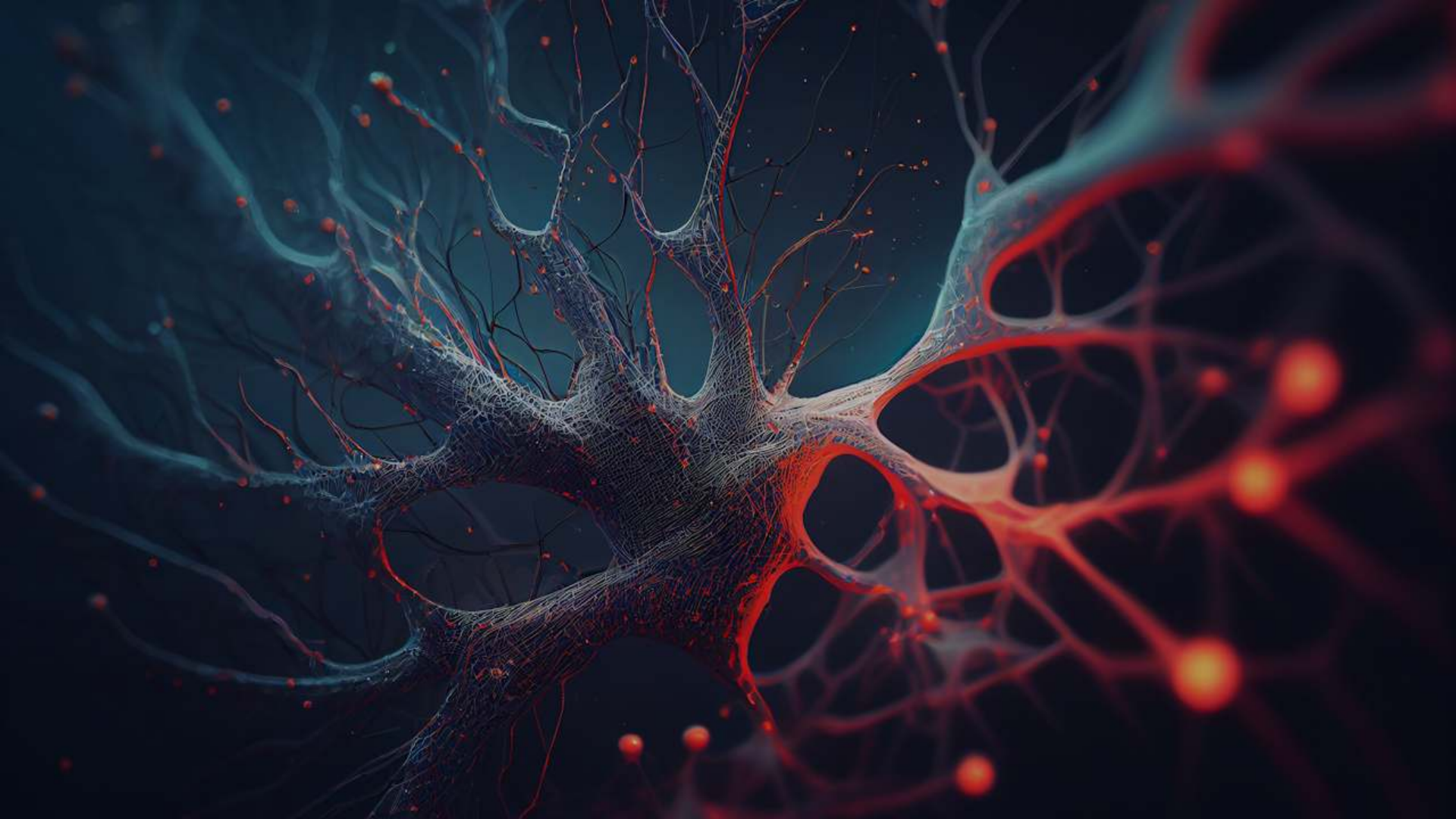
# AI'll Be Back

## Generative AI for Image and Video Production

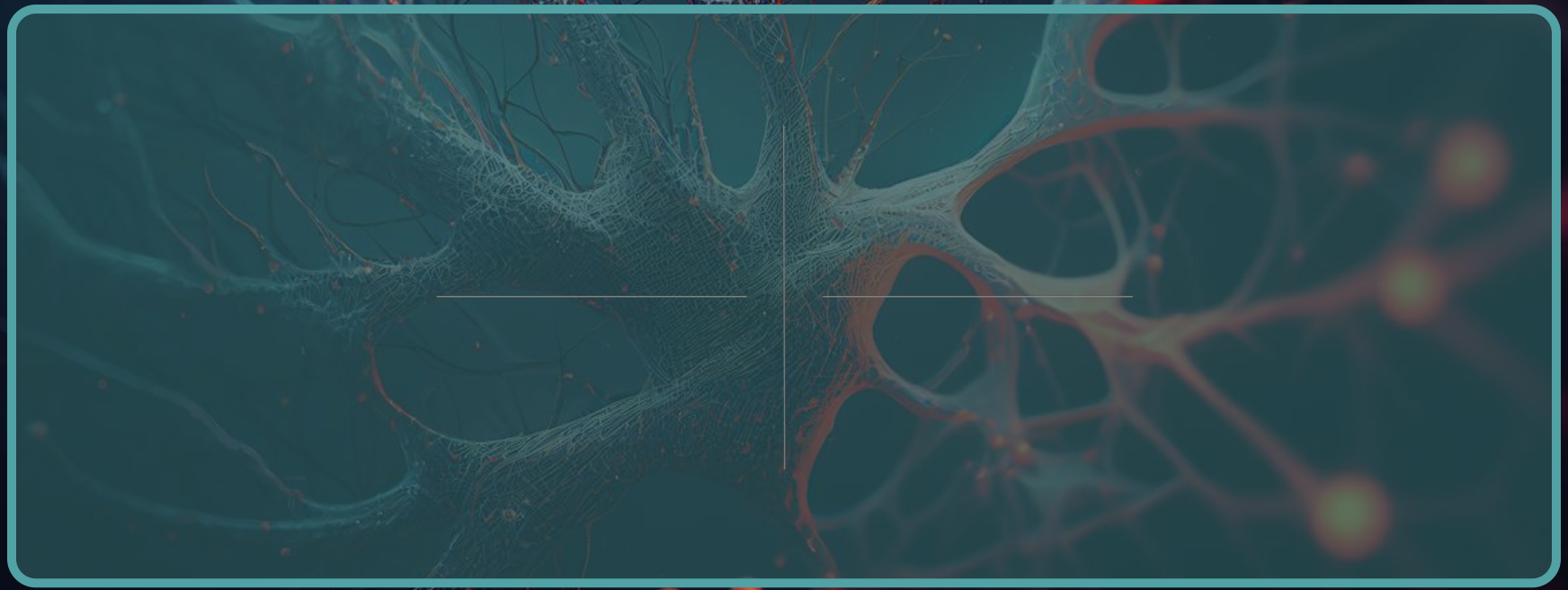
TNG TECHNOLOGY CONSULTING

AS SEEN ON TV



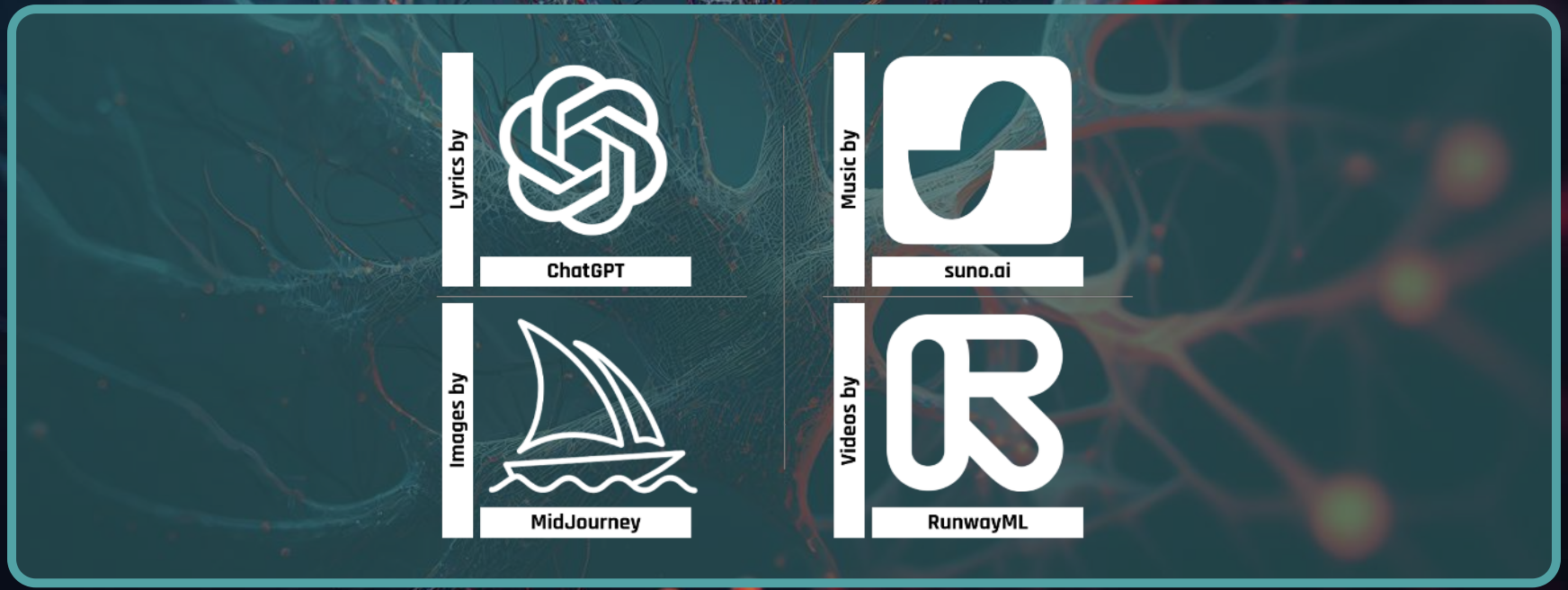
# Motivation

Question: What did you just see?



# Motivation

Entire music video created using Generative AI



# AI'll Be Back

## Generative AI for Image and Video Production

TNG TECHNOLOGY CONSULTING

AS SEEN ON TV

# Speakers

## Thomas Endres

### Managing Partner

Oracle® JavaOne Rockstar
Intel® Black Belt Software Developer
Intel® Software Innovator
Intel® Top Innovator

## Fabian Pottbäcker

### Software Consultant

ML Reverse Engineer
Space Nerd
(former) Competitive Programmer

# Overview

- Motivation
- Text → Text
- Text → Image
- Text → Audio
- Text → Video
- Ethics

# Motivation

## What Hollywood promised

AI Generated (Midjourney V6)

# Motivation

## What We Got

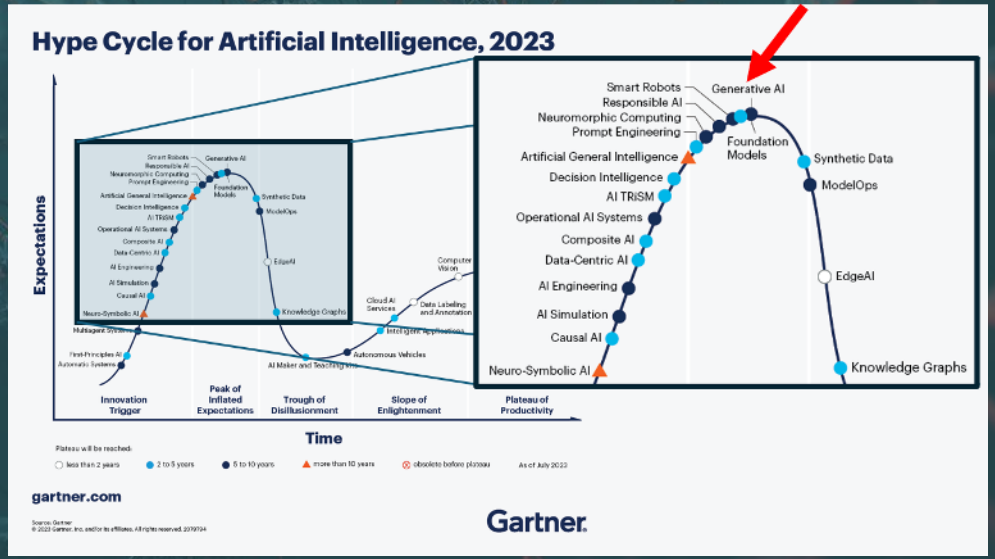
AI Generated (Midjourney V6)

Everyone: AI art will make designers obsolete

AI accepting the job:

# Motivation

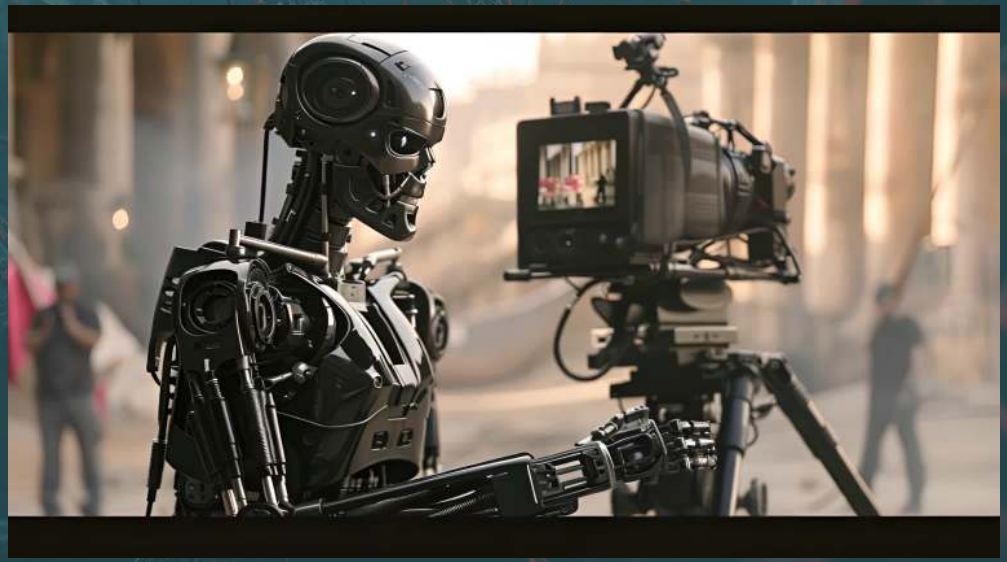
## Generative AI has reached the "Peak of Inflated Expectations"




https://www.gartner.com/en/articles/what-s-new-in-artificial-intelligence-from-the-2023-gartner-hype-cycle

# Motivation

## [Text | Image] → Video



AI Generated (RunwayML Gen-2)

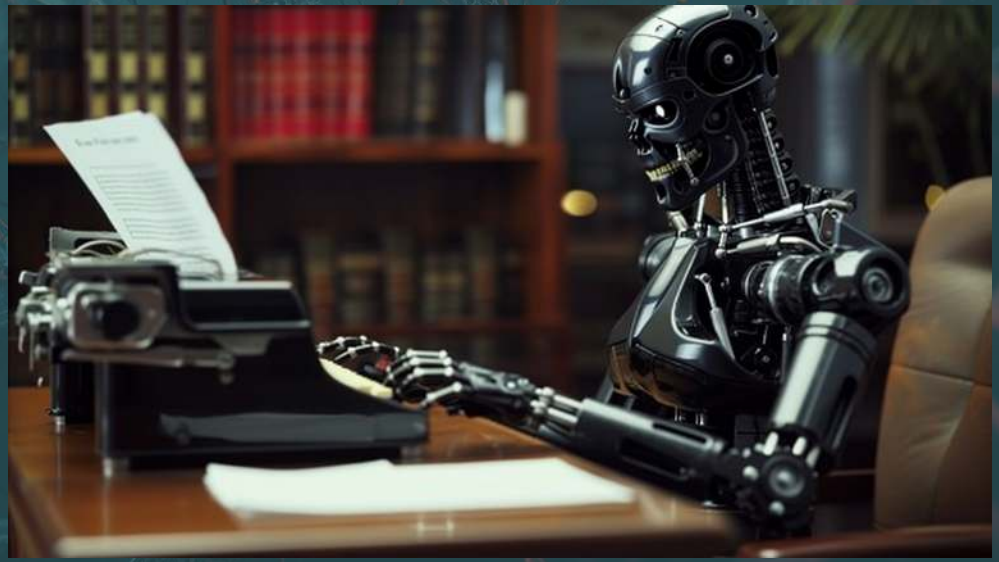
# Motivation

## Text → Audio

AI Generated (Midjourney V6)

# Motivation

## Text → Text



AI Generated (Midjourney V6)

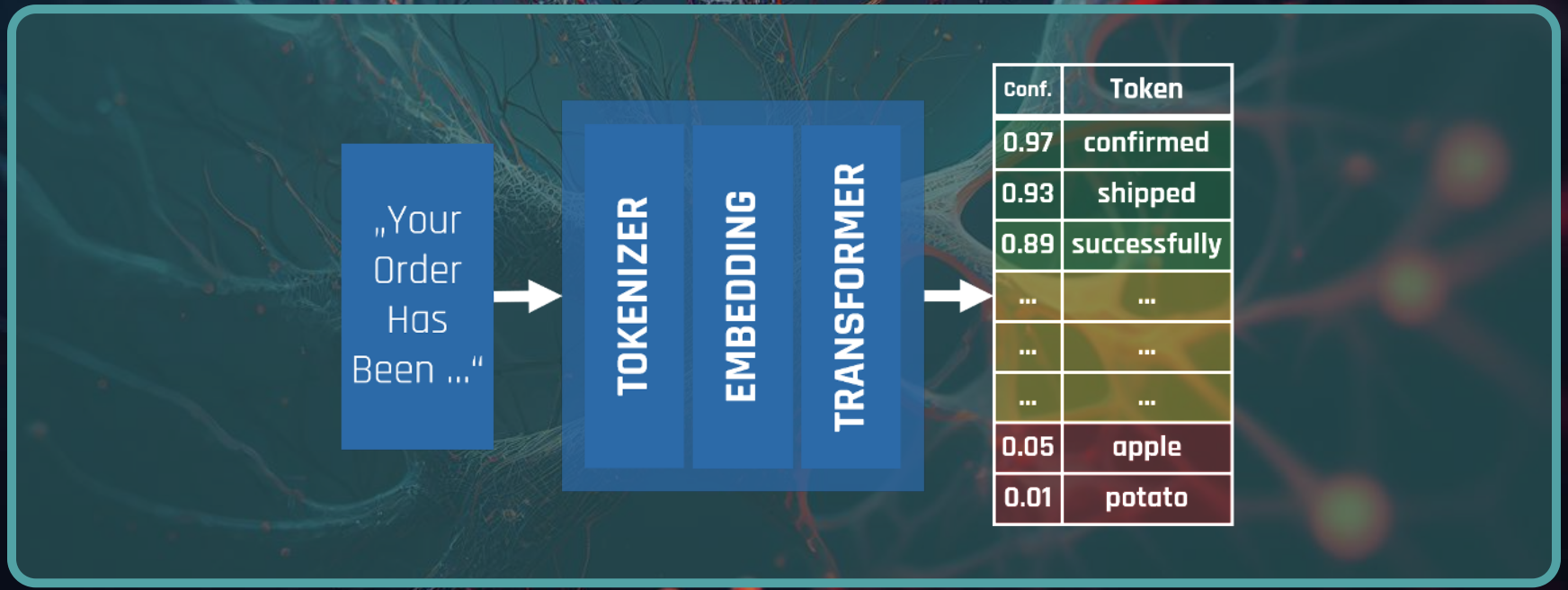
# Overview

- Motivation
- Text → Text
- Text → Image
- Text → Audio
- Text → Video
- Ethics

Say "HI!" to the listeners of our presentation about "Generative AI" and introduce yourself briefly.

# Generative Pretrained Transformer (GPT)

## Next Word Predictor

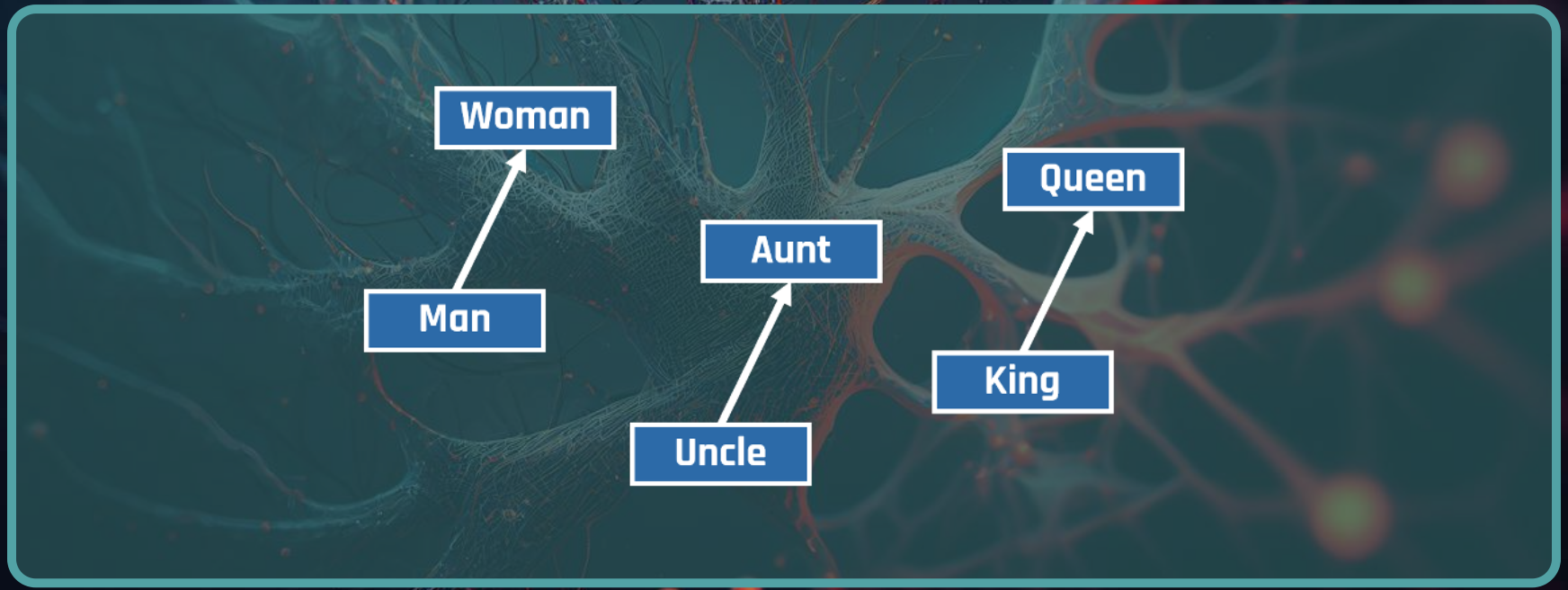


# Tokenization

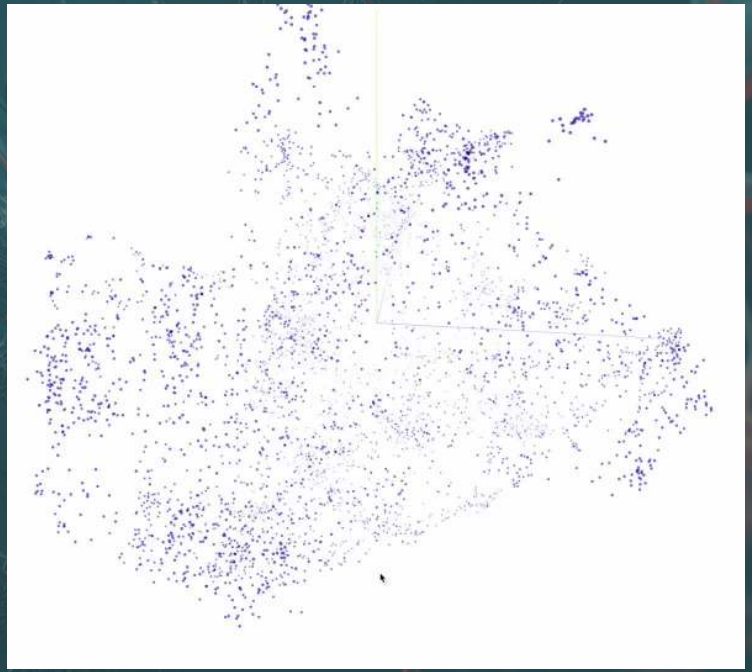
## Demo

# Embeddings

## Turning Numbers into Vectors



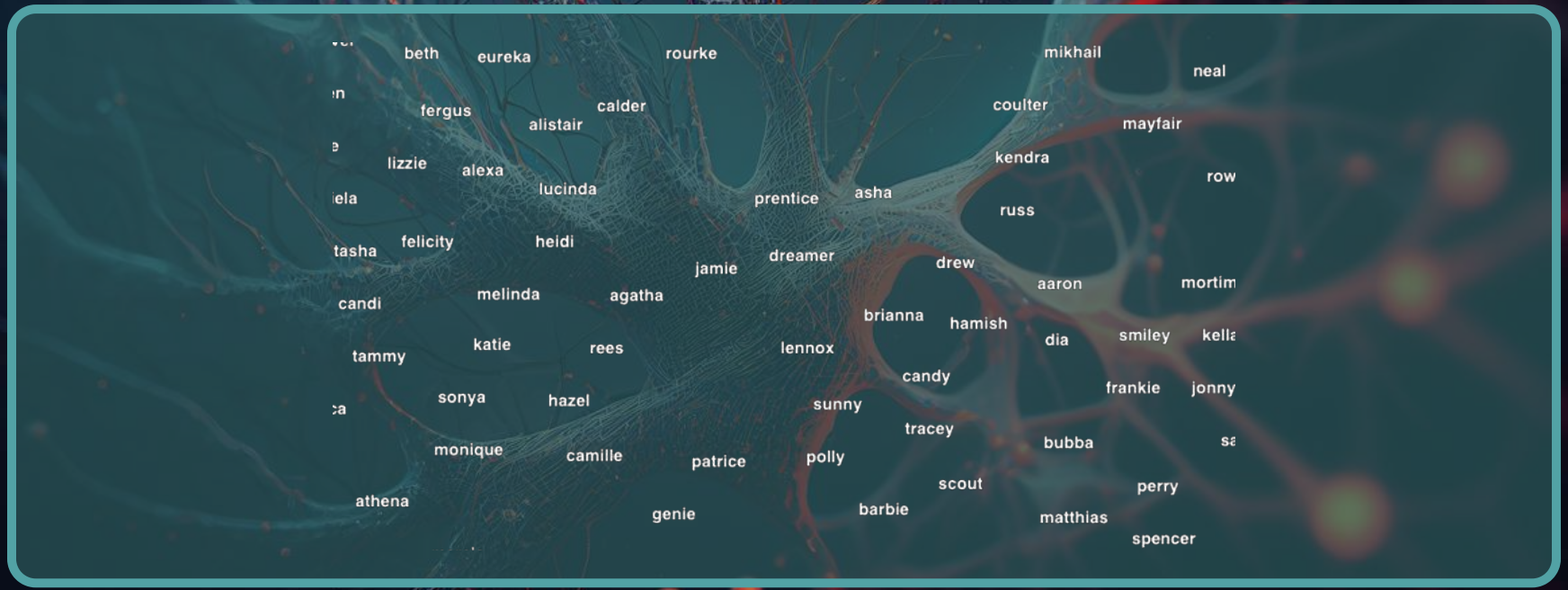
# Embeddings in 3D



https://projector.tensorflow.org/

# Embeddings

## Google BERT Embedding Cluster (First Names)



# Embeddings

## Google BERT Embedding Cluster (Place Names)

# Generative Pretrained Transformer (GPT)

## A Neural Network Architecture With a Secret Sauce


AI Generated (Midjourney V6)

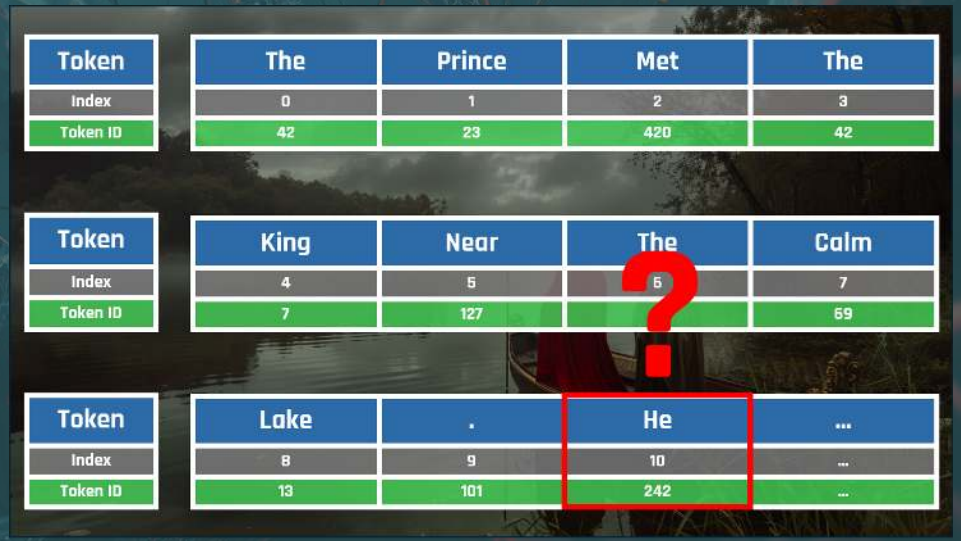
# Self-Attention

## The Mechanism Driving Context Awareness in Transformers

AI Generated (Midjourney V6)

# Self-Attention

## The Mechanism Driving Context Awareness in Transformers

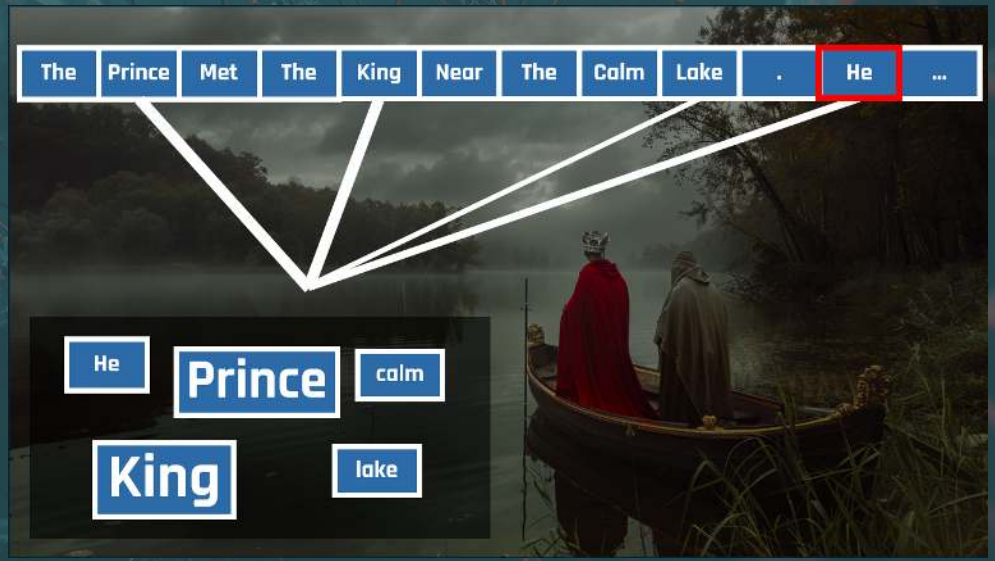



AI Generated (Midjourney V6)


He

# Self-Attention

## The Mechanism Driving Context Awareness in Transformers

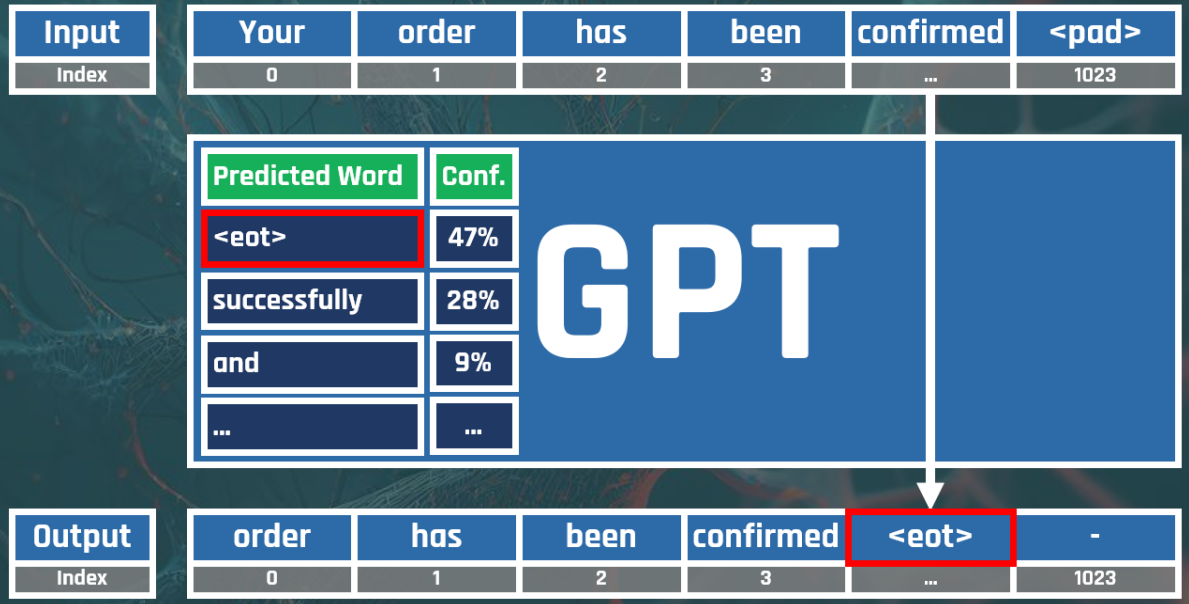



AI Generated (Midjourney V6)

# Transformers

## Putting it all Together

# GPT Example


Input
Index
Your
order
has
been
confirmed
<pad>
0
1
2
3
...
1023
Predicted Word
Conf.
<eot>
47%
successfully
28%
and
9%
...
...
GPT
Output
Index
order
has
been
confirmed
<eot>
-
0
1
2
3
...
1023

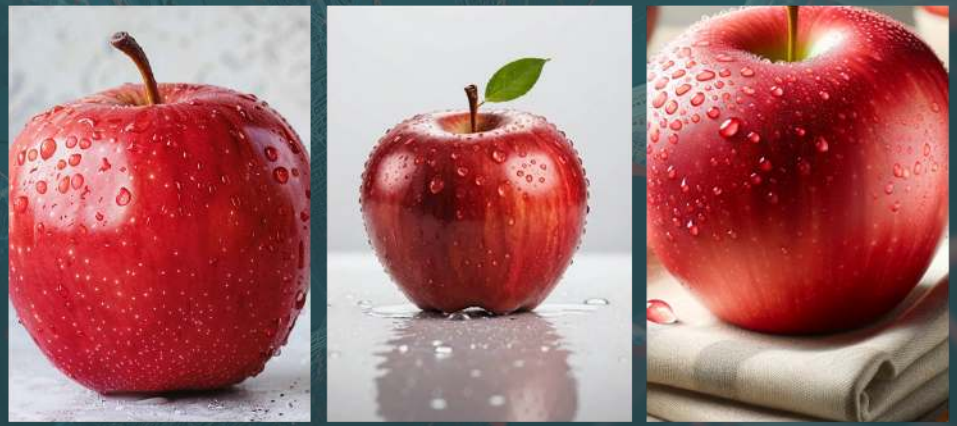
# Overview

- Motivation
- Text → Text
- Text → Image
- Text → Audio
- Text → Video
- Ethics

# Text → Image

MidJourney, Stable Diffusion, DALL-E, Imagen, ...

*"Red Apple With Water Droplets, Neutral Background"*



MidJourney | Stable Diffusion | DALL·E

# Text-to-Image AI

## MidJourney V1 Was Released on February 2022

AI Generated (Midjourney V1): Red Apple With Water Droplets, Neutral Background

# Text-to-Image AI

## MidJourney V2 Was Released on April 2022 (+2 Months)

AI Generated (Midjourney V2): Red Apple With Water Droplets, Neutral Background

# Text-to-Image AI

## MidJourney V3 Was Released on July 2022 (+3 Months)


AI Generated (Midjourney V3): Red Apple With Water Droplets, Neutral Background

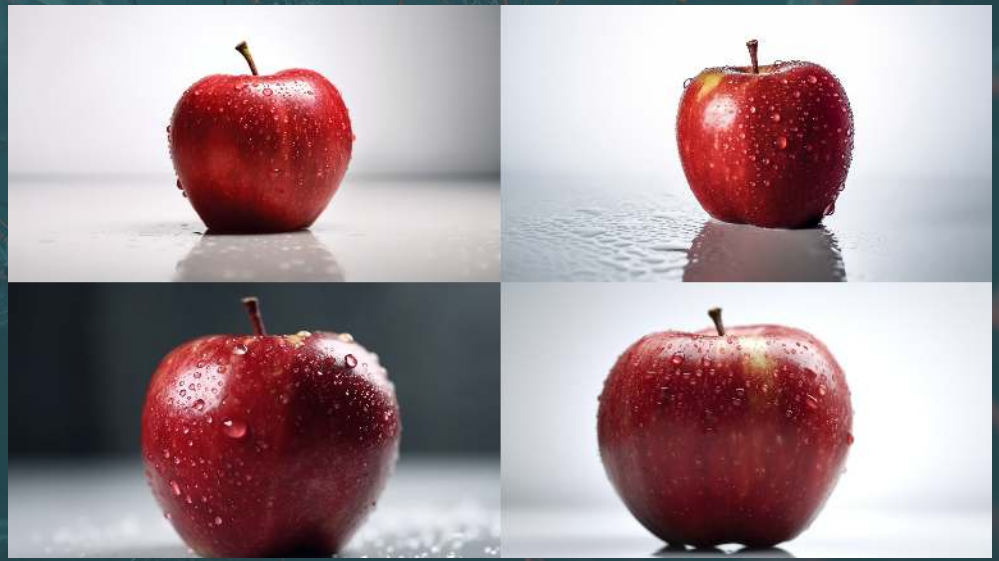
# Text-to-Image AI

## MidJourney V4 Was Released on November 2022 (+4 Months)

AI Generated (Midjourney V4): Red Apple With Water Droplets, Neutral Background

# Text-to-Image AI

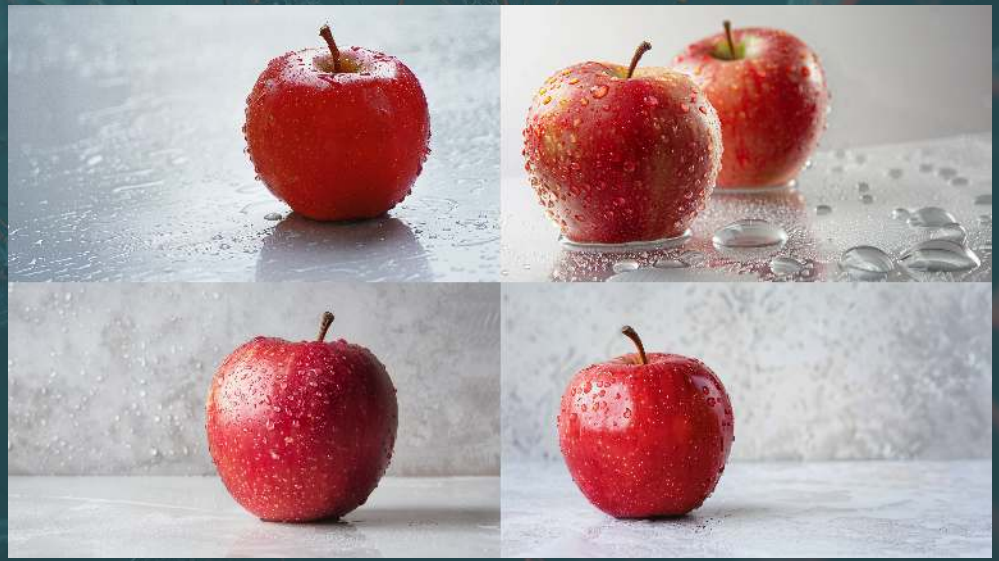
## MidJourney V5 Was Released on March 2023 (+4 Months)



AI Generated (Midjourney V5): Red Apple With Water Droplets, Neutral Background

# Text-to-Image AI

## MidJourney V6 Was Released on Dezember 2023 (+9 Months)



AI Generated (Midjourney V6): Red Apple With Water Droplets, Neutral Background

# CLIP

## Contrastive Language-Image Pre-training

„a red delicious apple“

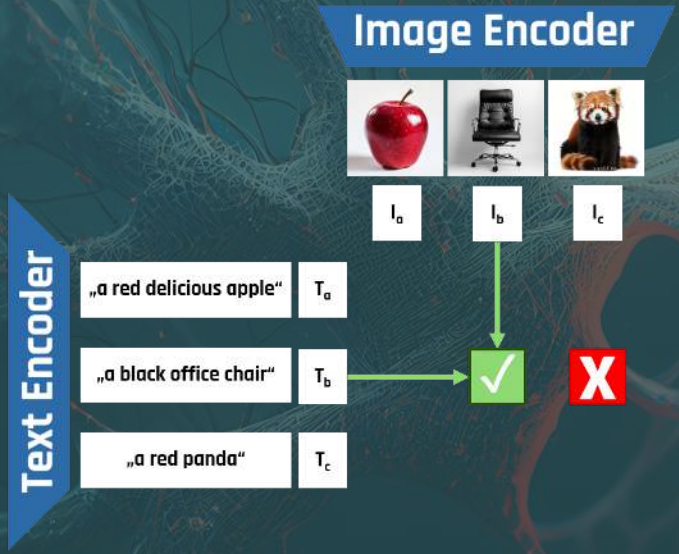
„a black office chair“

„a red panda“

AI Generated (Midjourney V6)

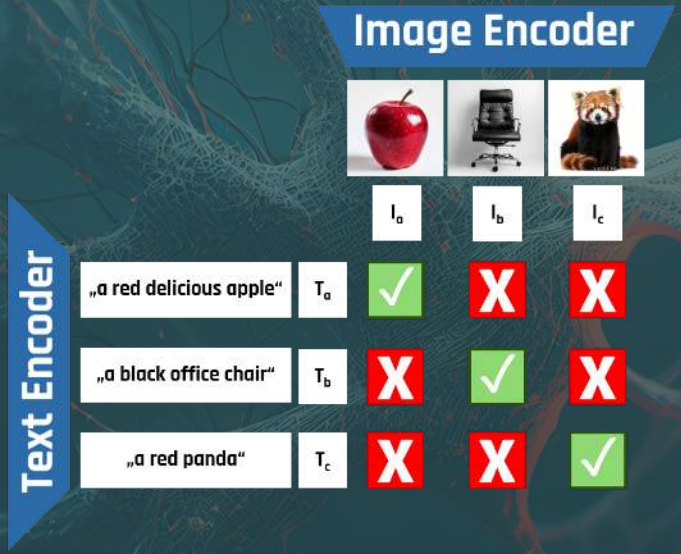
# CLIP

## Computes Similarity for each Text-Image Pair



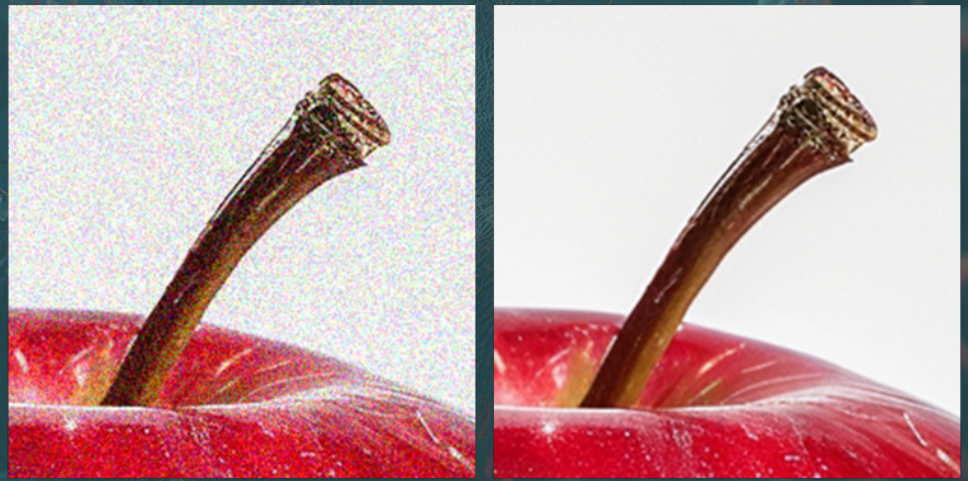
# CLIP

Maximizes Similarity for correct Pairs, Minimizes for others



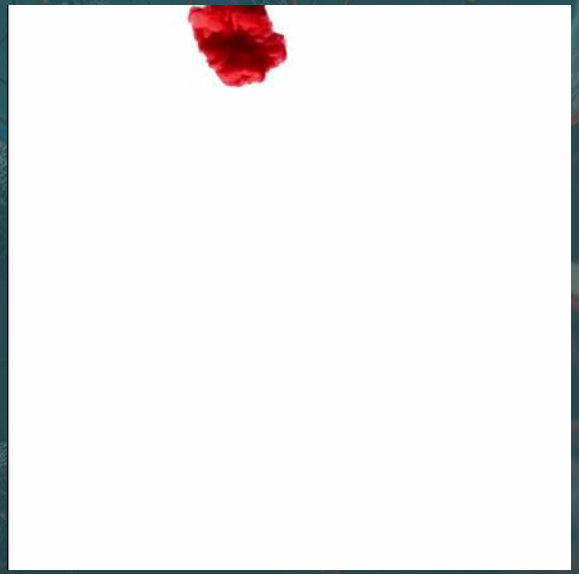
# Denoising

## Traditional Ways of Removing Noise in Pictures



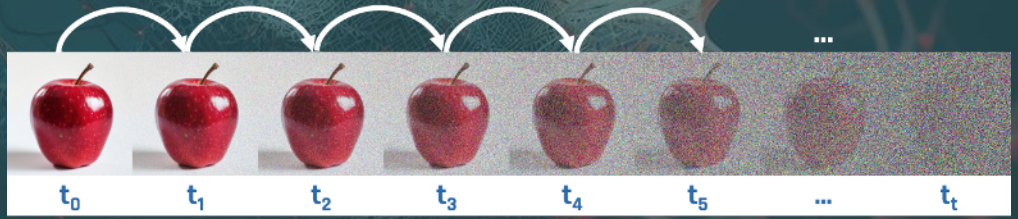
# Diffusion Models

## Learning using Nonequilibrium Thermodynamics



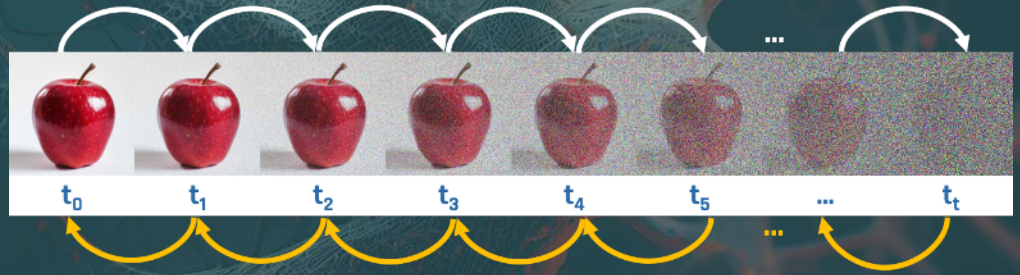
# Diffusion Process

## Add Noise to Image in multiple Steps



# Diffusion Models

## Learn to Reverse this Noise Adding Process



# Diffusion Process

## Starting With an Original Image

$t_0$

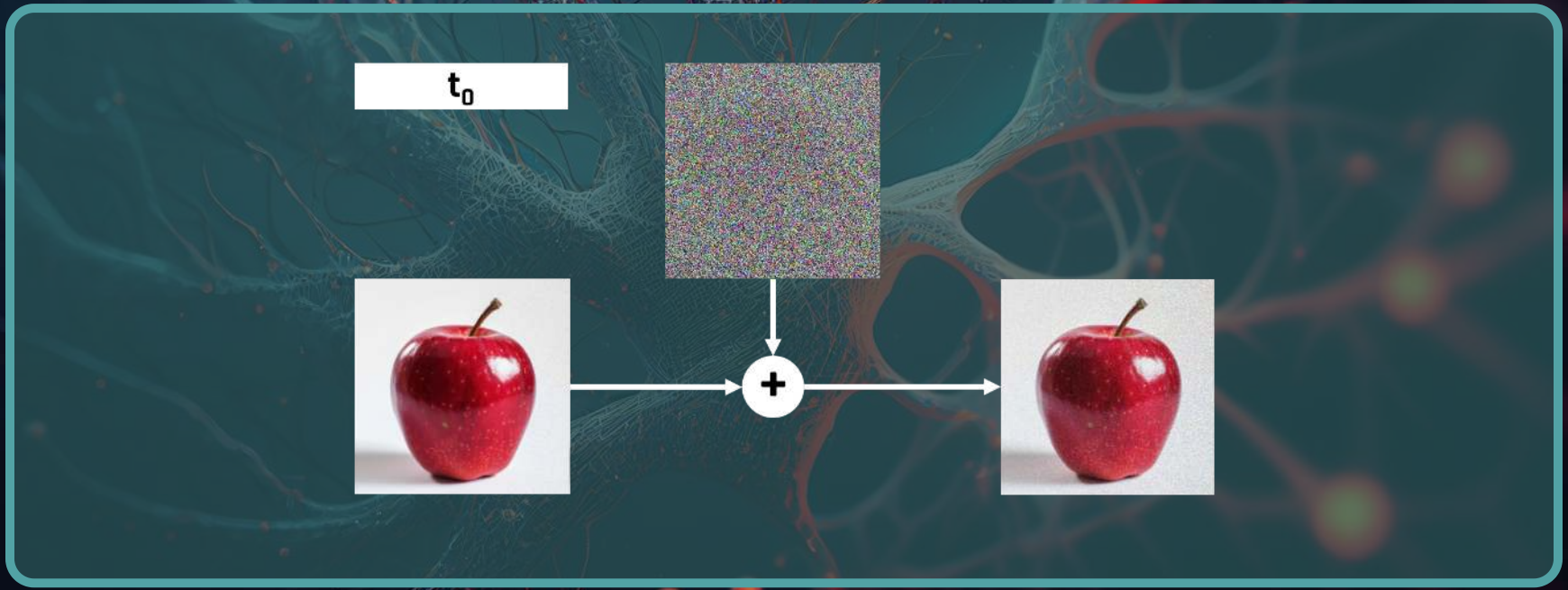
# Diffusion Process

## Generate Noise for Current Step t

$t_0$

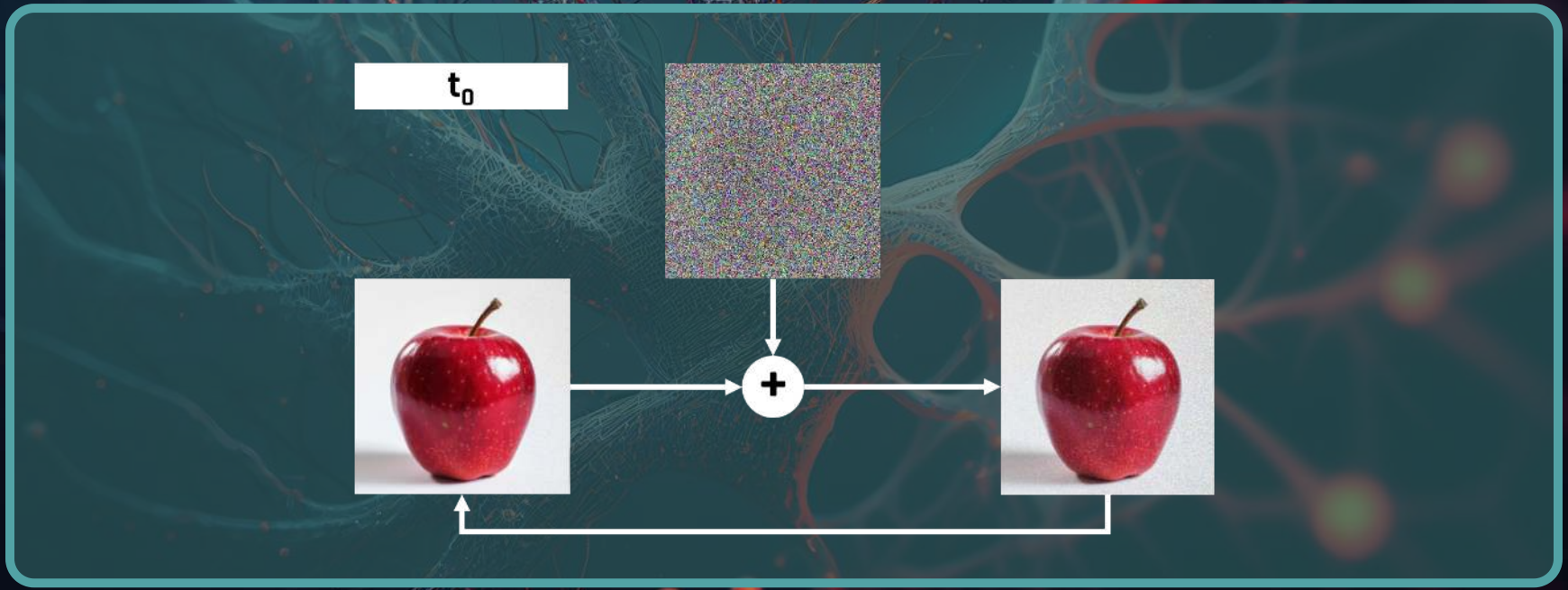
# Diffusion Process

## Add Noise to the Input Image Using Linear Scheduler



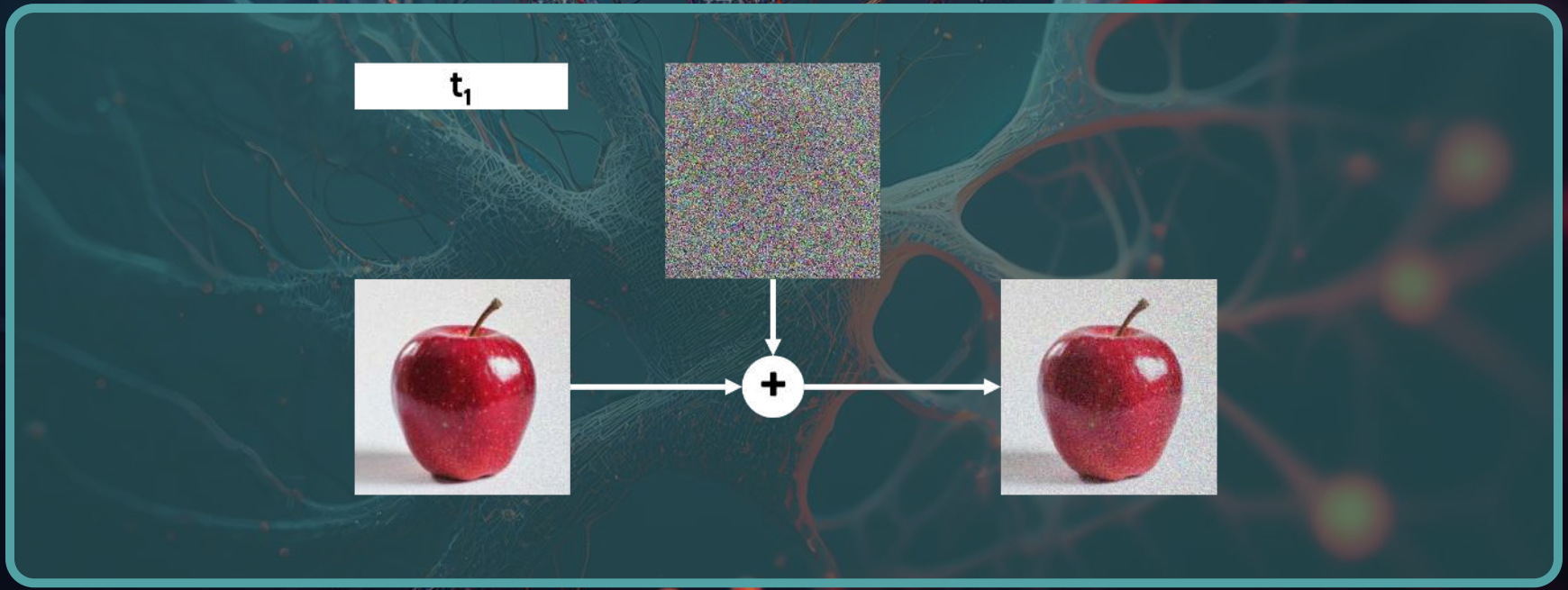
# Diffusion Process

## Output From the Step t Becomes Input in the t+1 Step



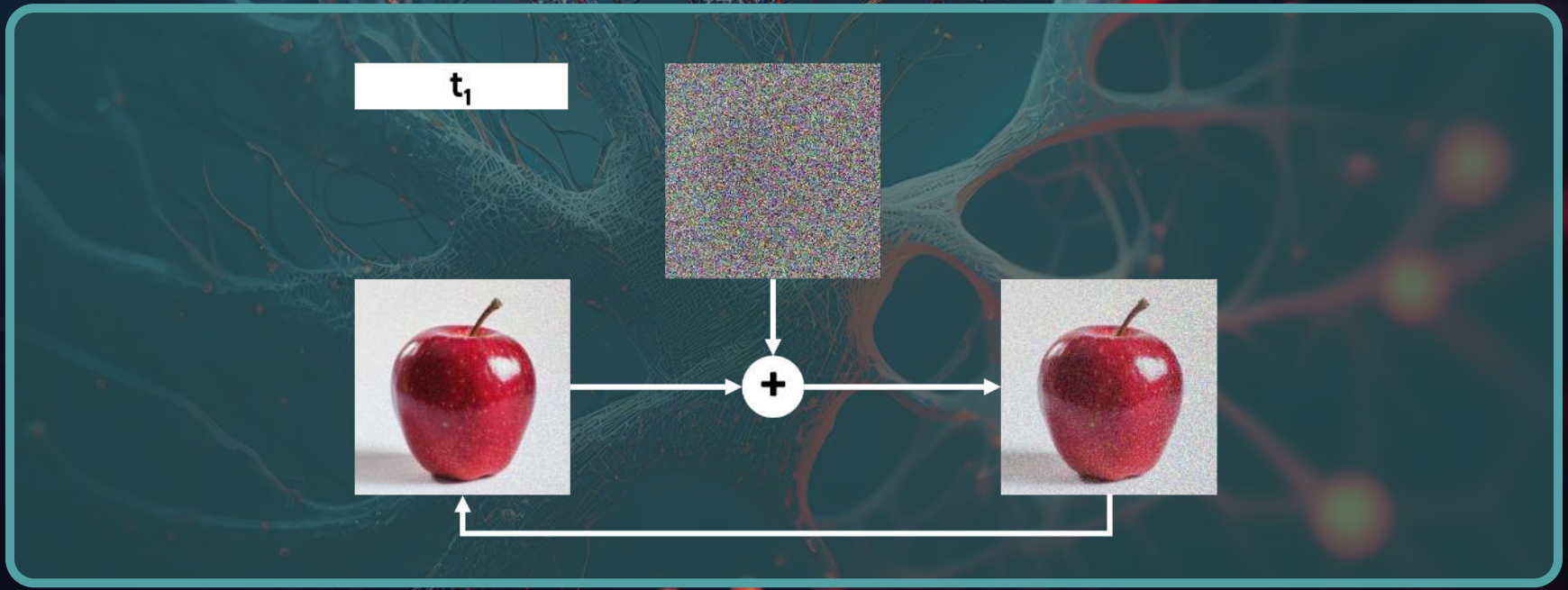
# Diffusion Process

## Add Noise to the Input Image Using Linear Scheduler



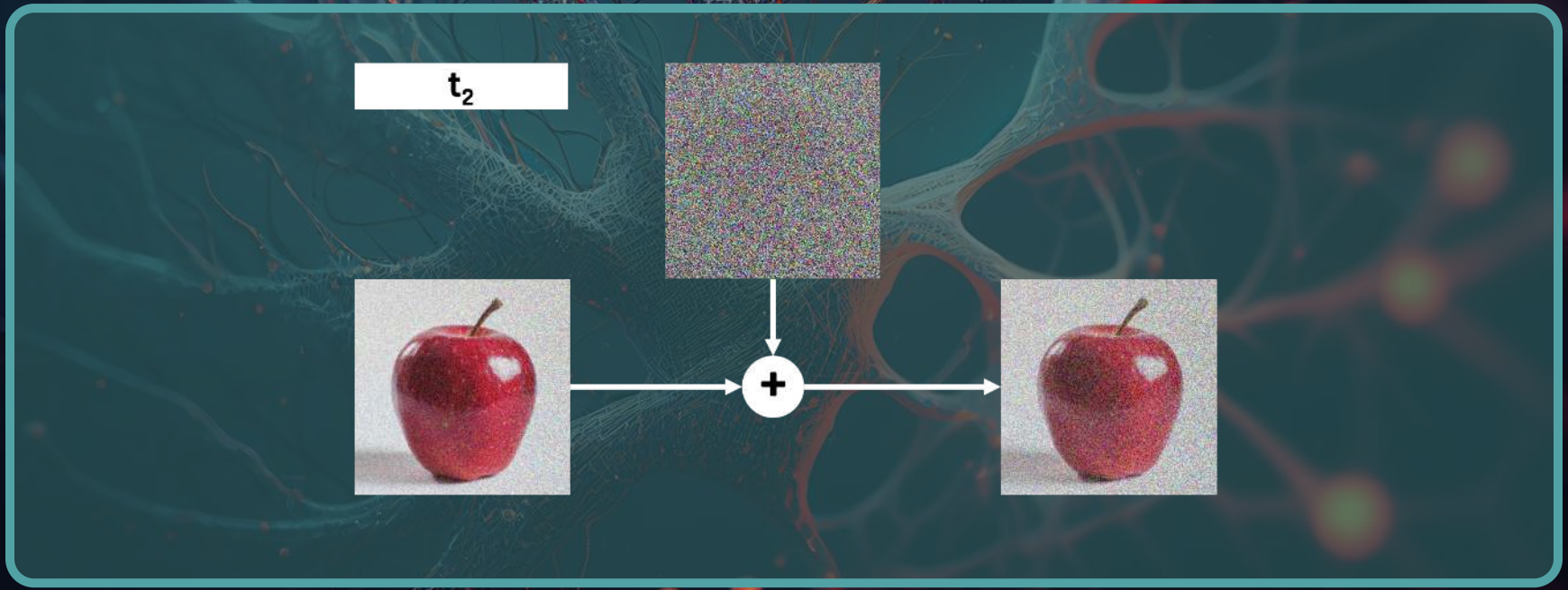
# Diffusion Process

## Output From the Step t Becomes Input in the t+1 Step



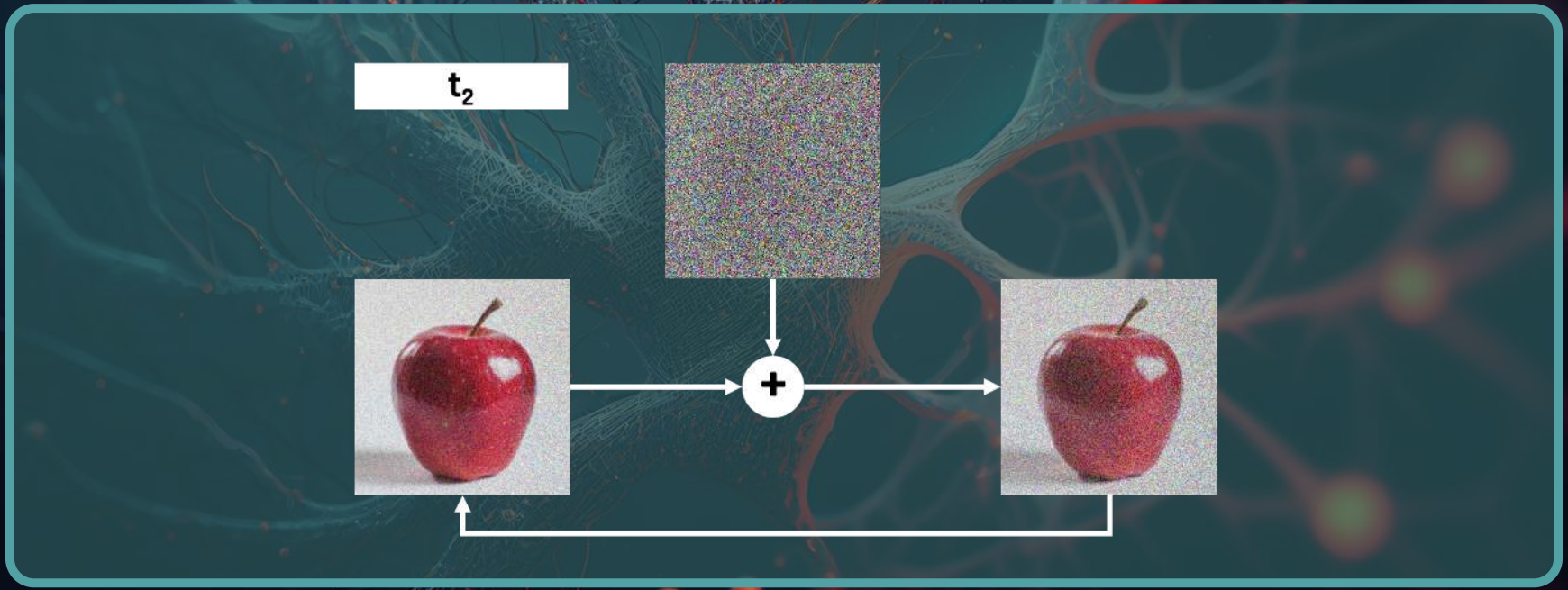
# Diffusion Process

## Add Noise to the Input Image Using Linear Scheduler



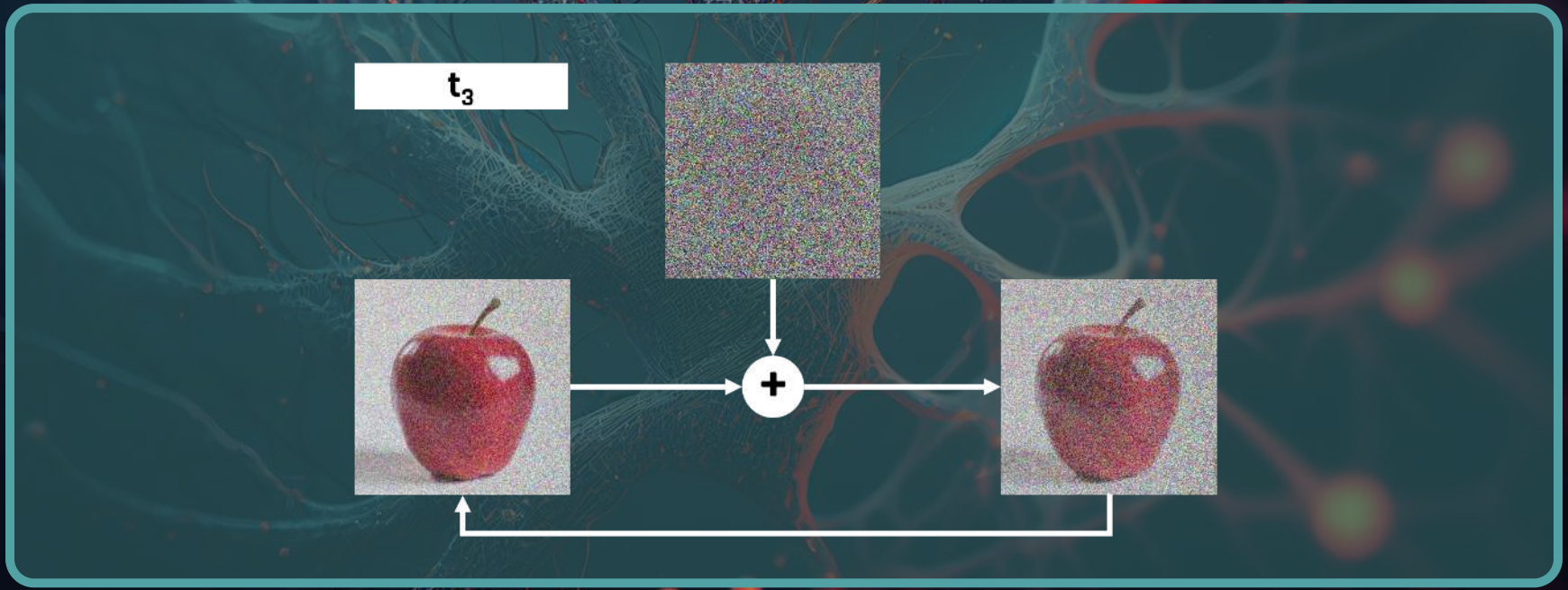
# Diffusion Process

## Output From the Step t Becomes Input in the t+1 Step



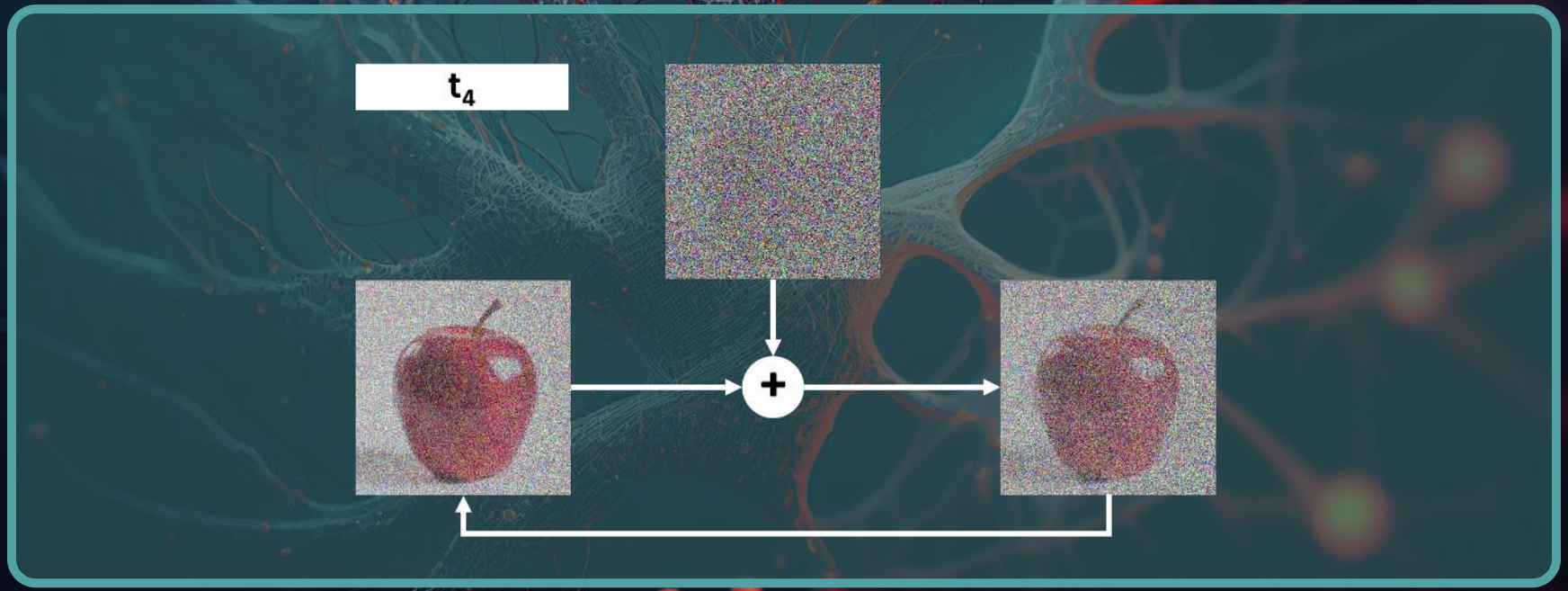
# Diffusion Process

## Output From the Step t Becomes Input in the t+1 Step



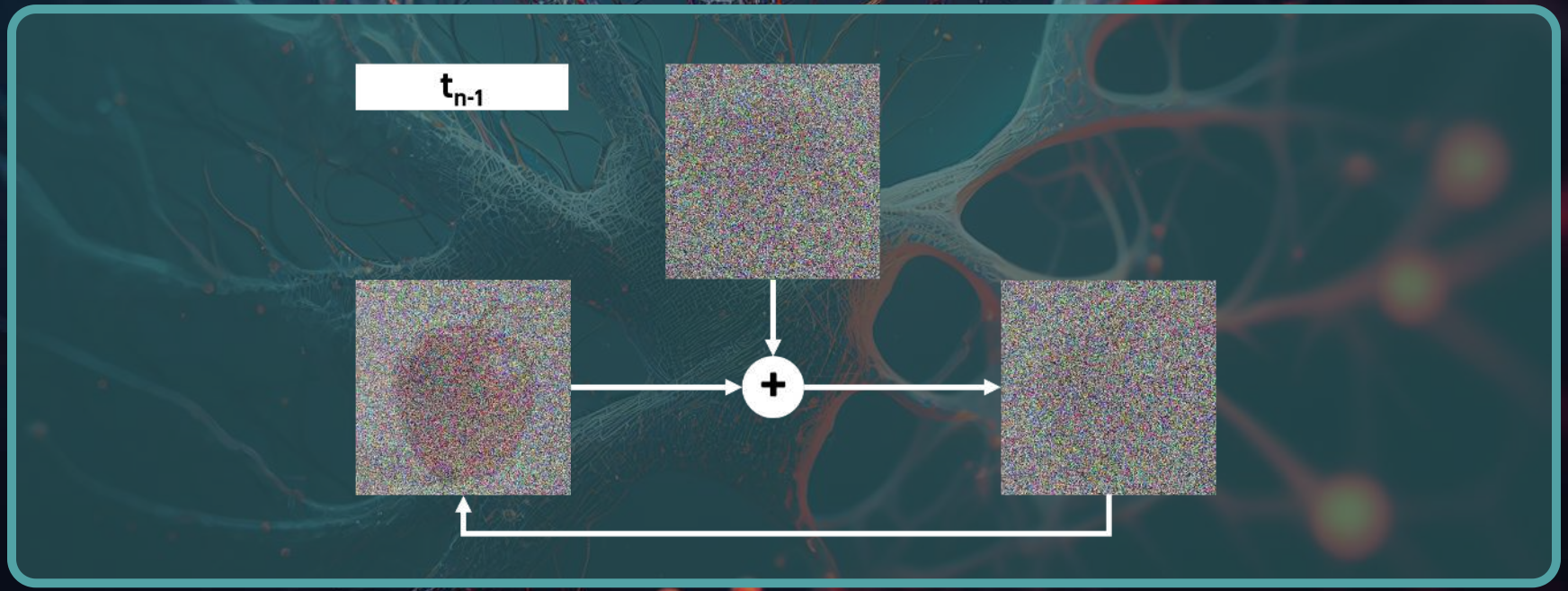
# Diffusion Process

## Using Linear Scheduler, Information is Lost Pretty Quickly



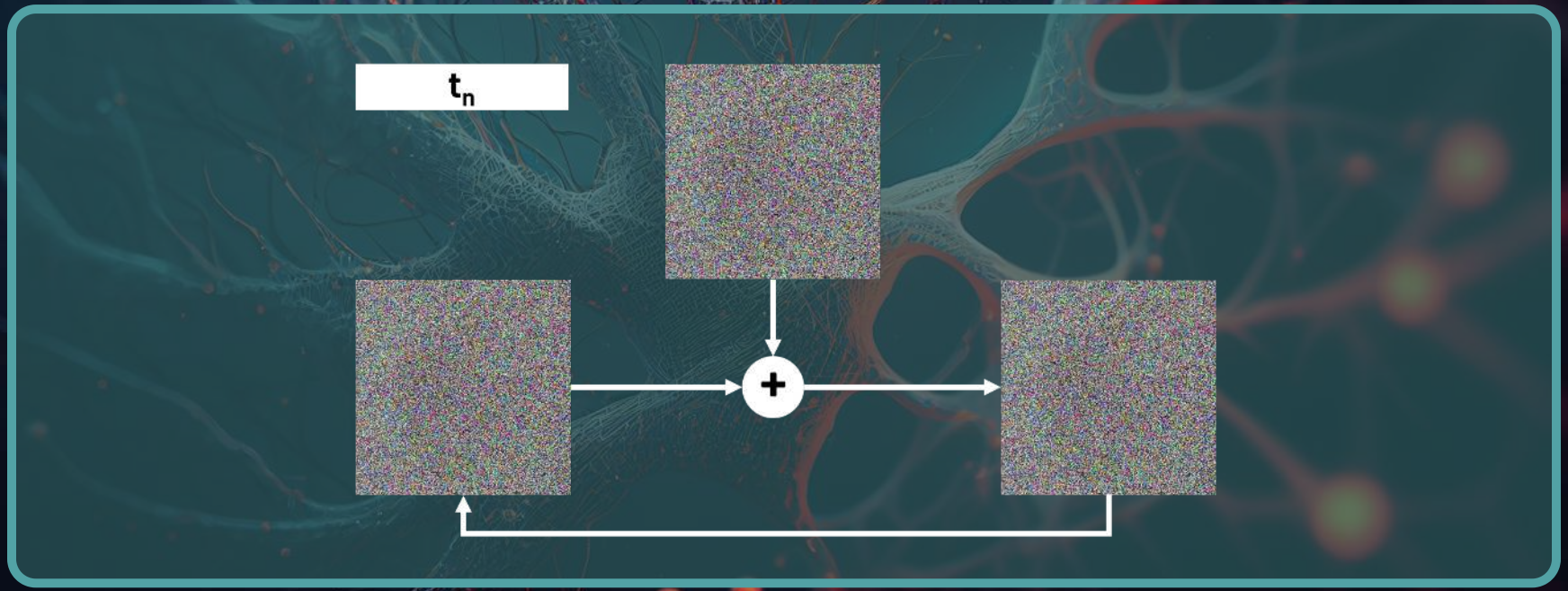
# Diffusion Process

## Output From the Step t Becomes Input in the t+1 Step



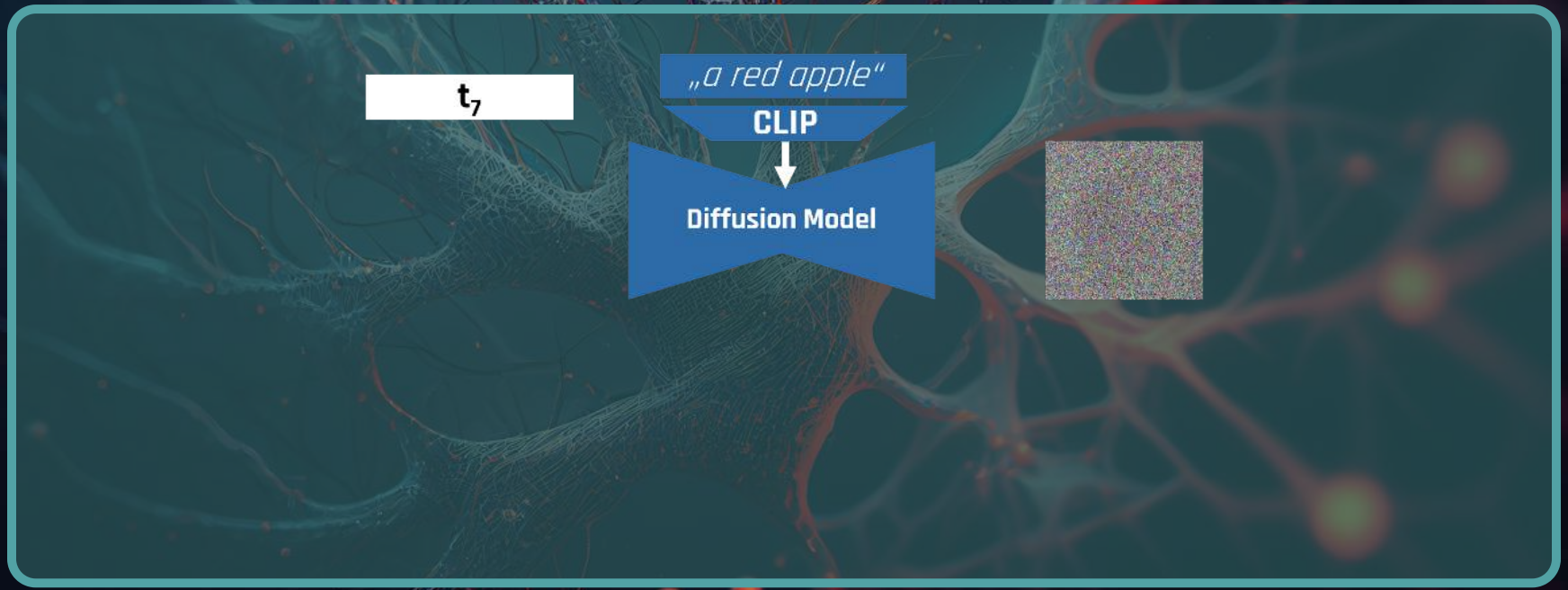
# Diffusion Process

## Output From the Step t Becomes Input in the t+1 Step



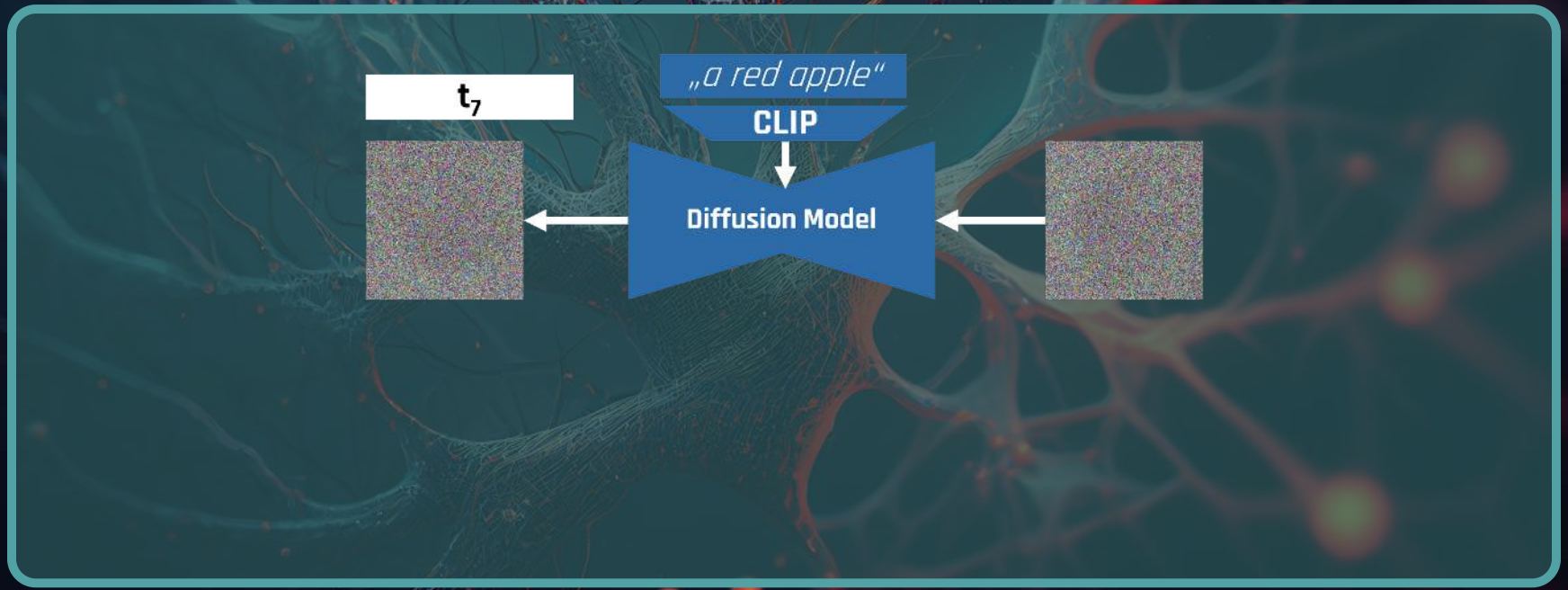
# Reverse Diffusion Process

## Given a pre-trained Model



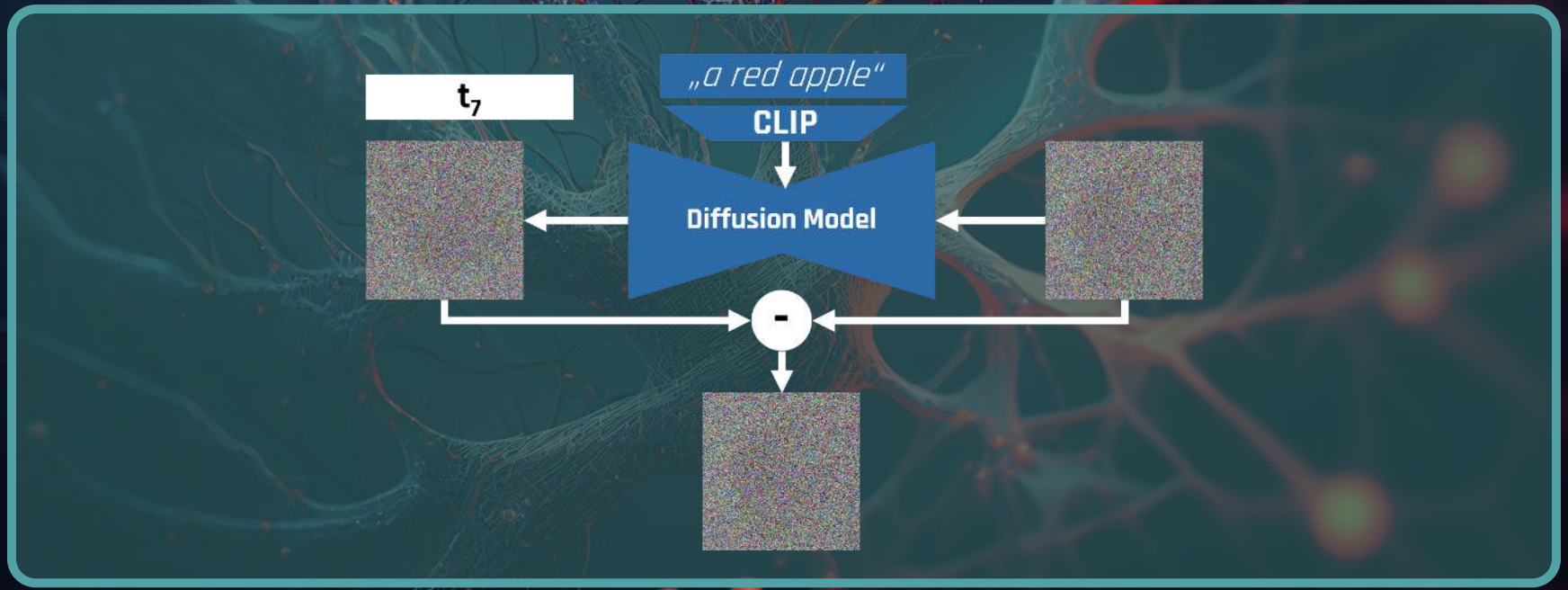
# Reverse Diffusion Process

## Output of the Model is an Entire Noise



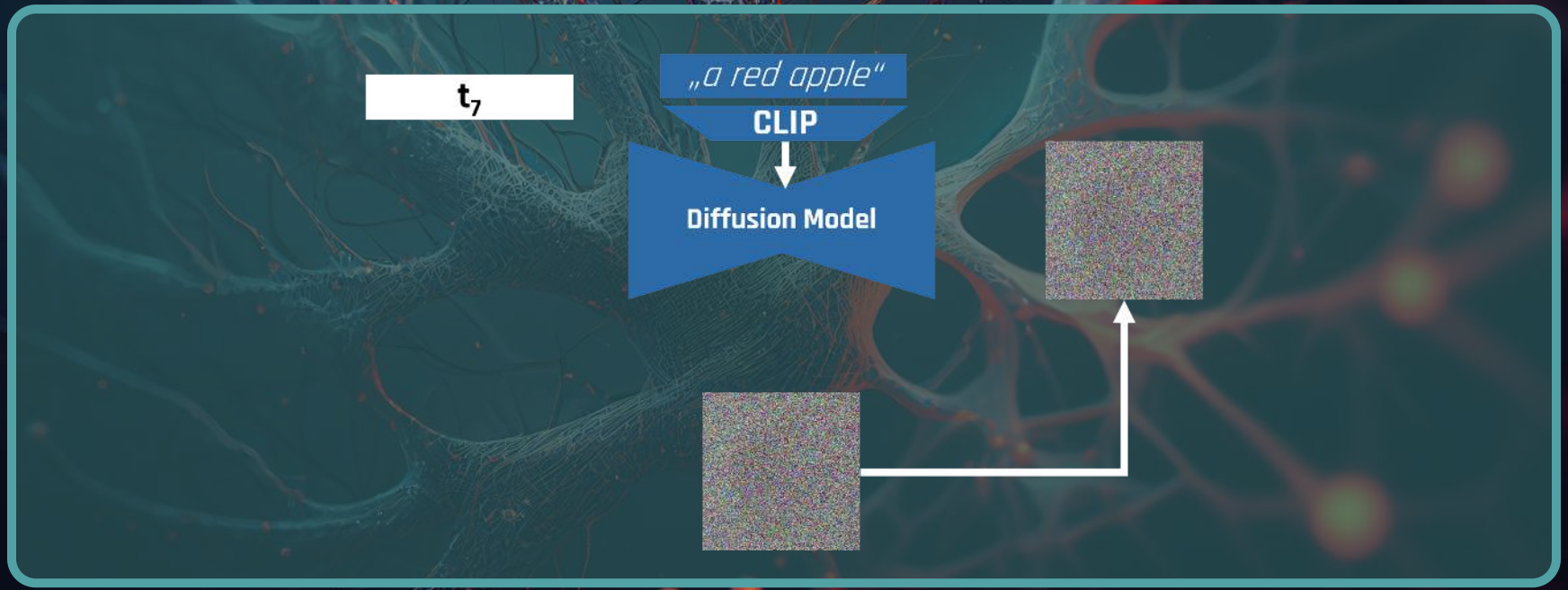
# Reverse Diffusion Process

## Substract Predicted Noise From the Input



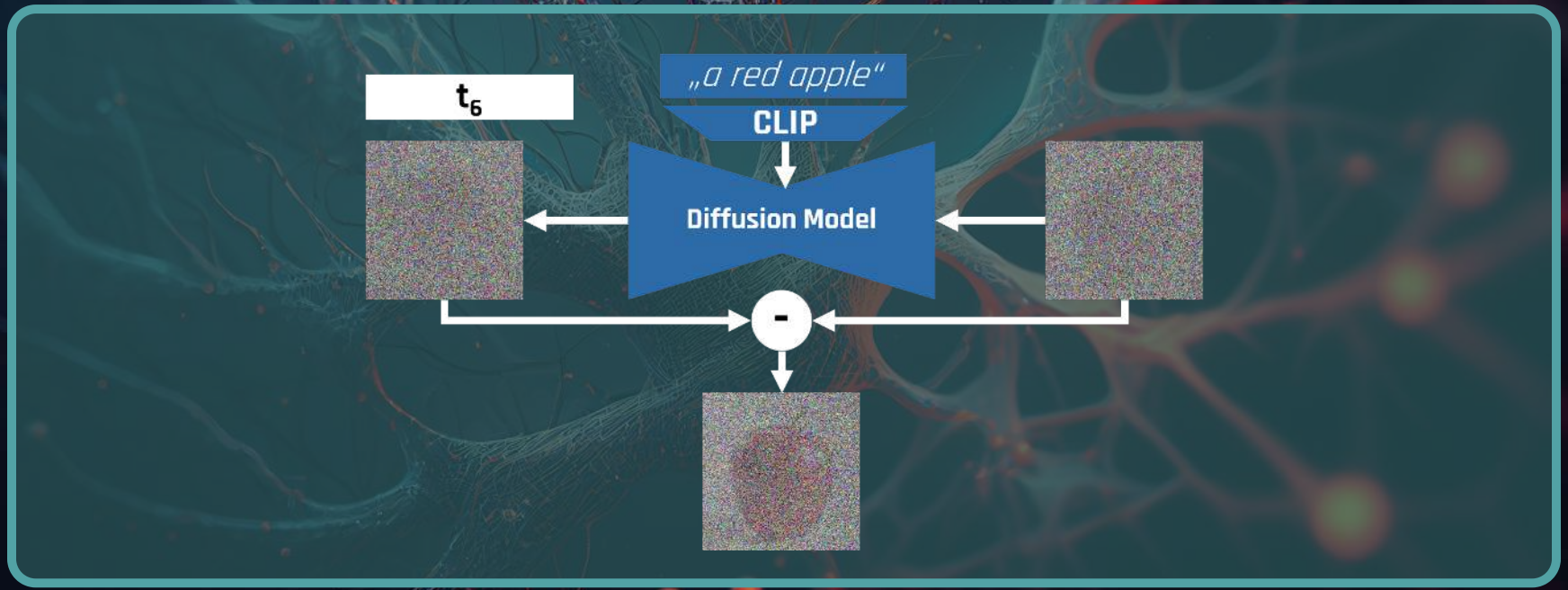
# Reverse Diffusion Process

## Output From the Timestamp t Becomes Input in Step of t-1



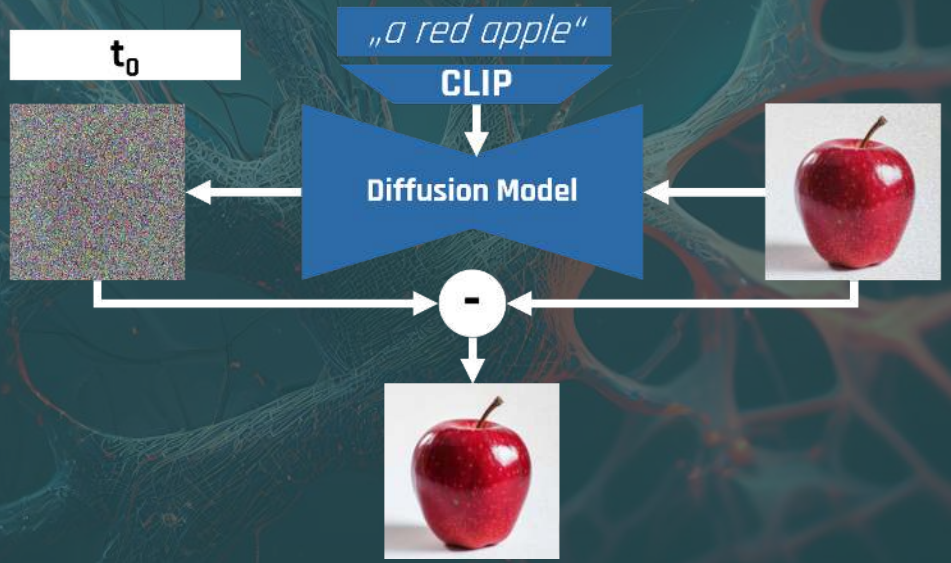
# Reverse Diffusion Process

## Continue With The Next Iteration



# Reverse Diffusion Process

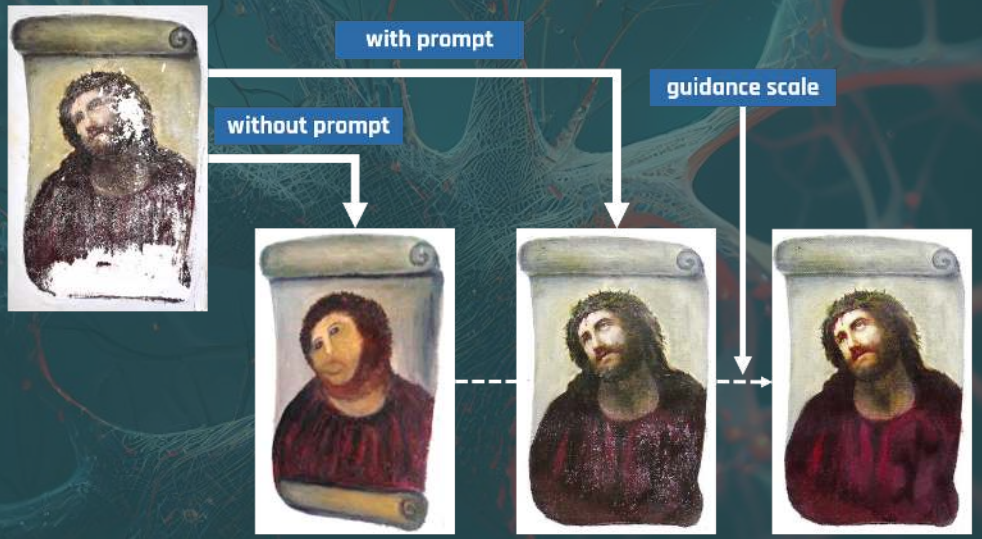
## Entire Noise Will be Substracted from Input Again



# Stable Diffusion Demo

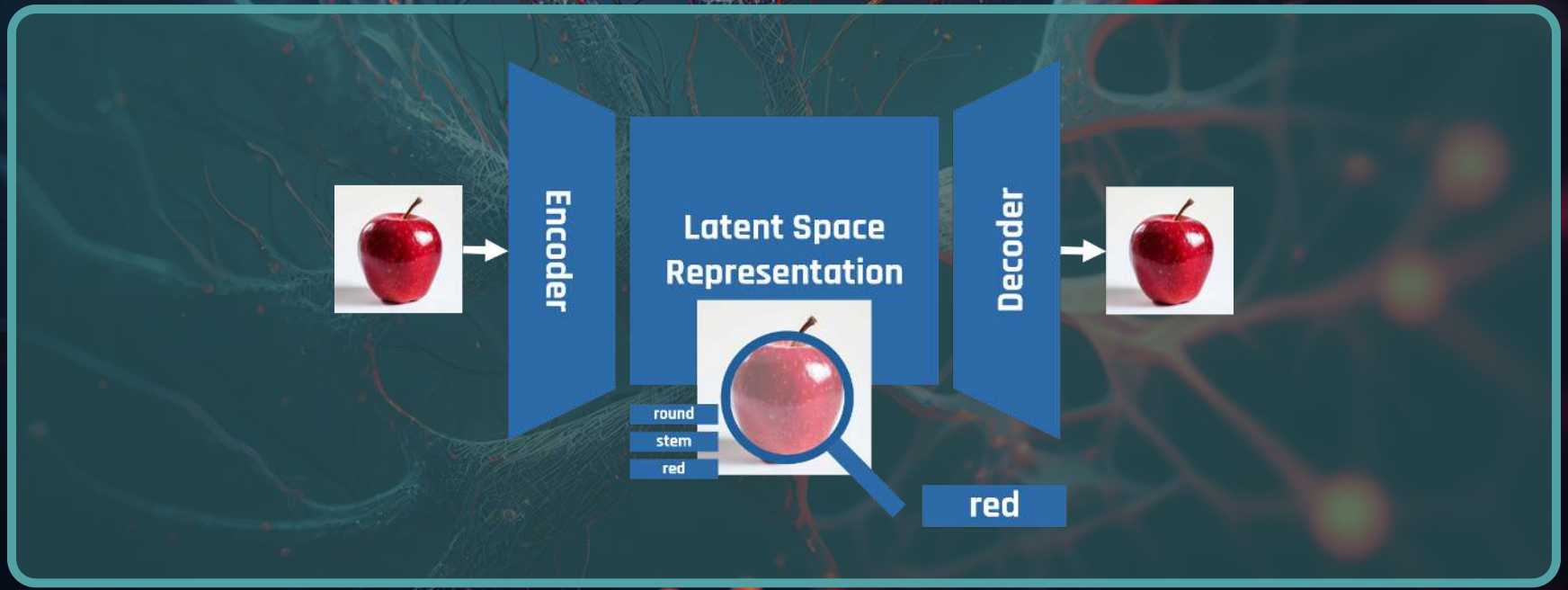
# Guidance

## Improving Artistic Control



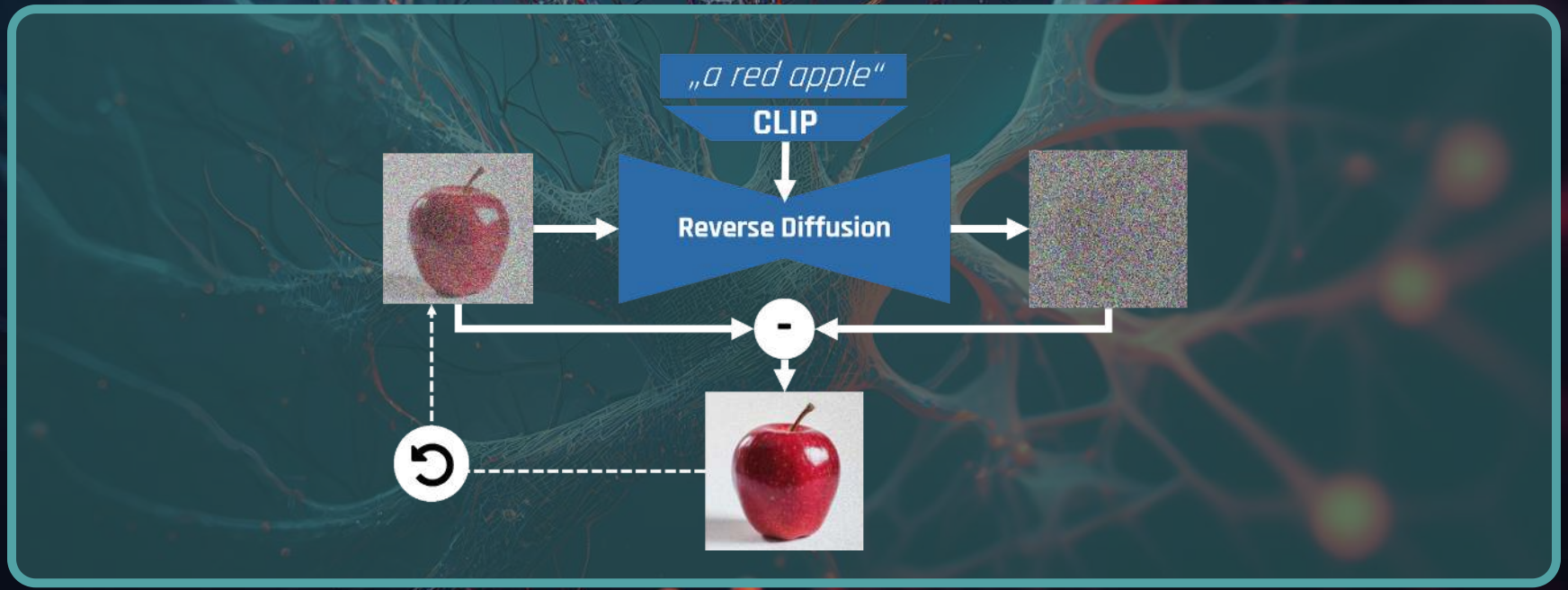
# Autoencoder

## Latent Space as a Compressed Representation of e.g. Images



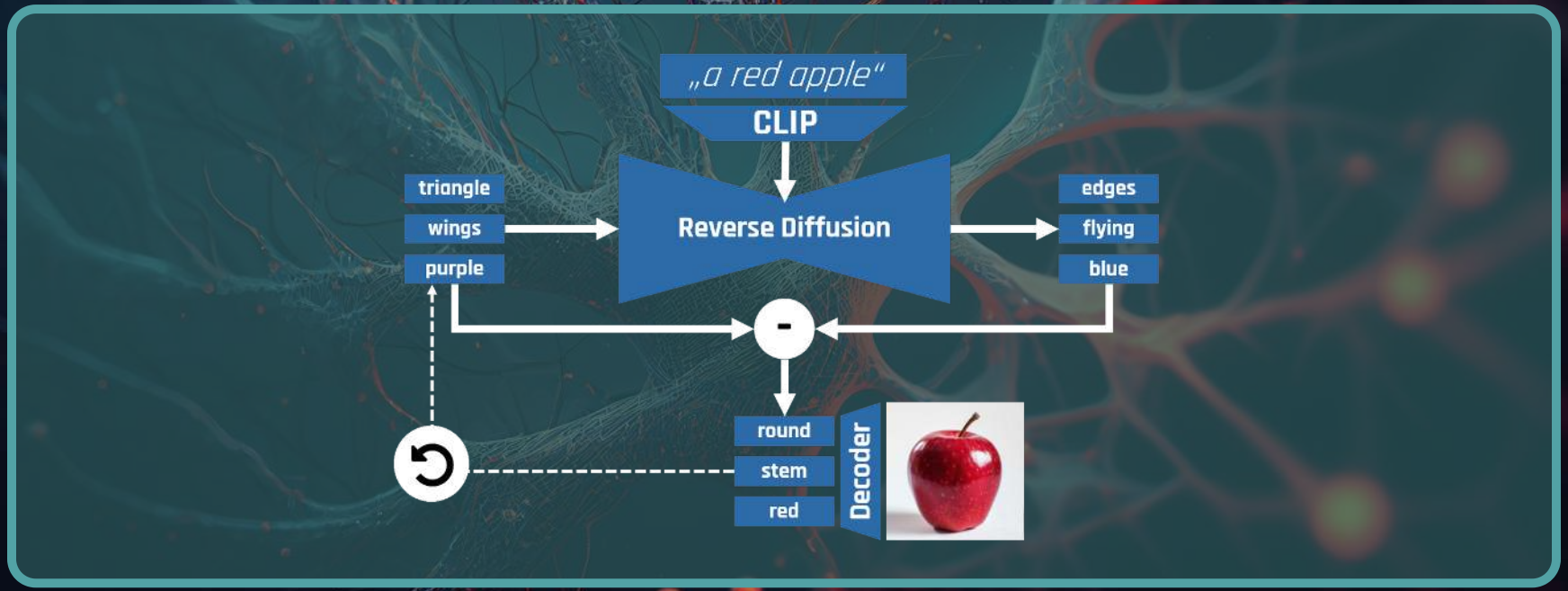
# Diffusion Models

## Pixel-Based Diffusion Process



# Diffusion Models

## Latent Diffusion Process

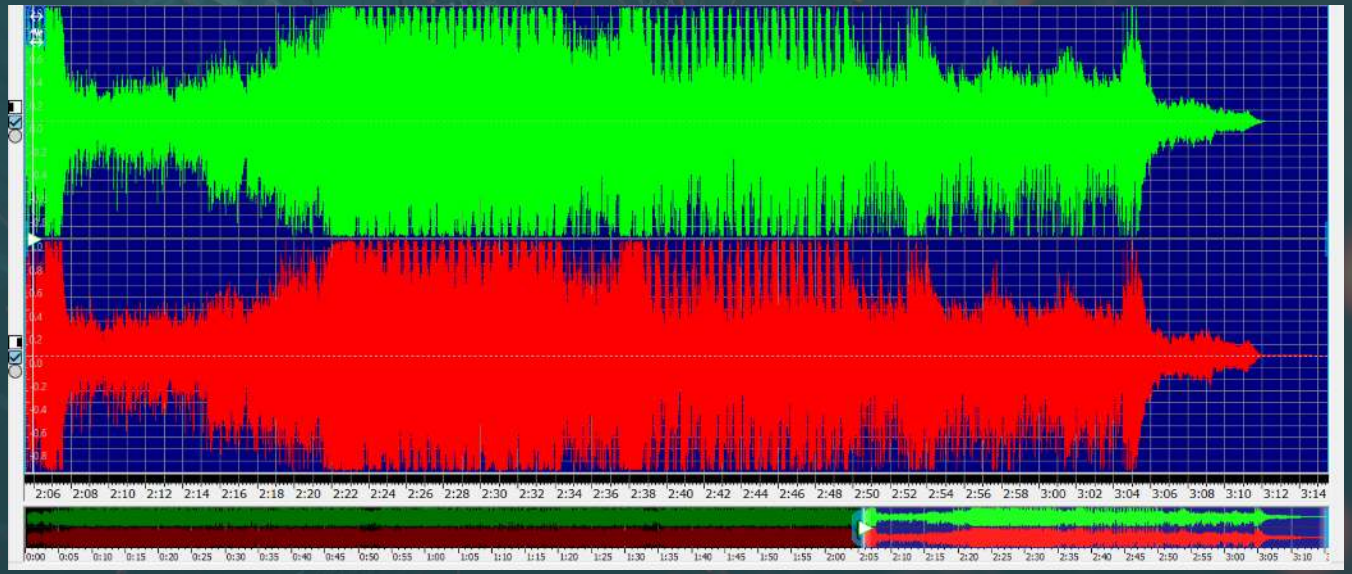


# Overview

- Motivation
- Text → Text
- Text → Image
- Text → Audio
- Text → Video
- Ethics

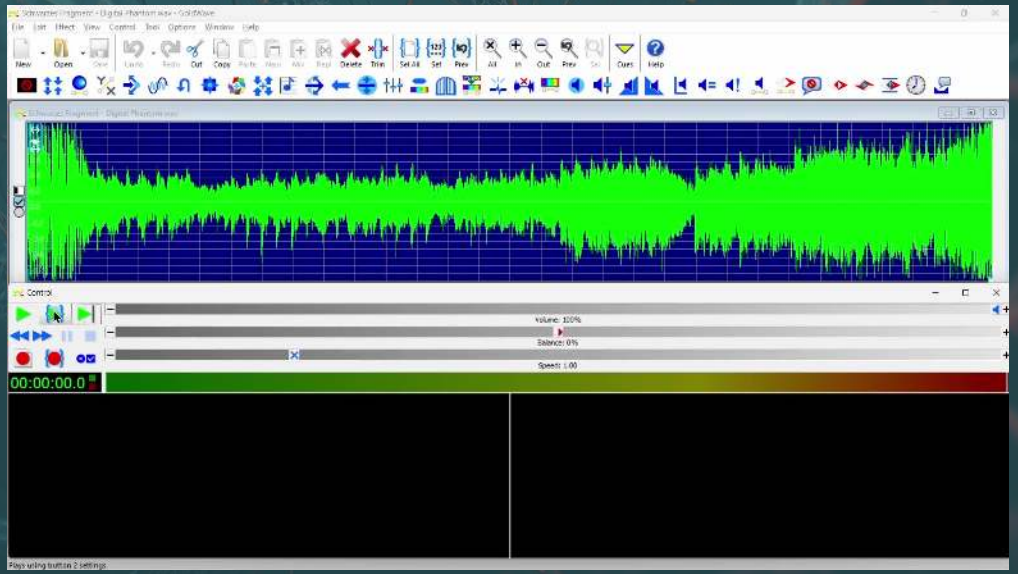
# Text → Audio

## Ways to Represent Audio Signals



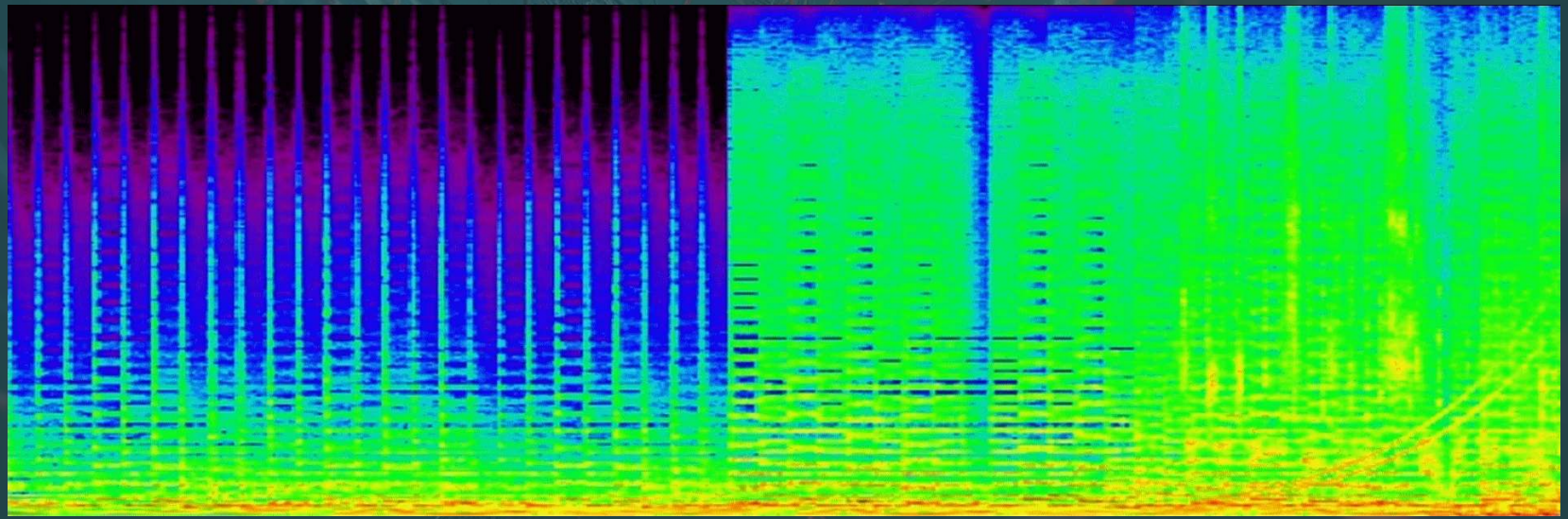
# Text → Audio

## Audio Can be Represented as Spectograms



# Text → Audio

What About Predicting the Continuation of the Spectrogram?



# Overview

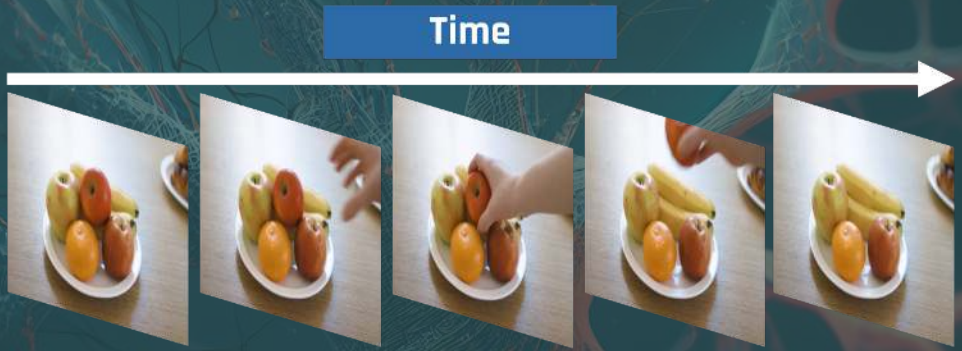
- Motivation
- Text → Text
- Text → Image
- Text → Audio
- Text → Video
- Ethics

# Text → Video

## Let's First Discuss a Typical Real Video

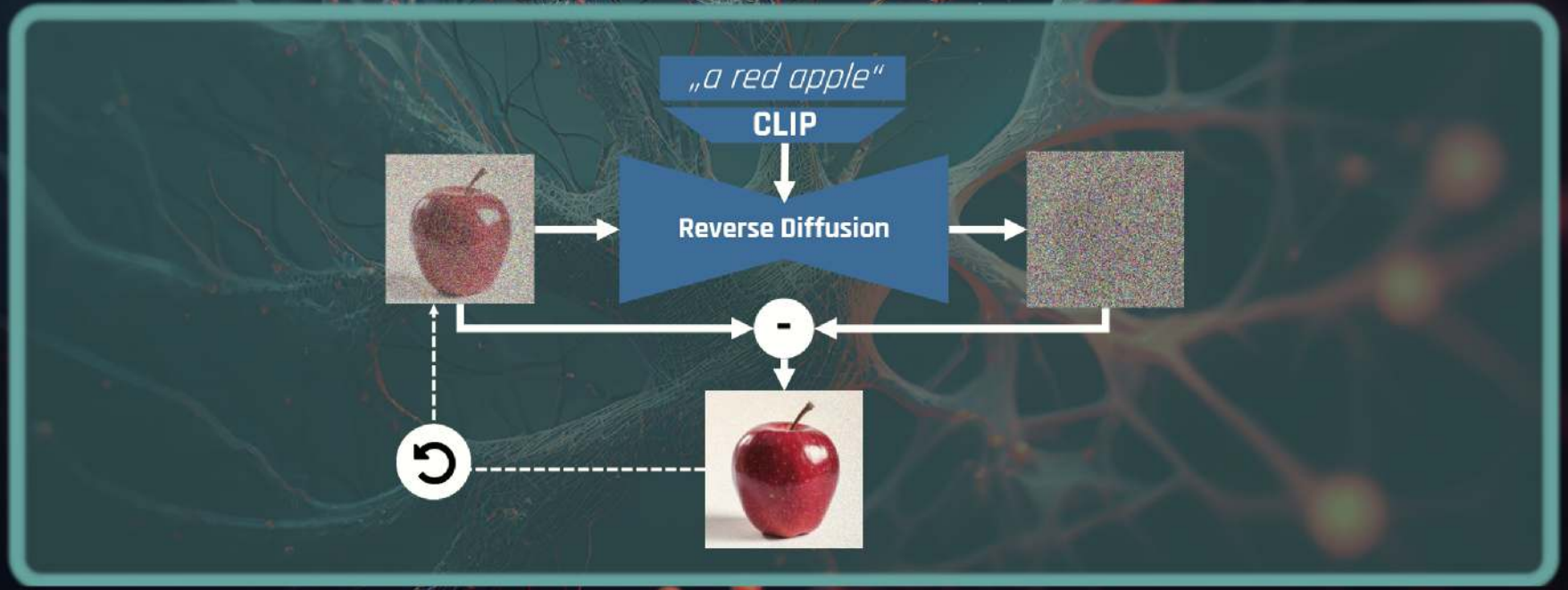
# Text → Video

A Video is a Sequence of Images (Frames)



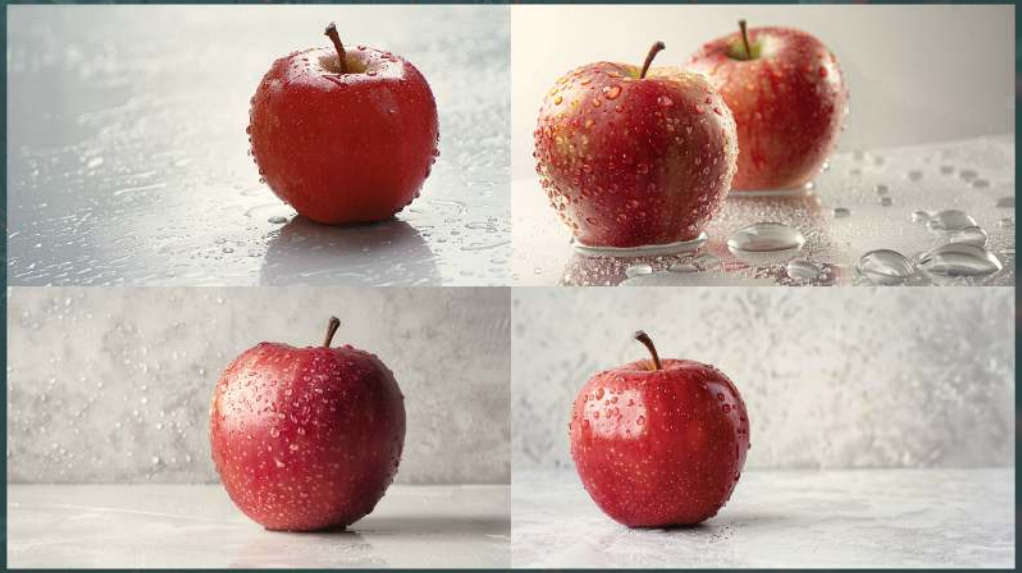
# Diffusion Models

## Pixel-Based Diffusion Process



# Diffusion Models

## Diffusion Model Outputs are not Coherent



# Text → Video

## Challenges: Temporal Coherence & Motion Dynamics

AI Generated (RunwayML Gen-2)

# Text → Video

## Challenges: Temporal Coherence & Motion Dynamics

AI Generated (RunwayML Gen-2)

# Text → Video

## Challenges: Logical Coherence & Physics

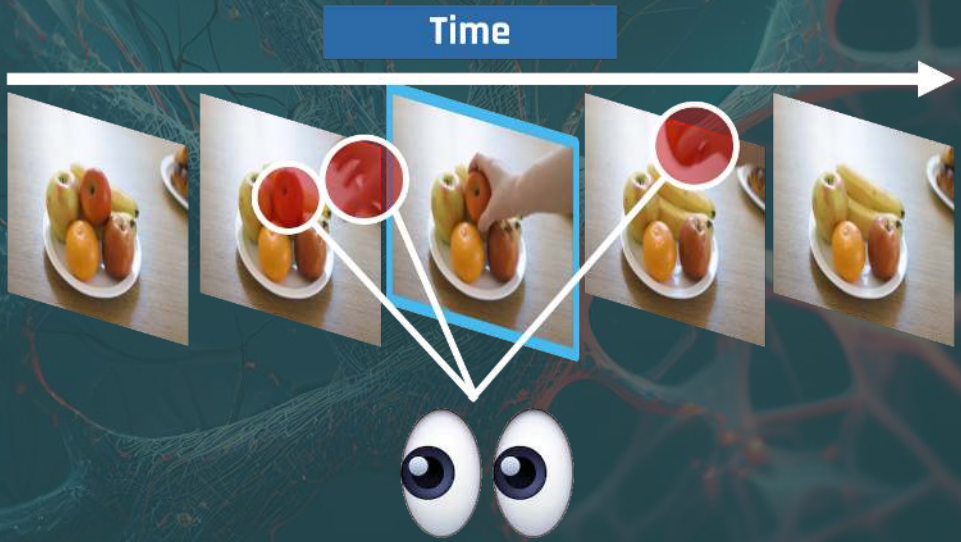
AI Generated (RunwayML Gen-2)

# Self-Attention

## Reminder: Attention in Text Transformers

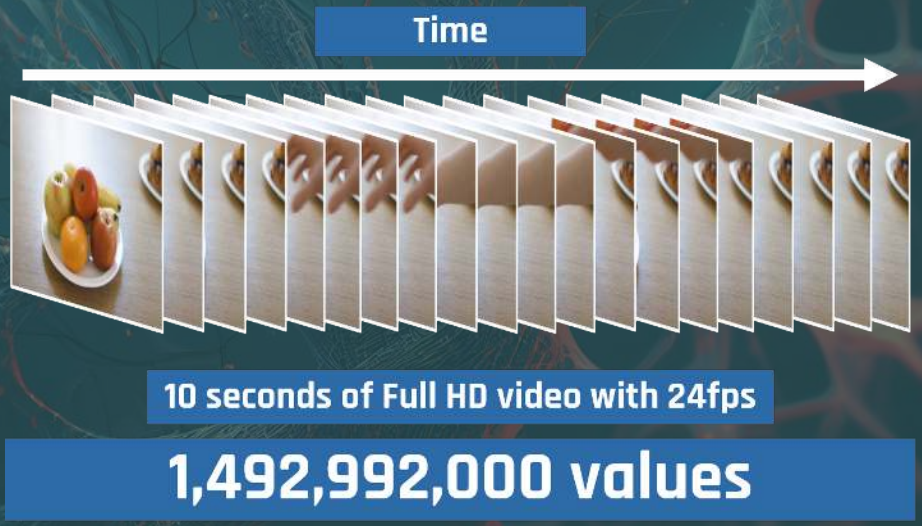
# Text → Video

## Attention Across Frames



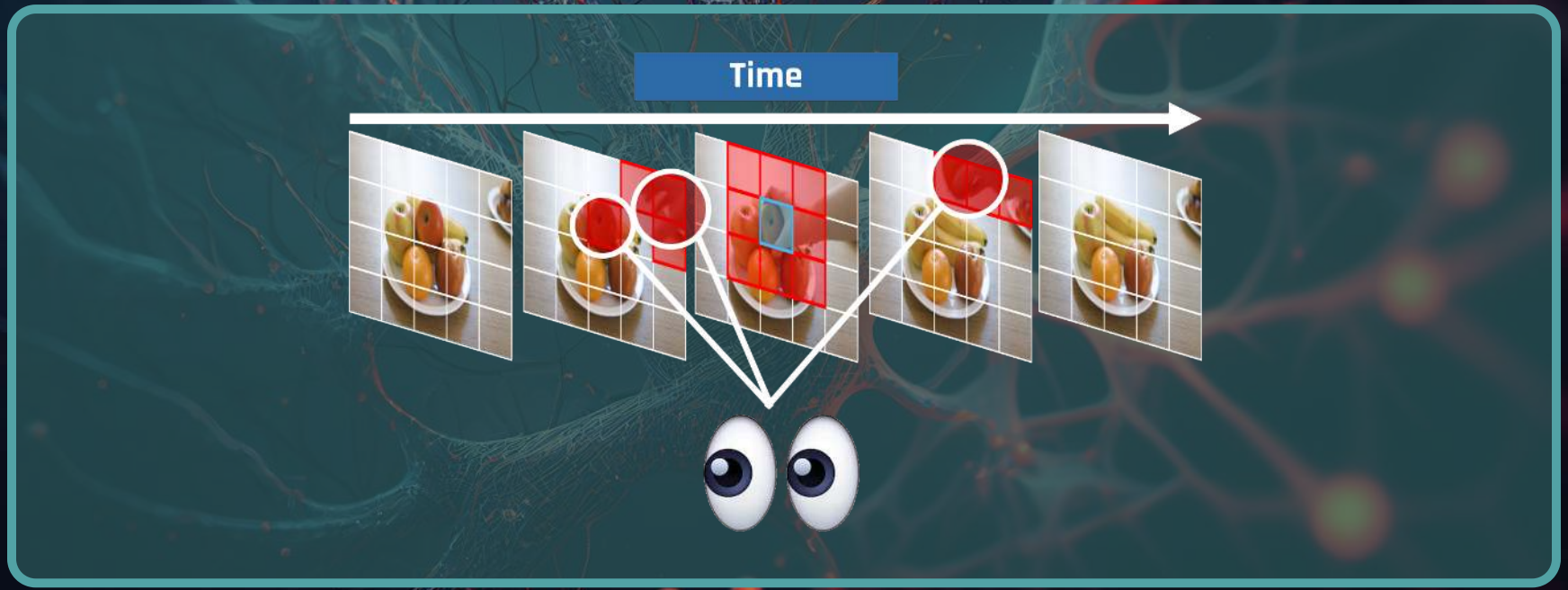
# Text → Video

## Challenges: Sheer Scale of Videos



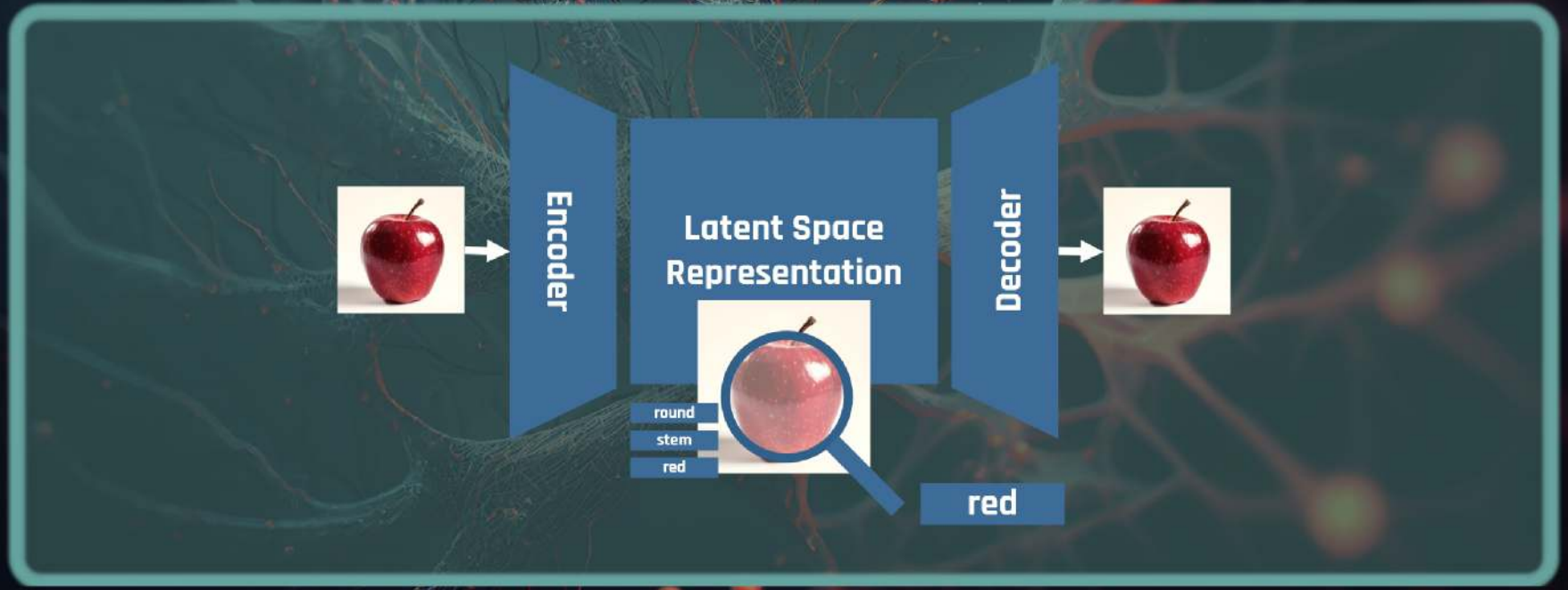
# Text → Video

## Solution: Divide and Conquer



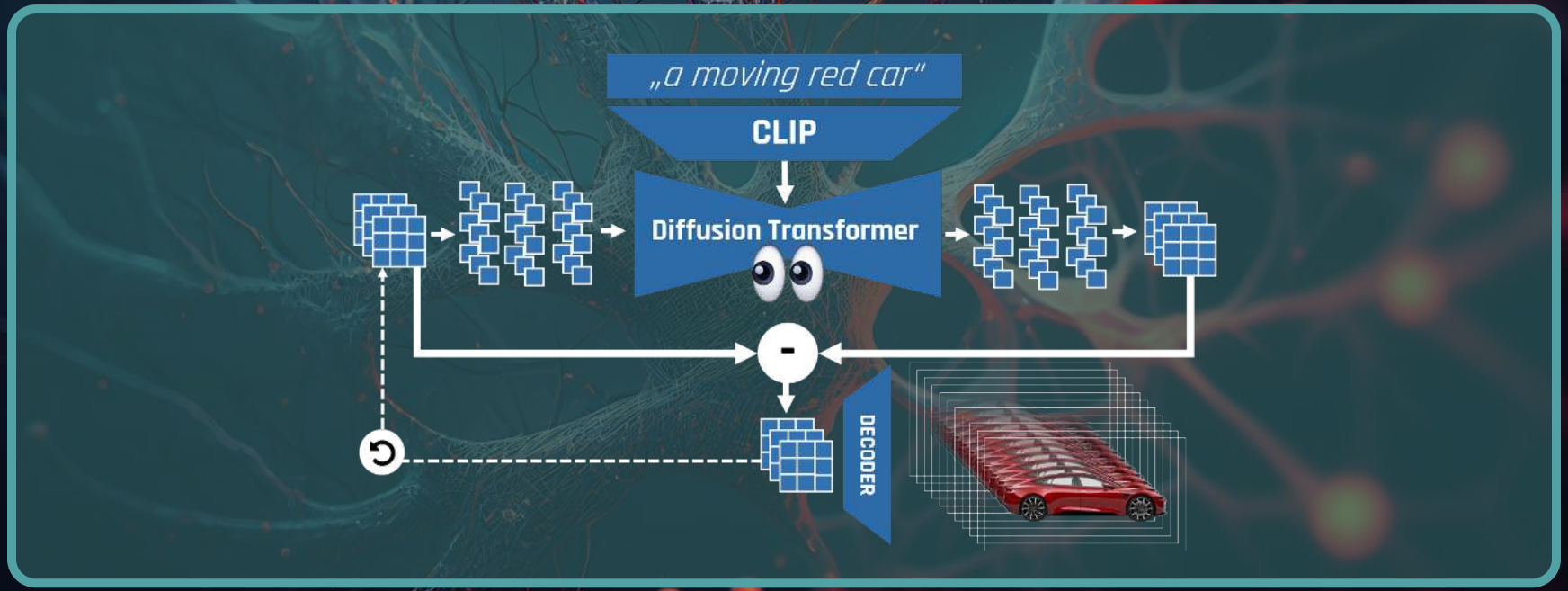
# Autoencoder

## Latent Space as a Compressed Representation of e.g. Images



# Text → Video

## Overview of Latent Video Diffusion Transformers



# OpenSora Demo

Current Step

Predicted Final Step

wide angle drone flyover footage of the TNG headquarters near Munich

# Text → Video

## Modern Open Weights Model Wan 2.1

# Text → Video

## Modern Open Weights Model Wan 2.1

# Text → Video

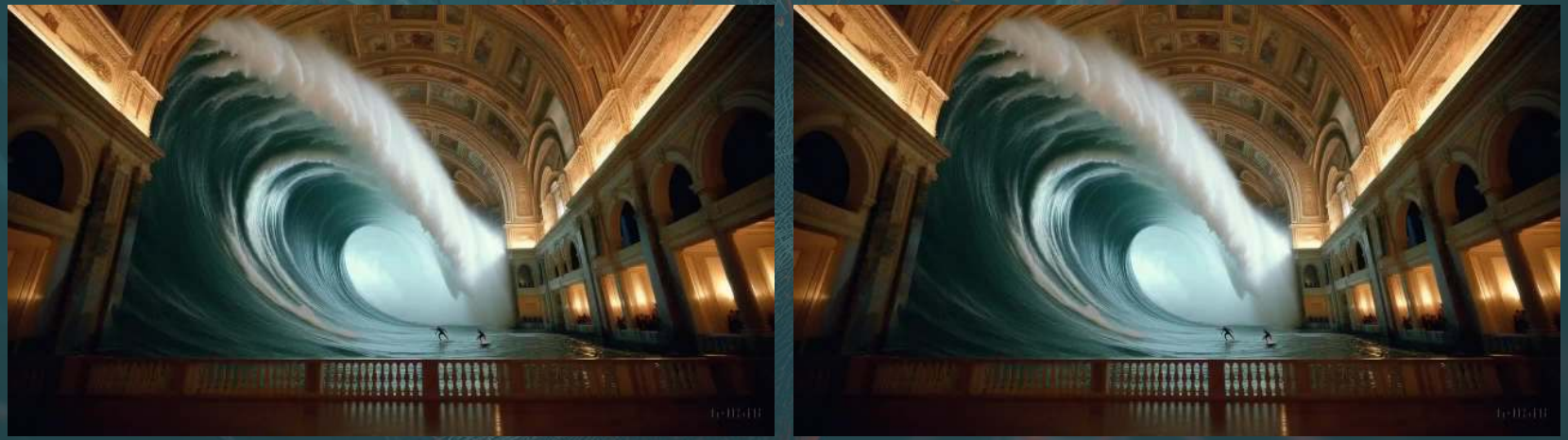
## Modern Open Weights Model Wan 2.1

# Video

## Applications of a Video Diffusion Transformer

# Text → Video

## Image to Video



AI Generated (OpenAI Sora Technical Report)

# Text → Video

## Applications of a Video Diffusion Transformer


# Text → Video

## Video Extension



AI Generated (OpenAI Sora Technical Report)

# Text → Video

## Applications of a Video Diffusion Transformer



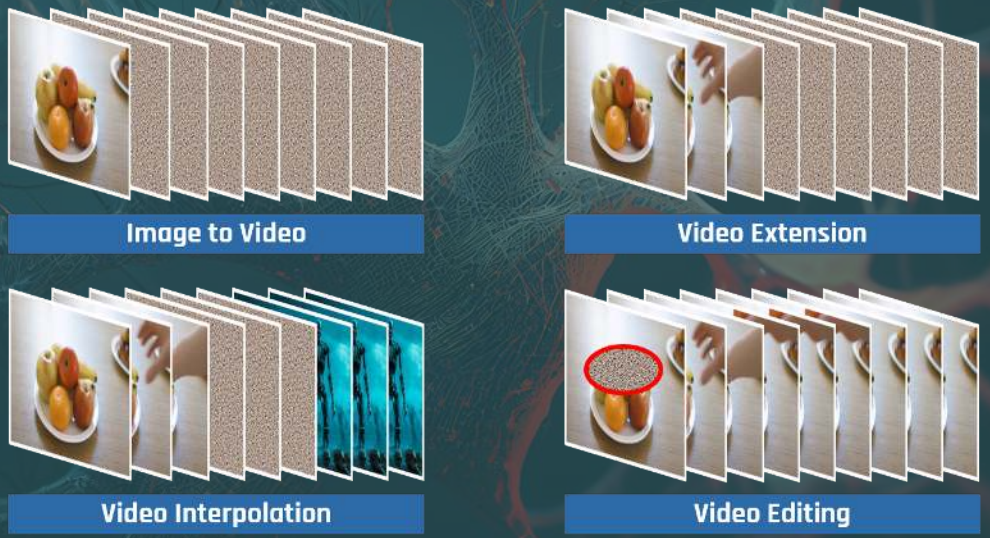
# Text → Video

## Applications of a Video Diffusion Transformer

AI Generated (OpenAI Sora Technical Report)

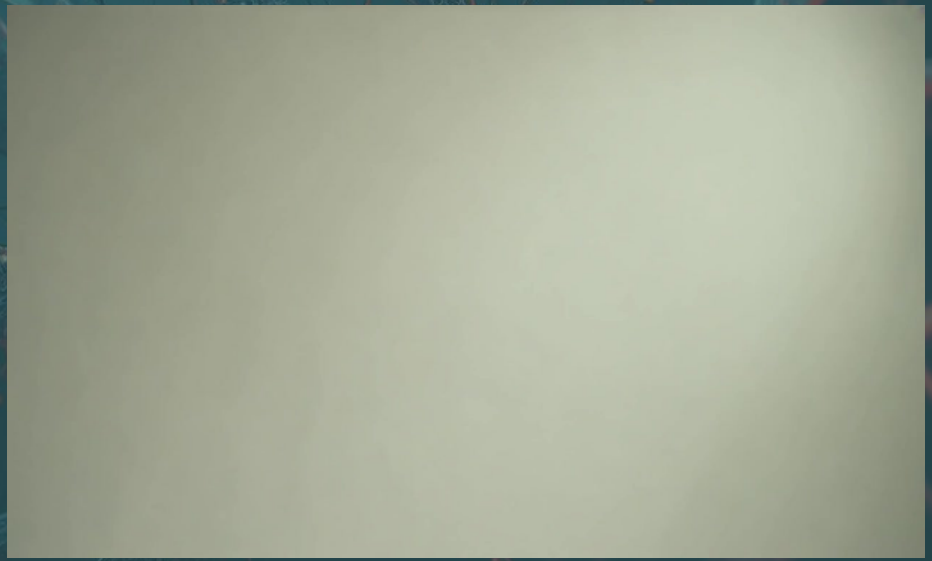
# Text → Video

## Applications of a Video Diffusion Transformer



# Limitations

## Understanding Gaps in AI's Grasp of Cause and Effect



AI Generated (RunwayML Gen-3 Alpha)

# Limitations

## Challenges in Simulating Complex Scene Physics

AI Generated (Open AI Sora Announcement)

# Limitations

## Tour de AI: How AI Imagines the Tour de France

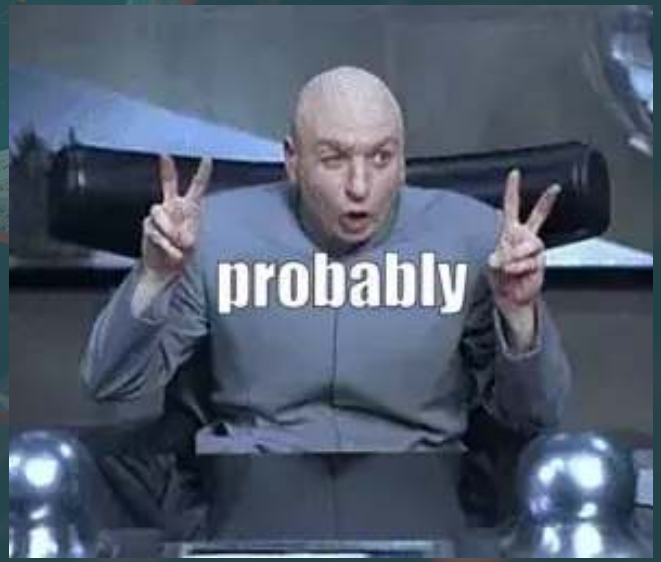
instagram.com/crit.meme

# Overview

- Motivation
- Text → Text
- Text → Image
- Text → Audio
- Text → Video
- Ethics

# Ethics

## Some AIs Have "Probably" Been Trained Without Owners' Consent




AI Generated (Midjourney V3)

# Ethics

## Which Training Data Was Used?

# Ethics

## Legal Challenges

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

THE NEW YORK TIMES COMPANY

Plaintiff,

v.

MICROSOFT CORPORATION, OPENAI, INC., OPENAI LP, OPENAI GP, LLC, OPENAI, LLC, OPENAI OPCO LLC, OPENAI GLOBAL LLC, OAI CORPORATION, LLC, and OPENAI HOLDINGS, LLC,

Defendants.

Civil Action No. _____

**COMPLAINT**

**JURY TRIAL DEMANDED**

# Ethics

## Legal Challenges

The Guardian

**Mark Zuckerberg**

### Zuckerberg approved Meta's use of 'pirated' books to train AI models, authors claim

Sarah Silverman and others file court case claiming CEO approved use of dataset despite warnings

Business live – latest updates

# Ethics

## Legal Challenges

# Ethics

## Out of Scope Topics

- Labour Rights in Labelling and after Job Displacement
- Energy Usage, Water Consumption, and Pollution
- Training Data Oligopoly
- Disinformation
- Right of Personality

# Ethics

## Will Smith Eating Spaghetti

March 2023

February 2024

reddit.com/user/chaindrop & instagram.com/willsmith

# Thank You!


**Thomas Endres**

**Managing Partner**

Oracle® JavaOne Rockstar
Intel® Black Belt Software Developer
Intel® Software Innovator
Intel® Top Innovator

**Fabian Pottbäcker**

**Software Consultant**

ML Reverse Engineer
Space Nerd
(former) Competitive Programmer