AI for Science

Hyungon Ryu | sr. Solutions Architect | NVIDIA AI Tech. Center (NVAITC)

2024 중력파 겨울학교 | 대전 KT인재개발원

AI Art at Christie’s Sells for $432K, Oct. 2018

OpenAI DALL-E 2 : The World's Smartest Artificial Intelligence Just Made Its First Magazine Cover, June 2022

MidJourney AI-generated artwork wins 1st prize at Colorado State Fair, Sep 2022

Introduction

Introduction

Brief History of Generative AI

Bolzmann machines

AE/VAEs

GAN

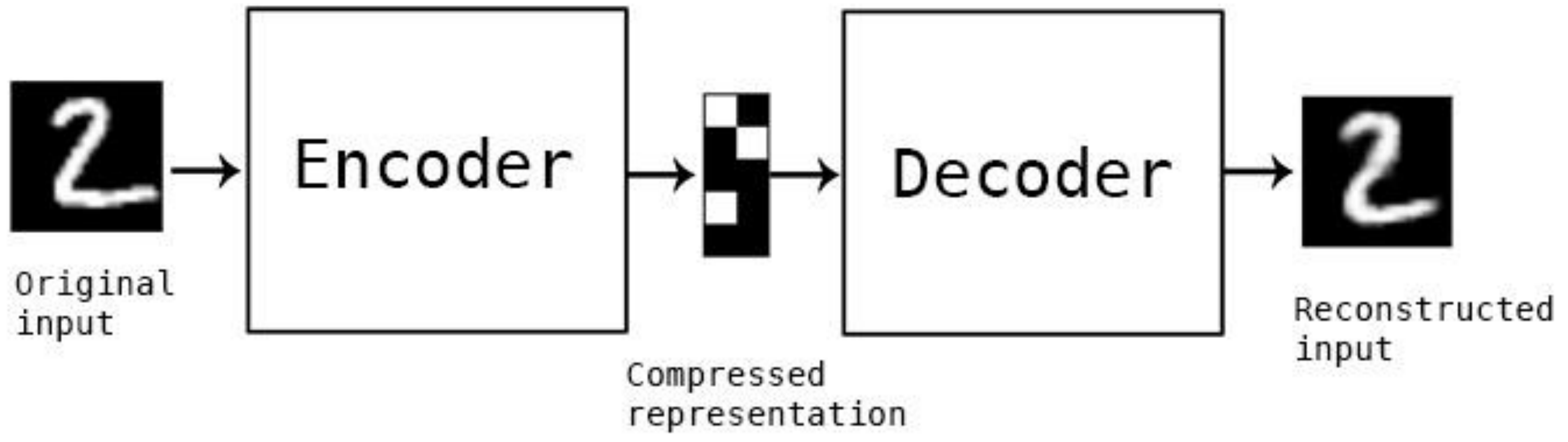
Diffusion Model

Introduction

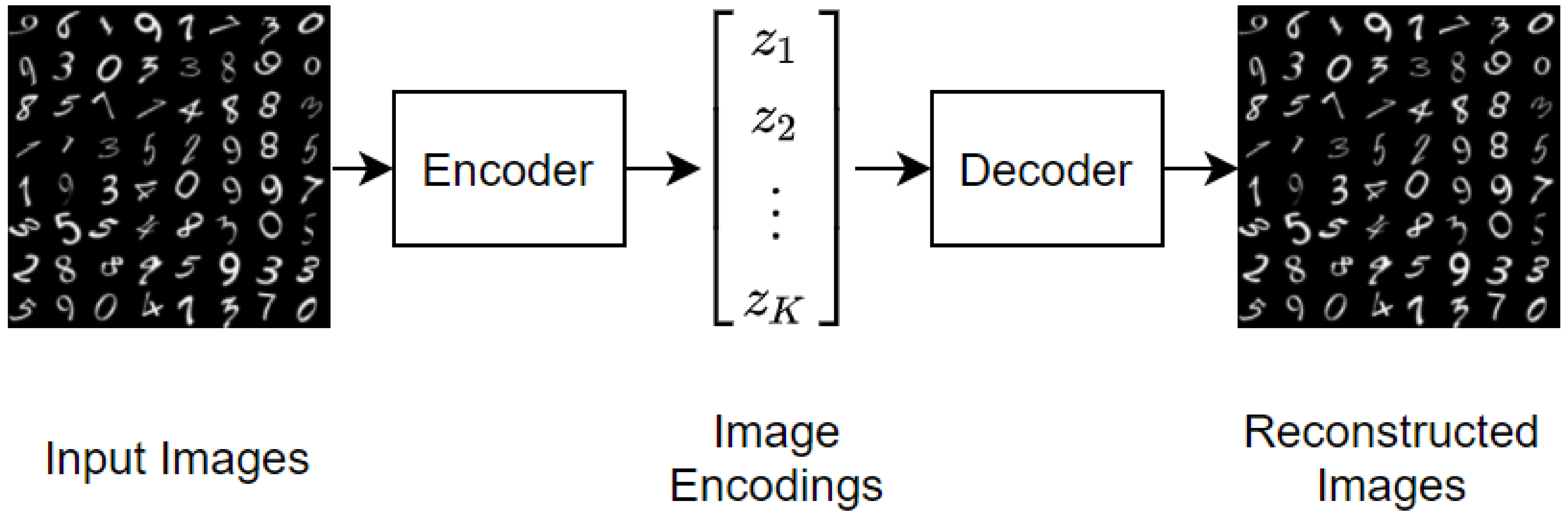
Fundamental Concepts and Techniques in Generative AI

Autoencoders and Variants(VAEs)

Autoencoder

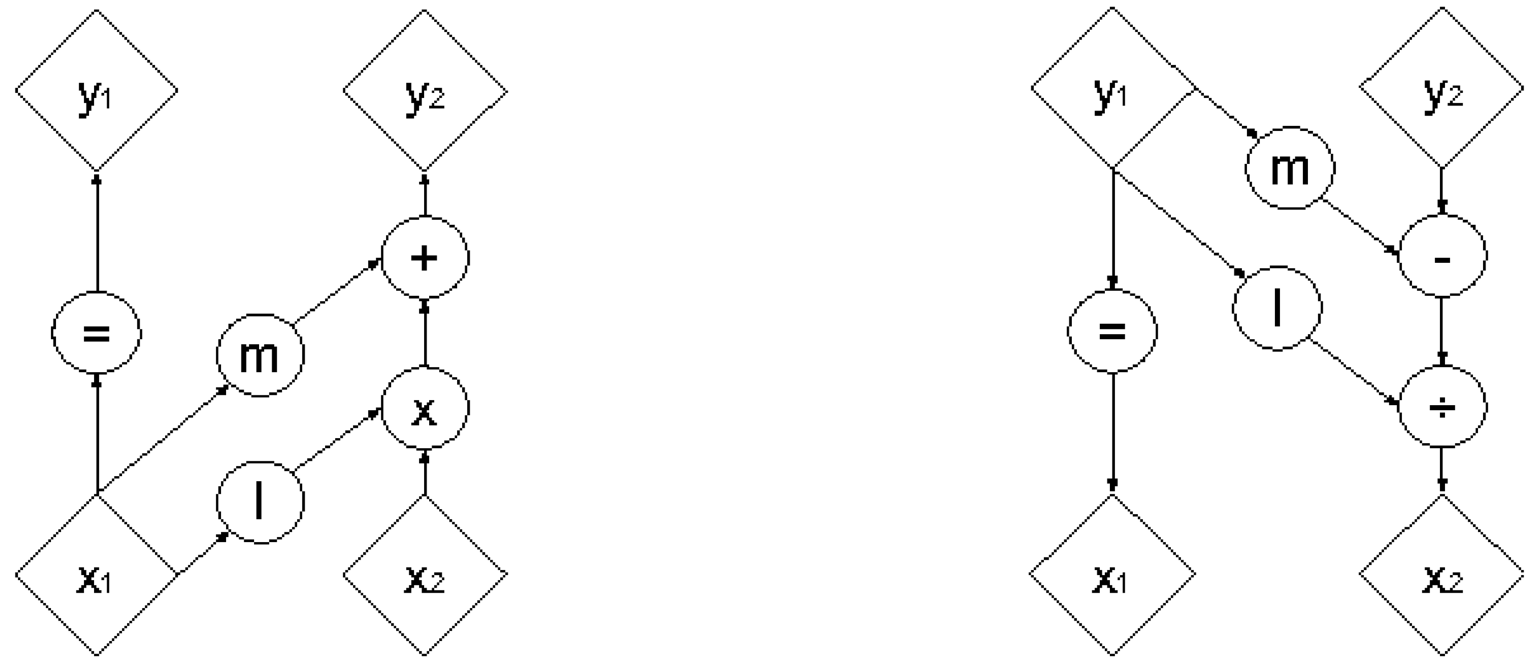


Variational Autoencoder



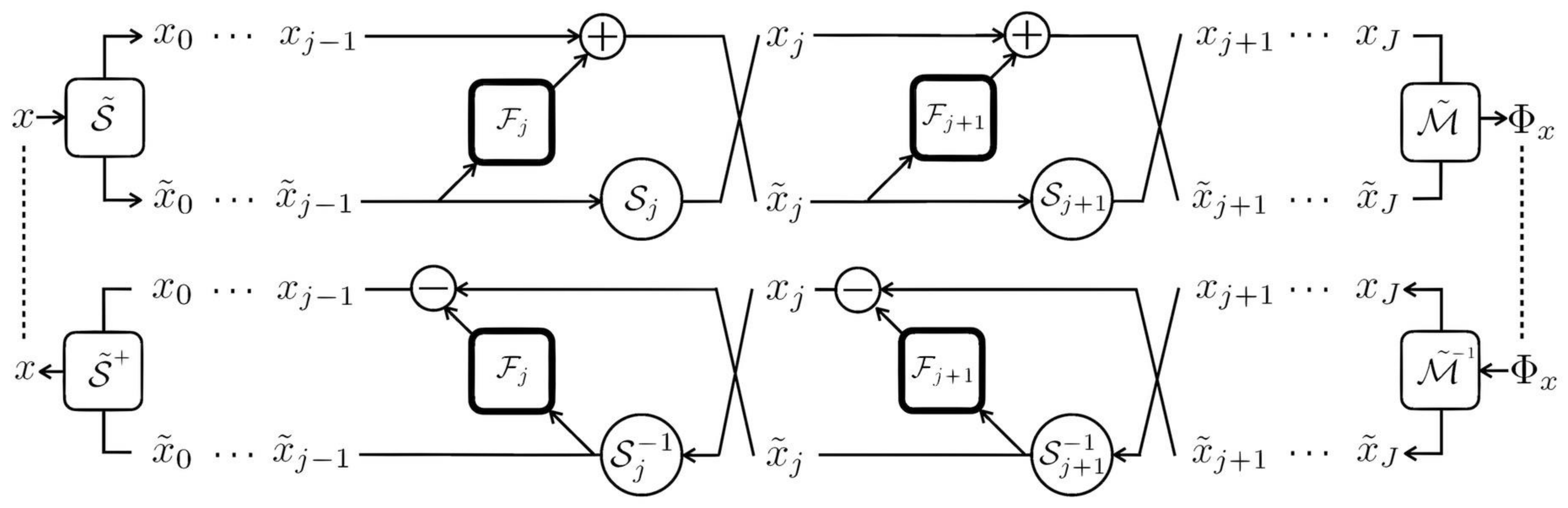
Normalization Flows

RealNVP

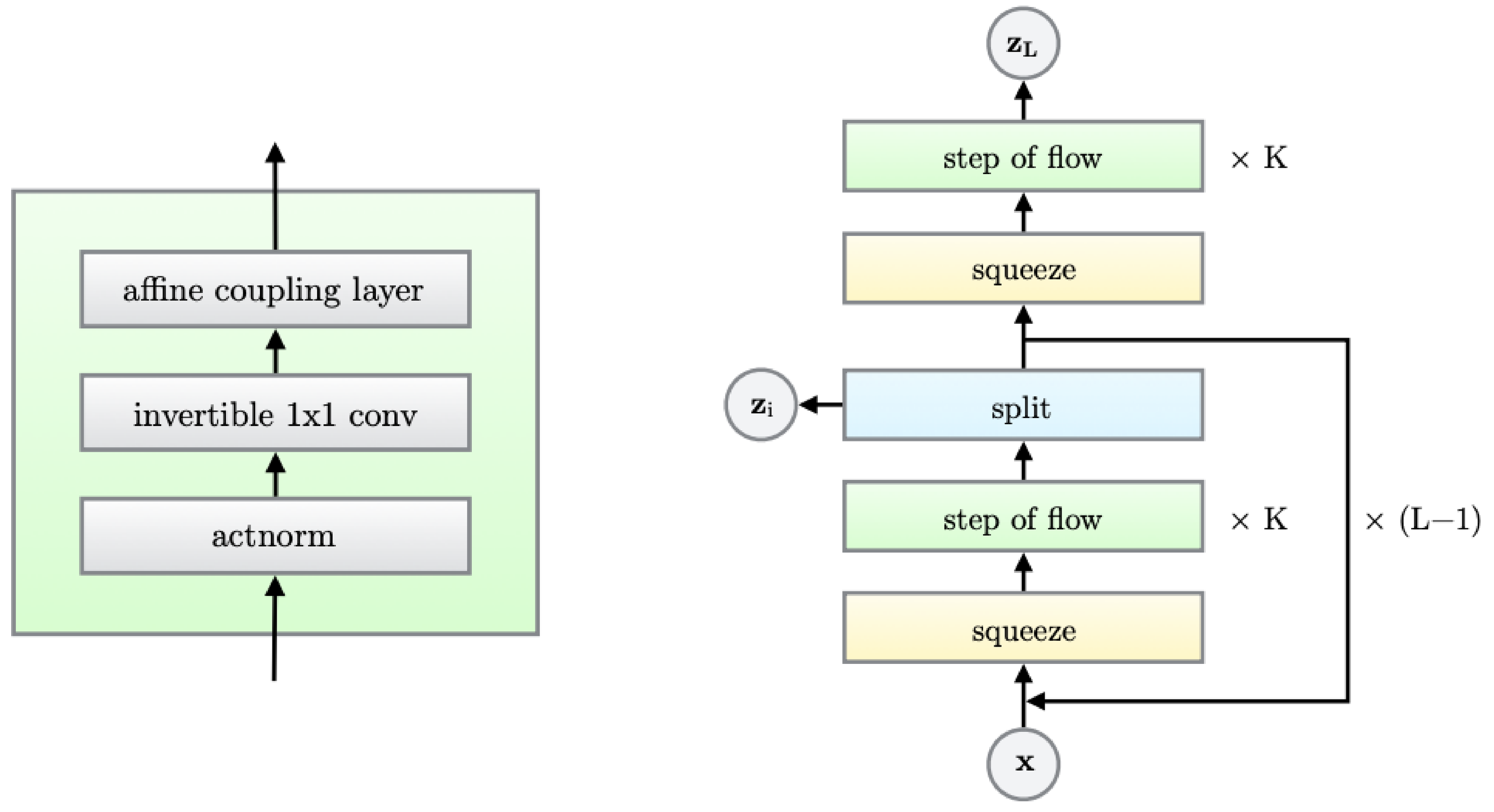


(a) Forward propagation (b) Inverse propagation

i-RevNet(Invertible ResNet)



GLOW

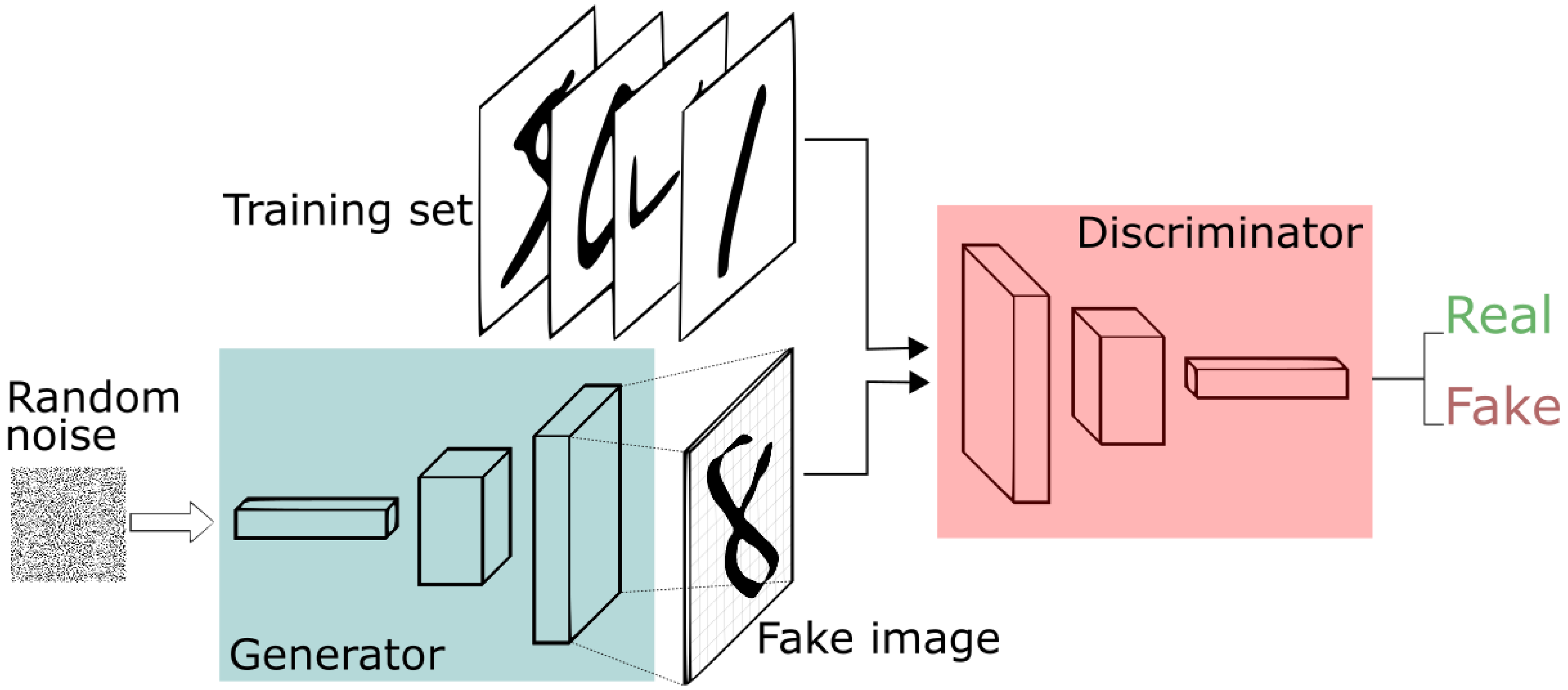


(a) One step of our flow. (b) Multi-scale architecture (Dinh et al., 2016).

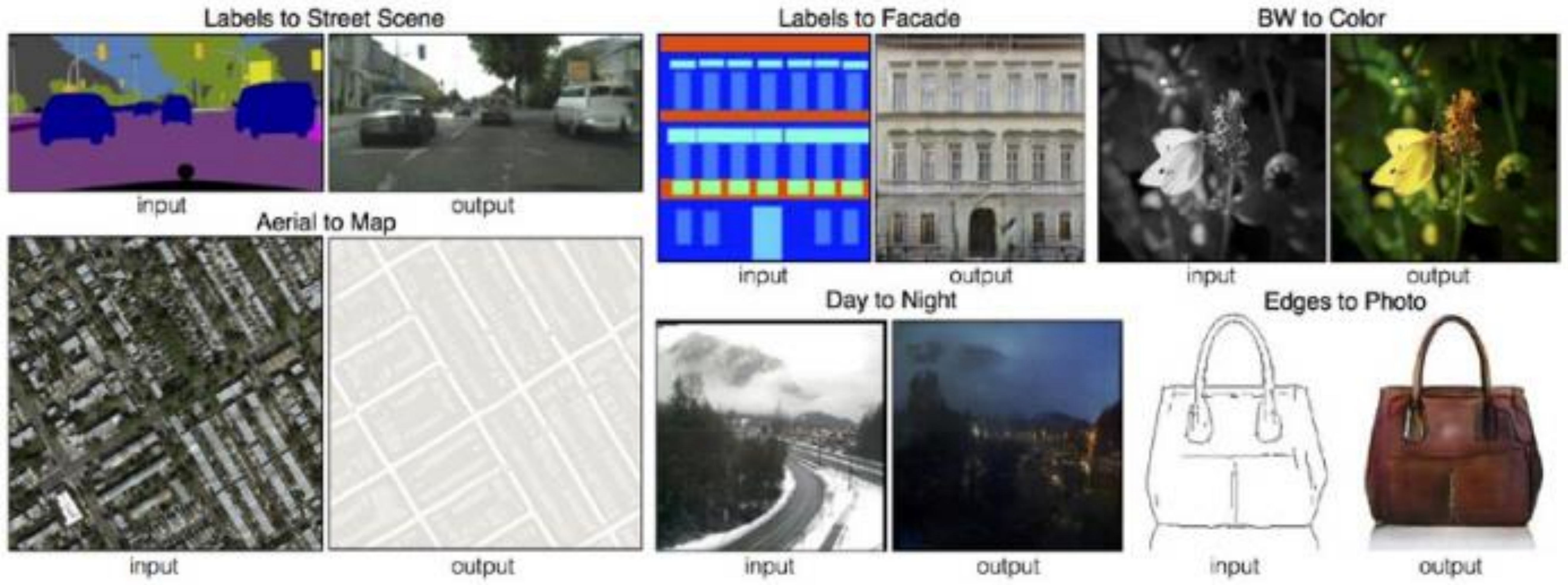
Figure 2: We propose a generative flow where each step (left) consists of an *actnorm* step, followed by an invertible 1×1 convolution, followed by an affine transformation (Dinh et al., 2014). This flow is combined with a multi-scale architecture (right). See Section 3 and Table 1.

Figure 1: Synthetic celebrities sampled from our model; see Section 3 for architecture and method, and Section 5 for more results.

Generative Adversarial Networks (GANs) and Variants

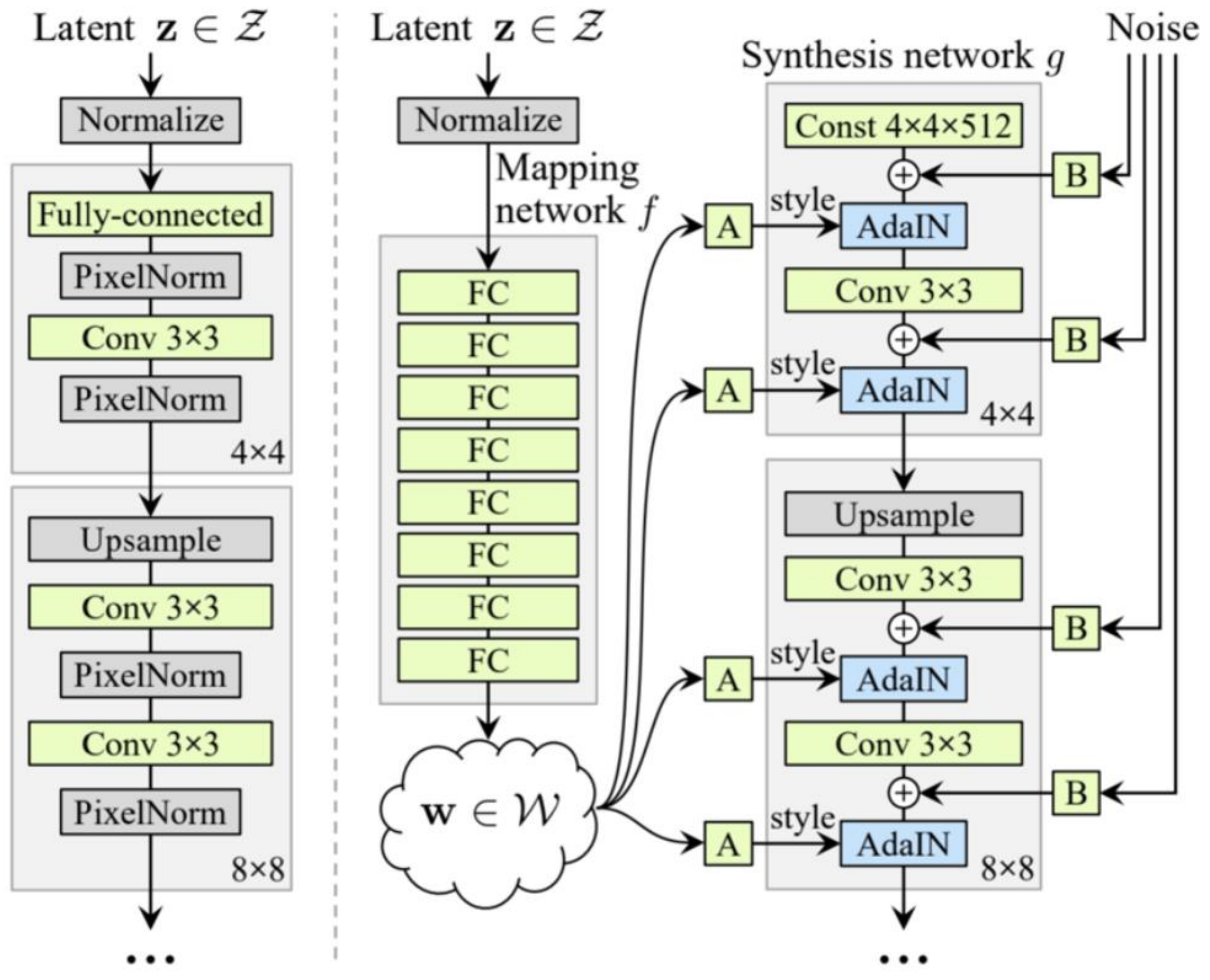


Pix2Pix



StyleGAN

Figure 2. Uncurated set of images produced by our style-based generator (config F) with the FFHQ dataset. Here we used a variation of the truncation trick [5, 29] with $\psi = 0.7$ for resolutions $4^2 - 32^2$. Please see the accompanying video for more results.

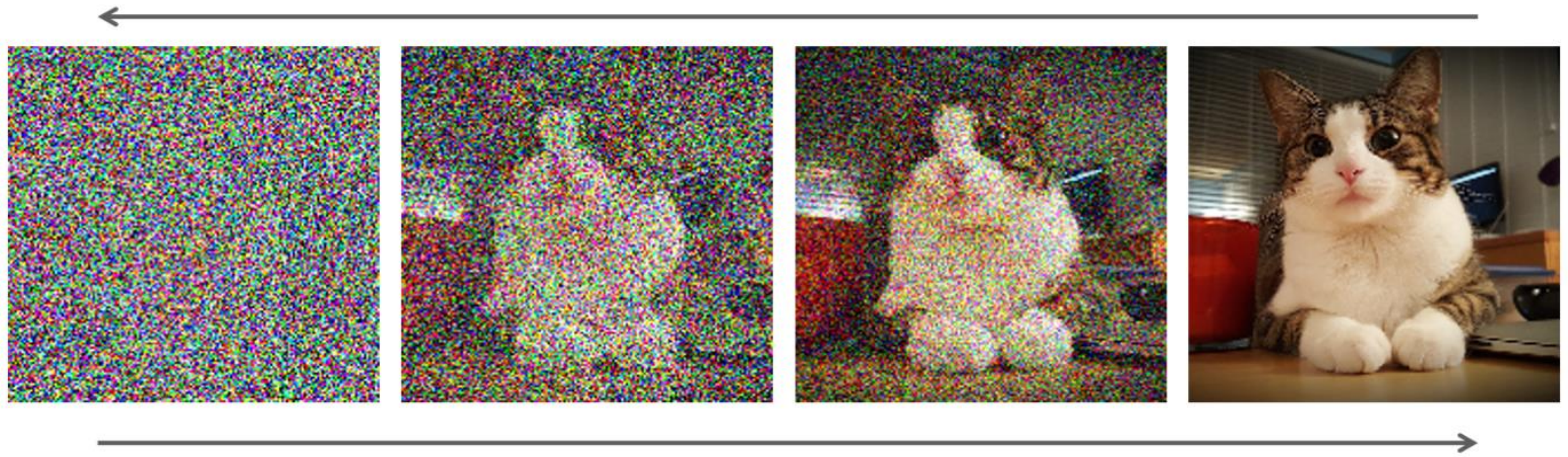


(a) Traditional

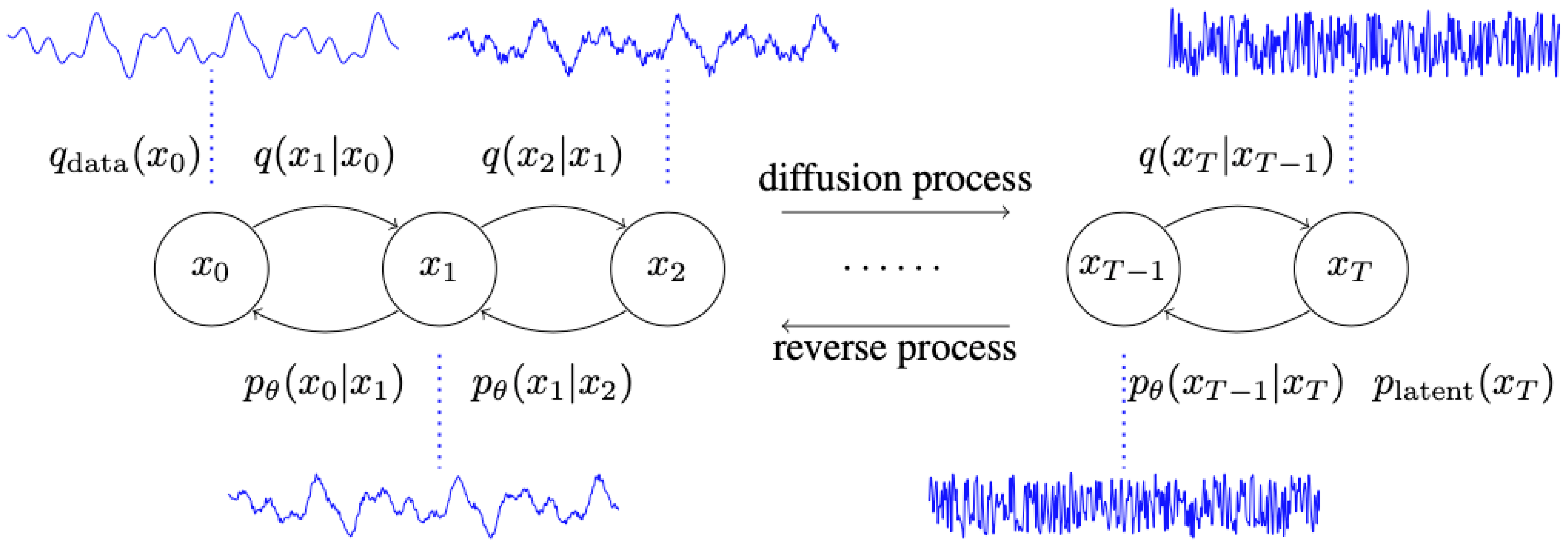
(b) Style-based generator

Diffusion Models

Diffusion Process

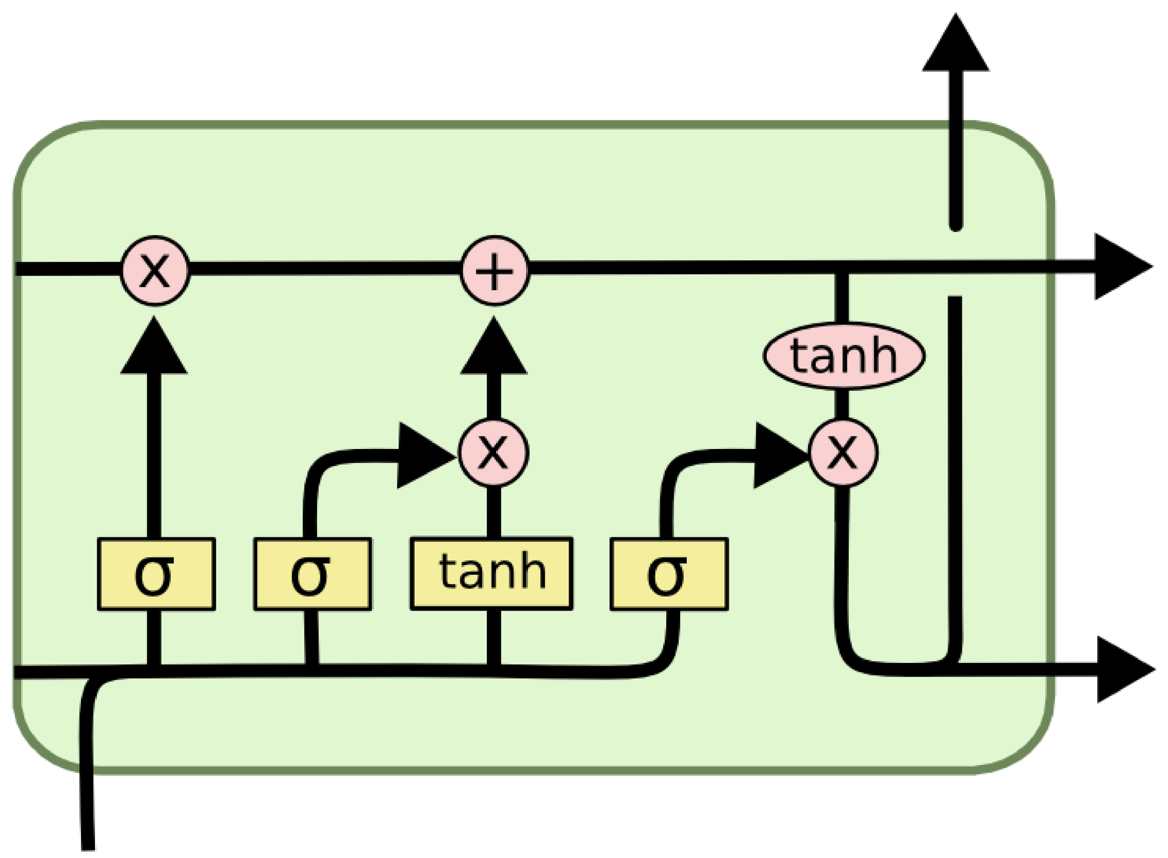


DiffWave

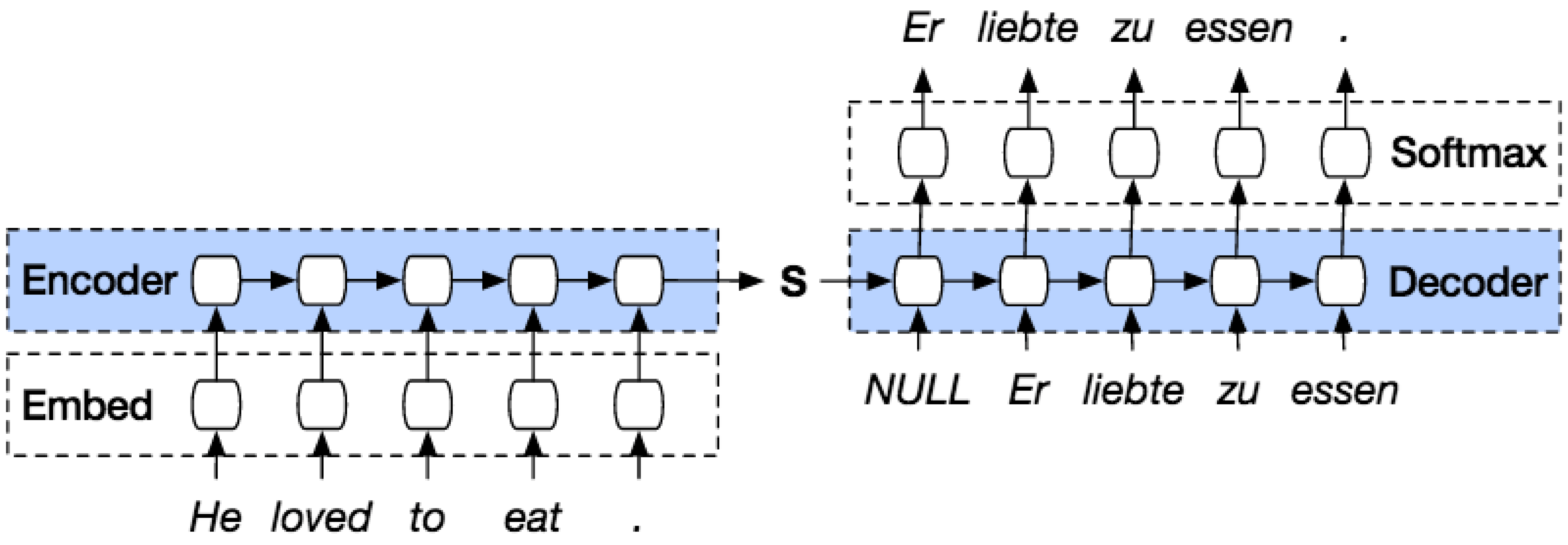


Seqence-to-Seqence models

LSTM

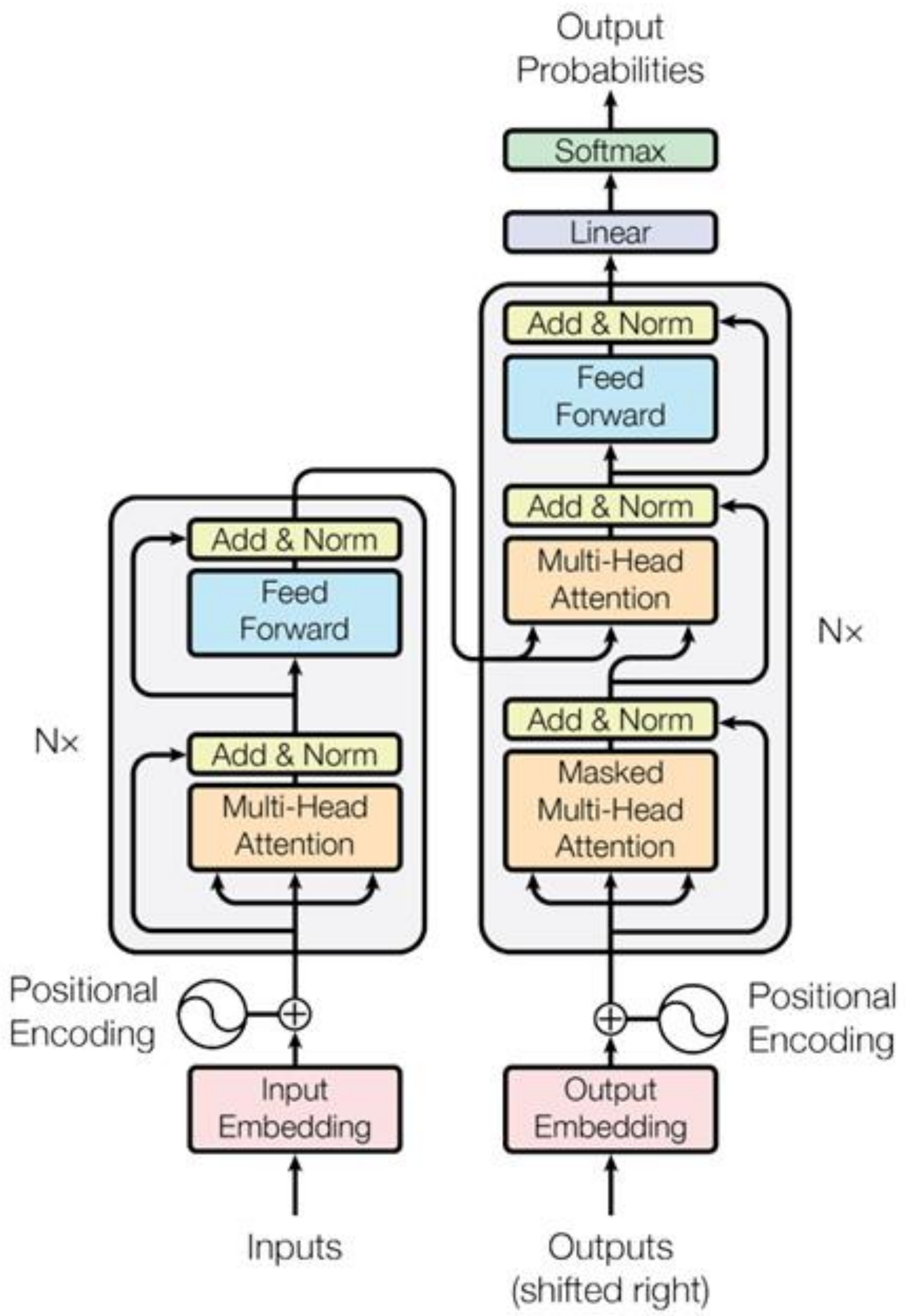


NMT(Neural Machine Translation)

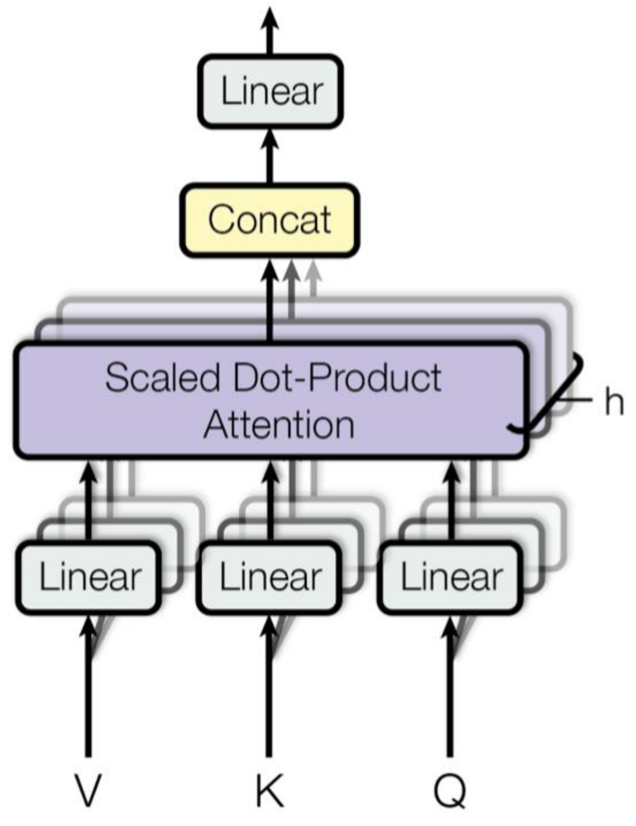


Transformers and Large Language Models

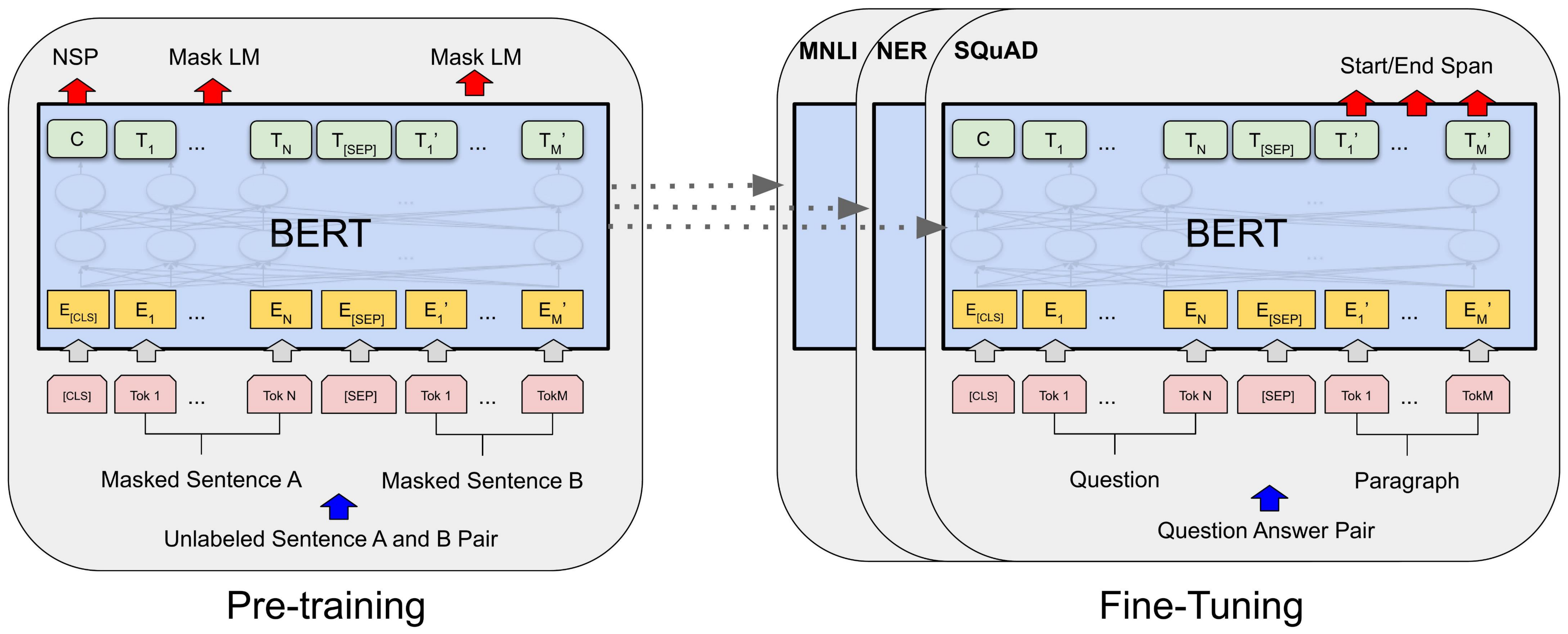
Transformer



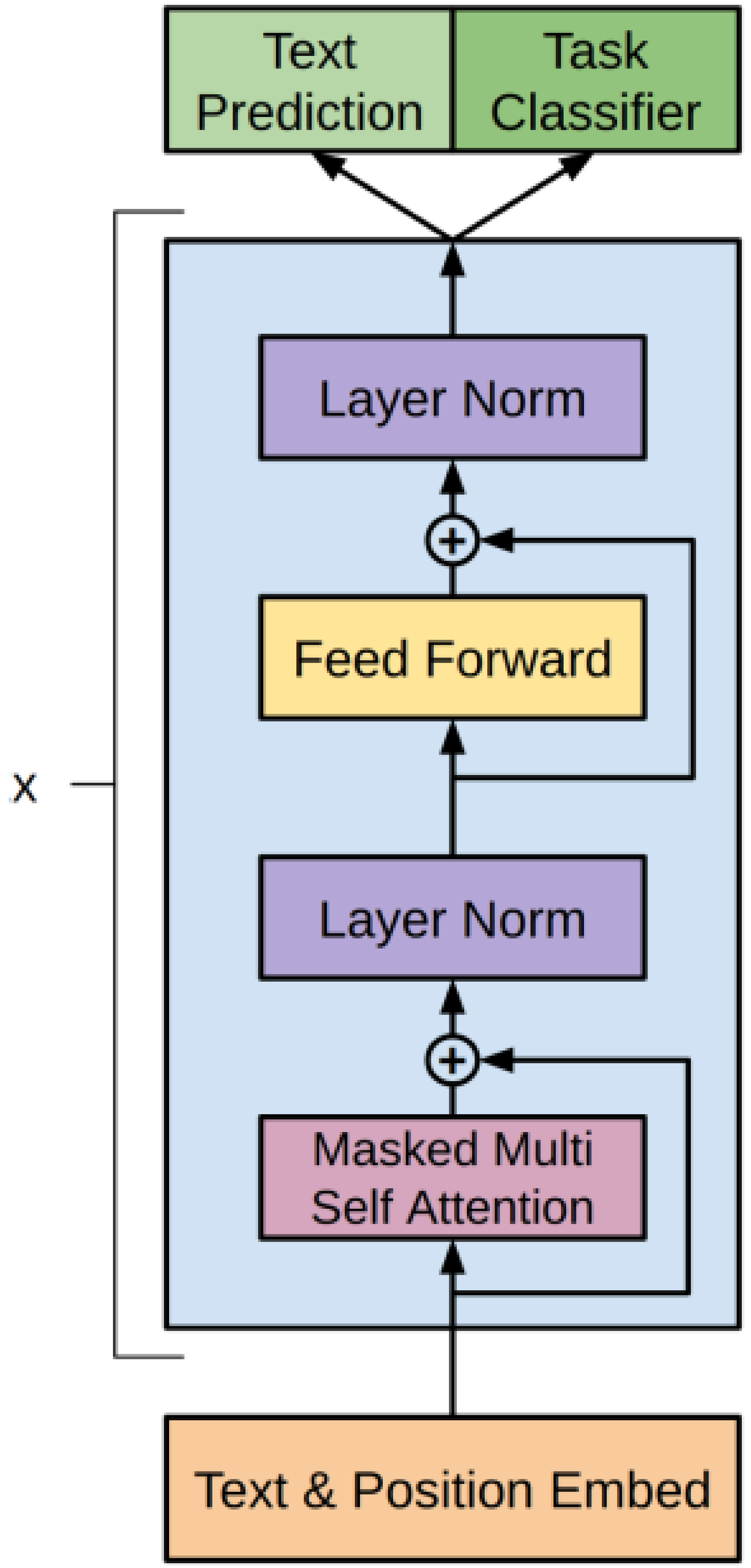
MHSA
(Multi Head Self Attention)



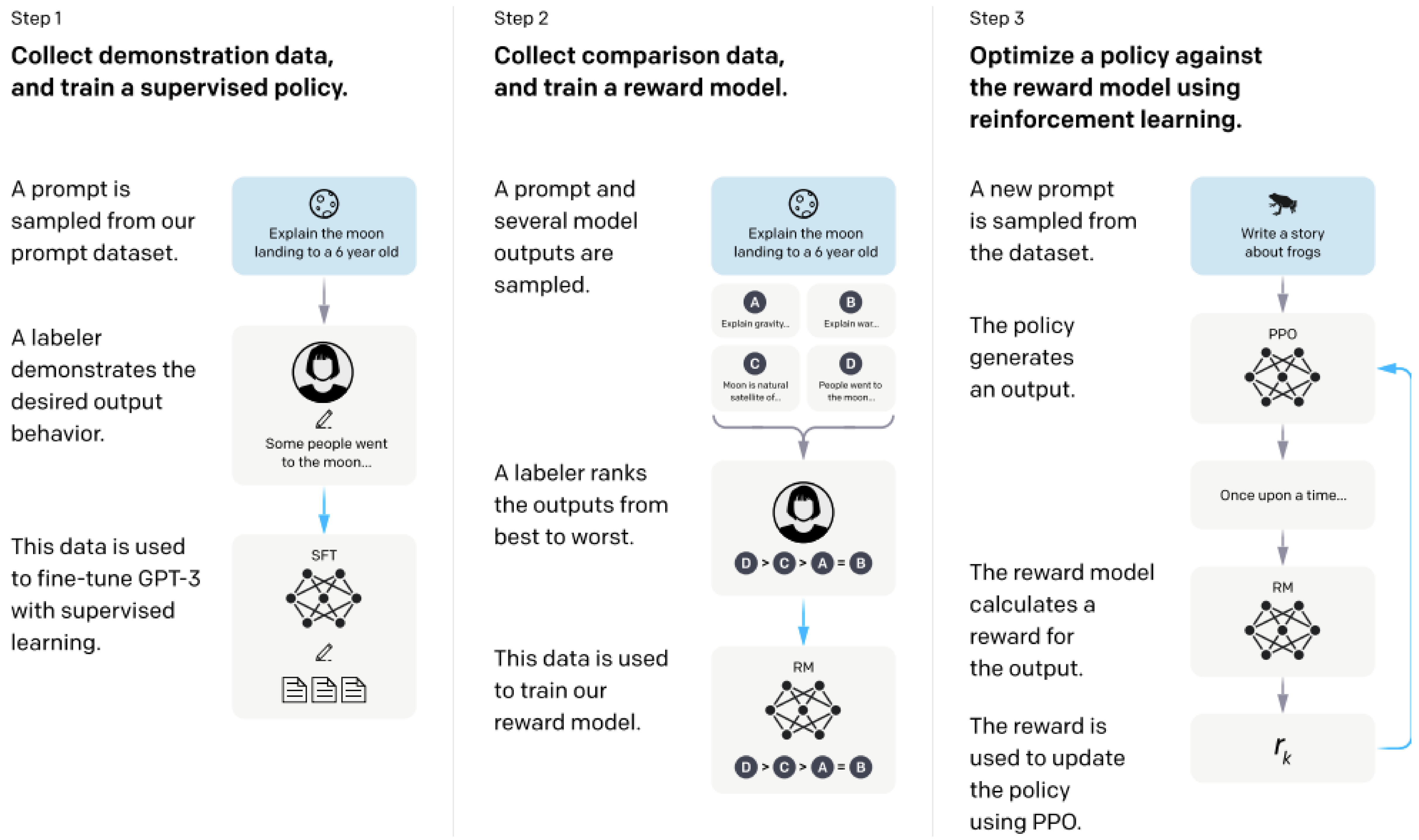
BERT



OpenAI GPT and It's Variant

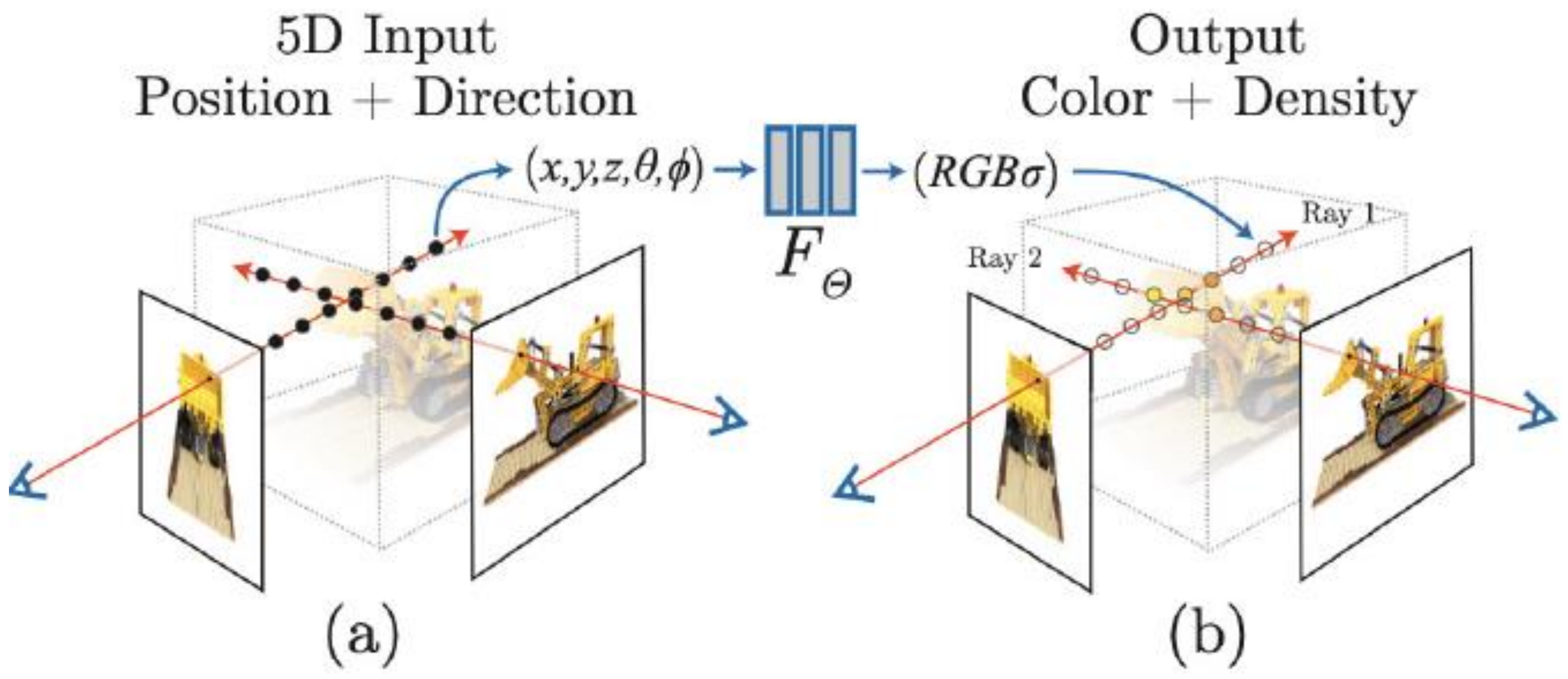


RLHF(InstructGPT)

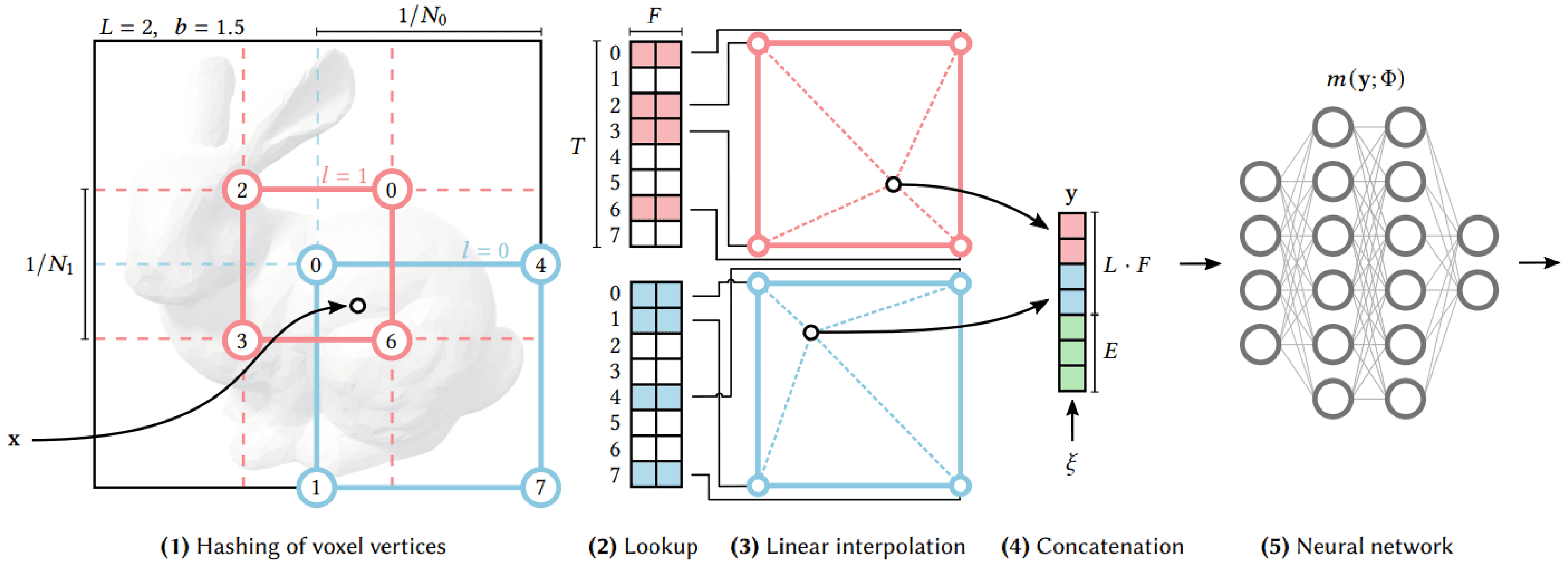


3D and Neural Rendering

NERF



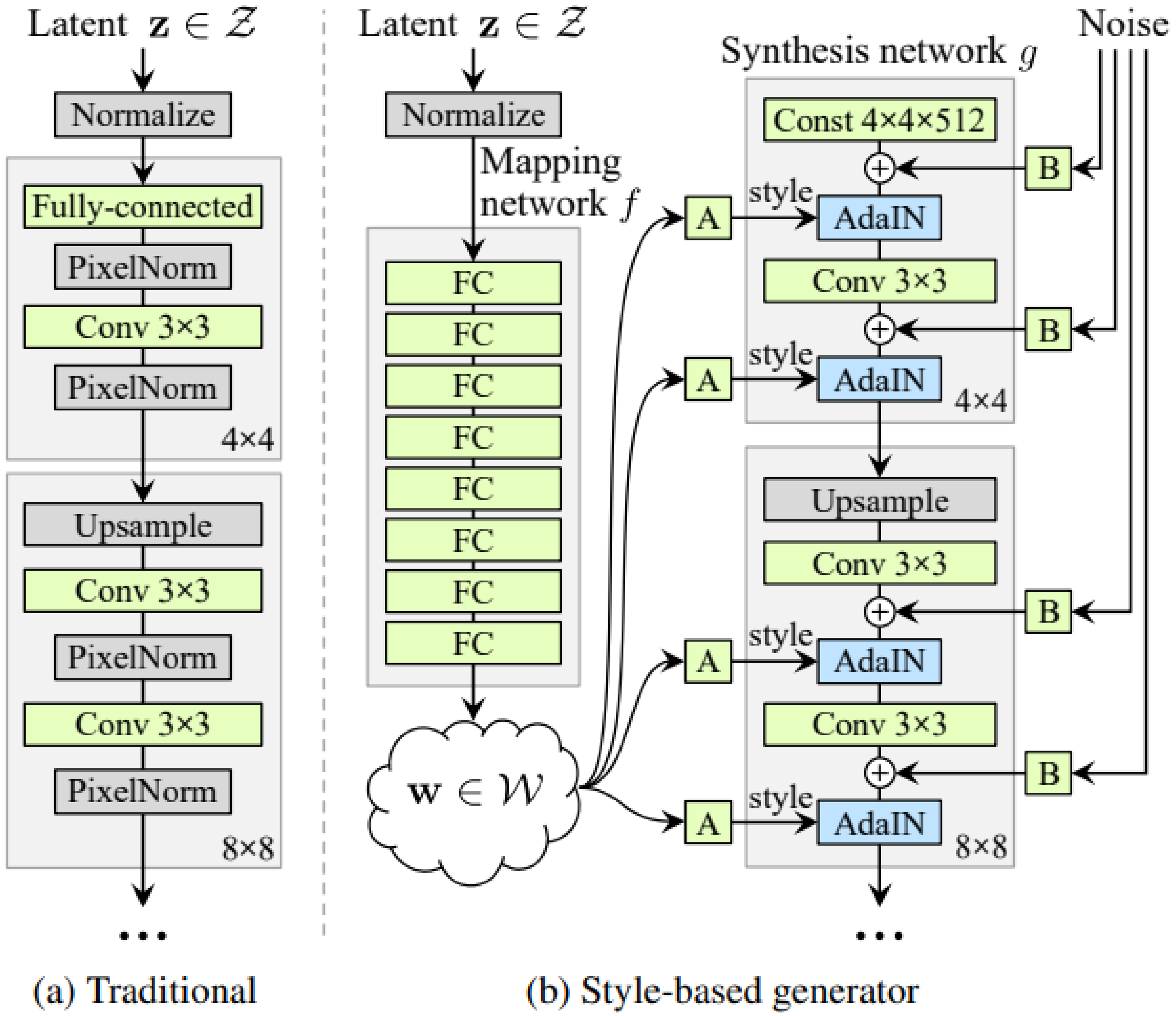
InstantNGP



Applications of Generative AI (Data types)

Image Generation

Figure 11. Four hand-picked examples illustrating the image quality and diversity achievable using StylegGAN2 (config F).



Text Synthesis

ChatGPT

Model: GPT-4

‹ 2 / 2 › R generate python dataloader for ERA5 surface level data

To create a Python DataLoader for ERA5 surface level data, you can use the following code. This example uses the `xarray` and `numpy` libraries to load and manipulate NetCDF files, which are commonly used for climate data. You can install them via pip:

bash Copy code

```bash
pip install xarray numpy
```

Now, you can create a custom DataLoader as follows:

python Copy code

```python
import xarray as xr
import numpy as np
from torch.utils.data import Dataset, DataLoader

class ERA5Dataset(Dataset):
    def __init__(self, data_path, transform=None):
        self.data_path = data_path
        self.transform = transform
        self.data = xr.open_dataset(data_path)

    def __len__(self):
        return len(self.data['time'])

    def __getitem__(self, idx):
        if isinstance(idx, slice):
```

BARD

R explain NVIDIA's picasso and Eddify (EDiff-I) model.

View other drafts

Sure.

Picasso is a text-to-image diffusion model developed by NVIDIA. It is trained on a massive dataset of text and images, and can be used to generate realistic images from text descriptions. For example, you could give Picasso the text description "a red apple on a blue table" and it would generate an image of a red apple sitting on a blue table.

Eddify (EDiff-I) is a differentiable neural computer architecture developed by NVIDIA. It is designed to be efficient and scalable, and can be used to train large language models. Eddify is currently being used to train Megatron-Turing NLG, a 530-billion parameter language model that is one of the largest language models ever created.

Both Picasso and Eddify are still under development, but they have the potential to revolutionize the way we interact with computers. By combining the power of text-to-image generation and differentiable neural computing, these technologies could make it possible for us to create realistic images and simulations from text descriptions. This could have a wide range of applications, from entertainment to scientific research.

Here are some additional details about each model:

Picasso

- Pros:

Enter a prompt here

Bard may display inaccurate or offensive information that doesn't represent Google's views.

Speech Synthesis

HifiGAN

Video Synthesis

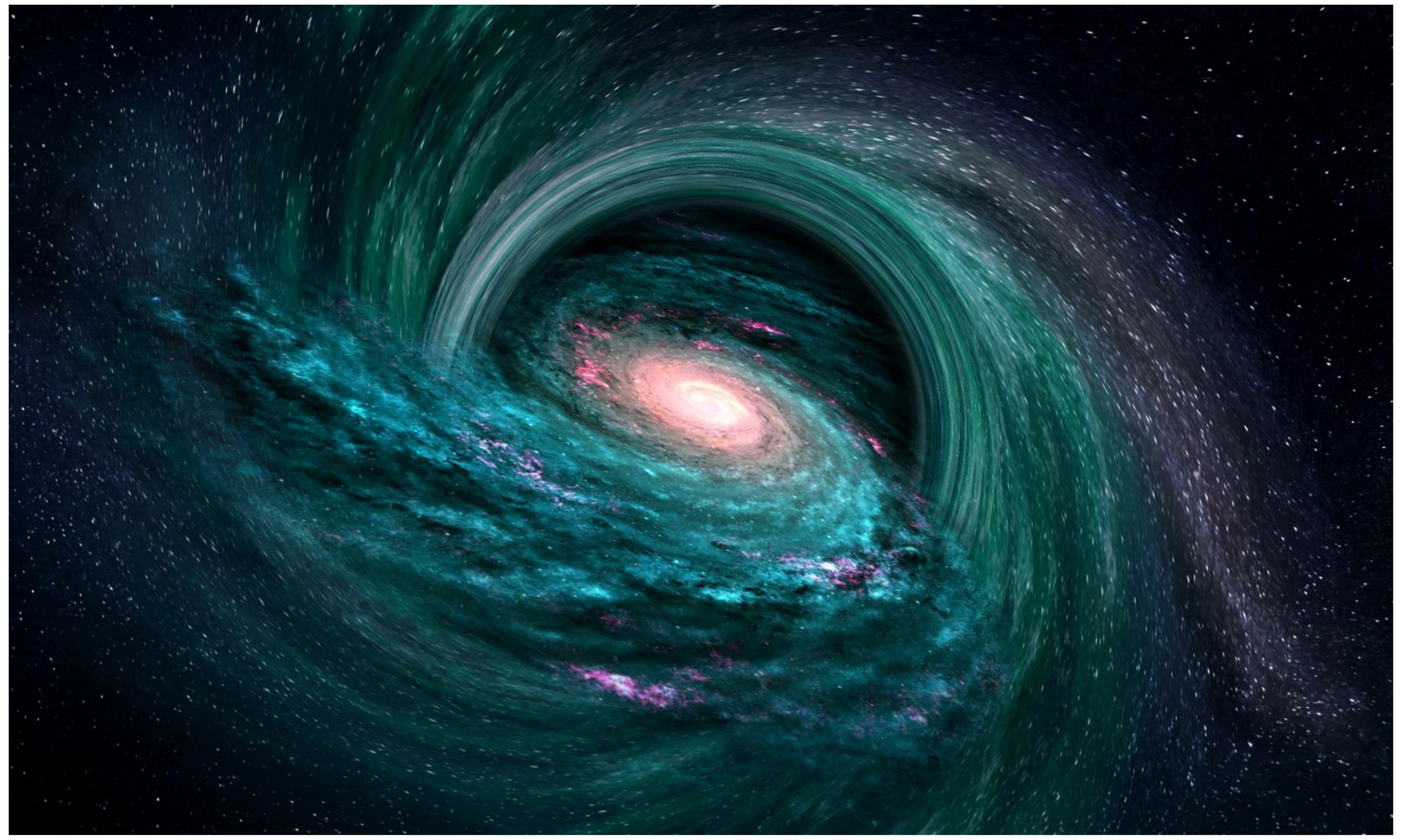
3D Mesh Generation

Avatars and Character Animation

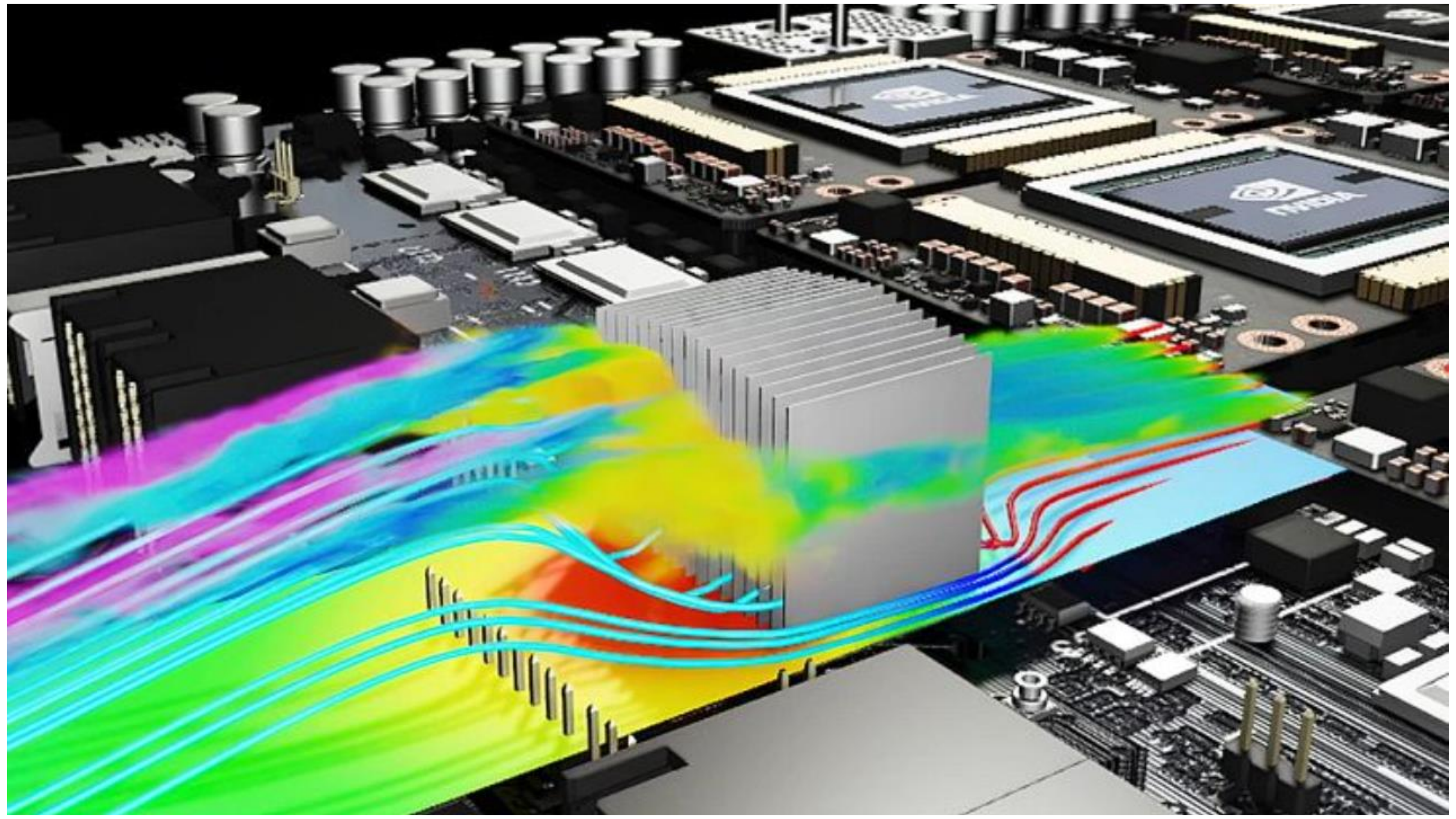
Applications of Generative AI (Industrial Domain)

Scientific Data Generation

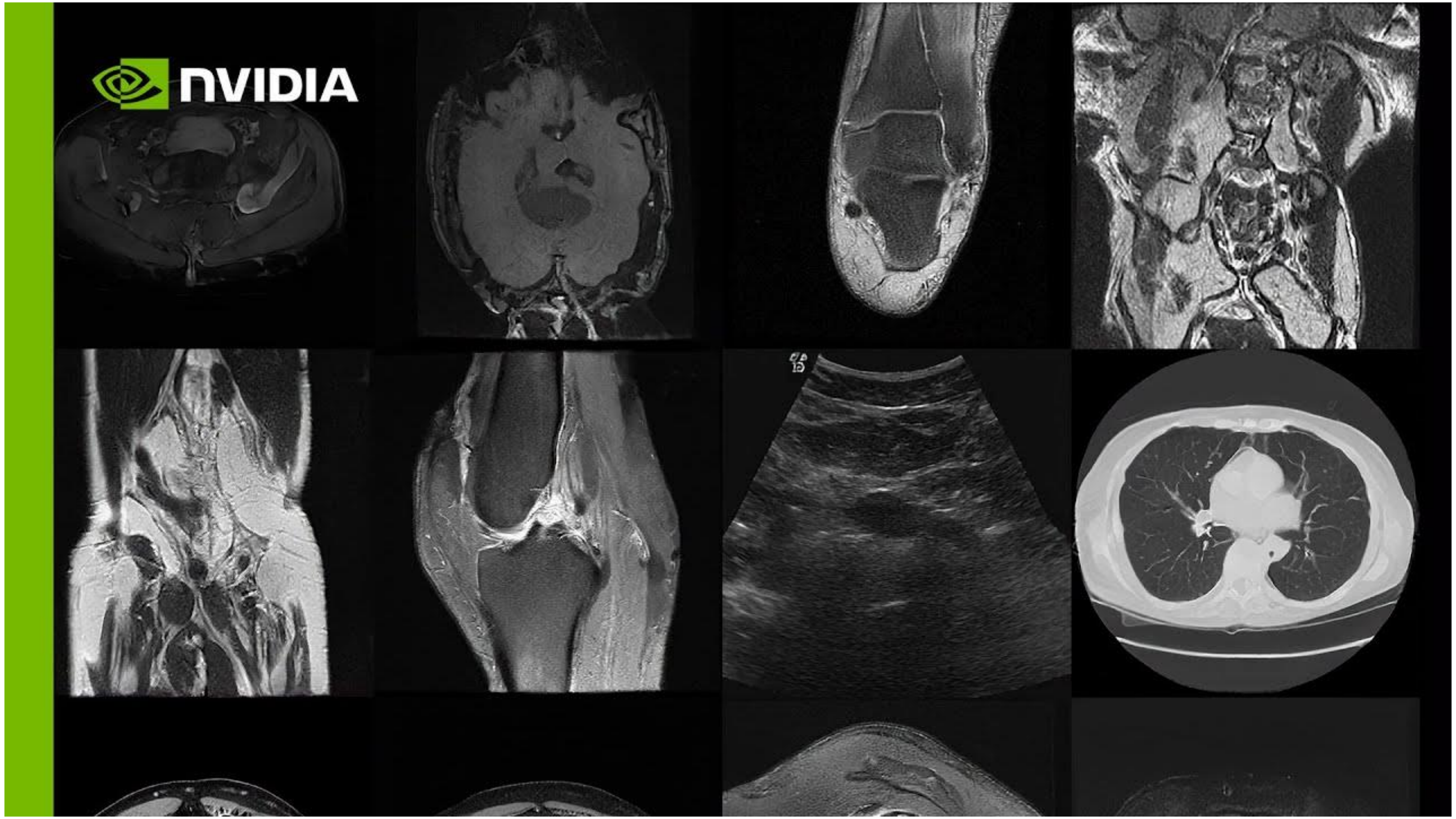
Gravitional Wave



CFD Flow with Modulus

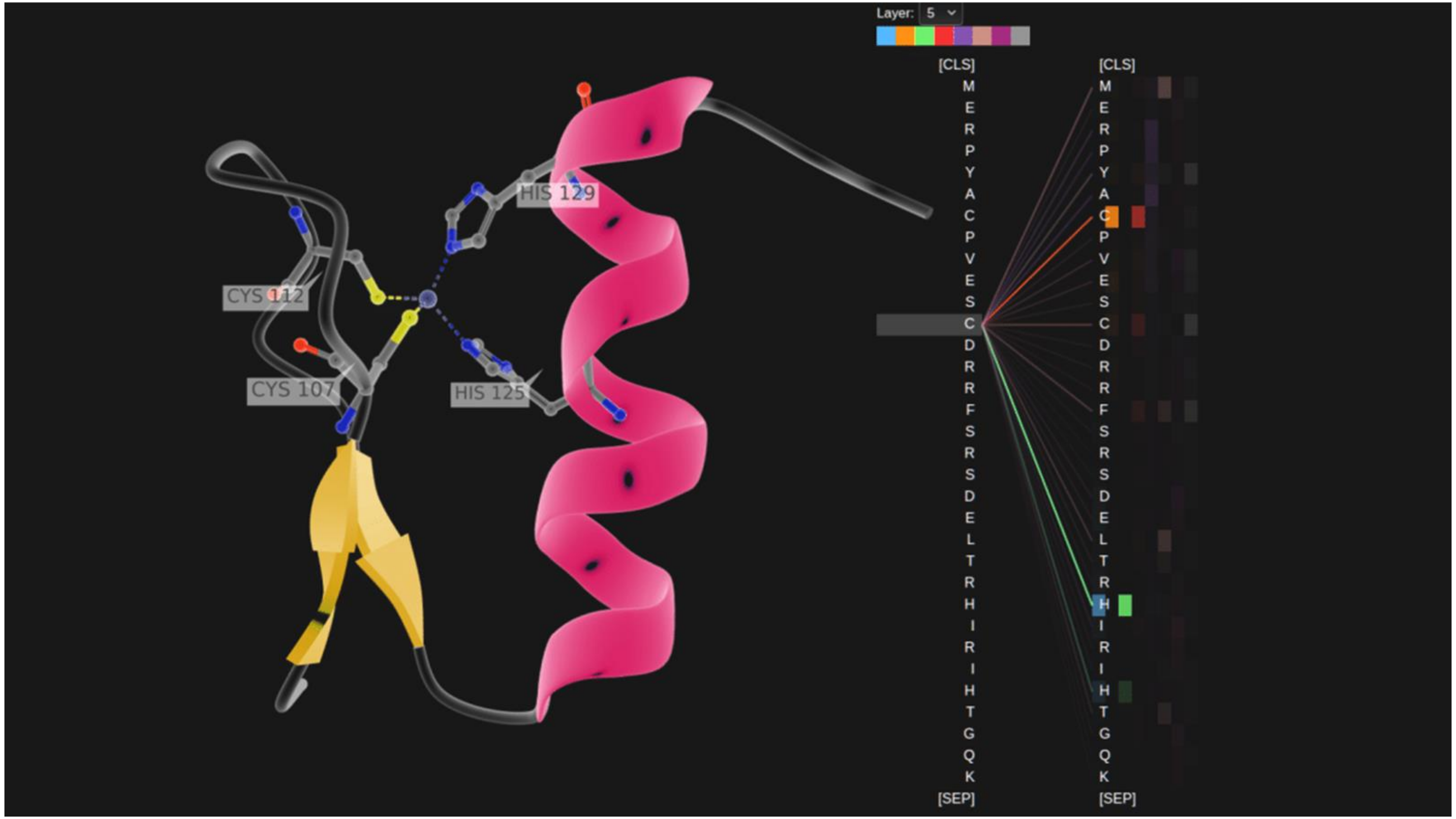


Generative AI in Healthcare

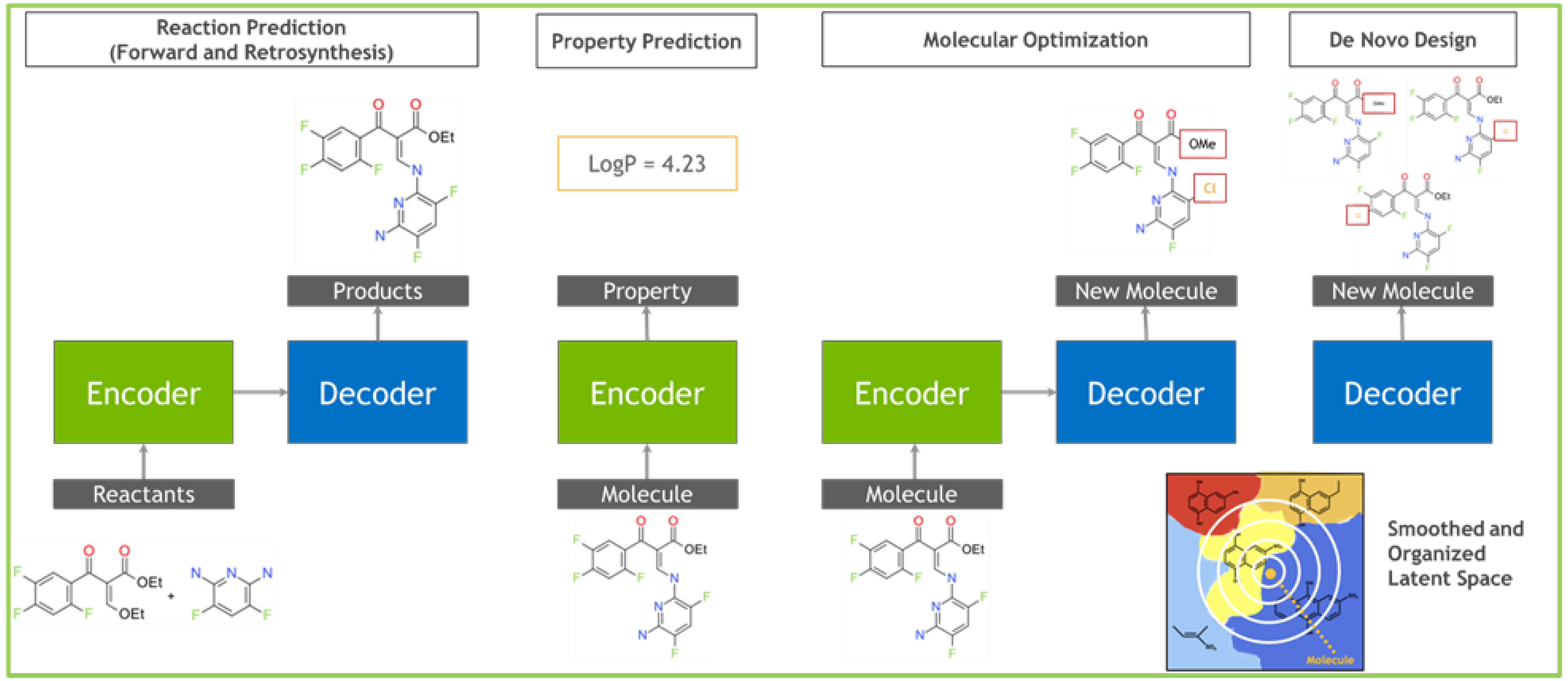


Generative AI in Protein Structure

BioNeMo

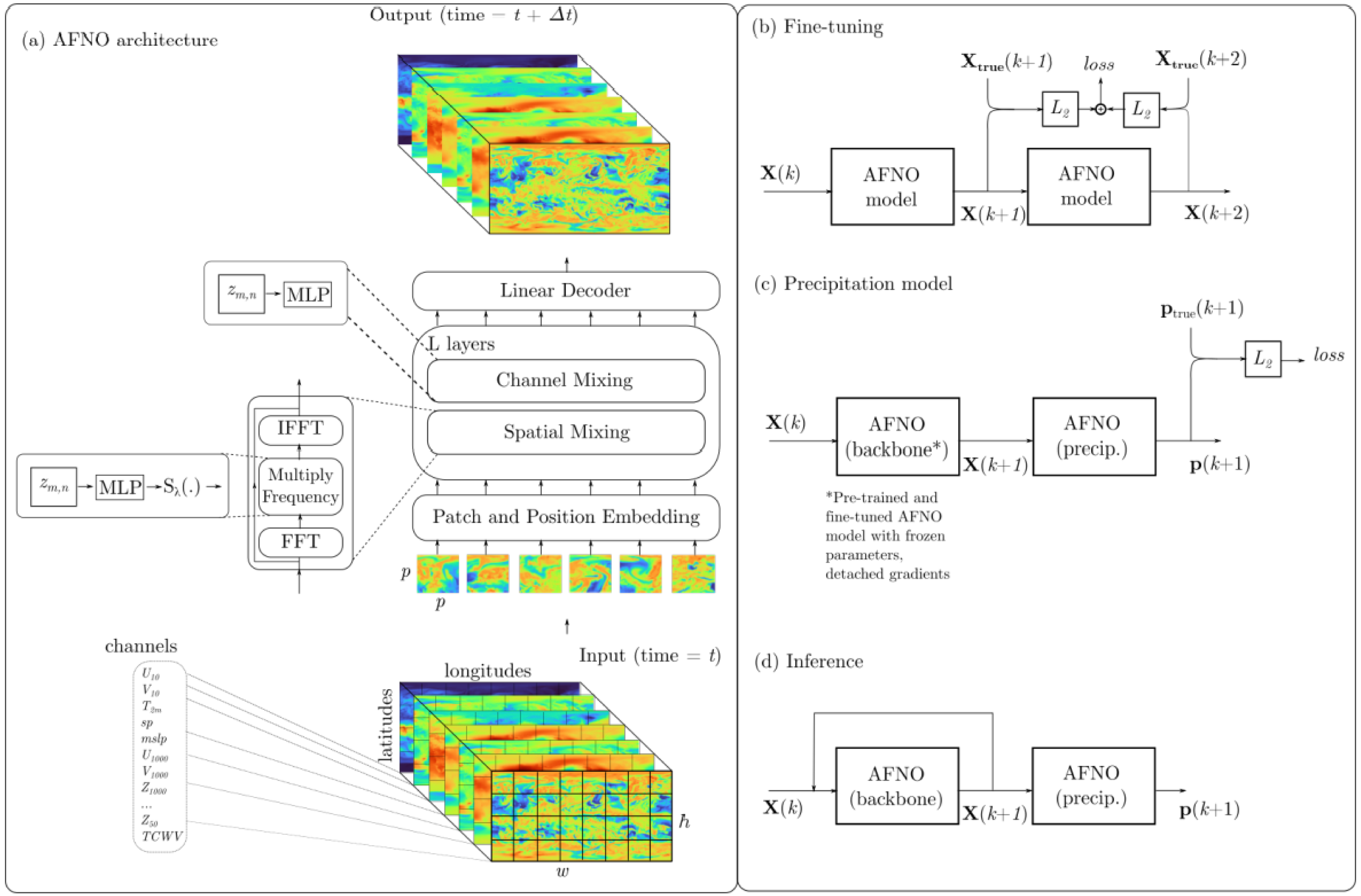


Generative AI in Smart Drug Discovery



Generative AI for AD

Generative AI for NWP

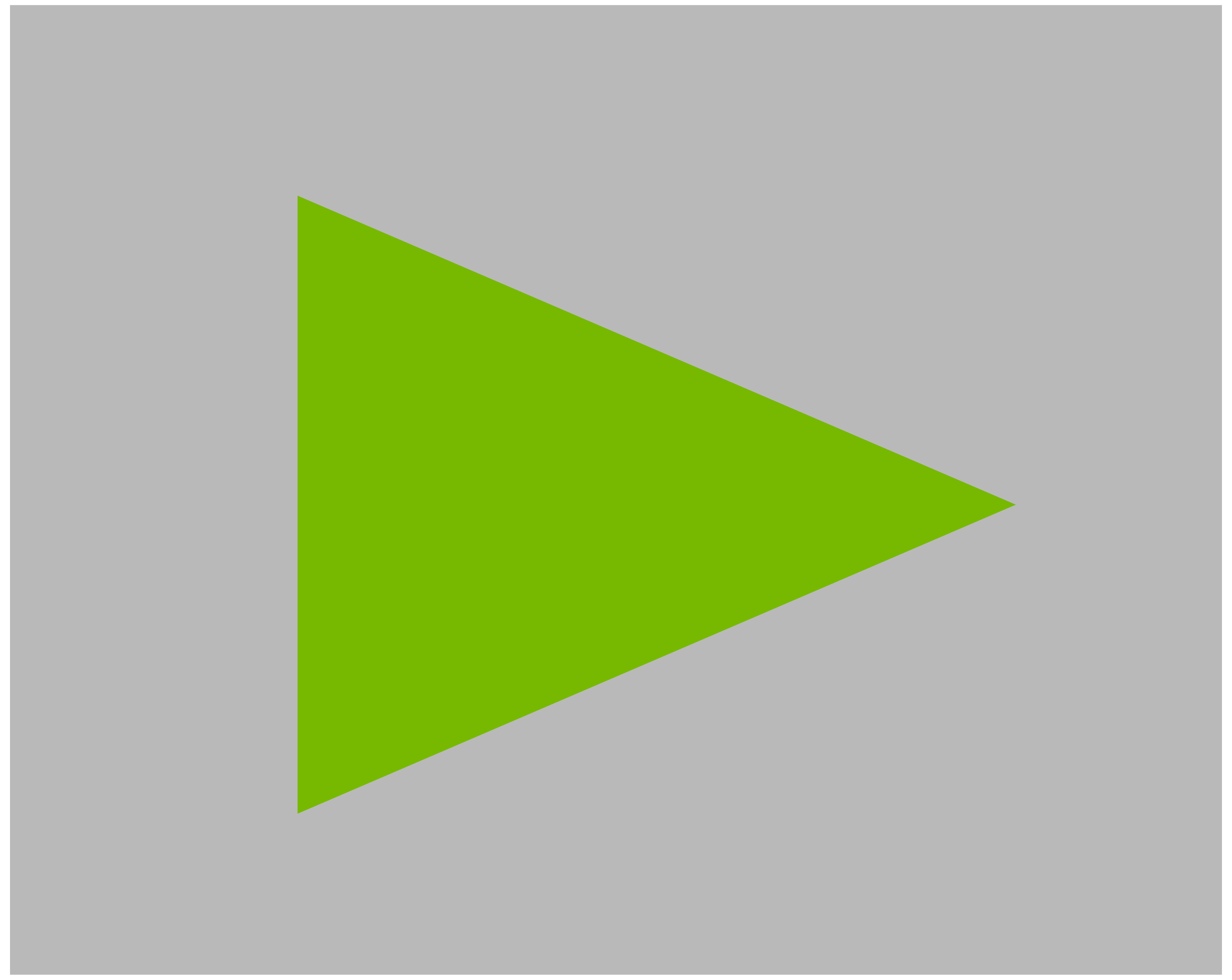


ASR - Openai Whisper model

- Whisper is an automatic speech recognition (ASR) system trained on 680,000 hours of multilingual and multitask supervised data collected from the web. We show that the use of such a large and diverse dataset leads to improved robustness to accents, background noise and technical language. Moreover, it enables transcription in multiple languages, as well as translation from those languages into English. We are open-sourcing models and inference code to serve as a foundation for building useful applications and for further research on robust speech processing.

- Other existing approaches frequently use smaller, more closely paired audio-text training datasets,[1 2,3] or use broad but unsupervised audio pretraining.[4,5,6] Because Whisper was trained on a large and diverse dataset and was not fine-tuned to any specific one, it does not beat models that specialize in LibriSpeech performance, a famously competitive benchmark in speech recognition. However, when we measure Whisper's zero-shot performance across many diverse datasets we find it is much more robust and makes 50% fewer errors than those models.

ASR test



ASR for GW domain

[
{ "idx": 1,
"text": "LIGO's data on graviton-mass bounds from GW150914 event necessitate quantum loop gravity theories.",
"NER": ["LIGO", "GW150914"],
"vocab": ["graviton-mass bounds", "quantum loop gravity"]   },

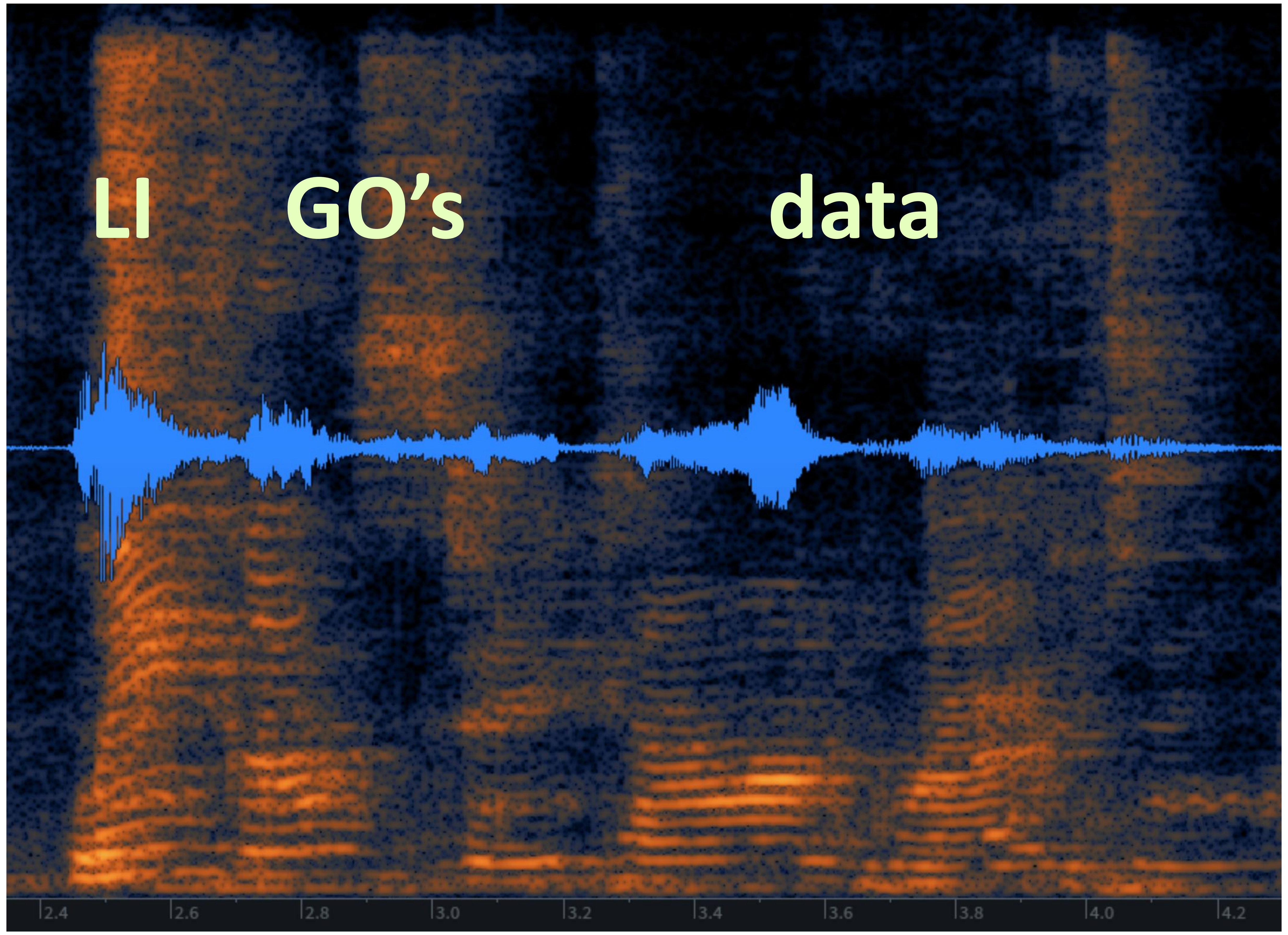
{ "idx": 2,
"text": "Prof. Cutler's work on Kerr-Newman black holes elucidates the gyroscopic precession in strong fields.",
"NER": ["Prof. Cutler", "Kerr-Newman black holes"],
"vocab": ["gyroscopic precession", "strong fields"]   },

{ "idx": 3,
"text": "The detection of B-mode polarization in CMB by BICEP2 poses questions on primordial gravitation waves.",
"NER": ["B-mode polarization", "CMB", "BICEP2"],
"vocab": ["primordial gravitation waves"]   },

{ "idx": 4,
"text": "VIRGO's analysis of gravitational wave chirality offers insights into parity violation in general relativity.",
"NER": ["VIRGO"],
"vocab": ["gravitational wave chirality", "parity violation", "general relativity"]   },

{ "idx": 5,
"text": "The LISA Pathfinder's success in measuring low-frequency gravitational waves revolutionizes space-based observatories.",
"NER": ["LISA Pathfinder"],
"vocab": ["low-frequency gravitational waves", "space-based observatories"]   },

TTS

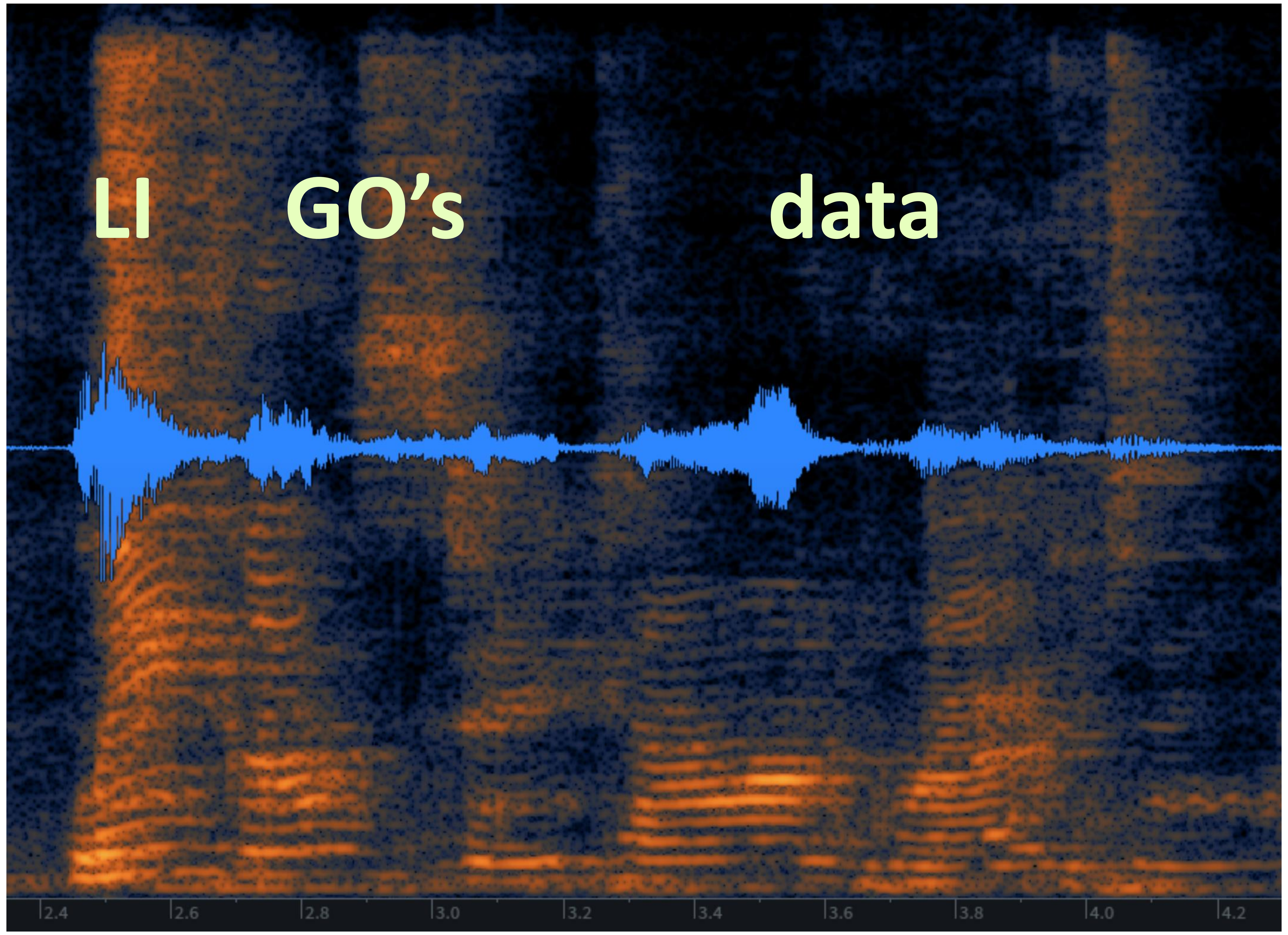


Text

spectrogram

waveform

ASR



Wave → spectrogram → Text

Research Idea

ASR

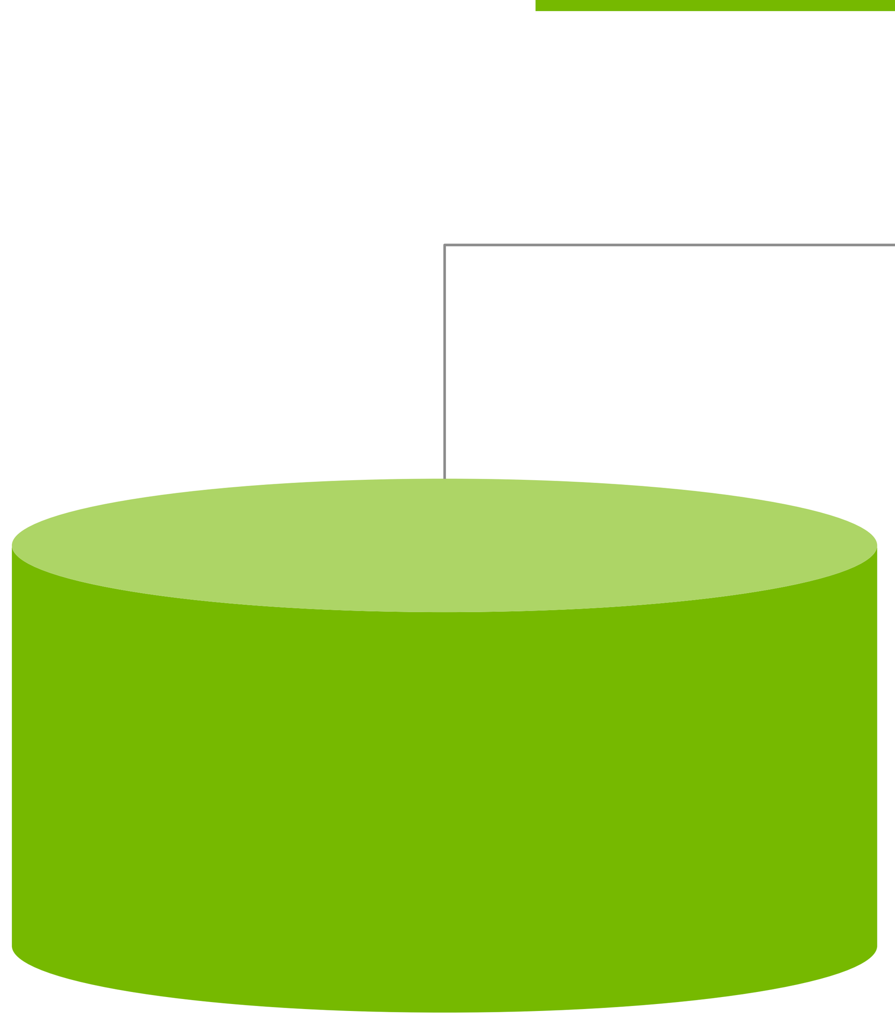
0.1% WER(Word Error Rate)

Noisy(robustness), domain

NMT

LLM

RAG



TTS

Emotion, fluent, personality

Low resources
Voice clone/Conversion