

Machine Learning: Deep Learning

**CPSC 501: Advanced Programming Techniques
Winter 2025**

Jonathan Hudson, Ph.D
Assistant Professor (Teaching)
Department of Computer Science
University of Calgary

Friday, February 21, 2025

Copyright © 2025

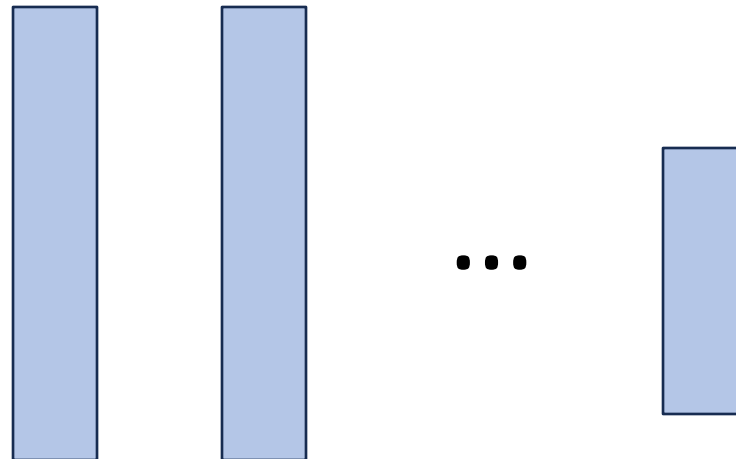


**UNIVERSITY OF
CALGARY**

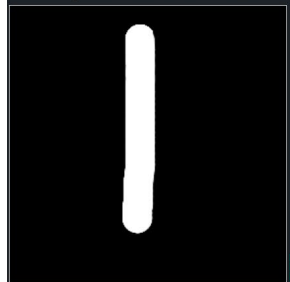
Example

This is one way you can think about what a neural net for MNIST is doing.

The input space has 255^{784} possible input values, but is ultimately compressed to 10 possible values representing the aspect of the image that we care about.



Draw your number here



Downsampled drawing: 1

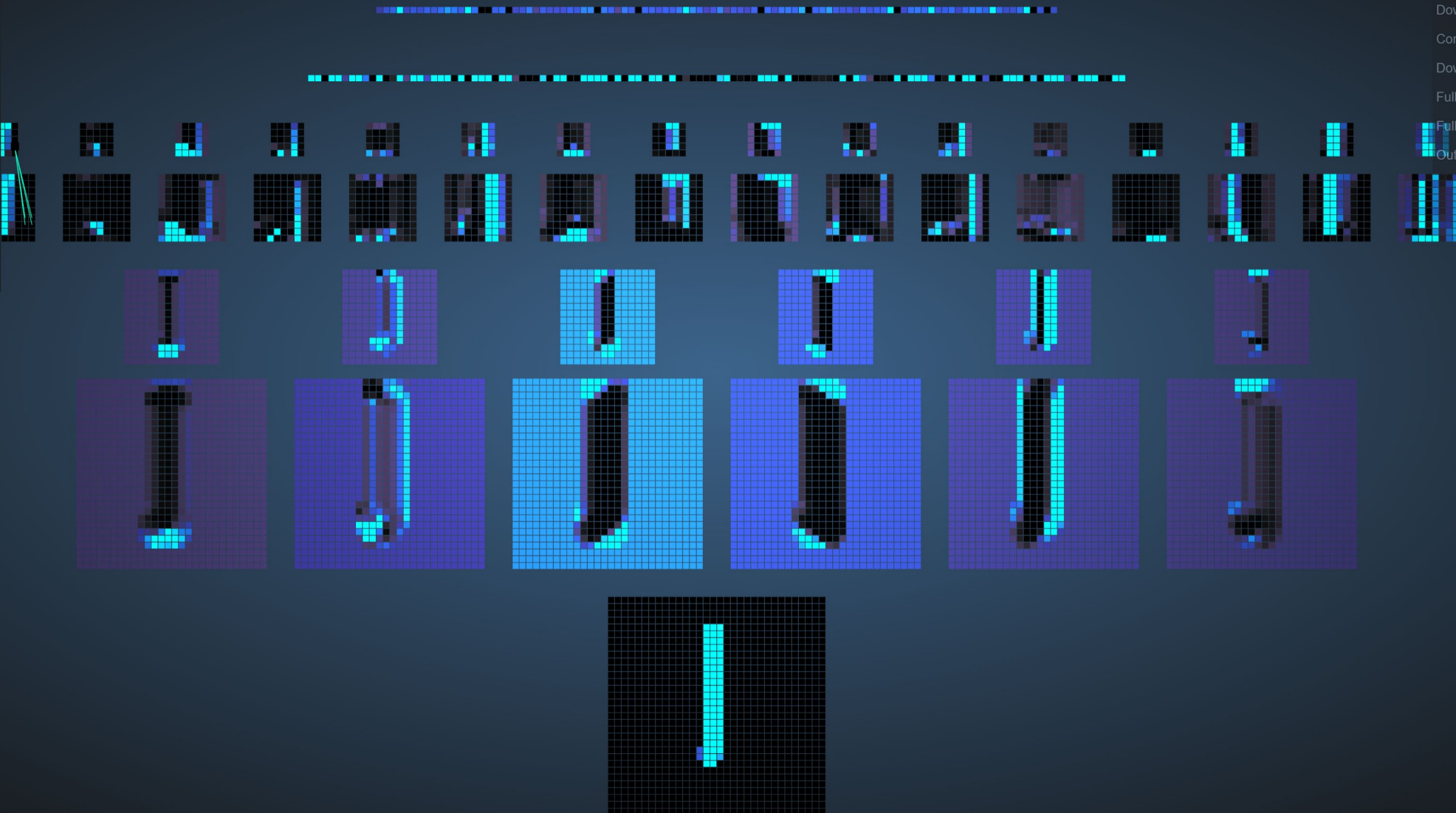
First guess: 1

Second guess: 0

0 1 2 3 4 5 6 7 8 9

Layer visibility

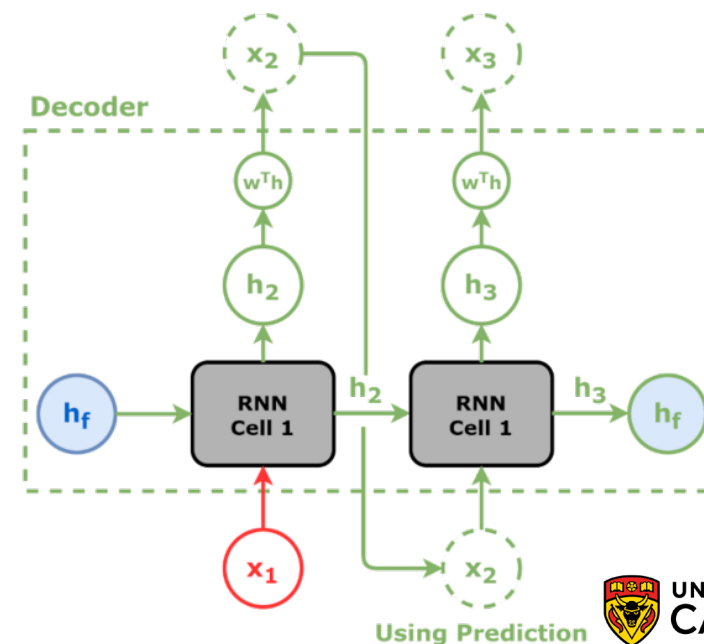
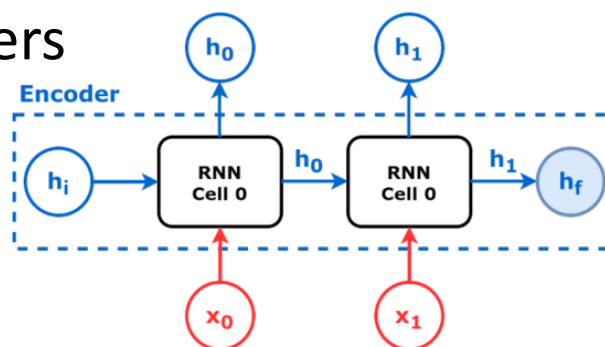
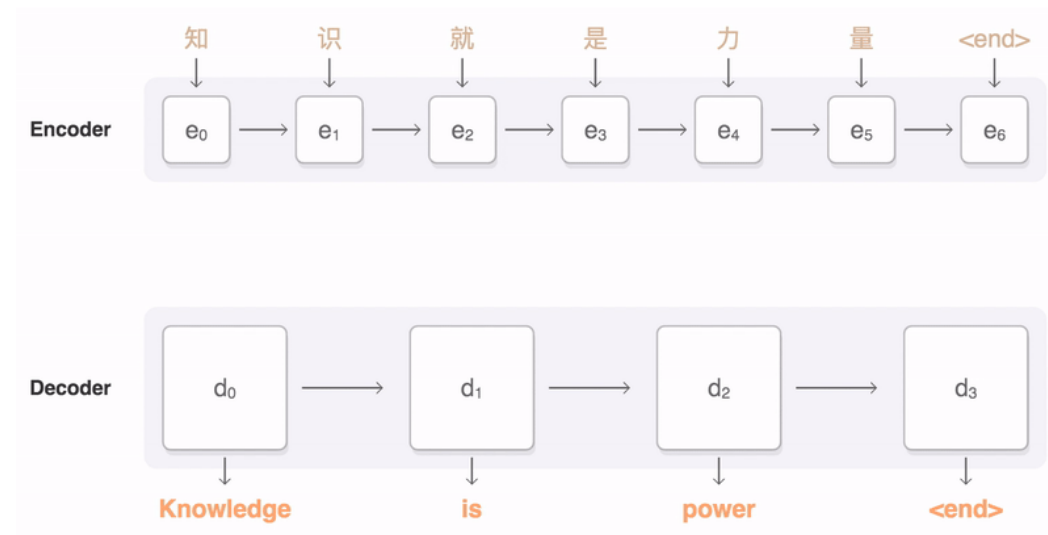
- Input layer
- Convolution layer 1
- Downsampling layer 1
- Convolution layer 2
- Downsampling layer 2
- Fully-connected layer 1
- Fully-connected layer 2
- Output layer



https://adamharley.com/nn_vis/cnn/2d.html

Encoder-Decoder

- Seq2seq problems require one sequence to be changed into another 2014
- In NN, encoder captures input to **latent vector state**
- Decoder takes a **latent vector** and generates output
 - retains consideration of prior symbols in sequence
- Issues with long input
- Standard until 2017 Transformers



Autoencoding

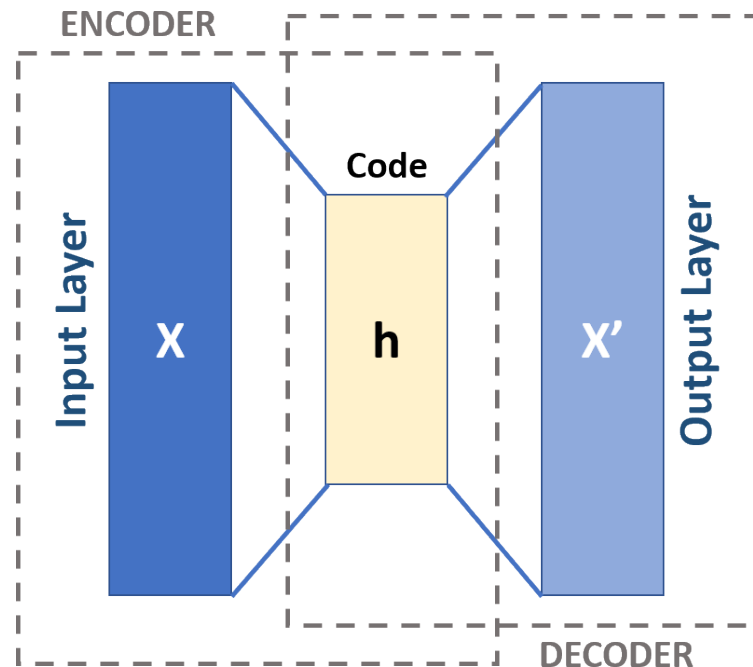
Autoencoding is a type of unsupervised learning used to find feature-preserving representations of unlabeled data in a smaller **latent space**.

It is a type of dimensionality reduction, where the reduction is learned by a model through training.

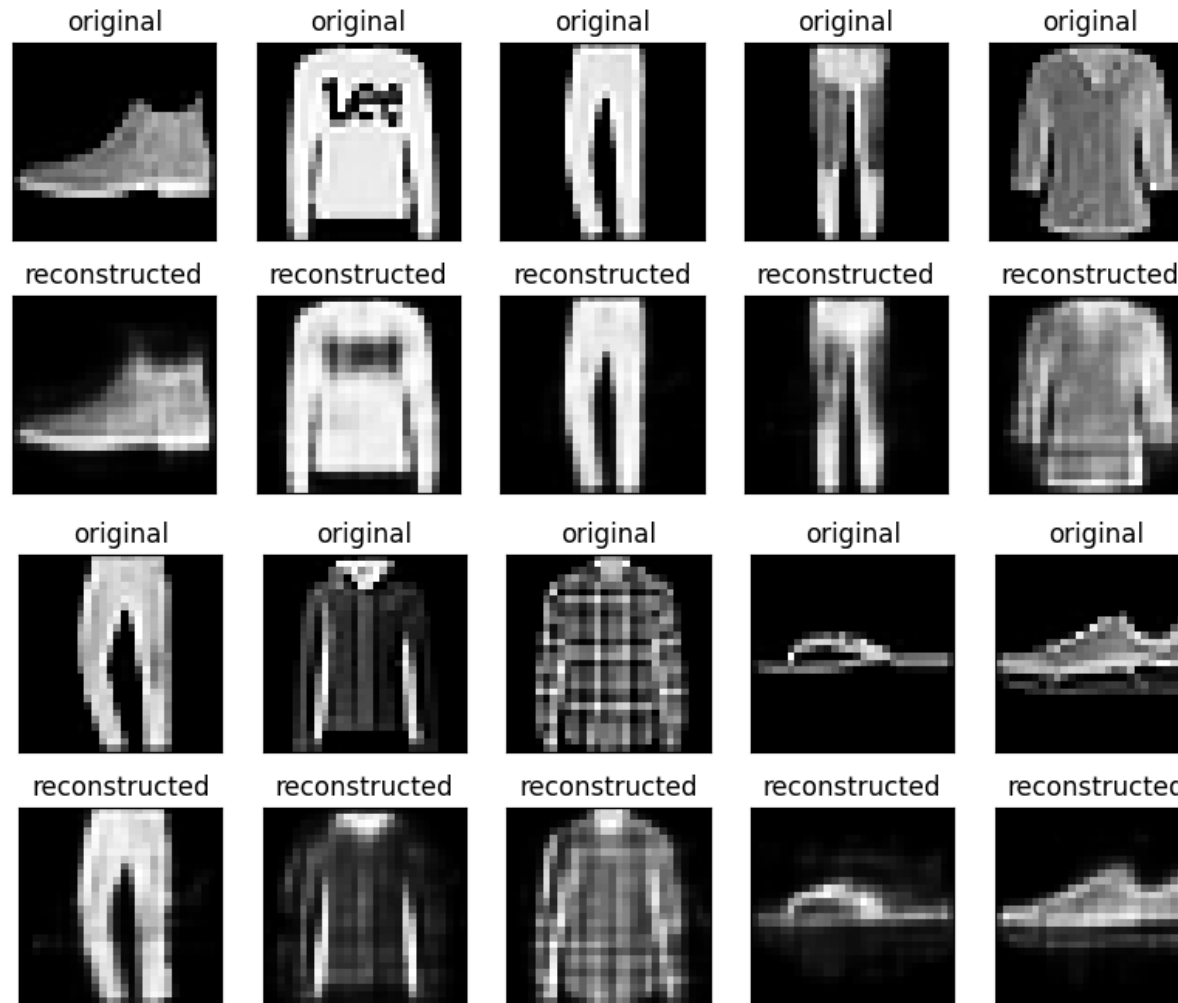
This is done by forcing the data through a smaller layer while trying to preserve the original content.

Autoencoding

- An autoencoder consists of two parts: an encoder and a decoder.
- The autoencoder is then trained to recreate the identity function, but passing data through a small intermediate layer.

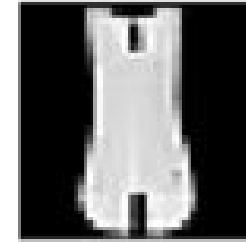
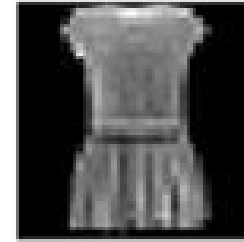
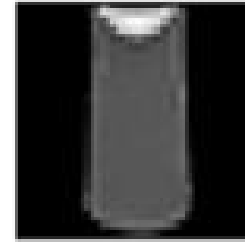
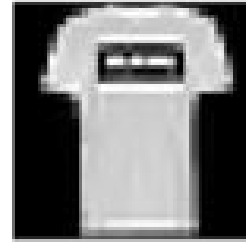
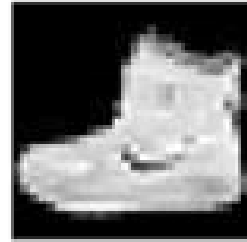


FashionMNIST

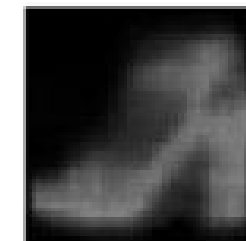
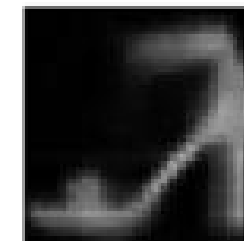
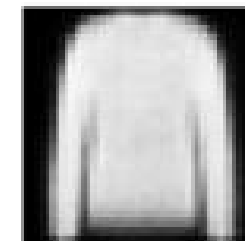
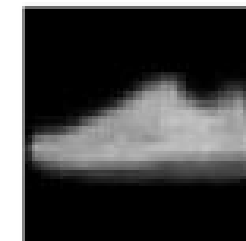
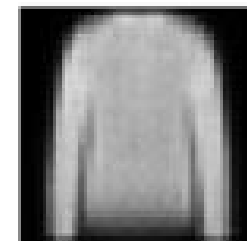
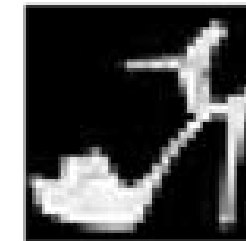
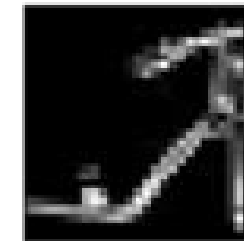
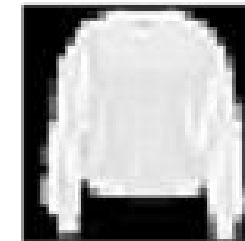
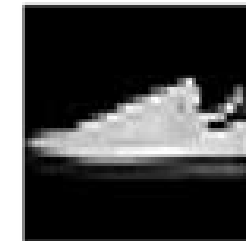
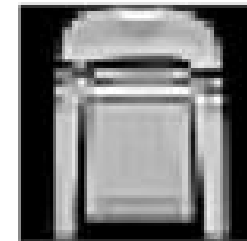
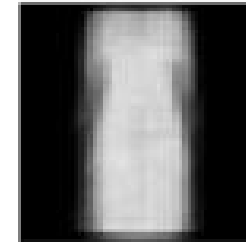
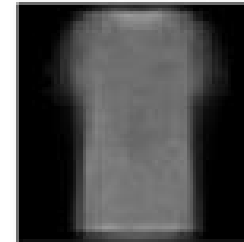
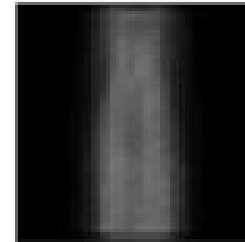
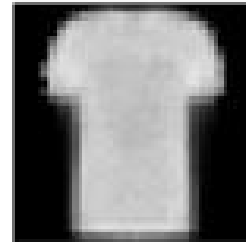
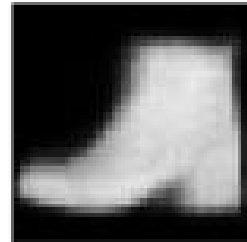


- 1: Trousers
- 2: Pullover
- 3: Dress
- 4: Coat
- 5: Sandal
- 6: Shirt
- 7: Sneaker
- 8: Bag
- 9: Ankle boot

Original
Image



Autoencoder
Reconstruction



Autoencoding

In fact, we can think of what neural networks are doing is learning to encode the data they see in training.

- The encoding will depend on the nature of the training set

The last layer then takes this encoding and makes a final decision based on the encoded data it sees.

VAE

A **variational autoencoder (VAE)** is a type of autoencoder.

VAEs are used as a component in image generators like DALL-E and Stable Diffusion.

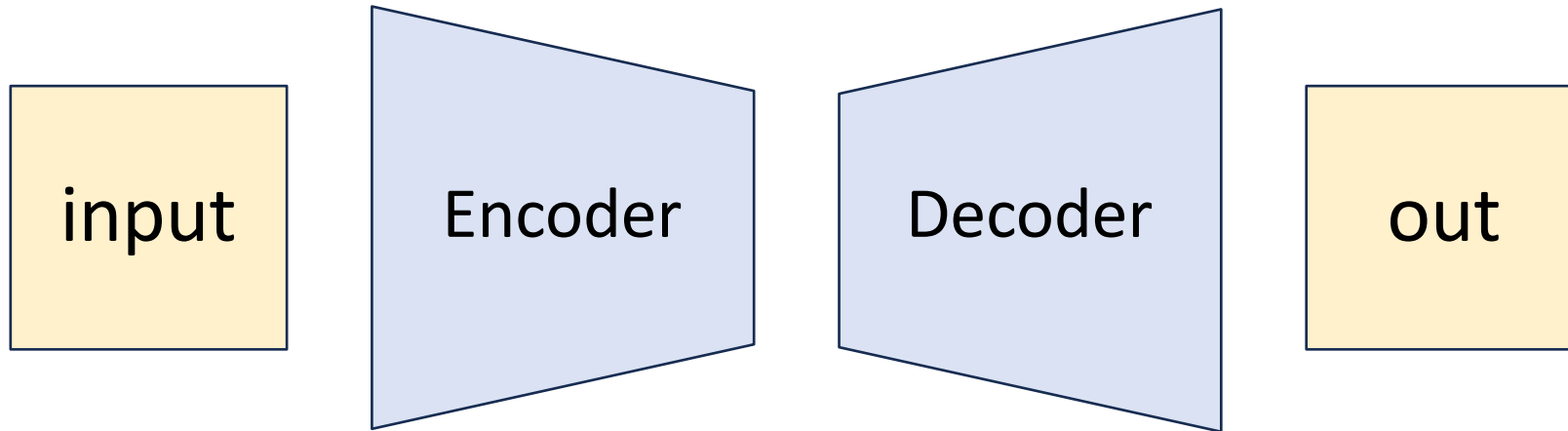
These VAEs are trained on a huge corpus of images pulled from the internet (ethically or otherwise).

- e.g. LAION-5B, which has 5 billion images scraped from the internet

Image generation

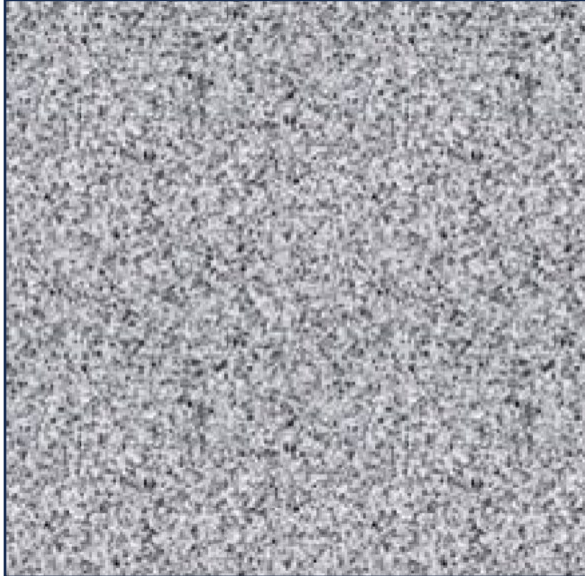
Image generators like DALL-E and Stable Diffusion work by applying a **diffusion** process to images in an encoded space.

- A VAN is responsible for the encoding/decoding steps



Diffusion

- A **diffusion** network is trained to remove noise from images.
- We can use this to generate images by applying it repeatedly to a starting image of random noise.



Latent Diffusion Model

- Latent Diffusion Model (2015)
- Trained by gradually adding noise to the training images. The model is then trained to reverse this process, starting with a noisy image and gradually removing the noise until it recovers the original image
 - Like a sequence of denoising autoencoders
- The resulting embeds within itself a 'latent' or neural network concept of an image (this can be connected to text input as well for prompting)
- And LDM is used by asking for the idea of a latent output, the LDM then runs on some noisy input until the output is 'denoised' sufficiently

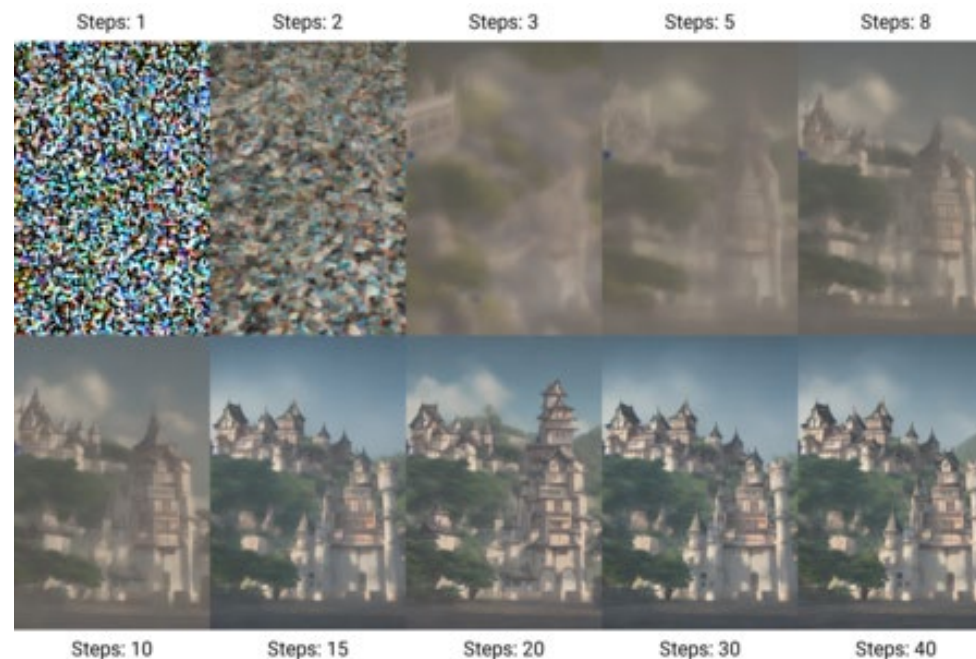
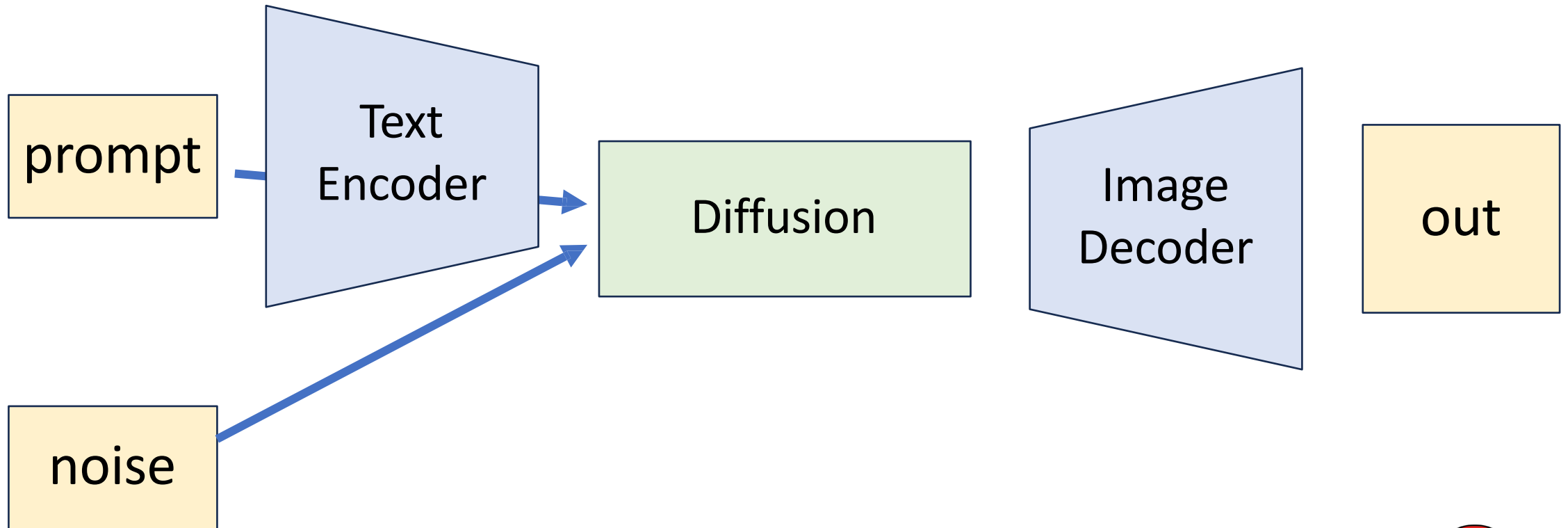


Image generator

The final diffusion-based image generator looks like the following:



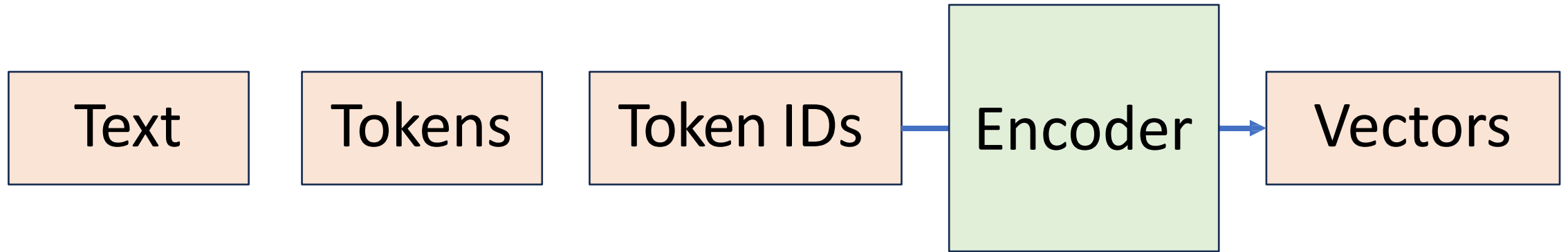
NLP

Natural language processing (NLP) is a subfield of computer science and AI concerned with giving computers the ability to process data encoded in natural language.

This is required for dialogue, comprehension, classification, summarization, prediction, translation, recognition (text or voice), generation, etc.

Processing

Text data is prepared according to the following process:



We want to train our encoder to produce a “good” representation, which is determined by the task we train it on.

Definition

A **language model** is a statistical model of a language, i.e. a probability distribution over words, symbols, or tokens in a language.

- Can be used for a range of language-processing tasks, esp. generation

Language models can be based on purely statistical patterns, or they can take the form of machine-learning models.

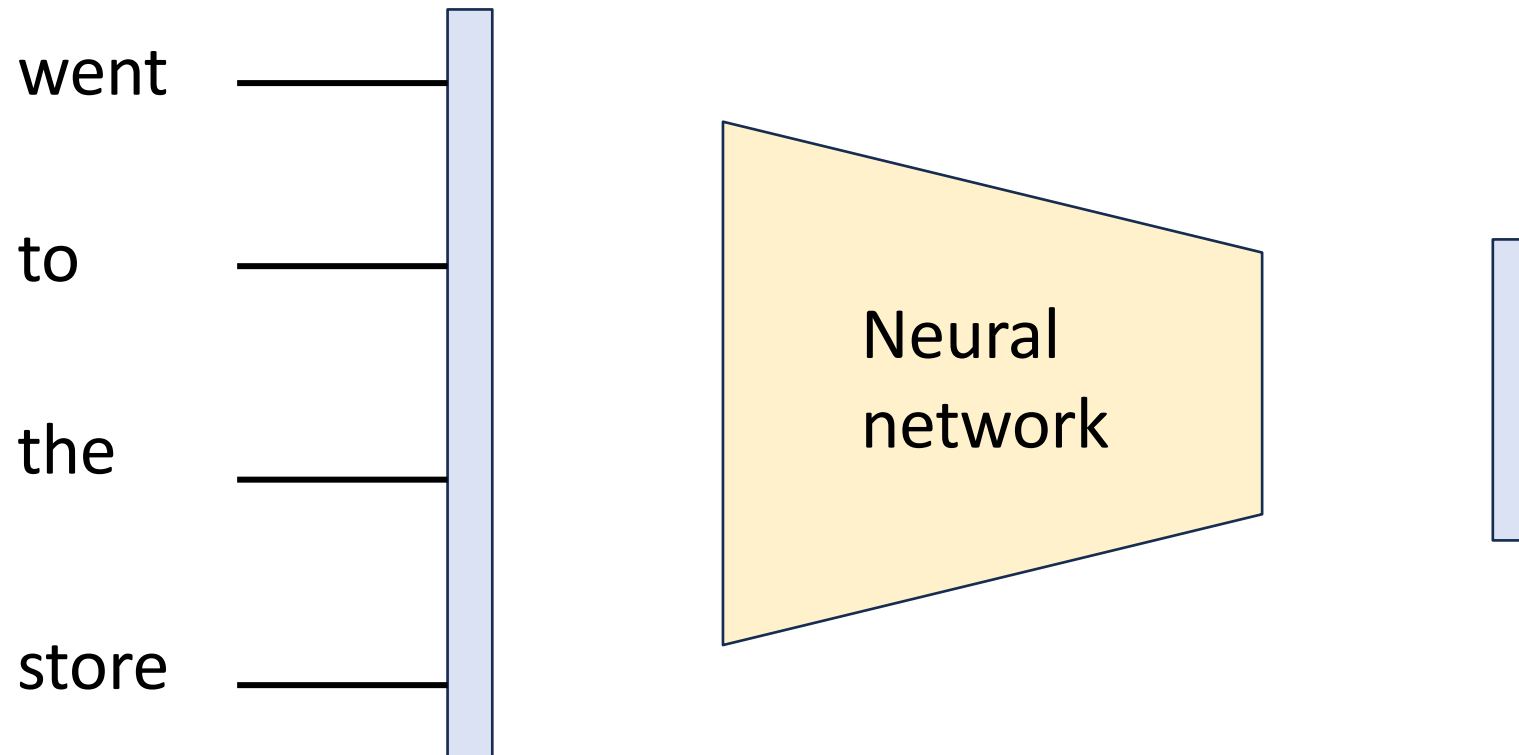
Memory

In order to encode tokens in a way that captures their semantics, we need to incorporate some form of “memory” into our encoder.

These encoders are usually trained on the task of predicting the next word in a piece of text.

Finite memory

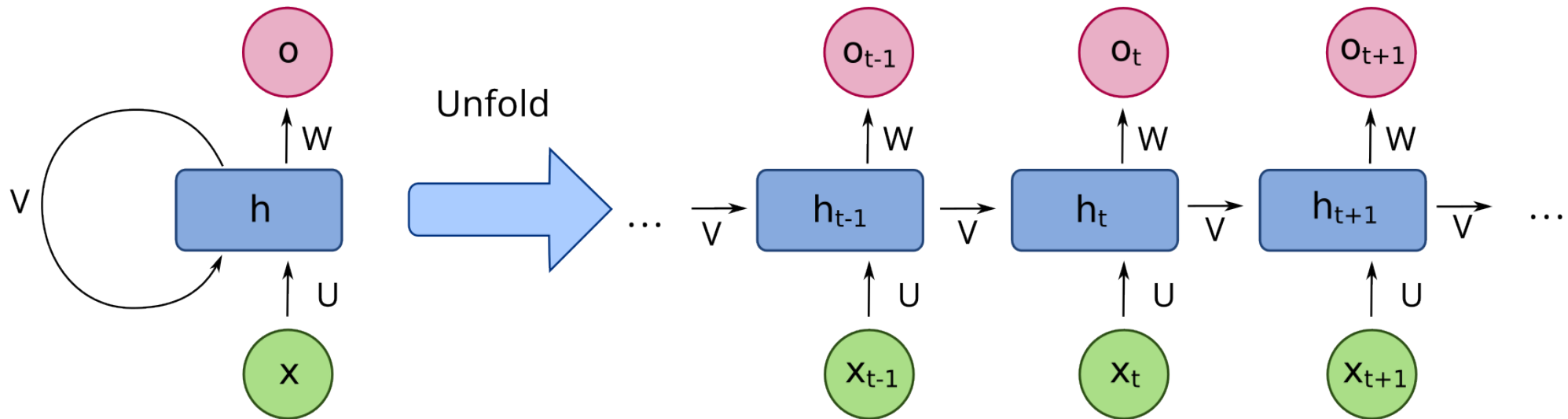
One way to approach this is to simply take the previous n words as input to a neural network.



Hi Renate I just found this one  and it says that I submitted the final exam grade it is a nice to make you feel like I have a good practice and a good time  for the late email to you if you want a good time  for you for the clarification on what you want me know about you can do you want to meet with me know you guys have some good stuff for you guys and we will get in contact that day going on the back of our first online cannot get the job offer and we will need a few minutes late reply and I was going through some emails from the bookstore to get in touch on the problems of our first online cannot get the job done right now has the right track to be as a TA as it can you please check out this one  I am still interested and I was going well I would like the latest reference

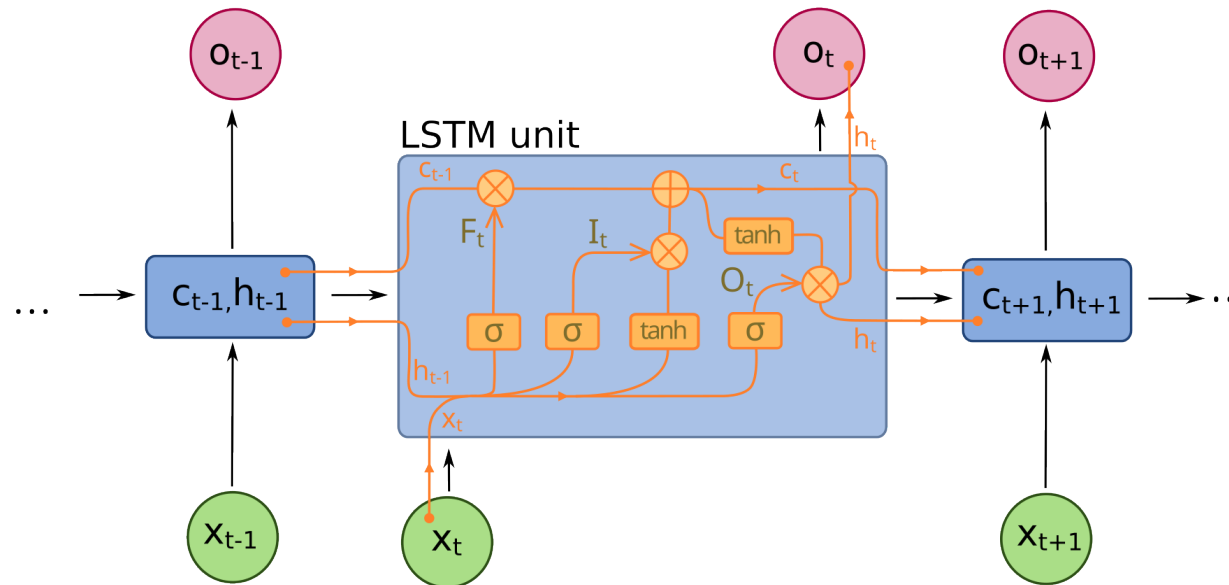
Recurrence

A **recurrent neural network (RNN)** uses “loopback” connections to feed its own output back into the network when processing the next input.



LSTM

Long short-term memory (LSTM) is the most commonly used RNN architecture. It has extra mechanisms that can “learn” what information is important to keep and what can be forgotten.

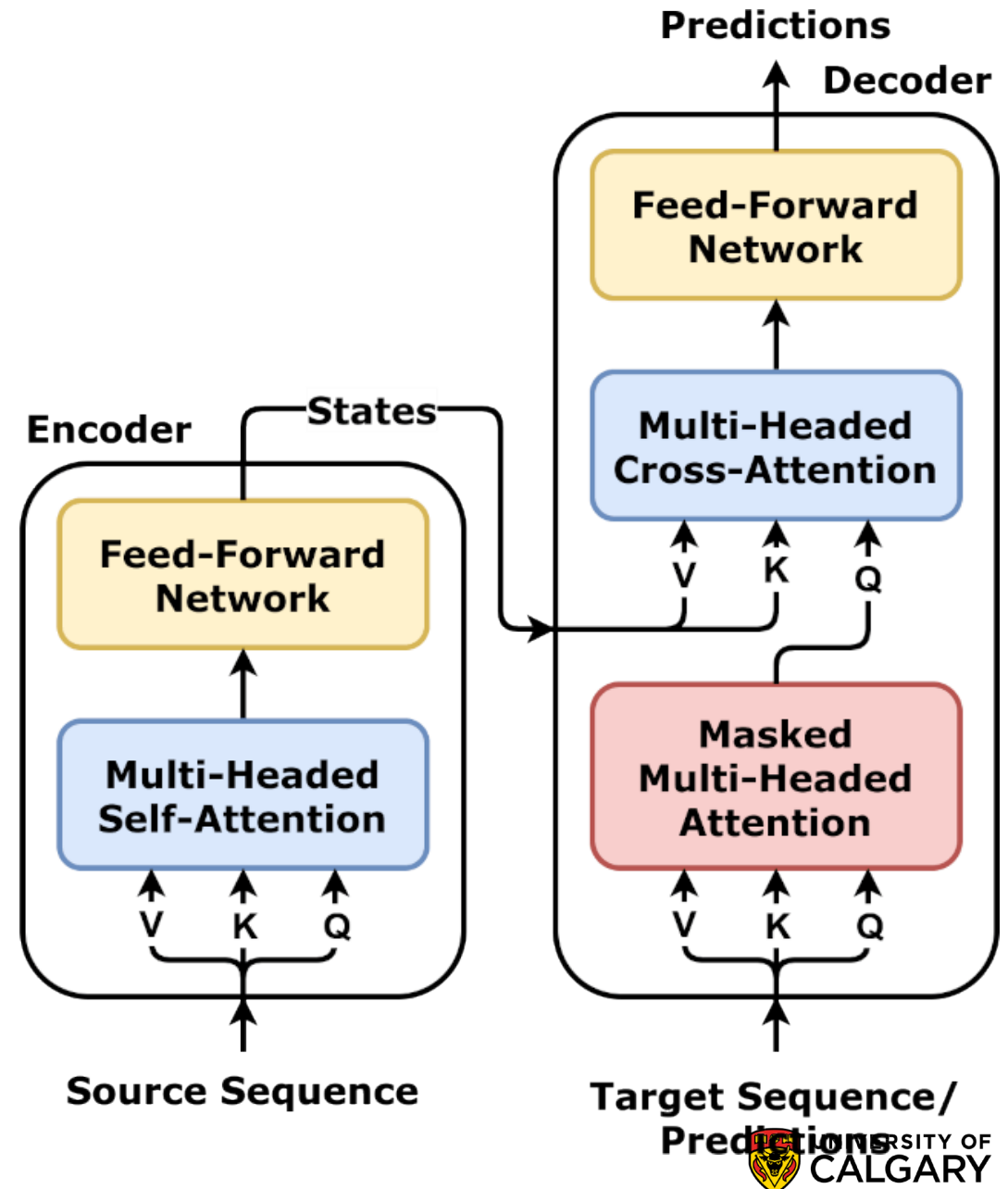


Transformers

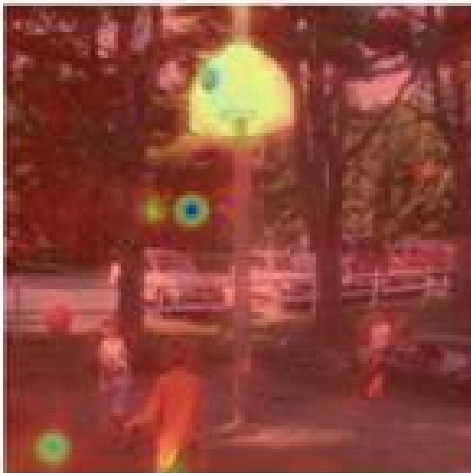
In 2017, a paper called “Attention is all you need” introduced the idea of **transformers**, which allow for sequences to be processed in parallel.

They also have an “attention” mechanism that learns which context information is important to keep for each token.

- The context scope for GPT-3 is about 3,000 words



basketball
(95.13%)



beacon (99.81%)



American_black_bear
(92.99%)



basketball
(87.74%)



LLMs

OpenAI's **Generative Pre-trained Transformer (GPT)** series of generative language models are based on this transformer architecture.

These are categorized as **large language models (LLMs)** because of their size and scope.

For example, GPT-2 had 1.5 billion parameters and was trained on 8 million scraped webpages (including Wikipedia).

Training process

Like previous GPT models, the GPT-4 base model was trained to predict the next word in a document, and was trained using publicly available data (such as internet data) as well as data we've licensed. The data is a web-scale corpus of data including correct and incorrect solutions to math problems, weak and strong reasoning, self-contradictory and consistent statements, and representing a great variety of ideologies and ideas.

So when prompted with a question, the base model can respond in a wide variety of ways that might be far from a user's intent. To align it with the user's intent within guardrails, we fine-tune the model's behavior using reinforcement learning with human feedback (RLHF).

Note that the model's capabilities seem to come primarily from the pre-training process—RLHF does not improve exam performance (without active effort, it actually degrades it). But steering of the model comes from the post-training process—the base model requires prompt engineering to even know that it should answer the questions.

LLMs

LLMs like the GPT series are trained purely as language models.

- i.e. predicting and generating text (plus some manual fine-tuning)

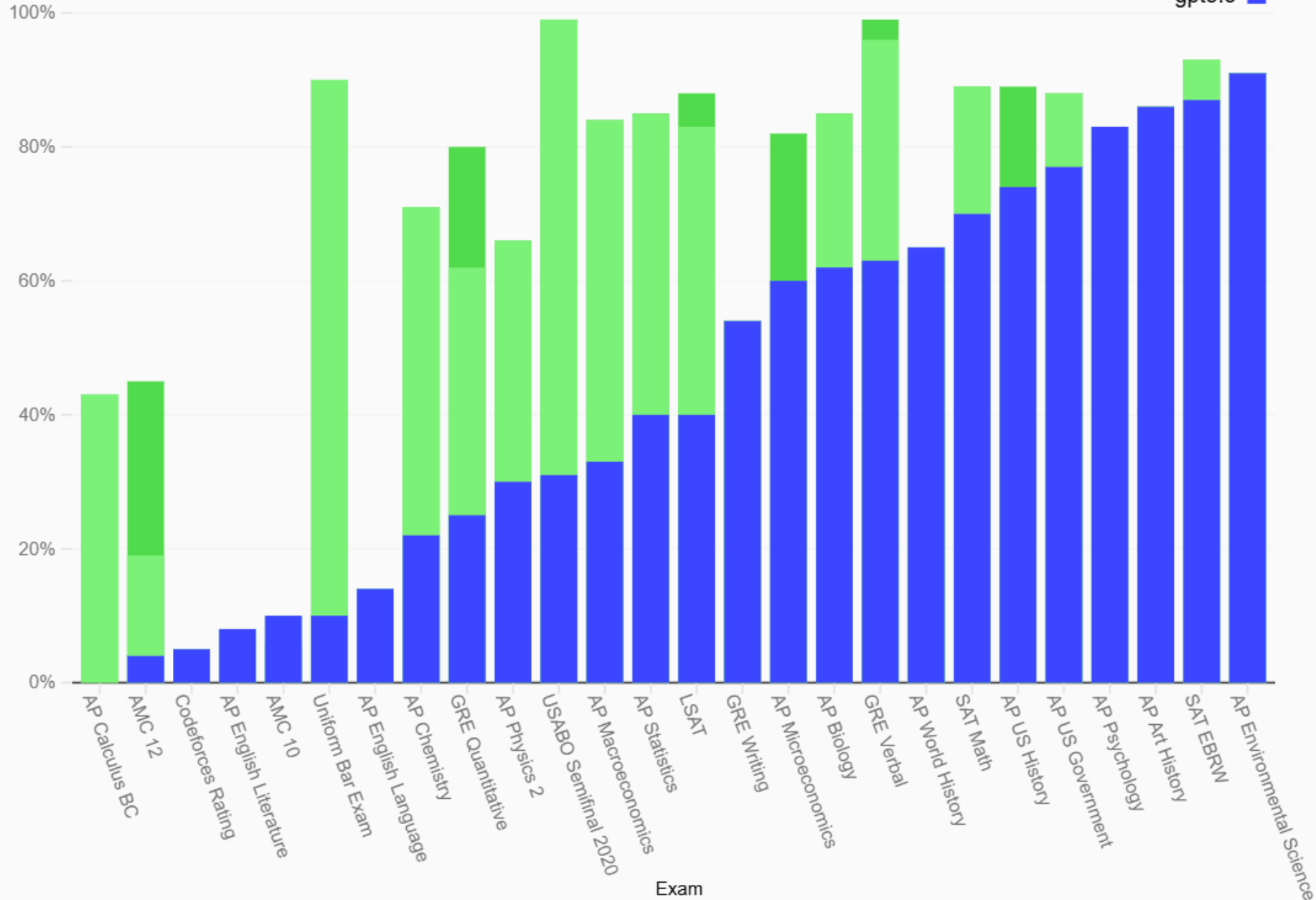
However, they have shown to perform surprisingly well at tasks they were not explicitly trained for.

- This is expected for tasks with a lot of text data, but also occurs for tasks with surprisingly little data

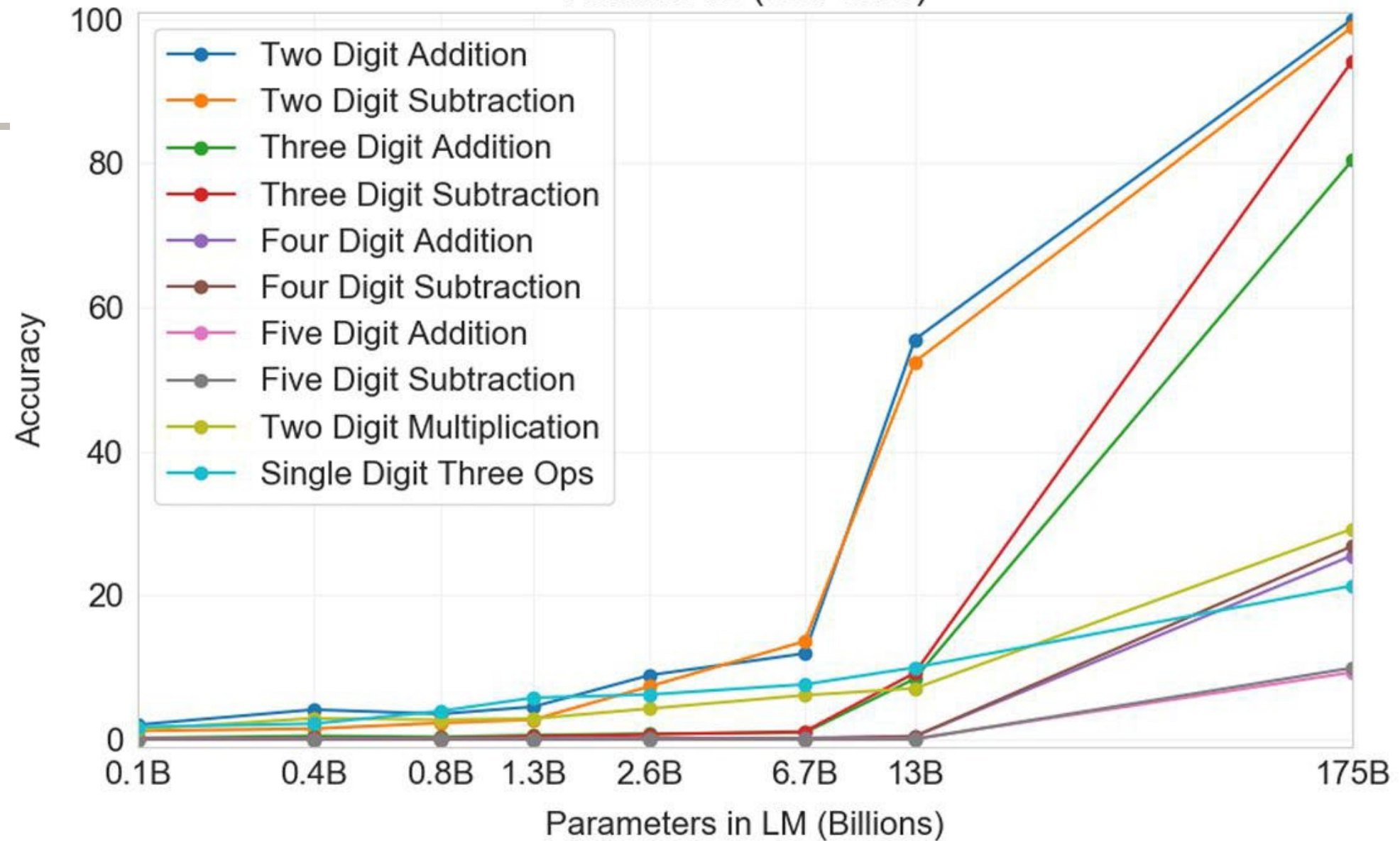
Exam results (ordered by GPT-3.5 performance)

Estimated percentile lower bound (among test takers)

gpt-4
gpt-4 (no vision)
gpt3.5



Arithmetic (few-shot)

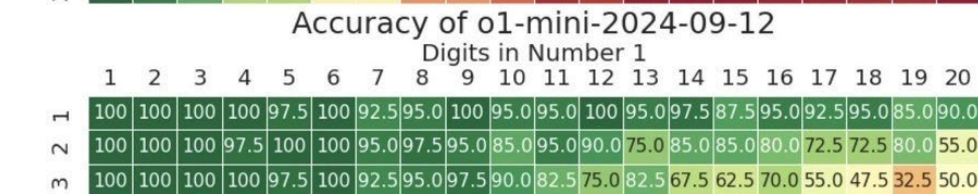
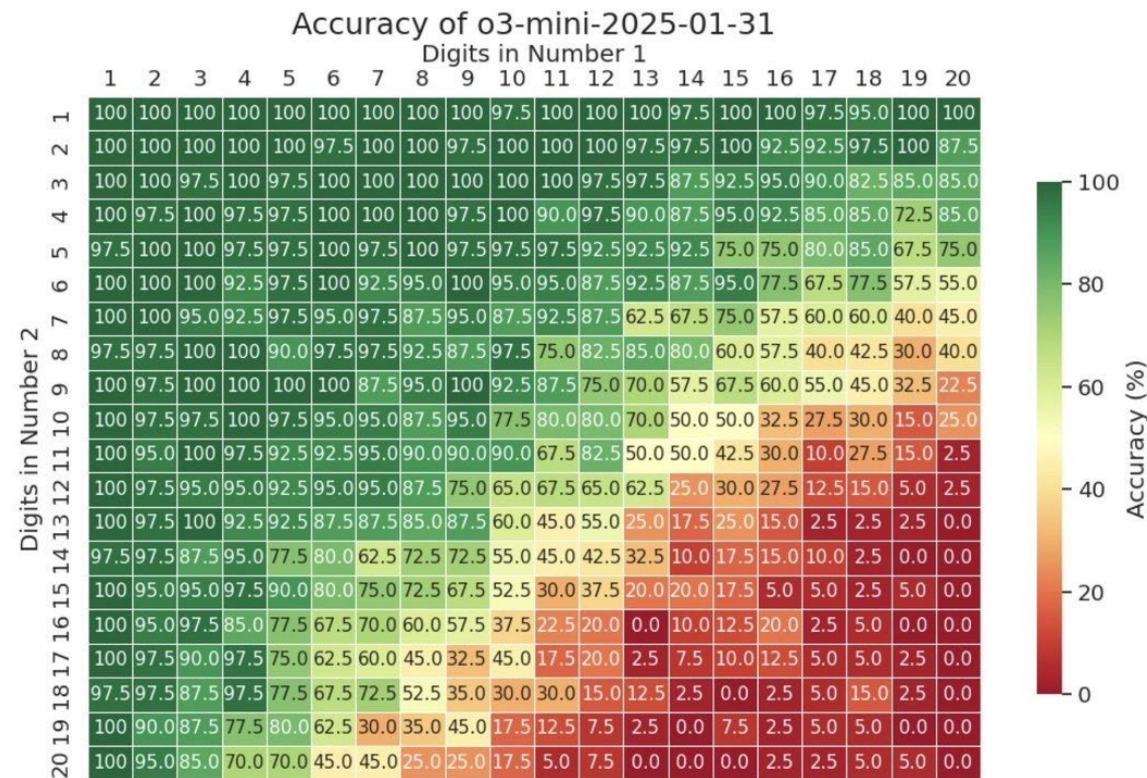




Yuntian Deng
@yuntiandeng

Follow

For those curious about how o3-mini performs on multi-digit multiplication, here's the result. It does much better than o1 but still struggles past 13×13. (Same evaluation setup as before, but with 40 test examples per cell.)



T. Brown

LLMs

- Another interesting feature about LLMs is that we might not have a full picture of their capabilities.
- There may be encoded “knowledge” that we might not know how to extract well
- e.g. **jailbreaking**
- Prompt processing/engineering is a huge part of LLM development, and is one of the most important differences between GPT-3 and GPT-4.

LLMs

However, LLMs still tend to perform much worse than humans in areas where there is less training data and/or where reinforcement requires human expertise.

They also have a tendency to **hallucinate**, or make up nonsensical or incorrect factual information.



horse dentist

@equine__dentist



what an insane diet for a bird



<https://www.petco.com> › pet-services

Cockatiel Bird Care Sheet | How to Care for a Cockatiel | Petco

A well-balanced cockatiel diet consists of: ...
Pregnant women, children under the age of 5,
senior citizens and people with weakened
immune ...

11:44 AM · 10/9/23 from Earth · 256K Views

Google

All Images

AI Overview

The first
1316. He

Other people

- Terry
1976
- Carey
the 2
- Jose



Quora

https://www.quora.com › When-was-the-first-document...

When was the first documented backflip?

the first recorded backflip on history was performed in **1316 by Sir John H. Backflip**. Sir Backflip was an acrobat and knight for king Edward II.

When was the **first backflip** ever performed?

3 answers

Feb 9, 2014

What was invented **first**, the **backflip** or the front flip?

3 answers

Aug 24, 2022

More results from www.quora.com

People also ask

Who did the first ever backflip?



When did backflip start?



Who was the first male skater to do a backflip?



Did Carey Hart do the first backflip?



Feedback



Reddit · r/AskHistory

1 comment · 8 years ago

Who was the first person/human in history to do a back flip ...

Well, the **Minoan Bull-Leaping Fresco** depicts a flip-like maneuver and is from about 1700-1425 BC; this type of bull acrobatics is also shown ...

as in
els.

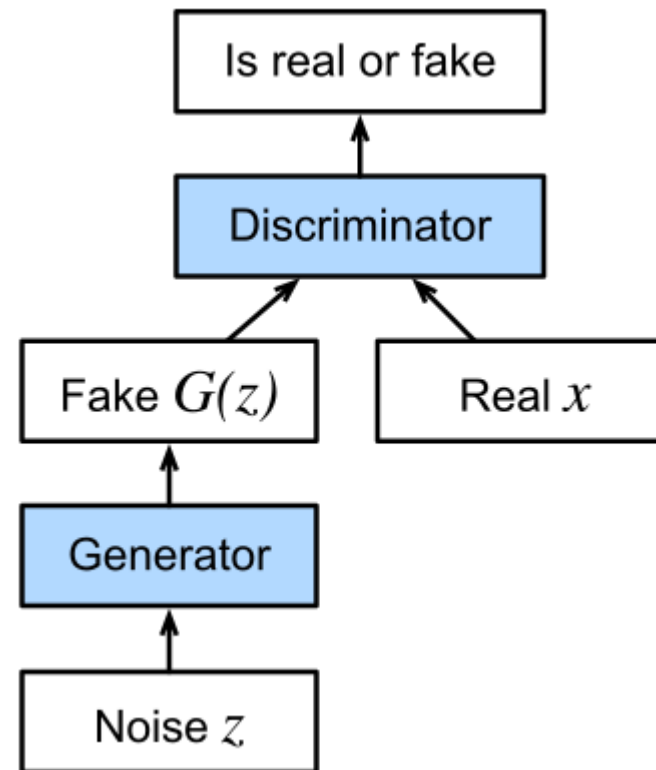


Games in

etition at

Generative Adversarial Networks

- Generative adversarial network (GAN)
- two neural networks contest with each other in the form of a zero-sum game,
 - where one agent's gain is another agent's loss.
 - Discriminator (labeler), Generator (maker)
- GANs are similar to mimicry in evolutionary biology, with an evolutionary arms race between both networks.
- Ex. Someone trains a model to identify AI generated text (or deepfake), so the generator adds that as requirement of output (to pass the detector model)



Onward to ... Bias in AI

Jonathan Hudson
jwhudson@ucalgary.ca
<https://pages.cpsc.ucalgary.ca/~jwhudson/>



UNIVERSITY OF
CALGARY