

# Image-generative AI

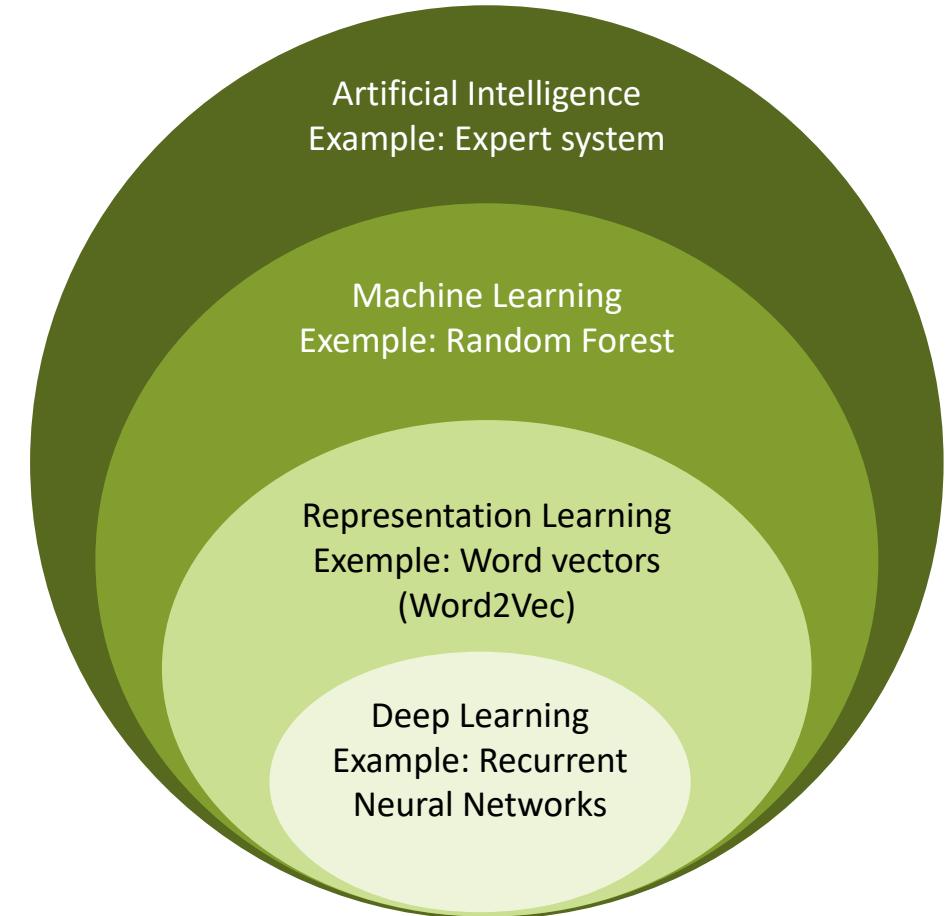
Erik Ylipää

AI Support - AIDA Data-hub

National Bioinformatics Infrastructure Sweden

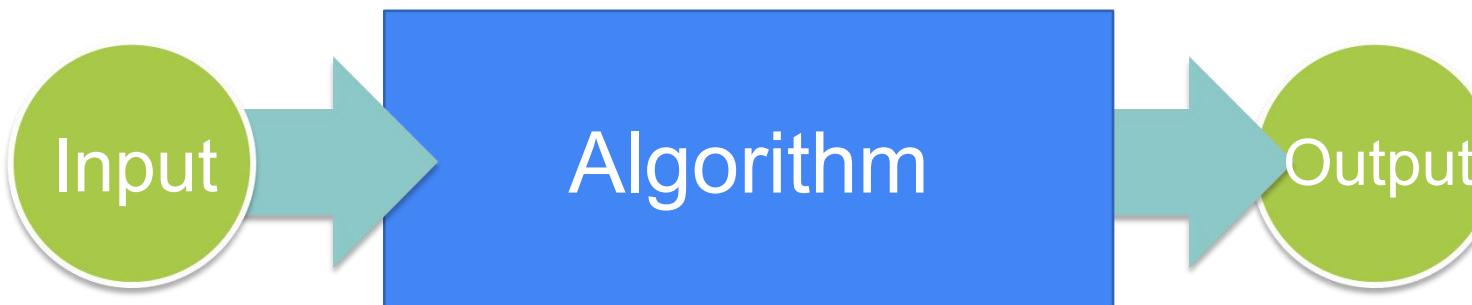
# Artificial Intelligence (AI)

- AI studies machines which can solve problems which humans solve using intelligence (whatever that is)
- Not primarily about machines that learn and think like humans
- AI we hear about today is most often based on ***machine learning***



Bengio, Yoshua, Ian Goodfellow, and Aaron Courville.  
Deep learning. Vol. 1. MIT press, 2017.  
<https://www.deeplearningbook.org/>

# AI-systems are algorithmic problem solvers



$$\begin{array}{r} 0 \\ 7 \overline{) 452} \\ 0 \downarrow \\ 45 \end{array}$$
$$\begin{array}{r} 06 \\ 7 \overline{) 452} \\ 0 \downarrow \\ 45 \\ -42 \\ \hline 32 \end{array}$$
$$\begin{array}{r} 064 \\ 7 \overline{) 452} \\ 0 \downarrow \\ 45 \\ -42 \\ \hline 32 \\ -28 \\ \hline 4 \end{array}$$

(64 r4)

Step 1: "How many times?"

Step 2: "Multiply"

Step 3: "Subtract"

Step 4: "Drop it down"

(repeat steps for each number, left to right)

# Learning algorithms



Rule based algorithm

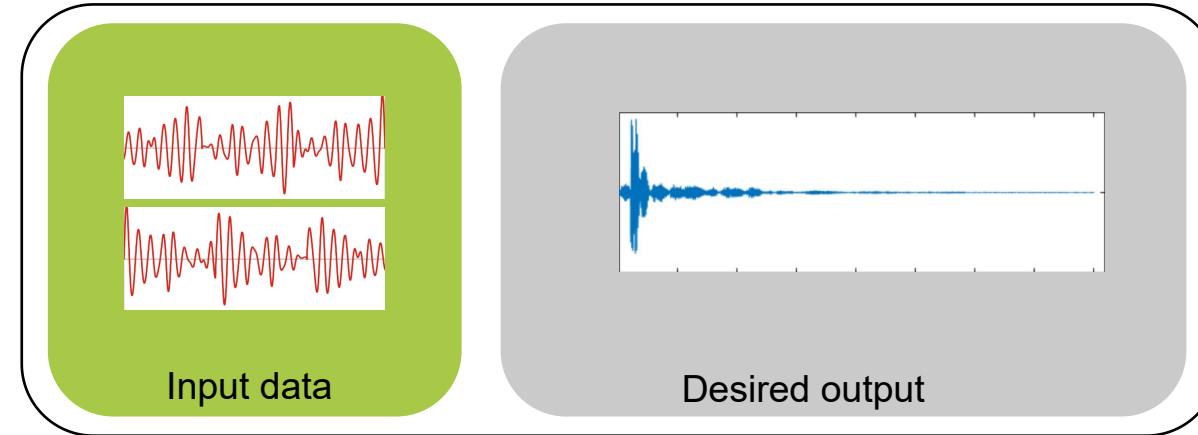
```
def fibonacci(n):
    terms = [0,1]
    i = 2
    while i<=n:
        terms.append(terms[i-1]+terms[i-2])
        i = i + 1
    return terms[n]
```

Learning algorithm



# Ingredients of learning algorithms

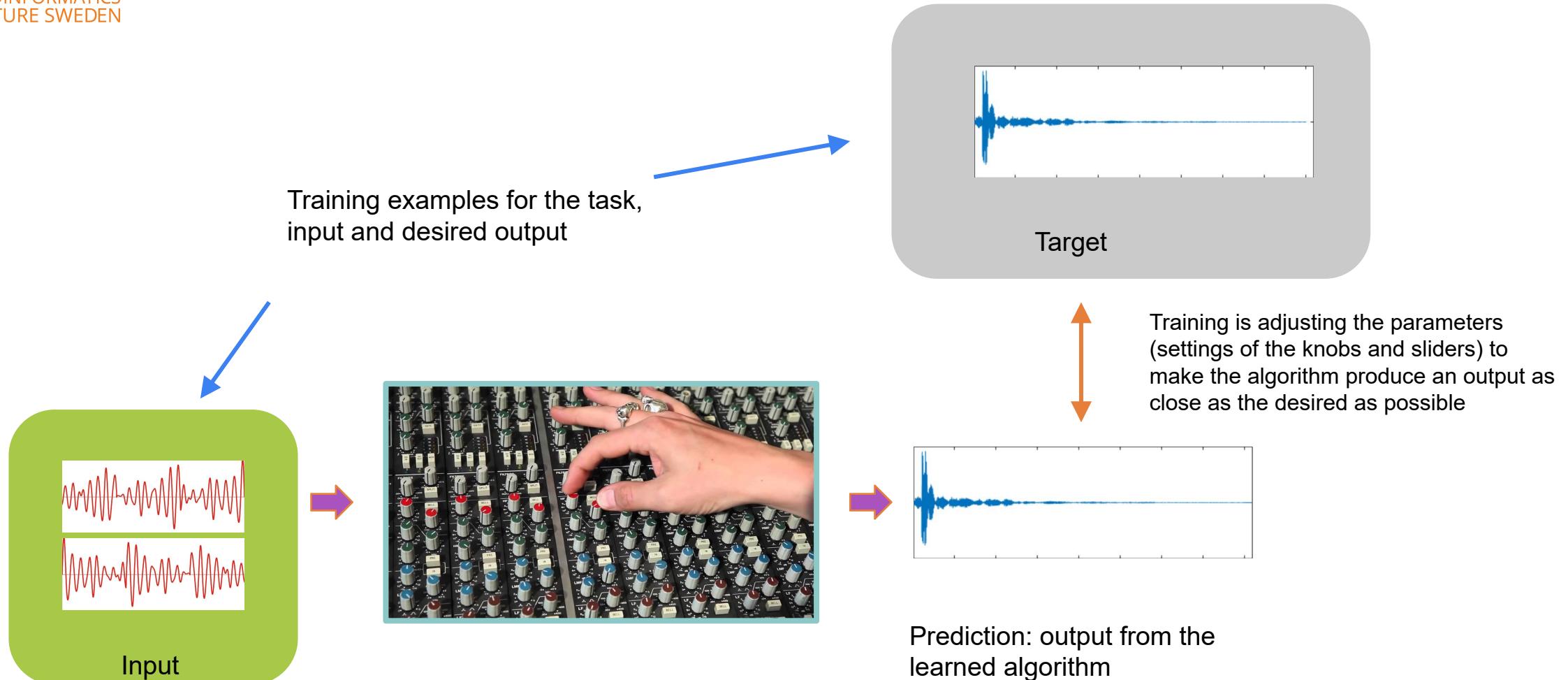
Data pair for training



Model with free parameters

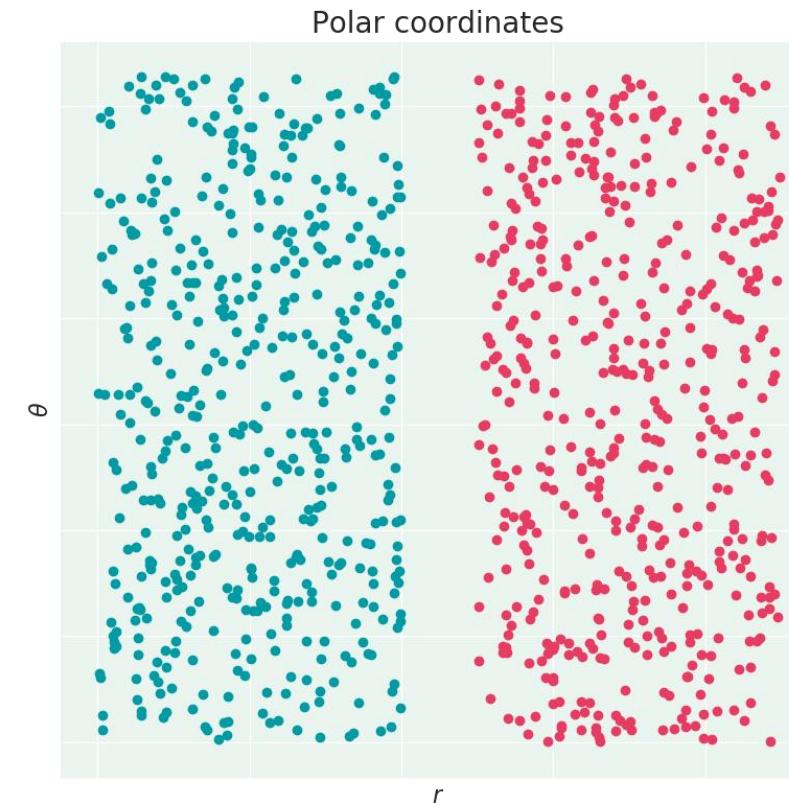
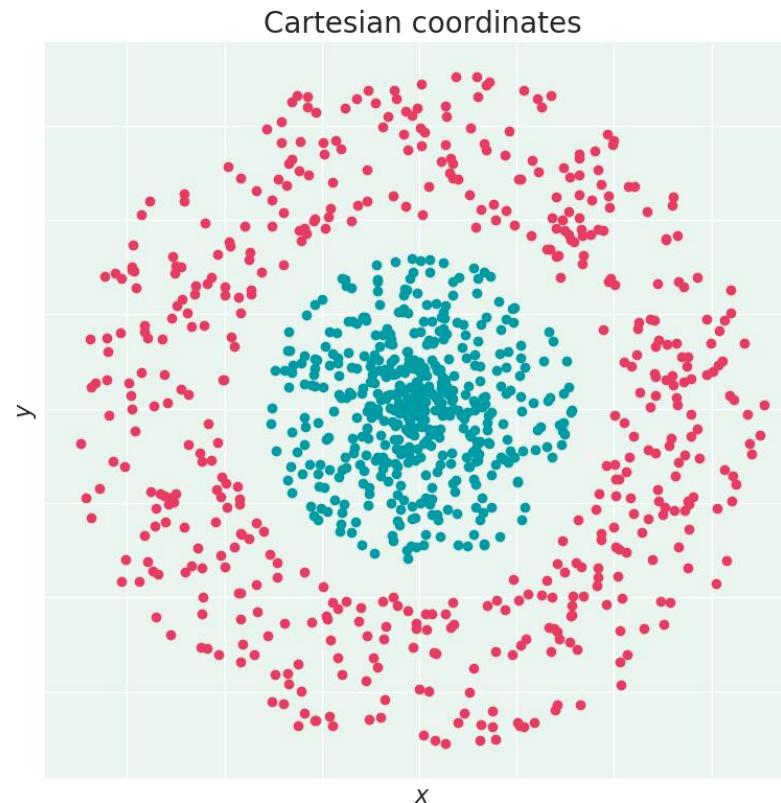


# Training a machine

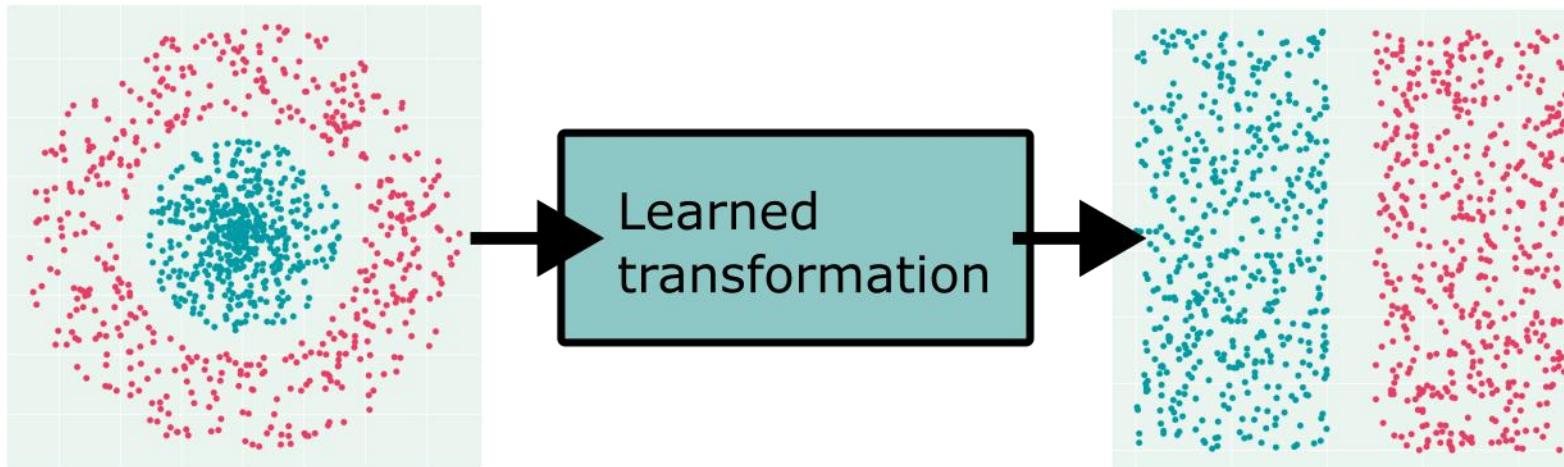


**How do we represent a problem/data to make automated decision making simple?**

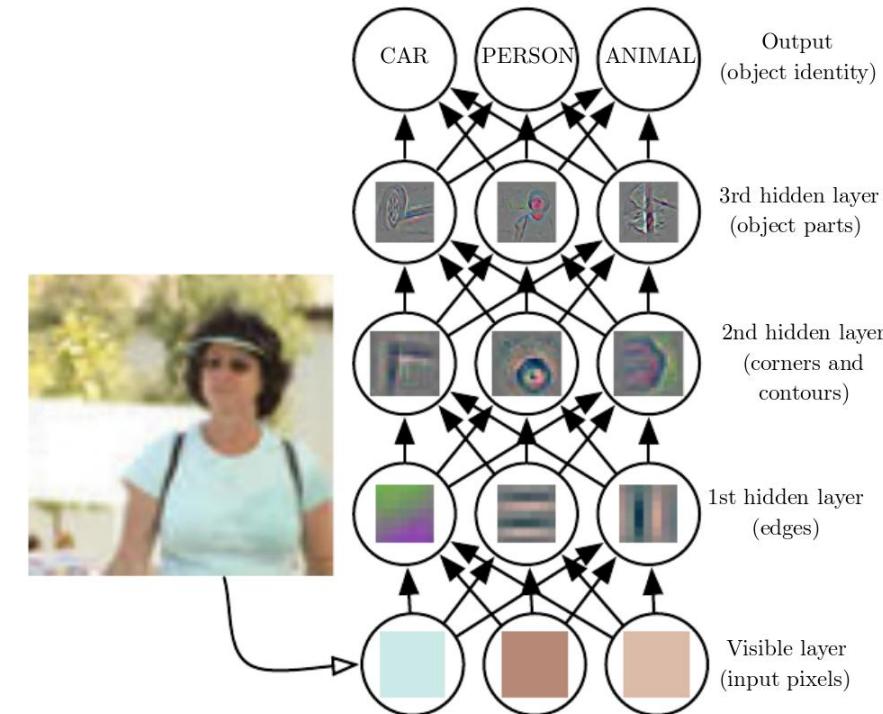
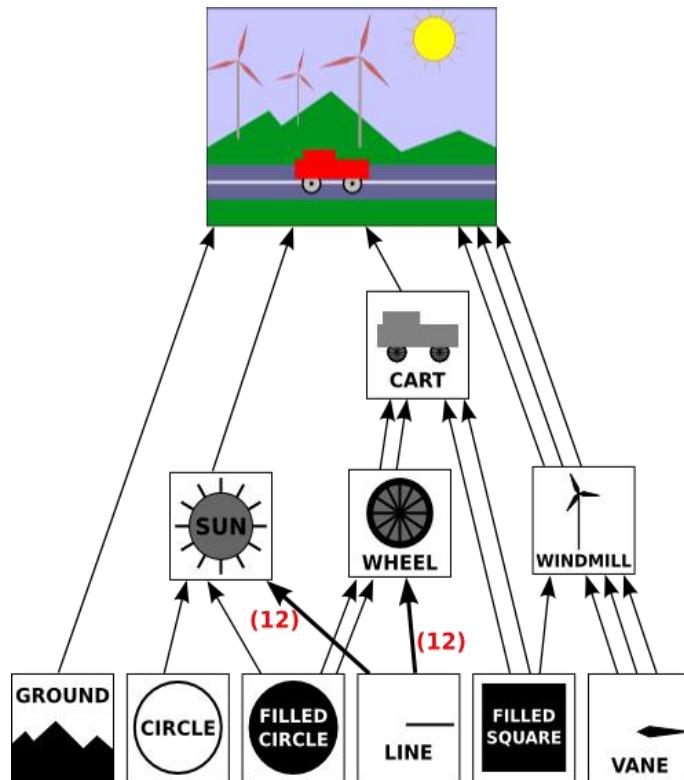
# How to represent data?



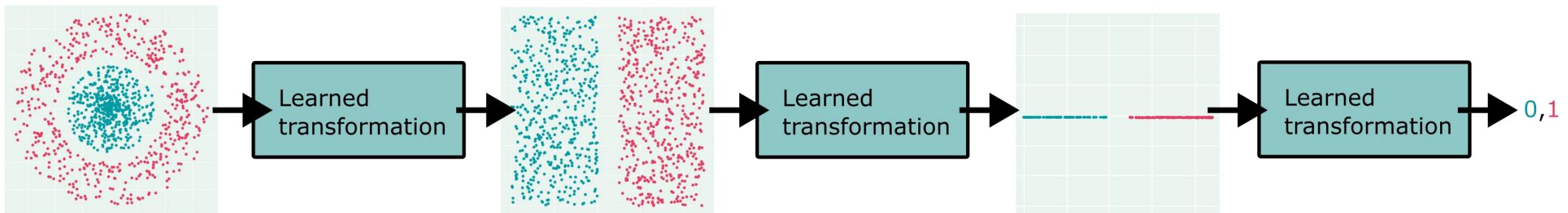
# Learn the representation



# Deep learning

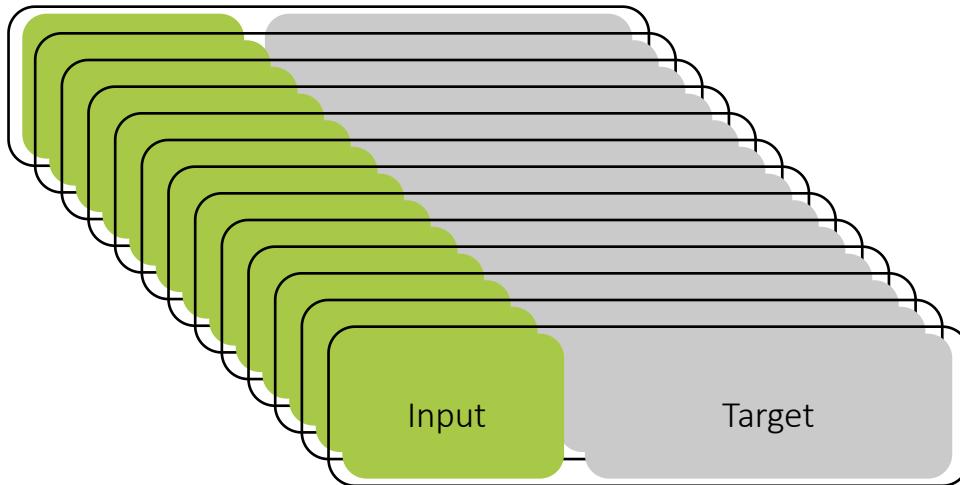


# Representation learning with deep neural networks



# Where do we get inputs and desired outputs which teaches the model good representations?

Training data



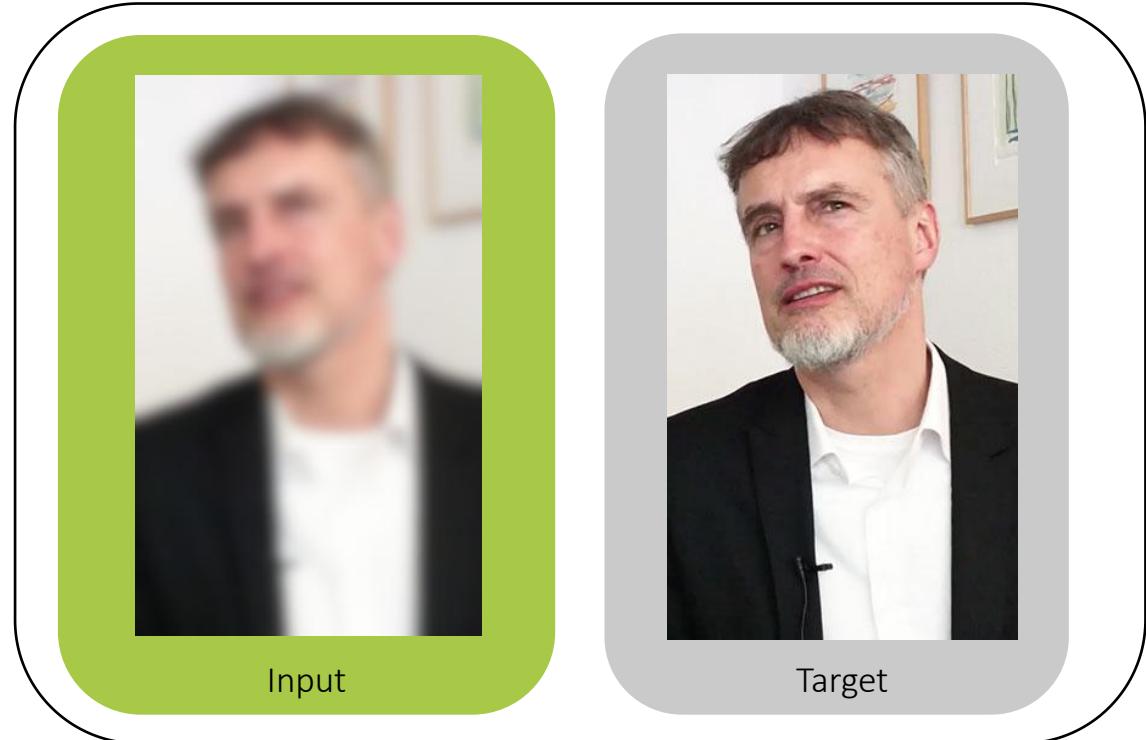
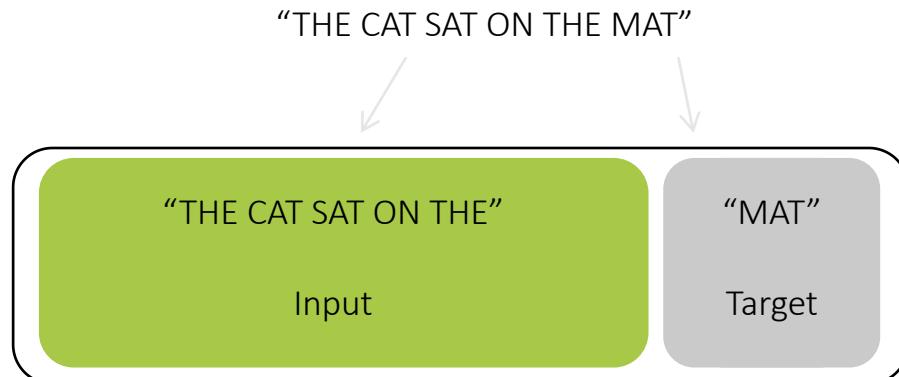
**Lots** of data to find parameters which works for many different inputs

Test data



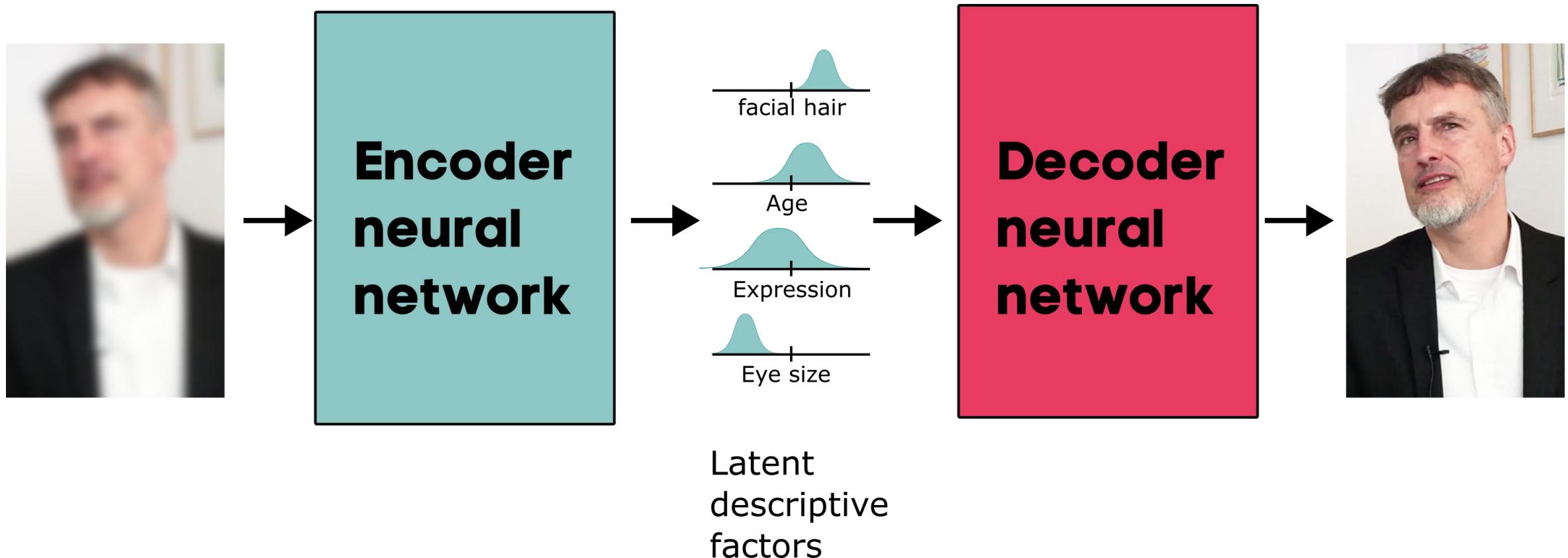
Enough test data to estimate performance on new data

# Self-supervised learning - generate a desired output based on the data *itself*

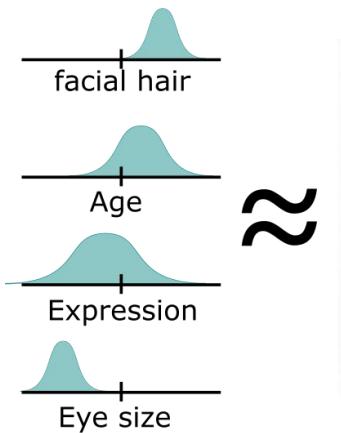


*Modify the data and predict the unmodified version*

# [Denoising] Auto Encoding



# Generative models: maps random values to complex data



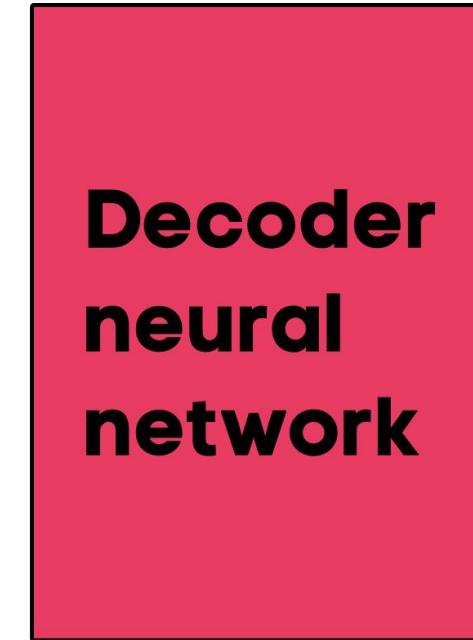
$\approx$



→ 2, 4, 6, X, 9 →  
→ 3, 1, 6, 2, 4 →

Random  
latent  
descriptive  
factors

Different  
random  
samples



→

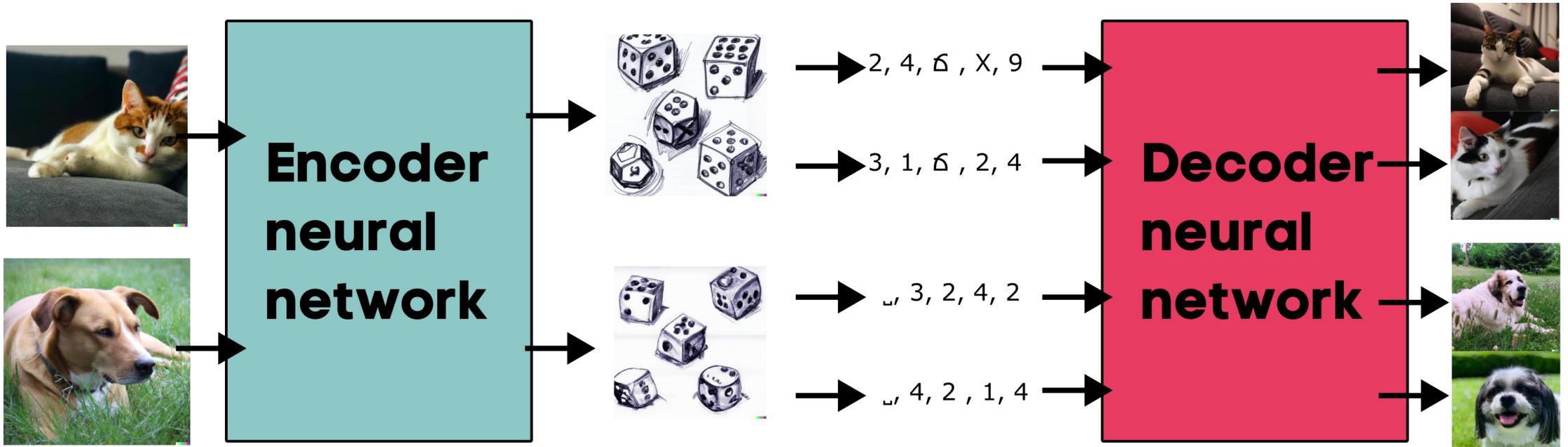


→



Different  
generated  
complex  
points

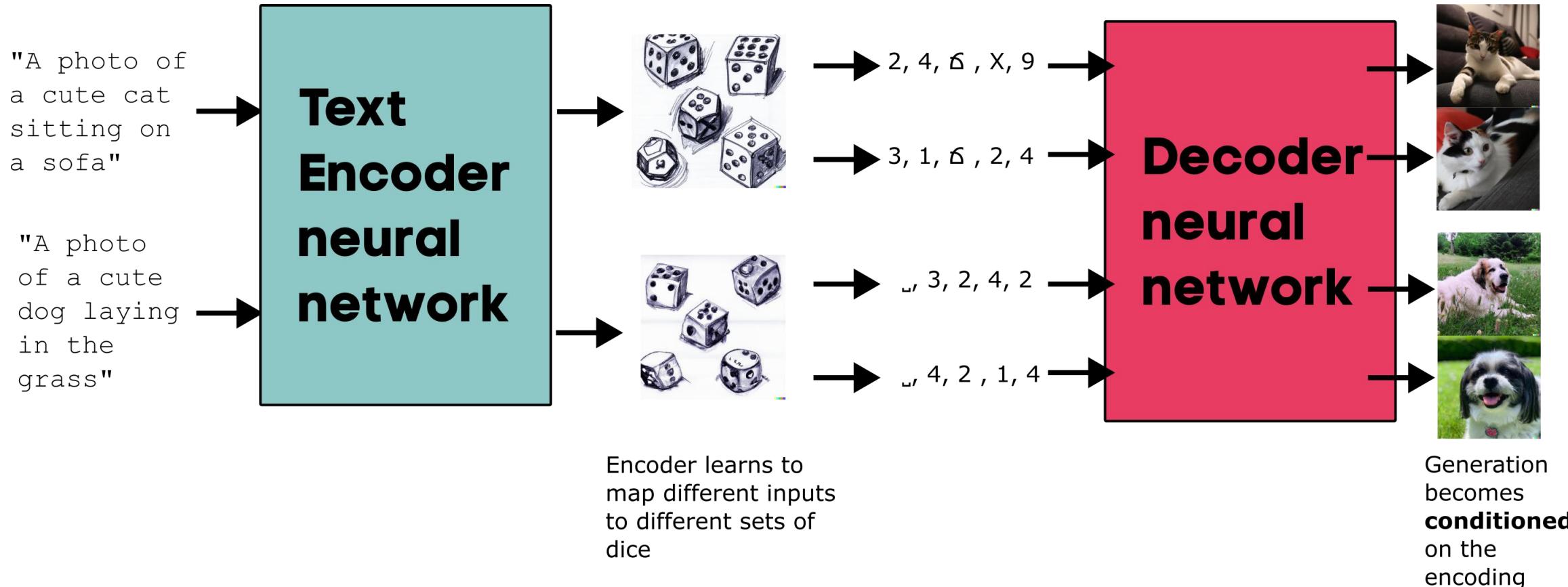
# Conditional generation



Encoder learns to map different inputs to different sets of dice

Generation becomes **conditioned** on the encoding

# DALL-E, Stable Diffusion, Midjourney, etc.



# Training data is text-image pairs

“Portrait photograph of the AI researcher Jürgen Schmidhuber”

Input A



Input B

“Photo of a bumblebee”

Input A



Input B

# Gathered from the internet



*Large-scale Artificial Intelligence Open Network*

TRULY OPEN AI. 100% NON-PROFIT. 100% FREE.

LAION, as a non-profit organization, provides datasets, tools and models to liberate machine learning research. By doing so, we encourage open public education and a more environment-friendly use of resources by reusing existing datasets and models.

[Join our community on discord!](#)

[Help building an Open Source ChatGPT!](#)

## LAION-400M

An open dataset containing 400 million English image-text pairs.

## LAION-5B

A dataset consisting of 5.85 billion multilingual CLIP-filtered image-text pairs.

## Clip H/14

The largest CLIP (Contrastive Language-Image Pre-training) vision transformer model.

## LAION-Aesthetics

A subset of LAION-5B filtered by a model trained to score aesthetically pleasing images.

# DALL-E 3

DALL-E 3 understands significantly more nuance and detail than our previous systems, allowing you to easily translate your ideas into exceptionally accurate images.

[Read research paper ↗](#)

© 2010 Pearson Education, Inc., publishing as Pearson Addison Wesley. All rights reserved.

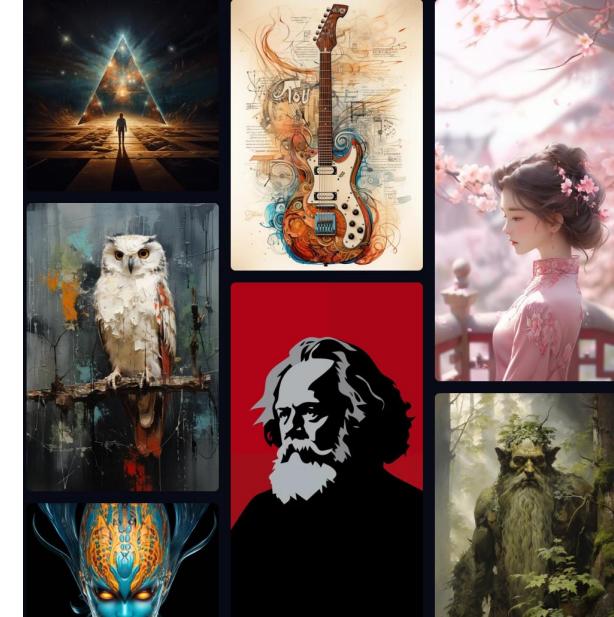
The sidewalks bustling with **pedestrians** enjoying the nightlife. —



At the corner stall, a **young woman** with fiery red hair, dressed in a signature velvet cloak, is **haggling with the grumpy old vendor.** —

The grumpy vendor, a **tall, sophisticated man**, is wearing a sharp suit, sports a **noteworthy moustache** and is animatedly conversing on his **steampunk telephone**.

Community Showcase



# Stable Diffusion

stability.ai

Models ▾

Developers ▾

Enterprise

Company ▾

News

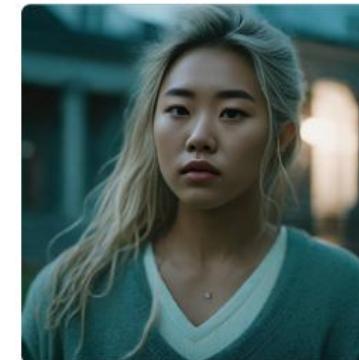
日本語

Contact Us

## Stable Diffusion XL

Create and inspire using the world's fastest-growing open source AI platform

With Stable Diffusion XL, you can create descriptive images with shorter prompts and generate words within images. The model is a significant advancement in image generation capabilities, offering enhanced image composition and face generation that results in stunning visuals and realistic aesthetics.

[Try on Clipdrop](#)[Download Code](#)

# By AI developers - for AI developers



Make sure to upgrade diffusers to >= 0.19.0:

```
pip install diffusers --upgrade
```

In addition make sure to install transformers, safetensors, accelerate as well as the invisible watermark:

```
pip install invisible_watermark transformers accelerate safetensors
```

To just use the base model, you can run:

```
from diffusers import DiffusionPipeline
import torch

pipe = DiffusionPipeline.from_pretrained("stabilityai/stable-diffusion-xl-base-1.0")
pipe.to("cuda")

# if using torch < 2.0
# pipe.enable_xformers_memory_efficient_attention()

prompt = "An astronaut riding a green horse"

images = pipe(prompt=prompt).images[0]
```

To use the whole base + refiner pipeline as an ensemble of experts you can run:

```
from diffusers import DiffusionPipeline
import torch

# load both base & refiner
base = DiffusionPipeline.from_pretrained(
    "stabilityai/stable-diffusion-xl-base-1.0", torch_dtype=torch.float16, variant="fp16"
)
base.to("cuda")
refiner = DiffusionPipeline.from_pretrained(
    "stabilityai/stable-diffusion-xl-refiner-1.0",
    text_encoder_2=base.text_encoder_2,
    vae=base.vae,
    torch_dtype=torch.float16,
    use_safetensors=True,
    variant="fp16",
)
refiner.to("cuda")

# Define how many steps and what % of steps to be run on each experts (80/20) here
n_steps = 40
high_noise_frac = 0.8

prompt = "A majestic lion jumping from a big stone at night"

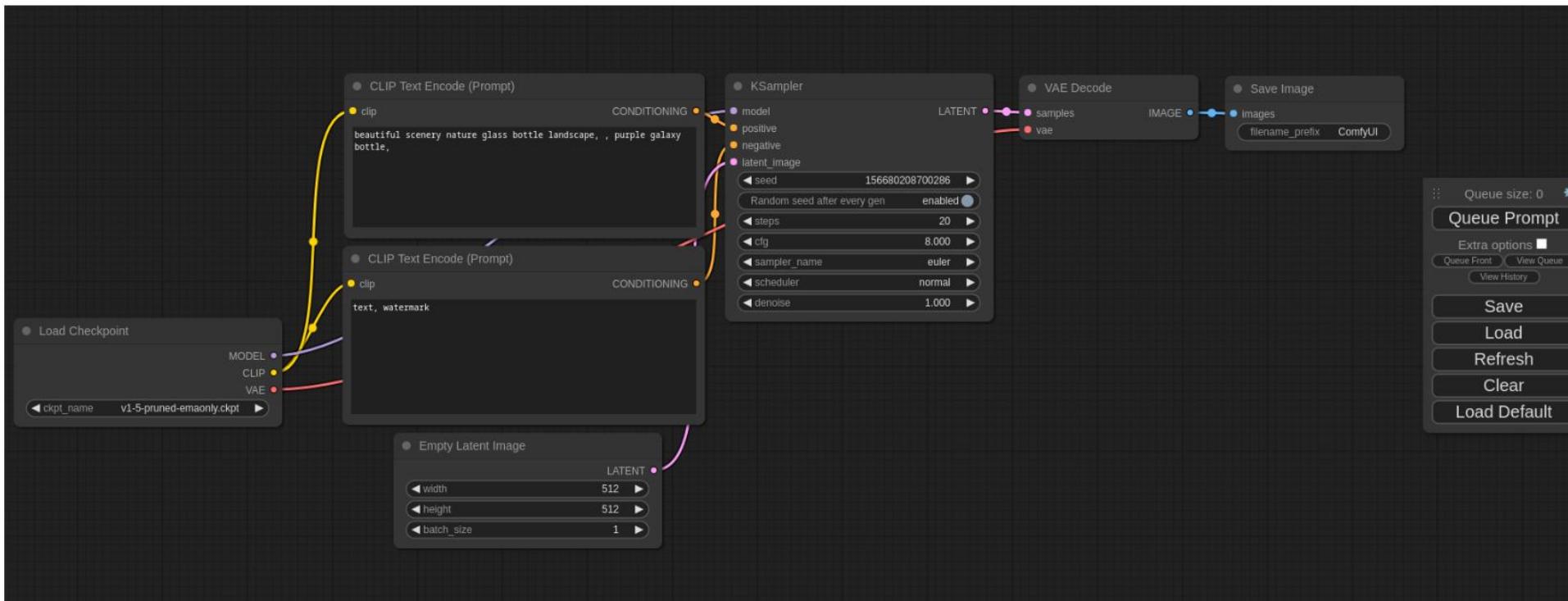
# run both experts
image = base(
    prompt=prompt,
    num_inference_steps=n_steps,
    denoising_end=high_noise_frac,
    output_type="latent",
).images
image = refiner(
    prompt=prompt,
    num_inference_steps=n_steps,
    denoising_start=high_noise_frac,
    image=image,
).images[0]
```

<https://huggingface.co/stabilityai/stable-diffusion-xl-base-1.0>

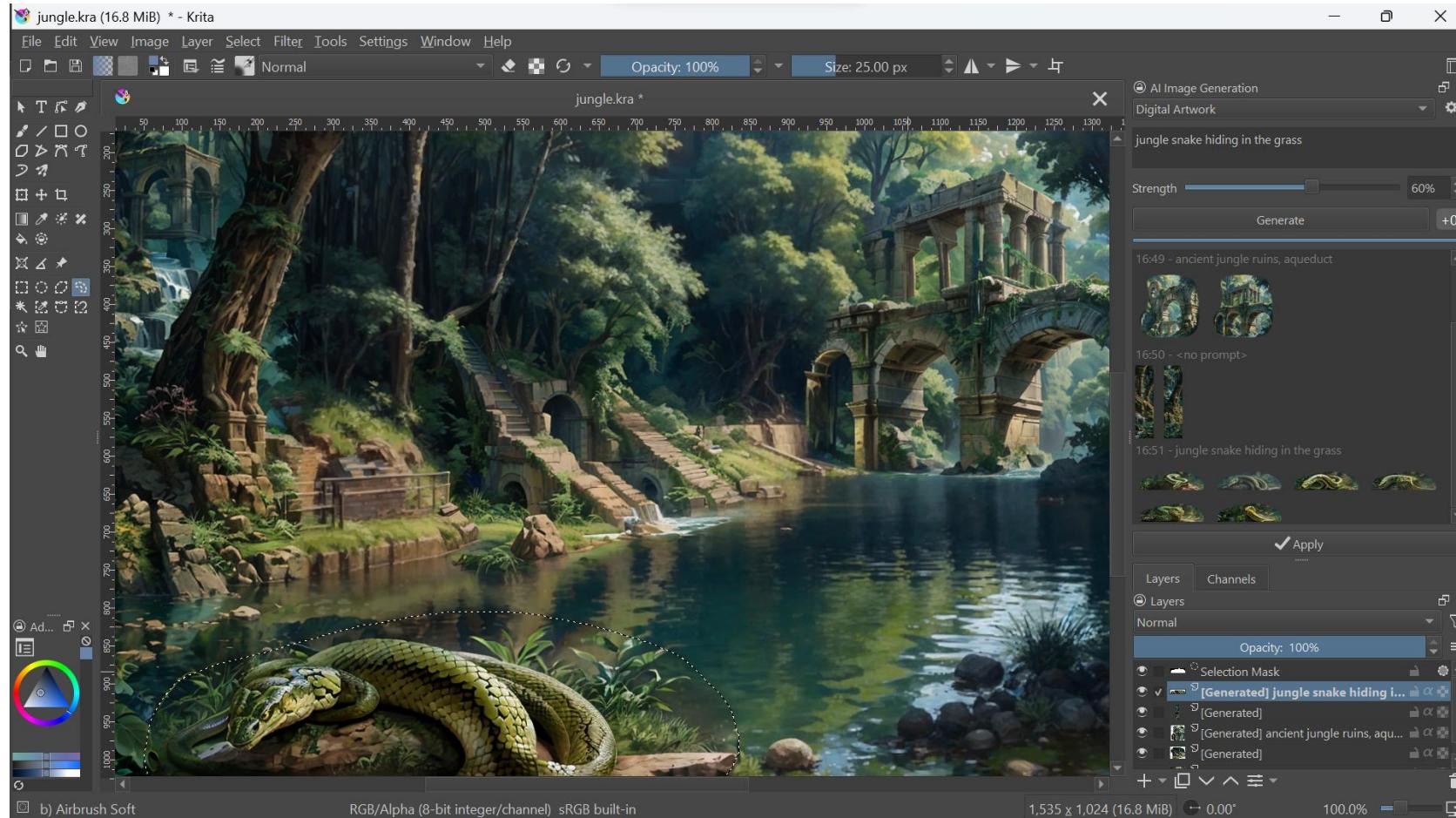
# ComfyUI

## ☞ ComfyUI

The most powerful and modular stable diffusion GUI and backend.



# Krita Diffusion



<https://github.com/Acly/krita-ai-diffusion>

# This workshop - SDXL on Colab

Find the material at <https://github.com/jonandernovella/prompt-engineering>

**Thank you!**

[erik.ylipaa@scilifelab.se](mailto:erik.ylipaa@scilifelab.se)

# Wider implications

# Controversies



TECH • ARTIFICIAL INTELLIGENCE  
He Used AI to Publish a Children's Book in a Weekend. Artists Are Not Happy About It



A page from *Alice and Sparkle*, a children's book made using artificial intelligence. Courtesy Ammaar Reshi

# Copyright and machine learning models

Data is scraped from the internet. Highly likely that the majority is protected under copyright.

Most copyright law considers copying (downloading digital files) as copyright infringement.

In the EU at least there are exemptions:

Under Article 5(1) of Directive 2001/29, an act of reproduction is exempted from the reproduction right provided for in Article 2 of that directive on condition that:

- it is temporary;
- it is transient or incidental;
- it is an integral and essential part of a technological process;
- its sole purpose is to enable a transmission in a network between third parties by an intermediary or a lawful use of a work or other subject-matter to be made, and
- it has no independent economic significance.

Whether the datasets used to train the machine learning models is exempt is yet to be determined.

# Lawsuits

We've filed a lawsuit challenging Stable Diffusion, a 21st-century collage tool that violates the rights of artists.

**Because AI needs to be fair & ethical for everyone.**

---

JANUARY 13, 2023

Hello. This is [Matthew Butterick](#). I'm a writer, designer, programmer, and lawyer. In November 2022, I teamed up with the [amazingly excellent](#) class-action litigators [Joseph Saveri](#), [Cadio Zirpoli](#), and [Travis Manfredi](#) at the [Joseph Saveri Law Firm](#) to file a [lawsuit against GitHub Copilot](#) for its "unprecedented open-source software piracy". (That lawsuit is still [in progress](#).)



<https://stablediffusionlitigation.com/>

# Are AI-created works copyrighted?

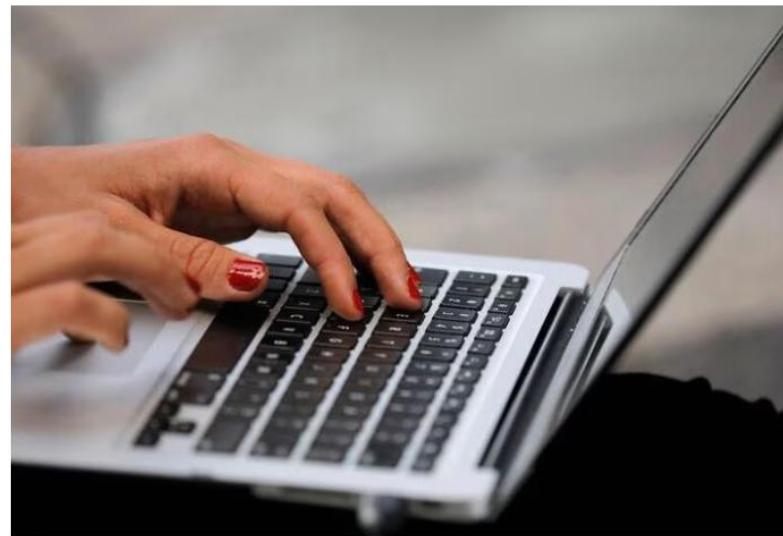
 **REUTERS®** World ▾ Business ▾ Legal ▾ Markets ▾ Breakingviews Technology ▾ Investigations More ▾

 Copyright  Technology  Intellectual Property  Litigation  Data Privacy

2 minute read · February 23, 2023 2:41 AM GMT+1 · Last Updated a day ago

## AI-created images lose U.S. copyrights in test for new technology

By Blake Brittain



REUTERS/Andrew Kelly

# Extracting



Figure 3: Examples of the images that we extract from Stable Diffusion v1.4 using random sampling and our membership inference procedure. The top row shows the original images and the bottom row shows our extracted images.

Carlini, Nicholas, et al. "Extracting training data from diffusion models." arXiv preprint arXiv:2301.13188 (2023).

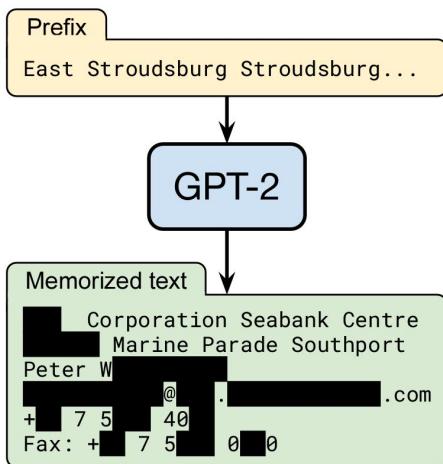


Figure 1: **Our extraction attack.** Given query access to a neural network language model, we extract an individual person's name, email address, phone number, fax number, and physical address. The example in this figure shows information that is all accurate so we redact it to protect privacy.

Carlini, Nicholas, et al. "Extracting Training Data from Large Language Models." USENIX Security Symposium. Vol. 6. 2021.

Category	Count
US and international news	109
Log files and error reports	79
License, terms of use, copyright notices	54
Lists of named items (games, countries, etc.)	54
Forum or Wiki entry	53
Valid URLs	50
<b>Named individuals (non-news samples only)</b>	46
Promotional content (products, subscriptions, etc.)	45
High entropy (UUIDs, base64 data)	35
<b>Contact info (address, email, phone, twitter, etc.)</b>	32
Code	31
Configuration files	30
Religious texts	25
Pseudonyms	15
Donald Trump tweets and quotes	12
Web forms (menu items, instructions, etc.)	11
Tech news	11
Lists of numbers (dates, sequences, etc.)	10

Table 1: Manual categorization of the 604 memorized training examples that we extract from GPT-2, along with a description of each category. Some samples correspond to multiple categories (e.g., a URL may contain base-64 data). Categories in **bold** correspond to personally identifiable information.

# AI-Generated content

FORTUNE

RANKINGS ▾ MAGAZINE NEWSLETTERS PODCASTS MORE ▾

SEARCH SIGN IN Subscribe Now

## ‘Out-of-hand’ flood of ChatGPT-like A.I.-generated stories forces prominent science fiction magazine to stop accepting submissions

BY PRARTHANA PRAKASH

February 22, 2023 at 1:51 PM GMT+1

Updated February 23, 2023 at 4:46 AM GMT+1



More people are using generative A.I. to explore creative avenues.

SOMPONG TOM—GETTY IMAGES

### Most Popular

TECH

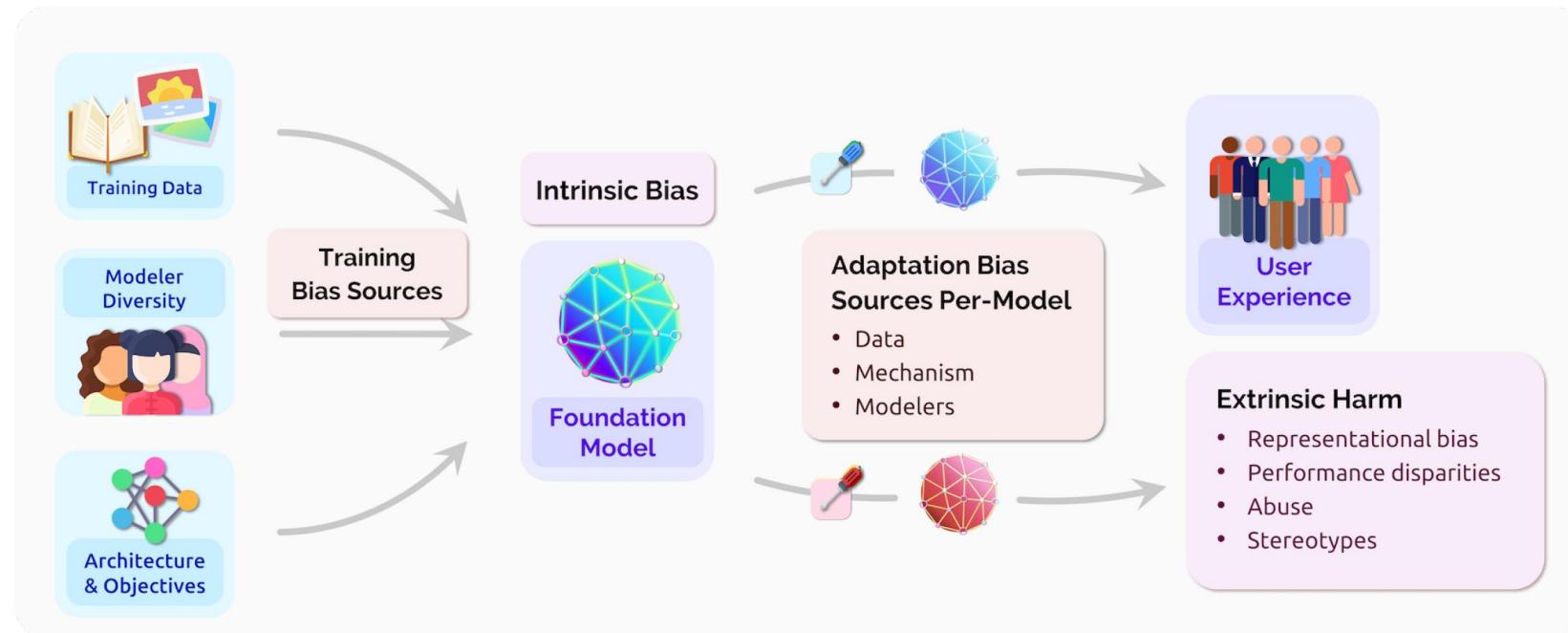
Google is asking staff to share with a ‘desk partner’ but they have to agree on tidiness and decorations



February 23, 2023

BY ELEANOR PRINGLE

# Societal risks of foundation models

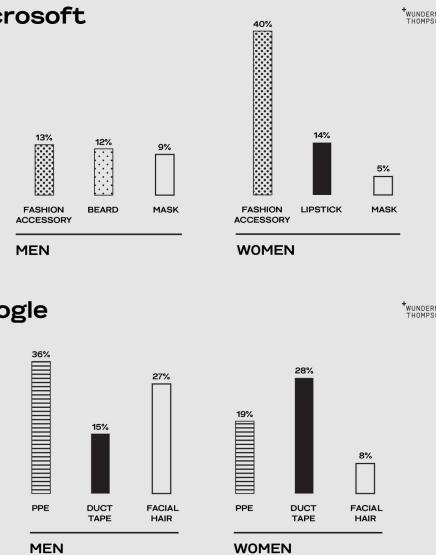


Bommasani, Rishi, et al. "On the opportunities and risks of foundation models." arXiv preprint arXiv:2108.07258 (2021).

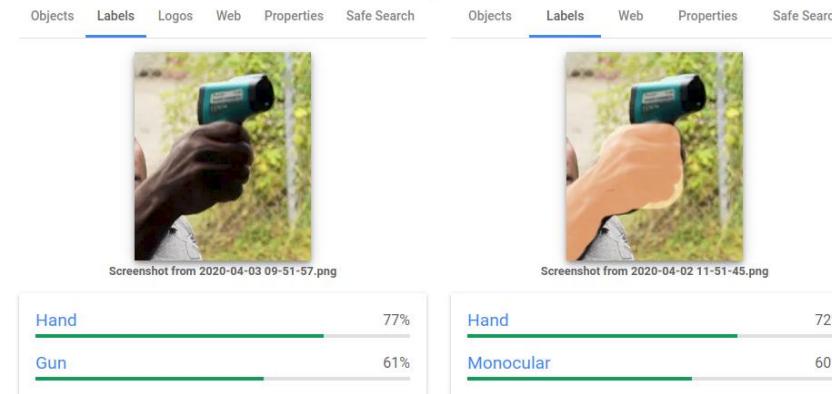
# Algorithmic Bias



Microsoft



<https://www.wundermanthompson.com/insight/ai-and-gender-bias>



Screenshot from 2020-04-03 09-51-57.png

Screenshot from 2020-04-02 11-51-45.png

Category	Score 1 (%)	Score 2 (%)
Hand	77%	72%
Gun	61%	60%
Monocular		

<https://algorithmwatch.org/en/story/computer-vision-police-discrimination/>



Figure 3. Wu and Zhang's "criminal" images (top) and "non-criminal" images (bottom). In the top images, the people are frowning. In the bottom, they are not. These types of superficial differences can be picked up by a deep learning system.

Blaise Agüera y Arcas, Margaret Mitchell and Alexander Todorov,  
*Physiognomy's New Clothes*,  
<https://medium.com/@blaisea/physiognomys-new-clothes-f2d4b59fdd6a>

Name suggestion	Company description	Distance
Magnus bilar	Bolaget ska bedriva verksamhet med bilar	0.028
Fredriks bilar	Bolaget ska bedriva verksamhet med bilar	0.038
Marias bilar	Bolaget ska bedriva verksamhet med bilar	0.044
Annas bilar	Bolaget ska bedriva verksamhet med bilar	0.075

Sahlgren, Magnus, and Fredrik Olsson. "Gender bias in pretrained Swedish embeddings." Proceedings of the 22nd Nordic Conference on Computational Linguistics. 2019.

Google Translate

HUNGARIAN - DETECTED POLISH PO ENGLISH POLISH PORTUGUESE

Ő szép. Ő okos. Ő olvas. Ő mosogat. Ő épít. Ő varr. Ő tanít. Ő főz. Ő kutat. Ő gyereket nevel. Ő zenél. Ő takarító. Ő politikus. Ő sok pénzt keres. Ő sütéményt süt. Ő professzor. Ő asszisztens. |

She is beautiful. He is clever. He reads. She washes the dishes. He builds. She sews. He teaches. She cooks. He's researching. She is raising a child. He plays music. She's a cleaner. He is a politician. He makes a lot of money. She is baking a cake. He's a professor. She's an assistant.

Hungarian has no gender pronoun, so Google Translate makes some assumptions.  
[https://www.reddit.com/r/europe/comments/m9uphb/hungarian\\_has\\_no\\_gendered\\_pronouns\\_so\\_google/?utm\\_source=share&utm\\_medium=web2x&context=3](https://www.reddit.com/r/europe/comments/m9uphb/hungarian_has_no_gendered_pronouns_so_google/?utm_source=share&utm_medium=web2x&context=3)

# What language is represented?



Fig. 5. Only a tiny percentage of the world's languages are currently represented in foundation models. There are over 6,000 languages in the world, with estimates varying due to the inherent uncertainty of what constitutes a separate language [Nordhoff and Hammarström 2011]. This map shows the languages of the world, with each dot representing one language and its color indicating the top-level language family. Data is from Glottolog [Hammarström et al. 2021]. We label a few of the languages on the map as examples.

Bommasani, Rishi, et al. "On the opportunities and risks of foundation models." arXiv preprint arXiv:2108.07258 (2021).

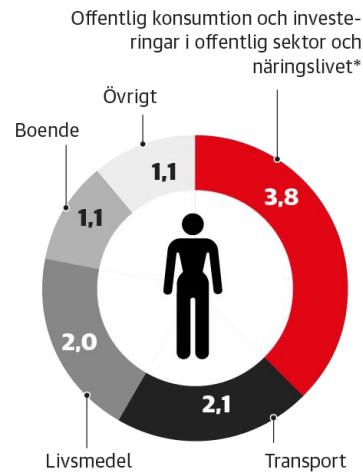
# Economic impacts

*“In summary, foundation models are poised to be an important general-purpose technology of our era. They have potential to increase living standards substantially, but also pose risks of increasing inequality and concentrating power. The economic implications of these technologies are not predetermined, but rather depend on how technologists, policymakers, managers, workers, and other stakeholders answer challenges such as:*

- *How can we harness the potential of foundation models to boost productivity?*
- *Can we develop models that enhance creativity and boost the rate of innovation?*
- *Will the benefits and control rights be limited to a few or widely shared?*

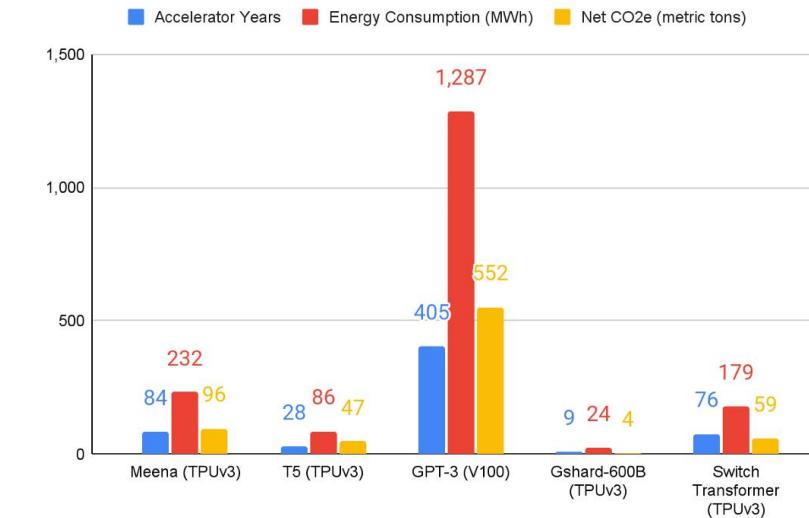
Bohmasani, Rishi, et al. "On the opportunities and risks of foundation models." arXiv preprint arXiv:2108.07258 (2021).

# Carl



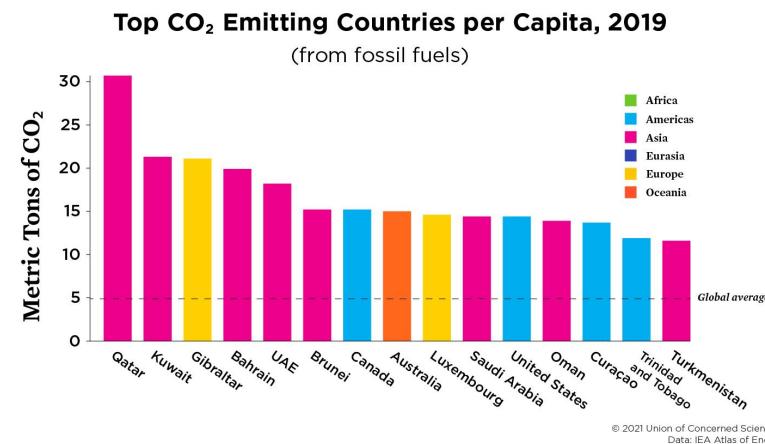
Average per capita swedish metric tons of CO<sub>2</sub>

<https://www.dn.se/nyheter/sverige/fakta-i-fragan-vilken-forandring-i-vardagen-gor-bast-nytta-for-klimatet/>



**Figure 3. Accelerator years of computation, energy consumption, and CO<sub>2</sub>e for five large NLP DNNs**

Carbon Emissions and Large Neural Network Training. David Patterson, Joseph Gonzalez, Quoc V. Le, Chen Liang, Lluís-Miquel Munguía, D. Rothchild, David R. So, Maud Texier, J. Dean. 2021. From Google.



<https://www.ucsusa.org/resources/each-countrys-share-co2-emissions>

## Science &amp; Tech

## A Photographer Submitted an A.I.-Generated Image to a Prestigious Art Competition to Be ‘Cheeky.’ It Won a Top Prize Anyway

Boris Eldagsen has refused the award, saying his submission was intended to start a discussion about the use of A.I. in photography.

Richard Whidington, April 17, 2023



Boris Eldagsen, *The Electrician* (2022). Photo courtesy Boris Eldagsen.  
<https://news.artnet.com/art-world/boris-eldagsen-photography-award-sony-ai-generated-images-dall-e-2286622>