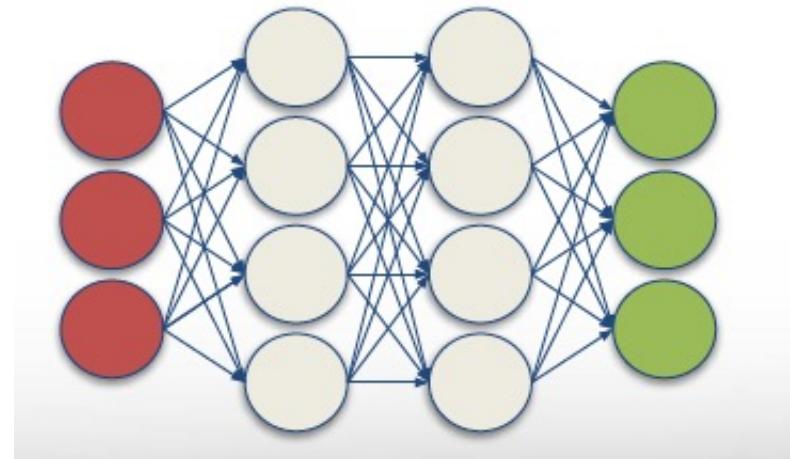


Neural Networks

Hualou Liang
Drexel University

<https://liang-lab.org/>



Free Credits for Neural Networks Class via the Google Cloud Platform (GCP) Education Grant



Google Cloud

A screenshot of the Google Cloud Platform Billing Overview page. The page has a blue header with the text "Google Cloud Platform". Below the header is a navigation bar with "Billing" selected. The main content area shows the project name "BMES-725-001. Neural Networks" and a "RENAME BILLING ACCOUNT" button. A red oval highlights this button. Below the project name, the billing account ID "018A99-B4AFA7-937F-47" is displayed. The "Credits" section shows a balance of "\$100.00" with a note "Out of \$100.00" and a "Days remaining" of "364" ending "Apr 1, 2020". At the bottom, it says "Projects linked to this billing account" and "There are no projects linked to this billing account."

<https://cloud.google.com/>

You will be asked for a name and email address, which needs to match the Drexel domain. A confirmation email will be sent to you with a coupon code.

Every day brings new headlines for how deep learning is changing the world around us

- Google DeepMind's **AlphaGo**
- Deep Learning Makes **Driverless Cars** Better at Spotting Pedestrians
- Deep learning algorithm **diagnoses skin cancer** as well as seasoned
- **AlphaFold**: a solution to a 50-year-old grand challenge in biology
- **ChatGPT** - optimizing language models for dialogue

AlphaGo → AlphaGo Zero → AlphaZero → AlphaStar



In March 2016, Deepmind's **AlphaGo** beat 18 times world champion Go player Lee Sedol 4–1 in a series watched by over 200 million people

AlphaGo → AlphaGo Zero → AlphaZero → AlphaStar

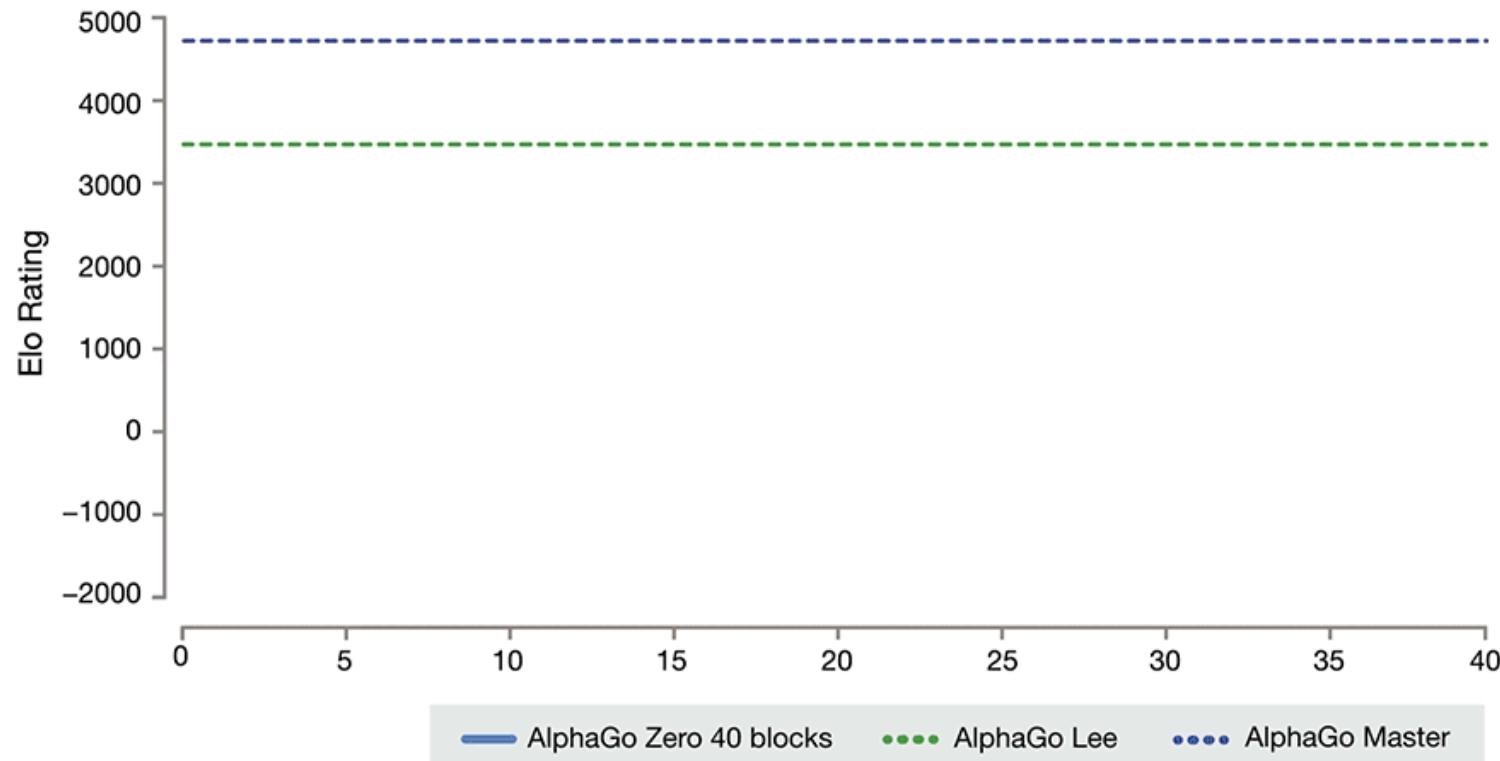
On 18th October 2017, DeepMind unveiled a new variant of the algorithm, **AlphaGo Zero**, that had defeated AlphaGo 100–0.

Mastering the game of Go without human knowledge, *Nature*, 2017

Jan 24, 2019, **AlphaStar**: Mastering the Real-Time Strategy Game StarCraft II
(DNN + raw game data + supervised learning + reinforcement learning)

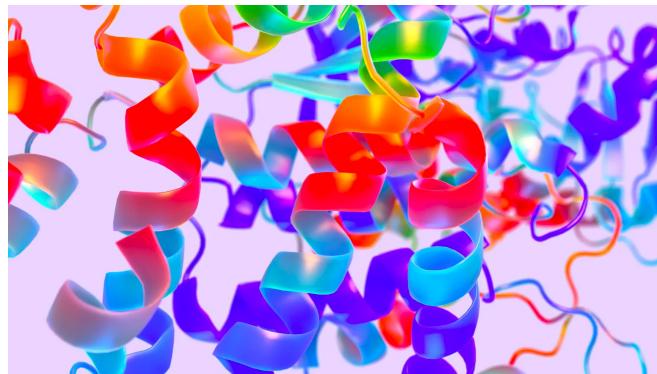
Grandmaster level in StarCraft II using multi-agent reinforcement learning,
Nature, 2019

DeepMind's AI has now outcompeted nearly all human players at StarCraft II, Oct 2019



Mastering the game of Go without human knowledge
Nature, 2017

AlphaFold: Using AI for scientific discovery



Predict the 3D structures of proteins from their amino-acid sequences:
A watershed moment for protein structure prediction

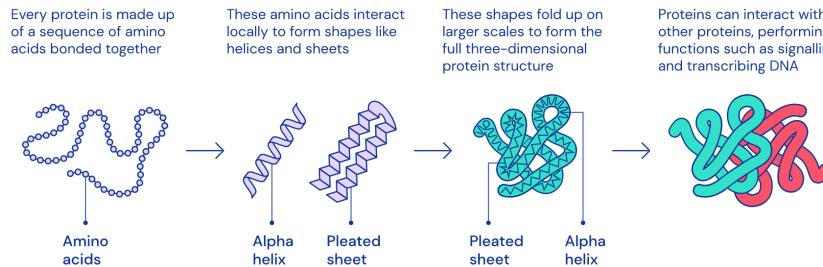


FIGURE 1: COMPLEX 3D SHAPES EMERGE FROM A STRING OF AMINO ACIDS.

Senior et al. Improved protein structure prediction using potentials from deep learning, Nature, 2020

AlphaFold: Timeline of a Breakthrough



2016 - As one team celebrates, another is formed

AlphaGo defeats legendary Go player Lee Sedol in Seoul.

2018 - The first public test of AlphaFold's performance

2020 - A solution to a 50-year-old grand challenge in biology

Senior et al. *Nature*, 2020

2021 - Putting the power of AlphaFold into the world's hands

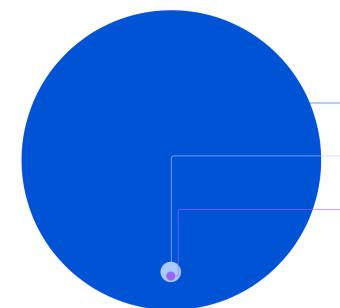
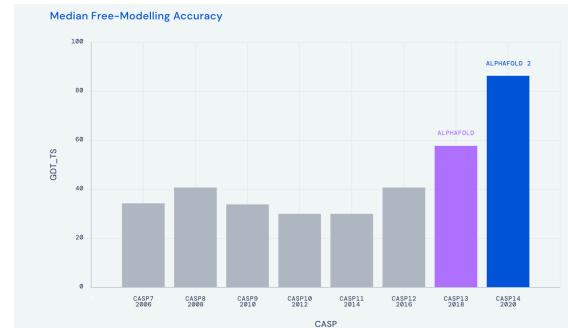
Jumper et al., *Nature*, 2021

2022 - Our database grows in orders of magnitude

The entire protein universe': AI predicts shape of nearly every known protein, Nature News

<https://doi.org/10.1038/d41586-022-02083-2>

<https://www.deepmind.com/blog/alphafold-reveals-the-structure-of-the-protein-universe>

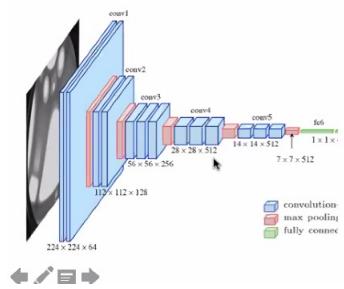


Number of Protein Structures

deep nets – a perfect storm

ca. 2012

Deep Architectures



Big Data



Big Computers

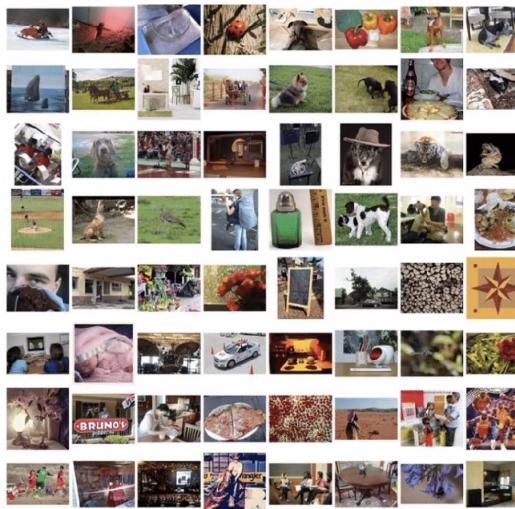
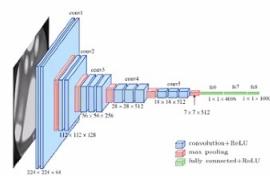


Selected Applications

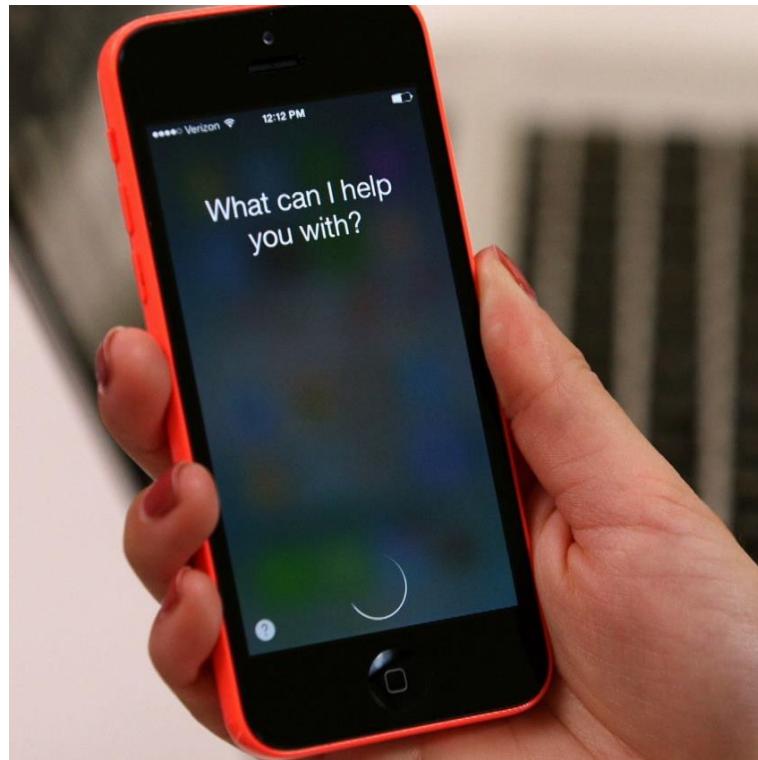
- Some Cool Applications using Deep Learning
 - From image to text
 - Self-driving cars
 - Computer Vision: object and activity recognition
 - Speech recognition and machine translation: Siri
 - Music and the arts
 - Natural Language Processing (NLP)
- A brief history of Neural Networks and Deep Learning
- Basics of Neural Networks

object classification

- **ImageNet** challenge
 - irregularly sized color images (often cropped to 256x256)
 - 20,000 classes of objects
 - 14 million images
- SOTA: % correct = 98.7% [top-5]
 - FixEfficientNet-L2 (FaceBook AI)
 - 480M parameters



Speech recognition and Machine translation



"Hey Siri, show me my Life Goals list."

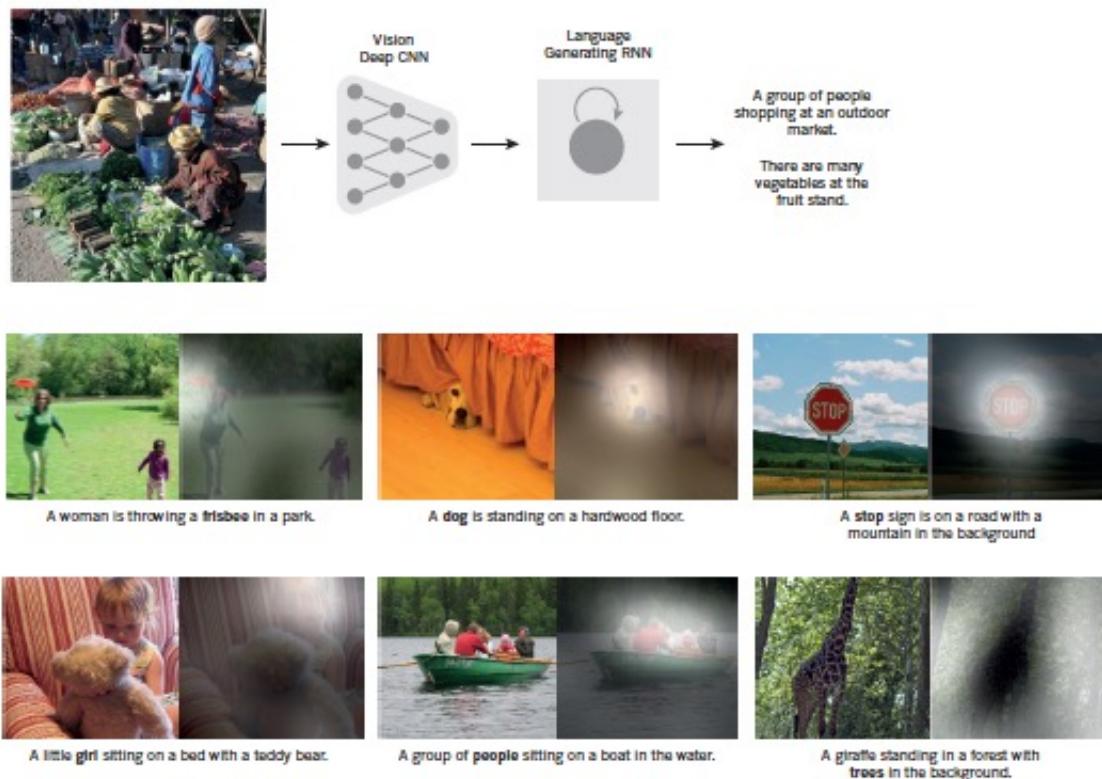
Deep Learning in Arts



Imitating famous painters:
transfer the artistic style of one input image onto another

A Neural Algorithm of Artistic Style" by L. Gatys, A. Ecker, and M. Bethge.
<http://arxiv.org/abs/1508.06576>

From image to text

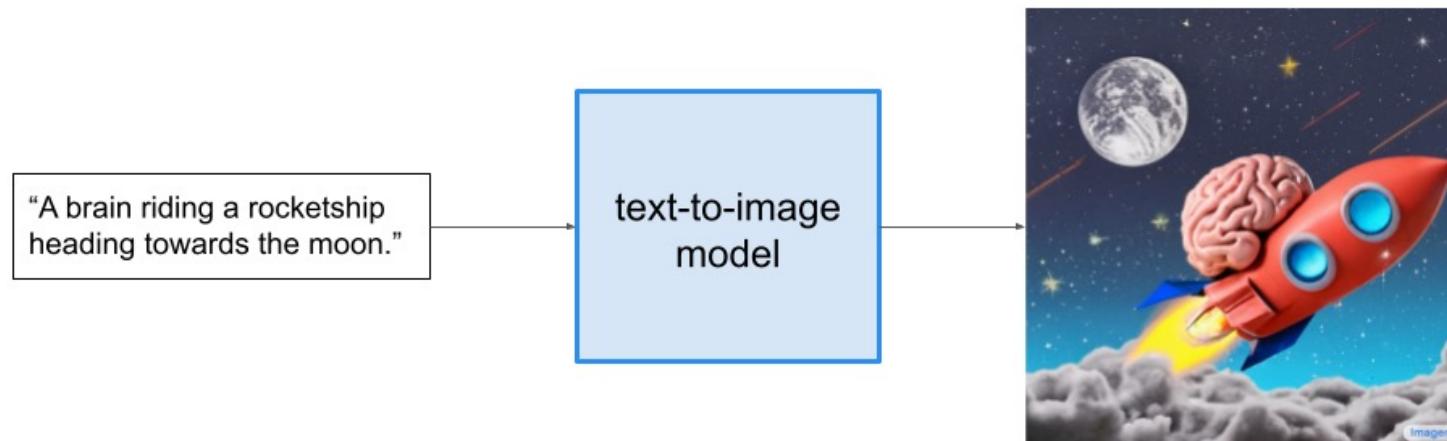


LeCun, Bengio & Hinton, Deep Learning, Nature, 2015

AI Generated Content (AIGC)

- **AI image:** text-to-image diffusion model ([Stable Diffusion](#))
- **AI music:** music generation with stable diffusion ([Riffusion](#))
- **AI text:** ChatGPT - AI chatbot writes smart essays ([ai.com](#))
- **AI code:** GitHub [Copilot](#): an AI pair programmer that offers autocomplete-style suggestions as you code

Diffusion Models underlying Text to Image



Text2Image

OpenAI's **DALLE-2**: Text to Image



A bowl of soup that is a portal to another dimension as digital art

Tryout: DALL-E mini (Craiyon): <https://www.craiyon.com/>

Text to Image: Google's **Imagen**



A dragon fruit wearing karate belt in the snow" and "a photo of a Corgi dog riding a bike in Times Square. It is wearing sunglasses and a beach hat"

Stable Diffusion: a latent text-to-image diffusion model

<https://thealgorithmicbridge.substack.com/p/stable-diffusion-is-the-most-important>

- Public Release: <https://stability.ai/blog/stable-diffusion-public-release>
- open-source

stability.ai



<https://github.com/CompVis/stable-diffusion>
<https://github.com/hlky/stable-diffusion> ?

[Lexica.art](#), a prompt/image/seed search engine

[DreamStudio](#) Lite: used for free for up to 200 image generations (to get a sense of what Stable Diffusion can do.

[paid subscription model](#): Stable Diffusion: £0.01/image whereas DALL-E: \$0.03/image

Midjourney: <https://www.midjourney.com/> 25 free

Rombach et al. High-Resolution Image Synthesis with Latent Diffusion Models, 2022

AI-Generated Picture Won an Art Price



Jason Allen's AI-generated work "Théâtre D'opéra Spatial" took first place in the digital category at the Colorado State Fair

https://www.vice.com/en/article/bvmvqm/an-ai-generated-artwork-won-first-place-at-a-state-fair-fine-arts-competition-and-artists-are-pissed?utm_medium=social&utm_source=vice_facebook&utm_source=VICE_Twitter&utm_medium=social+

TIME: The AI Arms Race Is Changing Everything

Feb 27 / Mar 6 2023



HU

Write an e-mail to congratulate new Drexel admits and mention hat you are excited about meeting all of them in-person.



Subject: Congratulations on Your Admission to Drexel University!



Dear [Name],

I am thrilled to congratulate you on your admission to Drexel University! Your hard work and dedication have paid off, and you have been accepted to one of the best universities in the country. I am sure that you will find your time at Drexel to be a fulfilling and transformative experience.

As a current member of the Drexel community, I am excited about the prospect of meeting you in-person. Drexel is known for its diverse and talented student body, and I am eager to get to know all of the new admits. I am confident that you will find the Drexel community to be welcoming, supportive, and inclusive.

As you prepare for the next chapter of your academic journey, I would like to offer my assistance in any way that I can. If you have any questions or concerns, please do not hesitate to reach out to me. I would be happy to share my experiences and offer advice to help you make the most of your time at Drexel.

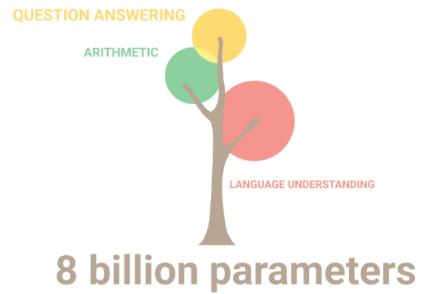
Again, congratulations on your admission to Drexel University. I look forward to meeting you soon and welcoming you to our community.

Best regards,

[Your Name]

Write an e-mail to congratulate new Drexel admits and mention hat you are excited about meeting all of them in-person.

Emergent Phenomena in Large Language Models



<https://ai.googleblog.com/2022/04/pathways-language-model-palm-scaling-to.html>

HU

what is large language model



A large language model is an artificial intelligence model that is trained to generate and understand natural language text. The "large" in the term refers to the size of the model, which is typically measured by the number of parameters it has. The more parameters a model has, the more data it has seen during training, and the more complex relationships between words, phrases, and sentences it can learn. This can lead to improved performance on natural language processing tasks such as text generation, machine translation, and question-answering.



Large language models like OpenAI's GPT-3 have been trained on massive amounts of text data and are capable of generating human-like text. This has led to much excitement in the AI research community, but also concerns about their potential impact on society, particularly with regards to issues such as bias, privacy, and control.

What is GPT-3? Finish Your Sentences for You

prompt

completion

HU What is GPT-3?

GPT-3 (Generative Pretrained Transformer 3) is a language model developed by OpenAI. It is one of the largest and most advanced language models available today, with over 175 billion parameters. It is based on a type of neural network called a Transformer, which is trained to generate text by predicting the next word in a sequence given the previous words.

GPT-3 has been trained on a diverse range of text data, including books, articles, and websites, and as a result, it has a vast knowledge of the world and the ability to generate human-like text. This has made it a popular tool for a variety of natural language processing tasks, such as text generation, machine translation, and question-answering.

However, despite its impressive abilities, GPT-3 is not without limitations. For example, like any AI model, it is only as good as the data it has been trained on, and it can sometimes generate biased or misleading text. Additionally, because of its size and computational requirements, it is not practical for many use cases. Nonetheless, it remains a remarkable example of the progress being made in the field of AI and natural language processing.

Given a set of words (**prompt**), find the words that are most likely to follow it (**completion**).



What is ChatGPT?

ai.com

HU

What is ChatGPT?



ChatGPT is a conversational model developed by OpenAI based on the GPT (Generative Pretrained Transformer) architecture. It is a type of language model that has been trained on a large dataset of text from the internet and is designed to generate human-like text.



In practice, this means that when given a prompt or context, ChatGPT can generate a response that continues the conversation in a natural and coherent manner. This has made ChatGPT a popular tool for building conversational AI systems, such as chatbots, virtual assistants, and customer service systems.

ChatGPT can answer a wide range of questions, provide explanations and information on a variety of topics, and even engage in casual conversation. However, like any AI model, it has limitations and can sometimes generate responses that are irrelevant or inappropriate. Nevertheless, it represents a significant advance in the field of conversational AI and has the potential to transform the way we interact with technology.

Holding conversations with the model, and rewarding it for good responses and punishing it for bad ones—a process called **Reinforcement Learning from Human Feedback (RLHF)**

GPT 3.5 + RLHF + some guardrails



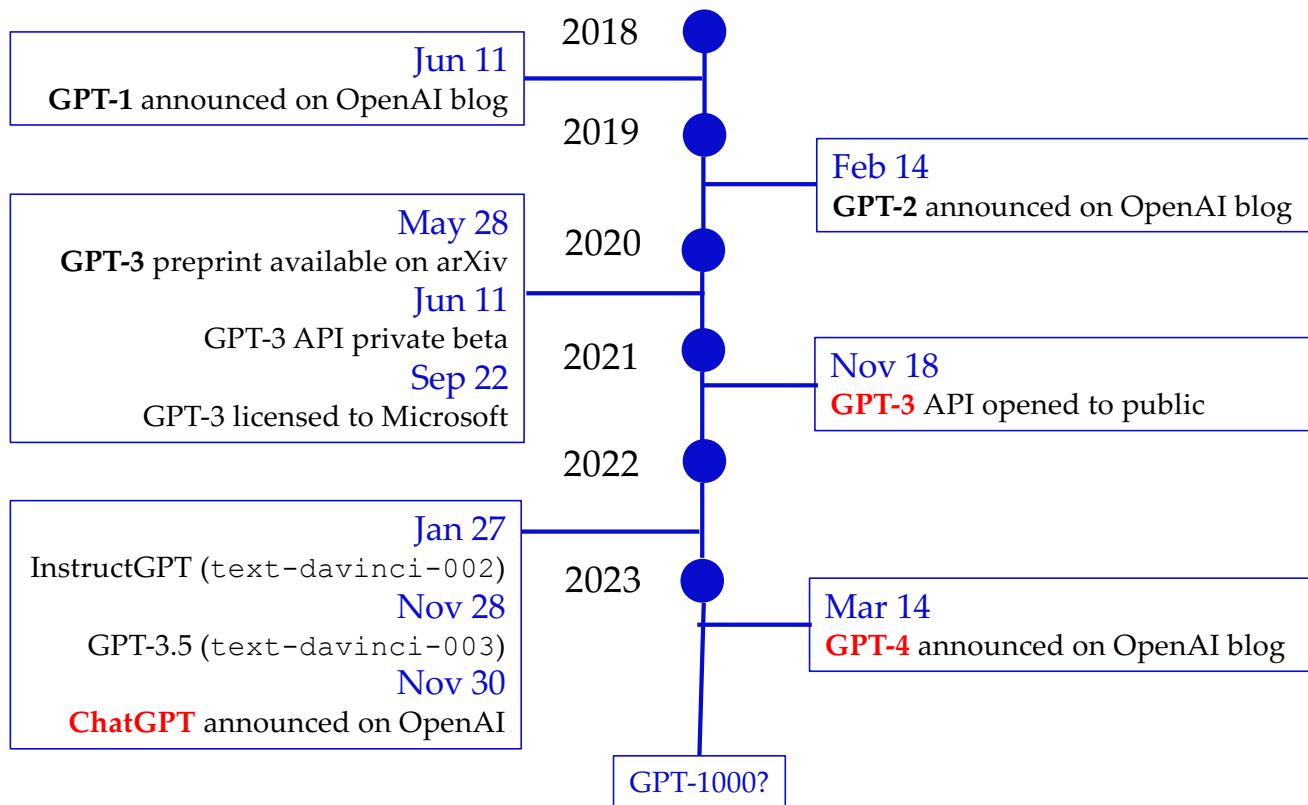
GPT-4: OpenAI's Most Advanced System

March 14, 2023

- Large multimodal: image + text as inputs, text outputs
- Human-level performance: outperform ChatGPT by scoring in the top 10% of test takers (90th percentile on BAR exam, SAT, GRE etc)
- Surpass ChatGPT in its advanced reasoning capabilities
- 25,000 words context: allows full documents to fit *within a single prompt*
- More creative & collaborative: generate, edit, and iterate with users on writing tasks
- Models available via API: gpt-4, gpt-4-0314, gpt-4-32k, and gpt-4-32k-0314

Model	Prompt Cost	Completion Cost	Context Window
gpt-4	\$0.03/1K prompt	\$0.06/1K completion	8K context window (~13 pages)
gpt-4-32k	\$0.06/1K prompt	\$0.12/1K completion	32K context window (~52 pages)

Timeline to ChatGPT



I hope that this class will give you the tools to build cool projects like those!

Goals

- Learn the fundamental concepts and techniques in neural networks and deep learning.
- Exposure to deep neural network techniques when dealing with big data sets in a wide range of biomedical applications.
- Provide hands-on experience in designing and programming deep learning algorithms to solve data science problems in biomedicine.

Course Schedule and Policies

- Weds, 6:00-8:50 pm, Zoom
- Lectures + presentation
- Office hours: Tues, 5-6 pm, Zoom Monell 103

Course Schedule and Policies

Grading:

- Class participation (10%),
- Paper review (20%),
- Paper presentation (teams of 2-3): 30%,
- Project (team size maximum of 3): 40%

Course Schedule and Policies

- Class participation (10%): A significant portion of the class will be based on reading and discussing the latest literature. Every student should read the assigned papers before class and should actively engage in paper discussions in class.

Course Schedule and Policies

- **Paper review (20%):** each team selects 2 or more other papers to review. The review concisely summarizes the key findings of the paper, highlight interesting ideas, weaknesses and potential extensions. No more than 5 pages. Format flexible, but this is with a minimum margin size of 1-inch minimum margins and a minimum font size of 11 points.

Course Schedule and Policies

- **Paper presentation (30%):** each team selects one or more of the suggested papers to present in detail to the class. Students presenting the same papers should work together to prepare about 40 minutes talk on the papers. The talk should be self-contained: give background/motivation, intuition and the key results from the papers that you think are the most interesting. Do not need to cover everything; present derivations if they convey insights. Clarity will be the main evaluation criterion. All students are expected to read the assigned paper and participate in the discussions. The topic of the presentation can be in one of the following categories:

computational neuroscience,

computational biology,

medical imaging,

electronic medical records and

neural data analysis

You are free to select your own topic and work in teams, but please check in with the instructor.

Course Schedule and Policies

- **Final project (40%):** This is the main component of the course; start early! This should be a research project that is related to the course material. You can even follow up with the topic of your paper presentation to further develop final project. The students will form teams of 2-4 and choose one or more datasets suggested or select their own project. Teams are expected to work on the research project throughout the second half of the quarter and produce final papers by following the format below (For details, see **Project Grading & Guidelines**). Each team will present the paper to the entire class at the end of the semester.

Course Schedule and Policies

Where to find datasets:

In todays "Big Data" world, there are an abundance of high-quality, free datasets to enjoy and explore. In addition to the data sometimes supplied by the papers provided in the class, below is a short list of websites that are great resources for data:

- [Kaggle](#)
- [FiveThirtyEight](#)
- [Awesome Public Datasets](#)
- [Datalad](#)
- [Datasets from Awesome Deep Learning](#)
- [COVID-19 in US and Canada \(1point3acres\)](#)

Key Dates

- Deadline for choosing project groups, with tentative project title, for paper review and presentation: **6 pm, 04/18/2023**
- **Paper review due** on the week (in the class 6pm) when the presentation will be given.
- Final Project deadline: **6 pm, 06/06/2023**

Policies

Academic Integrity Policy:

Drexel University is committed to a learning environment that embraces academic honesty. In order to protect members of our community from results of dishonest conduct, the University has adopted policies to deal with cases of academic dishonesty. Please read the definitions, understand, and follow the "Academic Integrity Policy", as written in the Official Student Handbook:

http://drexel.edu/studentaffairs/community_standards/studentHandbook/general_information/code_of_conduct/

Americans with Disabilities Act:

Students with documented disabilities who need course accommodations, have emergency medical information or require special arrangements for building evacuation should contact the instructor within the first two weeks of class. Verification of any special arrangements needs to be made through the *Office of Disability Services, 3201 Arch Street, Suite 210*. For further information visit:

<http://www.drexel.edu/oed/disabilityResources/Overview/.>

Appropriate Use of Course Materials

It is important to recognize that some or all of the course materials provided to you may be the intellectual property of Drexel University, the course instructor, or others. Use of this intellectual property is governed by Drexel University policies, including the policy found here: <https://drexel.edu/it/about/policies/policies/01-Acceptable-Use/>

Briefly, this policy states that course materials, including recordings, provided by the course instructor may not be copied, reproduced, distributed or re-posted. Doing so may be considered a breach of this policy and will be investigated and addressed as possible academic dishonesty, among other potential violations. Improper use of such materials may also constitute a violation of the University's Code of Conduct found here: <https://drexel.edu/cpo/policies/cpo-1/> and will be investigated as such.

Policies

- Assignments and other material will not be accepted late, except as explicitly permitted by the Professor with a discretionary penalty. Excused late assignments can be submitted for credit only, but not for a letter grade (they will be dropped from the average).
- Except when explicitly permitted and instructed to work as a group, students will be expected to work as individuals, with penalties at the discretion of the Professor (usually a failing grade).
- If students do work as a group, including on problem sets, the work they hand in should list the names of all group members. A group cannot have more than 3 (Three) members.
- Missing a test will generally not be permitted except under extraordinary circumstances as determined by the Professor. Missing an exam / homework will result in a failure of that exam/homework, there will be no make-ups allowed without the approval of the instructor for a major life events such as illness or injury. If a make-up is granted, in fairness to other students, there will be a grade point penalty.

Policies

- Regular attendance is required, as the course will include regular class participation. In class, group work will also be a part of the course. Students missing more than two classes must see the instructor to avoid a grade penalty. Any materials or announcements covered in class missed during an absence are the student's responsibility. Students are expected to arrive on time and stay the length of the class. Students who are late more than two times or who leave the class early more than twice will be required to meet with the instructor and may be penalized.
- Laptops for note-taking and programming only may be used in class. Any student using a laptop for other purposes will forfeit the right to use a laptop in subsequent classes.
- Cell phones must be put on silent mode or left closed. Only in the case of university emergency when all cell phones go off may they be opened. If a student has a family emergency and will need to make a call during class, notification of this possibility must be communicated in advance to the instructor so special arrangements may be made.

Recommended Literature

- Deep Learning, I. Goodfellow, Y. Bengio and A. Courville.
<http://www.deeplearningbook.org/>
- Neural networks and deep learning, Michael Nielsen, an online book <http://neuralnetworksanddeeplearning.com/>
- Deep Learning Resources : For a vast list of Deep Learning resources including links to online courses, books, lectures, datasets, and tutorials, visit
<https://github.com/ChristosChristofidis/awesome-deep-learning>

Weekly Schedule

1. Course introduction. Course logistics, overview, history of neural networks, and what is deep learning.
2. Introduction to Neural Networks (**Shallow models**). Feedforward neural networks, Backpropagation.
Demo: `Lect2_CNN_MNIST_2020.ipynb`, `tensorflow4prediction.ipynb`
3. What are Convolutional Neural Networks (CNN, **Deep models**)? Convolution, Non-Linearity (ReLU), Pooling and Classification.
Demo: Jupyter + Python + sklearn
4. Deep Learning with **MATLAB/Python/Jupyter Notebook**. Deep learning software, CNN architectures.
Demo: Train a two-layer network on random data with TensorFlow, Keras and PyTorch
5. Code time for deep learning. Demo (Google Colab/ GCP)
 - TensorFlow Intro step-by-step (`Tensorflow_Intro.ipynb`);
 - Keras implementation for MNIST with TensorBoard (`keras_mnist_callback.ipynb`)
 - Google Colab demonstration for image classification using **transfer learning** (`demo-image-classification-fastai.zip`)
6. Development of deep learning applications in the Cloud (Google Colab) / Deep neural networks in computational neuroscience
7. Deep learning for computational biology: protein structure, gene expression analysis and pharmacogenomics.
8. Deep learning in medical imaging and electronic medical records /clinical data.
9. Advanced topics: autoencoders and generative models (generative adversarial networks (GAN) and diffusion models
10. Large language models (GPT-3/4, ChatGPT) in digital health and predictive medicine
11. Student spotlight talks, conclusions.

Resources

[Python Tutorial](#)

[Jupyter notebook Tutorial](#)

[**Jupyter** is not a direct acronym for these languages (**Julia (Ju)**, **Python (Py)** and **R**), it does establish a firm alignment with them]

Deep Learning Resources : For a vast list of Deep Learning resources including links to online courses, books, lectures, datasets, and tutorials, visit <https://github.com/ChristosChristofidis/awesome-deep-learning>

Preparations

Install prerequisites

Instructions for Mac or Linux. You need to install a recent version of Python (>3), plus the packages [keras](#), [numpy](#), [matplotlib](#) and [jupyter](#).

We'll go over the very high-level ideas in class. If you're new to computer programming, Python, or Jupyter notebooks, you'll want to run through some tutorials to help you get started.

Introduction to [Python](#) (beginner)

Optional: Introduction to [Git](#) (beginner) and [GitHub tutorial](#).

Video introduction to [Jupyter Notebooks](#) (with [code](#)) (beginner)

Learning to code with [Python and Jupyter Notebooks](#) (beginner)

<https://www.tensorflow.org/install/>

Useful Python packages:

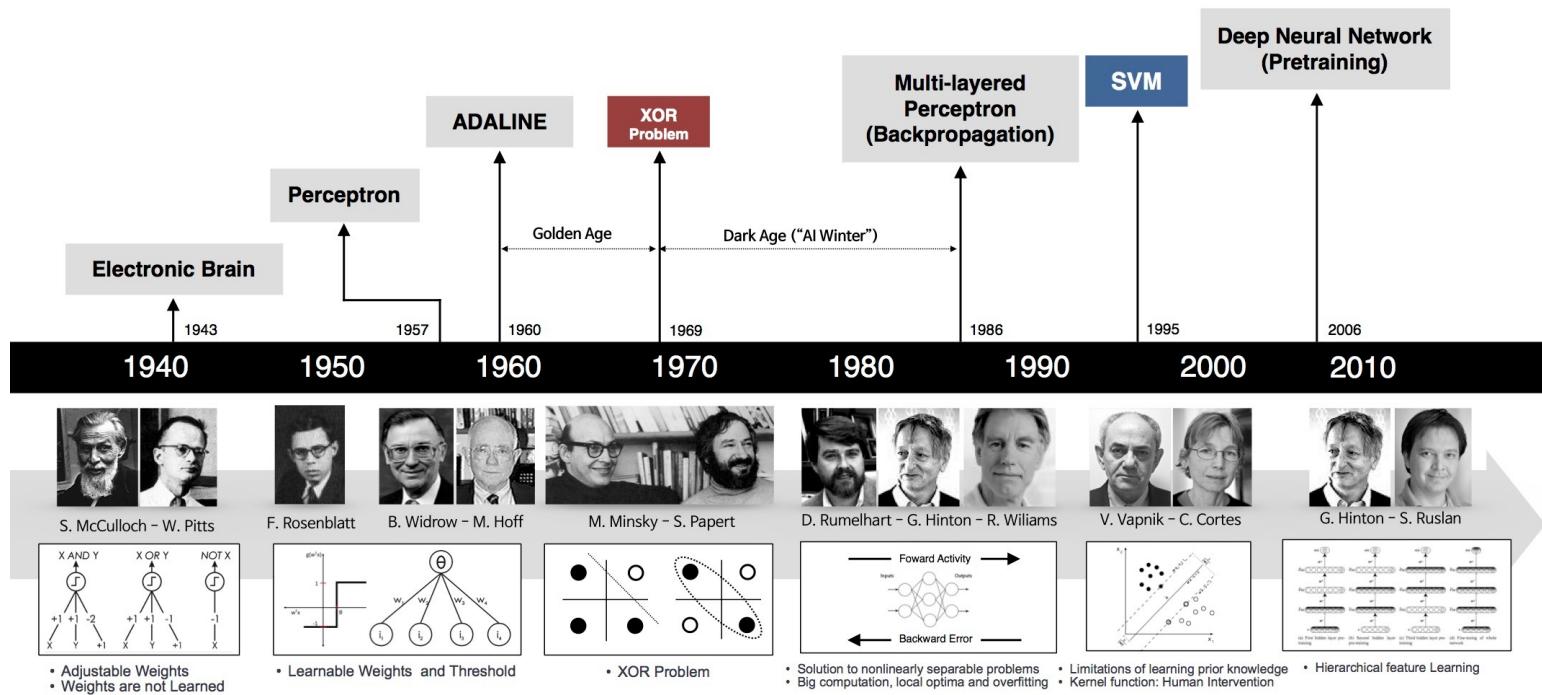
Introduction to [Pandas](#) (intermediate)

Practice your Pandas [skills](#) (intermediate)

Getting started with [Numpy](#) (intermediate)

Getting started with [Scipy](#) (intermediate)

A brief history of Neural Networks

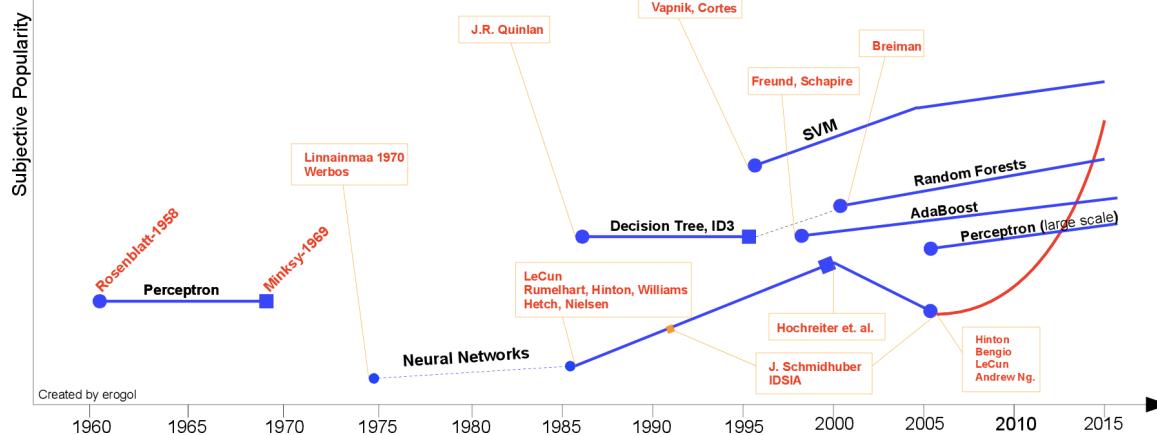


https://beamandrew.github.io/deeplearning/2017/02/23/deep_learning_101_part1.html

Turing Award 2018: Nobel Prize of Computing Given to the Godfathers of Deep Learning



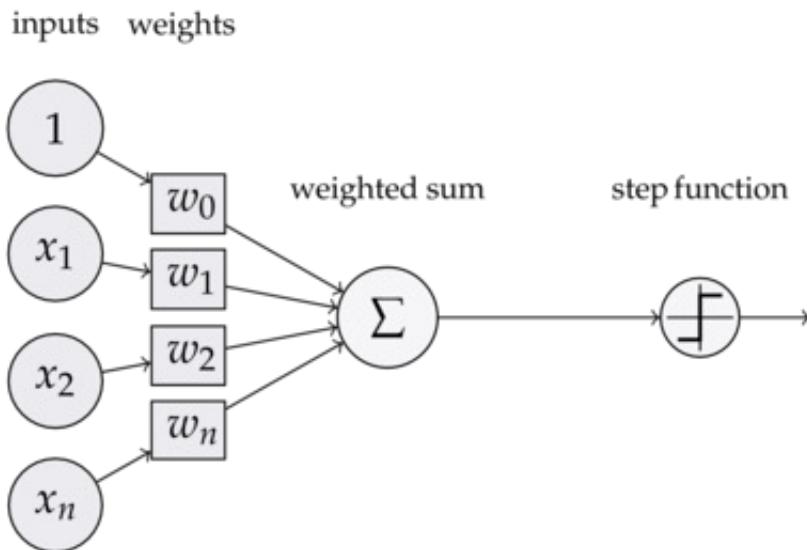
Popularity of Neural Networks



First appearance (roughly)



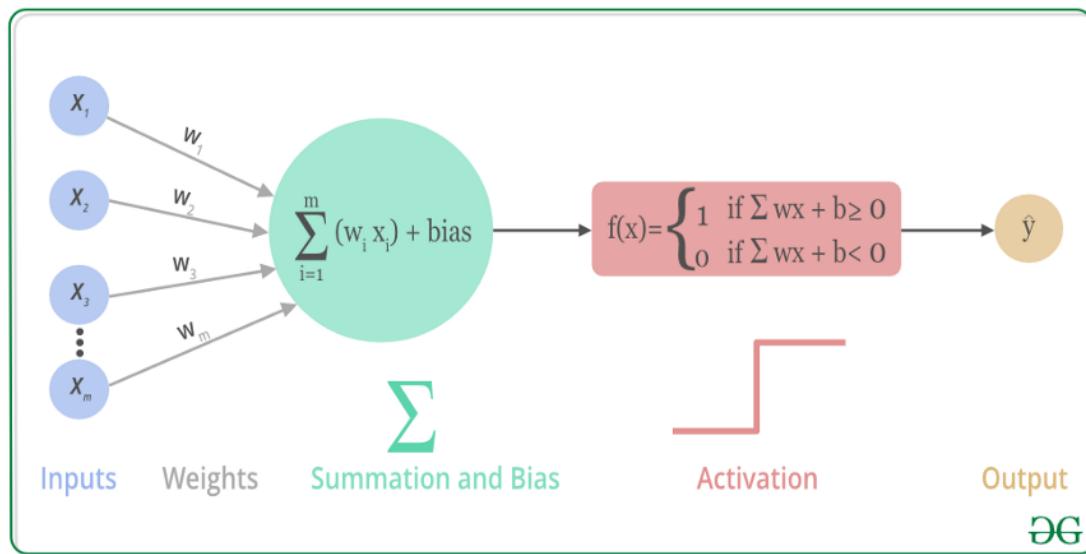
Rosenblatt's Perceptron for binary classification



Key idea:

- One weight w_i per input x_i
- Multiply weights w/ respective inputs and add bias $x_0=+/-1$
- If result large than threshold, return 1, otherwise 0

Rosenblatt's Perceptron for binary classification



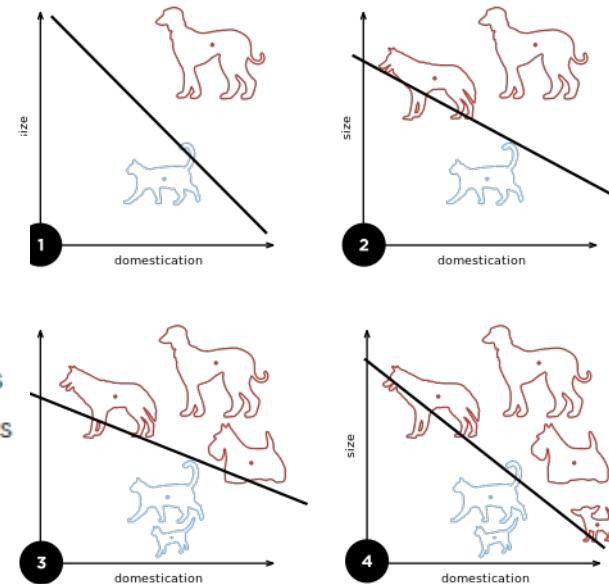
Key idea:

- One weight w_i per input x_i
- Multiply weights w/ respective inputs and add bias $x_0=+/-1$
- If result large than threshold, return 1, otherwise 0

Training a perceptron

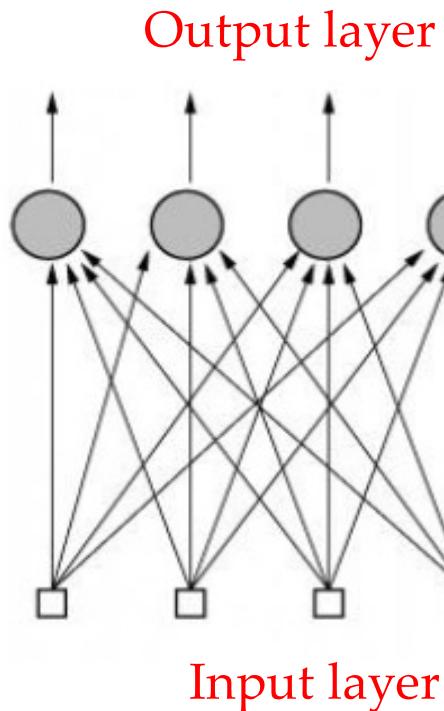
Main innovation was the learning algorithm for perceptron

- Learning algorithm
 - Initialize weights randomly
 - Take one sample x_i and predict y_i
 - For erroneous predictions update weights
 - If the output was $\hat{y}_i = 0$ and $y_i = 1$, increase weights
 - If the output was $\hat{y}_i = 1$ and $y_i = 0$, decrease weights
 - Repeat until no errors are made



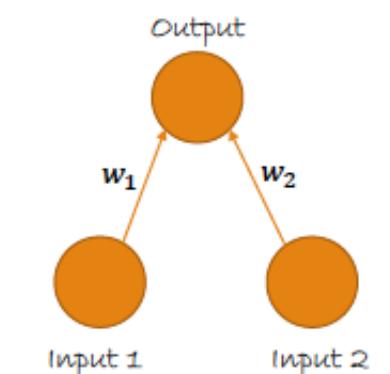
A perceptron updating its linear boundary as more training examples are added.

From a perceptron to neural network

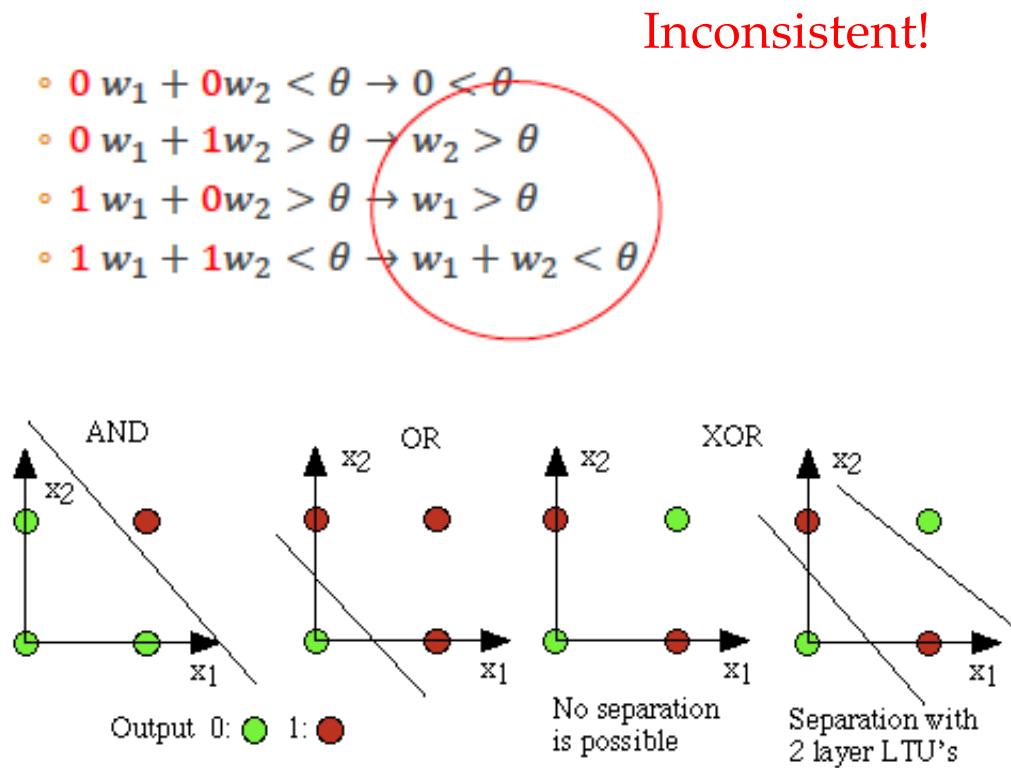


- 1 perceptron = 1 decision
- What about multiple decisions?
Handwritten digit
 - Stack as many outputs as the possible outcomes into a layer
(neural network)
- Use one layer as input to next layer
- **Multi-layer perceptron (MLP)**

Why can't the XOR-problem be solved by a one-layer perceptron?

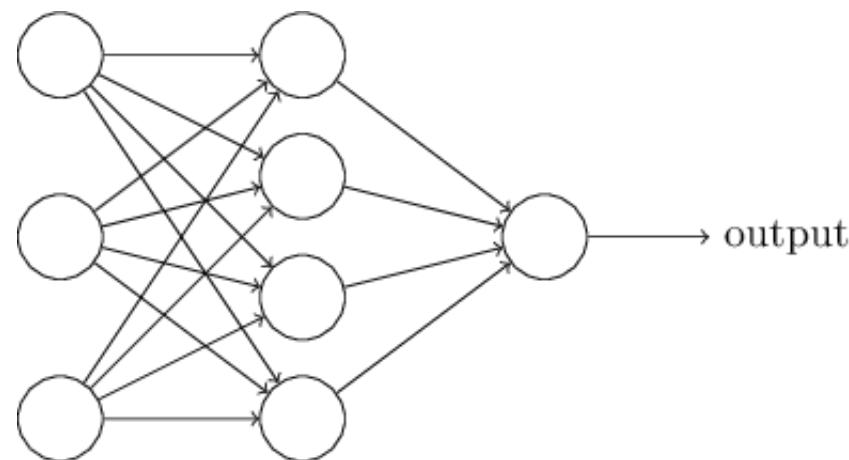


Input 1	Input 2	Output
1	1	0
1	0	1
0	1	1
0	0	0



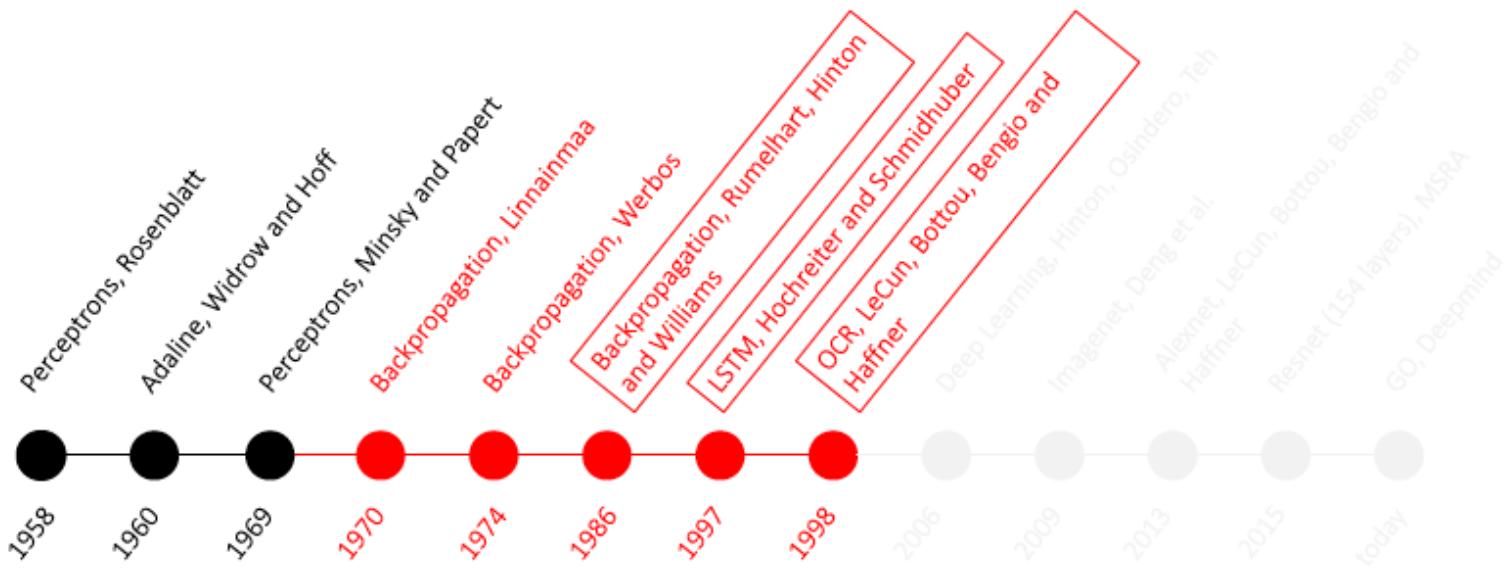
Minsky, M, and Papert S. "Perceptrons." (1969).

How to train multi-layer perceptron?



Rosenblatt's algorithm not applicable as it expects to know the desired target, which we do not know for the hidden layers

The “AI winter” despite notable successes



The first AI winter (1969)



If perceptron can't even solve simple XOR, why bother?
No results, no \$\$\$, no AI!

Backpropagation (BP) emerges (1986)

- Learn multi-layer perceptron with BP now possible: XOR and more complicated functions can be solved.

Universal approximation theorem: ability to learn any function

Rumelhart, Hinton and Williams,

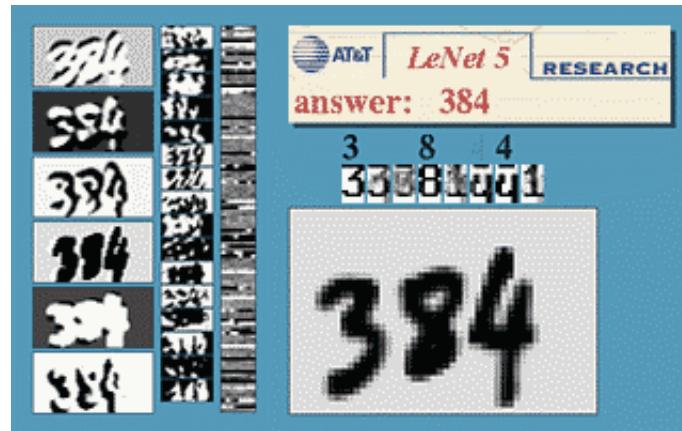
Learning representations by back-propagating errors, Nature, 1986

- **Backprop** works by taking the derivative of the network's loss function and back-propagating the errors to update the parameters in the lower layers.
- Digit recognition in cheques (OCR) solved before 2000

Who should be credited: Geoff Hinton /Yann Lecun/Yoshua Bengio
and Jürgen Schmidhuber

Backpropagation: OCR application

Digit recognition in cheques (OCR) solved w/
convolutional neural nets (CNN, convnets) by Yann Lecun



[Yann's convnet reading a check:](#)

Recurrent networks

- Traditional networks are “too plain”
Static Input → Processing → Static Output
- What about dynamic input
Temporal data, Language, Sequences
- Memory is needed to “remember” state changes
Recurrent feedback connections
- What kind of memory
 - Long, Short?
 - Both! Long-short term memory networks (LSTM),
Schmidhuber 1997
 - Attention 2017

The second AI winter

- Until 1998, key algorithm developments:
 - Backpropagation
 - Long-short term memory networks (LSTM)
 - OCR with convolutional neural nets
- At the same time, kernel machines (Support Vector Machine, SVM) emerge and become popular.

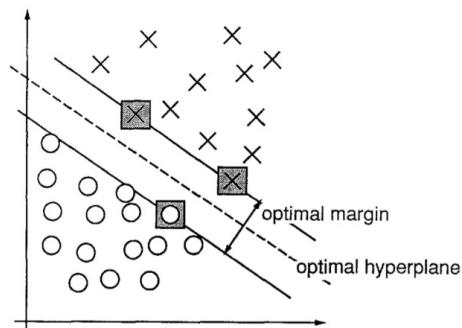
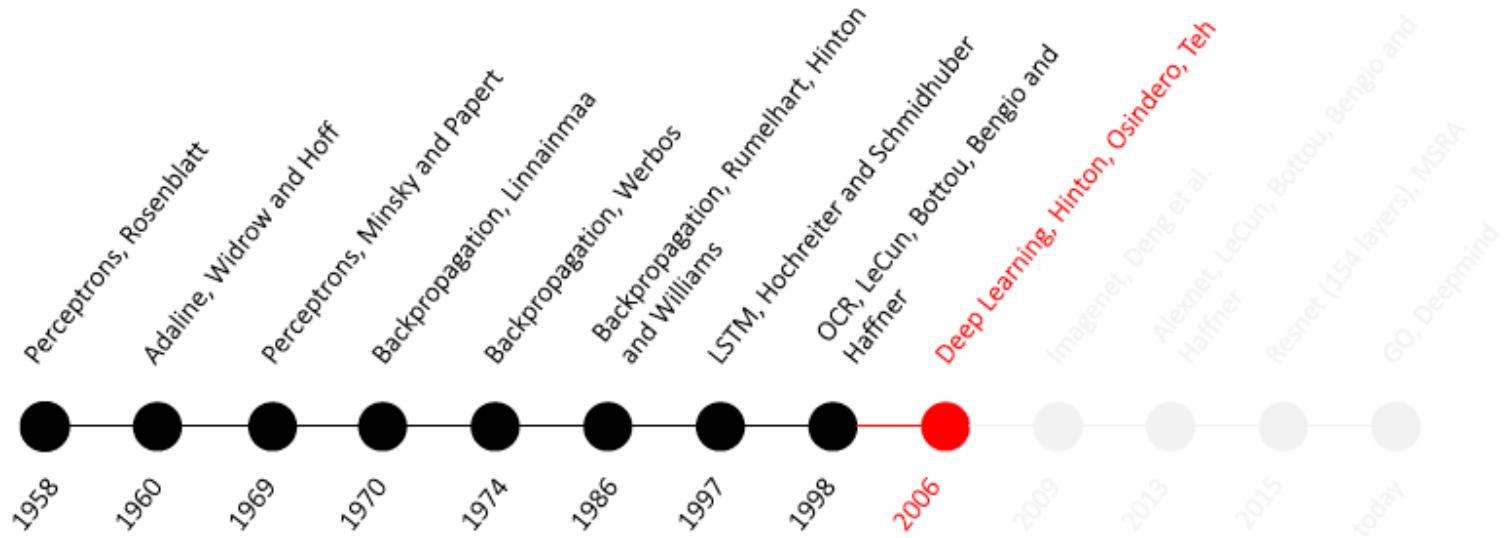


Figure 2. An example of a separable problem in a 2 dimensional space. The support vectors, marked with grey squares, define the margin of largest separation between the two classes.

As a result, AI community turns away from neural networks

The thaw of the “AI winter” - rebranding of ‘neural networks’ to **deep learning**



Problems with NNs

What was actually wrong with
backpropagation in 1986?

(Slide by Geoff Hinton)

- We all drew the wrong conclusions about why it failed.
The real reasons were:

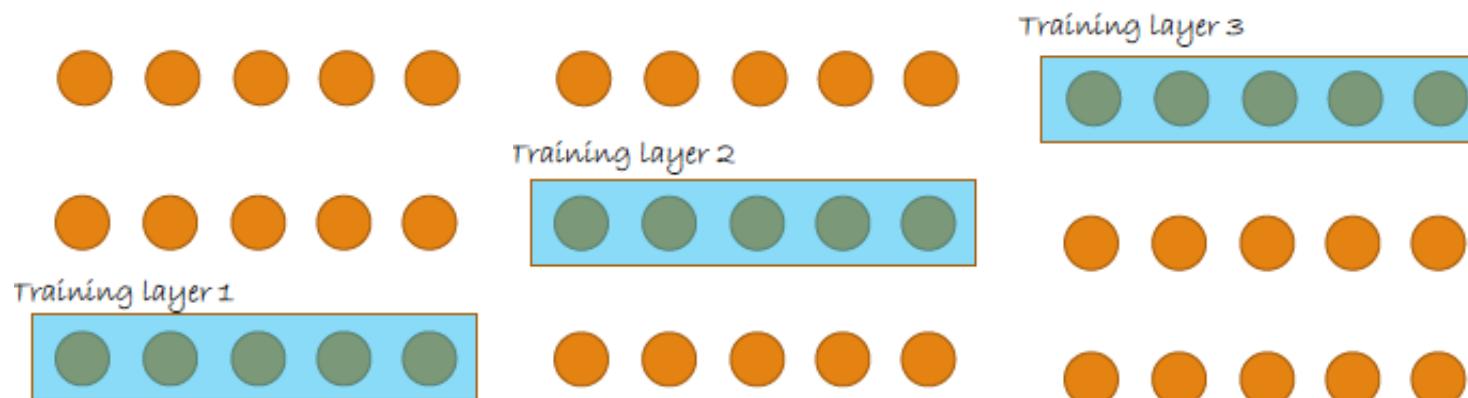
1. Our labeled datasets were thousands of times too small.
2. Our computers were millions of times too slow.
3. We initialized the weights in a stupid way.
4. We used the wrong type of non-linearity.

- **Overfitting**: models do not generalize well
- **Vanishing (exploding) gradient problem**: when learning w/ NN, you need to multiply some numbers. If all are 0.1, the result is 0.000000001 for n=10, too small for any learning
- **Selection of hyper-parameters**, e.g. how many hidden layers?

Deep learning moves beyond shallow NNs

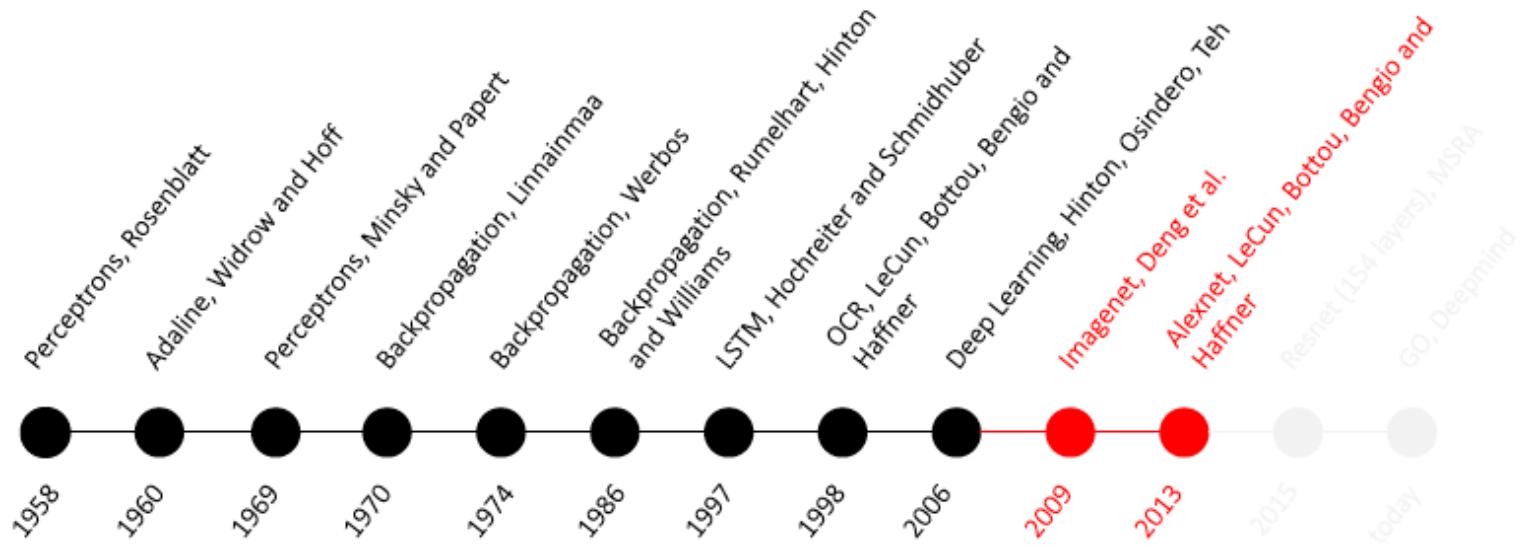
Layer-by-layer training

- It's easy to train each layer individually
- It became easier to train multi-layer NNs
- Parameters are initialized per-layer further training using contrastive divergence



[Hinton, Osindero & The \(2006\). A Fast Learning Algorithm for Deep Belief Nets. Neural Computation.](#) 18 (7): 1527–1554

Deep Learning Renaissance



Breakthrough 2012: ImageNet

- More data:
 - Imagenet database published (Deng 2009).
 - 16 million images annotated by humans
- ImageNet Large Scale Visual Recognition Competition (ILSVRC)
 - 1 million images
 - 1,000 classes
 - Top-5 and Top-1 error measured

[ImageNet](#)

Deng et al. ImageNet: A Large-Scale Hierarchical Image Database, CVPR 2009

Krizhevsky, A., Sutskever, I., & Hinton, G. E. ImageNet Classification with Deep Convolutional Neural Networks, 2012 - [Alexnet](#)

Alexnet

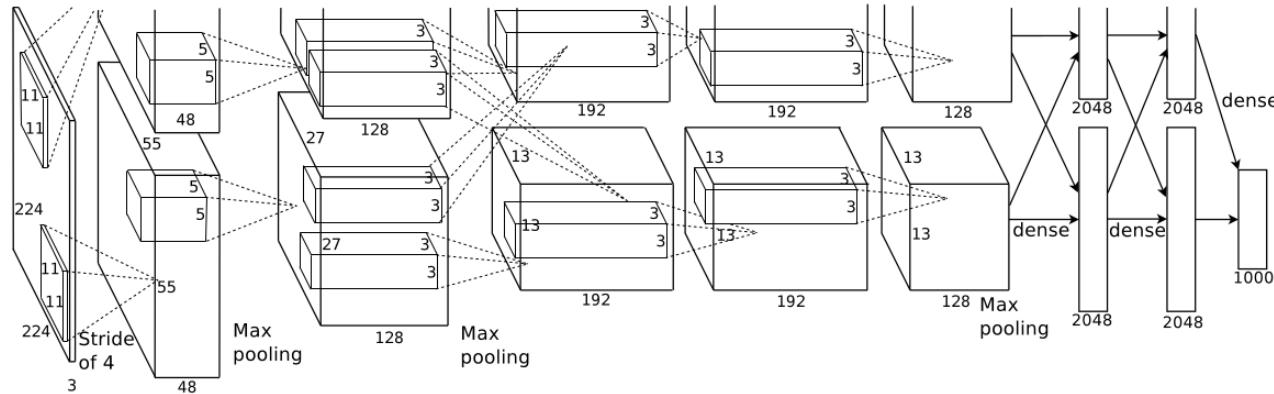
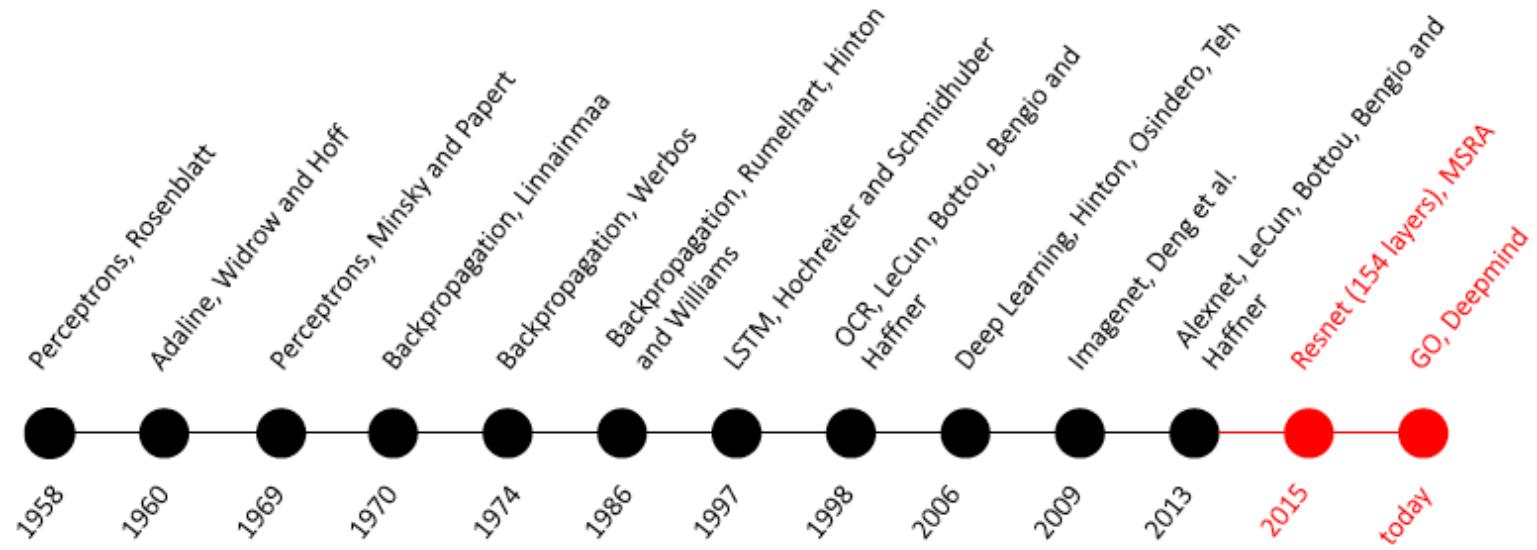


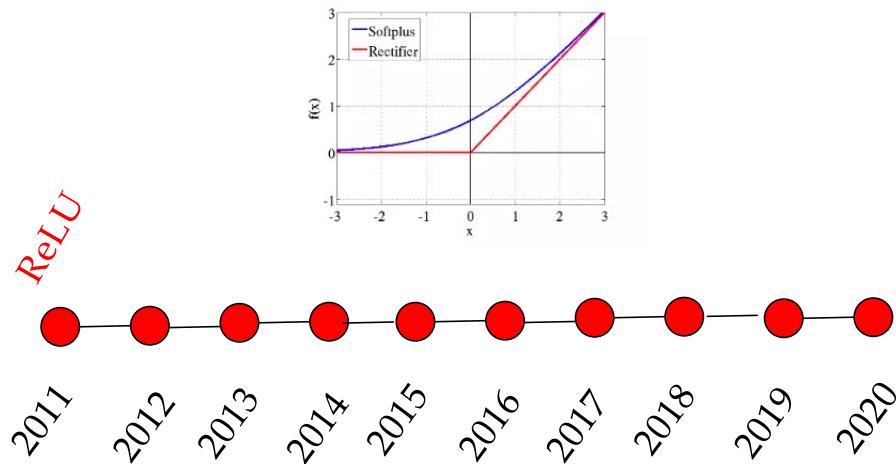
Figure 2: An illustration of the architecture of our CNN, explicitly showing the delineation of responsibilities between the two GPUs. One GPU runs the layer-parts at the top of the figure while the other runs the layer-parts at the bottom. The GPUs communicate only at certain layers. The network's input is 150,528-dimensional, and the number of neurons in the network's remaining layers is given by 253,440–186,624–64,896–64,896–43,264–4096–4096–1000.

- [Krizhevsky2013] re-implemented a convolutional neural network [LeCun1998] trained on Imagenet with two GPUs
- Further theoretical improvement: **Rectified Linear Units (ReLU)** (not sigmoid or tanh); **Dropout**; **Data augmentation**
- 2013 Imagenet Workshop a legendary turmoil (impressive 16% top-5 error)

Deep learning golden era



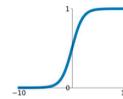
Milestone in DL: A 10-year Retrospective



Activation Functions

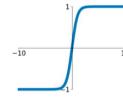
Sigmoid

$$\sigma(x) = \frac{1}{1+e^{-x}}$$



tanh

$$\tanh(x)$$



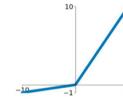
ReLU

$$\max(0, x)$$



Leaky ReLU

$$\max(0.1x, x)$$

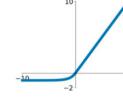


Maxout

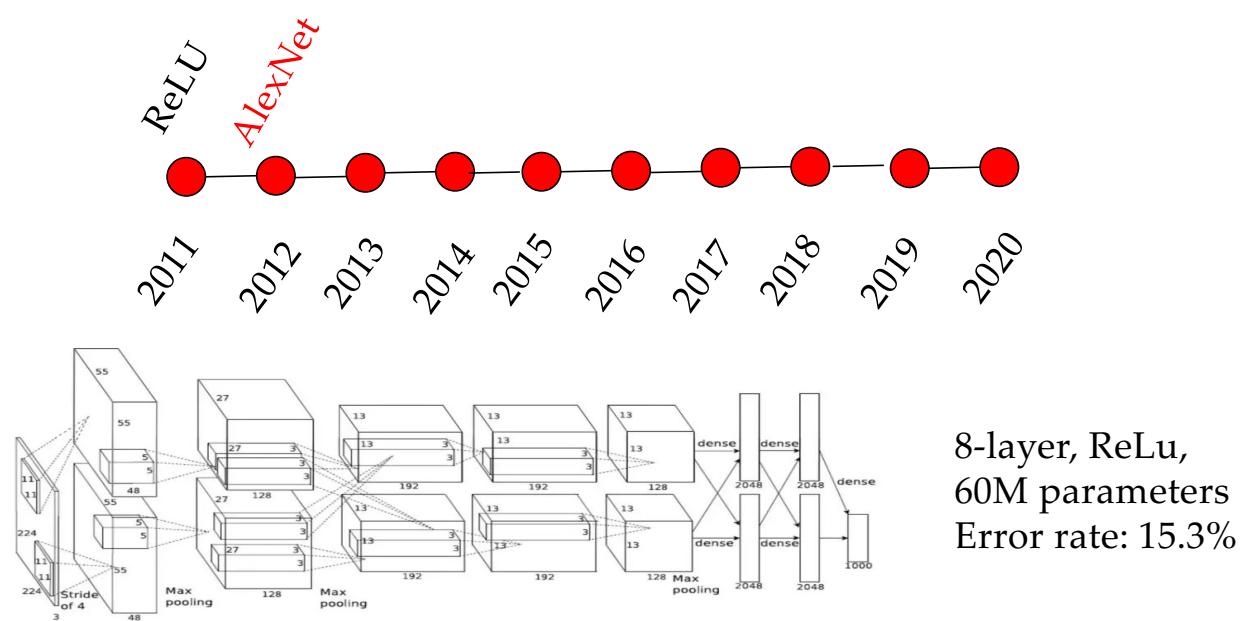
$$\max(w_1^T x + b_1, w_2^T x + b_2)$$

ELU

$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$



Milestone in Deep Learning: A 10-year Retrospective



Krizhevsky et al. ImageNet Classification with Deep Convolutional Neural Networks, NIPS 2012

Milestone in Deep Learning: A 10-year Retrospective

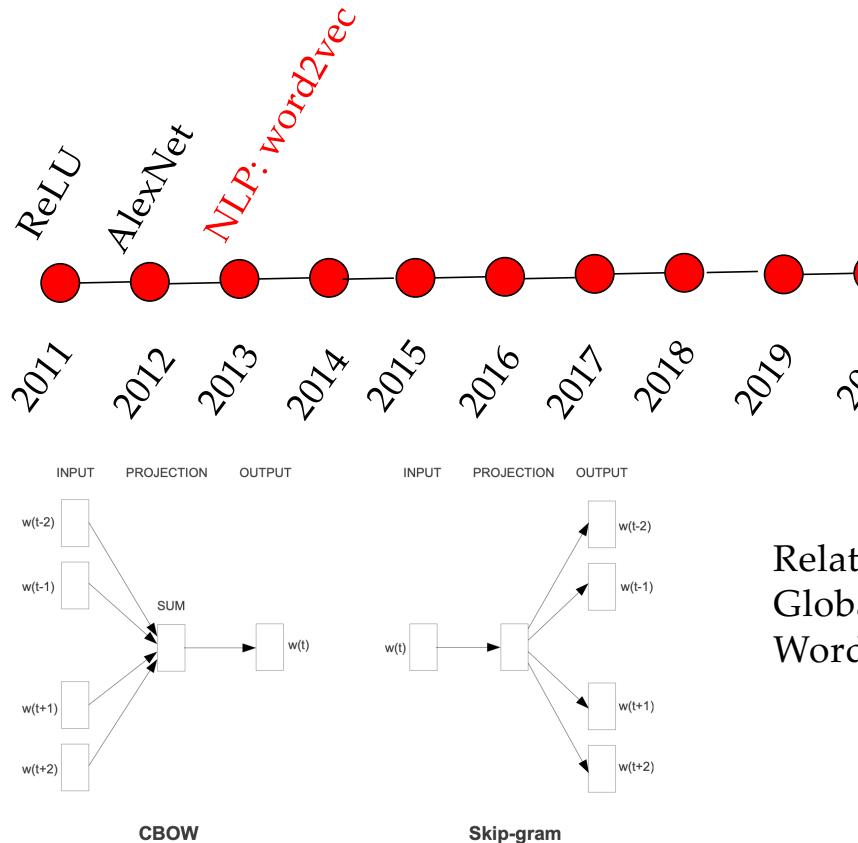
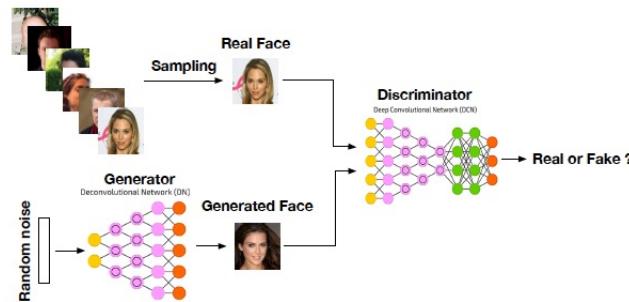
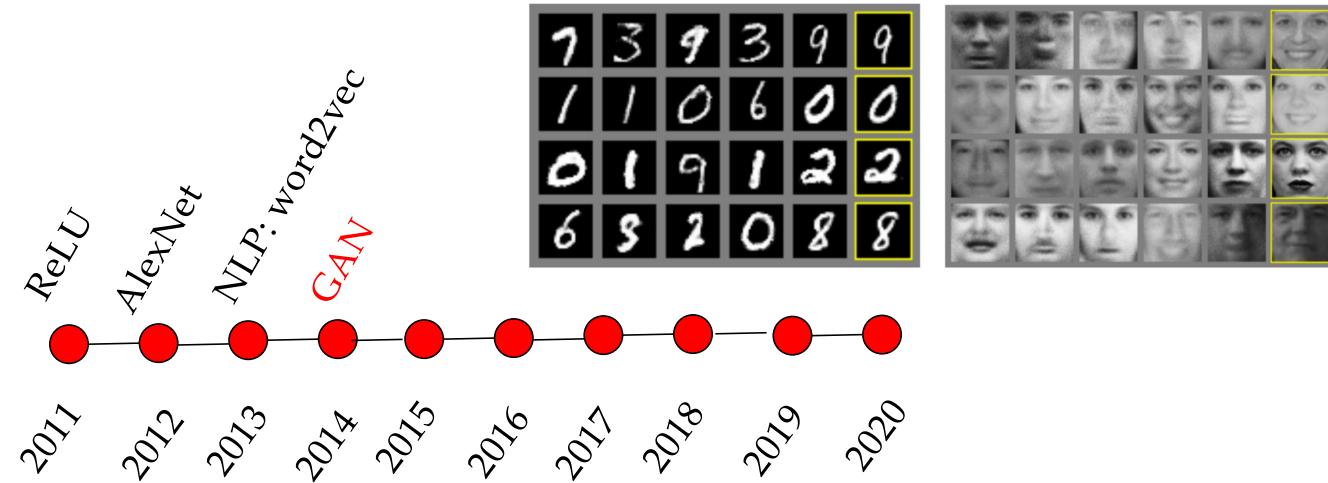


Figure 1: New model architectures. The CBOW architecture predicts the current word based on the context, and the Skip-gram predicts surrounding words given the current word.

Mikolov et al. Distributed Representations of Words and Phrases and their Compositionality, 2013

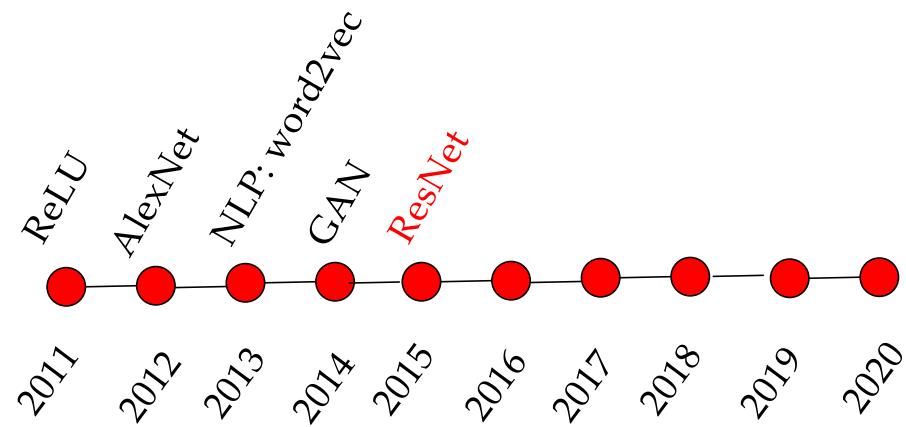
Milestone in Deep Learning: A 10-year Retrospective



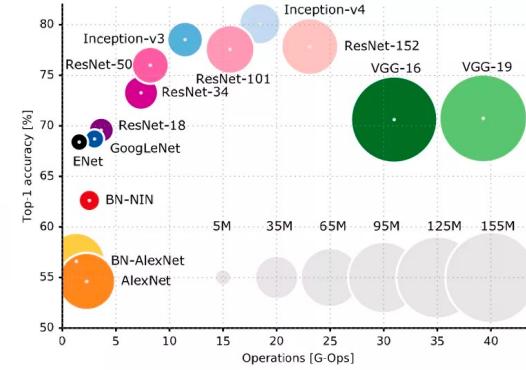
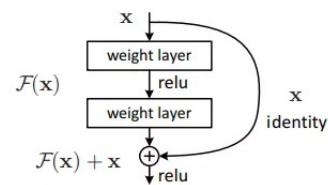
Related: Wasserstein
GAN, StyleGAN;
[Adam, Attention]

Goodfellow et al. Generative Adversarial Network

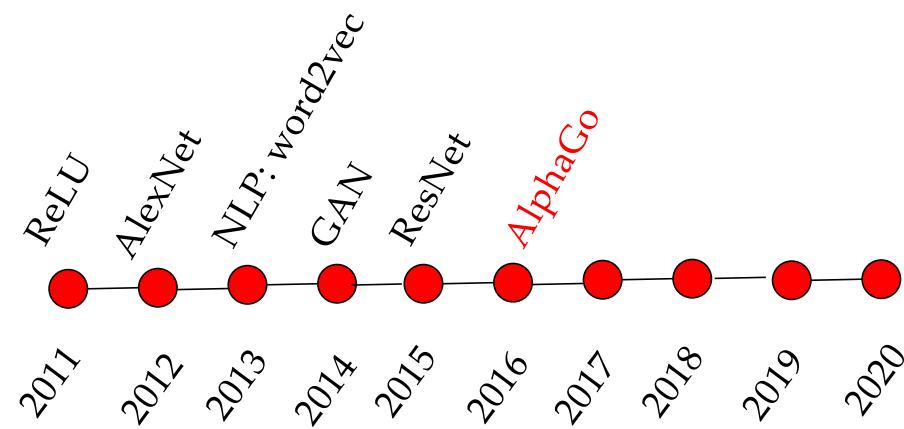
Milestone in Deep Learning: A 10-year Retrospective



Bypass human first time!



Milestone in Deep Learning: A 10-year Retrospective



Silver et al. Mastering the game of Go with deep neural networks and tree search, Nature, 2016

Milestone in Deep Learning: A 10-year Retrospective

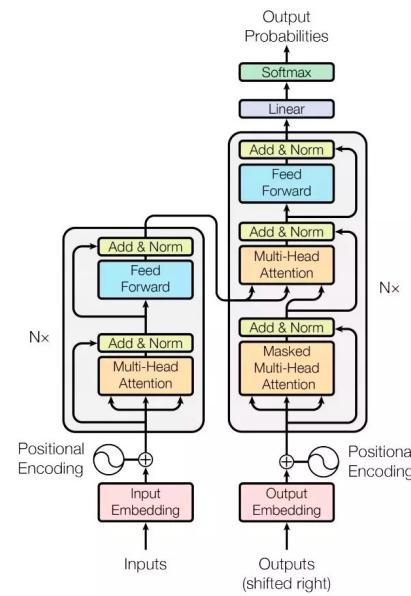
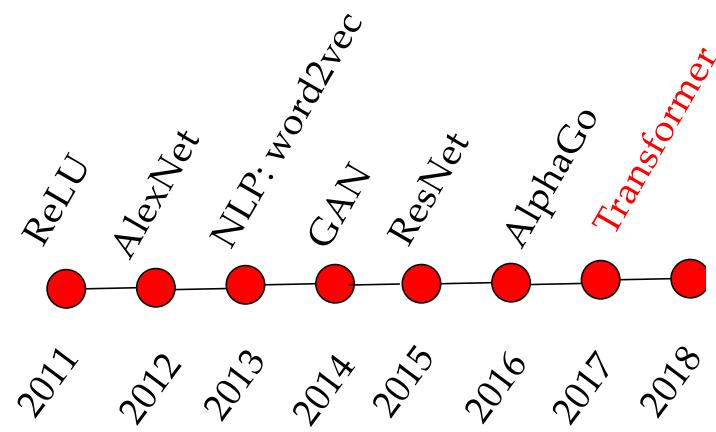
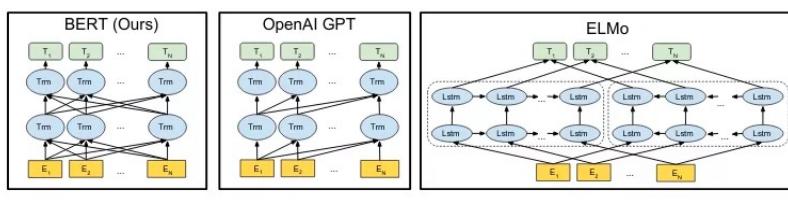
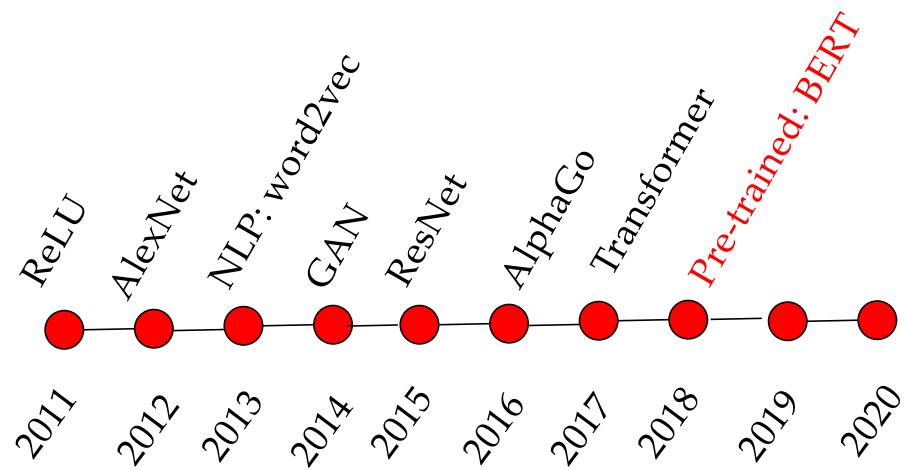


Figure 1: The Transformer - model architecture.

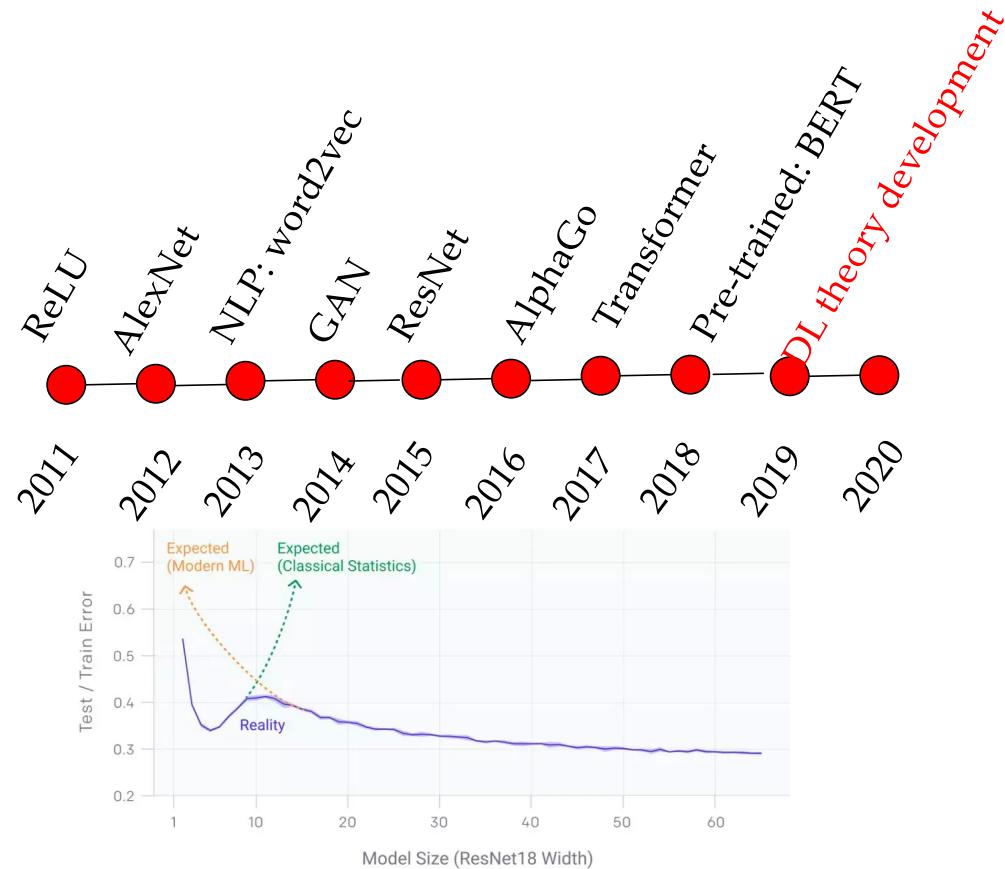
Vaswani et al. Attention Is All You Need, 2017

Milestone in Deep Learning: A 10-year Retrospective



GPT-2,
Transformer-XL
XLNet

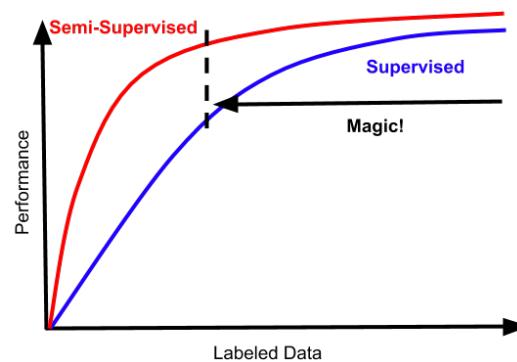
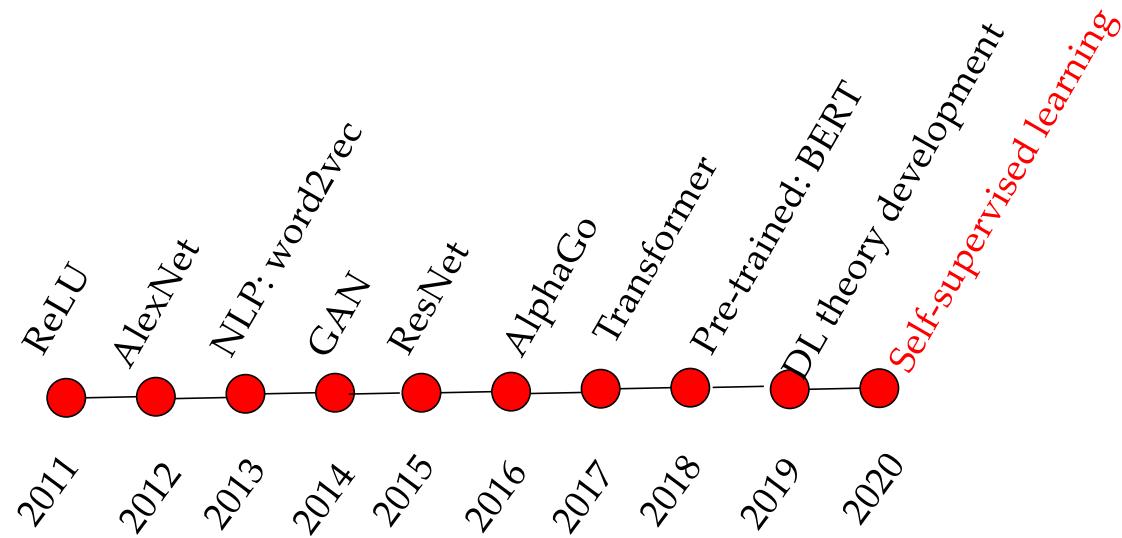
Milestone in Deep Learning: A 10-year Retrospective



Nakkiran et al, Deep Double Descent: Where Bigger Models and More Data Hurt, 2019

Frankle & Carbin, The Lottery Ticket Hypothesis: Finding Sparse, Trainable Neural Networks, 2019

Milestone in Deep Learning: A 10-year Retrospective



The Paradigm Shift of Self-Supervised Learning

Every day brings new headlines for how deep learning is changing the world around us

- Google DeepMind's **AlphaGo**
- Deep Learning Makes **Driverless Cars** Better at Spotting Pedestrians
- Deep learning algorithm **diagnoses skin cancer** as well as seasoned
- **Amazon Go:** How Deep Learning and AI will change Retailing

Deep learning is almost everywhere!

Why Deep Learning?

Limitations of traditional machine learning

- not good at handling high dimensional data.
- difficult to do feature extraction and object recognition.

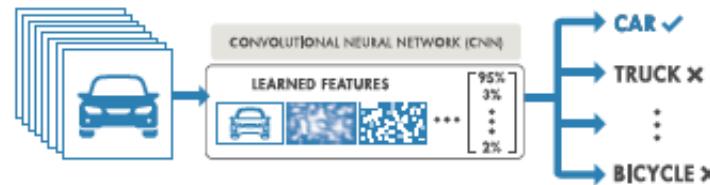
Advantages of deep learning

- DL is computationally expensive, but it is capable of handling high dimensional data.
- feature extraction is done automatically.

Why now?

- **Easy access to massive labeled data sets**

Data sets such as ImageNet and PASCAL VoC are freely available, and are useful for training on many different types of objects.



- **Increased computing power**

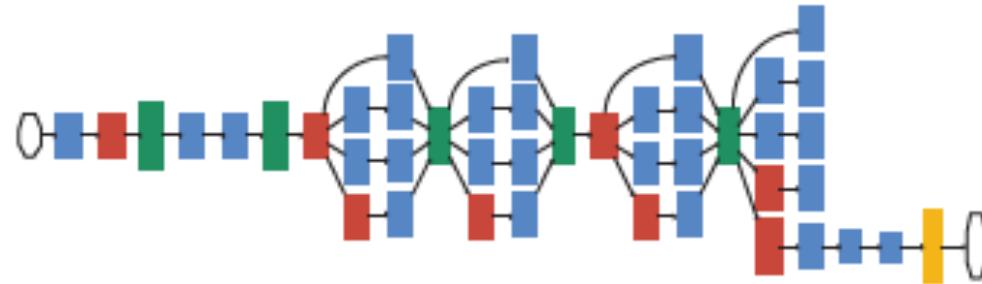
High-performance GPUs accelerate the training of the massive amounts of data needed for deep learning, reducing training time from weeks to hours.



Why now?

- **Pretrained models built by experts**

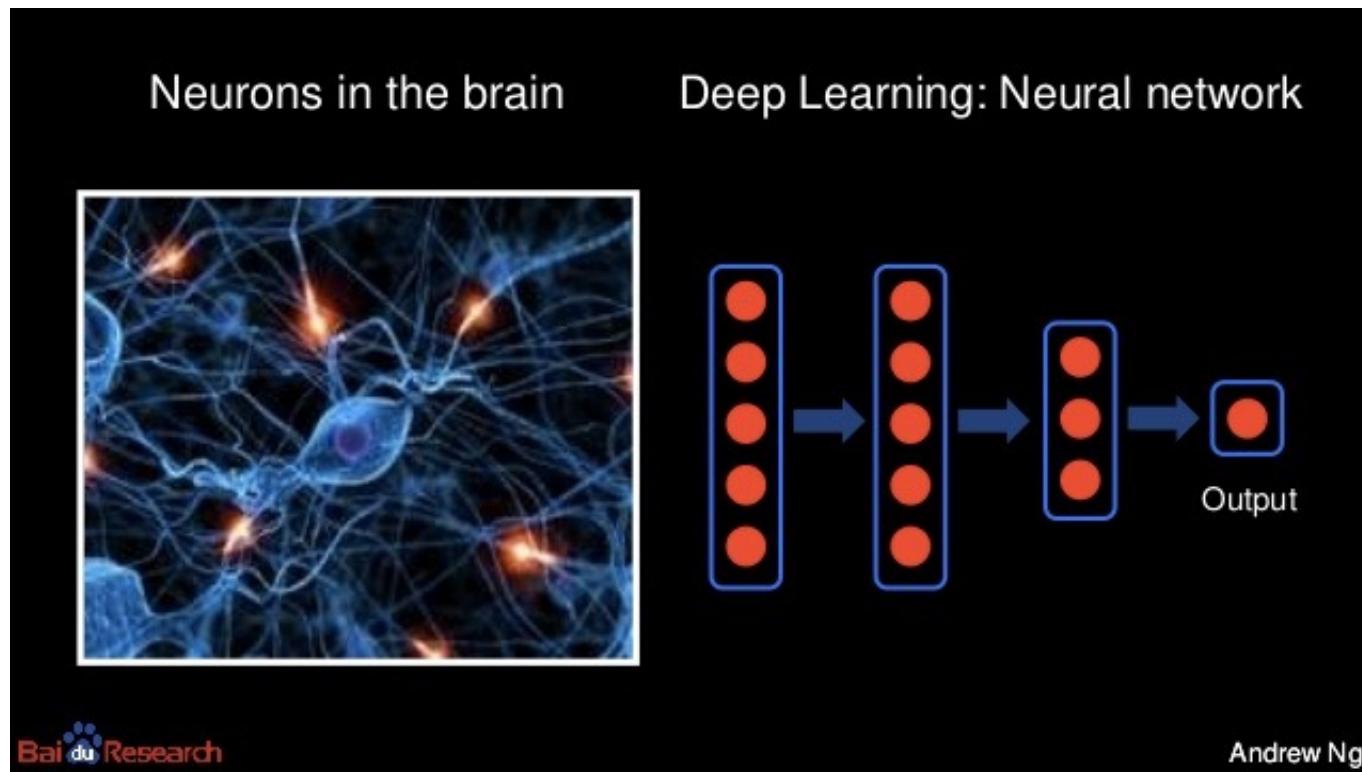
Models such as AlexNet can be retrained to perform new recognition tasks using a technique called transfer learning. While AlexNet was trained on 1.3 million high-resolution images to recognize 1000 different objects, accurate transfer learning can be achieved with much smaller datasets.



Why now?

- More data
- Bigger models with better algorithms
 - Better regularization method, such as dropout
 - Better optimization methods, such as batch normalization
- More computation, such as hardware

What is deep learning?

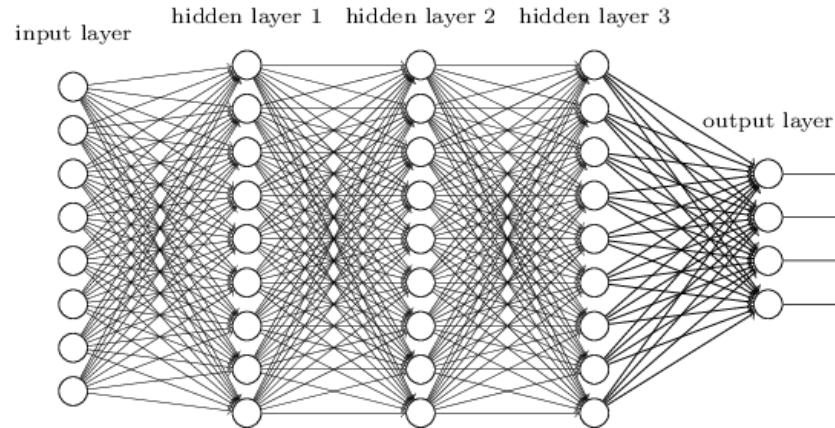


What is deep learning?

Deep learning is a class of machine learning algorithms that:

- use a cascade of multiple layers of nonlinear processing units for feature extraction and transformation. Each successive layer uses the output from the previous layer as input.
- learn in supervised (e.g., classification) and/or unsupervised (e.g., pattern analysis) manners.
- learn multiple levels of representations that correspond to different levels of abstraction; the levels form a hierarchy of concepts.

Deep Learning is Large Neural Networks



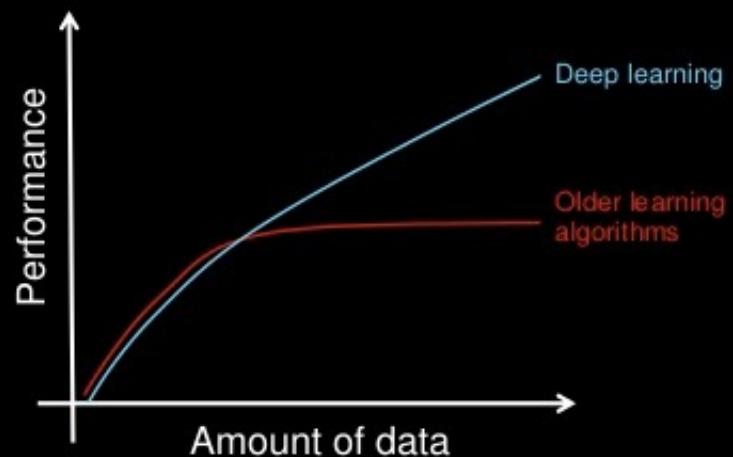
Deep learning is a type of machine learning in which a model learns to perform classification tasks directly from images, text, or sound. Deep learning is usually implemented using a neural network architecture. The term “deep” refers to the number of layers in the network—the more layers, the deeper the network. Traditional neural networks contain only 2 or 3 layers, while deep networks can have hundreds.

Long story short

- A family of parametric, non-linear and hierarchical representation learning functions, which are massively optimized with stochastic gradient descent to encode domain knowledge, i.e. domain invariances, stationarity.
- $a_L(x; \theta_{1,\dots,L}) = h_L(h_{L-1}(\dots h_1(x, \theta_1), \theta_{L-1}), \theta_L)$
 - x : input, θ_l : parameters for layer l , $a_l = h_l(x, \theta_l)$: (non-)linear function
- Given training corpus $\{X, Y\}$ find optimal parameters

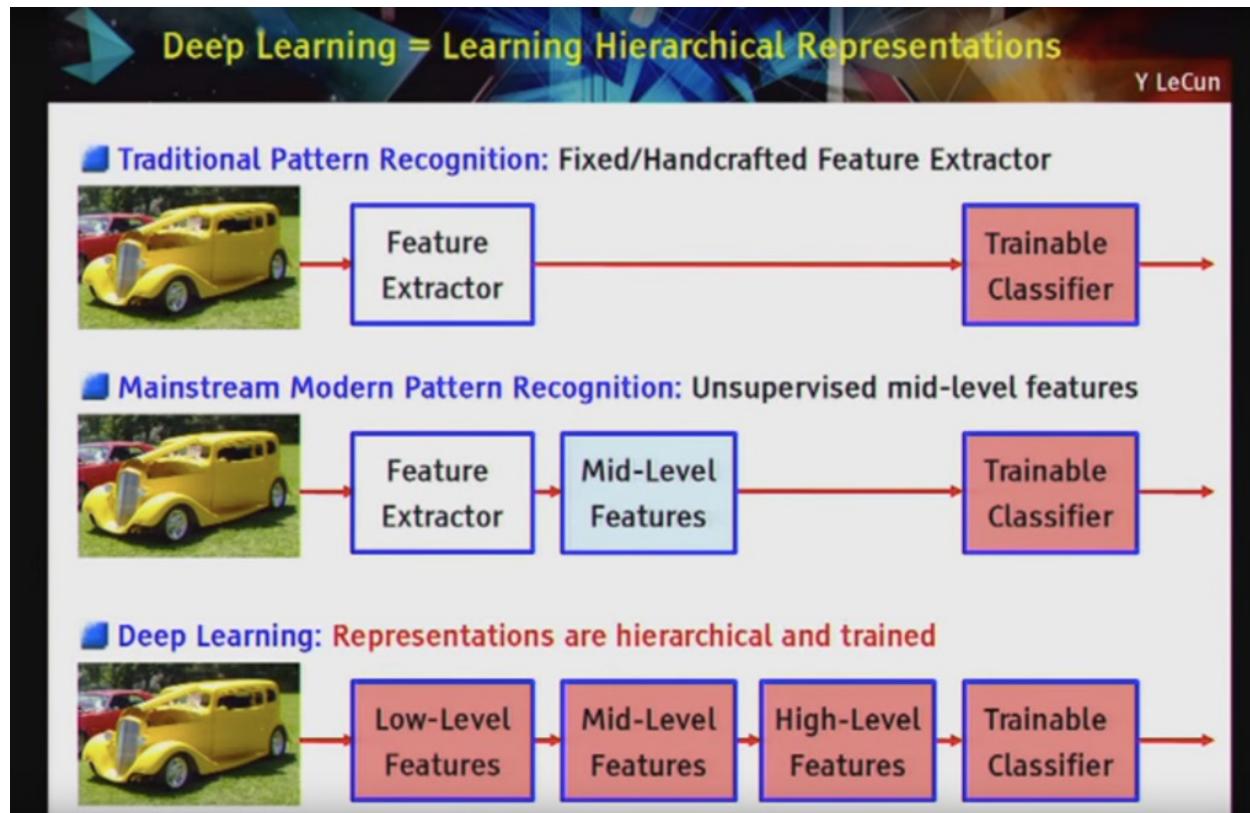
$$\theta^* \leftarrow \arg \min_{\theta} \sum_{(x,y) \in (X,Y)} \ell(y, a_L(x; \theta_{1,\dots,L}))$$

Why deep learning



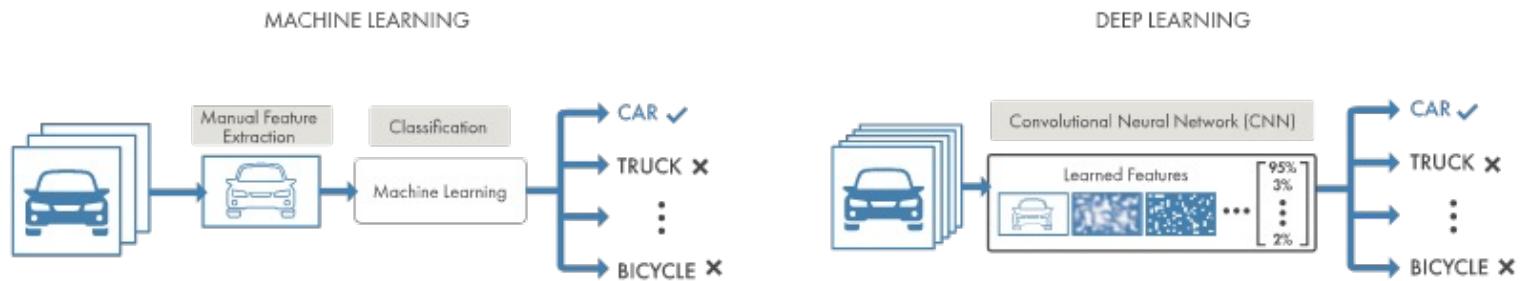
How do data science techniques scale with amount of data?

Learning Representations & Features



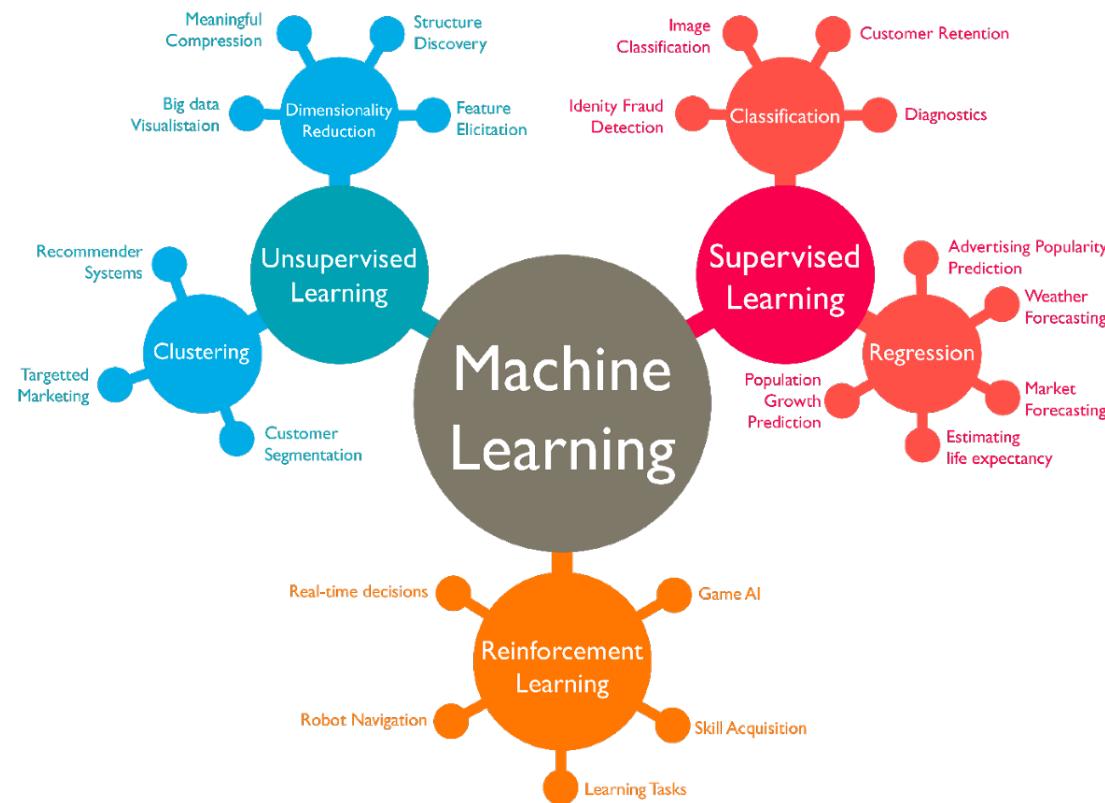
Deep learning is end-to-end learning;
features are also learned from data

Machine Learning vs Deep Learning



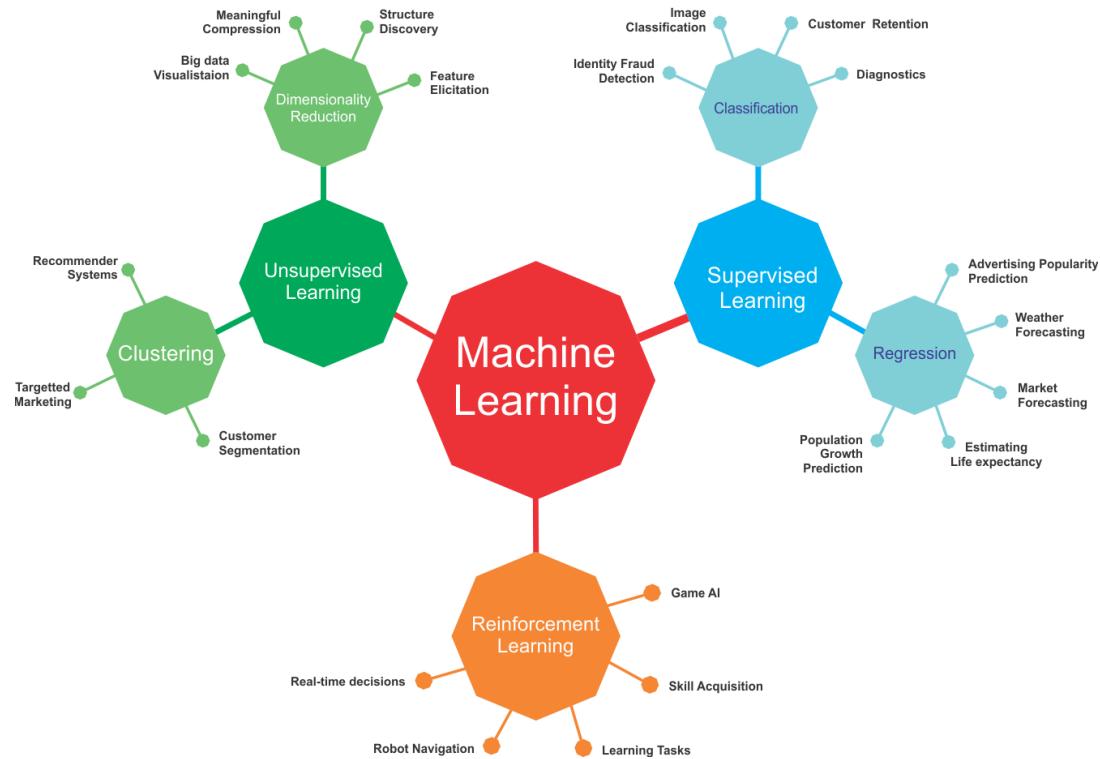
Machine Learning	Deep Learning
+ Good results with small data sets	- Requires very large data sets
+ Quick to train a model	- Computationally intensive
- Need to try different features and classifiers to achieve best results	+ Learns features and classifiers automatically
- Accuracy plateaus	+ Accuracy is unlimited

Types of learning



<https://wordstream-files-prod.s3.amazonaws.com/s3fs-public/machine-learning.png>

Types of learning



<https://wordstream-files-prod.s3.amazonaws.com/s3fs-public/machine-learning.png>

Types of learning

- **Supervised learning:**
trained with labeled data; including regression and classification problems. Convolutional neural networks (CNN)
- **Unsupervised learning**
trained with unlabeled data; clustering and association rule learning problems. Autoencoders, layer-by-layer training
- **Reinforcement learning**
 - no training data; stochastic Markov decision process; robotics and self-driving cars.
 - Learn from noisy, delayed rewards from your environment

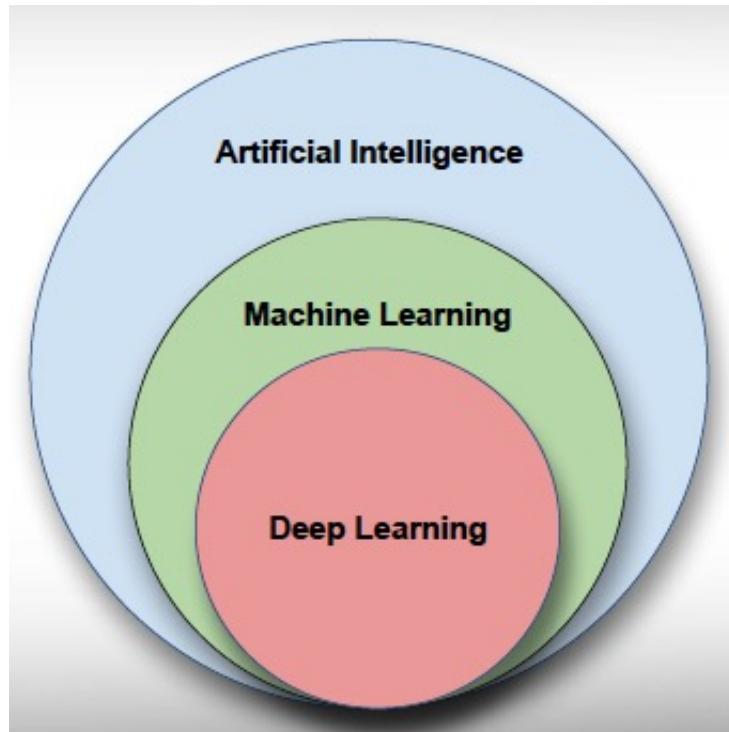
Deep architectures

- Feedforward
 - Convolutional NN
- Feedback
 - Deconvolutional networks
- Bi-directional
 - Deep Boltzmann Machines,
stacked autoencoders
- Sequence based
 - RNNs, LSTMs

Most Popular Neural Networks

- Multilayer Perceptron Networks (MLP)
- Convolutional Neural Networks (CNN)
- Recurrent Neural Networks: Long Short-Term Memory (LSTM) ; Gated recurrent unit (GRU)
- Generative Adversarial Network (GAN)
- Autoencoder: VAE

Relationship of AI, ML, and DL



- Artificial Intelligence (AI) is anything about man-made intelligence exhibited by machines.
- Machine Learning (ML) is an approach to achieve AI.
- Deep Learning (DL) is one technique to implement ML.

Reading

Chapter 4 from Machine Learning

Install Python and Jupyter etc if you have not done so