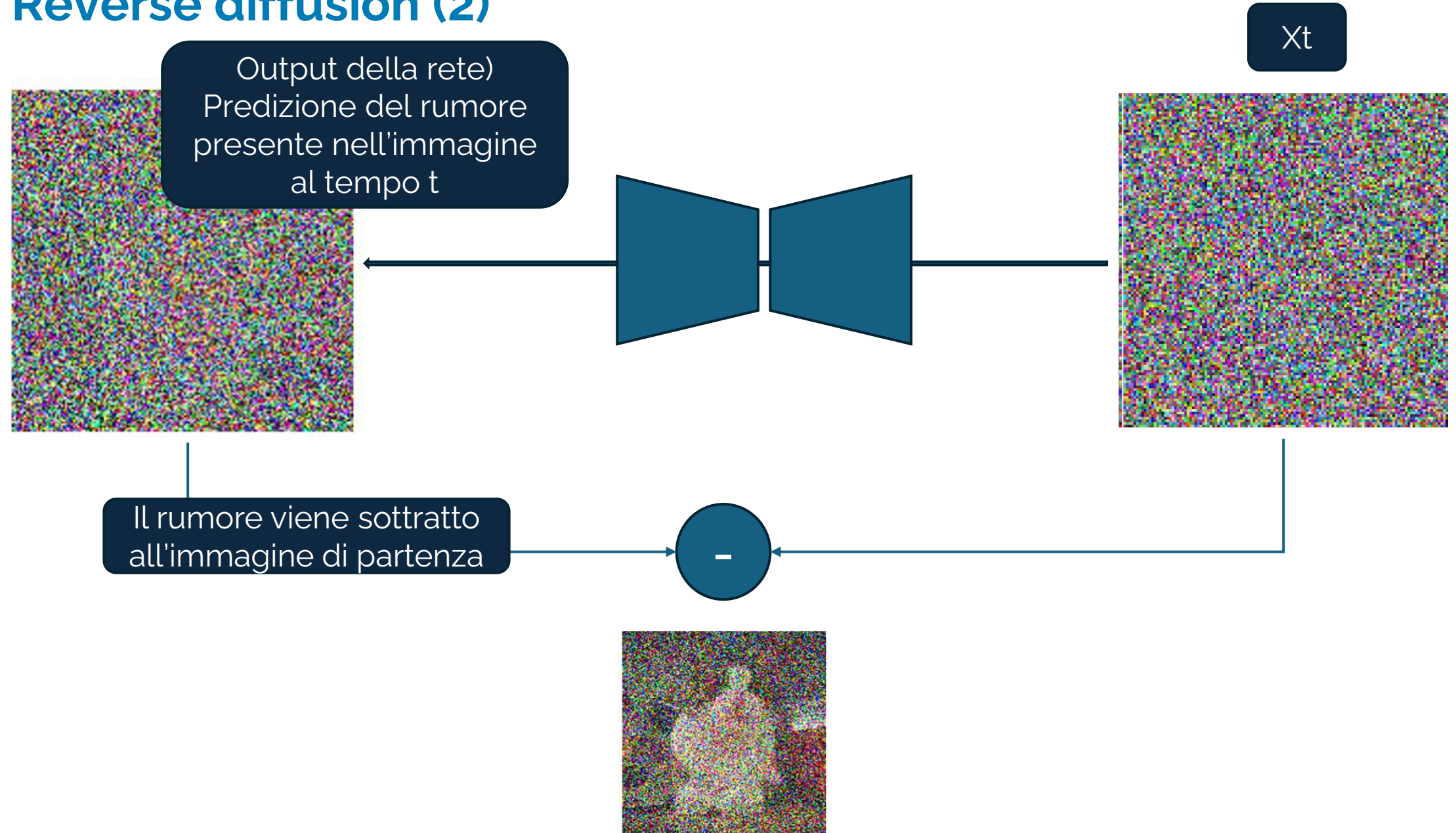
→ Questo è analogo a un processo fisico irreversibile (come la goccia che si diffonde nell'acqua).

Reverse diffusion (1)

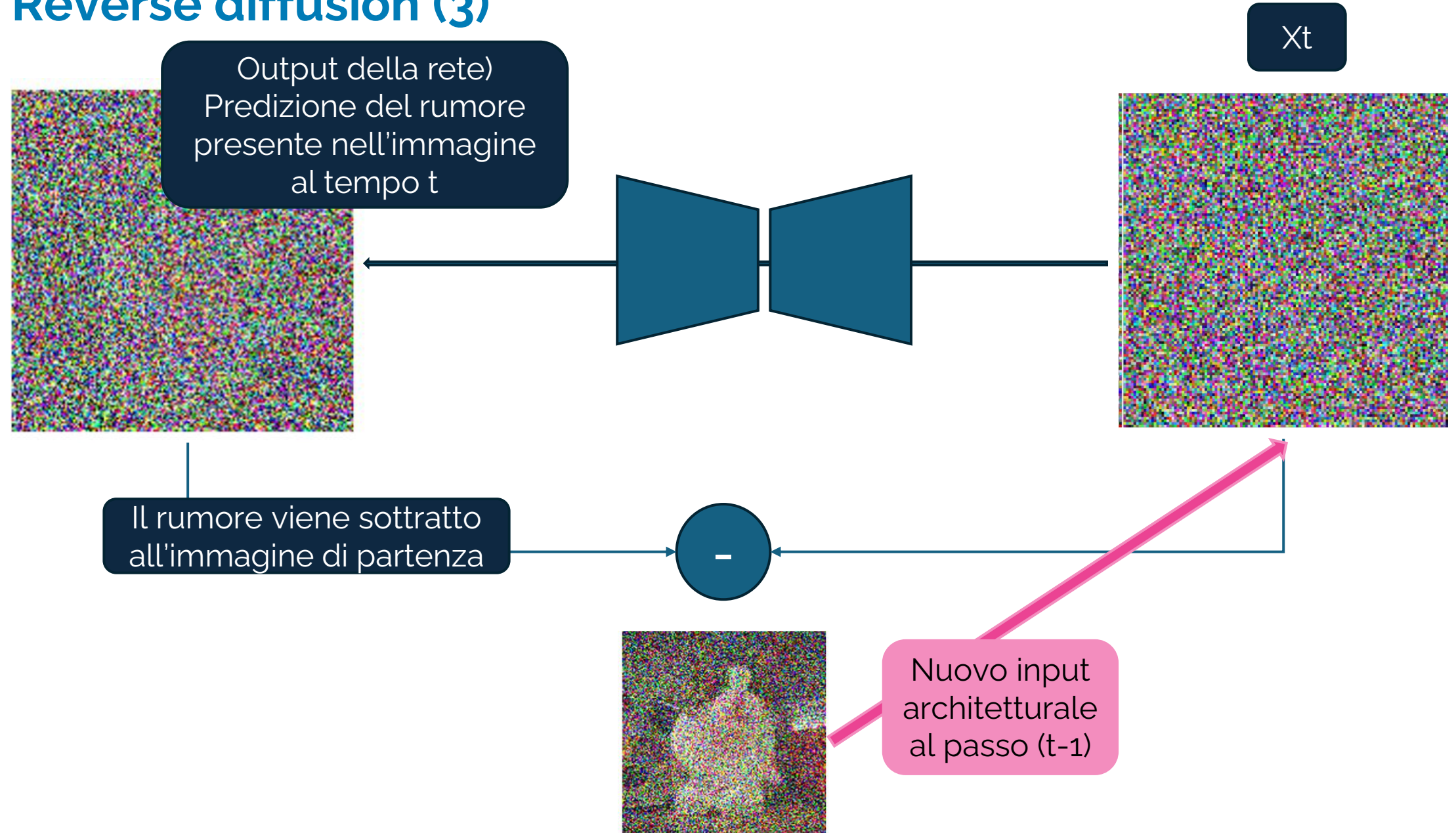
- **Reverse)** Successivamente, si **apprende (c'è addestramento!!)** un processo di diffusione inverso, che consente di ricostruire la struttura originale dei dati.



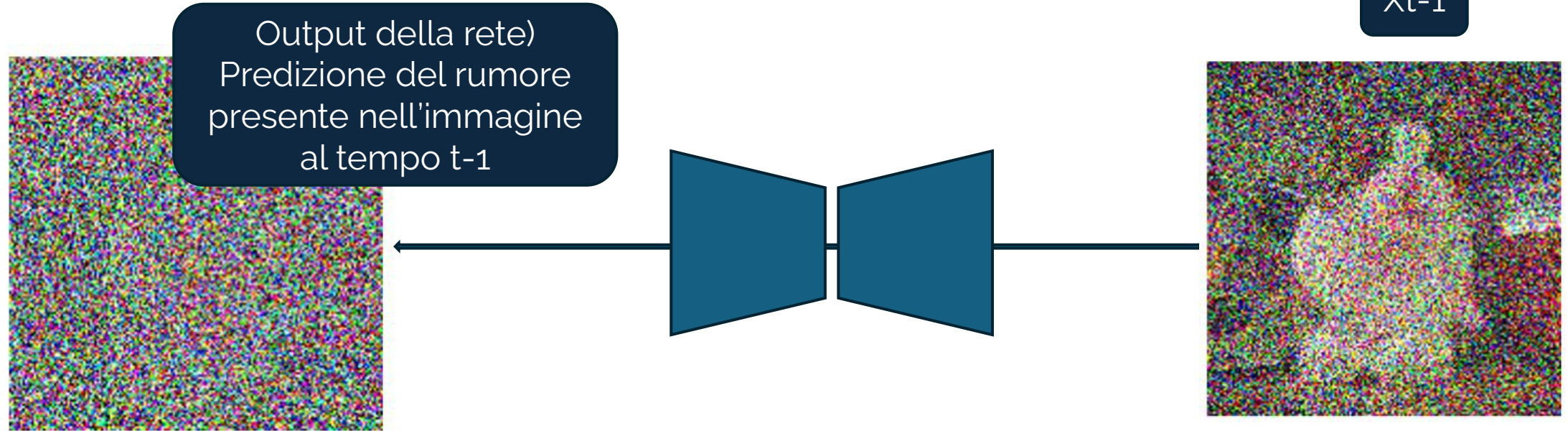
Reverse diffusion (2)



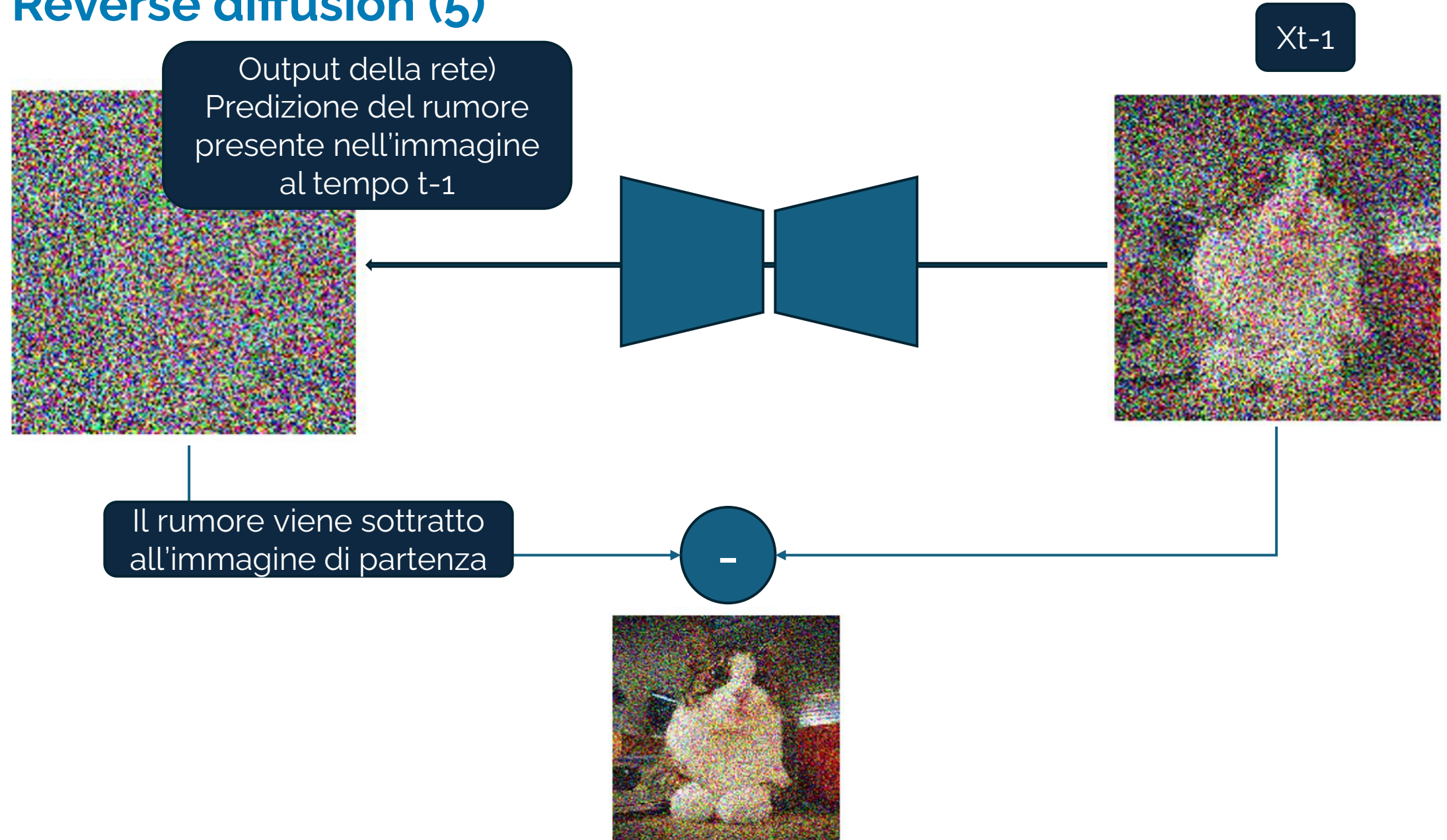
Reverse diffusion (3)



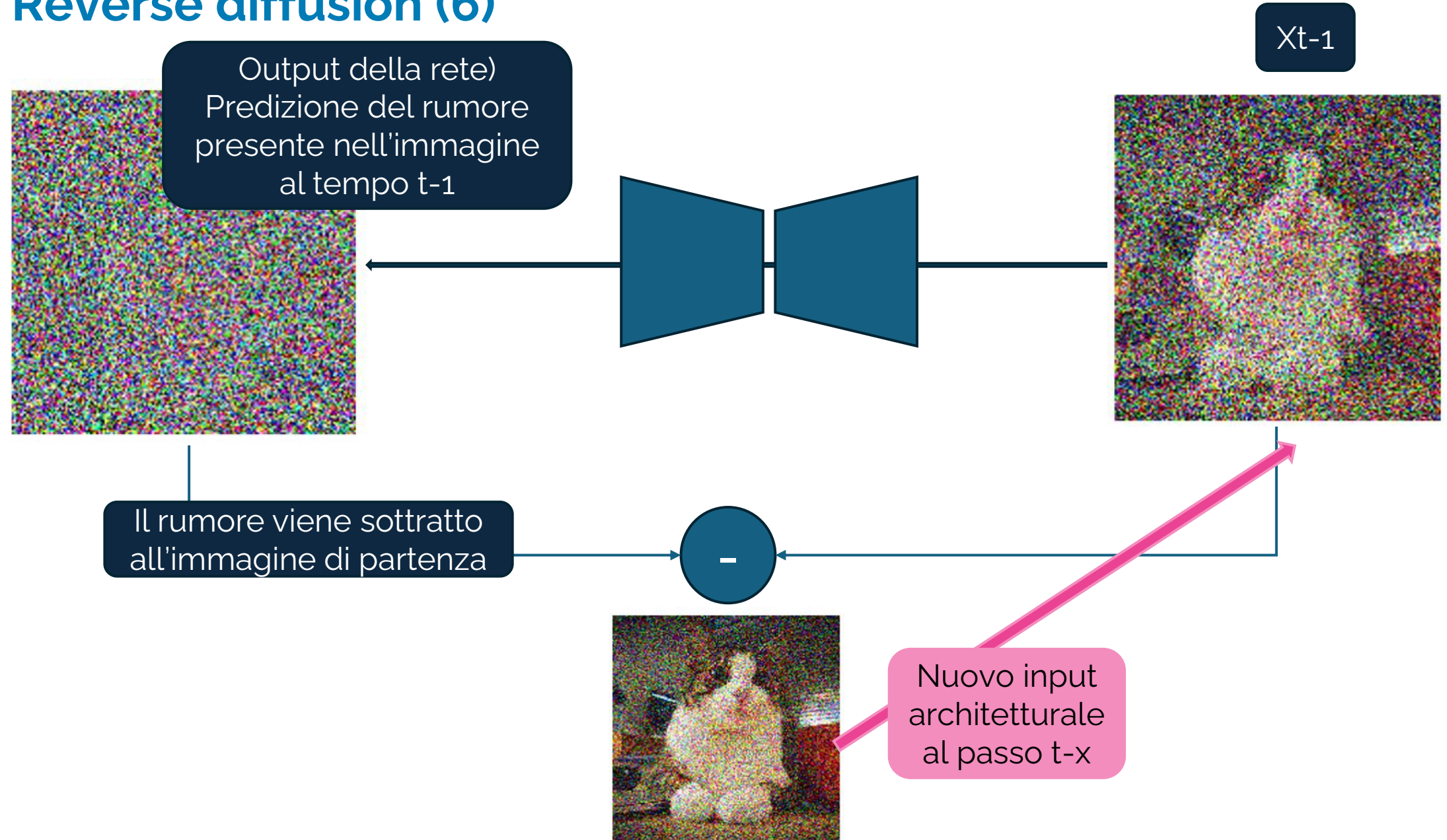
Reverse diffusion (4)



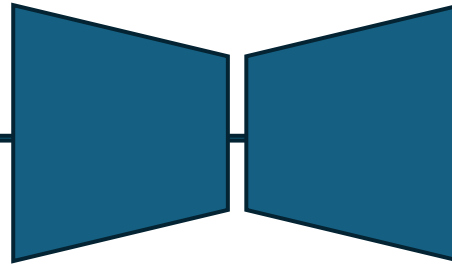
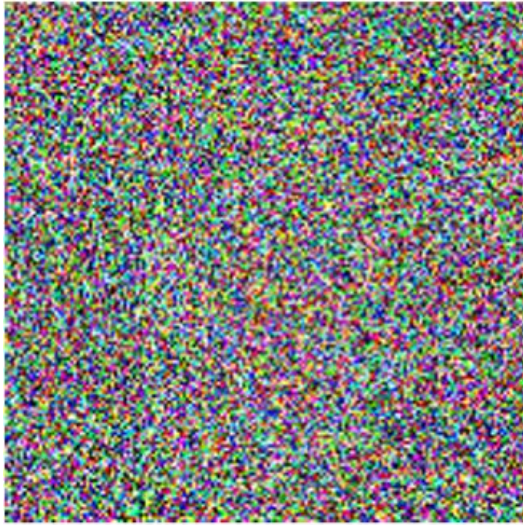
Reverse diffusion (5)



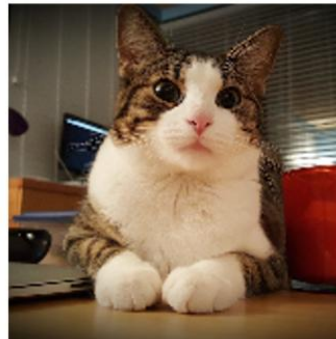
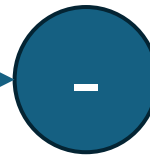
Reverse diffusion (6)



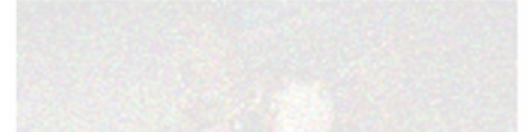
Reverse diffusion (7)



Il rumore viene sottratto
all'immagine di partenza



Reverse diffusion (8)

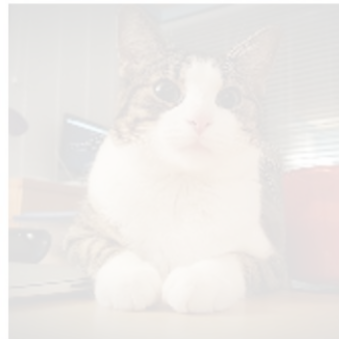
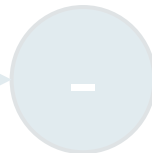


Diffusione inversa (reverse diffusion):

Ora il modello impara a invertire quel processo: cioè a partire da rumore puro, ricostruisce passo dopo passo un'immagine coerente e realistica (ad esempio, un gatto).

→ Il modello impara come riportare l'ordine nel caos.

Il rumore viene sottratto
all'immagine di partenza



Reverse diffusion (8)

