Generative AI In Life(&)Science

Viktoria Schuster
Postdoctoral Researcher
Department of Computer Science

IntroToML

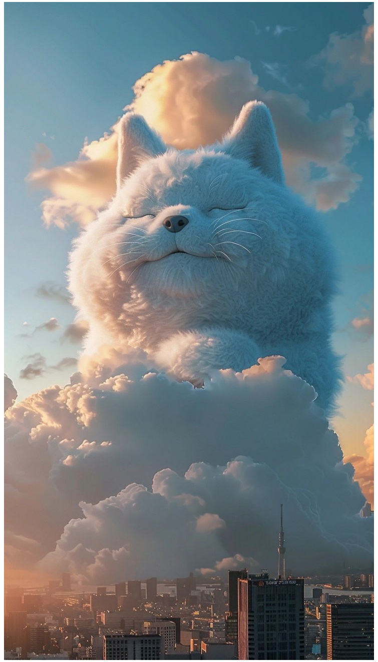
Generative AI in everyday life

- content creation:
    - **Text** (GPT, Gemini, Llama, Claude, ...)
    - Image (Midjourney, DALL-E, ...)
    - Video (Sora, ...)
- Speech recognition (Whisper)
- Vector graphics (Adobe)
- **Developer tools** (GitHub Copilot)
- **Literature review** (scite)

chatGPT

https://en.wikipedia.org/wiki/ChatGPT

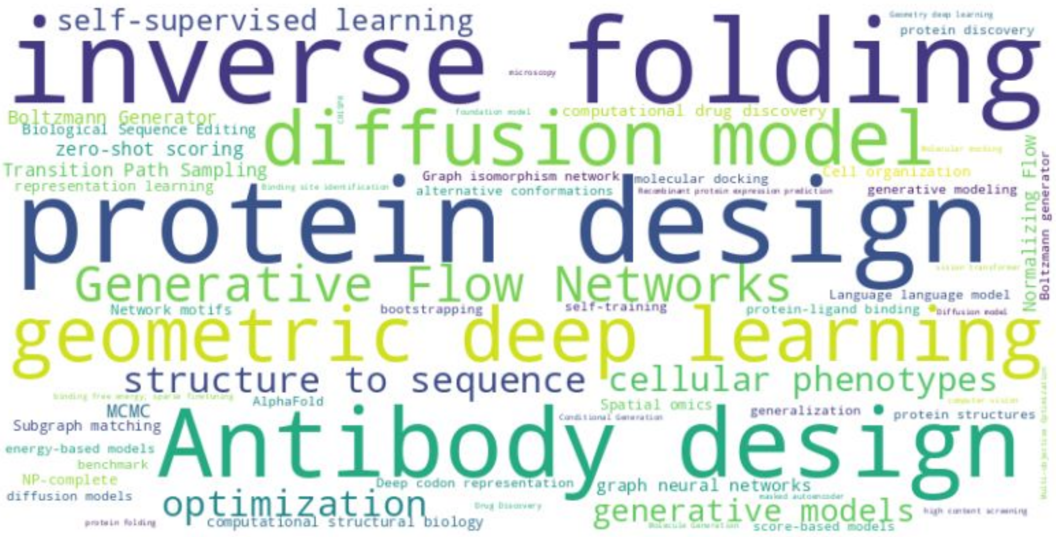
Midjourney



https://www.midjourney.com/jobs/7437a988-5ed3-4de1-a916-61c85f49af6c?index=0

Why are generative models interesting for us?

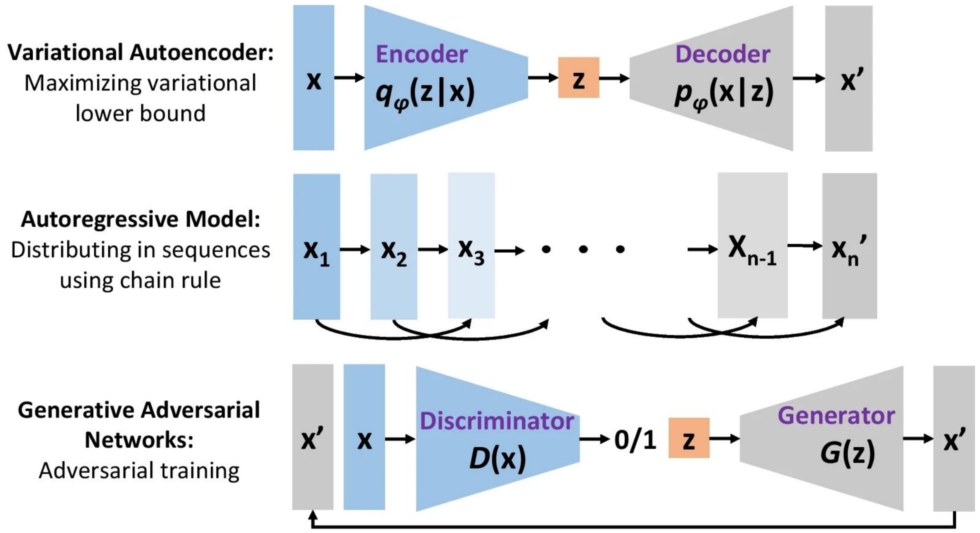
- **Synthetic data**
- **Drug discovery**
- **Molecular design** (protein, RNA, …)
- **Perturbation/mutation** effect modeling
- …



NeurIPS GenBio 2023 spotlight paper keywords

Common types of generative models

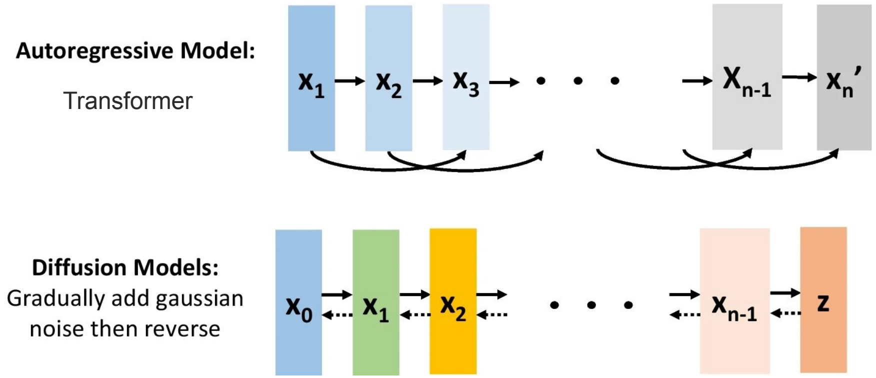
Once upon a time ...



https://link.springer.com/article/10.1007/s43503-023-00017-z

Common types of generative models

Transformers and Diffusion models are all the rage now and behind the impressive performance of many models.

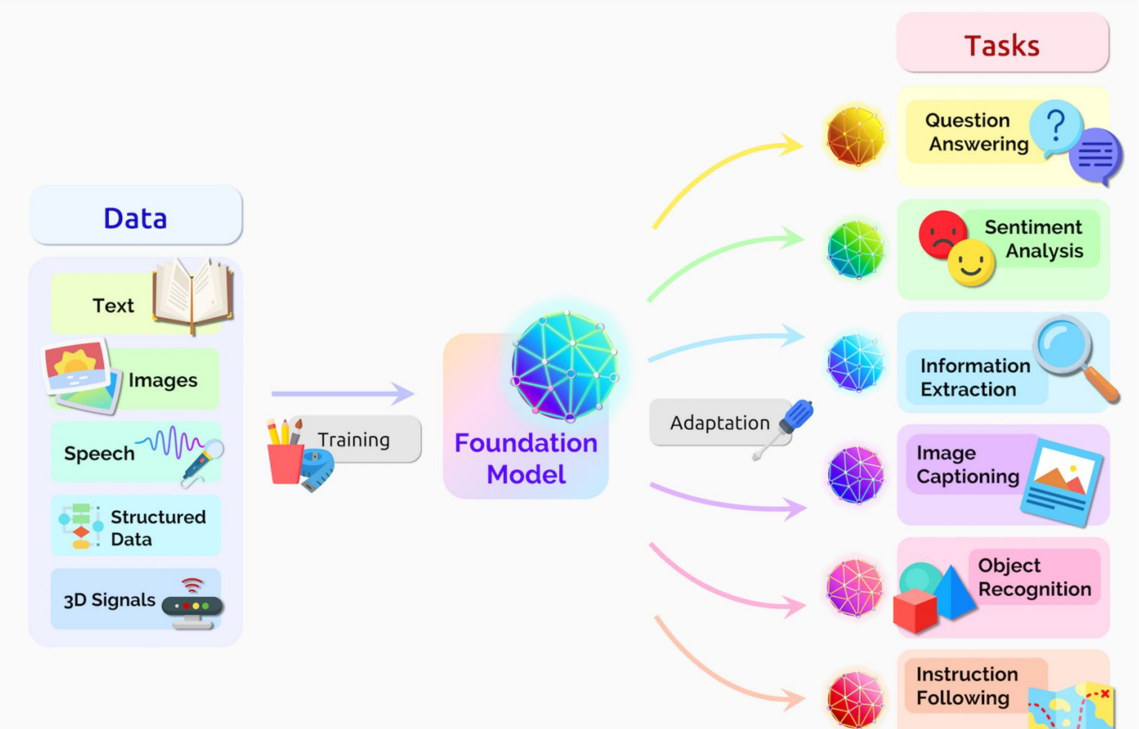


https://link.springer.com/article/10.1007/s43503-023-00017-z

Transformer

- Replaces many architectures for autoregressive and other tasks (CNNs and RNNs)

- Learns relationships between words in Sentences,
  AAs in proteins,
  …



https://blogs.nvidia.com/blog/what-is-a-transformer-model/

Transformer

What makes this model architecture so special?

Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com

Noam Shazeer* Google Brain noam@google.com

Niki Parmar* Google Research nikip@google.com

Jakob Uszkoreit* Google Research usz@google.com

Llion Jones* Google Research llion@google.com

Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu

Łukasz Kaiser* Google Brain lukaszkaiser@google.com

Illia Polosukhin* ‡ illia.polosukhin@gmail.com

[2] Vaswani et al. (2017) Attention Is All You Need, *NeurIPS*.

Transformer

What is attention? *(the ultra simplified version)*

Corpus

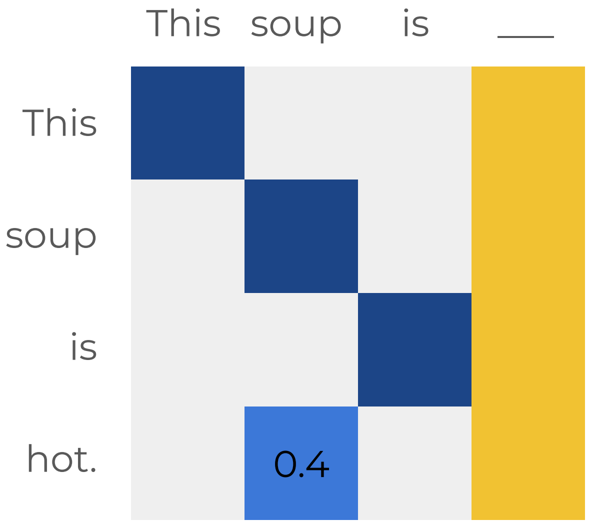
This soup is hot.
This soup is tasty.
This soup is hot.
This soup is hot.
This soup is cold.
This soup is cold.
This soup is hot.
This soup is tasty.
...
This soup is bat.

This is a token.

	This	soup	is	hot.
This	1			
soup		1		0.4
is			1	
hot.		0.4		1

Transformer

How can we use attention? *(the ultra simplified version)*



Next token prediction probabilities:

This soup is ___

hot	0.40
cold	0.20
tasty	0.20
...	...
bat	0.01

Transformer

How can we use attention? *(the ultra simplified version)*

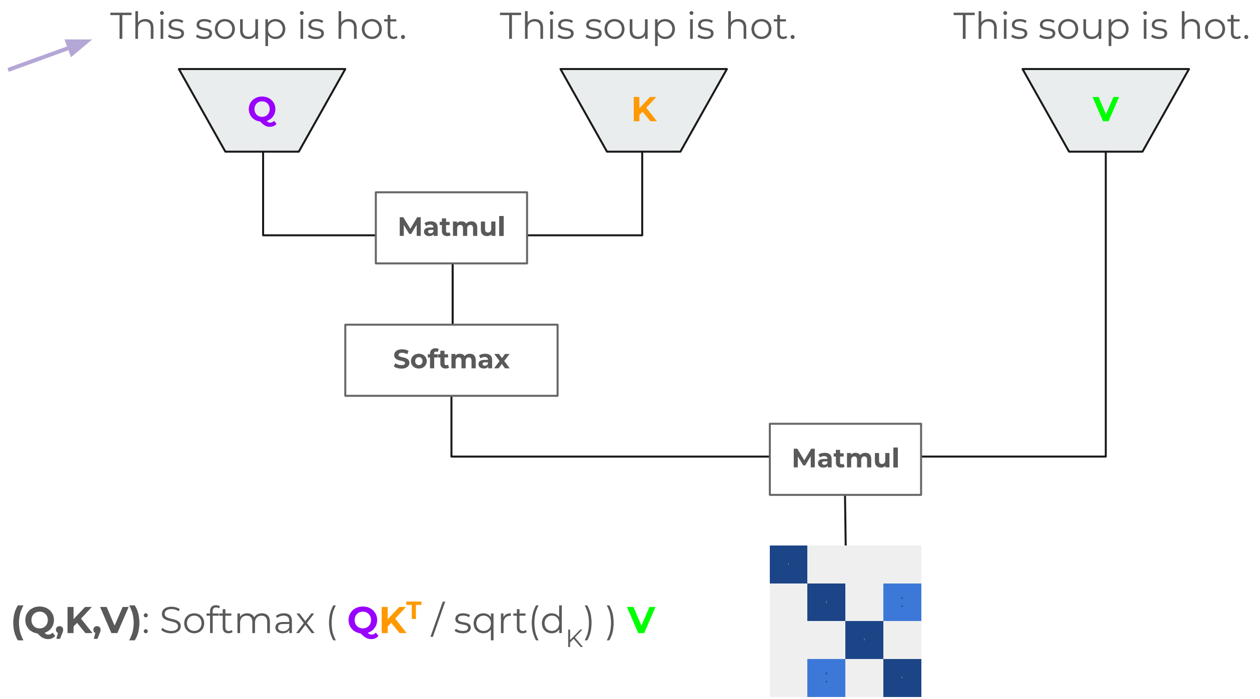
These probabilities can change with context:

It is winter. I like that this soup is ___

hot	0.70
cold	0.09
tasty	0.20
bat	0.01

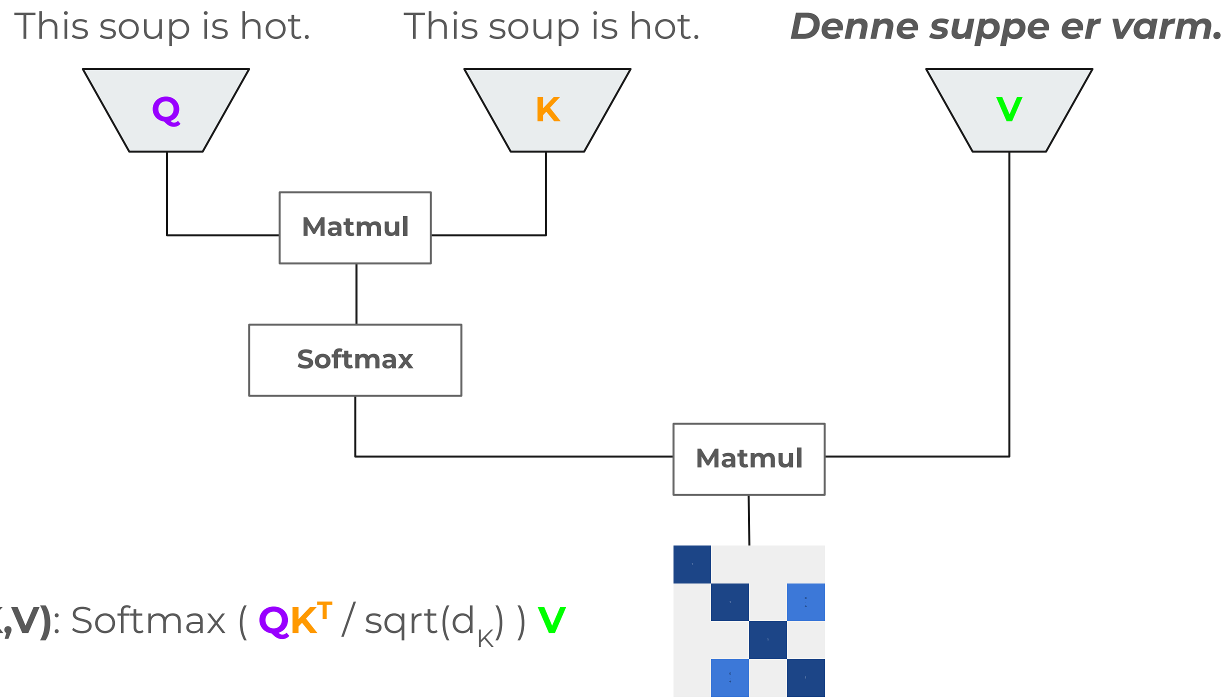
Transformer

This will be embedded



(Self-)Attention (Q,K,V): Softmax (QK^T / sqrt(d_K)) V

Transformer



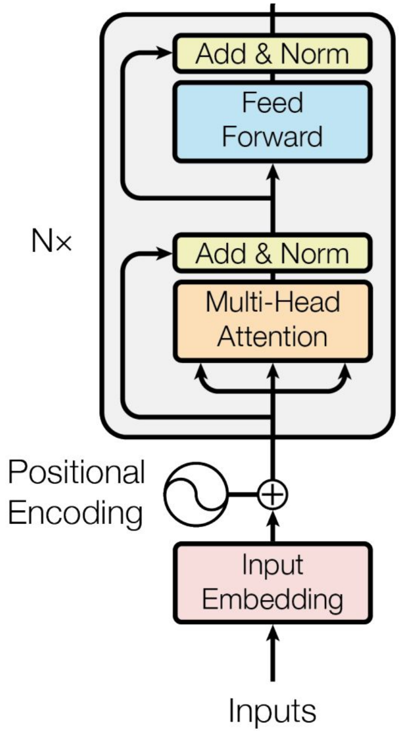
Cross-Attention (Q,K,V): Softmax (QK^T / sqrt(d_K)) V

Transformer

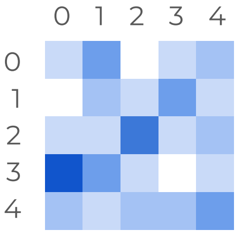
Apart from attention, the transformer architecture

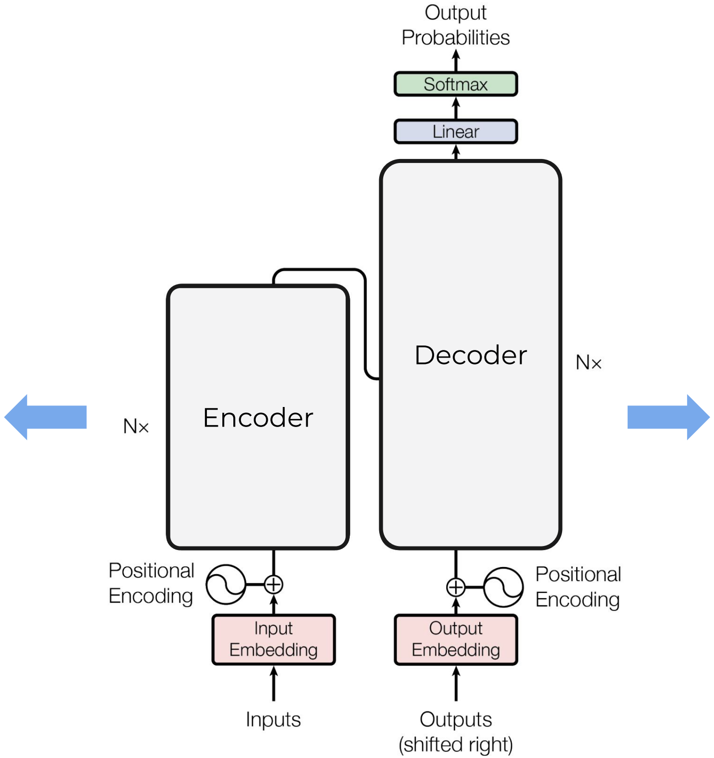
is pretty straightforward:

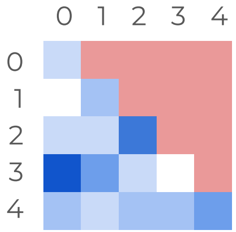
1. Attention
2. Normalization
3. Feed Forward NN
4. Normalization



Transformer



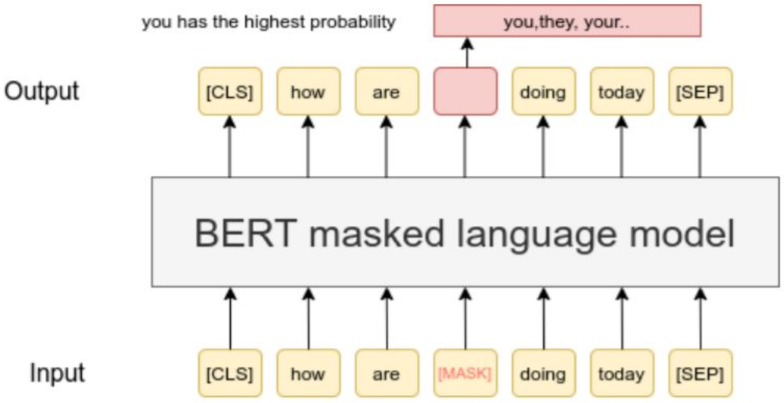




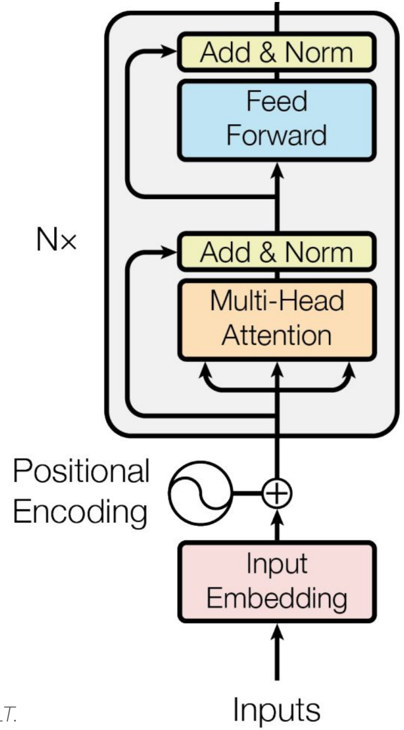
[2] Vaswani et al. (2017) Attention Is All You Need, *NeurIPS*.

Transformer

BERT: Bidirectional Encoder Representations from Transformers



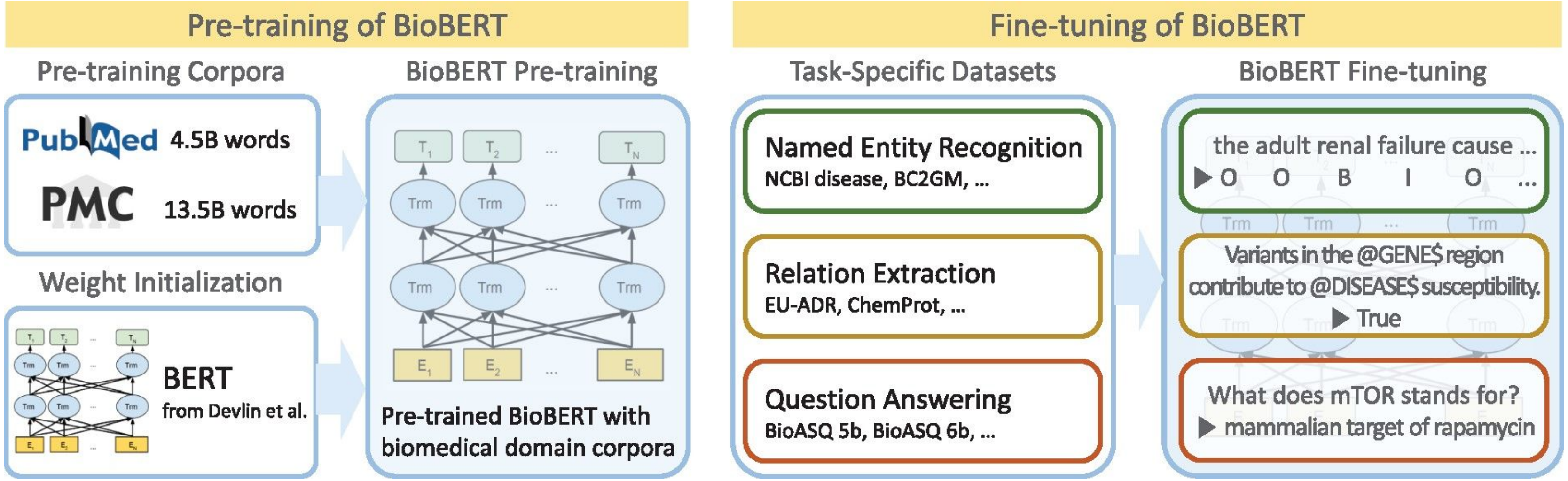
https://www.sbert.net/examples/unsupervised_learning/MLM/README.html



[3] Delvin et al. (2019) BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, *NAACL-HLT*.

Transformers in Biology

Many transformers out there for biological data: **- Medical data**



[4] Lee et al. (2019) BioBERT: a pre-trained biomedical language representation model for biomedical text mining, *Bioinformatics*.

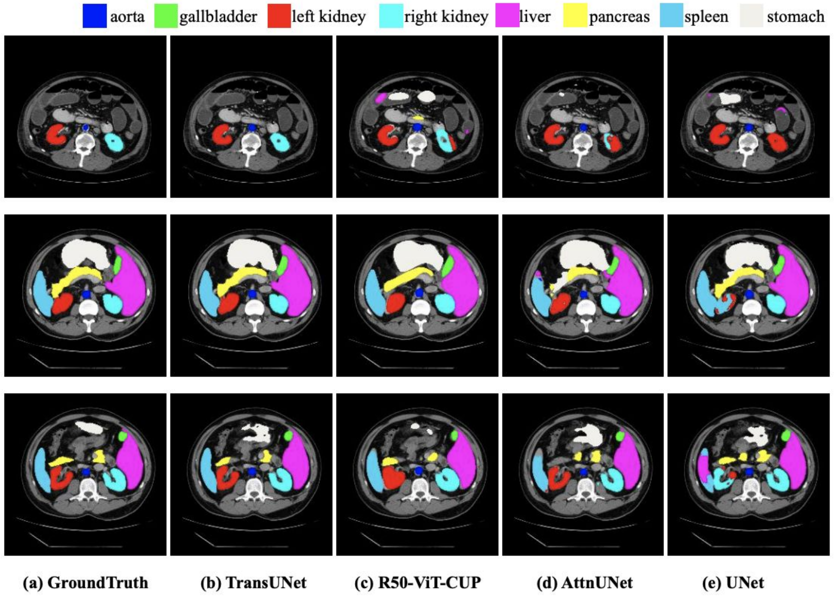
Transformers in Biology

Many transformers out there for biological data:

- **Medical data**
- RNA: CodonBert
- Proteins: AlphaFold, ESM

They can be used for many downstream tasks:

- **Medical image analysis**
- Mutational effect prediction
- Protein, antibody, vaccine, … design
- Gene therapy



[5] Chen et al. (2022) TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation, *CoRR*.

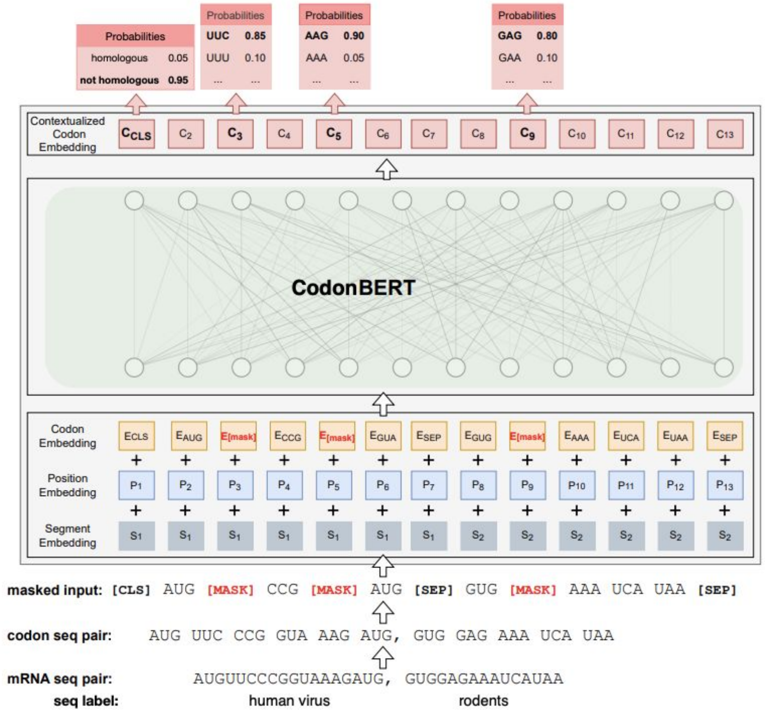
Transformers in Biology

Many transformers out there for biological data:

- Medical data
- **RNA: CodonBert**
- Proteins: AlphaFold, ESM (open source!)

They can be used for many downstream tasks:

- Medical image analysis
- **Mutational effect prediction**
- Protein, antibody, **vaccine**, ... **design**
- **Gene therapy**



[6] Li et al. (2023) CodonBERT: Large Language Models for mRNA design and optimization, *NeurIPS*.

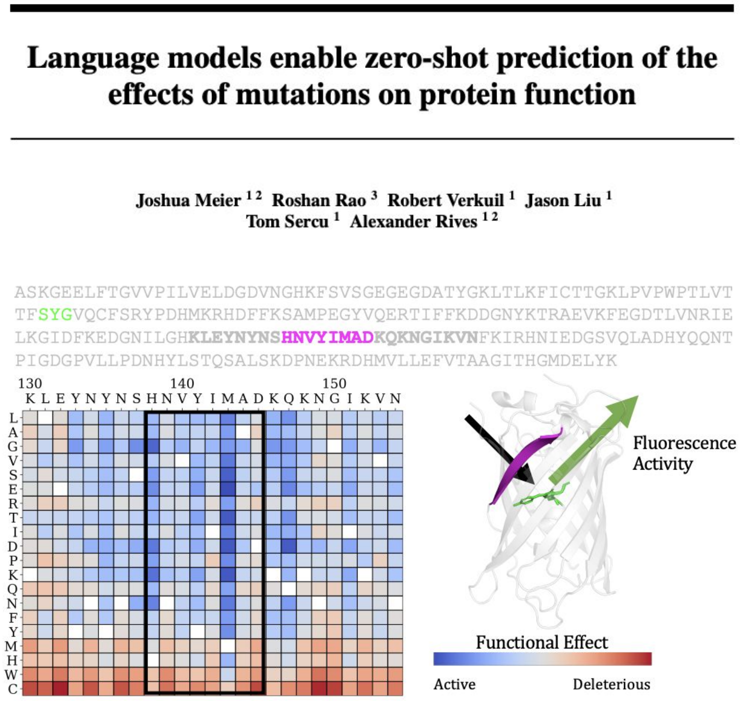
Transformers in Biology

Many transformers out there for biological data:

- Medical data
- RNA: CodonBert
- **Proteins**: AlphaFold, **ESM**

They can be used for many downstream tasks:

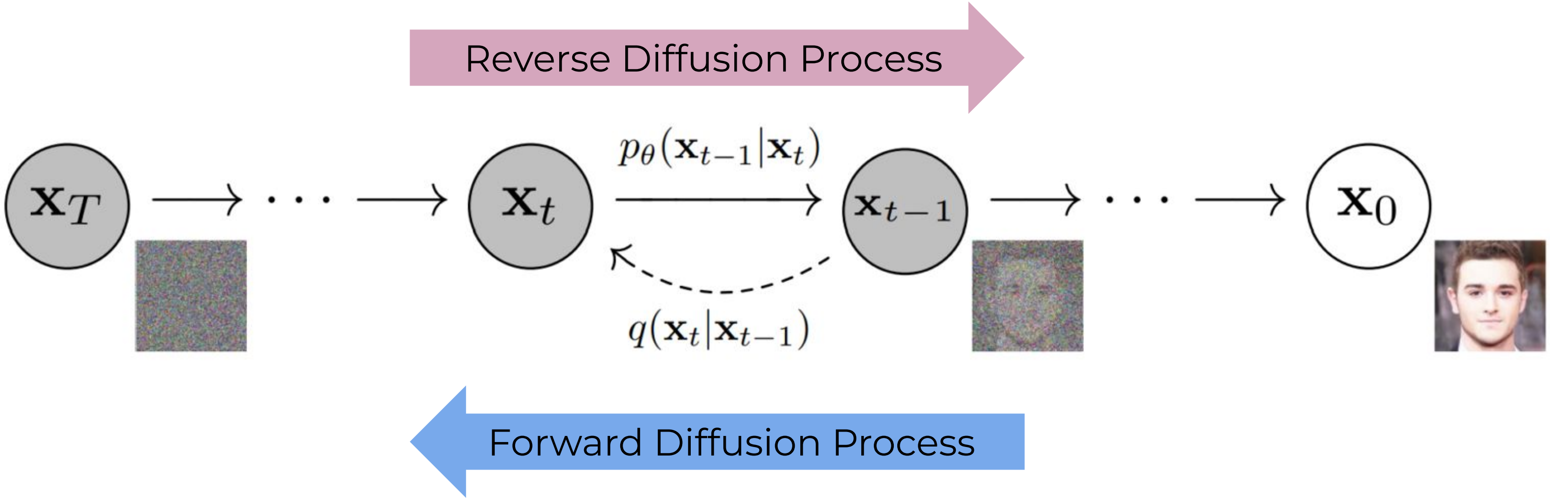
- Medical image analysis
- **Mutational effect prediction**
- **Protein**, **antibody**, vaccine, … **design**
- Gene therapy



[7] Meier et al. (2021) Language models enable zero-shot prediction of the effects of mutations on protein function, *NeurIPS*.

Diffusion - DDPM

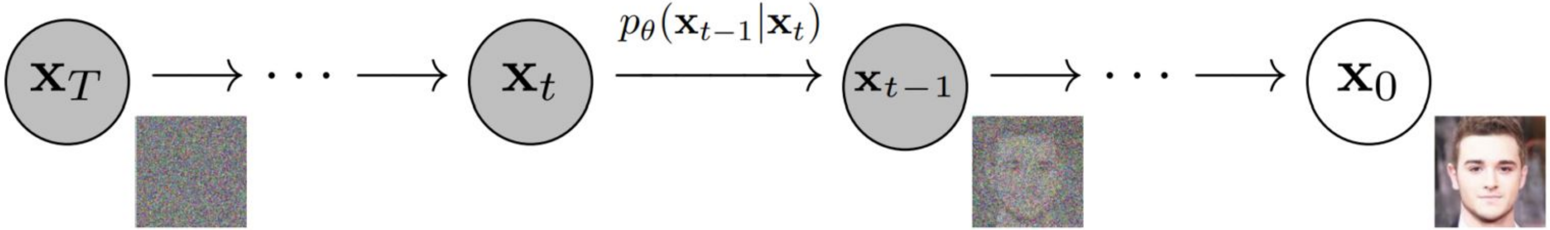
Denoising Diffusion Probabilistic Models



[8] Ho et al. (2020) Denoising Diffusion Probabilistic Models, *NeurIPS*.

Diffusion

How can we learn to remove noise?



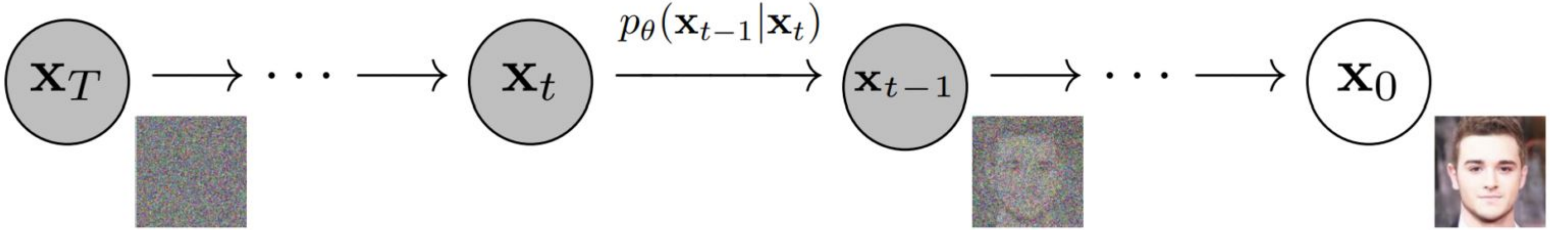
$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \underbrace{\boldsymbol{\mu}_\theta(\mathbf{x}_t, t)}_{\text{mean}}, \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$

$$\boldsymbol{\mu}_\theta(\mathbf{x}_t, t) = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\right)$$

[8] Ho et al. (2020) Denoising Diffusion Probabilistic Models, *NeurIPS*.

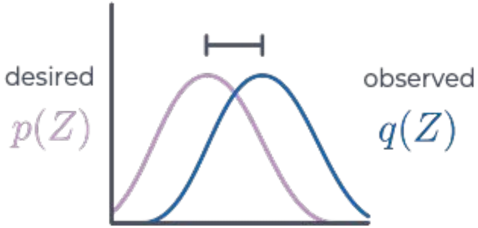
Diffusion

How can we learn to remove noise?



WE LEARN TO PREDICT THE SAMPLED NOISE!

$$L_{\text{simple}}(\theta) := \mathbb{E}_{t,\mathbf{x}_0,\boldsymbol{\epsilon}}\left[\left\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t)\right\|^2\right]$$

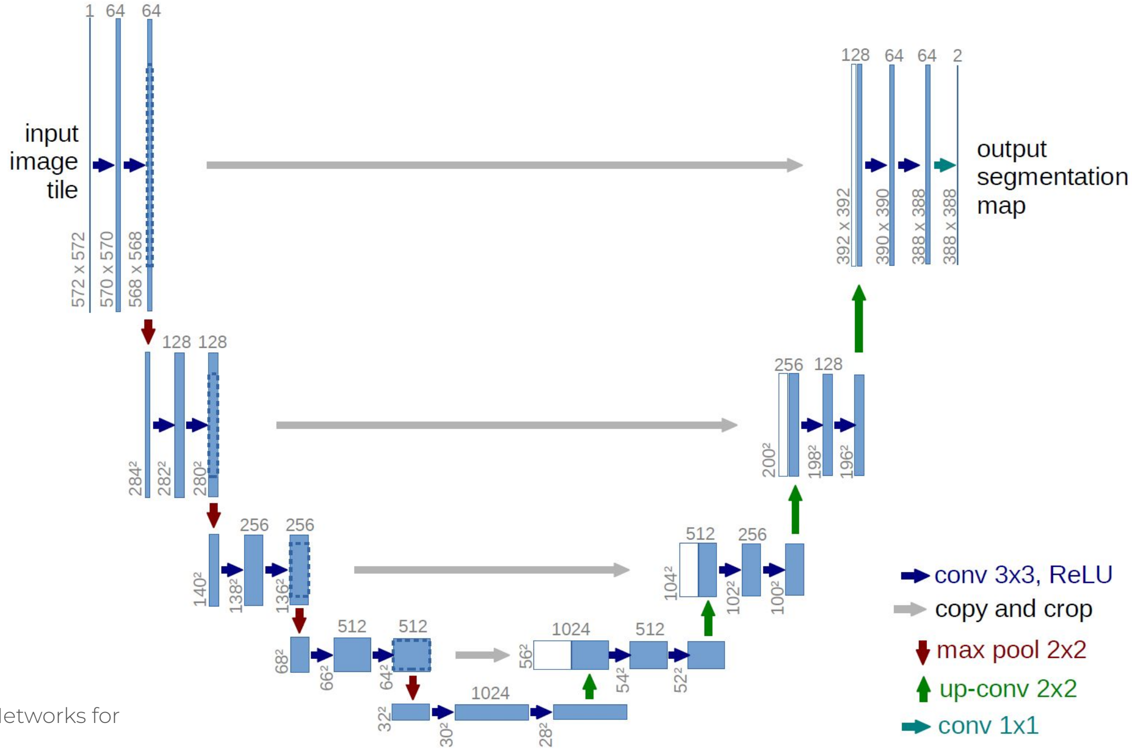


[8] Ho et al. (2020) Denoising Diffusion Probabilistic Models, *NeurIPS*.

Diffusion

The typical architecture:

UNets

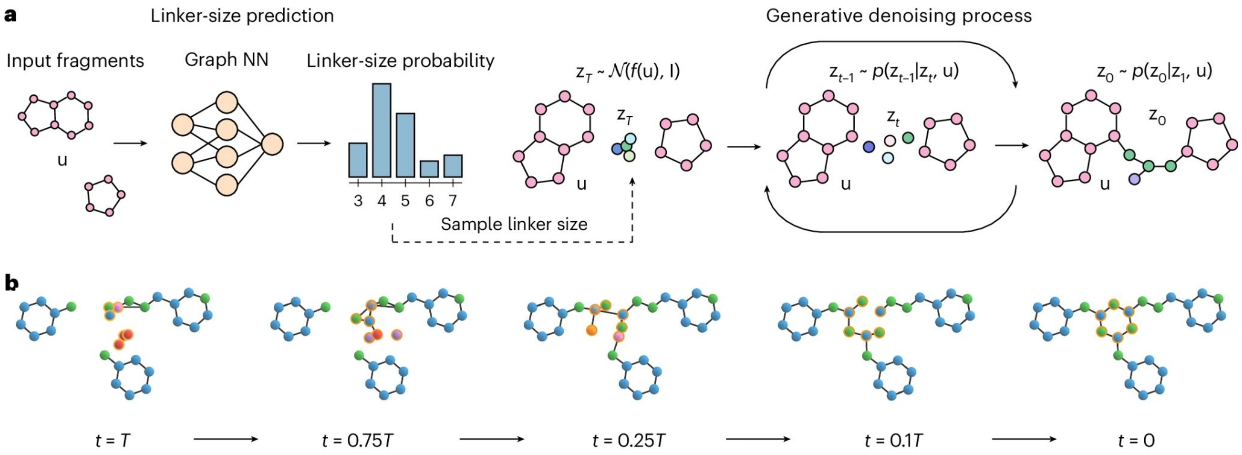


[9] Ronneberger et al. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation, *MICCAI*.

Diffusion in Biology

Applications:

- Protein and **molecule design**: TopoDiff, **DiffLinker**, …
- Protein-ligand interaction modeling: DiffDock, …
- …

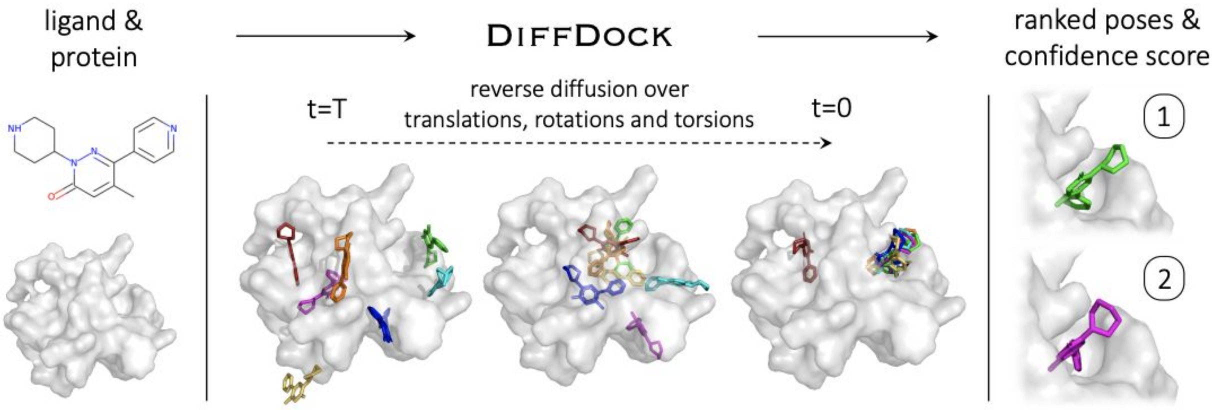


[10] Igashov et al. (2024) Equivariant 3D-conditional diffusion model for molecular linker design,, *Nature Machine Intelligence*.

Diffusion in Biology

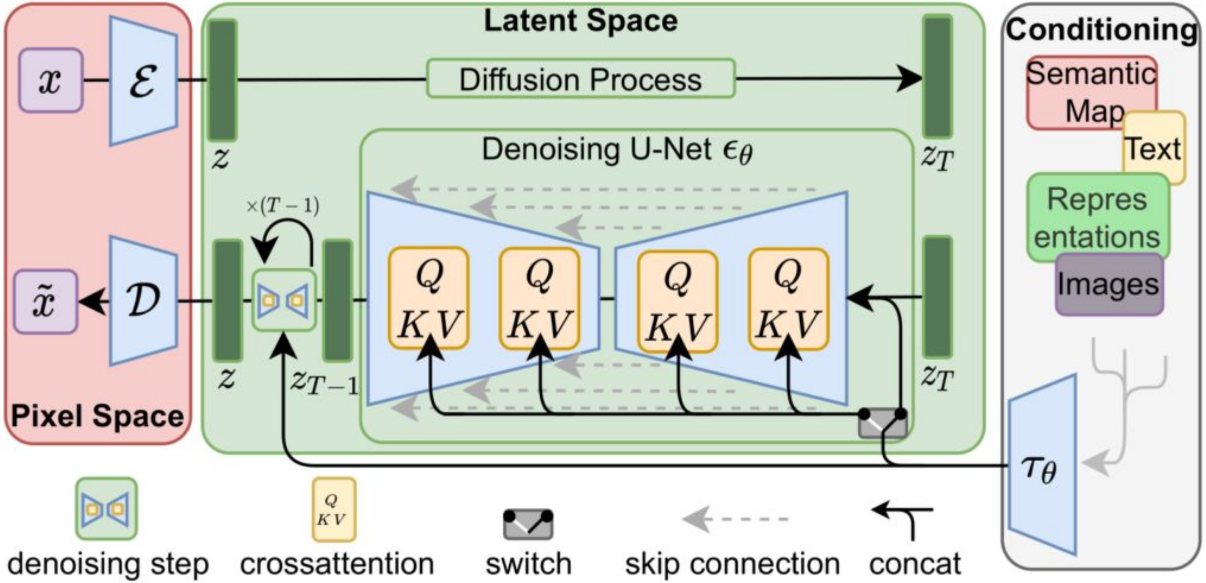
Applications:

- Protein and molecule design: TopoDiff, DiffLinker, …
- **Protein-ligand interaction** modeling: **DiffDock**, …
- …



[11] Corso et al. (2023) DIFFDOCK: Diffusion Steps, Twists, and Turns for Molecular Docking, *ICLR*.

Latent Diffusion



Rombach et al. High-Resolution Image Synthesis with Latent Diffusion Models. CVPR 2022.

Big generative AI

The **downside for us**:

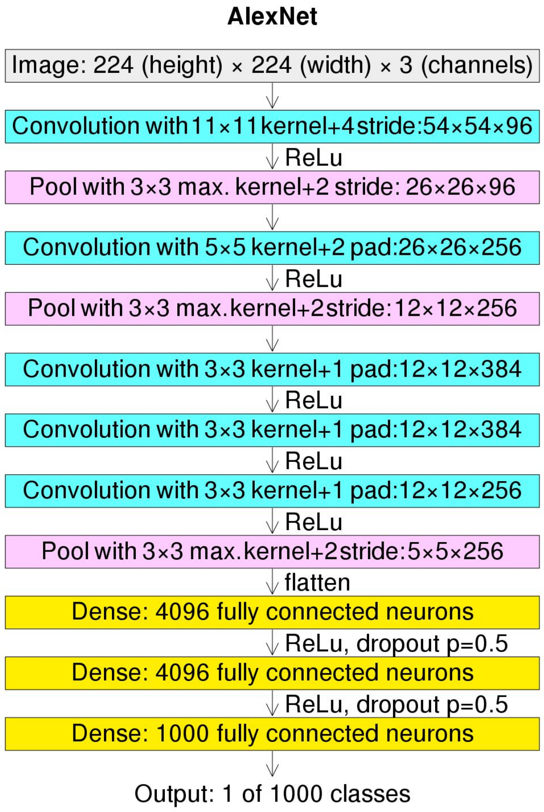
These big models need

a ton of **compute**!!!

AlexNet (2015):

62.3 M parameters

trained on 2 GPUs



https://en.wikipedia.org/wiki/AlexNet

Big generative AI

The **downside for us**:

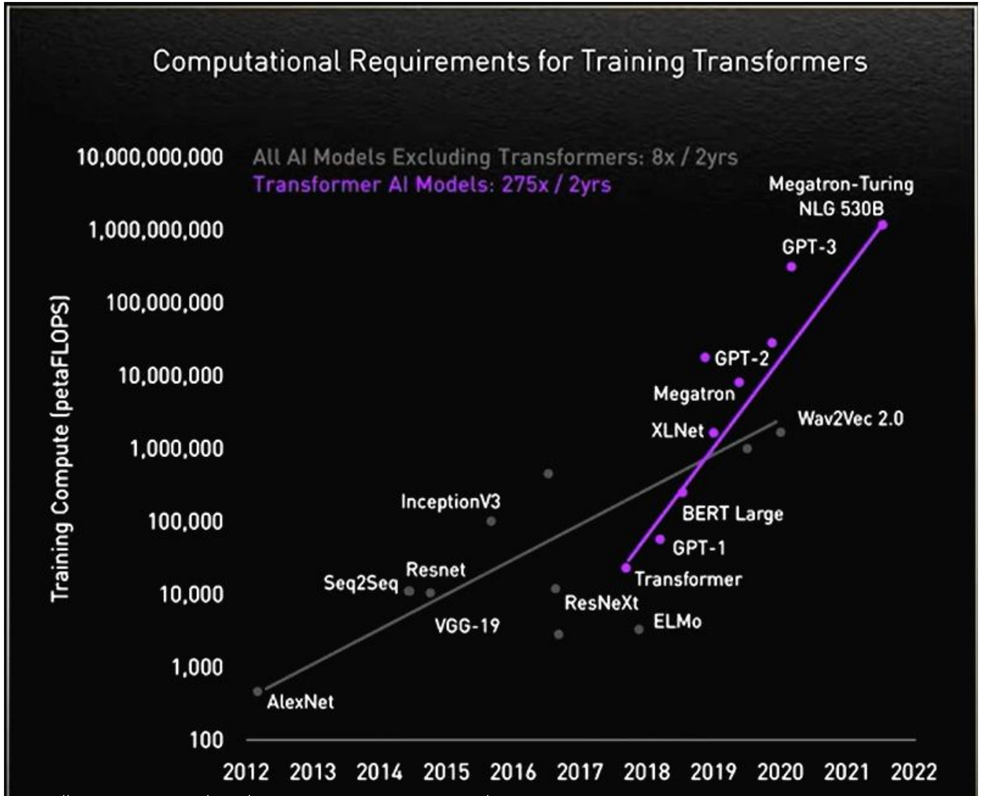
These big models need

a ton of **compute**!!!

Llama3 (2024):

8 B - 400 B parameters

trained on 16K H100 GPUs



https://blogs.nvidia.com/blog/what-is-a-transformer-model/