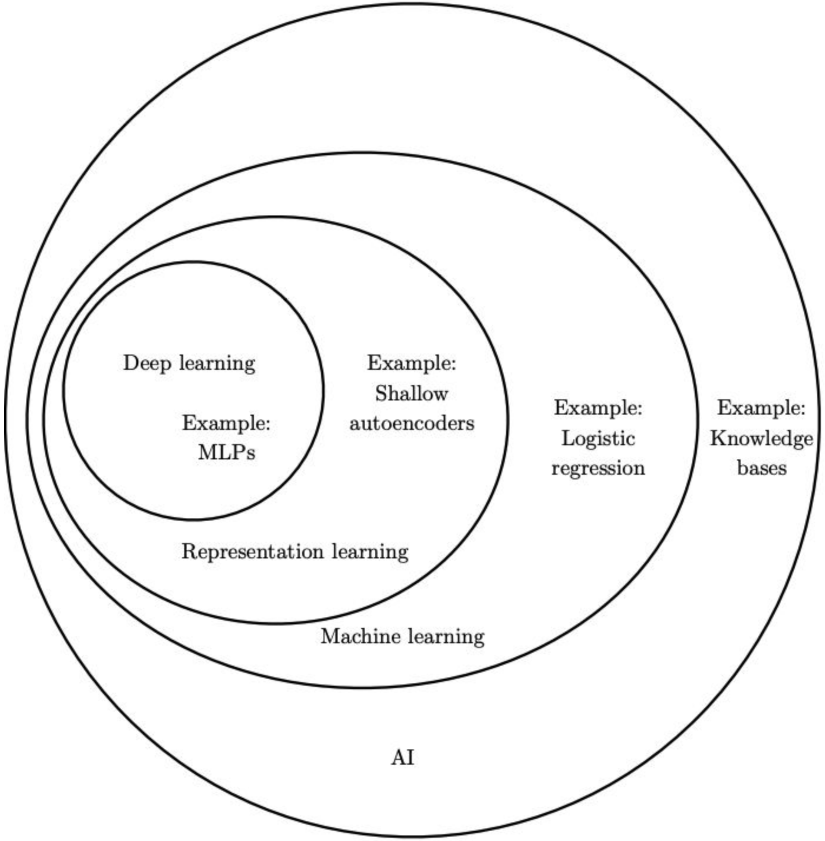
Intro to Deep Learning

Motivation and History

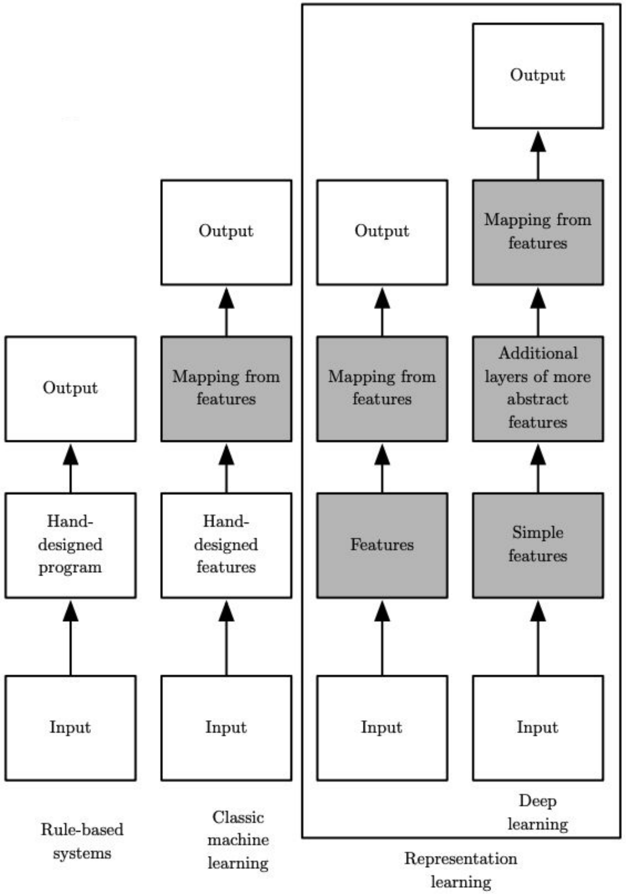
N. Rich Nguyen, PhD

SYS 6016

High-level Schematic of AI



How is deep learning different?

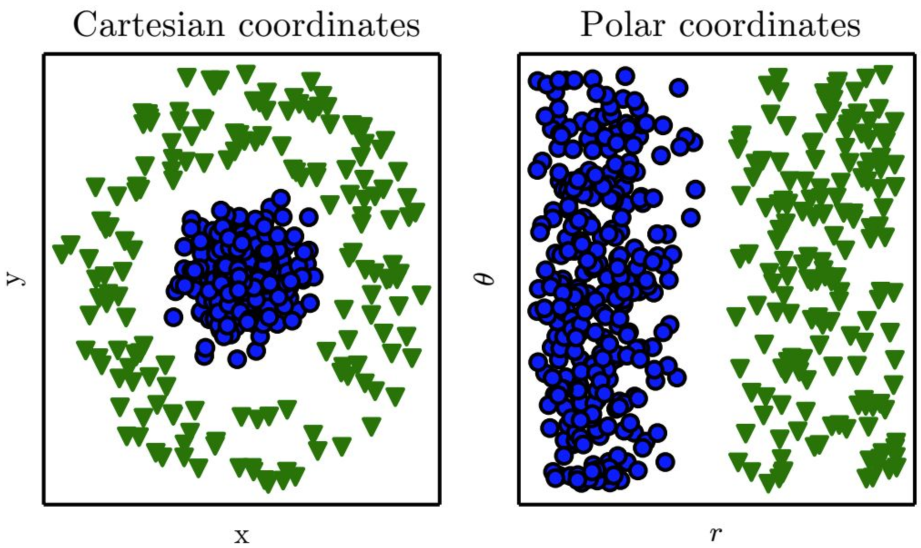


Presentations Matter

The performance of classical machine learning algorithms depends heavily on the representation of the data, aka **features**.

However, for many tasks, it is difficult to extract such high-level, abstract features from raw data.

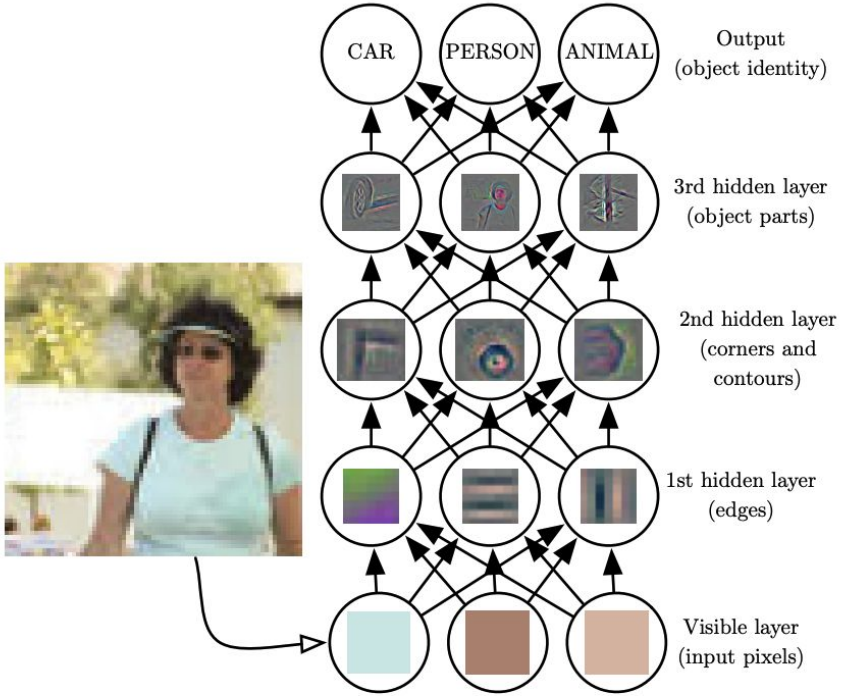
This process required as much sophisticated, near human-level understanding of the data → **representation learning**.



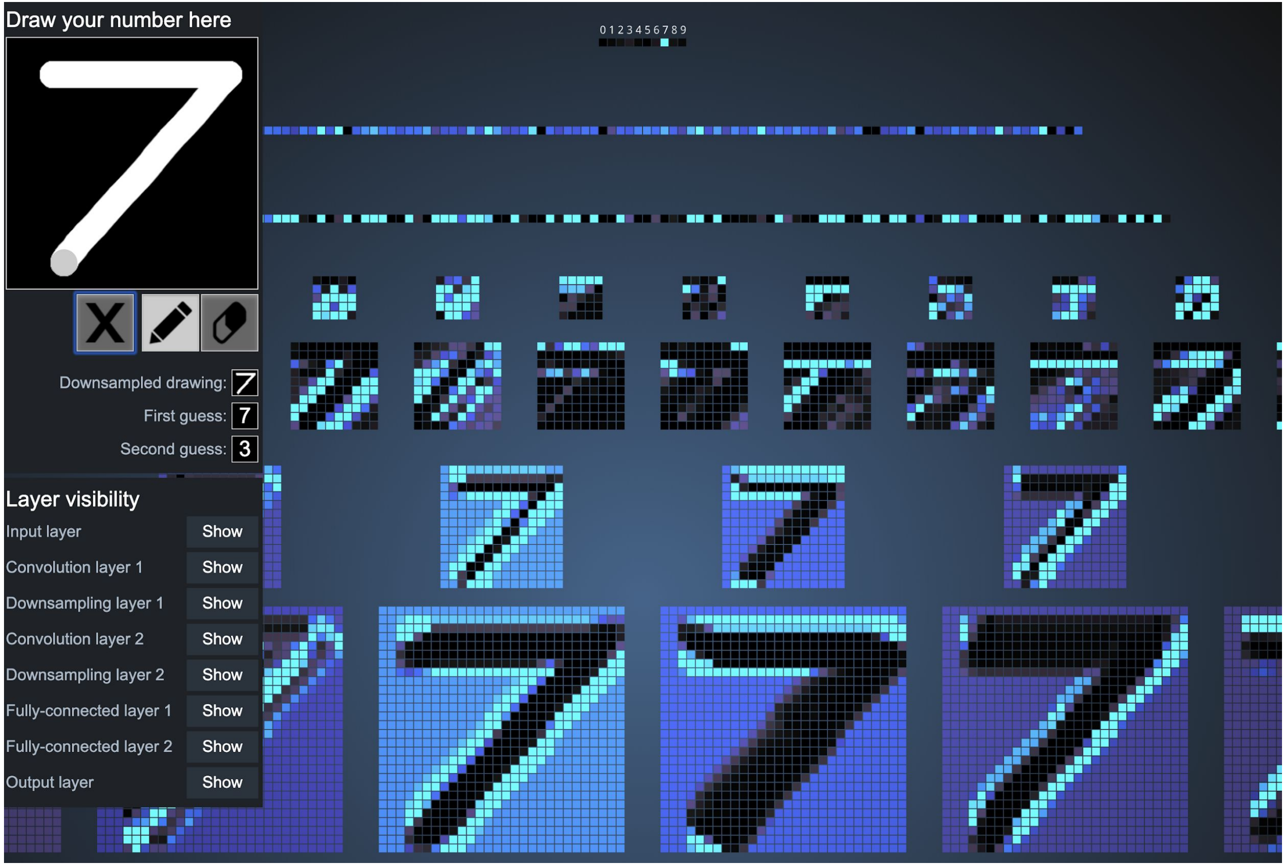
The rise of Deep Learning

As an approach to AI, Deep learning computes abstract representation in terms of other simpler representations, potentially at a multiple nested levels **in a hierarchy**.

If we draw a computational graph showing how these representations are build on top of each other, the graph is typically **deep** with many layers. This explains the “deep” in Deep Learning.

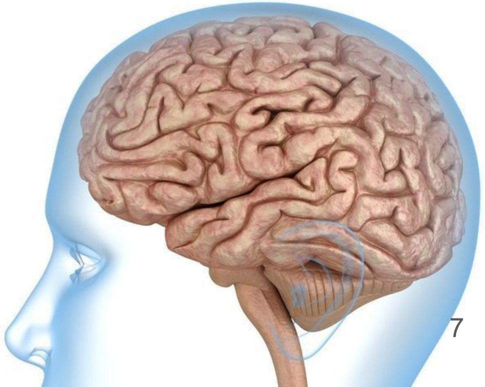


Example: Deep Learning on classifying numbers



The neural perspective of deep learning

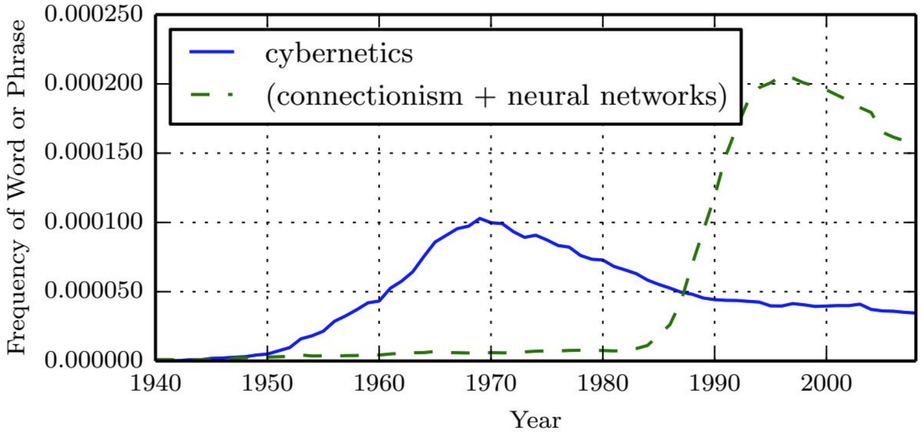
- Deep Learning is also known as **artificial neural network** as its early models are engineered systems *inspired* by the **brain** (biological neural network)
- However, artificial neural networks used for machine learning are generally **not** designed to be realistic models of biological function
- The modern term of “**deep learning**” is not necessarily neurally inspired, but appeals more to the principle of *learning at multiple levels of representation*
- Deep Learning has drawn heavily on our knowledge not only of the human brain architecture, but also in statistics and applied math



Historical Timeline* of Deep Learning

As a field, Deep Learning has been around for a long time and known under many names. Broadly speaking, there have been three waves of development:

- First wave: known as Cybernetics (1940s-1960s)
- Second wave: rebranded as Connectionism (1980s-1990s)
- Third wave: rebranded as Deep Learning (2006-present)



*These dates are for perspective only and not as definitive historical record of invention

First wave: Cybernetics (1940s-1960s)

- 1943: Inspired by the brain architecture, **neurons** and theories of biological learning proposed by McCulloch and Pitts
- 1958: **Perceptrons** created by Rosenblatt could learn the weights that defined the categories given inputs from each category.
- 1960: **Adaptive Linear Element** (ADALINE) by Widrow and Hoff could predict numeric values from data. Early success led to failed belief of its capability.
- 1969: **Backlash** against biologically inspired learning by Minsky and Papert mainly because its linear models could not learn the XOR function.
- 1970s: **First AI winter** as many promises would go unfulfilled

*These dates are for perspective only and not as definitive historical record of invention

Second wave: Connectionism (1980s-1990s)

- 1986: Revival of interest in research with **Backpropagation** by Rumelhart
- 1994: Advanced **modeling sequences** with recurrent neural nets by Bengio
- 1998: Early **convolutional neural nets** by LeCun (LeNet-5) performed well on MNIST dataset of handwritten digits
- 1990s-2000s: **Second AI winter** as deep networks were generally believed to be *very difficult to train*. Kernel methods (SVM) and graphical models were preferred by AI researchers and practitioners.
- 2004: The Canadian Institute for Advanced Research (**CIFAR**) helped *keep the research alive* with its Neural Computation and Adaptive Perception (**NCAP**) research initiative led by Hinton (U of Toronto), Bengio (U of Montreal) and LeCun (NYU).

*These dates are for perspective only and not as definitive historical record of invention

Third wave: Deep Learning (2006-present)

- 2006: A research breakthrough by Hinton, **deep belief network,** could be trained *efficiently* by a strategy called *greedy layer-wise pretraining*.
- 2007: Same training strategy adapted by other CIFAR-affiliated groups.
- 2012: First Deep Learning method, **AlexNet**, won the ImageNet Competition.
- 2014: Development of generative adversarial network (**GAN**)
- 2016: Achievement in deep reinforcement learning with **AlphaGo, AlphaZero**
- Present: Has seen tremendous growth in **popularity** and **usefulness**

Would this new wave of interest last?

*These dates are for perspective only and not as definitive historical record of invention

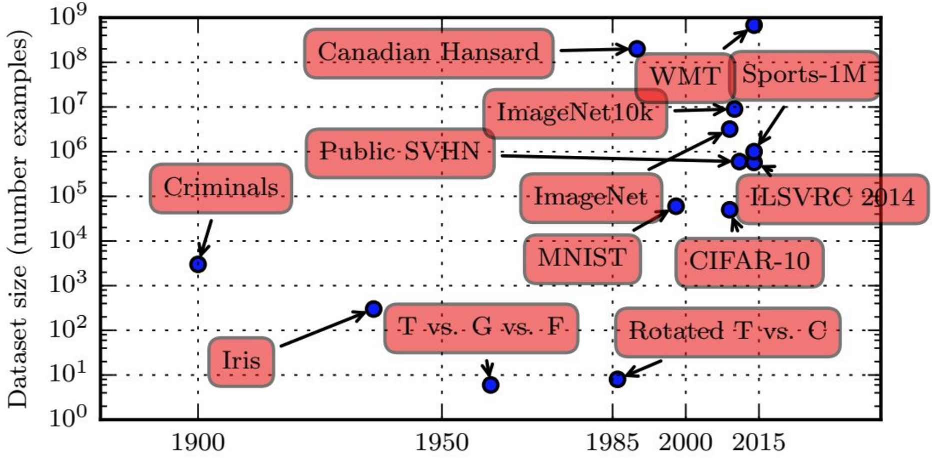
A few reasons why this time will have impact

1. **Data:** Huge *quantity of data* available to train
2. **Hardware:** Tremendous increase in *computing power*
3. **Algorithms:** Improved training procedures
4. **Investment:** A virtuous circle of *funding and progress*

Reason 1: Data

Game changer has been the rise of the Internet and the age of “Big Data” making it feasible to collect very large datasets for machine learning:

- User-generated tags on Flickr images and Youtube videos
- ImageNet dataset and annual competition
- Wikipedia for natural language processing



Reason 2: Hardware

Thanks to Moore's Law, we have the computational resources to run much larger models today:

- Availability of faster CPUs (5000X between 1990 and 2010)
- Advent of general purpose GPUs for parallel computing (neural network consisting of many small matrix multiplications are highly parallelizable).
- Beyond GPU: Google revealed its tensor processing unit (TPU), a new chip design developed from ground up to run deep neural networks (10X faster)
- Faster network connectivity

This trend is generally expected to continue well into the future

Reason 3: Algorithms

Previously, back-propagation through deep stack of layers had fade away feedback signal. Advent of several algorithmic improvements that allowed better gradient propagation:

- Better **activation function** for neural layers
- Improved **weight initialization** schemes
- Advanced **optimization schemes** (RMSProp and Adam)
- Many recent **algorithmic innovations** (batch normalization, residual connection, depth-wise separable convolutions)

Better software infrastructure and **libraries** (Theano, PyLearn, Torch, Caffee, MXNet, and TensorFlow)

Reason 4: Investment

As deep learning become the new state-of-the-art for computer vision and eventually all perceptual tasks, industry leaders take note.

Total venture capital investment in AI jumped from $19M (2011) to $384M (2014).

Many tech giants invested in R&D AI departments.

- Google acquired deep-learning startup DeepMind for $500M in 2013
- Baidu started a Deep Learning research center in Silicon Valley investing $300M in 2014
- Intel bought deep-learning hardware Nervana Systems for $400M in 2016

Research progress has reached a frenetic pace.

2018 Turing Award

- The ACM **Turing Award**, often referred to as the "**Nobel Prize** of Computing," carries a $1 million **prize**.
- 2018 Turing Award is given for "the conceptual and engineering breakthroughs that have made deep neural networks a critical component of computing."

Bonus Content

Sundar Pichai, CEO of Google, on ML in 2015

"Machine Learning is a core, transformative way by which we're rethinking how we're doing everything. We're thoughtfully applying it across all of our products, be it search, ads, YouTube, or Play. And we're in the early days, but you'll see us -- in a systematic way -- apply machine learning in all these areas" - Sundar Pichai, 2015

Will deep learning last?

Deep learning (DL) has several properties that justify its long-lasting status:

- **Simplicity**: DL replaces the need for feature engineering (complex engineering-heavy pipelines) with simple, end-to-end trainable models
- **Scalability**: DL models are highly parallelable, so they can take advantage of Moore's Law. They are also trained over small batch of data, allowing them to be trained on dataset with arbitrary size
- **Versatility**: DL models can be trained on additional data without restarting from scratch, making them viable for continuous online learning.
- **Reusability**: DL models are repurposable and reusable, which allow reinvest previous work into more complex and powerful models

What deep learning has achieved so far?

- Near human-level image classification
- Near human-level speech recognition
- Near human-level handwriting transcription
- Improved machine translation
- Improved text-to-speech conversion
- Near-human-level autonomous driving
- Improved ad targeting
- Improved search results
- Ability to answer natural language questions
- Super-human Go playing
- And much more...