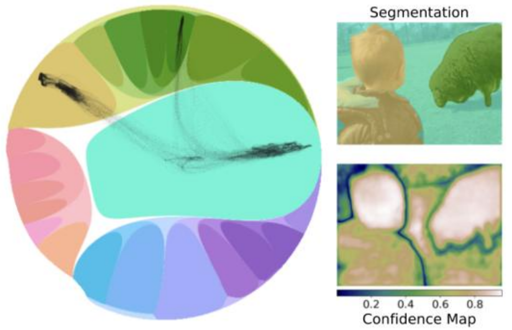
Deep Learning 1

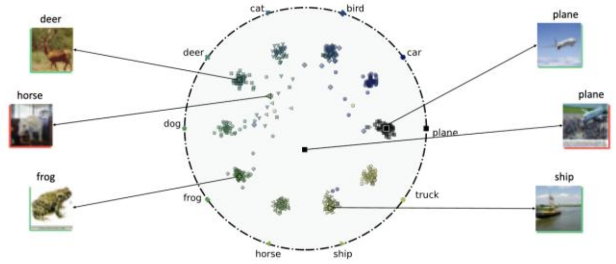
2025-2026 – Pascal Mettes

Lecture 1

Introduction and history of deep learning

A warm welcome from me



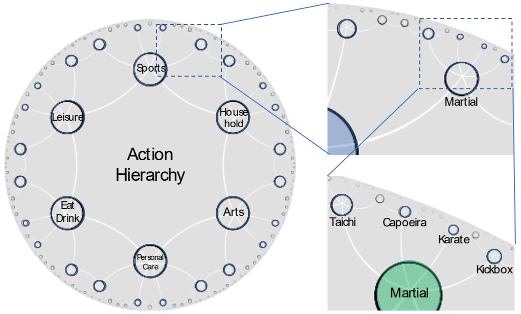


About me

Assistant Professor *at University of Amsterdam*

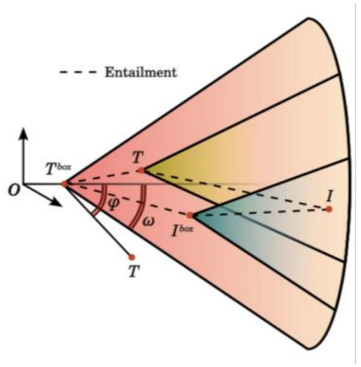
PI of *HypLab (Hyperbolic Deep Learning Lab)*

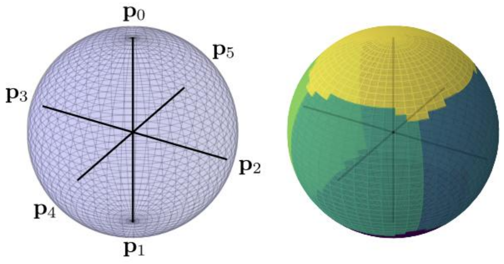
Scientific manager of *HAVA Lab (Human Aligned Video AI Lab)*



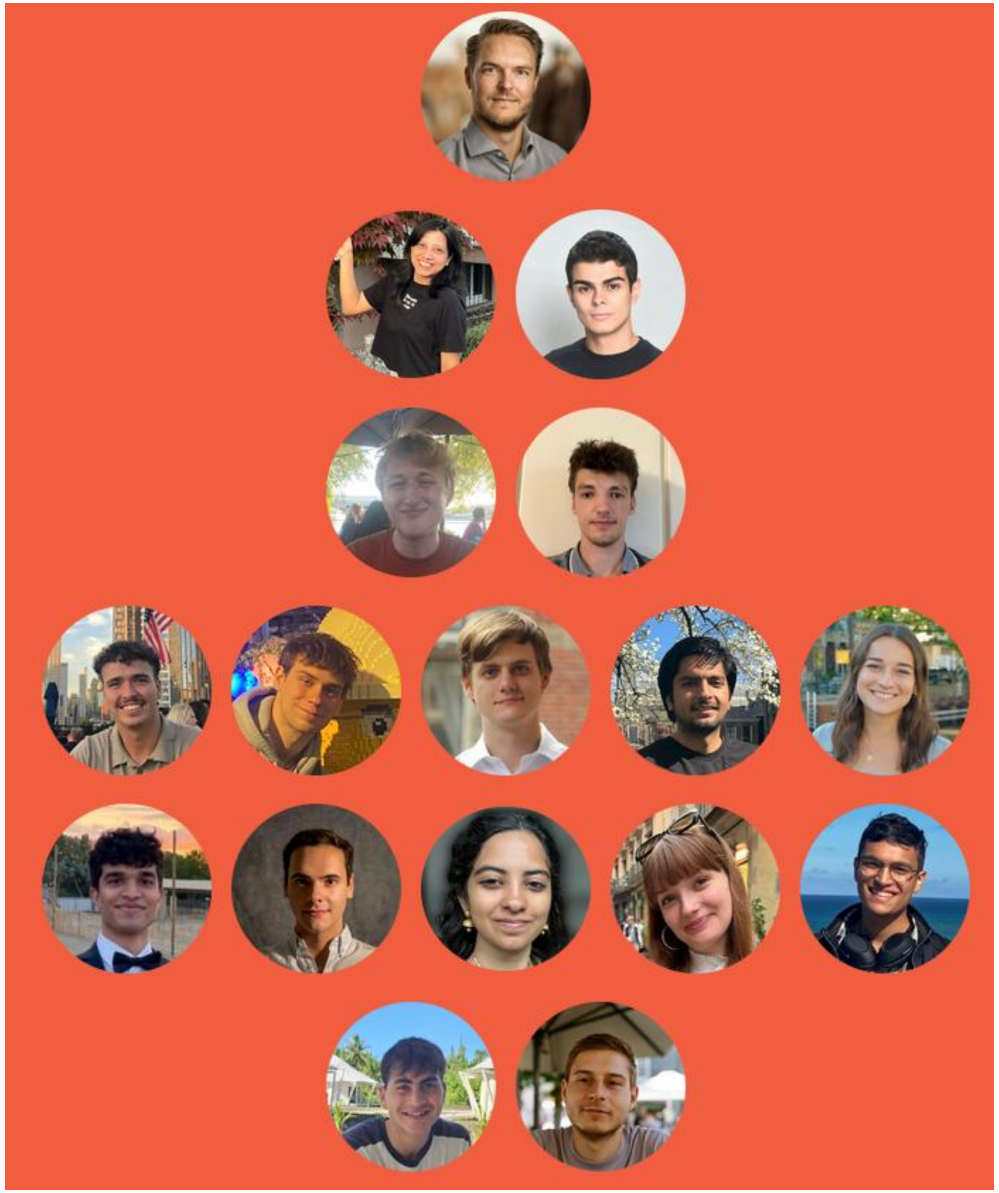
Our research mission

We question the geometric foundation of deep learning and we are leading a new direction of deep learning in hyperbolic space.





A warm welcome from our team



What is the course about?

Lecture	Title
1	Intro and history of deep learning
3	Deep learning optimization I
5	Convolutional deep learning
7	Graph deep learning
9	Multi-modal deep learning
11	What doesn't work in deep learning
13	Q&A

Lecture	Title
2	AutoDiff
4	Deep learning optimization II
6	Attention-based deep learning
8	From supervised to unsupervised deep learning
10	Generative deep learning
12	Non-Euclidean deep learning
14	Deep learning for videos

My goals with this course

Provide an intuitive foundation for understanding deep learning.

Challenge you to dive into the topic and learning by interacting and doing.

Give you the best starting point for following courses and the thesis.

Look at deep learning critically and understand what works and what doesn't.

Prerequisites

Machine Learning 1

Calculus and Linear Algebra

- Derivatives, integrals
- Matrix operations
- Computing lower bounds, limits

Some Probability Theory and Statistics

Advanced programming

Reading materials

Textbooks (most available online)

- Deep Learning by I. Goodfellow, Y. Bengio, A. Courville, 2016
- Dive Into Deep Learning, by A. Zhang, Z. Lipton, M. Li, A. Smola, 2019
- Neural Networks and Deep Learning: A Textbook, by C. Aggarwal, 2018
- https://udlbook.github.io/udlbook/ Understanding Deep Learning S.J.D. Prince 2022

Papers mentioned in the slides

At the end of each lecture, I list book chapters and papers to read

Tutorials

Alongside the lectures and the practicals, we have prepared tutorials.

In the tutorials we go step by step through fundamental concepts.

The tutorials include code, visualizations and explanations.

https://uvadlc-notebooks.readthedocs.io/

Held Tuesdays at 11:00-13:00.

Serves as extra service to you, to bridge theory and implementation.

At each tutorial, there will be some pen and paper questions: exam pre-practice.

Tutorial list

1: Intro to Snellius (!!!) and PyTorch

2: Activation functions + Tutorial 4: Optimization and Initialization

3: Inception, ResNet and DenseNet

4: Transformers and Multi-Head Attention

5: Graph Neural Networks

6: Self-Supervised Contrastive Learning with SimCLR

7: TBD

Exam and grading

Two parts: practicals (30%) + exam (70%).

Practicals grade = average grade of the three assignments.

Exam is classical written in-person assignment.

Practicals

Practical 1 [deadline: 9-11-2025]

Implement your own net modules, learn about core modules of deep learning.

Practical 2 [deadline: 30-11-2025]

Invariances in ConvNets, build your own nano-GPT, graph network theory.

Practical 3 [deadline: 14-12-2025]

Break networks with adversarial attacks and generative learning.

Vision of the practicals

Consists of both theory and implementation questions, handled through ANS.

Complementary view and information stream to the lectures.

Both a hands-on view to the lectures and its own reading materials!

On purpose with a view different interpretation and more theoretical background.

Snellius

We have secured a healthy amount of computation units for Deep Learning 1.

Tutorial 1 will cover the essentials of Snellius.

Lab 1 will also partially cover doing AI on clusters.

On grades

6

8

10

On boredom

Plagiarism (standard slide)

Plagiarism <u>**will not be**</u> tolerated

- Academic discussions are encouraged, and you can help each other
- What is plagiarism? Copying from each other is plagiarism. Sharing is plagiarism. Seeing someone else's code and retyping is plagiarism. Copying ideas & structure. Copying from existing GitHub accounts is plagiarism. Of course, re-using past answers is plagiarism. You devalue your own diploma, you make no good use of your money, and you take time we spend from teaching and reporting to do something we don't enjoy

Previously suspected cases

- Most got assignment nullified, some lost this/next year exam opportunities
- We check answer sheets from previous years, existing GitHubs and code repos, each other's answer sheets and other resources (we also use **plagiarism-detection software**)
- <u>Please, don't!</u>

Contact examiw-science@uva.nl for questions

Asking questions

1. During and after lectures, tutorials, and especially labs.
2. On Canvas Ed Discussions.
3. In case of technical/grade issues: contact Alejandro.
4. In case of personal/ethical/large-scale issues: contact me.

About Ed Discussions

Learning involves not understanding.

Asking questions is key to understanding.

Use Ed discussions to ask AND answer each other’s questions.

Multiples TAs will also check the platform regularly to answer questions.

Deep learning is an engineering science

This means:

- We do not have fundamental & general theories for lots of areas.
- Plenty of hands-on experience will lead to better development of intuition.

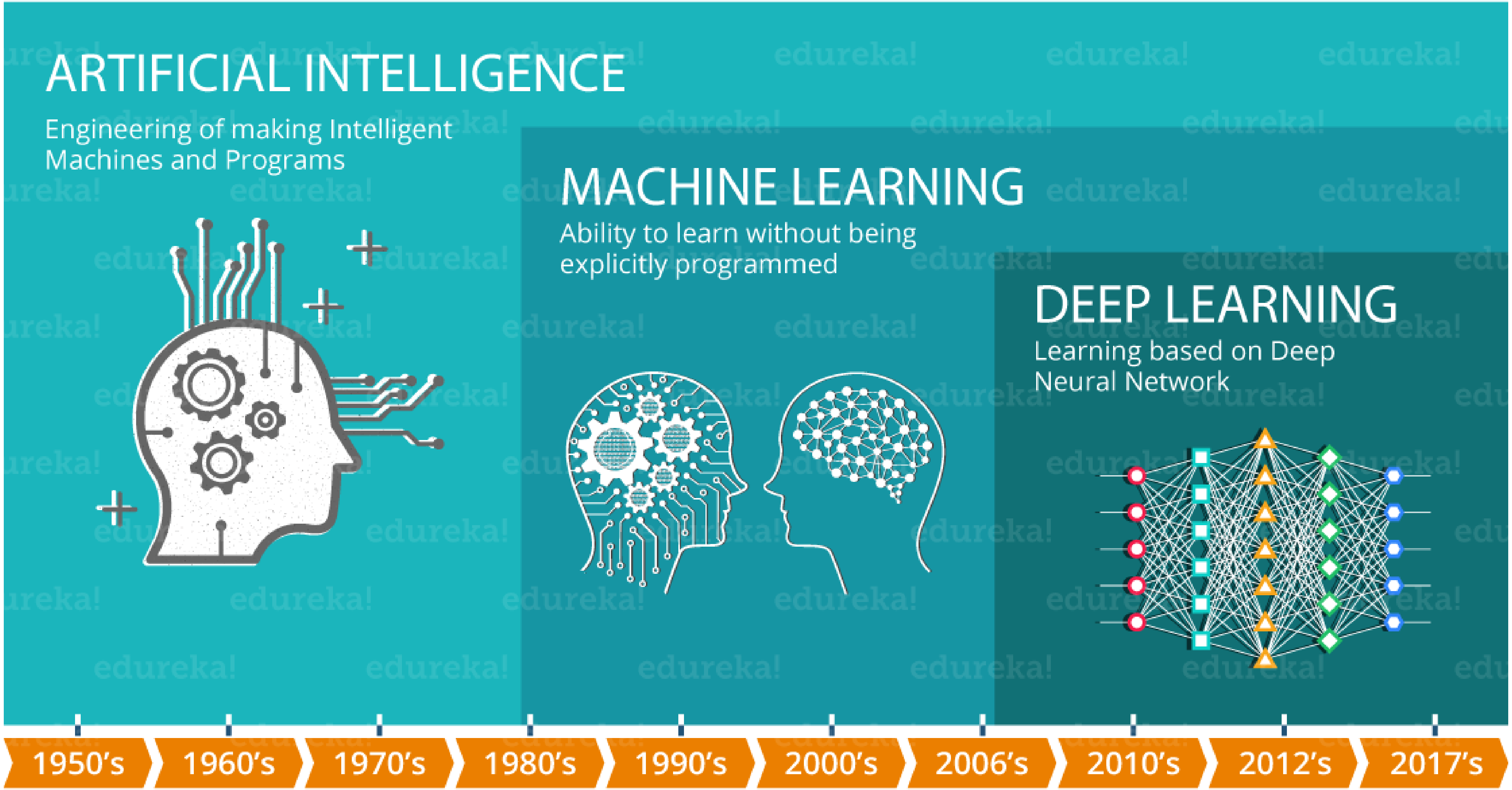
The field moves quickly.

Here we teach fundamentals, but some *will* be outdated very soon.

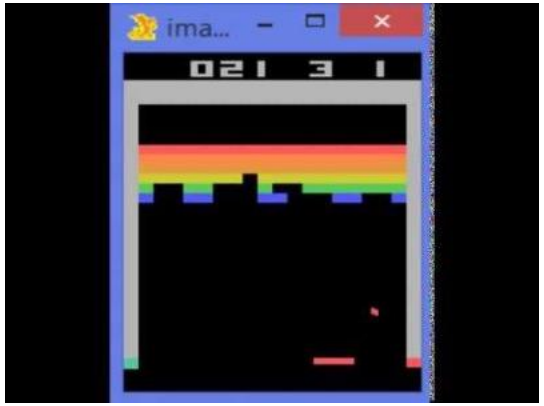
- Therefore, another goal of this course: develop tools/skillset to be able to independently delve into new topics.

Developing *intuition* is more important than memorizing.

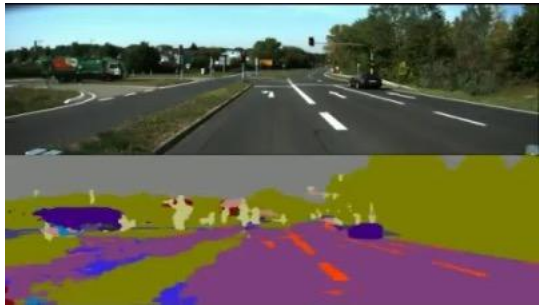
What is deep learning?



Highlights from first DL1 iteration (2016)



Playing Atari with Deep Reinforcement Learning. Mnih et al. 2013



SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. Badrinarayanan et al. 2015

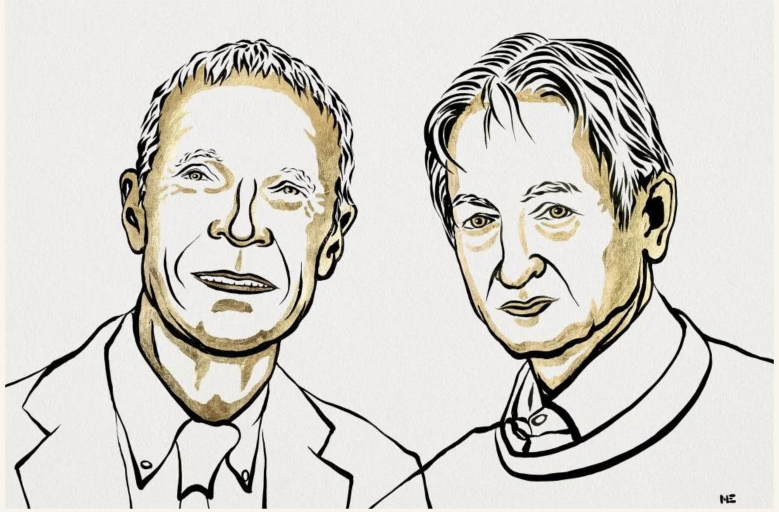
Large-scale Video Classification with Convolutional Neural Networks. Karpathy et al. 2014

https://github.com/karpathy/neuraltalk 2014

The 2024 physics laureates

The Nobel Prize in Physics 2024 was awarded to John J. Hopfield and Geoffrey E. Hinton "for foundational discoveries and inventions that enable machine learning with artificial neural networks."

John Hopfield created an associative memory that can store and reconstruct images and other types of patterns in data. Geoffrey Hinton invented a method that can autonomously find properties in data, and so perform tasks such as identifying specific elements in pictures.

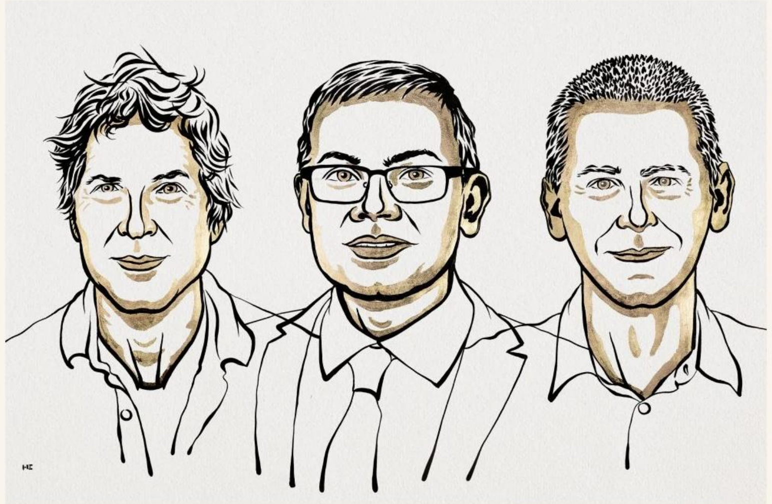


John Hopfield and Geoffrey Hinton. Ill. Niklas Elmehed © Nobel Prize Outreach

The 2024 chemistry laureates

The Nobel Prize in Chemistry 2024 was awarded with one half to David Baker "for computational protein design" and the other half jointly to Demis Hassabis and John M. Jumper "for protein structure prediction".

Demis Hassabis and John Jumper have successfully utilised artificial intelligence to predict the structure of almost all known proteins. David Baker has learned how to master life's building blocks and create entirely new proteins.



David Baker, Demis Hassabis and John Jumper. Ill. Niklas Elmehed © Nobel Prize Outreach

AI Holds Promise for Improving Diagnoses

With artificial intelligence, machines can see what many humans may have missed

BY NEIL PARMAR

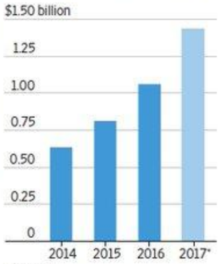
RADIOLOGISTS can scrutinize hundreds of images before identifying an area of concern in a patient's body. Mark Michalski likens the process to "playing 'Where's Waldo?' on each of those images."

But a type of artificial intelligence known as deep learning could soon help medical experts pinpoint problems faster and more accurately, says Dr. Michalski, executive director of the Boston-based Center for Clinical Data Science at Massachusetts General Hospital and Brigham and Women's Hospital.

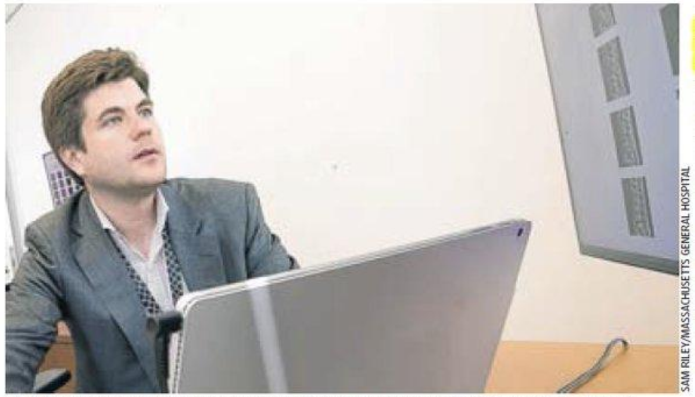
Deep learning includes algorithms, or computer programs, that search for, identify and an-

More Brainpower

Global spending on artificial intelligence and cognitive computing systems in health care

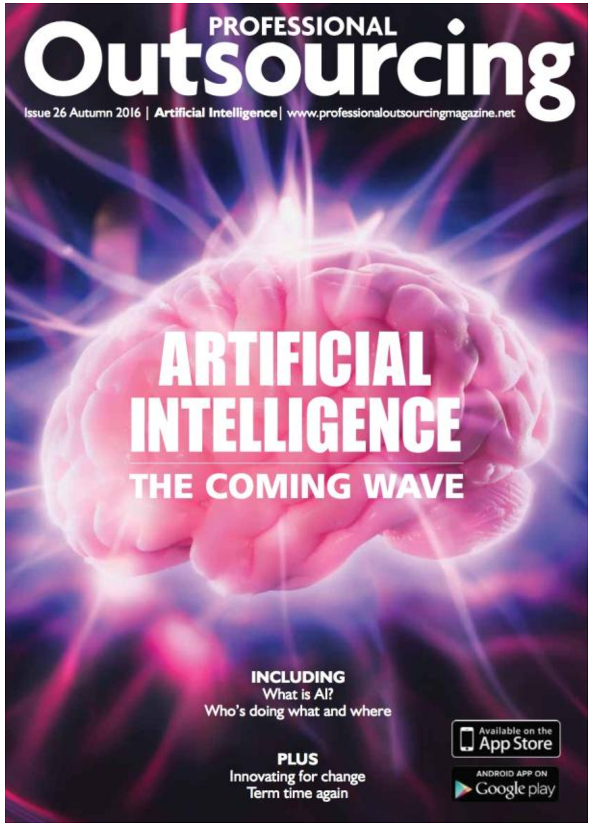


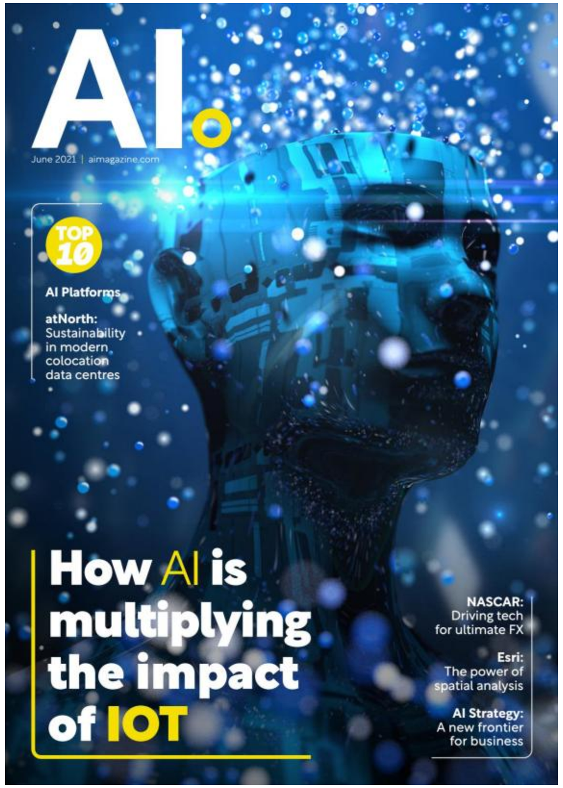
*Projected
Source: Frost & Sullivan
THE WALL STREET JOURNAL.



Mark Michalski works on a deep-learning algorithm for making diagnoses at Mass General.

SAM RILEY/MASSACHUSETTS GENERAL HOSPITAL





Artificial intelligence must not be allowed to hinder human rights

FROM time immemorial human beings have been interested in the world around them. Their curiosity has driven technological progress and led to a quality of life unimaginable even fifty years ago.

When my family arrived in Bryson Street in the Ballymacarrett-Short Strand area in 1963 there were three cars in the street of 50 houses, two telephones and a few televisions.

Forty years later members of my family had three cars in the driveway, several mobile phones in the house and a television in each room.

Technology has changed our lives – mostly, but not always – for the better.

The problem is not that technological advances are inherently bad. It is the way human beings use the advances which cause problems.

Nowhere can this be seen more clearly than in the threat to the planet's existence by global warming.

This existential threat forces us to review how humans should protect the planet and how they should change society, from an obsession with commodities as a benchmark of happiness, to a society which values human beings first and foremost.

It is around this notion of respect for human beings that a new society should be built, focused on human needs in areas such as health, education, economic security and human rights.

It is with this focus in mind that conversations about a new society, which harnesses human ingenuity and technological progress, should be conducted.

At this year's Féile an Phobal a fascinating lecture on the impact of rapid technological change, driven by developments in artificial intelligence (AI), examined the dilemma facing us today.

Unlike intelligence displayed by humans and animals AI is driven by computational power. It transforms input into output, and it advances

Jim
GIBNEY

automated and increasingly autonomous decision-making.

Dr Birgit Schippers, Senior Lecturer in Politics at St Mary's University College Belfast, spoke on the theme of 'Artificial Intelligence and the Challenges to Human Rights'.

Dr Schippers set the framework for the lecture by saying that it was "difficult to imagine one area of human life/activity that has not been touched by AI".

Of course, it is important to recognise the valuable contribution of AI across a broad spectrum: from global communication and global connection to medical diagnosis and treatment, environmental protection, manufacturing, education, social care and traffic control.

In our homes, our private lives and in the workplace, we use items such as mobile phones, robotic vacuum cleaners and lawnmowers, 'fit-bit' watches, smart cars and home security – tools that are propelled by artificial intelligence.

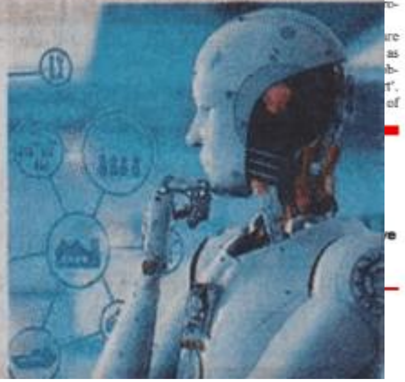
A principal concern is the growing shift from human decision-making to machine-generated "automated and autonomous" decisions

AI systems are also us
work. For example, they
publicise human rights
AI applications help t
and human rights activis
for justice and equality
borders.
But Dr Schippers emp
need to address AI's cap
undermine and threaten
A principal concern is
from human decision-ma
generated "automated a
decisions. It undermines
accountability and respo
contrary to the right not
automated decision mak
There is a real worry t
making capacity by intell
systems will be used to u
won progress of long and
Campaigns that fought fo
discrimination; the right
of expression, assembly
The use of AI applicati
boundaries between the
of democratic states and
interests of private comp
This is very worrying,
used in sensitive areas s
security and intelligence,
It is particularly concer
systems are used to unde
of democratic processes
One of the most contro
in the field of biometric t
They range from
cameras scanning c
facial images with th
recognition technol
emotional response
of a secondary scho
using facial recogni
monitor school atte
As Ms Schippers h
pointed out, human
technology are com
real need for robust and effective human rights protection.

AI workforce to boom in future

KANIZA GARARI | DC
HYDERABAD, MAY 21

The demand for Artificial Intelligence professionals is expected to be 2,38,000 in the next three years in India according to National Association of Software and Service Companies analysis released



The good, the bad, and the AV

From natural language and content creation, to advances in audio and the New Realities (augmented, virtual and mixed), AI in AV is here already, says *Amelia Kallman.*

At its core AI is meant to solve problems, using machines to perform tasks that would traditionally be too time-consuming or complex for humans. A subset of AI machine learning can help us make sense of large data pools, recognising patterns and user intent. Machine learning only becomes AI when the machines start to carry out tasks and interact with humans. Deep learning is a kind of machine learning

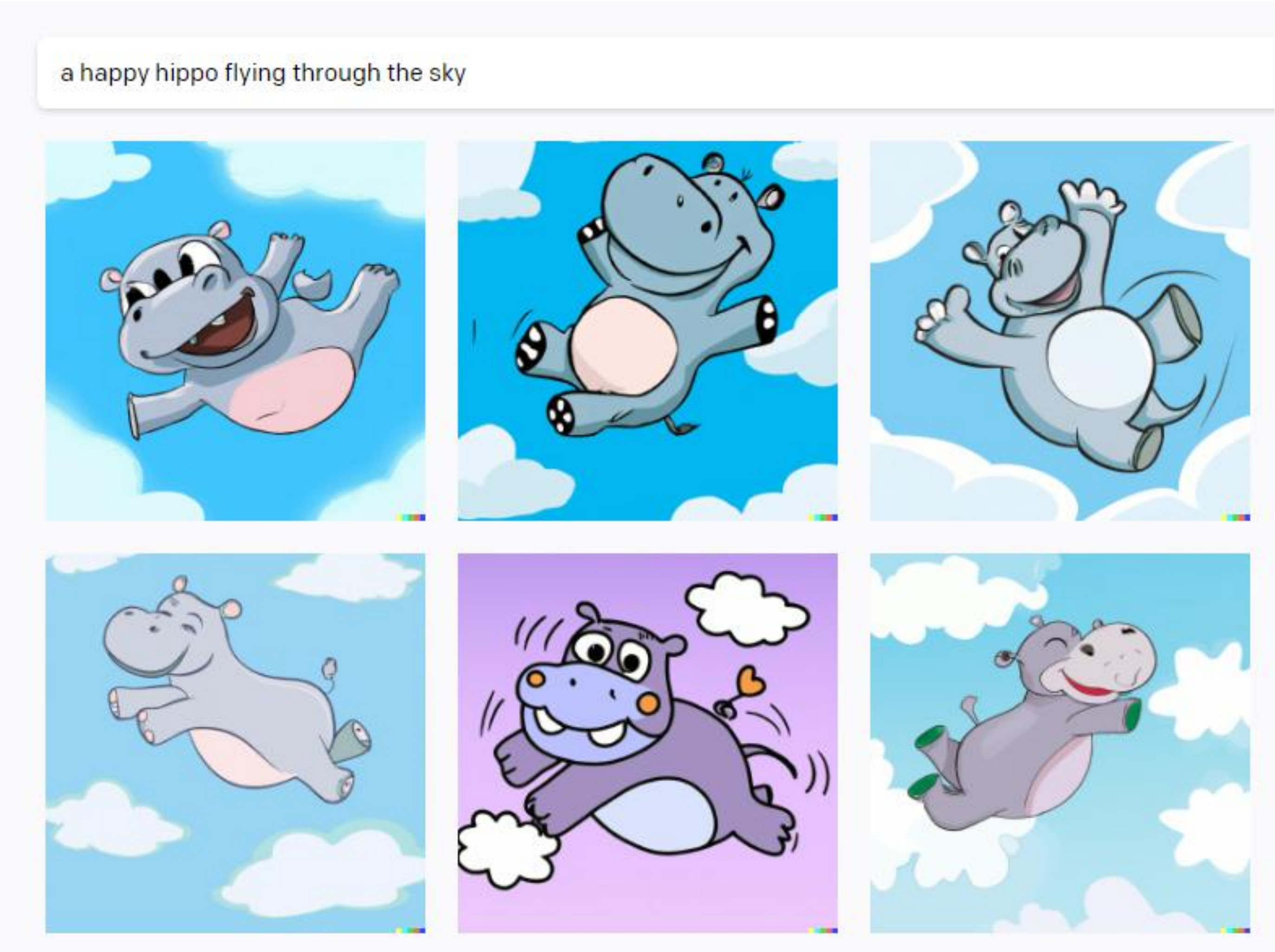
the future, today Alexa for Business is already allowing businesses to set up room-specific profiles that take the time and hassle out of things like room booking, setting up remote meetings, ordering supplies, and communicating alerts if something in the internet of things (IoT) is faltering or failing. Amazon hopes to create a continuously connected ecosystem, allowing people at work to connect Alexa for Business to Alexa at home and in your car. Cisco's Webex Teams Assistant, Microsoft's Cortana, Google Assistant, and Apple's Siri are also hoping to compete in the future of the smart conference room.

This high level of connectivity allows AI to get to know us and our business better than we know ourselves, predicting movements and preferences, being proactive and preemptive about supporting needs and desires.

However, with AI becoming ever more deeply integrated into IoT, this can also produce new problems and challenges. Every device connected into the IoT is a potential port of entry to access a whole system, meaning that the security of each connected device is vital to overall security. As enterprise networks start to expand from the office into the homes of employees, this could cause new vulnerabilities that many companies are not aware of or prepared for.

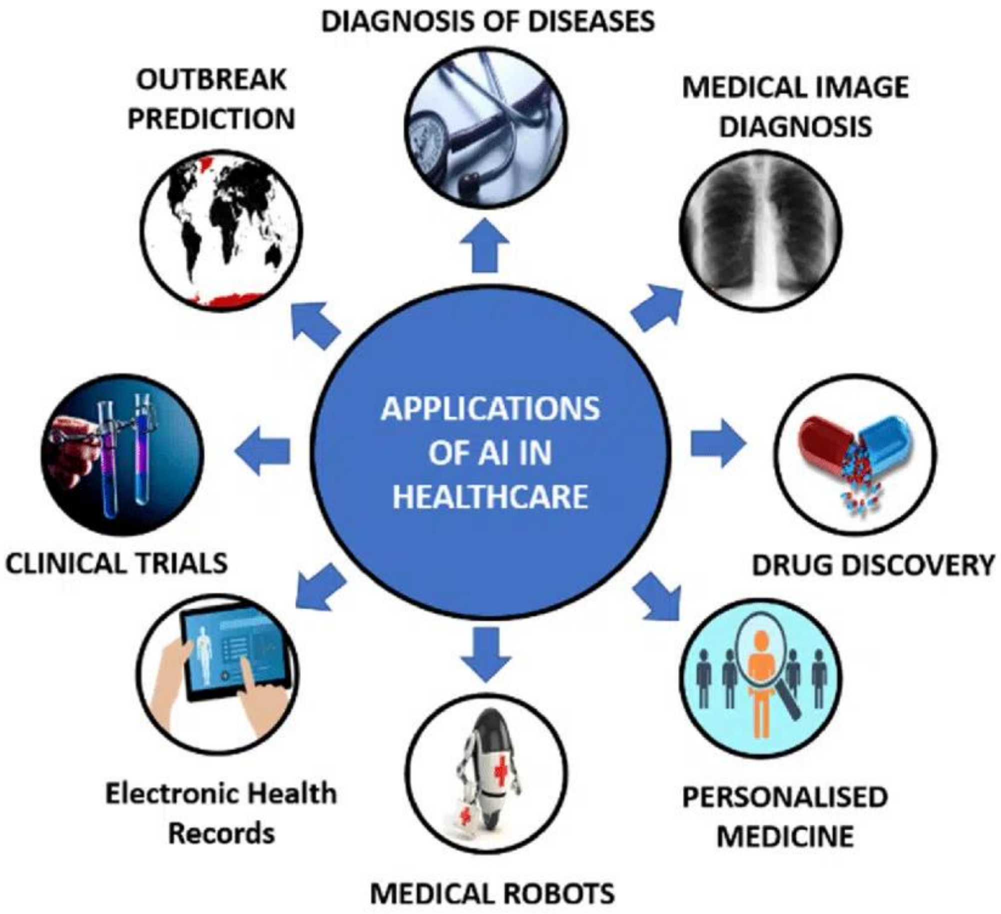
Taking the idea that a system is only as secure as its weakest link, this could result in unforeseen and increased costs to a company's cyber security infrastructure. Firewalls and data privacy are just a couple more obstacles that will need to be considered when integrating AI throughout businesses.

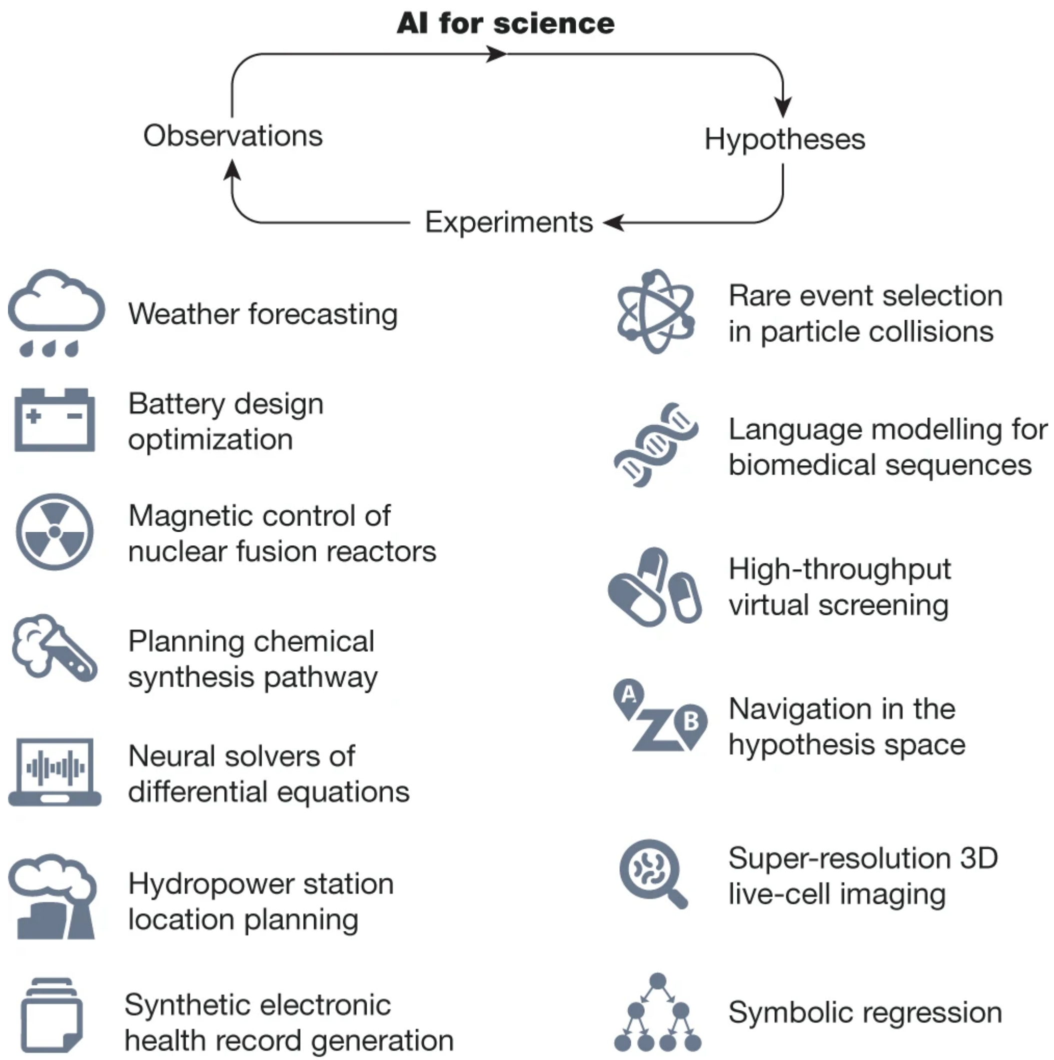
In retail we've been talking for years about how to bring together ecommerce and bricks-and-mortar, and AI integrated digital signage might be the key. Instead of pressing buttons to find your way, get price checks, or learn about complementing products or competitive offers, a person can simply talk to a screen and get an instant informed reply. AI has instant access to backroom stock, other store locations and their stock, as well as the ability to recognise customers, either through facial recognition »



DALL-E 2

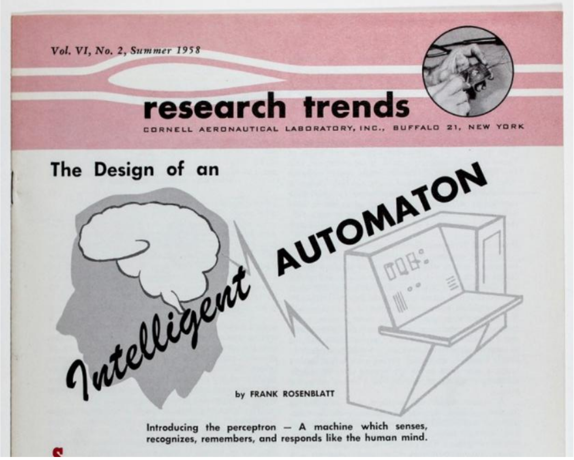

FIRST PLACE
FINE ARTS
COLORADO STATE FAIR
PUEBLO, CO
EMERGING
Jason Allen
Pueblo West
Théâtre D'opéra Spatial
$750
Colorado State Fair





Kareem Ullah – Medium June 6th 2023

https://ai4sciencecommunity.github.io/


Vol. VI, No. 2, Summer 1958

research trends

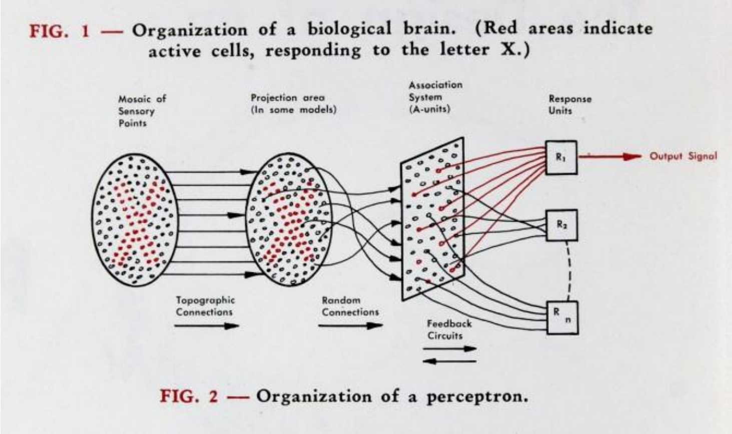
CORNELL AERONAUTICAL LABORATORY, INC., BUFFALO 21, NEW YORK

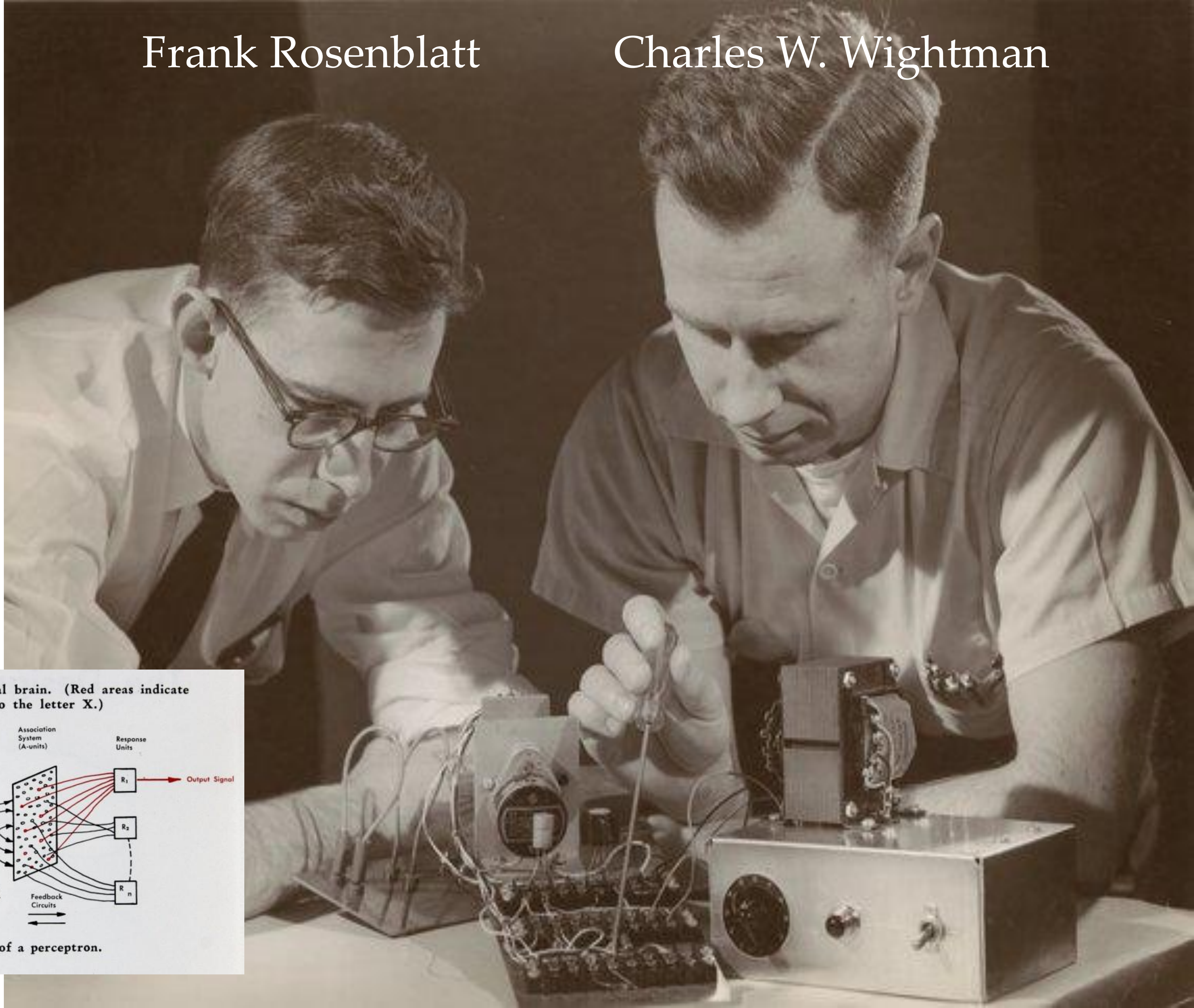
The Design of an Intelligent AUTOMATON

by FRANK ROSENBLATT

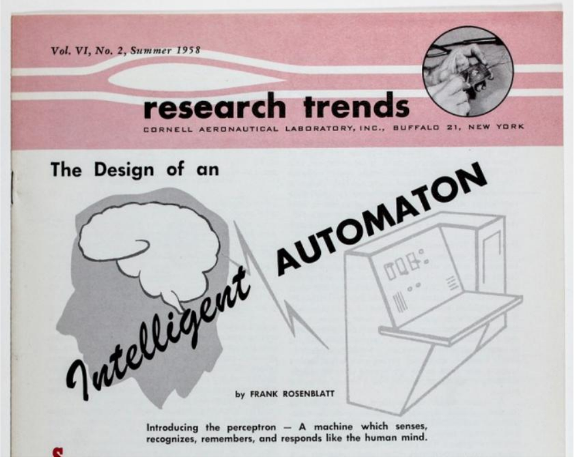
Introducing the perceptron — A machine which senses, recognizes, remembers, and responds like the human mind.

A brief history of deep learning

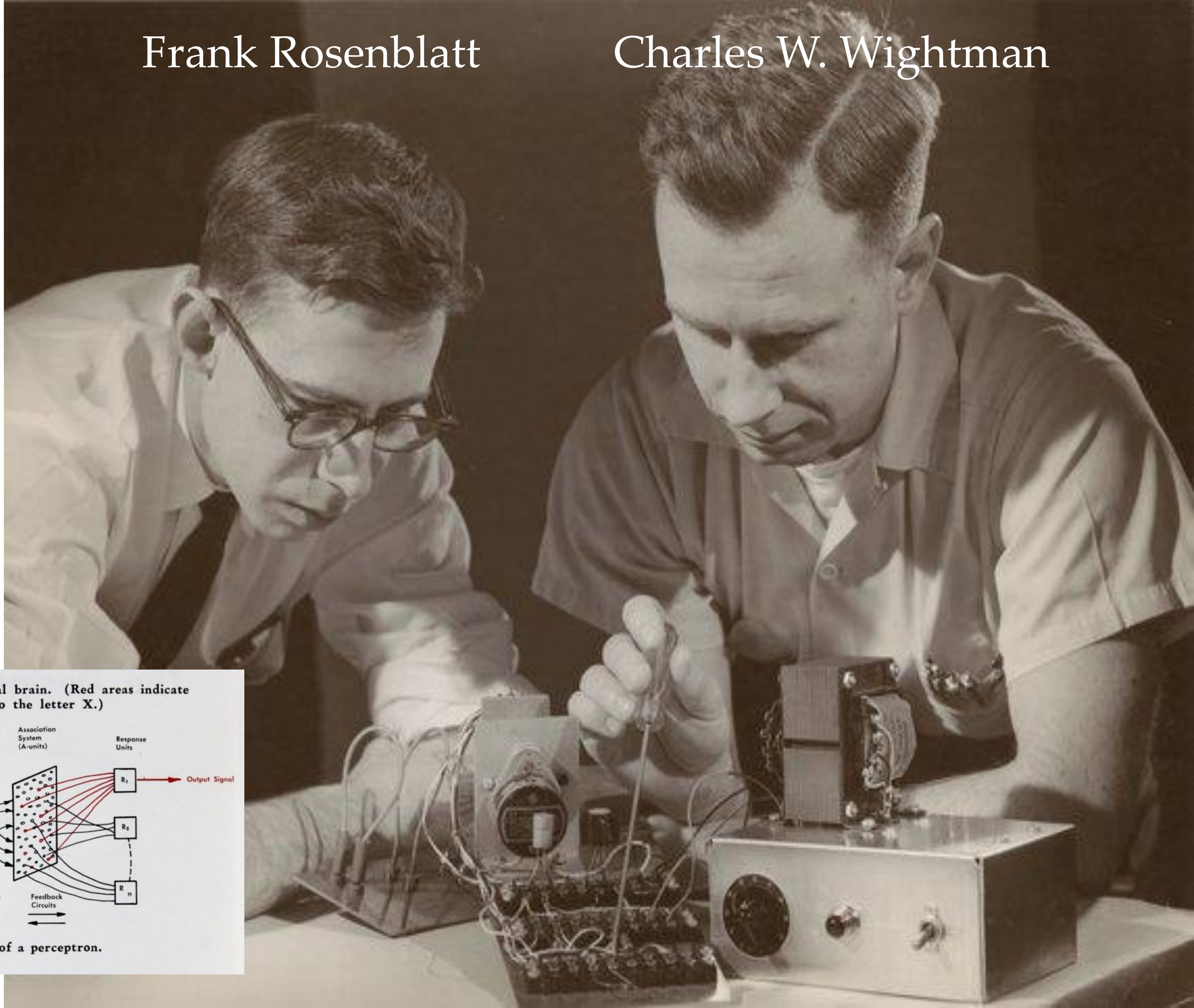


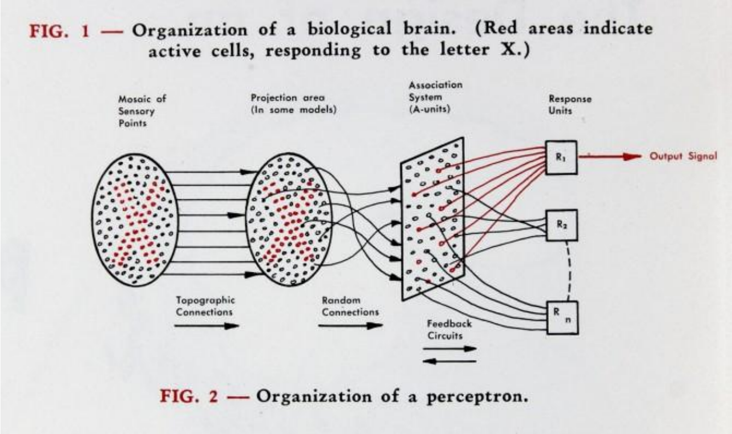


Break



A brief history of deep learning





Origins

1958 Perceptrons, Rosenblatt

1960 Adaline, Widrow and Hoff

1969 Perceptrons, Minsky and Papert

1970 Backpropagation, Linnainmaa

1974 Backpropagation, Werbos

1986 Backpropagation, Rumelhart, Hinton and Williams

1997 LSTM, Hochreiter and Schmidhuber

1998 OCR, LeCun, Bottou, Bengio and Haffner

2006 Deep Learning, Hinton, Osindero, Teh

2009 Imagenet, Deng et al.

2012 Alexnet, LeCun, Bottou, Bengio and Haffner

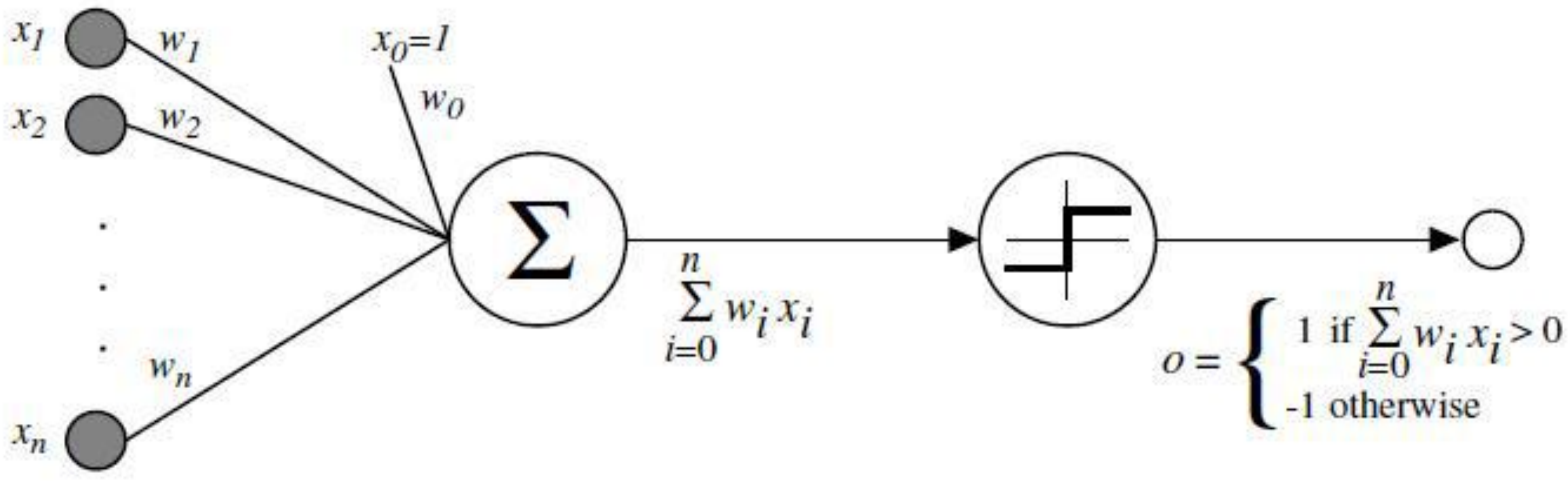
2015 Resnet (154 layers), MSRA

today GO, Deepmind

The perceptron

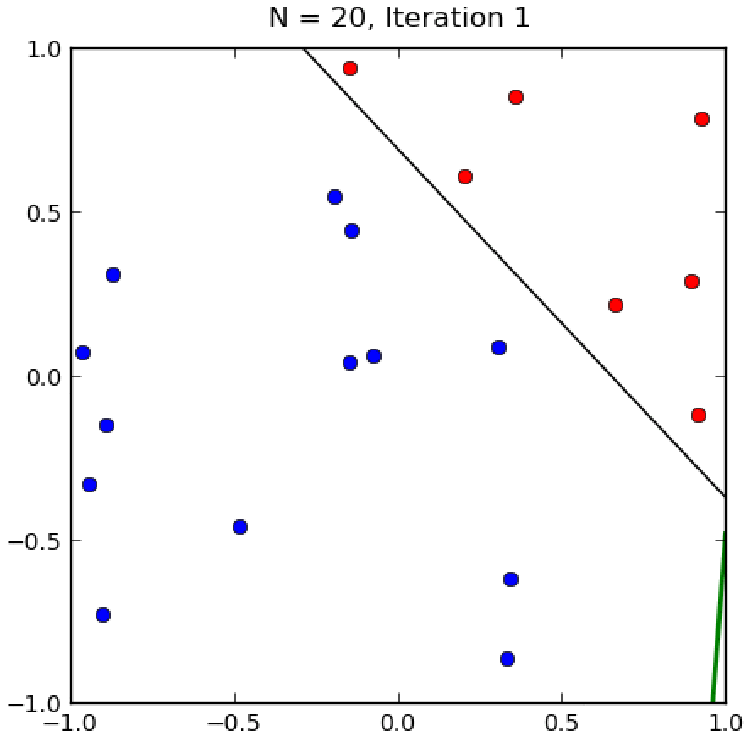
Single layer perceptron for binary classification.

- One weight w_i per input x_i
- Multiple each input with its weight, sum, and add bias
- If result larger than threshold, return 1, otherwise return 0

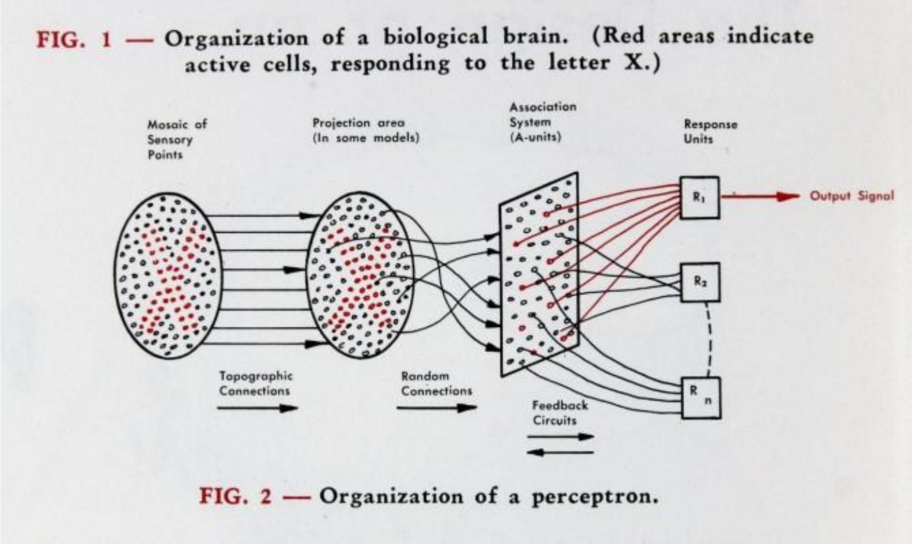


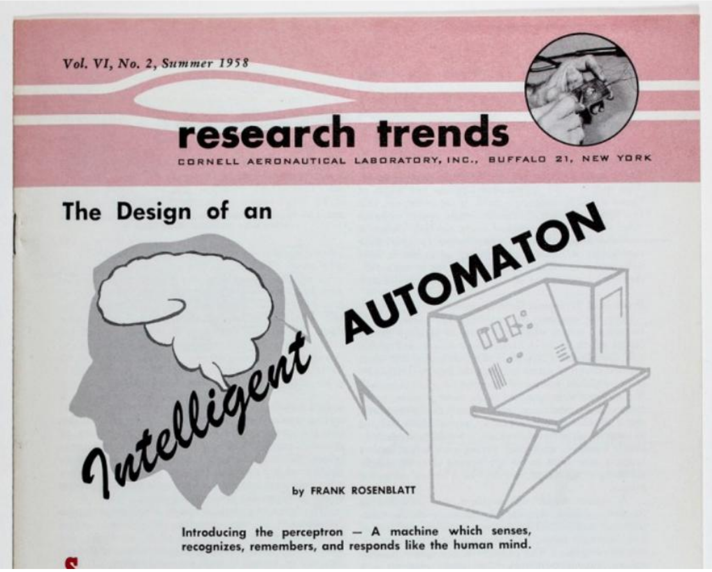
Training the perceptron

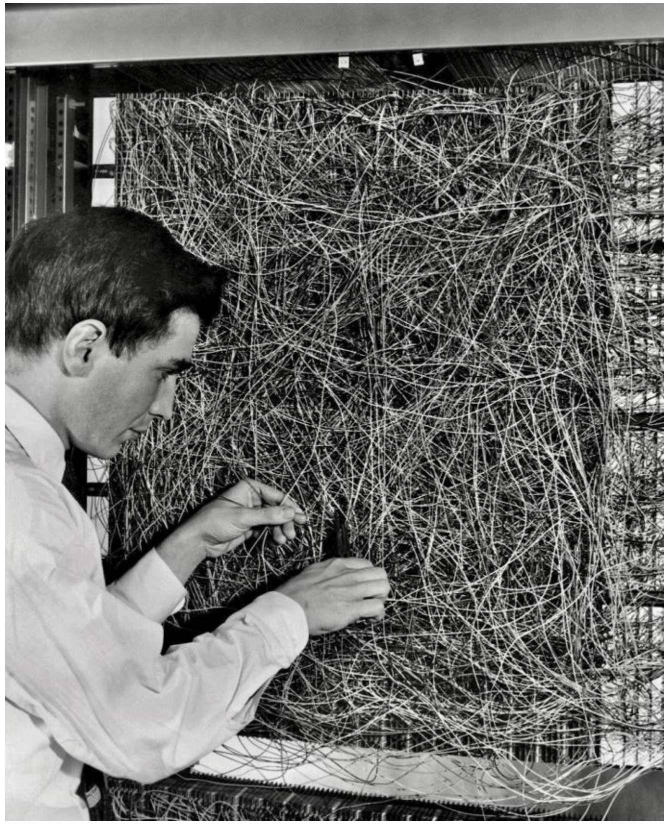
Perceptron learning algorithm	**Comments**
1. Set $w_j \leftarrow$ random	
2. Sample new (x_i, l_i)	New train image, label
3. Compute $y_i = [\![\sum w_i x_i > 0]\!]$	$[\![\cdot]\!]$: indicator function
4. If $y_i < 0, l_i > 0 \rightarrow w_i = w_i + \eta \cdot x_i$	Score too low. Increase weights!
5. If $y_i > 0, l_i < 0 \rightarrow w_i = w_i - \eta \cdot x_i$	Score too high. Decrease weights!
6. Go to 2	Repeat till happy ☺



Hyping up the perceptron



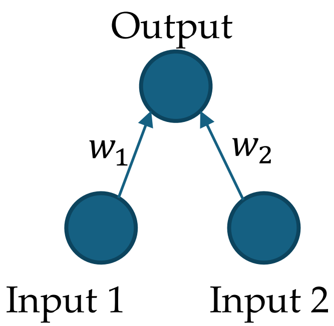




“a machine which senses, recognizes, remembers, and responds like a human mind”

XOR: The simplest hurdle

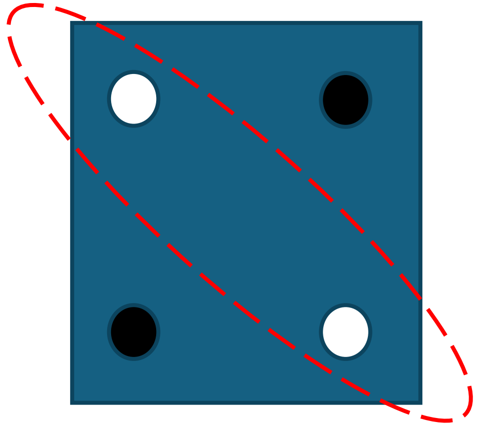
Input 1	Input 2	XOR
1	1	-1
1	0	+1
0	1	+1
0	0	-1



$$1 \cdot w_1 + 1 \cdot w_2 < \tau \rightarrow w_1 + w_2 < \tau$$
$$1 \cdot w_1 + 0 \cdot w_2 > \tau \rightarrow w_1 > \tau$$
$$0 \cdot w_1 + 1 \cdot w_2 > \tau \rightarrow w_2 > \tau$$
$$0 \cdot w_1 + 0 \cdot w_2 < \tau \rightarrow 0 < \tau$$

$$w_1 + w_2 > 2\tau$$
$$w_1 + w_2 < \tau$$

Inconsistent



No line can separate the white from the black

Moravec’s paradox

Reasoning requires little computation, perception from sensors a lot.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group — July 7, 1966

Vision Memo. No. 100.

THE SUMMER VISION PROJECT

Seymour Papert

The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

“AI Winter”

Perceptrons, Rosenblatt
Adaline, Widrow and Hoff
Perceptrons, Minsky and Papert
Backpropagation, Linnainmaa
Backpropagation, Werbos
Backpropagation, Rumelhart, Hinton and Williams
LSTM, Hochreiter and Schmidhuber
OCR, LeCun, Bottou, Bengio and Haffner
Deep Learning, Hinton, Osindero, Teh
Imagenet, Deng et al.
Alexnet, LeCun, Bottou, Bengio and Haffner
Resnet (154 layers), MSRA
GO, Deepmind
1958
1960
1969
1970
1974
1986
1997
1998
2006
2009
2012
2015
today

AI Winter

"If a perceptron cannot even solve XOR, why bother?"

Lack of funding due to overhyping.

Still, significant discoveries were made in this period

- Backpropagation → Learning algorithm for MLPs by Linnainmaa
- Recurrent networks → Varied-length inputs by Rumelhart
- CNNs → Neocognitron by Fukushima

The two paths of machine learning research

Path 1:

Fix perceptrons by making better features.

Path 2:

Fix perceptrons by making them more complex.

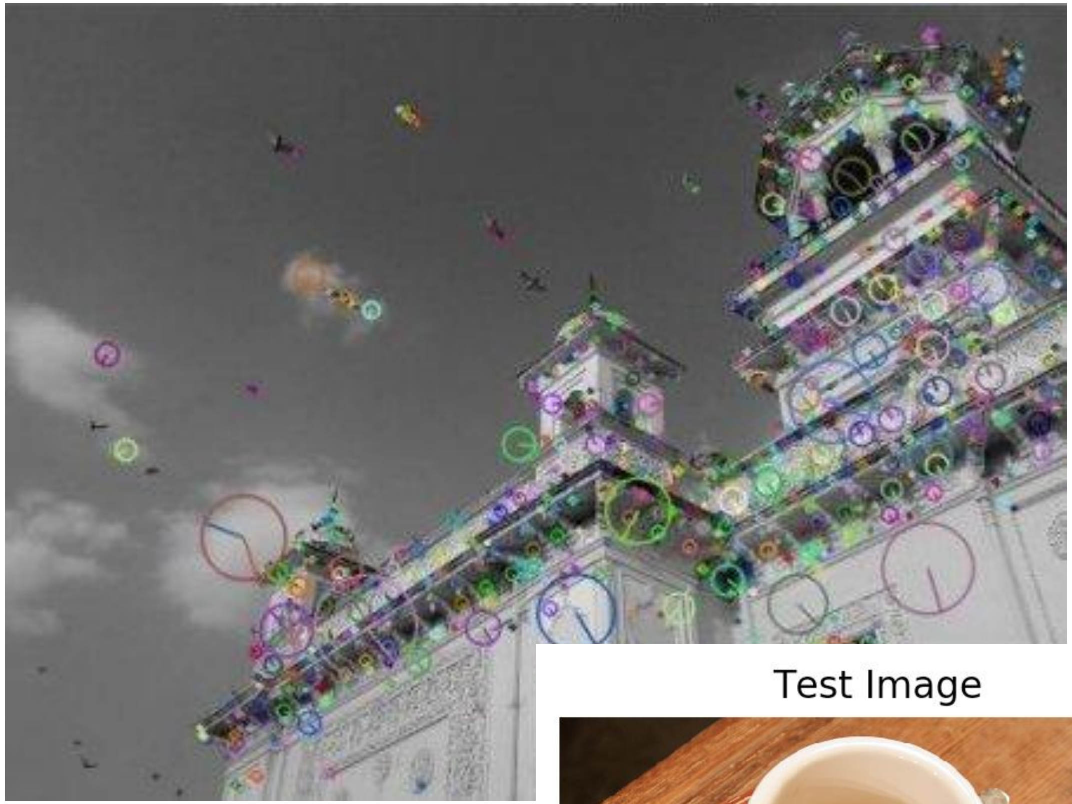
Path 1: Better inputs

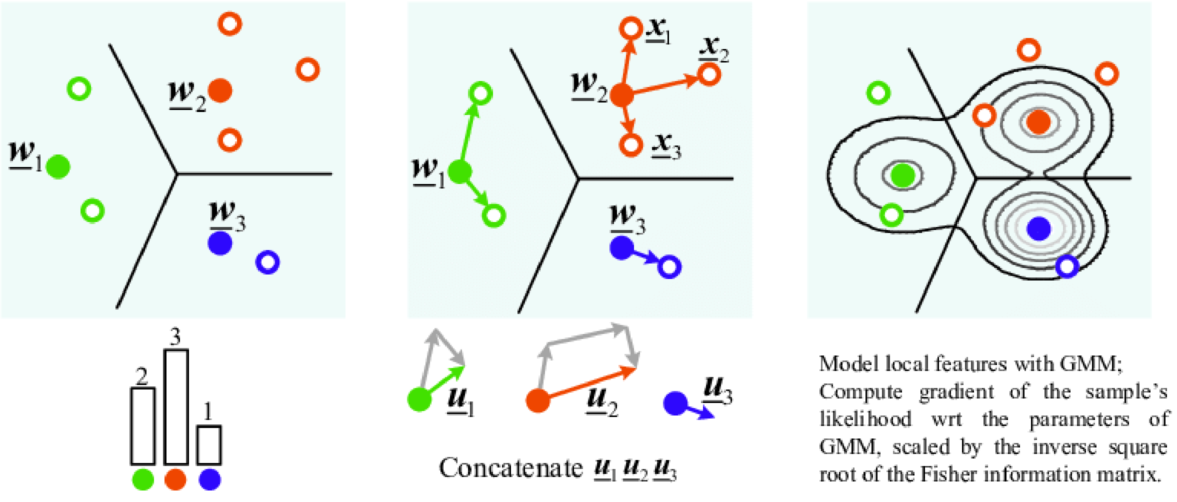
Philosophy: encode domain knowledge to help machine learning algorithms.

Classical image recognition pipeline:

1. Extract local features.
2. Aggregate local features over image.
3. Train classical models on aggregations.

Feature engineering: image domain

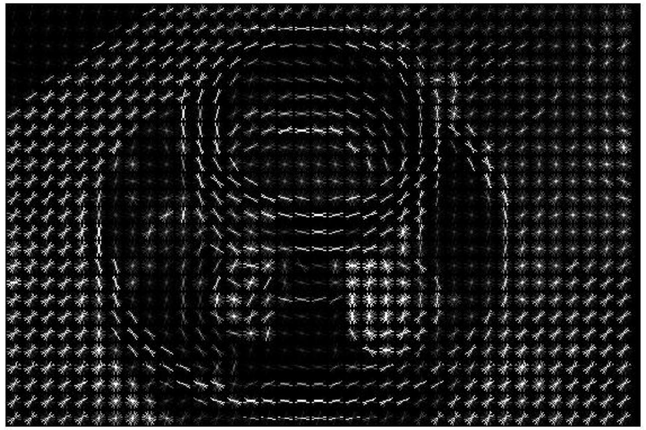




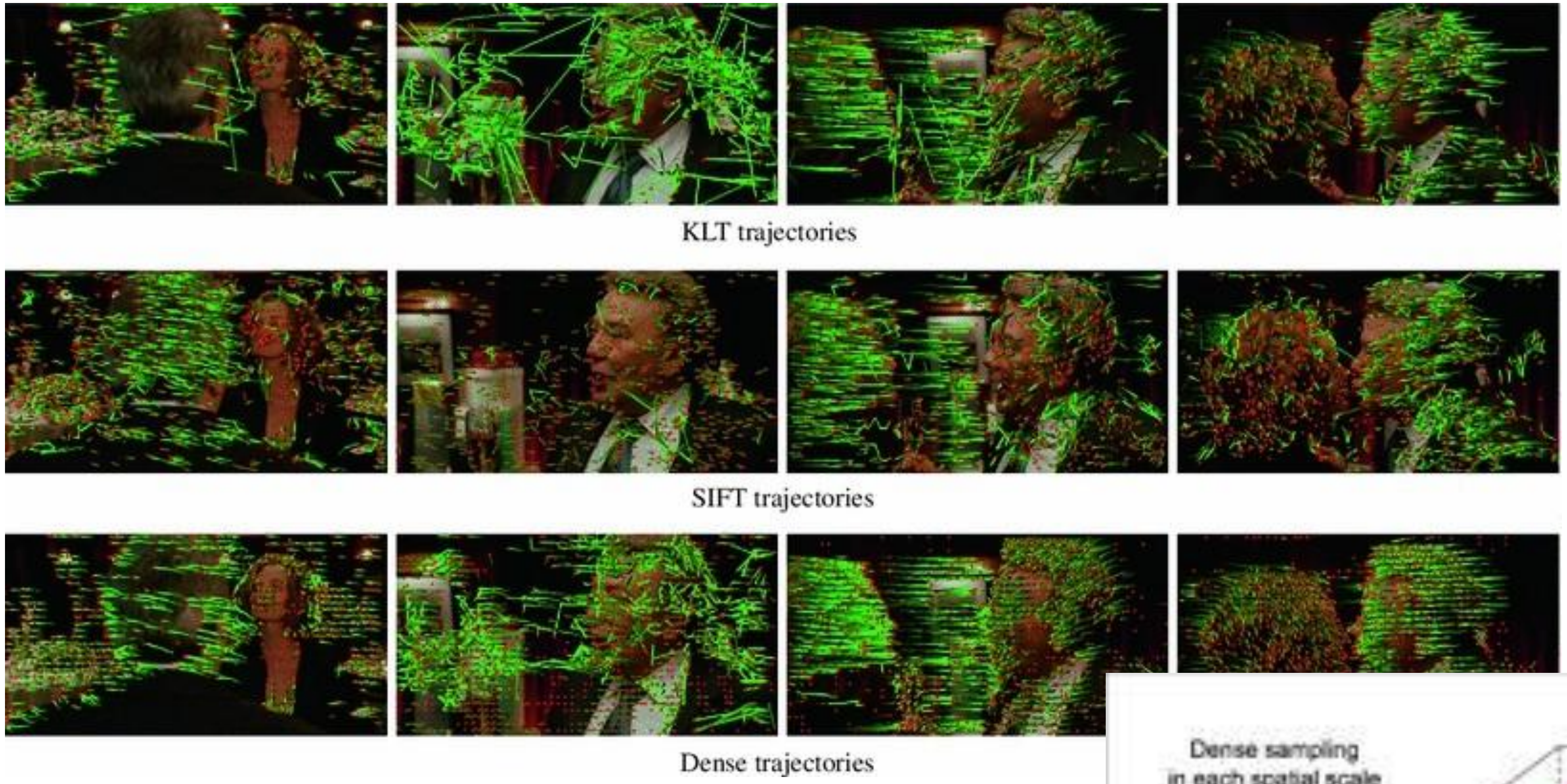
(a) BoW (b) VLAD (c) FV

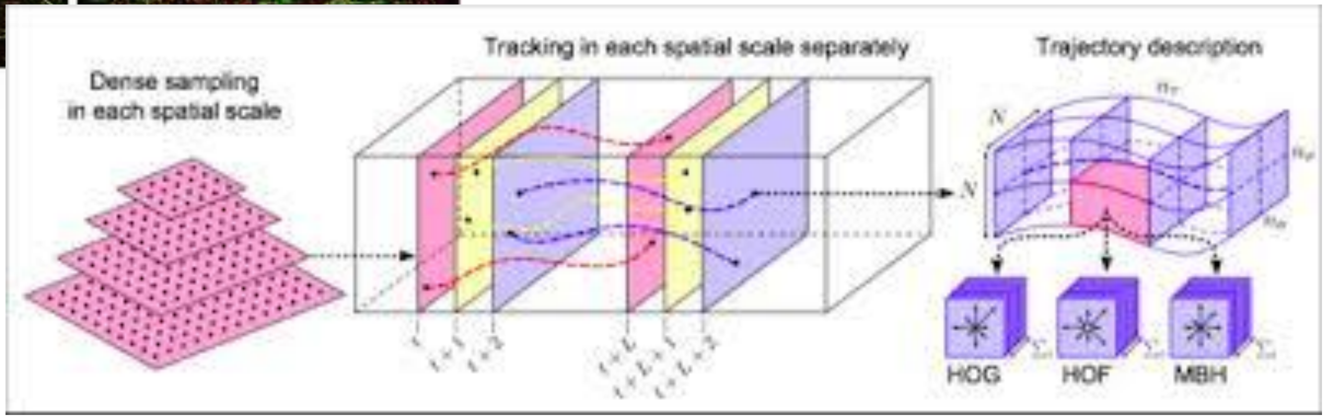
Test Image

Histogram of Oriented Gradients

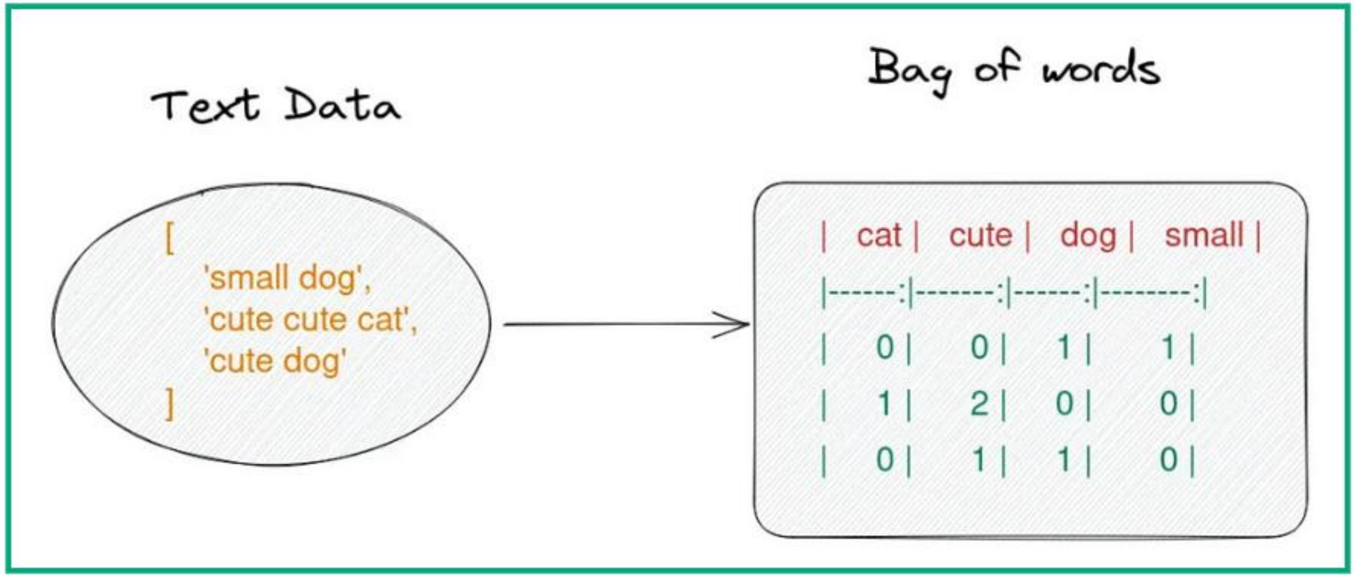


Feature engineering: video domain





Feature engineering: text domain



$$w_{x,y} = tf_{x,y} \times \log\left(\frac{N}{df_x}\right)$$

TF-IDF

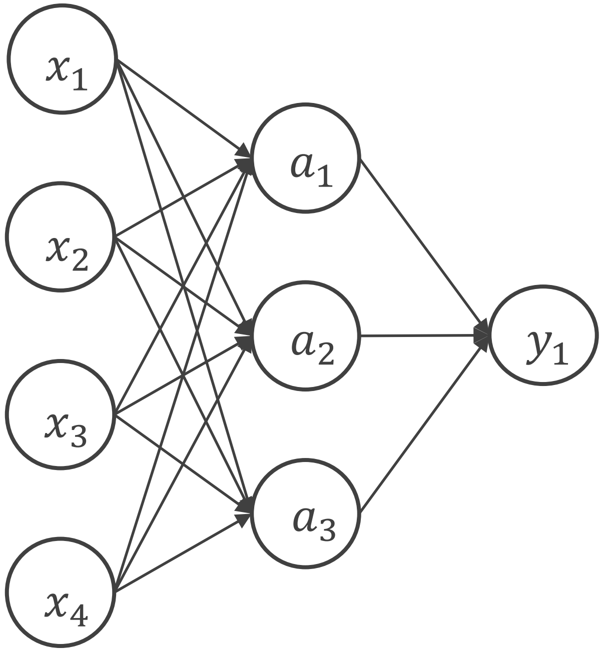
Term x within document y

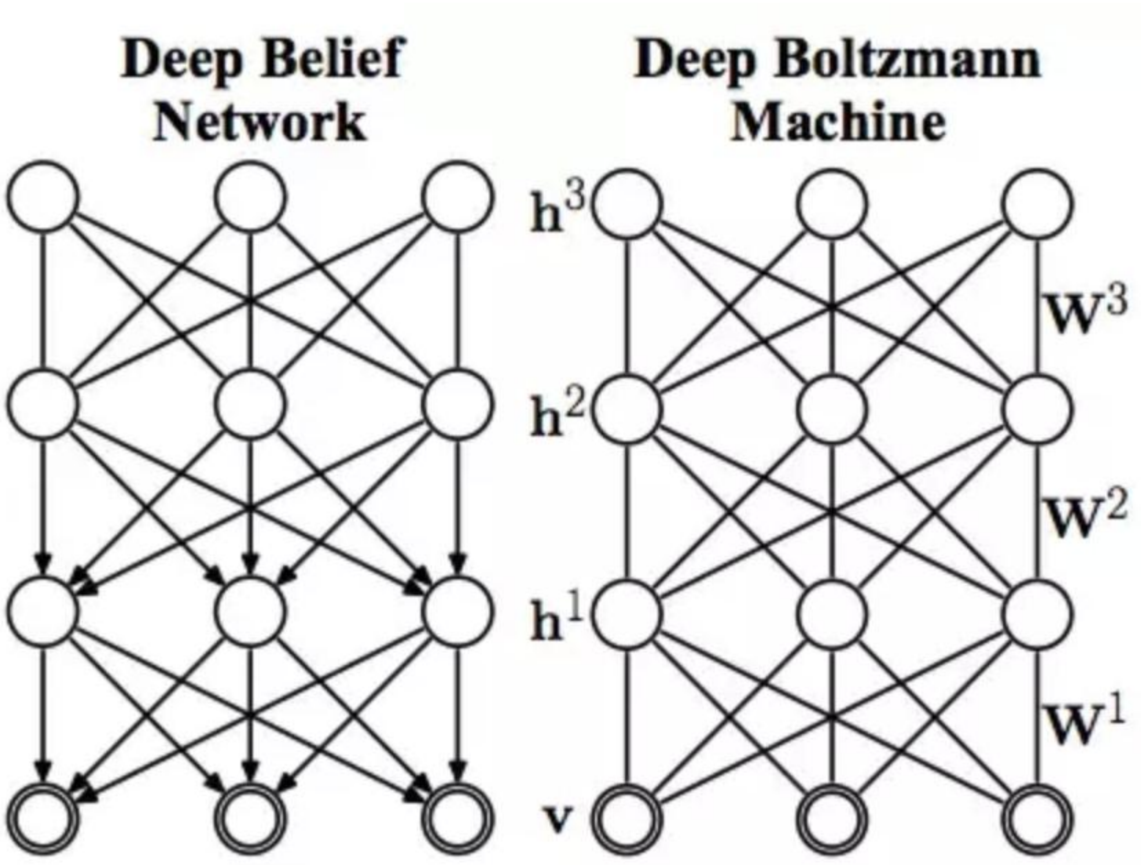
$tf_{x,y}$ = frequency of x in y

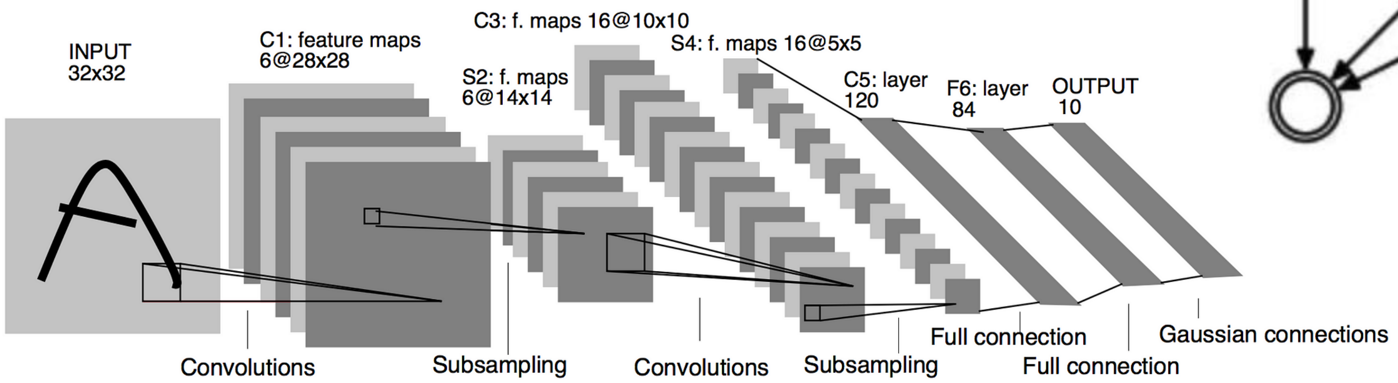
df_x = number of documents containing x

N = total number of documents

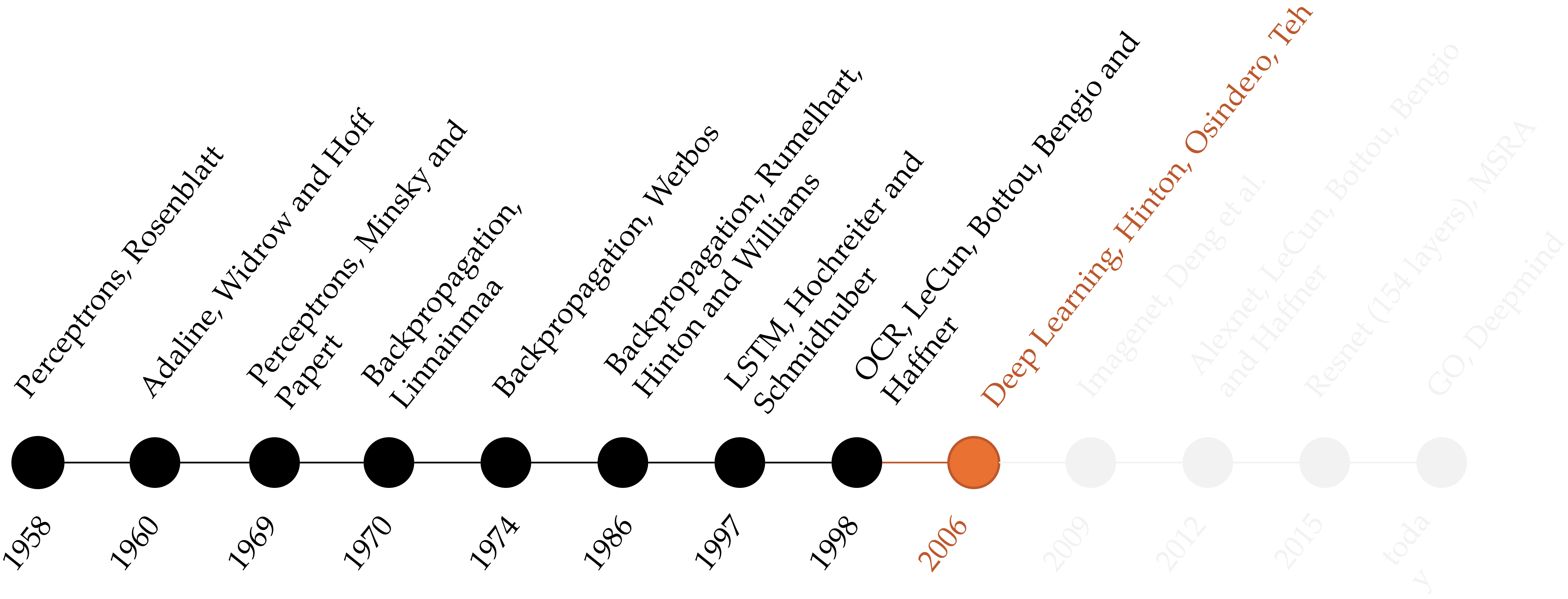
Path 2: Neural networks beyond a single layer







The thaw of AI Winter



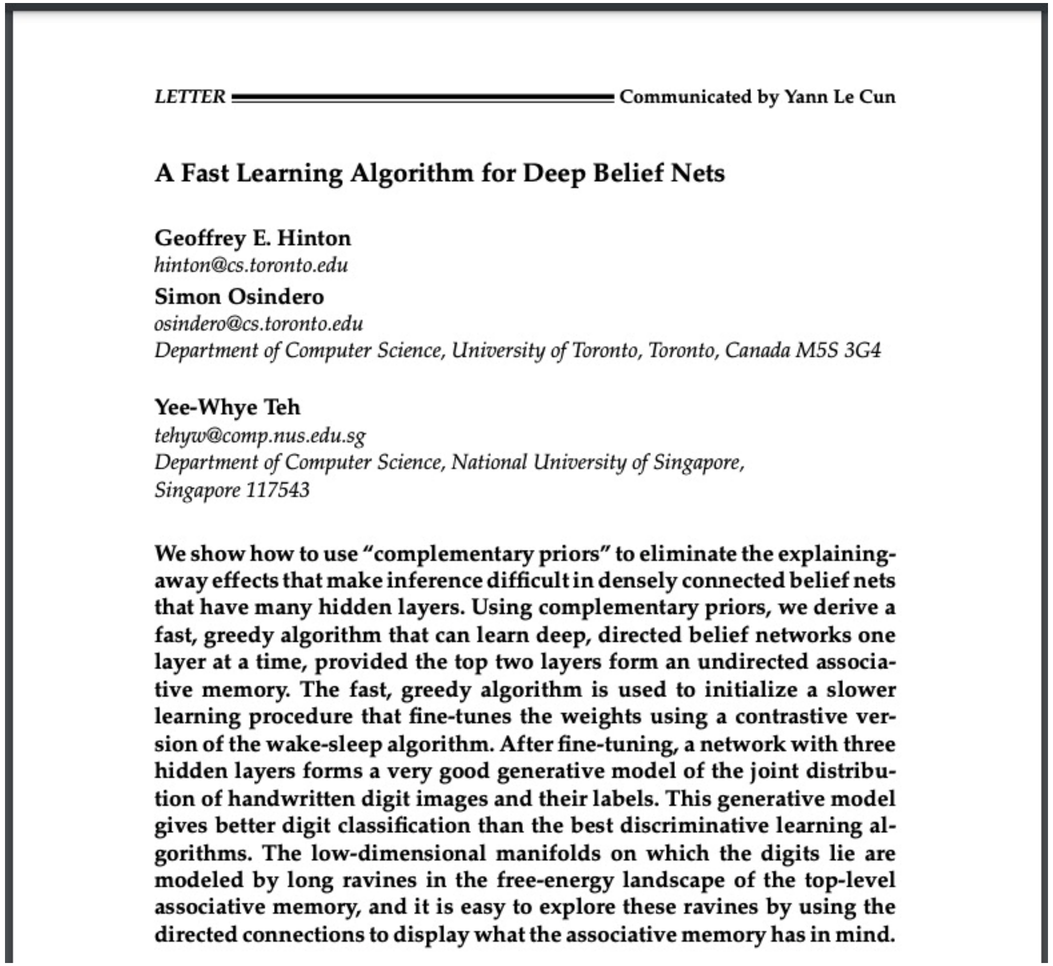
Signs of life of deep neural networks…

In 2006, Hinton and Salakhutdinov found multi-layer feedforward neural networks can be pretrained layer by layer.

Fine-tuned by backpropagation

Deep Belief Nets (DBNs)

(based on Boltzmann machines)



LETTER Communicated by Yann Le Cun

A Fast Learning Algorithm for Deep Belief Nets

Geoffrey E. Hinton
hinton@cs.toronto.edu
Simon Osindero
osindero@cs.toronto.edu
Department of Computer Science, University of Toronto, Toronto, Canada M5S 3G4

Yee-Whye Teh
tehyw@comp.nus.edu.sg
Department of Computer Science, National University of Singapore, Singapore 117543

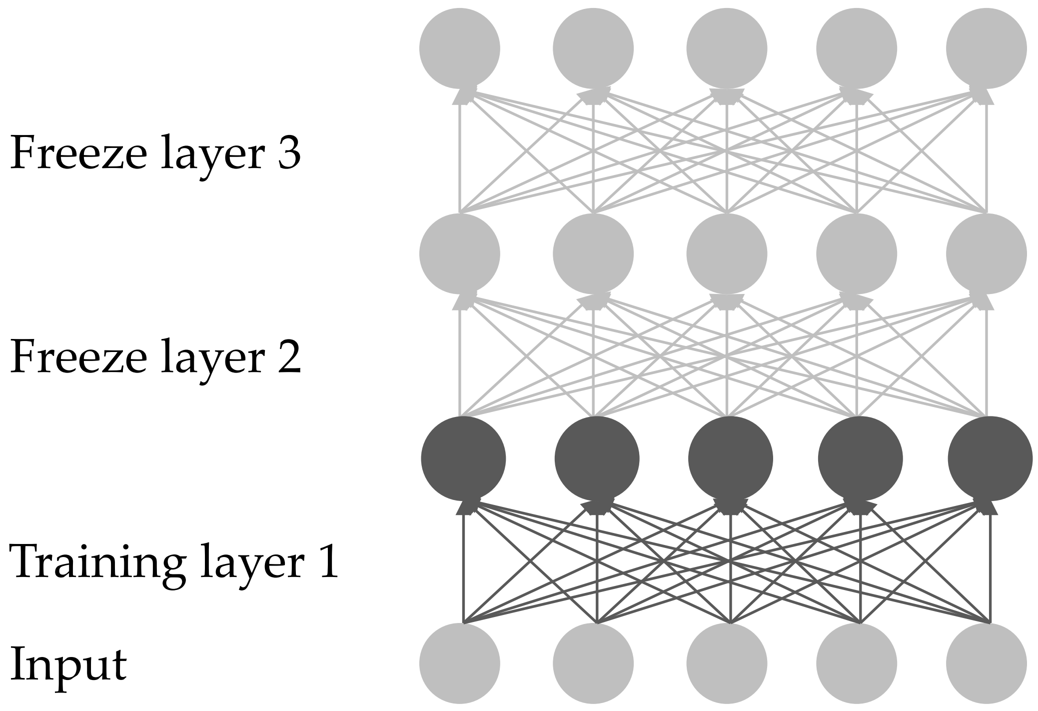
We show how to use "complementary priors" to eliminate the explaining-away effects that make inference difficult in densely connected belief nets that have many hidden layers. Using complementary priors, we derive a fast, greedy algorithm that can learn deep, directed belief networks one layer at a time, provided the top two layers form an undirected associative memory. The fast, greedy algorithm is used to initialize a slower learning procedure that fine-tunes the weights using a contrastive version of the wake-sleep algorithm. After fine-tuning, a network with three hidden layers forms a very good generative model of the joint distribution of handwritten digit images and their labels. This generative model gives better digit classification than the best discriminative learning algorithms. The low-dimensional manifolds on which the digits lie are modeled by long ravines in the free-energy landscape of the top-level associative memory, and it is easy to explore these ravines by using the directed connections to display what the associative memory has in mind.

Deep learning arrives

Easier to train one layer at a time → Layer-by-layer training

Training multi-layered neural networks became easier

Benefits of multi-layer networks, but single-layer easy of training

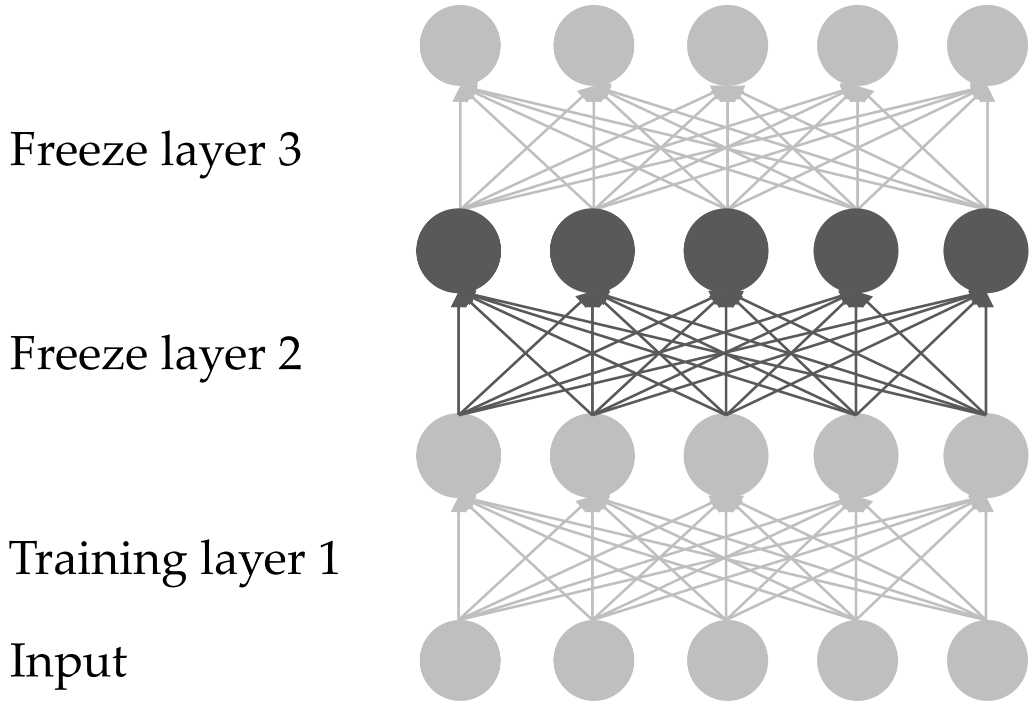


Deep learning arrives

Easier to train one layer at a time → Layer-by-layer training

Training multi-layered neural networks became easier

Benefits of multi-layer networks, but single-layer easy of training

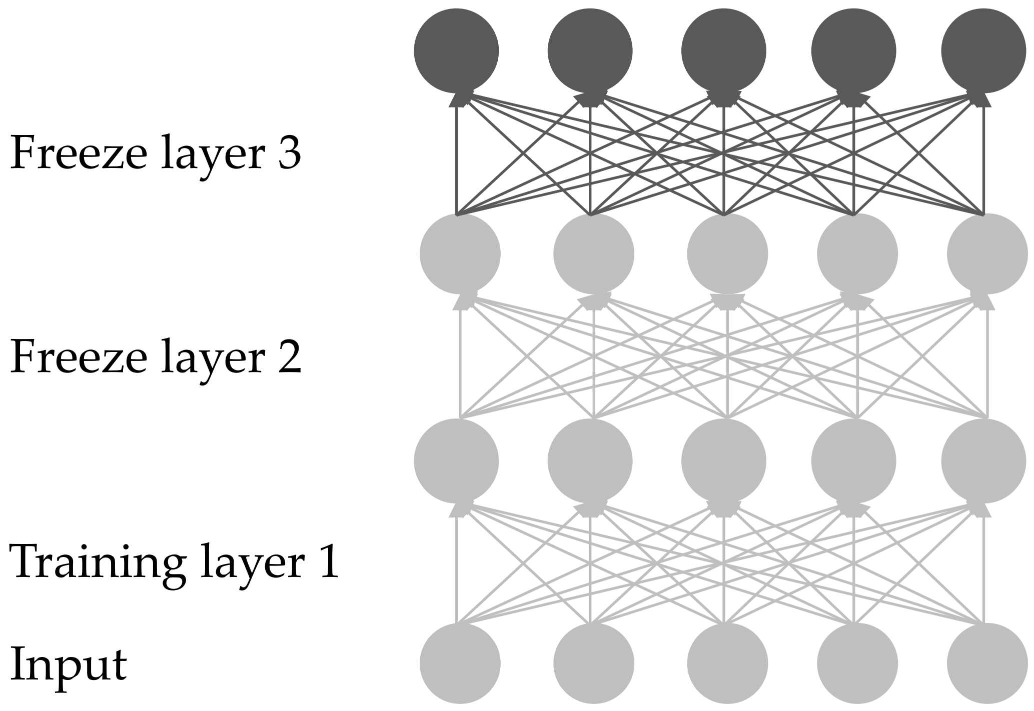


Deep learning arrives

Easier to train one layer at a time → Layer-by-layer training.

Training multi-layered neural networks became easier.

Benefits of multi-layer networks, but single-layer ease of training.



Challenges of neural networks then

Lack of processing power

Lack of data

Overfitting

Vanishing gradients

Experimentally, training multi-layer perceptrons was not that useful

"Are 1-2 hidden layers the best neural networks can do?"

…hence many skeptics remain

Do Deep Nets Really Need to be Deep?

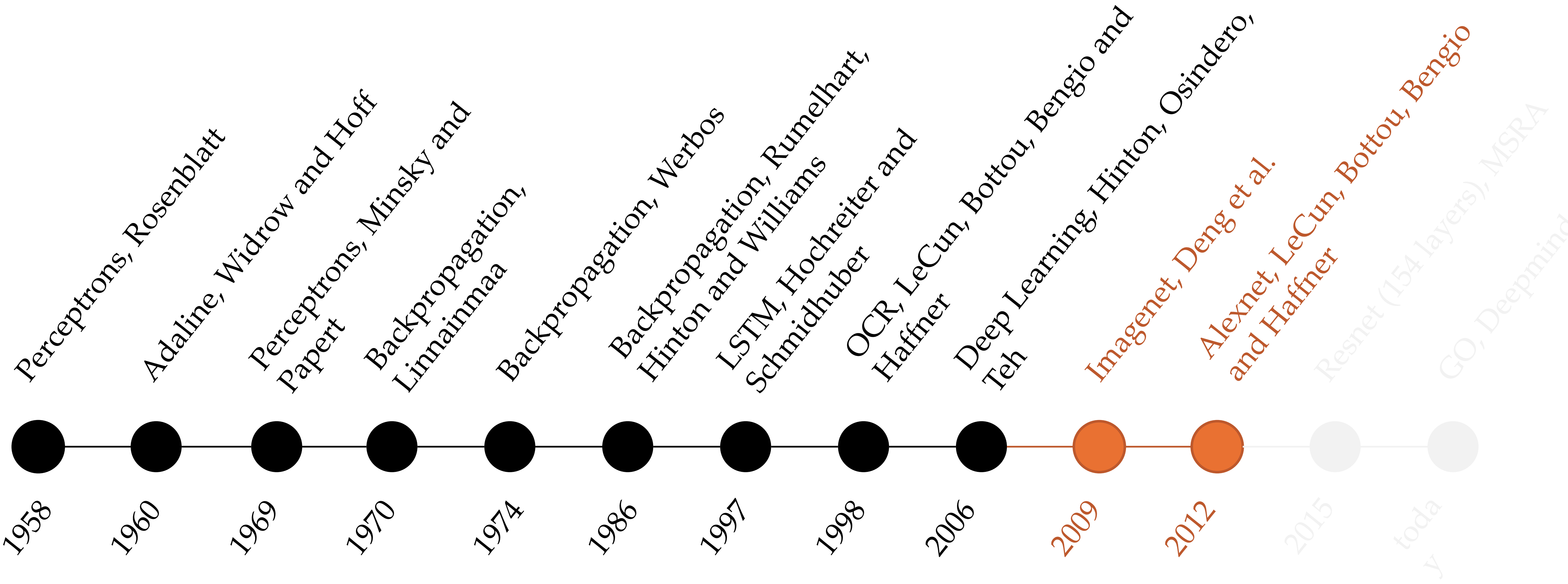
Lei Jimmy Ba
University of Toronto
jimmy@psi.utoronto.ca

Rich Caruana
Microsoft Research
rcaruana@microsoft.com

Deep Nets: What have They Ever Done for Vision?

Alan L. Yuille[1] · Chenxi Liu[1]

Deep learning arrives



Turns out that deep learning is data hungry

In 2009 the ImageNet dataset was published [Deng et al., 2009]

- Collected images for all 100K terms in Wordnet (16M images in total)
- Terms organized hierarchically: “Vehicle”→“Ambulance”

ImageNet Large Scale Visual Recognition Challenge (ILSVRC)

- 1 million images, 1,000 classes, top-5 and top-1 error measured

2012 Teams	%error
Supervision (Toronto)	15.3
ISI (Tokyo)	26.1
VGG (Oxford)	26.9
XRCE/INRIA	27.0
UvA (Amsterdam)	29.6
INRIA/LEAR	33.4

2013 Teams	%error
Clarifai (NYU spinoff)	11.7
NUS (singapore)	12.9
Zeiler-Fergus (NYU)	13.5
A. Howard	13.5
OverFeat (NYU)	14.1
UvA (Amsterdam)	14.2
Adobe	15.2
VGG (Oxford)	15.2
VGG (Oxford)	23.0

2014 Teams	%error
GoogLeNet	6.6
VGG (Oxford)	7.3
MSRA	8.0
A. Howard	8.1
DeeperVision	9.5
NUS-BST	9.7
TTIC-ECP	10.2
XYZ	11.2
UvA	12.1

CNN based, non-CNN based

ImageNet 2012 winner: AlexNet

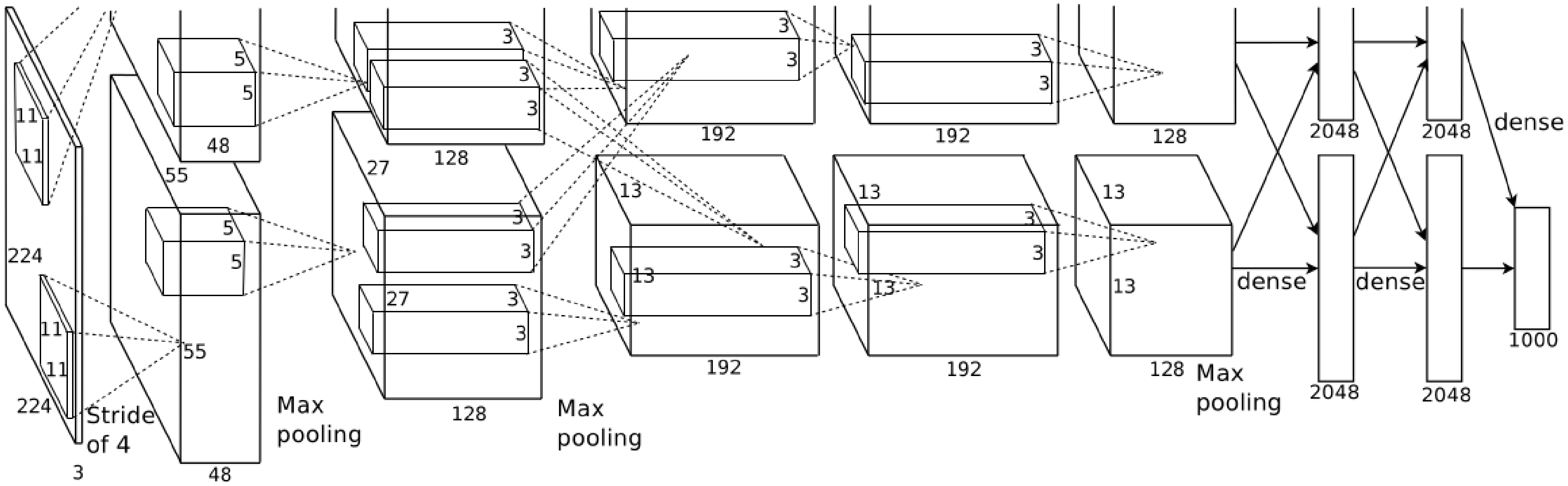


Figure 2: An illustration of the architecture of our CNN, explicitly showing the delineation of responsibilities between the two GPUs. One GPU runs the layer-parts at the top of the figure while the other runs the layer-parts at the bottom. The GPUs communicate only at certain layers. The network's input 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.

ImageNet

Most commonly used version: ImageNet-12: 1K categories, ~1.3M images, ~150GB. From Fei-Fei Li

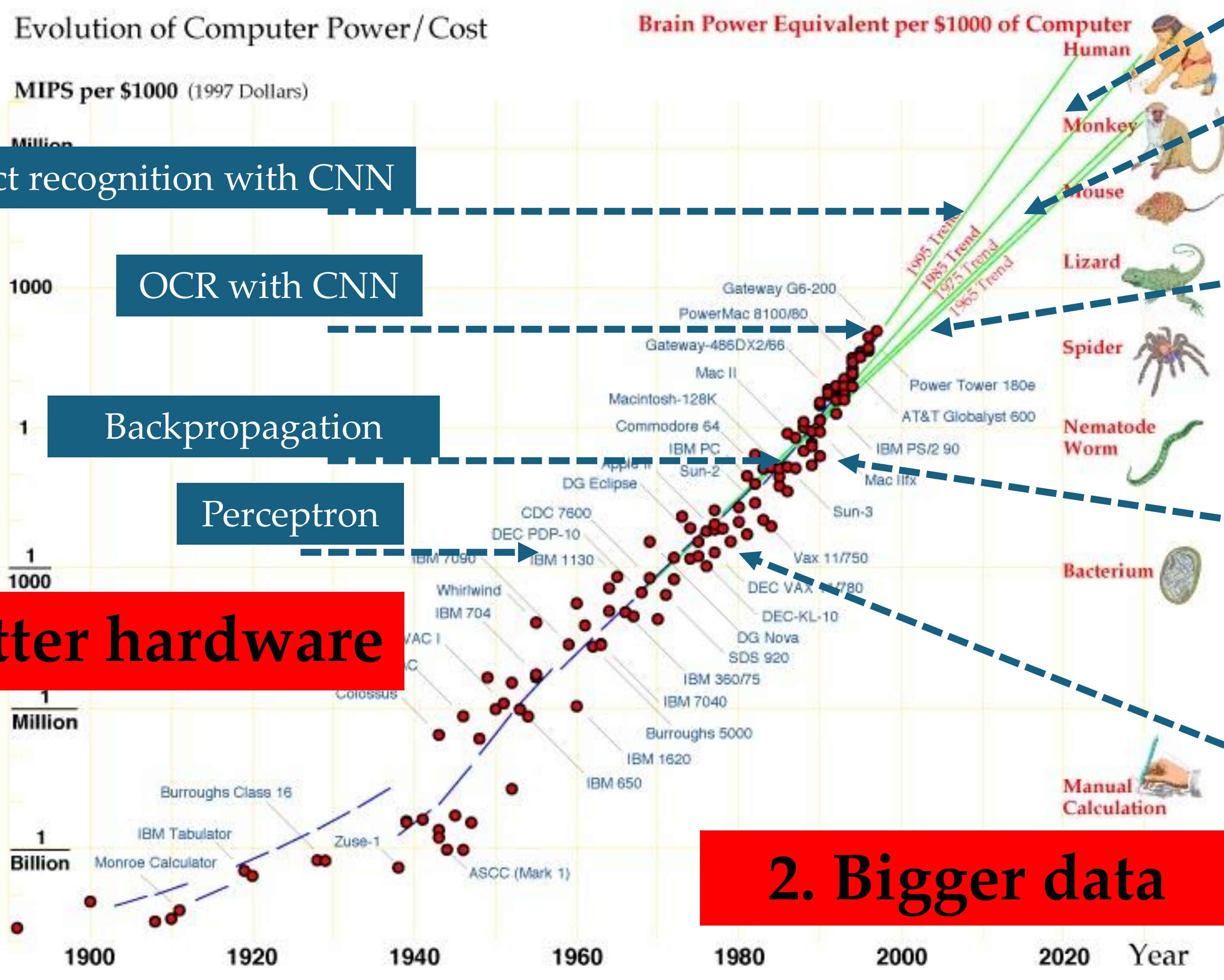
Explore them here: https://knowyourdata-tfds.withgoogle.com/#tab=STATS&dataset=imagenet2012

(Important to also "see" the data, do not just throw a neural network at it!)

Also check out: *On the genealogy of machine learning datasets: A critical history of ImageNet*. Denton et al. 2021

Evolution of scale

Datasets of everything (video, multi-modal, robots etc.)



Object recognition with CNN

OCR with CNN

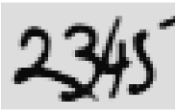
Backpropagation

Perceptron

1. Better hardware

Imagenet: 1000 classes from real images, 1M images

Bank cheques



Parity, negation problems

D1	D2	D3	Even-Parity
0	0	0	True
0	0	1	False
0	1	0	False
0	1	1	True
1	0	0	False
1	0	1	True
1	1	0	True
1	1	1	False

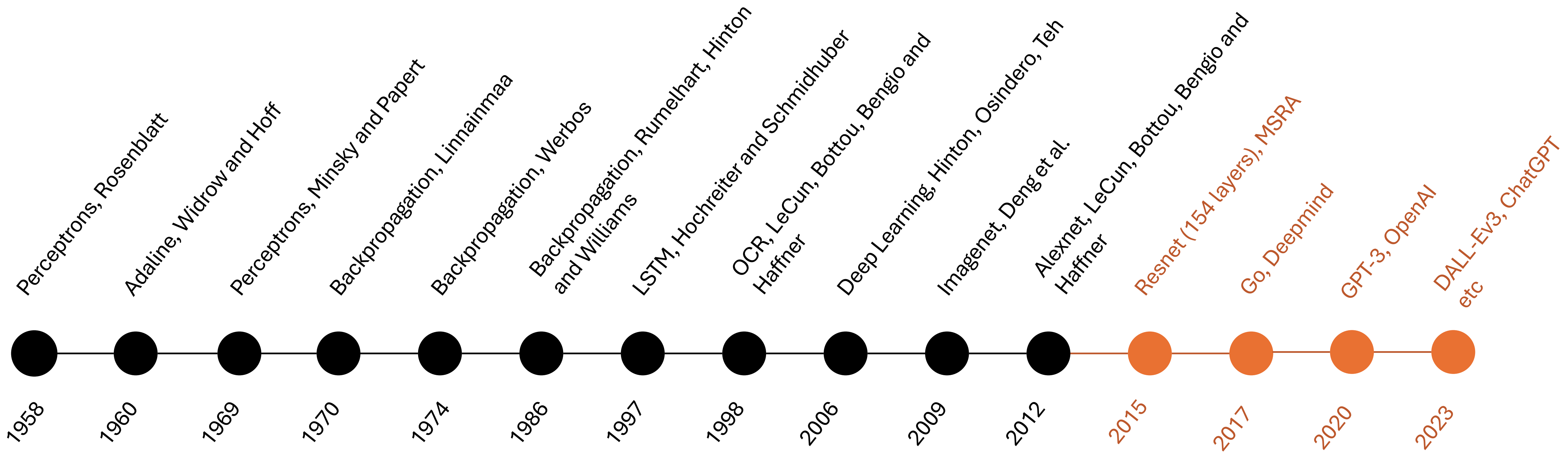
Mark I Perceptron

Potentiometers

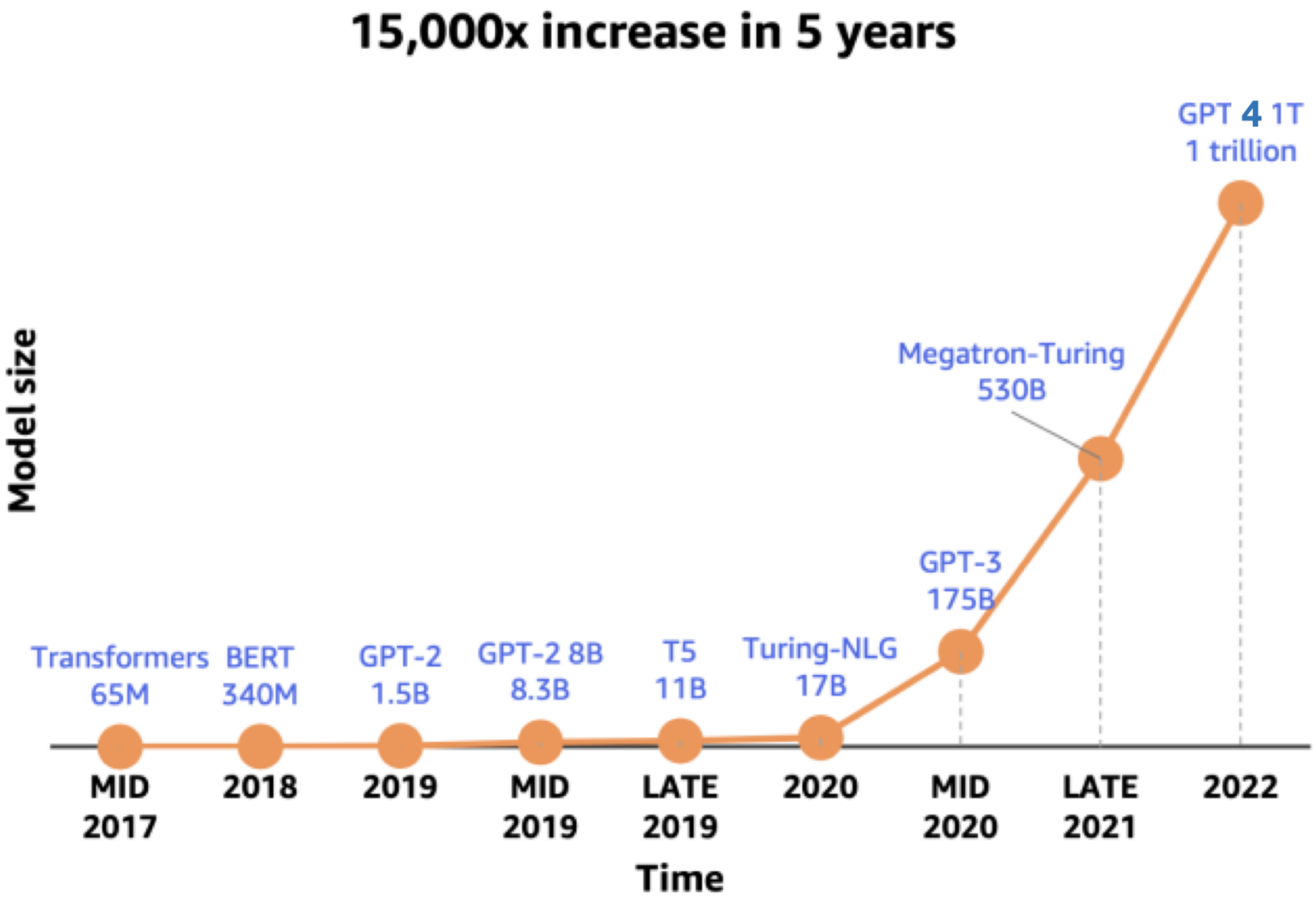
3. Better algorithms

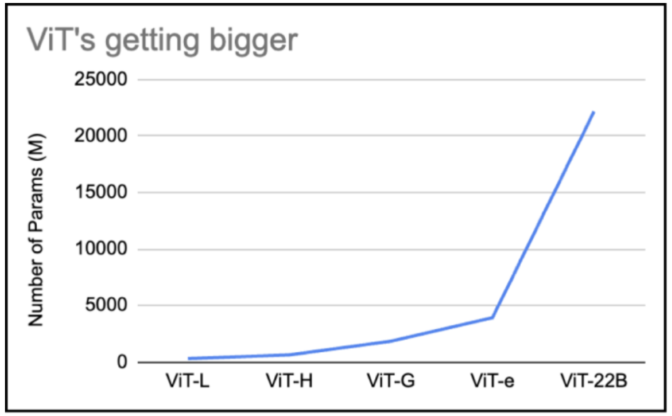
2. Bigger data

Golden era of deep learning



Scale, scale, scale





"Compute Requirements: ViT-22B was trained on 1024 TPU V4 chips [..]"

Current scaling of models

BERT model (354M parameters) ~ now $2K

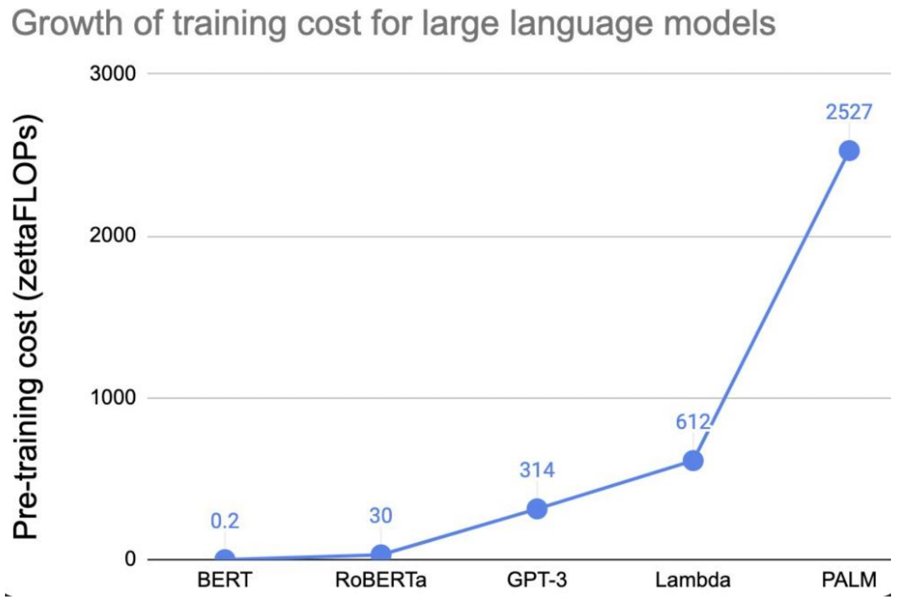
RoBERTa (1000 GPUs for a week) ~ now $350K

GPT-3 (175B parameters, 1500 GPUs for 2 months) ~ $3M

…

PaLM

- 6144 TPUs, ~$25M
- 3.2 million kWh ~~1000 Households for a year



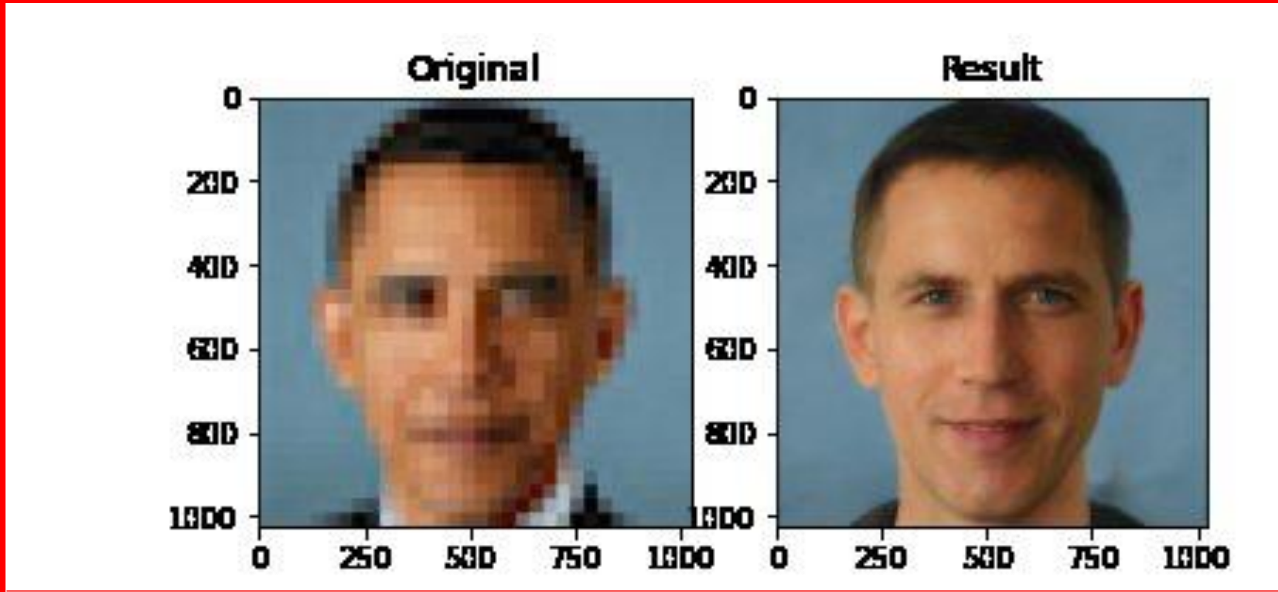
Source: https://twitter.com/tomgoldsteincs/status/1544370726119112704

Is deep learning ready for the real world?

Deadly accidents

Misinformation

Surveillance concerns



Bias

Deep learning: an ongoing effort

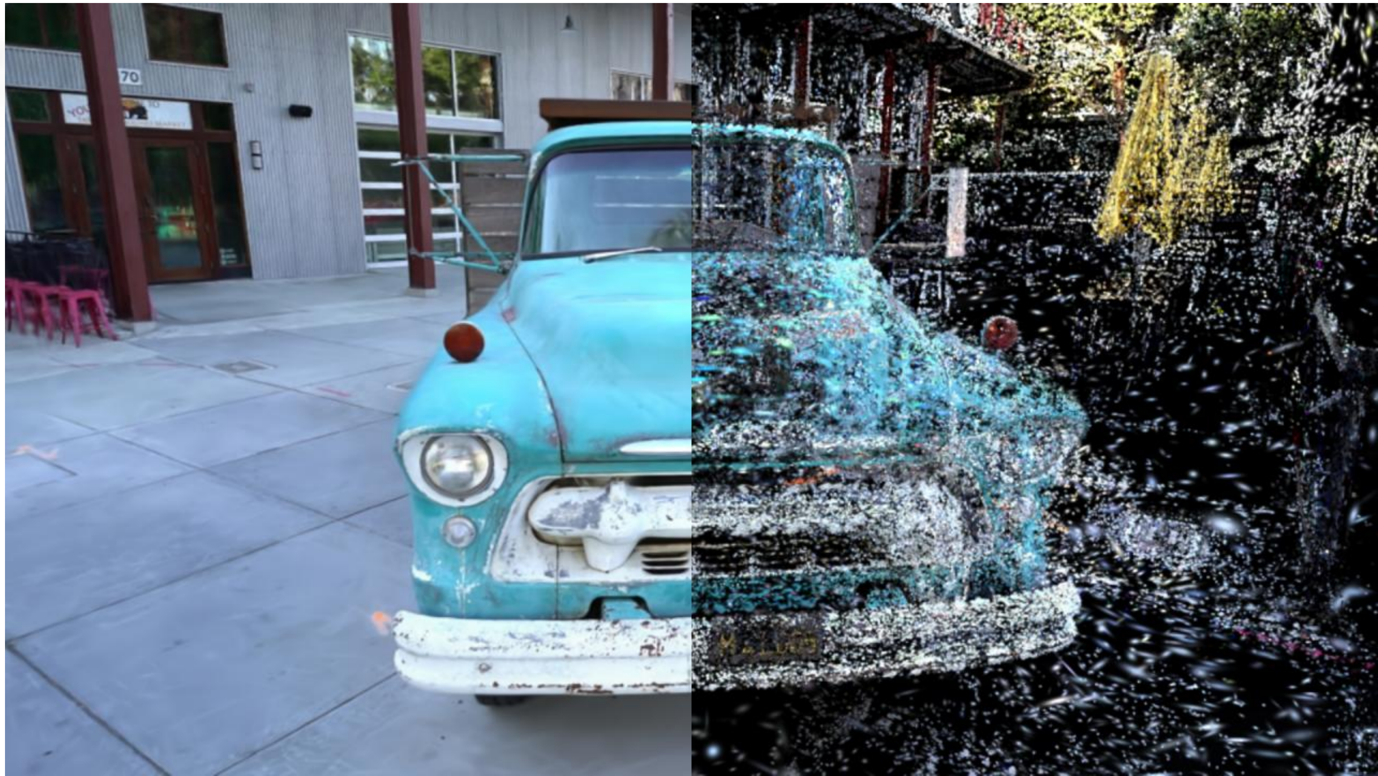
Rapid advances and quick strides.

AI is already assisting us with many basic efforts.

Many fundamental issues remain:

Causality, generalization, robustness, hierarchies, facts/knowledge, continuality, alignment, sustainability, ...

Is AI only deep learning?



Recap

Course overview

History of deep learning

Future of deep learning

Learning and reflection

Understanding Deep Learning: Chapter 1

Annotated History of Modern AI and Deep Learning – Jürgen Schmidhuber

All reading is in ”Reading Materials” under “Modules” in Canvas

Next lecture

Lecture	Title
1	Intro and history of deep learning
3	Deep learning optimization I
5	Convolutional deep learning
7	Graph deep learning
9	Multi-modal deep learning
11	What doesn't work in deep learning
13	Q&A

Lecture	Title
2	AutoDiff
4	Deep learning optimization II
6	Attention-based deep learning
8	From supervised to unsupervised deep learning
10	Generative deep learning
12	Non-Euclidean deep learning
14	Deep learning for videos