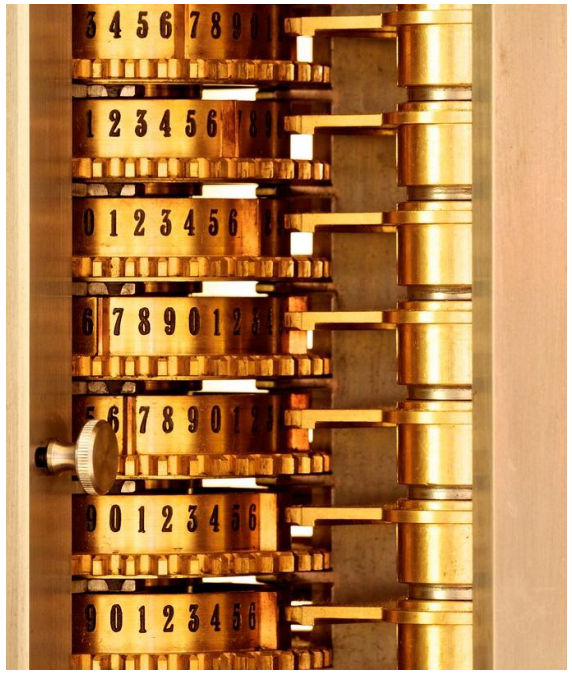
Computers, brains and AI

6: Deep learning and ChatGPT

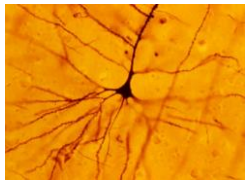




Difference engine by Marcin Wichary / CC BY
Ada Lovelace by Science Museum Group / CC0

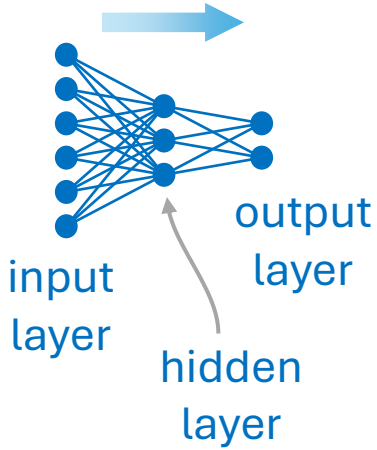
Neural networks …

… are inspired by biology



… are often thought of as blobs and lines

"Blob" is my word.



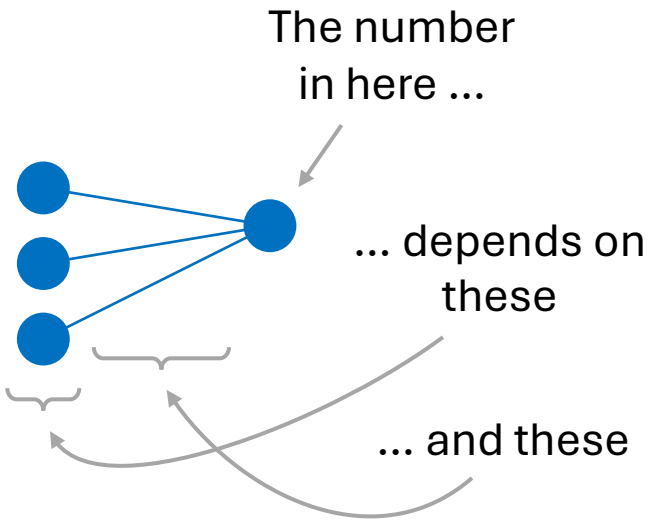
Each blob contains a number

Each line contains a number

… feed forward from the inputs to the outputs

… are usually made of layers

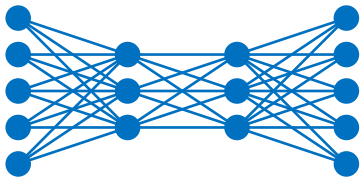
Arranging things neatly in layers makes the GPU calculations a lot simpler and faster.

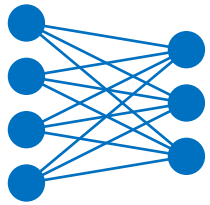


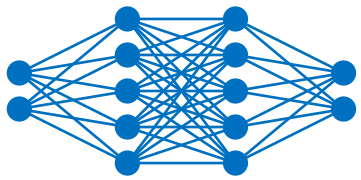
… can be many shapes and sizes

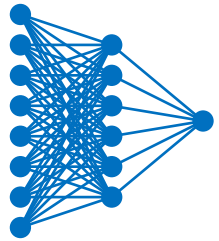
(The shape of a neural network is fixed)

A neural network can't change shape during training or once it has been trained.







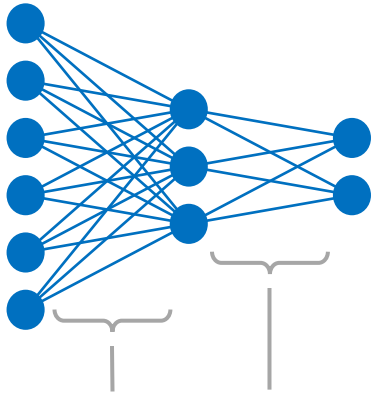


This neural network only has one output, so it isn't a classifier. Instead, it could produce a single prediction (like the temperature from a lot of weather data).

Neuron by Bob Jacobs / CC0

Learning

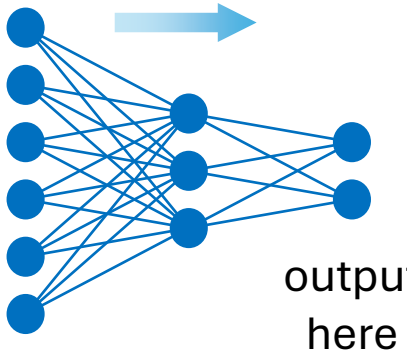
In neural networks, learning is a distinct initial phase...



adjusting these

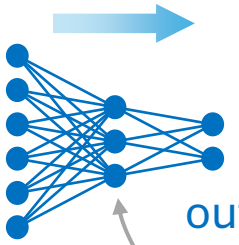
Doing

... before the neural network is put into production.



input here

output here



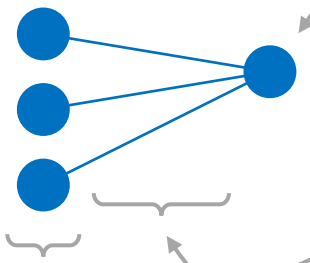
input layer

hidden layer

output layer

Each blob contains a number

Each line contains a number

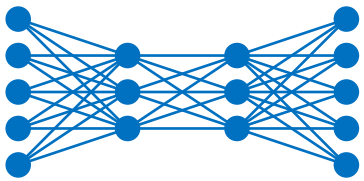


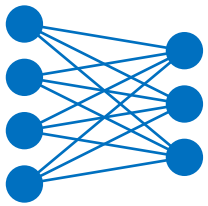
The number in here ...

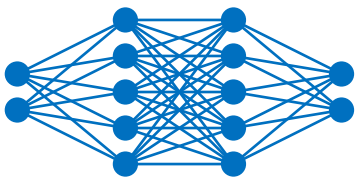
... depends on these

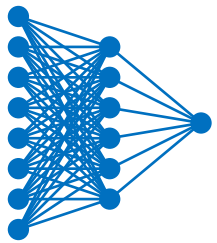
... and these







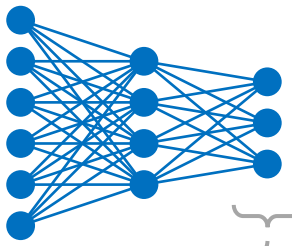




For this illustration, we imagine there are only three categories, namely, dog, cat and bus.

Three categorie

The neural network never sees the words, “dog”, “cat” and “bus”. Instead, it sees a simple numerical classification, 0, 1 or 2.



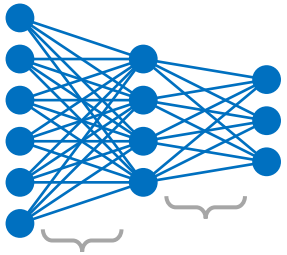
0 (dog)

1 (cat)

2 (bus)

Three “grandmother” cells

Learning

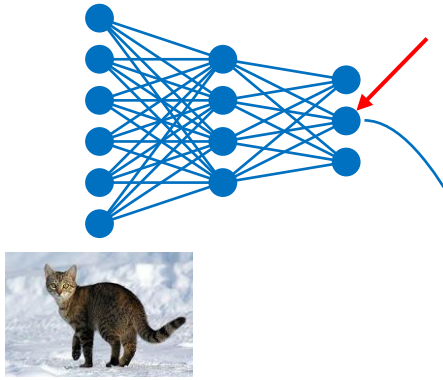


adjusting these

Training data

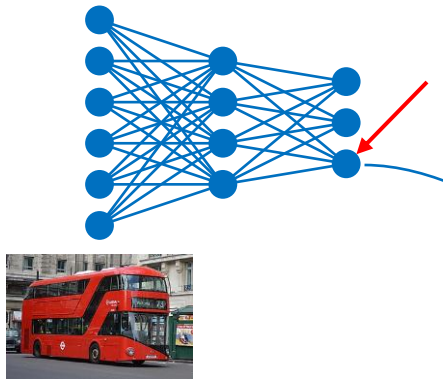
Image	Classification
	1
	2
	1

In the AlexNet contest, there were a million training images.



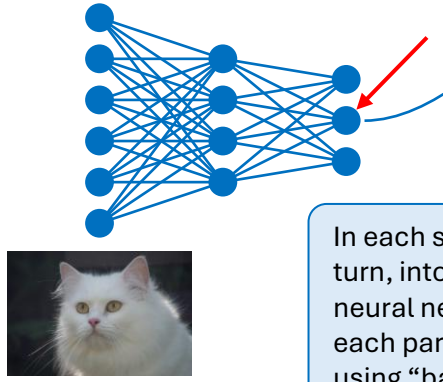
The numerical classifications make it easy for a computer to see how well the neural network is doing.

For example, in this first bit of training data, the answer is supposed to be 1, and so it can directly look at the second grandmother cell indicated by the red arrow. (It is the second one, because computers usually count from zero).



The total loss says how well the neural network is doing.

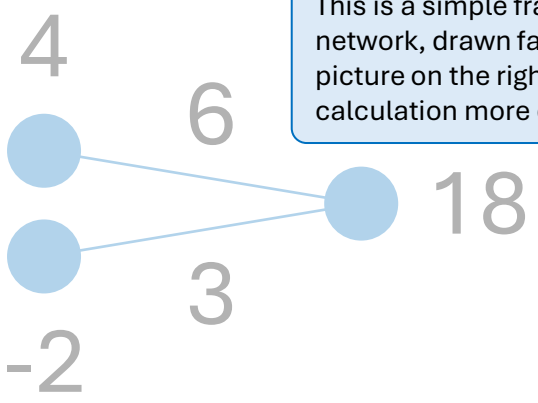
total loss



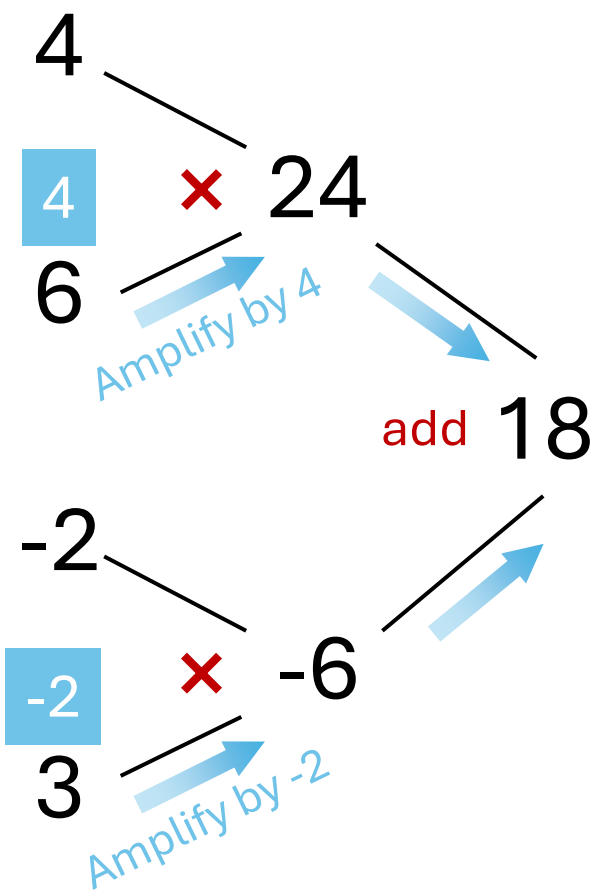
During training, the aim is to keep reducing the total loss, in a series of (expensive) steps. Training stops when the total loss doesn’t seem to be improving any more.

In each step, all one million images are fed, in turn, into the input layer. During training, the neural network keeps a note of what influence each parameter is having on the total loss, using “backpropagation” (see next slide).

Backpropagation



This is a simple fragment from a neural network, drawn faintly because the picture on the right shows the calculation more clearly.

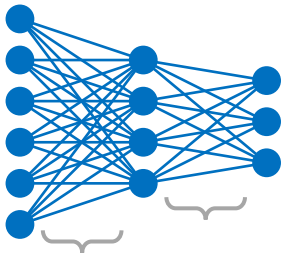


This "4", in blue, is a note made during backpropagation of what the effect of changing the "6" would be. It is saying that increasing the "6" by one would increase "18" by four.

(For people who like calculus, the "4" is the partial derivative with respect to that parameter).

This "18" is the hypothetical final part of the calculation of the loss function.

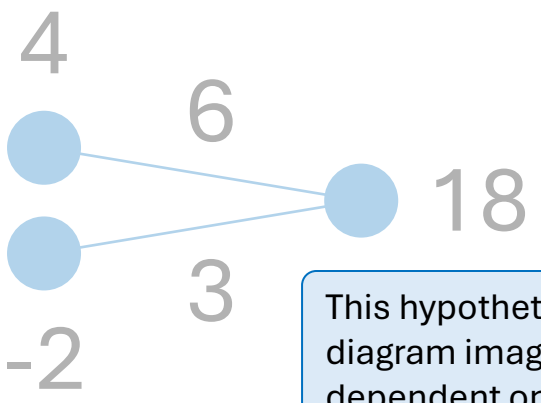
Learning



adjusting stuff

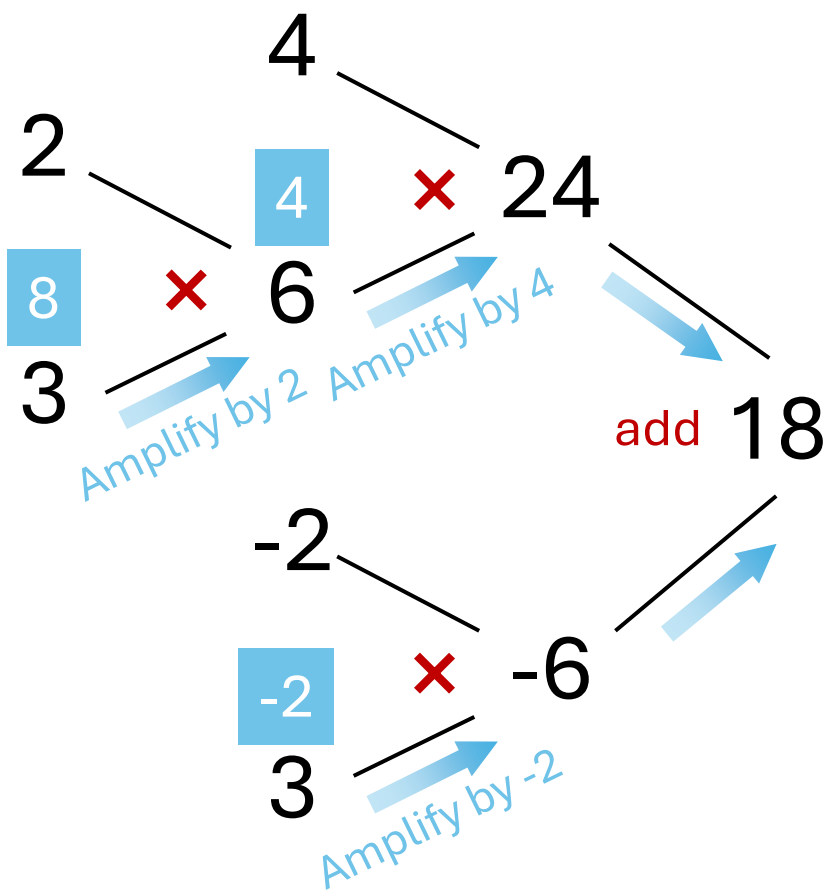
In summary...

With some (fiddly!) bookkeeping, the computer can see whether increasing a parameter would have a positive or negative effect, and therefore whether to adjust it up or down.

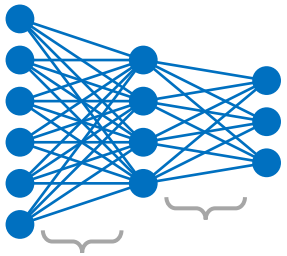


This hypothetical addition to the diagram imagines the “6” is dependent on more calculations.

It shows how the amplification factors are chained together, using the famous “chain rule” invented by Leibniz.



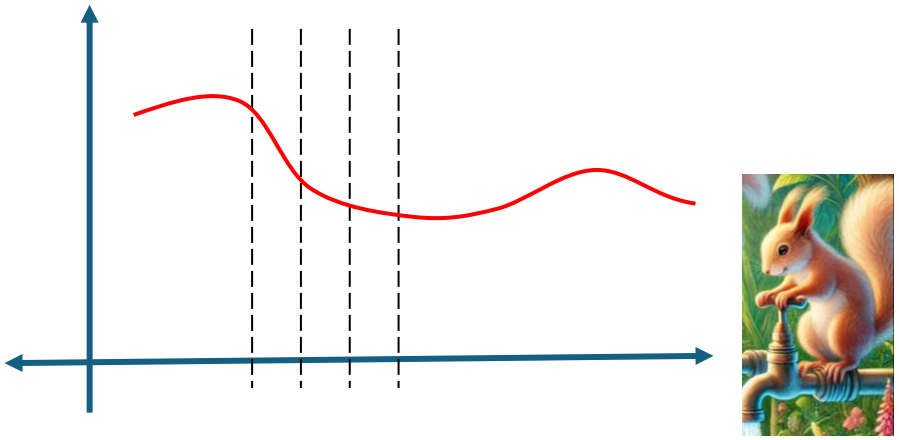
Learning



adjusting stuff

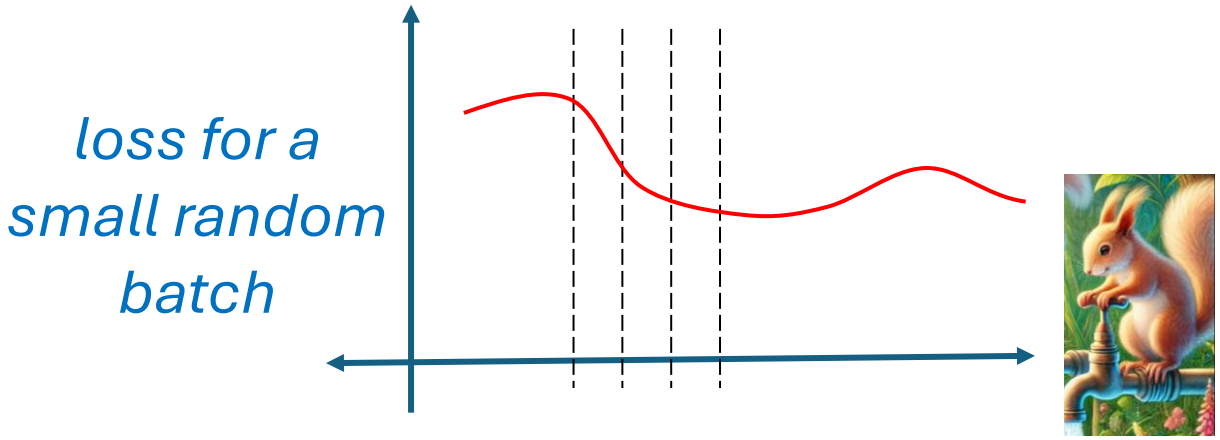
This very popular (but long!) video https://youtu.be/VMj-3S1tku0 has all the details.

Gradient descent



For each step, **gradient descent** calculates the loss using all the training images, a time-consuming exercise.

Stochastic gradient descent



For each step, **stochastic gradient descent** calculates the loss using a small random batch of training images. (For AlexNet, they were batches of 128 images, instead of the whole set of one million images). Each step uses a different random batch.

The steps are much faster, but less accurate and therefore somewhat chaotic. The chaos helps avoid getting stuck in local minima.

Dice by Brett Jordan / Unsplash

Deep learning surprises

It works at all

In the months before the ImageNet context, Hinton's group had papers on neural nets rejected because no one believed it was possible.

It doesn't get stuck in local minima

It generalizes

It takes some skill to train the neural network to generalize, rather locking too much onto the particular training data.



Shockingly good for such a simple idea

Hinton: growing a program

Hinton sometimes says that training a neural network is more like watching something grow than programming a computer.



Learning



A human child can learn the idea of, say, an aardvark from seeing one example.

Artificial neural networks, on the other hand, seem to need hundreds or maybe thousands of examples in order to generalize correctly.

Memorizing

Humans struggle to memorize data (like, for example, phone numbers), whereas computers store data accurately all the time.

Data

Aardvark by Kelly Abram / CC BY 4.0

More surprises

It doesn’t look back

When a neural network is being trained, no notes of the previous steps are taken, and so there isn’t any backtracking.

It doesn’t remember any pictures

During training, it sees the images many times but never stores them. Instead, it simply keeps adjusting its parameters.



More surprises

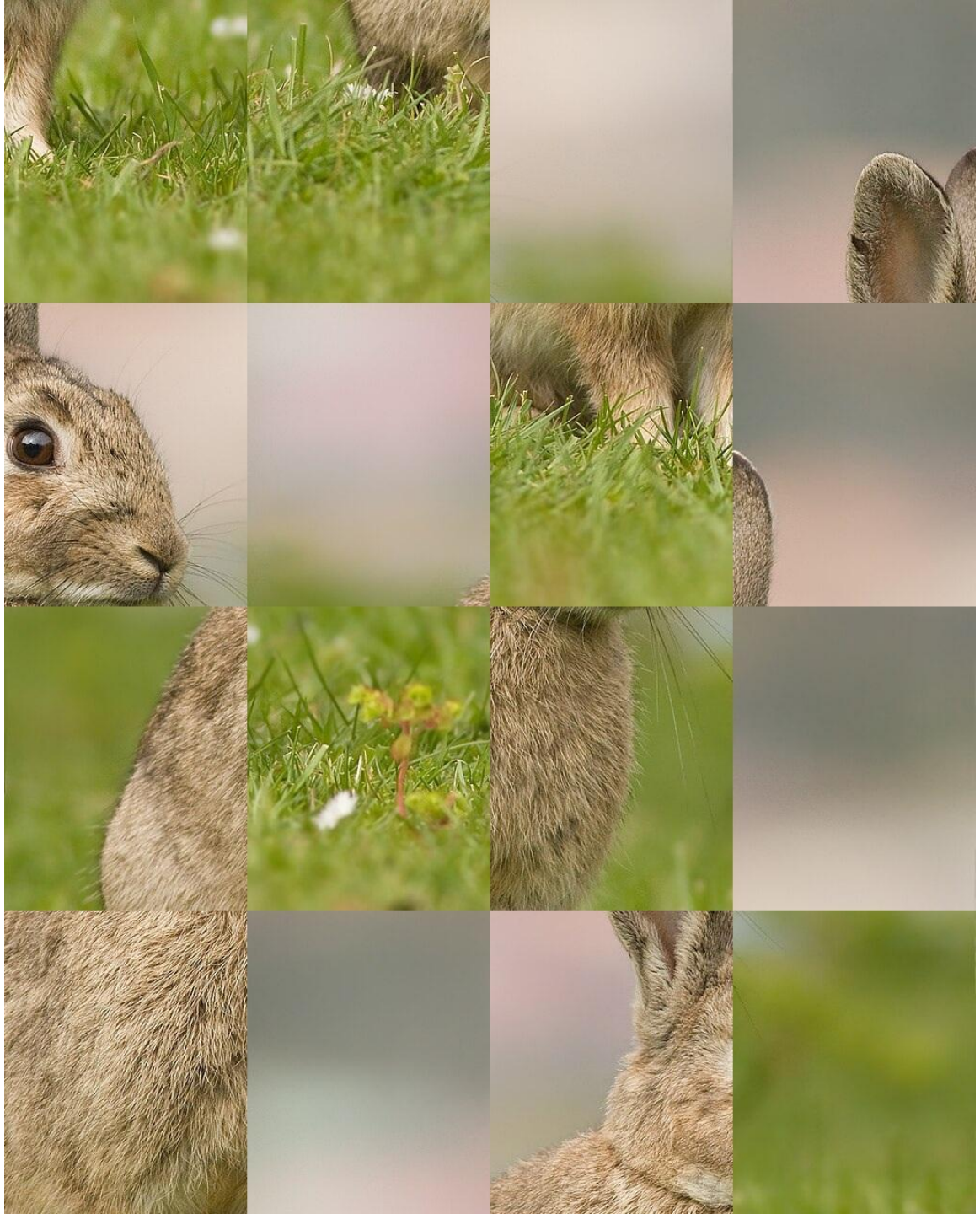
It can’t tell us what a dog is

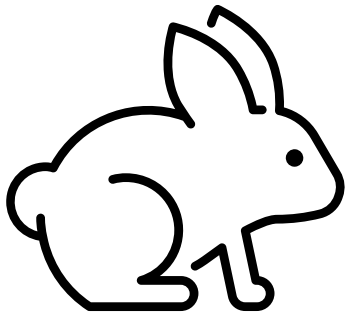
It is instinctive

It slowly learns all the categories at once









... but would struggle with this cartoon.

Unlike a child, a simple, early classifier might have thought this bus is a rabbit. Neural networks were notorious for fixating on low-level details like textures...













No one told AlexNet about textures, shadows, reflections, or that the world is 3D and includes occlusion and lighting.

But somehow, it learnt!



Medical scans

Many deep learning systems aren't told anything about the world (such as shadows or reflections, in this case) and are the better for it. A huge extra benefit is that such **end-to-end** systems can sometimes be retrained on something completely different, such as medical scans.

Earth from Apollo 8 by NASA / CC0
Shadow by © European Union, 2024 / CC BY 4.0
Mirror Lake By Oregon's Mt. Hood Territory / CC0

Is AlexNet intelligent?

Yes

Learns

Learns by itself

No

Narrow

Needs loads of data

Any child can do what it does



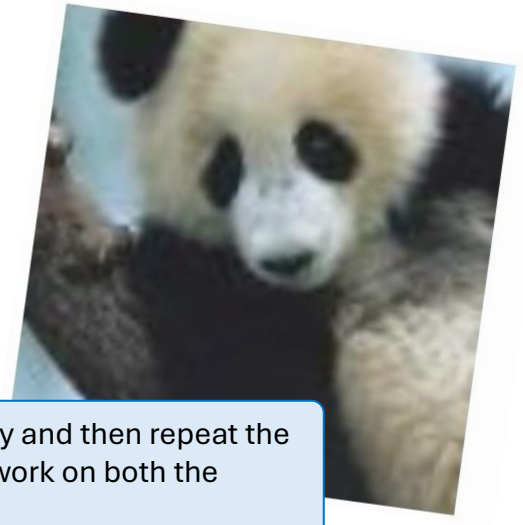
Adversarial machine learning

Shockingly, image classifiers can be fooled with tiny changes. It is shocking, not least, because it demonstrates that machine learning is weird and alien and inscrutable.



\+

carefully chosen tiny changes



=



panda

58% confidence

gibbon

99% confidence

A practical defence is to distort the image slightly and then repeat the classification. The attack is much less likely to work on both the original and the distorted image.

https://arxiv.org/pdf/1412.6572

Ian J. Goodfellow, Jonathon Shlens & Christian Szegedy Google Inc. (2015)

Where is the dog?

How does AlexNet know the furry thing is the dog? For all it knows, the flat expanse of water is the "dog".







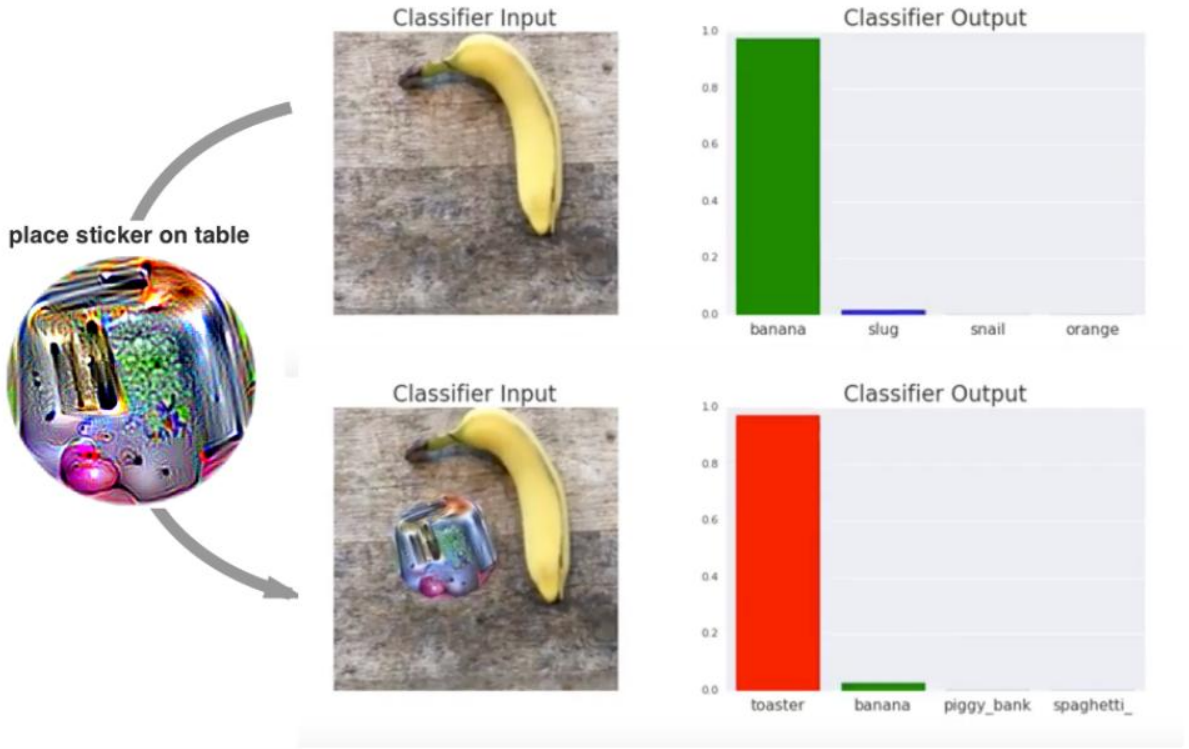
Inceptionism: Going Deeper into Neural Networks ✻

Feature Visualization ✻ How neural networks build up their understanding of images

Appendix: GoogLeNet Visualizations ✻ Visualizations of every channel in GoogLeNet

https://arxiv.org/pdf/1712.09665

Adversarial Patch



https://arxiv.org/pdf/1707.08945

Stop signs

End

chatgpt.com

ChatGPT

chatbot

The chatbot is the thing you talk to.

GPT

The "G" in GPT stands for "Generative", as in generating text. We will see the "P" (pre-trained) and "T" (transformer) later.

Generative

large language model (LLM)

The large language model is the part of the chatbot that knows about language.



If you want to play with ChatGPT, you just go to chatgpt.com in a browser.

These days, you don't even have to log in or sign up, unless you want some of the more advanced features.

chatgpt.com

chat.com

Why is the sky blue?

Here is a screenshot of ChatGPT's reply.

Why is the sky blue?

The sky appears blue due to a phenomenon called **Rayleigh scattering**. This scattering occurs because the Earth's atmosphere contains gases and particles that scatter sunlight. Here's a detailed explanation:

Key Factors

1. **Composition of Sunlight:**
    - Sunlight, or white light, is composed of all the colors of the visible spectrum (red, orange, yellow, green, blue, indigo, and violet).
    - Each color corresponds to a specific wavelength, with blue having shorter wavelengths and red having longer ones.
2. **Scattering in the Atmosphere:**
    - When sunlight enters Earth's atmosphere, it interacts with molecules of nitrogen, oxygen, and other small particles.

The screenshot continues on the next slide.

- Shorter wavelengths (blue and violet) are scattered more than longer wavelengths (red and yellow) because Rayleigh scattering is more effective at shorter wavelengths.

3. **Why Blue and Not Violet?**
    - Even though violet light is scattered more than blue, our eyes are less sensitive to violet. Additionally, some violet light is absorbed by the upper atmosphere.
    - As a result, the scattered light we perceive is predominantly blue.

During Sunrise and Sunset:

- At sunrise and sunset, sunlight passes through a thicker layer of the atmosphere. This causes most of the blue and violet light to be scattered out, leaving the longer wavelengths (reds and oranges) more visible, creating vibrant hues.

In summary, the blue appearance of the sky is a result of the preferential scattering of shorter wavelengths of light by the molecules in the atmosphere.

Why is the sky blue?

Here are some excerpts from ChatGPT's reply.

The sky appears blue due to a phenomenon called **Rayleigh scattering**. • • •

• • •

The blue dots are where I've cut most of the text.

Key Factors

The reply is fairly formal. It's like talking to Wikipedia!

1. **Composition of Sunlight:**

We will come back to the tone later.

• • •

~~The sky is usually grey...~~

It isn't pedantic. So, for example, it doesn't point out that the sky is usually grey or ask us which planet we are talking about!

Here is an excerpt from the formal reply from before.

Key Factors

1. **Composition of Sunlight:**

ChatGPT must figure out what “this” refers to.

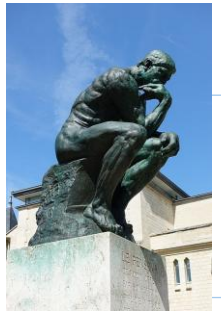
Why is the sky blue?

How would you explain this to a five year old?

In this reply, the tone has completely changed.

The sky looks blue because of sunlight and the air around us. Sunlight has all the colors of a rainbow, but when it shines through the air, the blue part gets scattered all around. That’s why, when you look up, the sky looks blue!

It’s like if you splash water, the little drops go everywhere. The blue color in sunlight "splashes" around in the sky more than the other colors!

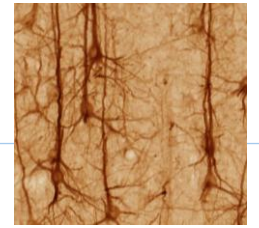


GOFAI

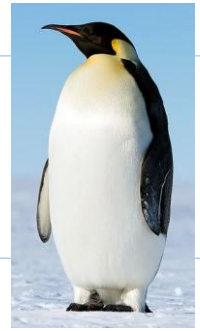
Machine learning

1960

ELIZA



1980





Good old fashioned AI

Birds fly





2000

AlexNet



2012

The AI hype went crazy when ChatGPT arrived.

ChatGPT November 2022

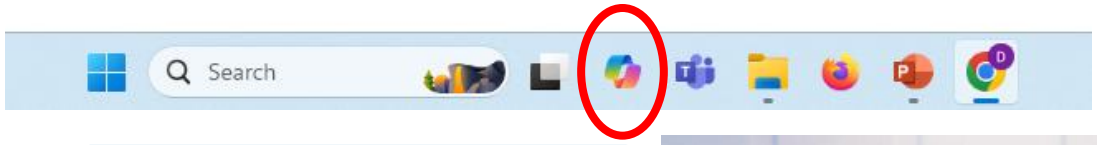
2020

Chess
Jeopardy
Book / CC0
Swan from pxhere / CC0
Bus by Martin49 / CC BY 2.0
Cat by Von.grzanka / CC BY-SA 3.0
Dog by Kyfishfishfish / CC BY-SA 4.0
The thinker by CrisNYCa / CC BY-SA 4.0
Tractor by Kharkiv Tractor Plant / CC BY 4.0
Penguin by Christopher Michel / CC BY 2.0
Neurons by UC Regents Davis campus / CC BY-3.0
Duckling by Luis Miguel Bugallo Sánchez / CC BY-SA 3.0

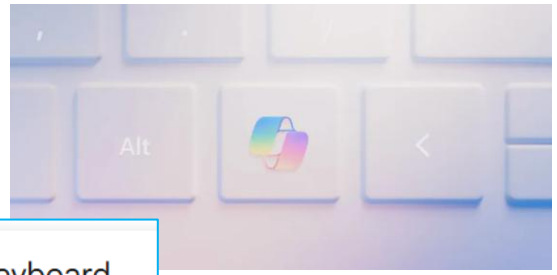
Here are some notable chatbots.

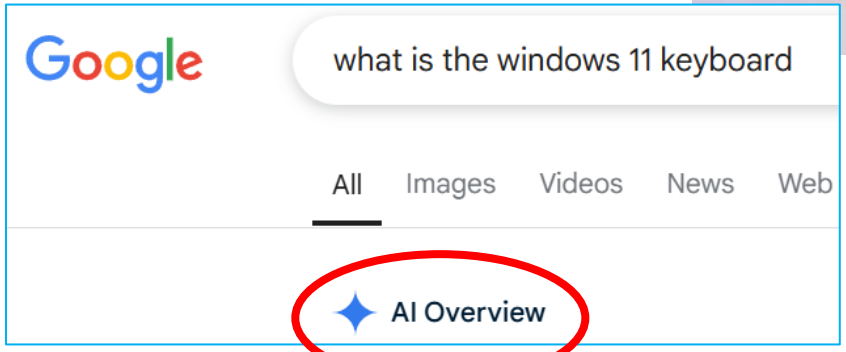
OpenAI	ChatGPT
Microsoft	Copilot
Google	Gemini
Anthropic	Claude
DeepSeek	V2, V3, R1

Copilot has insinuated itself onto this taskbar! I don't remember asking it to do this!

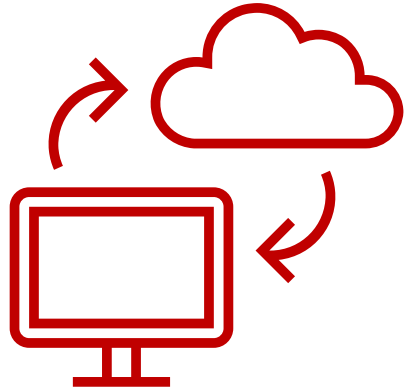


It also pops up all the time in Word and PowerPoint and even appears on some keyboards.





The big models run in the cloud, but small ones can run on your local computer, or even a Raspberry Pi.



Recently, Google search has started offering AI Overviews that are generated from its Gemini large language model.

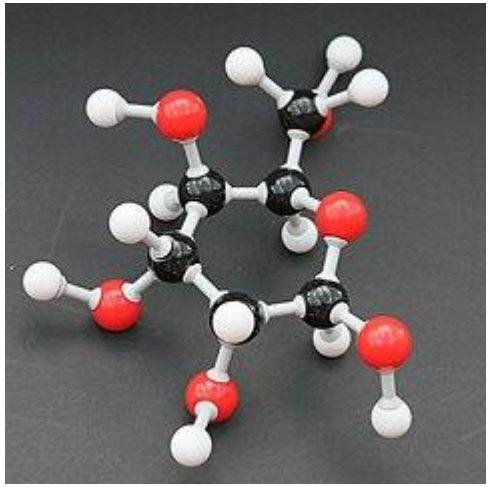
ChatGPT *chatbot*

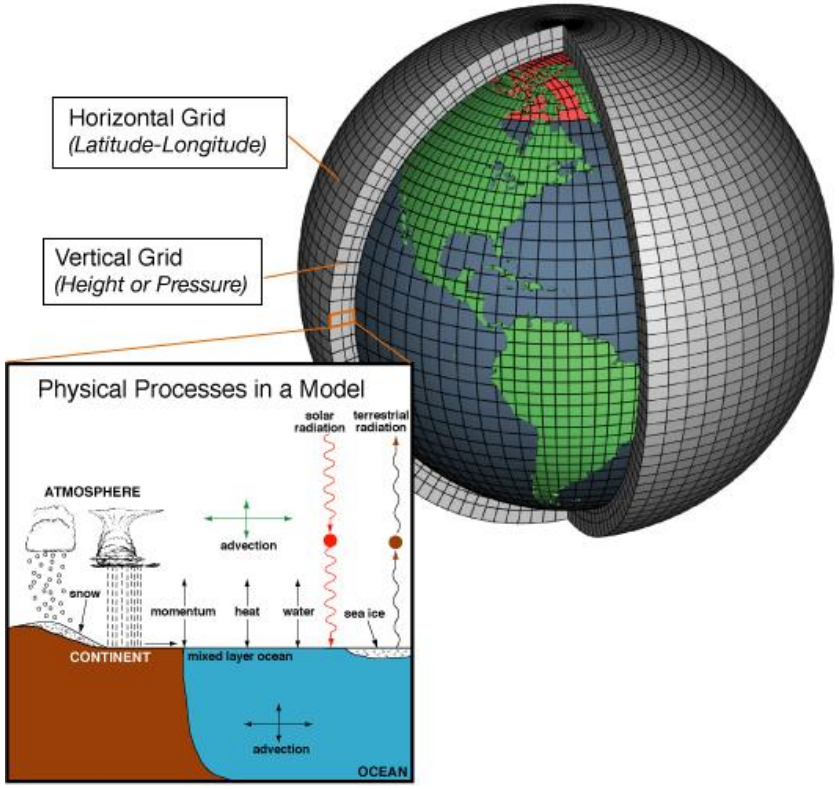
GPT

large language model (LLM)

Large language model

A model is an informative representation of something, either physical like the coloured balls of the molecule or abstract like the weather model.





Weather model by NOAA / CC0
Model by Bin im Garten / CC BY-SA 3.0

Large language model

the dog saw the ball ✓✓

the ball saw the dog ✓

saw ball the dog the ✗

In the era of GPT, language models say how likely bits of text are.

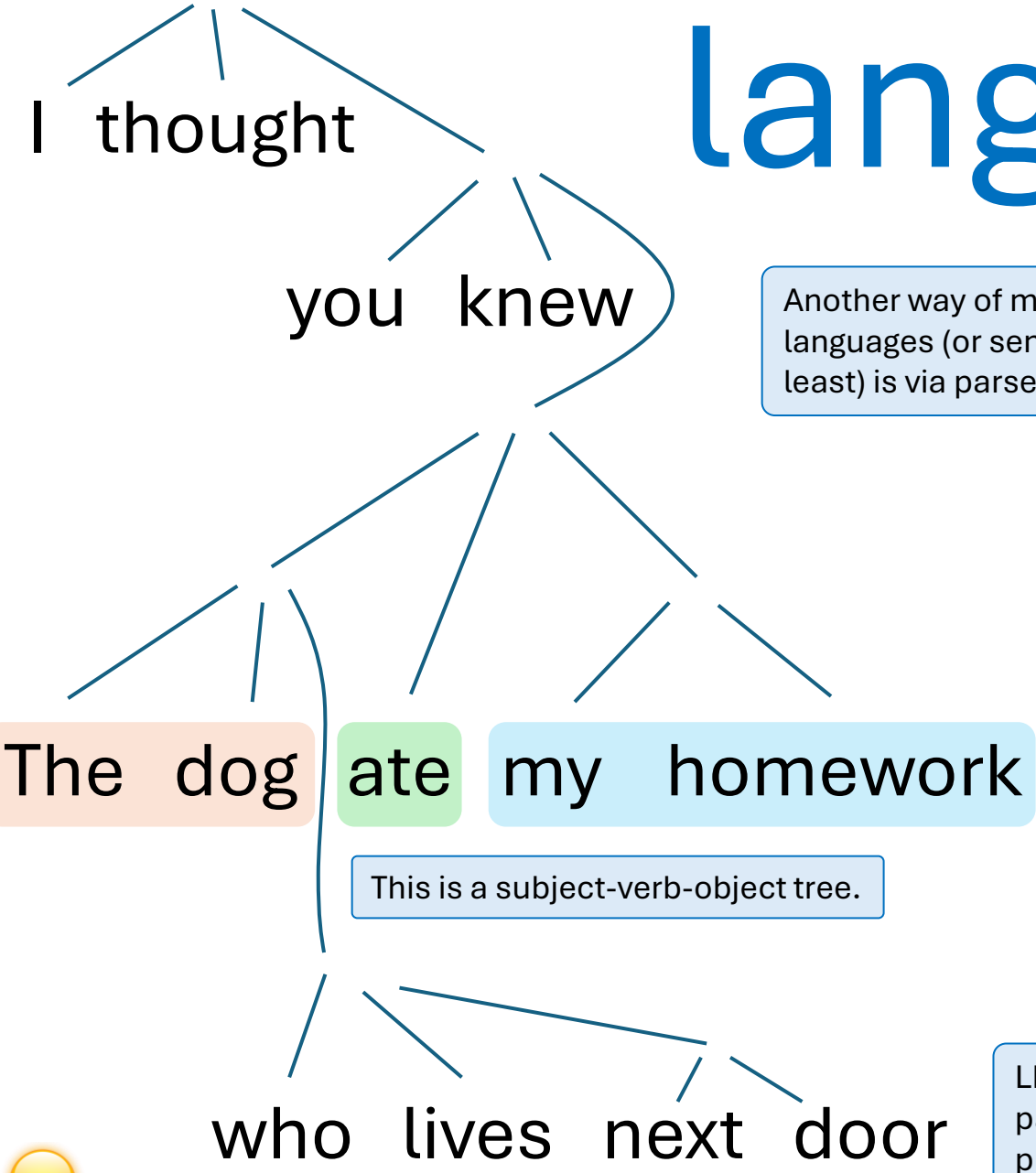
“The ball saw the dog” is unlikely (unless the ball is a cartoon character).

We (somehow!) instantly know that “saw ball the dog the” doesn’t make sense (maybe, in the same we instantly know the rabbit is a rabbit).



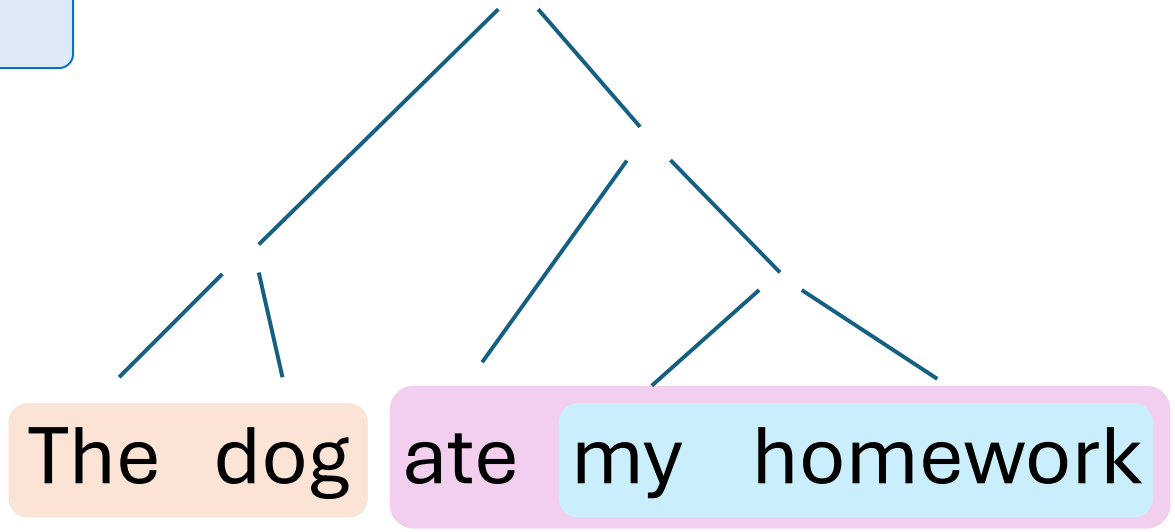
Rabbit by JJ Harrison / CC BY-SA 3.0

language model



Another way of modelling languages (or sentences, at least) is via parse trees.

This is a subject-verb-object tree.



This is a topic-predicate tree.

LLMs aren't told about grammar and parse trees. Instead, they learn purely from looking at a lot of text.

Large language model

a quick brown dog ✓✓

dog quick a brown ✗

apples and pears ✓✓

apples and pairs ✓

the big bad ???f

"Apples and pairs" is less likely, but not impossible because of phrases like "pairs of apples and pairs of bananas".

However, for an LLM to favour "apples and pears", it has to know things about the world, including fruit.

An early use of language models would have been speech recognition systems trying to decide whether to write "pears" or "pairs"...

... or trying to guess "wolf" from a muffled audio.



Large language model

Why is the sky blue?

- The sky is blue because ..... ✓✓
- The capital of France is ..... ✗

Language models must go beyond individual sentences. For example, “the capital of France” continuation would seem a bit odd.

In other words, language models must model more than just grammar.



Large language model



parameters

AlexNet had 60 million **parameters**. (Parameters are the things a neural network needs to learn during its training).

AlexNet

60 million

GPT-1

117 million

2018

In its day, GPT-2 was briefly considered dangerously powerful, but then just one year later they made an LLM over 100 times larger!

GPT-2

1.5 billion

2019

GPT-3 was a scaled-up version of **GPT-2**, which in turn was a scaled- up version of **GPT-1**.

GPT-3

175 billion

2020

The architecture of **GPT-4** is a secret but is believed to have over a trillion parameters.

The scaling laws