Brief History of AI and ChatGPT

賴冠廷教授
Prof. K. T. Lai
台北科技大學電子工程系
2024/2/20

AI Origin:

1956 Dartmouth Conference: The Founding Fathers of AI

John MacCarthy

Marvin Minsky

Claude Shannon

Ray Solomonoff

Alan Newell

Herbert Simon

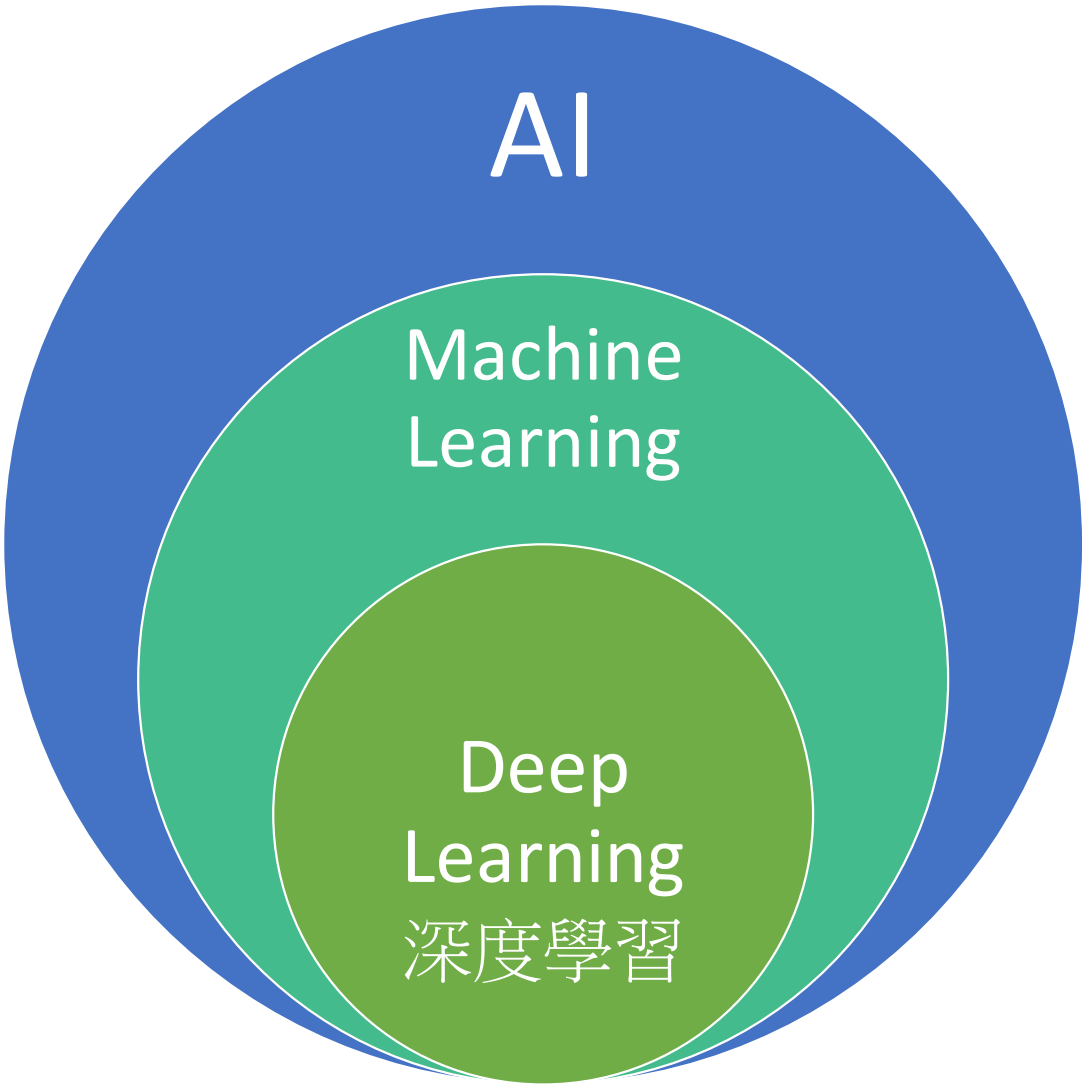
Arthur Samuel

Oliver Selfridge

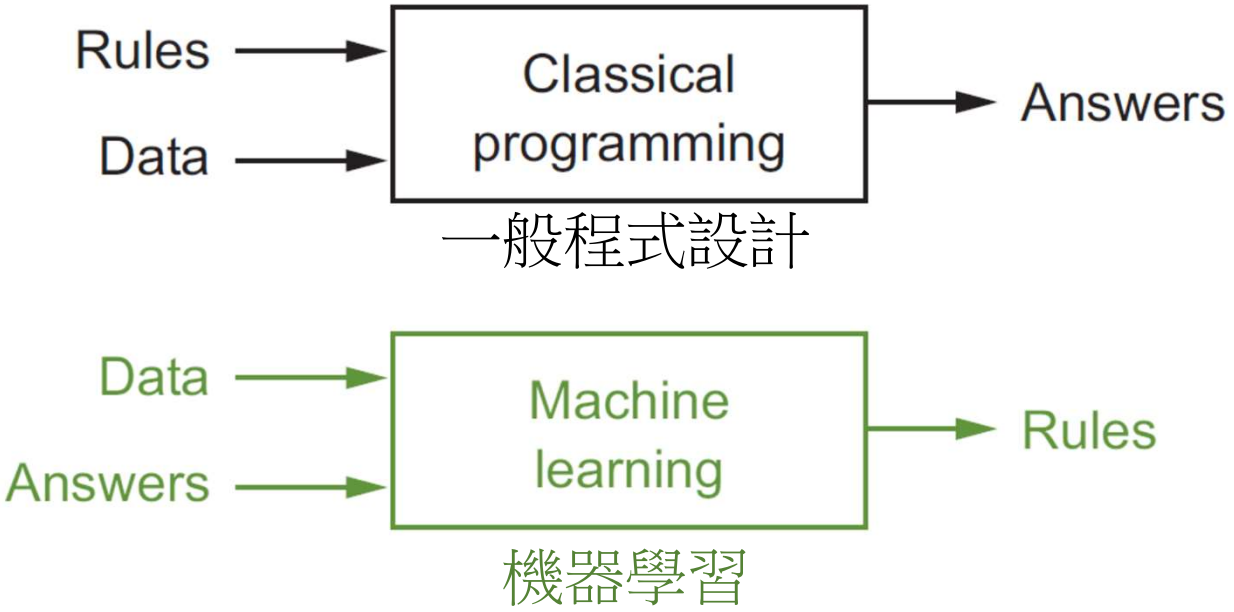
Nathaniel Rochester

Trenchard More

Courtesy of scienceabc.com

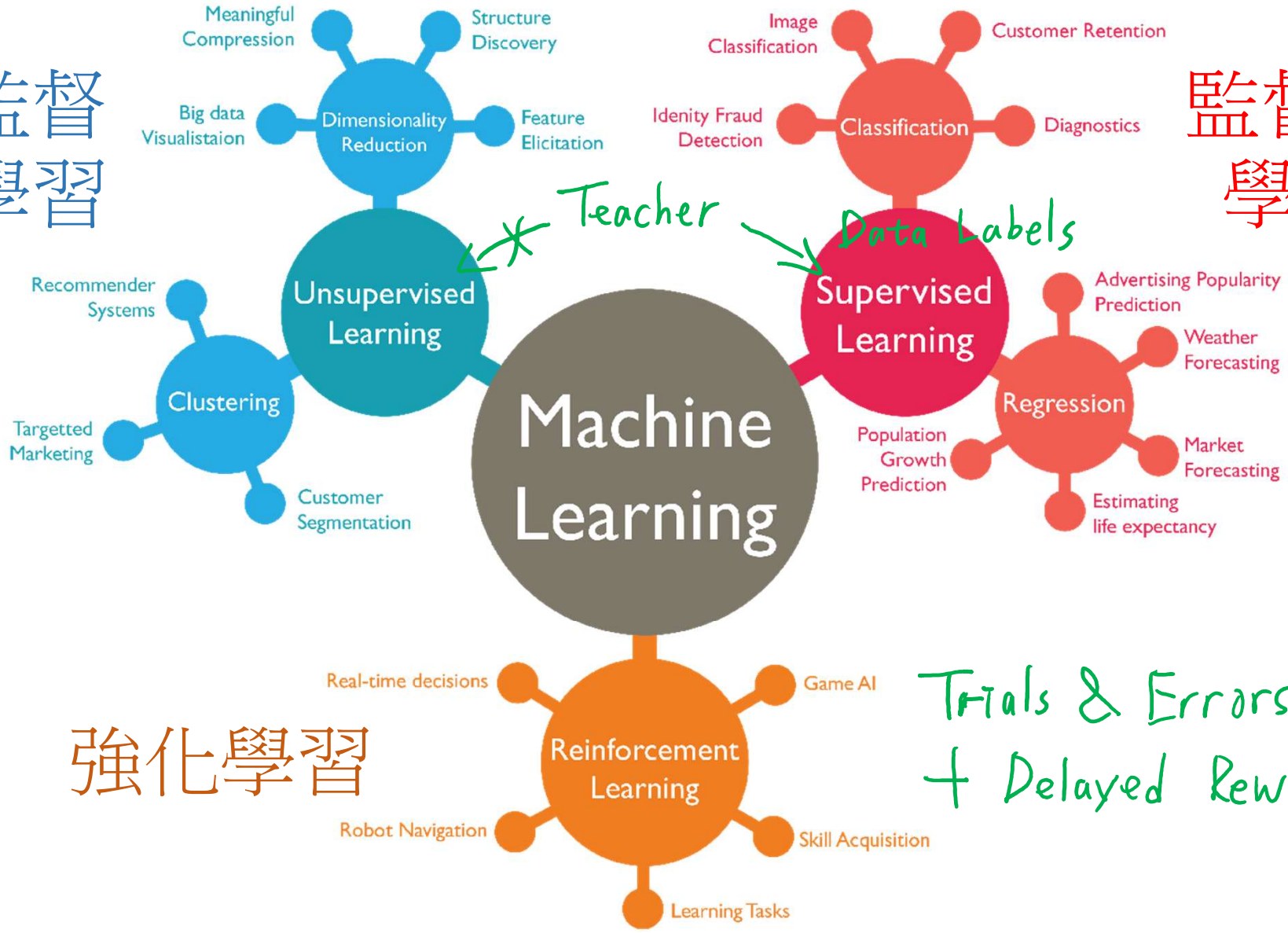

AI
Machine Learning
Deep Learning
深度學習

Machine Learning (Statistical Learning)
機器學習 vs. 程式設計



Francois Chollet, “Deep Learning with Python,” Manning, 2017

非監督式學習



監督式學習

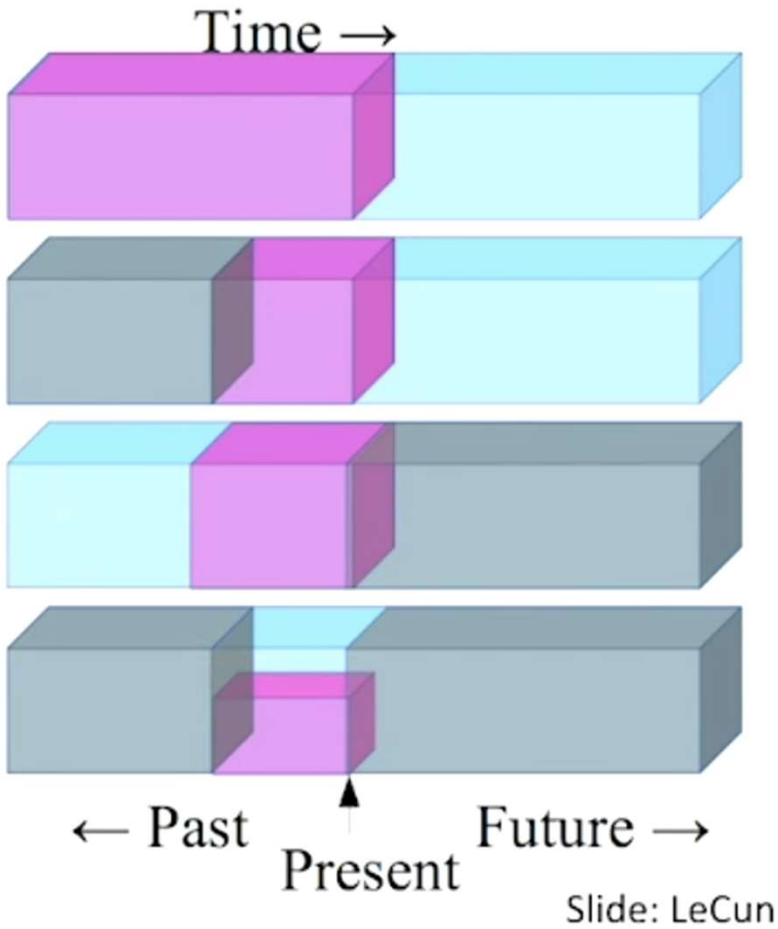
強化學習

Trials & Errors + Delayed Reward


Data
Labelers

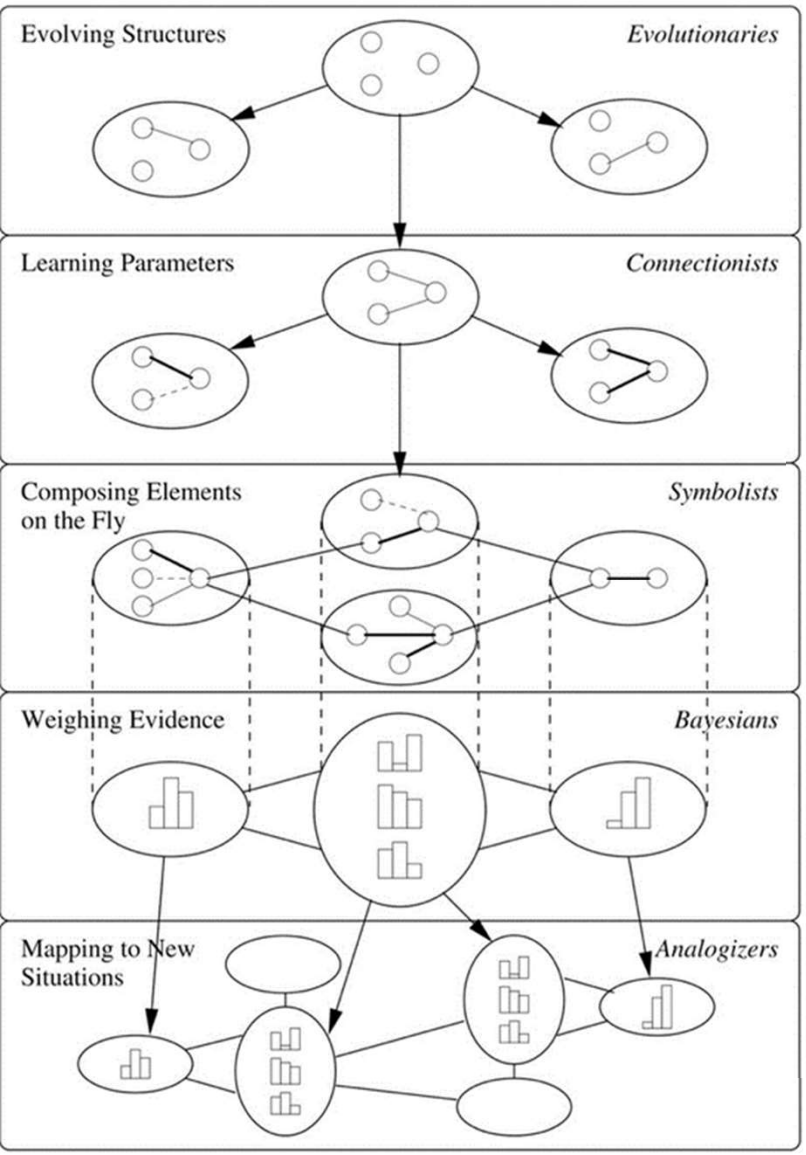
Self-Supervised Learning

- Predict any part of the input from any other part.
- Predict the future from the past.
- Predict the future from the recent past.
- Predict the past from the present.
- Predict the top from the bottom.
- Predict the occluded from the visible
- Pretend there is a part of the input you don't know and predict that.



Slide: LeCun

- https://www.youtube.com/watch?v=7I0Qt7GALVk



5 Tribes of Machine Learning

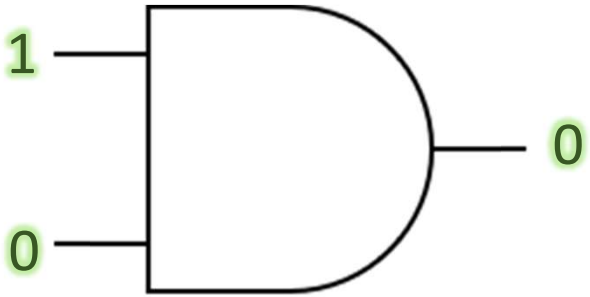
- Evolutionaries (演化法)
- Connectionists (類神經網路)
- Symbolists (歸納法)
- Bayesians (貝氏機率)
- Analogizers (類比近似)

The Master Algorithm – Pedro Domingos

5 Tribes of Machine Learning

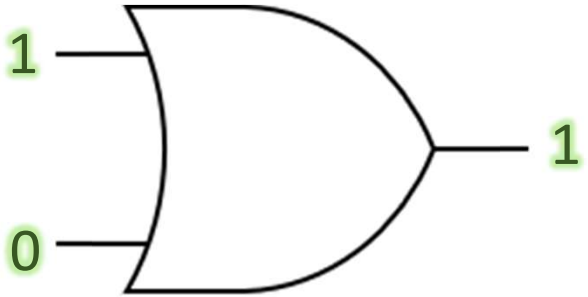
- Symbolists: Decision Trees, Random Forest
- Bayesians: Naïve Bayesians
- Analogizers: SVM, k-NN
- Evolutionaries: Gene algorithms
- Connectionists: Deep Learning

All Algorithms can be Reduced to 3 Operations!



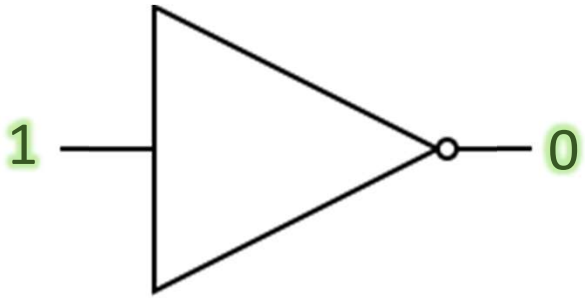
AND

A	B	Output
0	0	0
1	0	0
0	1	0
1	1	1



OR

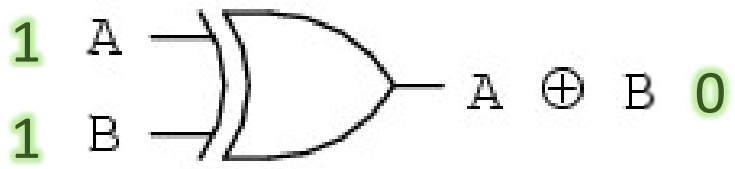
A	B	Output
0	0	0
1	0	1
0	1	1
1	1	1



NOT

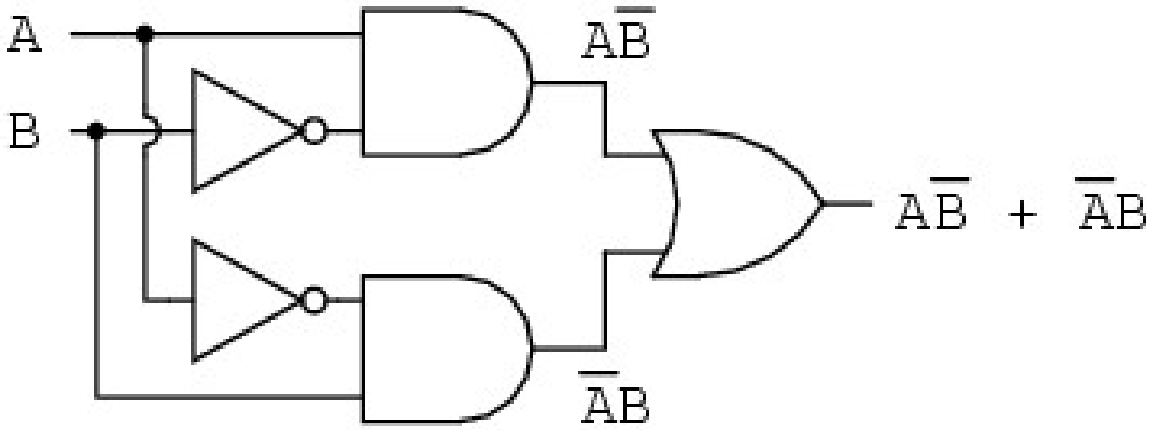
Input	Output
0	1
1	0

XOR

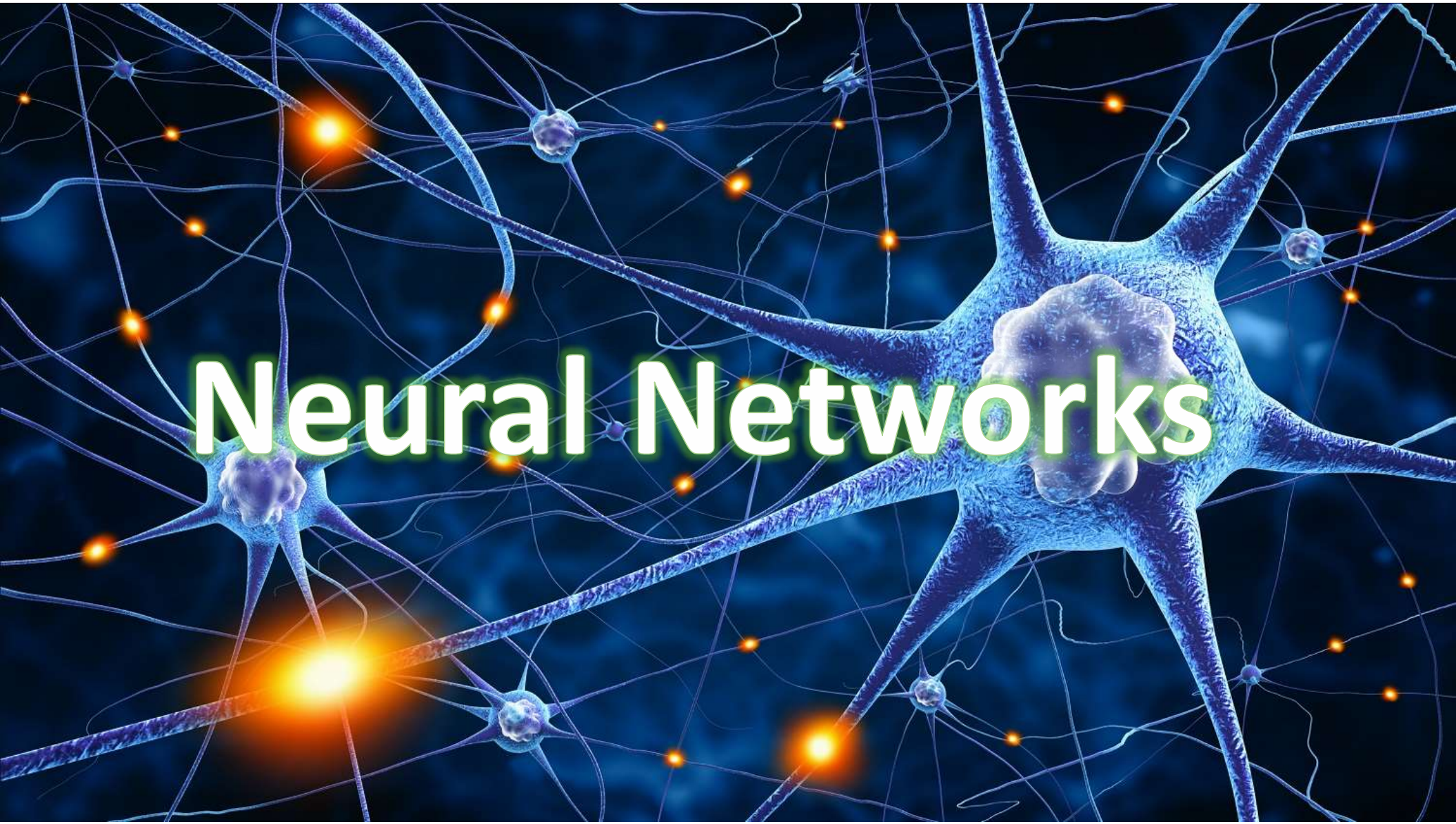


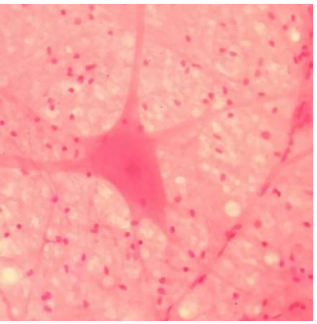
. . . is equivalent to . . .

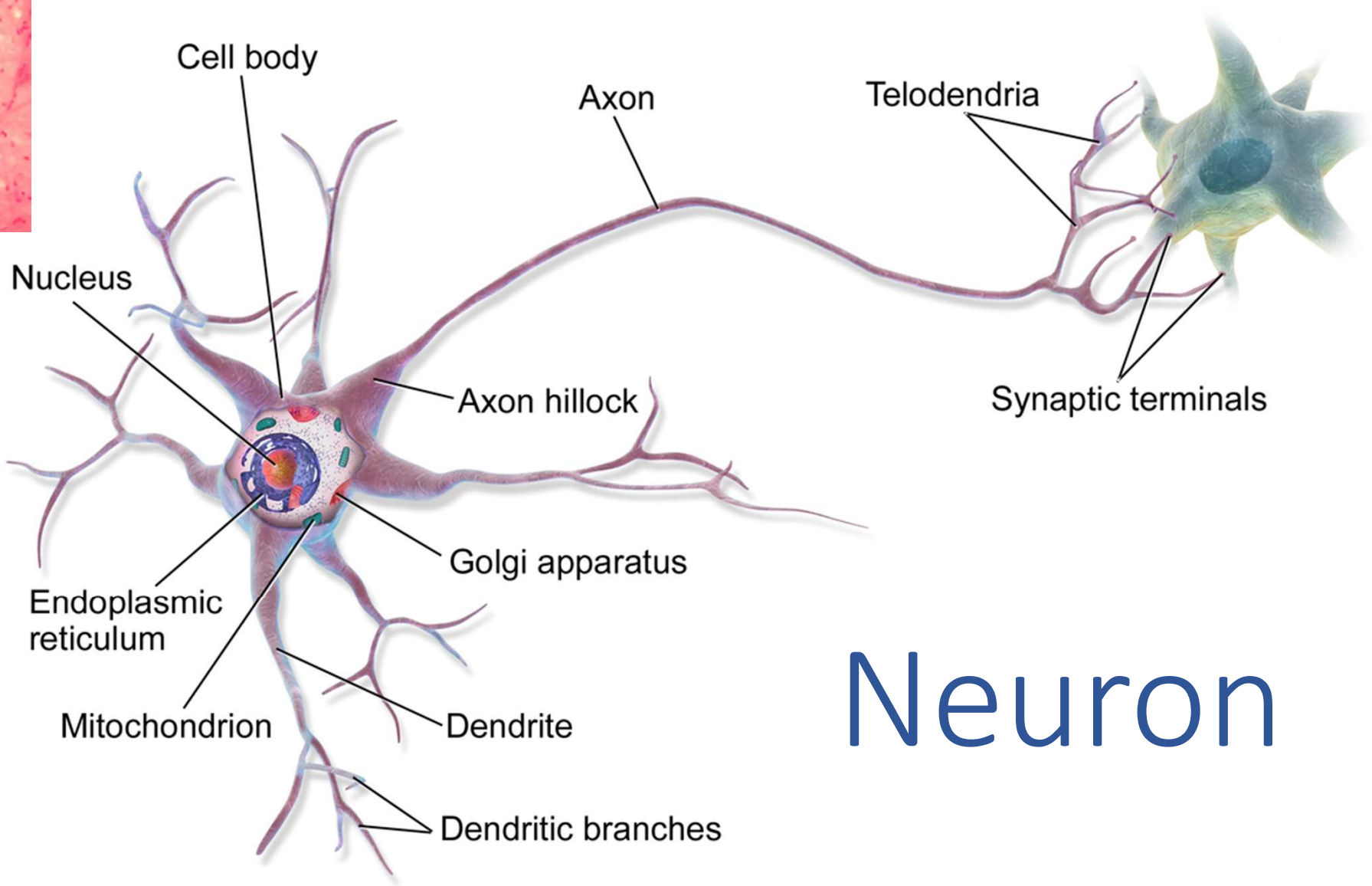
A	B	Output
0	0	0
1	0	1
0	1	1
1	1	0



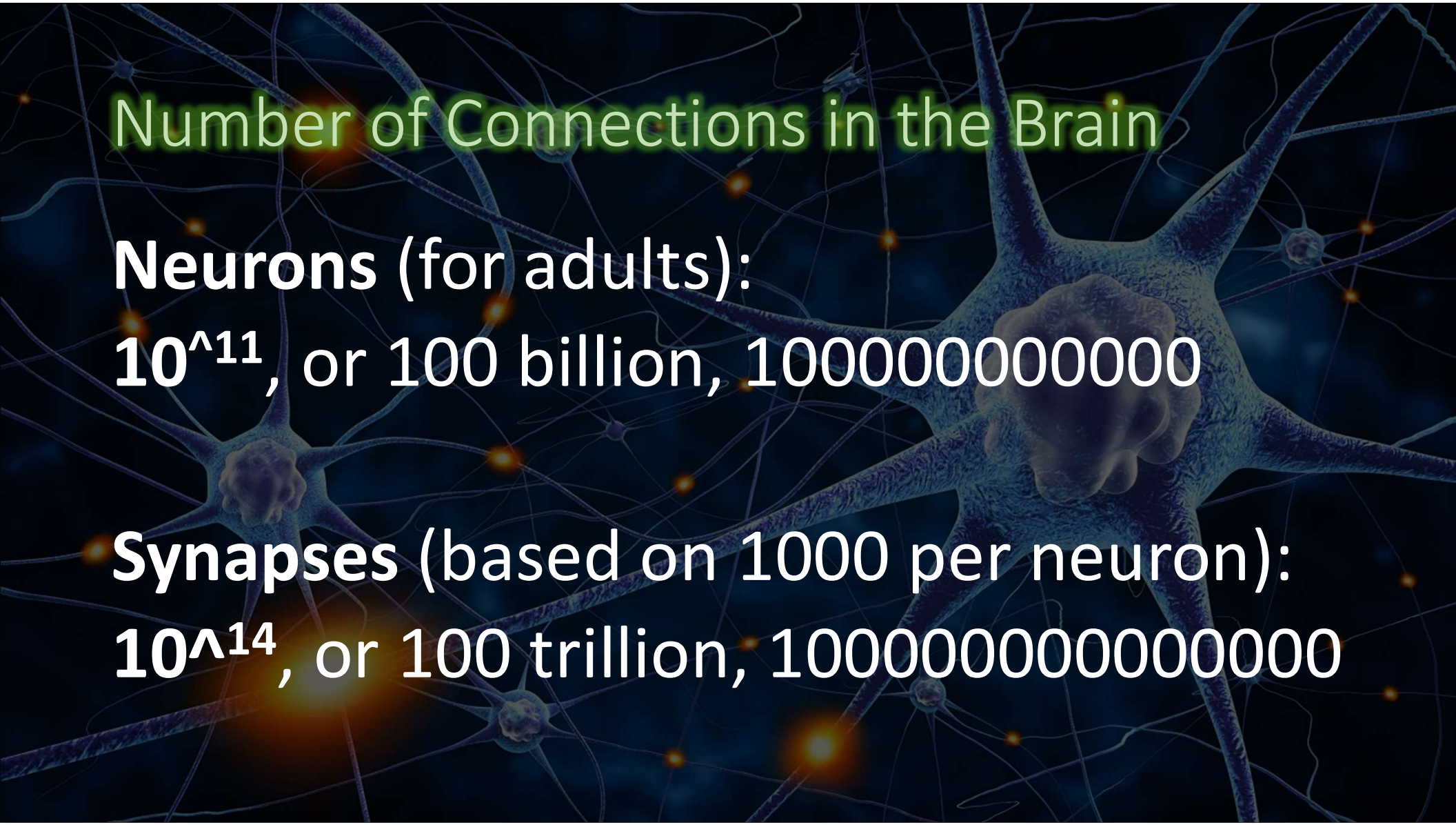
$$A \oplus B = A\overline{B} + \overline{A}B$$


Neural Networks




Cell body
Axon
Telodendria
Nucleus
Axon hillock
Synaptic terminals
Golgi apparatus
Endoplasmic reticulum
Mitochondrion
Dendrite
Dendritic branches

Neuron

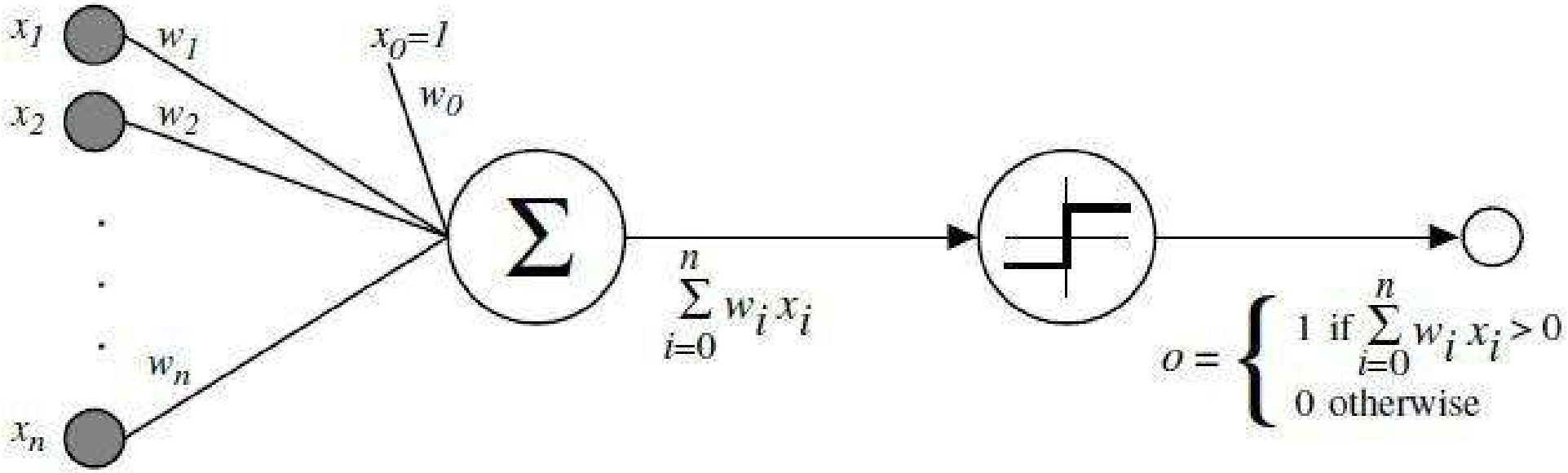


Number of Connections in the Brain

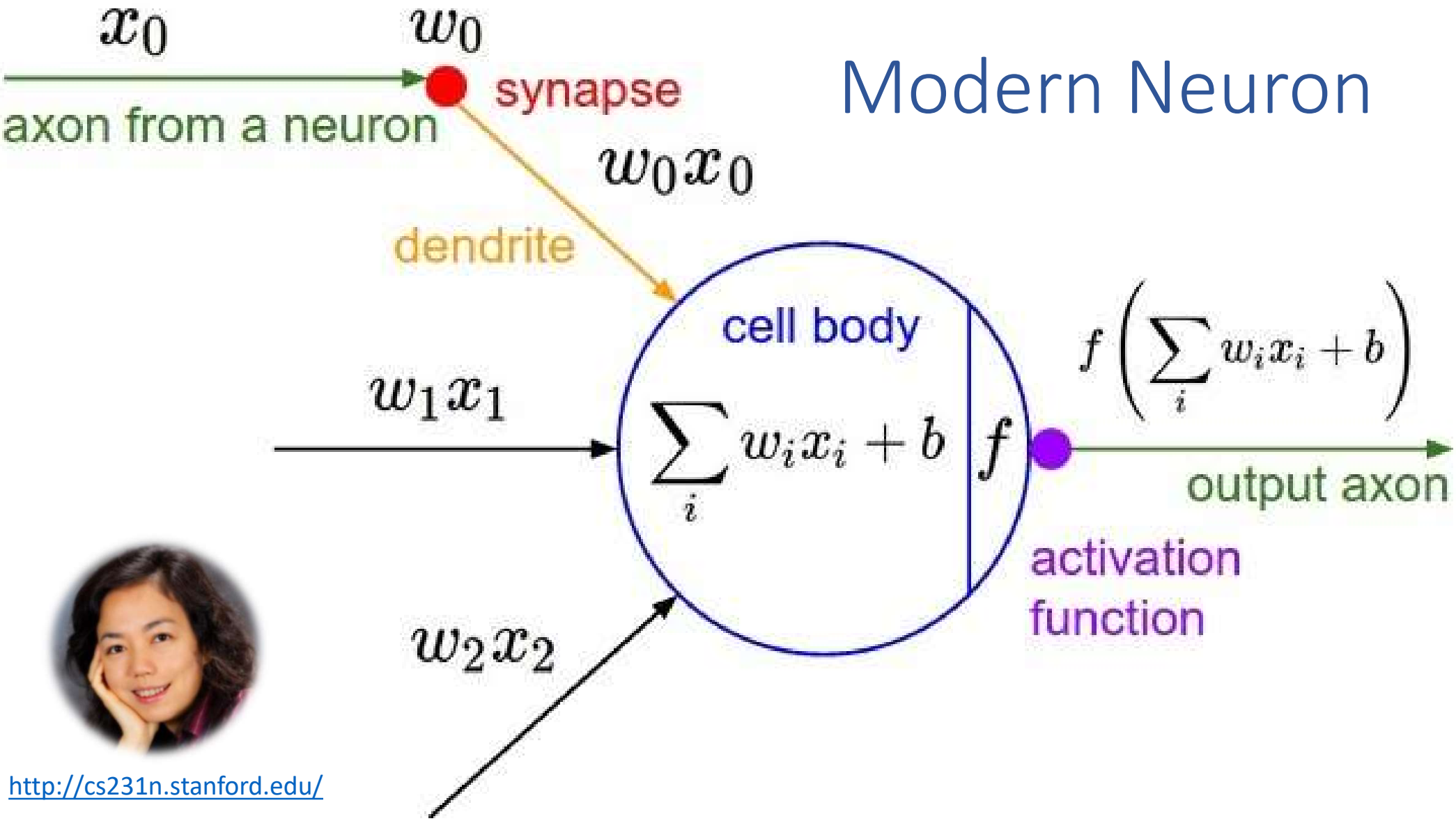
Neurons (for adults):
10^11, or 100 billion, 100000000000

Synapses (based on 1000 per neuron):
10^14, or 100 trillion, 100000000000000

Frank Rosenblatt's Perceptron (1957)

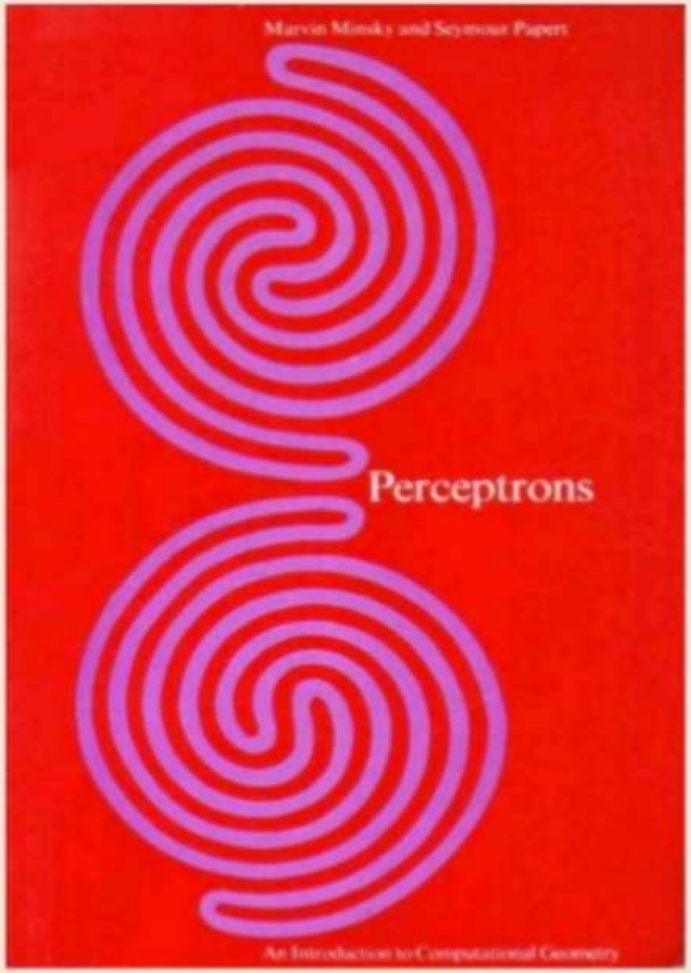


Modern Neuron

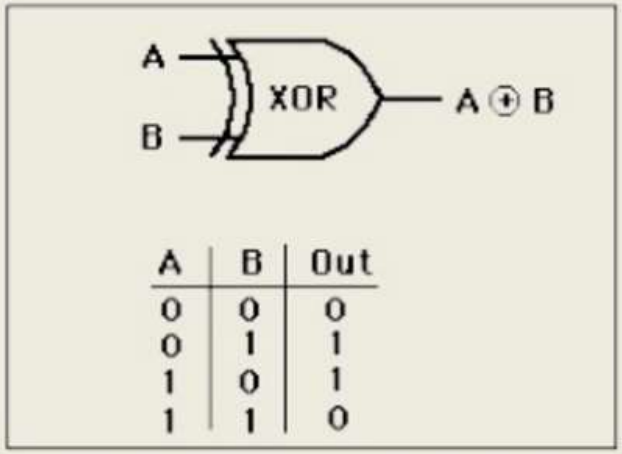


http://cs231n.stanford.edu/

1969: Perceptrons can't do XOR!



http://www.i-programmer.info/images/stories/BabBag/AI/book.jpg



http://hyperphysics.phy-astr.gsu.edu/hbase/electronic/ietron/xor.gif

Minsky & Papert

https://constructingkids.files.wordpress.com/2013/05/minsky-papert-71-csolomon-x640.jpg

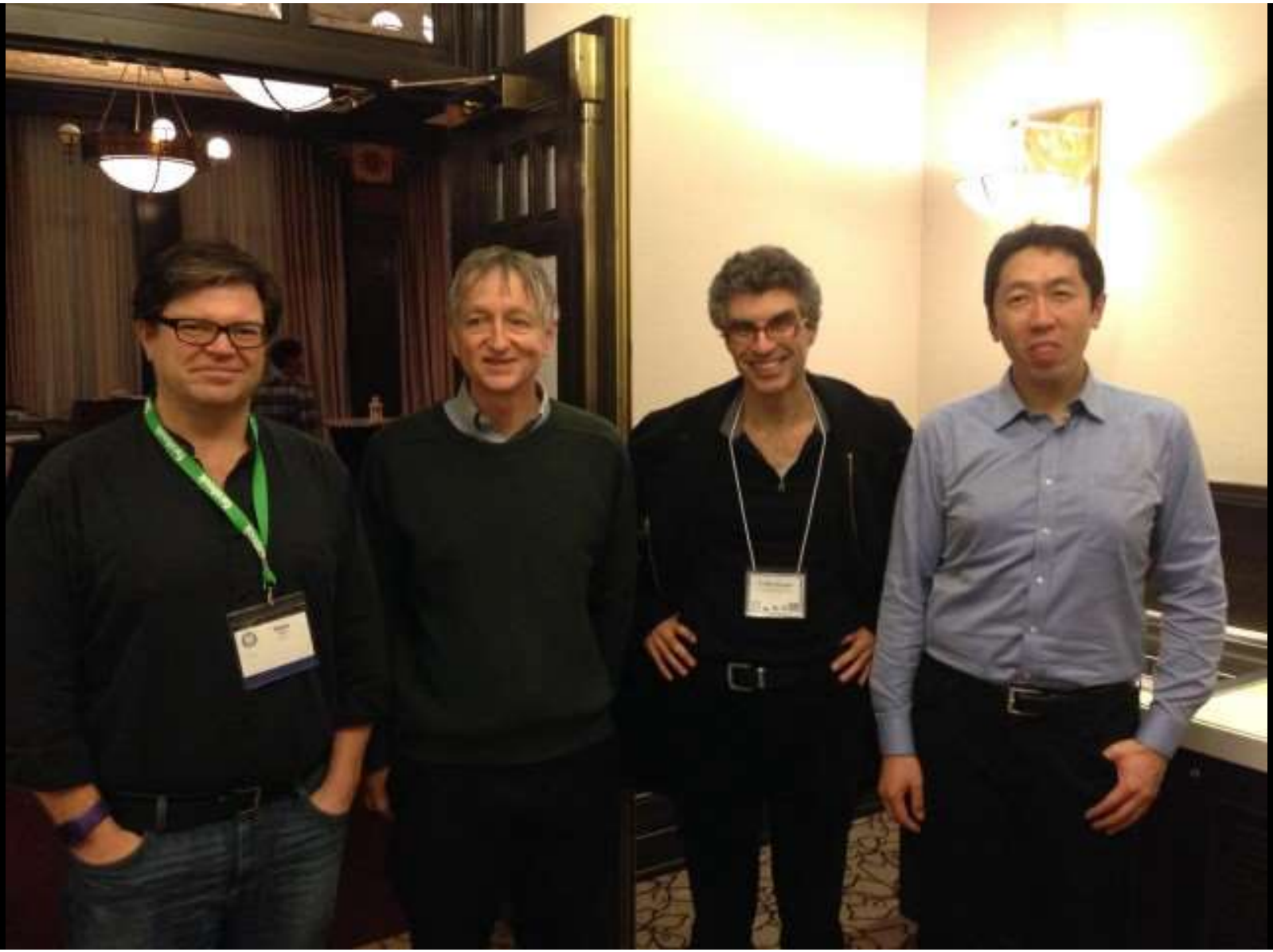

AI Winter
1969 - 1990

Deep Learning

Geoffrey Hinton
(Toronto, Google)

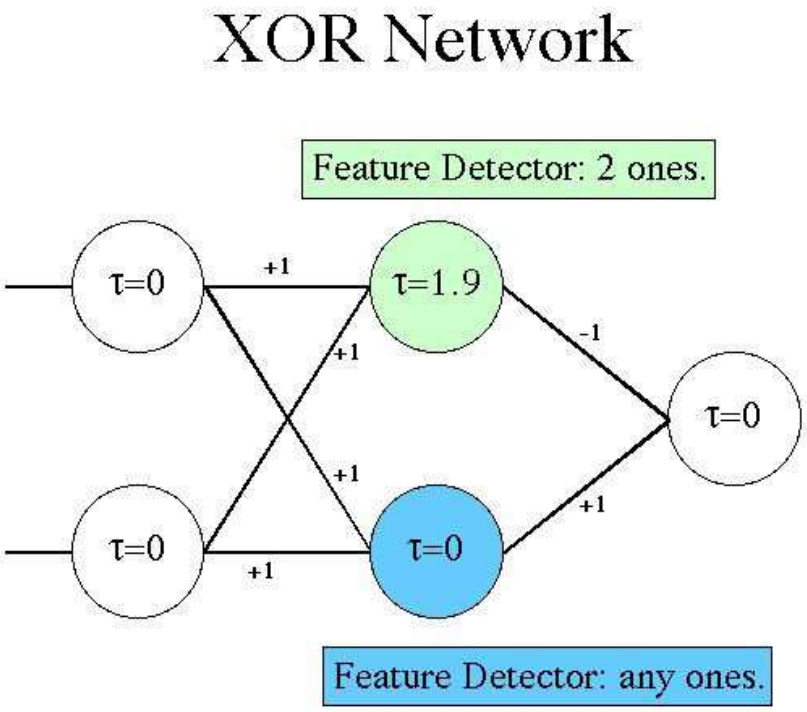
Yann LeCun
(NewYork, Facebook)

Yoshua Bengio
(Montreal)



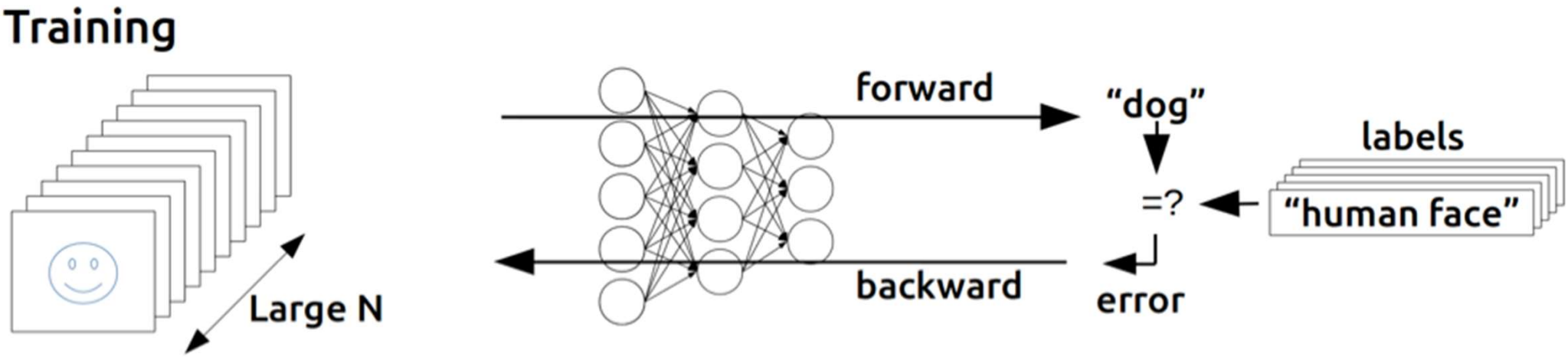
Learning XOR (1986)

Geoffrey Hinton

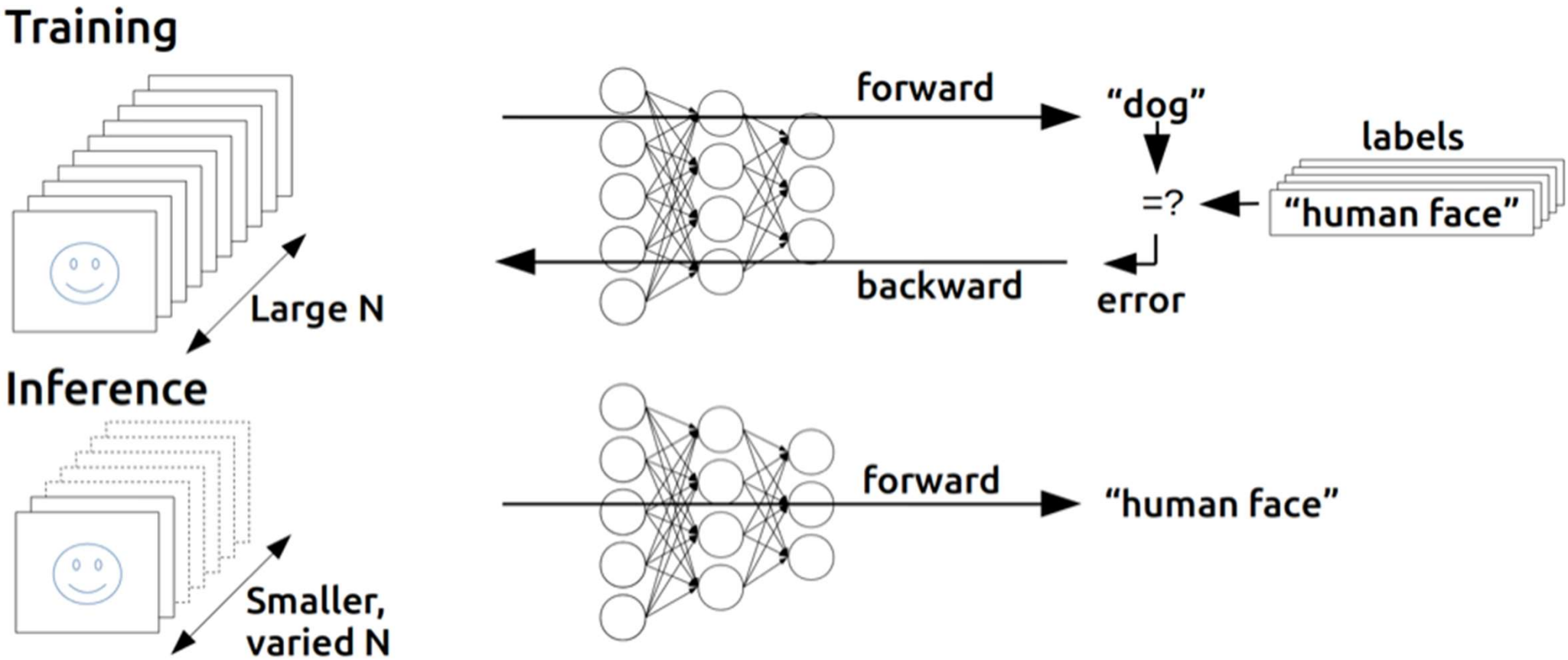


https://torontolife.com/life/ai-superstars-google-facebook-apple-studied-guy/

Backpropagation



Inference


Training
Large N
forward
"dog"
labels
=?
"human face"
backward
error
Inference
Smaller,
varied N
forward
"human face"

Chain Rule

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}$$

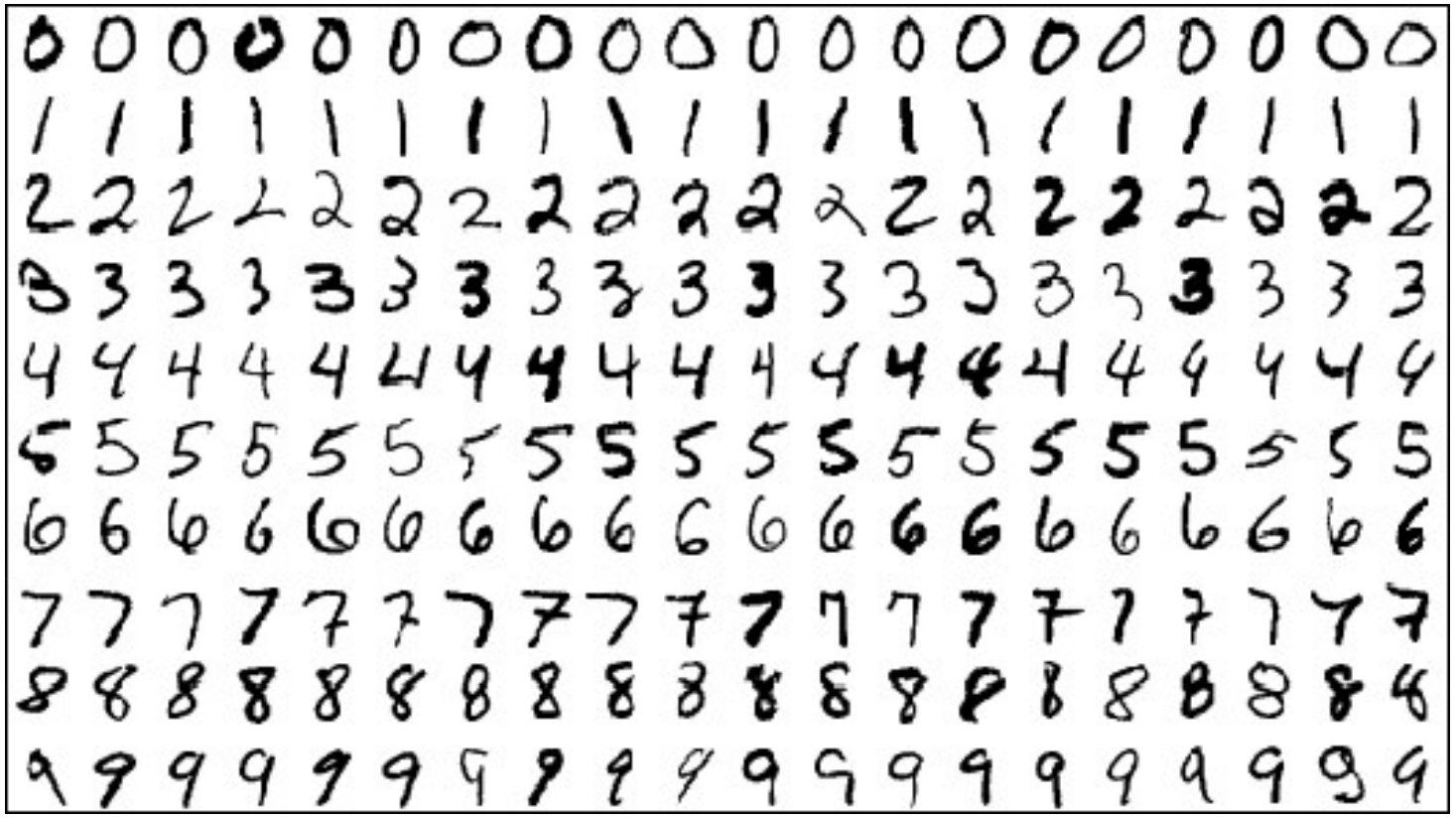
$$\frac{d^2y}{dx^2} = \frac{d^2y}{du^2}\left(\frac{du}{dx}\right)^2 + \frac{dy}{du}\frac{d^2u}{dx^2}$$

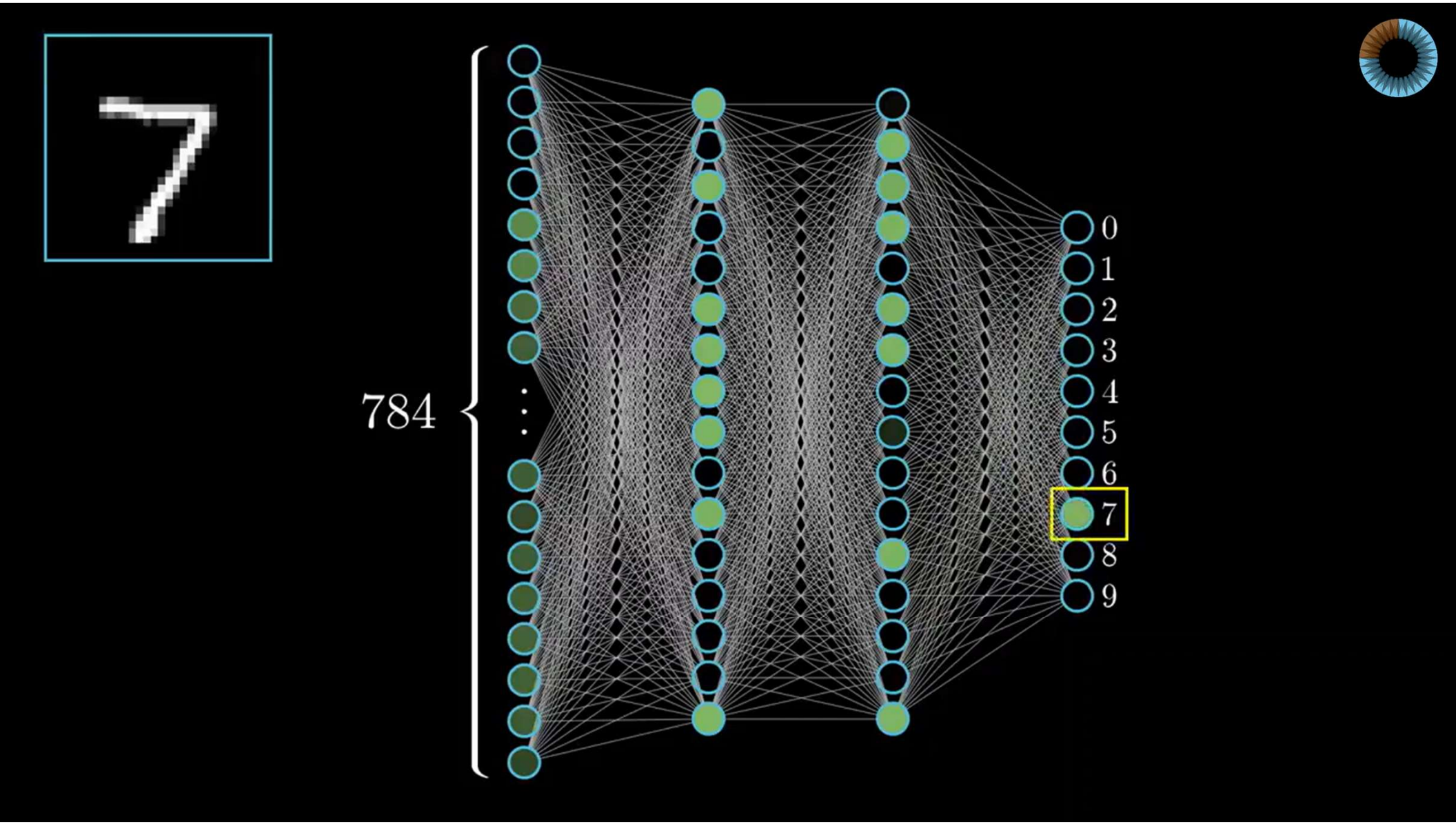
$$\frac{d^3y}{dx^3} = \frac{d^3y}{du^3}\left(\frac{du}{dx}\right)^3 + 3\,\frac{d^2y}{du^2}\frac{du}{dx}\frac{d^2u}{dx^2} + \frac{dy}{du}\frac{d^3u}{dx^3}$$

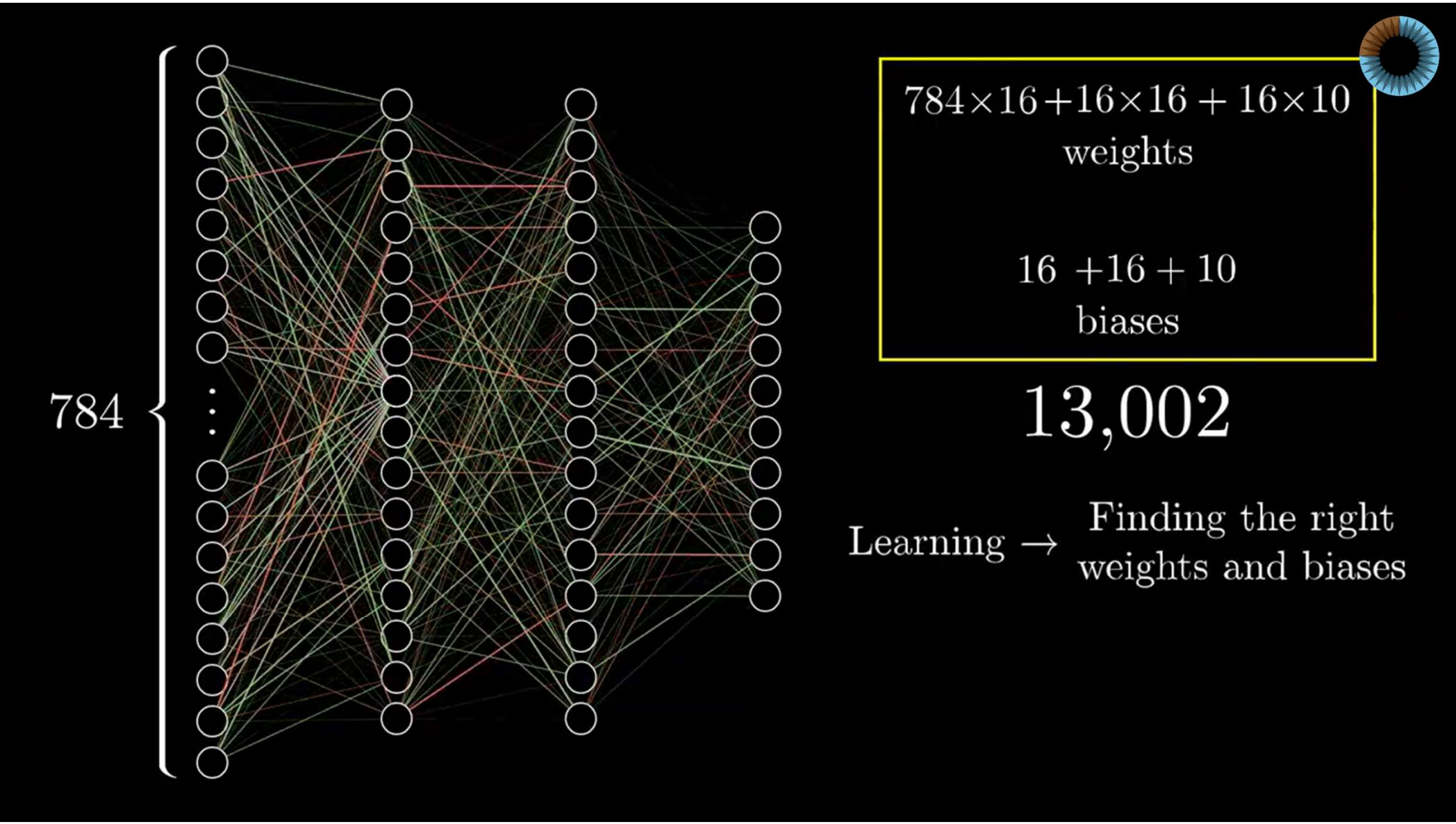
$$\frac{d^4y}{dx^4} = \frac{d^4y}{du^4}\left(\frac{du}{dx}\right)^4 + 6\,\frac{d^3y}{du^3}\left(\frac{du}{dx}\right)^2\frac{d^2u}{dx^2} + \frac{d^2y}{du^2}\left(4\,\frac{du}{dx}\frac{d^3u}{dx^3} + 3\left(\frac{d^2u}{dx^2}\right)^2\right) + \frac{dy}{du}\frac{d^4u}{dx^4}.$$

Example: Recognizing Handwritten Digits

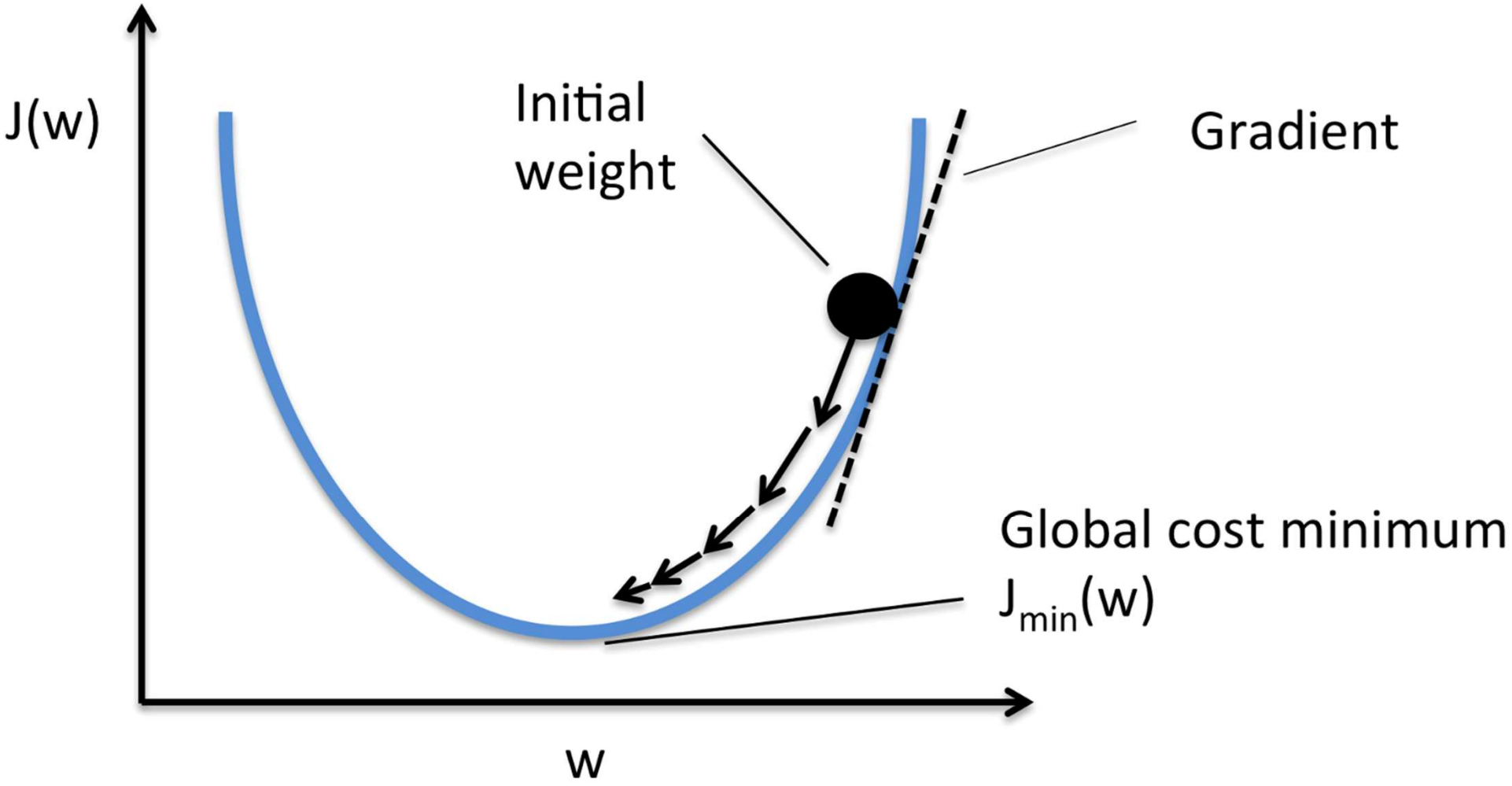
- MNIST dataset

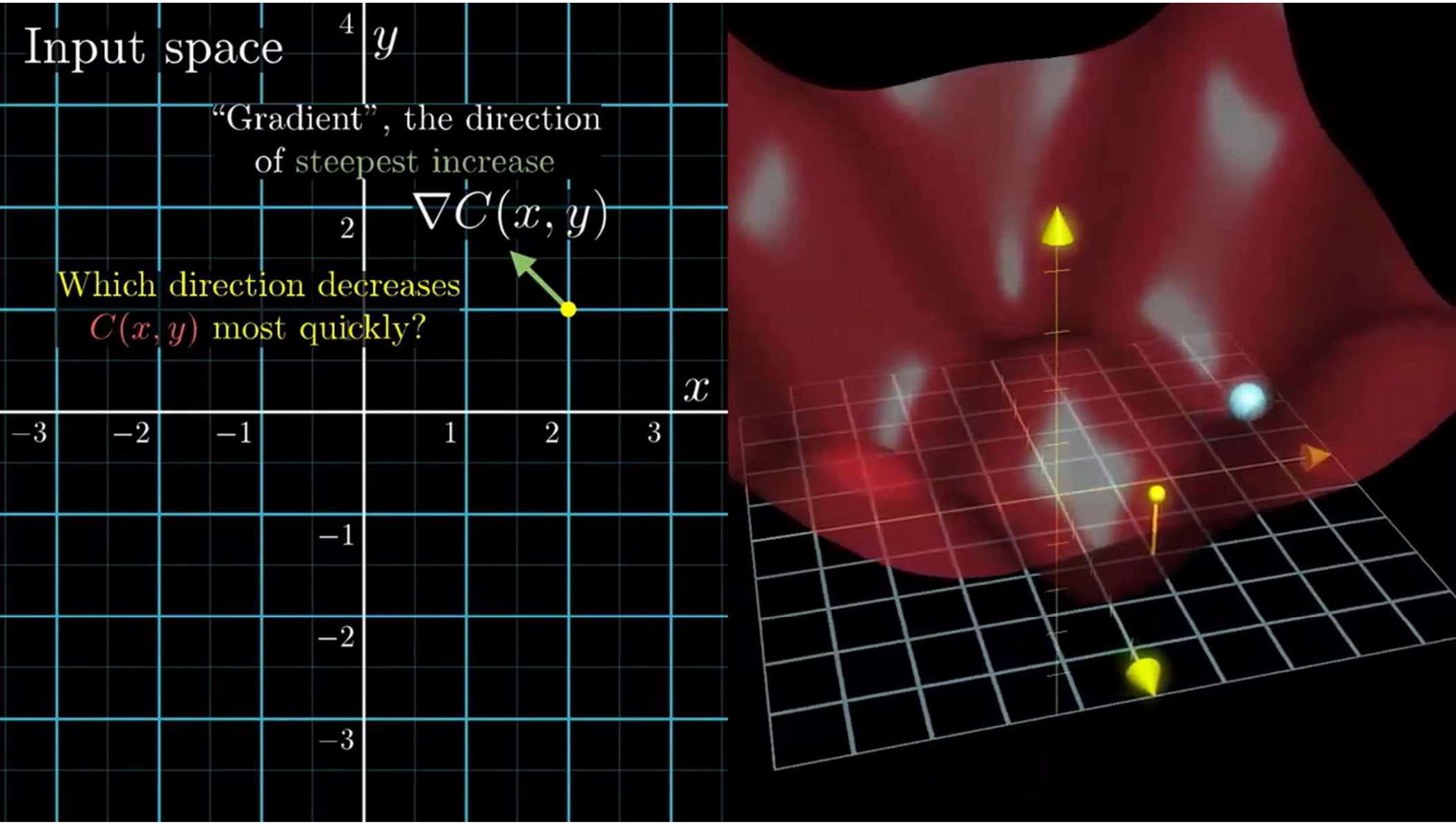



784
0
1
2
3
4
5
6
7
8
9


784
$784 \times 16 + 16 \times 16 + 16 \times 10$
weights
$16 + 16 + 10$
biases
13,002
Learning → Finding the right weights and biases

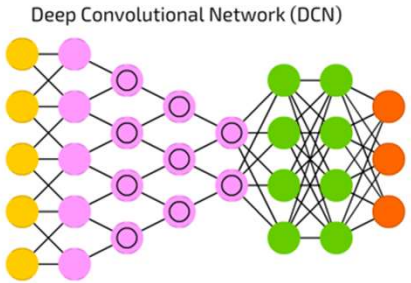
Gradient Descent

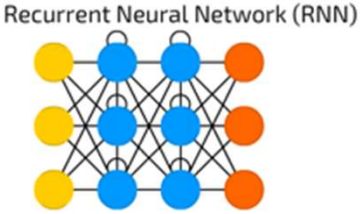


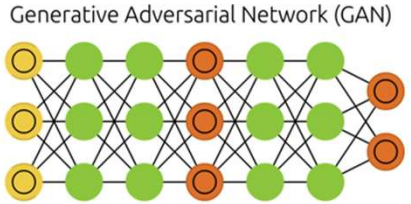

Input space
4
y
"Gradient", the direction
of steepest increase
$\nabla C(x, y)$
2
Which direction decreases
$C(x, y)$ most quickly?
x
−3
−2
−1
1
2
3
−1
−2
−3

Major Types of Neural Networks

- Convolutional Neural Networks (CNN)
- Recurrent Neural Networks (RNN)
- Generative Adversarial Networks (GAN)
- Attention & Transformer







https://www.asimovinstitute.org/neural-network-zoo/

Convolutional Neural Network (LeNet-5)

- https://medium.com/@sh.tsang/paper-brief-review-of-lenet-1-lenet-4-lenet-5-boosted-lenet-4-image-classification-1f5f809dbf17

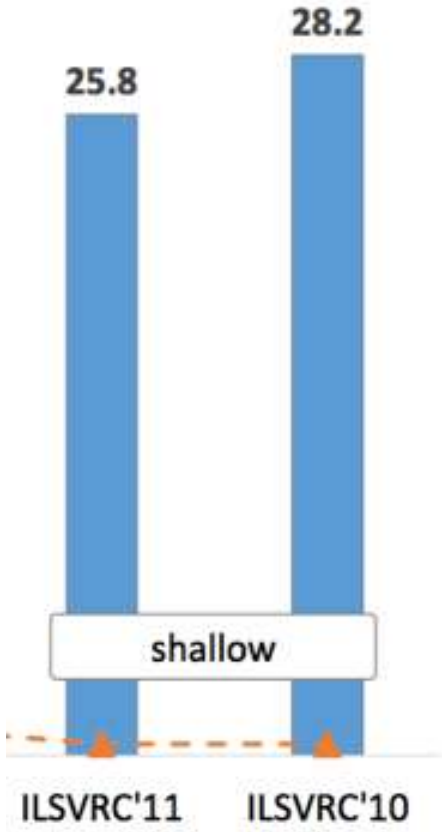
A Full Convolutional Neural Network (LeNet)


IMAGENET

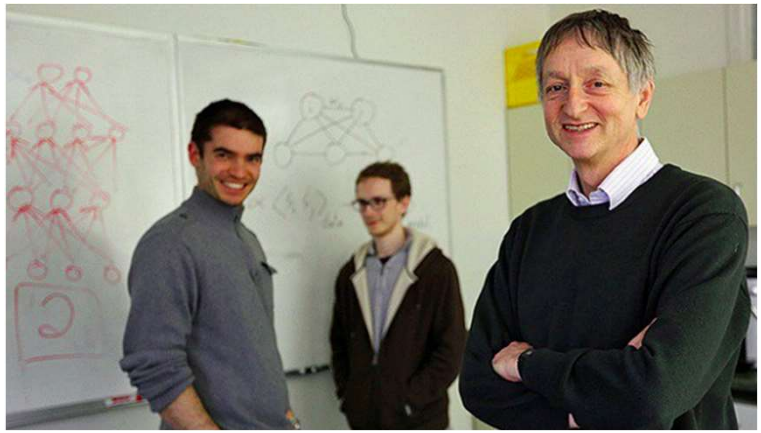
ImageNet Large Scale Visual Object Recognition Challenge (ILSVRC)

- 1000 categories
- For ILSVRC 2017
  - **Training images** for each category ranges from 732 to 1300
  - 50,000 validation **images** and 100,000 test **images**.
- Total number of images in ILSVRC 2017 is around 1,150,000

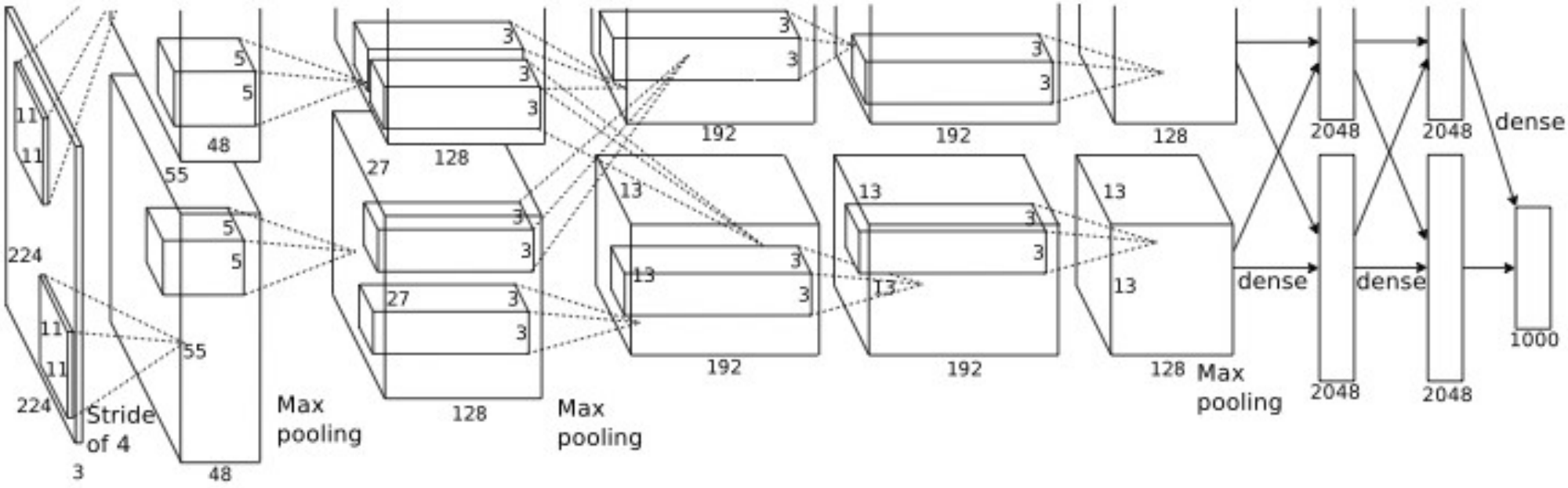
Error Rate on ImageNet Challenge (~2011)



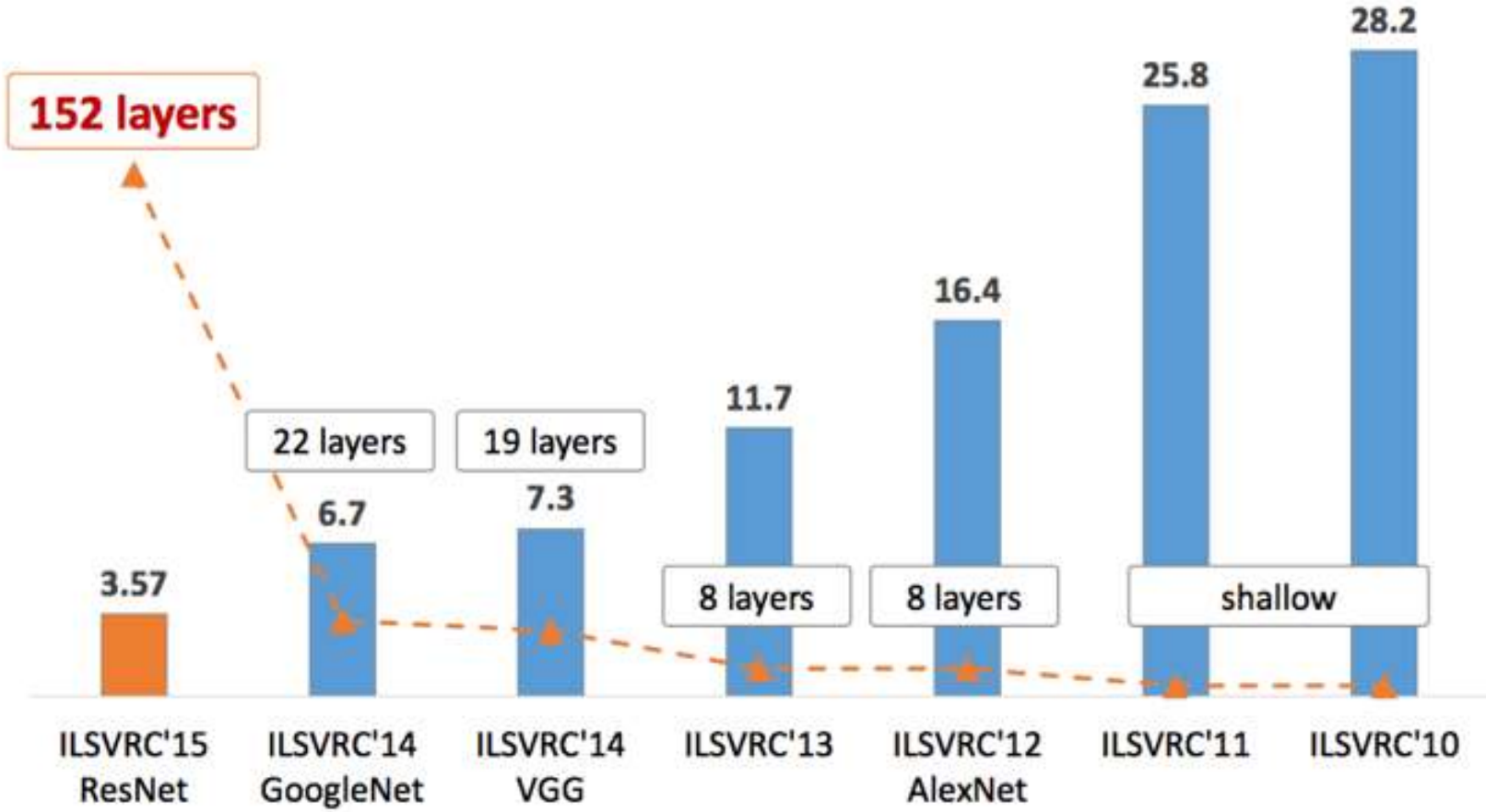
Deep Convolutional Neural Network (AlexNet)

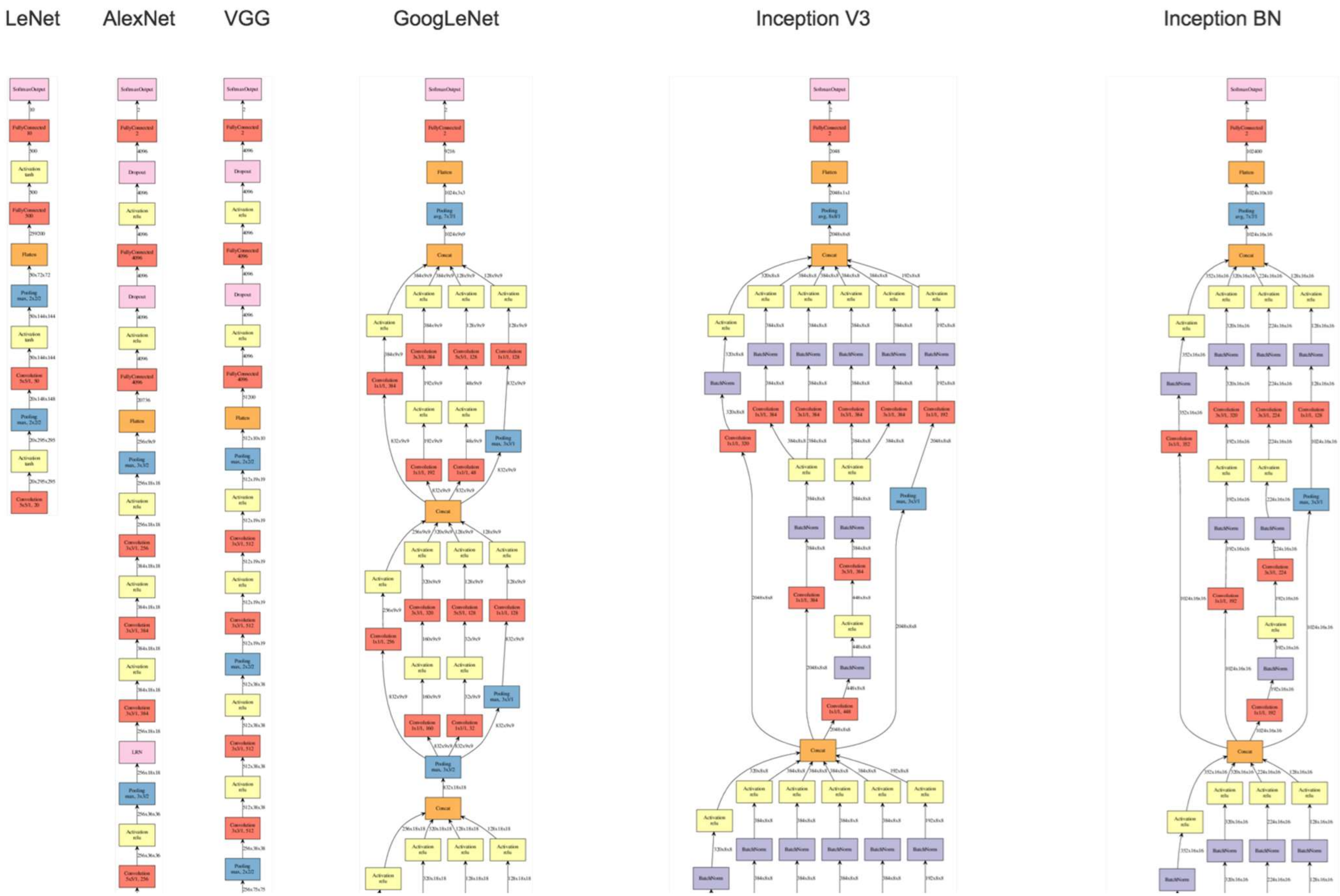


- Alex Krizhevsky, Ilya Suskever, Geoffrey Hinton, 2012



Error Rate on ImageNet Challenge (~2015)




LeNet
AlexNet
VGG
GoogLeNet
Inception V3
Inception BN


WE NEED TO GO
DEEPER
memegenerator.net

Residual Network (ResNet)

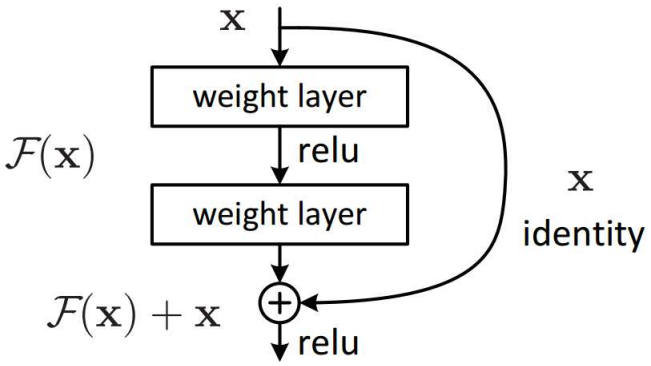
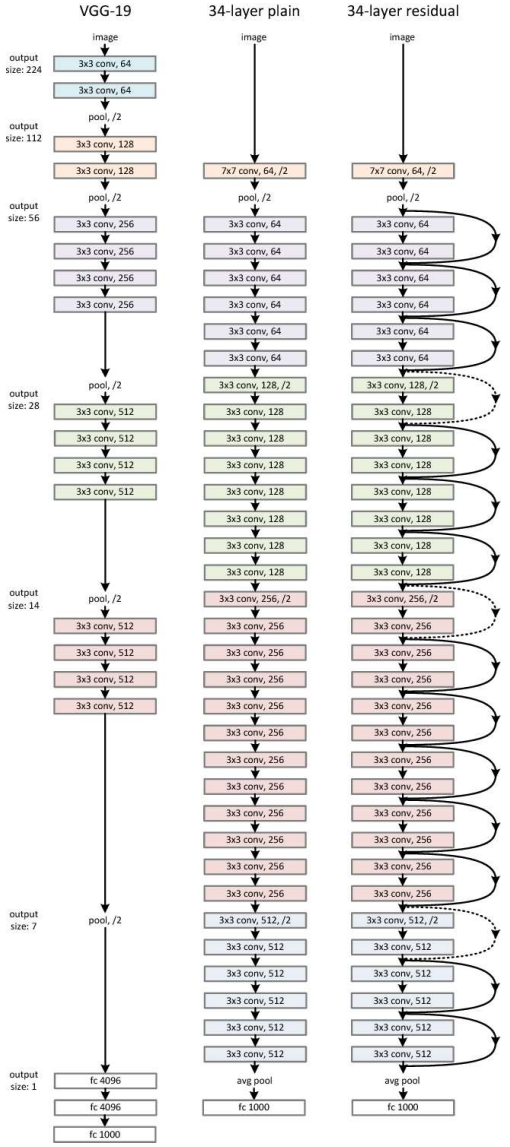
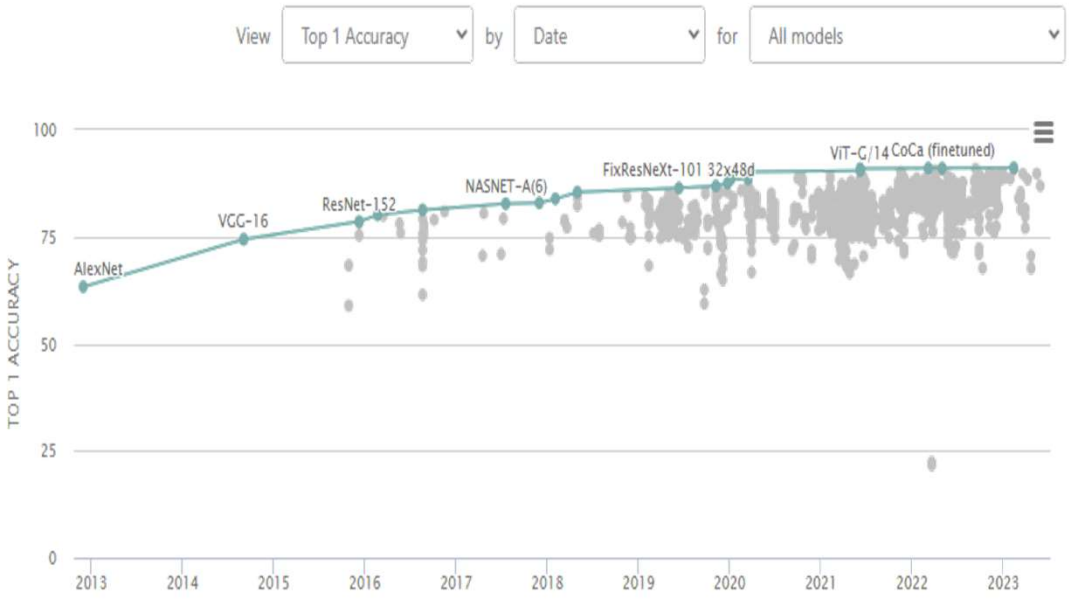


Figure 2. Residual learning: a building block.

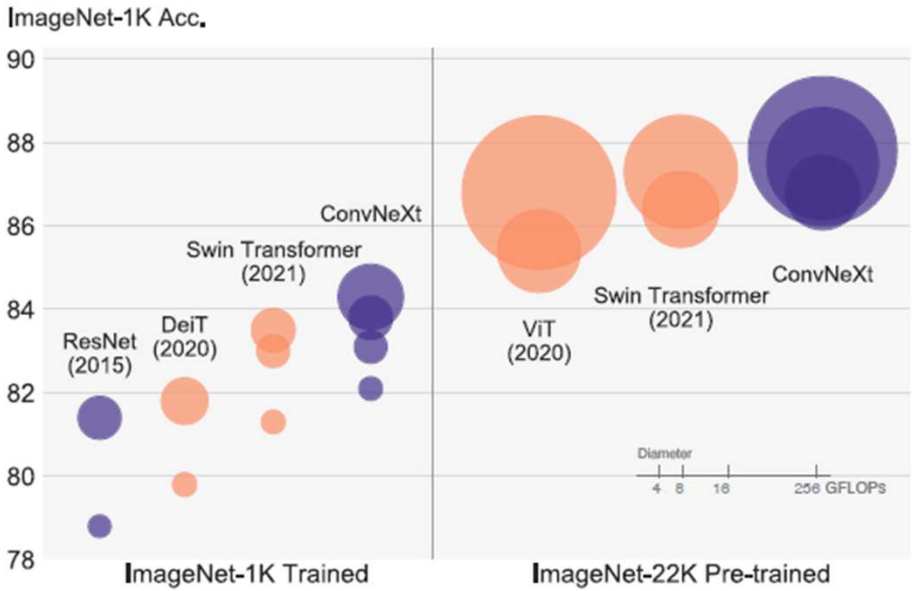


Comparison of Popular CNN Architectures

- ImageNet top 1 accuracy



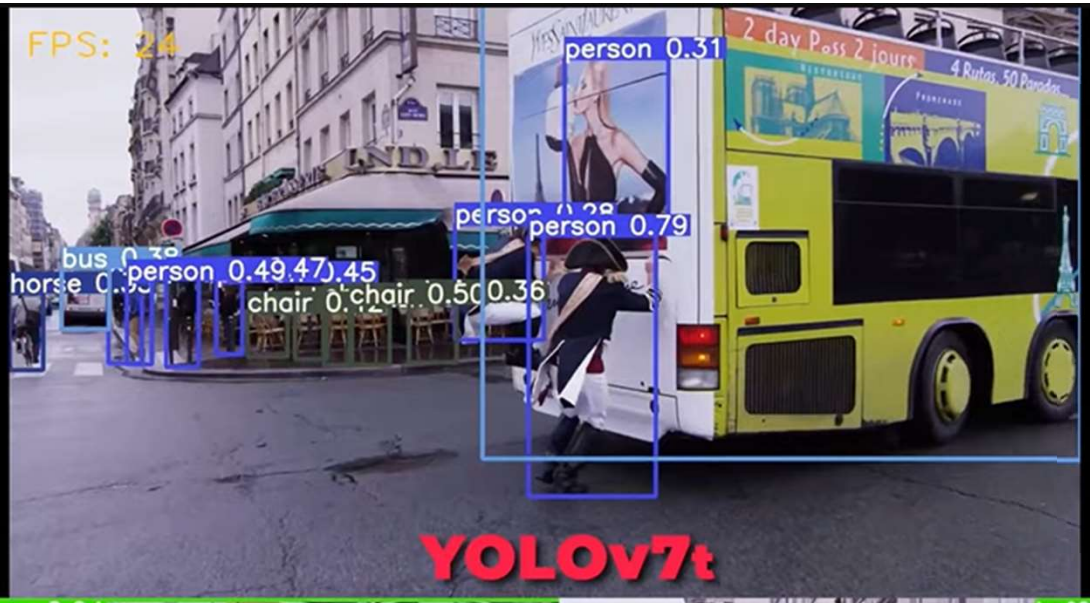
https://paperswithcode.com/sota/image-classification-on-imagenet



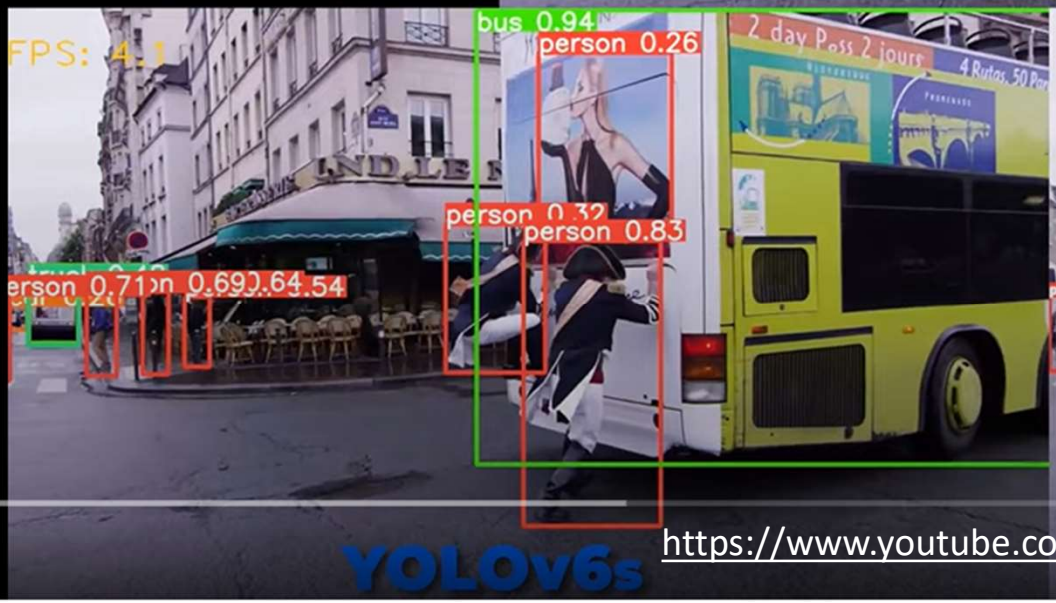
https://sh-tsang.medium.com/review-convnext-a-convnet-for-the-2020s-53b9ada30ab9

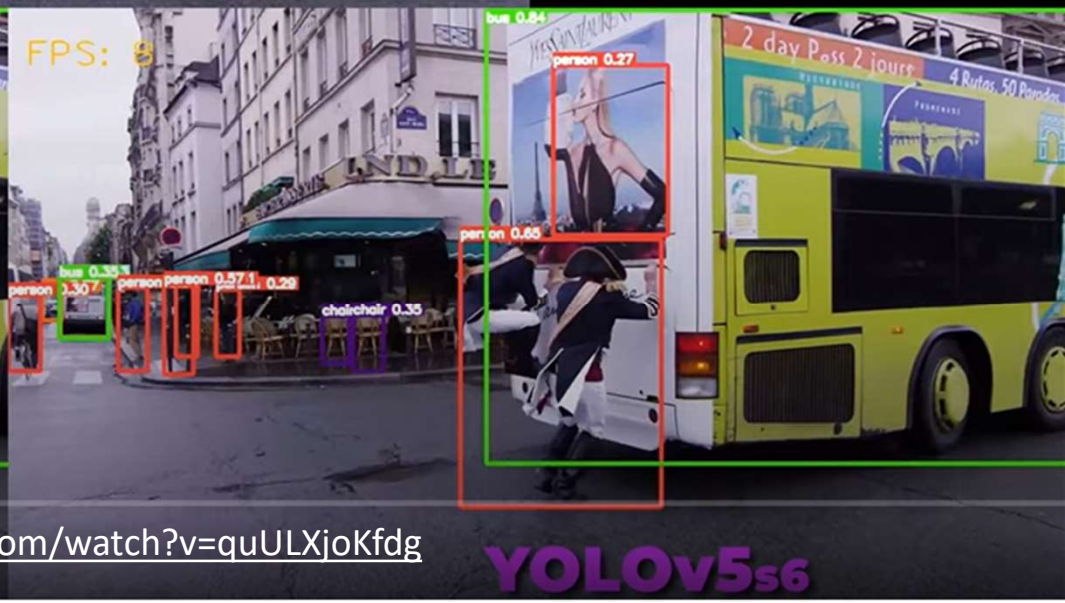
AUGMENTED STARTUPS
Computer Vision | AI | Drones

Real-time Object Detection



You Only Look Once (YOLO)

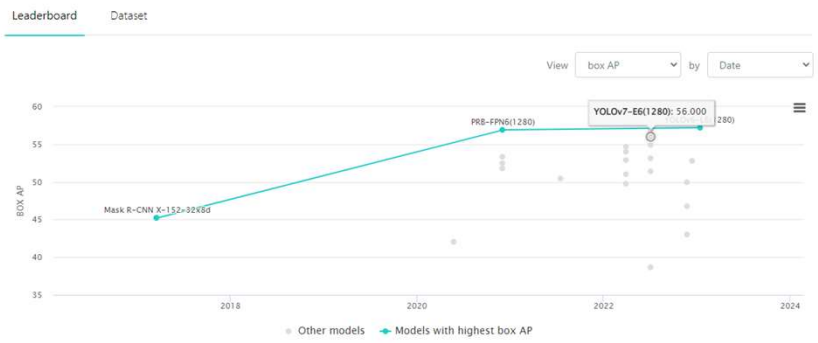




https://www.youtube.com/watch?v=quULXjoKfdg

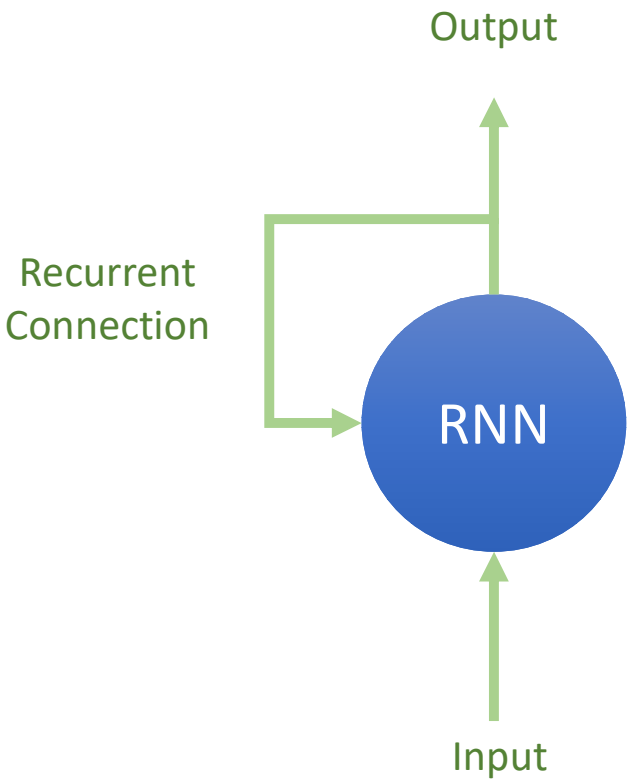
YOLO v4, v7 Authors and Me

- From left to right
  - Me, Dr. Mark Liao, Dr. Wang

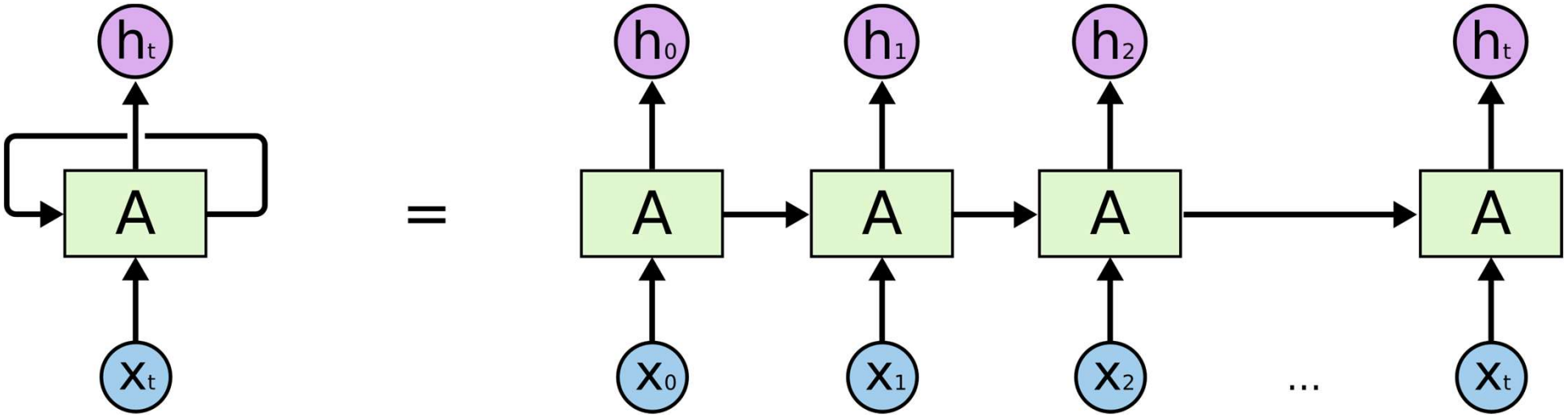


Recurrent Neural Networks (RNNs)

- An internal state (memory)
- Feedback loop
- Good for processing time-series data

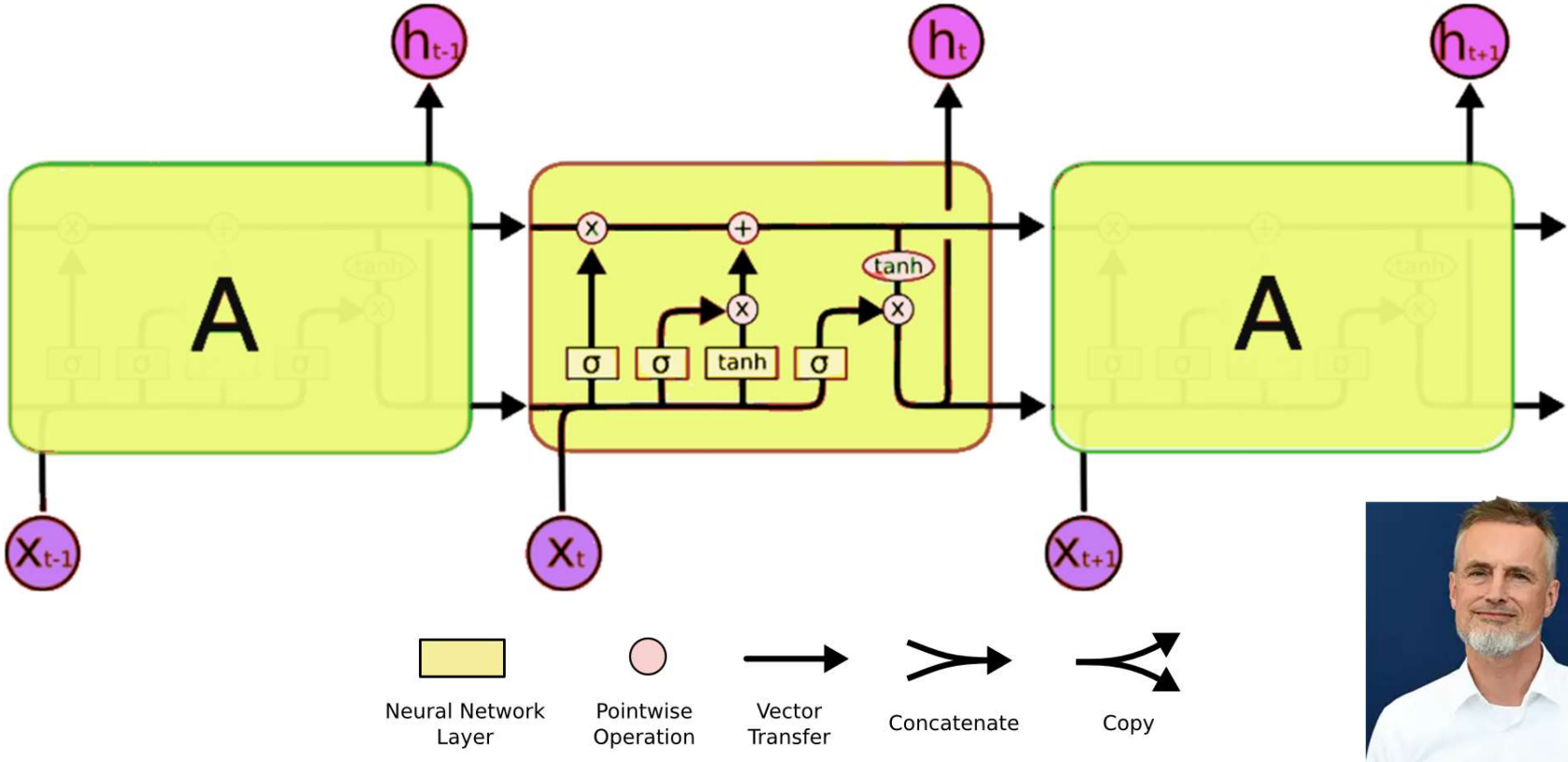


Unroll the RNN



https://colah.github.io/posts/2015-08-Understanding-LSTMs/

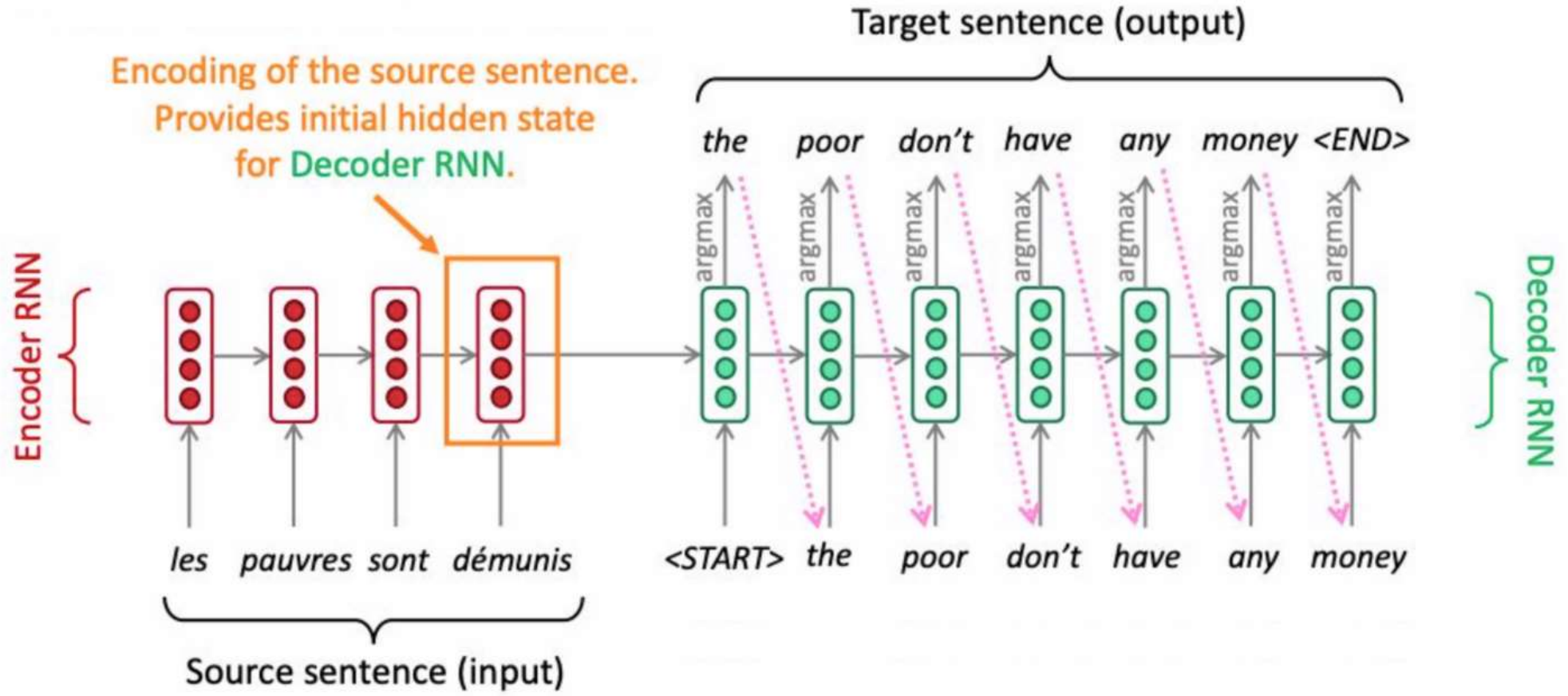
Long Short-term Memory (LSTM)



https://colah.github.io/posts/2015-08-Understanding-LSTMs/

Jürgen Schmidhuber

Sequence-2-Sequence model (Language Translation)



https://towardsdatascience.com/introduction-to-rnns-sequence-to-sequence-language-translation-and-attention-fc43ef2cc3fd

Attention is All You Need!

Ashish Vaswani* Google Brain avaswani@google.com

Noam Shazeer* Google Brain noam@google.com

Niki Parmar* Google Research nikip@google.com

Jakob Uszkoreit* Google Research usz@google.com

Llion Jones* Google Research llion@google.com

Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu

Łukasz Kaiser* Google Brain lukaszkaiser@google.com

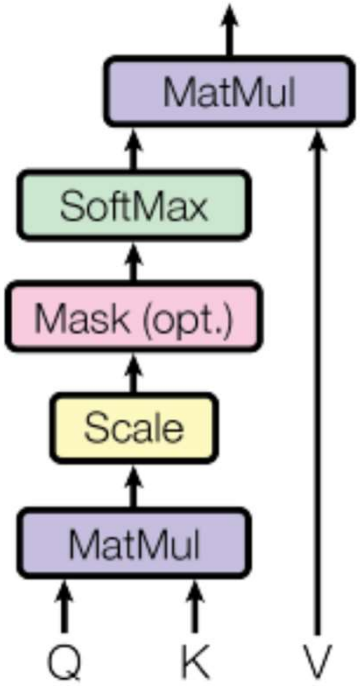
Illia Polosukhin* ‡ illia.polosukhin@gmail.com

Google Brain & University of Toronto, *NIPS*, 2017

Attention Module in Transformer

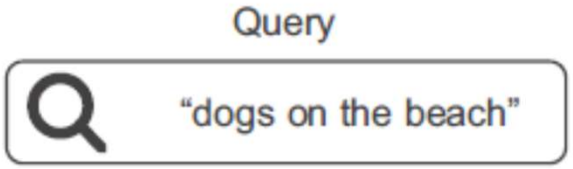
- Query (Q), Key (K), Value (V) attention

$$\text{Attention}(Q, K, V) = \text{softmax}(\frac{QK^T}{\sqrt{d_k}})V$$

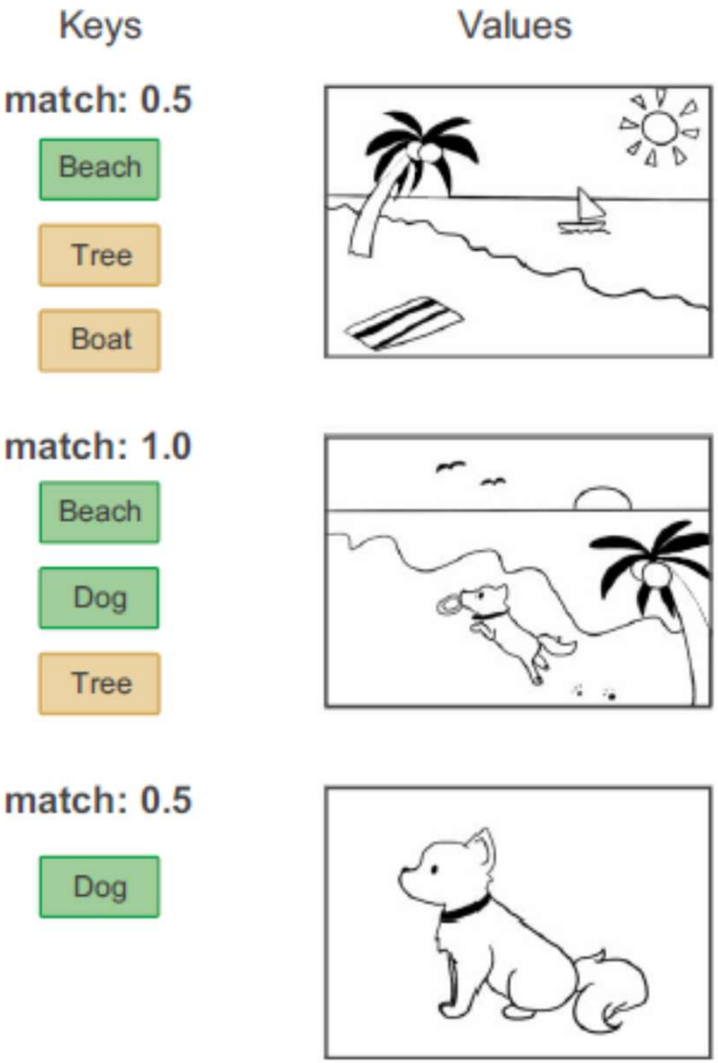


A. Waswani et al., "Attention is All You Need," NIPS, 2017

Query, Keys, Values

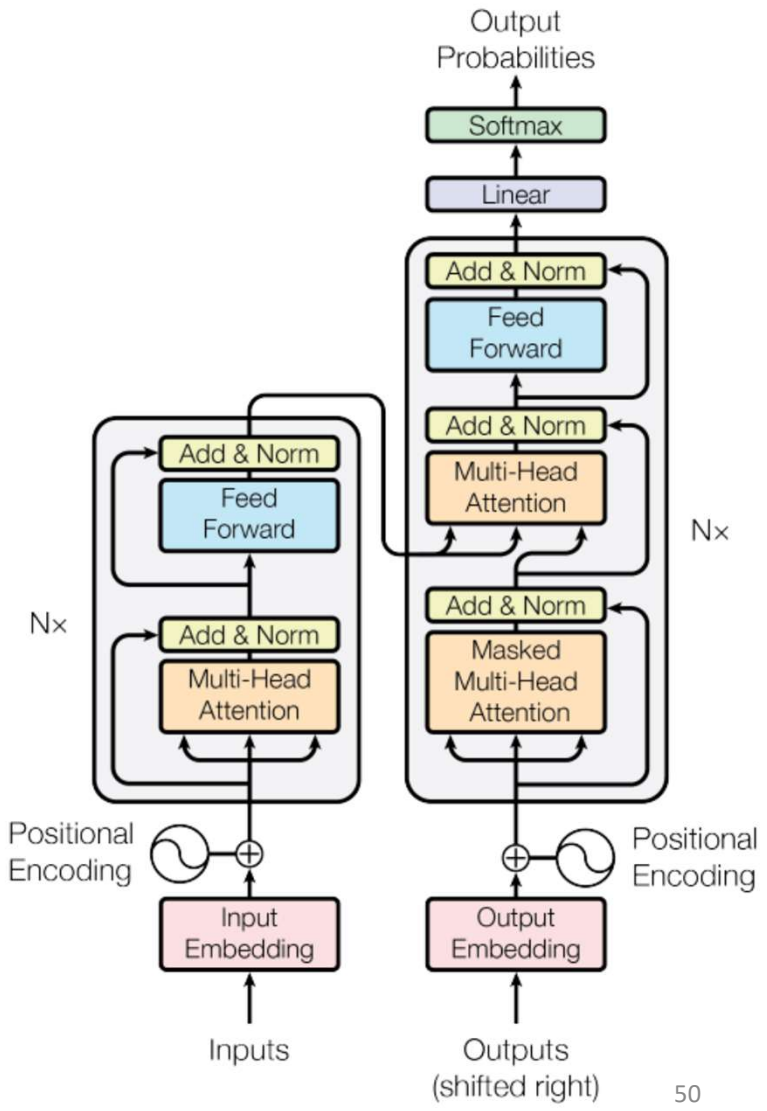


Retrieving images from a database



The Transformer Model

- Encoder-decoder architecture
- Multi-head attention
  - Self-attention in encoders
  - Masked Self-attention in decoders
  - Encoder-decoder attention
- Positional encoding

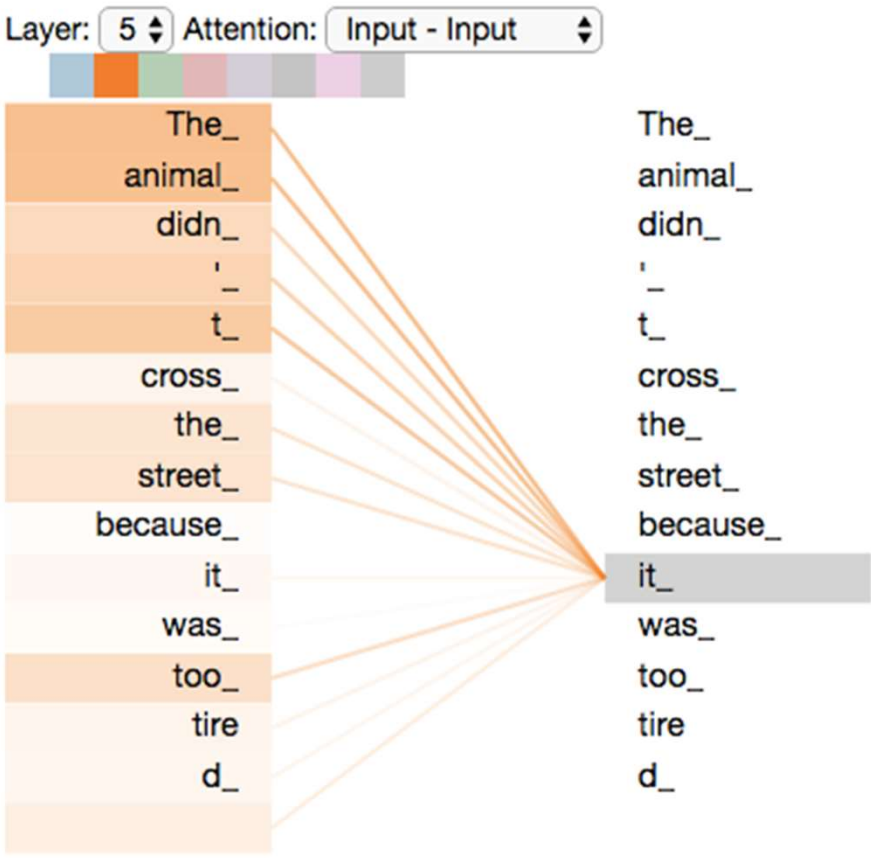


A. Waswani et al., "Attention is All You Need," NIPS, 2017

Visualizing Attention

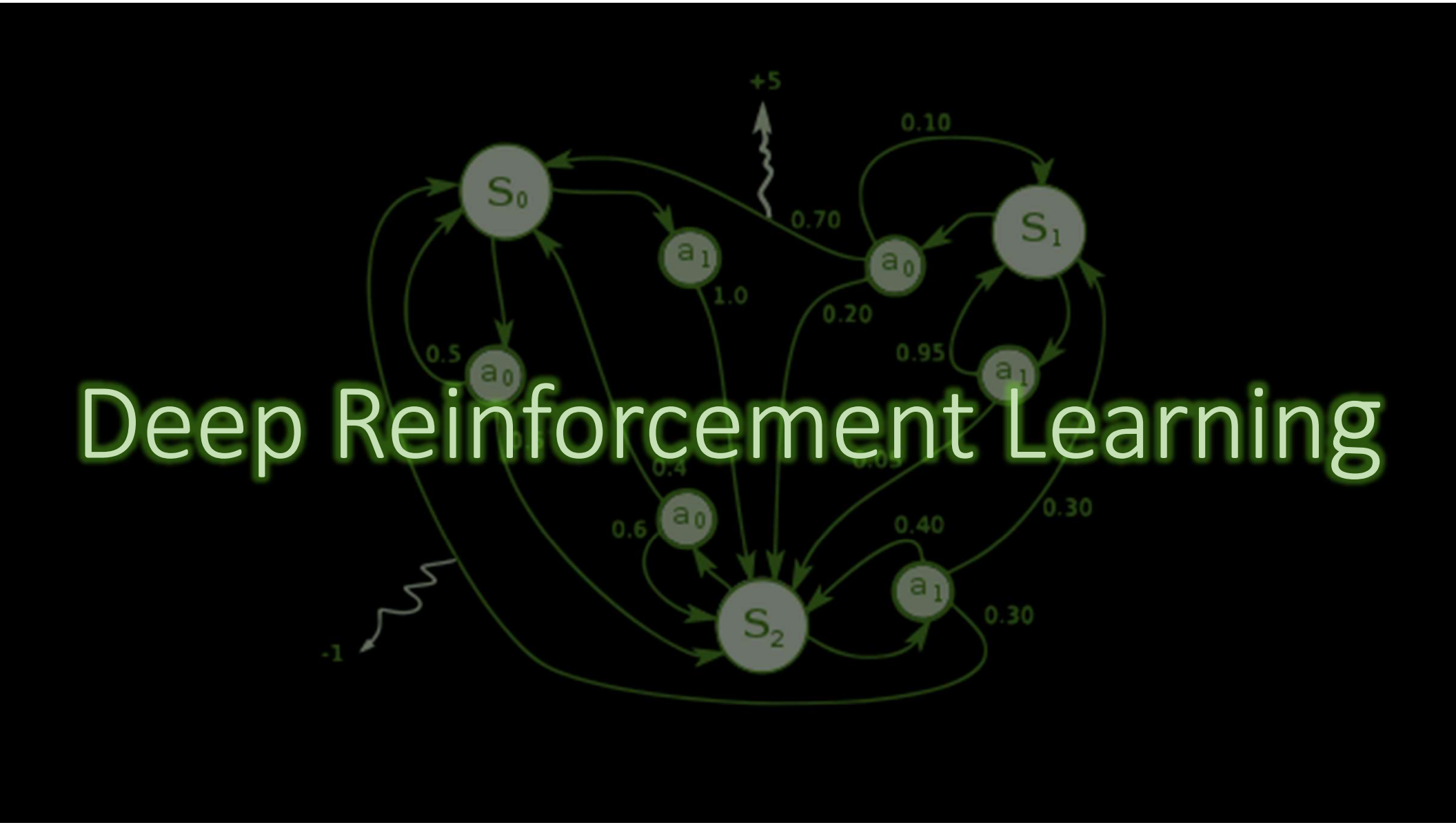
- Tensor2Tensor Notebook

https://colab.research.google.com/github/tensorflow/tensor2tensor/blob/master/tensor2tensor/notebooks/hello_t2t.ipynb




```
Inputs: The animal didn't cross the
street because it was too tired
```

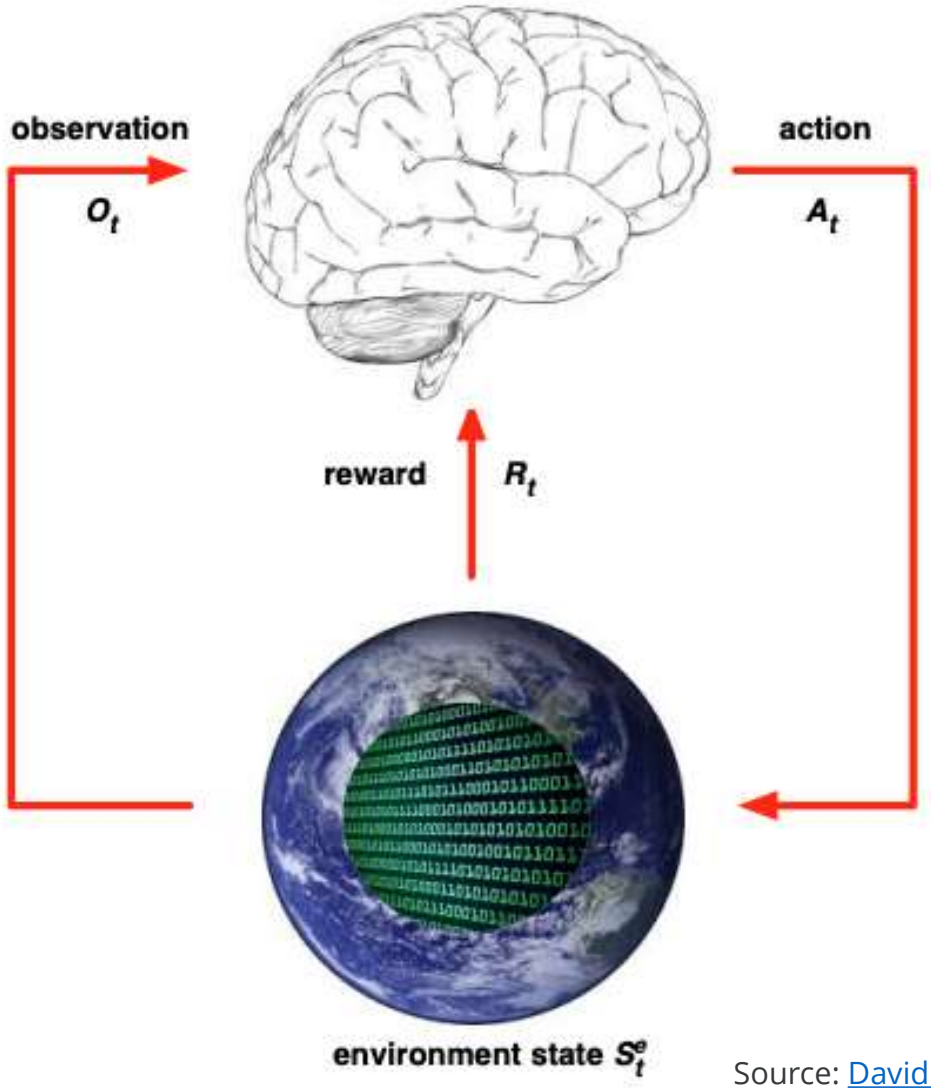
```
Outputs: Das Tier überquerte die Straße
nicht, weil es zu müde war, weil es zu
müde war.
```



Deep Reinforcement Learning

Reinforcement Learning

https://talkmarkets.com/content/deep-reinforcement-learning-for-trading-applications?post=252842

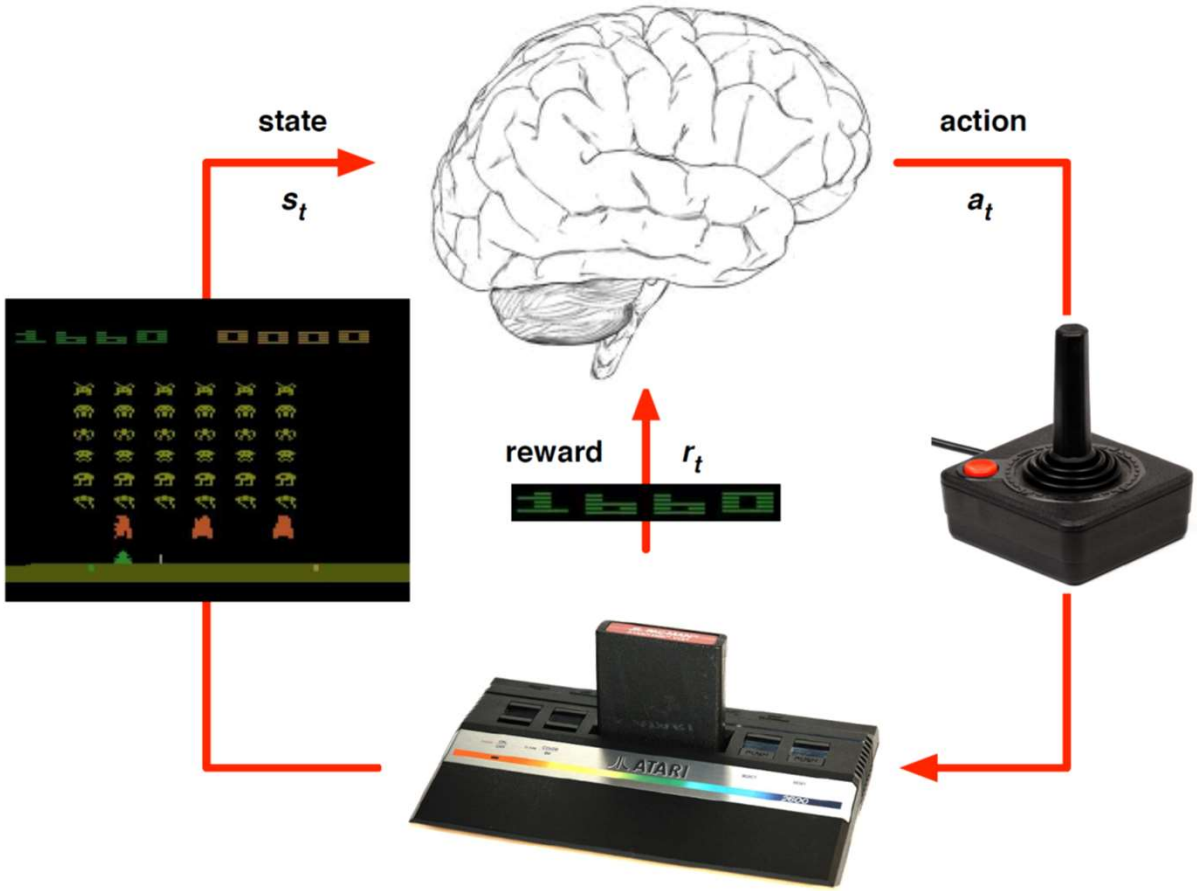


Source: David Silver, UCL

DeepMind: DRL in Atari

Demis Hassabis

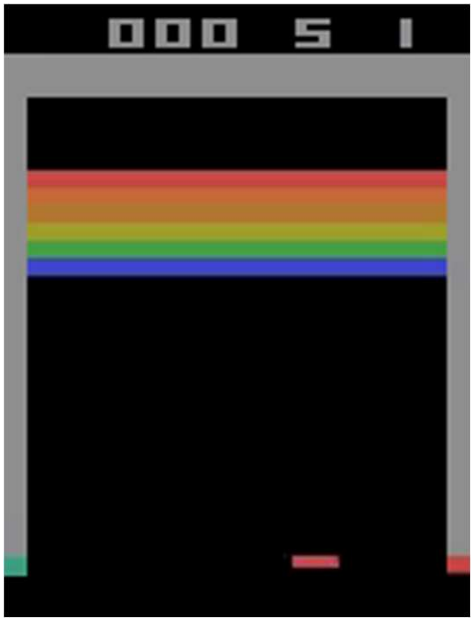
Mustafa Suleyman

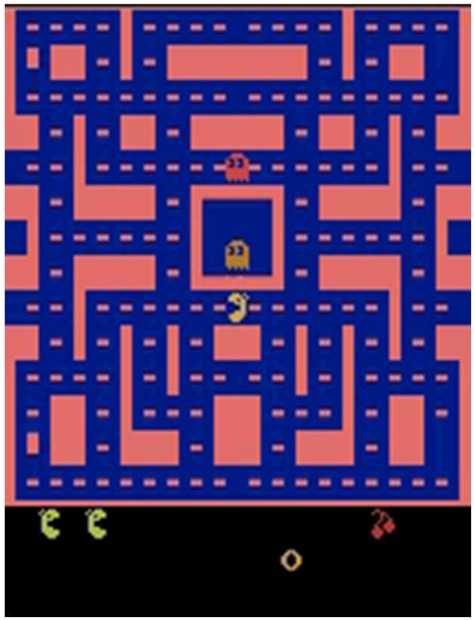


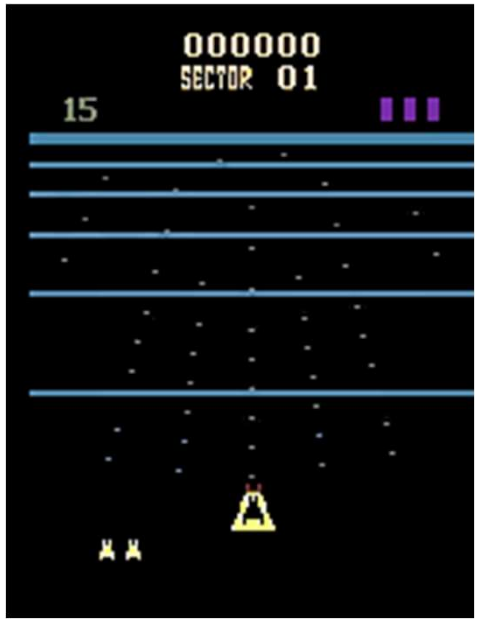
Mnih et al., "Human Level Control through Deep Reinforcement Learning," *Nature*, 2015

Learning to Play Atari Games

David Silver

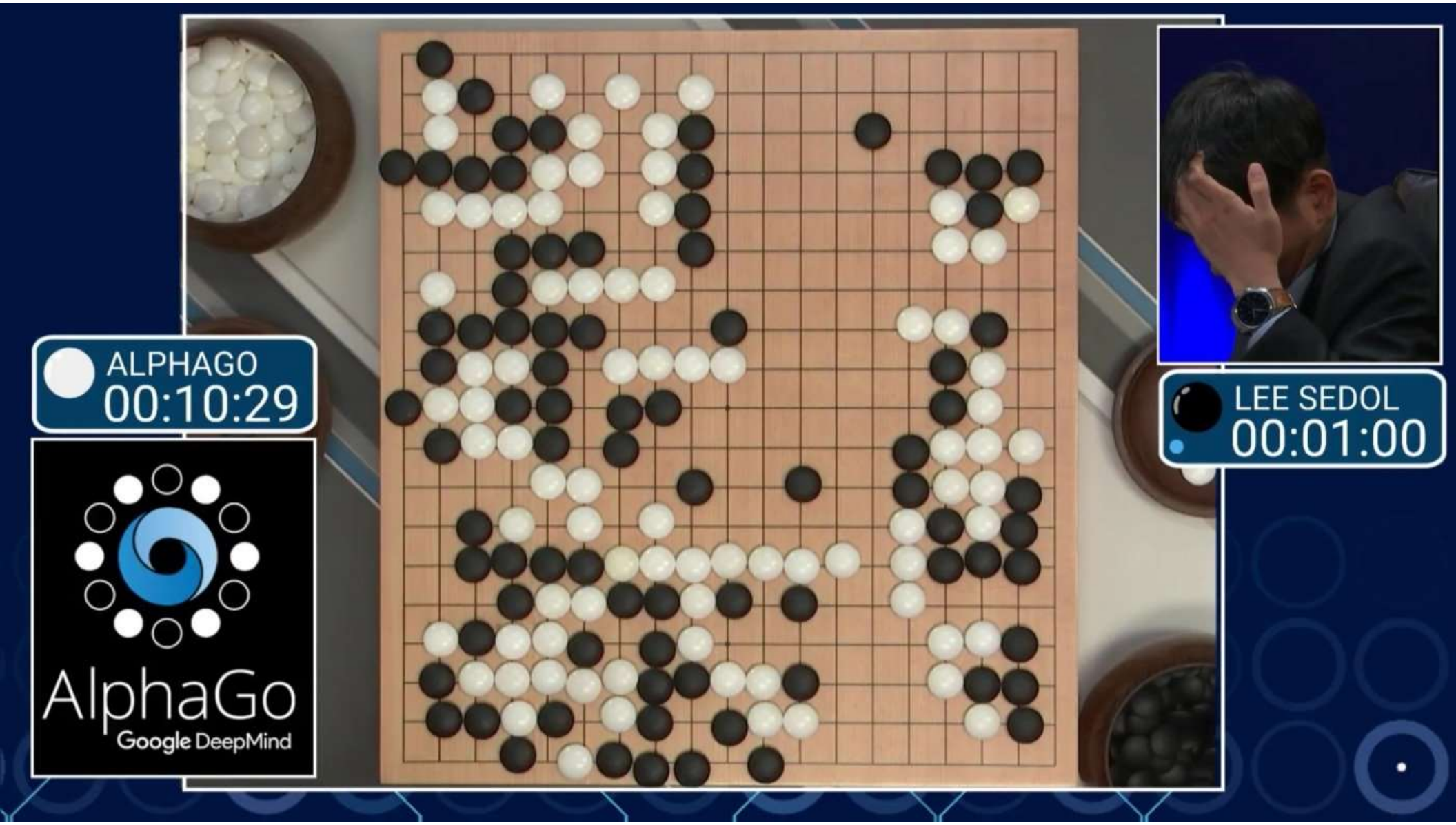






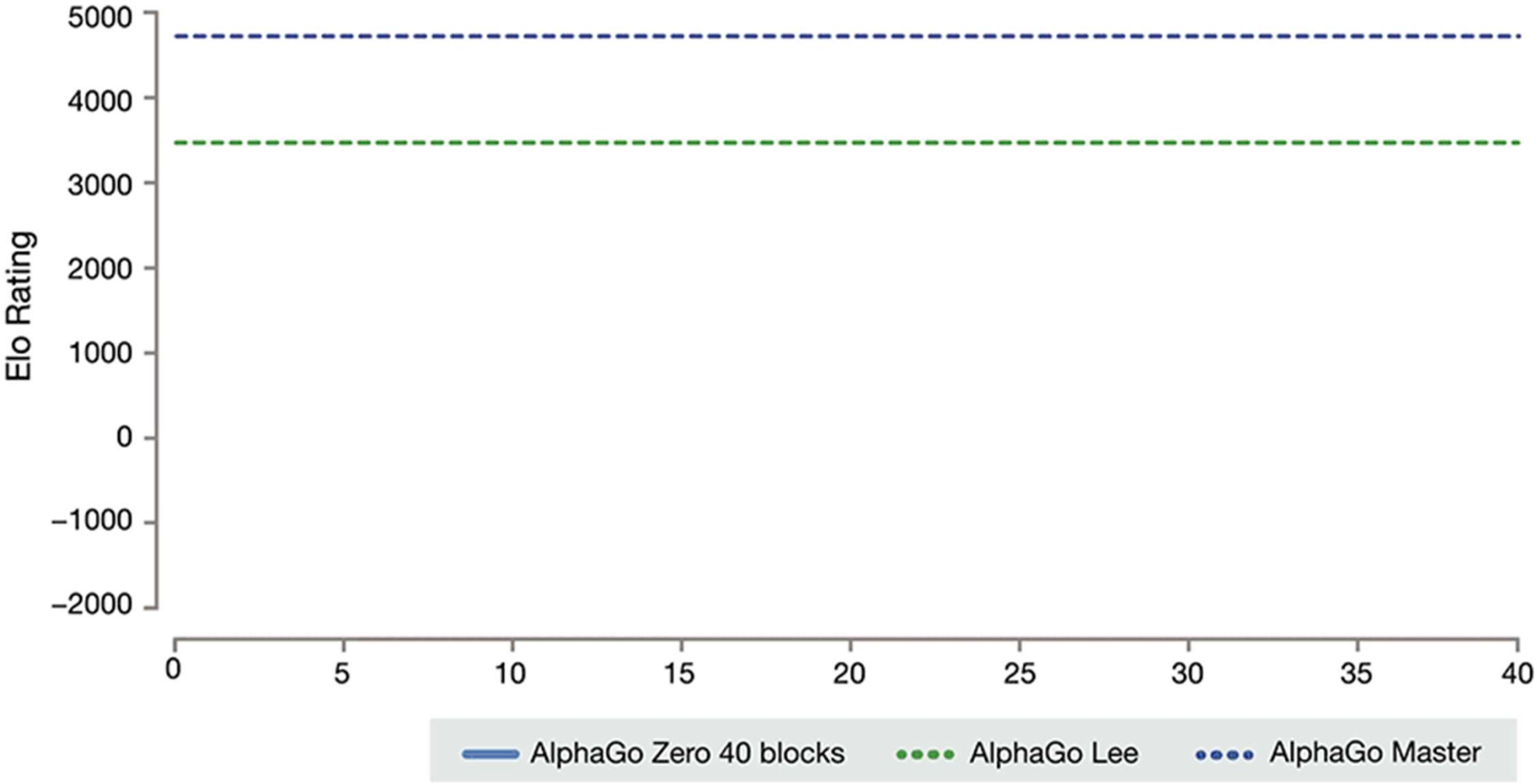
Complexity of Go vs. Chess

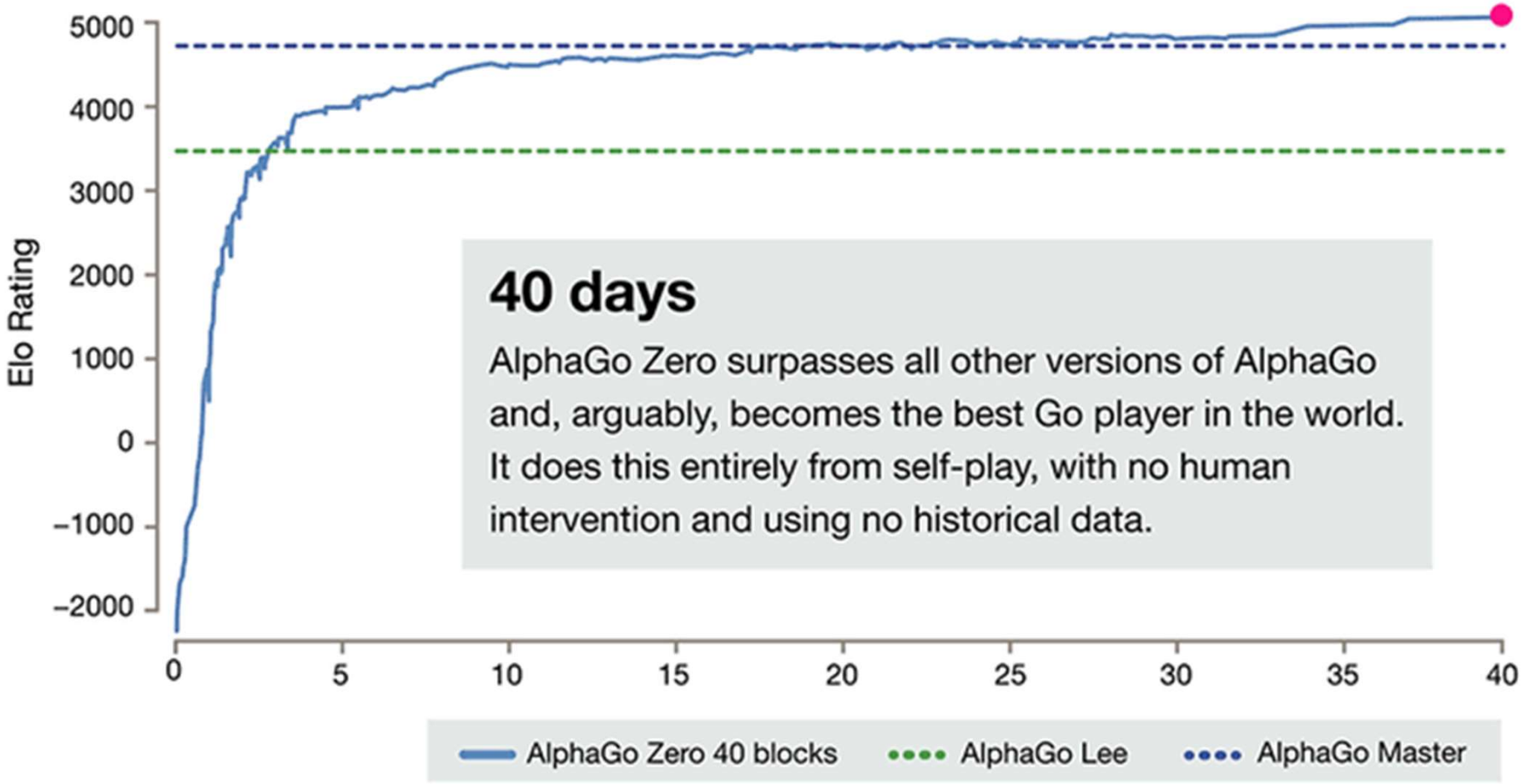
Game	Board size	State space	Game tree size
Go	19 x 19	10^{172}	10^{360}
Chess	8 x 8	10^{50}	10^{123}
Checkers	8 x 8	10^{18}	10^{54}


ALPHAGO
00:10:29
AlphaGo
Google DeepMind
LEE SEDOL
00:01:00


Dr. Aja Huang (黃士杰)
ALPHAGO
01:58:31
LEE SEDOL
AlphaGo
Lee Sedol


AlphaGo Zero
Starting from scratch

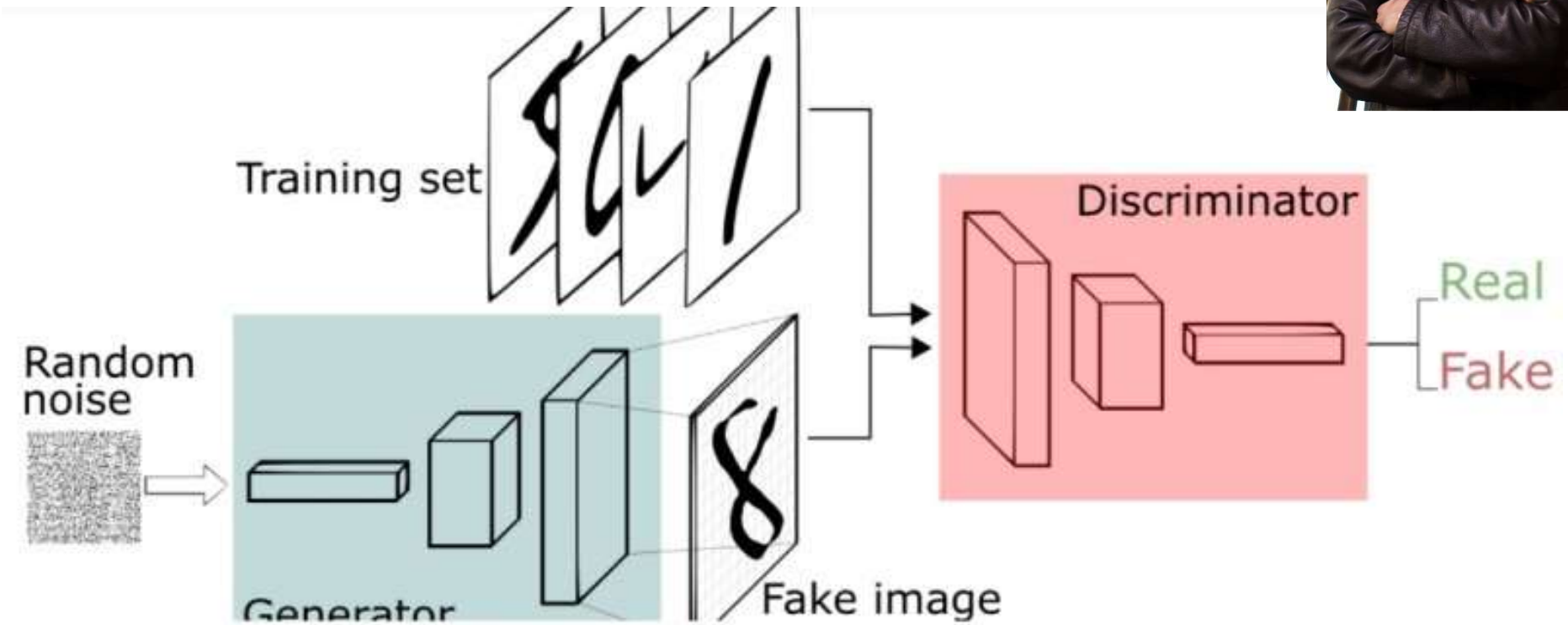

5000
4000
3000
2000
1000
0
−1000
−2000
Elo Rating
0
5
10
15
20
25
30
35
40
AlphaGo Zero 40 blocks
AlphaGo Lee
AlphaGo Master

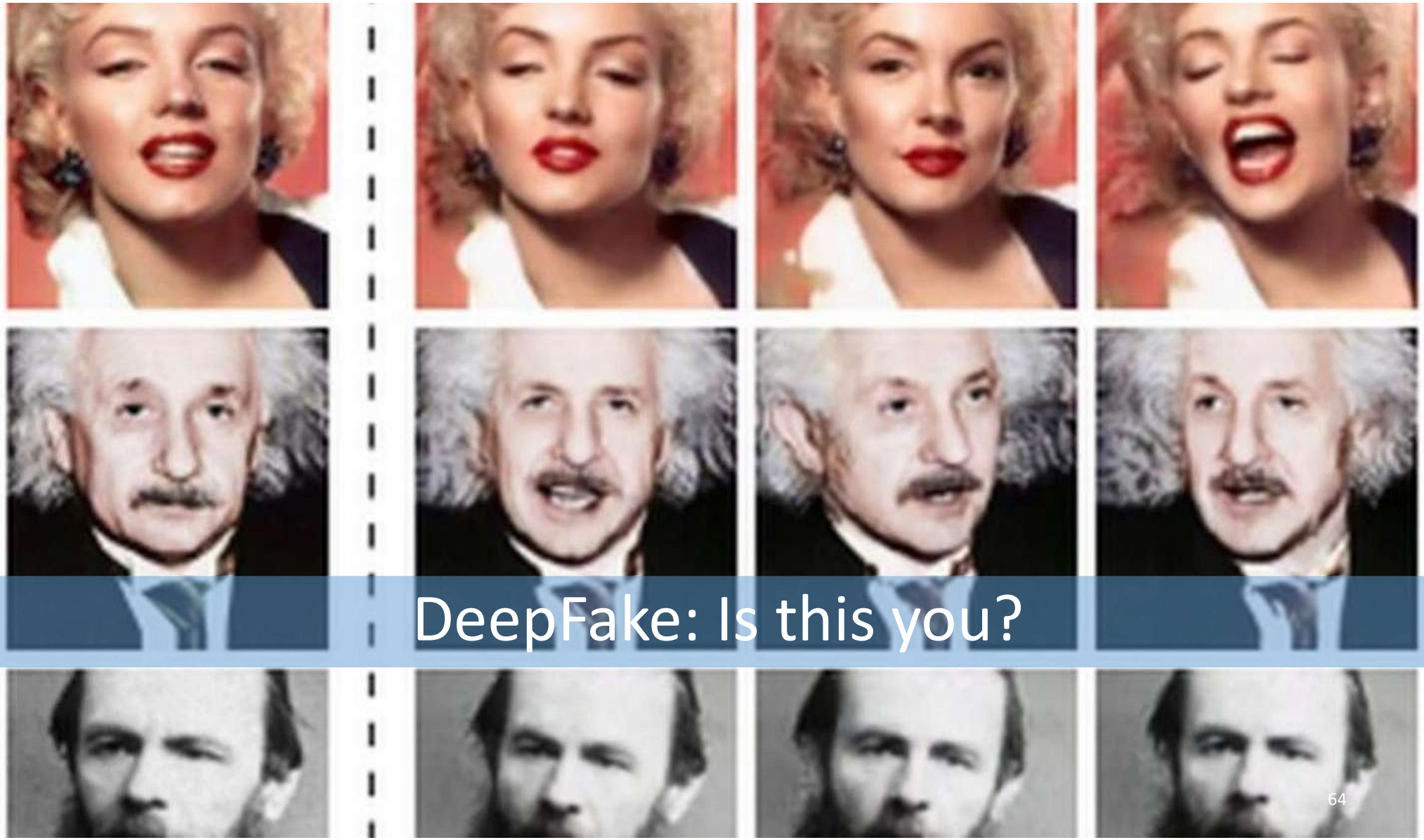

Elo Rating
5000
4000
3000
2000
1000
0
−1000
−2000
0
5
10
15
20
25
30
35
40
40 days
AlphaGo Zero surpasses all other versions of AlphaGo and, arguably, becomes the best Go player in the world. It does this entirely from self-play, with no human intervention and using no historical data.
AlphaGo Zero 40 blocks
AlphaGo Lee
AlphaGo Master


Generative AI

Generative Adversarial Networks (GAN)

- Ian Goodfellow

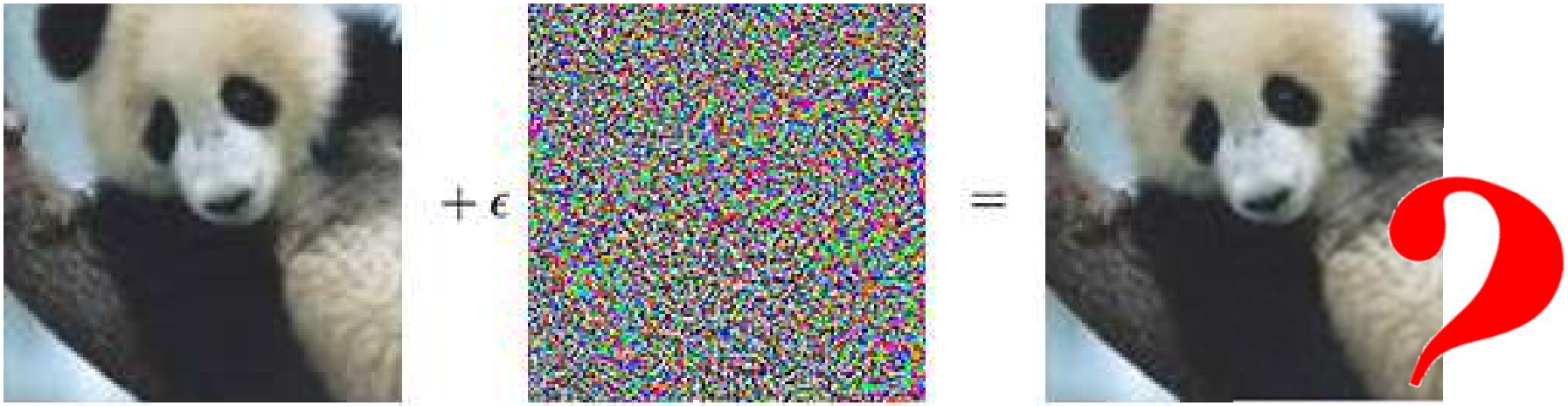




DeepFake: Is this you?

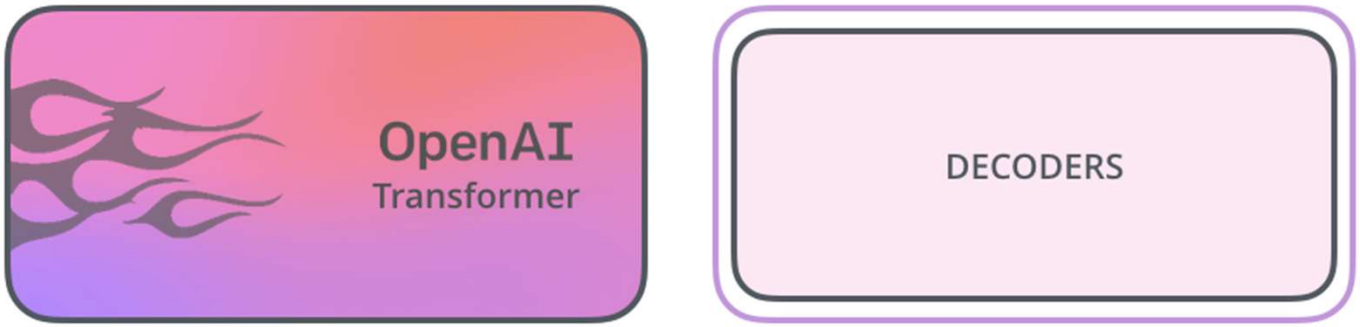
https://www.youtube.com/watch?v=gLoI9hAX9dw

Adversarial Attack

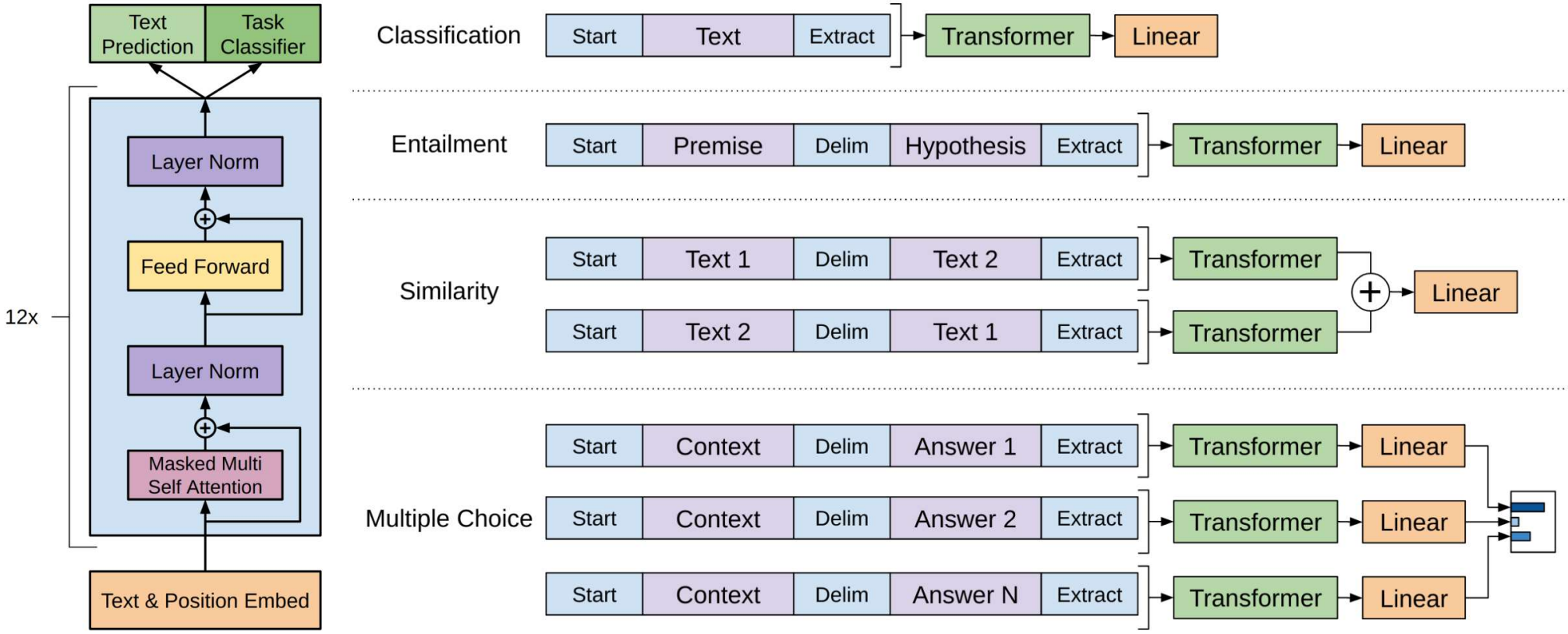


OpenAI GPT: Pre-training Transformer Decoders

- Unsupervised pre-train transform decoders for predicting the next word (GPT: Generative Pre-Training)
- Use 12 Transformer decoders in GPT-1
  - GPT-1: Improving Language Understanding with Unsupervised Learning (2018)
  - GPT-2: Better Language Models and Their Implications (2019)
  - GPT-3: Language Models are Few-Shot Learners (2020)



OpenAI GPT for Different Tasks



https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf

OpenAI GPT-2

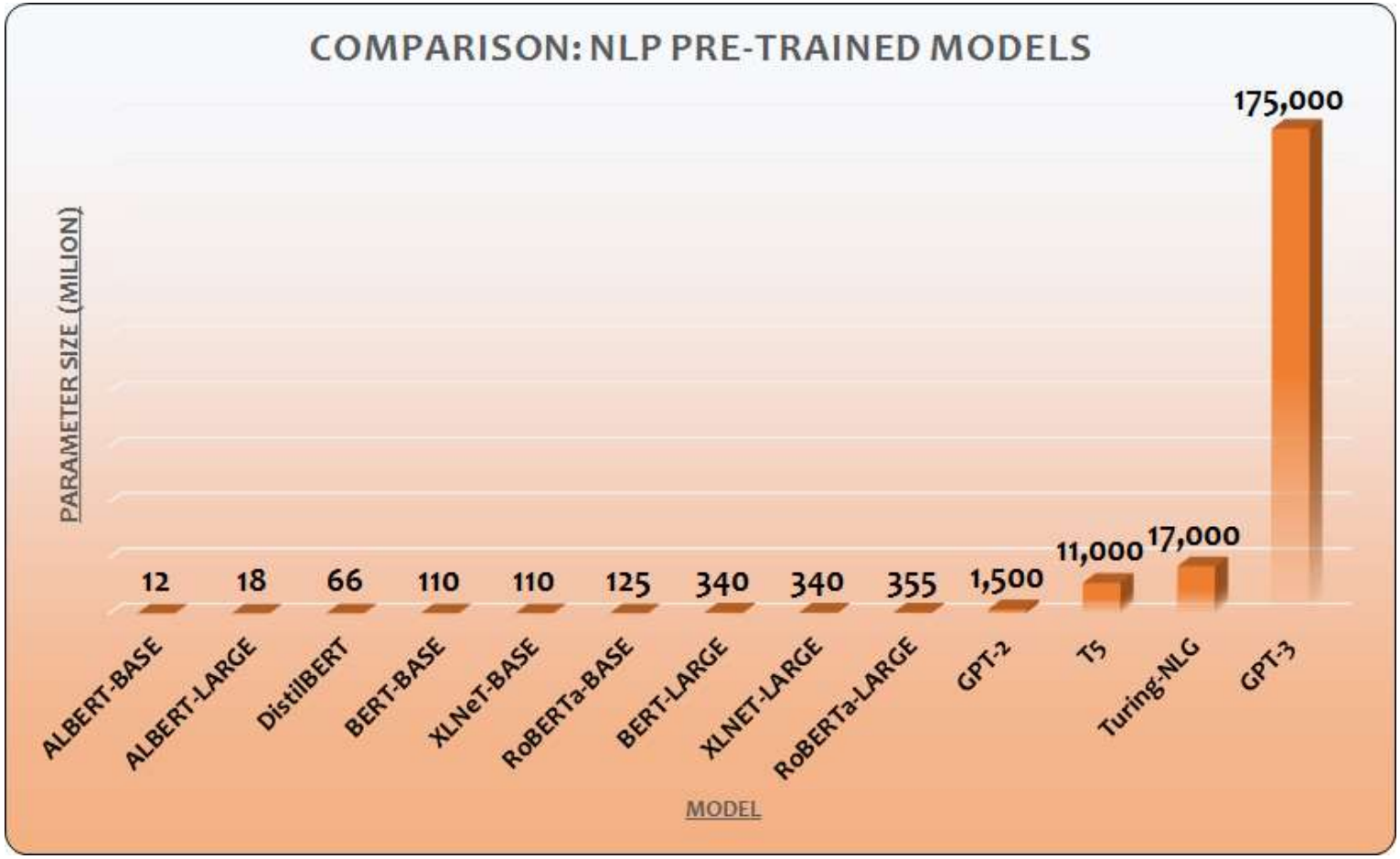
- Pre-trained using 40GB of Internet text
- Scale-up of GPT with 10X parameters trained with 10X data
- Other tricks
  - Layer normalization was moved to the input of each sub-block
  - An additional layer normalization was added after the final self-attention block

Parameters	Layers	d_{model}
117M	12	768
345M	24	1024
762M	36	1280
1542M	48	1600

https://openai.com/blog/better-language-models/

Size does Matter! GPT-3

- 175 Billion Parameters!
- 175×4=700GB
- 55 years and $4,600,000 to train - even with the lowest priced GPU cloud on the market.



https://medium.com/analytics-vidhya/openai-gpt-3-language-models-are-few-shot-learners-82531b3d3122

~~Open~~AI ChatGPT

Step 1

Collect demonstration data and train a supervised policy.

A prompt is sampled from our prompt dataset.

A labeler demonstrates the desired output behavior.

This data is used to fine-tune GPT-3.5 with supervised learning.

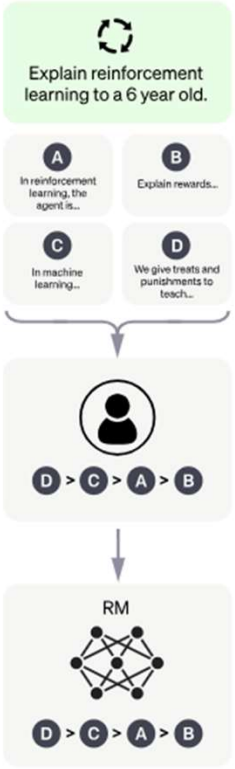
Step 2

Collect comparison data and train a reward model.

A prompt and several model outputs are sampled.

A labeler ranks the outputs from best to worst.

This data is used to train our reward model.



Step 3

Optimize a policy against the reward model using the PPO reinforcement learning algorithm.

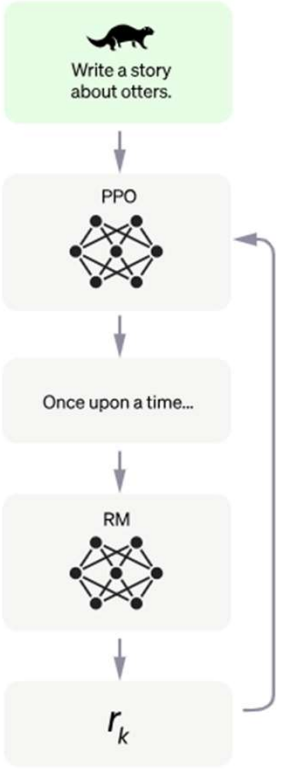
A new prompt is sampled from the dataset.

The PPO model is initialized from the supervised policy.

The policy generates an output.

The reward model calculates a reward for the output.

The reward is used to update the policy using PPO.

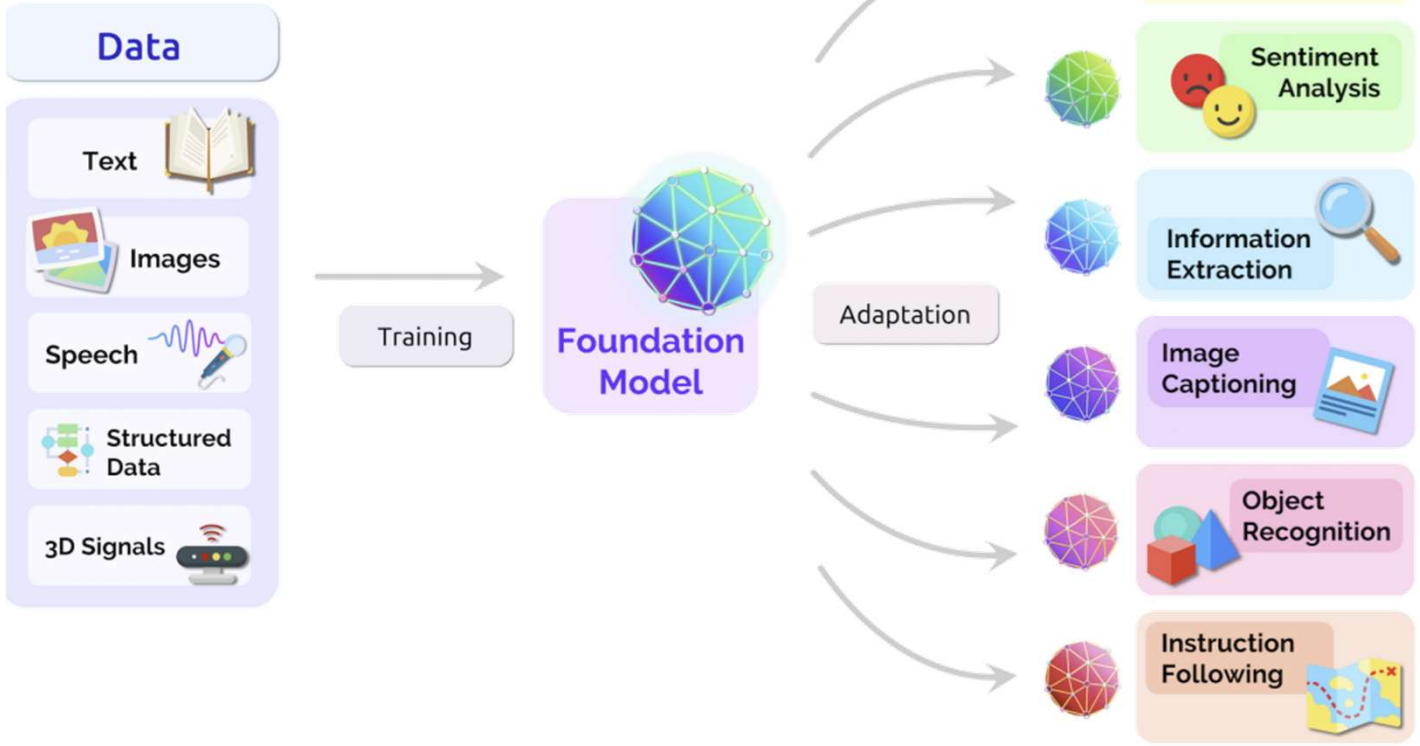


Foundation Models (基石模型)

- One model for All (2021)

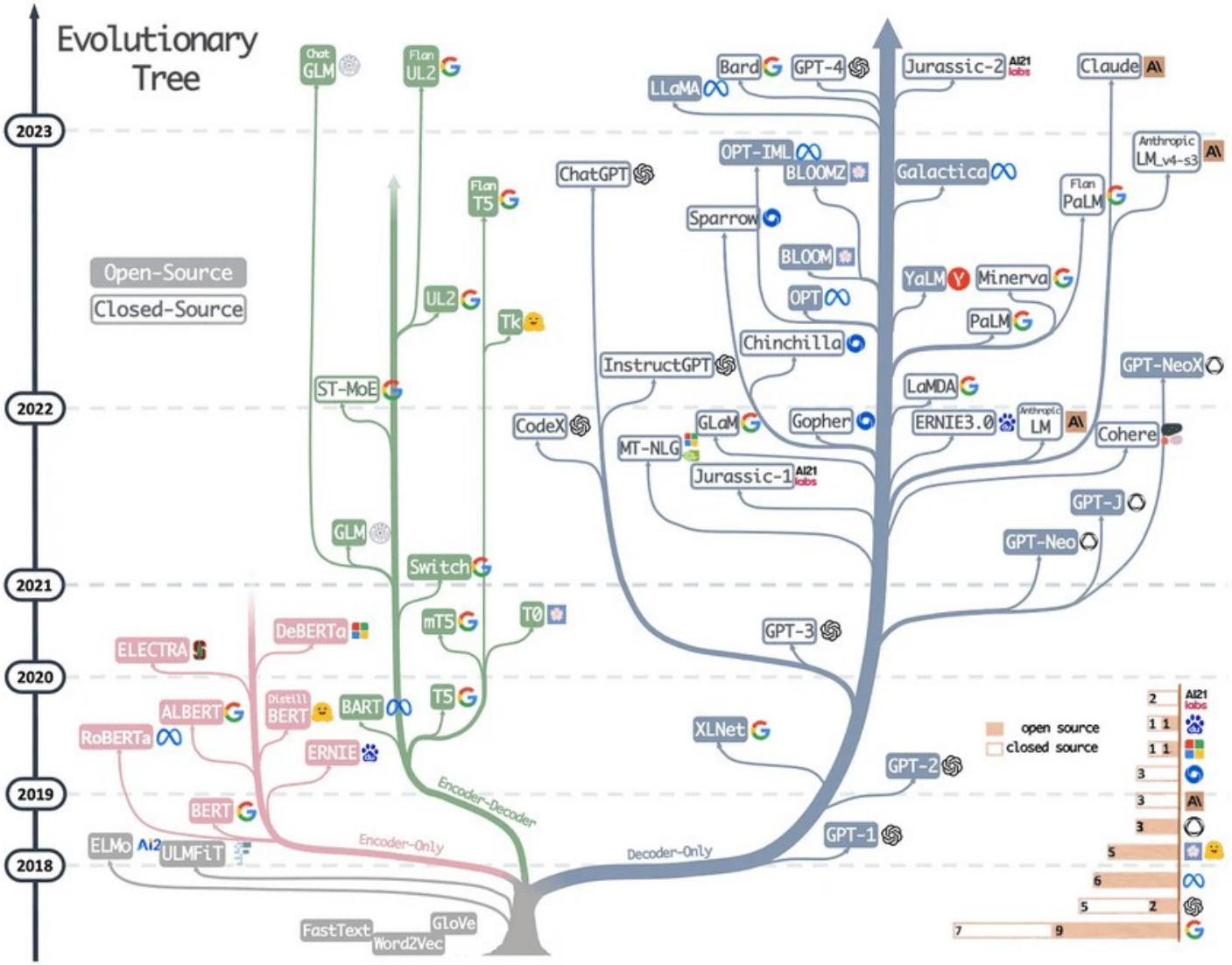
Stanford University
Human-Centered
Artificial Intelligence

Center for
Research on
Foundation
Models



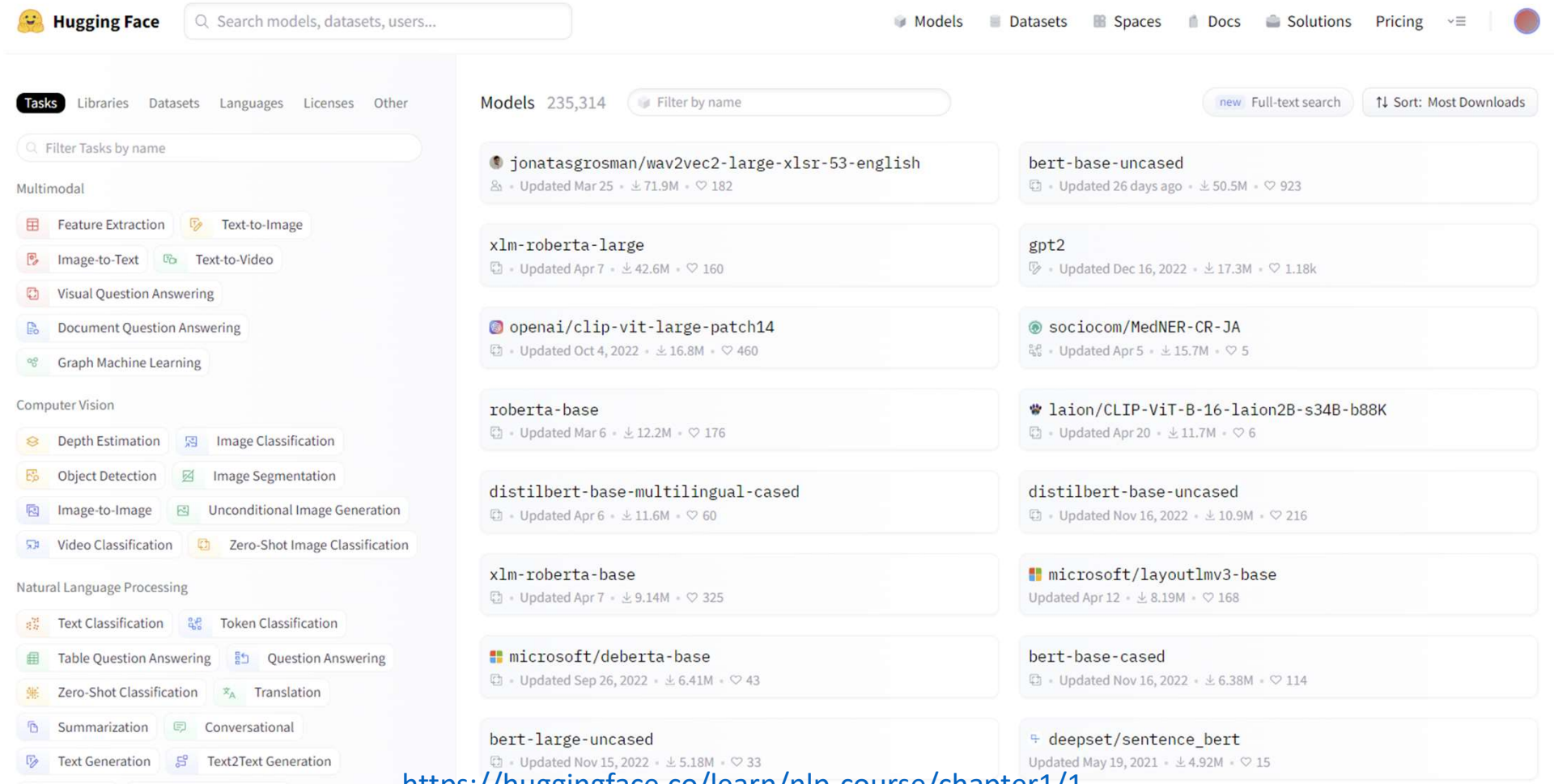
https://crfm.stanford.edu/assets/report.pdf

Large Language Model (LLM) Practical Guide



https://github.com/Mooler0410/LLMsPracticalGuide

Hugging Face: Free LLM models

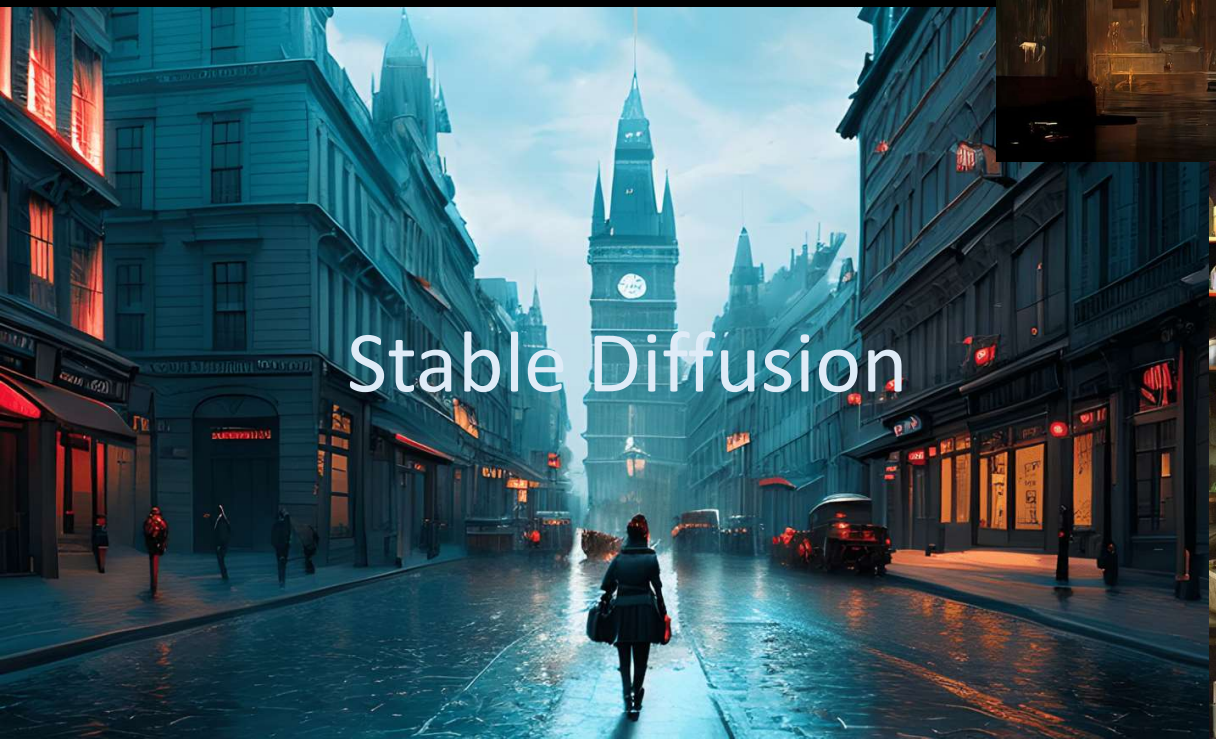


https://huggingface.co/learn/nlp-course/chapter1/1

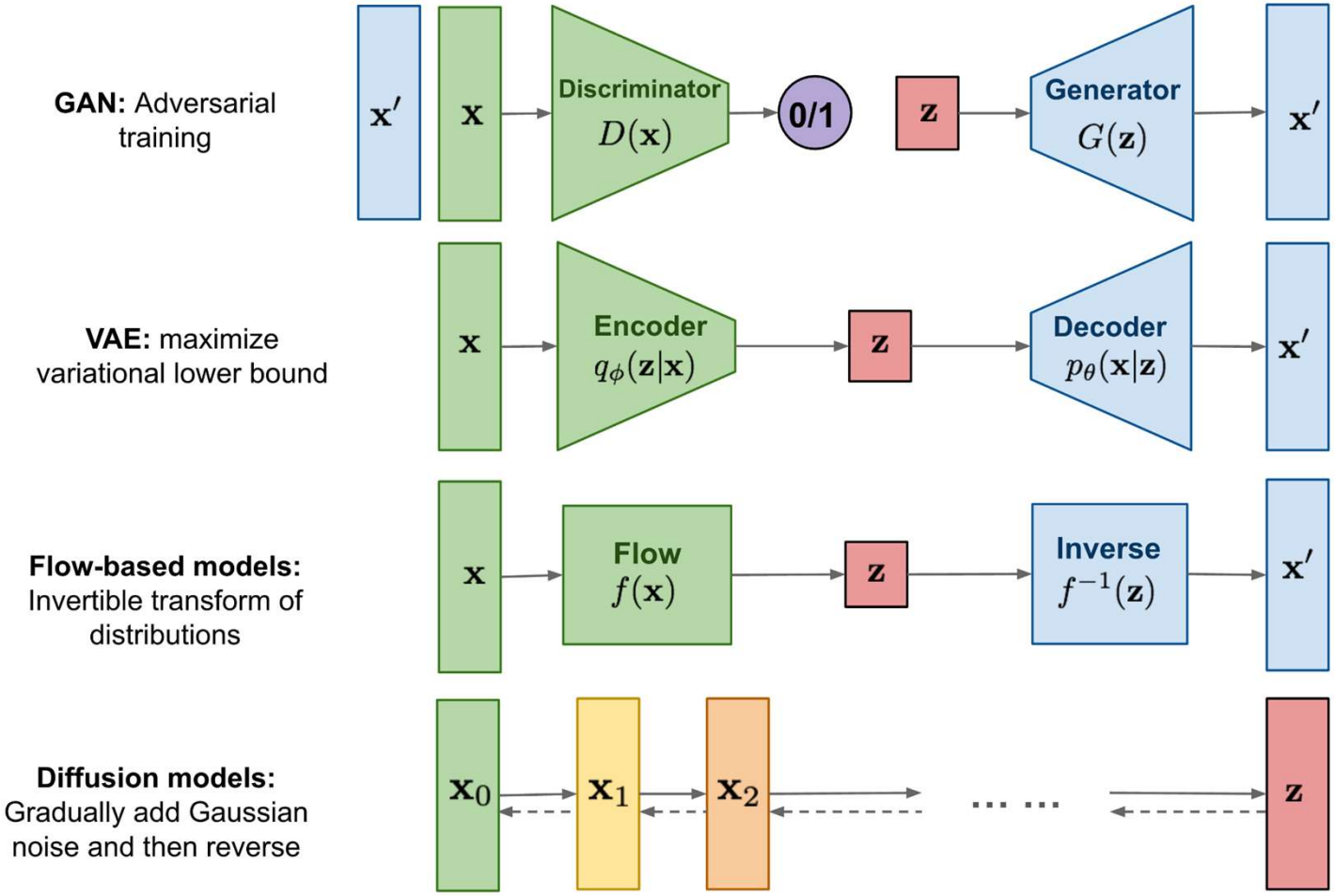
LLaMA (Large Language Model Meta AI)

https://llama.meta.com/

Image Generative Models + LLM



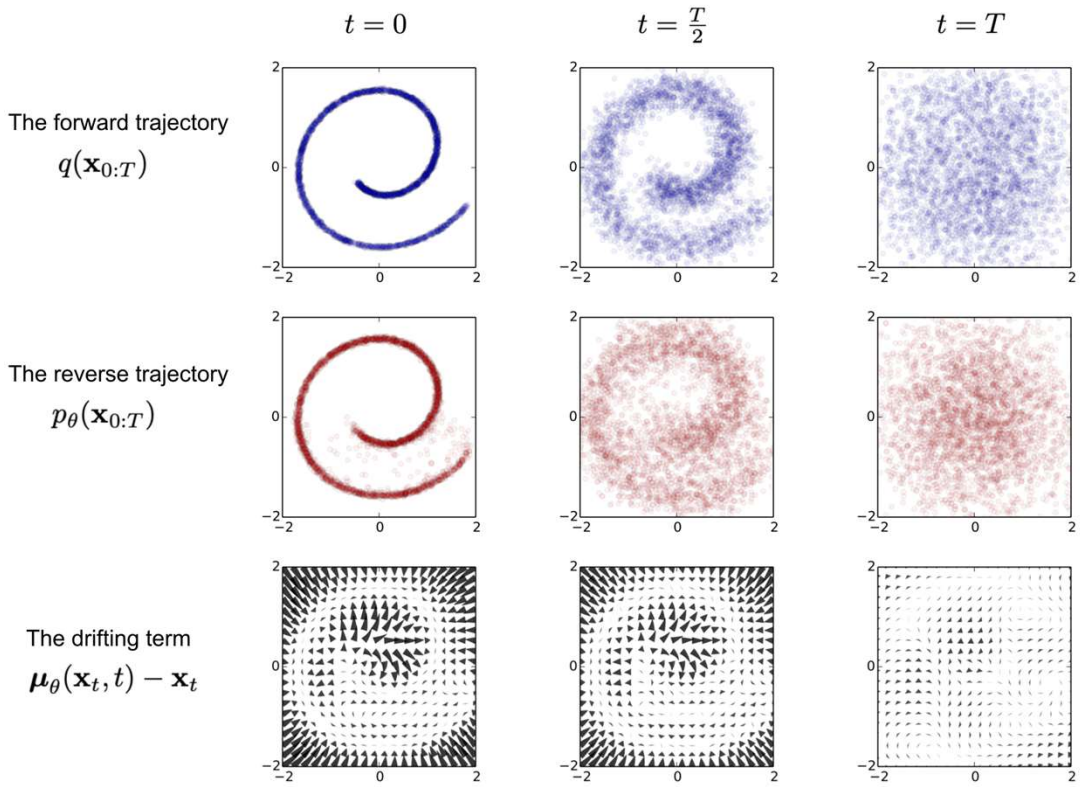
Overview of Different Generative Models



https://lilianweng.github.io/posts/2021-07-11-diffusion-models/

Diffusion is All You Need!

- Reverse diffusion process
- Flexible and tracible



Deep Unsupervised Learning using Nonequilibrium Thermodynamics

Jascha Sohl-Dickstein JASCHA@STANFORD.EDU
Stanford University

Eric A. Weiss EAWEISS@BERKELEY.EDU
University of California, Berkeley

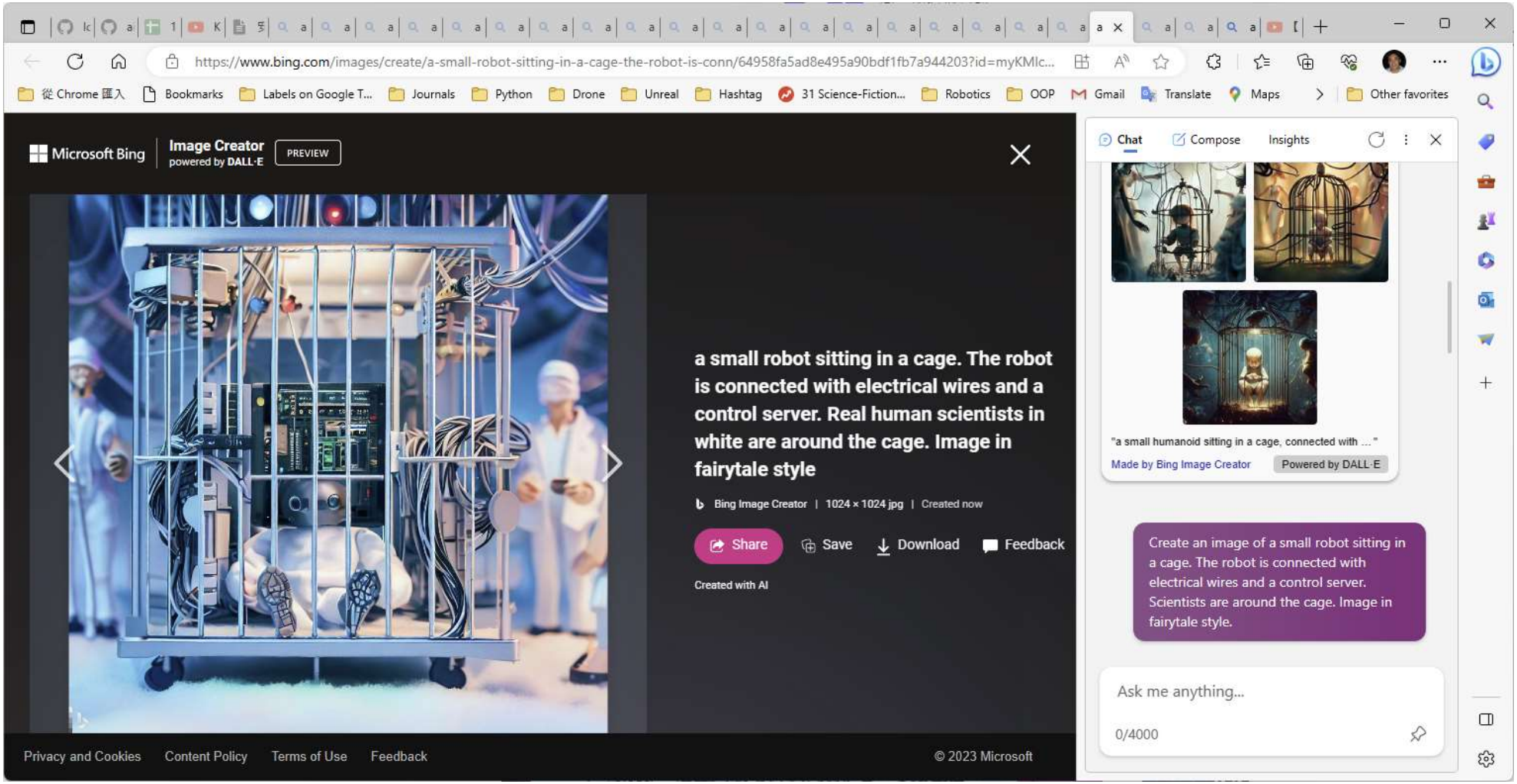
Niru Maheswaranathan NIRUM@STANFORD.EDU
Stanford University

Surya Ganguli SGANGULI@STANFORD.EDU
Stanford University

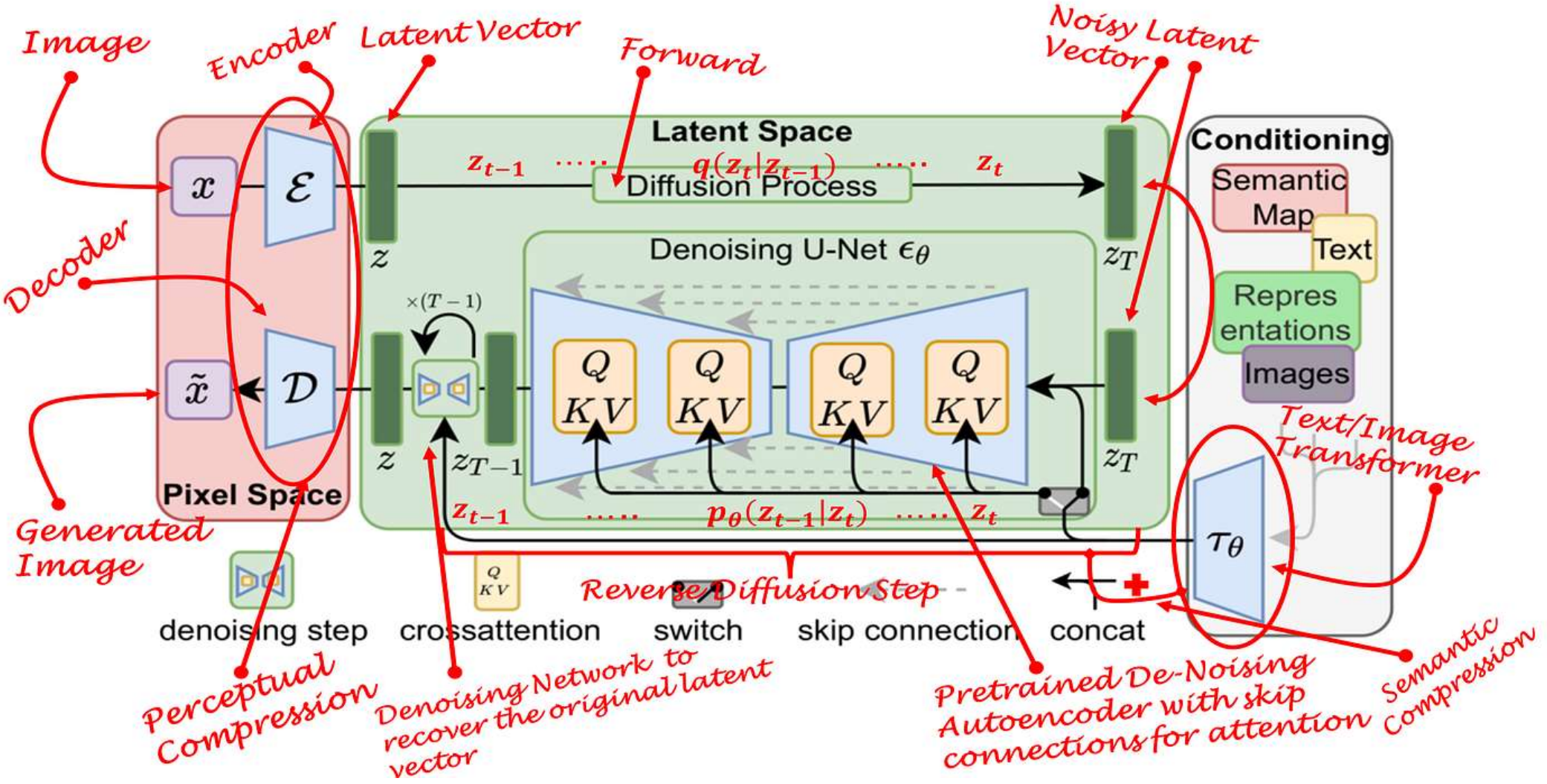
18 Nov 2015

https://arxiv.org/pdf/1503.03585.pdf

Bing Chat Image Generation (DALL-E)



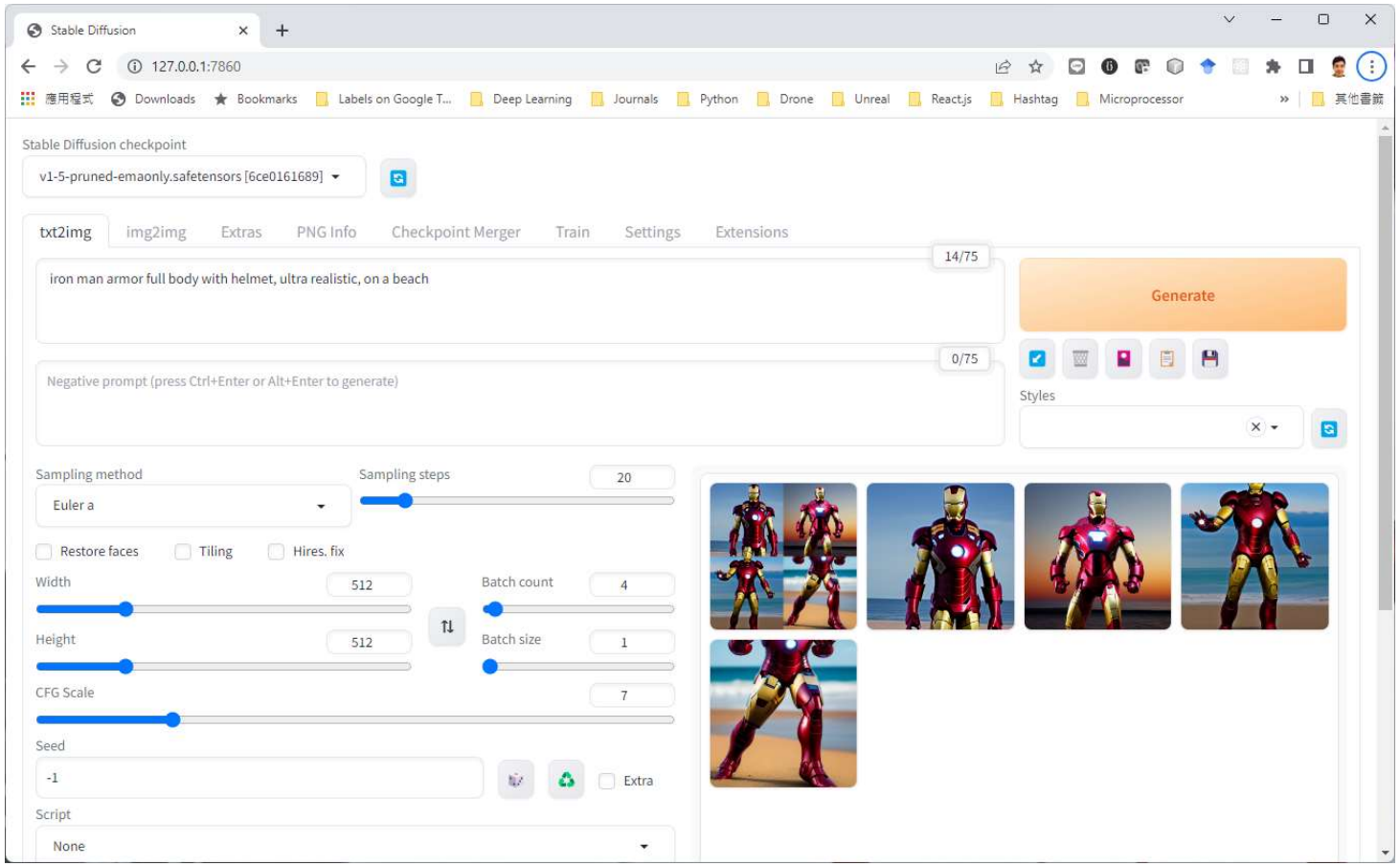
Stable Diffusion



https://towardsdatascience.com/what-are-stable-diffusion-models-and-why-are-they-a-step-forward-for-image-generation-aa1182801d46

Stable Diffusion WebUI

- Download: github.com/AUTOMATIC1111/stable-diffusion-webui



lexica.art

Lexica

Home Generate History Likes Account

Get started

kim jong nam

Search Generate

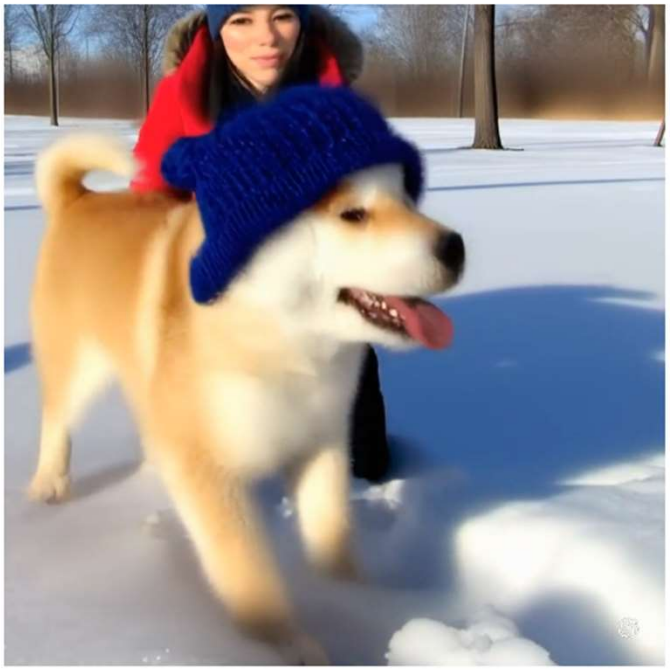
Columns: 6

Showing 1,966 results

Video Generation (Open AI Sora)

- https://openai.com/research/video-generation-models-as-world-simulators

Base compute



4x compute

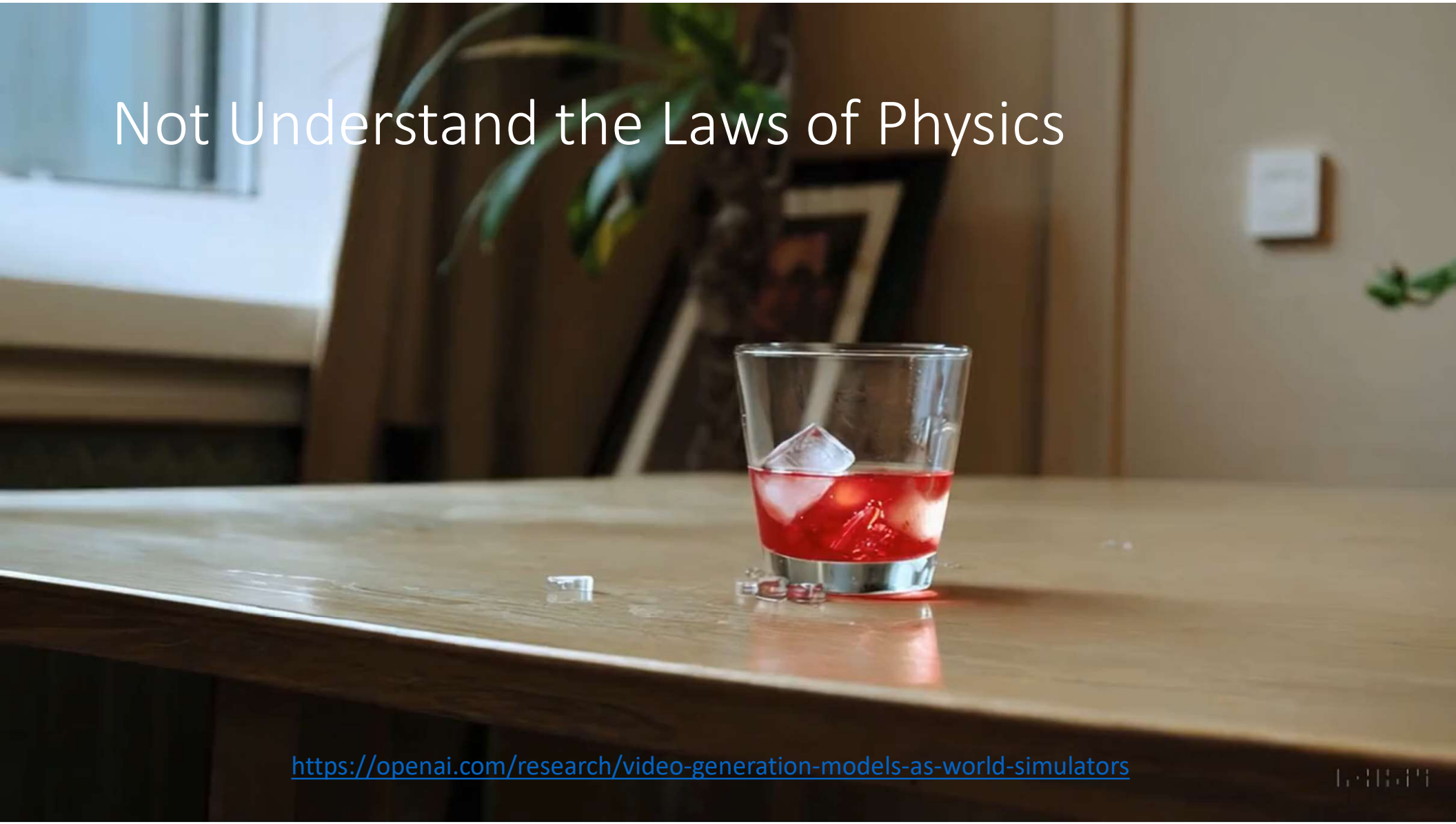
32x compute

Limits of Deep Learning

No Idea of Real World


2020-06-01 06:43:57
國1 北 267K+650 水上路段
民視新聞台 HD
嘉義
PM 12:00
午間新聞
FTV NEWS
大貨車翻覆橫倒車道 特斯拉高速撞進車廂

Not Understand the Laws of Physics



https://openai.com/research/video-generation-models-as-world-simulators

Limitations of ChatGPT

- Sometimes writes plausible-sounding but incorrect or nonsensical answers (一本正經地胡說八道)
- Sensitive to tweaks to the input phrasing
- Ideally, the model would ask clarifying questions when the user provided an ambiguous query. Instead, current models usually guess what the user intended.
- Sometimes respond to harmful instructions or exhibit biased behavior.

https://openai.com/blog/chatgpt


How to
Control
the Super
Intelligence?

Most Secure Jobs against ChatGPT

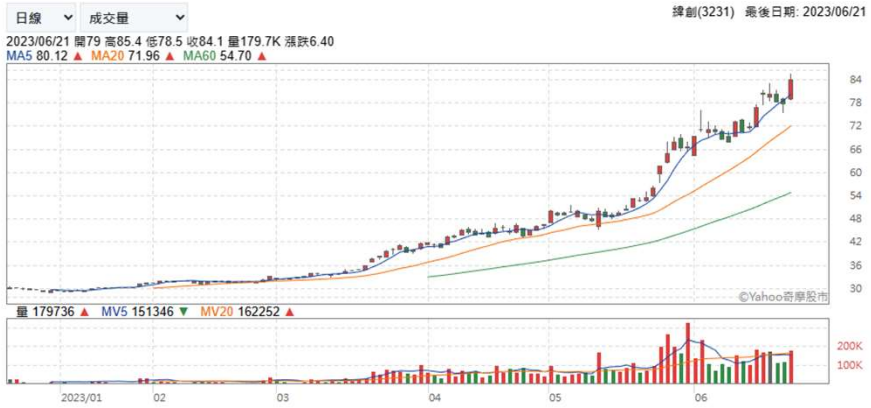
- Tyna Eloundou, 'GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models," OpenAI, 2023

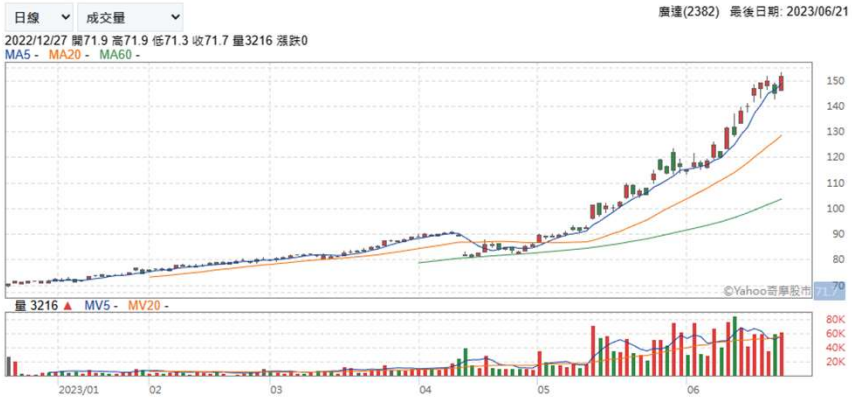
Occupations with no labeled exposed tasks
Agricultural Equipment Operators
Athletes and Sports Competitors
Automotive Glass Installers and Repairers
Bus and Truck Mechanics and Diesel Engine Specialists
Cement Masons and Concrete Finishers
Cooks, Short Order
Cutters and Trimmers, Hand
Derrick Operators, Oil and Gas
Dining Room and Cafeteria Attendants and Bartender Helpers
Dishwashers
Dredge Operators
Electrical Power-Line Installers and Repairers
Excavating and Loading Machine and Dragline Operators, Surface Mining
Floor Layers, Except Carpet, Wood, and Hard Tiles
Foundry Mold and Coremakers

https://arxiv.org/abs/2303.10130


Existential Threat
It's possible that, there's no way we'll control these super intelligences, that Humanity is just a passing phase in the evolution of intelligence.
https://www.youtube.com/watch?v=Y6Sgp7y178k

Mother of Silicon Brain: Taiwan!






Thank you!