

Deep Learning

Advent of Scale

Bryan Cardenas
Robert Jan Schlimbach
Caspar van leeuwen

High Performance Machine Learning Group

SURF

Why now?

Whose Revolution?

01.

Data Amount & Availability

Data generated by our devices continuously

02.

Compute Power

Easier available in the cloud
Frameworks to work with
accelerated hardware

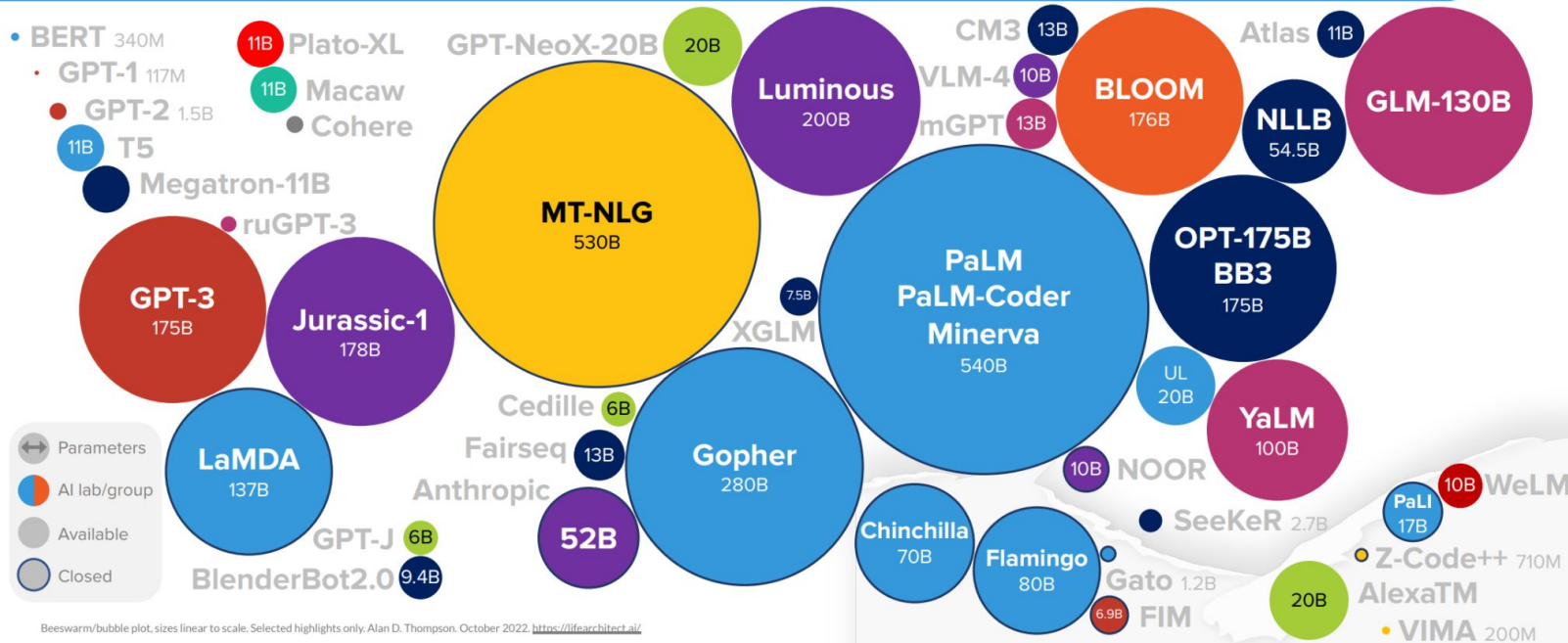
03.

Research & Funds

Define the function to
approximate
your desired solution

Large Neural Networks

LANGUAGE MODEL SIZES TO OCT/2022



The Speed of AI

2016.



A vintage photo of a cat

The Speed of AI

2016.



A vintage photo of a cat

2019.



A laptop sitting in a chair
next to a big TV

The Speed of AI

2016.



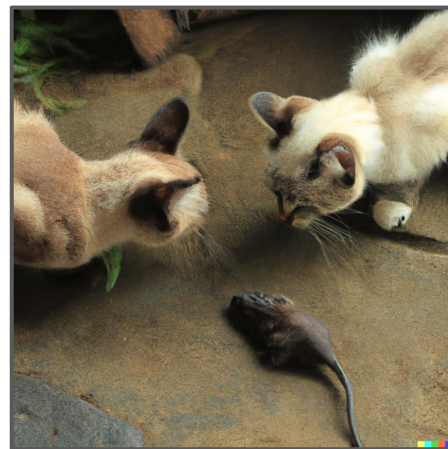
A vintage photo of a cat

2019.



A laptop sitting in a chair
next to a big TV

2022.



2 cats arguing about who's
gonna eat the mouse on
the floor

The Speed of AI



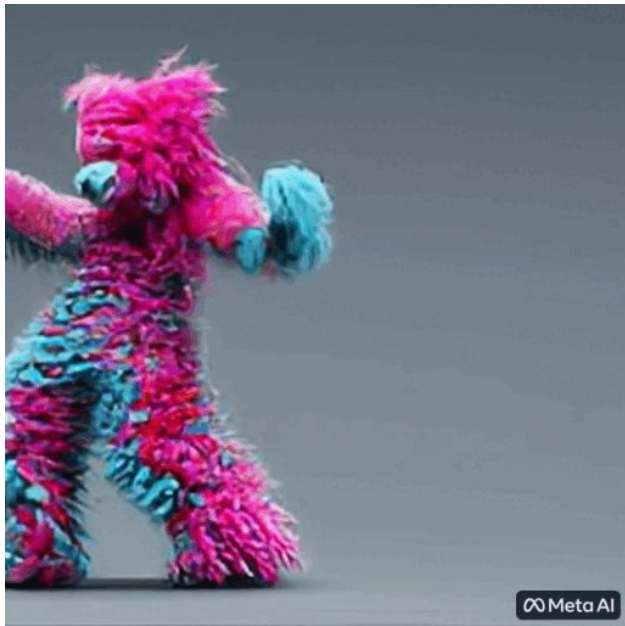
A teddy bear painting
a portrait

The Speed of AI



A fluffy baby sloth with an orange knitted hat trying to figure out a laptop close up highly detailed studio lighting screen reflecting in its eye

The Speed of AI

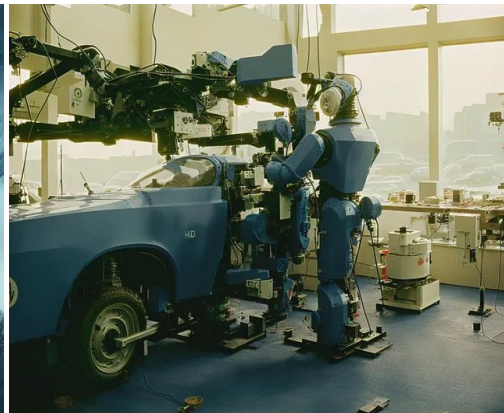


Stable Diffusion & DALL-E

Trained on 500M+ image and text pairs

Stability AI used 4000 A100 GPUs

But I can run the model on my home computer with at least 4 GB VRAM



How to handle large data?

Attention is all you need

(vaswani et al 2017)

Transformers

$$X \times W^Q = Q$$

$$X \times W^K = K$$

$$X \times W^V = V$$



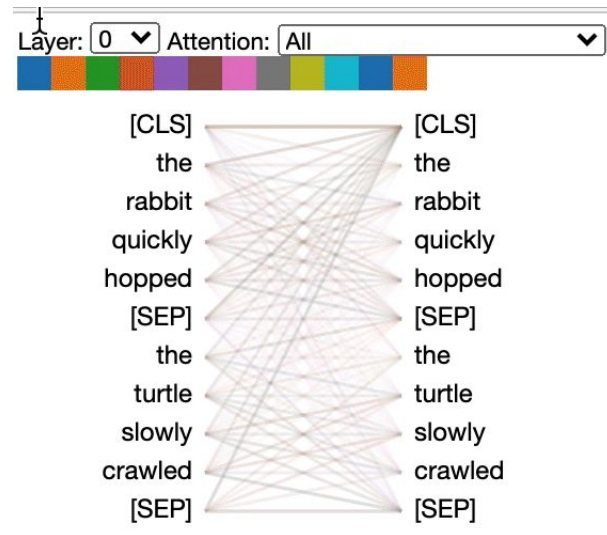
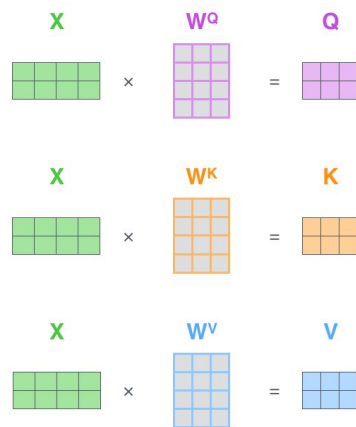
$$\text{softmax} \left(\frac{Q \times K^T}{\sqrt{d_k}} \right) \times V = Z$$

How to handle large data?

Attention is all you need

(vaswani et al 2017)

Transformers



$$\text{softmax}\left(\frac{Q \times K^T}{\sqrt{d_k}}\right) \times V$$

=

Z

Matrix Multiplications

AlphaTensor

Discover Algorithm to matrix multiply in fewer steps

Finds how to multiply 4x4 matrices in 47 steps instead of 49 steps

Able to improve in enormous action spaces

30 orders of magnitude larger than the game of Go

$$\begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} \times \begin{bmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \end{bmatrix} = \begin{bmatrix} c_{1,1} & c_{1,2} \\ c_{2,1} & c_{2,2} \end{bmatrix}$$

Standard algorithm

$$h_1 = a_{1,1} b_{1,1}$$

$$h_2 = a_{1,1} b_{1,2}$$

$$h_3 = a_{1,2} b_{2,1}$$

$$h_4 = a_{1,2} b_{2,2}$$

$$h_5 = a_{2,1} b_{1,1}$$

$$h_6 = a_{2,1} b_{1,2}$$

$$h_7 = a_{2,2} b_{2,1}$$

$$h_8 = a_{2,2} b_{2,2}$$

$$c_{1,1} = h_1 + h_3$$

$$c_{1,2} = h_2 + h_4$$

$$c_{2,1} = h_5 + h_7$$

$$c_{2,2} = h_6 + h_8$$

Strassen's algorithm

$$h_1 = (a_{1,1} + a_{2,2})(b_{1,1} + b_{2,2})$$

$$h_2 = (a_{2,1} + a_{2,2})b_{1,1}$$

$$h_3 = a_{1,1}(b_{1,2} - b_{2,2})$$

$$h_4 = a_{2,2}(-b_{1,1} + b_{2,1})$$

$$h_5 = (a_{1,1} + a_{1,2})b_{2,2}$$

$$h_6 = (-a_{1,1} + a_{2,1})(b_{1,1} + b_{1,2})$$

$$h_7 = (a_{1,2} - a_{2,2})(b_{2,1} + b_{2,2})$$

$$c_{1,1} = h_1 + h_4 - h_5 + h_7$$

$$c_{1,2} = h_3 + h_5$$

$$c_{2,1} = h_2 + h_4$$

$$c_{2,2} = h_1 - h_2 + h_3 + h_6$$

Matrix Multiplications

AlphaTensor

Discover Algorithm to matrix multiply in fewer steps

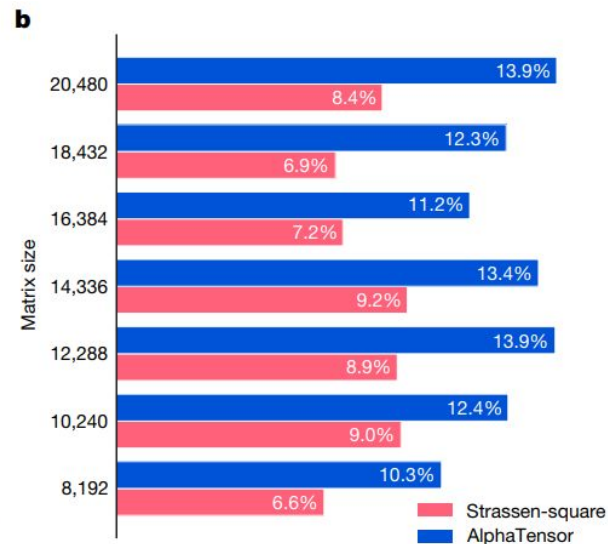
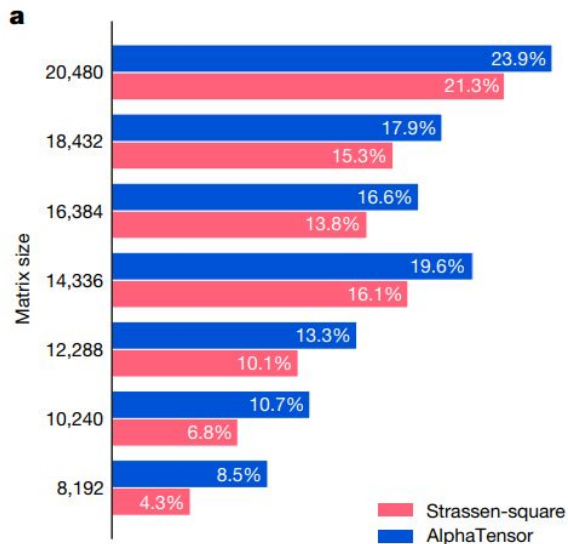
Finds how to multiply 4x4 matrices in 47 steps instead of 49 steps

Able to improve in enormous action spaces

30 orders of magnitude larger than the game of Go

Trained on 64 TPU V3 for a week

Transformer reigns supreme



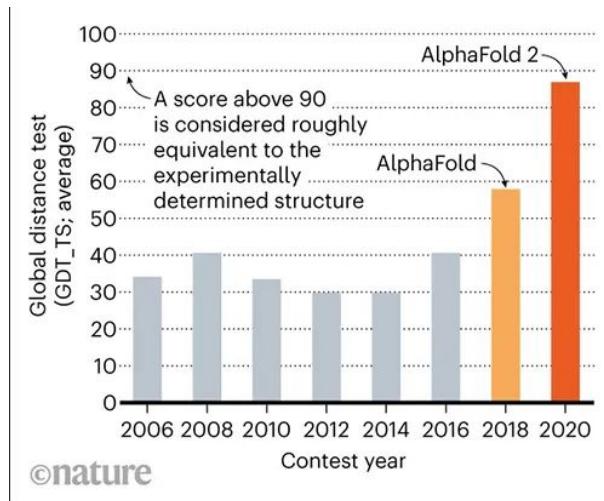
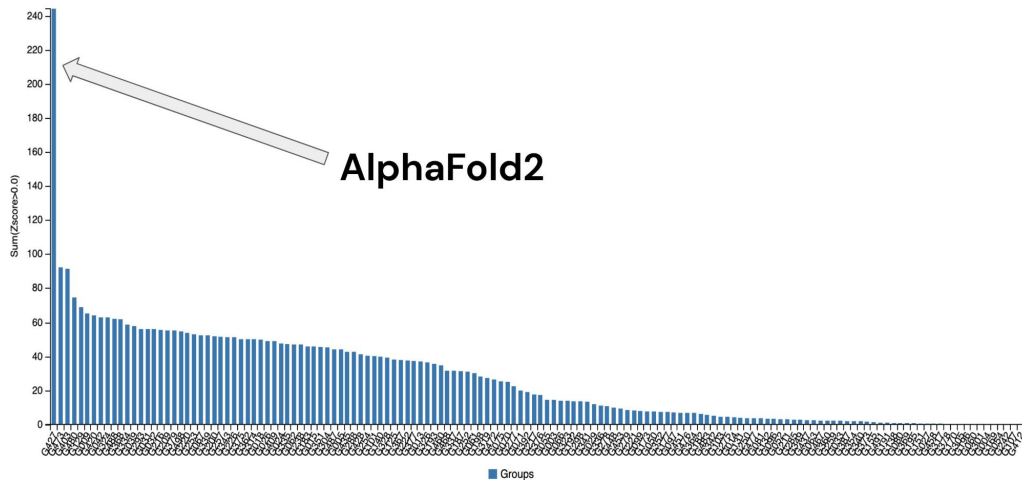
Protein Folding

AlphaFold2

How does a protein fold in space?

Crushed CASP14 scores

Trained on 128 TPUs for two weeks



Protein Folding

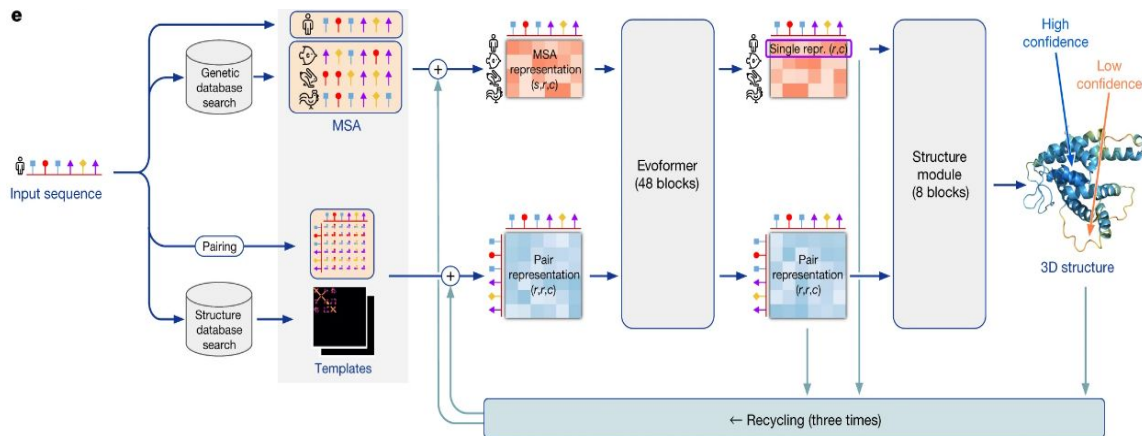
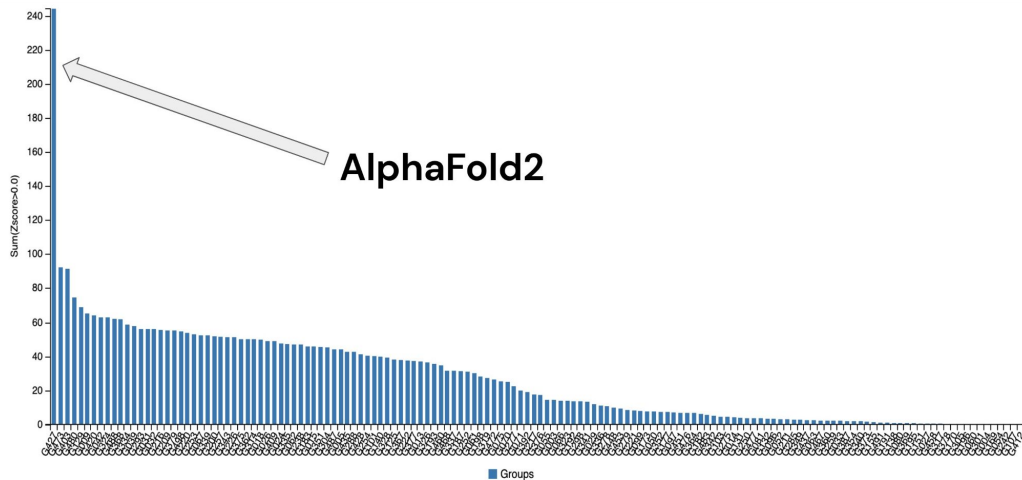
AlphaFold2

How does a protein fold in space?

Crushed CASP14 scores

Trained on 128 TPUs for two weeks

Evoformer is a transformer based module



Protein Folding

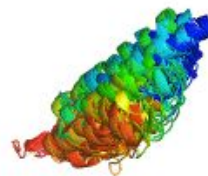
AlphaFold2

How does a protein fold in space?

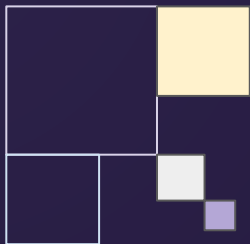
Crushed CASP14 scores

Trained on 128 TPUs for two weeks

Evoformer is a transformer based module



1



Thank You

