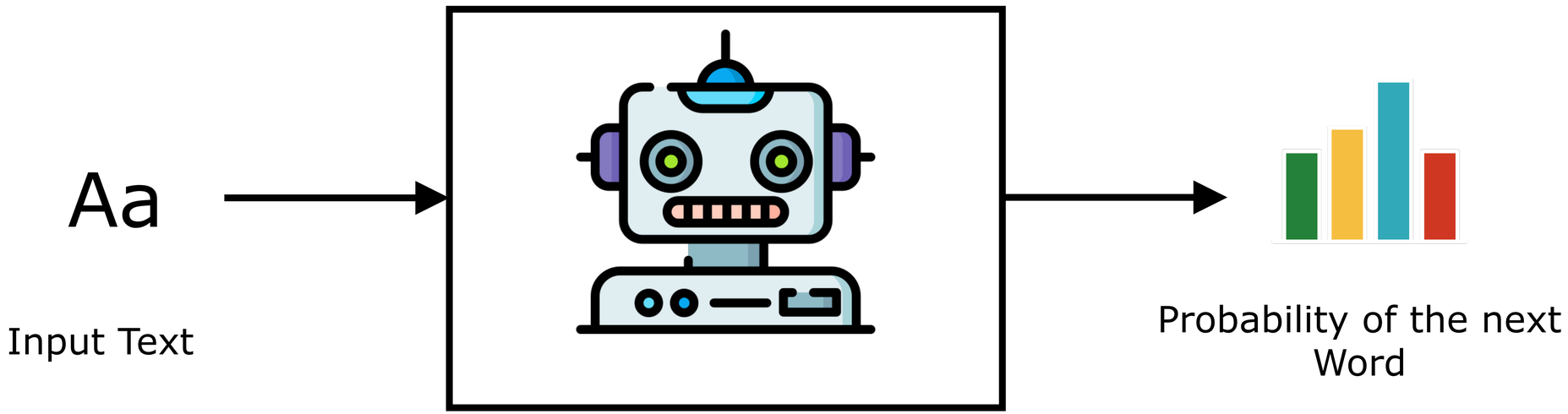
HSLU Hochschule Luzern

Current Topics in Generative Artificial Intelligence

Informatik
11. November 2023

FH Zentralschweiz

What is a Language Model?

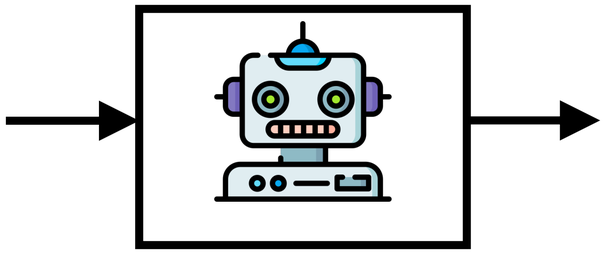


Bildquelle: Flaticon

Generate Text with a Language Model

<Q>Do white hens lay white eggs</Q>

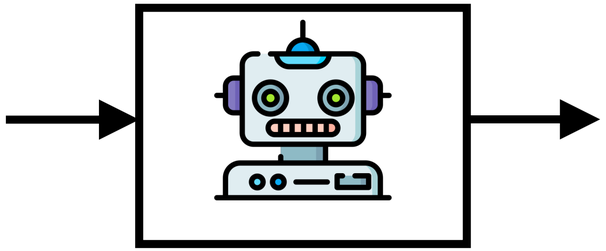
<A>



No 0.80
Yes 0.10
The 0.05
…

<Q>Do white hens lay white eggs</Q>

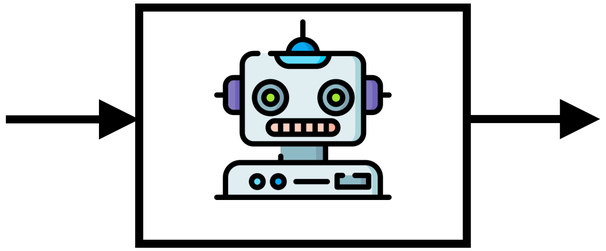
<A>No



the 0.60
hens 0.15
eggs 0.15
…

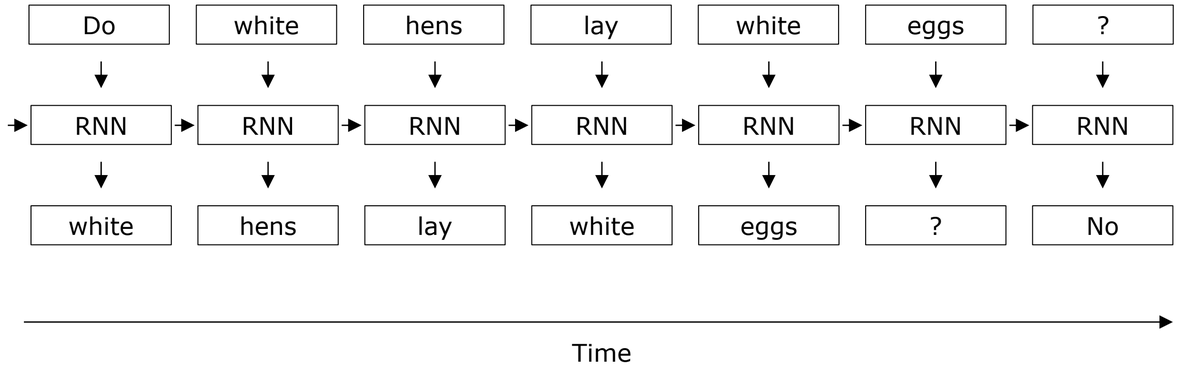
<Q>Do white hens lay white eggs</Q>

<A>No the



color 0.72
feather 0.21
farmer 0.09
…

Recurrent Neural Networks (RNNs)

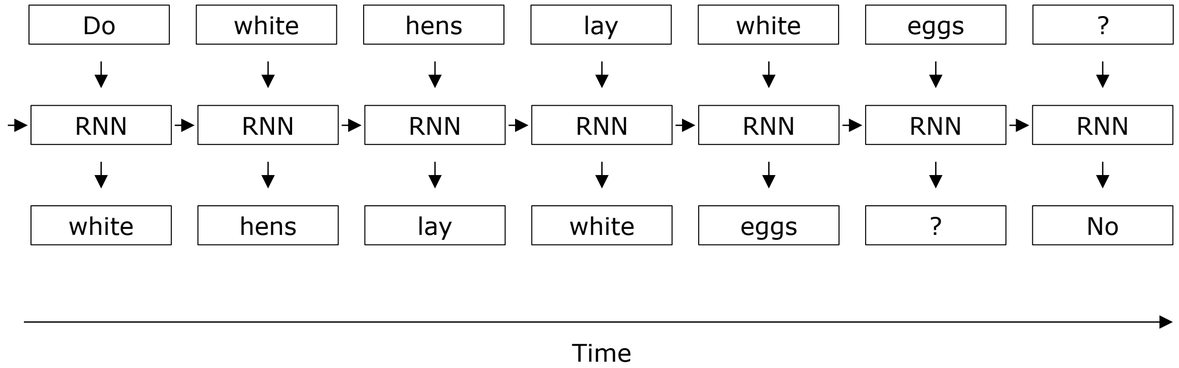


Idea: Train RNN to iteratively predict next word

Problem: Training a RNN is challenging.

- RNN is “unrolled” and trained by “backpropagation through time”
- Gradient is vanishing or exploding → training is unstable

Recurrent Neural Networks (RNNs)



Idea: Train RNN to iteratively predict next word

Problem: Training a RNN is challenging.

- RNN is “unrolled” and trained by “backpropagation through time”
- Gradient is vanishing or exploding → training is unstable

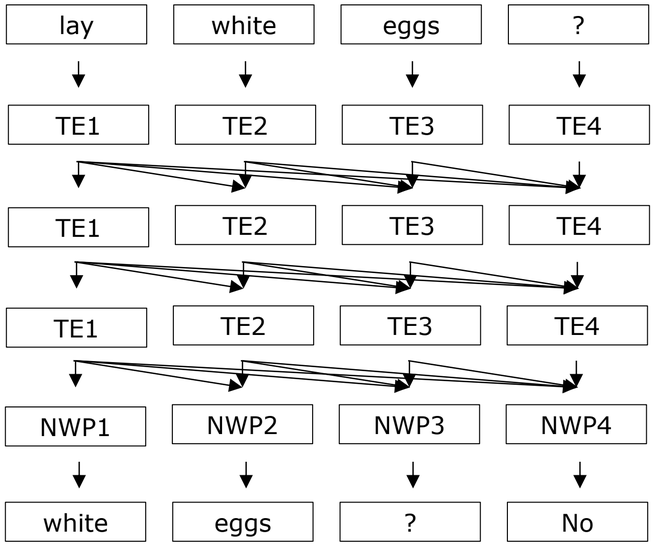
Transformers («Attention is All You Need», Vaswani et al. 2017)
GPT: Generative Pre-trained Transformer (Radford et al. 2018)

TE: Token embeddings

Note: We assume that tokens corresponds to words. **In practice, tokens are sub-words.**

NWP: Next word probability

Sampling



For each token:
- Transform into “key” K(TEn)
- Transform into “value” V(TEn)
- Transform into “query” Q(TEn)

Compute next token with self-attention:
Softmax(Q(TEn) x K(TE1)) x V(TE1) +
Softmax(Q(TEn) x K(TE2)) x V(TE2) +
… +
Softmax(Q(TEn) x K(TE4)) x V(TE4)

The softmax(Q(TEi) x K(Tej)) self-attention!

Analogy:
Imagine a “fuzzy” look-up table (Python dict):

```
dict[”purple”] = (
   similarity(“purple”, “red”) x dict[”red”] +
   similarity(“purple”, ”blue”) x dict[“blue”]
)
```

Training

RNN:
- **To predict n-th word, you HAVE to run the RNN n times**
  f(f(f(f(f(…), “lay”), “white”), “eggs”), “?”) → “No”

Transformer:
- **To predict n-th word, you run the transformer once**
  with words 1,2, .., n-1 as input
  f(“lay white eggs?”) → “No”
- **You can predict / train on words 1,2, .., in parallel!**

Inference

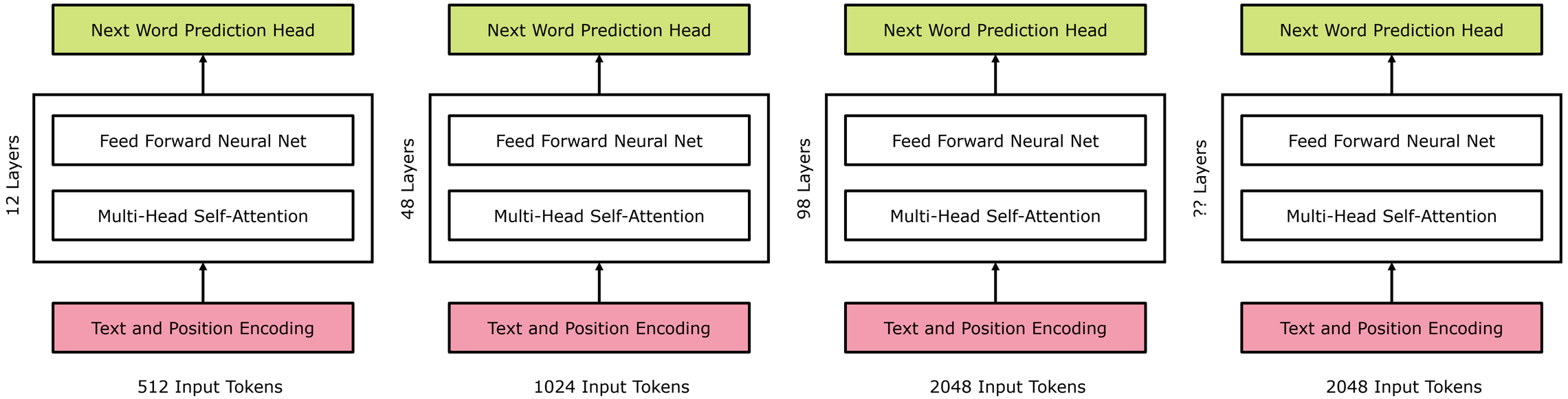
Both RNNs and transformers generate one word at a time.
Sentences are generated iteratively.

ChatGPT

ChatGPT is fine-tuned from GPT-3.5, a language model trained to produce text. ChatGPT was optimized for dialogue by using Reinforcement Learning with Human Feedback – a method that uses human demonstrations to guide the model toward desired behavior

ChatGPT General FAQ by OpenAI

GPT Architectures



GPT-1 from 2018
117 Million Parameters
General Purpose Model

GPT-2 from 2019
1.5 Billion Parameters
General Purpose Model

GPT-3 from 2020
175 Billion Parameters
General Purpose Model

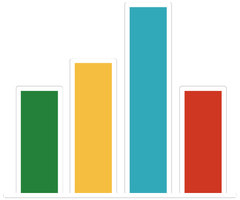
GPT-3.5 from 2022
we don't really know
Dialog Model

Architectures differ in skip connections and normalization layers, GPT-3 alternates dense and sparse attention

GPT-4 with over a trillion parameters is rumored to be coming out in 2023

Unsupervised Pre-Training of GPT Models

lecturer



Artificial Intelligence

Maybe	he	is	not	the	best	[MASK]

Maybe	he	is	not	the	best	lecturer	but	at	least	he	has	good	slides

What’s inside an LLM?

LLaMA: Open and Efficient Foundation Language Models

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet
Marie-Anne Lachaux, Timothee Lacroix, Baptiste Rozière, Naman Goyal
Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin
Edouard Grave, Guillaume Lample

Meta AI

Model	Size	Training data
LLaMA (base model)	7B, 13B, 33B, 65B	Various
Alpaca	7B, 13B	52k GPT-3 instructions
Vicuna	7B, 13B	70k ChatGPT conversations
Koala-distill	7B, 13B	117k cleaned ChatGPT conversations
GPT4-x-Alpaca	13B	20k GPT4 instructions
WizardML	7B	70k instructions synthesized with ChatGPT/GPT-3
OpenAssistant LLaMA	13B, 30B	600k human interactions (OpenAssistant Conversations)

Dataset	Sampling prop.	Epochs	Disk size
CommonCrawl	67.0%	1.10	3.3 TB
C4	15.0%	1.06	783 GB
Github	4.5%	0.64	328 GB
Wikipedia	4.5%	2.45	83 GB
Books	4.5%	2.23	85 GB
ArXiv	2.5%	1.06	92 GB
StackExchange	2.0%	1.03	78 GB

Could we train a GPT Language Model at HSLU?

Performance
GPT-3 from 2020

Learning Capacity
175 Billion Parameters

Data Hunger
45 TB Text Data

Pre-Training Cost *
4.6M $ | 85’000 kg CO_2

Fine-tuning is, however, possible ☺

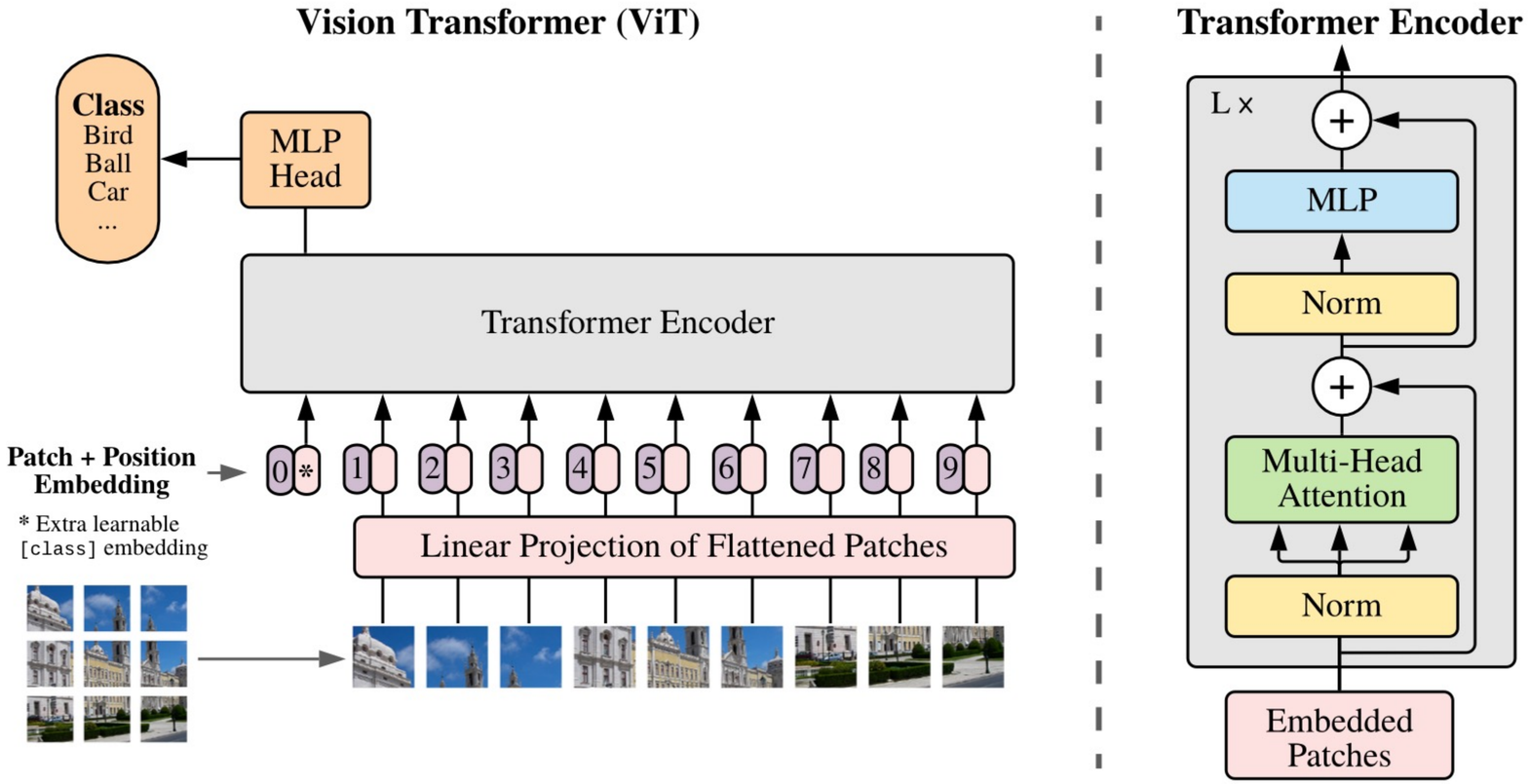
Numbers from 2020 on Tesla V-100, lambdalabs.com

Rocket launches generate between 50-75t of CO_2 per passenger

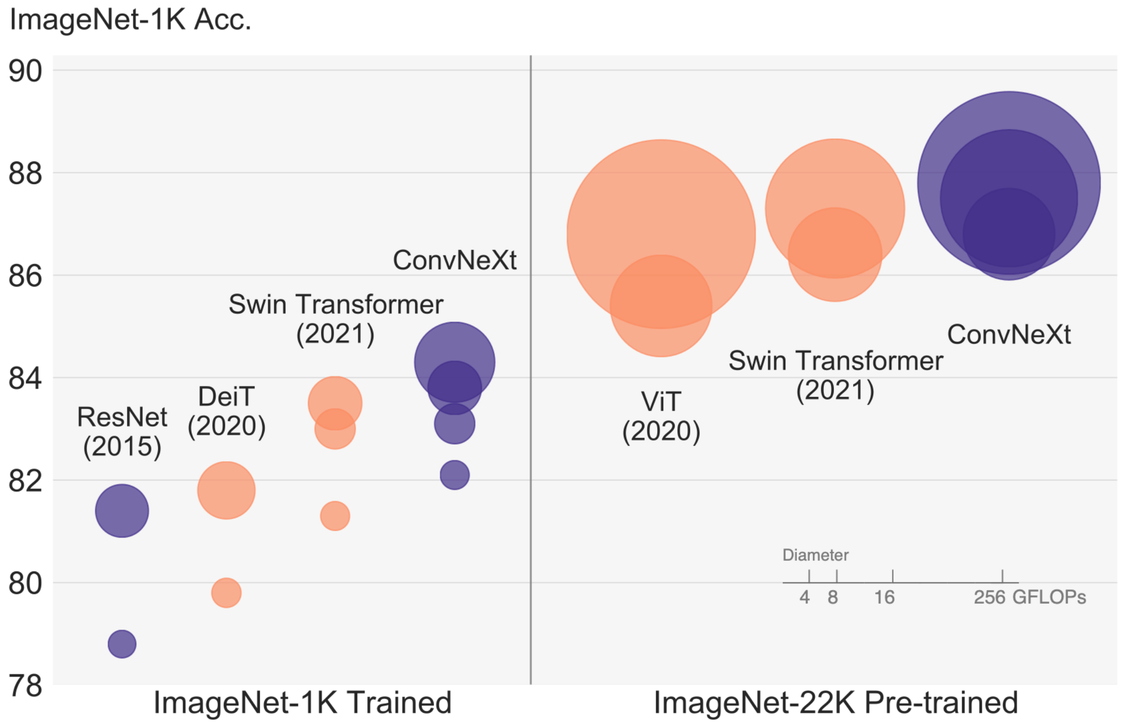
Bildquelle: The Franklin Institute

Transformers for Computer Vision Applications

Vision Transformers (ViT; Dosovitskiy et al. 2021)



ConvNeXt (Liu et al. 2022) – a modernized ResNet



While Transformers clearly outperform competing methods for natural language processing (NLP), modern convolutional neural networks (CNNs) are competitive with / slightly outperform vision transformers.

Generating Images from Text

Generative Adversial Networks (GAN)

- Deep learning architecture that allows to generate outputs similar to the data in the training set (Goodfellow et al. 2014)

- Consist of a generator and a discriminator
- The generator is trained to generate new images
- The discriminator is trained to distinguish between real and fake images
- The model is trained until a local nash equilibrium is reached: no player can improve without changing the parameters of the other party
- Allow to generate new images that are similar to the data in the training set just from noise

Dataset: A huge set of unlabeled images

Diffusion Models

- Similar to GANs, diffusion models allow to generate new data that is similar to the training data
- They are trained by adding noise to an image (noising process) and then reconstructing the image from the noise (denoising)

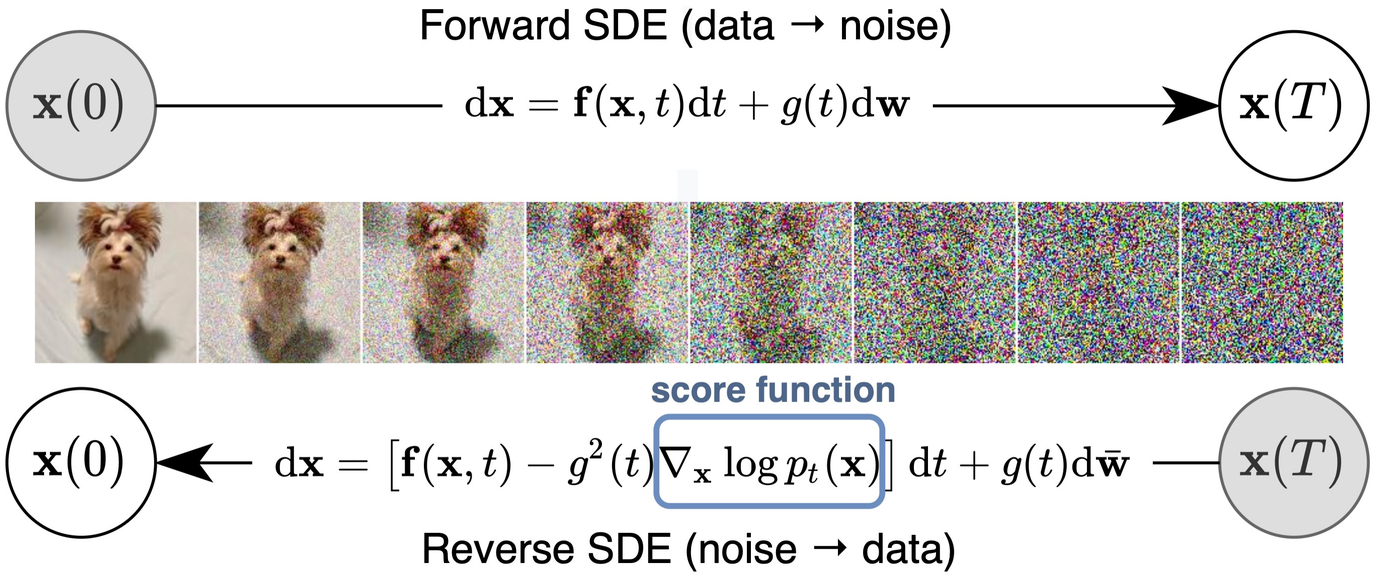


Image Source: https://www.assemblyai.com/blog/how-imagen-actually-works/

Text-to-Image Models

During the de-noising process, the decoder is conditioned on the input prompt

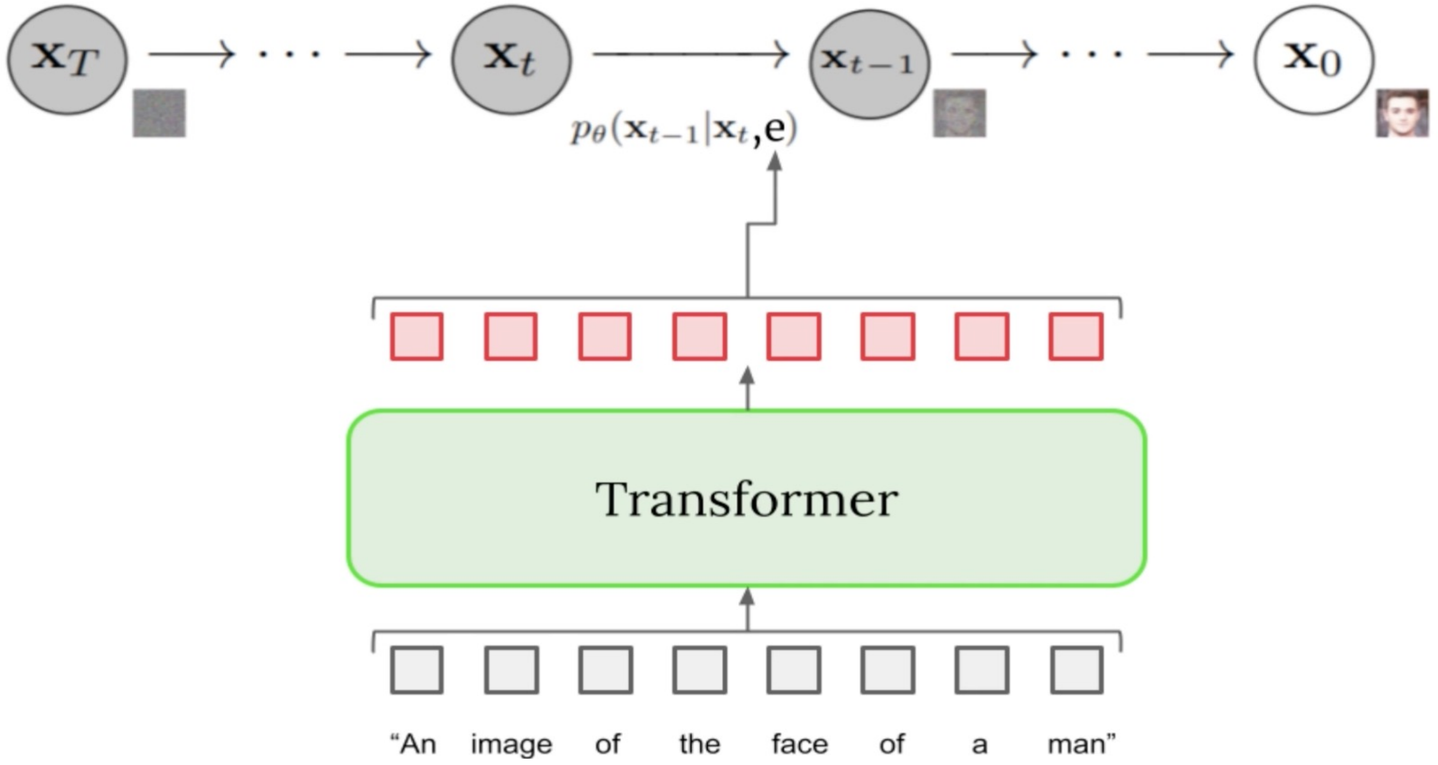


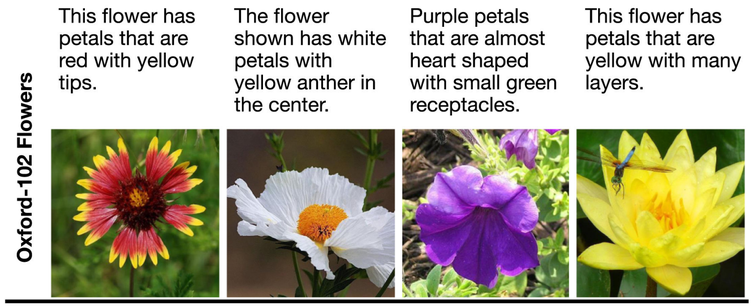
Image Source: https://www.assemblyai.com/blog/how-imagen-actually-works/

Text-to-Image Models

- Instead of generating new images from noise, they use a *textual input* to guide the generation process

- Typically consists of an encoder model and a decoder model
- The *encoder model* transforms the input text into a representation (e.g., using Transformer)
- The *decoder model* generates the image conditioned by the encoded input (e.g., using a GAN or diffusion model)

Dataset: Images paired with a caption



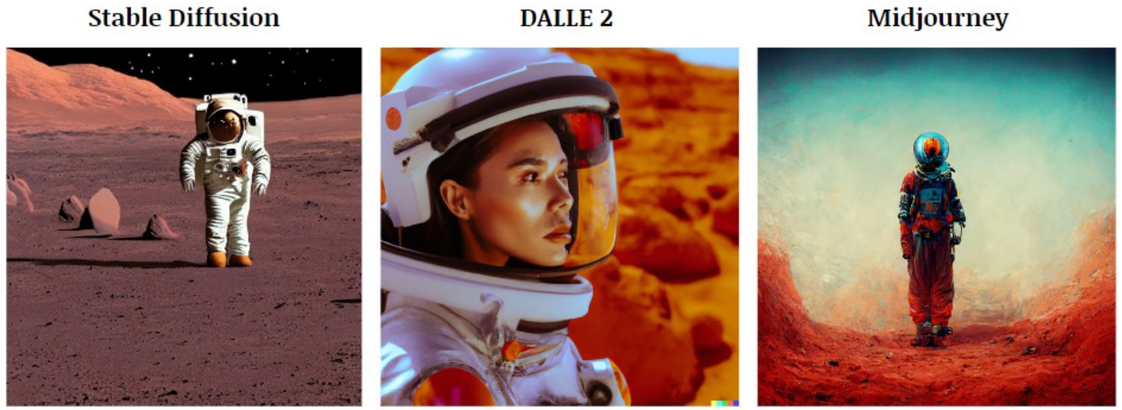
Mansimov et al. 2015

A stop sign is flying in blue skies.

A herd of elephants flying in the blue skies.

A toilet seat sits open in the grass field.

Examples from 2022



Alone astronaut on Mars, mysterious, colorful, hyper realistic

Text-to-Image Models

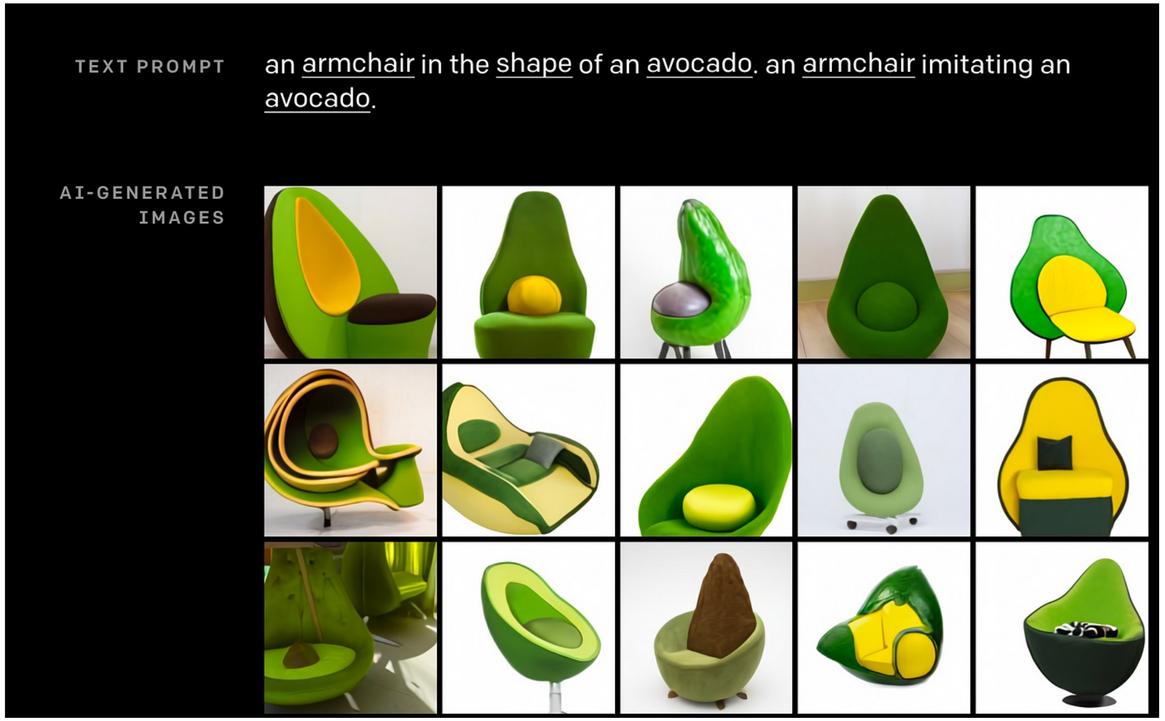
Furniture GAN (2019) @ HSLU

Generate furniture based on short descriptions

Input Text	Generated Samples
Stuhl aus Holz Birke	
Tisch	
Bücherregal BILLY in schwarz	
Stuhl weiß hell	
Stuhl schwarz dunkel	
Kleiderschrank PAX mit Türen aus Holz	

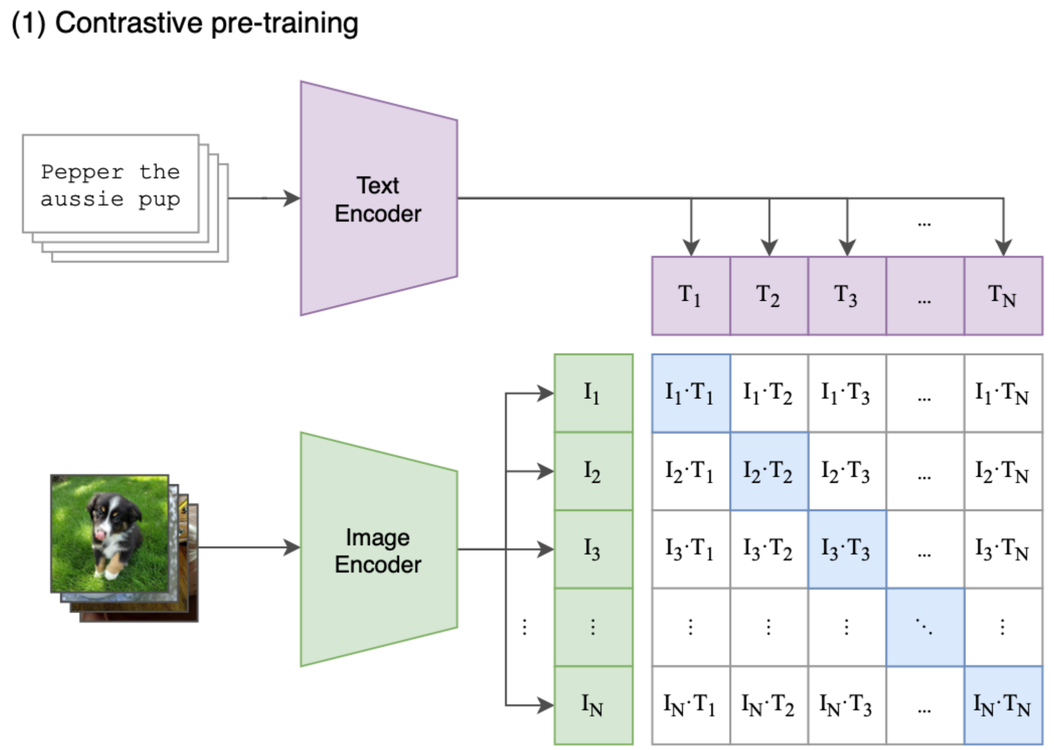
DALL-E 1 (2021) @ OpenAI

Generate images based on text



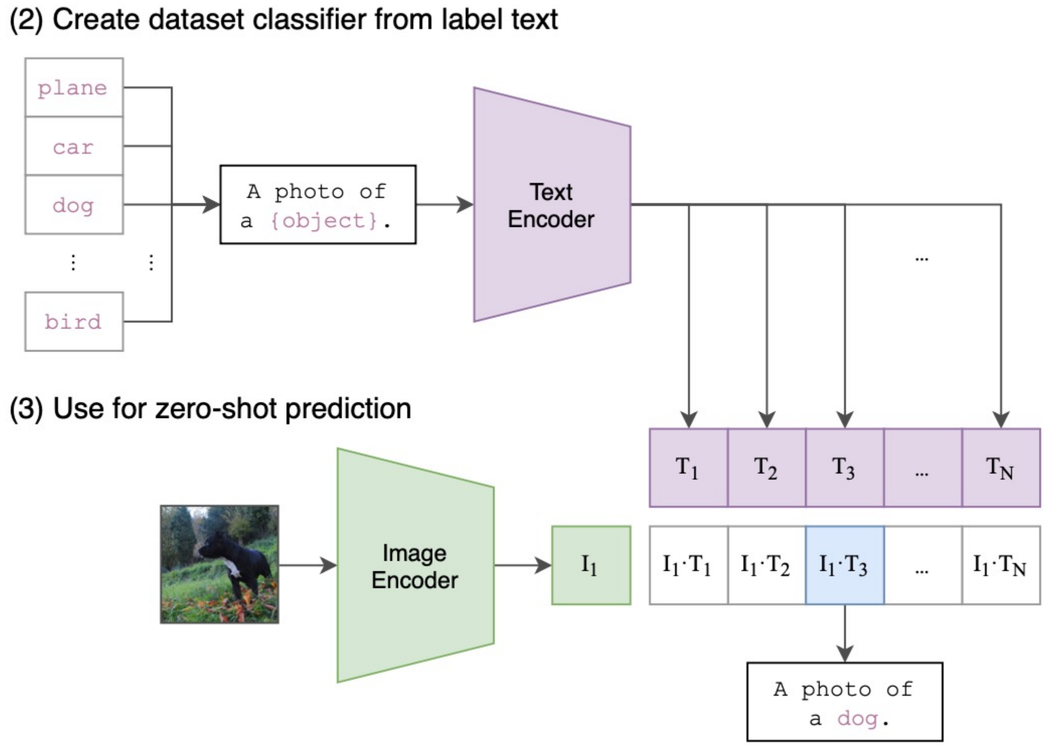
https://www.linkedin.com/pulse/how-develop-new-products-artificial-intelligence-kevin-kuhn

Learning Visual Models from Natural Language (Radford et al. 2021)



Use (image, description in natural language) pairs to train an image model
Much easier to scrape such a dataset than to generate (image, label) pairs

Learning Visual Models from Natural Language (Radford et al. 2021)



You can calculate probabilities for arbitrary classes!

Image-to-Text Models @ HSLU

2020:Deep Neural Inverse Cooking
https://cooking.abiz.ch/

- Spaghetti
- Olive oil
- Garlic
- Tomatoes
- Basil
- ...

Remove the stem from the tomatoes and make a cross incision on the bottom. Immerse them for a few seconds in boiling water, then place in iced water to stop the cooking process. Peel them and remove the seeds …

Deep Neural Inverse Cooking - Motivation

- Pigment network
- Negative network
- Streaks
- Milia-like cyst
- Globules
- ...

Benign pigmented macule (solar lentigo) appearing on fair-skinned individuals that related to ultraviolet radiation exposure, typically from the sun. Must be distinguished from …

Sketch of an Image-to-Text Model

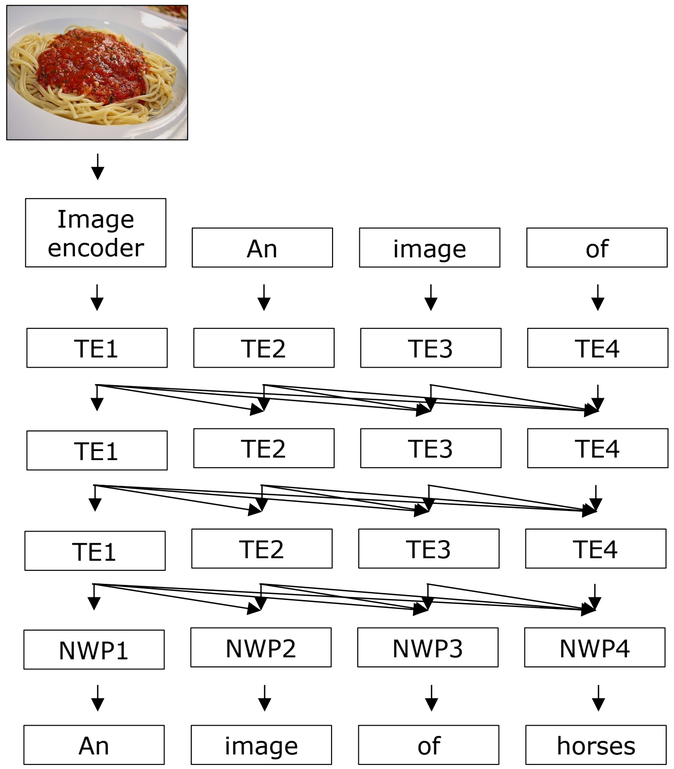


Image-to-Code