STATS / DATA SCI 315

Course logistics and introduction to deep learning

Course logistics

What is the course about?

- **Deep learning (DL):** A branch of machine learning that uses multilayer neural networks to solve problems such as object recognition and playing board games (Chess, Go)
  - **Machine learning (ML):** A branch of artificial intelligence (AI) that seeks to endow machines with the ability to learn from experience
  - **Statistics (Stats)**: The science of learning from data. ML and Stats have been coming increasingly close over the past two decades
  - **Data science (DS)**: An emerging discipline that seeks to marry statistical thinking with computational thinking to solve difficult real world problems.
- **Tensorflow (TF)/Keras**: We will use popular software libraries that make it very easy to build deep learning models (which has its pros and cons)

Prerequisites

- Calculus: derivatives, gradients, chain rule
  - We will try to get by *without* multivariable calculus
- Linear algebra: vectors, matrices, norms
  - Covered in boot camp
- Prob/Stats: random variables, expectations, linear regression
- Programming: variables, data structures, loops, functions, classes, objects
- We will NOT assume prior exposure to:
  - Machine learning
  - Python
- The first few weeks can feel intense as we cover the necessary Stats/ML and Python material to get you started with deep learning using Tensorflow

Books

- *Deep Learning with Python* (2nd edition) by Chollet
  - Emphasizes the programming side
- *Dive into Deep Learning* by Zhang, Lipton, Li and Smola
  - Comprehensive, covers math background as well
- *Deep Learning* by Goodfellow, Bengio and Courville
  - Graduate student/researcher level, just for reference
- *Understanding Deep Learning* by Simon J.D. Prince
  - Intended as successor of Goodfellow et al.'s book, just for reference

You don't have to buy anything! Materials not on the web will be provided via Canvas

Course communication tools

- Website: https://www.ambujtewari.com/stats315-winter2024/
- Canvas: https://umich.instructure.com/courses/657959
- Slack: https://um-wn24-stats315.slack.com/

Instructional team

Faculty Instructor: Ambuj Tewari, tewaria@umich.edu

GSI: Sahana Rayan, srayan@umich.edu

GSI: Jacob (Jake) Trauger, jtrauger@umich.edu

GSI: Abhiti Mishra, abhiti@umich.edu

Grading will be on a curve

- Canvas quizzes (20%): Times, multiple choice, open book. Will drop two lowest scores
- Homeworks (30%): Assigned roughly every other week. Will drop one lowest score
- Midterm Exam (20%): In class, timed, multiple choice, open book
- Final Exam (30%): In class, timed, multiple choice, open book

- A-/B+ boundary will be close to the median overall weighted score

Academic Integrity

- You can discuss homeworks (but not quizzes/exams) with your classmates
- But all submitted work, including code, must be your own
- Misconduct will be reported to the Dean’s office
- Generative AI is:
  - Allowed in HWs but you have to describe how you used it
  - *Not allowed* in quizzes and exams
- When in doubt, ask!

Accommodation for Students with Disabilities

- Submit your VISA form as soon as possible
  - Electronic submission strongly encouraged
- Talk to me privately if you need any other accommodations

Mental Health and Well-Being

- Be aware of available resources
- Seeking help when needed is courageous!
- If this course is adding to your stress, talk to me privately

Rough Course Outline

- Intro to Python, Numpy, TF2, Keras, Jupyter, Colab
- Simple models that are precursors to DL: linear regression
- Fully connected, multilayer Neural Networks (NNs)
- Convolutional neural networks and vision
- Transformers and language
- (Experimental, not graded) Machine Olfaction
  - Fun lab activity
  - Guest Lecture (Apr 11) by Alex Wiltschko, UM alum and founder of osmo.ai

Reading assignments

- Course schedule is at: https://www.ambujtewari.com/stats315-winter2024/
  - Parts of it are still *under construction*
- Every lecture has associated reading assignments
- Material not available on the web will be under “Files” in Canvas
- **It is your responsibility to read the required materials**
- Recommendation: read it at least twice, preferably thrice. At least once before lecture and once afterwards

Introduction to deep learning

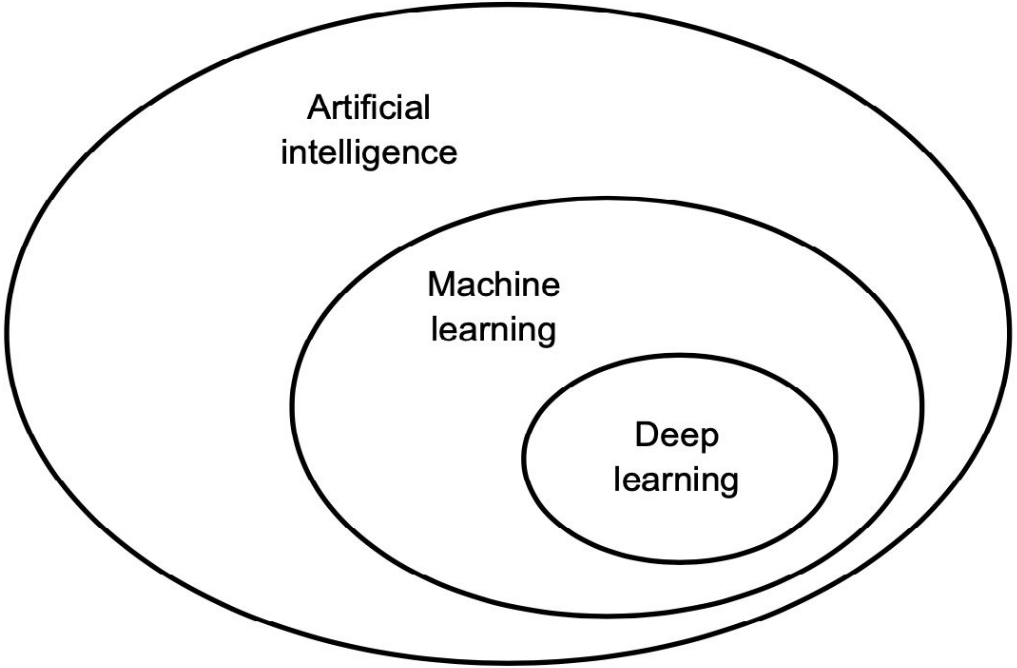


Figure 1.1 Artificial intelligence, machine learning, and deep learning

Beginnings of AI

Alan Turing's seminal papers

- *Intelligent Machinery*, report written for the Nat. Physical Lab, 1948 (published only in 1970)
- *Computing Machinery and Intelligence*, MIND, Vol. 59, 1950

Dartmouth summer workshop proposal, 1956

The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.

Symbolic AI

- Relied on hand-crafted rules for manipulating knowledge stored in databases
    - Cyc had over 24 million rules about over 1 million objects in its ontology in 2017
- Dominant approach in the 1950s to 1980s
- Reached its peak with the *expert systems* boom of the 1980s
- Tasks that are easy for people to do but hard to formalize proved challenging
    - Recognizing spoken words or recognizing faces in images

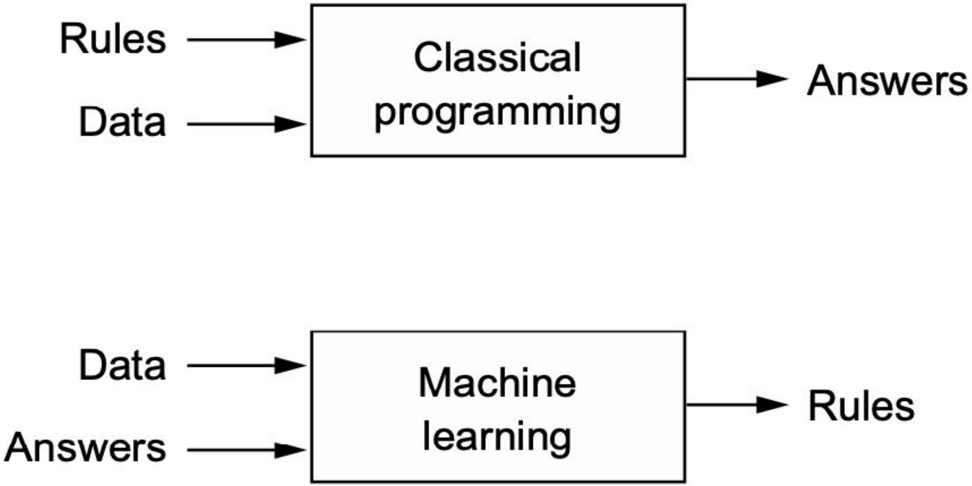


Figure 1.2 Machine learning: a new programming paradigm

Machine learning

- An ML system is *trained* instead of being explicitly programmed
- Started in the 1990s
- Has had explosive growth driven by faster hardware and larger datasets
- Example: wake word detection in voice assistants (Alexa, Siri, Google Assistant)

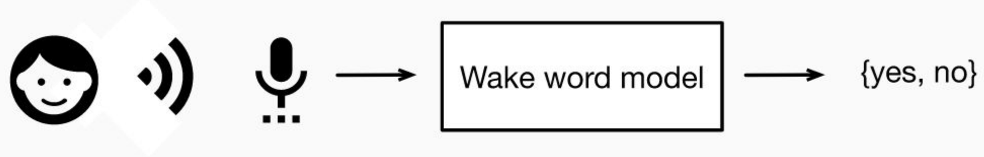


Fig. 1.1.1 Identify a wake word.

Wake word detection

- You do not know how to program a computer to recognize the word “Alexa”
- You yourself are able to recognize it!
- We can collect a huge dataset containing examples of:
    - Audio, and
    - Label, i.e. whether or not the audio contains the wake word

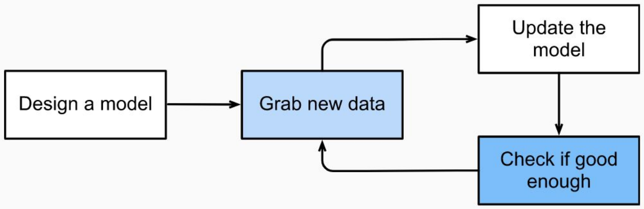


Fig. 1.1.2 A typical training process.

Representations

- An ML model *transforms* the input data (audio) into meaningful output (is the wake word in it?)
- Different representations are useful for different tasks: consider images in RGB (red-green-blue) vs HSV (hue-saturation-value) formats
  - “Select all red pixels” is easier in RGB, “Make the image less saturated” is easier in HSV
- We can make tasks easier by choosing better representations

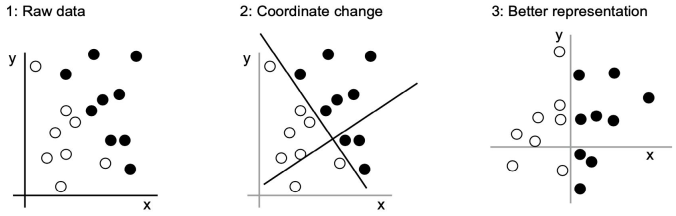


Figure 1.4 **Coordinate change**

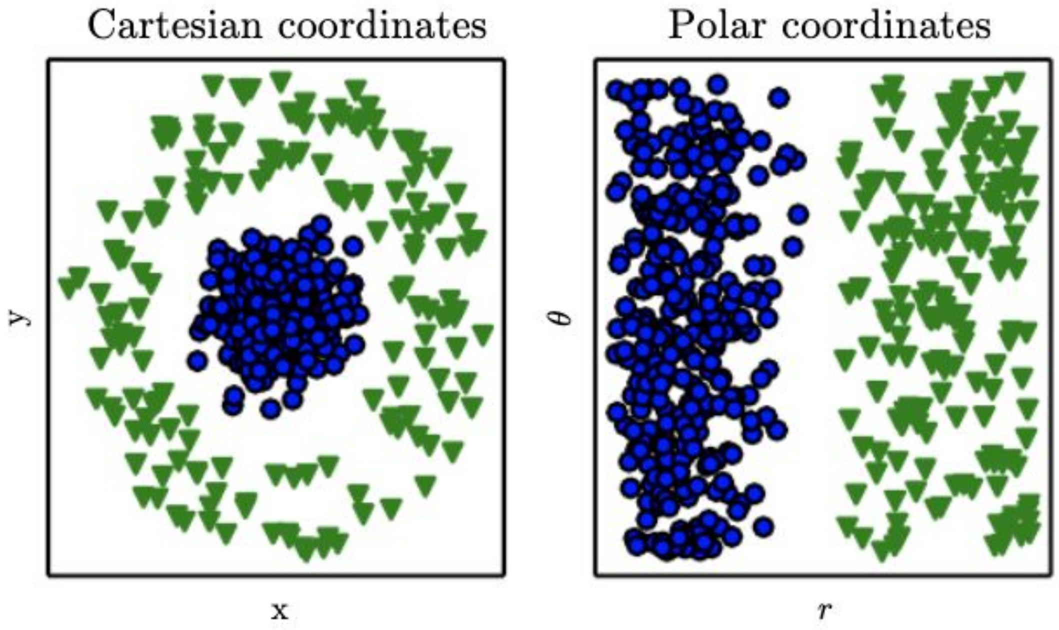
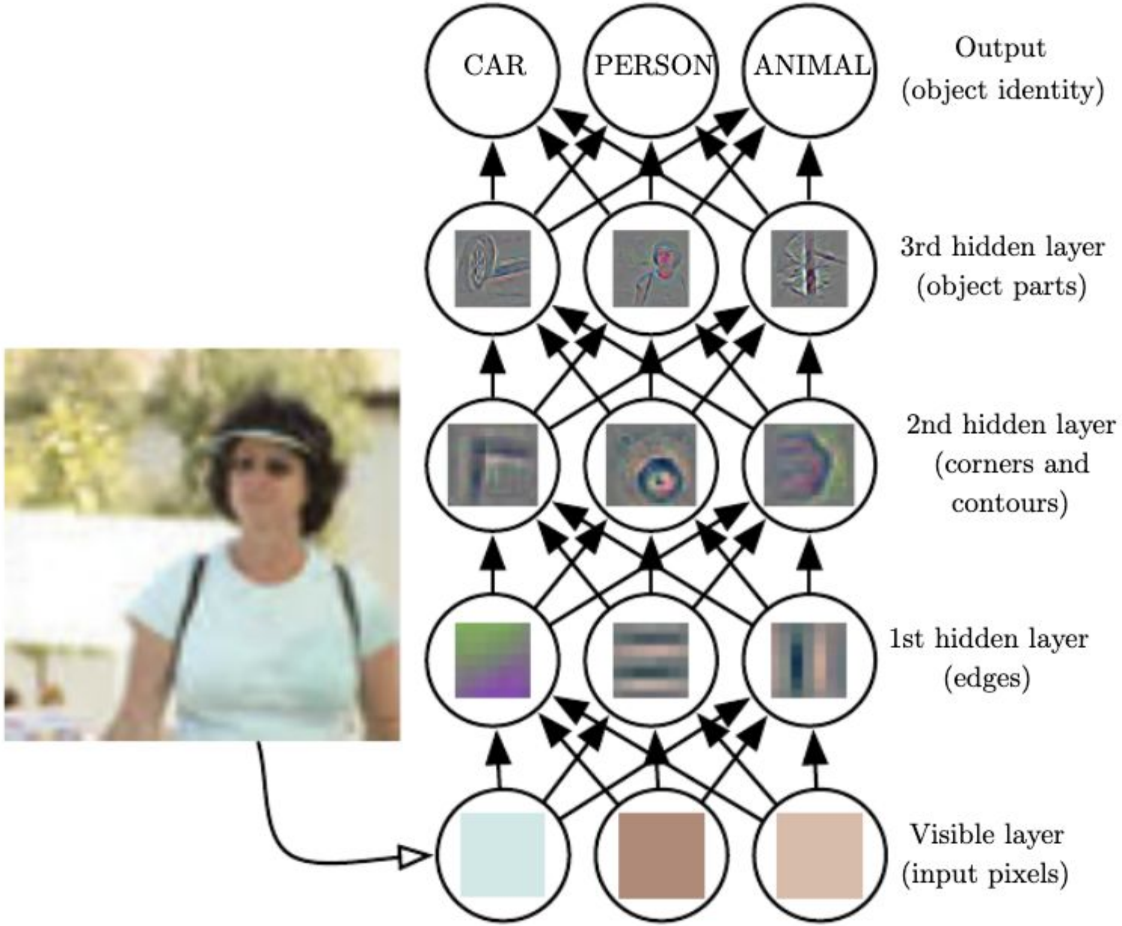


Figure 1.1: Example of different representations: suppose we want to separate two categories of data by drawing a line between them in a scatterplot. In the plot on the left, we represent some data using Cartesian coordinates, and the task is impossible. In the plot on the right, we represent the data with polar coordinates and the task becomes simple to solve with a vertical line. (Figure produced in collaboration with David Warde-Farley.)

Deep learning

- Learn successive layers of increasingly meaningful representations
- Enables the computer to build complex concepts out of simpler concepts
- Modern DL involves tens and sometimes hundreds of layers
- All of the parameters in these layers are learned automatically from data


CAR
PERSON
ANIMAL
Output
(object identity)
3rd hidden layer
(object parts)
2nd hidden layer
(corners and
contours)
1st hidden layer
(edges)
Visible layer
(input pixels)

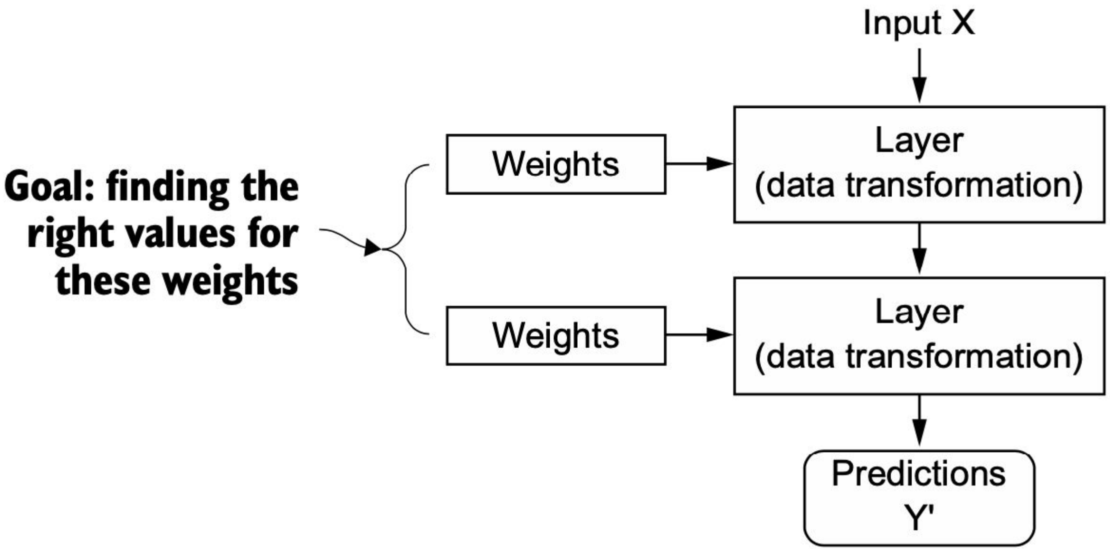


Figure 1.7 A neural network is parameterized by its weights.

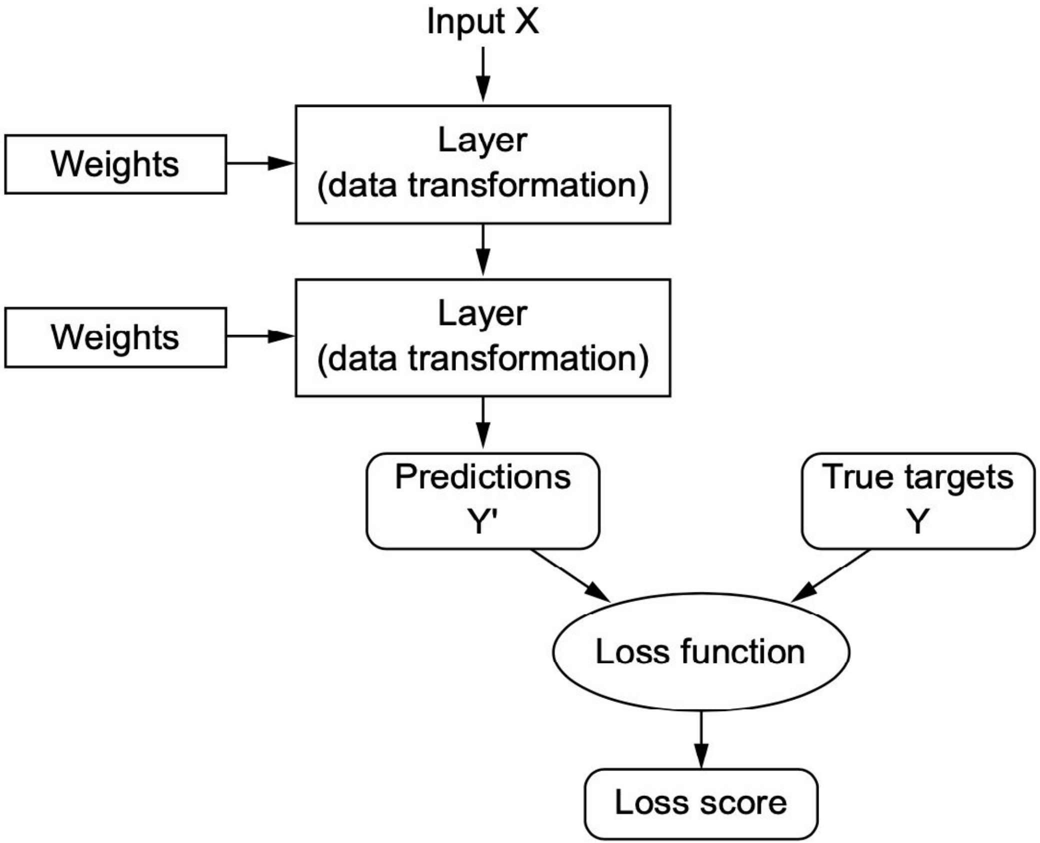


Figure 1.8 A loss function measures the quality of the network's output.

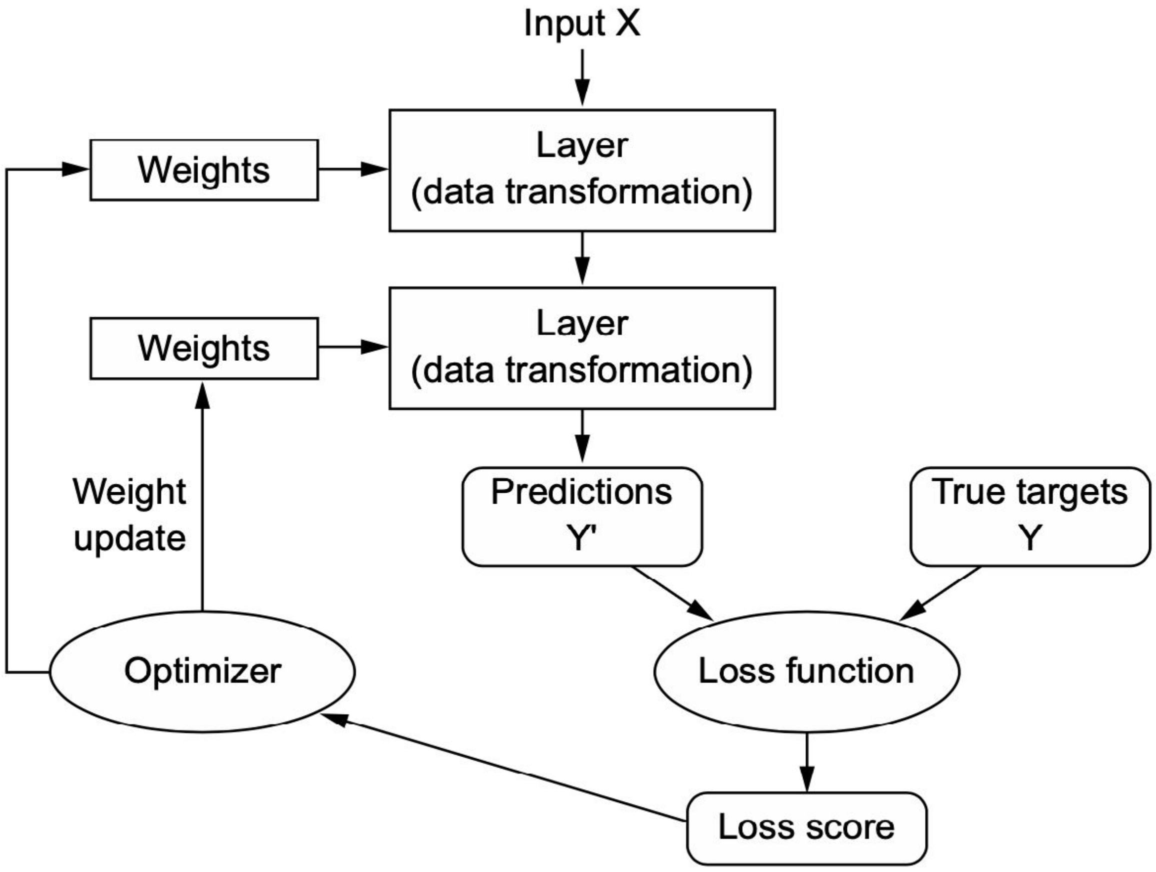
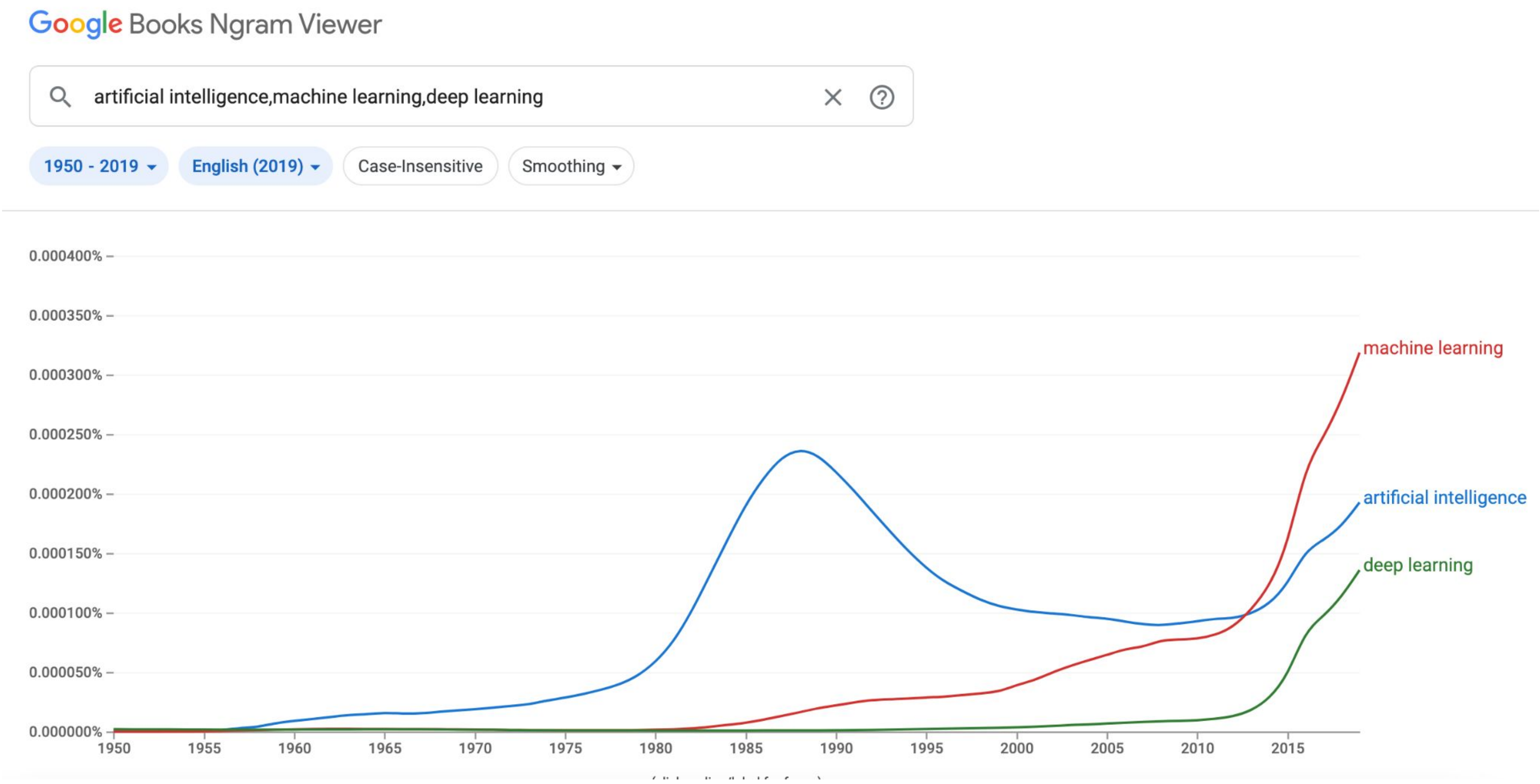


Figure 1.9 The loss score is used as a feedback signal to adjust the weights.

Deep learning: why now?



Google Books Ngram Viewer
artificial intelligence,machine learning,deep learning
1950 - 2019
English (2019)
Case-Insensitive
Smoothing
0.000400%
0.000350%
0.000300%
0.000250%
0.000200%
0.000150%
0.000100%
0.000050%
0.000000%
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
machine learning
artificial intelligence
deep learning

Historical roots of deep learning

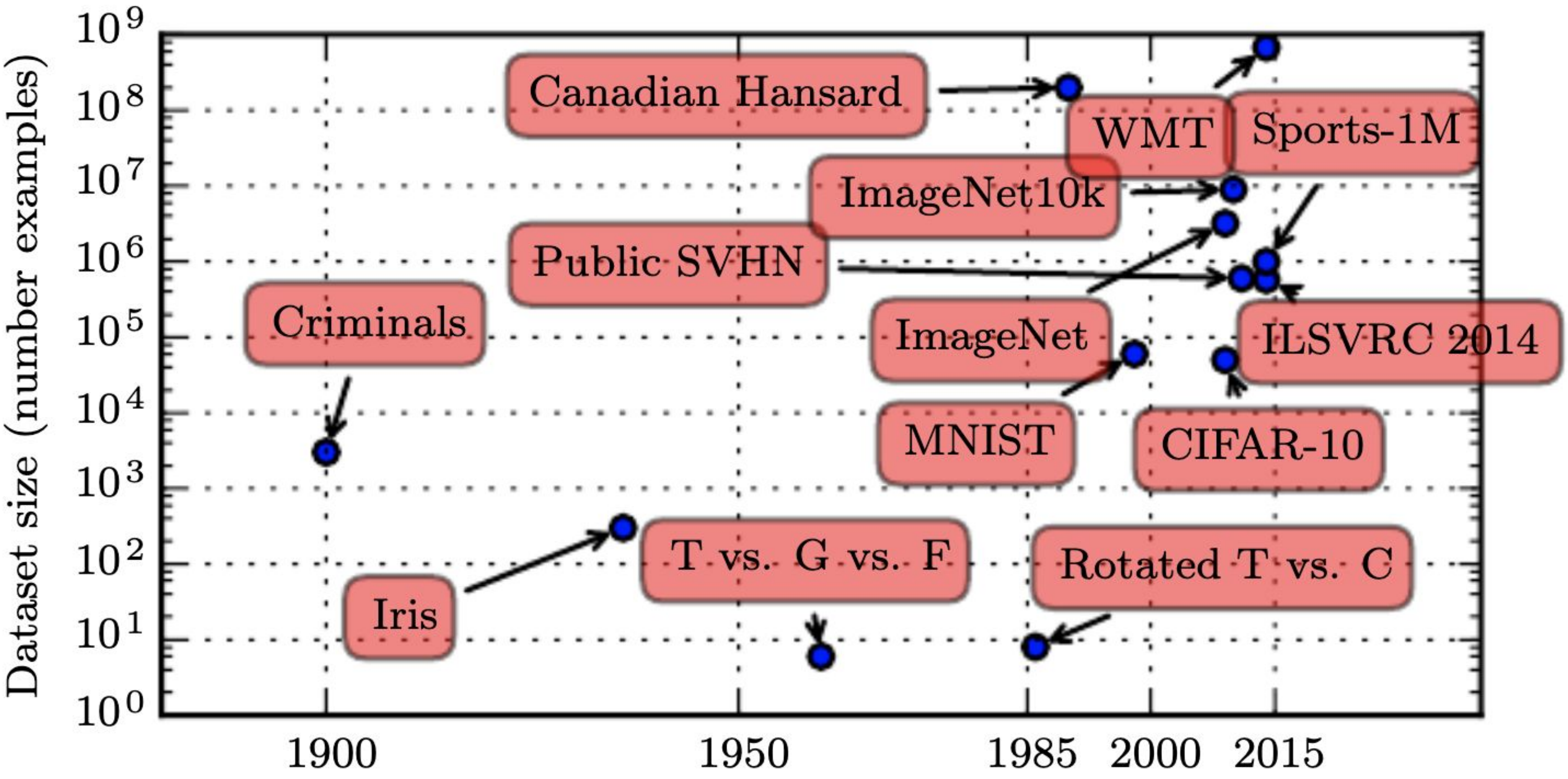
- The apparent novelty of deep learning is deceptive
- “Deep learning” is a recent rebranding of neural networks
- NN research has had three waves:
    - 1940s-60s: cybernetics
    - 1980s-90s: connectionism
    - Since 2006: deep learning
- (Artificial) neural networks are engineered systems inspired by the brain
- However, the models and algorithms we will use are NOT realistic biological models of brain functions
    - People are trying to develop biologically plausible versions of deep learning algorithms

What happened around 2010?

- Prior to 2010, algorithms were too computationally costly for available hardware
- In 2006, Geoff Hinton showed that a kind of neural network called a “deep belief network” could be efficiently trained
- This wave of neural networks research popularized the use of the term “deep learning”
- Researchers were now able to train deeper neural networks than had been possible before
- They focused their attention on the theoretical importance of depth

Table 1.5.1 Dataset vs. computer memory and computational power

Decade	Dataset	Memory	Floating point calculations per second
1970	100 (Iris)	1 KB	100 KF (Intel 8080)
1980	1 K (House prices in Boston)	100 KB	1 MF (Intel 80186)
1990	10 K (optical character recognition)	10 MB	10 MF (Intel 80486)
2000	10 M (web pages)	100 MB	1 GF (Intel Core)
2010	10 G (advertising)	1 GB	1 TF (Nvidia C2050)
2020	1 T (social network)	100 GB	1 PF (Nvidia DGX-2)


Dataset size (number examples)
10^0
10^1
10^2
10^3
10^4
10^5
10^6
10^7
10^8
10^9
1900
1950
1985
2000
2015
Canadian Hansard
WMT
Sports-1M
ImageNet10k
Public SVHN
Criminals
ImageNet
ILSVRC 2014
MNIST
CIFAR-10
T vs. G vs. F
Rotated T vs. C
Iris

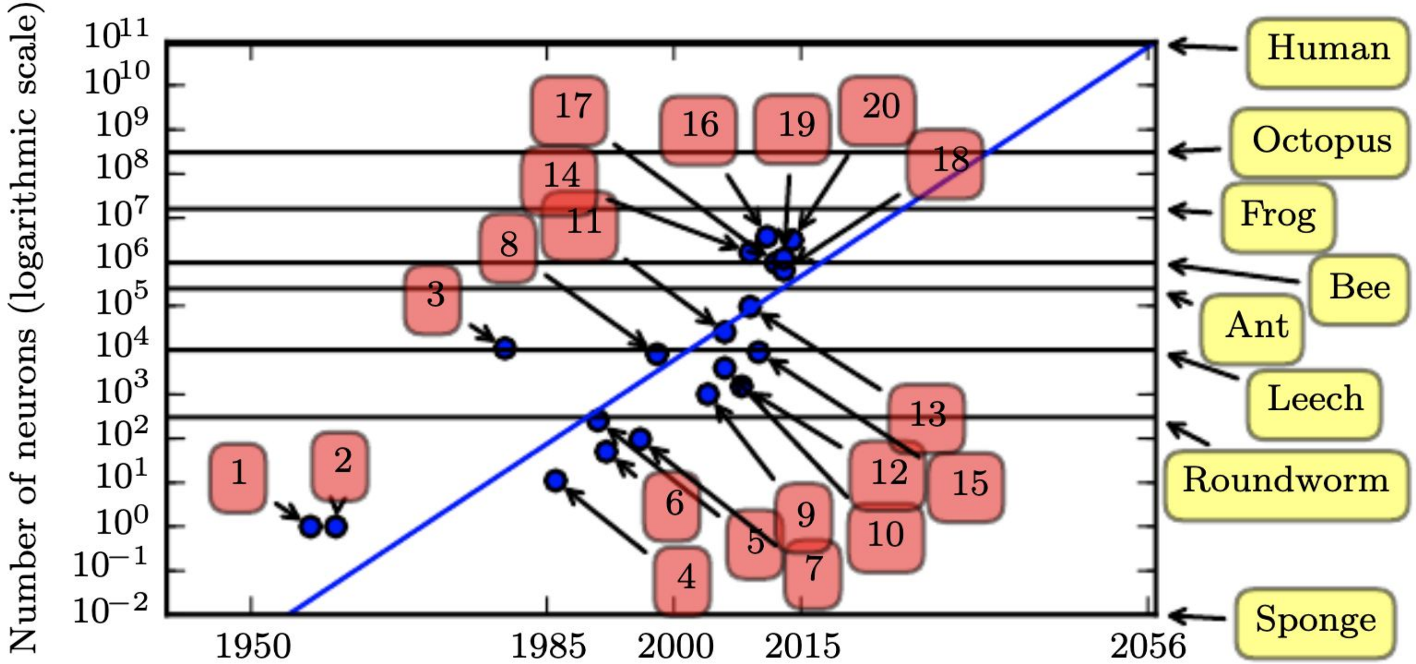
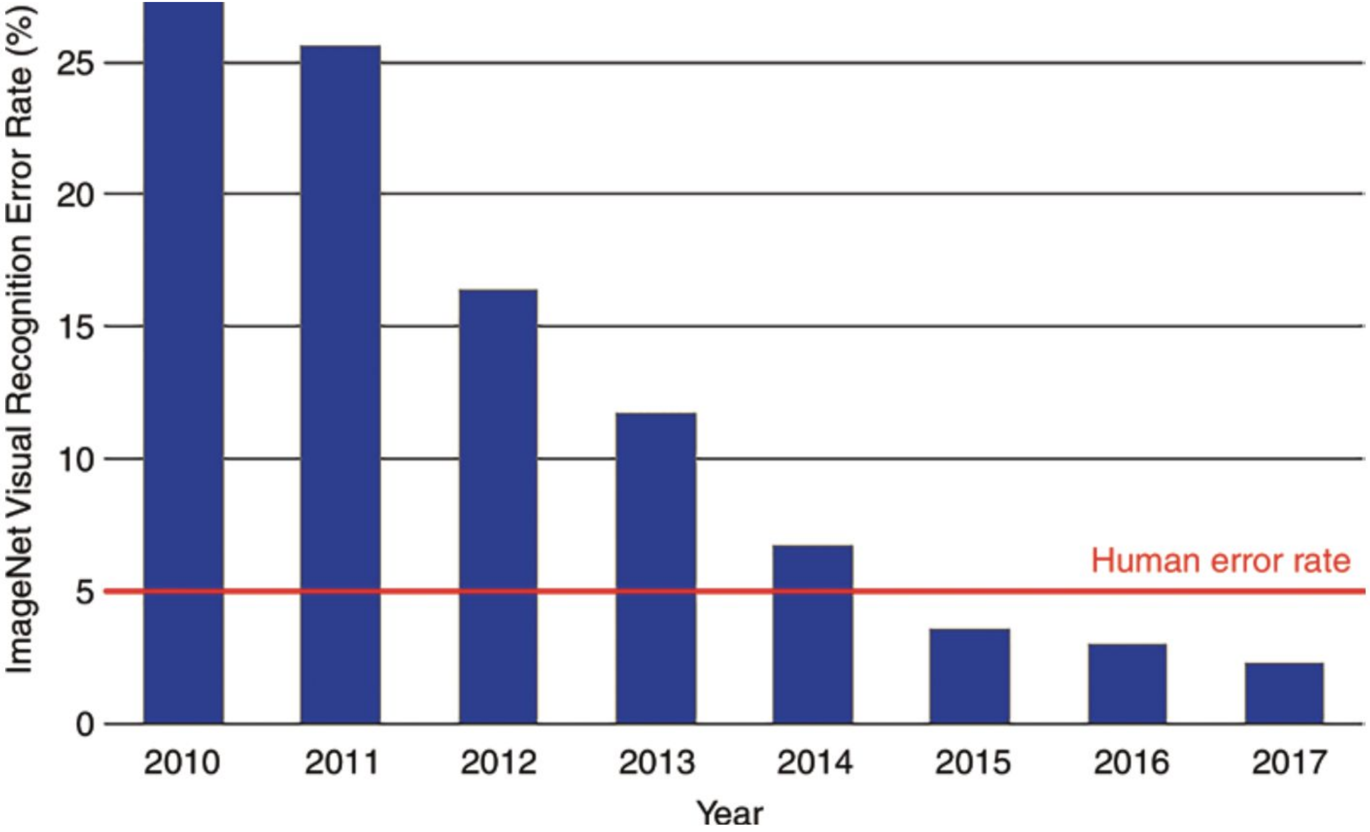
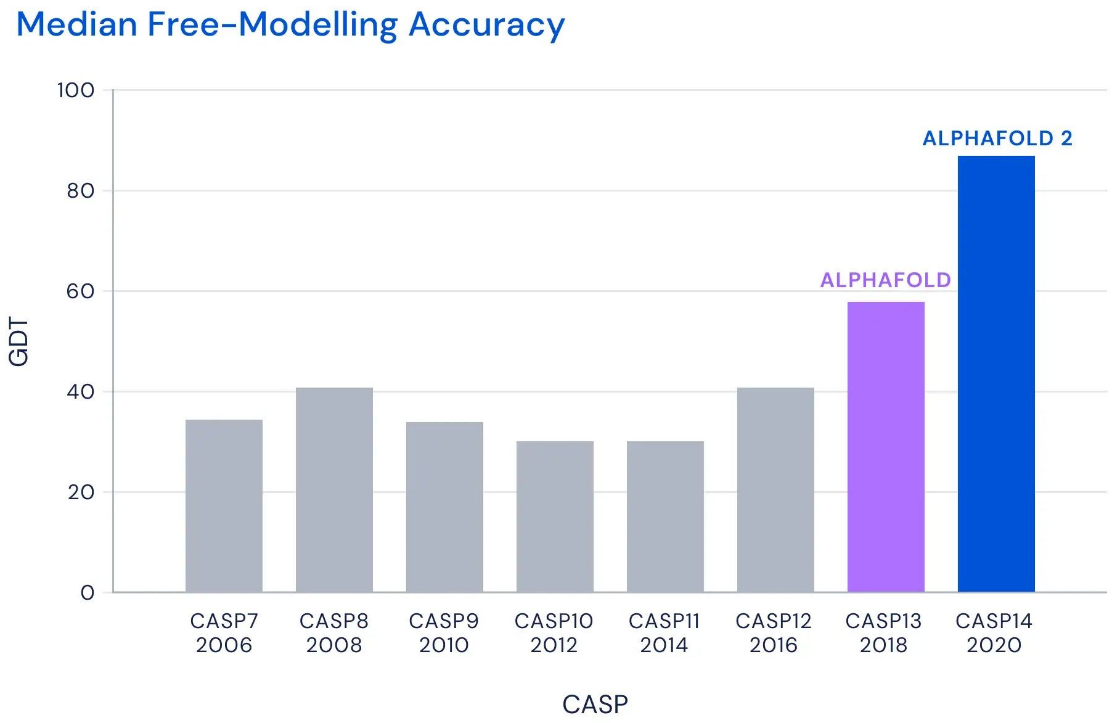


Figure 1.11: Increasing neural network size over time. Since the introduction of hidden units, artificial neural networks have doubled in size roughly every 2.4 years. Biological neural network sizes from Wikipedia (2015).



Error rates on the ImageNet Large-Scale Visual Recognition Challenge. Accuracy dramatically improved with the introduction of deep learning in 2012 and continued to improve thereafter. Humans perform with an error rate of approximately 5%.



Improvements in the median accuracy of predictions in the free modelling category for the best team in each CASP. A score of around 90 GDT is informally considered to be competitive with results obtained from experimental methods.

"CASP14 (2020) saw an enormous jump in the accuracy of single protein and domain models such that many are competitive with experiment. That advance is largely the result of the successful application of deep learning methods, particularly by the AlphaFold and, since that CASP, RosettaFold. As a consequence, computed protein structures are becoming much more widely used in a broadening range of applications. CASP has responded to this new landscape with a revised set of modeling categories."

Source: https://predictioncenter.org/casp15/

AlphaFold's success forced CASP to revise its competition structure and categories!

Success stories

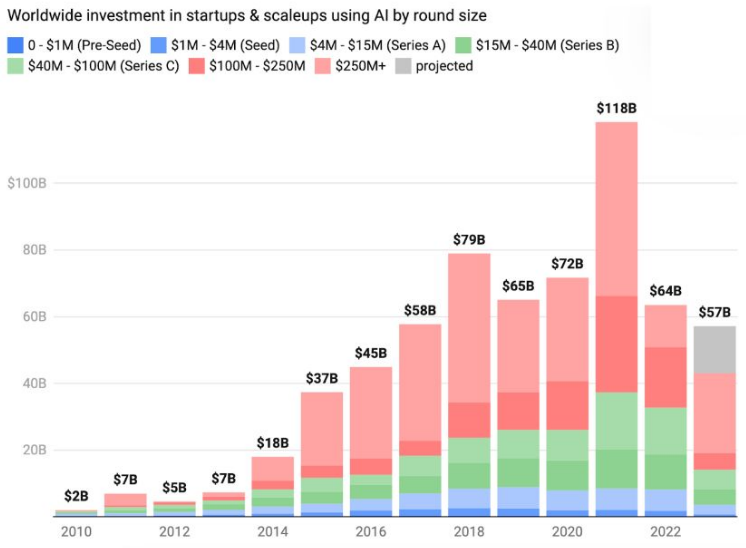
Near-human-level image classification
Near-human-level speech transcription
Near-human-level handwriting transcription
Dramatically improved machine translation
Dramatically improved text-to-speech conversion
Digital assistants such as Google Assistant and Amazon Alexa
Near-human-level autonomous driving
Improved ad targeting, as used by Google, Baidu, or Bing
Improved search results on the web
Superhuman Go playing
Dramatically improved protein structure prediction
New antibiotics to tackle antibiotic resistance
Large Language Models (LLMs) like GPT

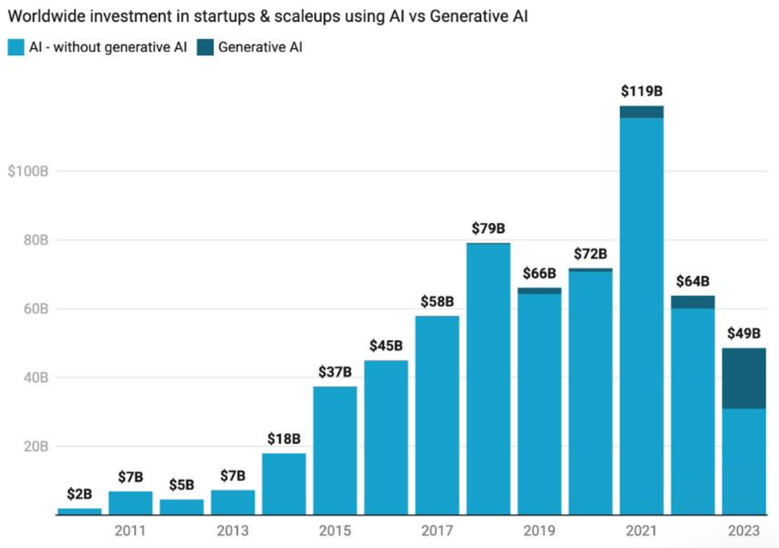
Real world impact of computing and AI

- In 2009 the top ten companies by market cap included only one big tech company, Microsoft
- As of Jan 2024, there are six: Apple, Microsoft, Alphabet (Google), Amazon, NVIDIA, Meta (Facebook)
  - Seven, if you consider Tesla a tech company

“GenAI” is the new “new” thing: AI investments are stable vs. 2022, powered by GenAI

- **Funding for startups using AI H1 2023 was nearly on par with H1 2022…without capital pouring into GenAI, overall AI investments would have suffered a 40% drop compared to last year vs. 54% drop across all startups.**





Source: state of AI report 2023 at https://www.stateof.ai/

Short term hype vs long term vision

- Expectations for what the field will be able to achieve in the next decade tend to run much higher than what will likely be possible
- Believable dialogue systems, human-level machine translation across arbitrary languages, and human-level natural language understanding may remain elusive for a long time
- We may be currently witnessing the third cycle of AI hype and disappointment, and we’re still in the phase of intense optimism. Might another AI Winter be lurking around?
- Long term vision: AI will help humanity as a whole move forward, by assisting human scientists in new breakthrough discoveries across all scientific fields, from genomics to mathematics
- Don’t believe the short-term hype, but do believe in the long-term vision!

Turing Award Won by 3 Pioneers in Artificial Intelligence

From left, Yann LeCun, Geoffrey Hinton and Yoshua Bengio. The researchers worked on key developments for neural networks, which are reshaping how computer systems are built. From left, Facebook, via Associated Press; Aaron Vincent Elkaim for The New York Times; Chad Buchanan/Getty Images

Ethical and societal issues in AI

- Energy consumption, sustainability, climate change
- Social, economic and political inequality
- Unemployment
- Autonomous weapons
- Surveillance and loss of privacy
- Safety and the problem of control