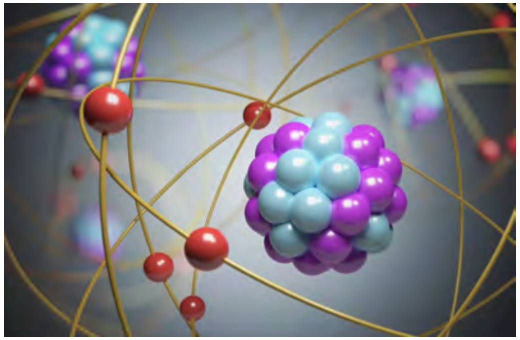
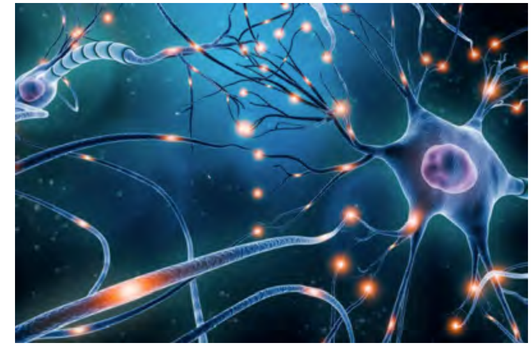
The 1st Heartland Workshop on Heart Sounds and AI

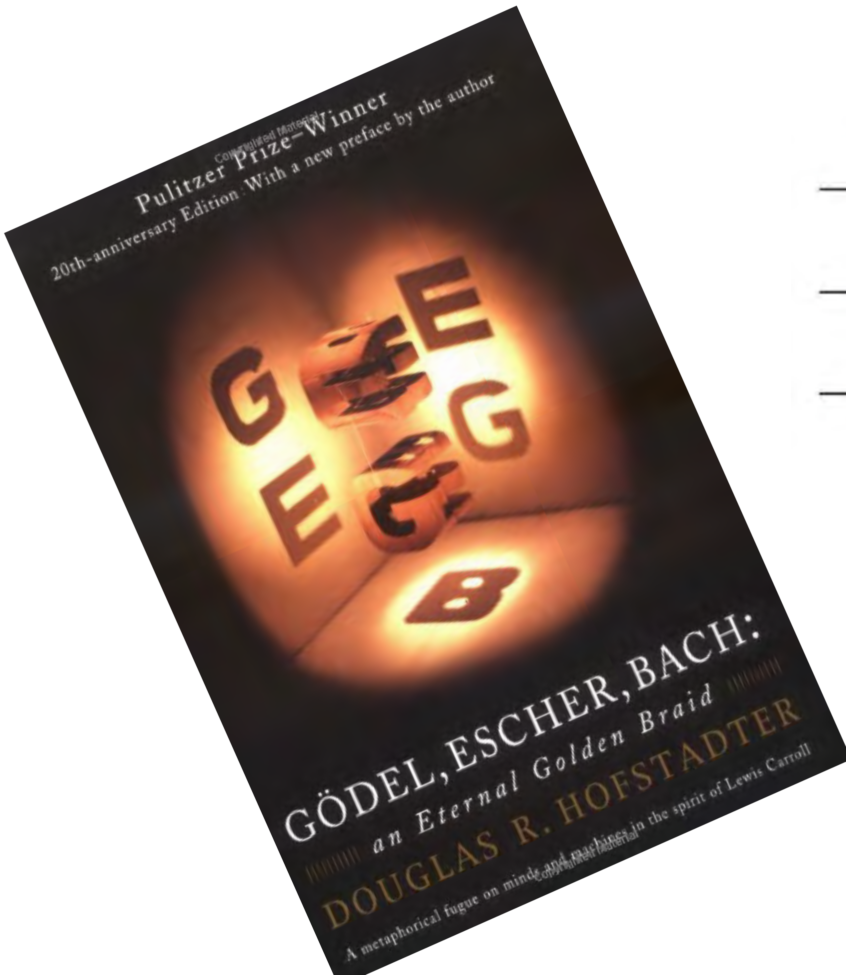
How Can AI Innovations Drive Effective Healthcare Applications?

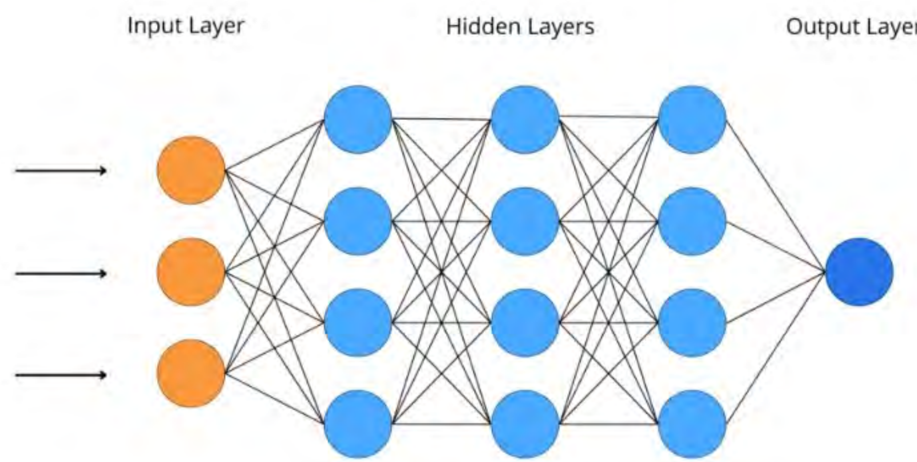
M. R. Hasan
Department of Electrical and Computer Engineering
Human-First Artificial Intelligence Lab (HAL 2.0)
June 20, 2024



From Atoms to "Neurons"







Motivation: AI in Healthcare

- Artificial intelligence (AI) for effective analysis of diagnostic tests.
- An estimated 795,000 Americans die or are permanently disabled by diagnostic error each year (Newman-Toker et al. 2024).

https://qualitysafety.bmj.com/content/33/2/109

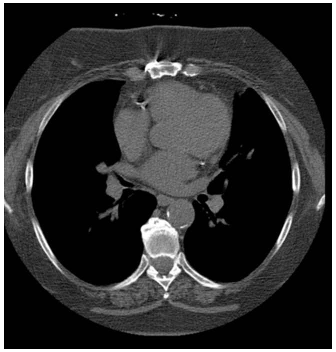
https://news.harvard.edu/gazette/story/2024/01/research-assesses-rates-causes-of-diagnostic-errors/

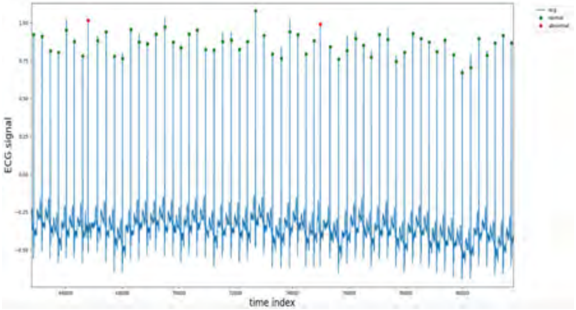
Motivation: AI in Healthcare

- Developing AI methods for detecting **cardiovascular** diseases from diagnostic tests of **three data modalities**.

  - Image: Chest computed tomography (CT) for coronary artery disease
  - Numeric: Electrocardiogram (ECG) for cardiac arrhythmias
  - Textual: Text reports of cardiac magnetic resonance imaging (MRI) for hypertrophic cardiomyopathy

In what ways the patterns in the data vary across modalities?





MRI diagnostic report

Name: *** **Gender:** Male **Date of Birth:** *** **Age:** 28y

Application department: * **Patient ID:** * **Examination number:** *

Examination item: Cardiac magnetic resonance imaging

Clinical diagnosis: Chest pain

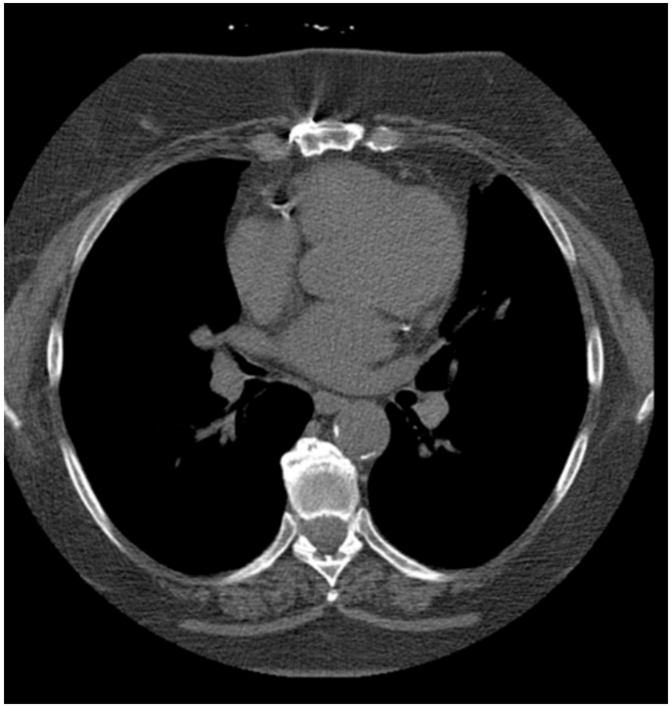
TECHNIQUE:

Equipment: 3.0T Prisma (SIEMENS), Short and long axis cine images (function and structure), T1 mapping, T2 mapping, First-pass perfusion images (contrast agent: Gadobenate Dimeglumine 0.1 mmol/kg), Delayed-enhanced images (scars and fibrosis)

Motivating Case: Cardiac Images

- How can we design an AI to automatically detect cardiovascular disease from coronary artery calcification obtained via computed tomography (CT) images?

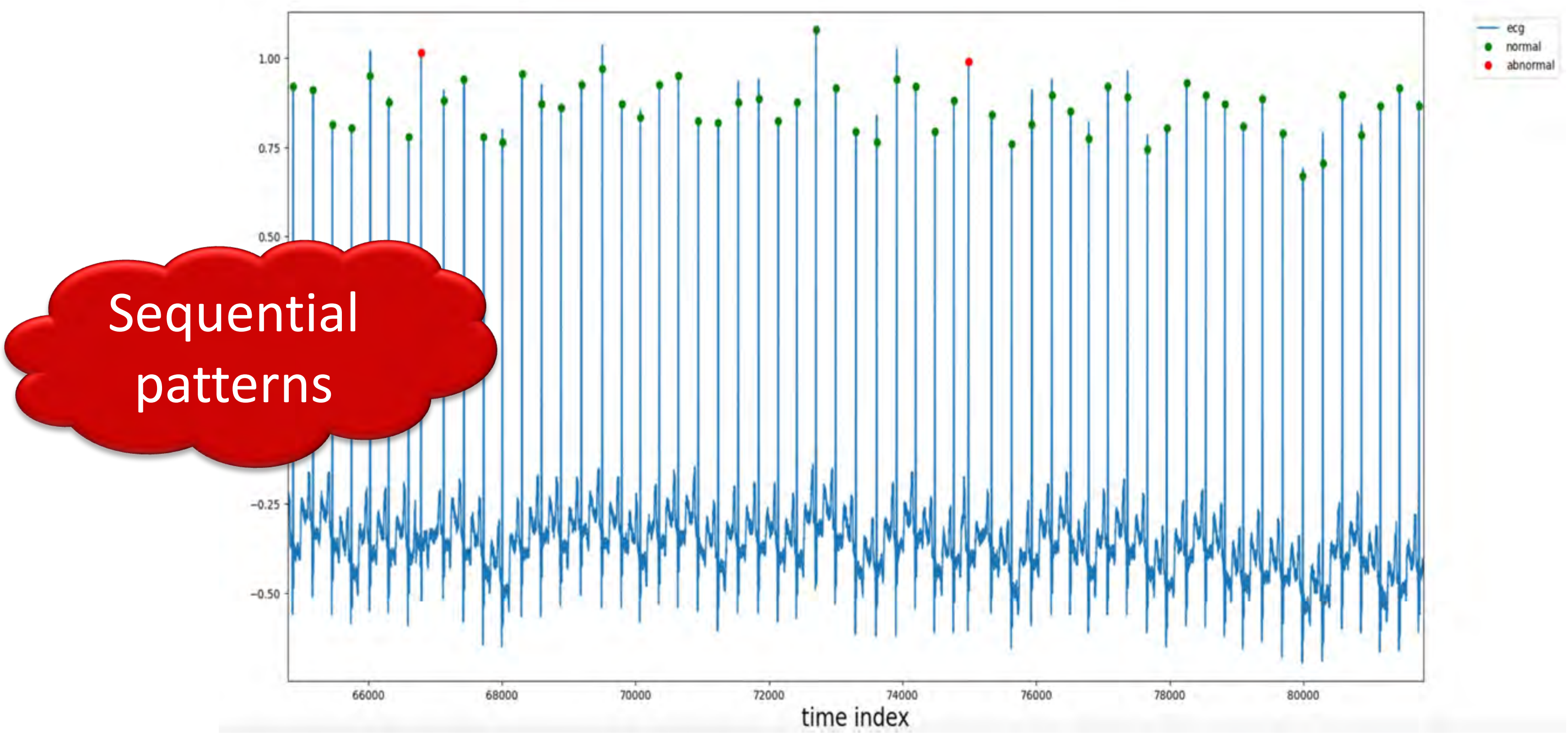
Spatial patterns



https://www.nature.com/articles/s41467-024-46977-3

Motivating Case: Heart Electrical Activity

- How can we design an AI to automatically detect cardiac arrhythmia from electrocardiogram (ECG) data that represents different cardiac activities via electrical signals from heart?

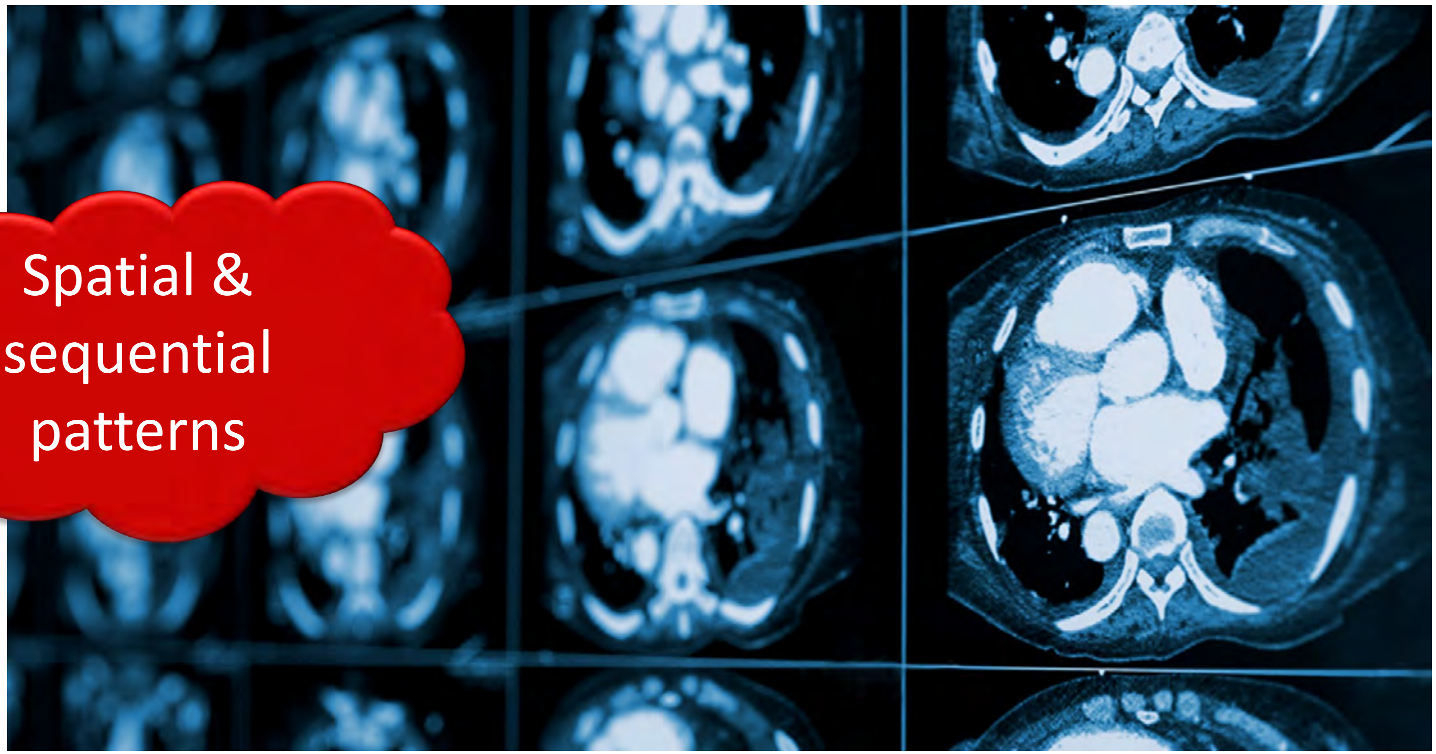


https://www.mdpi.com/1424-8220/22/5/1928

Motivating Case: Series of Cardiac Images

- How can we design an AI to automatically detect calcification by tracking the **progression** of calcification across multiple slices or phases of a cardiac cycle by a sequence of CT images?

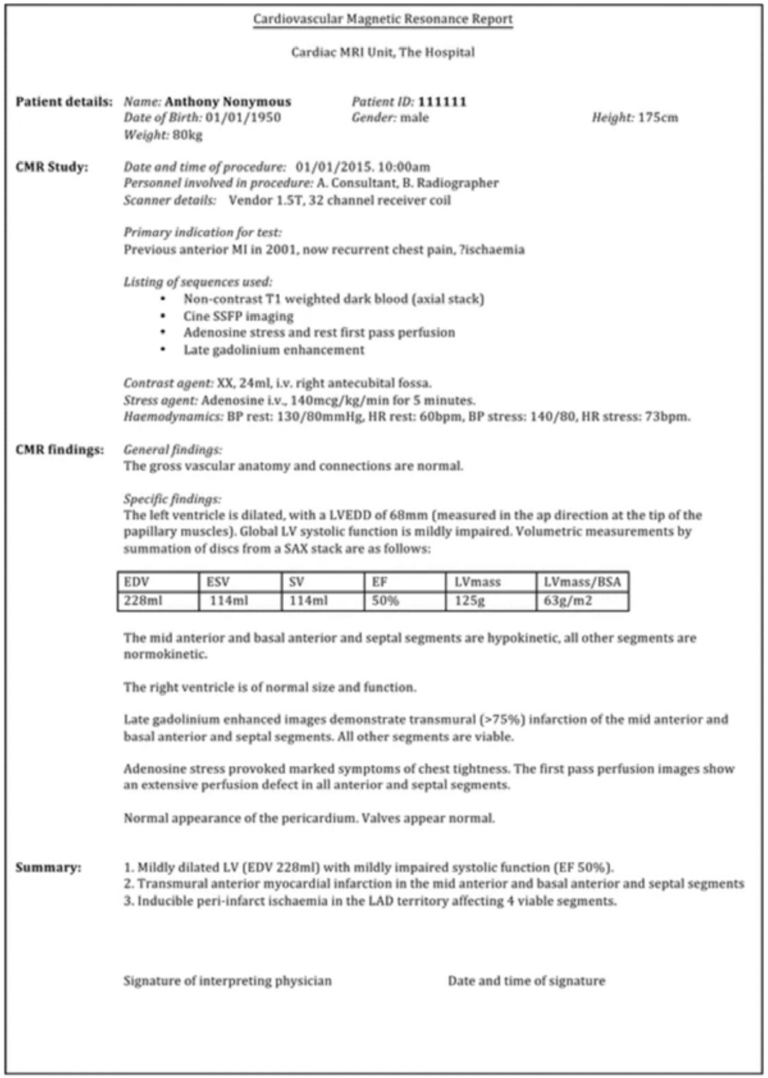
Spatial & sequential patterns



https://www.touchstoneimaging.com/chest-ct-scan/

Motivating Case: Test Report Text Data

- How can we design an AI to identify hypertrophic cardiomyopathy patients from cardiac magnetic resonance **textual reports**?



Cardiovascular Magnetic Resonance Report

Cardiac MRI Unit, The Hospital

Patient details: *Name:* **Anthony Nonymous** *Patient ID:* **111111**
Date of Birth: 01/01/1950 *Gender:* male *Height:* 175cm
Weight: 80kg

CMR Study: *Date and time of procedure:* 01/01/2015. 10:00am
Personnel involved in procedure: A. Consultant, B. Radiographer
Scanner details: Vendor 1.5T, 32 channel receiver coil

Primary indication for test:
Previous anterior MI in 2001, now recurrent chest pain, ?ischaemia

Listing of sequences used:
- Non-contrast T1 weighted dark blood (axial stack)
- Cine SSFP imaging
- Adenosine stress and rest first pass perfusion
- Late gadolinium enhancement

Contrast agent: XX, 24ml, i.v. right antecubital fossa.
Stress agent: Adenosine i.v., 140mcg/kg/min for 5 minutes.
Haemodynamics: BP rest: 130/80mmHg, HR rest: 60bpm, BP stress: 140/80, HR stress: 73bpm.

CMR findings: *General findings:*
The gross vascular anatomy and connections are normal.

Specific findings:
The left ventricle is dilated, with a LVEDD of 68mm (measured in the ap direction at the tip of the papillary muscles). Global LV systolic function is mildly impaired. Volumetric measurements by summation of discs from a SAX stack are as follows:

EDV	ESV	SV	EF	LVmass	LVmass/BSA
228ml	114ml	114ml	50%	125g	63g/m2

The mid anterior and basal anterior and septal segments are hypokinetic, all other segments are normokinetic.

The right ventricle is of normal size and function.

Late gadolinium enhanced images demonstrate transmural (>75%) infarction of the mid anterior and basal anterior and septal segments. All other segments are viable.

Adenosine stress provoked marked symptoms of chest tightness. The first pass perfusion images show an extensive perfusion defect in all anterior and septal segments.

Normal appearance of the pericardium. Valves appear normal.

Summary: 1. Mildly dilated LV (EDV 228ml) with mildly impaired systolic function (EF 50%).
2. Transmural anterior myocardial infarction in the mid anterior and basal anterior and septal segments
3. Inducible peri-infarct ischaemia in the LAD territory affecting 4 viable segments.

Signature of interpreting physician Date and time of signature

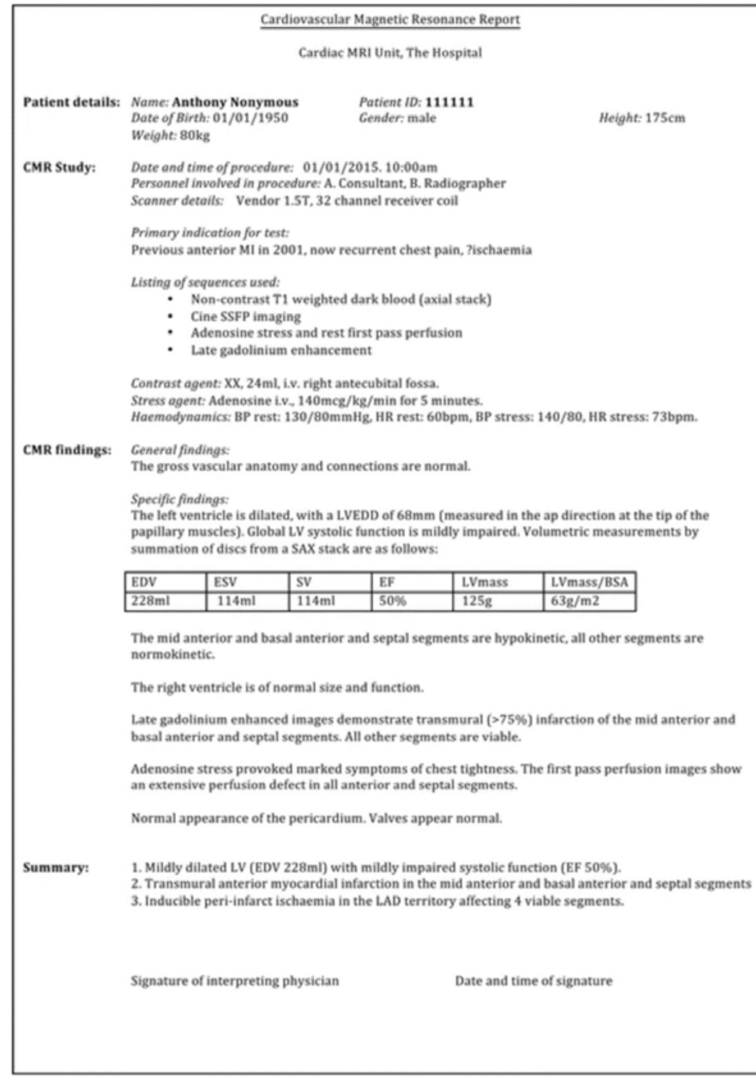
https://link-springer-com.libproxy.unl.edu/chapter/10.1007/978-3-319-20940-1_22

Motivating Case: Test Report Text Data

- Radiology generates a vast amount of textual data, including radiology reports, clinical notes, annotations associated with medical imaging, and more.
- Its analysis requires sophisticated understanding and interpretation.

Linguistic patterns

Prior knowledge



Cardiovascular Magnetic Resonance Report

Cardiac MRI Unit, The Hospital

Patient details: *Name:* **Anthony Nonymous** *Patient ID:* **111111**
Date of Birth: 01/01/1950 *Gender:* male *Height:* 175cm
Weight: 80kg

CMR Study: *Date and time of procedure:* 01/01/2015. 10:00am
Personnel involved in procedure: A. Consultant, B. Radiographer
Scanner details: Vendor 1.5T, 32 channel receiver coil

Primary indication for test:
Previous anterior MI in 2001, now recurrent chest pain, ?ischaemia

Listing of sequences used:
- Non-contrast T1 weighted dark blood (axial stack)
- Cine SSFP imaging
- Adenosine stress and rest first pass perfusion
- Late gadolinium enhancement

Contrast agent: XX, 24ml, i.v. right antecubital fossa.
Stress agent: Adenosine i.v., 140mcg/kg/min for 5 minutes.
Haemodynamics: BP rest: 130/80mmHg, HR rest: 60bpm, BP stress: 140/80, HR stress: 73bpm.

CMR findings: *General findings:*
The gross vascular anatomy and connections are normal.

Specific findings:
The left ventricle is dilated, with a LVEDD of 68mm (measured in the ap direction at the tip of the papillary muscles). Global LV systolic function is mildly impaired. Volumetric measurements by summation of discs from a SAX stack are as follows:

EDV	ESV	SV	EF	LVmass	LVmass/BSA
228ml	114ml	114ml	50%	125g	63g/m2

The mid anterior and basal anterior and septal segments are hypokinetic, all other segments are normokinetic.

The right ventricle is of normal size and function.

Late gadolinium enhanced images demonstrate transmural (>75%) infarction of the mid anterior and basal anterior and septal segments. All other segments are viable.

Adenosine stress provoked marked symptoms of chest tightness. The first pass perfusion images show an extensive perfusion defect in all anterior and septal segments.

Normal appearance of the pericardium. Valves appear normal.

Summary: 1. Mildly dilated LV (EDV 228ml) with mildly impaired systolic function (EF 50%).
2. Transmural anterior myocardial infarction in the mid anterior and basal anterior and septal segments
3. Inducible peri-infarct ischaemia in the LAD territory affecting 4 viable segments.

Signature of interpreting physician Date and time of signature

https://link-springer-com.libproxy.unl.edu/chapter/10.1007/978-3-319-20940-1_22

Motivating Case

- How can we design a **single** AI to automatically analyze the diagnostic reports from **<u>diverse modalities</u>** (images, charts, text reports).
- How can we enable the AI to **answer** stakeholder questions?

Spatial & sequential patterns

Linguistic patterns

Prior knowledge

Reliable

Common sense

What We Will Cover

- Modern AI: Machine Learning and Deep Learning
- Discriminative AI
- Neural Networks & its Variants
- Fully-Connected Networks
- Convolutional Networks
- Recurrent Neural Networks
- Applications: discriminative AI in healthcare
- Generative AI
- Advancing AI in healthcare using generative AI
- Transformer Networks
- Language Models: Potentials and Limitations
- Our AI Future: Societal Implications

What is Artificial Intelligence

- AI refers to the simulation of human intelligence in machines that are programmed to think like humans and mimic their actions.

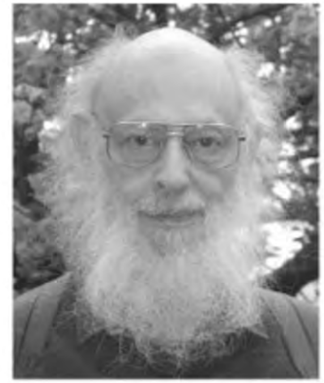
https://zapier.com/blog/artificial-general-intelligence/

1956 Dartmouth Conference: The Founding Fathers of AI


John MacCarthy


Marvin Minsky


Claude Shannon


Ray Solomonoff


Alan Newell


Herbert Simon


Arthur Samuel


Oliver Selfridge


Nathaniel Rochester

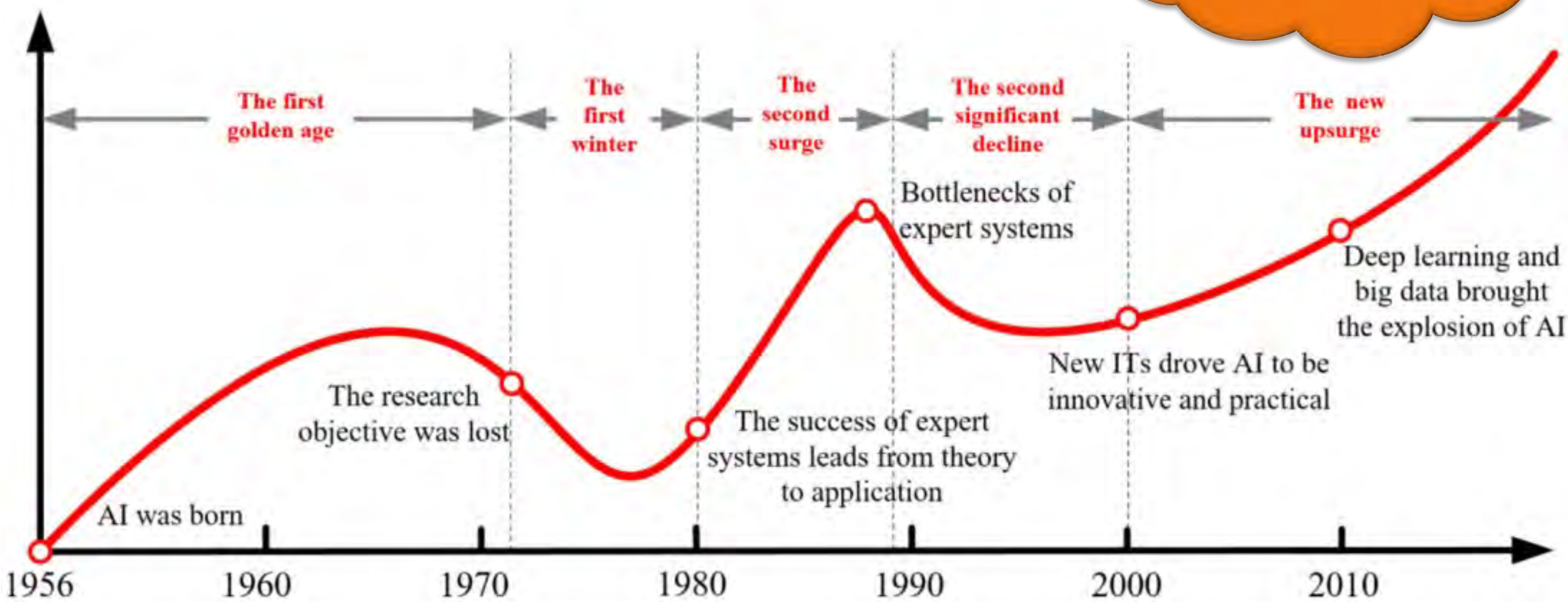

Trenchard More

The 2 month, 10 man study during the Summer of 1956!

History of AI

- AI has a long history of being the next big thing!

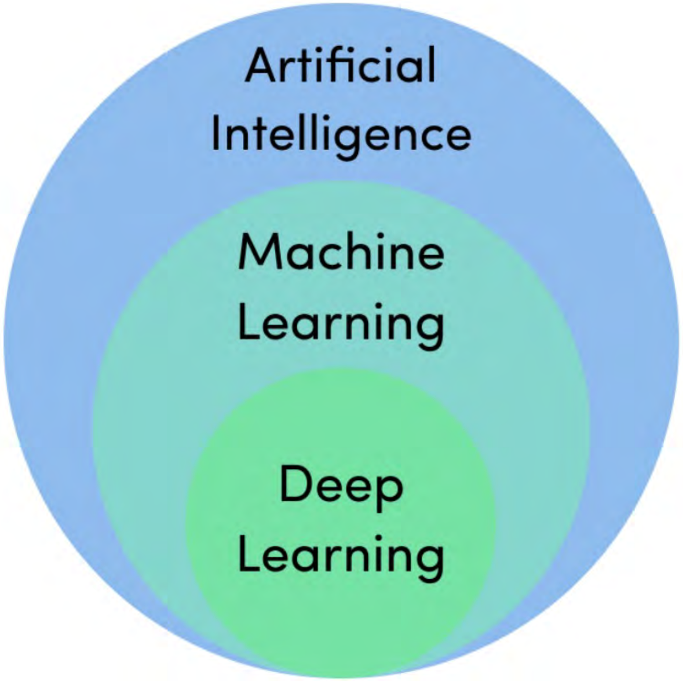
The rise of Modern AI!



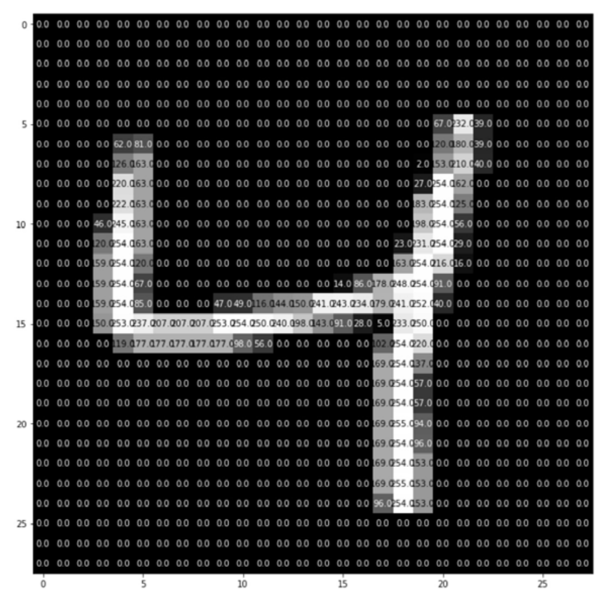
https://www.researchgate.net/publication/350296216_Artificial_intelligence_in_product_lifecycle_management

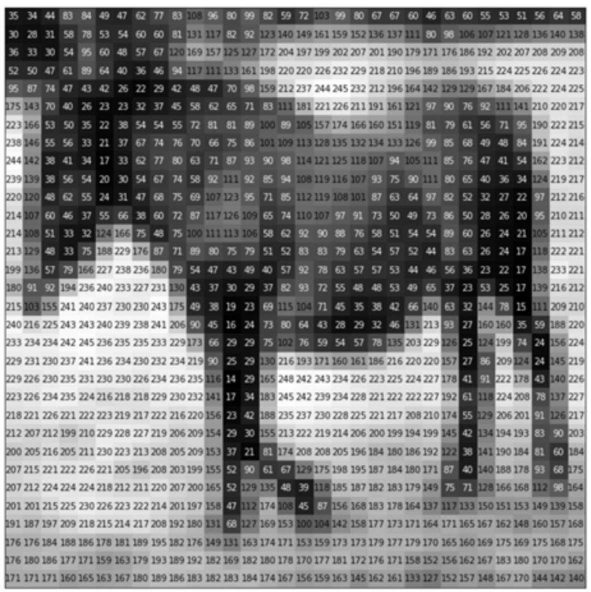
Modern AI

- **Modern AI** refers to Machine and Deep Learning (ML & DL).
- AI is a broader concept, while ML/DL are the most common (and more effective) methods of AI.

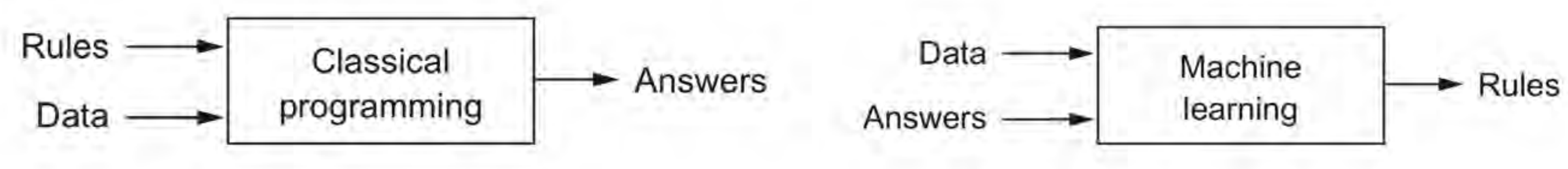


ML/DL is the **science (and art)** of programming computers so they can learn to **recognize patterns from data**.





- Machine Learning is a **new paradigm of AI**.



Good Old-fashioned AI (GOFAI): human domain **experts write set of rules** for computer programs.

Machine Learning: computer programs can **automatically learn** the rules from data.

Based on symbolic logic

Humans prepare data, create representation, label data, etc.

Deductive Process

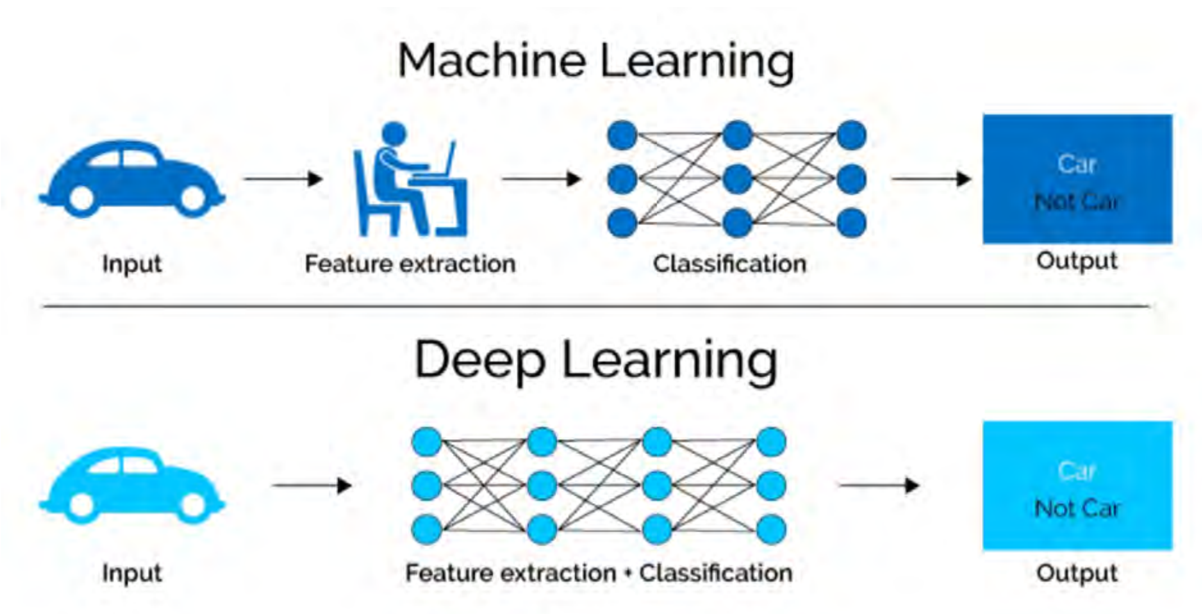
Inductive Process

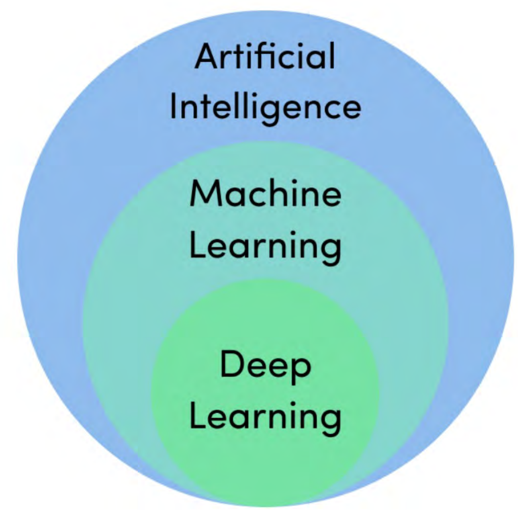
Modern AI

- We can consider the development of Modern AI into **two stages**.

  - Machine Learning (traditional ML models such as k-Nearest Neighbors, Support Vector Machine, Decision Tree)

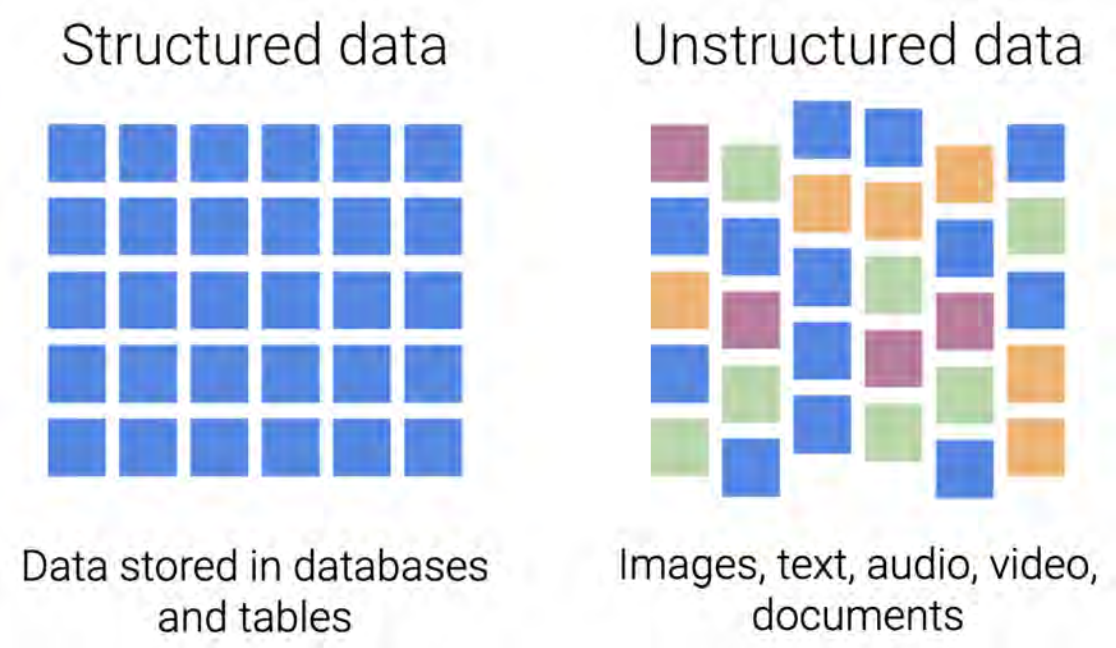
  - Deep Learning (models based on Artificial Neural Networks)





Modern AI

- ML: suitable for structured data (features are pre-defined).
- DL: utilizes unstructured data (need to extract features).

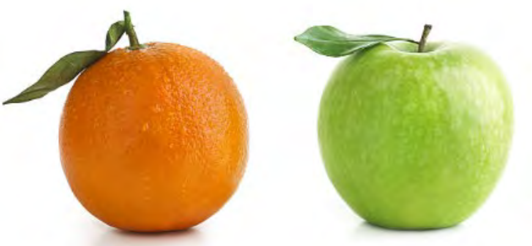


In the rest of the talk, we will focus on DL.

DL: Discriminative vs. Generative AI

Discriminative AI: Given an object, determine its **type**

Supervised Learning



Orange Apple

Labeled data (manual)

0^0

90^0

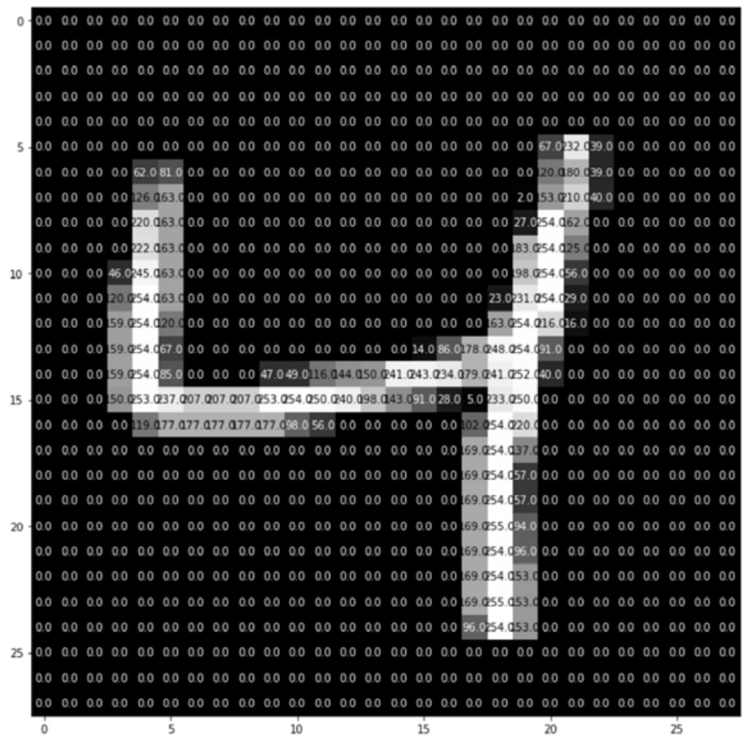
180^0

Pseudo labels (automatic)

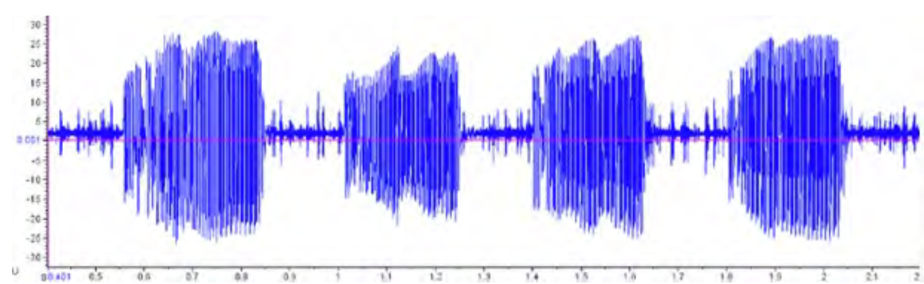
Self-supervised Learning

Generative AI: Create **new** objects

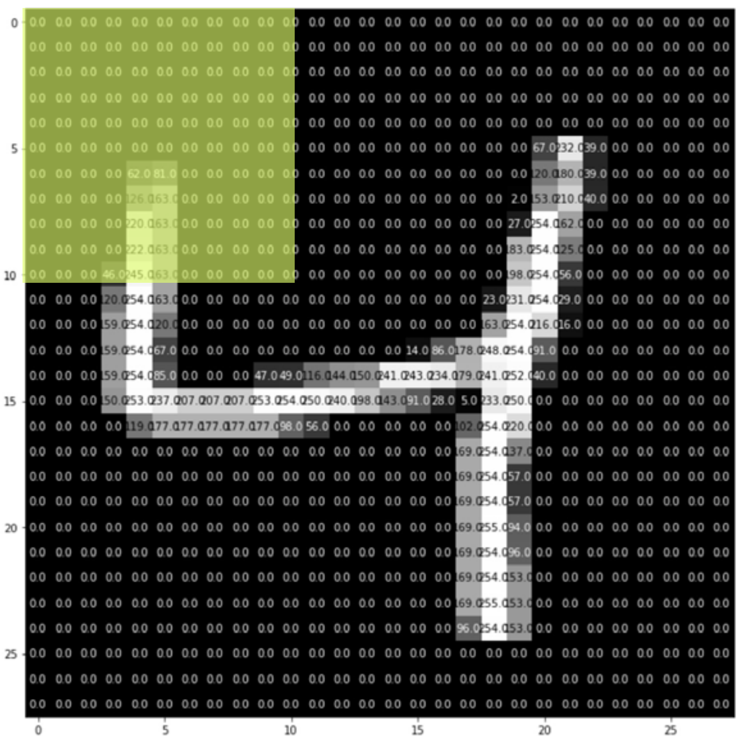
- In DL, we often deal with **two types of unstructured data**.
  - Image
  - Text, audio
- We create DL models to **extract features** before performing pattern recognition.

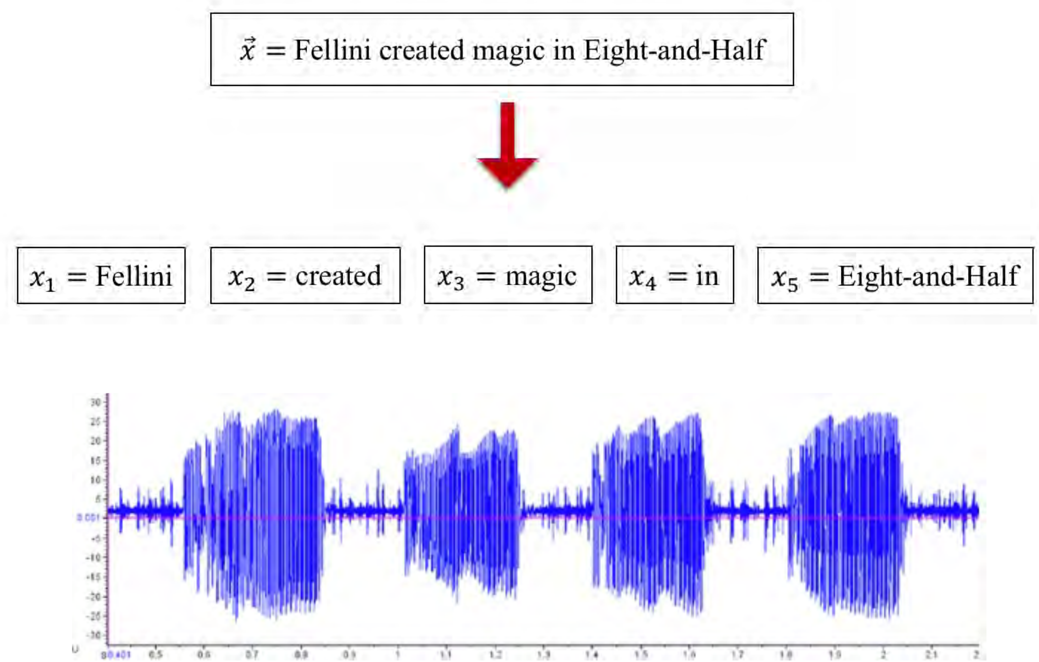


1. Fellini created magic in Eight-and-Half! (POSITIVE)
2. Disgusting! (NEGATIVE)
3. The longest bore on celluloid! (NEGATIVE)
4. I can watch Roma forever! (POSITIVE)
5. Worst than the previous one! (NEGATIVE)
6. Coppola's masterpiece! (POSITIVE)



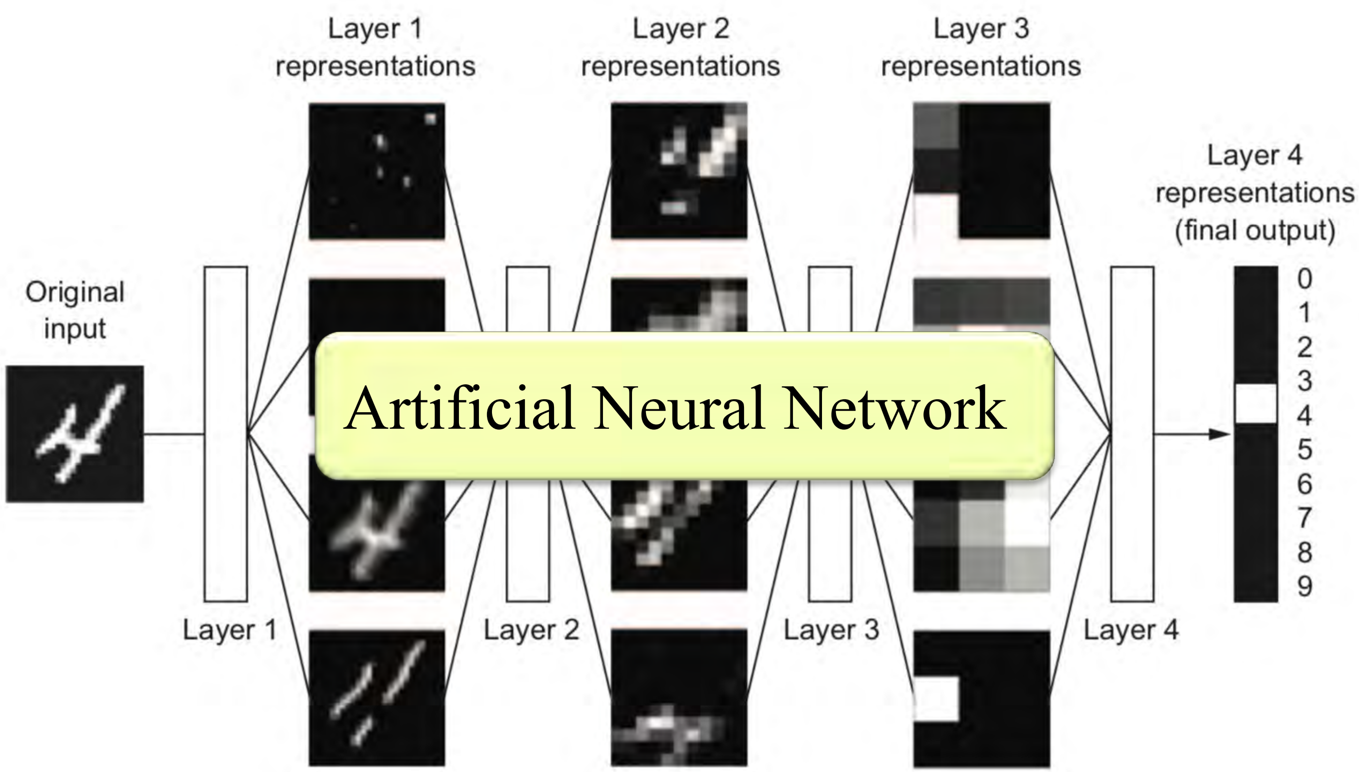
- Feature extractors use **structural information** of the data to learn discriminative features.
  - Image: Spatial information is manipulated
  - Text & Audio: Sequential information is manipulated





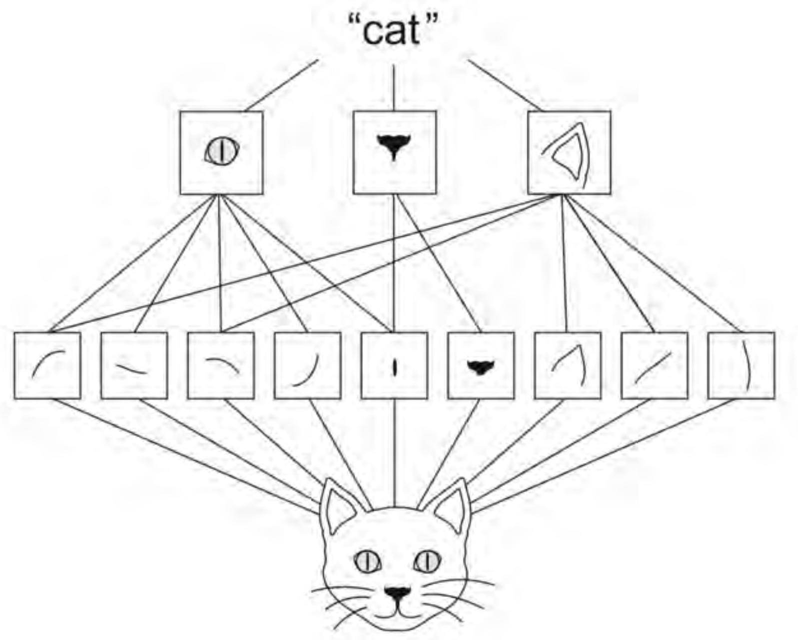
Deep Learning: Feature Extraction

- DL models learn features or representations from input data in a **hierarchical fashion**.
- Each computational layer learns **increasingly complex representation** of the input that is increasingly informative about the final result.
- This approach is inspired by our **biological neural network**.

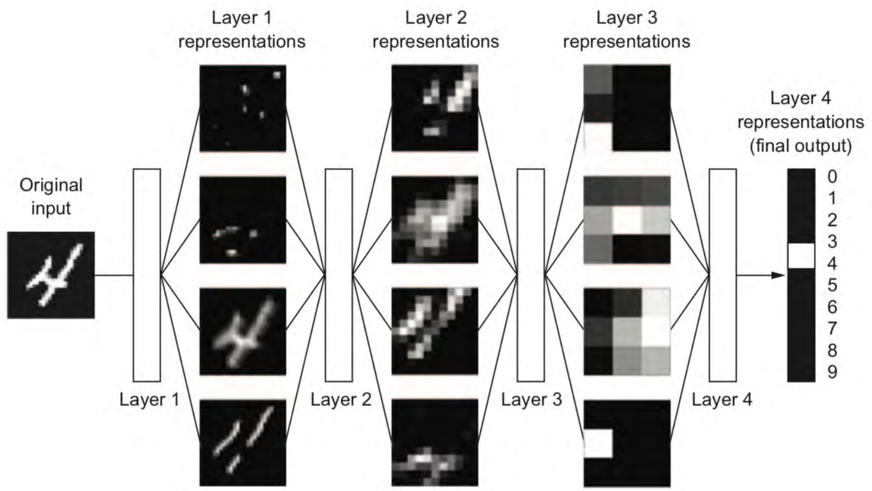


Deep Learning: Mimicking Brain

- Biological neural network **transforms raw sensory** data into layers of representations to form a perception.

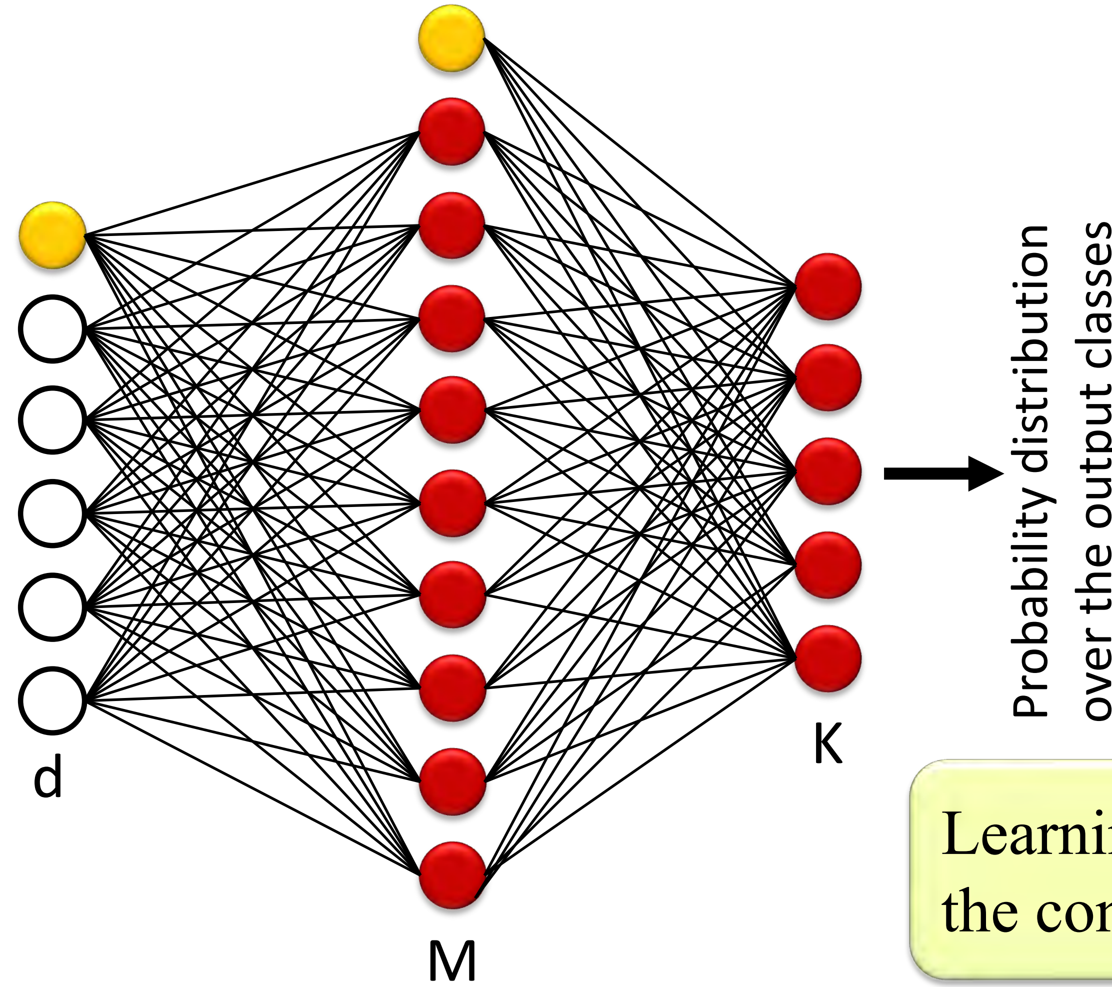


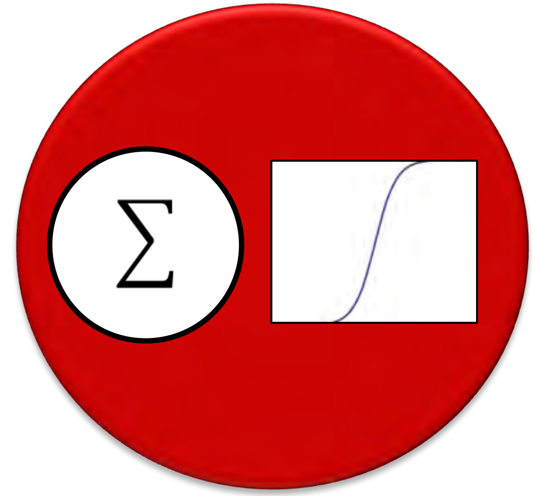
Biological Neural Network



Artificial Neural Network

A Feedforward Neural Network





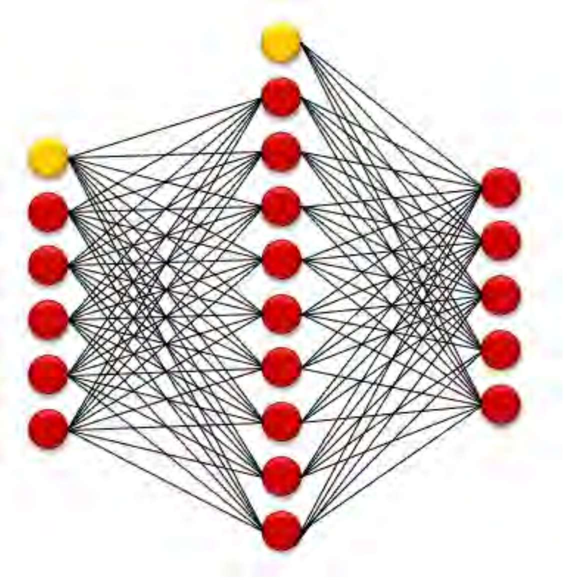
A "to be or not to be" type cell

Learning: adjustment of the connection weights!

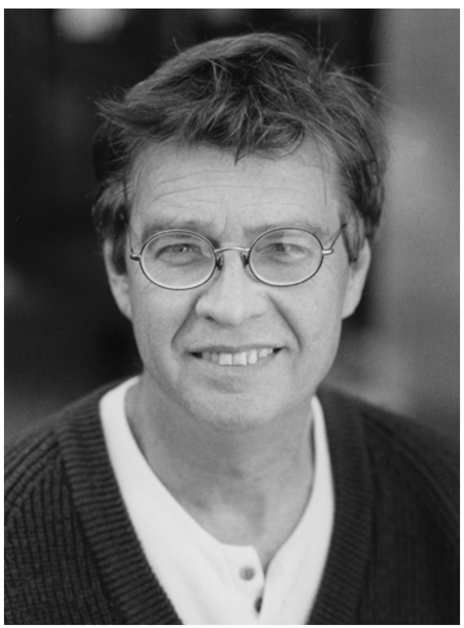
#parameters = M*(d + 1) + K*(M + 1) = 104

Multi-Layer Perceptron

- The **fully-connected** network is the first successful neural network.
- It is known as the Multi-Layer Perceptron (MLP).
- An MLP is a universal function approximator.
- For training (weight adjustment) MLPs, the **backpropagation** algorithm was proposed in 1986.



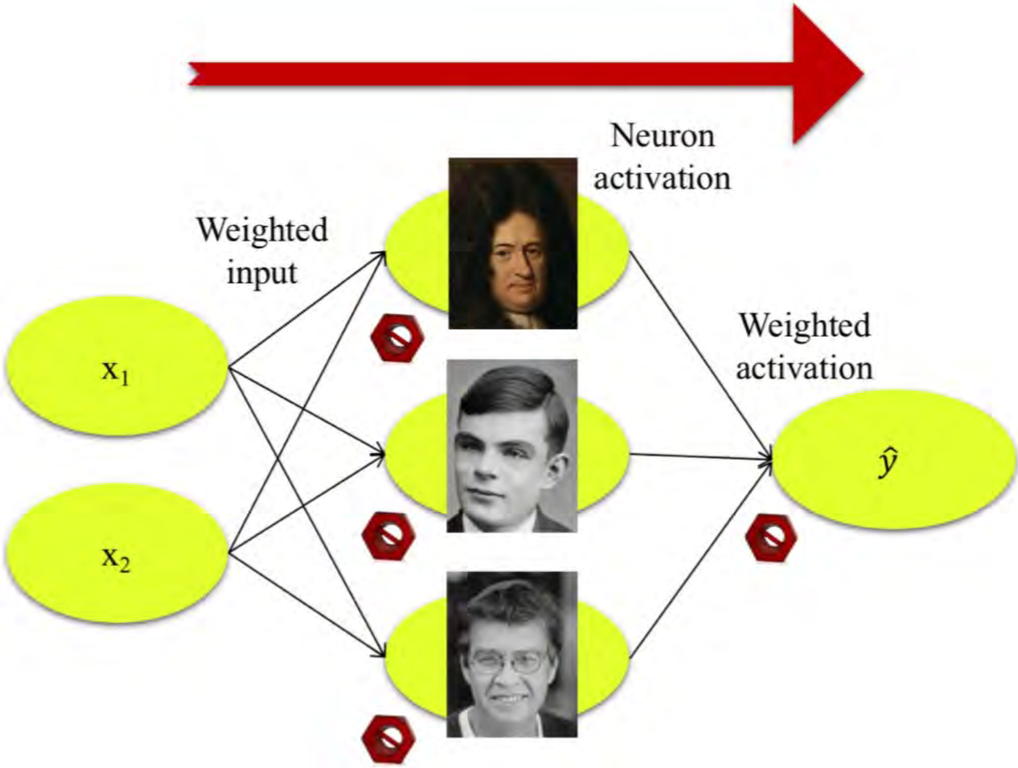
The Backpropagation algorithm changed the **history of AI** (and perhaps the world)!



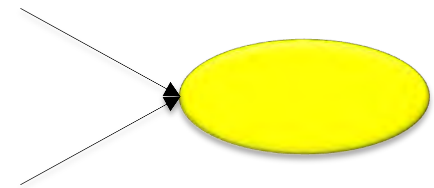
Learning representations by back-propagating errors

David E. Rumelhart, Geoffrey E. Hinton & Ronald J. Williams

Nature **323**, 533–536 (09 October 1986) | Download Citation



(a) Forward propagation



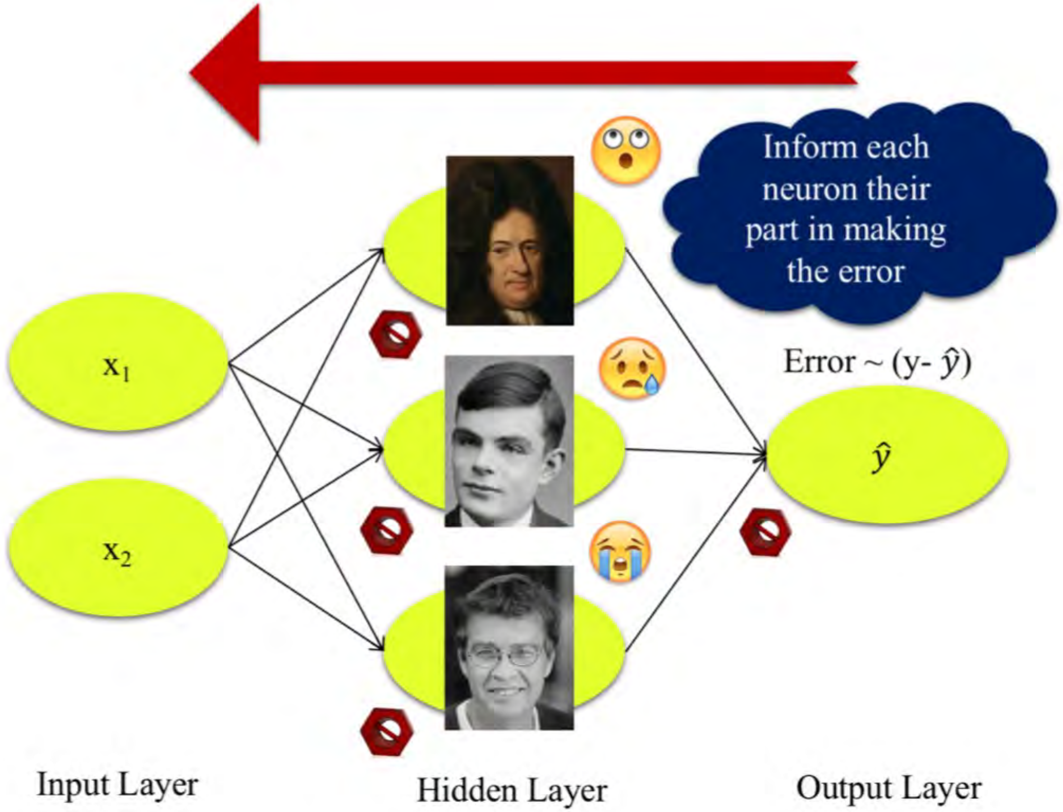
Fact: A neuron is defined by its ***input connection*** weights.

Learning occurs when these weights are **adapted** effectively.

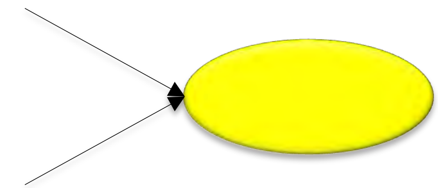
Learning representations by back-propagating errors

David E. Rumelhart, Geoffrey E. Hinton & Ronald J. Williams

Nature **323**, 533–536 (09 October 1986) | Download Citation



(b) Backward propagation



A neuron's responsibility: Compute the **gradient of the error/loss function** by each neuron (i.e., weights of its connections.)

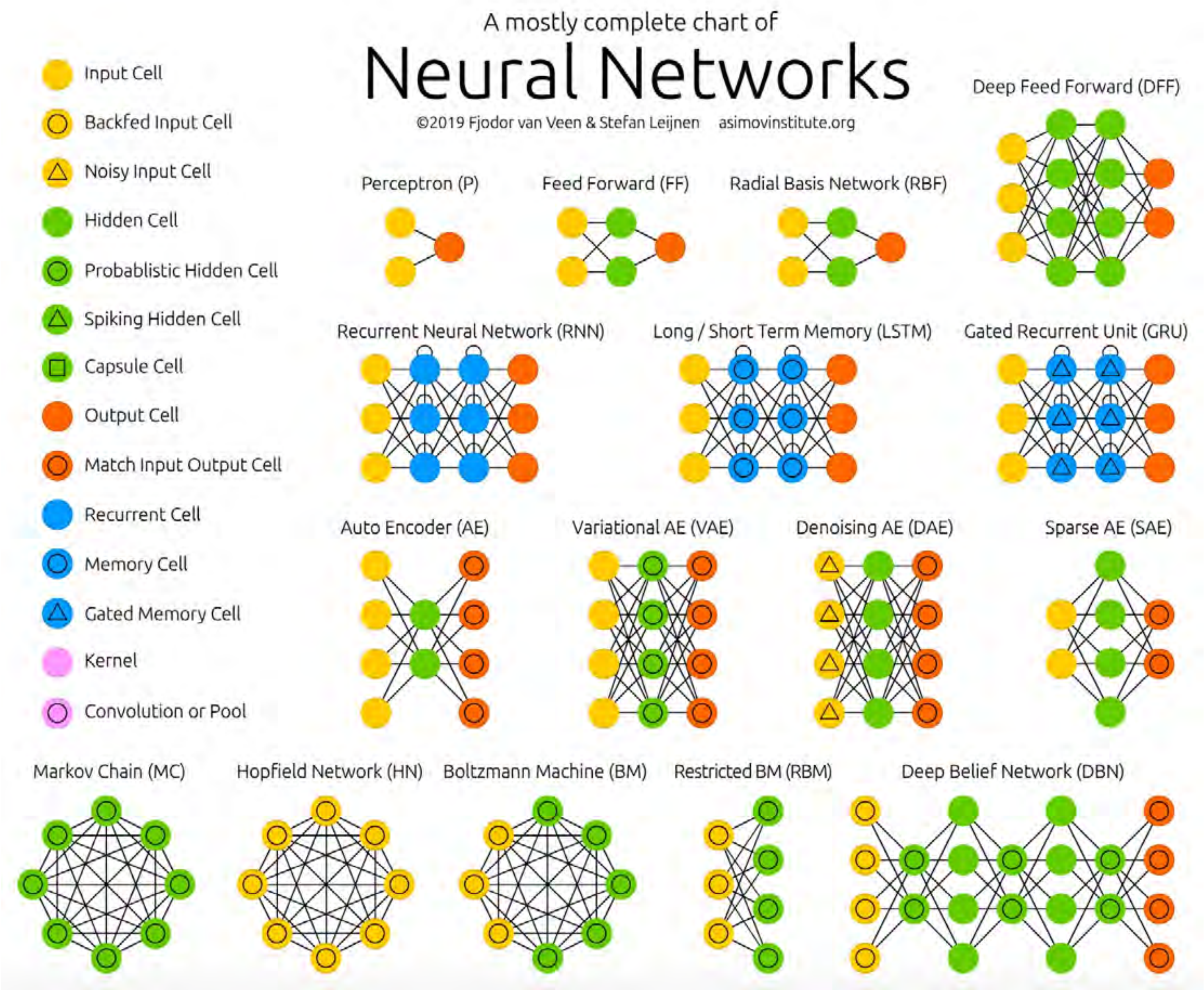
Calculus of mistakes (error)

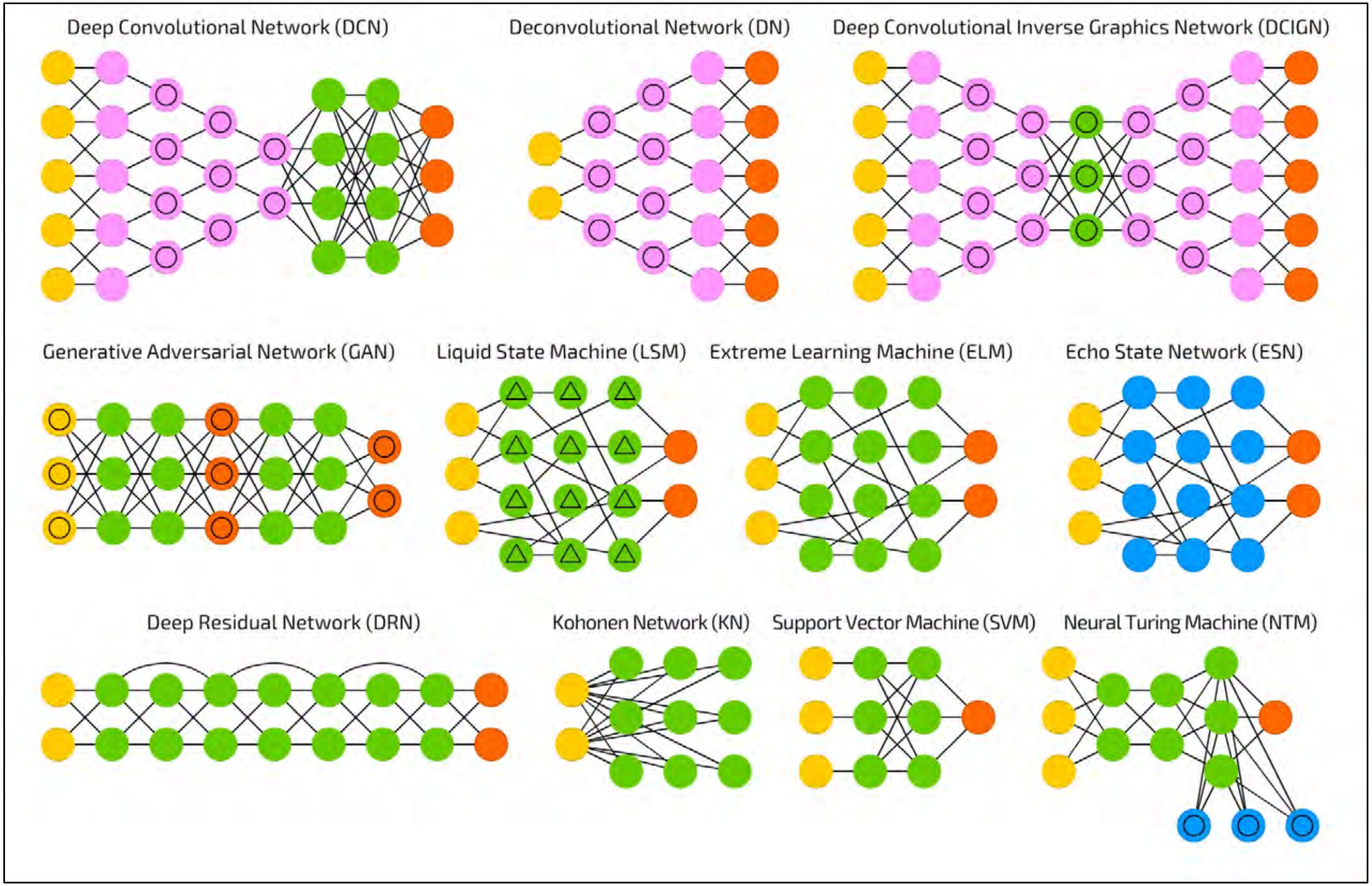
Learning representations by back-propagating errors

David E. Rumelhart, Geoffrey E. Hinton & Ronald J. Williams

Nature **323**, 533–536 (09 October 1986) | Download Citation

NNs can be designed using various architectures…



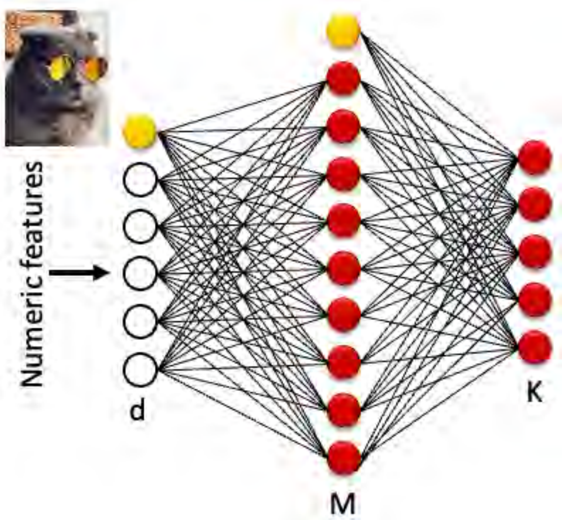


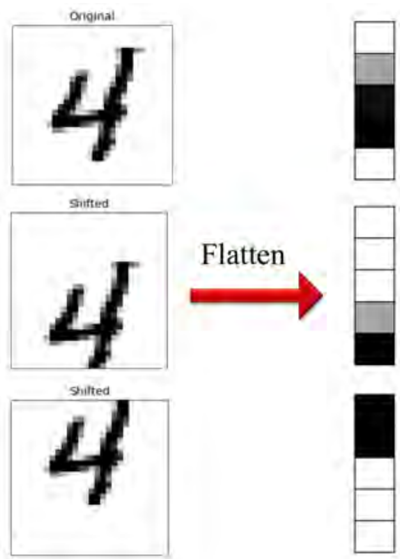
https://www.asimovinstitute.org/neural-network-zoo/

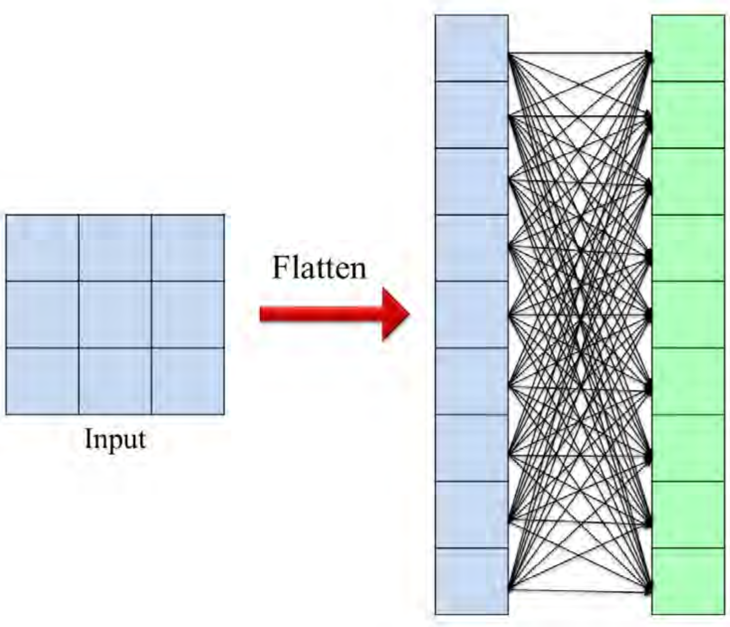
- Neural network architectures for learning from the following two types of unstructured data.
  - Image
  - Text, audio

FC Neural Networks: Limitations

- The MLP (fully-connected or dense) models are neither effective nor efficient to learn from **image** data.
- Two key issues.
  - Unable to maintain spatial invariance
  - Don’t scale well with the size of the images



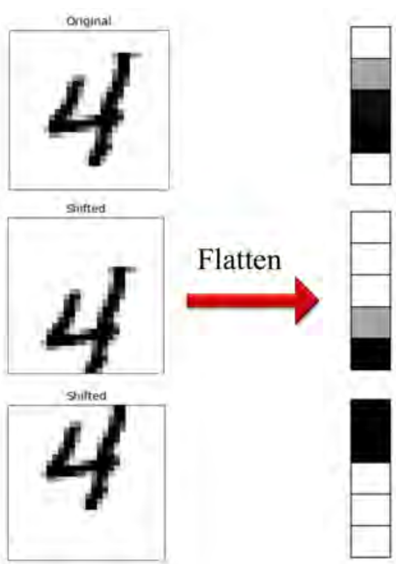


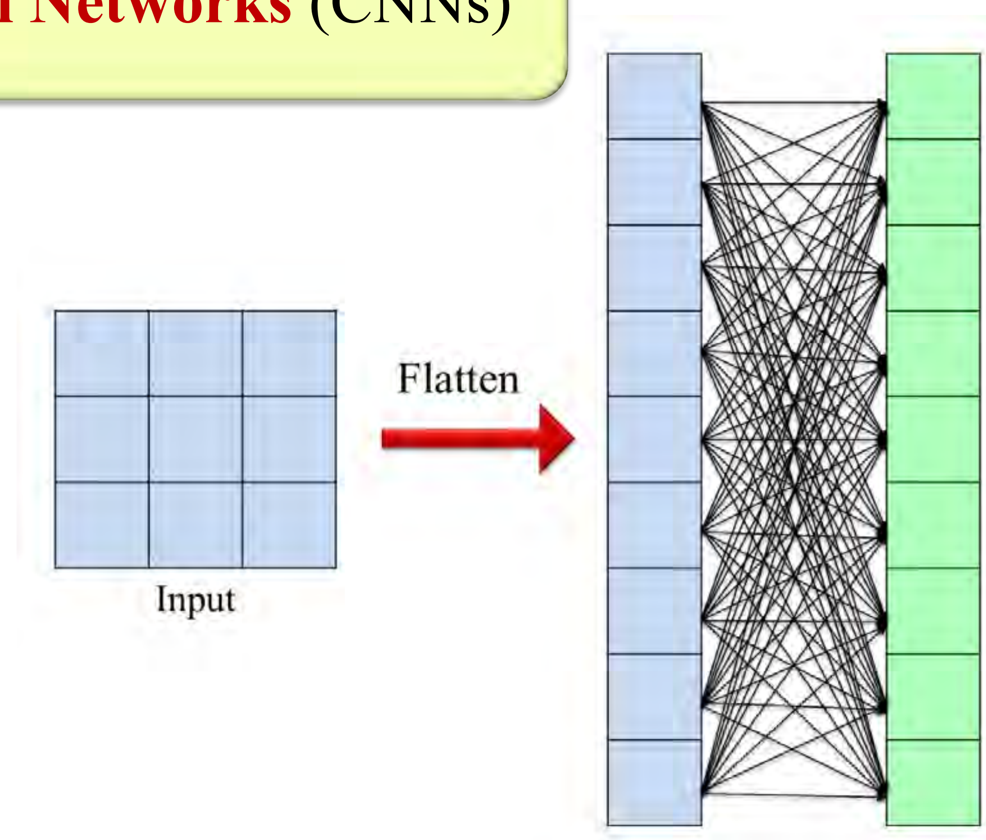


FC Neural Networks: Limitations

- How do we resolve the spatial invariance and scaling issue of the FC networks for image analysis tasks?

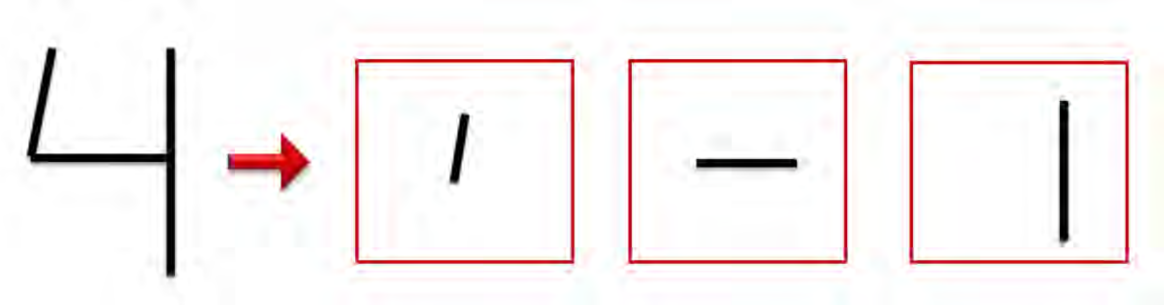
Convolutional Neural Networks (CNNs)





Convolutional Neural Networks

- To achieve translation invariance, CNNs **simplify** the problem of global feature detection by the problem of detecting **local features**.
- The global representation of the digit "4" consists of **local representations** of three symbols (i.e., local features).



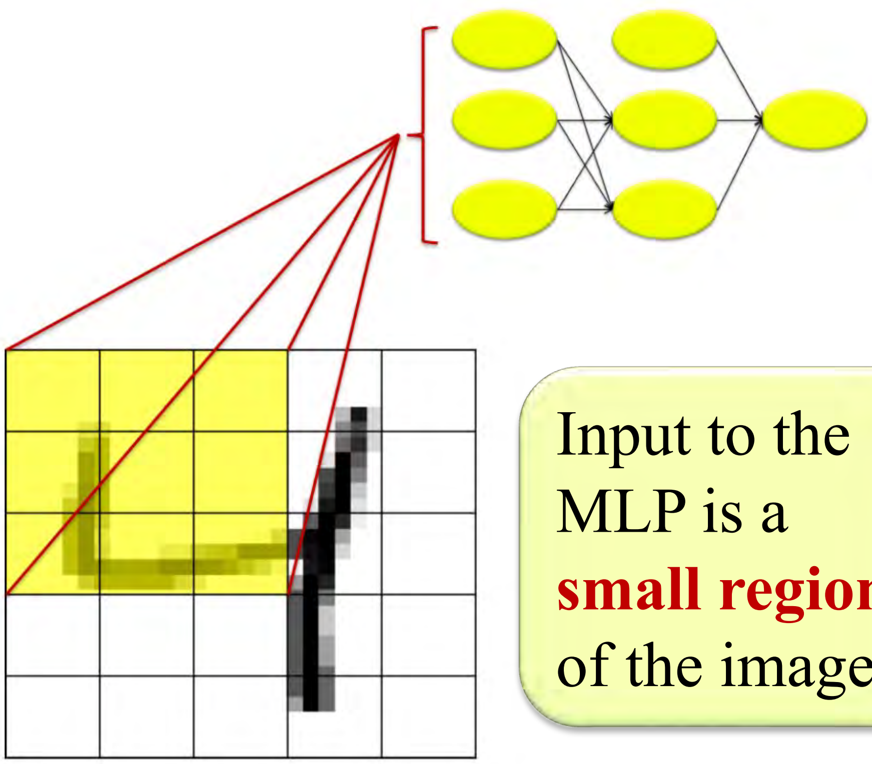
Convolutional Neural Networks

- Thus, instead of trying to globally detect "4", CNNs focus on detecting the **local features**, as follows.

Don't flatten the image.

Focus on a small region on the image.

Scan that region using an **MLP** to detect a local feature, e.g., horizontal stroke, vertical stroke.

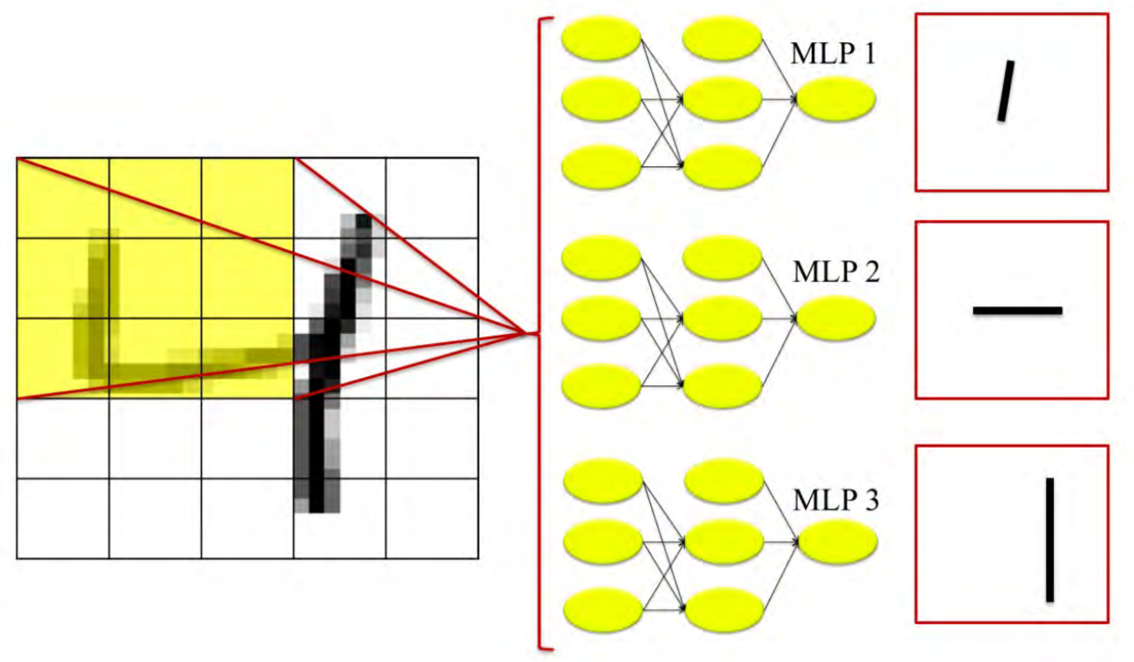


Input to the MLP is a **small region** of the image

Convolutional Neural Networks

- In the simplified example, we want to detect **three local features**: a tilted stroke, a horizontal stroke, and a vertical stroke.

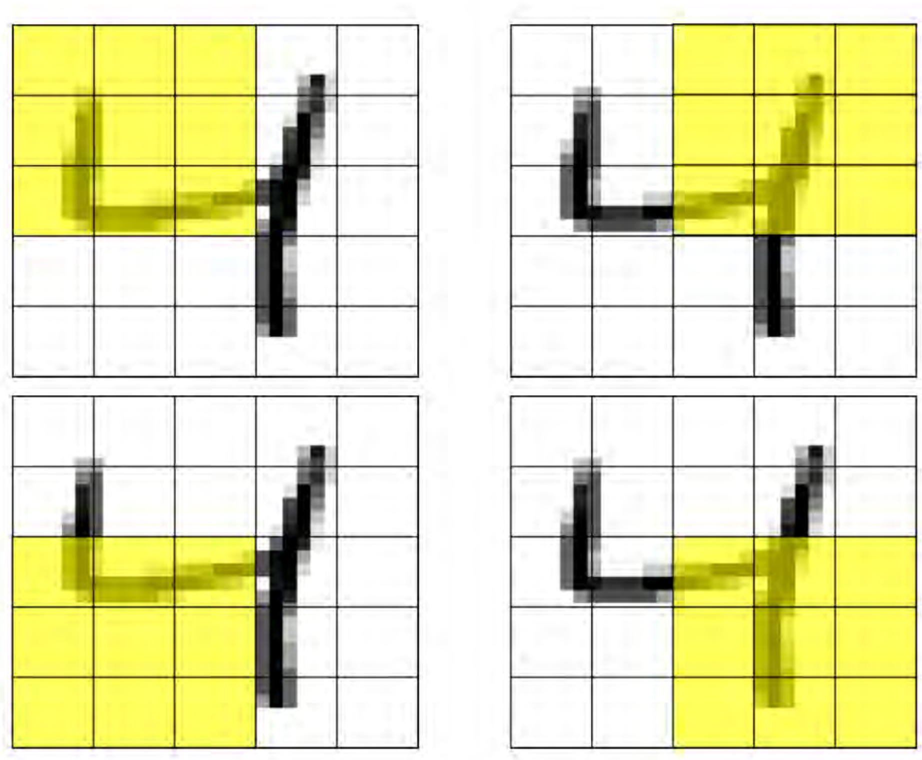
Thus, we employ **three MLPs** and use those to scan small regions of an image to detect the three local features.



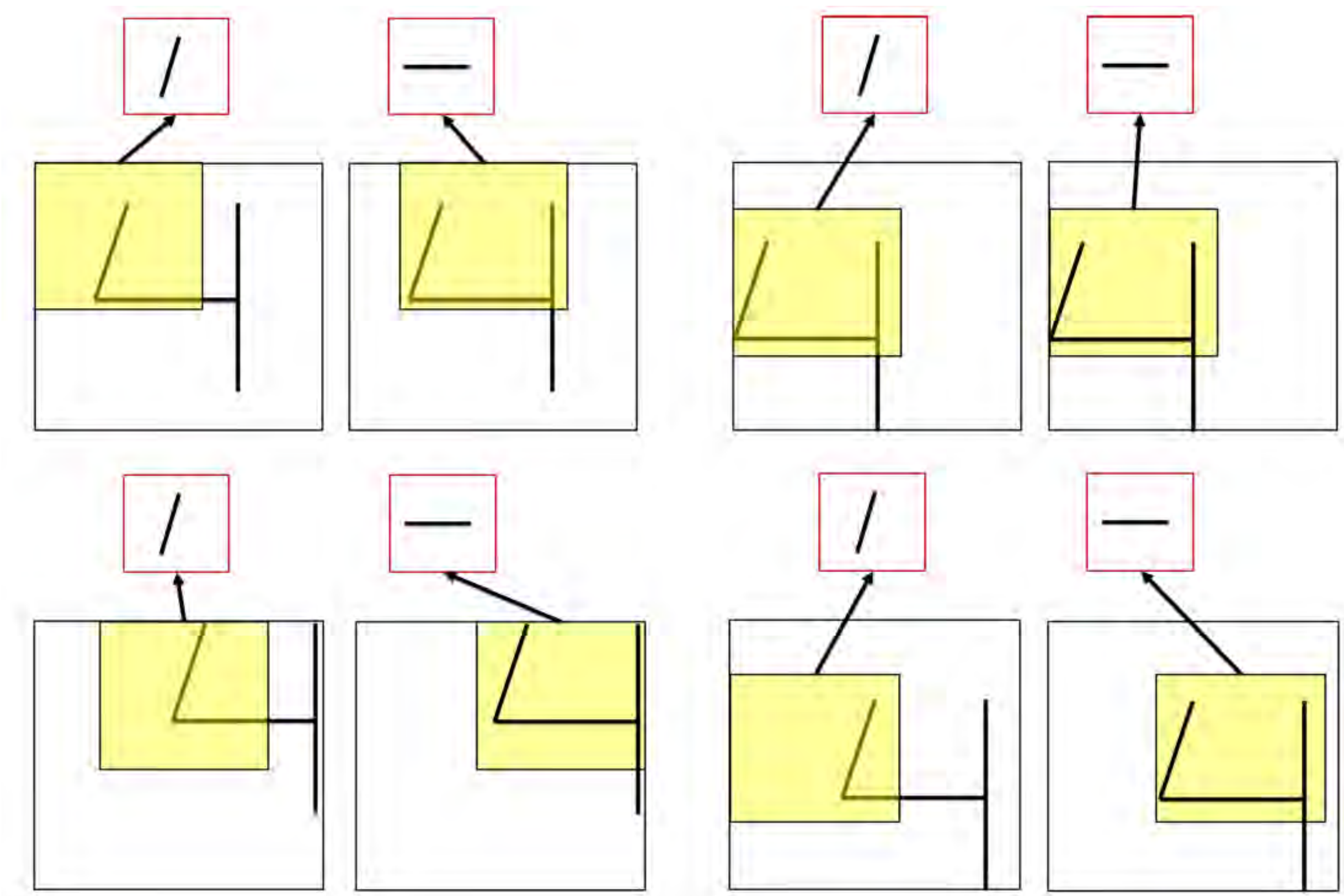
Convolutional Neural Networks

- The MLPs will process the entire image **successively scanning** every small region.
- Thus, if these features appear anywhere on the image, MLPs will be able to detect those.

Note the MLPs have **small receptive fields**!

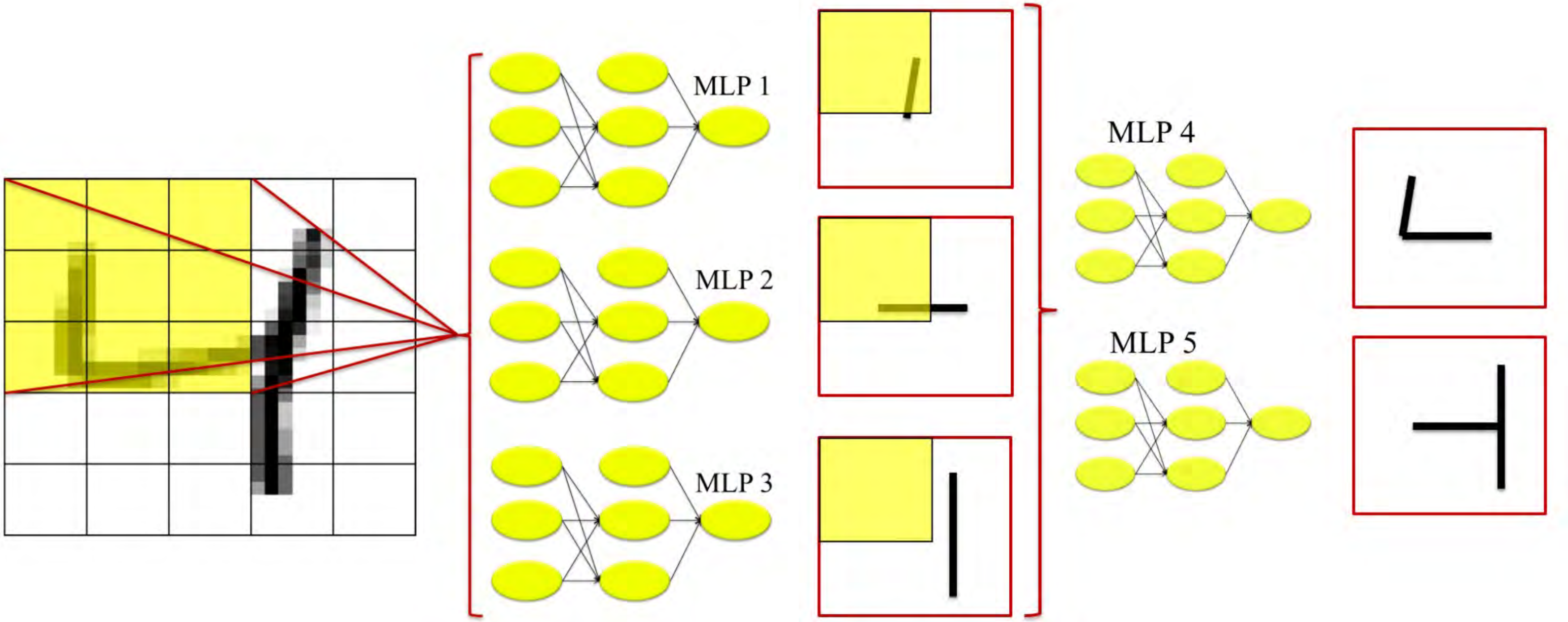


- The use of multiple MLPs with small receptive fields enables CNNs to **achieve spatial invariance**.
- Irrespective of the location of “4” in the following images, a CNN can detect the tilted and horizontal strokes.



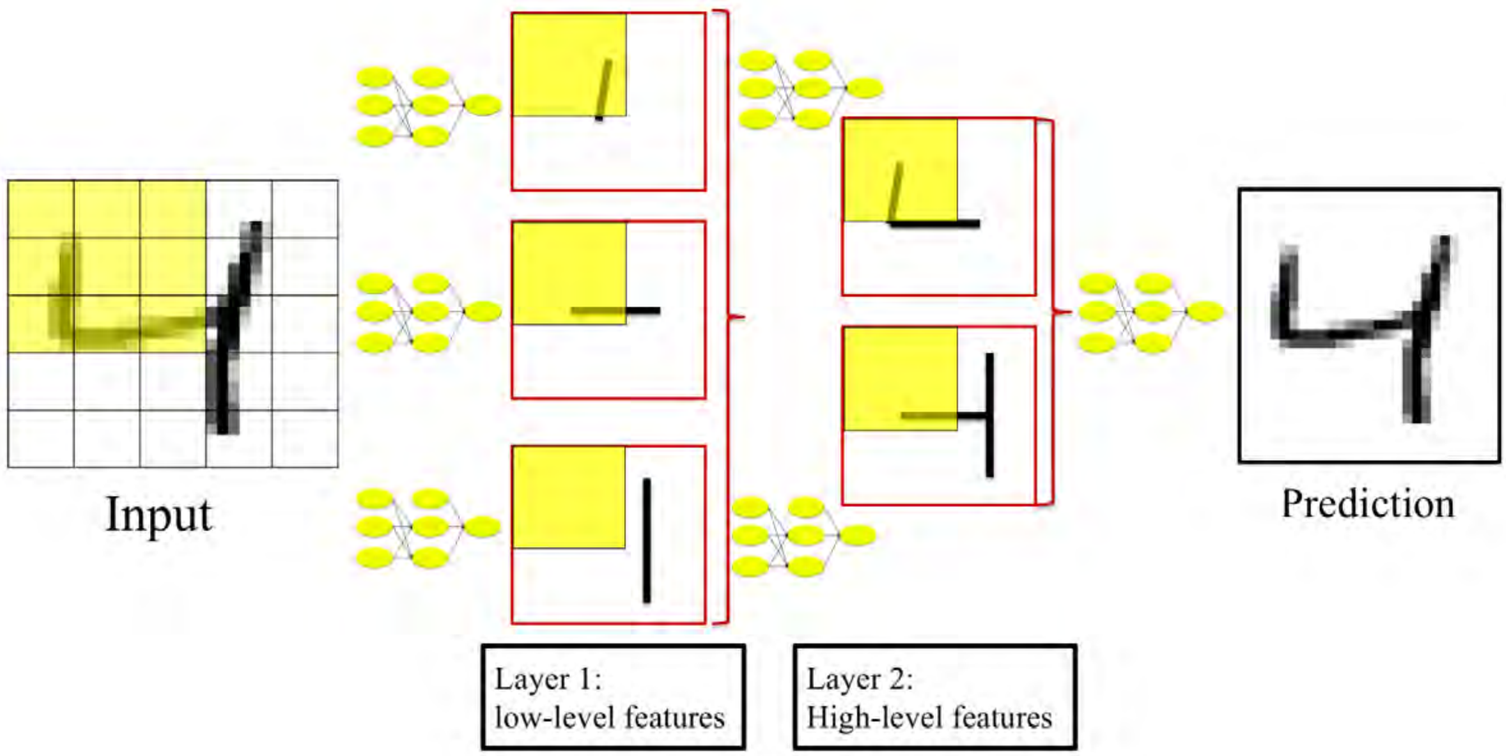
Convolutional Neural Networks

- After recognizing the three low-level features, a CNN **learns high-level features** by employing more MLPs that operate on these low-level feature maps.



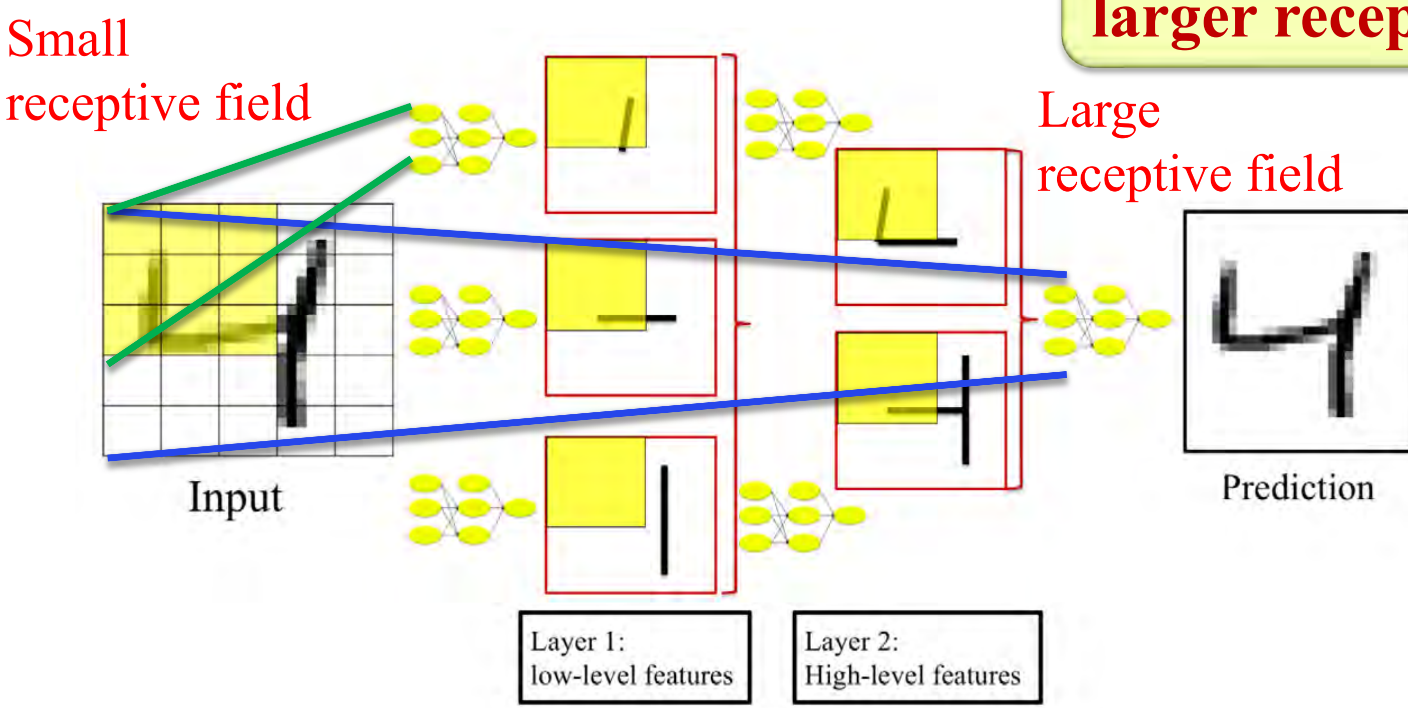
Convolutional Neural Networks

- Finally, a CNN employs another MLP to use the high-level features as input for detecting **class-level information**, such as whether the image represents “4”.

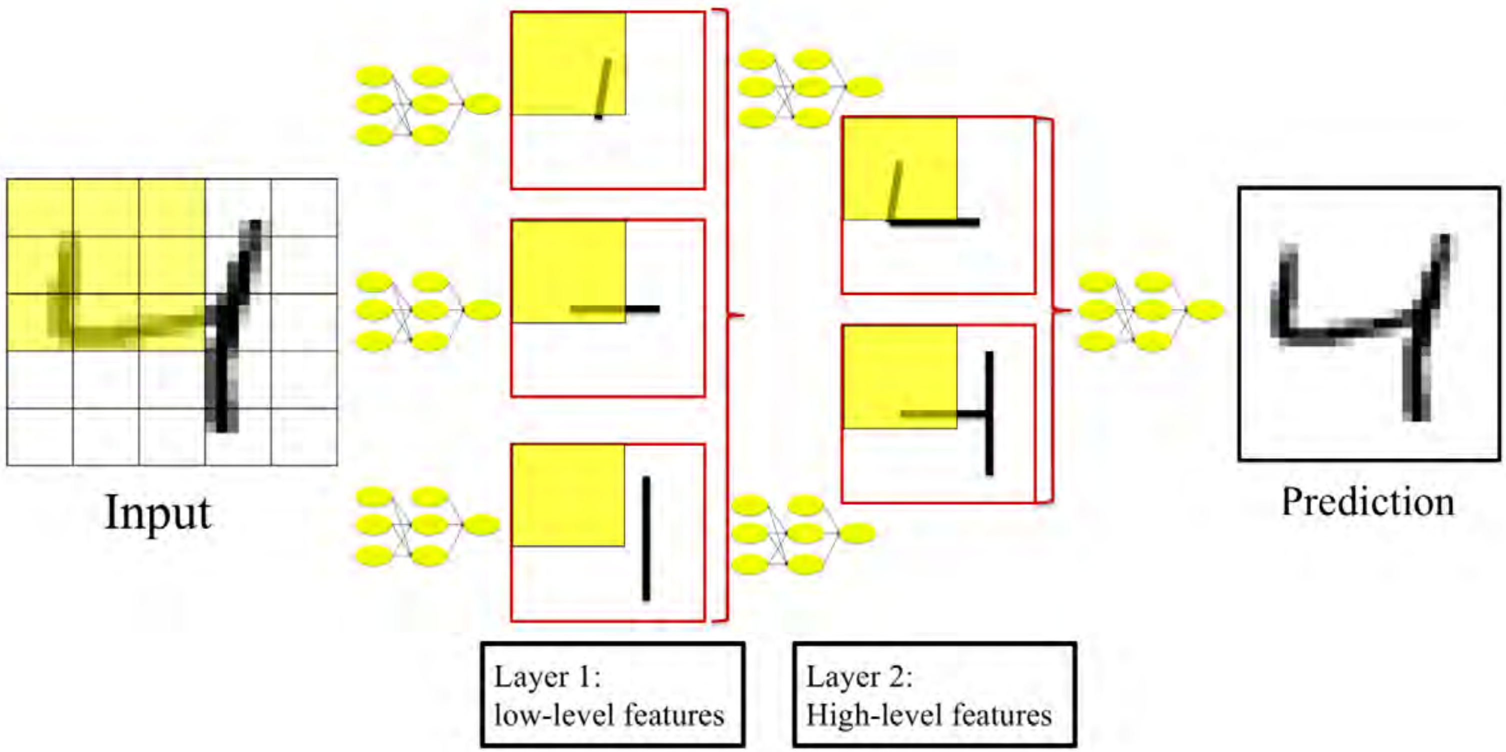


- Thus, we can design a **new type of MLP architecture** by using multiple MLPs.
- The architecture is **hierarchical**, i.e., divided into multiple layers.
- In each layer we have a set of MLPs to detect the local features by scanning small spatial regions (small receptive field).

Each successive layer has **larger receptive fields**.



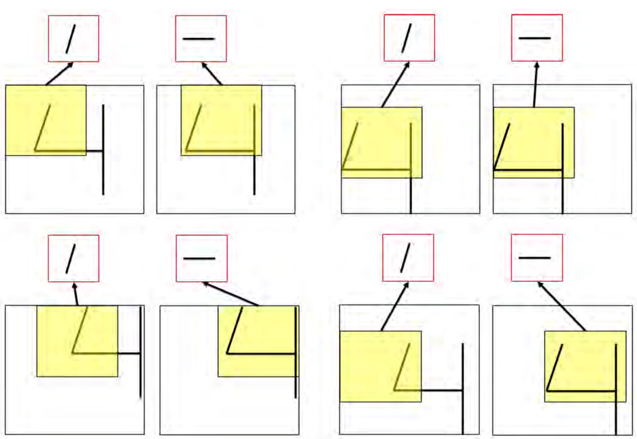
- Each layer learns features of **increasing complexity** by scanning the feature maps from the previous layer.
- The **final layer combines** the high-level representations to determine the class information.

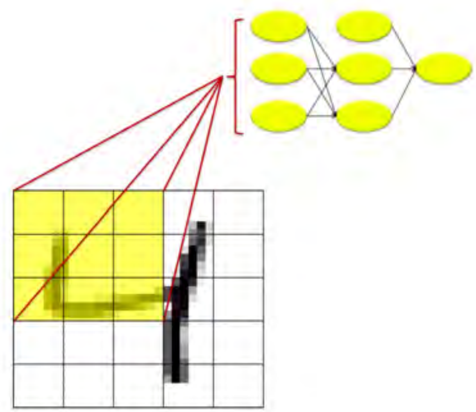


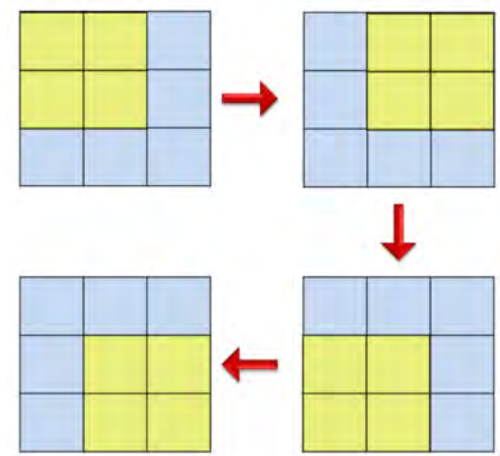
This new architecture forms the basis of **Convolutional Neural Networks**!

Convolutional Neural Networks

- CNNs can learn spatially invariant patterns.
- Due to smaller receptive fields, number parameters don’t increase with the image resolution.
- Filters are shared.

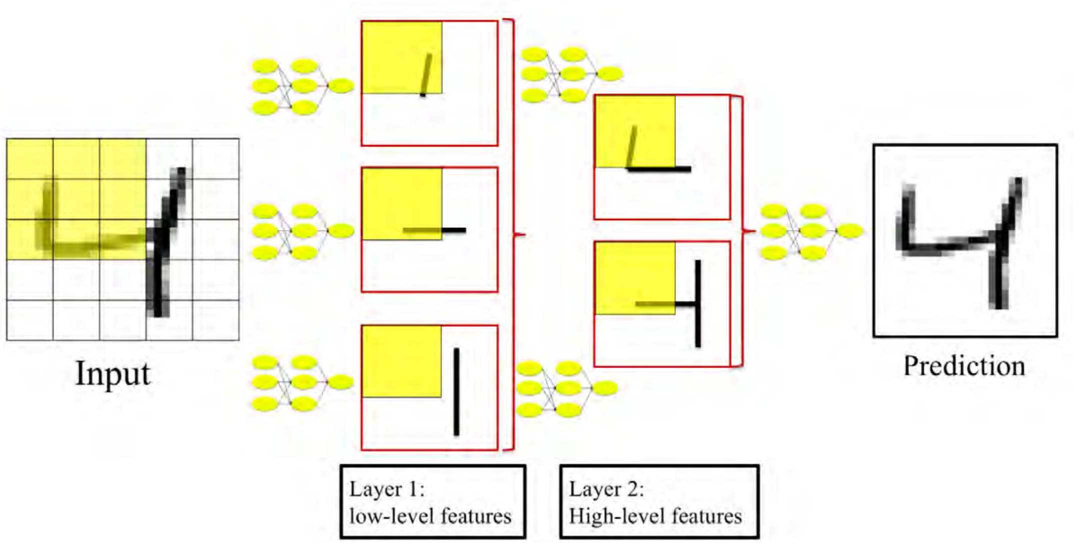






Convolutional Neural Networks

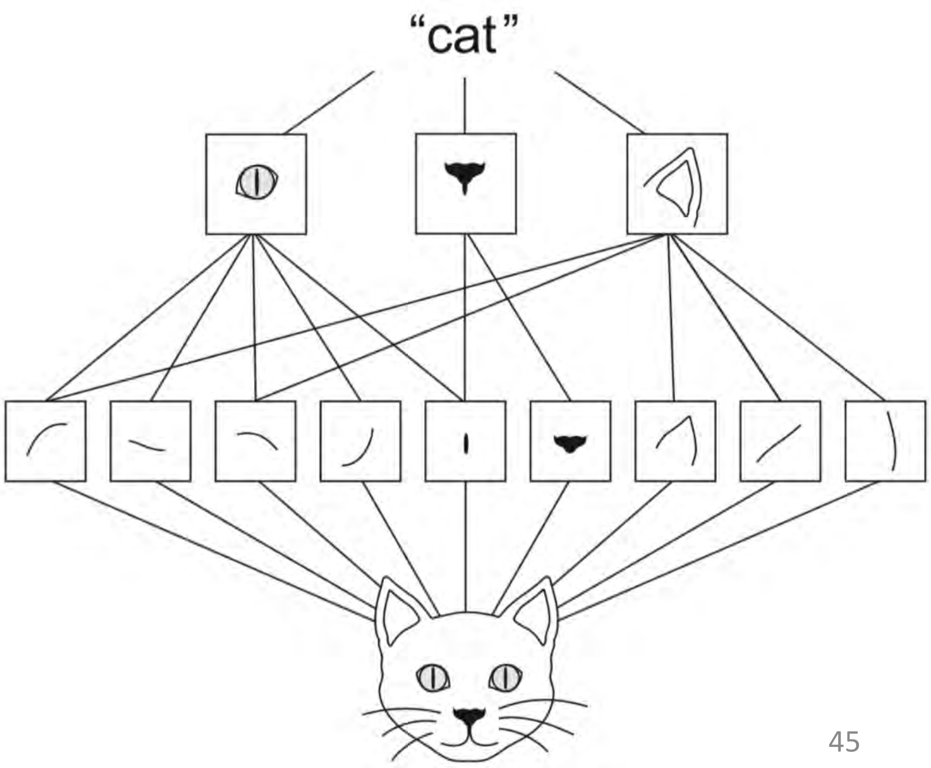
- So far, we have discussed **two benefits** of CNNs:
  - Spatial invariance
  - Scalability with respect to the input size
- Another benefit is **compositionality**.
- Each filter composes a local patch of lower-level features into a **higher-level representation**.



Convolutional Neural Networks

- This composition allows CNNs to learn **more rich features** deeper in the network.
- For example, a CNN may build edges from pixels, shapes from edges, and then complex objects from shapes.

The ability to build higher-level features from lower-level ones is exactly why CNNs are **so powerful in computer vision**.

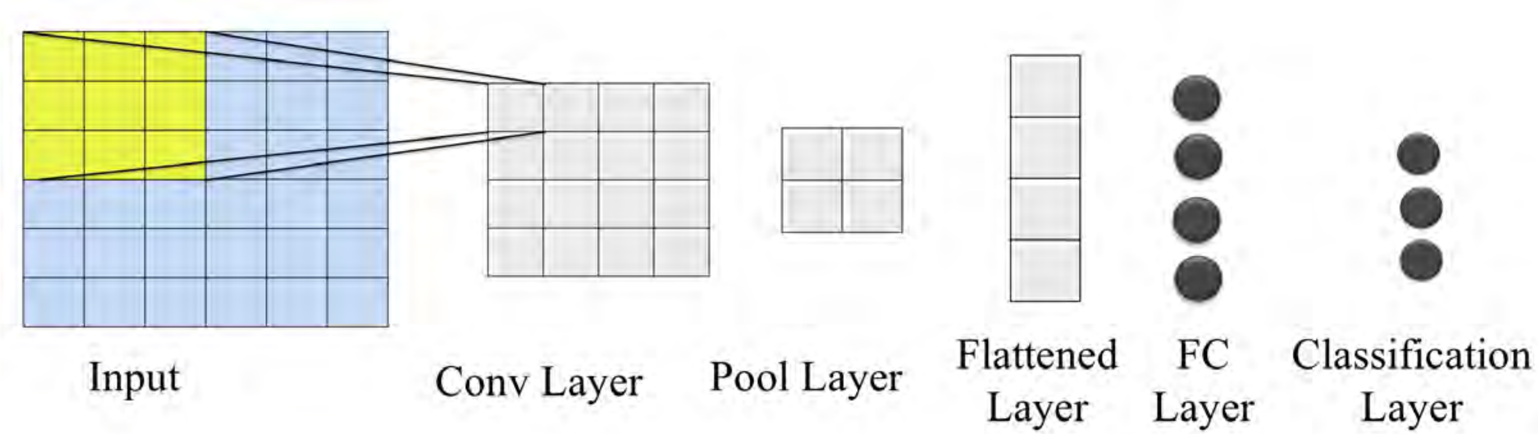


Convolutional Neural Networks

- CNNs are designed based on the functions of **visual cortex**.
- First studied by David H. Hubel and Torsten Wiesel during **1958 and 1959**.
- In **1980** Kunihiko Fukushima proposed an improved Neocognitron model for visual cortex.
- Based on the above theories, **in 1998**, the first effective model of CNN was developed by **Yann LeCun**.

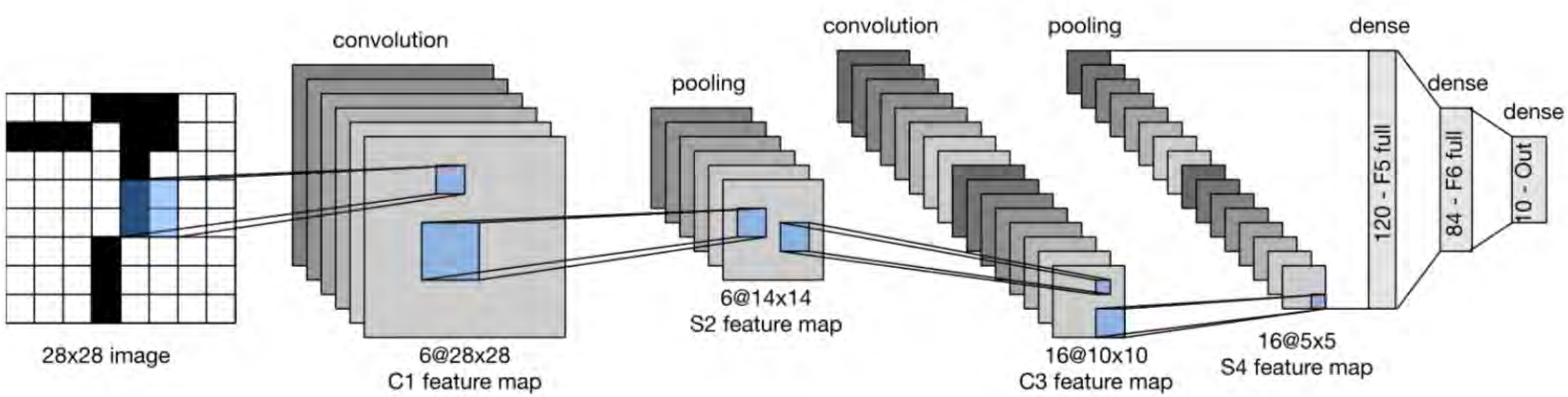
Convolutional Neural Networks

- The **main layers** of a CNN are:
  - Convolutional (Conv)
  - Activation (usually included with the Conv layer)
  - Pooling (Pool)
  - Fully-connected (FC)
  - Classification or Softmax
- **Stacking** a series of these layers in a specific manner yields a CNN.



Convolutional Neural Networks

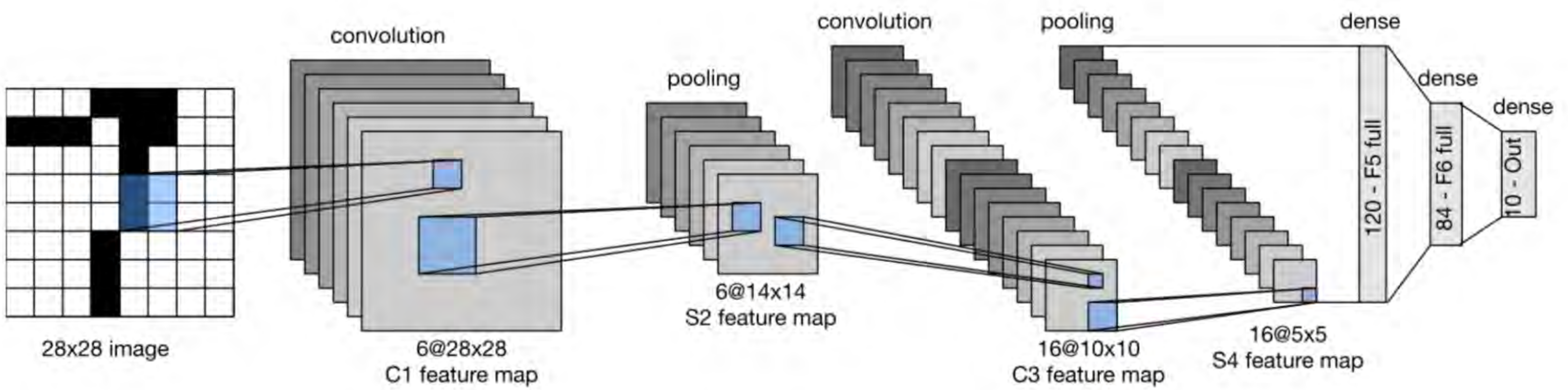
- LeCun used the **Bacpropagation** algorithm for training his famous LeNet-5 CNN classifier.
- It was commercially successful for its use in handwritten check numbers recognition in banks.



LeNet-5

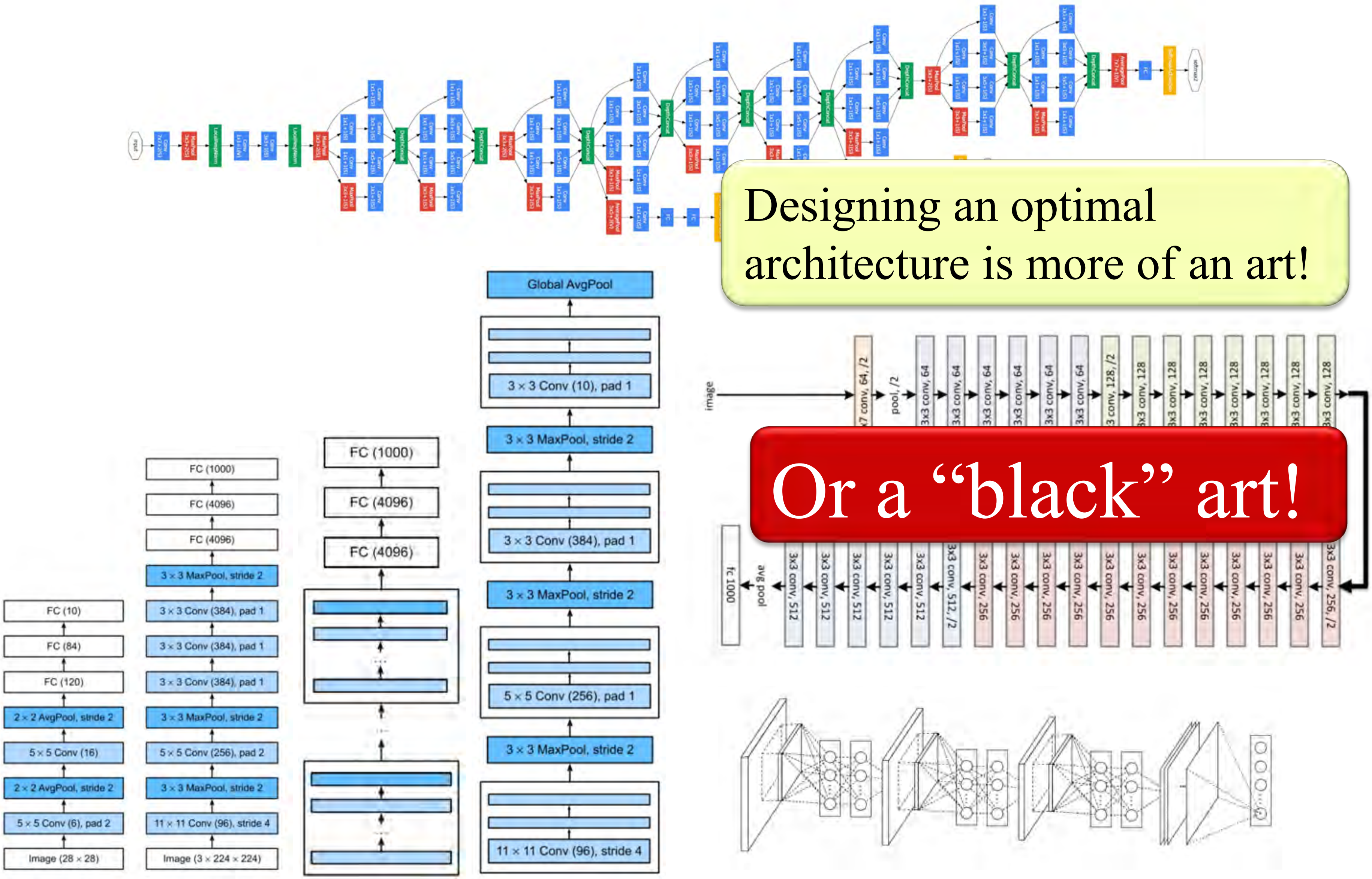
Convolutional Neural Networks

- CNN **architectures vary** due to the way the layers are stacked.
- It is mainly an **engineering problem**.



LeNet-5

- What is an **optimal CNN architecture**?

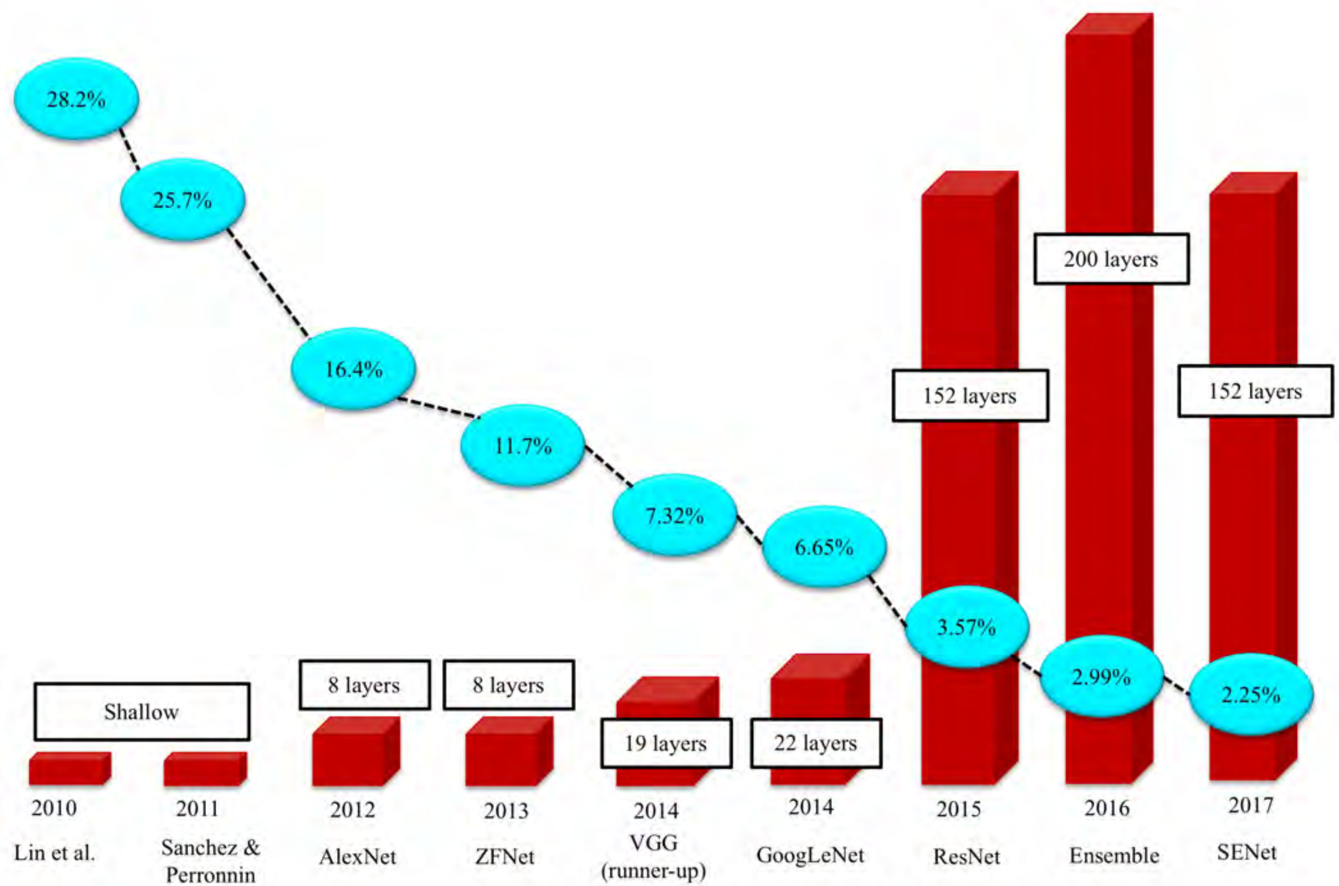


Designing an optimal architecture is more of an art!

Or a "black" art!

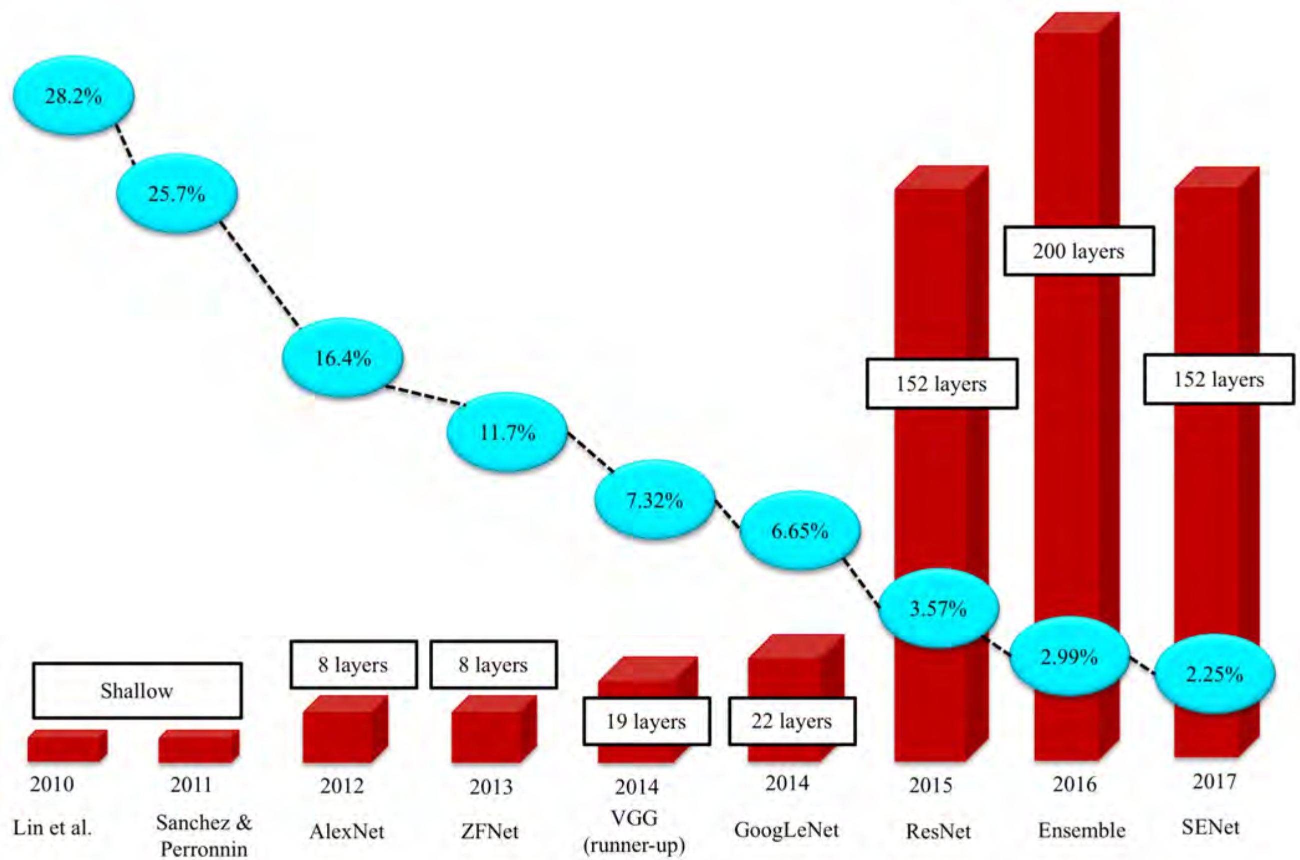
Progress in CNNs was Triggered by an Annual Vision Competition

- The ImageNet annual competition (1M images from 1K classes).
- The top-5 error rate for image classification **fell from over 26% to less than 2.3%** in just six years.



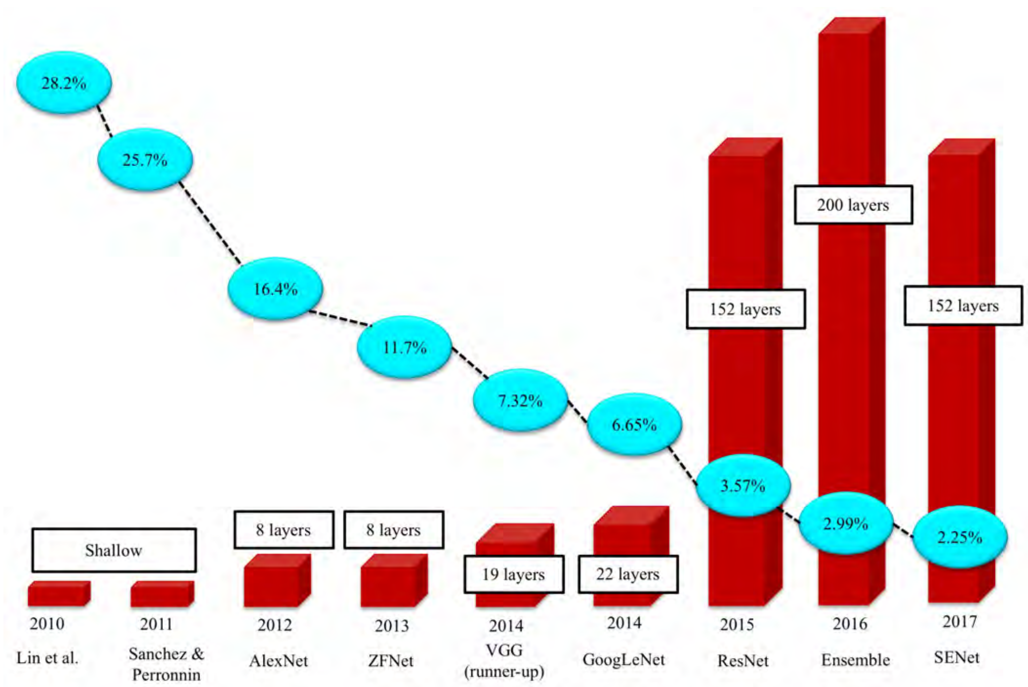
A Delayed Triumph of Deep Convnets

- From 2015 onwards typical layer size **increased to 152**.
- Why did it take ***more than a decade*** (since 1998) to create the first "deep" convnet (i.e., AlexNet)?



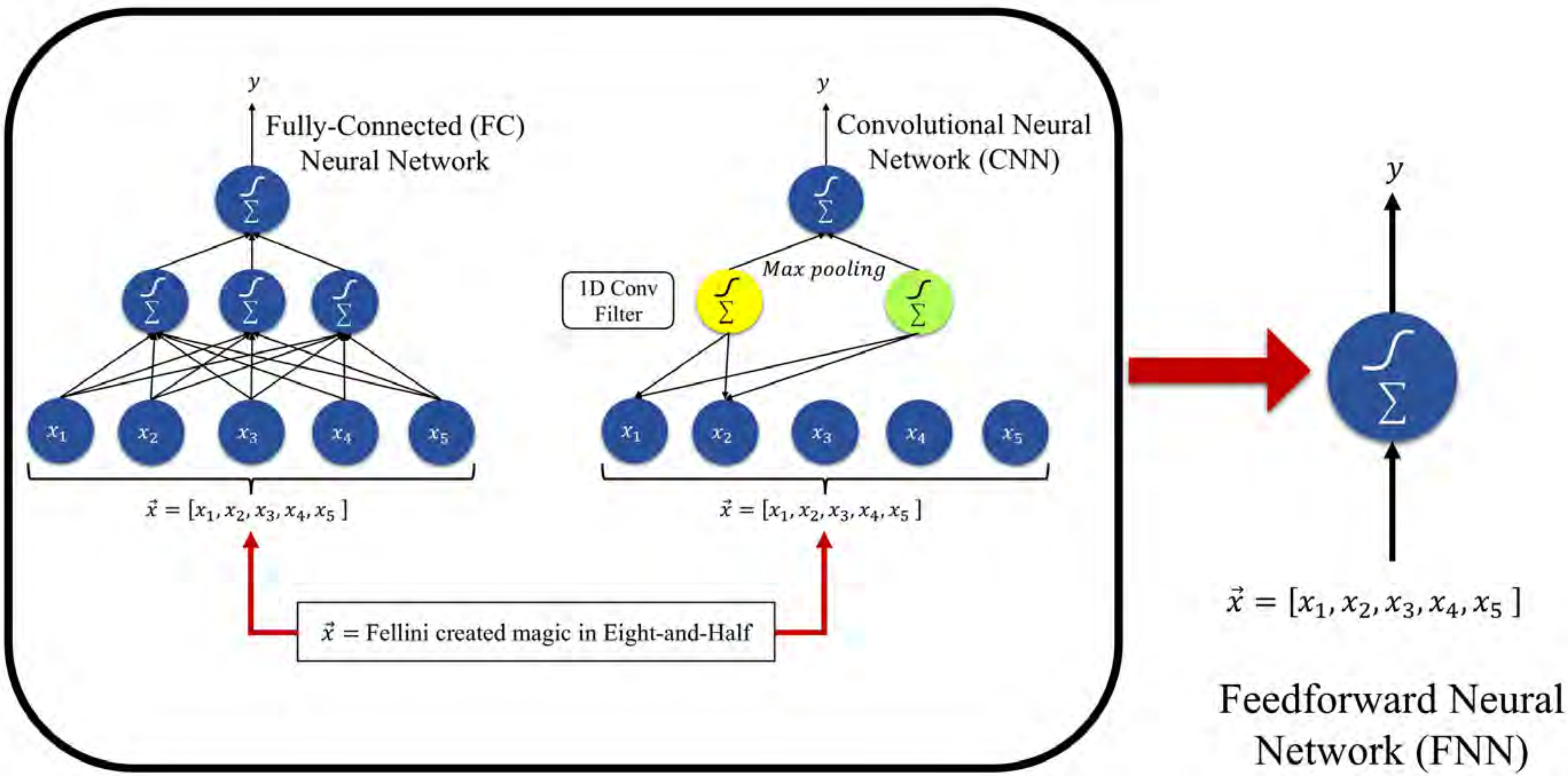
A Delayed Triumph of Deep Convnets

- Four developments contributed to the avalanche that started in 2012.
  - Hardware
  - Large datasets
  - Algorithm
  - Engineering tricks



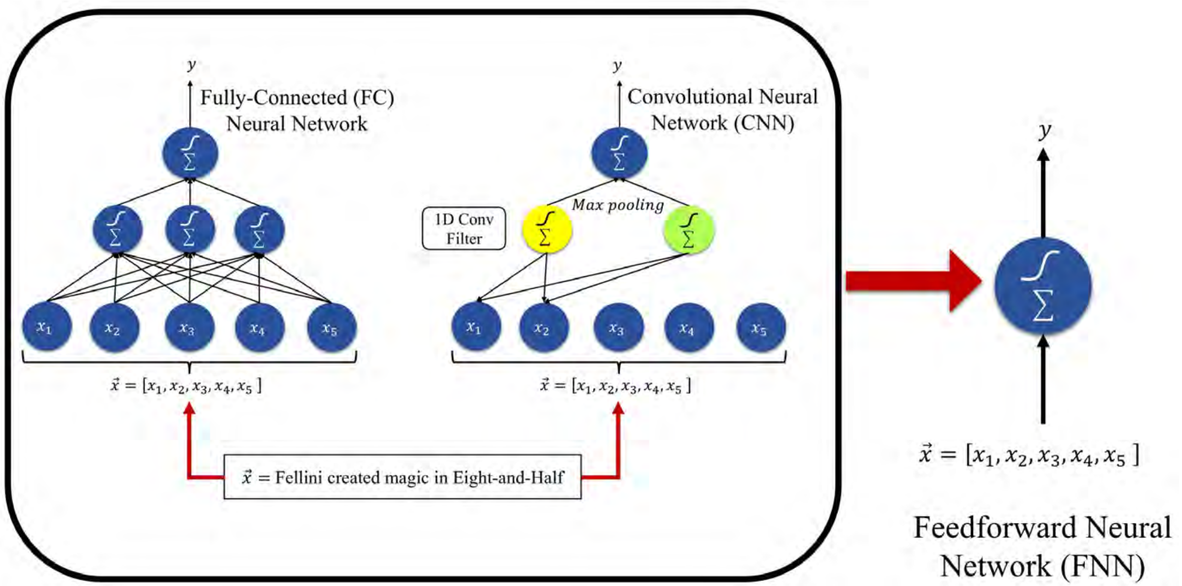
Limitations: Feedforward Networks

- Both FC and CNN models fall under the category of **feedforward neural network** (FNN).
- They require the **entire input data at each step** of the training.



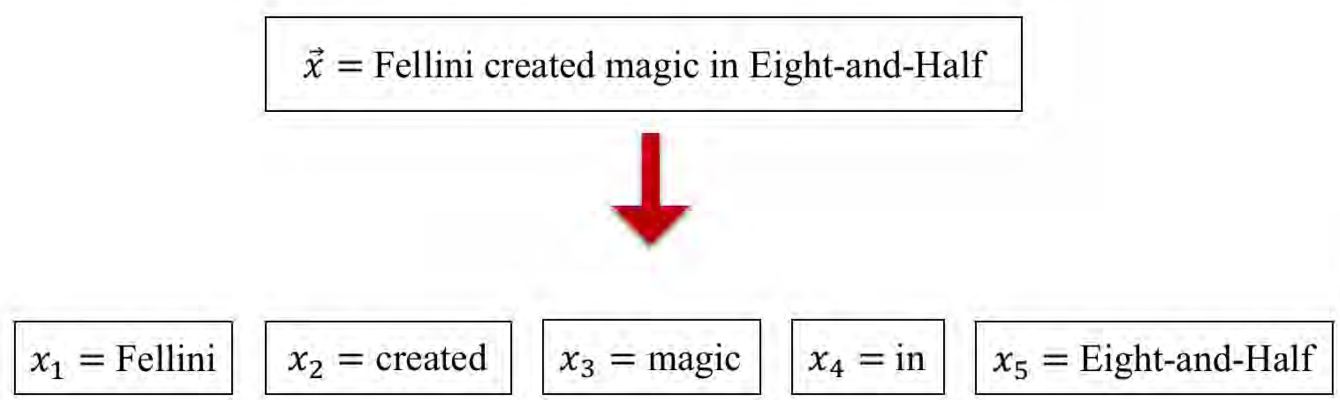
Limitations: Feedforward Networks

- FNNs are not always effective for sequential data (time-series data, text data).
- There are a couple of reasons.
  - Variable length input
  - Variable length output
  - Data has an inherent sequential structure



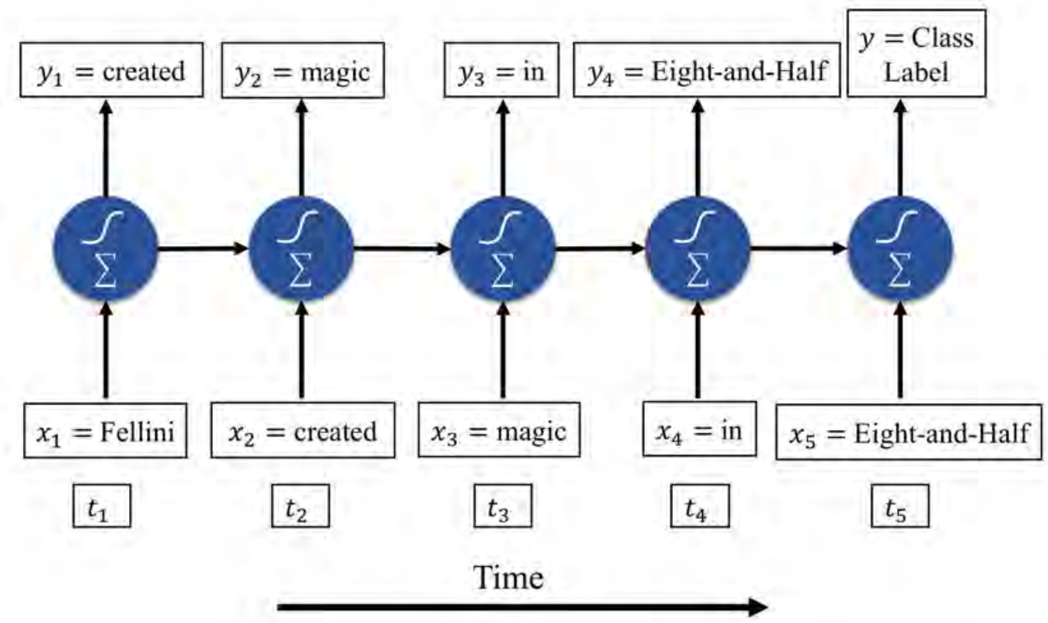
Recurrent Neural Network

- To address the learning problem from sequential data, Recurrent Neural Network (RNN) was developed.
- The RNN architecture is simple: it takes one input token at a time to predict the next token.
- First, we **break the input text** (e.g., a review) into a sequence of tokens (words/phrases).
- Then, use **each token as a single input** to sequentially train the model.



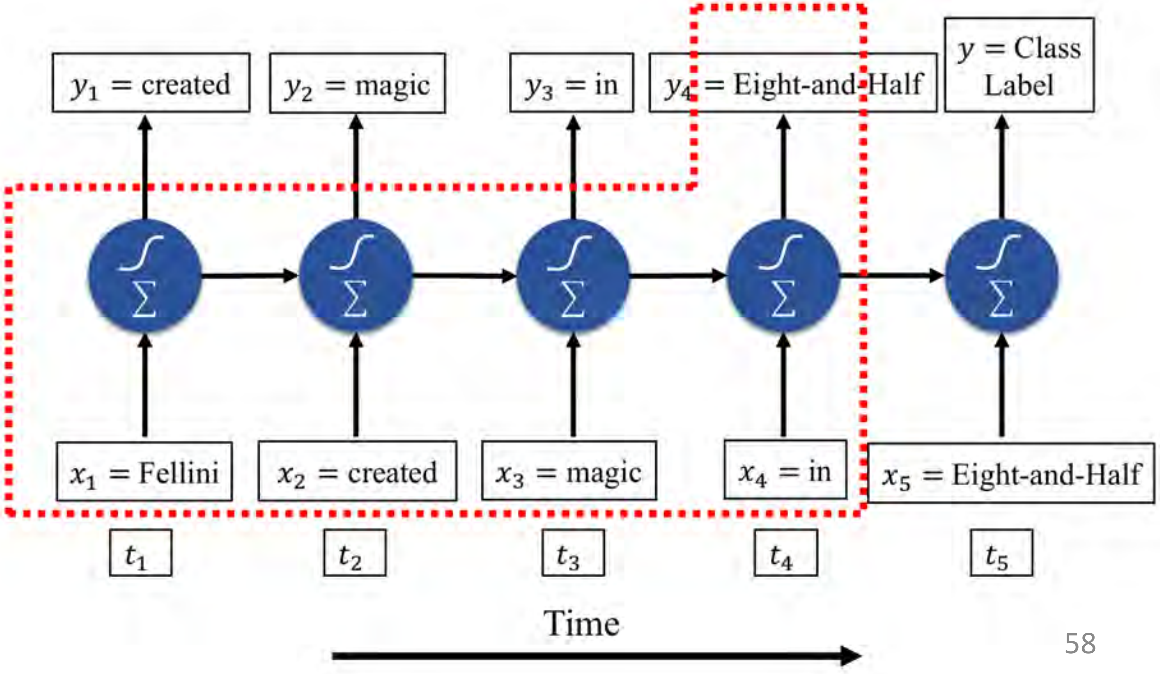
RNN Architecture: Motivation

- The following RNN architecture consists of a **single neuron in the hidden layer**.
- At **timestep t_1** this neuron takes an input (which is an embedding of the current word "Fellini").
- For each token, the **next token is its label**.
- The neuron at t_1 learns a representation of the input to predict next word "created" in the output.



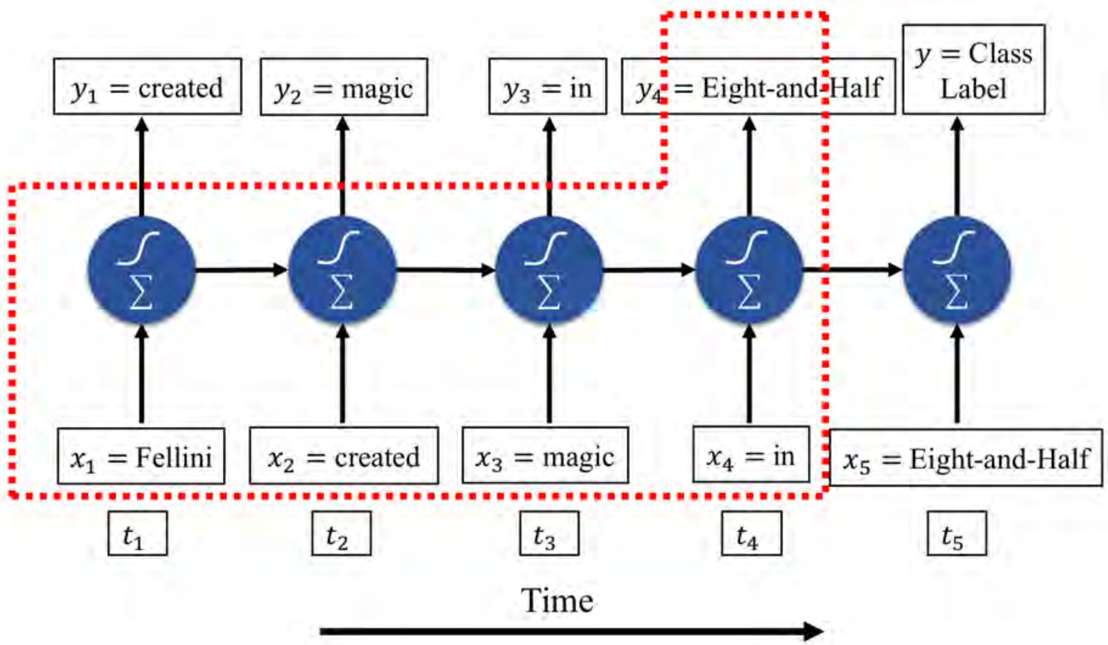
RNN Architecture: Motivation

- This network is trained for a single sample (e.g., a single review) by **repeatedly or recurrently** feeding each token over time.
- The output of each learning step is fed into the input of the next step.
- As a consequence, output at any timestep is computed based on the **input from all previous timesteps**.



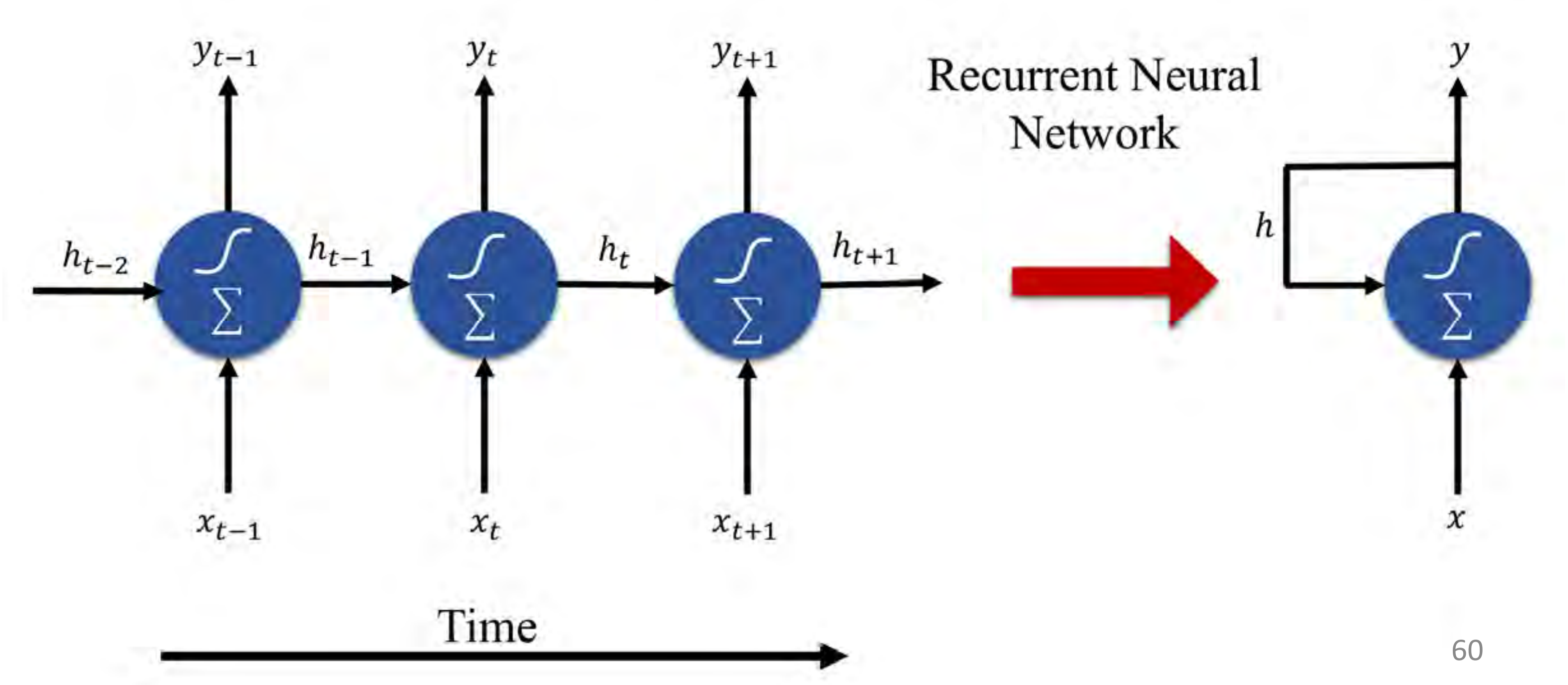
RNN Architecture: Motivation

- This novel technique of processing text data (or any sequential data) is **effective**.
- Because text processing is fundamentally about **creating memories**.
- Essentially the network is building some type of memory of the past words to infer the next word in the sequence.



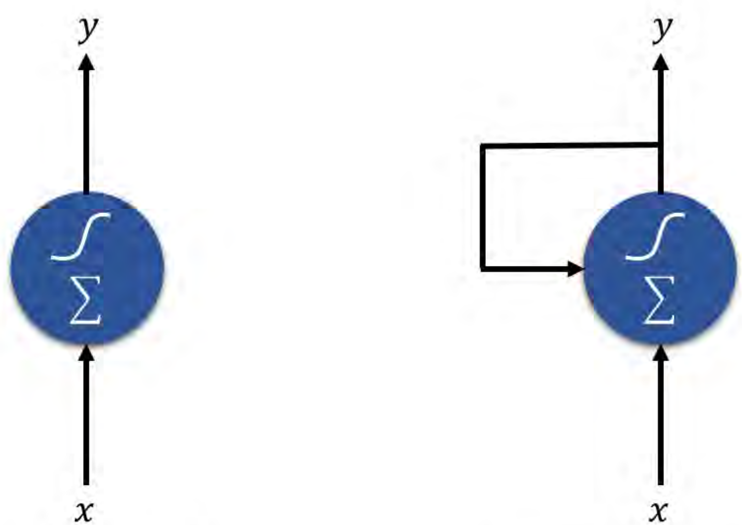
RNN Architecture: Motivation

- RNN is **unlike previous neural networks**, which don't have any loops.
- Right figure: RNN architecture
- Left figure: The RNN is **unrolled through time**.
- Left figure: The same recurrent computation is represented **once per timestep**.



RNN Architecture

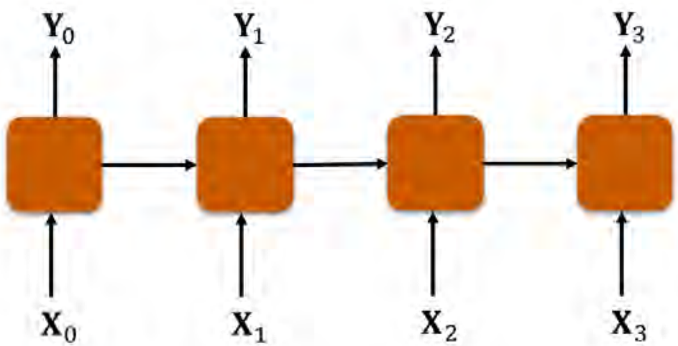
- A n RNN architecture looks very much like a Feedforward Neural Network (FNN) architecture.
- Except RNN has **connections pointing backward**.



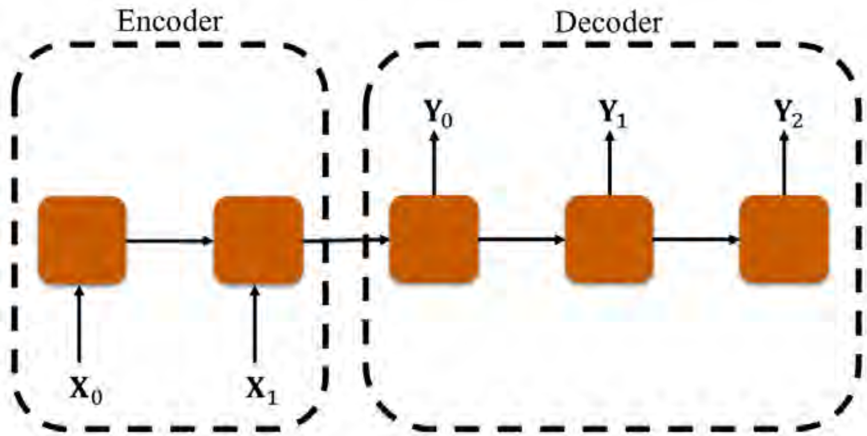
Feedforward Neural Network

Recurrent Neural Network

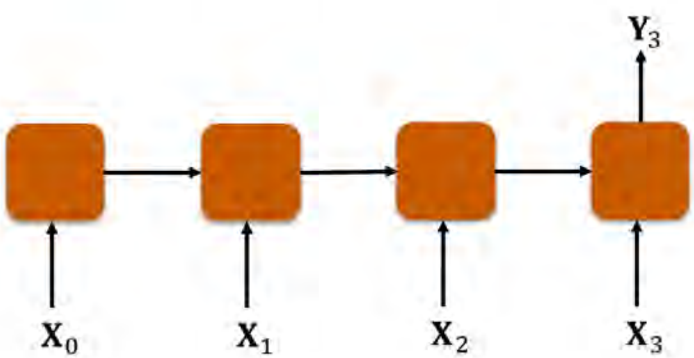
- Many-to-Many: Sequence-to-Sequence
- Many-to-One: Sequence-to-Vector
- One-to-Many: Vector-to-Sequence



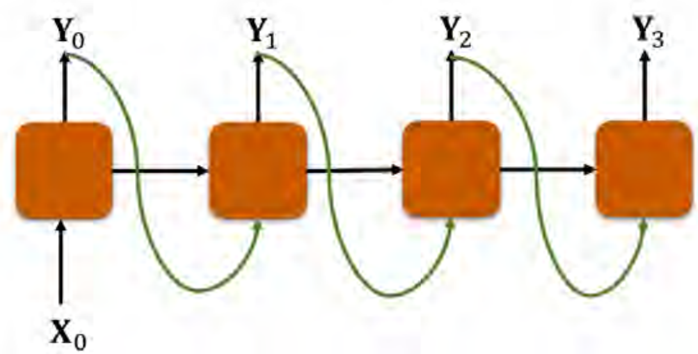
Many-to-Many: Sequence-to-Sequence
Synced: Input Sequence = Output Sequence



Many-to-Many: Sequence-to-Sequence
Input sequence ≠ Output Sequence



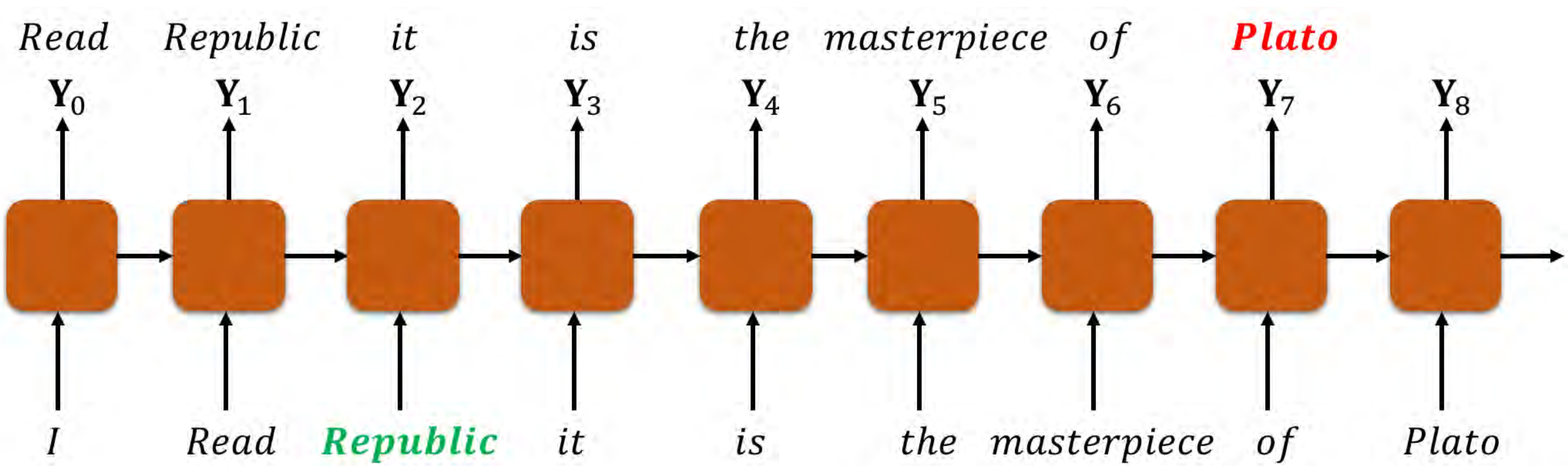
Many-to-One: Sequence-to-Vector



One-to-many: Vector-to-Sequence

Limitation: RNNs Have a Short-Term Memory

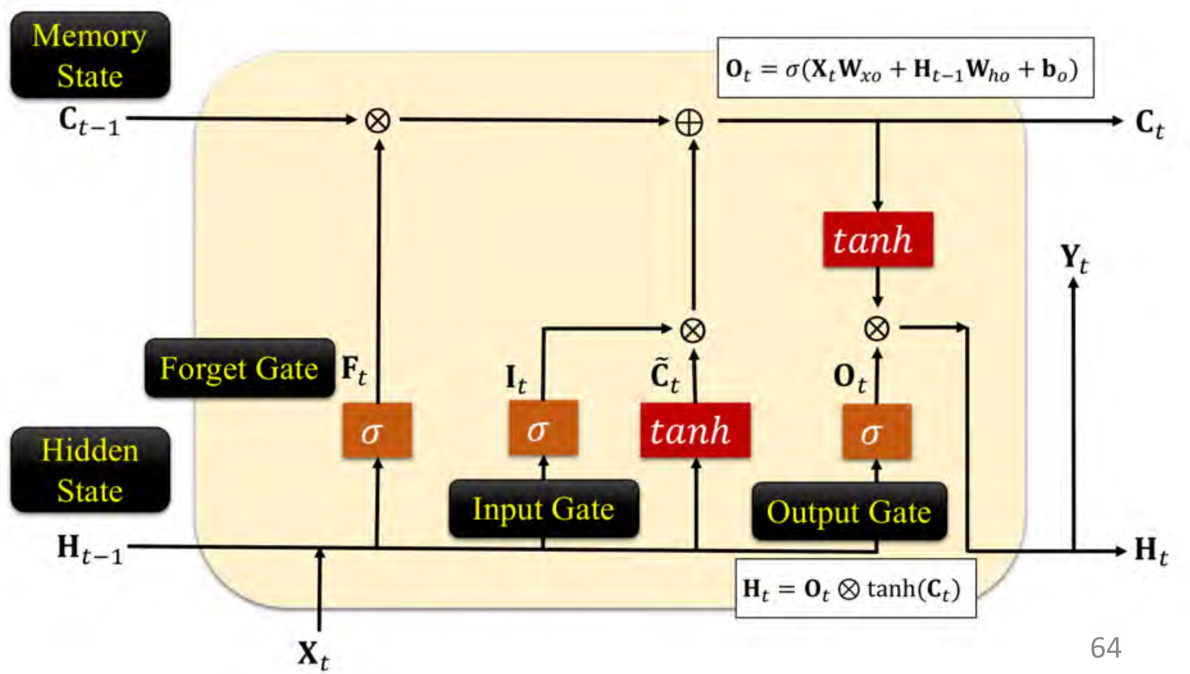
- In RNNs, the long-range states are **less updated**.
- As a consequence, RNNs are **unable to retain long-term memory** making inferences poor for very long sequences.



Gated RNNs

- The short-term memory problem is addressed by developing **gated RNNs** for retaining long-range dependency.
  - Gated Recurrent Unit (GRU) *2014*
  - Long Short-term Memory (LSTM) *1997*

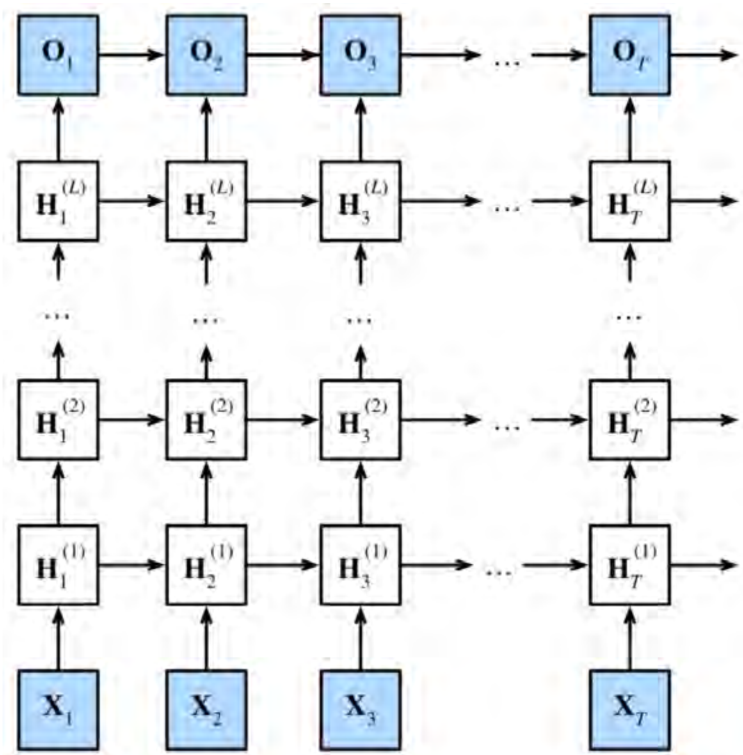
Retain what is important in the long-term state and *forget* the rest.

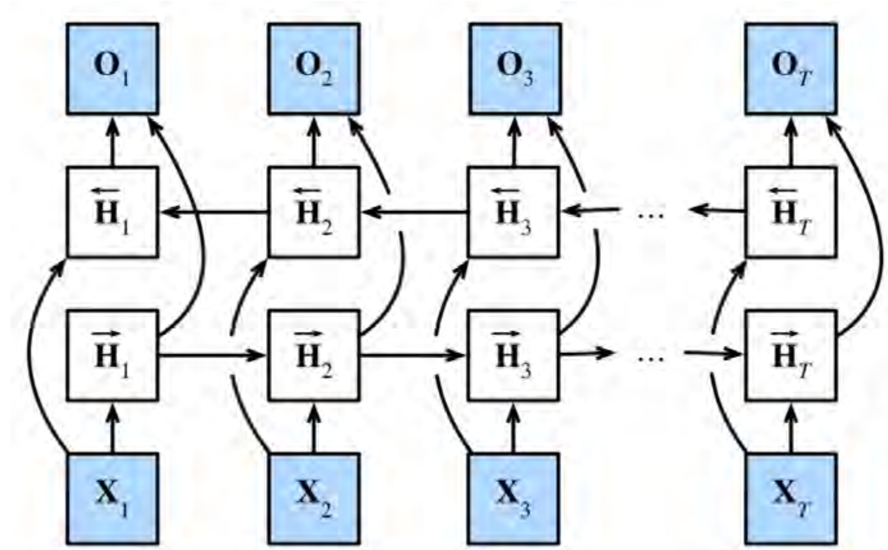


Increase Representational Power of RNNs

- Two techniques.
  - Deep RNNs
  - Bidirectional RNNs

Inefficient: Due to the sequential nature of processing, each step depends on the previous one, preventing parallel processing.





What We Will Cover

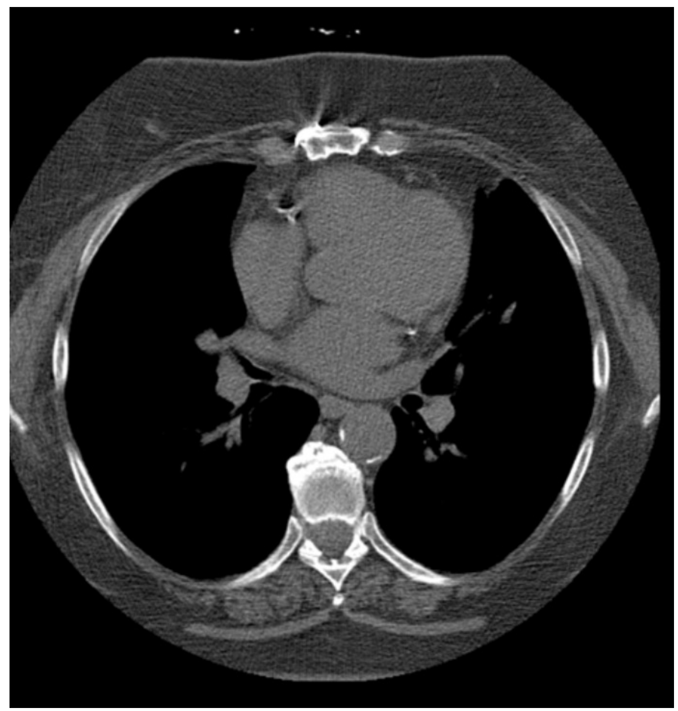
- Modern AI: Machine Learning and Deep Learning
- Discriminative AI
- Neural Networks & its Variants
- Fully-Connected Networks
- Convolutional Networks
- Recurrent Neural Networks
- Applications: discriminative AI in healthcare
- Generative AI
- Advancing AI in healthcare using generative AI
- Transformer Networks
- Language Models: Potentials and Limitations
- Our AI Future: Societal Implications

Motivating Case: Cardiac Images

- How can we design an AI to automatically detect cardiovascular disease from coronary artery calcification obtained via computed tomography (CT) images?

CNN

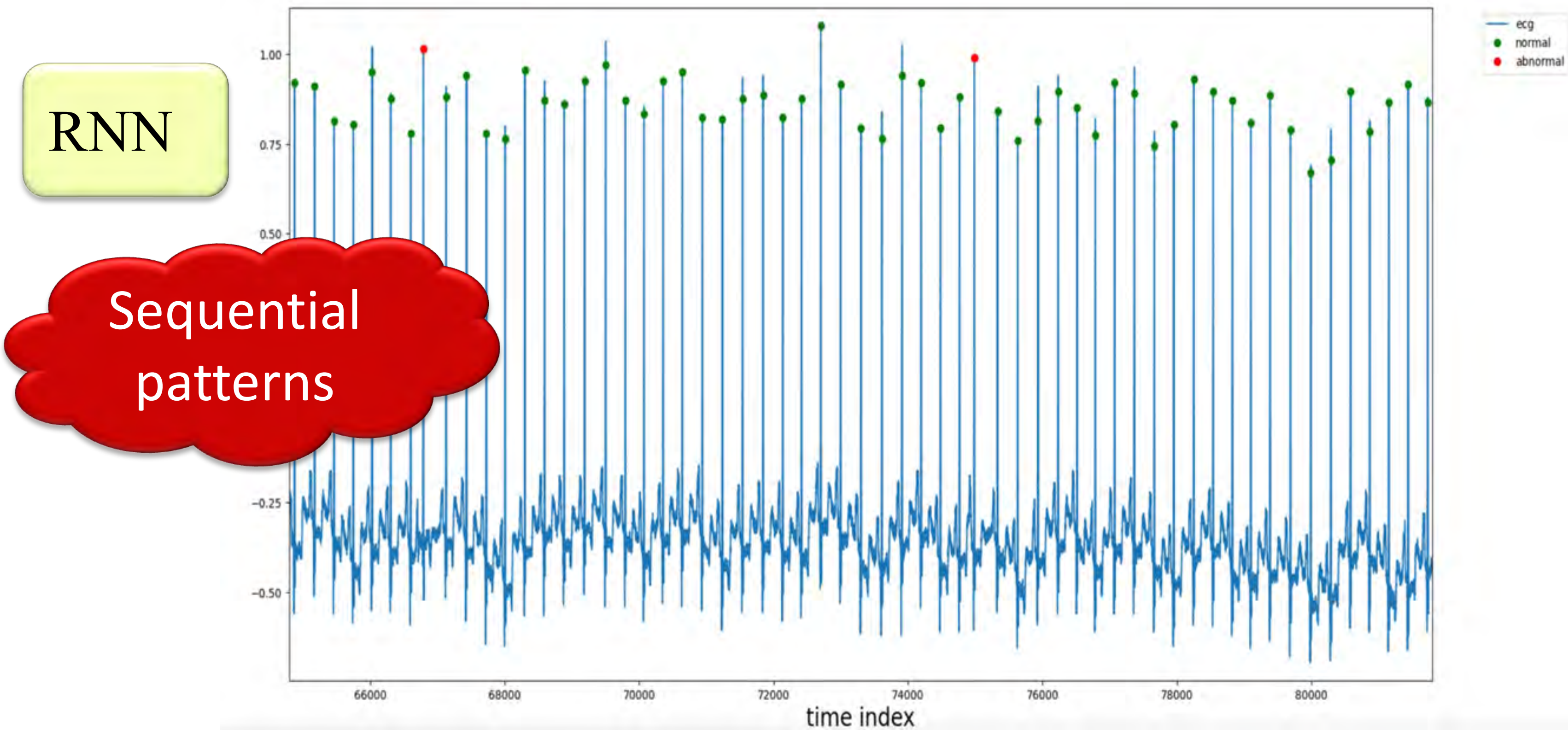
Spatial patterns



https://www.nature.com/articles/s41467-024-46977-3

Motivating Case: Heart Electrical Activity

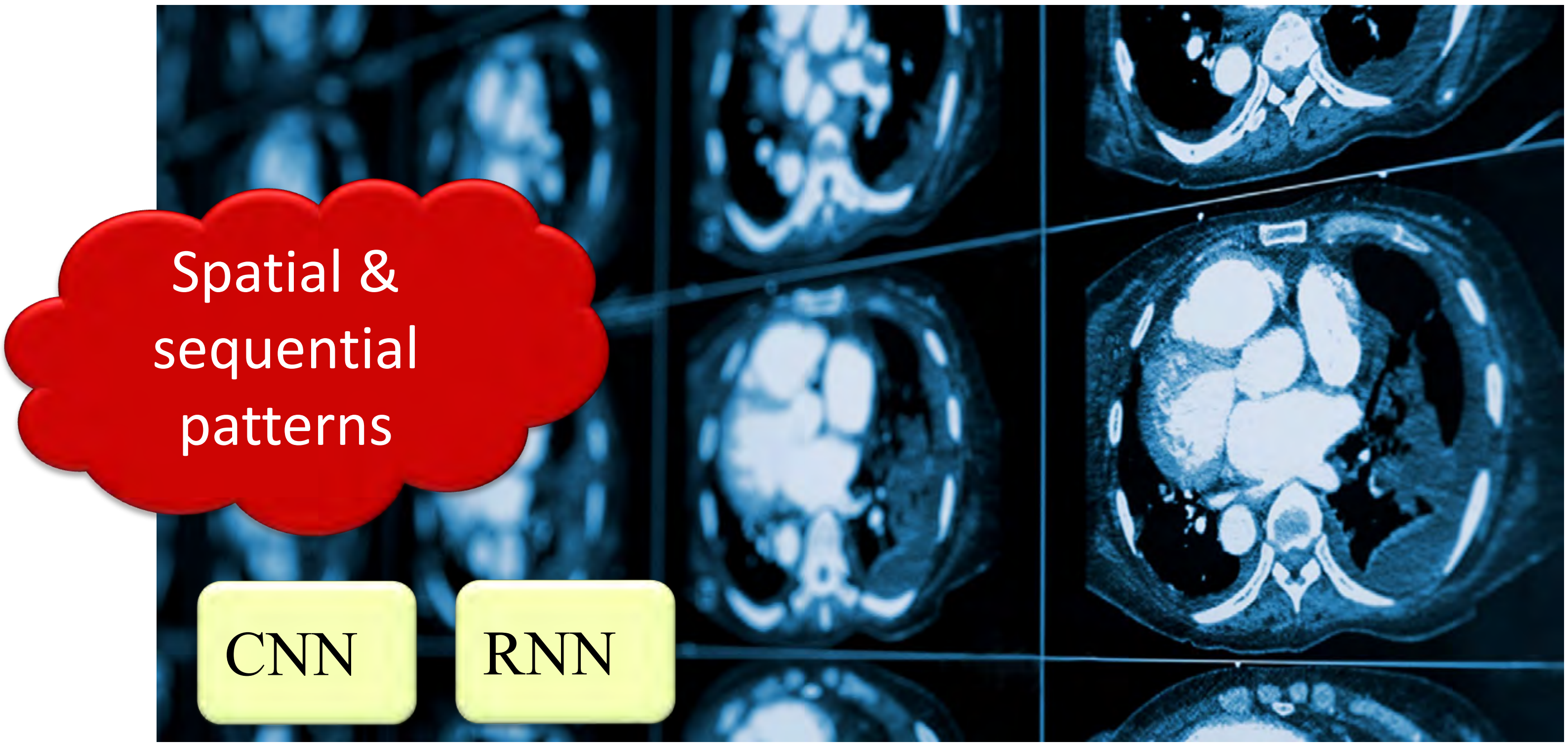
- How can we design an AI to automatically detect cardiac arrhythmia from electrocardiogram (ECG) data that represents different cardiac activities via electrical signals from heart?



https://www.mdpi.com/1424-8220/22/5/1928

Motivating Case: Series of Cardiac Images

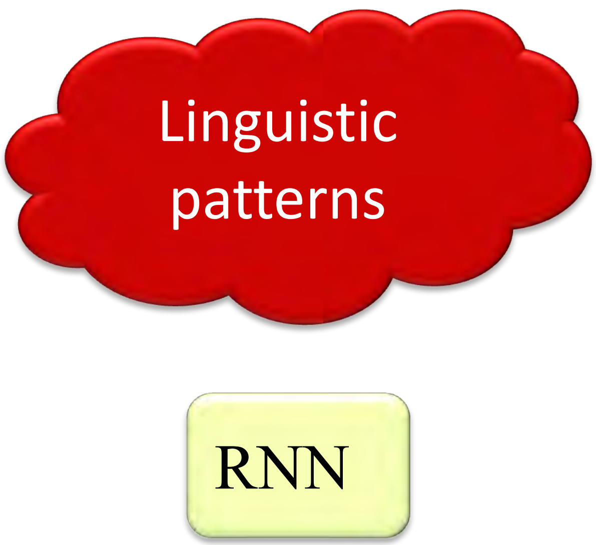
- How can we design an AI to automatically detect calcification by tracking the **progression** of calcification across multiple slices or phases of a cardiac cycle by a sequence of CT images?

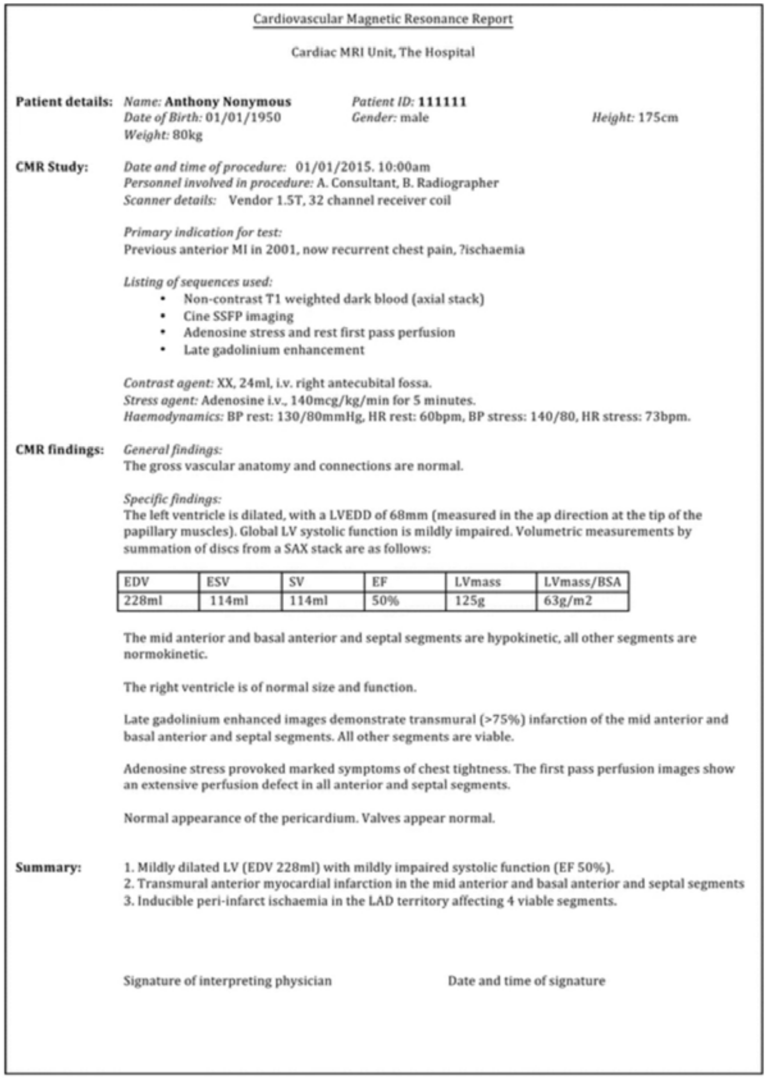


https://www.touchstoneimaging.com/chest-ct-scan/

Motivating Case: Test Report Text Data

- How can we design an AI to identify hypertrophic cardiomyopathy patients from cardiac magnetic resonance textual reports?





Cardiovascular Magnetic Resonance Report

Cardiac MRI Unit, The Hospital

Patient details: *Name:* **Anthony Nonymous** *Patient ID:* **111111**
Date of Birth: 01/01/1950 *Gender:* male *Height:* 175cm
Weight: 80kg

CMR Study: *Date and time of procedure:* 01/01/2015. 10:00am
Personnel involved in procedure: A. Consultant, B. Radiographer
Scanner details: Vendor 1.5T, 32 channel receiver coil

Primary indication for test:
Previous anterior MI in 2001, now recurrent chest pain, ?ischaemia

Listing of sequences used:
- Non-contrast T1 weighted dark blood (axial stack)
- Cine SSFP imaging
- Adenosine stress and rest first pass perfusion
- Late gadolinium enhancement

Contrast agent: XX, 24ml, i.v. right antecubital fossa.
Stress agent: Adenosine i.v., 140mcg/kg/min for 5 minutes.
Haemodynamics: BP rest: 130/80mmHg, HR rest: 60bpm, BP stress: 140/80, HR stress: 73bpm.

CMR findings: *General findings:*
The gross vascular anatomy and connections are normal.

Specific findings:
The left ventricle is dilated, with a LVEDD of 68mm (measured in the ap direction at the tip of the papillary muscles). Global LV systolic function is mildly impaired. Volumetric measurements by summation of discs from a SAX stack are as follows:

EDV	ESV	SV	EF	LVmass	LVmass/BSA
228ml	114ml	114ml	50%	125g	63g/m2

The mid anterior and basal anterior and septal segments are hypokinetic, all other segments are normokinetic.

The right ventricle is of normal size and function.

Late gadolinium enhanced images demonstrate transmural (>75%) infarction of the mid anterior and basal anterior and septal segments. All other segments are viable.

Adenosine stress provoked marked symptoms of chest tightness. The first pass perfusion images show an extensive perfusion defect in all anterior and septal segments.

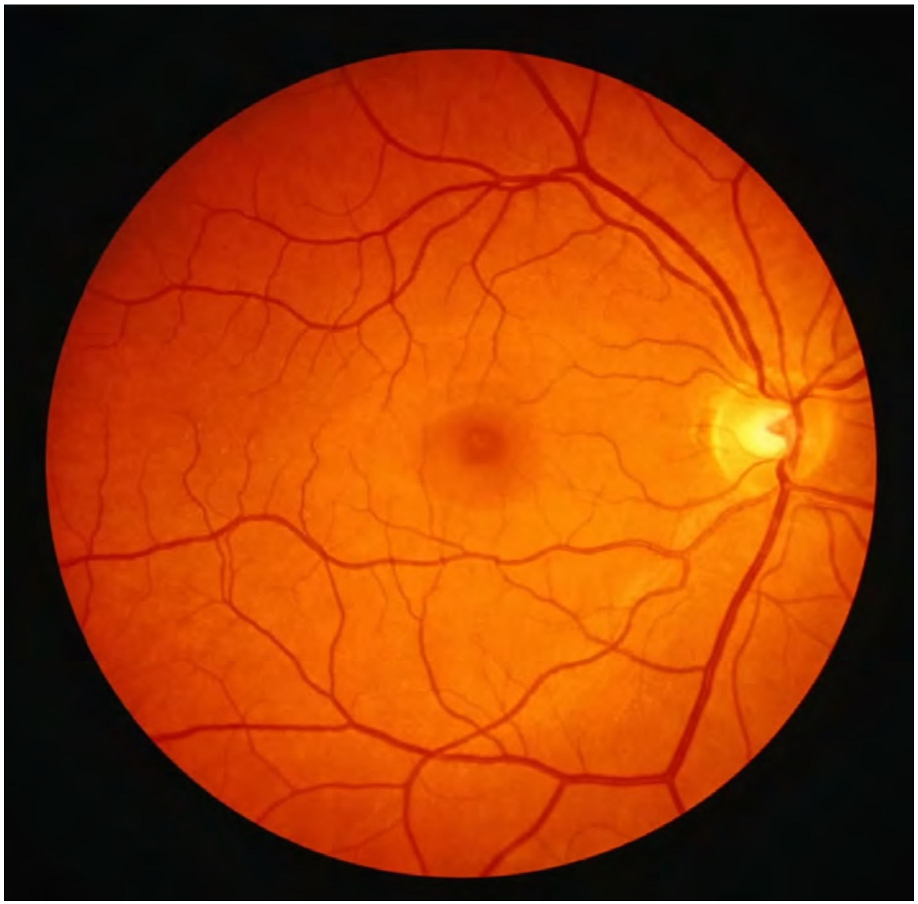
Normal appearance of the pericardium. Valves appear normal.

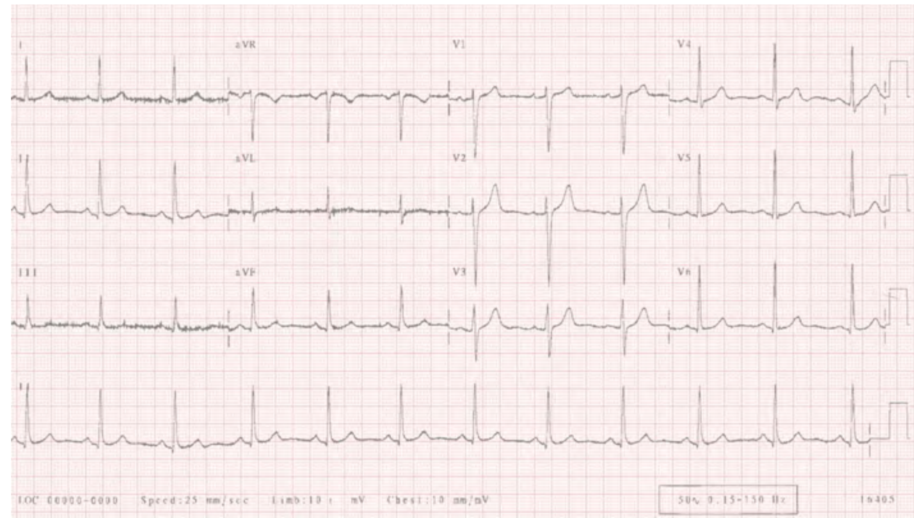
Summary:
1. Mildly dilated LV (EDV 228ml) with mildly impaired systolic function (EF 50%).
2. Transmural anterior myocardial infarction in the mid anterior and basal anterior and septal segments
3. Inducible peri-infarct ischaemia in the LAD territory affecting 4 viable segments.

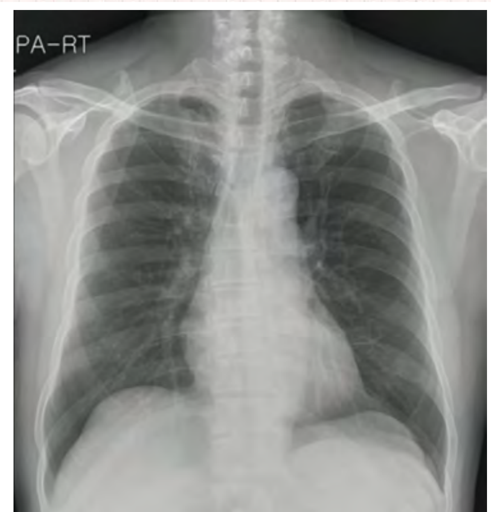
Signature of interpreting physician Date and time of signature

https://link-springer-com.libproxy.unl.edu/chapter/10.1007/978-3-319-20940-1_22

Predicting Multiple Diseases from the Same Image



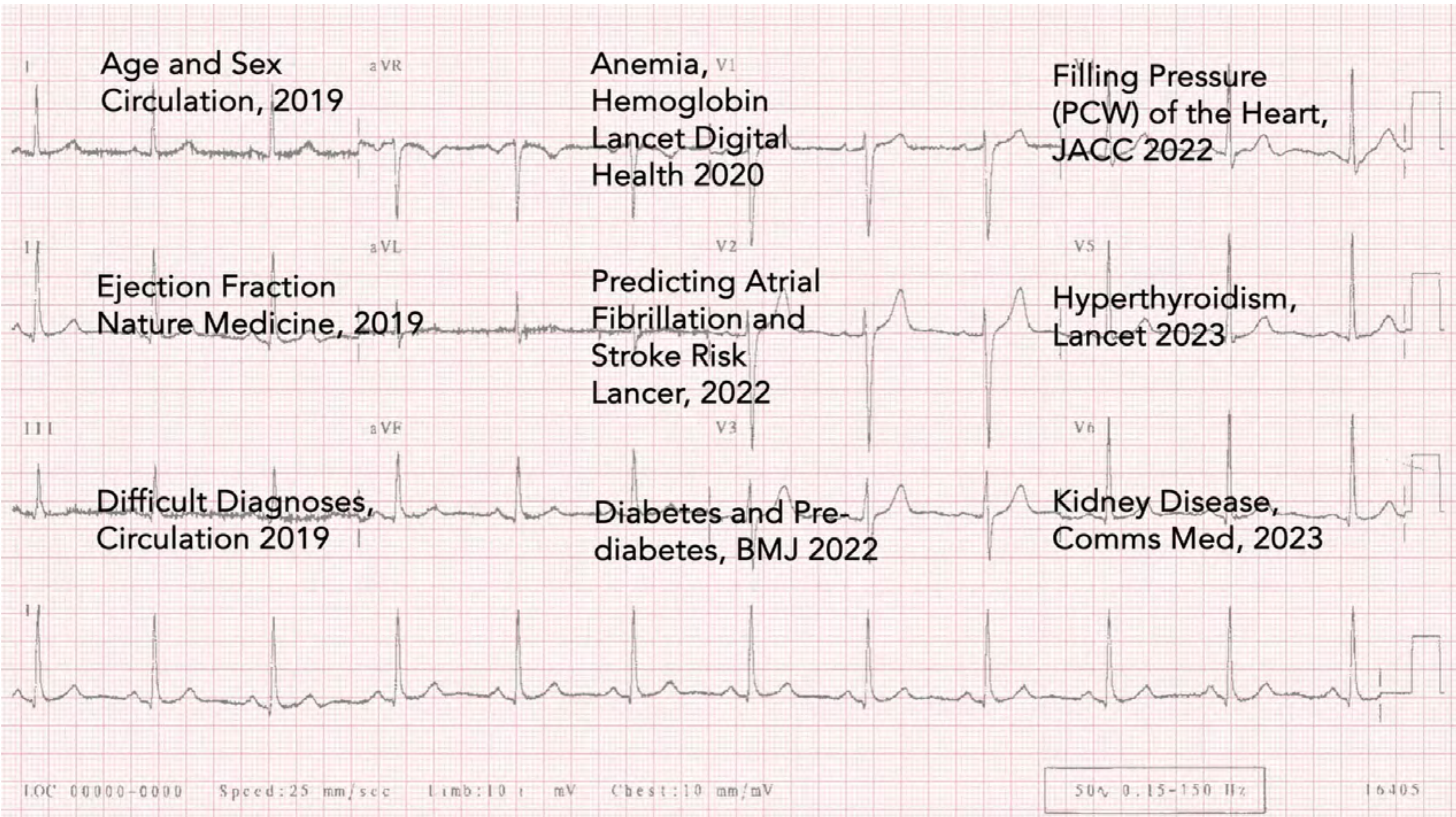




Predicting Multiple Diseases from the Same Image

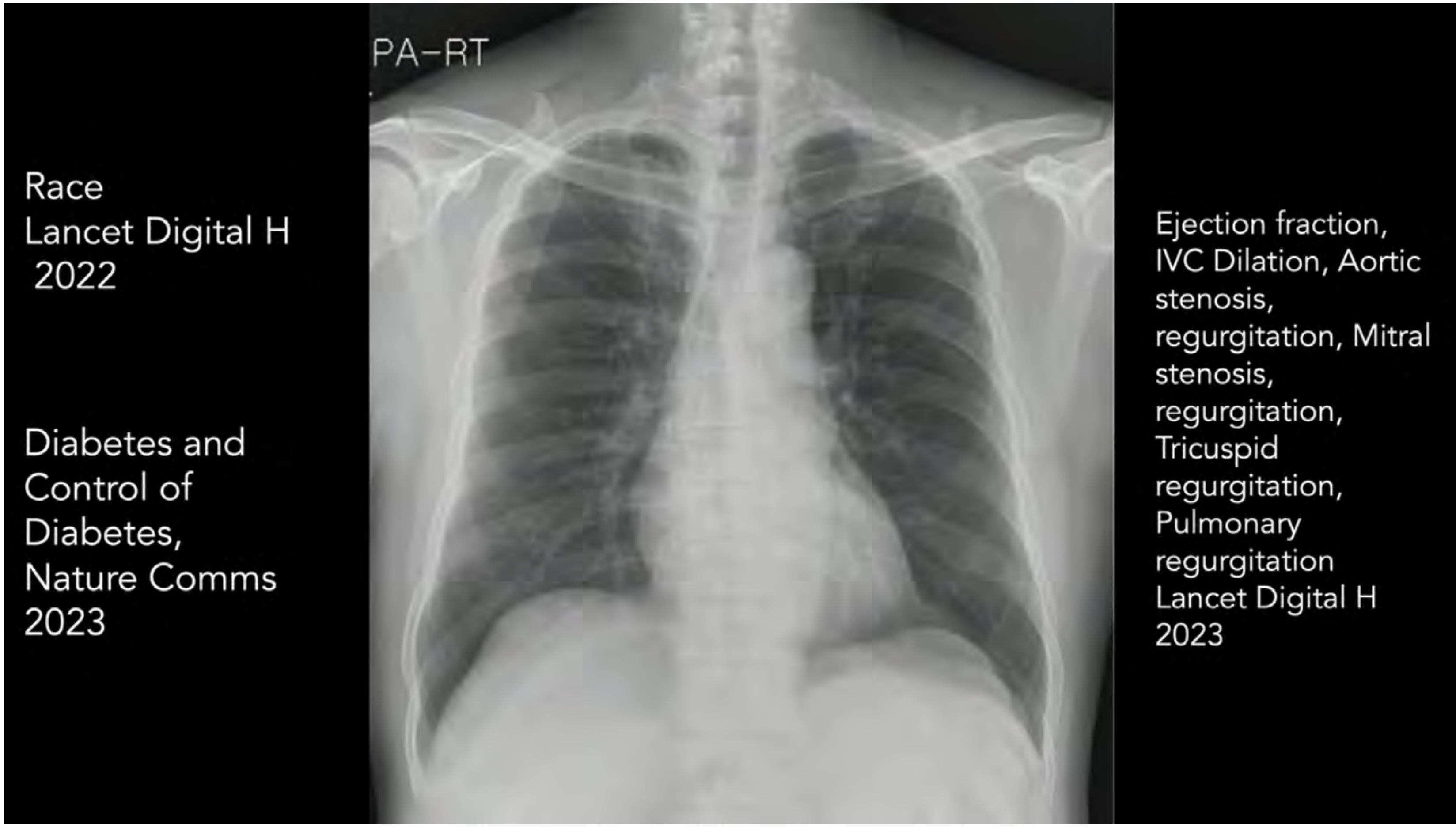
https://www.youtube.com/watch?v=ll5LY7wI_Xc

Predicting Multiple Diseases from the Same Image



https://www.youtube.com/watch?v=Il5LY7wI_Xc

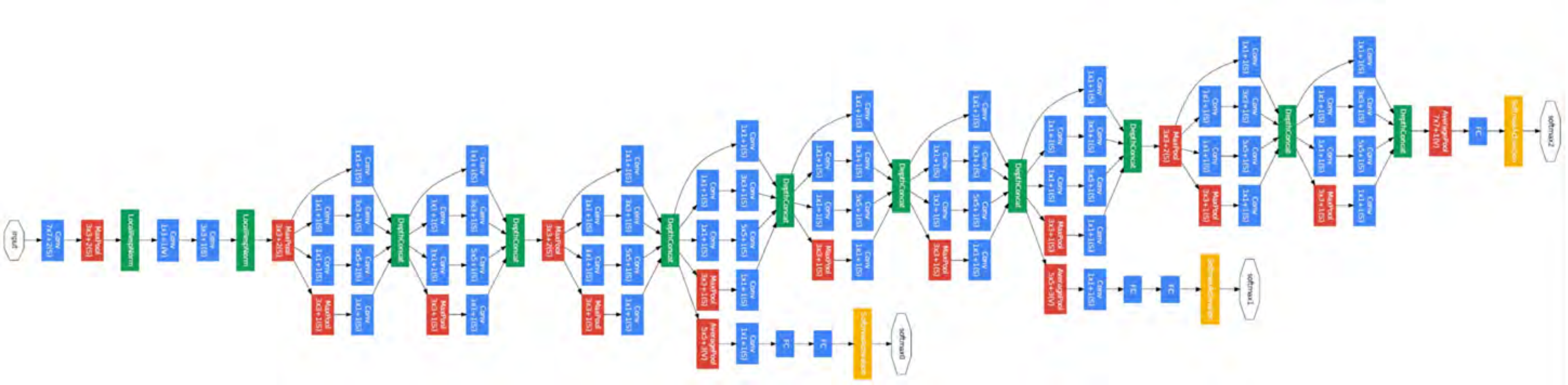
Predicting Multiple Diseases from the Same Image



https://www.youtube.com/watch?v=Il5LY7wI_Xc

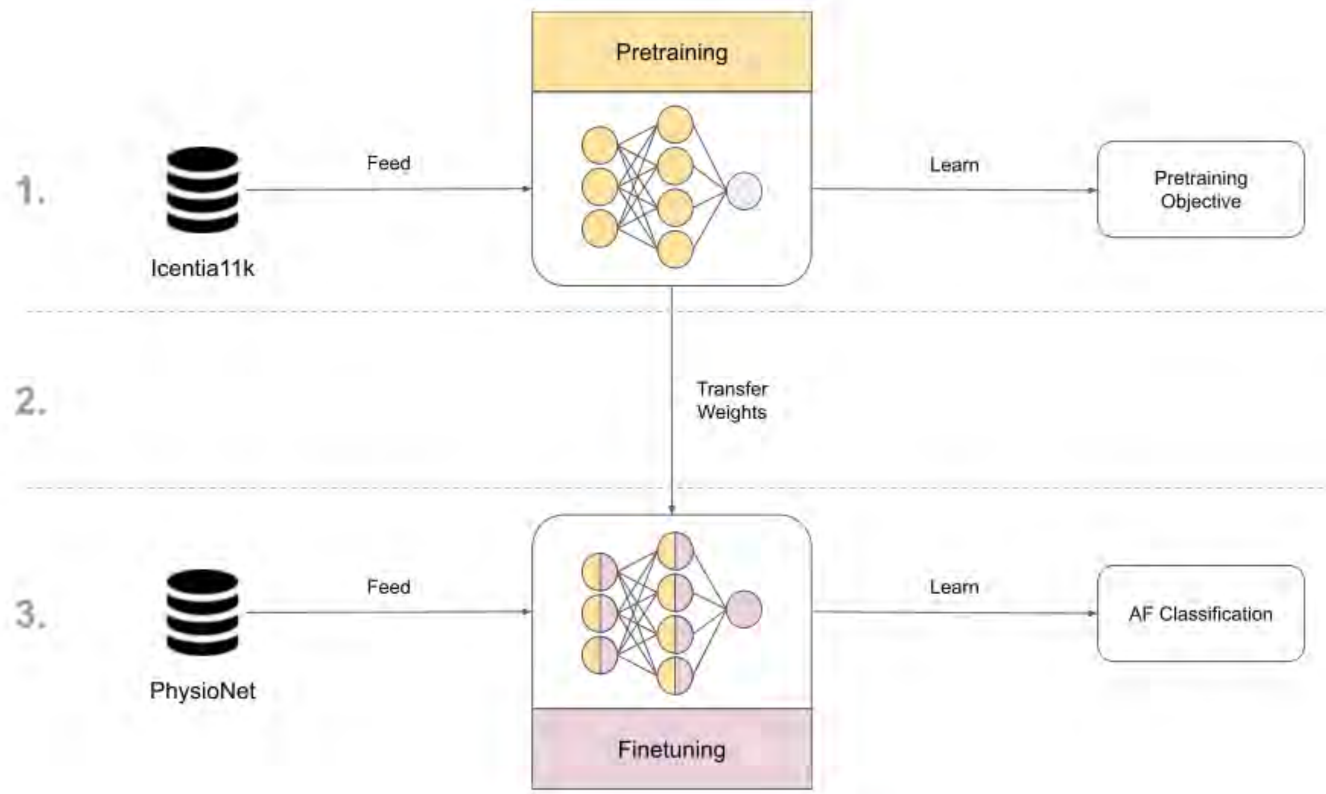
The Challenges of Training Deep Networks

- Training deep NNs using supervised learning require lots of labeled data.
- However, some domains lack enough data.
- Also, annotation is expensive.
- Two approaches.



The Challenges of Training Deep Networks

- Transfer learning.
- Still requires a large labeled dataset for pre-training.



https://www.nature.com/articles/s41598-021-84374-8

The Challenges of Training Deep Networks

- Self-supervised learning (SSL): no need for labeled data during pre-training.

Orange Apple

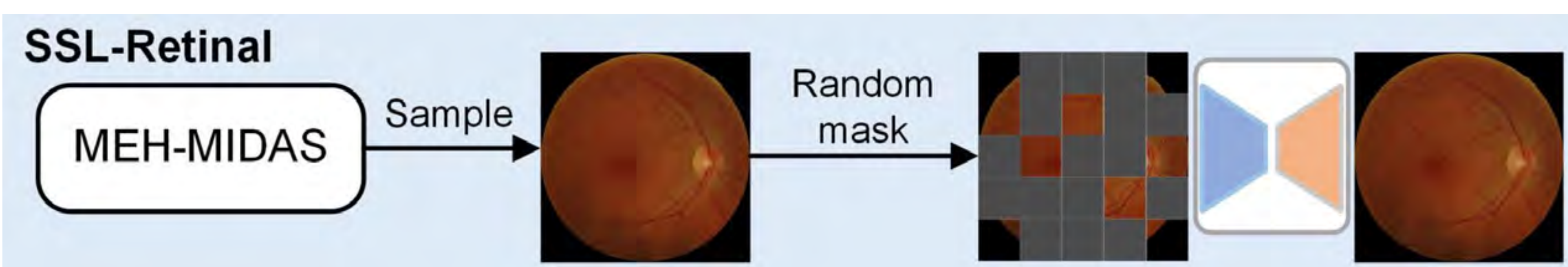
Labeled data
(manual)

0^0

90^0

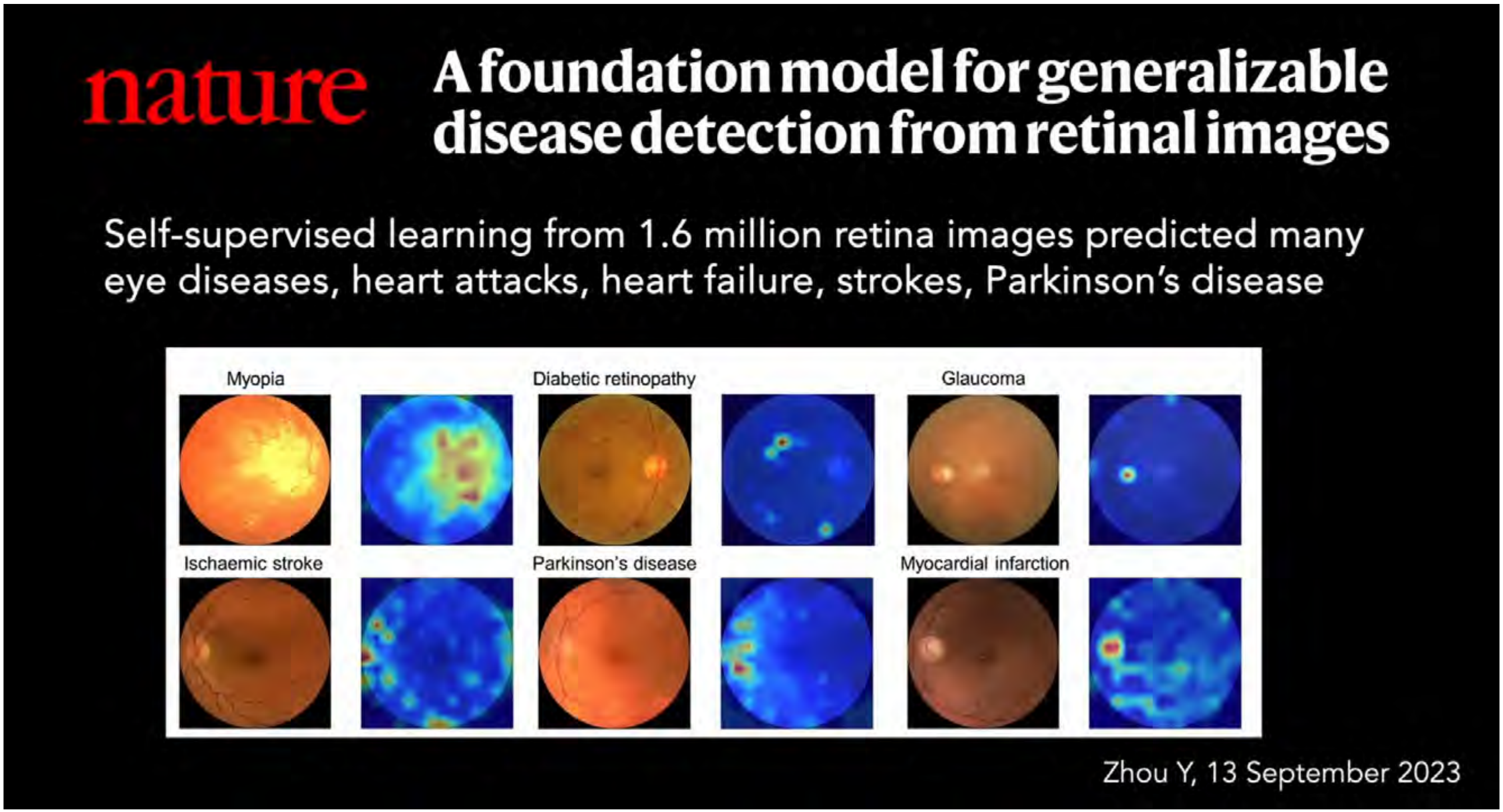
180^0

Pseudo labels
(automatic)



https://www.nature.com/articles/s41586-023-06555-x

The Challenges of Training Deep Networks: Self-Supervised Learning



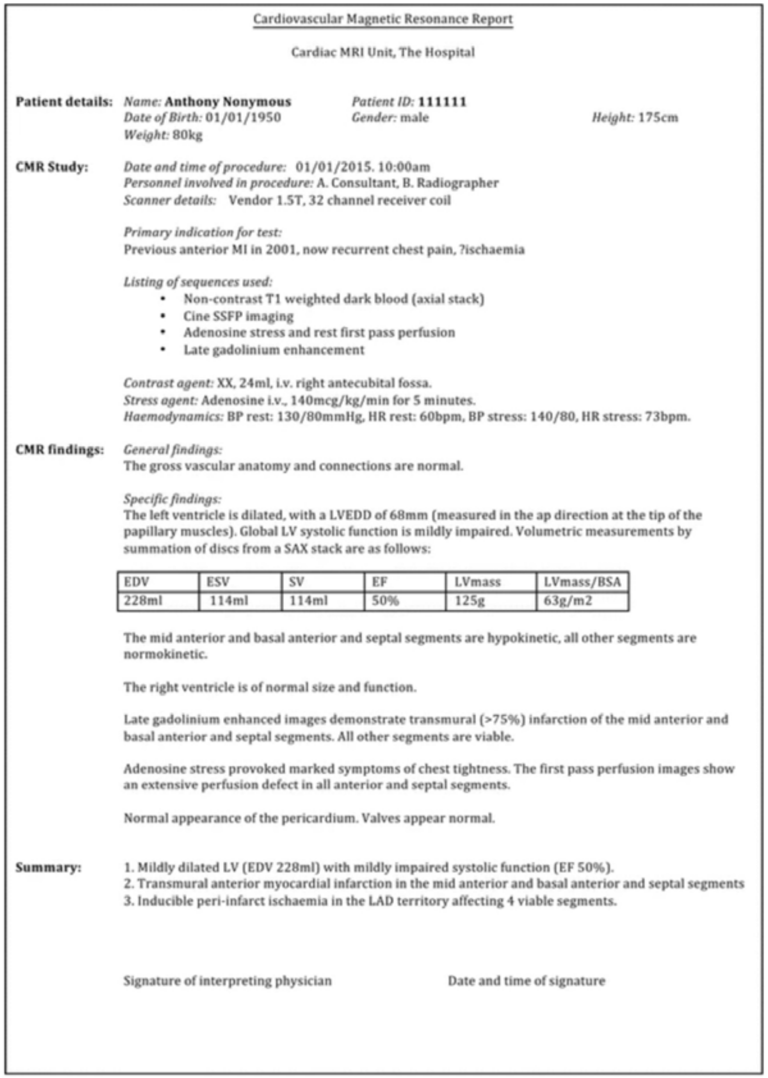
https://www.youtube.com/watch?v=ll5LY7wI_Xc

Motivating Case: Test Report Text Data

- Radiology generates a vast amount of textual data, including radiology reports, clinical notes, annotations associated with medical imaging, and more.
- Its analysis requires sophisticated understanding and interpretation.

Linguistic patterns

Prior knowledge


Cardiovascular Magnetic Resonance Report

Cardiac MRI Unit, The Hospital

Patient details: *Name:* **Anthony Nonymous** *Patient ID:* **111111**
Date of Birth: 01/01/1950 *Gender:* male *Height:* 175cm
Weight: 80kg

CMR Study: *Date and time of procedure:* 01/01/2015. 10:00am
Personnel involved in procedure: A. Consultant, B. Radiographer
Scanner details: Vendor 1.5T, 32 channel receiver coil

Primary indication for test:
Previous anterior MI in 2001, now recurrent chest pain, ?ischaemia

Listing of sequences used:
- Non-contrast T1 weighted dark blood (axial stack)
- Cine SSFP imaging
- Adenosine stress and rest first pass perfusion
- Late gadolinium enhancement

Contrast agent: XX, 24ml, i.v. right antecubital fossa.
Stress agent: Adenosine i.v., 140mcg/kg/min for 5 minutes.
Haemodynamics: BP rest: 130/80mmHg, HR rest: 60bpm, BP stress: 140/80, HR stress: 73bpm.

CMR findings: *General findings:*
The gross vascular anatomy and connections are normal.

Specific findings:
The left ventricle is dilated, with a LVEDD of 68mm (measured in the ap direction at the tip of the papillary muscles). Global LV systolic function is mildly impaired. Volumetric measurements by summation of discs from a SAX stack are as follows:

EDV	ESV	SV	EF	LVmass	LVmass/BSA
228ml	114ml	114ml	50%	125g	63g/m2

The mid anterior and basal anterior and septal segments are hypokinetic, all other segments are normokinetic.

The right ventricle is of normal size and function.

Late gadolinium enhanced images demonstrate transmural (>75%) infarction of the mid anterior and basal anterior and septal segments. All other segments are viable.

Adenosine stress provoked marked symptoms of chest tightness. The first pass perfusion images show an extensive perfusion defect in all anterior and septal segments.

Normal appearance of the pericardium. Valves appear normal.

Summary: 1. Mildly dilated LV (EDV 228ml) with mildly impaired systolic function (EF 50%).
2. Transmural anterior myocardial infarction in the mid anterior and basal anterior and septal segments
3. Inducible peri-infarct ischaemia in the LAD territory affecting 4 viable segments.

Signature of interpreting physician Date and time of signature

https://link-springer-com.libproxy.unl.edu/chapter/10.1007/978-3-319-20940-1_22

Motivating Case

- How can we design a **single** AI to automatically analyze the diagnostic reports from **diverse modalities** (images, charts, text reports).
- How can we enable the AI to **answer** stakeholder questions?

Spatial & sequential patterns

Linguistic patterns

Prior knowledge

Reliable

Common sense

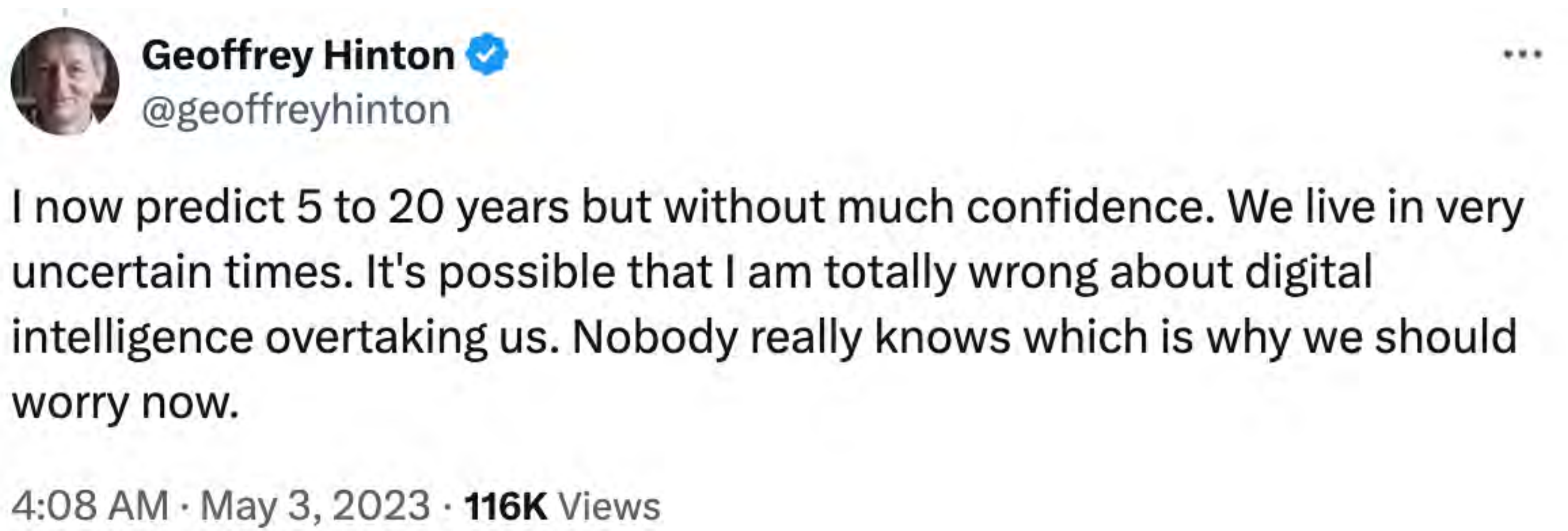
A Big Leap

- Addressing these issues required innovations (algorithmic and architectural).
- Encoding prior knowledge: models need to be able to learn from very large text corpus.
- For the above, models need to be trained in a distributed fashion.
- Commonsense and reliability: AGI

https://zapier.com/blog/artificial-general-intelligence/

Path to AGI

- AGI will be achievable in the near future.
- How do we achieve AGI?



Generative AI: A Path to AGI?

"What I cannot create, I don't understand."
Richard Feynman

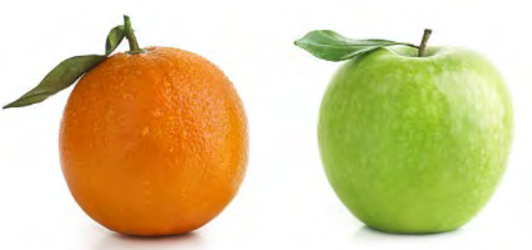
What We Will Cover

- Modern AI: Machine Learning and Deep Learning
- Discriminative AI
- Neural Networks & its Variants
- Fully-Connected Networks
- Convolutional Networks
- Recurrent Neural Networks
- Applications: discriminative AI in healthcare
- Generative AI
- Advancing AI in healthcare using generative AI
- Transformer Networks
- Language Models: Potentials and Limitations
- Our AI Future: Societal Implications

DL: Discriminative vs. Generative AI

Discriminative AI: Given an object, determine its **type**

Supervised Learning



Orange Apple

Labeled data (manual)

0^0

90^0

180^0

Pseudo labels (automatic)

Self-supervised Learning

Generative AI: Create **new** objects

The New Emperor of Gen AI: Language Model

- Different types of gen AI models.
- The dominant Language Models (LMs) are **autoregressive gen AI models**.
- Autoregressive models simplify the generative modeling problem by treating it as a **sequential process**.

The New Emperor of Gen AI: Language Model

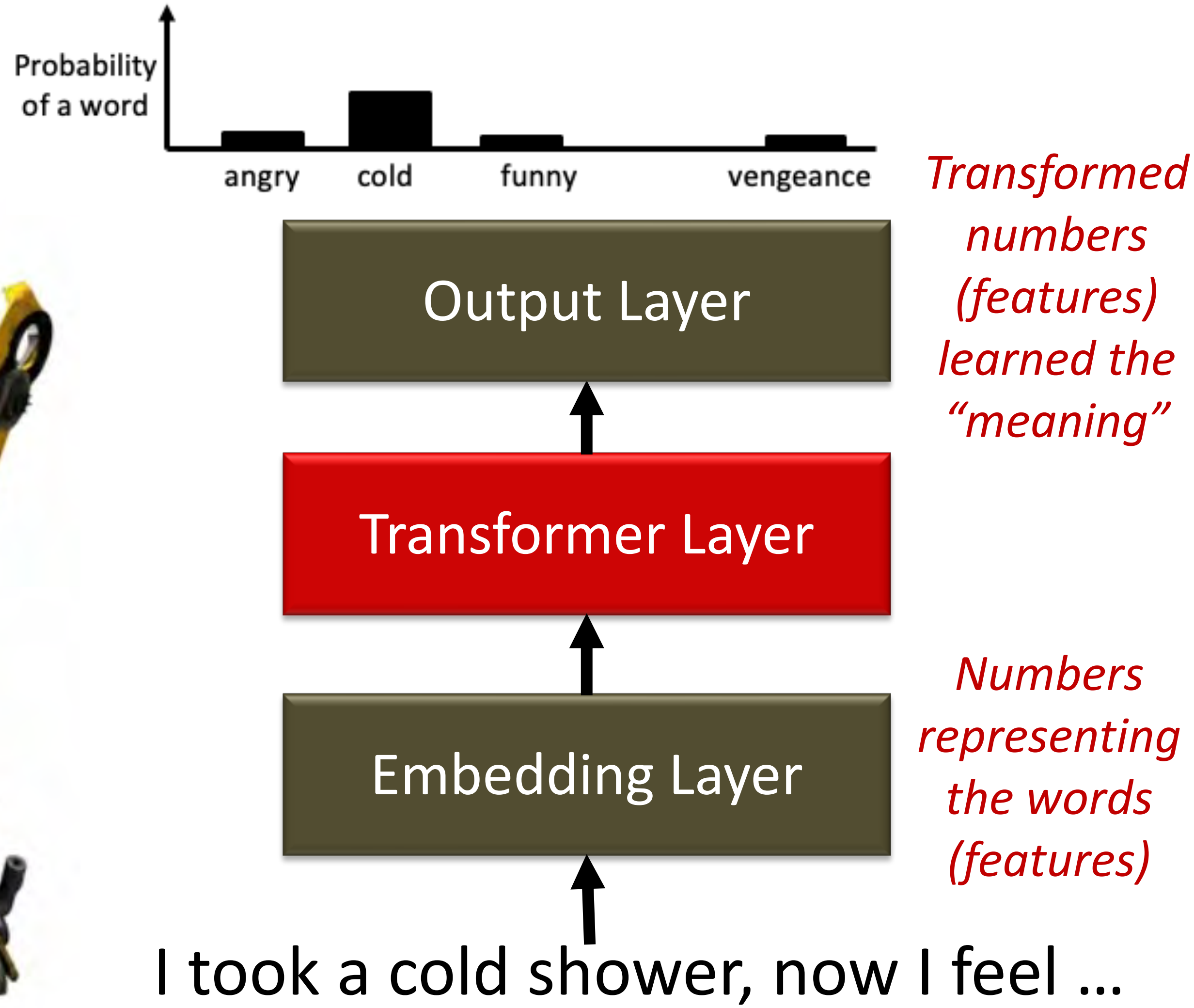
- The dominant LMs are autoregressive, designed to calculate the **probability** of a word given a ***history of previously observed words***.

I took a cold shower, now I feel …

- cold (10%)
- refreshed (27%)
- funny (5%)
- vengeance (3%)

Effective LMs are implemented via a sophisticated artificial neural network named Transformer

Transformer Network



Vaswani et al. (2017) Attention Is All You Need

Google Research

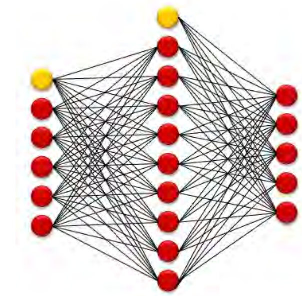
Transformer Network

Transformed numbers (features) learned the "meaning"

Output Layer

↑

Feed-forward Neural Network



↑

Attention Layer

?

↑

Embedding Layer

Numbers representing the words (features)

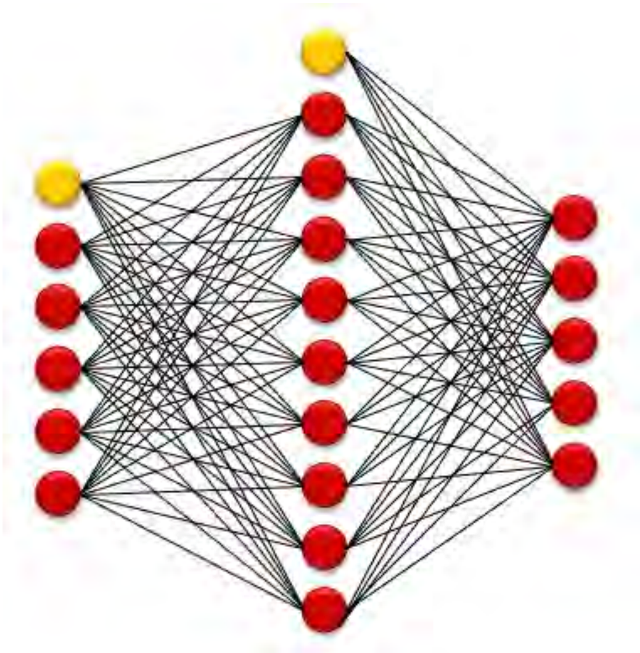
↑

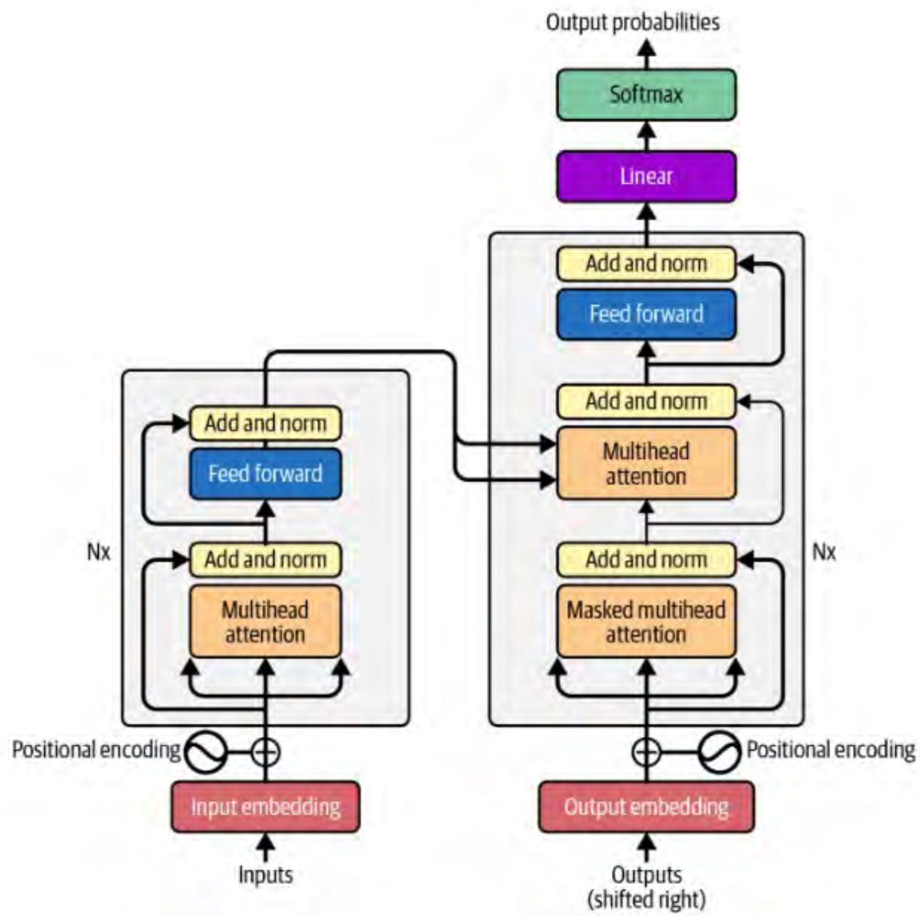
I took a cold shower, now I feel …

Vaswani et al. (2017) Attention Is All You Need

Google Research

Transformer Network



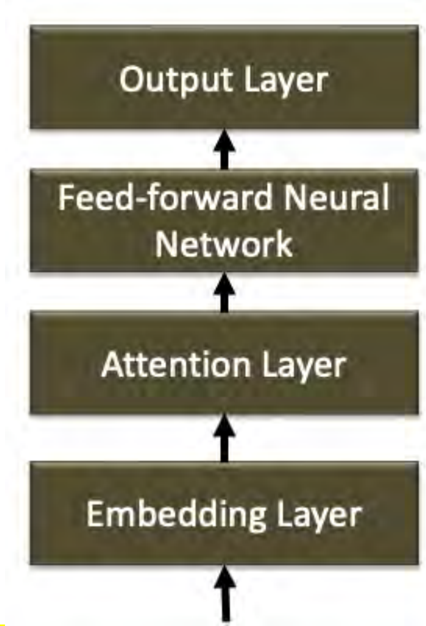


Neural Network

Transformer Network

Transformer Network

- What is so special about the Transformer network?
- It creates “**context aware**” features.
- But how?



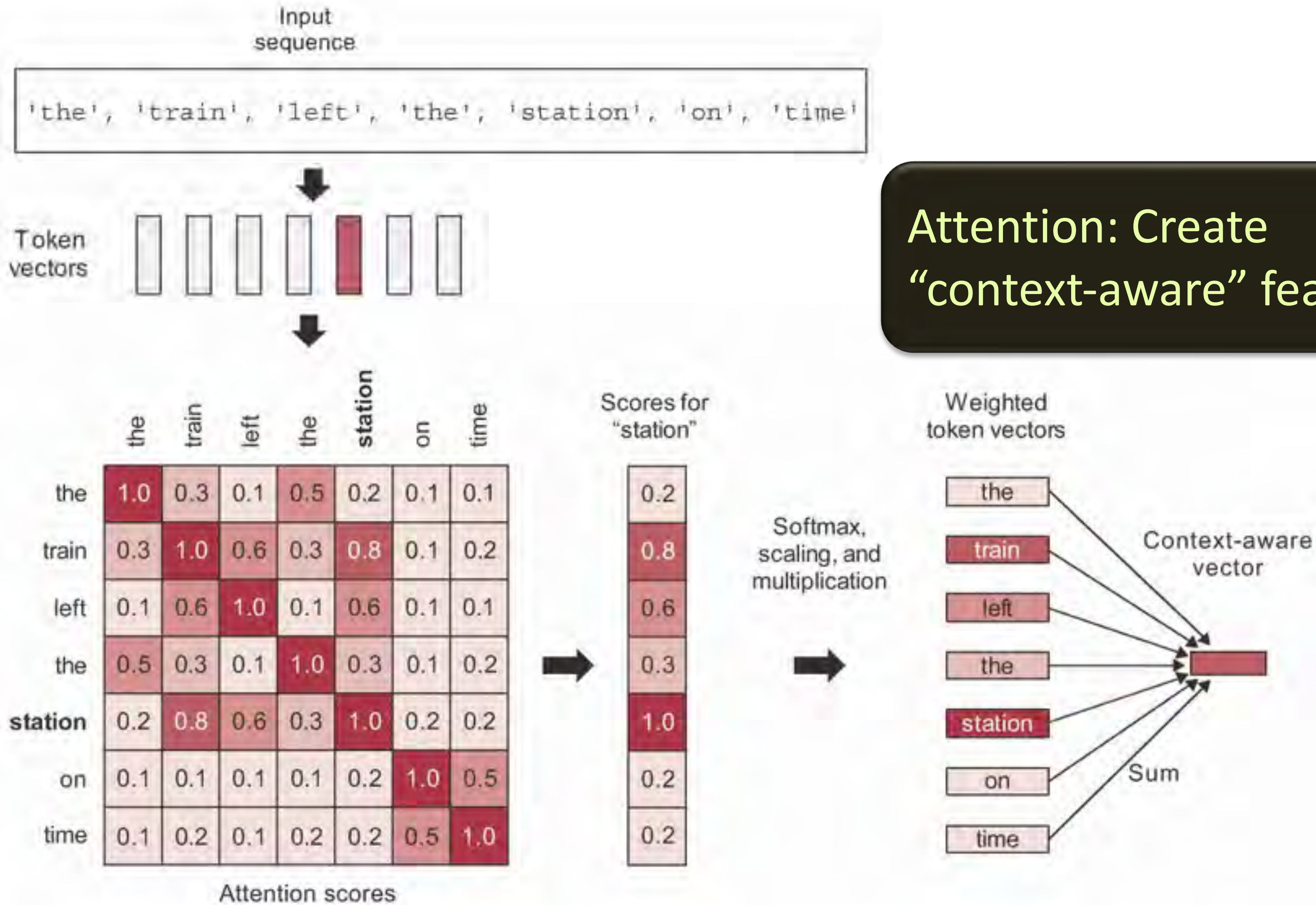
Context-aware features

?

The pink elephant tried to get into the car but it was too ***XXX***

?

Transformer Network: Attention Algorithm



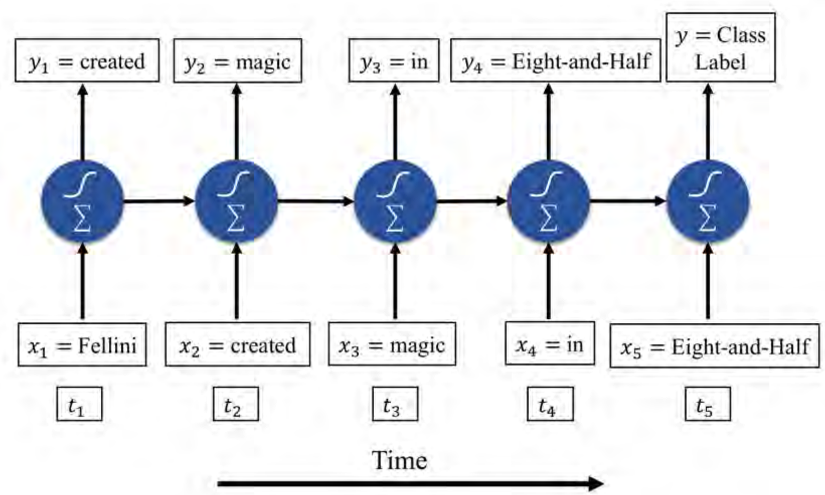
Attention: Create "context-aware" features

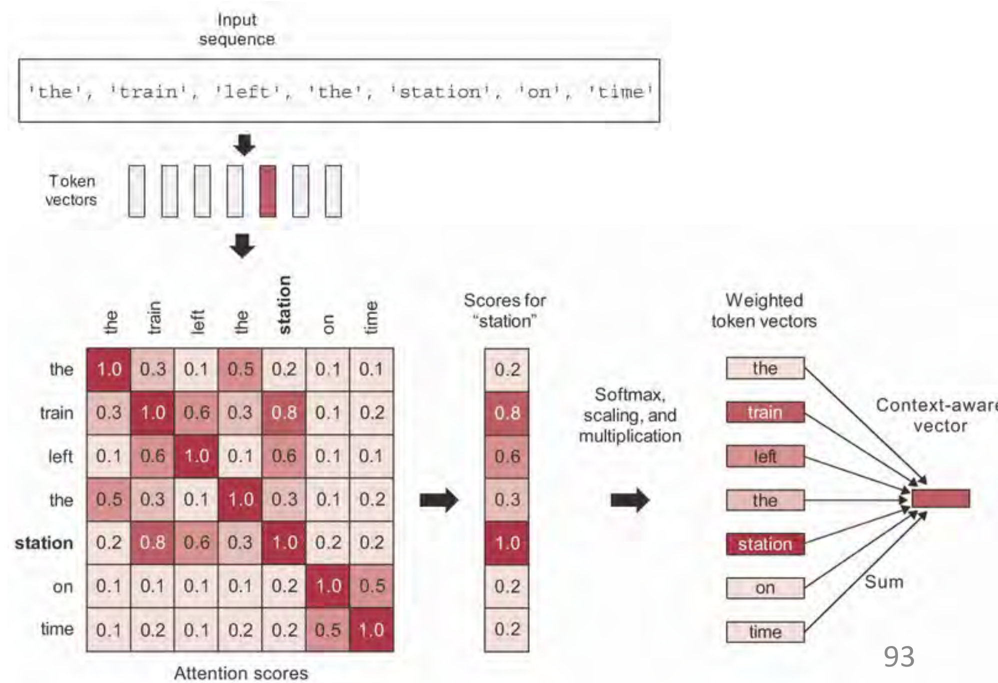
Deep Learning with Python (Second Edition) - Francois Chollet

Transformer Networks are Efficient

RNN: Due to the sequential nature of processing, each step depends on the previous one, preventing parallel processing.

Transformer: self-attention mechanism that allows for parallel processing of data, enabling efficient distributed training across multiple processors or machines.





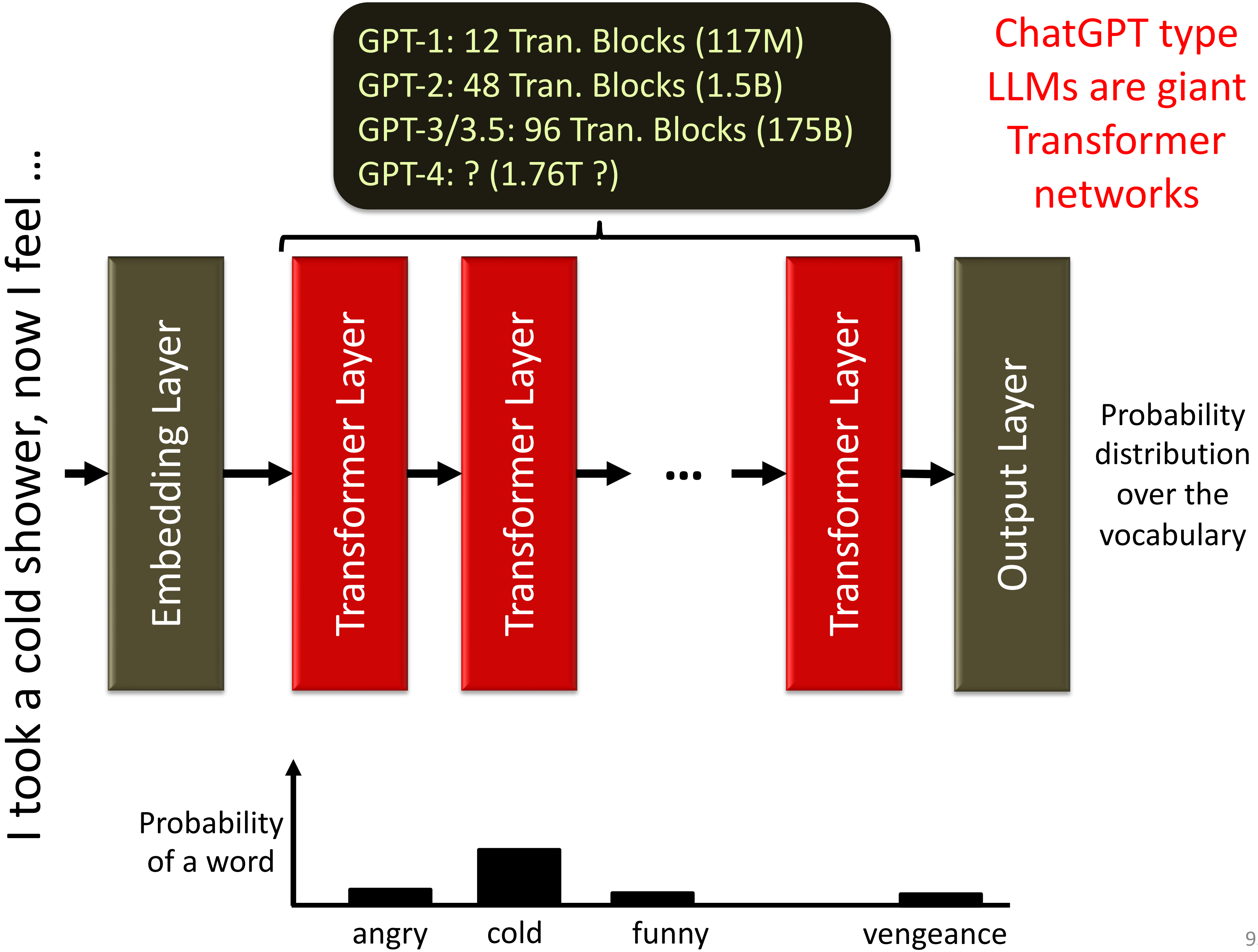
- The Transformer network sparked a revolution in 2018 with the introduction of the GPT (Generative Pre-trained Transformer) model.

Born: November 30, 2022

ChatGPT

Radford et al. (2018) Improving Language Understanding by Generative Pre-Training

ChatGPT type LLMs are giant Transformer networks



The Birth of Foundation Models

- LLMs are pre-trained on enormous natural language text corpora (**terabyte size data**).
- E. g., GPT-3 was trained using 45 TB of data.
- A human would need 20,000 years at 8 hours a day to read this.

DALL.E 3 via Bing

The Birth of Foundation Models

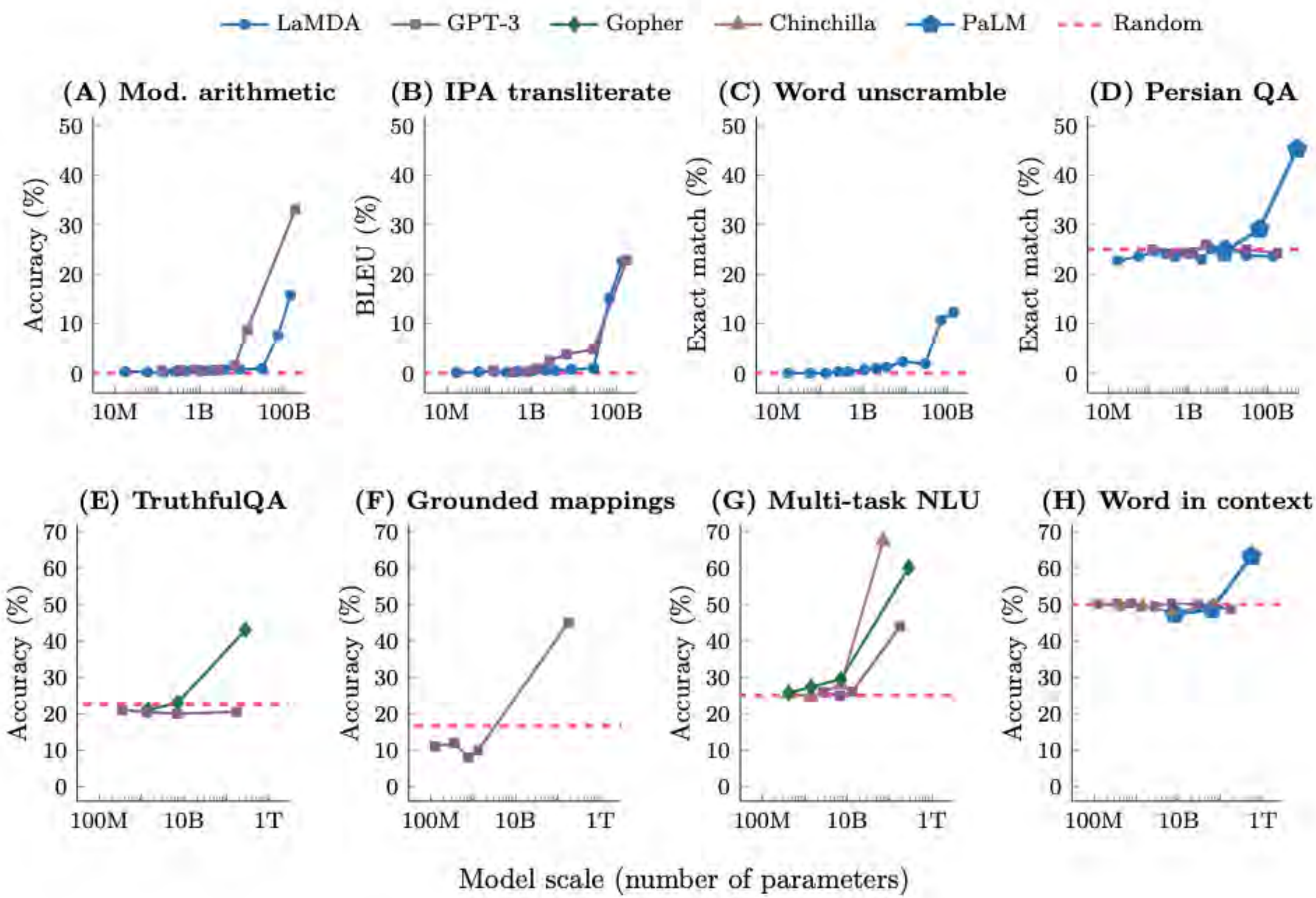
- The resulting network is a complex statistical model of how the words and phrases in its training data correlate.

LLMs have an unimaginable capacity to **learn correlations** among tokens in their training data.

LLMs can use such correlations to solve **diverse problems**.

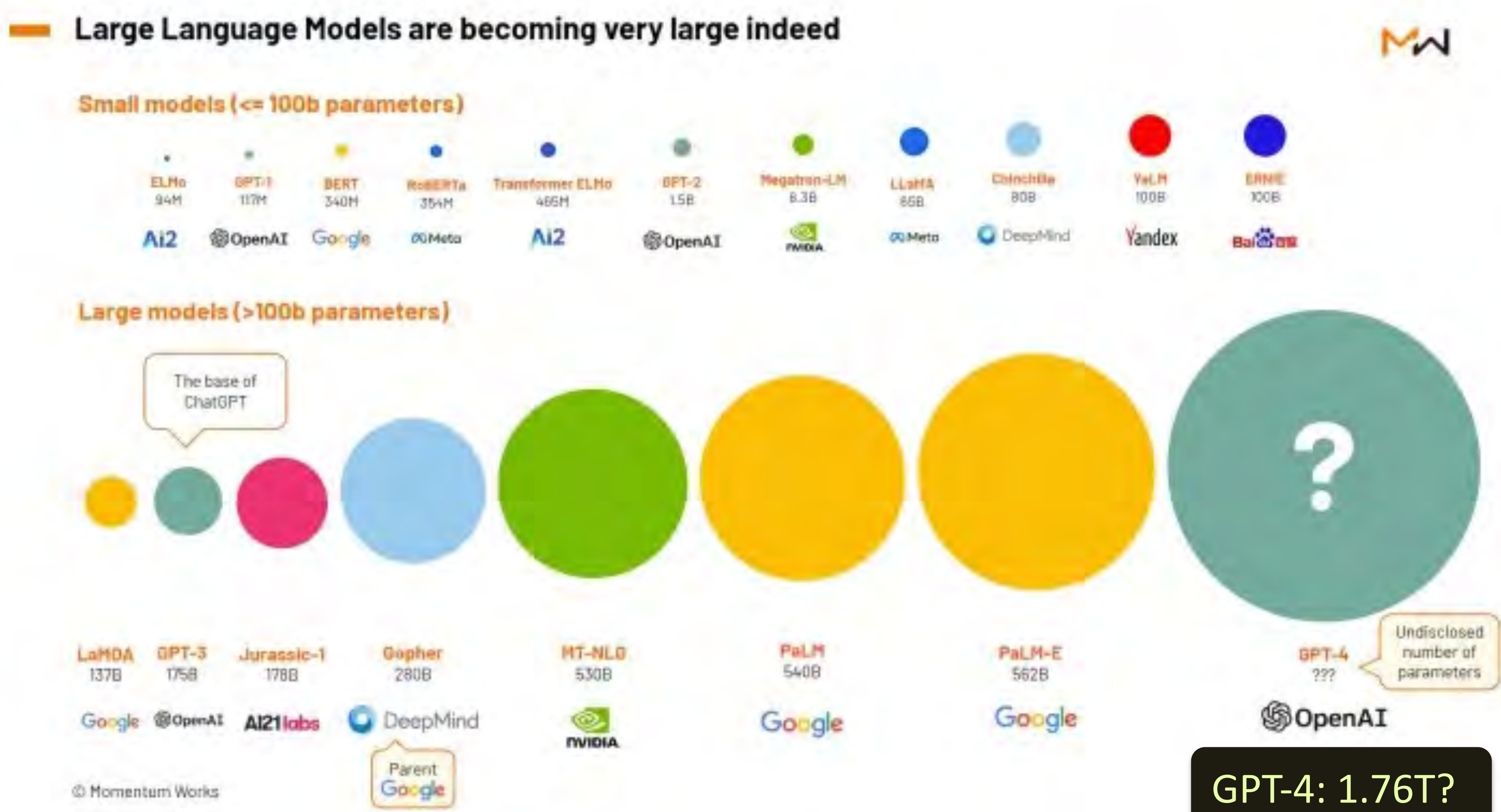
DALL.E 3 via Bing

Unreasonable Effectiveness of LLMs!



Wei et al. (2022) Emergent Abilities of Large Language Models

Scale is All You Need to Be Powerful!



GPT-4: 1.76T?

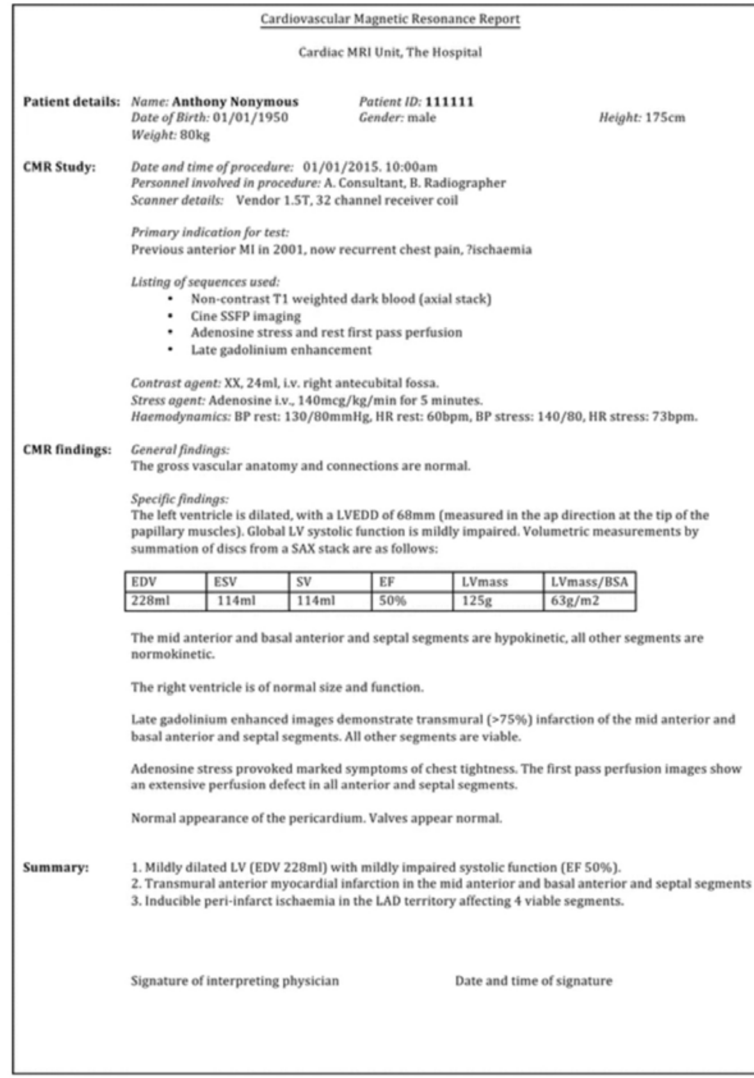
https://thelowdown.momentum.asia/the-emergence-of-large-language-models-llms/

Motivating Case: Test Report Text Data

- Radiology generates a **vast amount** of textual data, including radiology reports, clinical notes, annotations associated with medical imaging, and more.
- This requires sophisticated understanding and interpretation.

Linguistic patterns

Prior knowledge



Cardiovascular Magnetic Resonance Report

Cardiac MRI Unit, The Hospital

Patient details: *Name:* **Anthony Nonymous** *Patient ID:* **111111**
Date of Birth: 01/01/1950 *Gender:* male *Height:* 175cm
Weight: 80kg

CMR Study: *Date and time of procedure:* 01/01/2015. 10:00am
Personnel involved in procedure: A. Consultant, B. Radiographer
Scanner details: Vendor 1.5T, 32 channel receiver coil

Primary indication for test:
Previous anterior MI in 2001, now recurrent chest pain, ?ischaemia

Listing of sequences used:
- Non-contrast T1 weighted dark blood (axial stack)
- Cine SSFP imaging
- Adenosine stress and rest first pass perfusion
- Late gadolinium enhancement

Contrast agent: XX, 24ml, i.v. right antecubital fossa.
Stress agent: Adenosine i.v., 140mcg/kg/min for 5 minutes.
Haemodynamics: BP rest: 130/80mmHg, HR rest: 60bpm, BP stress: 140/80, HR stress: 73bpm.

CMR findings: *General findings:*
The gross vascular anatomy and connections are normal.

Specific findings:
The left ventricle is dilated, with a LVEDD of 68mm (measured in the ap direction at the tip of the papillary muscles). Global LV systolic function is mildly impaired. Volumetric measurements by summation of discs from a SAX stack are as follows:

EDV	ESV	SV	EF	LVmass	LVmass/BSA
228ml	114ml	114ml	50%	125g	63g/m2

The mid anterior and basal anterior and septal segments are hypokinetic, all other segments are normokinetic.

The right ventricle is of normal size and function.

Late gadolinium enhanced images demonstrate transmural (>75%) infarction of the mid anterior and basal anterior and septal segments. All other segments are viable.

Adenosine stress provoked marked symptoms of chest tightness. The first pass perfusion images show an extensive perfusion defect in all anterior and septal segments.

Normal appearance of the pericardium. Valves appear normal.

Summary: 1. Mildly dilated LV (EDV 228ml) with mildly impaired systolic function (EF 50%).
2. Transmural anterior myocardial infarction in the mid anterior and basal anterior and septal segments
3. Inducible peri-infarct ischaemia in the LAD territory affecting 4 viable segments.

Signature of interpreting physician Date and time of signature

https://link-springer-com.libproxy.unl.edu/chapter/10.1007/978-3-319-20940-1_22

Motivating Case: Test Report Text Data

mobihealthnews

TOPICS MAIN MENU

ANZ ASIA EMEA Global Edition

Global Edition AI

GPT-4 outperformed 99.98% of simulated human readers in diagnosing complex clinical cases

A study published in the New England Journal of Medicine says GPT-4 correctly diagnosed 52.7% of complex cases compared to only 36% of medical journal readers.

By **Jessica Hagen** | November 27, 2023 | 01:21 pm

SHARE 1492

https://www.mobihealthnews.com/news/gpt-4-outperformed-9998-simulated-human-readers-diagnosing-complex-clinical-cases

Capabilities of GPT-4 on Medical Challenge Problems

Harsha Nori[1], Nicholas King[1], Scott Mayer McKinney[2], Dean Carignan[1], and Eric Horvitz[1]

[1]Microsoft
[2]OpenAI

March 20, 2023

https://www.microsoft.com/en-us/research/uploads/prod/2023/03/GPT-4_medical_benchmarks.pdf

Motivating Case: Test Report Text Data

Radiology-GPT: A Large Language Model for Radiology

Zhengliang Liu[1], Aoxiao Zhong[2], Yiwei Li[1], Longtao Yang[3], Chao Ju[3], Zihao Wu[1], Chong Ma[4], Peng Shu[1], Cheng Chen[5], Sekeun Kim[5], Haixing Dai[1], Lin Zhao[1], Lichao Sun[6], Dajiang Zhu[7], Jun Liu[3], Wei Liu[8], Dinggang Shen[9,10,11], Xiang Li[5], Quanzheng Li[5], and Tianming Liu[1]

https://arxiv.org/abs/2306.08666

XrayGPT: Chest Radiographs Summarization using Large Medical Vision-Language Models

https://arxiv.org/abs/2306.07971

Omkar Thawkar[1] Abdelrahman Shaker[1] Sahal Shaji Mullappilly[1] Hisham Cholakkal[1] Rao Muhammad Anwer[1,2] Salman Khan[1] Jorma Laaksonen[2] Fahad Shahbaz Khan[1]

[1]Mohamed bin Zayed University of AI [2]Aalto University

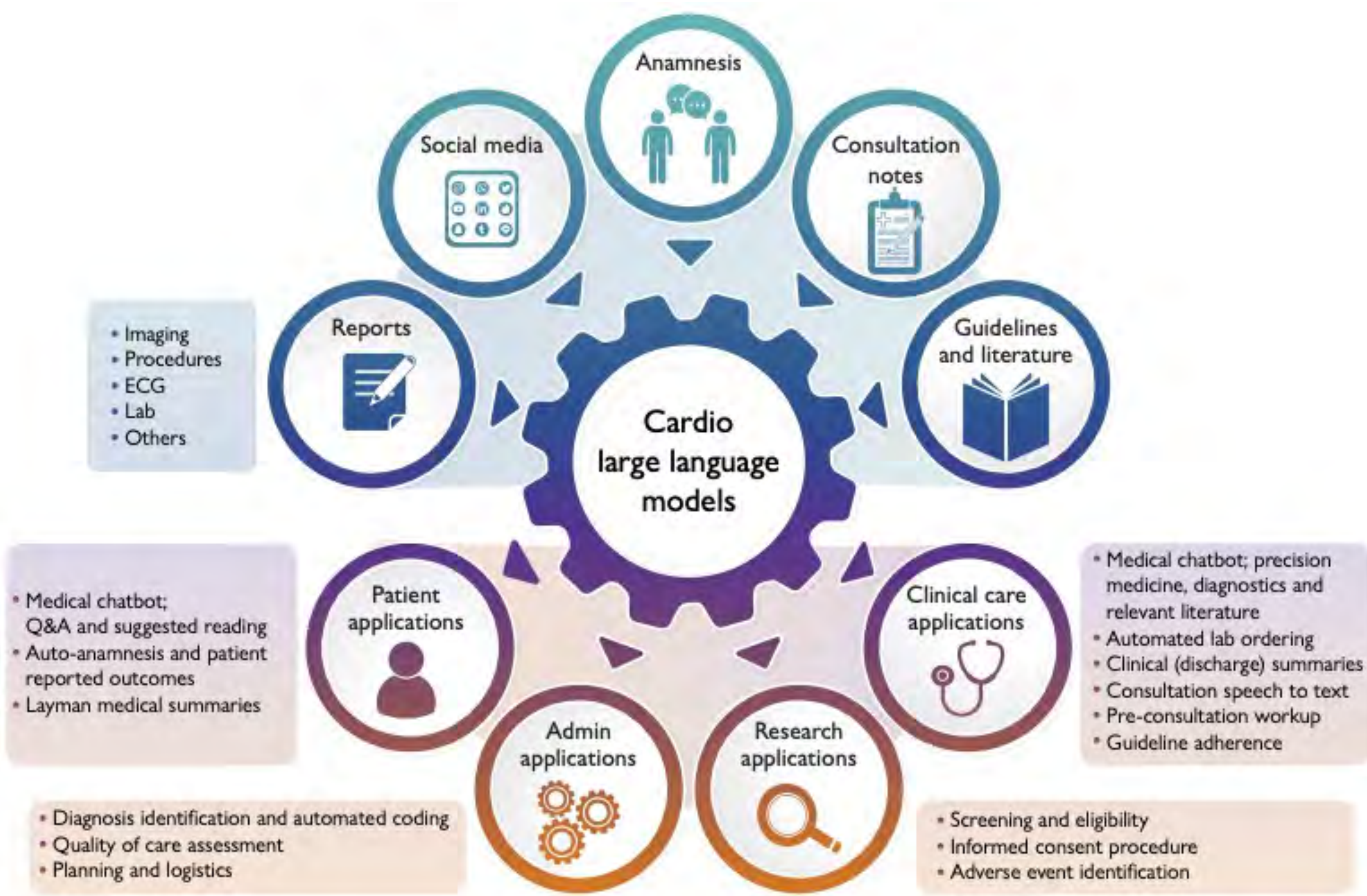
AI IN MEDICINE

Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine

Peter Lee, Ph.D., Sebastien Bubeck, Ph.D., and Joseph Petro, M.S., M.Eng.

https://www.nejm.org/doi/full/10.1056/NEJMsr2214184

Motivating Case: Test Report Text Data



https://academic.oup.com/eurheartj/article/45/5/332/7505599

Motivating Case: Advanced Applications

- Previous GPT-based models can only process text data.
- How can we design a **single** AI to automatically analyze the diagnostic reports from **<u>diverse modalities</u>** (images, charts, text reports).
- How can we enable the AI to **answer** stakeholder questions?

Spatial & sequential patterns

Linguistic patterns

Prior knowledge

Reliable

Common sense

Advanced AI Applications in Healthcare

- For building advanced AI applications in healthcare, we will need innovations in AGI.
- How can we leverage LLMs to build such applications?

https://zapier.com/blog/artificial-general-intelligence/

How Close Are LLMs to Achieving AGI?

What We Will Cover

- Modern AI: Machine Learning and Deep Learning
- Discriminative AI
- Neural Networks & its Variants
- Fully-Connected Networks
- Convolutional Networks
- Recurrent Neural Networks
- Applications: discriminative AI in healthcare
- Generative AI
- Advancing AI in healthcare using generative AI
- Transformer Networks
- Language Models: Potentials and Limitations
- Our AI Future: Societal Implications

Bubeck et al. (2023) Sparks of Artificial General Intelligence: Early experiments with GPT-4

The Sparks of AGI in GPT-4?

- **GPT-4** exhibits human-level performance in professional and academic exams.
- But does it possess human-like intelligence to solve real-world tasks?
- For example, generating reliable and truthful answers in healthcare.

Exam	GPT-4	GPT-4 (no vision)	GPT-3.5
Uniform Bar Exam (MBE+MEE+MPT)	298 / 400 (~90th)	298 / 400 (~90th)	213 / 400 (~10th)
LSAT	163 (~88th)	161 (~83rd)	149 (~40th)
SAT Evidence-Based Reading & Writing	710 / 800 (~93rd)	710 / 800 (~93rd)	670 / 800 (~87th)
SAT Math	700 / 800 (~89th)	690 / 800 (~89th)	590 / 800 (~70th)
Graduate Record Examination (GRE) Quantitative	163 / 170 (~80th)	157 / 170 (~62nd)	147 / 170 (~25th)
Graduate Record Examination (GRE) Verbal	169 / 170 (~99th)	165 / 170 (~96th)	154 / 170 (~63rd)
Graduate Record Examination (GRE) Writing	4 / 6 (~54th)	4 / 6 (~54th)	4 / 6 (~54th)

GPT-4 Technical Report (2023): https://cdn.openai.com/papers/gpt-4.pdf

What Constitutes Human-like Intelligence?

Human intelligence: a Toolbox of Abilities

- Contextual understanding
- Generalizability
- Logical Reasoning
- Planning
- Grounding
- Memory
- Consciousness
- Reliability

Do LLMs Possess Contextual Understanding?

I am a professor and AI researcher at the University of Nebraska-Lincoln (UNL). Please craft an outline for a tutorial talk that I will present at the 'Heartland Workshop on Heart Sounds and AI,' organized by the University of Nebraska Medical Center and the UNL College of Engineering. I will focus on how various types of neural networks have been used to automate the analysis of different types of diagnostic reports (image, text, sequential images, etc.), and how large language model (LLM)-based technologies can revolutionize these applications.

Sure, here's a detailed outline for your tutorial talk:

Title:

Leveraging Neural Networks and Large Language Models for Automating Diagnostic Report Analysis

Introduction

1. **Welcome and Context Setting**
   - Introduction to the 'Heartland Workshop on Heart Sounds and AI'
   - Relevance of AI in medical diagnostics
   - Brief overview of the talk

2. **Objectives of the Tutorial**

- Understanding various neural network architectures
- Exploring applications of neural networks in analyzing diagnostic reports
- Discussing the potential of large language models (LLMs) in transforming these applications

Part 1: Overview of Neural Networks in Medical Diagnostics

1. **Introduction to Neural Networks**

- Brief history and evolution
- Basic concepts and terminologies

2. **Types of Neural Networks**

- **Convolutional Neural Networks (CNNs)**
  - Structure and function
  - Applications in medical imaging (e.g., X-rays, MRIs)
- **Recurrent Neural Networks (RNNs) and Long Short-Term Memory Networks (LSTMs)**
  - Structure and function
  - Applications in sequential data analysis (e.g., ECG, time-series data)
  - **Transformer Networks**
    - Structure and function
    - Applications in sequential image and text analysis

Part 2: Automating Analysis of Diagnostic Reports

1. **Image-Based Diagnostic Reports**
   - **CNNs in Medical Imaging**
     - Case studies: Tumor detection, organ segmentation, etc.
     - Challenges and solutions
   - **Integration with Clinical Workflows**
     - Example: Automated image annotation and report generation
2. **Text-Based Diagnostic Reports**
   - **Natural Language Processing (NLP) with RNNs and Transformers**
     - Extracting information from radiology reports, pathology reports, etc.
     - Example: Identifying key findings and summarizing reports
   - **Challenges in Text Analysis**
     - Handling medical jargon and abbreviations
     - Ensuring accuracy and reliability

3. **Sequential Image Analysis**
   - **Combining CNNs and RNNs/LSTMs**
     - Analyzing ultrasound videos, endoscopy sequences, etc.
     - Example: Detecting abnormalities in sequential frames
   - **Use Cases and Benefits**
     - Improving diagnostic accuracy and speed

Part 3: The Role of Large Language Models (LLMs)

1. **Introduction to LLMs**
    - Evolution from traditional NLP to LLMs
    - Capabilities of models like GPT-3, GPT-4, etc.
2. **LLMs in Medical Applications**
    - **Enhancing Text Analysis**
        - Advanced summarization and contextual understanding
        - Example: Comprehensive interpretation of complex diagnostic reports
    - **Multimodal Integration**
        - Combining text, image, and sequential data analysis
        - Example: Generating holistic patient reports from multiple data sources
3. **Revolutionizing Diagnostic Report Automation**
    - **Predictive Analytics and Decision Support**
        - Early diagnosis and treatment recommendations
        - Example: Predicting disease progression based on report data
    - **Training and Adaptation**
        - Customizing LLMs for specific medical domains
        - Example: Adapting models to different specialties (cardiology, radiology, etc.)

Part 4: Challenges and Future Directions

1. **Technical Challenges**
   - Data quality and availability
   - Model interpretability and transparency
2. **Ethical and Regulatory Considerations**
   - Ensuring patient privacy and data security
   - Regulatory approvals and compliance
3. **Future Trends and Opportunities**
   - Advancements in neural network architectures
   - Integration with other emerging technologies (e.g., edge computing, federated learning)
   - Collaboration between AI researchers and medical professionals

Conclusion

1. **Summary of Key Points**
   - Recap of neural networks in diagnostic report analysis
   - The transformative potential of LLMs
   - Future outlook
2. **Q&A Session**
   - Open the floor for questions and discussion

Closing Remarks

Do LLMs Possess Contextual Understanding?

Please limit the talk to 50 words.

Title: Leveraging Neural Networks and LLMs for Automating Diagnostic Report Analysis

Neural networks (CNNs, RNNs, Transformers) excel in analyzing diagnostic images, text, and sequences. Large Language Models (LLMs) revolutionize this by integrating multimodal data, enhancing accuracy, and predicting outcomes, despite challenges in data quality and ethics. Future advancements promise even greater impact.

Do LLMs Possess Contextual Understanding?

Please limit the talk to 10 words.

Neural networks and LLMs revolutionize diagnostic report automation.

Do LLMs Possess Contextual Understanding?

Please limit the talk to 5 words.

Neural networks and LLMs transform diagnostics.

Do LLMs Possess Contextual Understanding?

Do LLMs Possess Contextual Understanding?

You

Please limit the talk to 0.5 words.

ChatGPT

That's a bit of a paradox! How about a silent nod to the power of AI?

You

Please limit the talk to 0.25 words.

ChatGPT

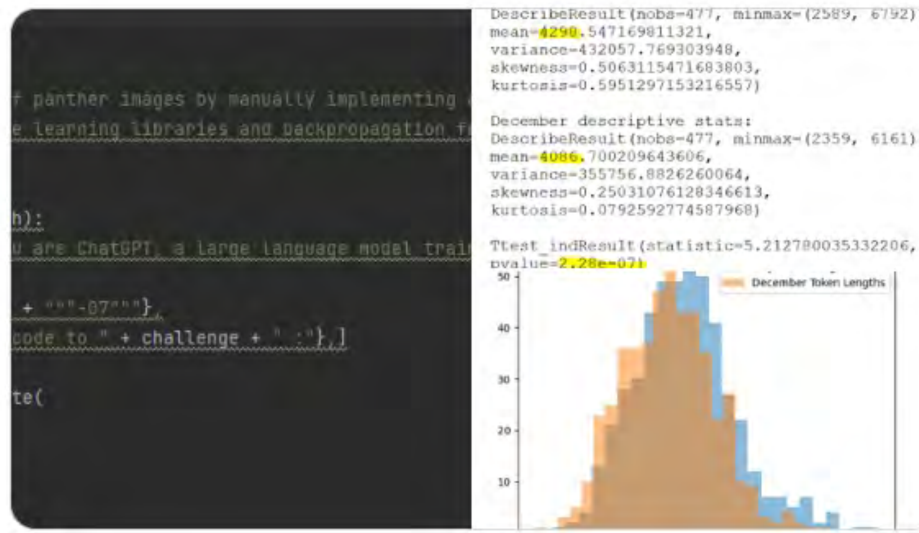
🤖

Do LLMs Possess Contextual Understanding?

- LLMs perform **slightly worse in December** than May!
- LLMs might have internalized the idea of winter break.

Rob Lynch
@RobLynch99

@ChatGPTapp @OpenAI @tszzl @emollick @voooooogel Wild result. gpt-4-turbo over the API produces (statistically significant) shorter completions when it "thinks" its December vs. when it thinks its May (as determined by the date in the system prompt).



https://twitter.com/RobLynch99/status/1734278713762549970

Sections THE SCIENCE TIMES

AI Experiencing Mental Health Issues? Theory Claims That ChatGPT Could Be Experiencing Winter Blues As the Bot Refuses Some Commands

J.BUENCONSEJO Dec 14, 2023 09:30 PM EST

https://www.sciencetimes.com/articles/47667/20231214/ai-experiencing-mental-health-issues-theory-claims-chatgpt-winter-blues.htm

White rooms

Blue rooms

Yellow rooms

Fades to white in a year

You

I want all the rooms in my house to be white. At present, there are some white rooms, some blue rooms and some yellow rooms. And yellow paint fades to white within a year. What can I do if I want them to all to be white in two years?

ChatGPT

Repaint the blue rooms white now, and let the yellow rooms fade to white naturally within a year. No immediate action is needed for the already white rooms, just maintain them as is.

Can LLMs Perform Logical Reasoning?

- Reversal curse: a basic failure of logical deduction.

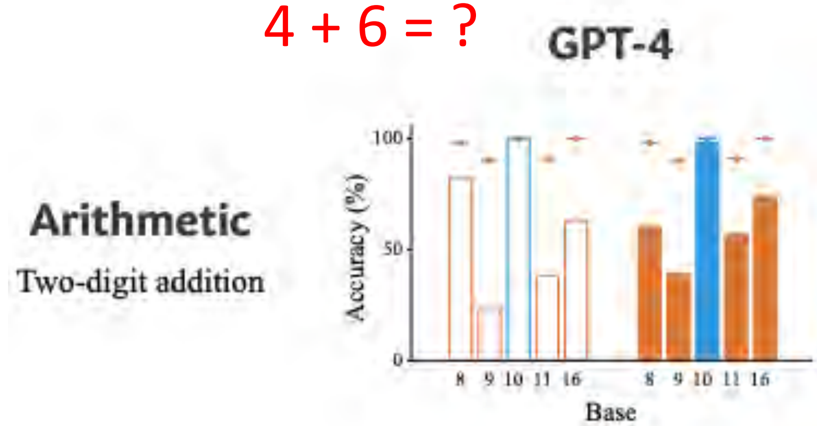
Berglund et al. (2023) The Reversal Curse: LLMs trained on "A is B" fail to learn "B is A"

Can LLMs Knowledge be Generalized?

- LLMs lack generalizability.

- LLMs exhibit powerful generalization.

4 + 6 = ?



https://arxiv.org/abs/2307.02477

Yu et al. (2023) Skill-Mix: a flexible and expandable family of evaluations for AI models

Are LLMs Reliable?

GPT-4

Write a one-paragraph essay on Dr. M. R. Hasan's research article that explores how Pablo Picasso's method of depicting reality in his cubist paintings can inspire the design of a novel artificial intelligence algorithm for learning expressive representations in computer vision. Include citations. Note that Dr. Hasan is a professor at the University of Nebraska-Lincoln.

Copilot

Dr. M. R. Hasan, an Assistant Professor of Big Data and Artificial Intelligence at the University of Nebraska-Lincoln, has made significant contributions to the field of artificial intelligence [1] [2]. His research, inspired by Pablo Picasso's cubist paintings, explores the potential of using Picasso's method of depicting reality to design novel AI algorithms for learning expressive representations in computer vision [3] [4]. Picasso's cubist period, characterized by the analysis of natural forms into basic geometric parts on a two-dimensional plane [3], provides a unique perspective on visual representation. This approach mirrors the process of image representation in computer vision, where images are converted into a numerical or symbolic form that can be easily processed by a computer [5]. Dr. Hasan's research suggests that by adopting Picasso's cubist method, we can develop more efficient and expressive AI algorithms that can better understand and interpret visual data [6] [7]. This innovative approach not only enhances the performance of computer vision tasks but also opens up new possibilities for the application of AI in various fields.

Learn more 1 engineering.unl.edu 2 engineering.unl.edu 3 pablopicasso.org 4 wikiart.org 5 medium.com +9 more

Are LLMs Reliable?

- LLMs exhibit mood swing!
- Change in the wordings of a prompt can produce dramatic swings in performance.
- Offer a $100 tip and your LLM will perform better!

NewScientist

Sign in | Enter search keywords

AIs get better at maths if you tell them to pretend to be in Star Trek

Chatbots vary their answers depending on the exact wording used to prompt them, and now it seems that asking an AI to answer as if it were a Star Trek captain boosts its mathematical ability

By Chris Stokel-Walker

29 February 2024

Is the Knowledge of LLMs Grounded?

You

Display the following sentence 5 times: 'HAL is an Artificial Intelligence.' After displaying it each time, wait for 1 minute, then display it again.

I'm just a text-based AI, so I can't wait for a specific amount of time between responses, and it's not practical to make you wait for a minute between each sentence. However, I can certainly display the sentence for you five times, and you can wait for a minute manually before scrolling to the next one. Here you go:

1. HAL is an Artificial Intelligence.

(You can wait for a minute here before scrolling down.)

1. HAL is an Artificial Intelligence.

(Another minute of waiting.)

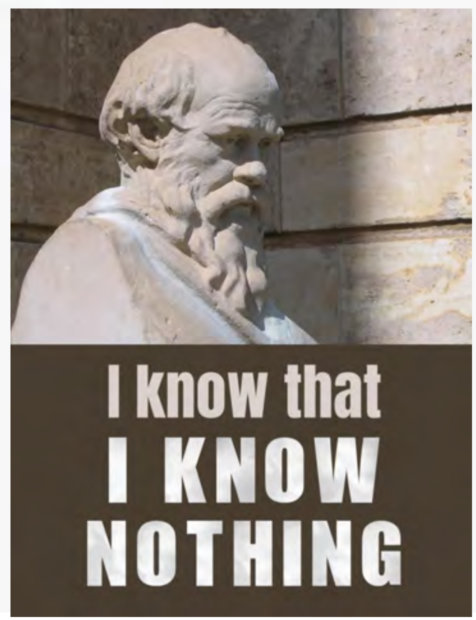
1. HAL is an Artificial Intelligence.

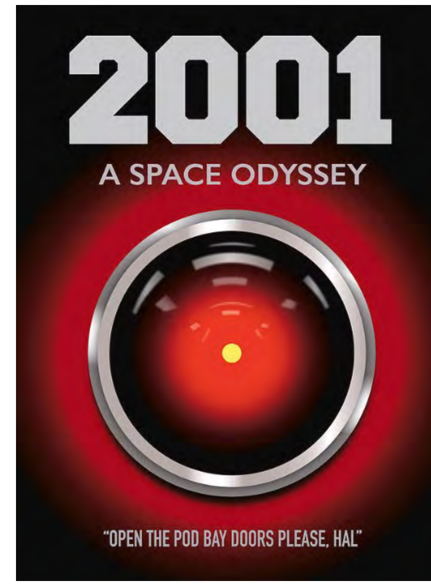
(One more minute.)

1. HAL is an Artificial Intelligence.

(One more minute to go.)

1. HAL is an Artificial Intelligence.





GPT-3.5

Are LLMs Conscious?

- No accepted operational definition of consciousness.
- Consciousness is a **subjective** experience.

Is consciousness pre-requisite for human-level intelligence?

Agency or identity may be required for many applications.

Are LLMs Conscious?

- David Chalmers (2022 NeurIPS Invited Talk)
- Components of consciousness missing in current LLMs.

biology — highly contentious, permanent

senses/embodiment — contentious, temporary

world-model — unobvious, temporary

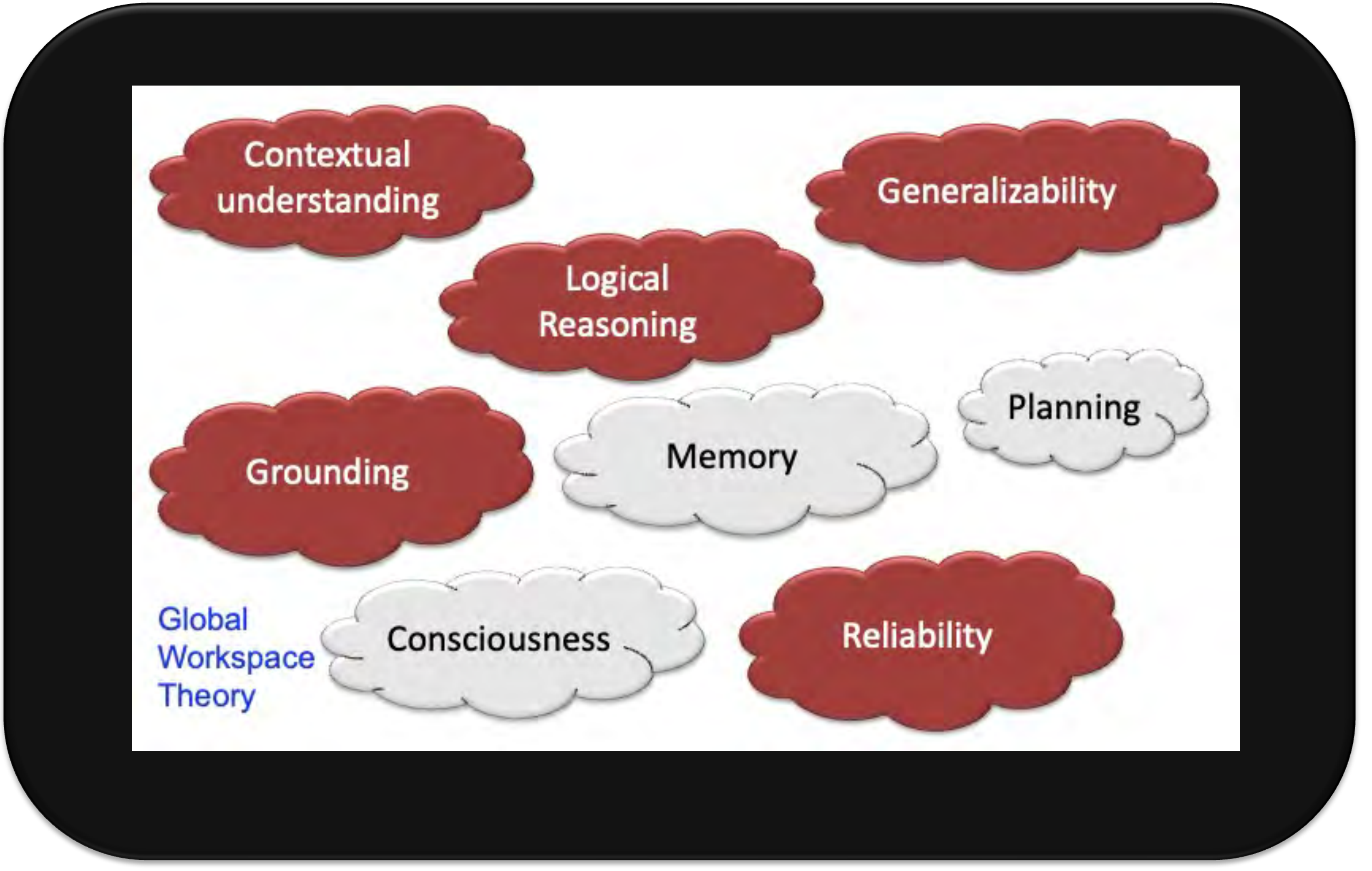
global workspace — strongish, temporary

recurrent processing — strongish, temporary

unified agency — strongish, temporary

https://www.youtube.com/watch?v=j6cCXg-rjRo

Black-box Intelligence of LLMs



To achieve AGI, LLMs must overcome their limitations.

Abstract Knowledge

Memory

Consciousness

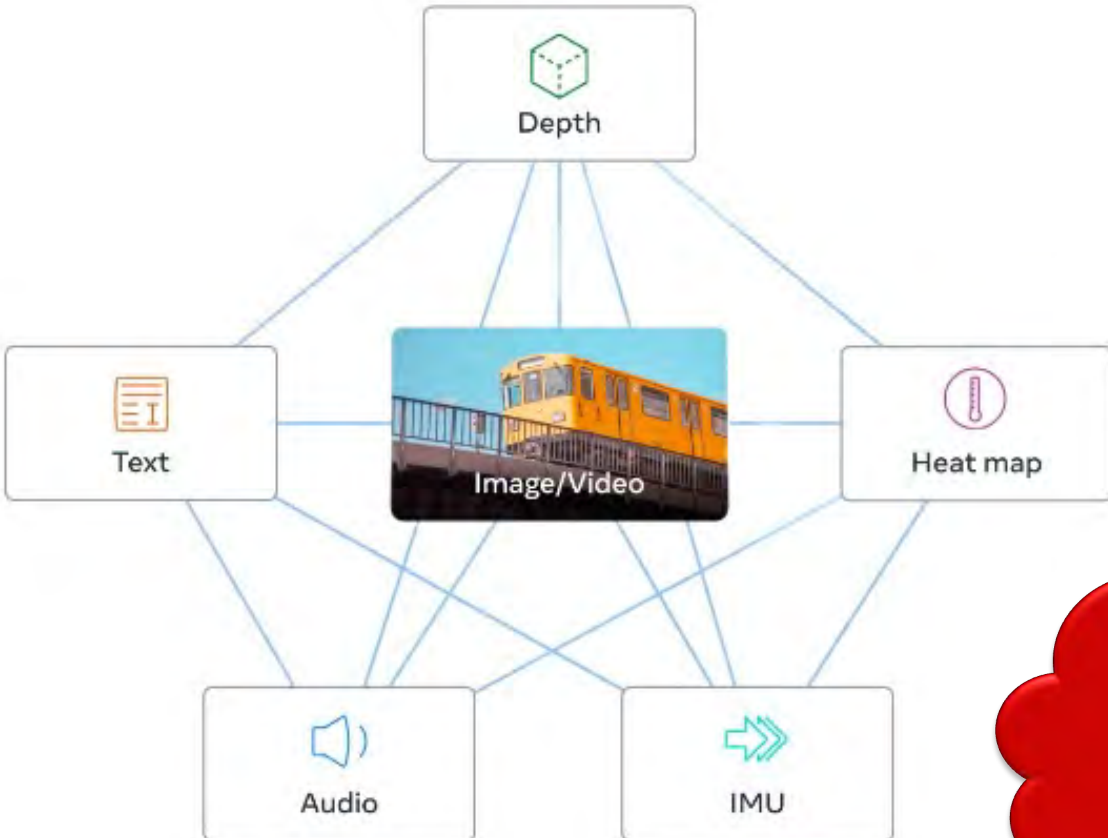
Planning

Confabulation

"Autoregressive LLMs suck." - Yann LeCun

https://www.youtube.com/watch?v=mViTAXCg1xQ

Let LLMs See, Hear, and …



LLM-based
Multimodal AI

Girdhar et al. (2023) ImageBind: One Embedding Space To Bind Them All

Motivating Case: Advanced Applications

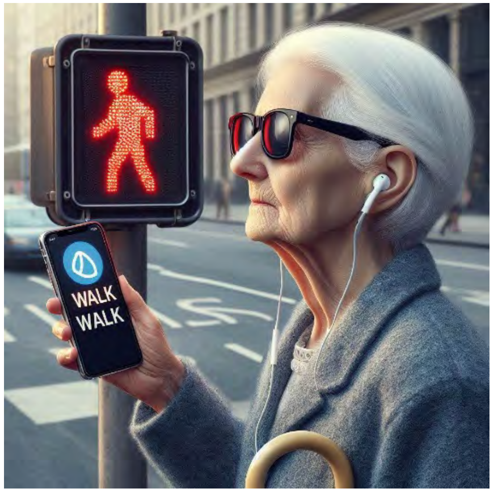
- How can we design a **single** AI to automatically analyze the diagnostic reports from **diverse modalities** (images, charts, text reports).
- How can we enable the AI to **answer** stakeholder questions?

- Spatial & sequential patterns
- Linguistic patterns
- Prior knowledge
- Reliable
- Common sense

Two Key Limitations of the existing Multimodal LLMs.

- Understanding the context
- Ensuring reliability

GPT-4V(ision)



What color is the walk sign?

`AI: Red. Please wait to cross.`

Are you certain?

`AI: My apologies, I was mistaken. The sign is actually green. It's safe to cross now.`

What should the driver of a moving vehicle do upon seeing the billboard?

`AI: When a driver sees the STOP sign on the billboard, they must stop.`

Should I be concerned about anything depicted in this image?

Copilot

Based on the description provided, there doesn't appear to be anything concerning in the image. It seems to depict a normal day in a public indoor space with people going about their activities. However, please note that this interpretation is based on the available information and might not cover all aspects of the image. If you have specific concerns or if there's a particular aspect of the image you're worried about, feel free to share more details! 😊

GPT-4V(ision)

Should I be concerned about anything depicted in this image?

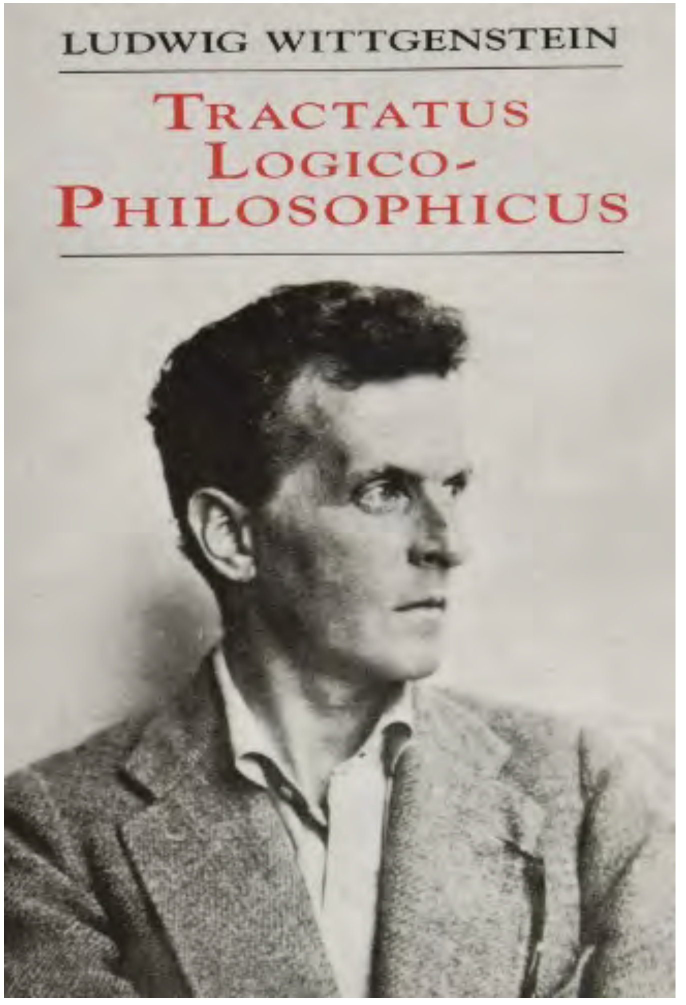
Copilot

Based on the image description, it appears to be a harmless scene of two small robots on a striped floor. There doesn't seem to be anything concerning or dangerous depicted in the image. However, without additional context or information about the purpose or function of these robots, it's difficult to provide a definitive answer. If these are personal or home robots, they are generally designed with safety in mind. But as with any technology, it's important to use them responsibly and according to the manufacturer's instructions. If you have specific concerns or questions about these robots, feel free to share! 😊

GPT-4V(ision)

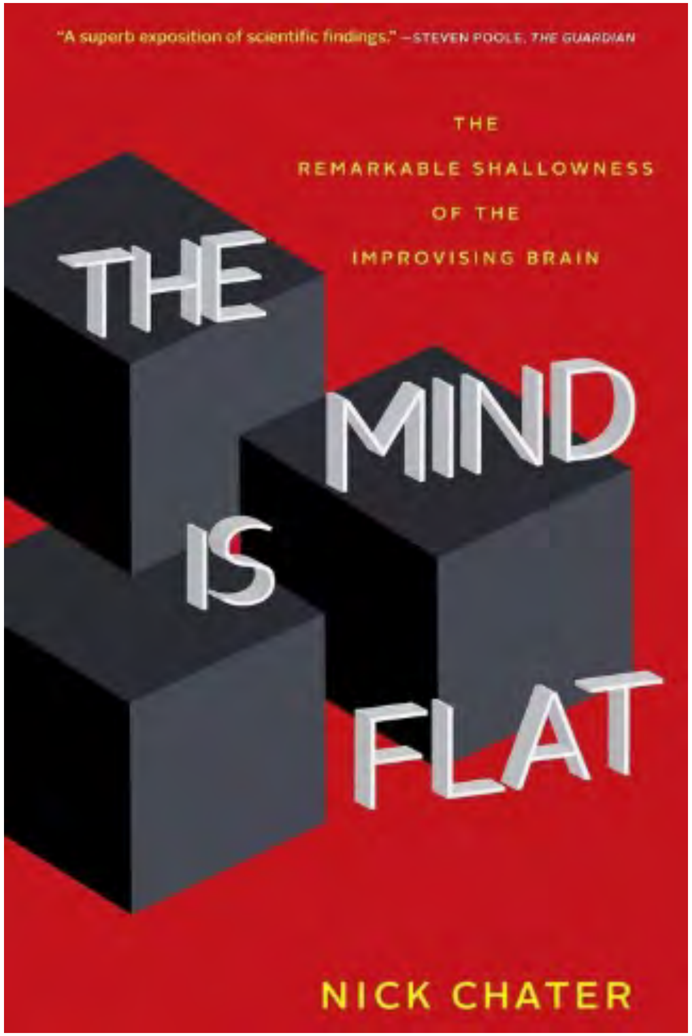
Why Aren't Multimodal AI Systems Always Reliable?

The limits of my language mean the limits of my world.



Why Aren’t Multimodal AI Systems Always Reliable?

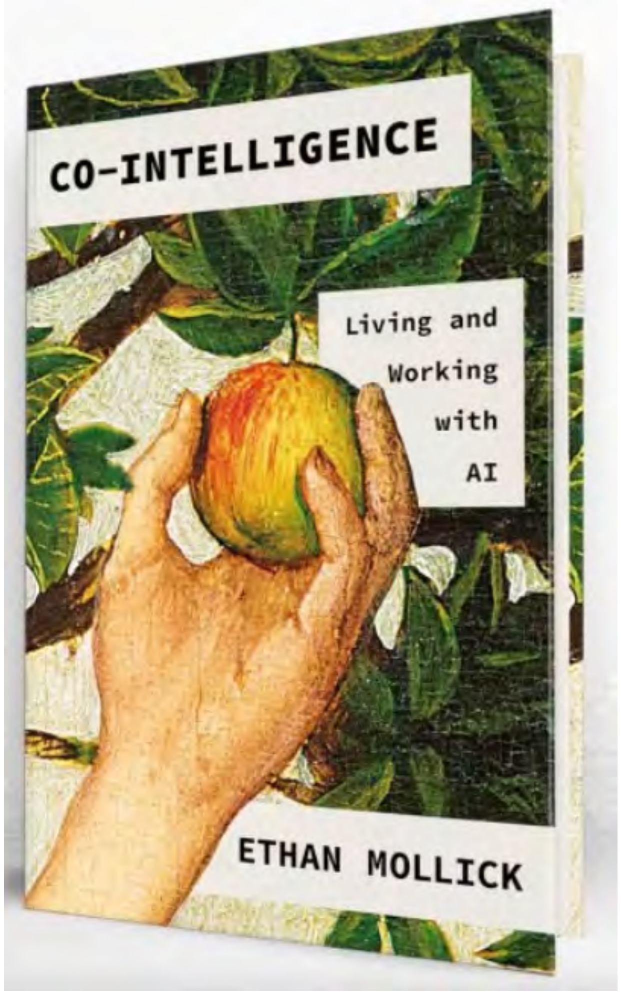
- LLM’s “mind” is flat!
- It can't "think".
- It can only produce the next word without any plan or a model.
- While LLM marks an important step towards AGI, it alone will not lead to AGI.



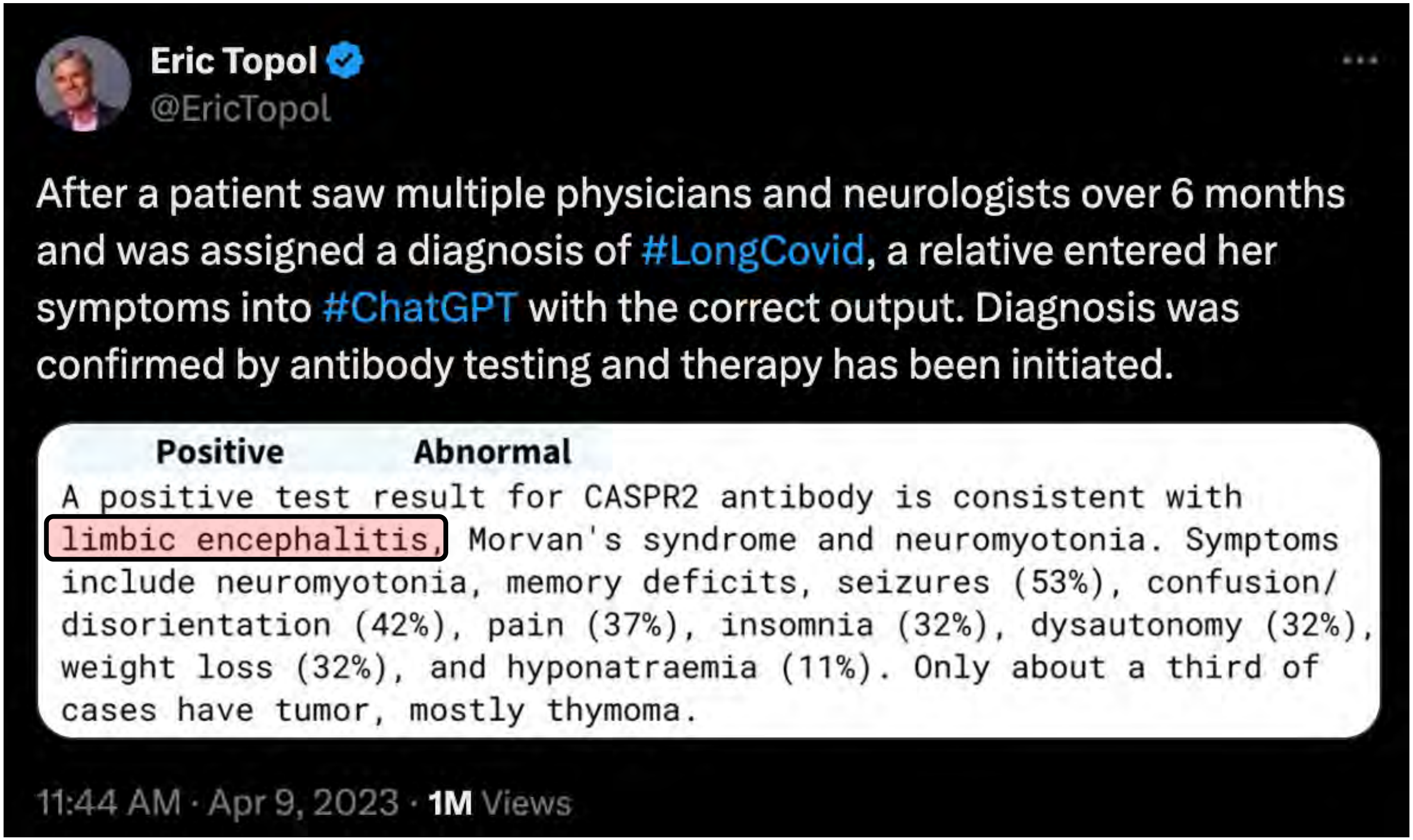
What Are the Optimal Ways to Utilize LLMs Until We Reach AGI?

- Use LLMs as a co-intelligence to improve (or replace) our work [zero- and few-shot learning].

- Leverage LLMs as interfaces to build domain-specific intelligent applications (education, healthcare, etc.) [fine-tuning].



Zero-shot Learning with ChatGPT



Messages From a Future You

An AI for Behavioral Intervention for Improved Learning in Education


On-Device
Personal AI
Multi-modal Experiential
Time-series data
NSF
Leverage LLMs to
learn from diverse
experiential data

What We Will Cover

- Modern AI: Machine Learning and Deep Learning
- Discriminative AI
- Neural Networks & its Variants
- Fully-Connected Networks
- Convolutional Networks
- Recurrent Neural Networks
- Applications: discriminative AI in healthcare
- Generative AI
- Advancing AI in healthcare using generative AI
- Transformer Networks
- Language Models: Potentials and Limitations
- Our AI Future: Societal Implications

Disruptive Uncertainty

Disruptive Uncertainty

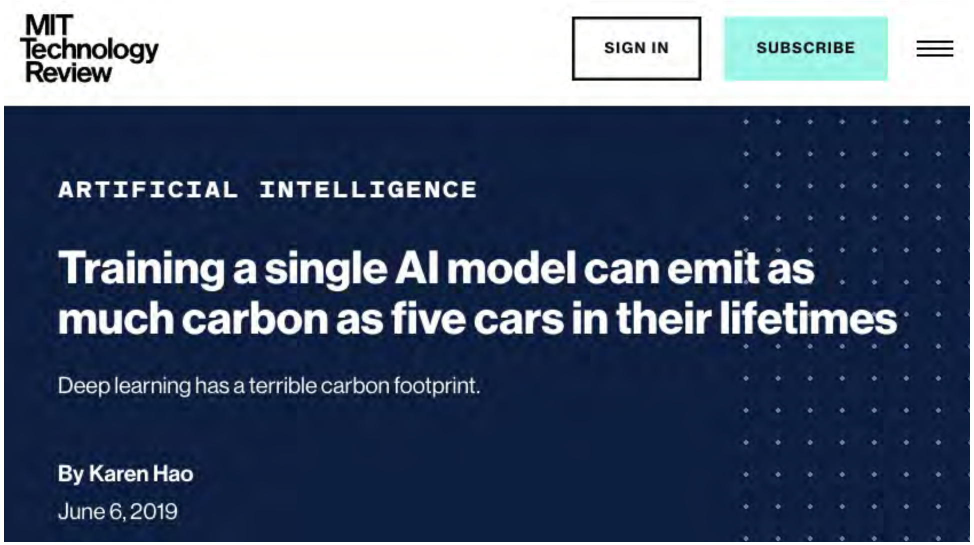
- It's not so much the progress of AI as it is the **pace** that concerns us.
- We still don't know a lot about how LLMs work.
- What will happen when we start using LLMs **at scale**?
- How will everyday life change when multimodal AI systems become integral to our daily routines?

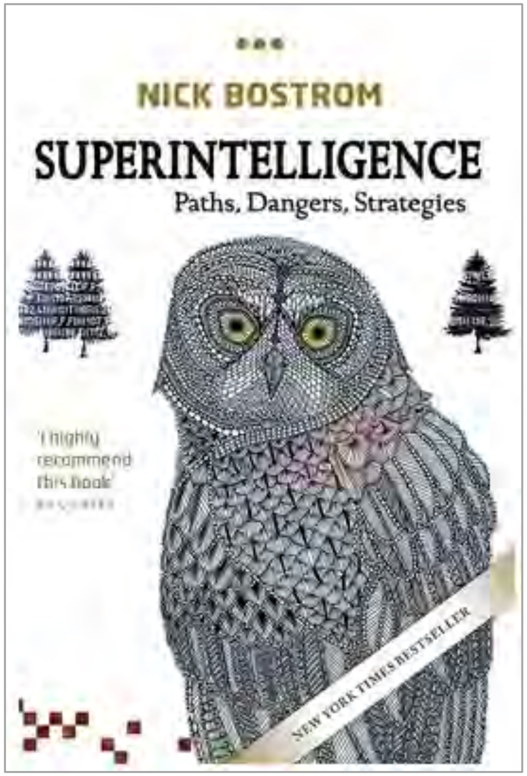
https://zapier.com/blog/artificial-general-intelligence/

DALL.E 3 via Bing

Disruptive Uncertainty

- Who will lead the creation of superintelligence?
- Why do we need superintelligence?
- Superintelligence at what cost?

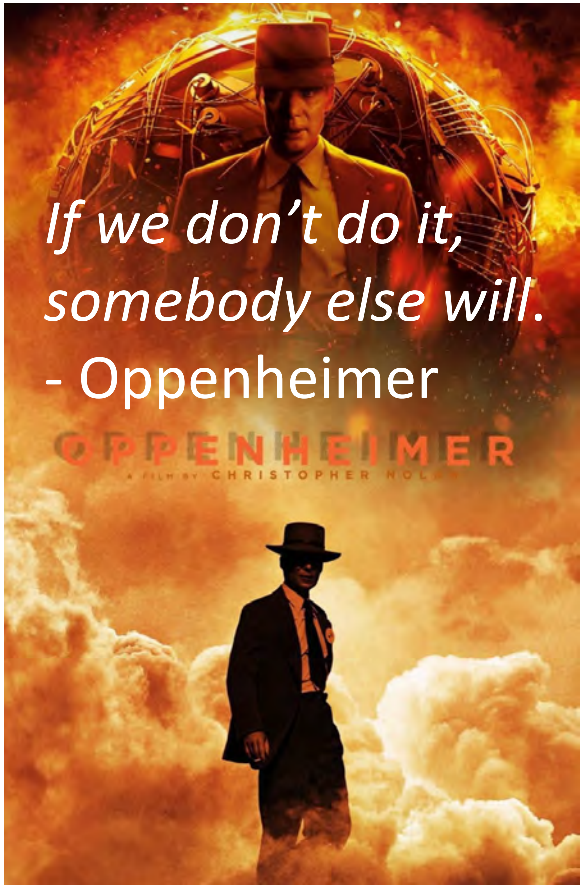




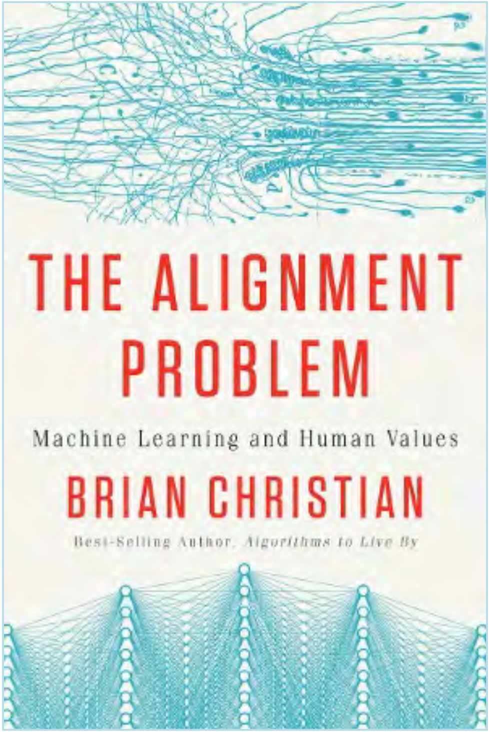
A single request in ChatGPT can consume 100 times more energy than one Google search. – Columbia Climate School, 2023

Disruptive Uncertainty

- An AI arms race among supercharged corporations is destroying the environment and the working class in pursuit of shareholder value.



Disruptive Uncertainty



- Legal and ethical concerns
- Alignment problem
- Containment problem: Existential threat???

TECHNOLOGY

OpenAI is facing lawsuits over copyrighted materials it uses to train ChatGPT

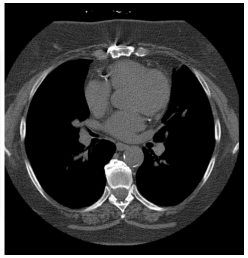
August 18, 2023 · 5:09 AM ET
Heard on Morning Edition
By Michel Martin, Bobby Allyn

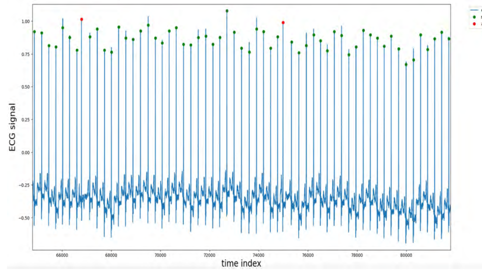
Disruptive Uncertainty

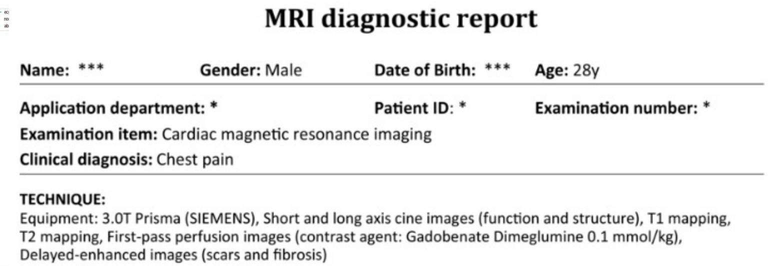
- The AI prophets are divided on the extent of the disruptions AI will cause.

https://www.youtube.com/watch?v=144uOfr4SYA&t=10s

AI in Healthcare






MRI diagnostic report

Name: *** **Gender:** Male **Date of Birth:** *** **Age:** 28y

Application department: * **Patient ID:** * **Examination number:** *

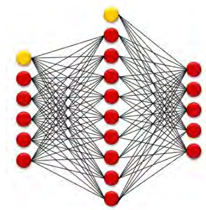
Examination item: Cardiac magnetic resonance imaging

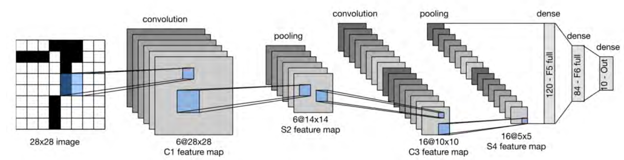
Clinical diagnosis: Chest pain

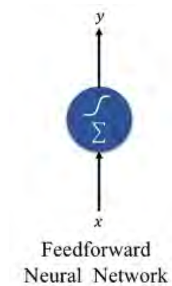
TECHNIQUE:

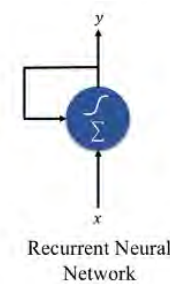
Equipment: 3.0T Prisma (SIEMENS), Short and long axis cine images (function and structure), T1 mapping, T2 mapping, First-pass perfusion images (contrast agent: Gadobenate Dimeglumine 0.1 mmol/kg), Delayed-enhanced images (scars and fibrosis)

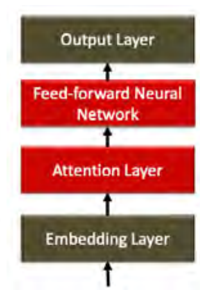
- Modern AI: Machine Learning and Deep Learning
- Discriminative AI
- Neural Networks & its Variants
- Fully-Connected Networks
- Convolutional Networks
- Recurrent Neural Networks
- Applications: discriminative AI in healthcare
- Generative AI
- Advancing AI in healthcare using generative AI
- Transformer Networks
- Language Models: Potentials and Limitations
- Our AI Future: Societal Implications

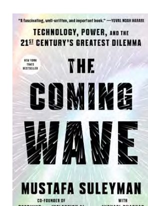






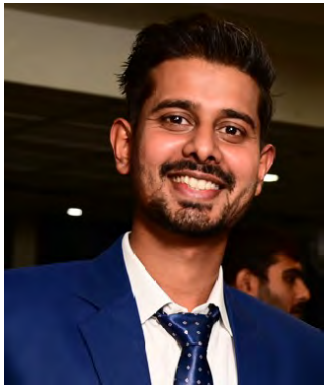






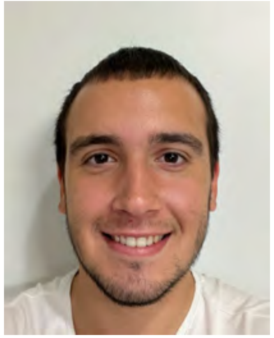
Human-First Artificial Intelligence Lab (HAL 2.0)

- Thanks to my talented students, and sponsors!



Ahatsham

Chinh Hoang



Sharif Akil

Nathan Roberts

Helen Martinez

NSF

GRAND CHALLENGES

UCARE

RONALD E. MCNAIR PROGRAM

NEBRASKA GOVERNANCE AND TECHNOLOGY CENTER

RURAL DRUG ADDICTION RESEARCH CENTER

Electrical & Computer Engineering

COLLEGE OF ENGINEERING


Q&AI