DEEP LEARNING APPLICATIONS FOR BUSINESS

Introduction

HOW AI/ML AND RELATED TECHNOLOGIES ARE CHANGING BUSINESS?

The Changing Context

Data is being produced and consumed at an increasingly phenomenal rate.

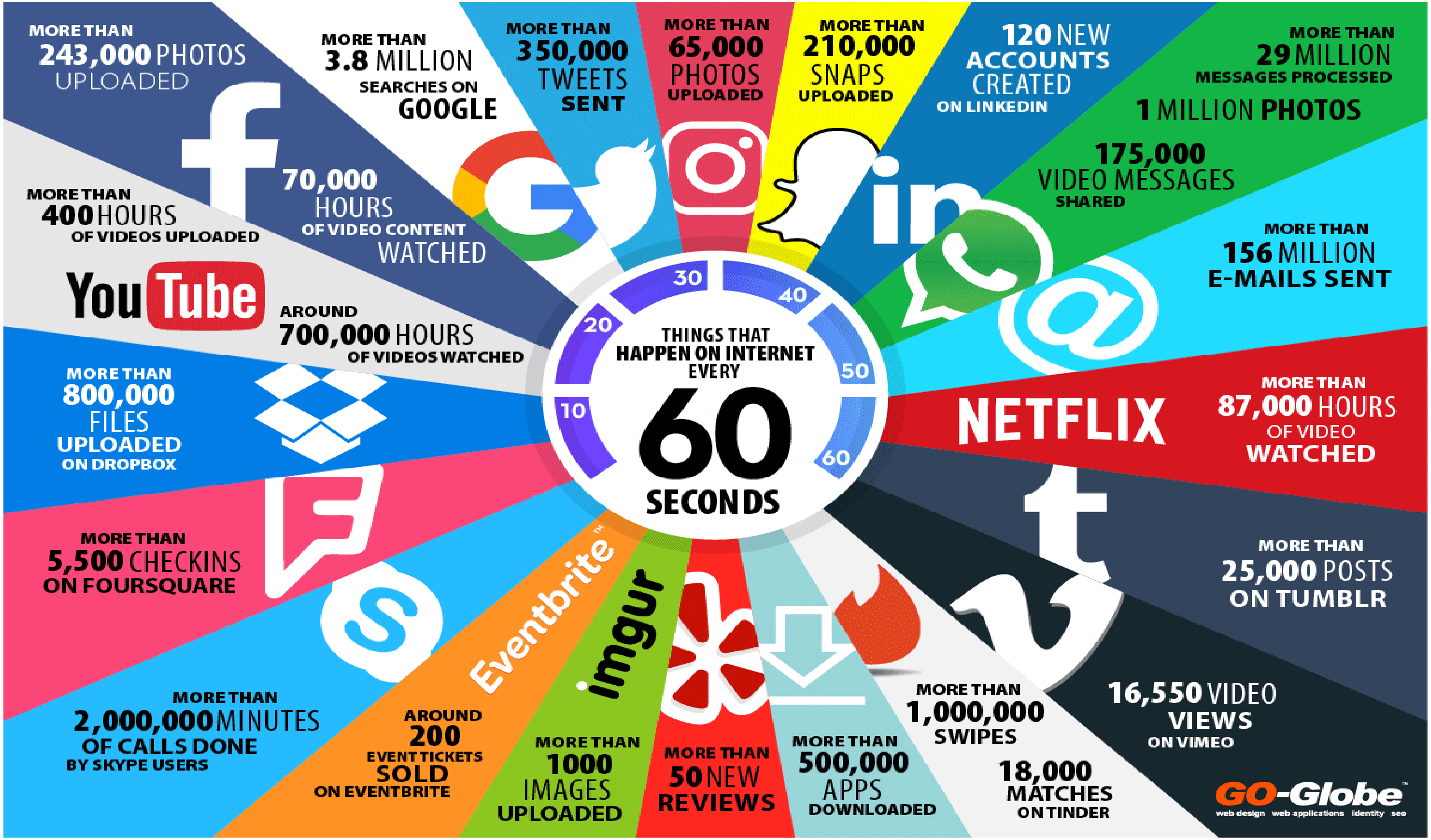
Technology trends

- Improved data creation and collection technologies
- Computing has become faster but cheaper
- Data storage and retrieval has become faster but cheaper
- New tools and techniques to analyze data

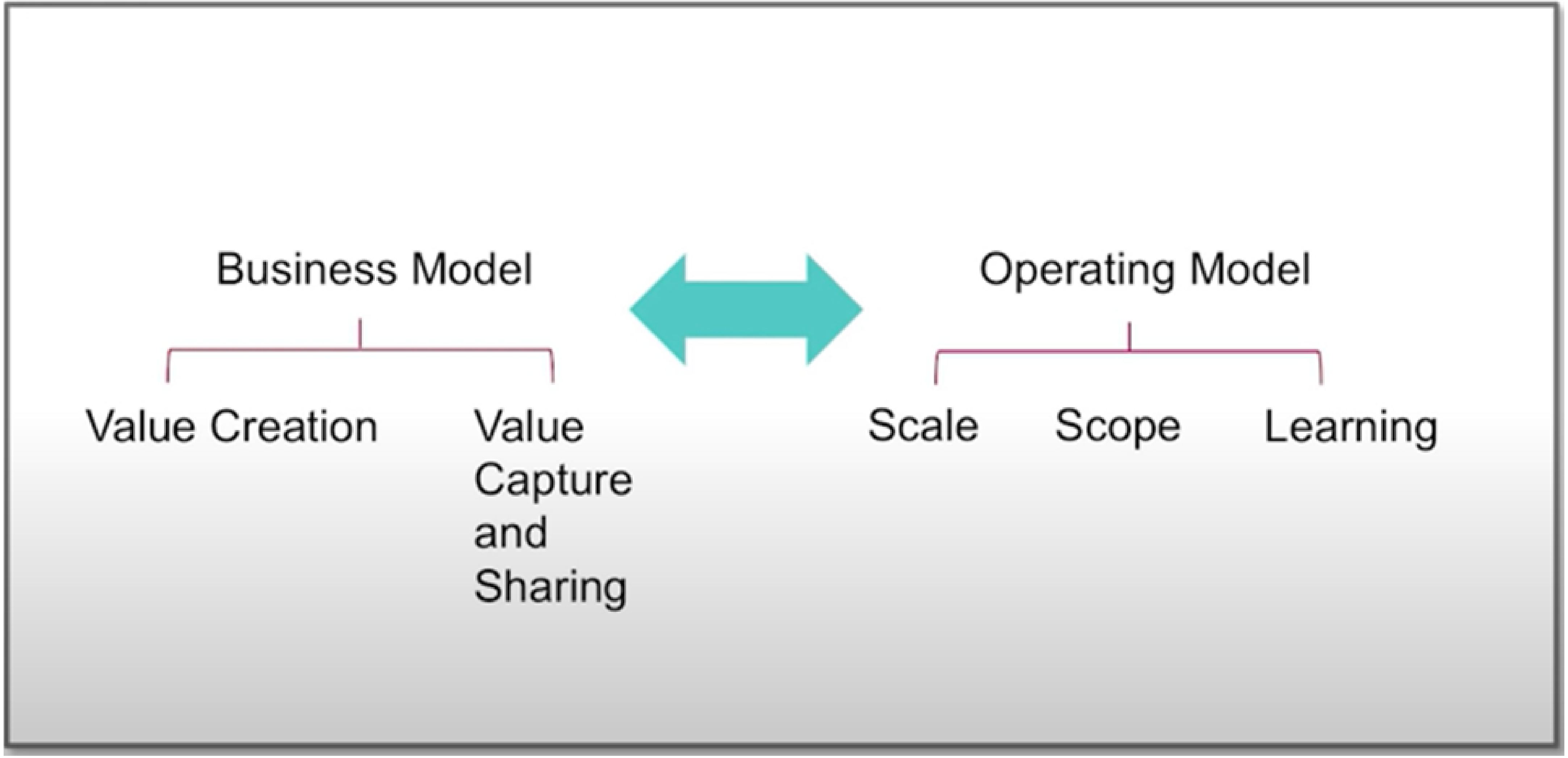
Business trends

- Technology has created a level-playing field for organizations
- Economies of Scale and Scope no longer give businesses the key competitive advantage
- Firm size does not have the same competitive advantage as before
- Increasingly "Data" has been looked upon as a source of competitive advantage

THE DATA DELUGE

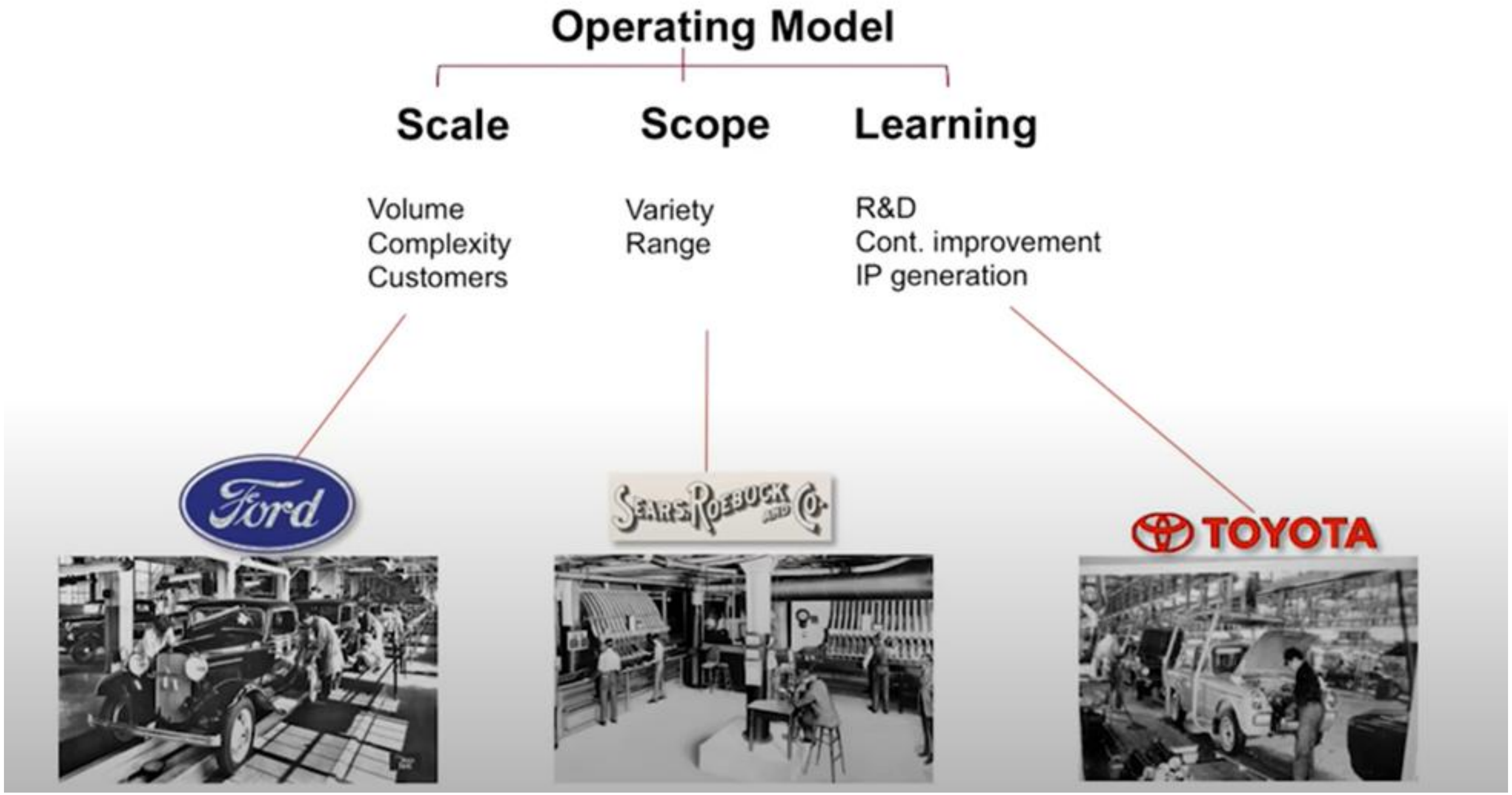


HOW FIRMS CREATE AND DELIVER VALUE.



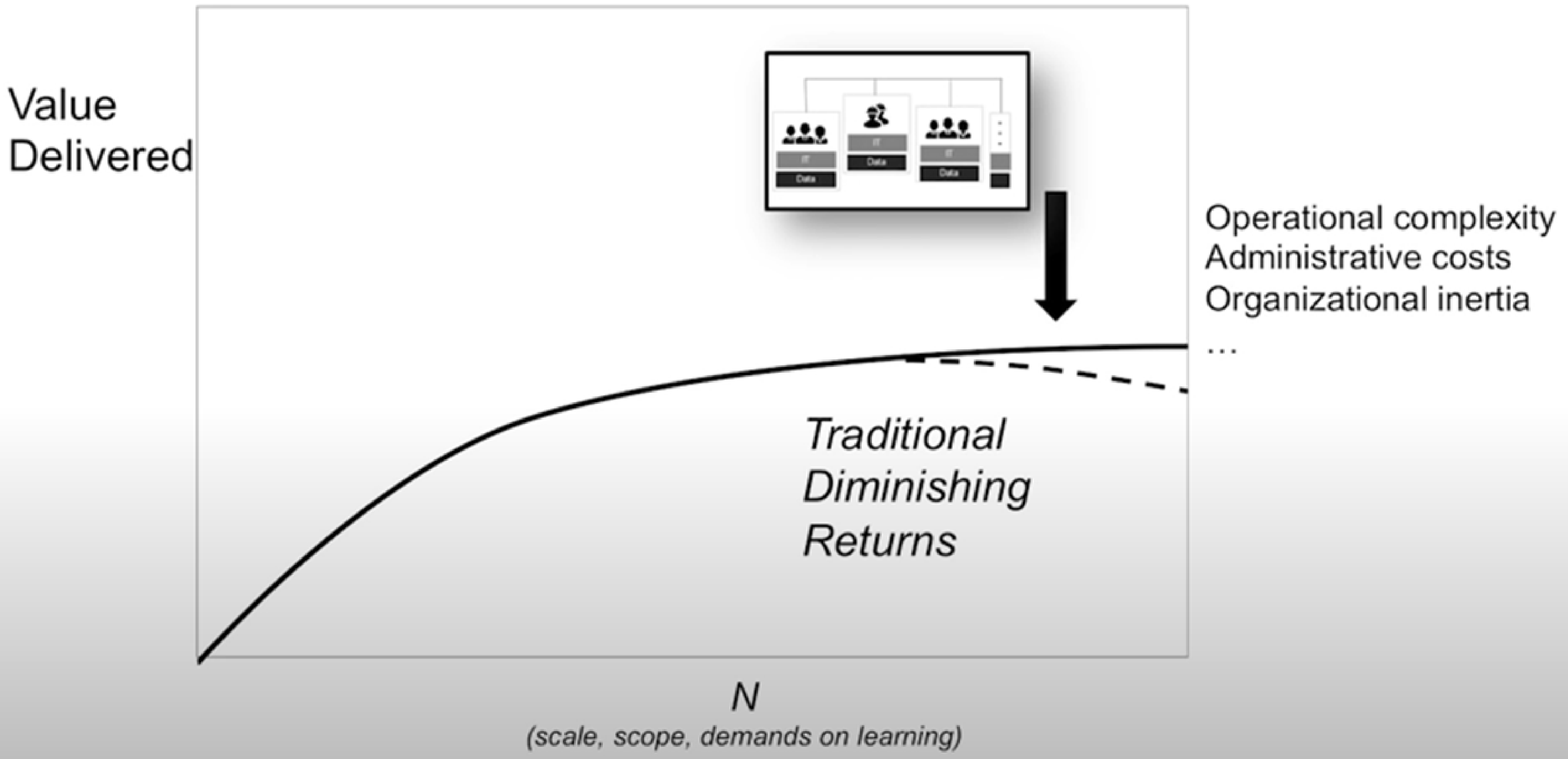
Source: *New rules in the age of AI, Karim R. Lakhani, Harvard Business School*

TRADITIONAL OPERATING PERFORMANCE DRIVERS



Source: *New rules in the age of AI, Karim R. Lakhani, Harvard Business School*

Traditional Operating Models Can Constrain the Value of the Firm

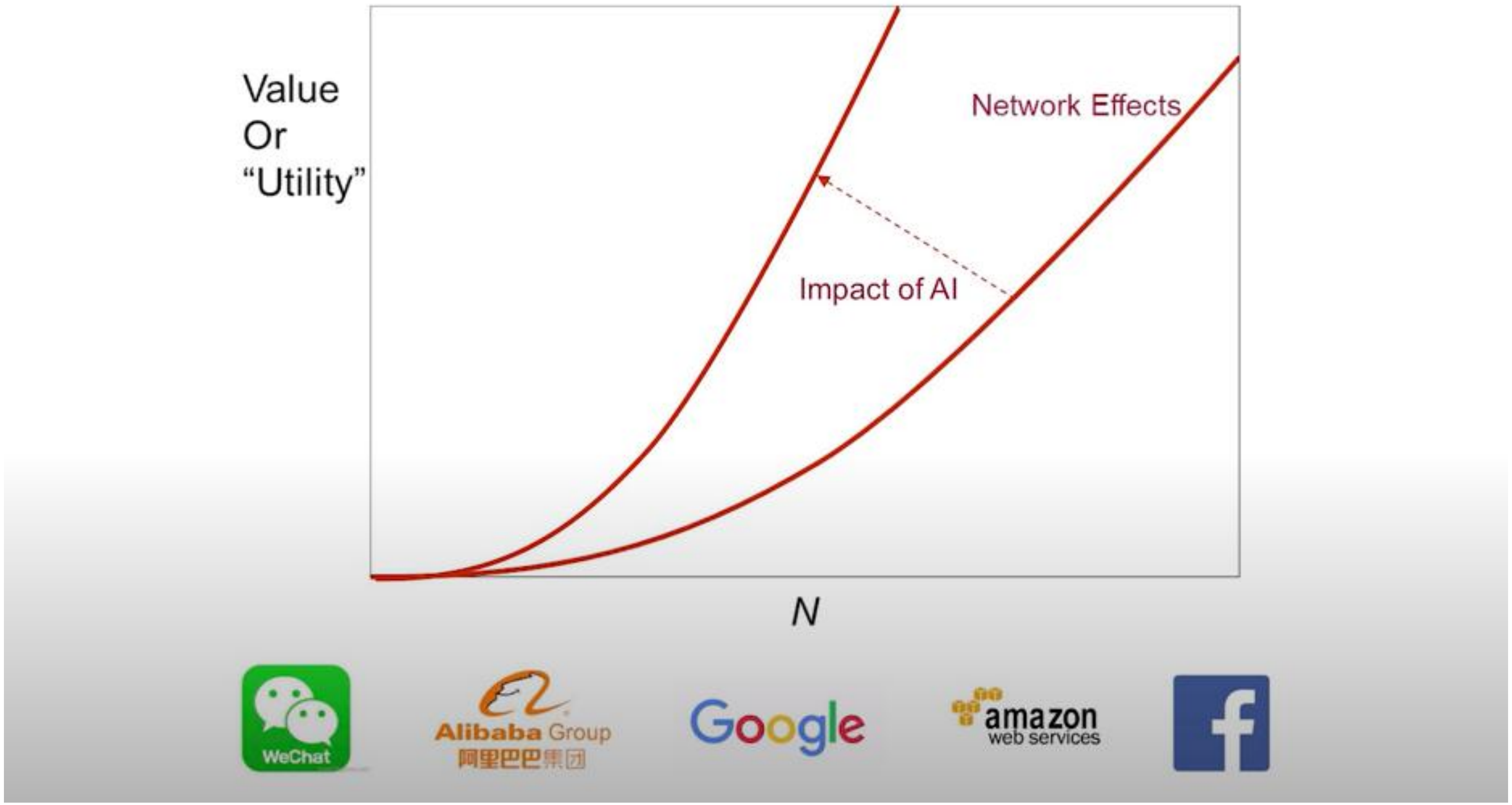


Source: *New rules in the age of AI, Karim R. Lakhani, Harvard Business School*

DIGITAL OPERATING MODELS

- Software programs and algorithms ***replace humans in the critical path*** in the way the firm delivers value
- More scalable than traditional processes
- Enable greater scope or variety
  - Easily connect with a myriad of other digitized business processes
- Create powerful opportunities for learning and improvement
- Ability to produce even more accurate, complex and sophisticated predictions
- Gain fundamental understanding of customers

Digital Operating Models have Near Zero Marginal Costs and can Generate “Network” and “Learning” Effects



Source: *New rules in the age of AI, Karim R. Lakhani, Harvard Business School*

Digital Operating Models

…move human labor off the critical path and remove traditional constraints to scale, scope and learning

Operating Model

Scale

"Zero" marginal cost, network effects

amazon web services

Scope

Aggregation and modularity across markets,

Alibaba Group 阿里巴巴集团

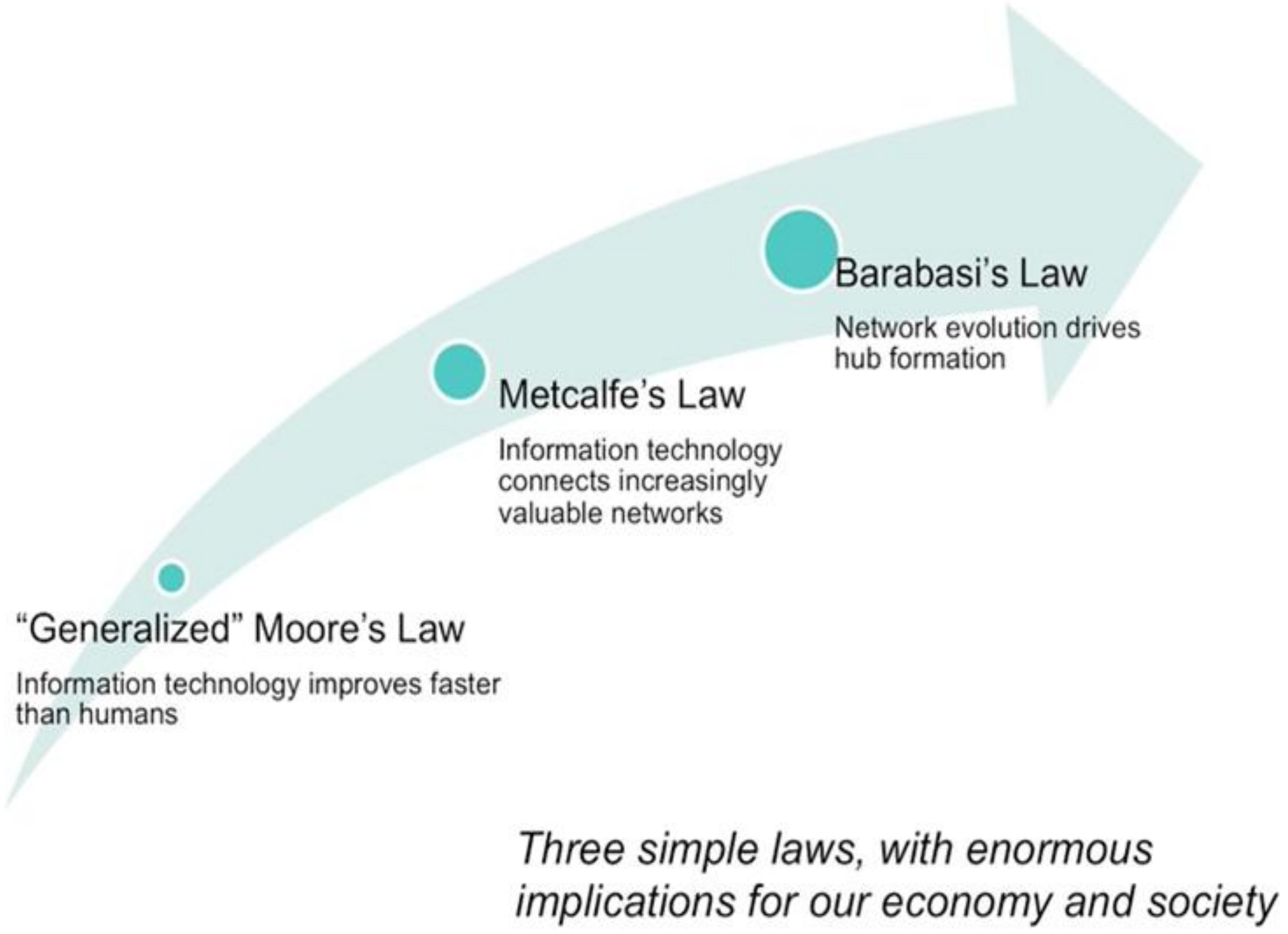
Learning

Constant innovation and AI/ML driven improvement

Google

Source: *New rules in the age of AI, Karim R. Lakhani, Harvard Business School*

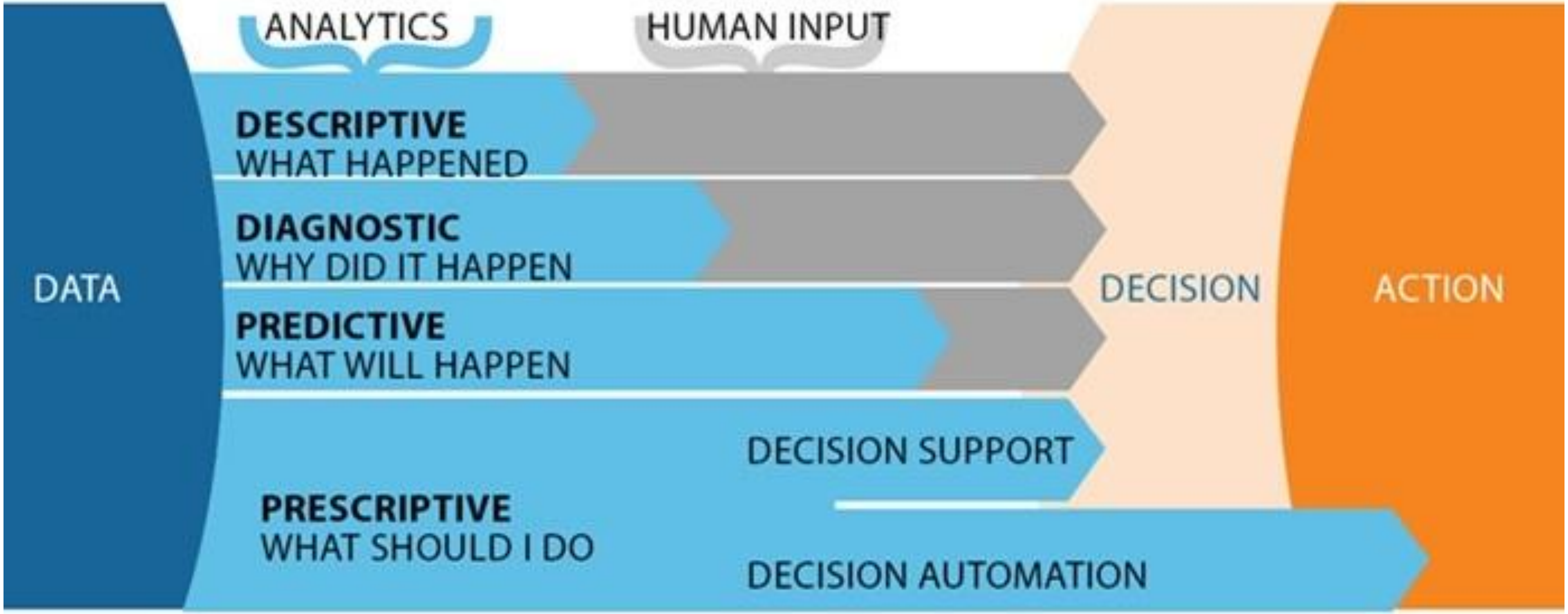
SYSTEMIC CHANGE: DIGITAL TECHNOLOGY IS EVERYWHERE AND ITS EVOLUTION IS HITTING THE ENTIRE ECONOMY AT THE SAME TIME.



Source: *New rules in the age of AI, Karim R. Lakhani, Harvard Business School*

CATEGORIZATION BASED ON BUSINESS QUESTIONS

From Data to Actions



Source: https://blog.dcrworkforce.com/hindsight-foresight-prescriptive-analytics-work

INTRODUCTION & HISTORY OF NEURAL NETWORKS

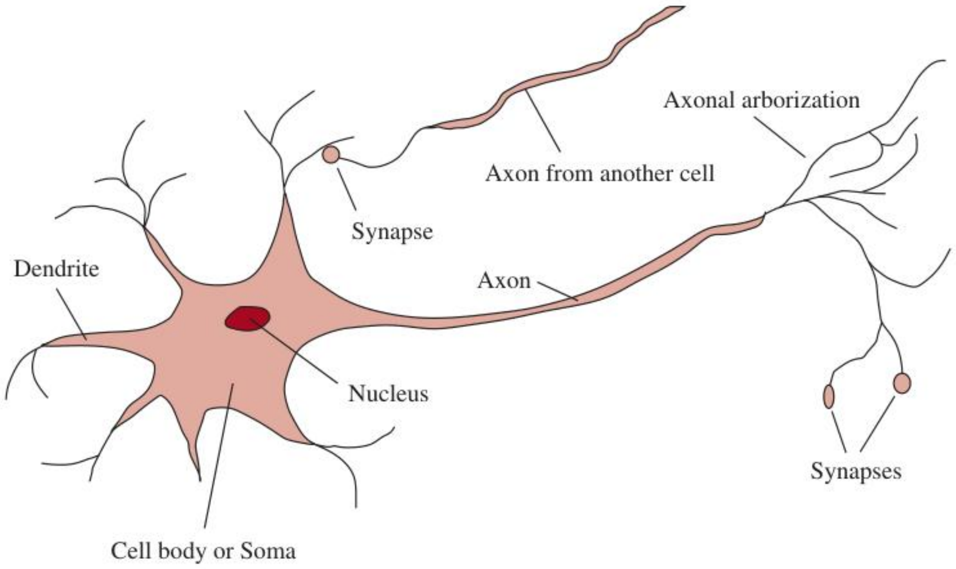
MOTIVATION

- Humanity for thousands of years has tried to understand how we think and act.
- Ability to perceive, understand, predict and manipulate a world that is far larger and more complicated than the humanity that inhabits it.
- Artificial intelligence, or AI, is concerned with both understanding and also building intelligent entities.
- Researchers have pursued different versions of AI based on their own opinions of what should be the subject matter of Artificial Intelligence

MOTIVATION

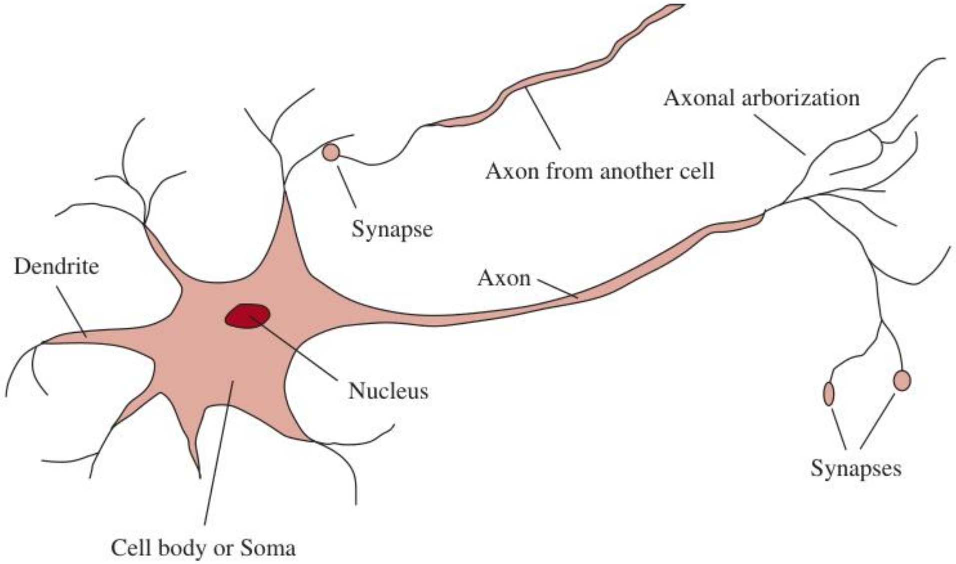
- Human intelligence possesses robust attributes for information processing and decision-making
  - Complex sensory, control and affective (emotional processes)
  - Cognitive Processes (Thought)
- Central Nervous Systems – biological neurons
- Continuous stream of information through our five natural sensory mechanisms
- The brain integrates information and provides an interpretation through cognitive computing
- Limited understanding of the human nervous system has led to emulation of certain human learning behaviours

The parts of a nerve cell or neuron [1]



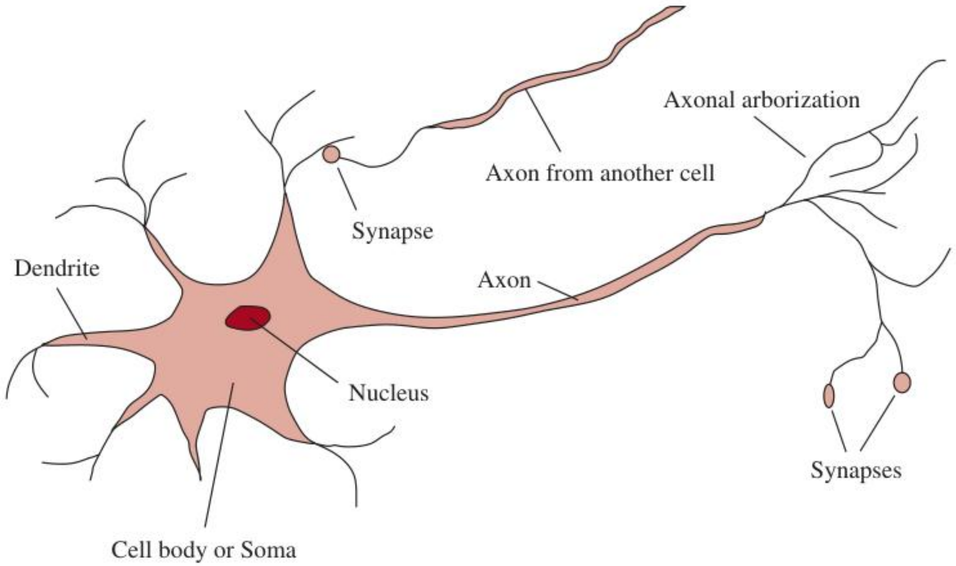
- A neuron consists of a cell body, or soma, that contains a cell nucleus.
- Dendrites are a number of fibers that branch out of the cell body.
- The single long fiber coming out of the cell body is called the axon.
- **Length of axon:** Typically, an axon is 1 cm long (100 times the diameter of the cell body), but can reach up to 1 meter.
- Each neuron makes connections with 10 to 100,000 other neurons at junctions called synapses.

The parts of a nerve cell or neuron [2]



- Signals are propagated from neuron to neuron by a complicated electrochemical reaction.
- The signals control brain activity in the short term and also enable long-term changes in the connectivity of neurons.
- These mechanisms are thought to form the basis for learning in the brain.
- Most information processing happens in the cerebral cortex, the outer layer of the brain.

The parts of a nerve cell or neuron [3]



- **Basic organizational unit**
  - a column of tissue about 0.5 mm in diameter,
  - containing about 20,000 neurons
  - and extending the full depth of the cortex (about 4 mm in humans)
- **Total number of neurons**
  - Over 100 billion

Total number of pathways in the brain exceeds the number of atoms in the universe !!

TWO PARADIGMS FOR ARTIFICIAL INTELLIGENCE

The Logic inspired approach – traditional AI (Since 1950s)

- The essence of intelligence is using symbolic rules to manipulate symbolic expressions.
- Focus on Reasoning

The Biologically Inspired Approach

- The essence of intelligence is learning the strength of connections in a neural network
- Focus on Learning and Perception

Source: G. Hinton, "The Deep Learning Revolution", Turing Award Lecture

TWO VIEWS OF INTERNAL REPRESENTATION

The Logic inspired approach

- Internal Representations are Symbolic Expressions
  - A programmer can give them to the computer using an unambiguous language
  - New representations can be derived by applying rules to existing representations

The Biologically Inspired Approach

- Internal Representations are nothing like language
  - Large Vectors of Neural Activity
  - They have direct causal effect on other vectors of neural activity
  - These vectors are learned from data

Source: G. Hinton, "The Deep Learning Revolution", Turing Award Lecture

TWO WAYS TO MAKE A COMPUTER DO WHAT YOU WANT

Intelligent Design/Programming

- Figure out consciously how exactly you would manipulate symbolic representations to perform the task at hand
- Describe to the computer in excruciating detail exactly what to do.

Learning

- Show the computer lots of examples of inputs together with the desired outputs
- Let the computer learn how to map the inputs to output using a general purpose learning procedure

Source: G. Hinton, "The Deep Learning Revolution", Turing Award Lecture

CENTRAL QUESTION

- Large neural networks containing millions of weights and many layers of non-linear neurons are very powerful computing devices
- But can a neural network learn a difficult task (like object recognition or machine translation) by starting from random weights and acquiring all of its knowledge from the training data?

Source: G. Hinton, “The Deep Learning Revolution”, Turing Award Lecture

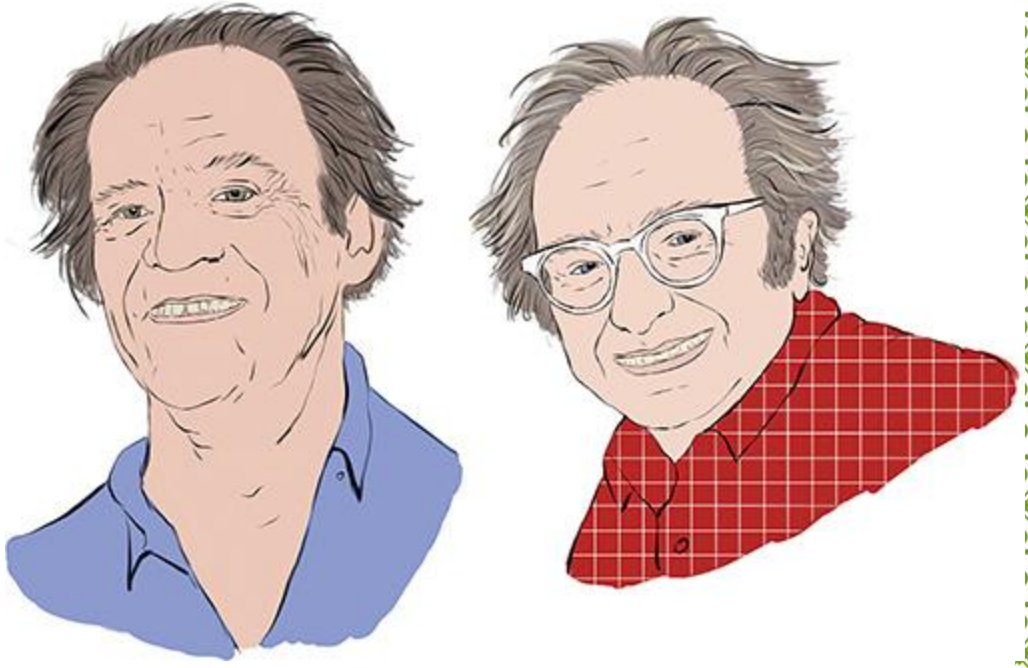
THE OBVIOUS LEARNING ALGORITHM

- Early researchers like Turing and Selfridge proposed that neural networks with initially random connections could be trained by reinforcement learning
  - This is extremely inefficient

Source: G. Hinton, "The Deep Learning Revolution", Turing Award Lecture

VISUAL INFORMATION PROCESSING [1]

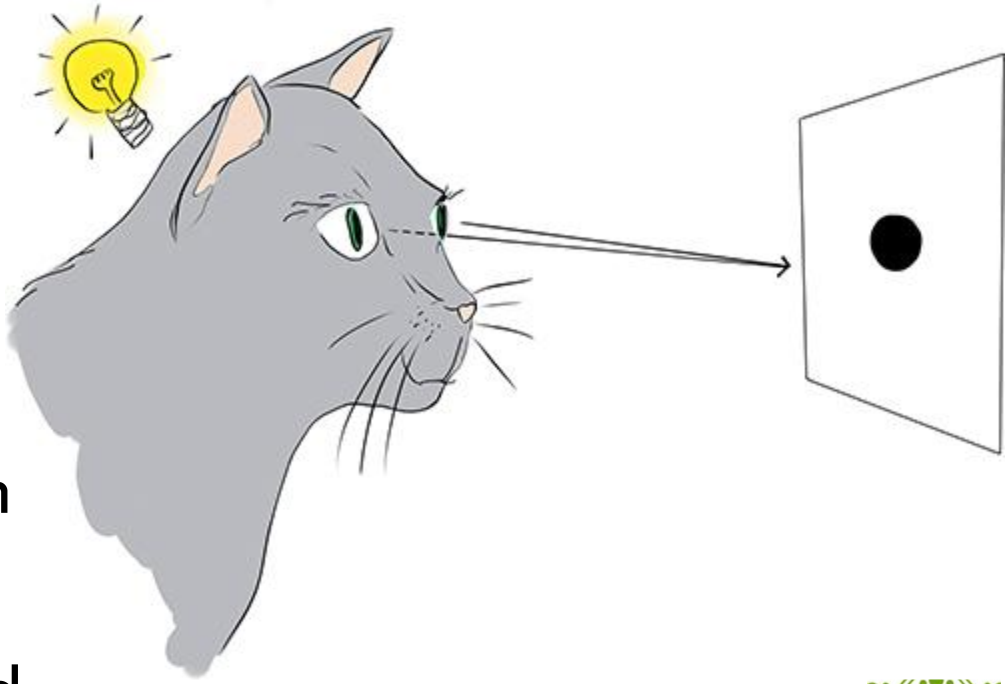
- In modern mammals, the brain's cerebral cortex (the outer gray matter of the brain) is involved in much visual perception.
- At Johns Hopkins University in the late 1950s, the physiologists David Hubel and Torsten Wiesel (left) worked on how visual information is processed in the mammalian cerebral cortex.
- This work later was awarded a Nobel Prize.



Source: J. Krohn, G. Beyleveld and A. Bassens, "Deep Learning Illustrated", Pearson

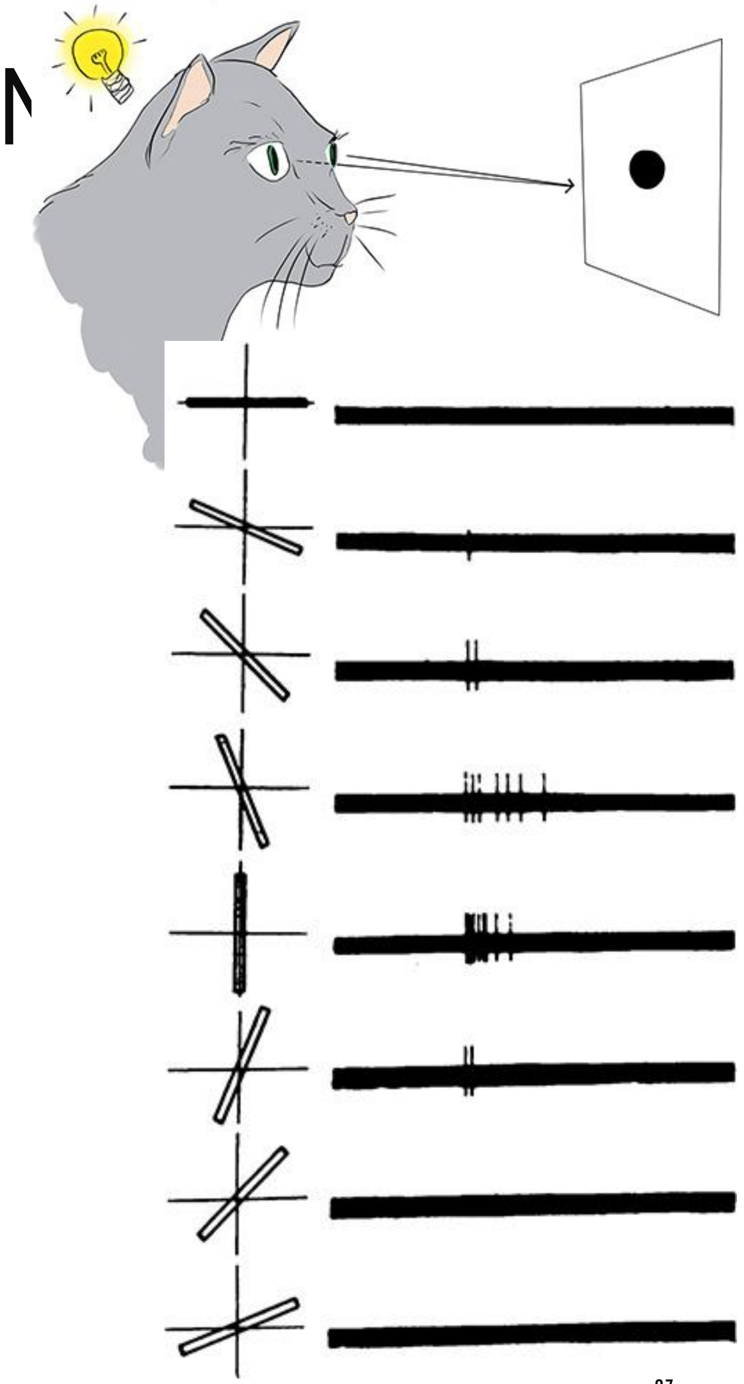
VISUAL INFORMATION PROCESSING [2]

- Images were shown to anesthetized cats.
- Simultaneously, the activities of individual neurons from the ***primary visual cortex*** *was recorded.*
- The primary visual cortex is the first part of the cerebral cortex to receive visual input from the eyes.
- Electrical recording equipment was implanted within the cat's skull
- Disheartening initial results – no response to visual stimuli
  - Jumping and waving hands before the cat elicited no response



Source: J. Krohn, G. Beyleveld and A. Bassens, "Deep Learning Illustrated", Pearson

VISUAL INFORMATION PROCESSIN

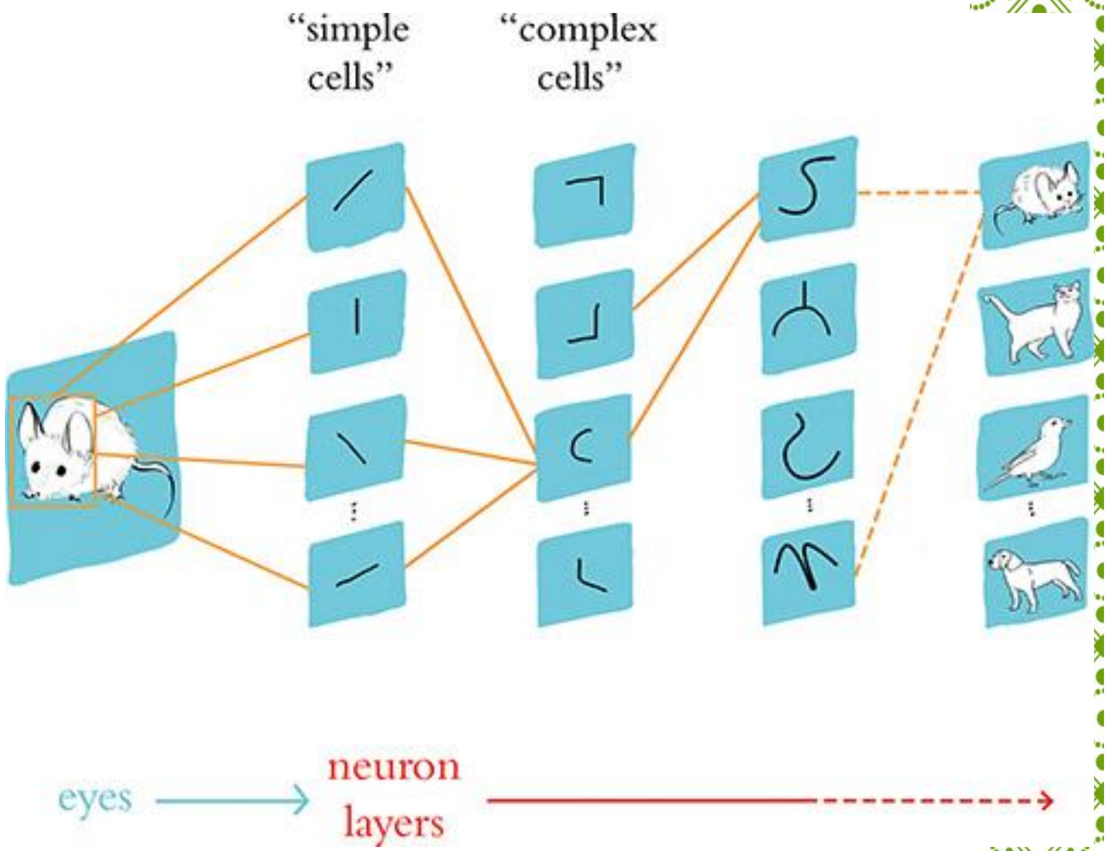


- As they removed one of their slides from the projector, its straight edge elicited a response in their recording equipment
- Turns out that the neurons that receive visual input from the eye are most responsive to simple straight edges
- Simple Neurons
  - Responds optimally to an edge at a particular specific orientation
  - A group of neurons together are able to represent the complete 360 degrees of orientation

Source: J. Krohn, G. Beyleveld and A. Bassens, "Deep Learning Illustrated", Pearson

VISUAL INFORMATION PROCESSING [4]

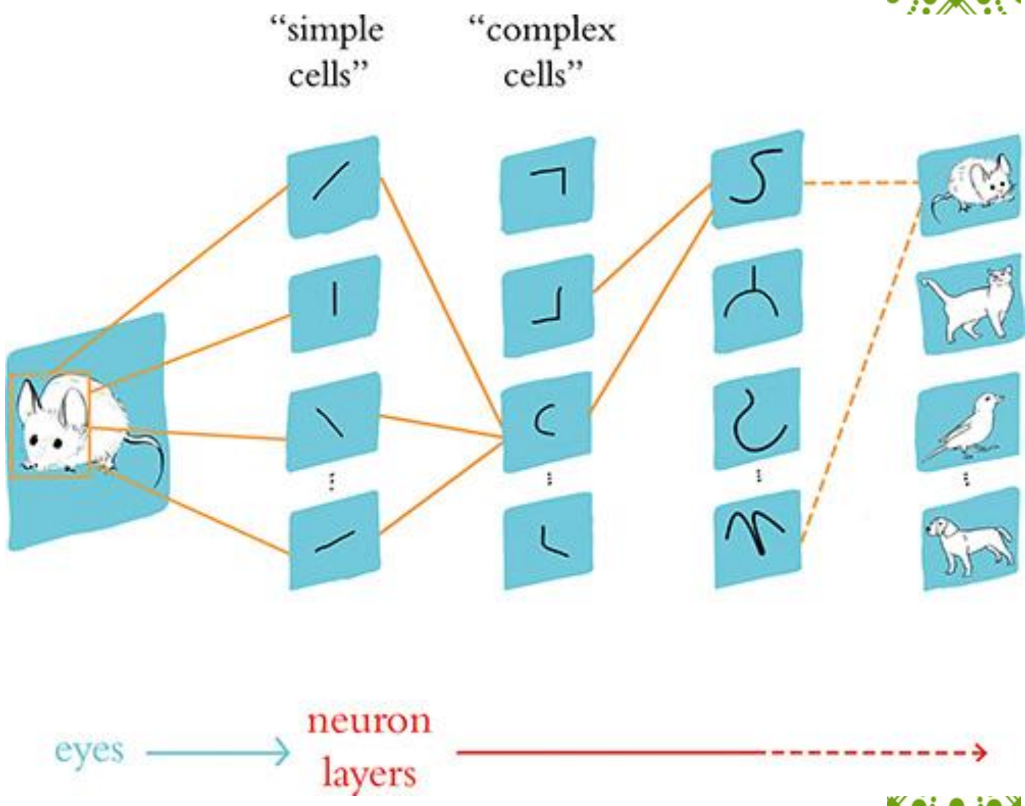
- Edge detecting simple cells pass on the information to a large number of so-called "complex neurons."
- A complex neuron recombines multiple line orientations into more complex shape like a corner or a curve
- Hierarchically organized layers of neurons feeding information into increasingly higher-order neurons
- Complex stimuli represented by the brain



Source: J. Krohn, G. Beyleveld and A. Bassens, "Deep Learning Illustrated", Pearson

VISUAL INFORMATION PROCESSING [5]

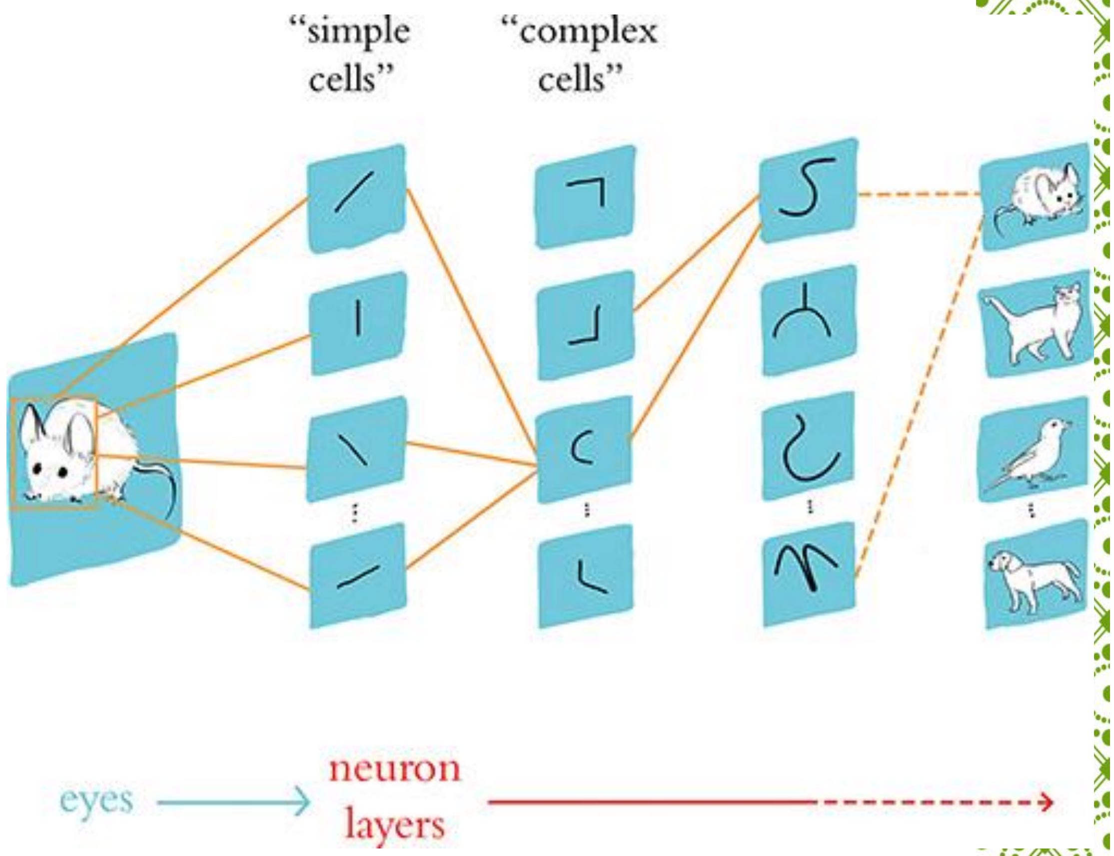
- Simple neurons relay information about the ***presence or absence of lines*** at particular orientations to a subsequent layer of *complex cells.*
- The complex cells **assimilate and recombine the information**, enabling the representation of **more complex visual stimuli** such as the curvature of the mouse's head.
- As information is passed through several subsequent layers, representations of visual stimuli can **incrementally become more complex and more abstract.**
- The far-right layer of neurons, following many hierarchical layers, the brain is ultimately able to represent visual concepts as abstract as a mouse, a cat, a bird, or a dog.



Source: J. Krohn, G. Beyleveld and A. Bassens, "Deep Learning Illustrated", Pearson

VISUAL INFORMATION PROCESSING [6]

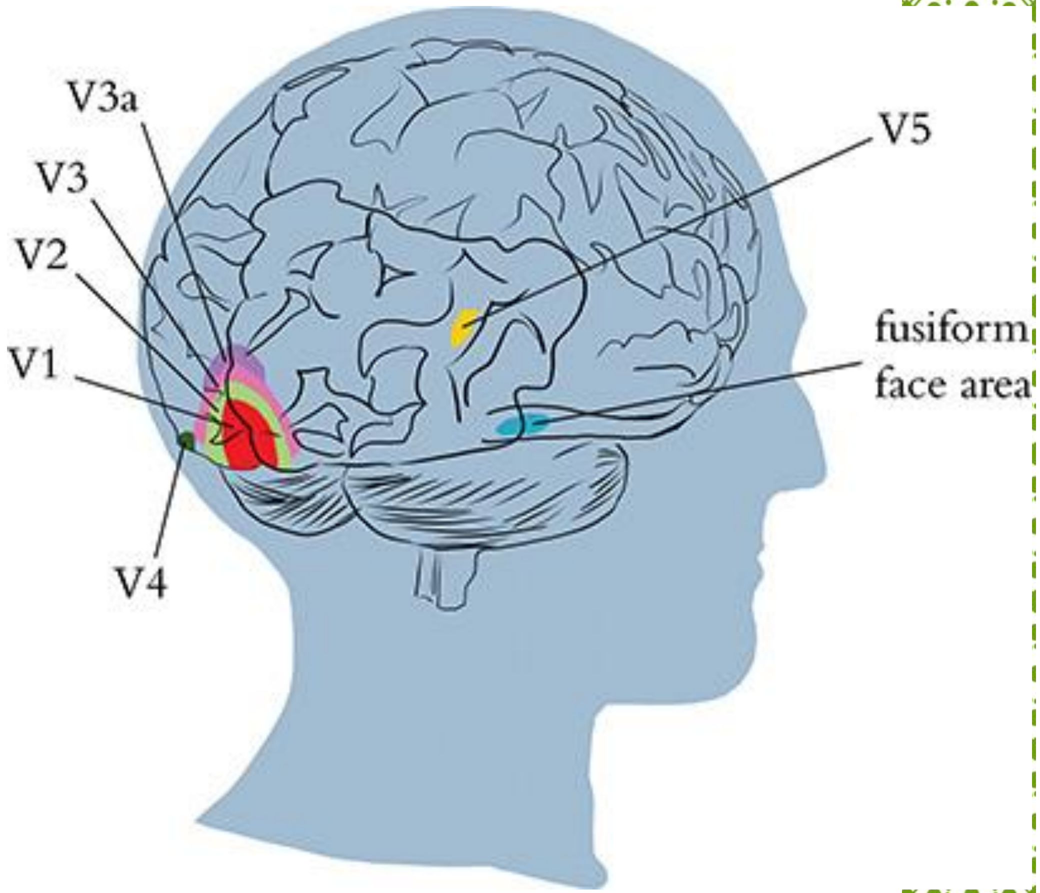
❖Kunihiko Fukushima (late 1970s) proposed an architecture similar to the human brain for machine vision - **Neocognitron**



Source: J. Krohn, G. Beyleveld and A. Bassens, “Deep Learning Illustrated”, Pearson

VISUAL INFORMATION PROCESSING [7]

- The V1 region receives input from the eyes and contains the simple cells that detect edge orientations.
- Through the recombination of information via myriad subsequent layers of neurons (including within the V2, V3, and V3a, regions), increasingly abstract visual stimuli are represented.
- In the human brain there are regions containing neurons with concentrations of specializations.
  - Colour (V4)
  - Motion (V5)
  - People's faces (fusiform face area)



Source: J. Krohn, G. Beyleveld and A. Bassens, "Deep Learning Illustrated", Pearson

PERCEPTRONS

- 1960s – Rosenblatt introduced a simple, efficient learning procedure that could figure out how to weight features of the input in order to classify inputs correctly
  - But perceptrons could not learn the features
- 1969 – Minsky and Papert showed that perceptrons had some very strong limitations on what they could do.
  - Minsky and Papert also implied that having deeper networks would not help.
- 1970s – The first neural net winter

Source: G. Hinton, "The Deep Learning Revolution", Turing Award Lecture

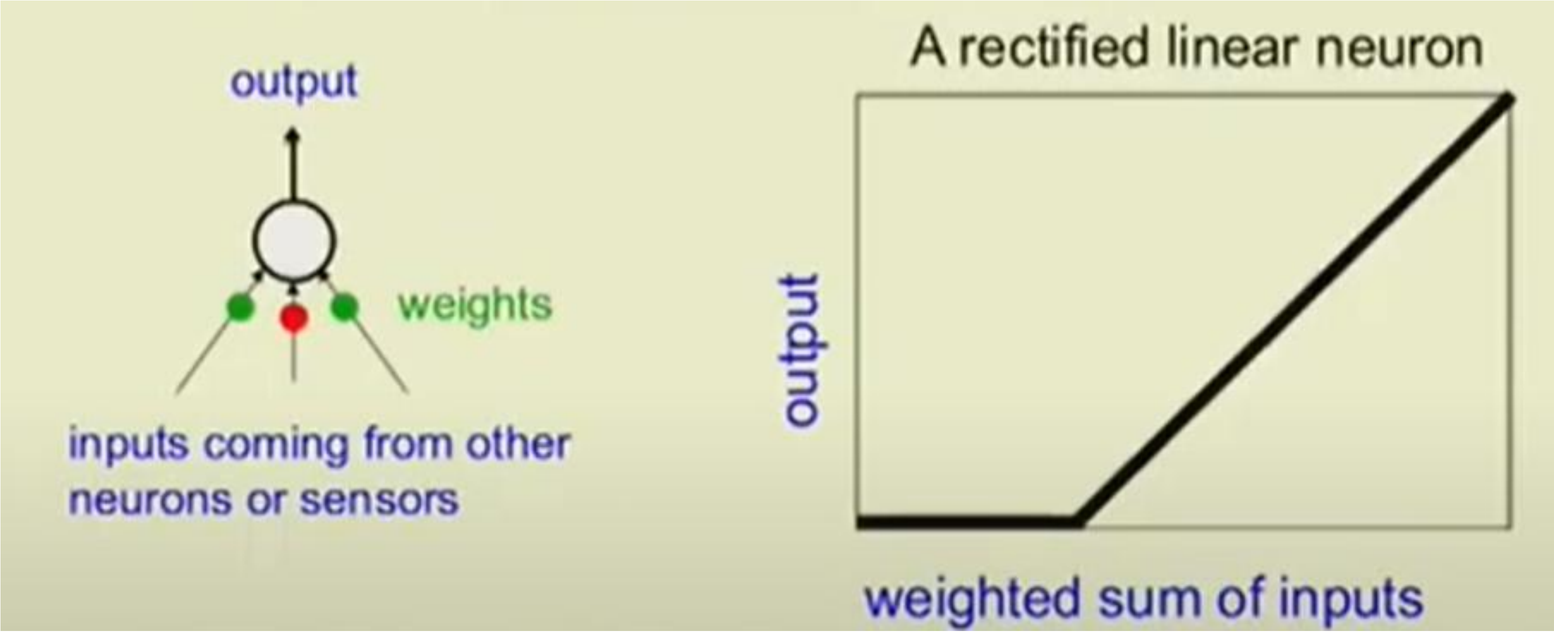
BACKPROPAGATION

- 1980s – The back-propagation algorithm allows neural networks to design their own features and to have multiple layers of features
  - Backpropagation created a lot of excitement
  - It could learn ***vector embeddings*** that captured the meanings of words just by trying to ***predict the next word*** in a string
  - Looked as if it would solve tough problems like speech recognition and shape recognition
  - Did very well for some forms of shape recognition

Source: G. Hinton, "The Deep Learning Revolution", Turing Award Lecture

WHAT IS AN ARTIFICIAL NEURON?

- Gross idealization of a real neuron
- Enables us investigate how neurons can collaborate to do computations that are too difficult to program such as:
  - Convert the pixel intensity values of an image into a string of words that describe the image



Source: G. Hinton, "The Deep Learning Revolution", Turing Award Lecture

A close-up of a child holding a stuffed animal.

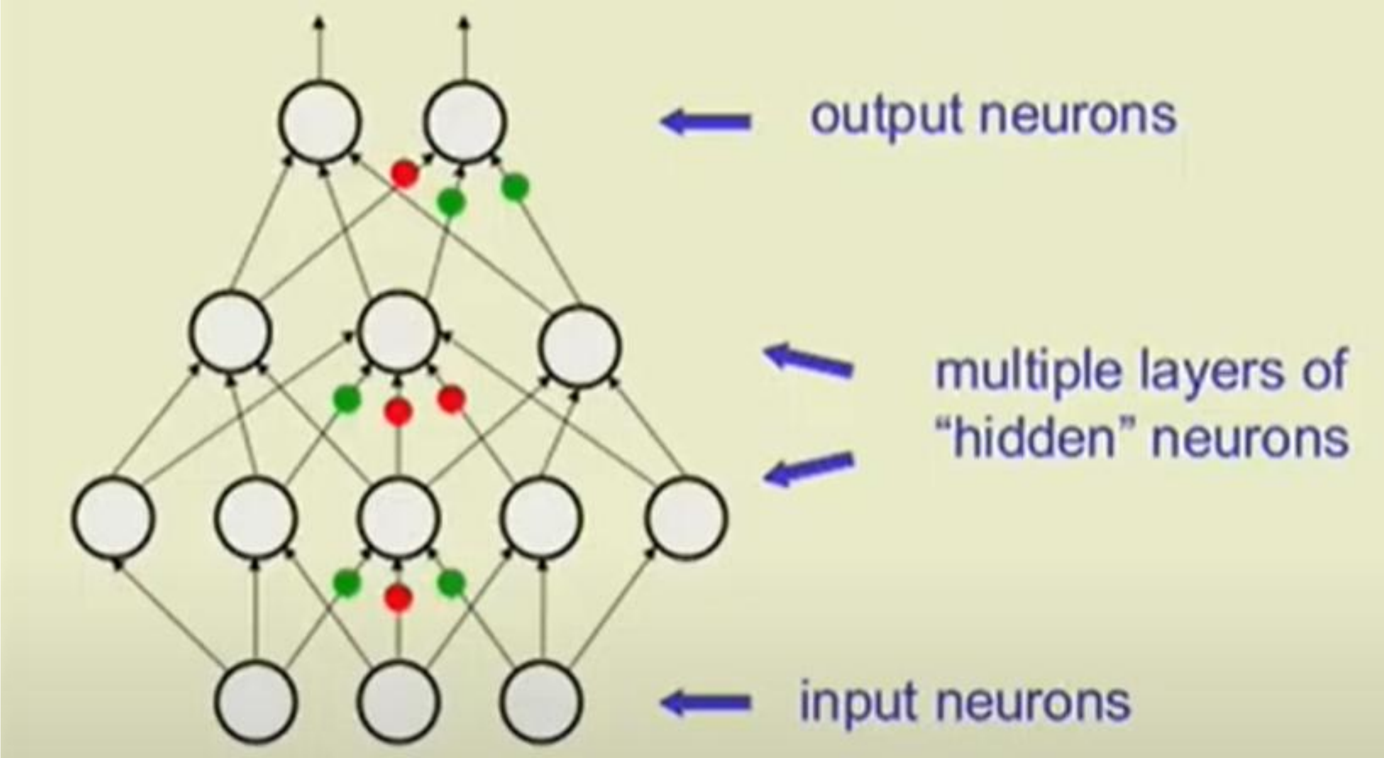
Input is an image

Output is a caption

Source: G. Hinton, “The Deep Learning Revolution”, Turing Award Lecture

WHAT IS AN ARTIFICIAL NEURAL NETWORK?

- If we connect the neurons in layers with no cycles, we get a feed forward neural network
- Change the Weights → change the features that the NN would learn
- Learning weights → Learning Features



Source: G. Hinton, "The Deep Learning Revolution", Turing Award Lecture

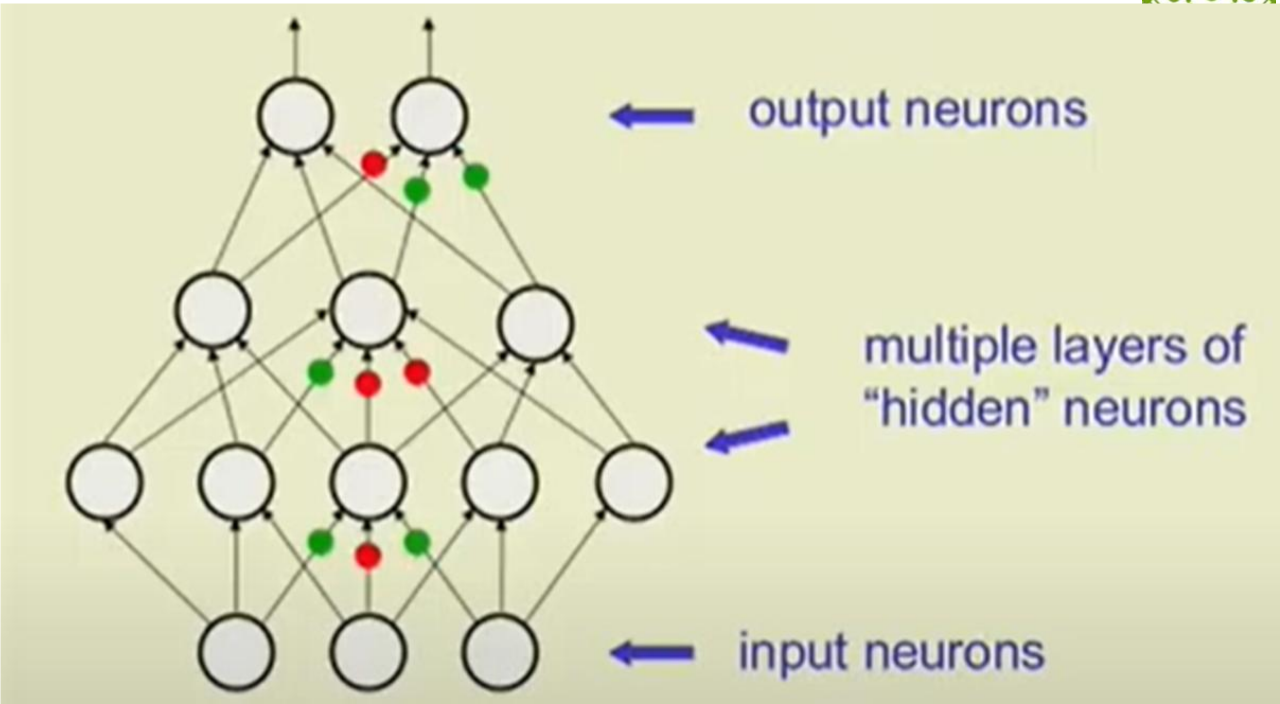
HOW DO WE TRAIN ARTIFICIAL NEURAL NETWORKS?

Supervised Learning: Show the network an input vector and tell it the correct output

- Adjust the weights of the inputs to reduce the discrepancy between the correct output and actual output

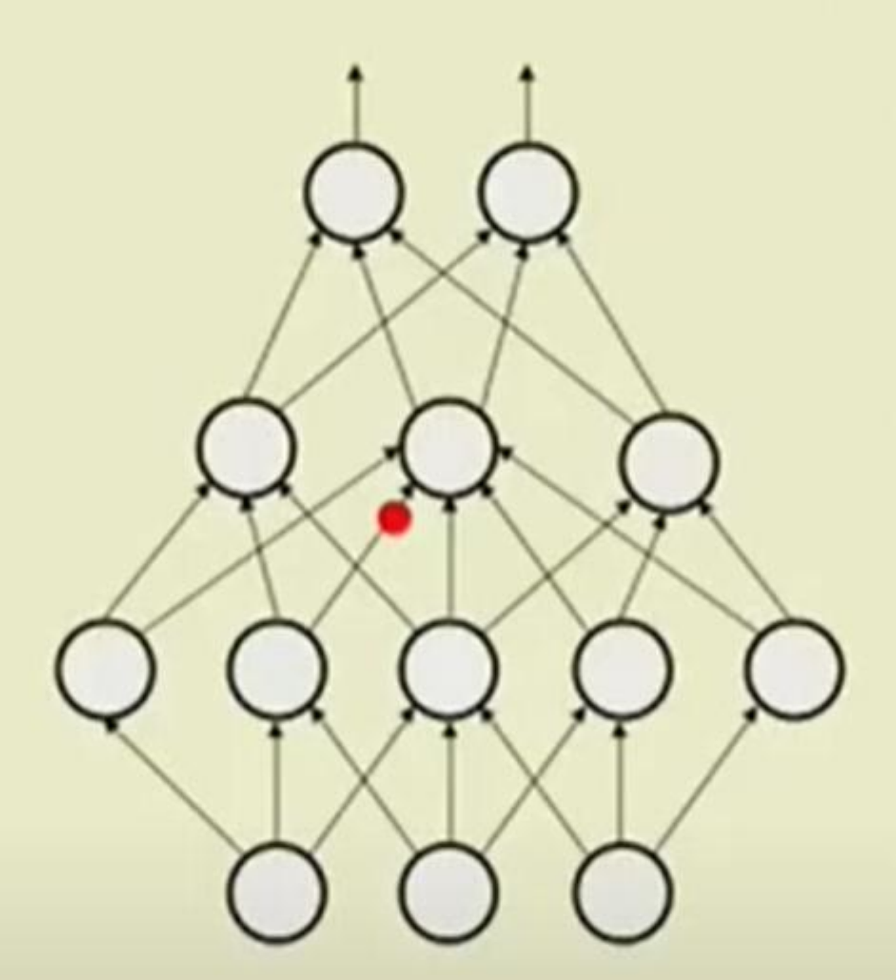
Unsupervised Learning: Show the network only the input

- Adjust the weights to get better at reconstructing the input (or parts of the input) from the activities of the hidden neurons



Source: G. Hinton, "The Deep Learning Revolution", Turing Award Lecture

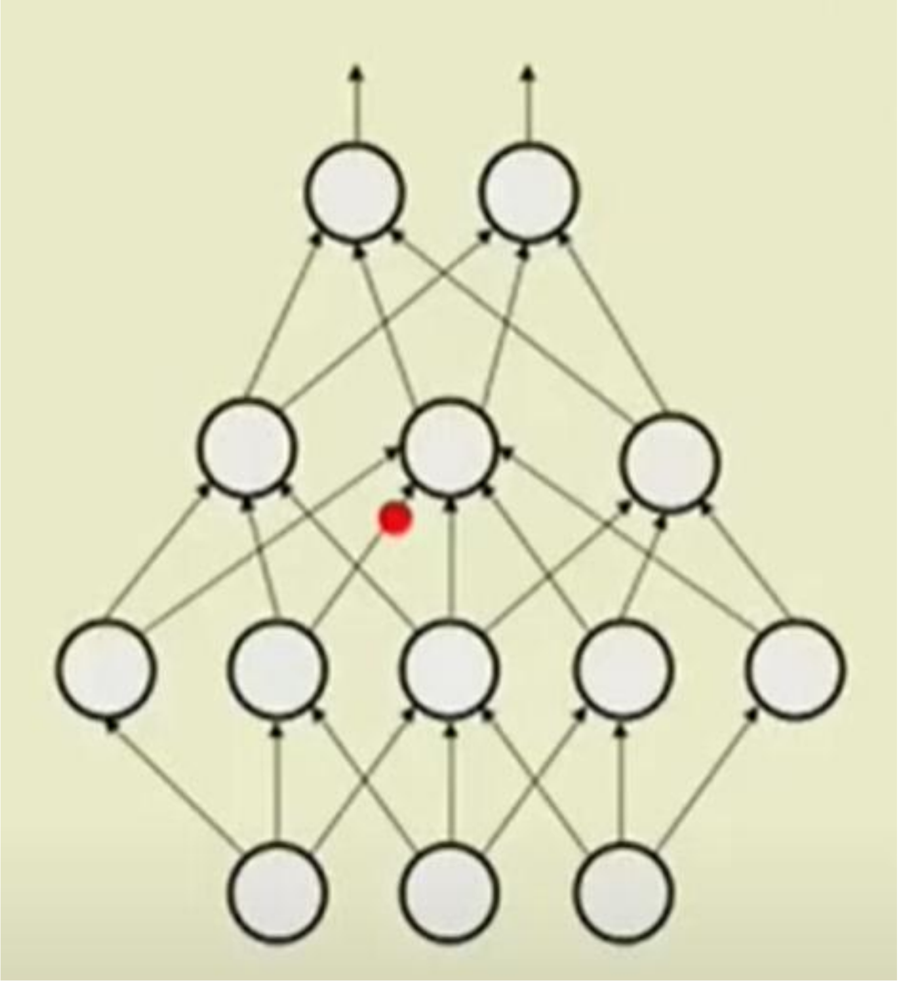
SUPERVISED TRAINING: AN EASY TO UNDERSTAND BUT INEFFICIENT MUTATION METHOD



- Take a small random sample of the training cases
- Measure how well the network does on this sample
- Pick one of the weights
- Increase or decrease the weights slightly and measure how well the network does now
- If the change helped, keep it

Source: G. Hinton, "The Deep Learning Revolution", Turing Award Lecture

SUPERVISED TRAINING: AN EASY TO UNDERSTAND BUT INEFFICIENT MUTATION METHOD



- Achieves the result, but…
- Takes a long time
- Very inefficient

Source: G. Hinton, "The Deep Learning Revolution", Turing Award Lecture

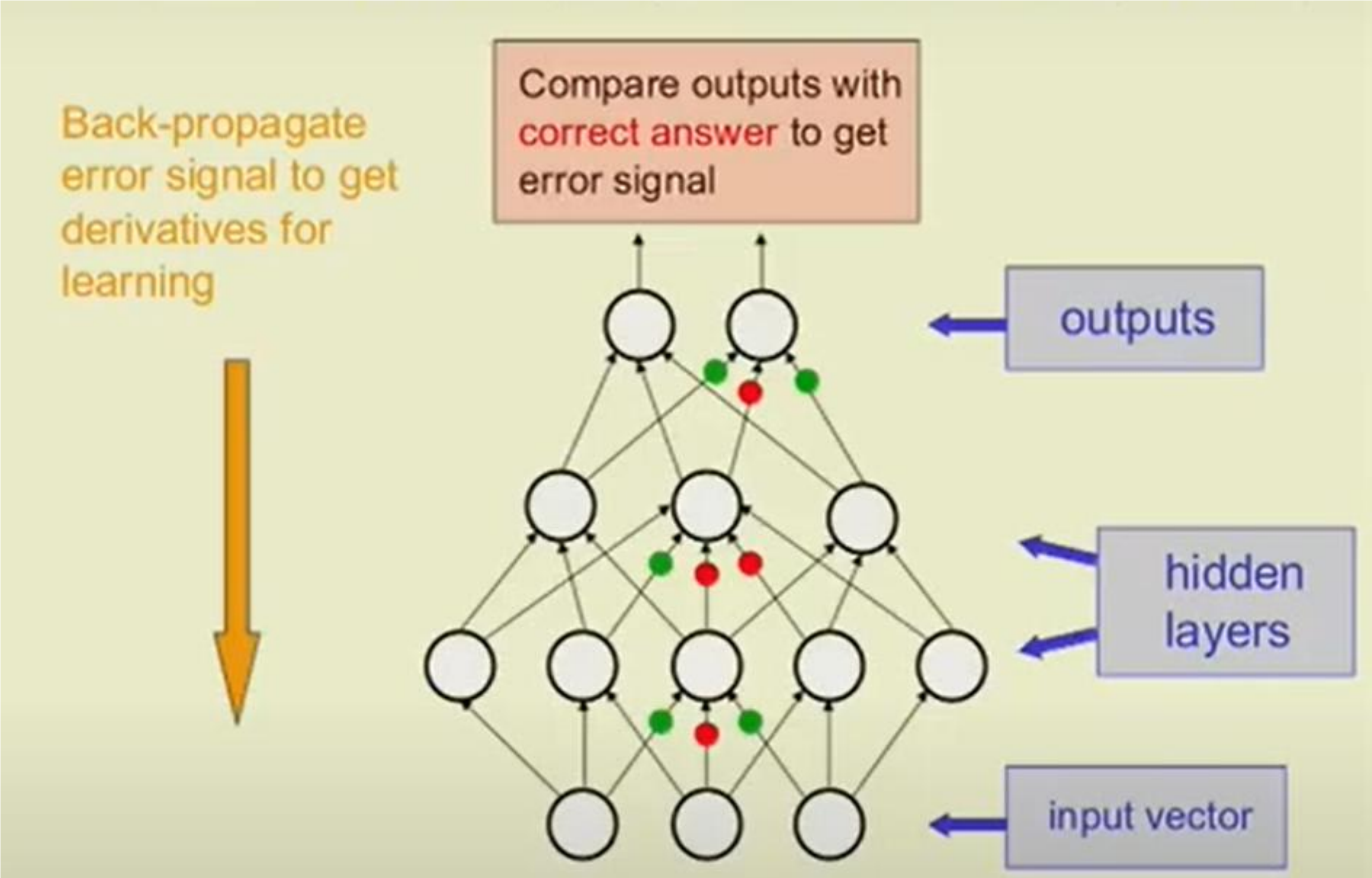
THE BACKPROPAGATION ALGORITHM

- Backpropagation is just an efficient way of computing how the change in a weight will affect the output error
- Instead of perturbing the weights one at a time and measuring the effect
  - Use calculus to **compute the error gradients for all the weights at the same time**
- This can be done because the effect of changing a weight on the output is known
- With a million weights, this is more efficient than the mutation method by a factor of a million

Source: G. Hinton, "The Deep Learning Revolution", Turing Award Lecture

HOW TO LEARN MANY LAYERS OF FEATURES (~1985)

- Stochastic gradient descent works very well at scale i.e. it needs a lot of inputs to work.



Source: G. Hinton, "The Deep Learning Revolution", Turing Award Lecture

A BIG DISAPPOINTMENT

- 1990s: Backpropagation works pretty well, but underperforms the expectations of its proponents.
  - It is hard to train deep neural networks (But why?)
- On modest-sized datasets, some other machine learning methods work better than backpropagation
  - The second neural network winter (in the machine learning community) begins
- Symbolic AI researchers claim that it is unrealistic to expect to learn difficult tasks in big deep neural networks that start with random connections and no prior knowledge.

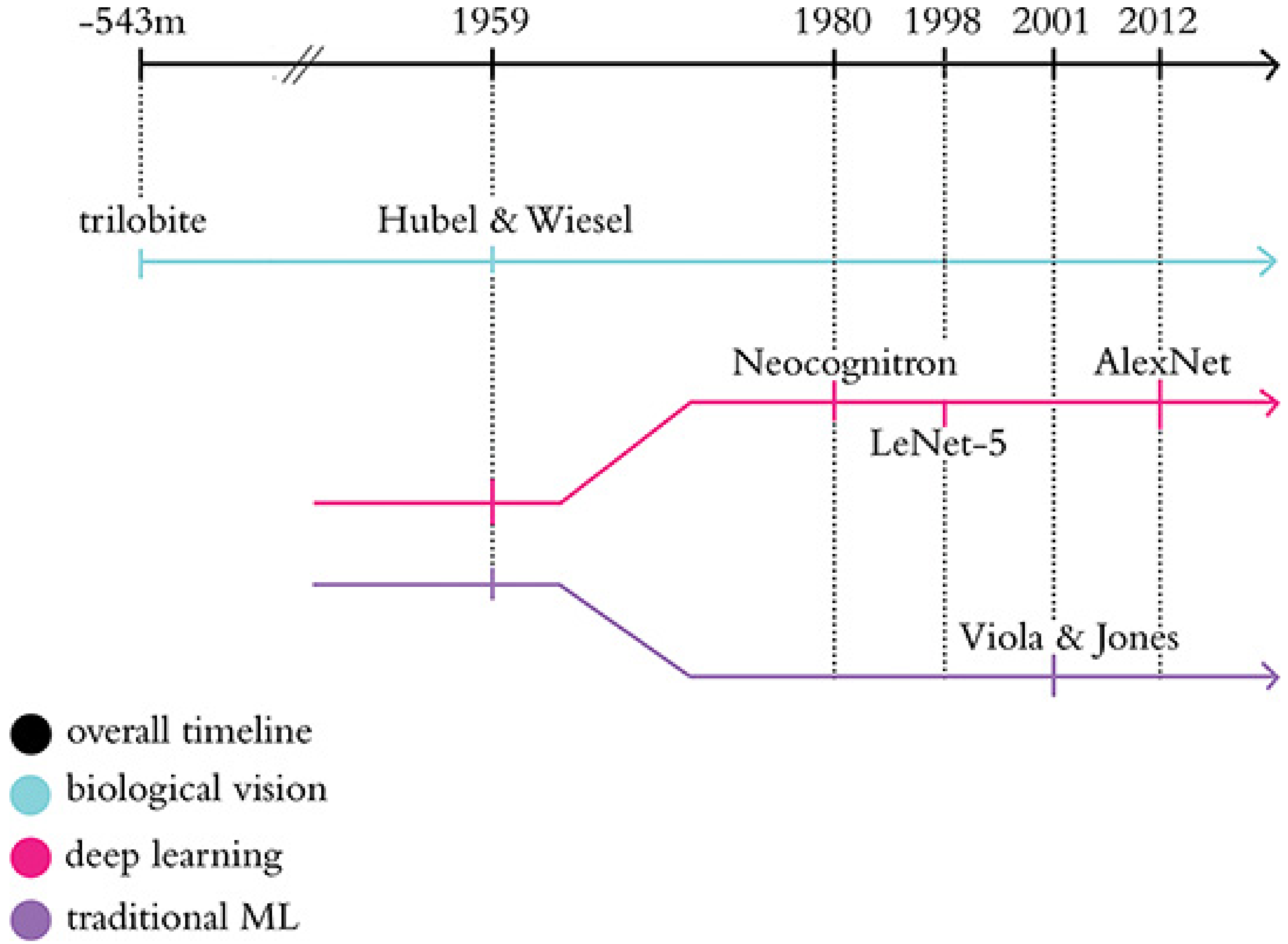
WHAT WAS ACTUALLY WRONG WITH BACKPROPAGATION IN 1986?

- Wrong conclusions were drawn about why it failed.

The real reasons (according to Geoff Hinton) were:

- Labeled datasets were thousands of times too small
- Computers were millions of times too slow
- The weights were initialized in an inappropriate or wrong fashion
- Wrong type of non-linearity was used

ABRIDGED TIMELINE OF BIOLOGICAL AND MACHINE VISION



Source: J. Krohn, G. Beyleveld and A. Bassens, “Deep Learning Illustrated”, Pearson

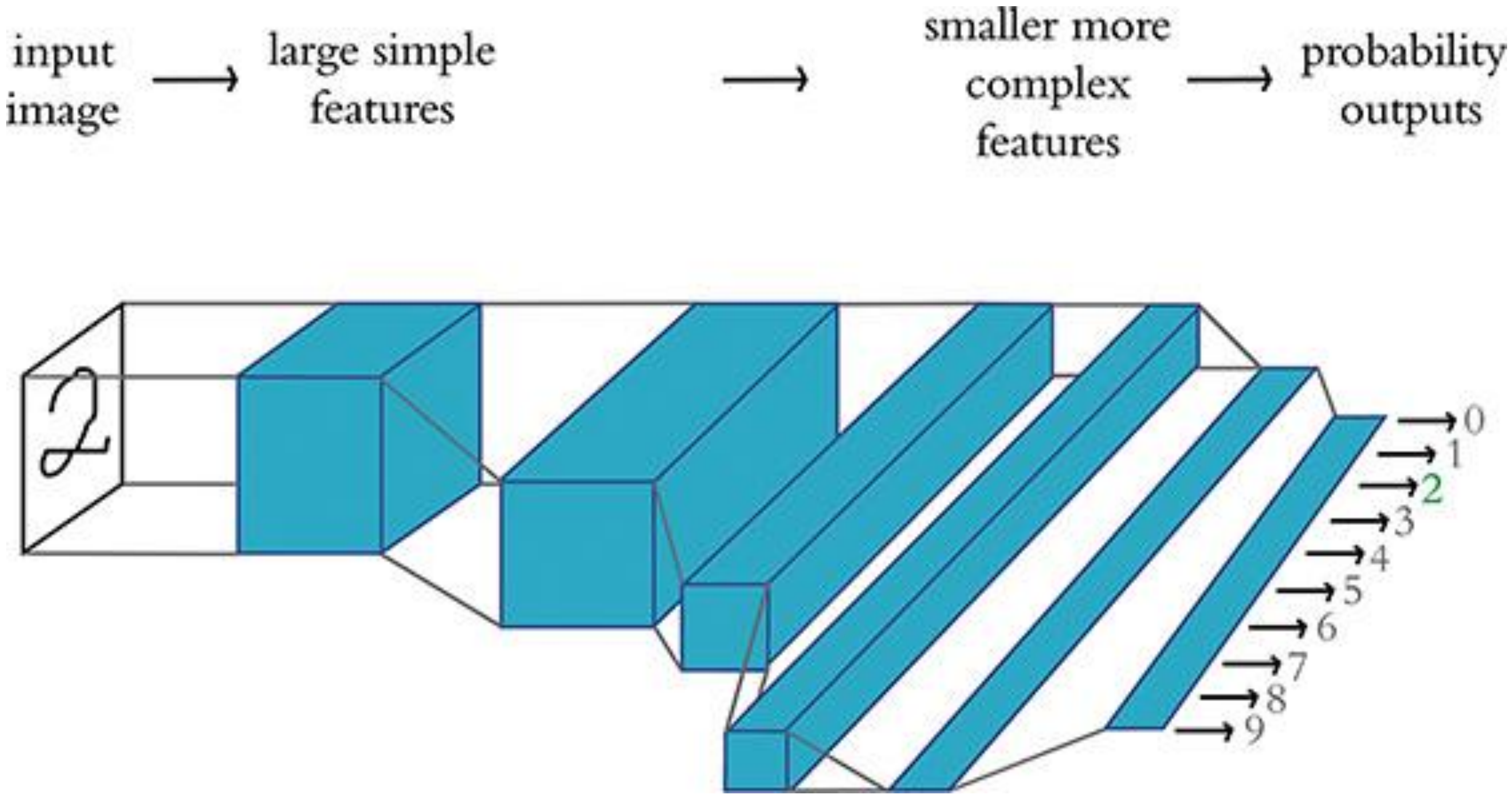
LENET-5

- LeNet-5, Yann LeCun and Yoshua Bengio (1989)
  - Gradient-based learning applied to document recognition
- LeNet-5's hierarchical architecture was built on Fukushima's lead and the biological inspiration uncovered by Hubel and Wiesel
- Foundations
  - Superior data for training their models
  - Faster processing power
  - Backpropagation algorithm
- Sufficiently reliable for an early commercial application of Deep Learning
  - Automation of reading of ZIP codes by US Postal Service
- LeNet-5 was correctly able to predict the handwritten digits without the need to include any expertise about handwritten digits in their code

Source: J. Krohn, G. Beyleveld and A. Bassens, "Deep Learning Illustrated", Pearson

LENET-5 ARCHITECTURE

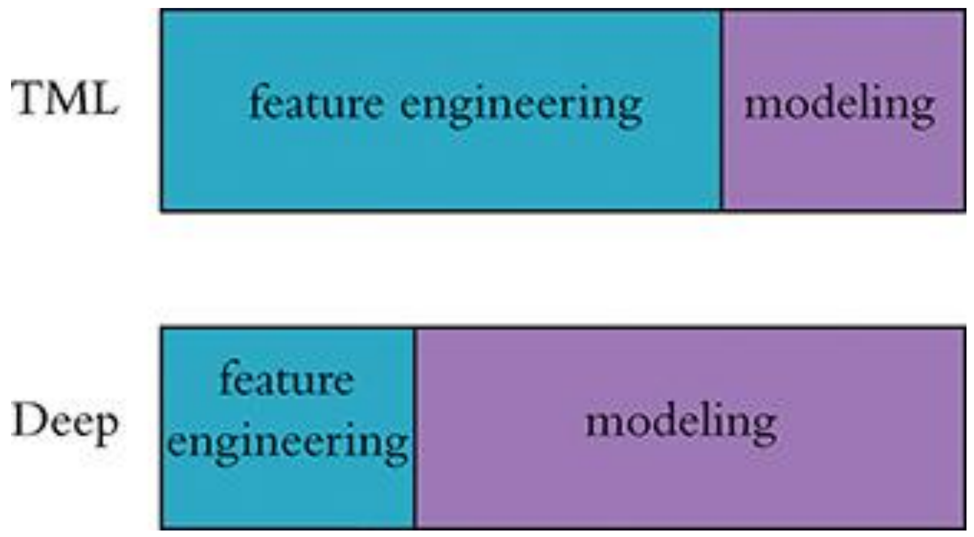
- The leftmost layer represents simple edges
- Successive layers represent increasingly complex features
- A handwritten "2" can be correctly recognized as the number two



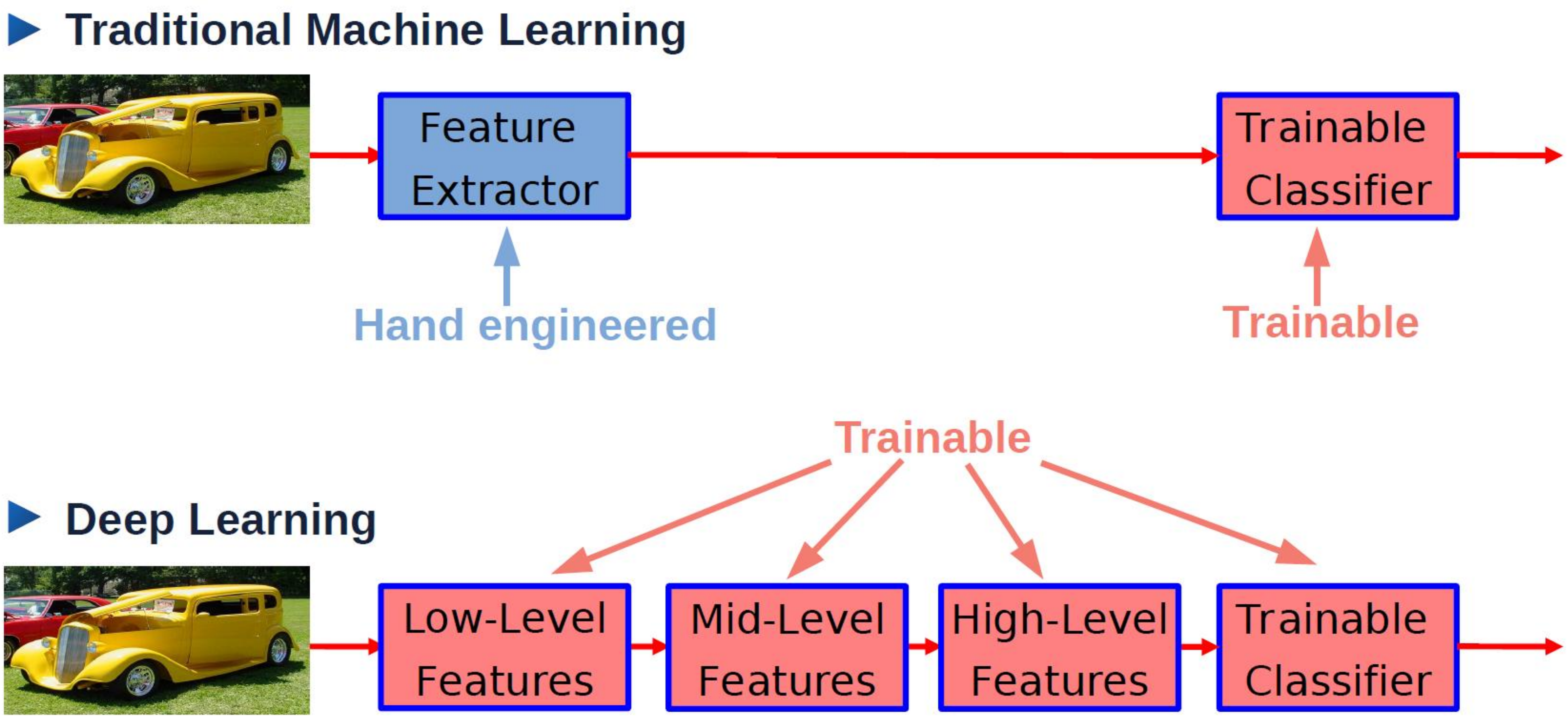
Source: J. Krohn, G. Beyleveld and A. Bassens, "Deep Learning Illustrated", Pearson

TRADITIONAL ML VS DEEP LEARNING

- **Traditional ML algorithms:** Feature engineering—the transformation of raw data into thoughtfully transformed input variables
- **Deep Learning:** Often involves little to no feature engineering
  - majority of time spent instead on the design and tuning of model architectures.



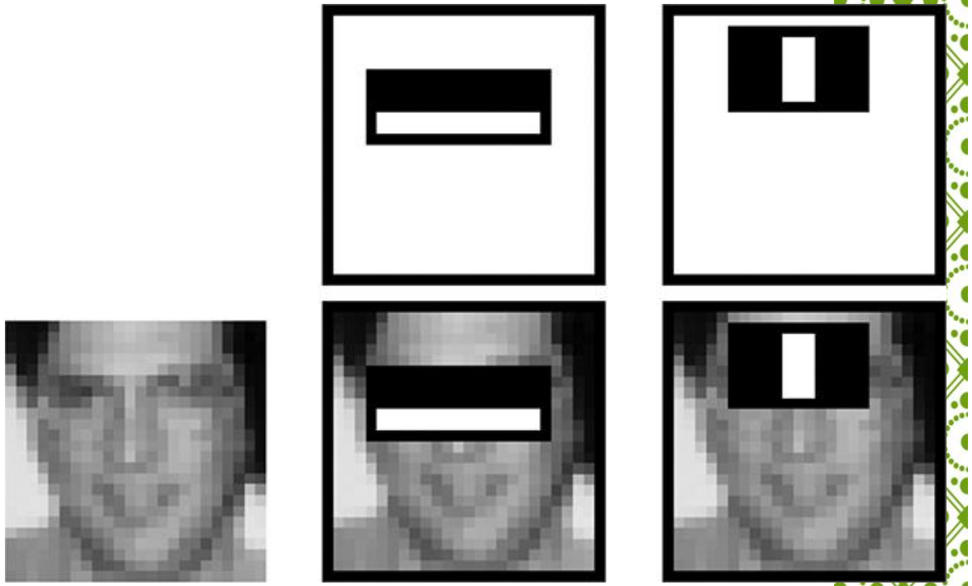
Source: J. Krohn, G. Beyleveld and A. Bassens, "Deep Learning Illustrated", Pearson



Source: Alfredo Canziani, Yann LeCun, Deep Learning Course Slides

FEATURE ENGINEERING – TRADITIONAL ML

- Paul Viola and Michael Jones (2001)
- Employed rectangular filters such as the vertical or horizontal black-and-white bars
- Features generated by passing these filters over an image can be fed into machine learning algorithms to reliably detect the presence of a face.
- Notable work - algorithm was efficient enough to be the first real-time face detector outside the realm of biology –
  - Fujifilm cameras – real-time auto-focus
- Required domain knowledge
  - Devising clever face-detecting filters to process raw pixels into features for input into a machine learning model was accomplished via years of research and collaboration on the characteristics of faces.



Source: J. Krohn, G. Beyleveld and A. Bassens, "Deep Learning Illustrated", Pearson

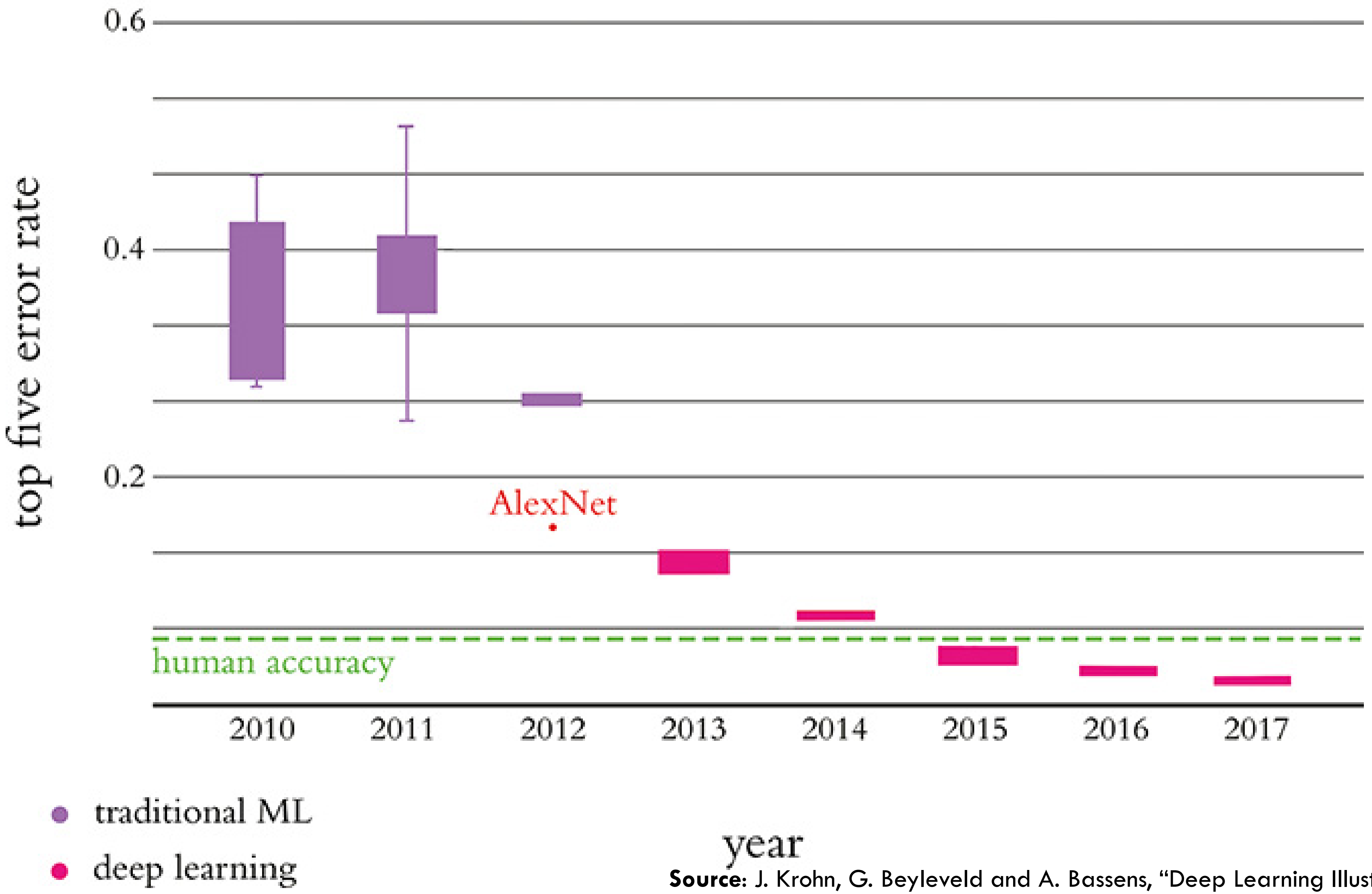
IMAGENET AND ILSVRC

- ImageNet - a labeled index of photographs devised by Fei-Fei Li armed machine vision researchers with an immense catalog of training data
  - LeNet-5 was trained on tens of thousands of images
  - ImageNet contains tens of millions
    - 14 million images across 22,000 categories
- ILSVRC (the ImageNet Large Scale Visual Recognition Challenge) on a subset of the ImageNet data that has become the premier ground for assessing the world's state-of-the-art machine vision algorithms
  - Open Challenge run by Li since 2010
  - Contains 1.4 million images across 1,000 categories
    - Dogs and their Breeds

The **ImageNet** dataset was the brainchild of Chinese-American computer science professor Fei-Fei Li and her colleagues at Princeton in 2009. Now a faculty member at Stanford University, Li is also the chief scientist of A.I./ML for Google's cloud platform.

Source: J. Krohn, G. Beyleveld and A. Bassens, "Deep Learning Illustrated", Pearson

ILSVRC Results

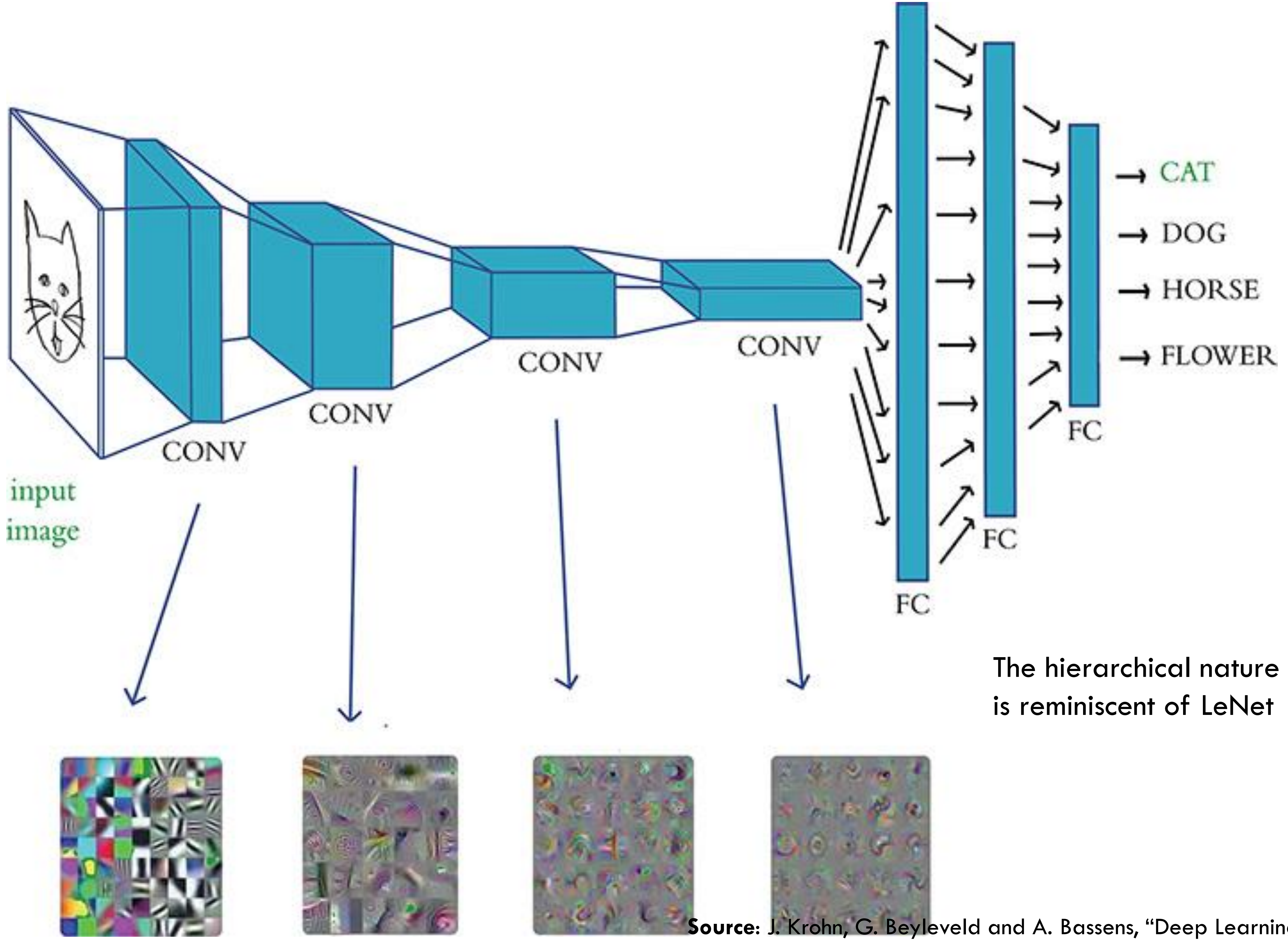


Source: J. Krohn, G. Beyleveld and A. Bassens, "Deep Learning Illustrated", Pearson

ALEXNET

- Devised by Alex Krizhevsky, Ilya Sutskever and Geoffrey Hinton —working out of the University of Toronto lab performed much better that the then existing benchmarks with their deep learning-based submission – AlexNet
- Watershed moment - deep learning architectures emerged from the fringes of machine learning to its forefront
- Opensource libraries, etc.

Source: J. Krohn, G. Beyleveld and A. Bassens, "Deep Learning Illustrated", Pearson

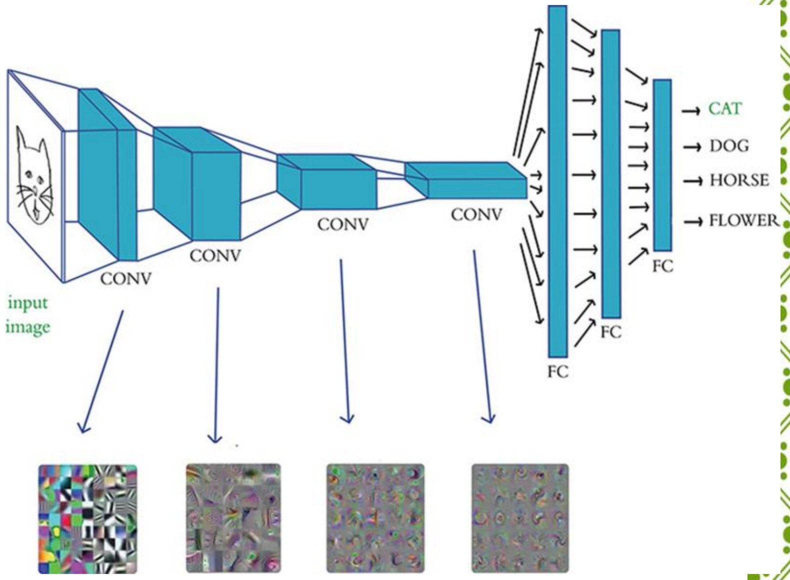


The hierarchical nature of AlexNet is reminiscent of LeNet

Source: J. Krohn, G. Beyleveld and A. Bassens, “Deep Learning Illustrated”, Pearson

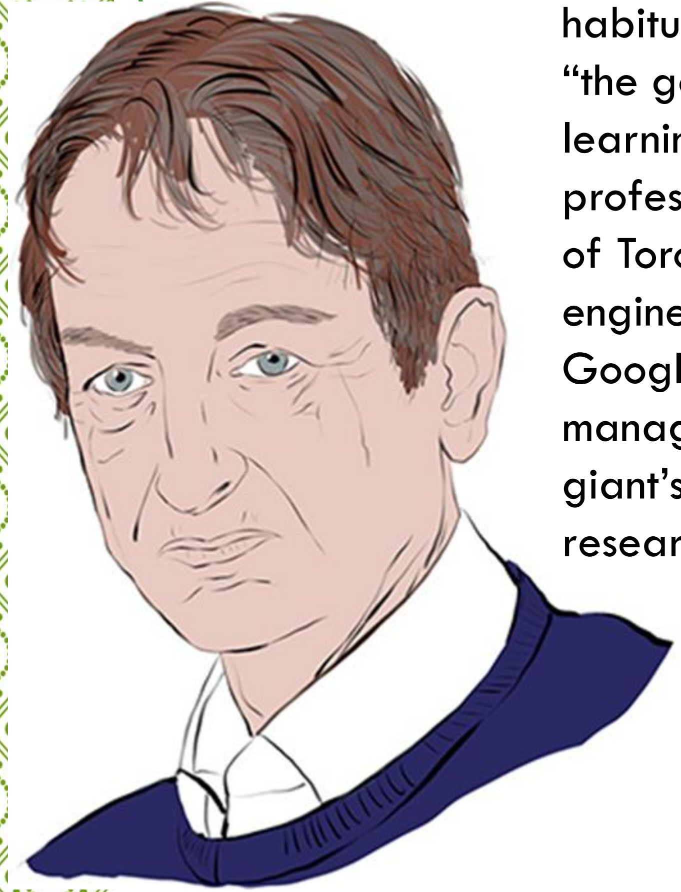
SUCCESS OF ALEXNET

- Training Data
  - Access to the massive ImageNet index
  - Artificially expanded the data available to them by applying transformations to the training images
- Processing power
  - Computing power per unit of cost increased dramatically from 1998 to 2012
  - Programmed two GPUs to train their large datasets efficiently
- Architectural advances
  - AlexNet is deeper (has more layers) than LeNet-5
  - Advantage of both a new type of artificial neuron (CNNs) and a nifty trick that helps generalize deep learning models beyond the data they're trained on.

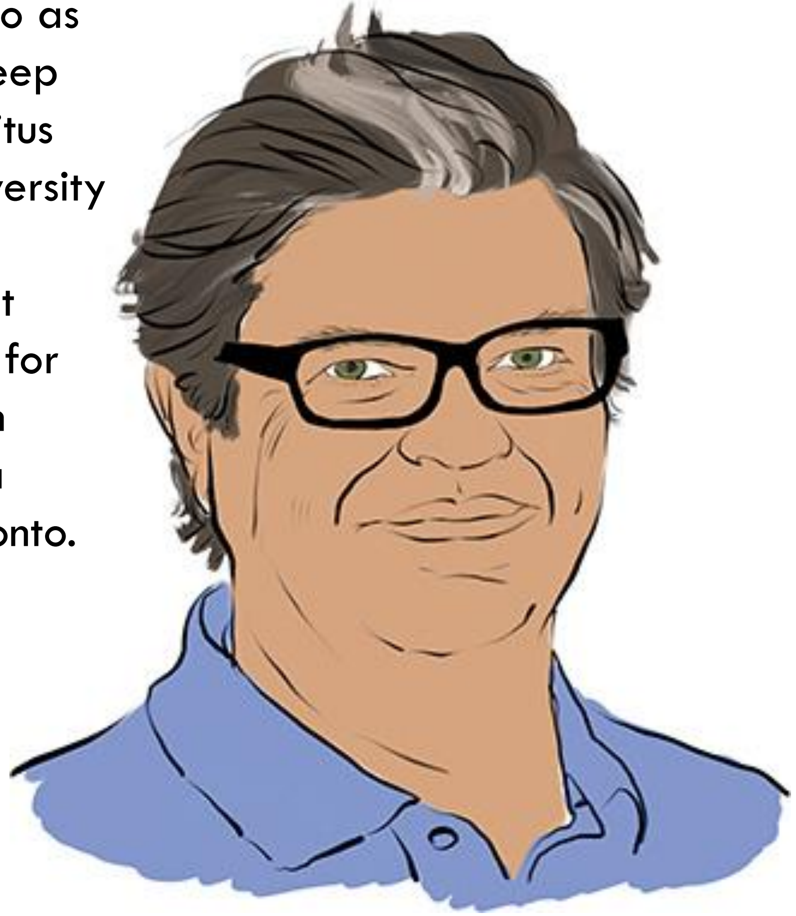


Source: J. Krohn, G. Beyleveld and A. Bassens, "Deep Learning Illustrated", Pearson

IN 2019, HINTON, YANN LECUN AND YOSHUA BENGIO WERE JOINTLY RECOGNIZED WITH THE TURING AWARD—THE HIGHEST HONOR IN COMPUTER SCIENCE—FOR THEIR WORK ON DEEP LEARNING.



Geoffrey Hinton, habitually referred to as "the godfather of deep learning" is an emeritus professor at the University of Toronto and an engineering fellow at Google, responsible for managing the search giant's Brain Team, a research arm, in Toronto.



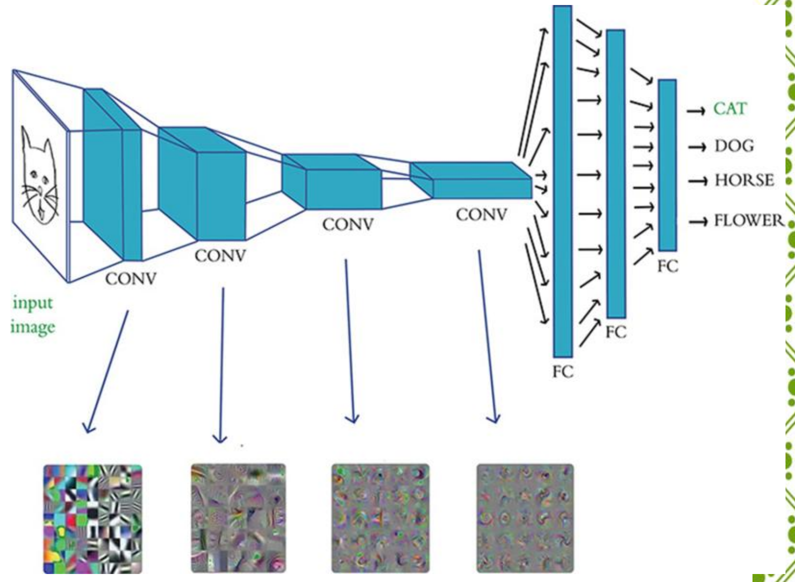
Yoshua Bengio: Computer science professor at the University of Montreal and codirects the renowned Machines and Brains program at the Canadian Institute for Advanced Research

Yann LeCun is the founding director of the New York University Center for Data Science as well as the director of AI research at the social network Facebook.

Source: J. Krohn, G. Beyleveld and A. Bassens "Deep Learning Illustrated", Pearson

DISRUPTIVE NATURE OF DEEP LEARNING MODELS

- Dramatically reduce the subject-matter expertise required to build highly accurate predictive models
  - Across industries – insurance, medicine
- Automatic feature generation in Deep Learning
- For a rapidly growing list of use cases, one's ability to apply deep learning techniques outweighs the value of domain-specific proficiency.
- https://playground.tensorflow.org/



Source: J. Krohn, G. Beyleveld and A. Bassens, "Deep Learning Illustrated", Pearson

THANKS