Introduction

M. Soleymani

Deep Learning

Sharif University of Technology

Spring 2024

Course Info

- Course Number: 40-719 (Time: Sun-Tue 9:00-10:30 Location: CE 201)
- Instructor: Mahdieh Soleymani (soleymani@sharif.edu)
- Head TAs: **Mohammad Mozafari** (mozafari.mmd@gmail.com) & **Fahimeh Hosseini** (fahim.hosseini.77@gmail.com)
- Head of Assignments: Ali Abdollahi (aliabdollahi024a@gmail.com)
- Office hours: set appointment through email

Communication

- Quera
  - Policies and rules
  - Tentative schedule
  - Slides and notes
  - Assignments
  - Discussions
    - Ask questions about homework, grading, logistics
    - communication with course staff
- Email
  - Private questions

Materials

- (Optional) text books
  - **Deep Learning: Foundations and Concepts**, by C. Bishop and H. Bishop, 2023.
  - **Deep Learning**, by Ian Goodfellow, Yoshua Bengio, and Aaron Courville, 2016. (free version)
  - **Understanding Deep Learning**, by Simon J.D. Prince, 2023. (free draft version)
  - **Dive into deep learning,** by Zhang et al., 2023 (free version)

- Some papers

- Notes, lectures, and demos

Marking Scheme

• Midterm Exam:	20%
• Final Exam:	25%
• Homework assignments:	55%

About homework assignments

- 5 assignments (both theoretical and practical)

- HWs are implementation-heavy
  - A lot of coding and experimenting
  - In some assignments, you deal with large datasets

- Language of choice: Python

- Toolkit of choice: Pytorch

Assignments: Late policy

- Everyone gets up to 5 slack days for theoretical problems and 10 total slack days for practical ones
- You can distribute them across your assignments but use up to 3 late days per assignment
- Once you use up your slack days, all subsequent late submissions will accrue a 10% penalty daily (on top of any other penalties)

Collaboration policy

- We follow the CE Department Honor Code – read it carefully.
- Don’t look at solutions or code of others; everything you submit should be your own work
- Don’t share your solution or code with others although discussing general ideas is fine and encouraged
- Indicate in your submissions anyone you worked with

Prerequisites

- Machine Learning
- Knowledge of calculus and linear algebra
  - You can review it from this book: Mathematics of deep learning
- Programming (Python)
- Time and patience

TA Section

- Hands-on tutorials, with more practical details than the main lecture
- Workshops on some topics

Course Objectives

- Understanding neural networks and training issues
- Comprehending several popular networks for various tasks
- Fearlessly design, develop, and train networks
  - Hands-on practical experience.
- Gain an understanding of where the field is
  - Familiarity with recent models

40719 Deep Learning

- Deep Learning Basics (Lectures 1-6)
- Deep Learning Architectures (Lectures 7-14)
- Generative Models (Lectures 15-19)
- More recent models and topics (Lectures 20-27)

40719 Deep Learning

- Deep Learning Basics (Lectures 1-6)
  - MLP, back propagation, optimization and Generalization techniques, training issues and techniques
- Deep Learning Architectures (Lectures 7-14)
  - CNN, RNN, Attention, and Transformer
- Generative Models (Lectures 15-19)
  - VAE, GAN, diffusion models
- More recent models and topics (Lectures 20-27)

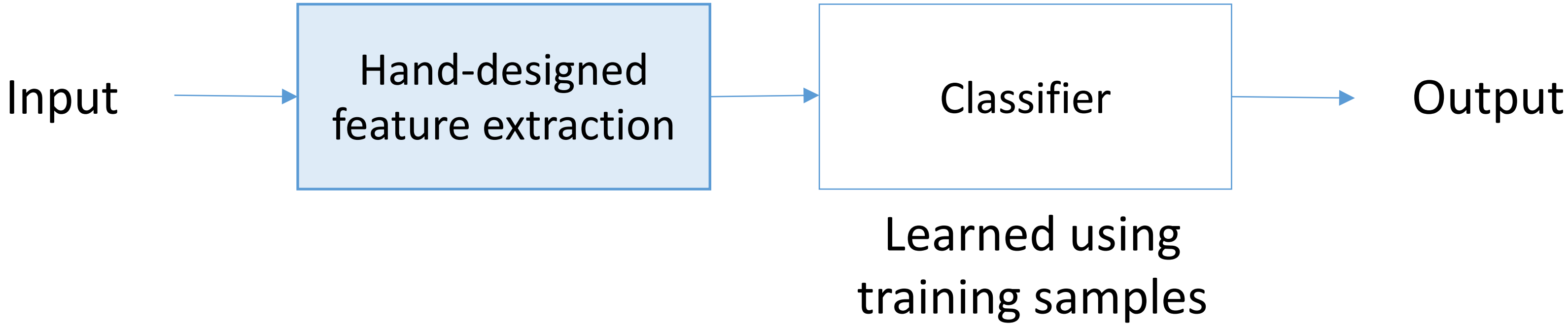
Deep learning

- Learning a computational models consists of multiple processing layers
  - learn representations of data with multiple levels of abstraction.

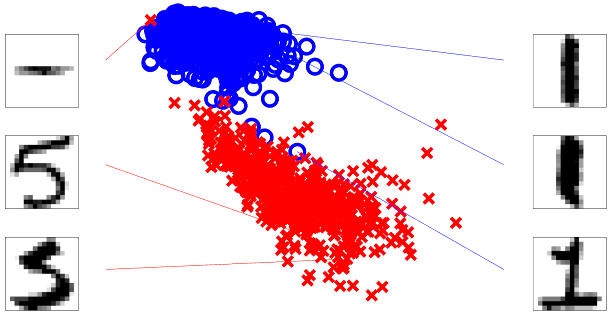
- Dramatically improved the state-of-the-art in many speech, vision and NLP tasks (and also in many other domains like bioinformatics)

Machine Learning Methods

- Conventional machine learning methods:
  - try to learn the mapping from the input features to the output by samples
  - However, they need appropriately designed hand-designed features



x_1: intensity
x_2: symmetry



[Abu Mostafa, 2012]

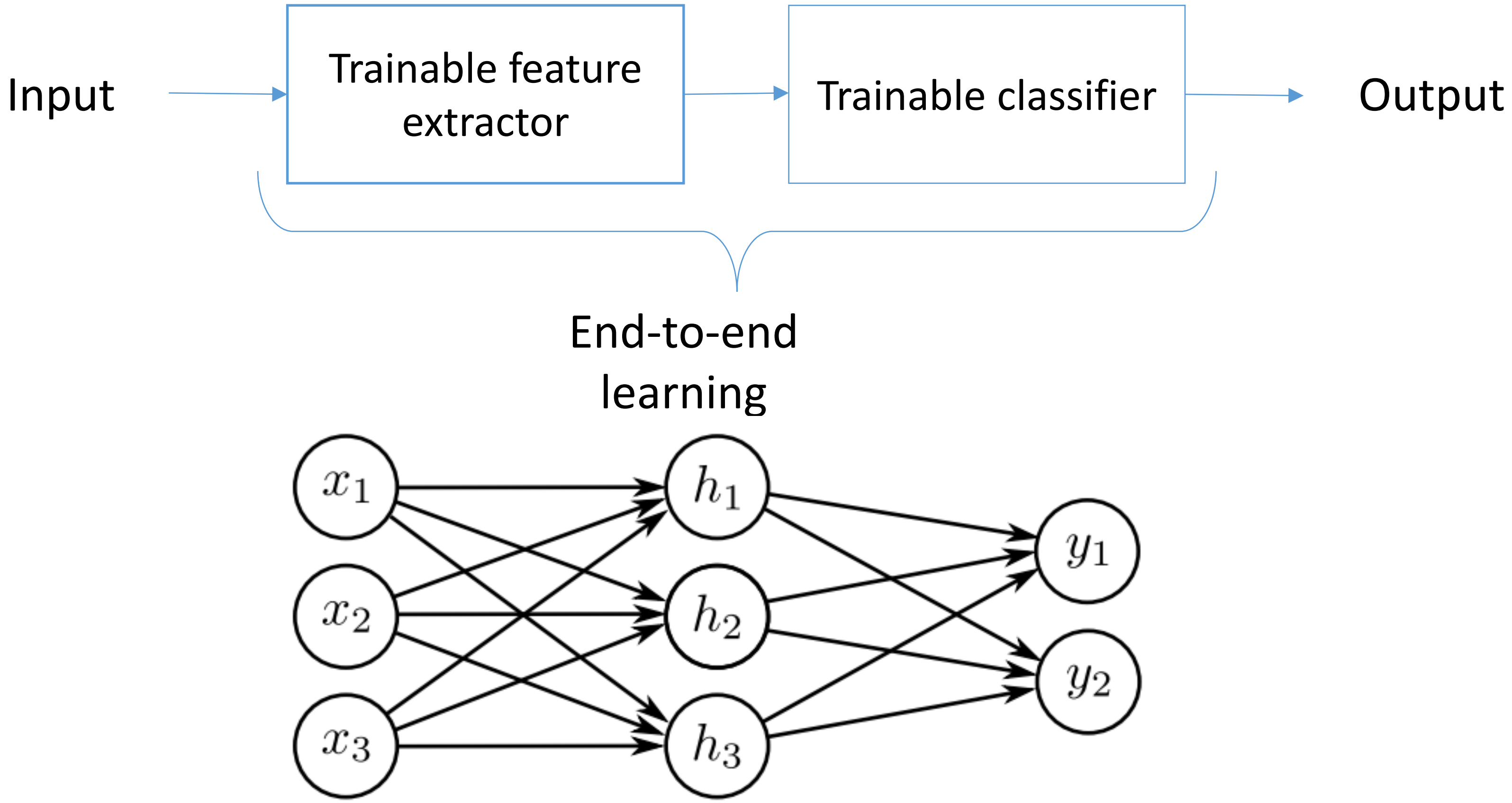
Representation of Data

- Performance of traditional learning methods depends heavily on the representation of the data.
  - **Most efforts were on designing proper features**

- However, designing hand-crafted features for inputs like image, videos, time series, and sequences is not trivial at all.
  - It is difficult to know which features should be extracted.
    - Sometimes, it needs long time for a community of experts to find (an incomplete and over-specified) set of these features.

Representation Learning

- Using learning to discover both:
  - the representation of data from input features
  - and the mapping from representation to output



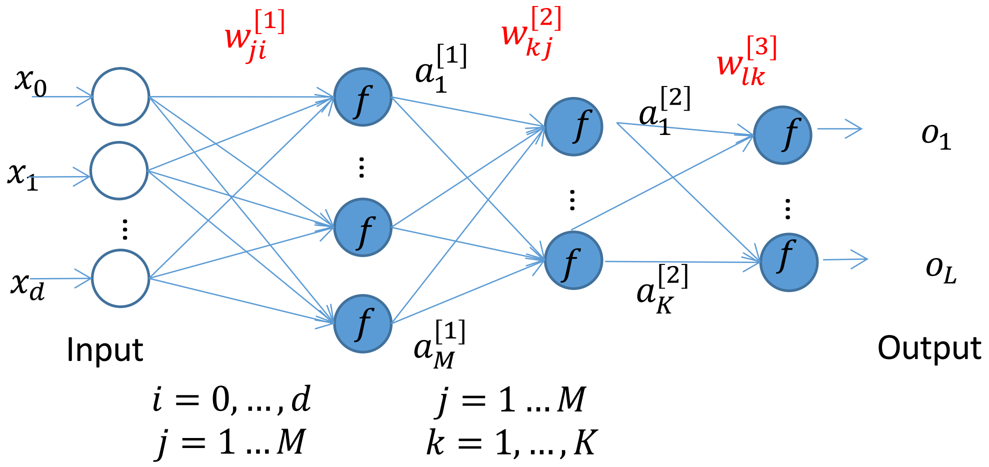
Previous Representation Learning Methods

- Although metric learning and kernel learning methods attempted to solve this problem, they were shallow models for feature (or representation) learning

- Deep learning finds representations that are expressed in terms of other, simpler representations
  - Usually hierarchical representation is meaningful and useful

Multi-layer Neural Network

- A multilayer perceptron is just a mapping input values to output values.
  - The function is formed by composing many simpler functions.
  - These middle layers are not given in the training data must be determined



Example of f functions:

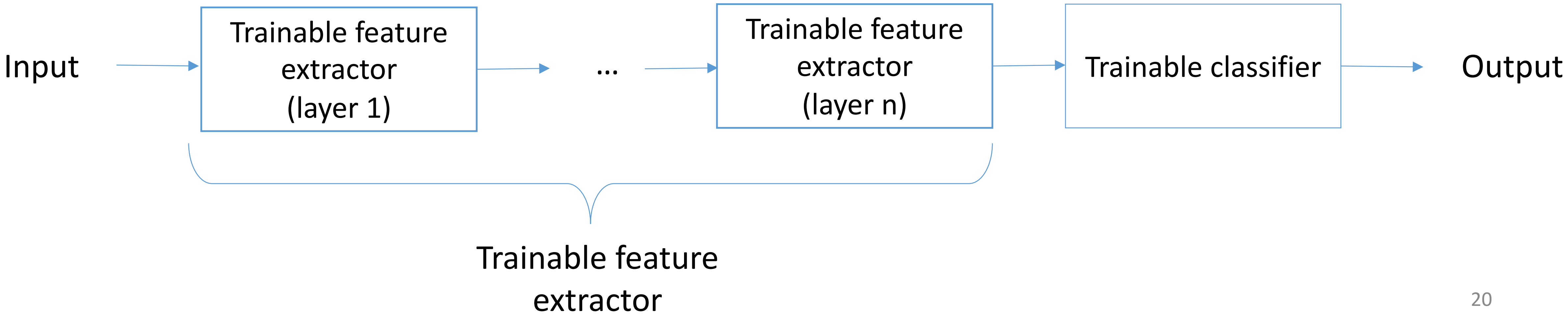
$$f(z) = \max(0, z)$$

$$a_k^{[l]} = f\left(\sum_{i=0}^{M} w_{ki}^{[l]} a_i^{[l-1]}\right)$$

Deep Learning Approach

- Deep breaks the desired complicated mapping into a series of nested simple mappings
  - each mapping described by a layer of the model.
  - each layer extracts features from output of previous layer

- shows impressive performance on many Artificial Intelligence tasks



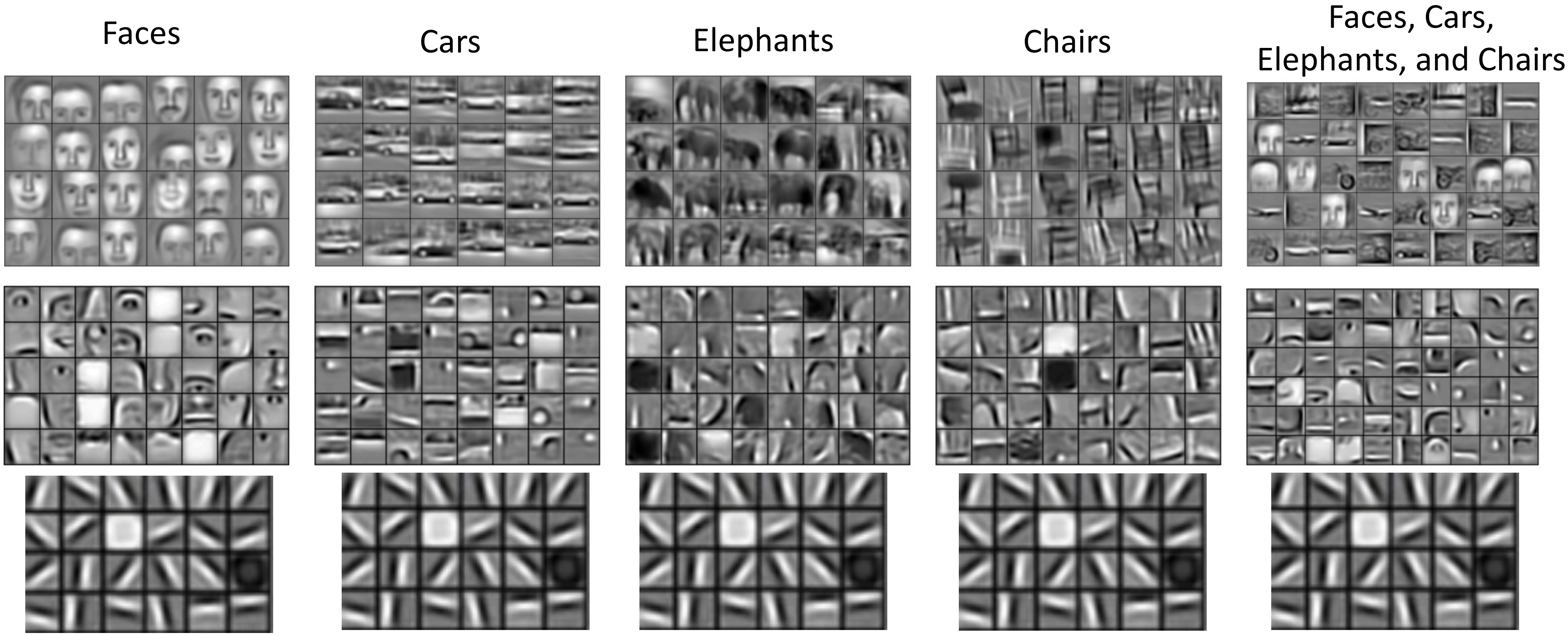
Deep Representations: The Power of Compositionality

- Compositionality is useful to describe the world around us efficiently
  - Learned function seen as a composition of simpler operations
  - Hierarchy of features, concepts, leading to more abstract factors enabling better generalization
    - each concept defined in relation to simpler concepts
    - more abstract representations computed in terms of less abstract ones.
  - Again, theory shows this can be exponentially advantageous

- Deep learning has great power and flexibility by learning to represent the world as a nested hierarchy of concepts

Slice credit: Yoshua Bengio

Example of Nested Representation



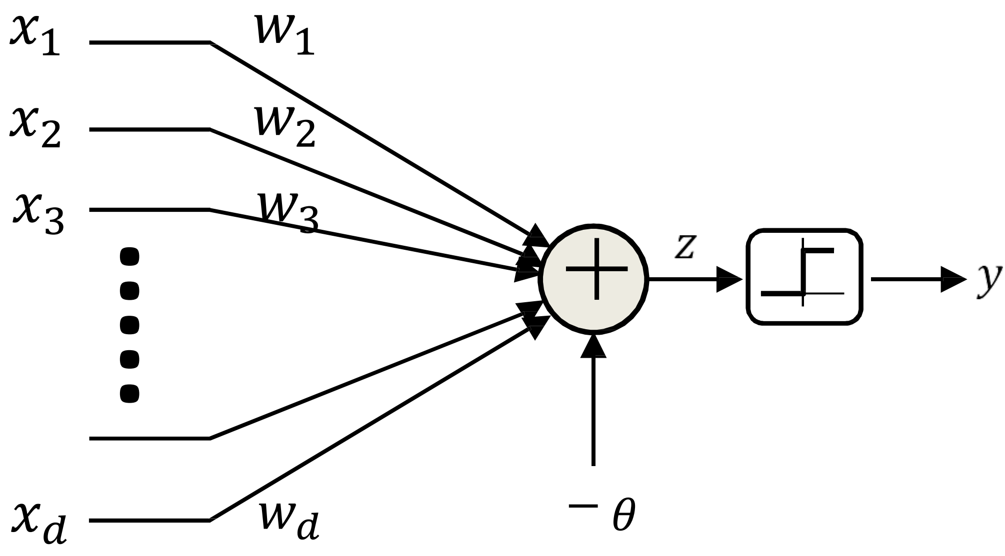
Lee et al., ICML 2009

Deep Learning Brief History

- 1940s–1960s:
  - development of theories of biological learning
  - implementations of the first models
    - perceptron (Rosenblatt, 1958) for training of a single neuron.
- 1980s-1990s: back-propagation algorithm to train a neural network with more than one hidden layer
  - too computationally costly to allow much experimentation with the hardware available at the time.
  - Small datasets
- 2006 “Deep learning” name was selected
  - ability to train deeper neural networks than had been possible before
    - Although began by using unsupervised representation learning, later success obtained usually using large datasets of labeled samples

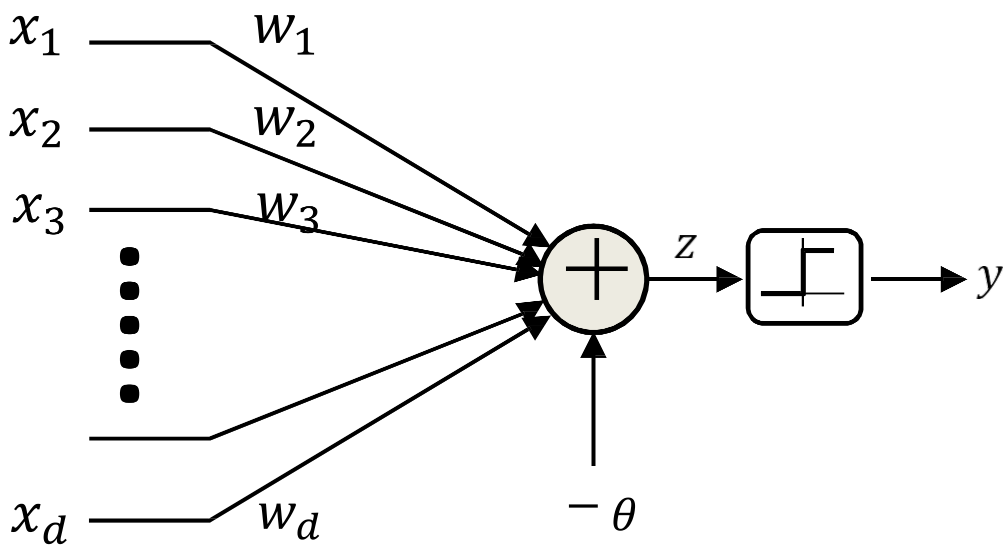
Deep Learning History: Timeline

- **1943 Artificial Neuron**
- 1957 Perceptron
- 1969 Limitations of Neural Networks
- 1979 Neocognitron (inspires CNNs)
- 1982 Recurrent Neural Networks (RNNs)
- 1986 Back propagation
- 1997 LSTM
- 1998 LeNet (Neocognitron+Backprop)
- 2006 Deep Learning
- 2009 ImageNet
- 2012 AlexNet



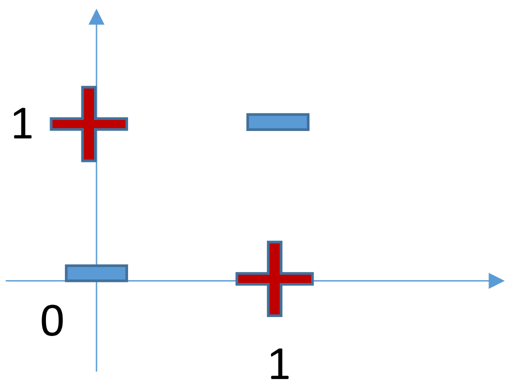
Deep Learning History: Timeline

- 1943 Artificial Neuron
- **1957 Perceptron**
- 1969 Limitations of Neural Networks
- 1979 Neocognitron (inspires CNNs)
- 1982 Recurrent Neural Networks (RNNs)
- 1986 Back propagation
- 1997 LSTM
- 1998 LeNet (Neocognitron+Backprop)
- 2006 Deep Learning
- 2009 ImageNet
- 2012 AlexNet



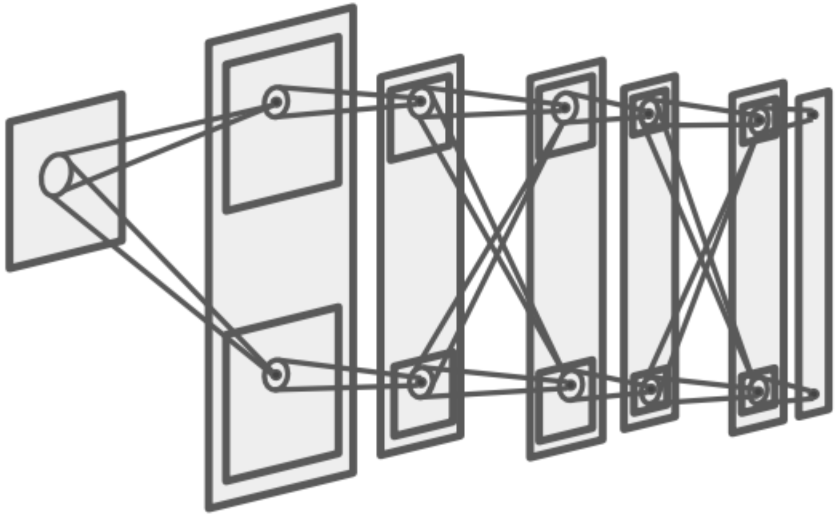
Deep Learning History: Timeline

- 1943 Artificial Neuron
- 1957 Perceptron
- **1969 Limitations of Neural Networks**
- 1979 Neocognitron (inspires CNNs)
- 1982 Recurrent Neural Networks (RNNs)
- 1986 Back propagation
- 1997 LSTM
- 1998 LeNet (Neocognitron+Backprop)
- 2006 Deep Learning
- 2009 ImageNet
- 2012 AlexNet



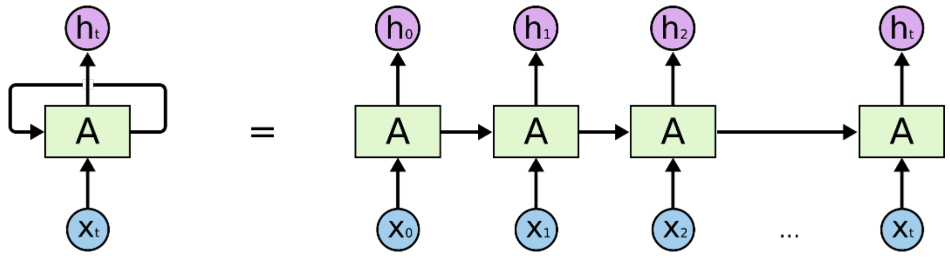
Deep Learning History: Timeline

- 1943 Artificial Neuron
- 1957 Perceptron
- 1969 Limitations of Neural Networks
- **1979 Neocognitron (inspires CNNs)**
- 1982 Recurrent Neural Networks (RNNs)
- 1986 Back propagation
- 1997 LSTM
- 1998 LeNet (Neocognitron+Backprop)
- 2006 Deep Learning
- 2009 ImageNet
- 2012 AlexNet



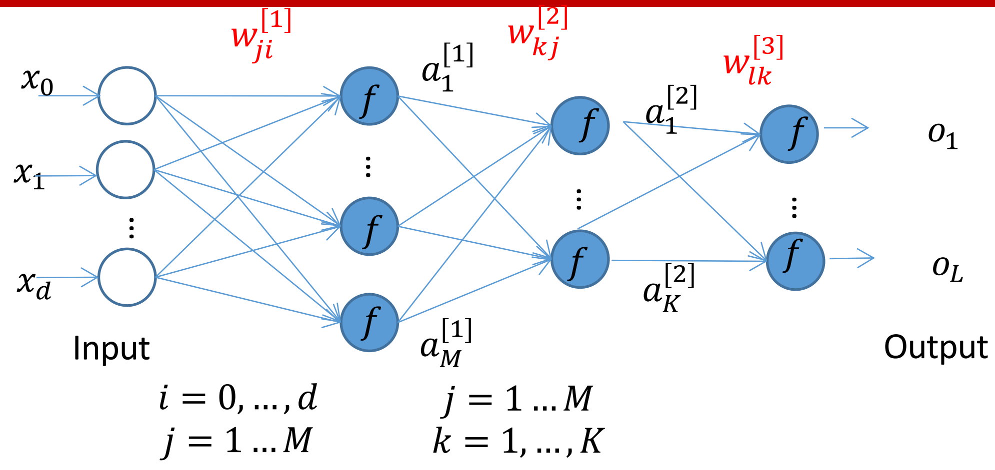
Deep Learning History: Timeline

- 1943 Artificial Neuron
- 1957 Perceptron
- 1969 Limitations of Neural Networks
- 1979 Neocognitron (inspires CNNs)
- **1982 Recurrent Neural Networks (RNNs)**
- 1986 Back propagation
- 1997 LSTM
- 1998 LeNet (Neocognitron+Backprop)
- 2006 Deep Learning
- 2009 ImageNet
- 2012 AlexNet



Deep Learning History: Timeline

- 1943 Artificial Neuron
- 1957 Perceptron
- 1969 Limitations of Neural Networks
- 1979 Neocognitron (inspires CNNs)
- 1982 Recurrent Neural Networks (RNNs)
- **1986 Back propagation (previously invented as automatic differentiation in 1970)**
- 1997 LSTM
- 1998 LeNet (Neocognitron+Backprop)
- 2006 Deep Learning
- 2009 ImageNet
- 2012 AlexNet

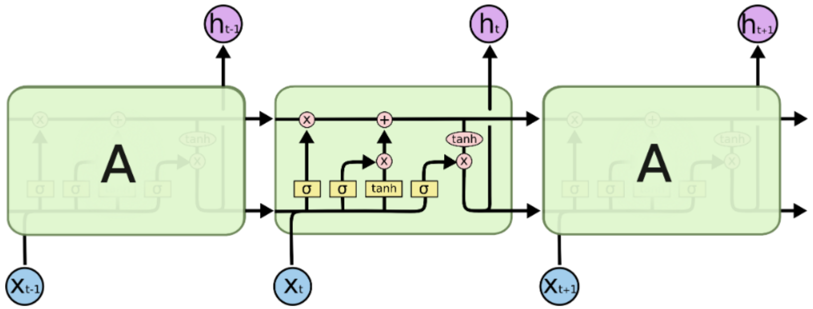


$$\frac{\partial E}{\partial w_{kj}^{[l]}} = \frac{\partial E}{\partial a_k^{[l]}} \frac{\partial a_k^{[l]}}{\partial w_{kj}^{[l]}}$$

$$\frac{\partial E}{\partial a^{[l]}} = \frac{\partial a^{[l+1]}}{\partial a_l^{[l]}} \frac{\partial E}{\partial a^{[l+1]}}$$

Deep Learning History: Timeline

- 1943 Artificial Neuron
- 1957 Perceptron
- 1969 Limitations of Neural Networks
- 1979 Neocognitron (inspires CNNs)
- 1982 Recurrent Neural Networks (RNNs)
- 1986 Back propagation
- **1997 LSTM**
- 1998 LeNet (Neocognitron+Backprop)
- 2006 Deep Learning
- 2009 ImageNet
- 2012 AlexNet

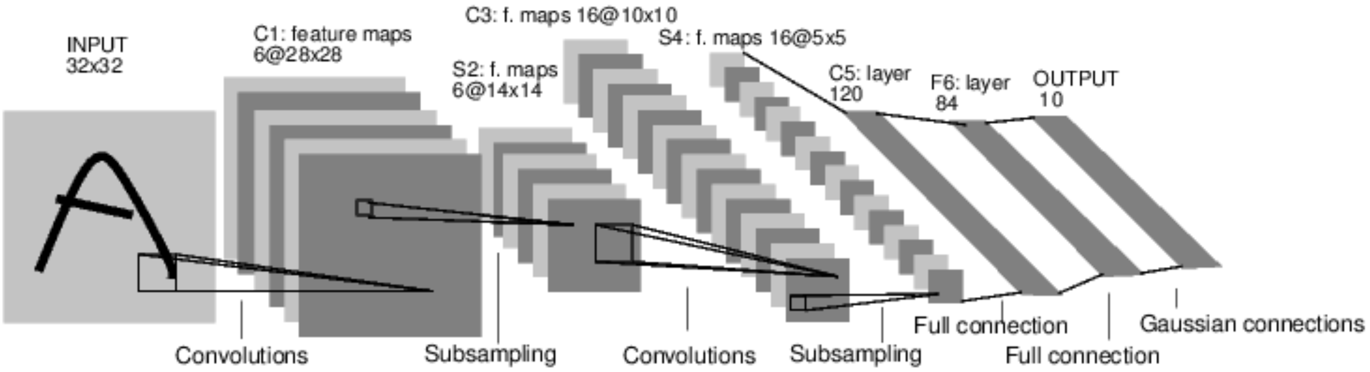


Source: Colah's blog

Deep Learning History: Timeline

- 1943 Artificial Neuron
- 1957 Perceptron
- 1969 Limitations of Neural Networks
- 1979 Neocognitron (inspires CNNs)
- 1982 Recurrent Neural Networks (RNNs)
- 1986 Back propagation
- 1997 LSTM
- **1998 LeNet (Neocognitron+Backprop)**
- 2006 Deep Learning
- 2009 ImageNet
- 2012 AlexNet

LeNet: Handwritten Digit Recognition (recognizes zip codes)
Training Sample : 9298 zip codes on mails



[LeNet, Yann Lecun, et. al, 1989]

Deep Learning History: Timeline

- 1943 Artificial Neuron
- 1957 Perceptron
- 1969 Limitations of Neural Networks
- 1979 Neocognitron (inspires CNNs)
- 1982 Recurrent Neural Networks (RNNs)
- 1986 Back propagation
- 1997 LSTM
- 1998 LeNet (Neocognitron+Backprop)
- **2006 Deep Learning**
- 2009 ImageNet
- 2012 AlexNet

The training of each layer individually is an easier undertaking
- Training multi layered neural networks became easier
- Per-layer trained parameters initialize further training

ImageNet

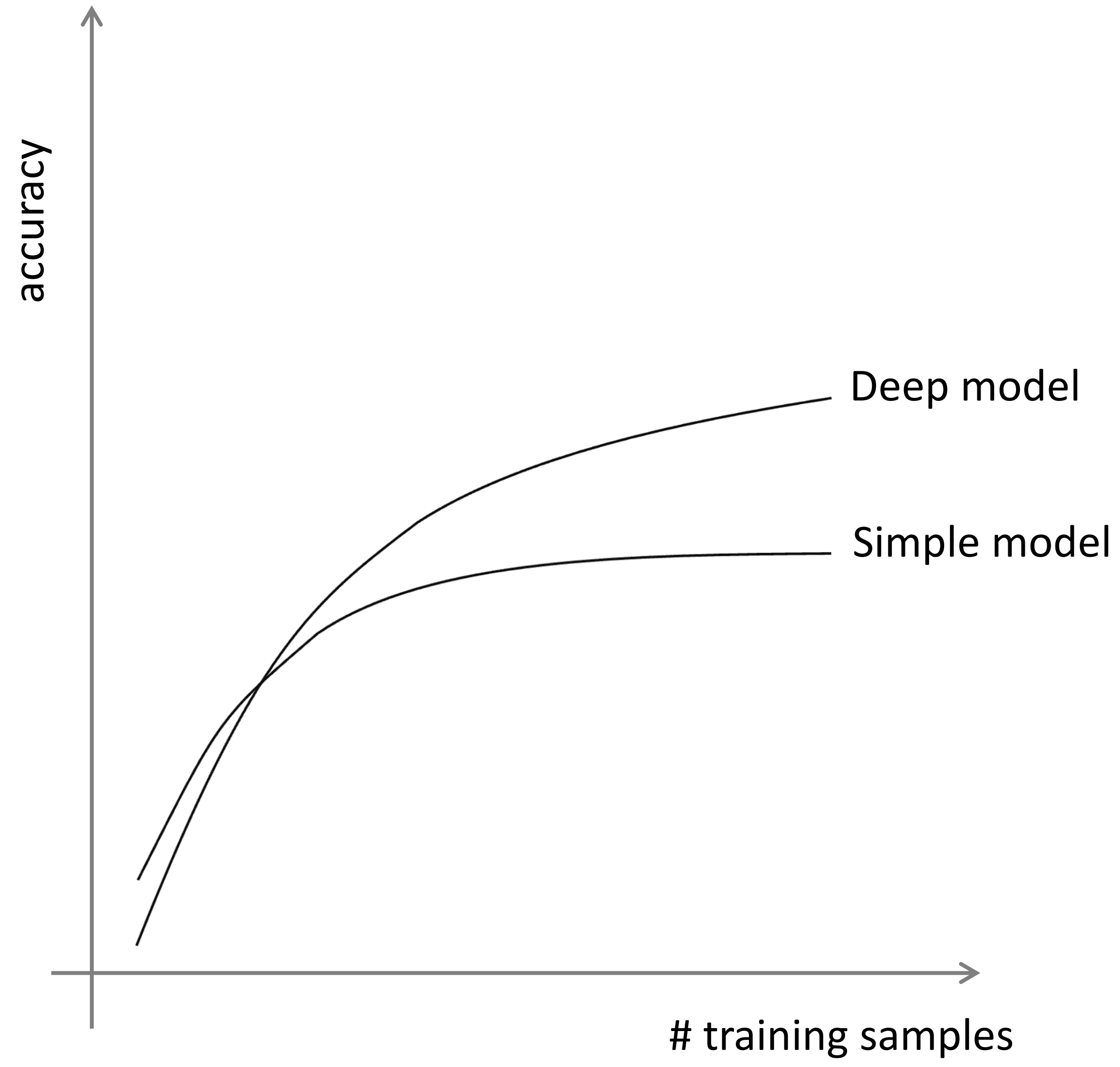
[Deng, Dong, Socher, Li, Li, & Fei-Fei, 2009]

- 22K categories and 14M images
  - Collected from web & labeled by Amazon Mechanical Turk

- The Image Classification Challenge:
  - Imagenet Large Scale Visual Recognition Challenge (ILSVRC)
  - 1,000 object classes
  - 1,431,167 images

- Much larger than the previous datasets of image classification

Large Datasets



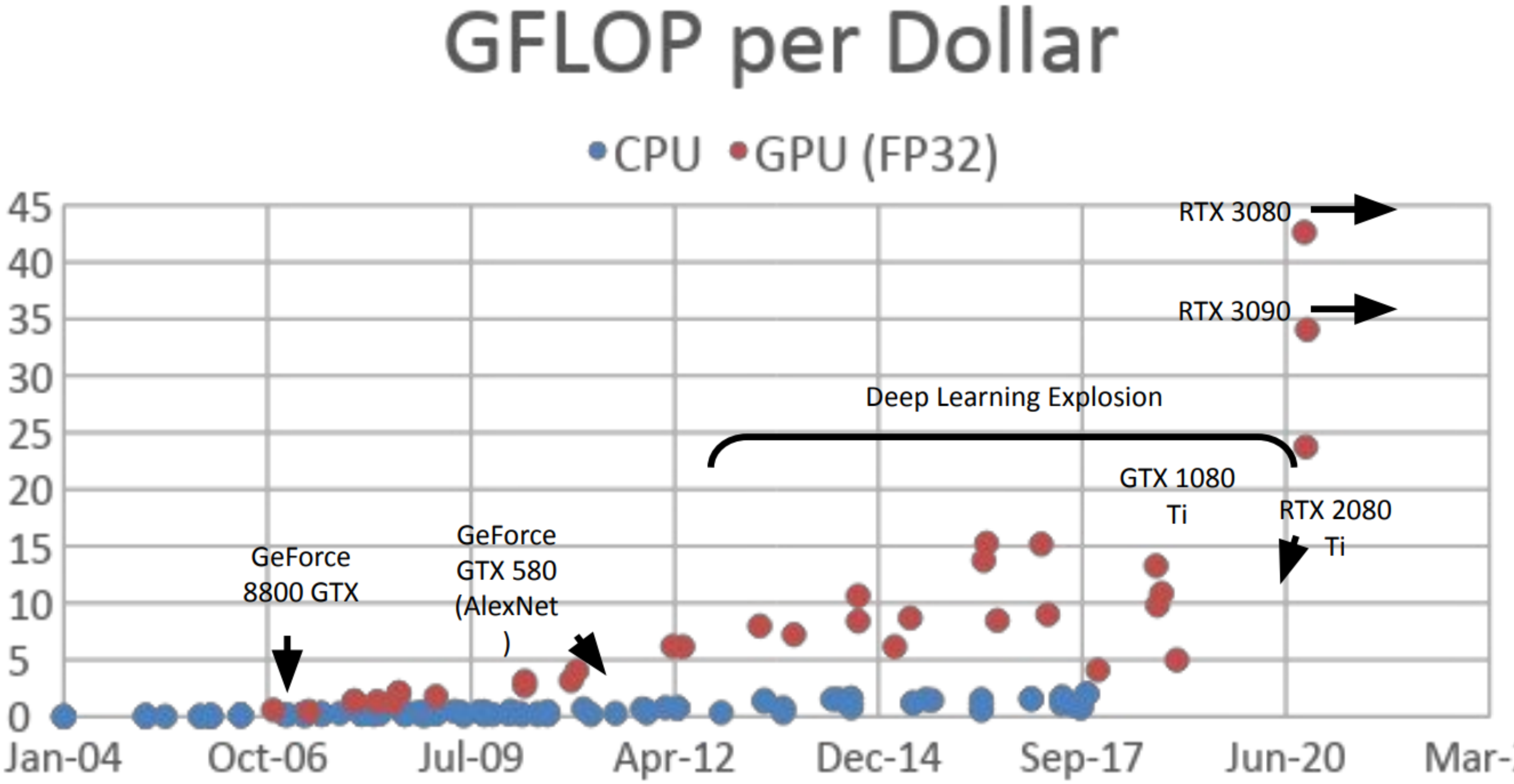
Why does deep learning become popular?

- Data: Large datasets

- Hardware: Availability of the computational resources to run much larger models

- Algorithm
  - New training techniques
  - New models
  - Frameworks

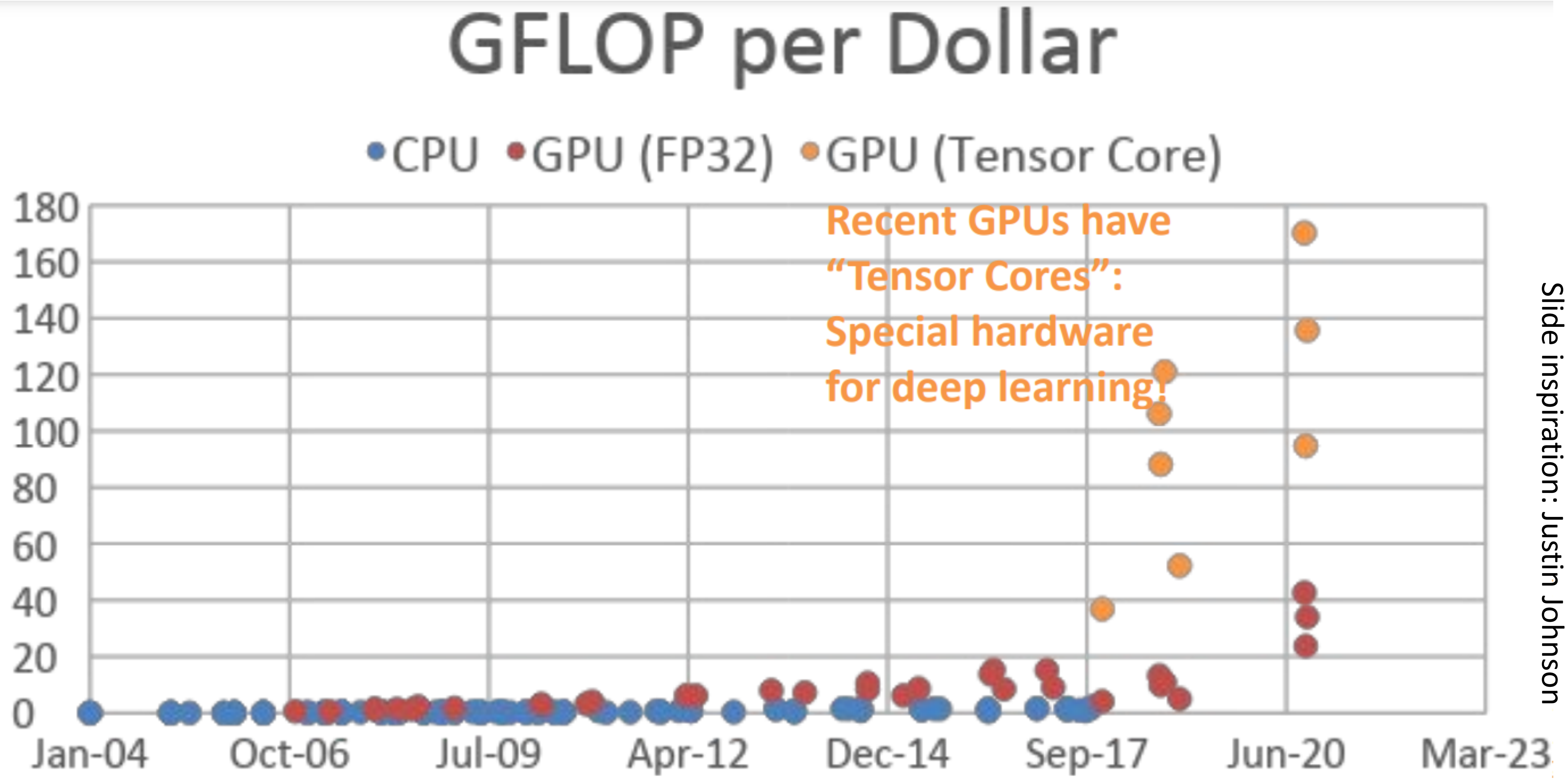
Computational Resources



Slide inspiration: Justin Johnson

Slide credit: Fei-Fei Li & Jiajun Wu & Ruohan Gao, CS231n

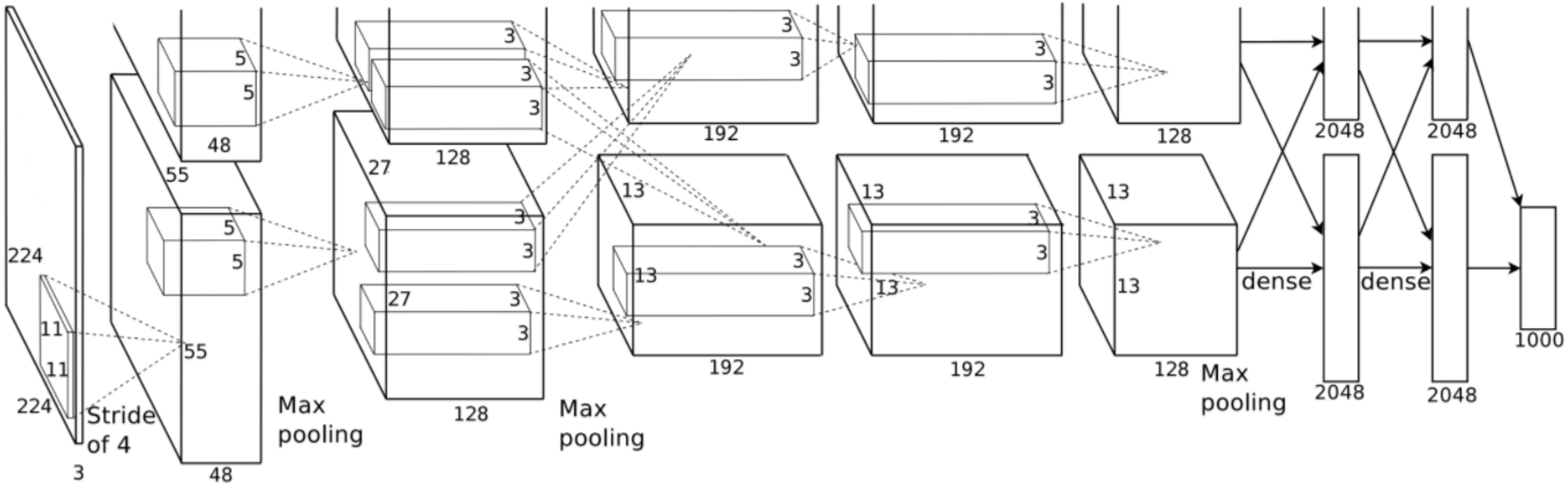
Computational Resources



Slide inspiration: Justin Johnson

Slide credit: Fei-Fei Li & Jiajun Wu & Ruohan Gao, CS231n

Alexnet (2012)

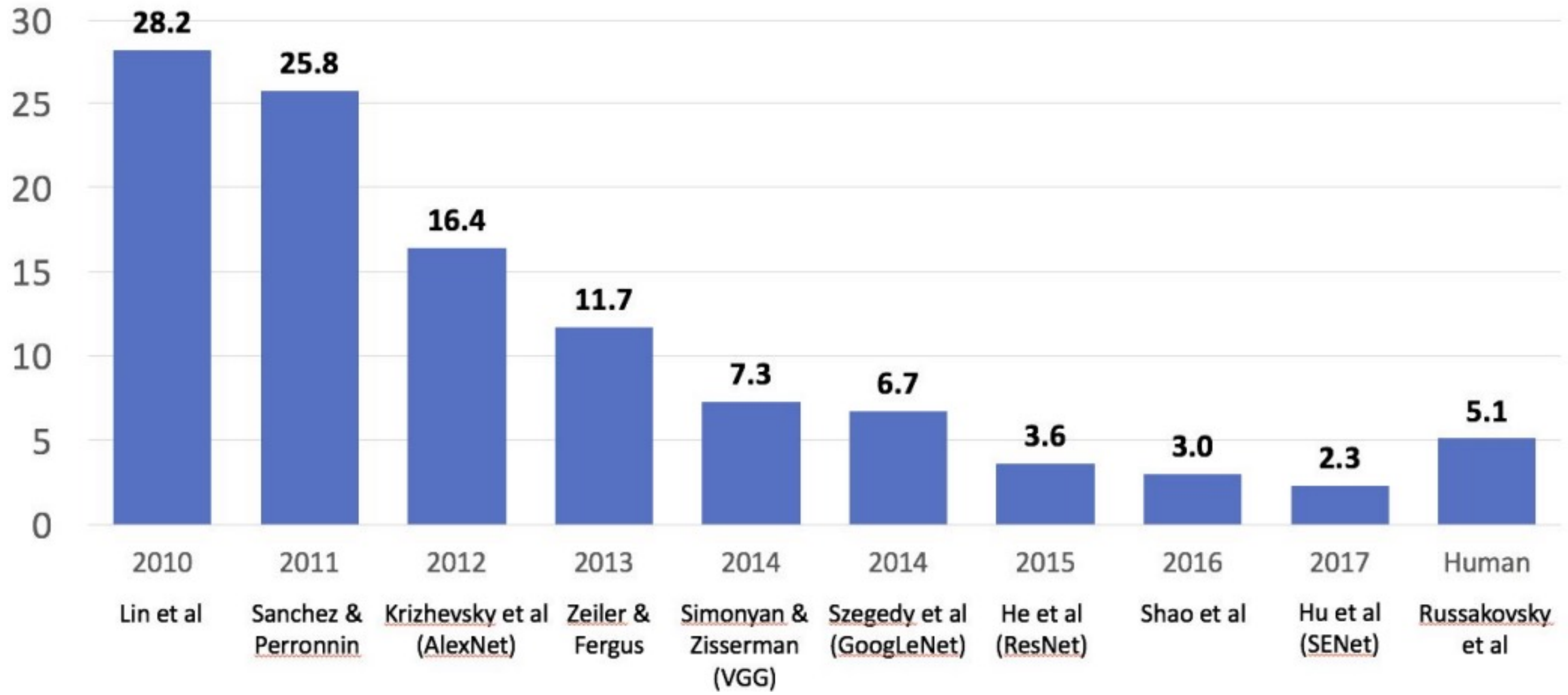


Krizhevsky, Alex, Sutskever, and Hinton, Imagenet classification with deep convolutional neural networks, NIPS 2012

- Reduces 25.8% top 5 error of the winner of 2011 challenge to 16.4%

Deep Models for Image Classification (ILSVRC)

- 5.1% is the performance of human on this data set



Using Pre-trained Models

- We don’t have large-scale datasets on all image tasks and also we may not time to train such deep networks from scratch

- On the other hand, learned weights for popular networks (on ImageNet) are available.

- Use pre-trained weights of these networks (other than final layers) as generic feature extractors for images

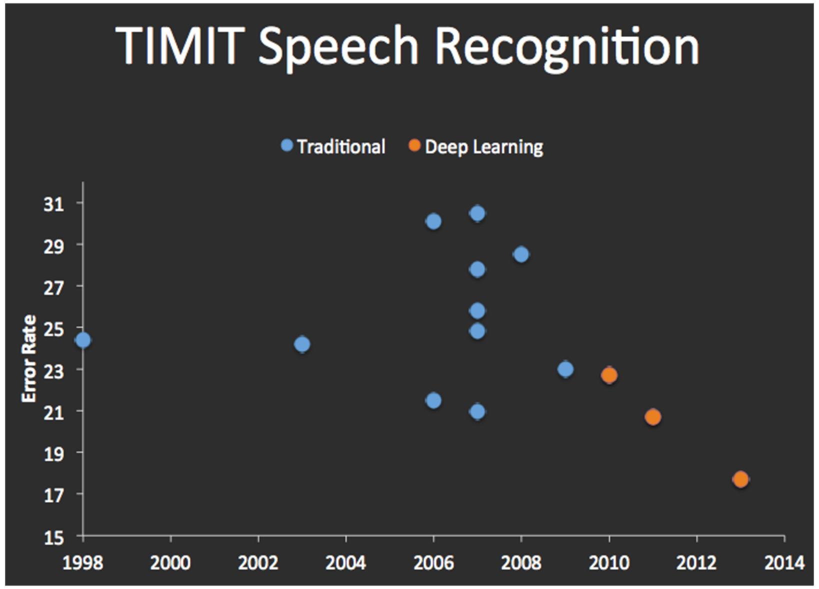
- Works better than handcrafted feature extraction on natural images

Other Vision Tasks

- After image classification, achievements were obtained in other vision tasks:
  - Object detection
  - Segmentation
  - Image captioning
  - Visual Question Answering (VQA)
  - …

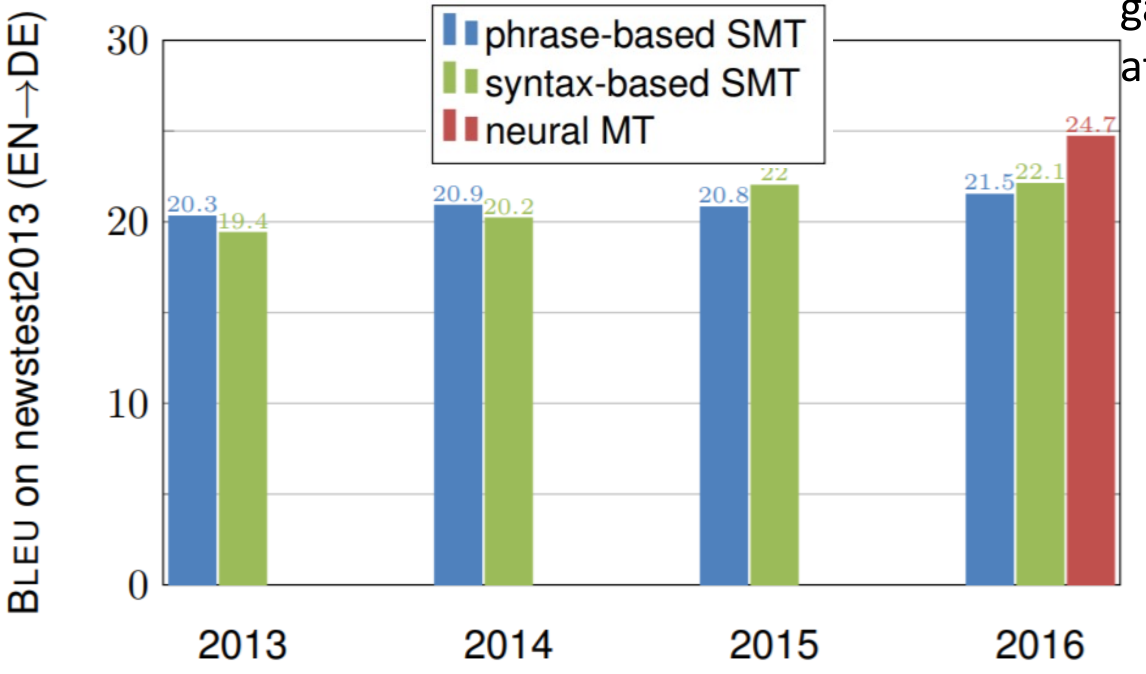
Similar Trends also in Speech and NLP

- Deep learning became SOTA also in speech and NLP tasks



Source: clarifai

Edinburgh's WMT Results Over the Years

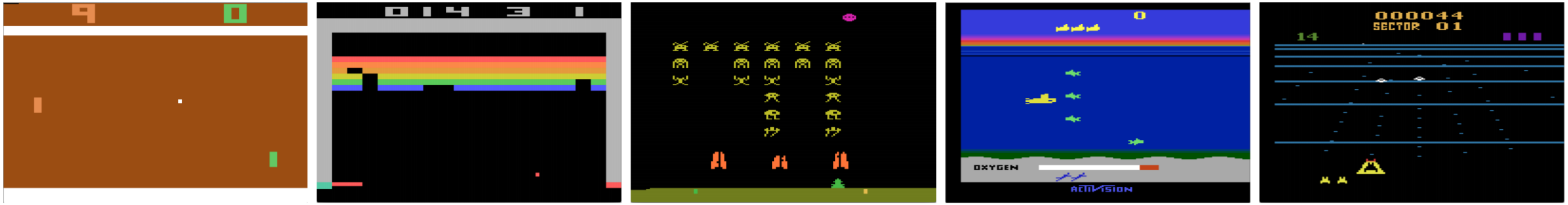


RNN with gating units + attention

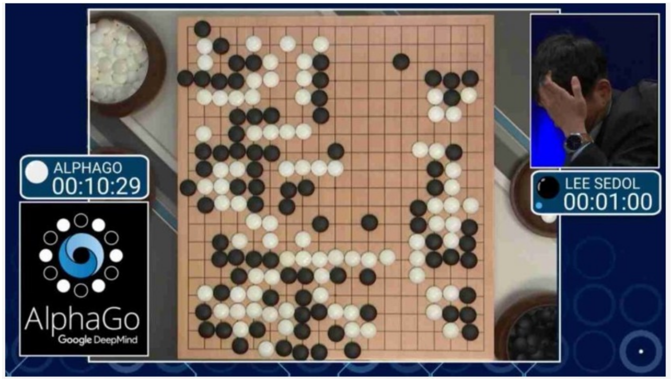
Source: http://www.meta-net.eu/events/meta-forum2016/slides/09_sennrich.pdf

Games

- DQN (2013): Atari 2600 games
  - neural network agent that is able to successfully learn to play as many of the games as possible without any hand-designed feature.



- Deep Mind’s alphaGo defeats former world champion in 2016.



Source: https://gogameguru.com/alphago-shows-true-strength-3rd-victory-lee-sedol/

Deep Learning is Everywhere

- Vision
- NLP
- Speech
- Games
- Bioinformatics
- …

nature

View all Nature Researc

Explore content ⌄ Journal information ⌄ Publish with us ⌄ Subscribe

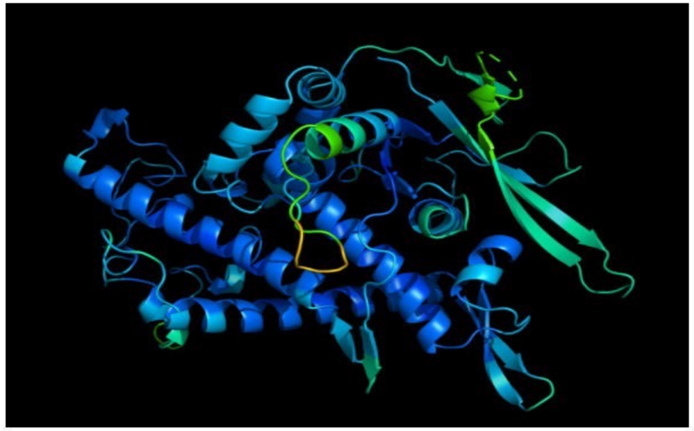
nature > news > article

NEWS · 30 NOVEMBER 2020

‘It will change everything’: DeepMind’s AI makes gigantic leap in solving protein structures

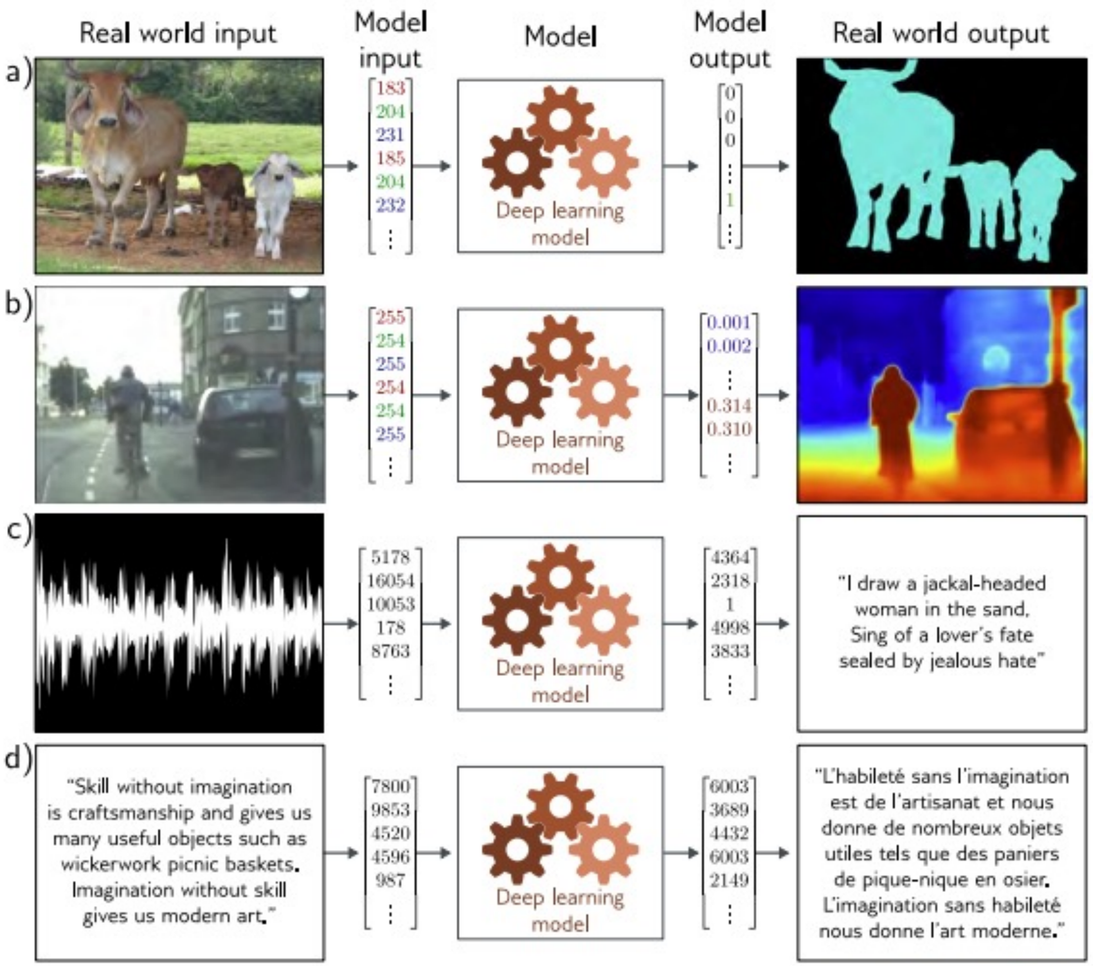
Google’s deep-learning program for determining the 3D shapes of proteins stands to transform biology, say scientists.

Ewen Callaway



A protein’s function is determined by its 3D shape. Credit: DeepMind

Some Examples



Understanding Deep Learning, 2023

AI’s Explosive Growth & Impact

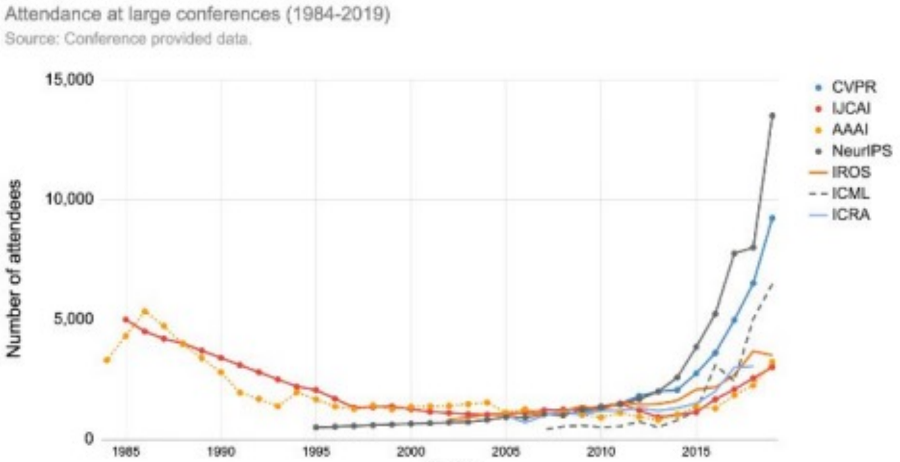
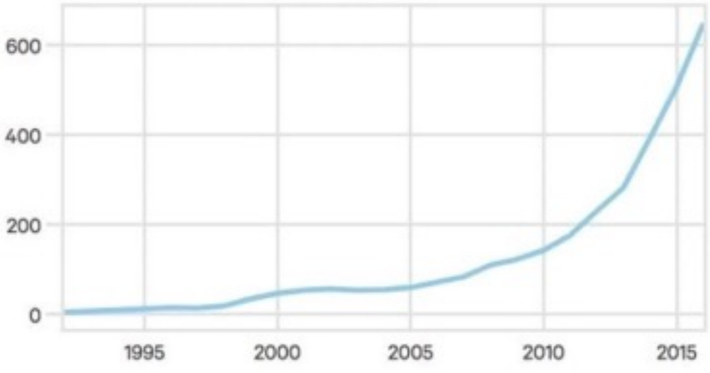
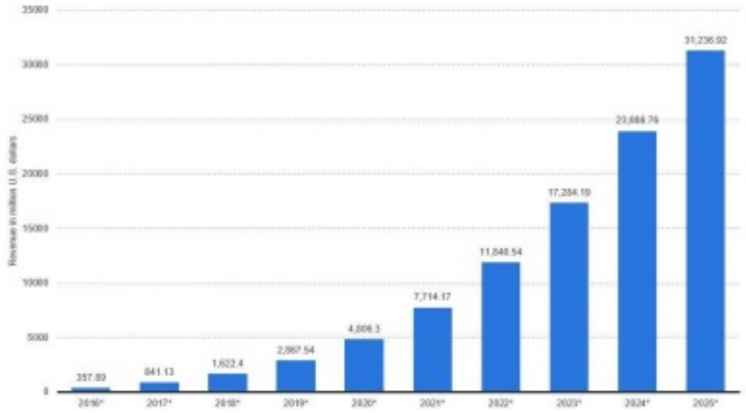


Fig. 2.1a



Year



Number of attendance At AI conferences

Source: The Gradient

Startups Developing AI Systems

Source: Crunchbase, VentureSource, Sand Hill Econometrics

Enterprise Application AI Revenue

Source: Statista

Slide credit: Fei-Fei Li & Jiajun Wu & Ruohan Gao, CS231n

Language: Transformer

- In 2017, transformer was introduced for NLP tasks.
- Achieved state-of-the-art results on eleven NLP tasks
- Pre-trained transformers (like BERT) can be fine-tuned for a wide range of tasks, such as question answering and language inference
  - without substantial task-specific architecture modifications
  - with just one additional output layer

Attention Is All You Need, NeurIPS 2017

Self-supervised Models

- How do we use unlabeled data for learning representations?
  - Predict next word / patch of image
  - Predict missing word / patch of image
  - Predict if two images are related (contrastive learning)

- Making poweful foundation models using this approach

Multi-modal Models

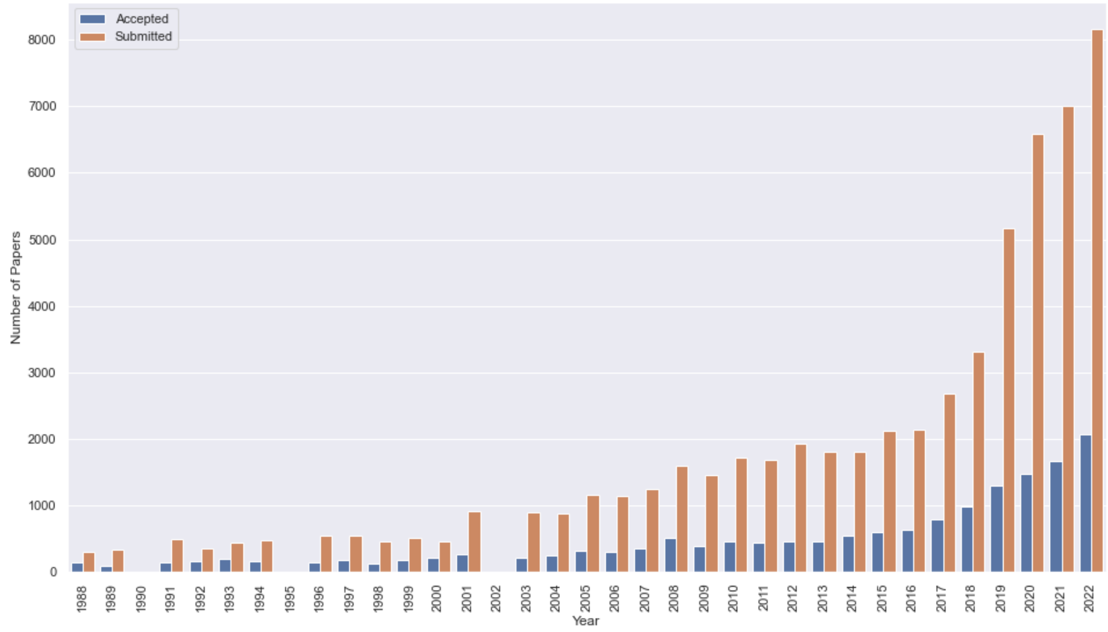
- Using the available large amount of multi-modal data

- CLIP: Learns a multi-modal embedding space by jointly training an image encoder and text encoder

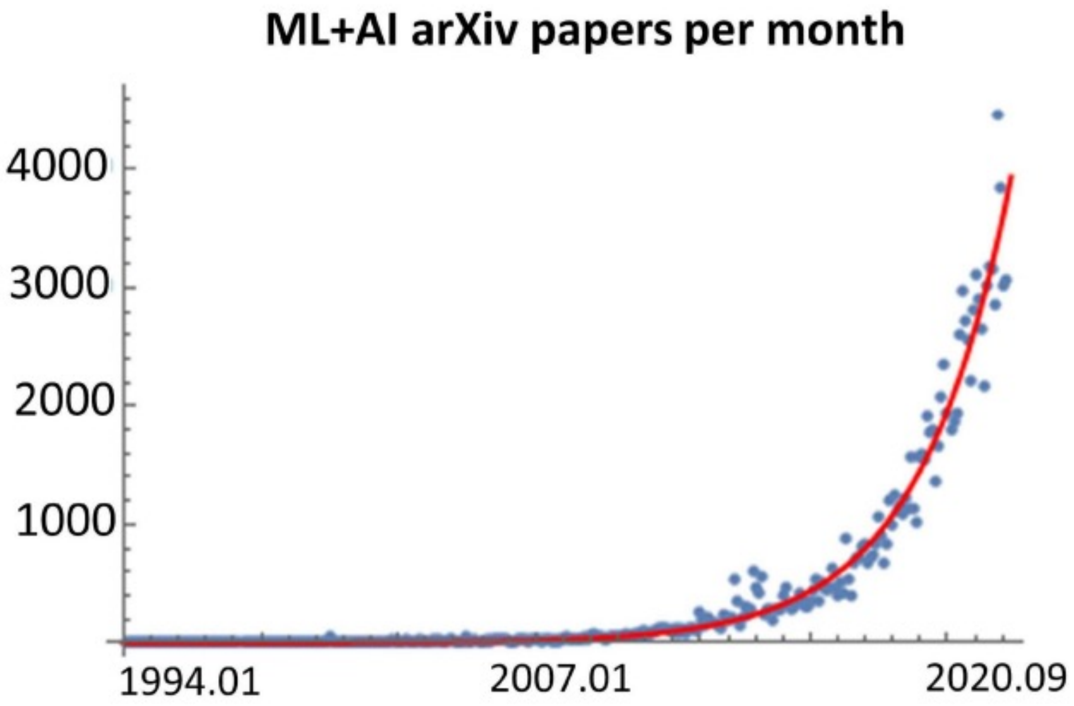
- Zero-shot classification

A. Radford, J.W. Kim, et al., “Learning Transferable Visual Models From Natural Language Supervision”, 2021

2012 to Present: Deep Learning Explosion



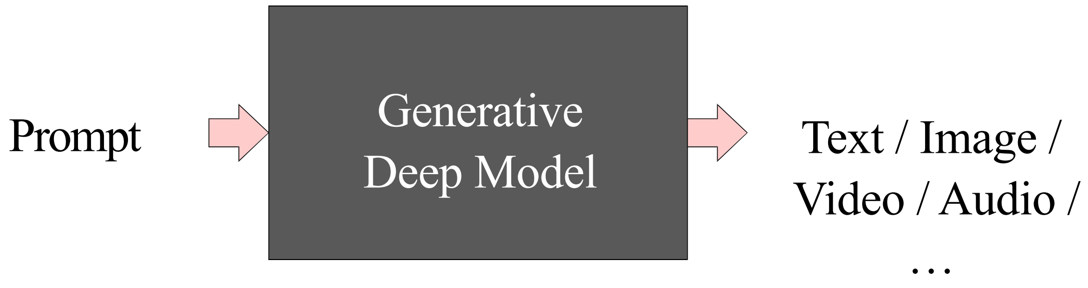
CVPR papers



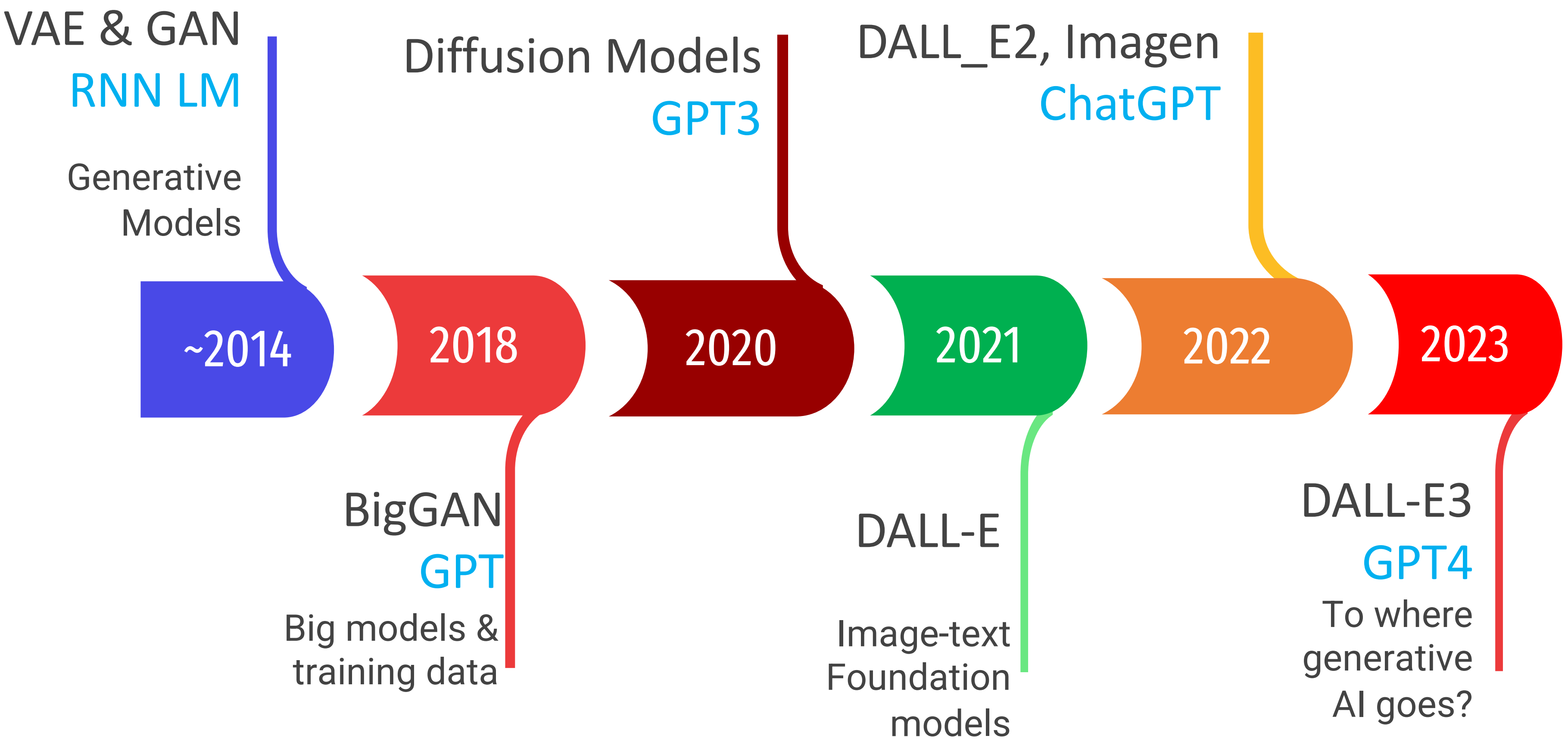
ML+AI papers (source)

Generative Models

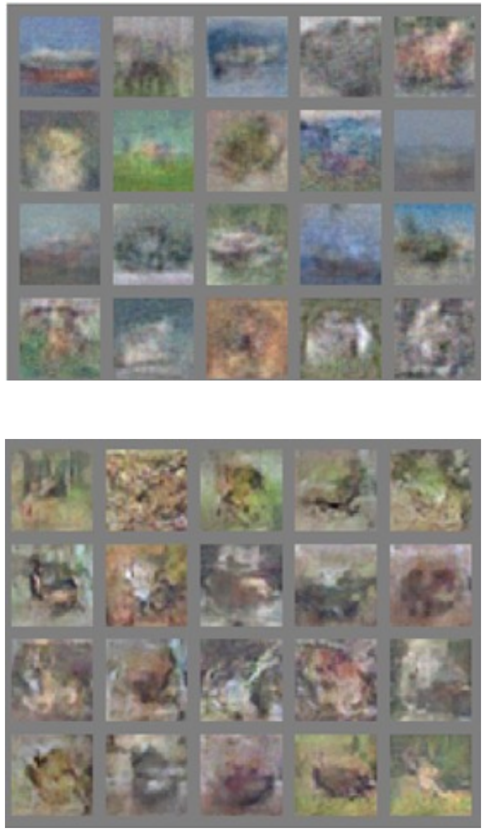
Output text, image, video, audio, …. given no condition or a partial guidance or prompt



Timeline of Generative Image/text models

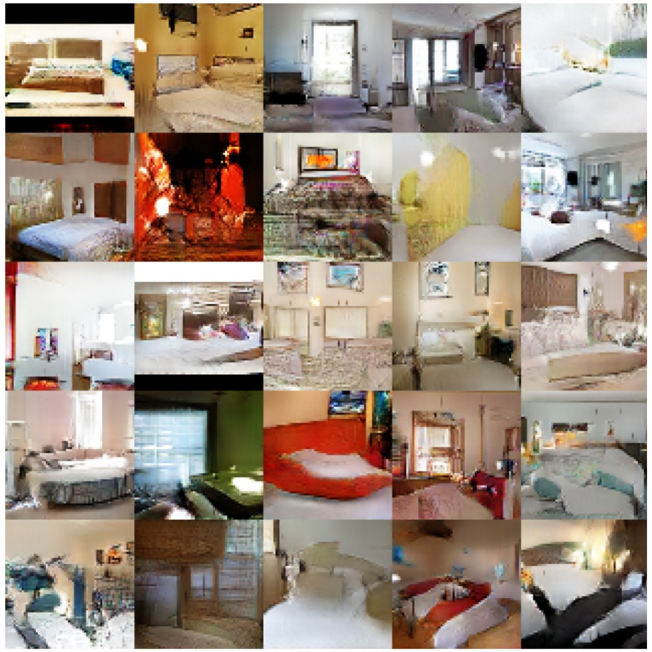


Generate samples by GAN



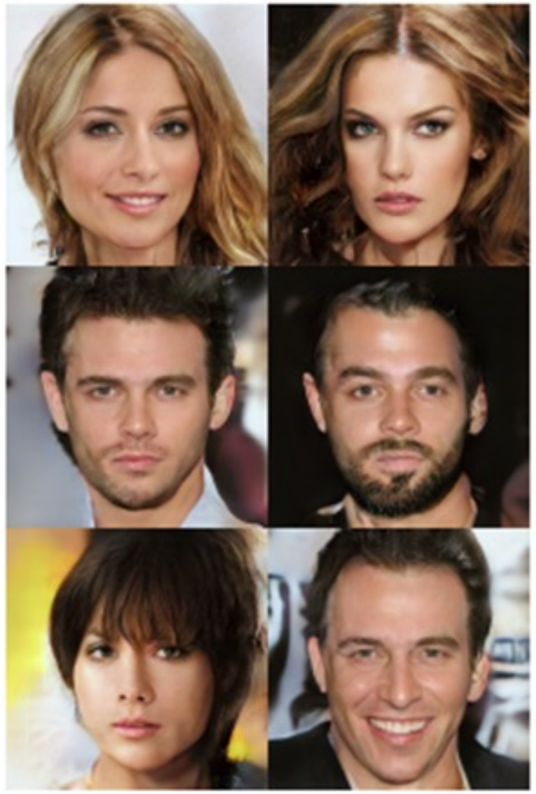
CIFAR-10

Goodfellow et al., GAN, NIPS 2014



LSUN

Radford et al., DCGAN, ICLR 2016



CelebA

Karras et al., Progressive GAN, ICLR 2018

ImageNet

Brocks et al., BigGAN, ICLR 2019

Denoising Diffusion Models

Emerging as powerful generative models, outperforming GANs

Dhariwal & Nichol, Diffusion Models Beat GANs, OpenAI, 2021

Ho et al., Cascaded Diffusion Models, Google, 2021

DALL-E

Text prompt: an armchair in the shape of an avocado. An armchair imitating an avocado

Ramesh et al, "DALL·E: Creating Images from Text", 2021.
https://openai.com/blog/dall-e/

DALL-E2 & Imagen

OpenAI DALL·E 2

a teddy bear on a skateboard in times square

[Ramesh et al., Hierarchical Text-Conditional Image Generation with CLIP Latents, 2022]

Google Imagen

A group of teddy bears in suit in a corporate office celebrating the birthday of their friend. There is a pizza cake on the desk.

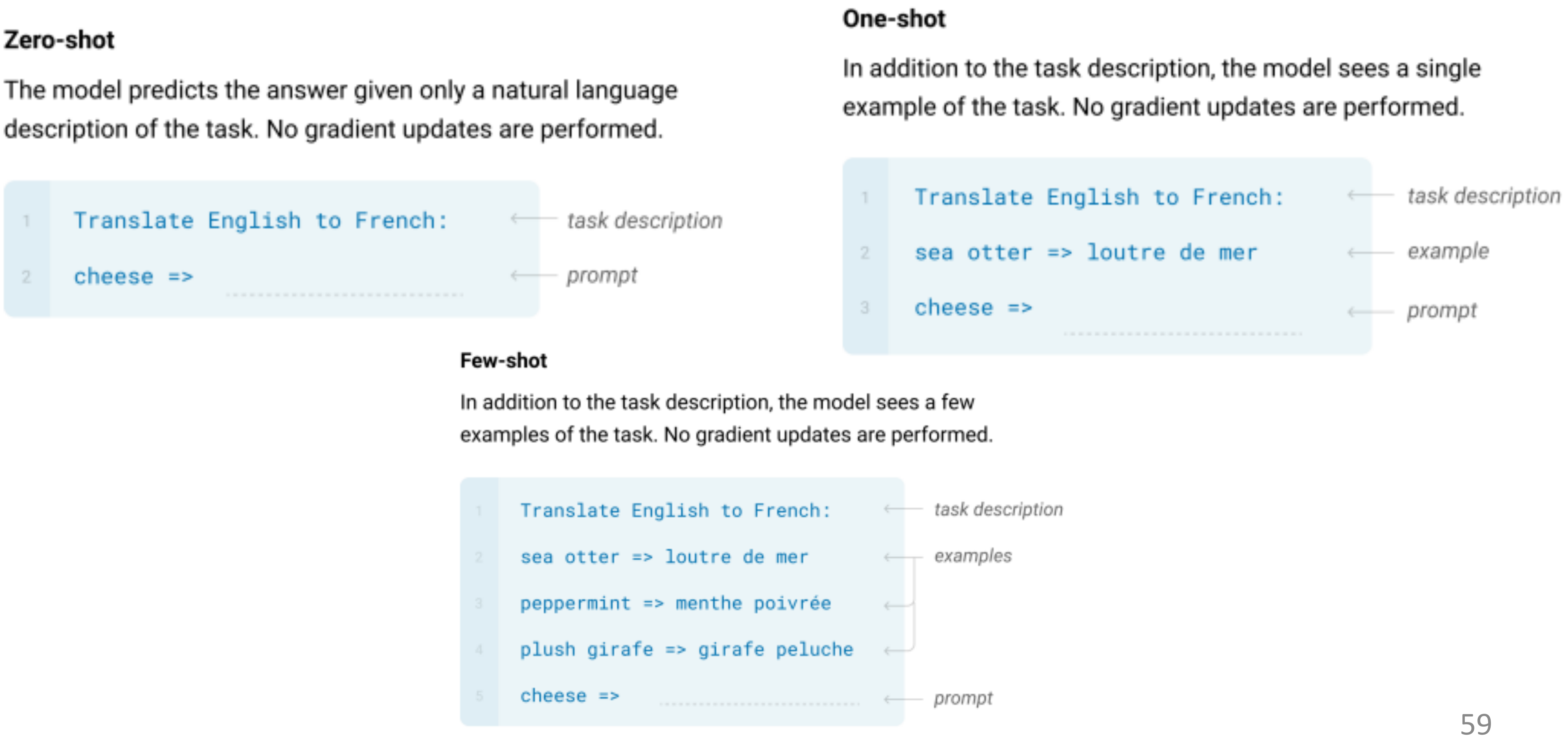
[Saharia et al., Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding, 2022]

StableDiffusion

Two snowmen are drinking some tea in the winter by the way they are in a cottage with a beautiful fireplace.

DALL-E 3

OpenAI GPT3: few-shot and zero-shot learning



[Brown et al., 2020]()

Large Language Models (LLMs)

- Large language models are one of the most successful applications of transformer models.

- Using massive unlabeled text datasets to learn LLMs (as self-supervised pre-trained models)
  - recognize, summarize, translate, predict and generate text and other content based on knowledge gained from massive datasets.

Recent Large Language Models (LLMs)

OpenAI ChatGPT (GPT3.5) & GPT4

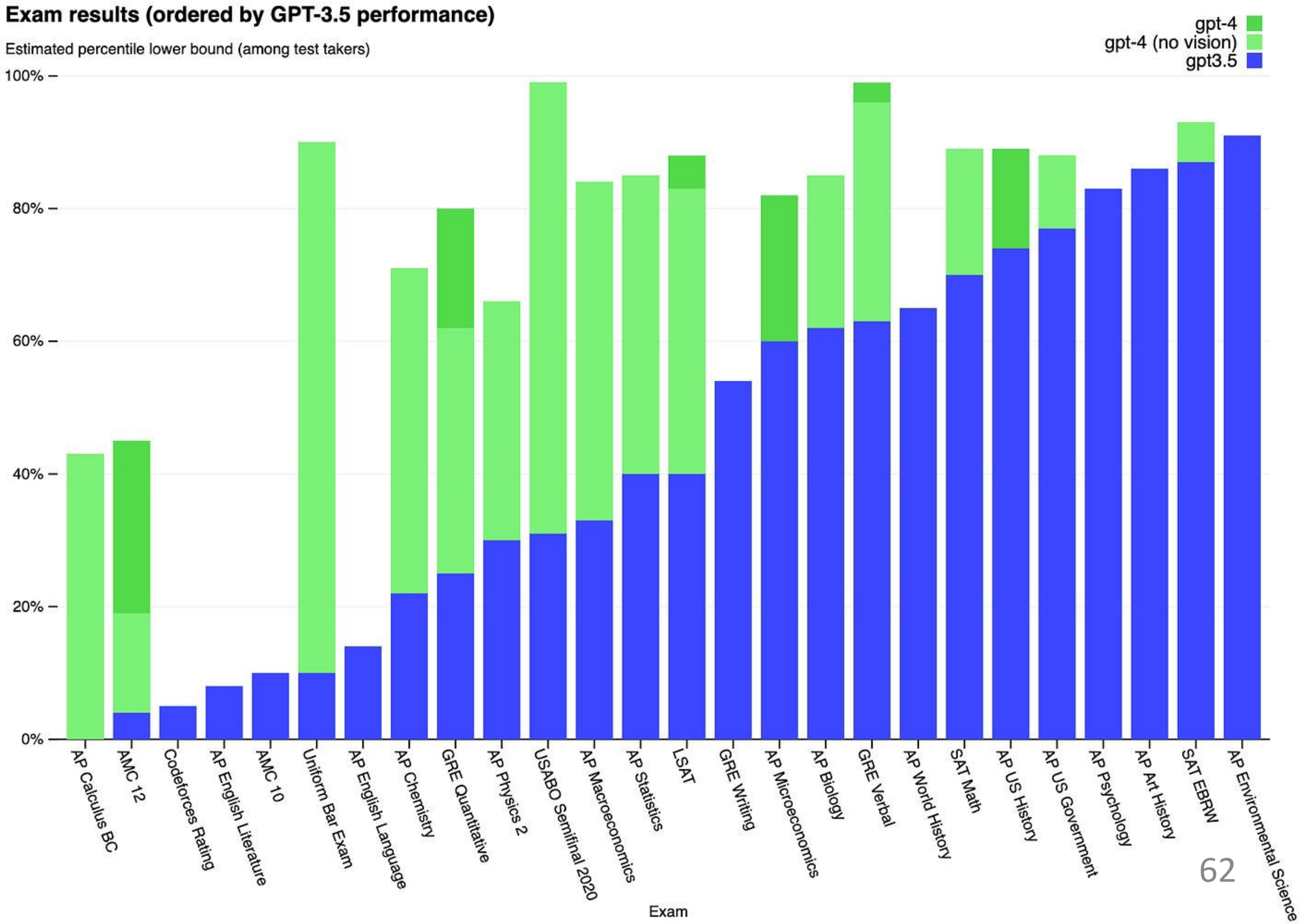
Google Bard & PaLM2

Conversational generative AI chatbots:

- generate response to a wide range of prompts and questions
- answer questions, translate languages, write different kinds of creative content, generate codes, provide summaries of topics or create stories

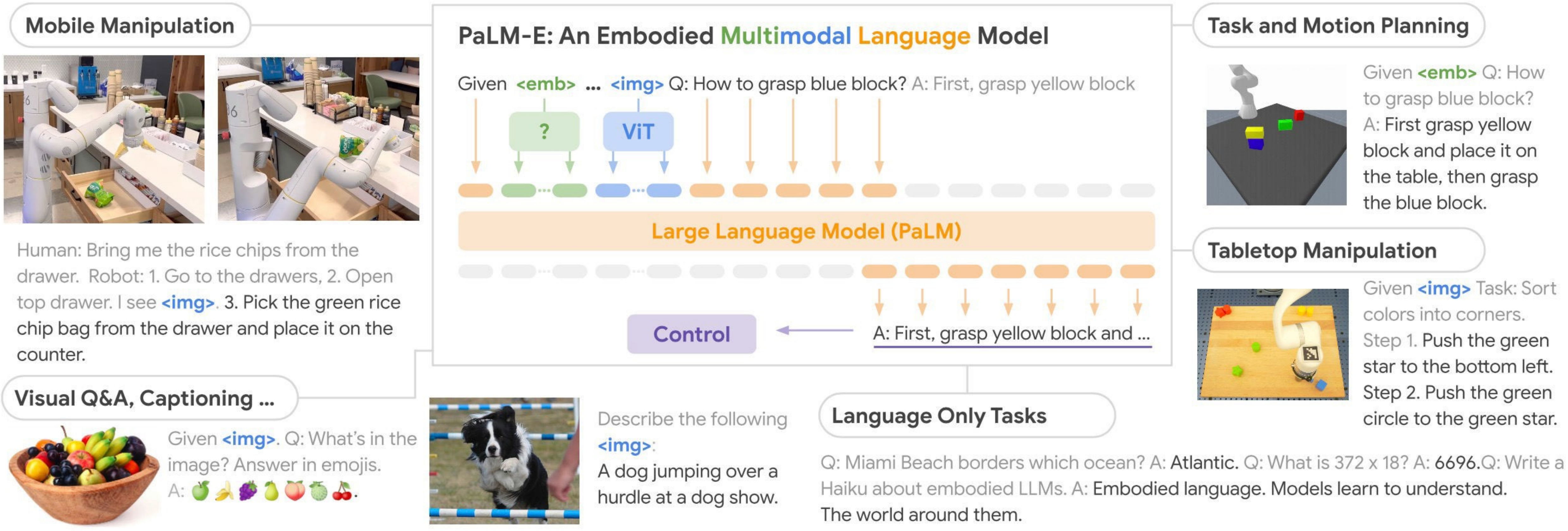
Tracking Progress

- How well AI can do human tasks



[OpenAI]

Can we build a single model of all data types?



[PaLM-E, Driess et al, 2023]

ChatGPT

HN

How will deep learning affect Sharif University of Technology and its students in the coming years?

In the coming years, deep learning is likely to have a significant impact on Sharif University of Technology and its students in the following ways:

1. Research: Deep learning is an active area of research and Sharif University of Technology is likely to continue to be at the forefront of this field, producing cutting-edge research and contributing to the development of deep learning algorithms and applications.
2. Curriculum: Deep learning is likely to become an increasingly important part of the curriculum at Sharif University of Technology, with new courses and programs being developed to meet the demand for expertise in this field.
3. Career Opportunities: The demand for experts in deep learning is growing rapidly, and graduates of Sharif University of Technology with deep learning skills will be well positioned to take advantage of these opportunities.
4. Industry Collaboration: Sharif University of Technology is likely to strengthen its relationships with industry partners to bring deep learning solutions to market and provide students with opportunities for hands-on experience.

Overall, deep learning is likely to have a positive impact on Sharif University of Technology and its students, helping to advance the field, provide new career opportunities, and improve the quality of education.

ChatGPT Jan 30 Version

ChatGPT

HN

Is there a danger that deep learning will be able to do what junior engineers do cheaply, thus making the income levels of Sharif graduates fall drastically?

It is true that deep learning has the potential to automate certain tasks that are currently performed by junior engineers, and this could result in some job displacement. However, it is important to note that deep learning is not a replacement for all types of engineers, but rather a tool that can be used to support and augment their work.

Additionally, the development of new technologies often leads to the creation of new jobs, as well as the need for people with skills in those technologies. For example, there will likely be a growing demand for experts in deep learning who can design, build, and implement these systems.

Graduates from Sharif University of Technology with deep learning skills will be well positioned to take advantage of these new job opportunities and play a key role in the development and deployment of deep learning systems. Furthermore, deep learning is a rapidly evolving field, and new applications and use cases are being discovered all the time, which could lead to further job opportunities.

In conclusion, while automation brought about by deep learning may disrupt certain jobs, it is also likely to create new job opportunities and markets, particularly in the field of deep learning. Graduates from Sharif University of Technology with deep learning skills will be well positioned to take advantage of these opportunities and succeed in the changing job market.

ChatGPT Jan 30 Version

ChatGPT

HN Which skills will have the most market value in the coming years for Sharif graduates?

The market value of skills for Sharif graduates will likely evolve over time in response to changes in technology and industry demand. However, some skills that are likely to have high market value in the coming years include:

1. Deep Learning: As deep learning continues to revolutionize various industries, expertise in this field is likely to be in high demand. This includes skills in areas such as computer vision, natural language processing, and reinforcement learning.
2. Data Science: The ability to collect, process, analyze, and interpret large amounts of data will continue to be a valuable skill. This includes expertise in areas such as data visualization, data cleaning, and machine learning.
3. Software Engineering: As more and more companies adopt deep learning and other advanced technologies, the need for software engineers who can design, build, and maintain these systems will increase.
4. Cloud Computing: The increasing use of cloud computing platforms such as Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP) has led to a growing demand for experts in cloud computing.
5. Cybersecurity: As more systems and devices become connected, the need for experts in cybersecurity will continue to grow. This includes skills in areas such as network security, threat intelligence, and incident response.
6. Interdisciplinary Skills: In addition to technical skills, graduates with interdisciplinary skills such as communication, teamwork, project management, and leadership will be well positioned to succeed in a rapidly changing job market.

It is important to note that the market value of skills will continue to evolve over time, and it is important for Sharif graduates to continually update their skills and stay current with the latest technologies and industry trends.

ChatGPT Jan 30 Version

ChatGPT

HN

Will Sharif graduates be better served with a rigorous, challenging course in deep learning that has a lot of difficult theoretical and practical exercises? Or will they be better off with a course that teaches them the fundamentals without overloading them, allowing them to self-study further interests, and improve their GPA?

The answer to this question depends on the individual goals and aspirations of each Sharif graduate.

For those who are interested in pursuing a career in deep learning or related fields, a rigorous and challenging course that covers both the theoretical foundations and practical applications of deep learning will likely provide the best preparation. This type of course will provide a strong foundation of knowledge and skills, as well as the opportunity to gain hands-on experience with the technology.

On the other hand, for those who may not be interested in a career in deep learning but still want to have a general understanding of the technology, a course that covers the fundamentals without overloading them may be a better fit. This approach will allow them to gain a basic understanding of deep learning without dedicating a significant amount of time and effort, and will also leave them more time and resources to focus on other areas of interest.

Ultimately, the choice between a rigorous and challenging course or a more basic course will depend on the individual goals and interests of each Sharif graduate. Both approaches have their benefits, and it is up to each student to determine which approach will best serve their needs and career aspirations.

ChatGPT Jan 30 Version