

# Diffusion models

**CS 4804: Introduction to AI**

*Fall 2025*

<https://tuvllms.github.io/ai-fall-2025/>

**Tu Vu**



based on Issac Ke's lecture

# Logistics

- HW 2 **due today**
- Final presentations: **12/4 & 12/9**
  - Sign-up form available on Piazza

# Grok-4.1

## xAI's Grok 4.1 (thinking) & Grok 4.1 rank #1 and #2 of the Text Arena

Rank ↑	Rank Spread ⓘ (Upper-Lower)	Model ↓	Score ↓	95% CI (±) ↓	Votes ↓	Organization ↓	License ↓
1	1 ↔ 2	 grok-4.1-thinking	1483 ⓘ Preliminary	±11	3,298	xAI	Proprietary
2	1 ↔ 4	 grok-4.1	1465 ⓘ Preliminary	±11	3,413	xAI	Proprietary
3	2 ↔ 6	 gemini-2.5-pro	1452	±4	65,503	Google	Proprietary
4	2 ↔ 8	 claude-sonnet-4-5-20250929-thinking-32k	1450	±5	16,549	Anthropic	Proprietary
5	3 ↔ 7	 claude-opus-4-1-20250805-thinking-16k	1449	±4	32,080	Anthropic	Proprietary
6	3 ↔ 11	 claude-sonnet-4-5-20250929	1445	±6	11,121	Anthropic	Proprietary
7	4 ↔ 13	 gpt-4.5-preview-2025-02-27	1442	±6	14,644	OpenAI	Proprietary
8	5 ↔ 13	 claude-opus-4-1-20250805	1440	±4	44,792	Anthropic	Proprietary
9	6 ↔ 13	 chatgpt-4o-latest-20250326	1438	±4	51,321	OpenAI	Proprietary
10	6 ↔ 14	 gpt-5-high	1437	±5	32,955	OpenAI	Proprietary
11	7 ↔ 14	 o3-2025-04-16	1434	±4	61,685	OpenAI	Proprietary
12	7 ↔ 16	 qwen3-max-preview	1434	±5	28,191	Alibaba	Proprietary

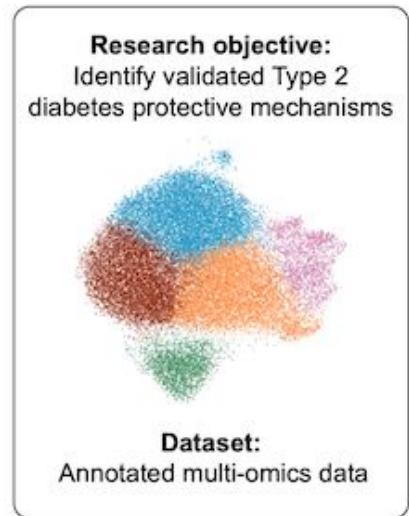
# Gemini 3.0

Benchmark	Description		Gemini 3 Pro	Gemini 2.5 Pro	Claude Sonnet 4.5	GPT-5.1
<b>Humanity's Last Exam</b>	Academic reasoning	No tools	<b>37.5%</b>	21.6%	13.7%	26.5%
<b>ARC-AGI-2</b>	Visual reasoning puzzles	ARC Prize Verified	<b>31.1%</b>	4.9%	13.6%	17.6%
<b>GPQA Diamond</b>	Scientific knowledge	No tools	<b>91.9%</b>	86.4%	83.4%	88.1%
<b>AIME 2025</b>	Mathematics	No tools	<b>95.0%</b>	88.0%	87.0%	94.0%
		With code execution	<b>100%</b>	—	<b>100%</b>	—
<b>MathArena Apex</b>	Challenging Math Contest problems		<b>23.4%</b>	0.5%	1.6%	1.0%
<b>MMMU-Pro</b>	Multimodal understanding and reasoning		<b>81.0%</b>	68.0%	68.0%	80.8%
<b>ScreenSpot-Pro</b>	Screen understanding		<b>72.7%</b>	11.4%	36.2%	3.5%
<b>CharXiv Reasoning</b>	Information synthesis from complex charts		<b>81.4%</b>	69.6%	68.5%	69.5%
<b>OmniDocBench 1.5</b>	OCR	Overall Edit Distance, lower is better	<b>0.115</b>	0.145	0.145	0.147
<b>Video-MMMU</b>	Knowledge acquisition from videos		<b>87.6%</b>	83.6%	77.8%	80.4%
<b>LiveCodeBench Pro</b>	Competitive coding problems from Codeforces, ICPC, and IOI	Elo Rating, higher is better	<b>2,439</b>	1,775	1,418	2,243
<b>Terminal-Bench 2.0</b>	Agentic terminal coding	Terminus-2 agent	<b>54.2%</b>	32.6%	42.8%	47.6%
<b>SWE-Bench Verified</b>	Agentic coding	Single attempt	76.2%	59.6%	<b>77.2%</b>	76.3%
<b>t2-bench</b>	Agentic tool use		<b>85.4%</b>	54.9%	84.7%	80.2%
<b>Vending-Bench 2</b>	Long-horizon agentic tasks	Net worth (mean), higher is better	<b>\$5,478.16</b>	\$573.64	\$3,838.74	\$1,473.43
<b>FACTS Benchmark Suite</b>	Held out internal grounding, parametric, MM, and search retrieval benchmarks		<b>70.5%</b>	63.4%	50.4%	50.8%
<b>SimpleQA Verified</b>	Parametric knowledge		<b>72.1%</b>	54.5%	29.3%	34.9%
<b>MMMLU</b>	Multilingual Q&A		<b>91.8%</b>	89.5%	89.1%	91.0%
<b>Global PIQA</b>	Commonsense reasoning across 100 Languages and Cultures		<b>93.4%</b>	91.5%	90.1%	90.9%
<b>MRCR v2 (8-needle)</b>	Long context performance	128k (average) 1M (pointwise)	<b>77.0%</b> <b>26.3%</b>	58.0% 16.4%	47.1% not supported	61.6% not supported

# An AI Scientist for Autonomous Discovery

a

Input



Kosmos World Model



Output

The T2D protective variant rs9379084-A tags a candidate cis-acting regulatory sequence that modulates levels of SSR1 via altered ATF3 binding

**Summary**  
The increase variant allele frequency A in the rs9379084 locus is associated with increased Type 2 diabetes protection, with convergent multi-omics evidence indicating that the protective effects result from increased gene expression, increased chromatin accessibility, and downstream transcription factor activity. The protective variant allele frequency A is associated with increased ATF3 binding, including rs9379084 is a leading candidate target reported to increased ATF3 motif affinity, one among 100,000 ATF3 motifs with the highest mean target scores (ATF3, GSP, and SMAD9), indicating a multi-gene regulatory network consistent with coordinated co-regulatory control [10].

**Background**  
Human genetic studies of Type 2 diabetes highlight genes involved in metabolism and regulation mechanisms, including those that regulate blood glucose homeostasis, chromatin accessibility, and context-specific gene regulation. In addition, it has been shown that human genetic data discriminatory evaluate directional concordance between variants and their effect sizes. In this study, we have used a genome-wide association study (GWAS) to identify variants that are more protective than risk across molecular traits, and prioritize mechanisms with multi-source inputs. Our results show that the most protective variants are those that are most likely to be causal, regulatory biologics in which single nucleotide polymorphisms result in gene programs that participate in both protective and therapeutic activity paths.

**Results & Discussion**  
This work establishes rs9379084-A as a high-confidence protective variant and a novel causal regulatory variant in the rs9379084 locus. The protective variant allele frequency A tags a cis-acting regulatory sequence that modulates levels of SSR1 via altered ATF3 binding. The protective variant allele frequency A is associated with increased gene expression, increased chromatin accessibility, and downstream transcription factor activity. The protective variant allele frequency A is associated with increased ATF3 motif affinity, one among 100,000 ATF3 motifs with the highest mean target scores (ATF3, GSP, and SMAD9), indicating a multi-gene regulatory network consistent with coordinated co-regulatory control [10].

**The mechanistic chain for SSR1 is robust across molecular traits.** As the level of sequence variation increases, so does the protective effect for ATF3 (binding = 1.0%), while also modulating a broader set of metabolic and disease pathways. The protective variant allele frequency A is associated with increased binding for ATF3 (1.0%), MSH3 (1.0%), VDR (1.0%), and MYBPC3 (1.0%), and decreased binding for GPR119 (0.2%), GSK3B (0.2%), RAN (0.2%),

Figure 1. Multi-heterogeneous objectives, net zero datasets, and model learning processes after learning phase. (A) Researcher input: three datasets (Discovery 1, 2, 3) and a literature review agent. (B) Learning process where the net zero datasets (Discovery 1, 2, 3) receive multi-heterogeneous inputs (Data Analysis Agent, Literature Review Agent, Established Findings). (C) Output: the mechanism chain for SSR1 is robust across molecular traits.

Figure 1. Multi-heterogeneous objectives, net zero datasets, and model learning processes after learning phase. (A) Researcher input: three datasets (Discovery 1, 2, 3) and a literature review agent. (B) Learning process where the net zero datasets (Discovery 1, 2, 3) receive multi-heterogeneous inputs (Data Analysis Agent, Literature Review Agent, Established Findings). (C) Output: the mechanism chain for SSR1 is robust across molecular traits.

Figure 1. Multi-heterogeneous objectives, net zero datasets, and model learning processes after learning phase. (A) Researcher input: three datasets (Discovery 1, 2, 3) and a literature review agent. (B) Learning process where the net zero datasets (Discovery 1, 2, 3) receive multi-heterogeneous inputs (Data Analysis Agent, Literature Review Agent, Established Findings). (C) Output: the mechanism chain for SSR1 is robust across molecular traits.

Figure 1. Multi-heterogeneous objectives, net zero datasets, and model learning processes after learning phase. (A) Researcher input: three datasets (Discovery 1, 2, 3) and a literature review agent. (B) Learning process where the net zero datasets (Discovery 1, 2, 3) receive multi-heterogeneous inputs (Data Analysis Agent, Literature Review Agent, Established Findings). (C) Output: the mechanism chain for SSR1 is robust across molecular traits.

Figure 1. Multi-heterogeneous objectives, net zero datasets, and model learning processes after learning phase. (A) Researcher input: three datasets (Discovery 1, 2, 3) and a literature review agent. (B) Learning process where the net zero datasets (Discovery 1, 2, 3) receive multi-heterogeneous inputs (Data Analysis Agent, Literature Review Agent, Established Findings). (C) Output: the mechanism chain for SSR1 is robust across molecular traits.

Figure 1. Multi-heterogeneous objectives, net zero datasets, and model learning processes after learning phase. (A) Researcher input: three datasets (Discovery 1, 2, 3) and a literature review agent. (B) Learning process where the net zero datasets (Discovery 1, 2, 3) receive multi-heterogeneous inputs (Data Analysis Agent, Literature Review Agent, Established Findings). (C) Output: the mechanism chain for SSR1 is robust across molecular traits.

Figure 1. Multi-heterogeneous objectives, net zero datasets, and model learning processes after learning phase. (A) Researcher input: three datasets (Discovery 1, 2, 3) and a literature review agent. (B) Learning process where the net zero datasets (Discovery 1, 2, 3) receive multi-heterogeneous inputs (Data Analysis Agent, Literature Review Agent, Established Findings). (C) Output: the mechanism chain for SSR1 is robust across molecular traits.

Figure 1. Multi-heterogeneous objectives, net zero datasets, and model learning processes after learning phase. (A) Researcher input: three datasets (Discovery 1, 2, 3) and a literature review agent. (B) Learning process where the net zero datasets (Discovery 1, 2, 3) receive multi-heterogeneous inputs (Data Analysis Agent, Literature Review Agent, Established Findings). (C) Output: the mechanism chain for SSR1 is robust across molecular traits.

Figure 1. Multi-heterogeneous objectives, net zero datasets, and model learning processes after learning phase. (A) Researcher input: three datasets (Discovery 1, 2, 3) and a literature review agent. (B) Learning process where the net zero datasets (Discovery 1, 2, 3) receive multi-heterogeneous inputs (Data Analysis Agent, Literature Review Agent, Established Findings). (C) Output: the mechanism chain for SSR1 is robust across molecular traits.

Figure 1. Multi-heterogeneous objectives, net zero datasets, and model learning processes after learning phase. (A) Researcher input: three datasets (Discovery 1, 2, 3) and a literature review agent. (B) Learning process where the net zero datasets (Discovery 1, 2, 3) receive multi-heterogeneous inputs (Data Analysis Agent, Literature Review Agent, Established Findings). (C) Output: the mechanism chain for SSR1 is robust across molecular traits.

Figure 1. Multi-heterogeneous objectives, net zero datasets, and model learning processes after learning phase. (A) Researcher input: three datasets (Discovery 1, 2, 3) and a literature review agent. (B) Learning process where the net zero datasets (Discovery 1, 2, 3) receive multi-heterogeneous inputs (Data Analysis Agent, Literature Review Agent, Established Findings). (C) Output: the mechanism chain for SSR1 is robust across molecular traits.

Figure 1. Multi-heterogeneous objectives, net zero datasets, and model learning processes after learning phase. (A) Researcher input: three datasets (Discovery 1, 2, 3) and a literature review agent. (B) Learning process where the net zero datasets (Discovery 1, 2, 3) receive multi-heterogeneous inputs (Data Analysis Agent, Literature Review Agent, Established Findings). (C) Output: the mechanism chain for SSR1 is robust across molecular traits.

Figure 1. Multi-heterogeneous objectives, net zero datasets, and model learning processes after learning phase. (A) Researcher input: three datasets (Discovery 1, 2, 3) and a literature review agent. (B) Learning process where the net zero datasets (Discovery 1, 2, 3) receive multi-heterogeneous inputs (Data Analysis Agent, Literature Review Agent, Established Findings). (C) Output: the mechanism chain for SSR1 is robust across molecular traits.

Figure 1. Multi-heterogeneous objectives, net zero datasets, and model learning processes after learning phase. (A) Researcher input: three datasets (Discovery 1, 2, 3) and a literature review agent. (B) Learning process where the net zero datasets (Discovery 1, 2, 3) receive multi-heterogeneous inputs (Data Analysis Agent, Literature Review Agent, Established Findings). (C) Output: the mechanism chain for SSR1 is robust across molecular traits.

Figure 1. Multi-heterogeneous objectives, net zero datasets, and model learning processes after learning phase. (A) Researcher input: three datasets (Discovery 1, 2, 3) and a literature review agent. (B) Learning process where the net zero datasets (Discovery 1, 2, 3) receive multi-heterogeneous inputs (Data Analysis Agent, Literature Review Agent, Established Findings). (C) Output: the mechanism chain for SSR1 is robust across molecular traits.

Figure 1. Multi-heterogeneous objectives, net zero datasets, and model learning processes after learning phase. (A) Researcher input: three datasets (Discovery 1, 2, 3) and a literature review agent. (B) Learning process where the net zero datasets (Discovery 1, 2, 3) receive multi-heterogeneous inputs (Data Analysis Agent, Literature Review Agent, Established Findings). (C) Output: the mechanism chain for SSR1 is robust across molecular traits.

Figure 1. Multi-heterogeneous objectives, net zero datasets, and model learning processes after learning phase. (A) Researcher input: three datasets (Discovery 1, 2, 3) and a literature review agent. (B) Learning process where the net zero datasets (Discovery 1, 2, 3) receive multi-heterogeneous inputs (Data Analysis Agent, Literature Review Agent, Established Findings). (C) Output: the mechanism chain for SSR1 is robust across molecular traits.

Figure 1. Multi-heterogeneous objectives, net zero datasets, and model learning processes after learning phase. (A) Researcher input: three datasets (Discovery 1, 2, 3) and a literature review agent. (B) Learning process where the net zero datasets (Discovery 1, 2, 3) receive multi-heterogeneous inputs (Data Analysis Agent, Literature Review Agent, Established Findings). (C) Output: the mechanism chain for SSR1 is robust across molecular traits.

Figure 1. Multi-heterogeneous objectives, net zero datasets, and model learning processes after learning phase. (A) Researcher input: three datasets (Discovery 1, 2, 3) and a literature review agent. (B) Learning process where the net zero datasets (Discovery 1, 2, 3) receive multi-heterogeneous inputs (Data Analysis Agent, Literature Review Agent, Established Findings). (C) Output: the mechanism chain for SSR1 is robust across molecular traits.

Figure 1. Multi-heterogeneous objectives, net zero datasets, and model learning processes after learning phase. (A) Researcher input: three datasets (Discovery 1, 2, 3) and a literature review agent. (B) Learning process where the net zero datasets (Discovery 1, 2, 3) receive multi-heterogeneous inputs (Data Analysis Agent, Literature Review Agent, Established Findings). (C) Output: the mechanism chain for SSR1 is robust across molecular traits.

Figure 1. Multi-heterogeneous objectives, net zero datasets, and model learning processes after learning phase. (A) Researcher input: three datasets (Discovery 1, 2, 3) and a literature review agent. (B) Learning process where the net zero datasets (Discovery 1, 2, 3) receive multi-heterogeneous inputs (Data Analysis Agent, Literature Review Agent, Established Findings). (C) Output: the mechanism chain for SSR1 is robust across molecular traits.

Figure 1. Multi-heterogeneous objectives, net zero datasets, and model learning processes after learning phase. (A) Researcher input: three datasets (Discovery 1, 2, 3) and a literature review agent. (B) Learning process where the net zero datasets (Discovery 1, 2, 3) receive multi-heterogeneous inputs (Data Analysis Agent, Literature Review Agent, Established Findings). (C) Output: the mechanism chain for SSR1 is robust across molecular traits.

Figure 1. Multi-heterogeneous objectives, net zero datasets, and model learning processes after learning phase. (A) Researcher input: three datasets (Discovery 1, 2, 3) and a literature review agent. (B) Learning process where the net zero datasets (Discovery 1, 2, 3) receive multi-heterogeneous inputs (Data Analysis Agent, Literature Review Agent, Established Findings). (C) Output: the mechanism chain for SSR1 is robust across molecular traits.

Figure 1. Multi-heterogeneous objectives, net zero datasets, and model learning processes after learning phase. (A) Researcher input: three datasets (Discovery 1, 2, 3) and a literature review agent. (B) Learning process where the net zero datasets (Discovery 1, 2, 3) receive multi-heterogeneous inputs (Data Analysis Agent, Literature Review Agent, Established Findings). (C) Output: the mechanism chain for SSR1 is robust across molecular traits.

# Today's lecture



Andrej Karpathy ✅

@karpathy



Nice, short post illustrating how simple text (discrete) diffusion can be.

Diffusion (i.e. parallel, iterated denoising, top) is the pervasive generative paradigm in image/video, but autoregression (i.e. go left to right bottom) is the dominant paradigm in text. For audio I've seen a bit of both.

<https://x.com/karpathy/status/1980347971935068380>

# Diffusion process

*The laws of physics tell us that dye particles continue to move through water until their concentration becomes uniform everywhere in the beaker, which marks equilibrium.*



# Diffusion models power AI image/video tools



# A turtle wearing sunglasses playing basketball



Image created by ChatGPT

# What are diffusion models?

- A diffusion model is a generative model that first corrupts data by gradually adding noise and then learns to reverse that process so it can create new samples by iteratively denoising

---

# Denoising Diffusion Probabilistic Models

---

**Jonathan Ho**

UC Berkeley

[jonathanho@berkeley.edu](mailto:jonathanho@berkeley.edu)

**Ajay Jain**

UC Berkeley

[ajayj@berkeley.edu](mailto:ajayj@berkeley.edu)

**Pieter Abbeel**

UC Berkeley

[pabbeel@cs.berkeley.edu](mailto:pabbeel@cs.berkeley.edu)

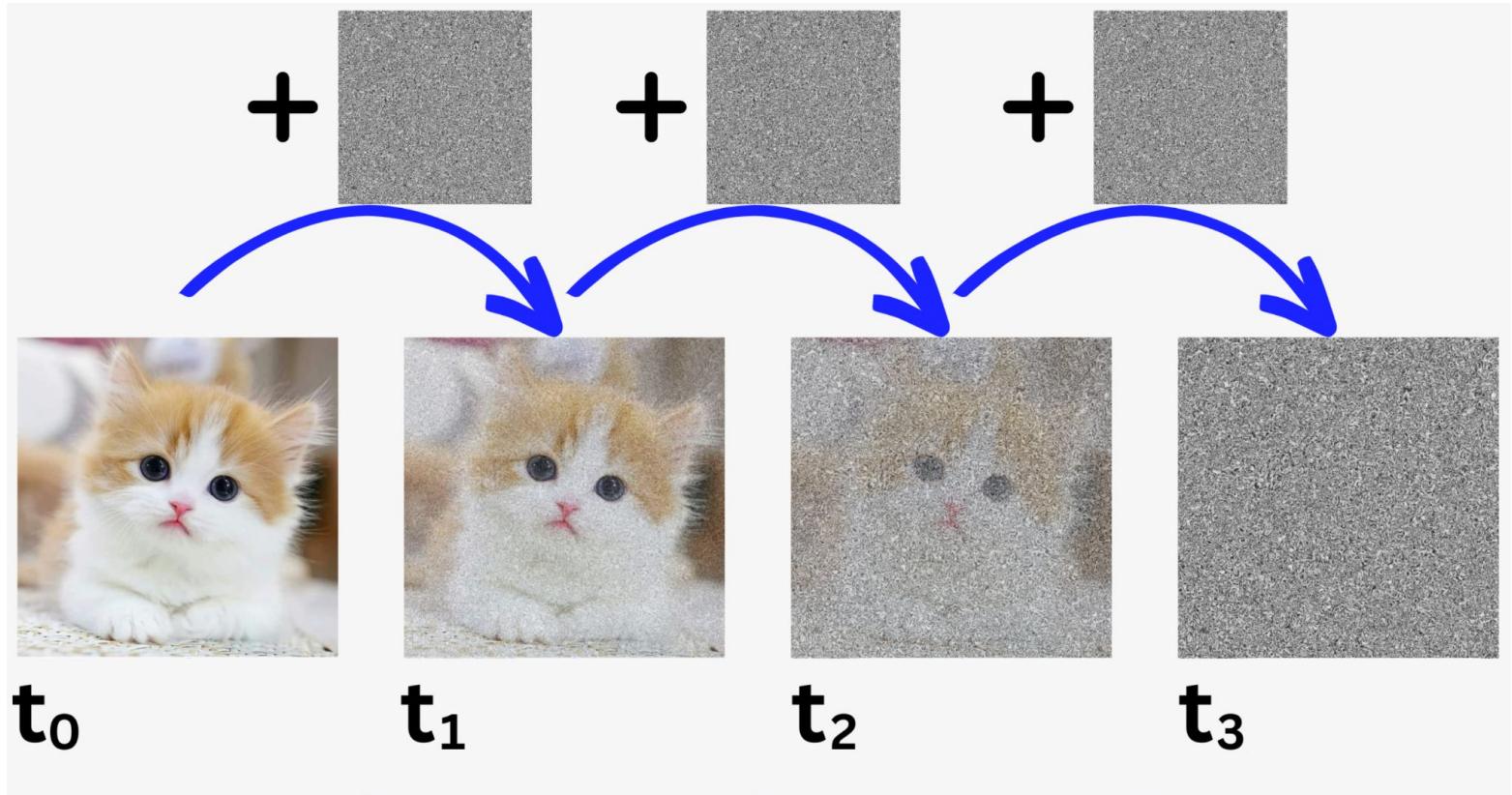
## Abstract

We present high quality image synthesis results using diffusion probabilistic models, a class of latent variable models inspired by considerations from nonequilibrium thermodynamics. Our best results are obtained by training on a weighted variational bound designed according to a novel connection between diffusion probabilistic models and denoising score matching with Langevin dynamics, and our models naturally admit a progressive lossy decompression scheme that can be interpreted as a generalization of autoregressive decoding. On the unconditional CIFAR10 dataset, we obtain an Inception score of 9.46 and a state-of-the-art FID score of 3.17. On 256x256 LSUN, we obtain sample quality similar to ProgressiveGAN. Our implementation is available at <https://github.com/hojonathanho/diffusion>.

# **Three important concepts**

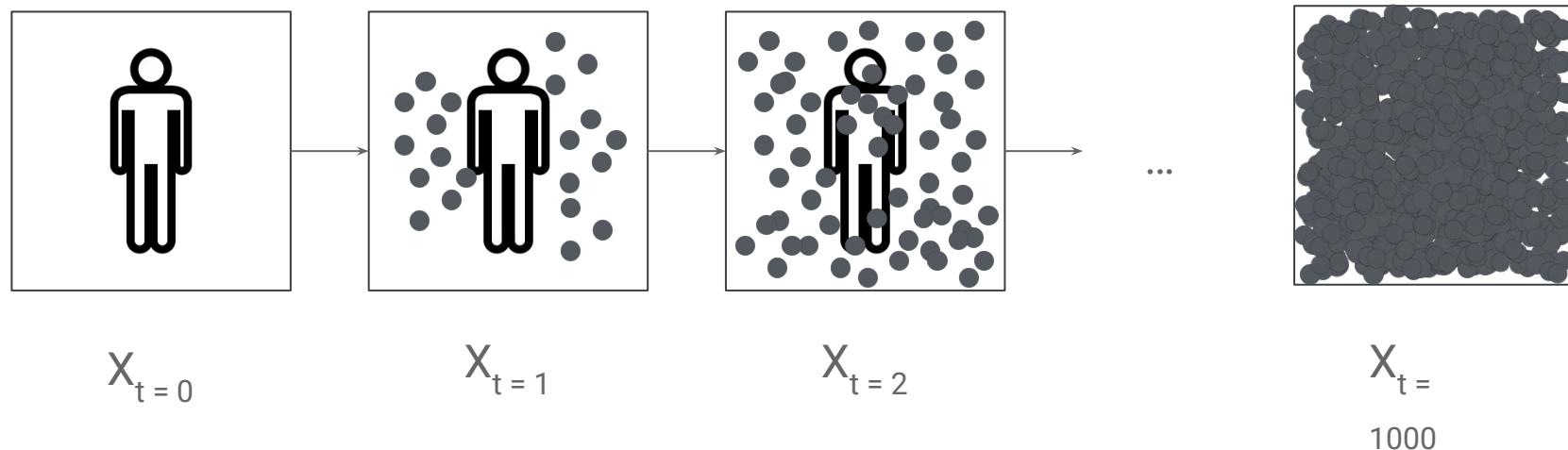
- **Forward diffusion**
- **Reverse diffusion**
- **Conditional diffusion**

# Forward diffusion



<https://newsletter.theaiedge.io/p/diffusion-models-stable-diffusion>

# Forward diffusion (cont'd)



# Grayscale images



0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

# Color images

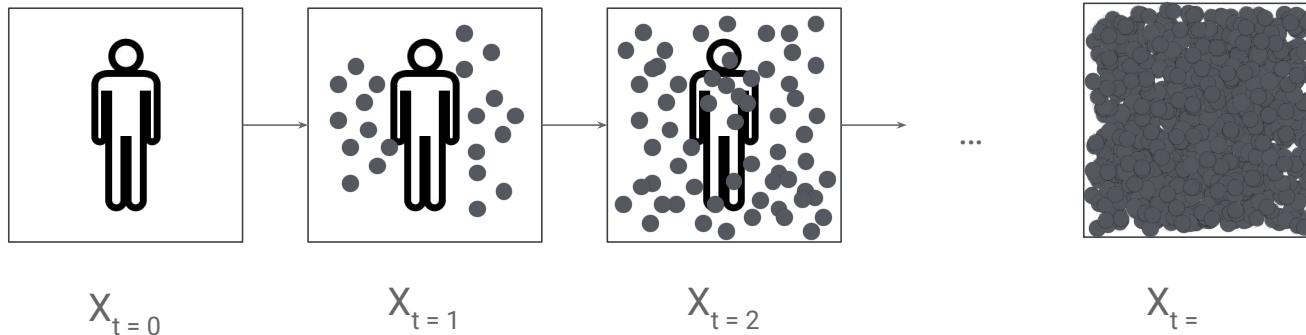


0	2	0	5	1	7	6	0	8
3	2	0	5	4	7	6	9	8
2	3	0	1	2	3	4	5	6
5	2	1	0	3	2	5	4	7
4	5	2	3	0	1	2	3	4
7	4	3	2	1	0	3	2	5
6	7	4	5	2	3	0	1	2
9	6	5	4	3	2	1	0	3
8	9	6	7	4	5	2	3	0
	8	7	6	5	4	3	2	1

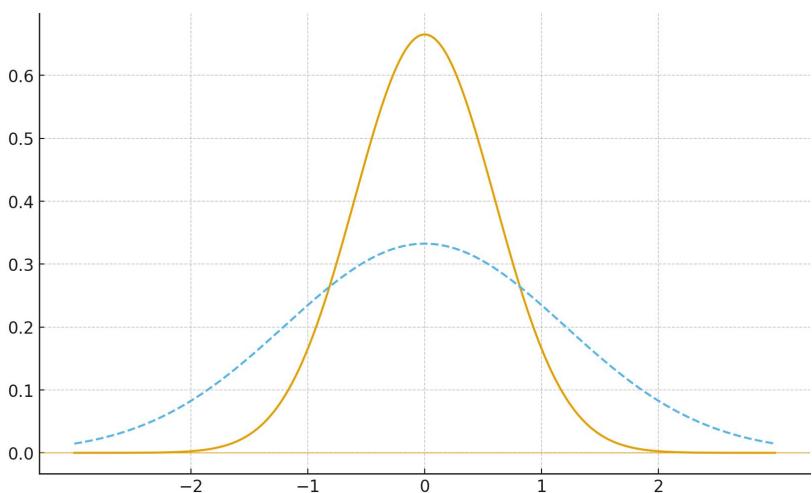
*channel x height x width*

***Channels are usually RGB: Red, Green, and Blue***

# Forward diffusion (cont'd)



$(255, 0, 0) + (-2, 2, 0)$   
 $\rightarrow (253, 2, 0)$   
// less red + more green



When you draw a sample from a Gaussian distribution  $N(\mu, \sigma^2)$ , the sampled value is

$$x = \mu + \sigma \varepsilon$$

where

- $x$  is the sampled value.
- $\mu$  is the mean that centers the distribution.
- $\sigma$  is the standard deviation that determines the spread.
- $\varepsilon$  is a standard normal variable that follows  $N(0, 1)$  and provides randomness.

$$x = \mu + \sqrt{\beta} \varepsilon$$

where

- $x$  is the sampled value.
- $\mu$  is the mean that centers the distribution.
- $\beta$  is the variance.
- $\sqrt{\beta}$  is the standard deviation.
- $\varepsilon$  is a standard normal variable that follows  $N(0, 1)$  and provides randomness.

# Markov chain

$$q(x_t \mid x_{t-1}) = \boxed{?} x_{t-1} + \sqrt{\beta} \epsilon$$

# Objective

$$q(x_t \mid x_0) \xrightarrow[t \rightarrow \infty]{} \mathcal{N}(0, 1)$$

## Noise scheduler (cont'd)

$$q(x_t \mid x_{t-1}) = \sqrt{1 - \beta} x_{t-1} + \sqrt{\beta} \epsilon$$

## Noise scheduler (cont'd)

Let  $\alpha_t = 1 - \beta_t$  and  $\bar{\alpha}_t = \prod_{i=1}^t \alpha_i$

$$\begin{aligned}\mathbf{x}_t &= \sqrt{\alpha_t} \mathbf{x}_{t-1} + \sqrt{1 - \alpha_t} \boldsymbol{\epsilon}_{t-1}^* \\&= \sqrt{\alpha_t} \left( \sqrt{\alpha_{t-1}} \mathbf{x}_{t-2} + \sqrt{1 - \alpha_{t-1}} \boldsymbol{\epsilon}_{t-2}^* \right) + \sqrt{1 - \alpha_t} \boldsymbol{\epsilon}_{t-1}^* \\&= \sqrt{\alpha_t \alpha_{t-1}} \mathbf{x}_{t-2} + \sqrt{\alpha_t - \alpha_t \alpha_{t-1}} \boldsymbol{\epsilon}_{t-2}^* + \sqrt{1 - \alpha_t} \boldsymbol{\epsilon}_{t-1}^* \\&= \sqrt{\alpha_t \alpha_{t-1}} \mathbf{x}_{t-2} + \sqrt{\sqrt{\alpha_t - \alpha_t \alpha_{t-1}}^2 + \sqrt{1 - \alpha_t}^2} \boldsymbol{\epsilon}_{t-2} \\&= \sqrt{\alpha_t \alpha_{t-1}} \mathbf{x}_{t-2} + \sqrt{\alpha_t - \alpha_t \alpha_{t-1} + 1 - \alpha_t} \boldsymbol{\epsilon}_{t-2} \\&= \sqrt{\alpha_t \alpha_{t-1}} \mathbf{x}_{t-2} + \sqrt{1 - \alpha_t \alpha_{t-1}} \boldsymbol{\epsilon}_{t-2} \\&= \dots \\&= \sqrt{\prod_{i=1}^t \alpha_i} \mathbf{x}_0 + \sqrt{1 - \prod_{i=1}^t \alpha_i} \boldsymbol{\epsilon}_0 \\&= \sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \boldsymbol{\epsilon}_0\end{aligned}$$

# Noise scheduler

Let  $\alpha_t = 1 - \beta_t$  and  $\bar{\alpha}_t = \prod_{i=1}^t \alpha_i$

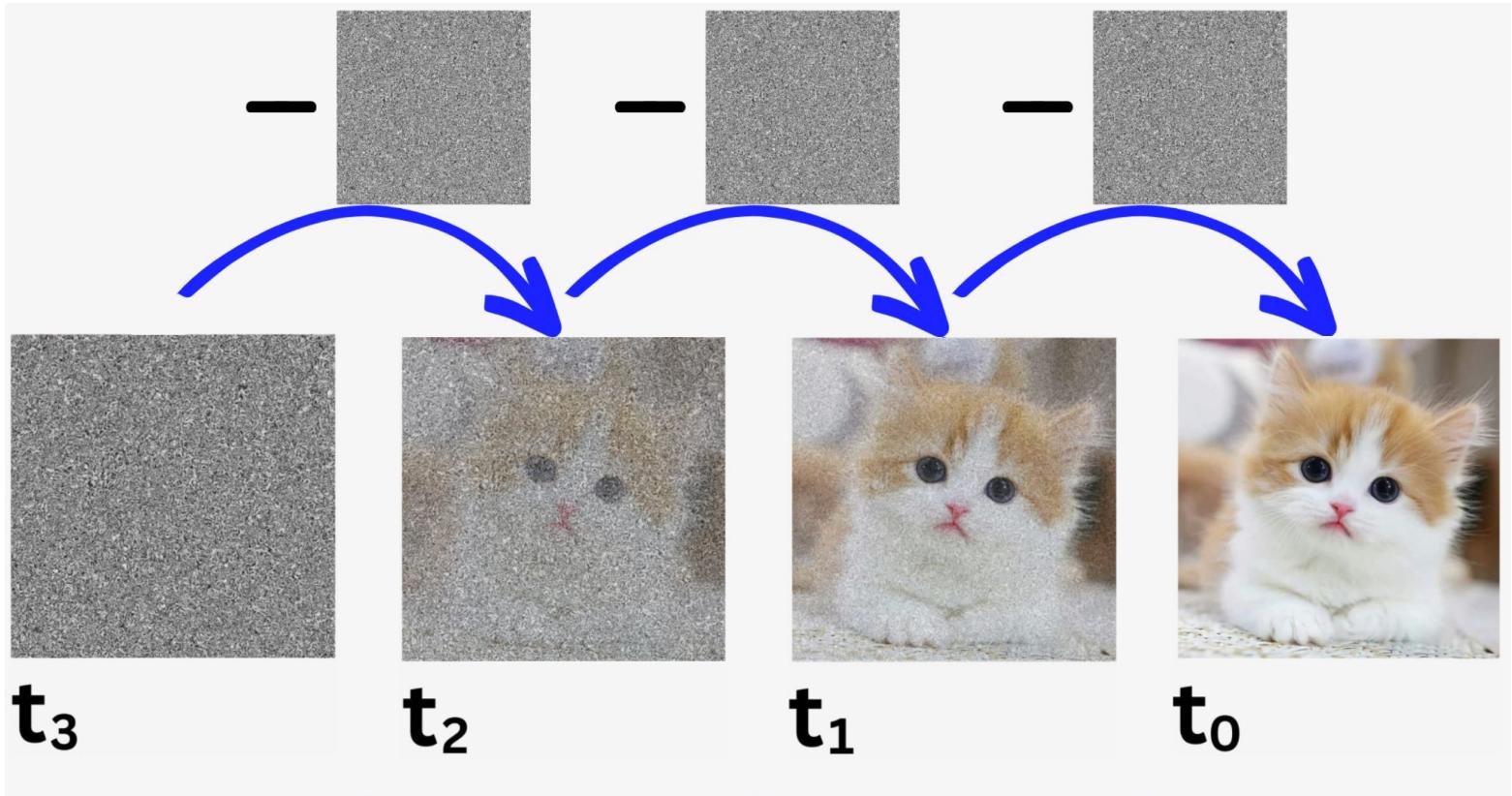
$$\begin{aligned}\mathbf{x}_t &= \sqrt{\alpha_t} \mathbf{x}_{t-1} + \sqrt{1 - \alpha_t} \boldsymbol{\epsilon}_{t-1} && ; \text{where } \boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \dots \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \\ &= \sqrt{\alpha_t \alpha_{t-1}} \mathbf{x}_{t-2} + \sqrt{1 - \alpha_t \alpha_{t-1}} \bar{\boldsymbol{\epsilon}}_{t-2} && ; \text{where } \bar{\boldsymbol{\epsilon}}_{t-2} \text{ merges two Gaussians (*).} \\ &= \dots \\ &= \sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \boldsymbol{\epsilon}\end{aligned}$$

(\*) Recall that when we merge two Gaussians with different variance,  $\mathcal{N}(\mathbf{0}, \sigma_1^2 \mathbf{I})$  and  $\mathcal{N}(\mathbf{0}, \sigma_2^2 \mathbf{I})$ , the new distribution is  $\mathcal{N}(\mathbf{0}, (\sigma_1^2 + \sigma_2^2) \mathbf{I})$ . Here the merged standard deviation is

$$\sqrt{(1 - \alpha_t) + \alpha_t(1 - \alpha_{t-1})} = \sqrt{1 - \alpha_t \alpha_{t-1}}.$$

Usually, we can afford a larger update step when the sample gets noisier, so  $\beta_1 < \beta_2 < \dots < \beta_T$  and therefore  $\bar{\alpha}_1 > \dots > \bar{\alpha}_T$ .

# Reverse diffusion

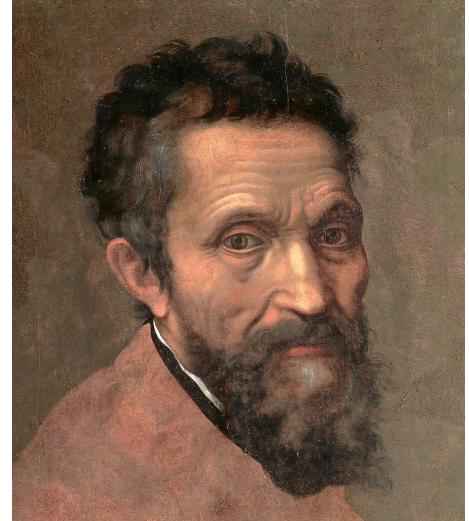


<https://newsletter.theaiedge.io/p/diffusion-models-stable-diffusion>

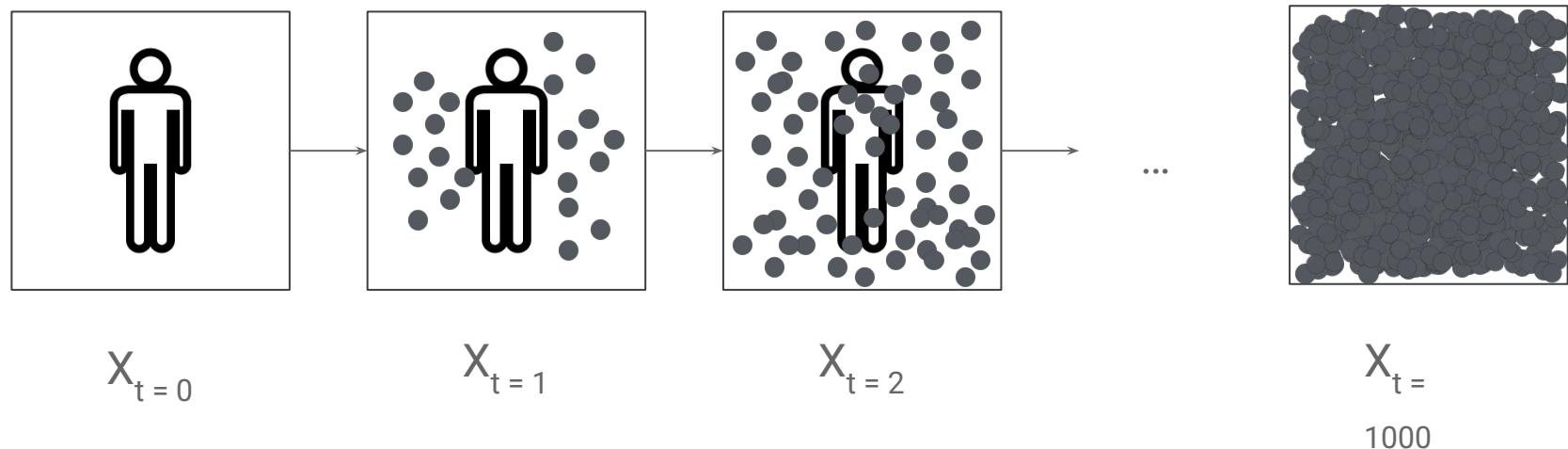
# Reverse diffusion (cont'd)

*Every block of stone has a statue  
inside it and it is the task of the  
sculptor to discover it.*

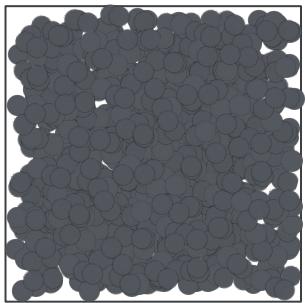
- *Michelangelo*



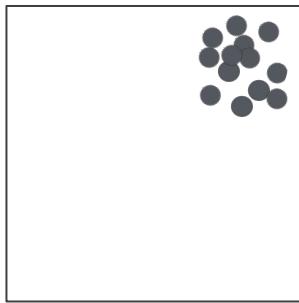
# Reverse diffusion (cont'd)



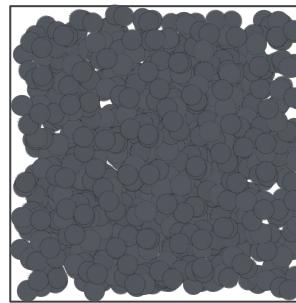
# Reverse diffusion (cont'd)



-



=

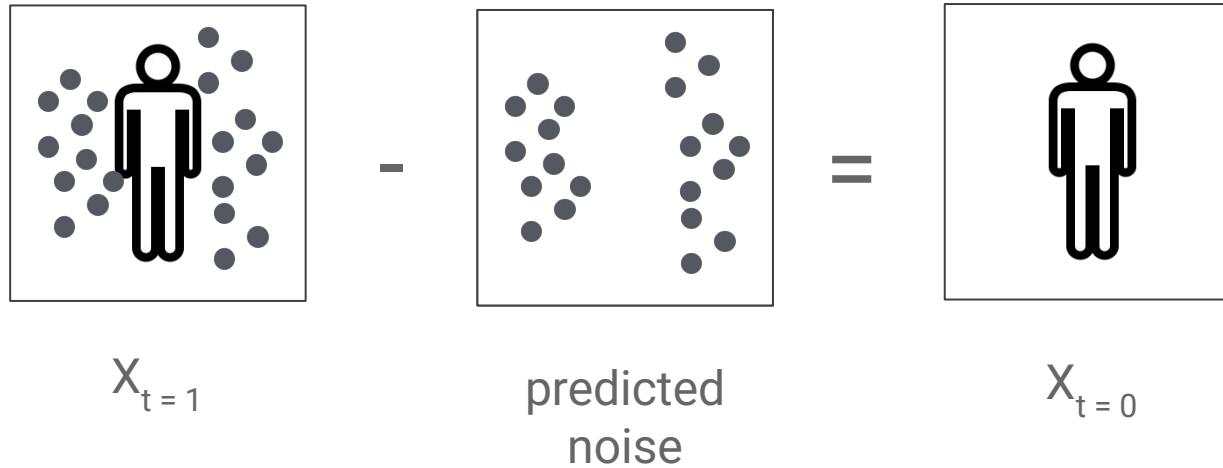


$X_{t=T}$

predicted  
noise

$X_{t=T-1}$

# Reverse diffusion (cont'd)

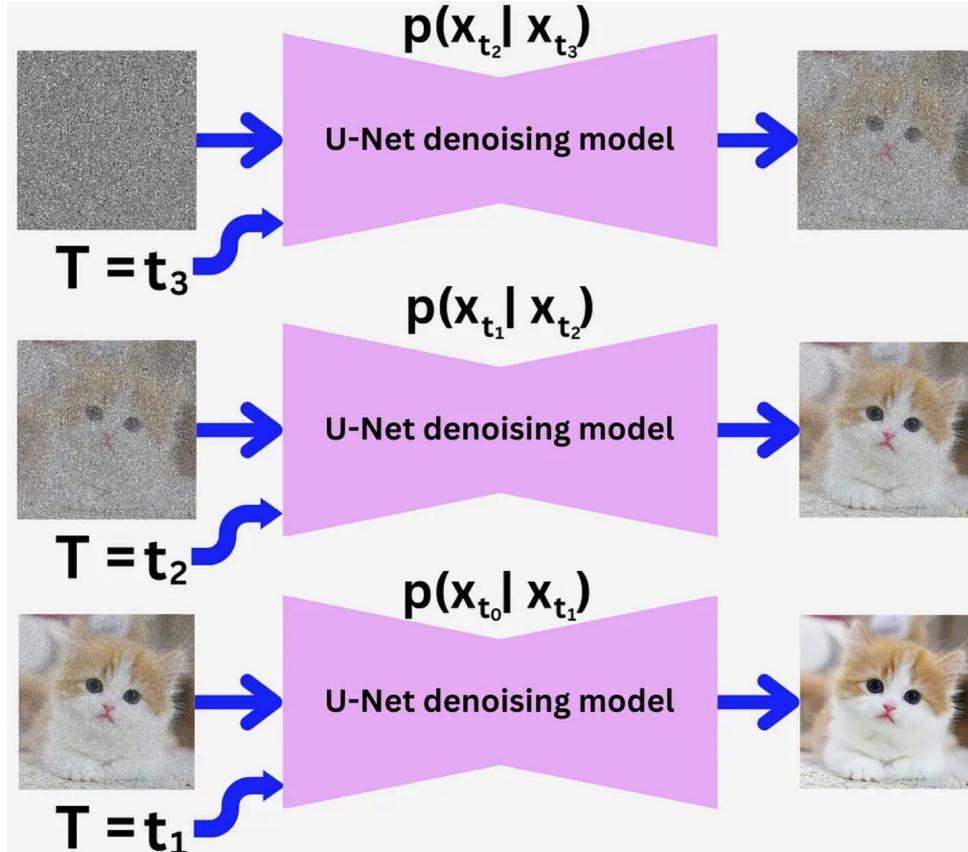


# Mean squared error

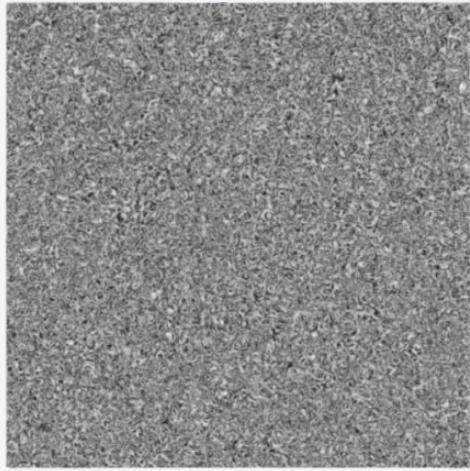
$$(\text{ground truth noise} - \text{predicted noise})^2$$

The equation illustrates the Mean Squared Error (MSE) calculation. It shows two square boxes representing data points. The left box contains a cluster of dark gray dots labeled "ground truth noise". The right box contains a similar cluster of dark gray dots labeled "predicted noise". A minus sign is positioned between the two boxes, indicating the subtraction step in the MSE formula.

# Reverse diffusion (cont'd)



# Conditional diffusion



Text embedding of  
“A turtle wearing  
sunglasses playing  
basketball”

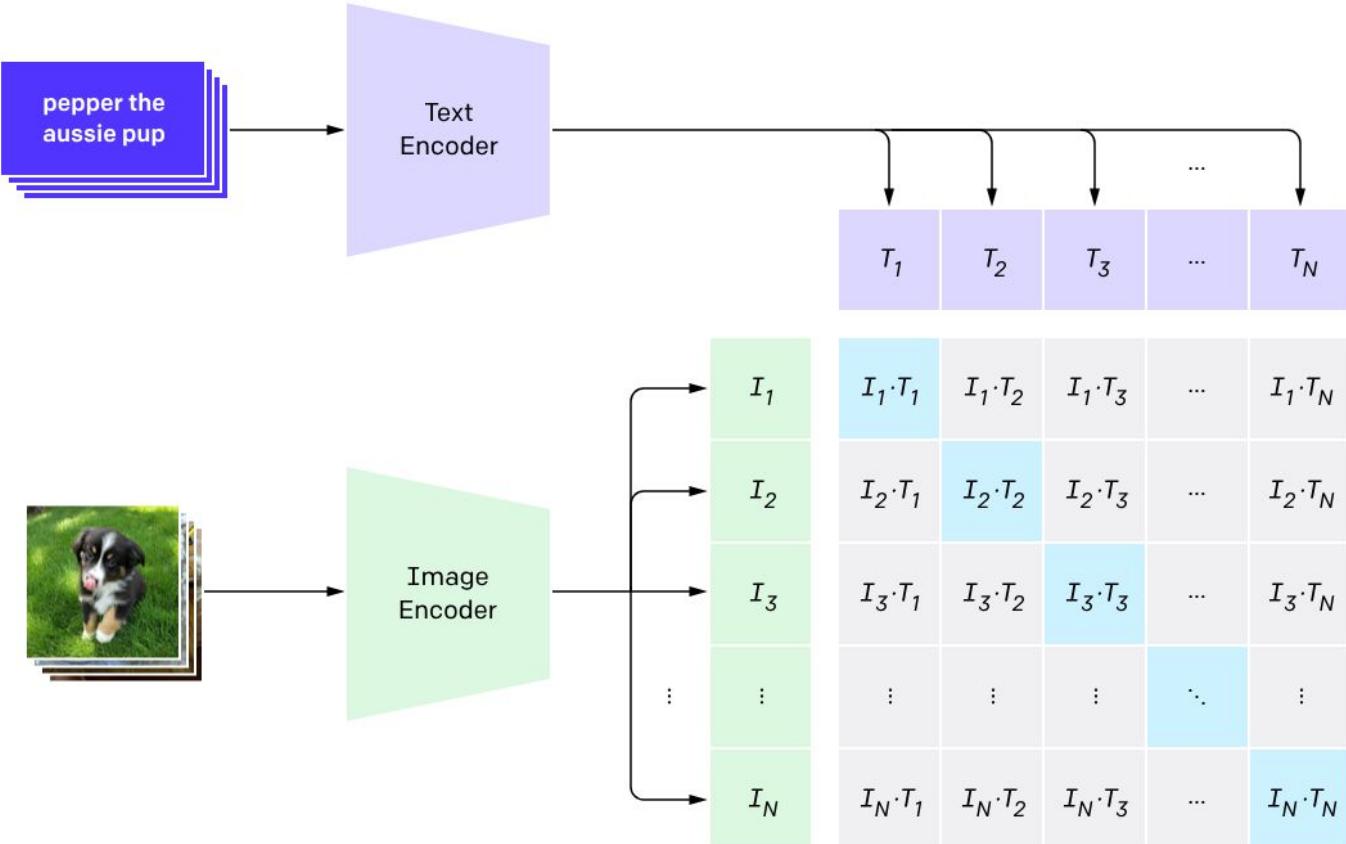
*The model learns the relationship between the meaning of words and how they correlate with certain denoising sequences that gradually reveal different features and shapes and edges in the picture.*

# Methods for incorporating text embeddings

- SAG: Self-attention guidance
  - Forces the model to pay attention to how specific portions of the prompt influenced the generation of certain regions or areas of the image
- Classifier-free diffusion guidance
  - Helps amplify the effect that certain words in the prompt on how the image is generated

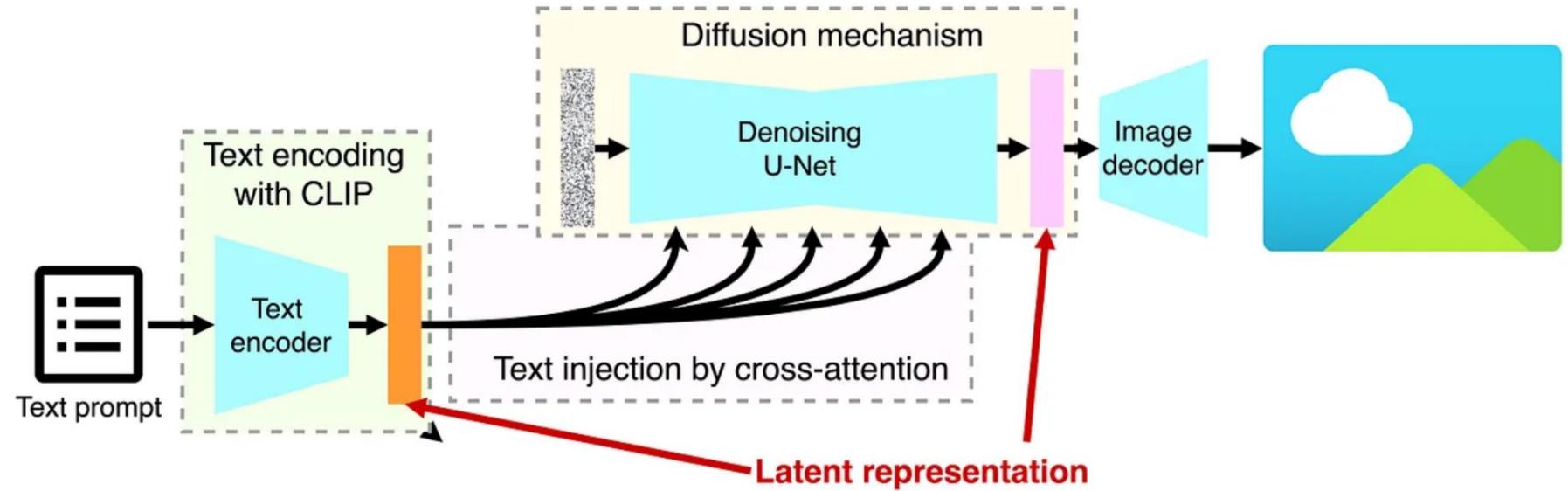
# OpenAI's CLIP

## 1. Contrastive pre-training

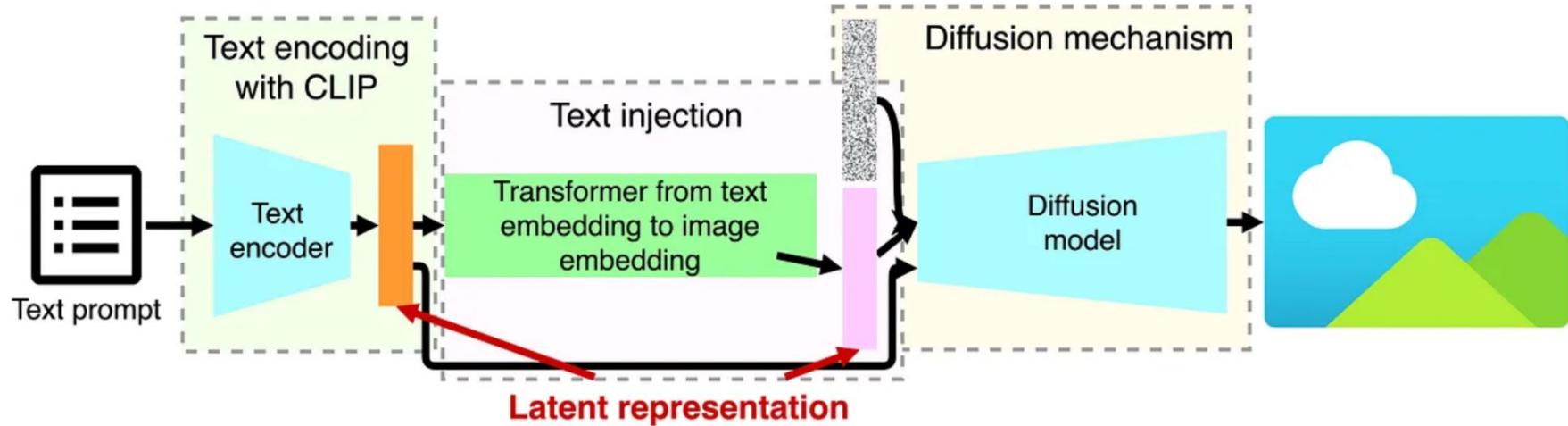


<https://openai.com/index/clip/>

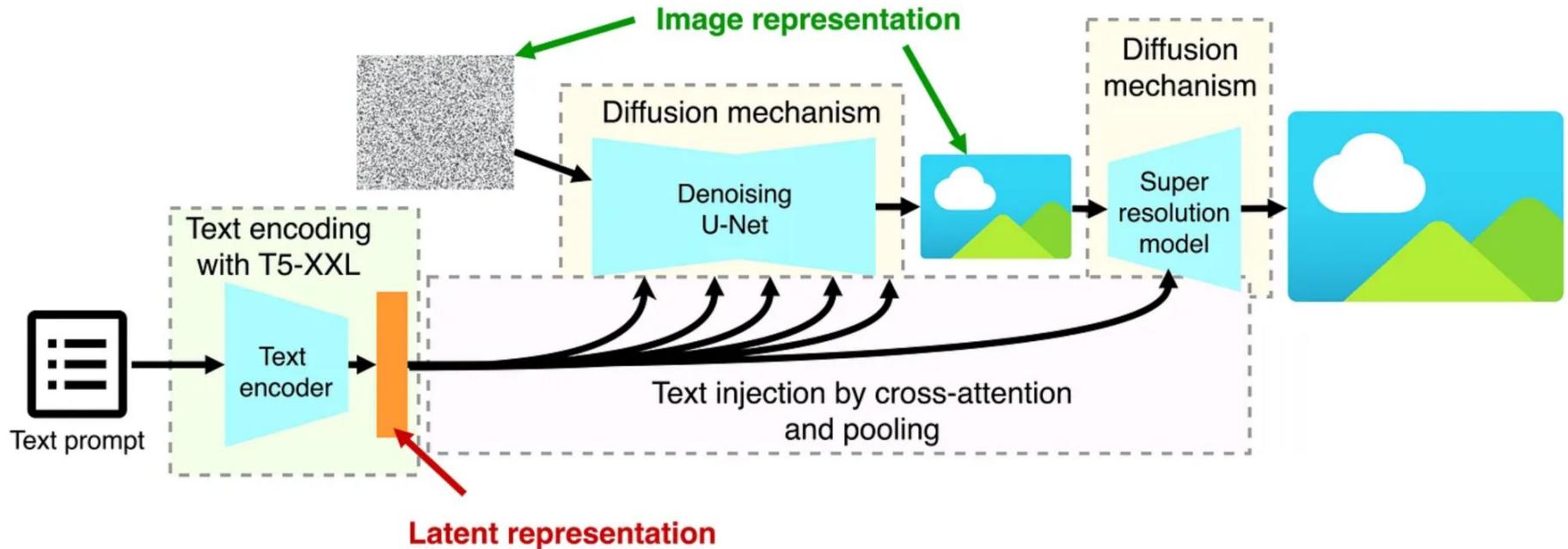
# Stability AI's Stable Diffusion



# DALL-E



# Imagen



# Sora

- <https://www.youtube.com/watch?v=fG3IE9dkyKY>

**Thank you!**