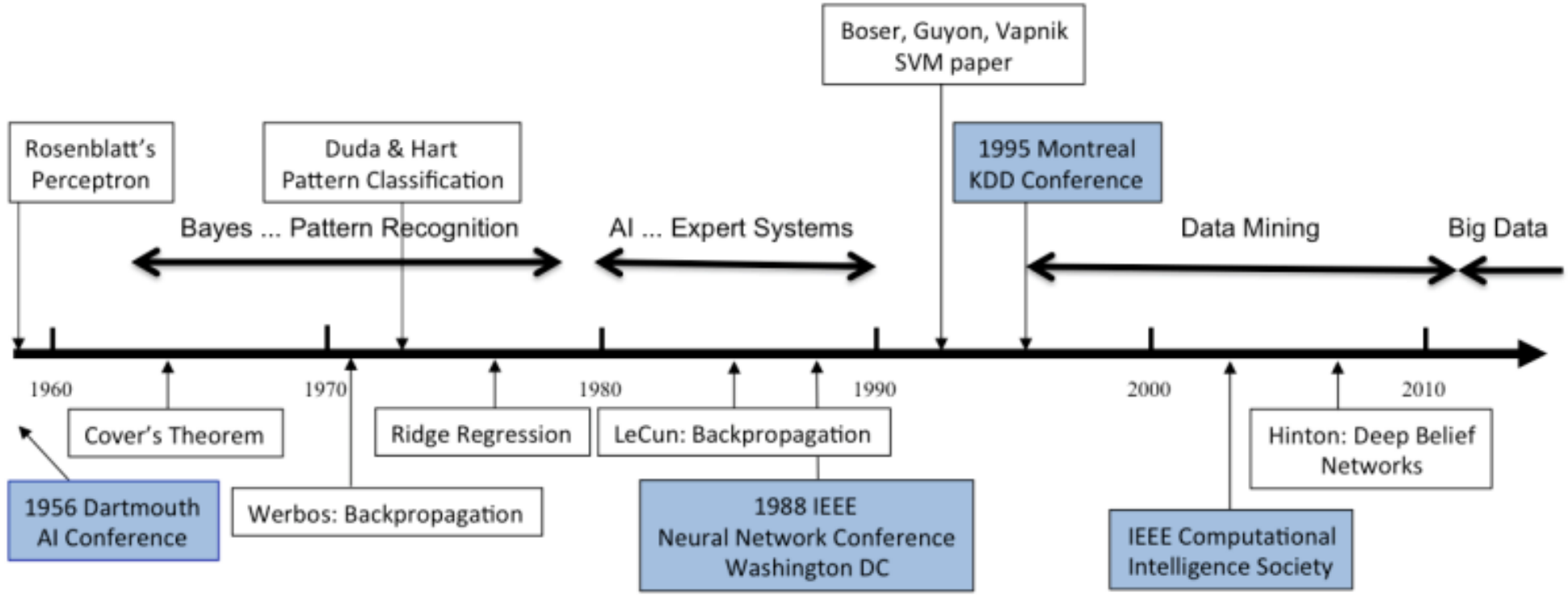
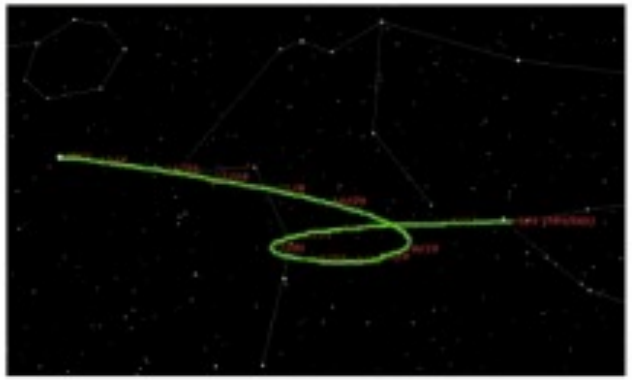
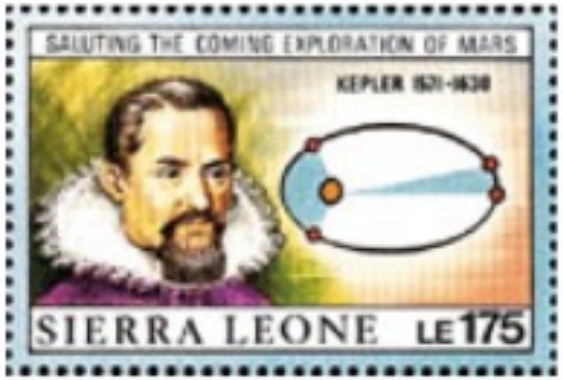
A Brief History of Artificial intelligence

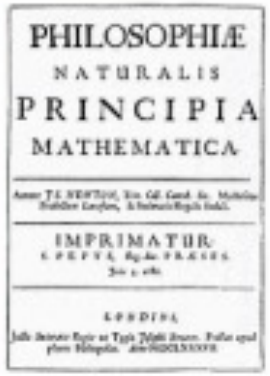
Since it's Inception AI is continuously being redefined



1609 Kepler's Law: Early History of Data-Driven Science



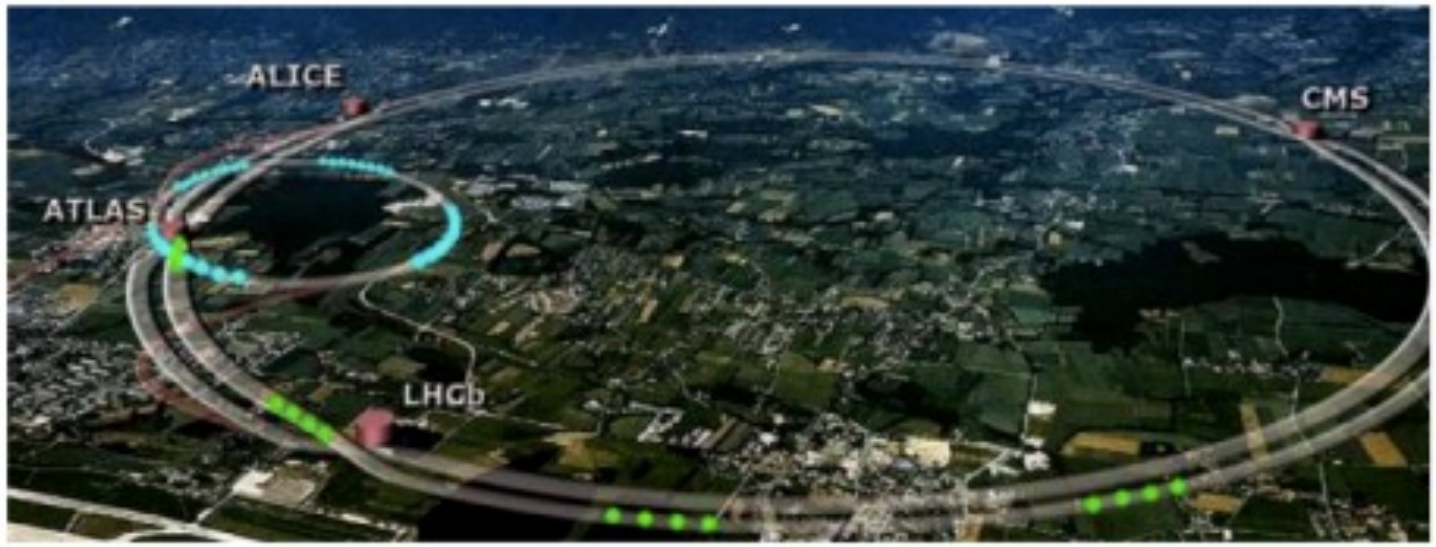




Lord Snow: Data-driven approach to identify cause of Cholera (1854)

The London physician john Snow traced the cause of cholera to the communal water pumps by plotting the number of cholera instances on London street maps. (Map drawn and lithographed by Charles Cheffins)

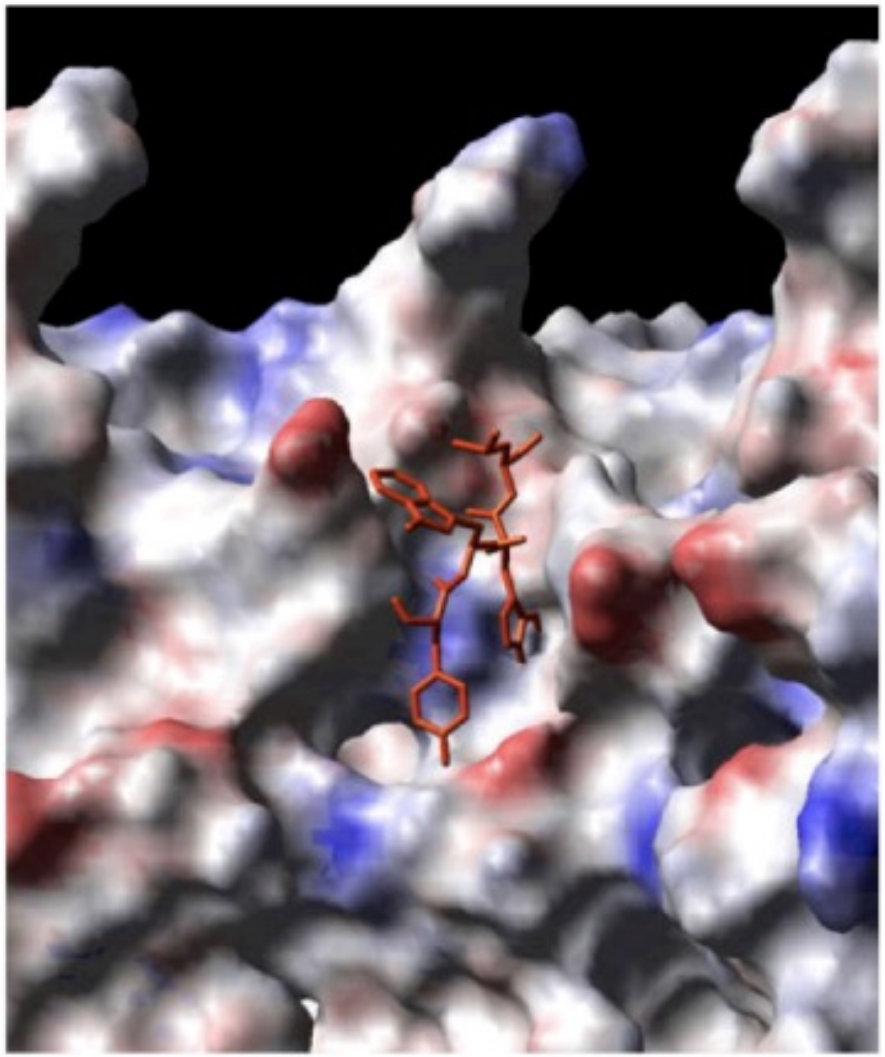
2012: The confirmation of the Higg's boson was data-driven



CERN's Large Hadron Collider and stamp commemorating the Large Hadron Collider which led to the confirmation of the existence of Higgs boson in 2012. An example of data-driven science.

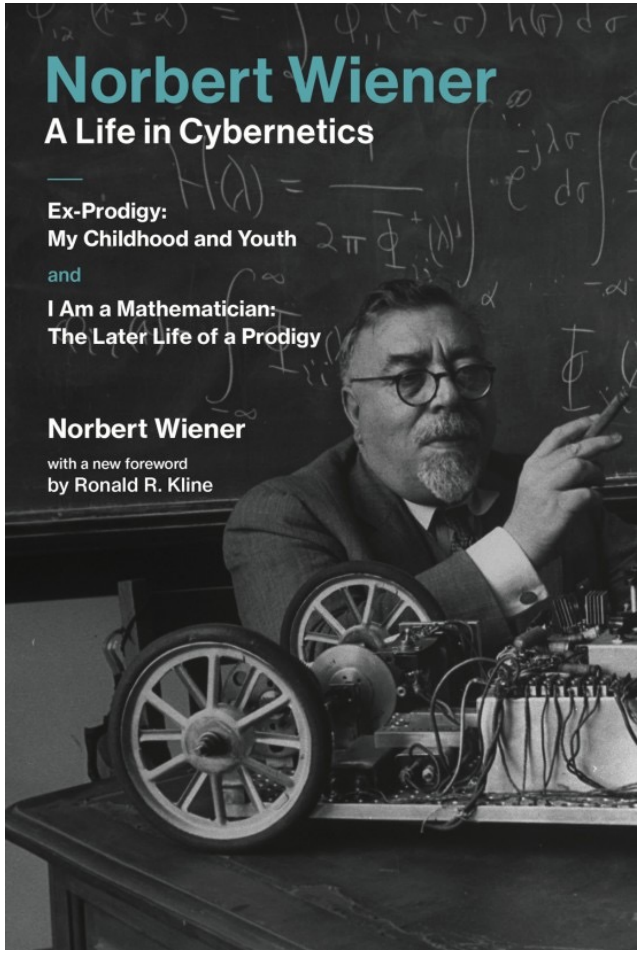
François Baron Englert and Peter W. Higgs are jointly awarded the Nobel Prize in Physics 2013 for the theory on how particles acquire mass

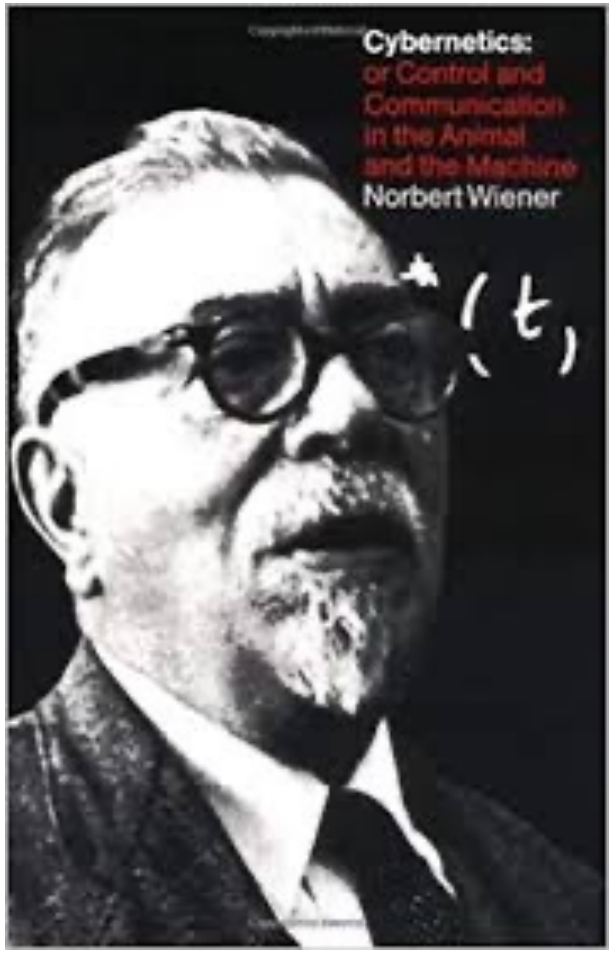
1860-1960: QSAR (quantitative structure activity relationship) as a data-driven approach for drug design

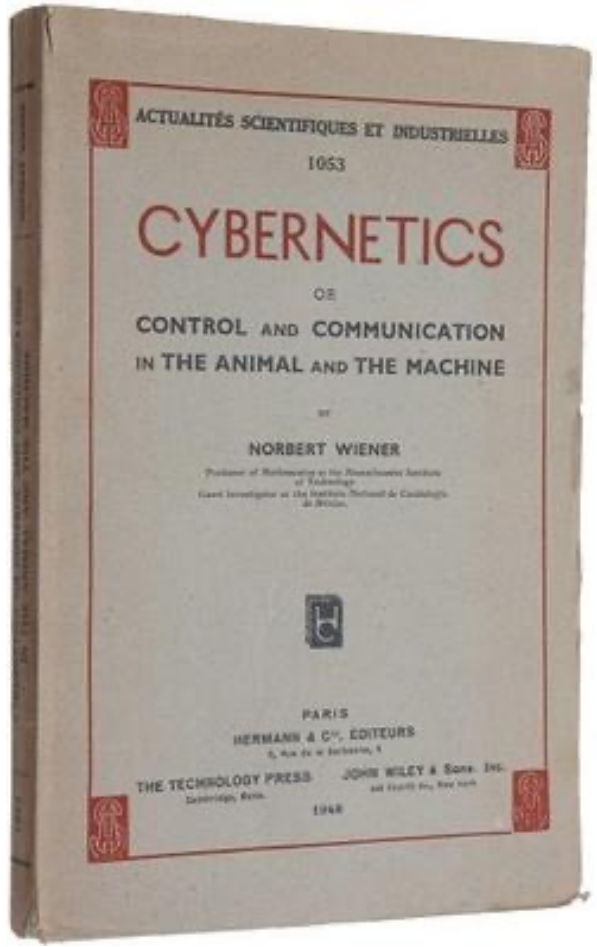


Hugo Kubinyi [2002] From narcosis to hyperspace: The history of QSAR. Quantitative Structure-Activity Relationships, Vol. 21, pp. 348-356.

1948: Norbert Wiener publishes Cybernetics







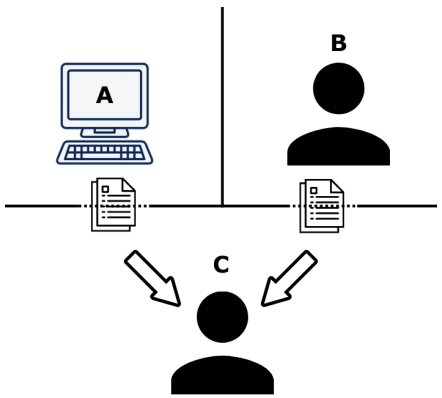
- Second part of the book talks about the future of neural networks
- Often cited as one of the most influential books of the 20th century

Norbert Wiener [1948] Cybernetics: Control and Communication in the Animal and the Machine, The Technology Press

Alan Turing: 1950 Paper Computing machinery and Intelligence

We may hope that machines will eventually compete with men in all purely intellectual fields.
But which are the best ones to start with? Even this is a difficult decision.
Many people think that a very abstract activity, like the playing of chess, would be best
It can also be maintained that it is best to provide the machine with the best sense organs that money can buy, and then teach it to understand and speak English.
This process could follow the normal teaching of a child.
Things would be pointed out and named, etc.
Again I do not know what the right answer is, but I think both approaches should be tried.

- Introduces the Turing Test



Born	Alan Mathison Turing 23 June 1912 Maida Vale, London, United Kingdom
Died	7 June 1954 (aged 41) Wilmslow, Cheshire, United Kingdom
Cause of death	Suicide (disputed) by cyanide poisoning
Resting place	Ashes scattered in gardens of Woking Crematorium
Residence	Wilmslow, Cheshire, United Kingdom
Nationality	English
Education	Sherborne School
Alma mater	University of Cambridge (BA, MA) Princeton University (PhD)
Known for	Cryptanalysis of the Enigma Turing's proof Turing machine Turing test Unorganised machine Turing pattern Turing reduction The Chemical Basis of Morphogenesis
Partner(s)	Joan Clarke (engaged in 1941; did not marry)
Awards	Smith's Prize (1936)
	Scientific career
Fields	Logic Mathematics Cryptanalysis Computer science Mathematical and theoretical biology[1]

A. M. Turing [1950] Computing Machinery and Intelligence. *Mind*, Vol. 49, pp. 433-460

1956: Darthmouth Summer Research Project on Artificial Intelligence (DSRPAI)

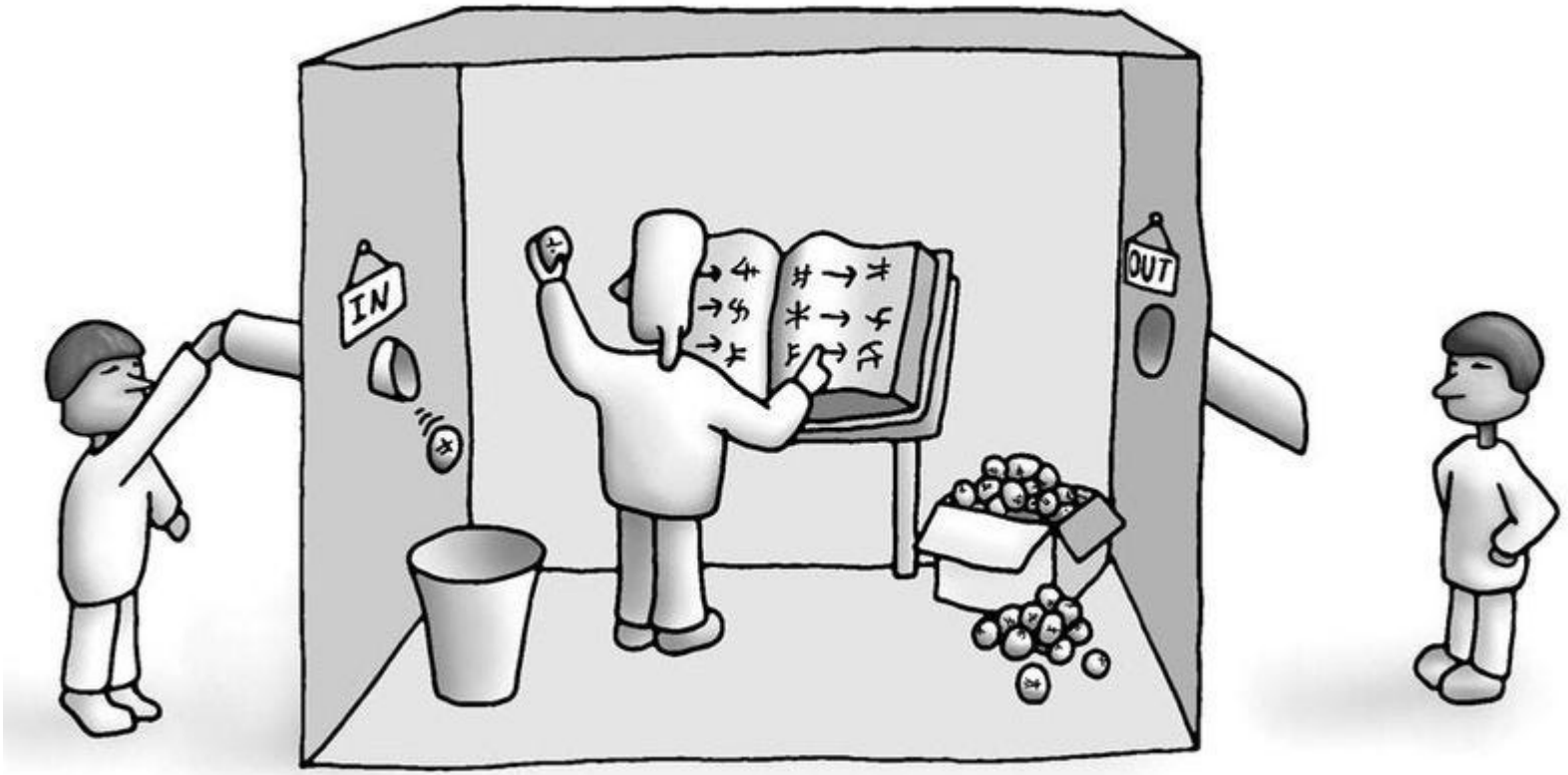
In the summer of 1956, John McCarthy, Marvin Minsky and Claude Shannon organized a conference at Dartmouth College on the subject of what they called "artificial intelligence" (a term coined by McCarthy for the occasion).

1. Ray Solomonoff
2. Marvin Minsky
3. John McCarthy
4. Claude Shannon
5. Trenchard More
6. Nat Rochester
7. Oliver Selfridge
8. Julian Bigelow
9. W. Ross Ashby
10. W.S. McCulloch
11. Abraham Robinson
12. Tom Etter
13. John Nash
14. David Sayre
15. Arthur Samuel
16. Kenneth R. Shoulders
17. Shoulders' friend
18. Alex Bernstein
19. Herbert Simon
20. Allen Newell

- John McCarthy coins the term Artificial Intelligence

J. Moor [2006] The Dartmouth College Artificial Intelligence Conference: The Next Fifty years. *AI Magazine*, Vol 27(4), pp. 87-89

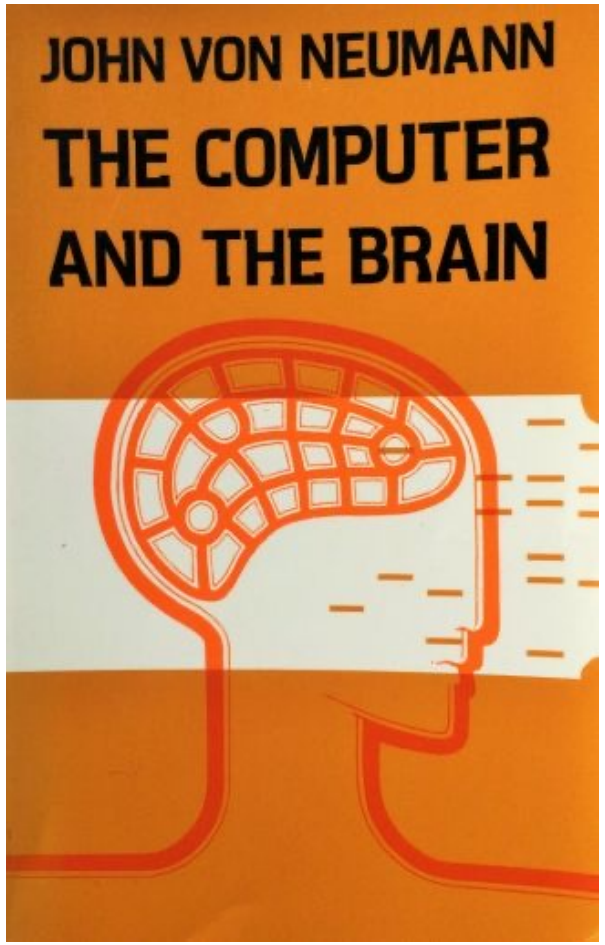
1980: John Searle Chinese room experiment: Does AI really understand Chinese?

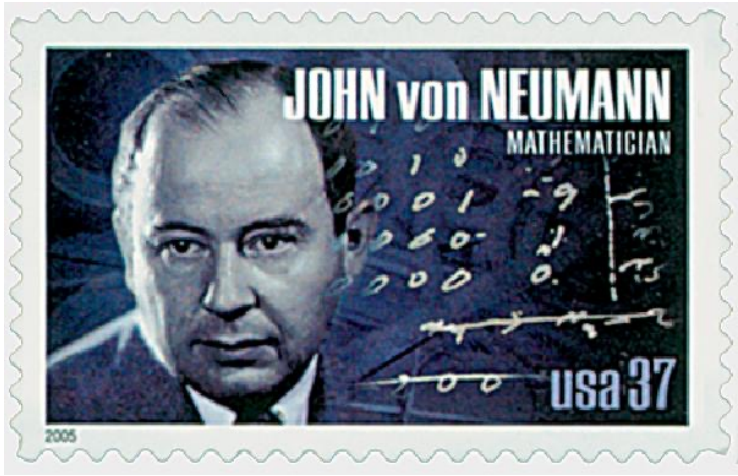


The question Searle wants to answer is: Does the machine *really* understand Chinese?
Or is it merely *simulating* the ability to understand Chinese?
Searle calls the first position "strong AI" and the latter "weak AI"

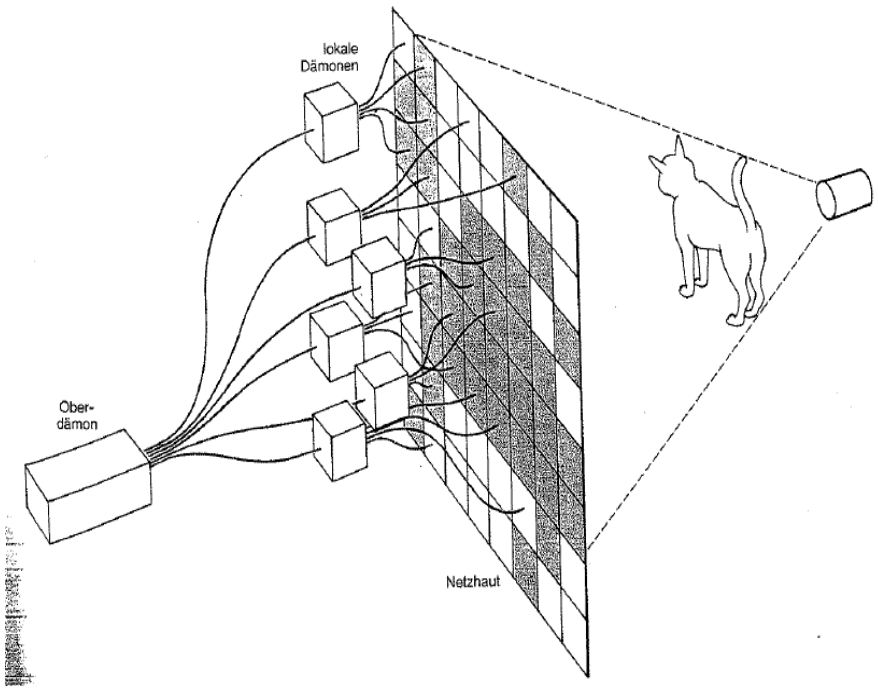
*John Searle [1980] Minds, Brains and Programs. Behavioral and Brain Sciences Vol **3** (3), pp. 417–457.*

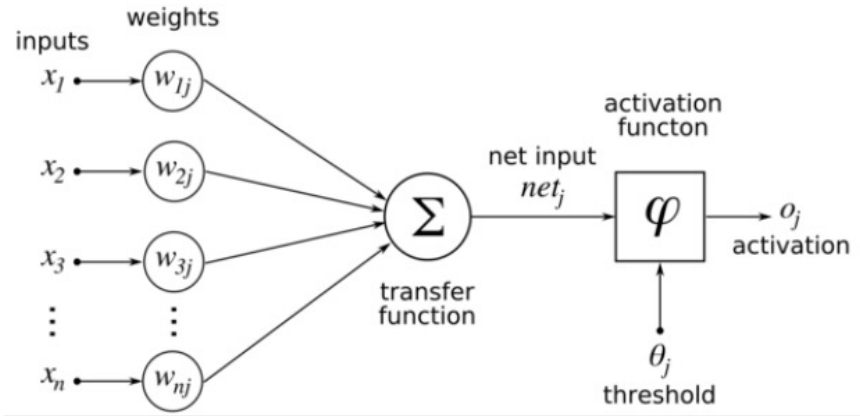
1956: John von Neumann – The computer and the brain

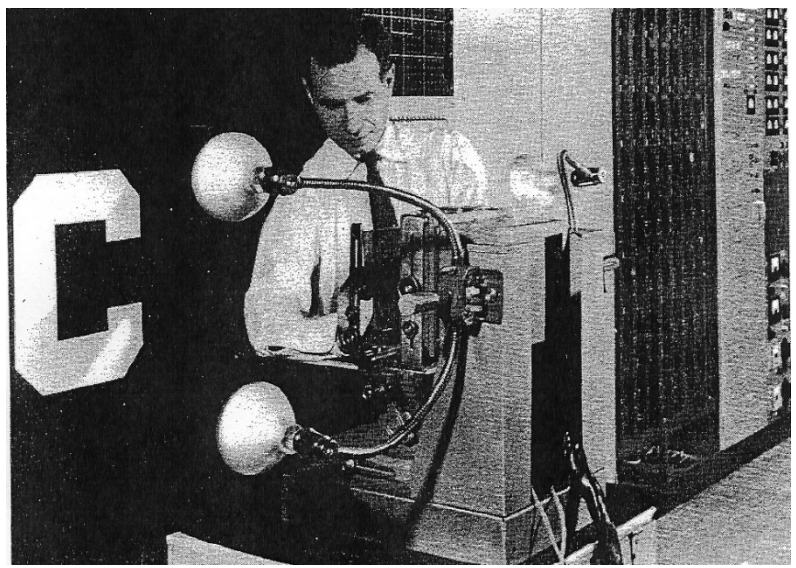


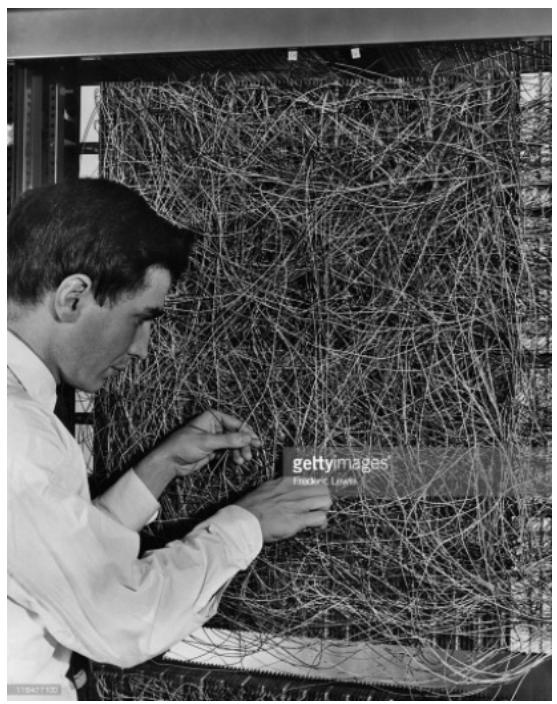


1957: Rosenblatt's Perceptron





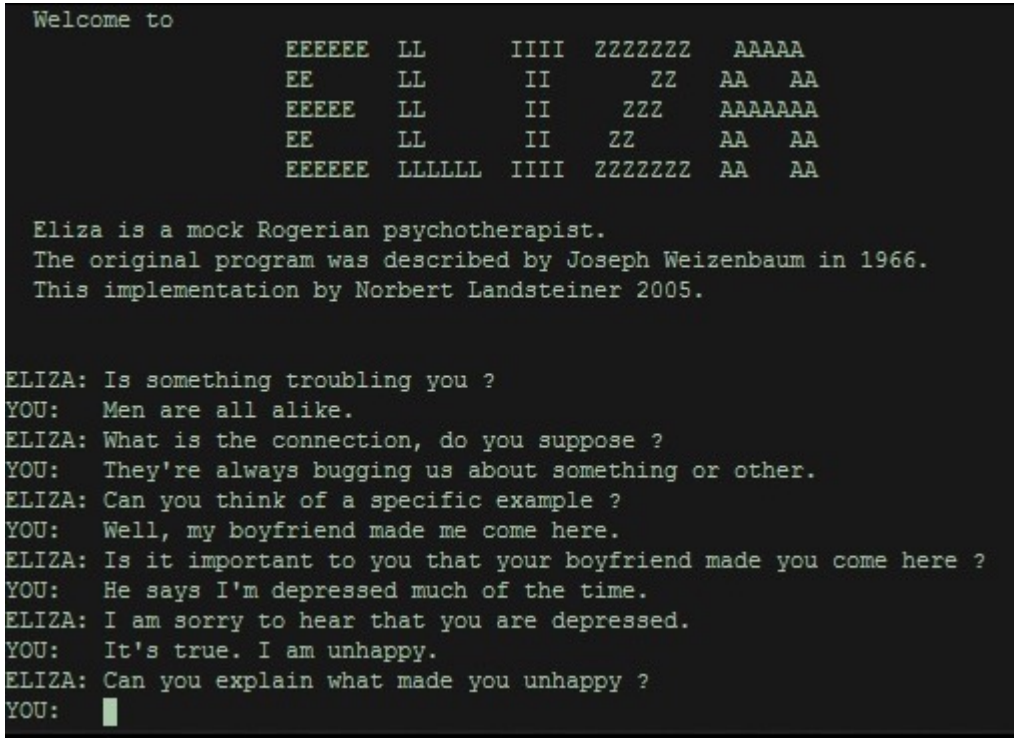




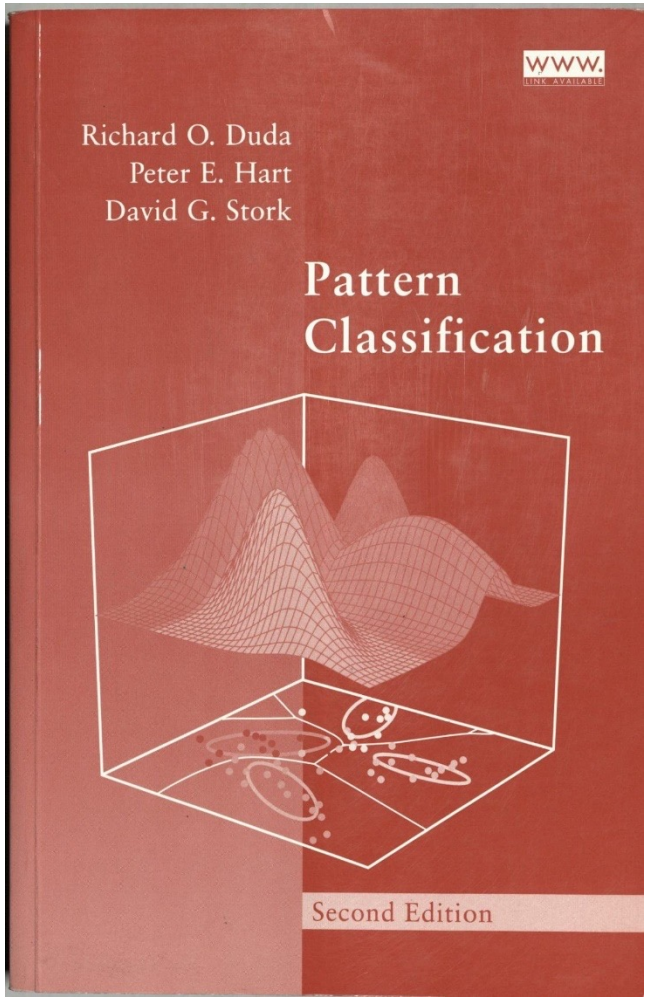
Rosenblatt, Frank (1957), The Perceptron--a perceiving and recognizing automaton. Report 85-460-1, Cornell Aeronautical Laboratory
Frank Rosenblatt [1958] The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. Psychological Review, Vol. 65(6), pp. 386–408.

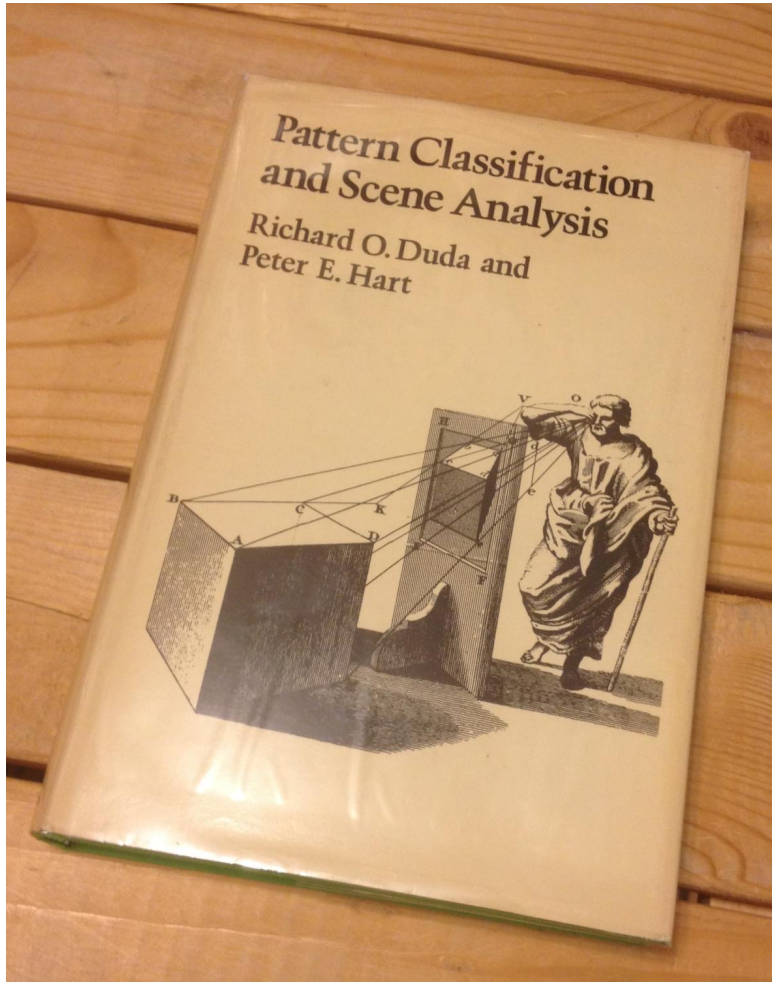
1964-1966: Joseph Weizenbaum's *ELIZA*

Joseph Weizenbaum's ELIZA, running the DOCTOR script, was created to provide a parody of "the responses of a non-directional psychotherapist in an initial psychiatric interview" and to "demonstrate that the communication between man and machine was superficial[“.] While ELIZA is most well known for acting in the manner of a psychotherapist, this mannerism is due to the data and instructions supplied by the DOCTOR script. ELIZA itself examined the text for keywords, applied values to said keywords, and transformed the input into an output; the script that ELIZA ran determined the keywords, set the values of keywords, and set the rules of transformation for the output. Weizenbaum chose to make the DOCTOR script in the context of psychotherapy to "sidestep the problem of giving the program a data base of real-world knowledge“.



1970's: AI = Pattern Recognition





1980 Eward Feigenbaum: AI = Expert Systems

Expert systems are computer programs aiming to model human expertise in one or more specific knowledge areas.
They usually consist of three basic components:

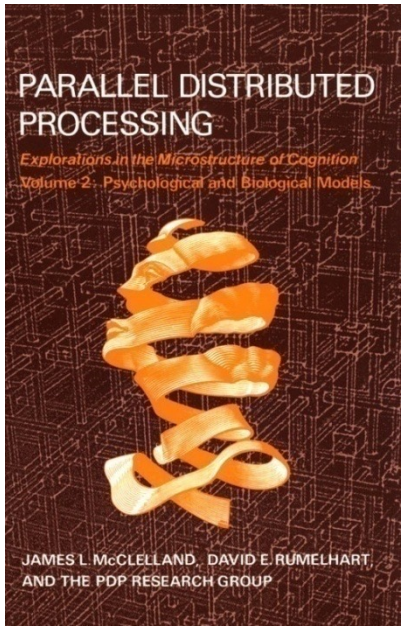
- a knowledge database with facts and rules representing human knowledge and experience
- an inference engine processing consultation and determining how inferences are being made
- an input/output interface for interactions with the user

Expert systems can be characterized by:

- using symbolic logic rather than only numerical calculations
- the processing is data-driven;
- a knowledge database containing explicit contents of certain area of knowledge
- the ability to interpret its conclusions in the way that is understandable to the user.

Symbolics Lisp Machine: an early platform for expert systems

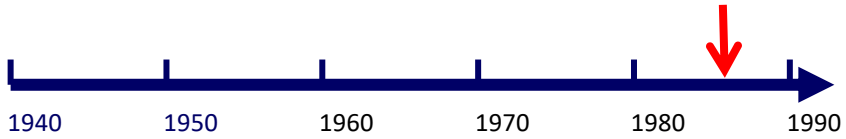
1986: Rediscovery of backpropagation

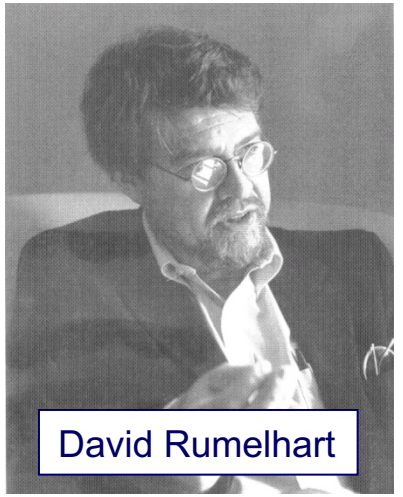


Learning representations by back-propagating errors

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

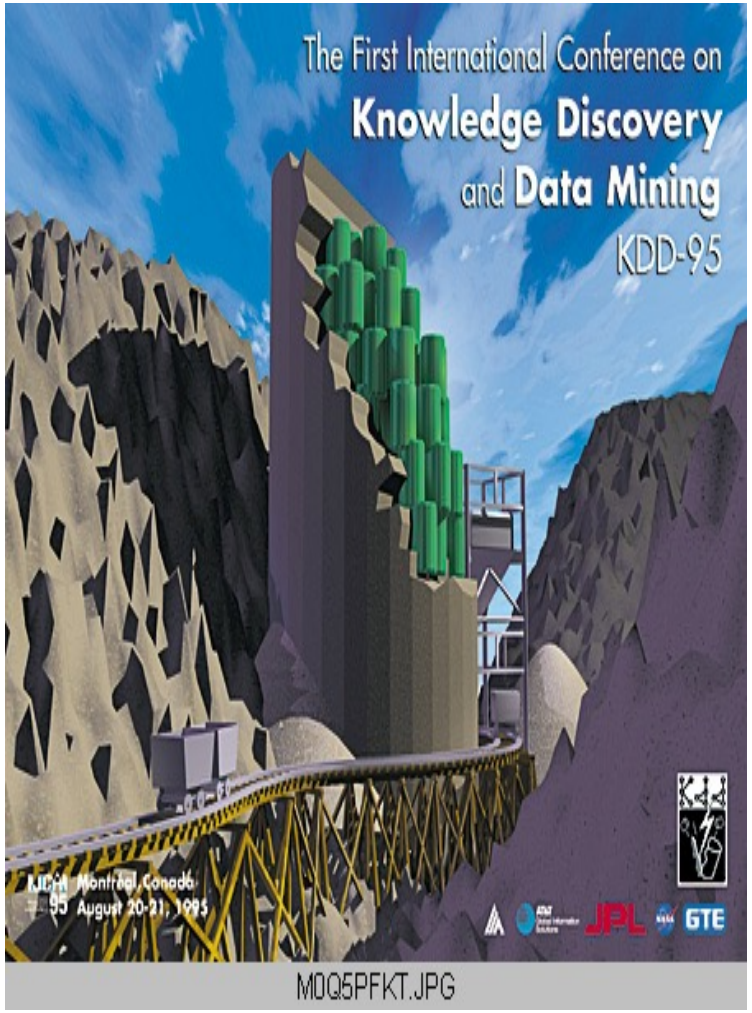
* Institute for Cognitive Science, C-015, University of California, San Diego, La Jolla, California 92093, USA
† Department of Computer Science, Carnegie-Mellon University, Pittsburgh, Philadelphia 15213, USA





David E. Rummelhart, Geoffrey E. Hinton, and Ronald J. Williams [1986] Learning representations by backpropagating errors. Nature, Vol. 323 (9), pp. 533-536

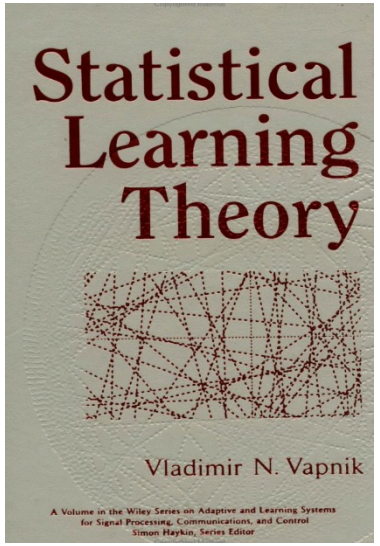
1990's: Data Mining and Machine Learning

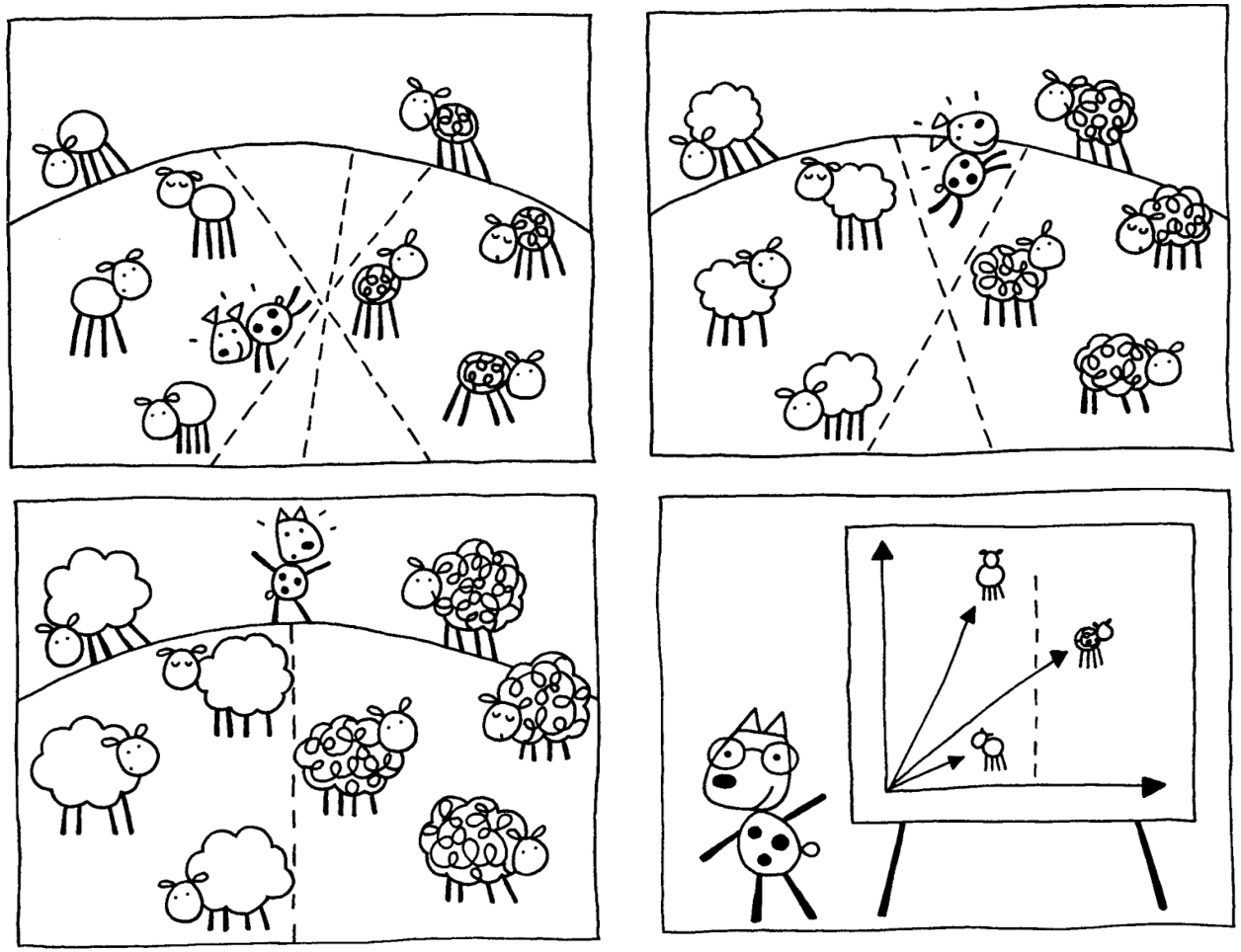


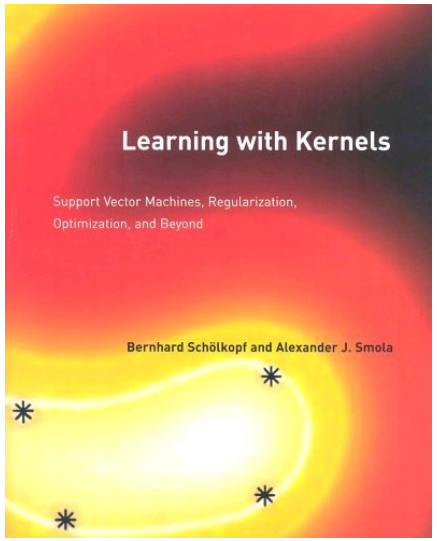
- Knowledge discovery in large data bases
- Rather than let the experts speak, let the data speak

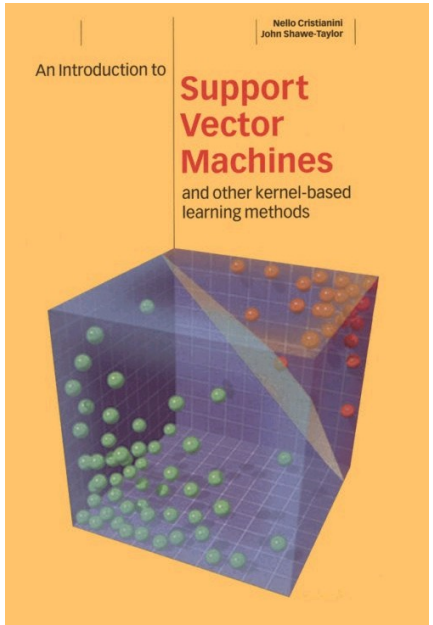
Data mining is the process of automatically extracting valid, novel, potentially useful and ultimately comprehensible information from very large databases

1992 Vapnik-Guyon-Boser: Support Vector Machines









B. Boser, I. Guyon,and V. Vapnik [1992] A training algorithm for optimal margin classifiers. In: Proceedings of the Fifth Annual Workshop on Computational Learning Theory, Pittsburgh.

1992 TD-Gammon: Neural Networks champion in computer-based backgammon

- IBM's Gerads Tesauro
- Based on reinforcement learning

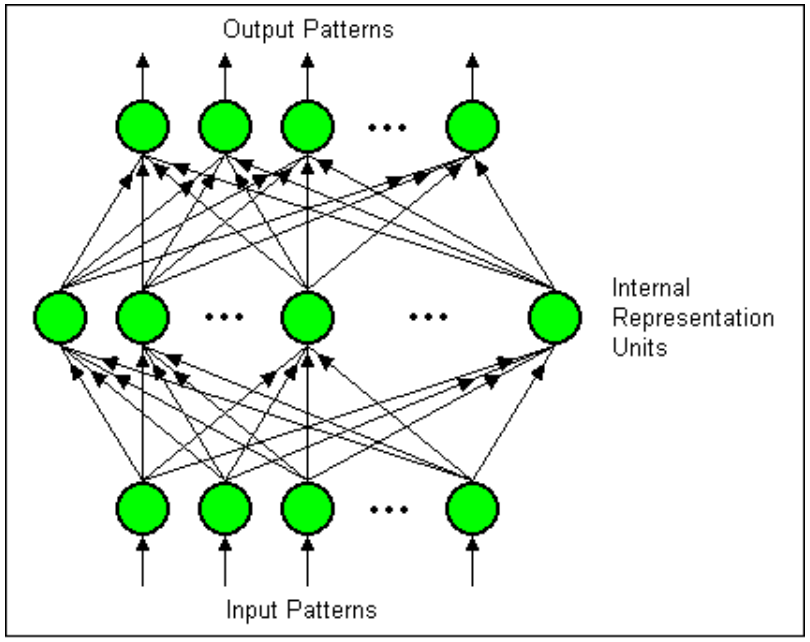


Figure 1. An illustration of the multilayer perception architecture used in TD-Gammon's neural network. This architecture is also used in the popular backpropagation learning procedure. Figure reproduced from [9].

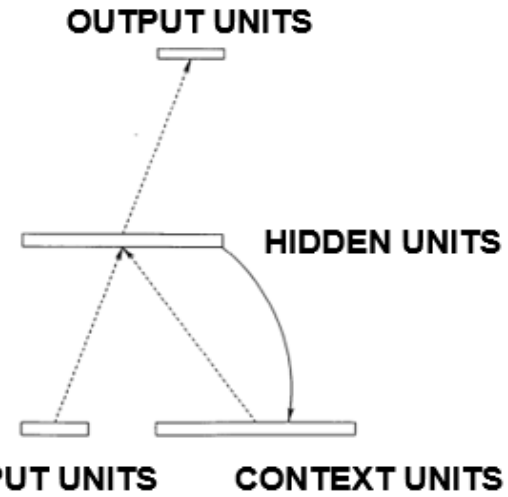
Schmidhuber-Hochreiter (1997): Long Short-Term Memory(LSTM) for recurrent networks: Basis for speech processing and language translation with RNNS (Recurrent neural Nets)

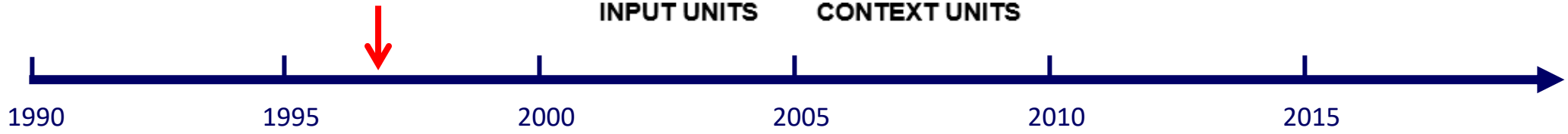
LONG SHORT-TERM MEMORY

NEURAL COMPUTATION 9(8):1735–1780, 1997

Sepp Hochreiter
Fakultät für Informatik
Technische Universität München
80290 München, Germany
hochreit@informatik.tu-muenchen.de
http://www7.informatik.tu-muenchen.de/~hochreit

Jürgen Schmidhuber
IDSIA
Corso Elvezia 36
6900 Lugano, Switzerland
juergen@idsia.ch
http://www.idsia.ch/~juergen





Sepp Hochreiter and Jurgen Schmidhuber [1997] Long Short-Term Memory. Neural Computation, Vol. 9(8), pp. 1735-1780.

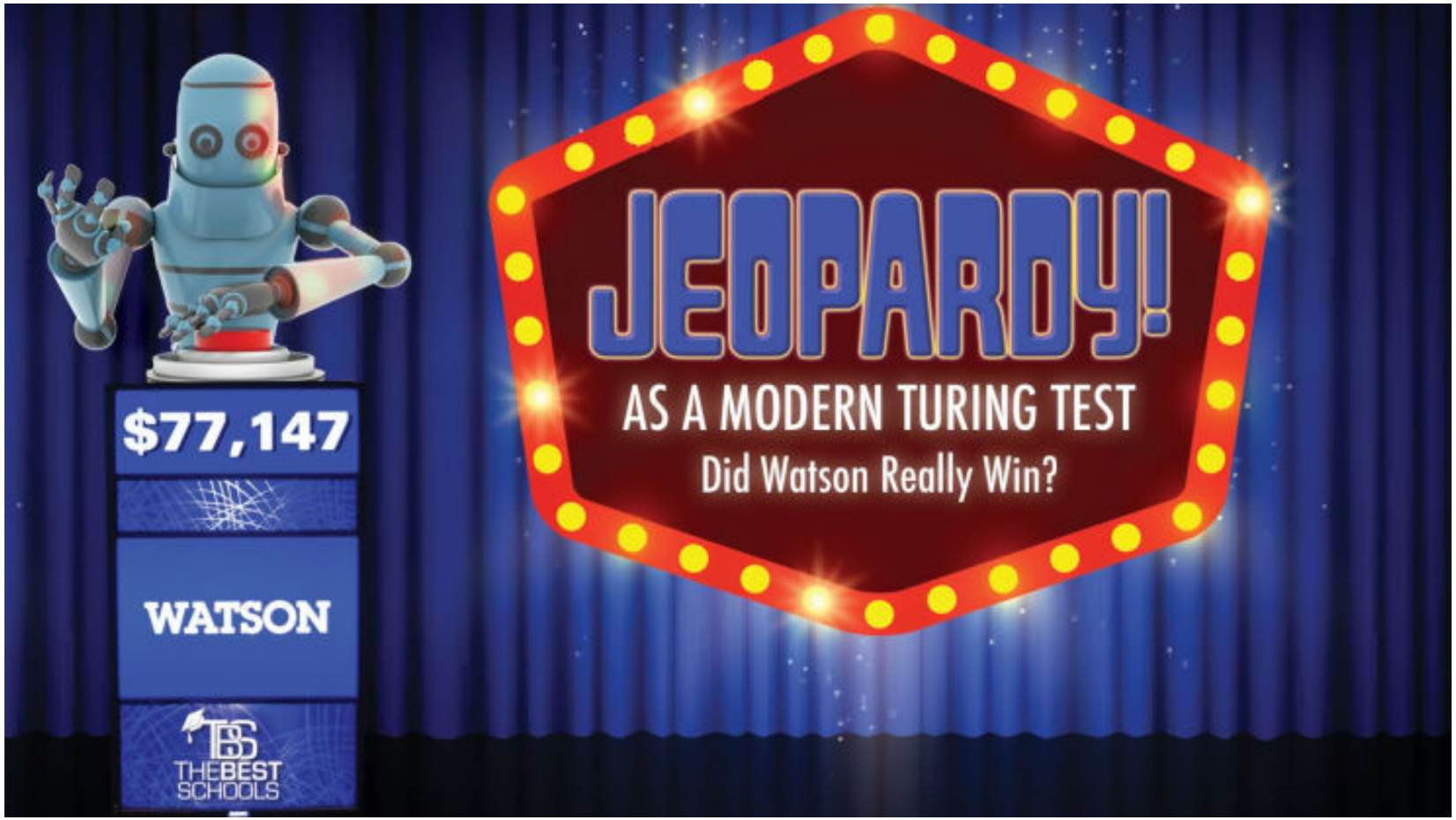
1995: Sebastian Thrun's Autonomous Vehicle wins the DARPA Grand Challenge

- Autonomously across Mojave Desert for 131 miles
- Thrun later develops autonomous vehicles for Google
- Thrun later starts Coursera

May 11 1997: Deep Blue (IBM defeats Gary Kasparov)

Deep Blue at the Computer History Museum

February 16, 2011: IBM's Watson "wins" Jeopardy



- IBM's Watson won just a single game

October 14, 2011: Apple Introduces Siri

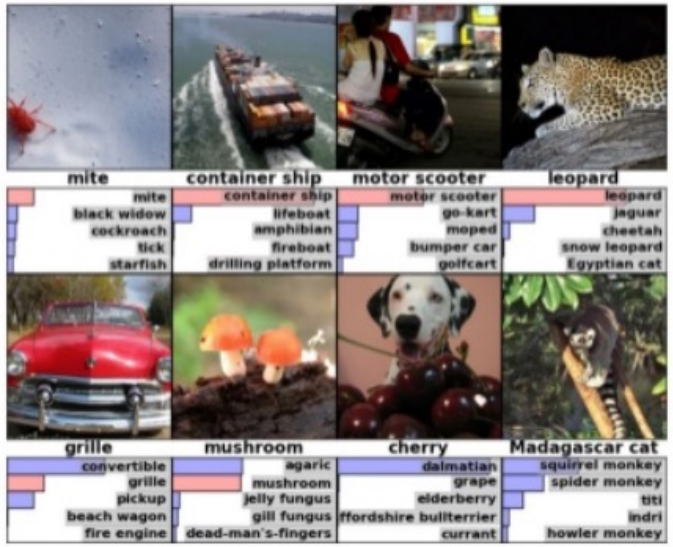
2012 ImageNet: Computer-vision contest

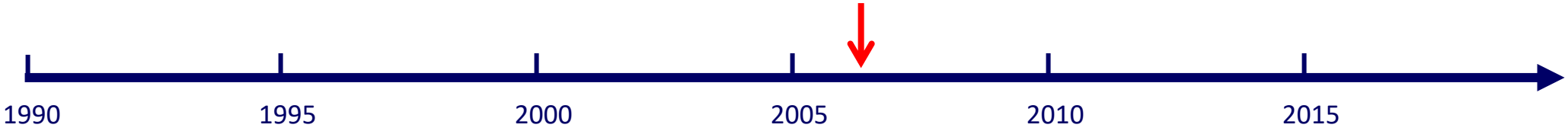
- Fei-Fei Li (Stanford): launched ImageNet in 2007, assembling a free database of more than 14 million labeled images
- Competition went live in 2009
  - 1010: an annual contest to incentivize and publish computer-vision breakthroughs
- Hinton and Students won competition by a large margin in 2012 using deep learning

ImageNet Challenge

IMAGENET

- 1,000 object classes (categories).
- Images:
  - 1.2 M train
  - 100k test.





L. Fei-Fei, ImageNet: crowdsourcing, benchmarking & other cool things, *CMU VASC Seminar, March, 2010*.

2012: Merck Molecular Activity Challenge

kaggle

Merck: Revolutionizing R&D for Safe, Effective Medicines

Modern computing has revolutionized the pharmaceutical industry, making it easier than ever before to discover new medicines that are both safe and effective. The Merck Molecular Activity Challenge offered $40,000 for Kaggle's data science community to outperform medicine discovery techniques (called QSAR models) used by pharmaceutical companies.

MERCK

- Recent computer-aided design competitions were won with deep learning
- Hinton's team (University of Toronto) won using Deep Belief Networks

2008: Toxicity challenge data as an example of real-world QSAR data

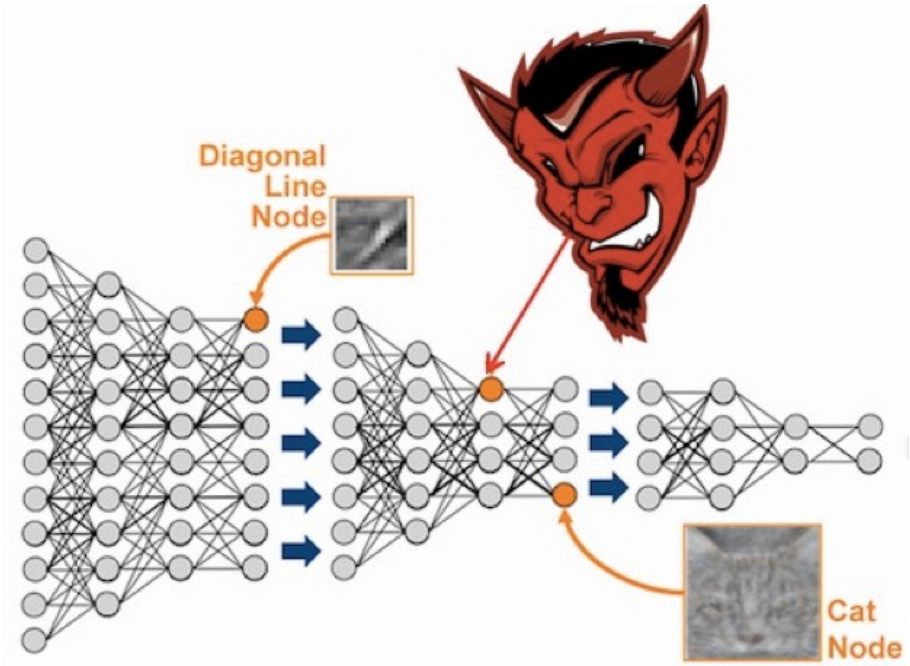
- 1093 training data and 120 test data
- 2223 descriptors

Rank	Username		Diagram	RMSE on blind test set	RMSE on known test set	Status
1	Dr Bunsen Honeydew	Final Winner		0.741	0.353	
2	embrechts	First Pass Winner		0.742	0.367	
3	oobrezanova	Final Winner		0.756	0.41	
4	Lajkonik	First Pass Winner		0.76	0.292	
5	MJE	First Pass Winner		0.765	0.395	
6	mb	First Pass Winner		0.778	0.288	

[1] Hao Zhu, Alexander Tropsha, Denis Fourches, Alexandre Varnek, Ester Papa Paola Gramatica, Tomas Oberg, Phuong Dao, Artem Cherkasov, and Igor V. Tetko [2008] Combinatorial QSAR Modeling of Chemical Toxicants Tested against Tetrahymena pyriformis. Journal of Chemical Information and Modeling, Vol. 48 pp. 766-784

[2] Igor V. Tetko, Iurii Sushko, Anil Kumar Pandey, Hao Zhu, Alexander Tropsha, Ester Papa, Tomas Oberg, Robrto Todeschini, Denis Fourches, and Alexandre Varnek [2008] Critical Assessment of QSAR Models of Environmental Toxicity against Tetrahymena pyriformis: Focusing on Applicability Domain of Overfitting by Variable Selection. Journal of Chemical Information and Modeling, Vol. 48, pp. 1733 - 1746

2012: NY Times Article on Deep Learning

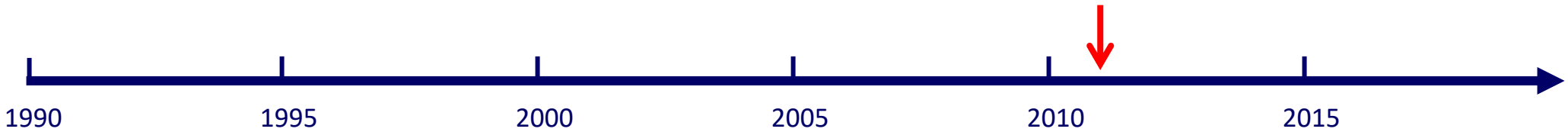


NATURE | INSIGHT | REVIEW

Deep learning

Yann LeCun, Yoshua Bengio & Geoffrey Hinton

Nature **521**, 436–444 (28 May 2015) doi:10.1038/nature14539
Received 25 February 2015 Accepted 01 May 2015 Published online 27 May 2015



SCIENCE

Scientists See Promise in Deep-Learning Programs

By JOHN MARKOFF NOV. 23, 2012

A voice recognition program translated a speech given by Richard F. Rashid, Microsoft's top scientist, into Mandarin Chinese. Hao Zhang/The New York Times

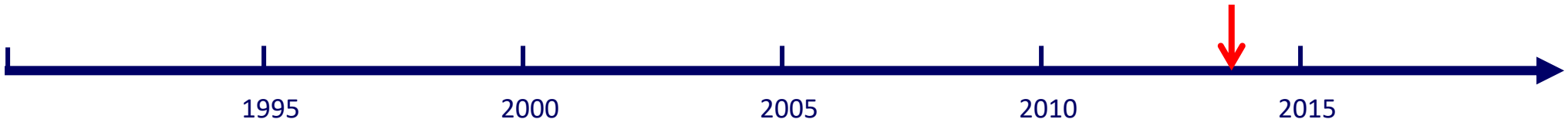
Using an artificial intelligence technique inspired by theories about how

2019 Turing Award in Artificial intelligence (1M Prize)

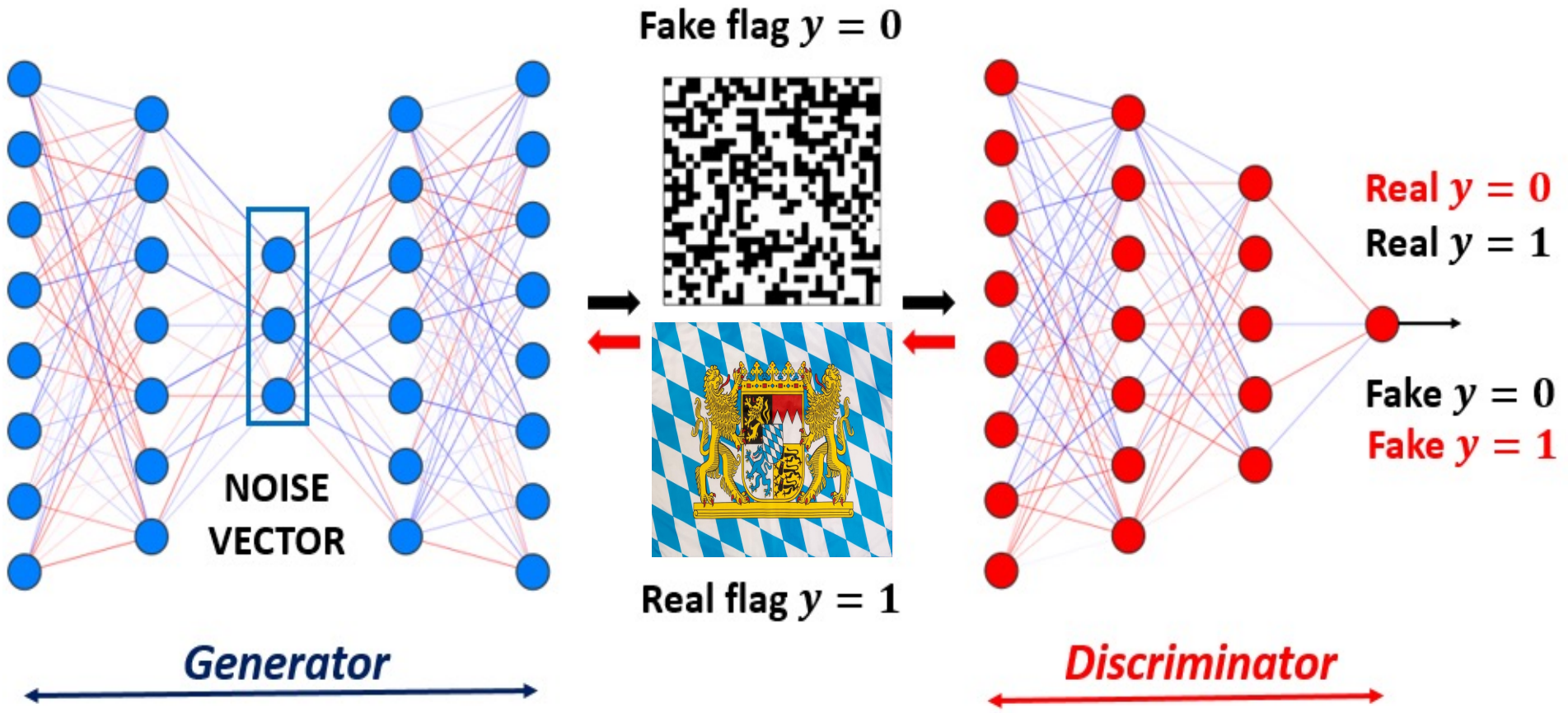
Drs. LeCun and Bengio in 2017 with Dr. Hinton, who created a research program dedicated to "neural computation and adaptive perception" in 2004. Re•Work

Cade Metz [2019] Turing Award won by 3 pioneers in artificial intelligence, New York Times, March 27, 2019.

2014: GAN (Generative Adversarial Neural Network)



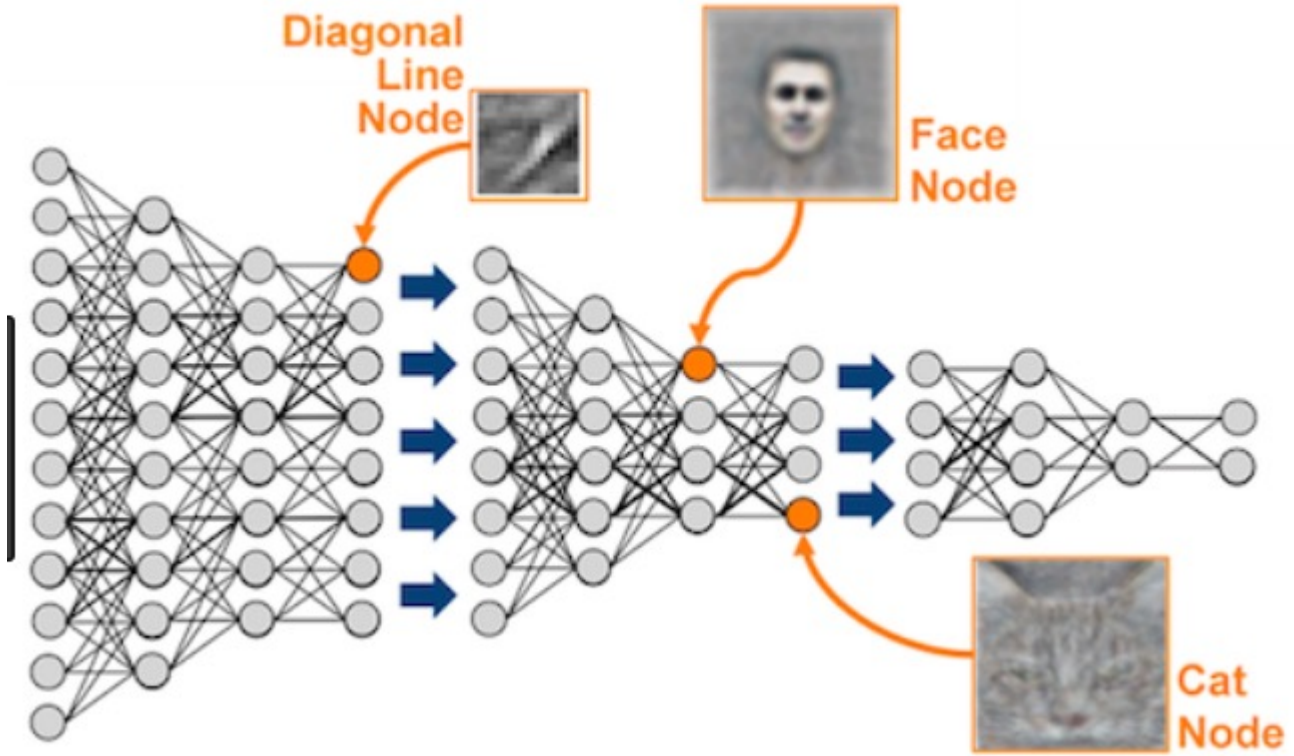
2014: Generative Adversarial Nets (GANs)



Schematic for training a General Adversarial Neural Network with backpropagation. As a sample for a real flag image, the flag of Albania is shown.

Goodfellow, I. J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A Courvolle, and Y. Bengio [2014] Generative adversarial nets. Proceedings of Advanced in Neural Information Processing Systems, pp. 2672-2680.

2015: Google Tensorflow for deep learning

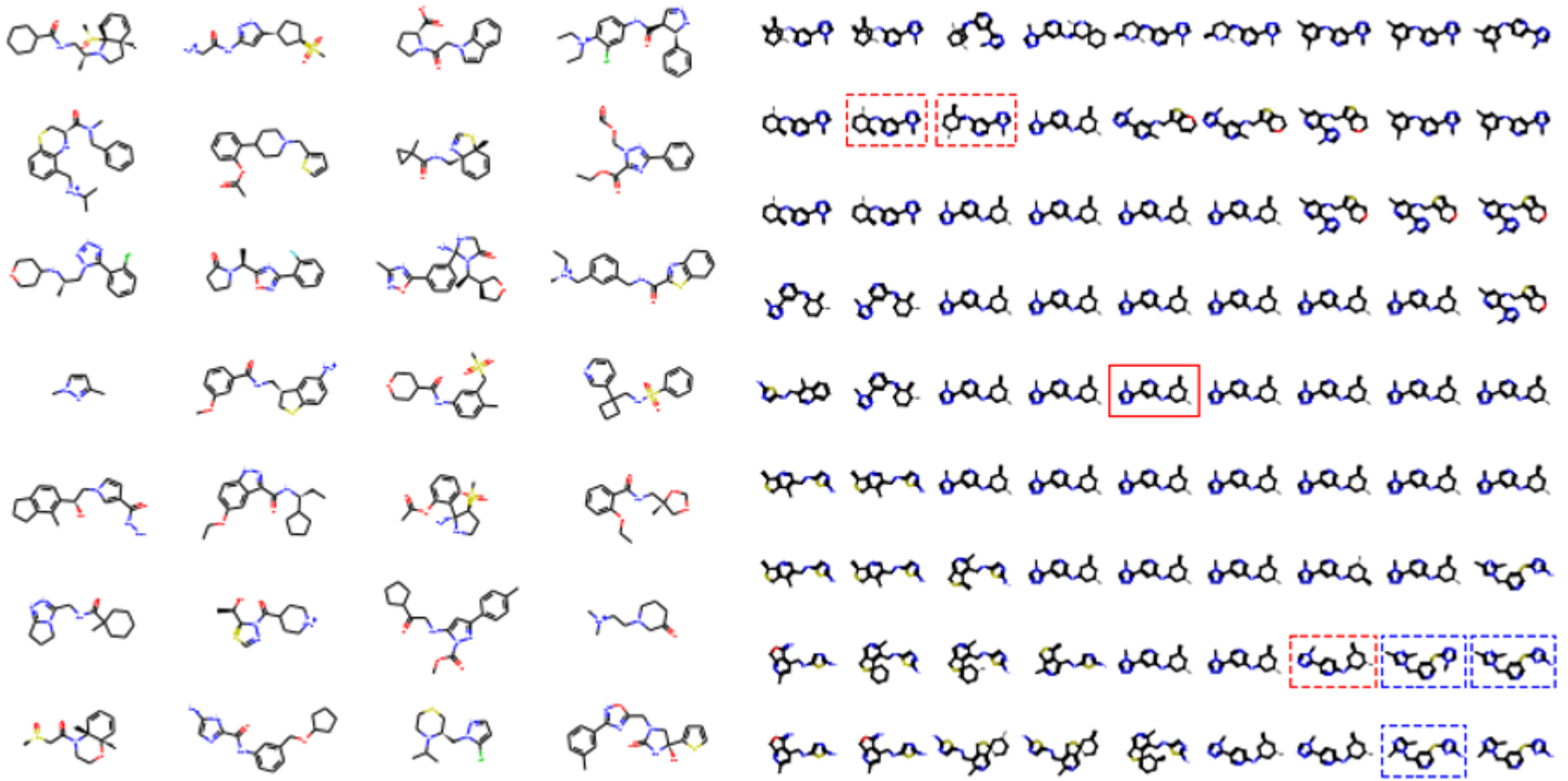


Jeff Dean

Born	1968 (age 49–50)
Residence	United States
Nationality	American
Alma mater	University of Minnesota B.S. Computer Science and Economics (1990) University of Washington, Ph.D. Computer Science (1996);
Known for	MapReduce, Bigtable, Spanner
	Scientific career
Fields	Computer Technology
Institutions	Google; Digital Equipment Corporation
Thesis	*Whole-program optimization of object-oriented languages* (1996)
Doctoral advisor	Craig Chambers

Martin Abadi, Jeffrey Dean, and Rajat Monga et al. [2015] TensorFlow: Large-Scale Machine Learning on Hetergeneous Distributed Systems, Preliminary White paper, Google, November 9, 2015.

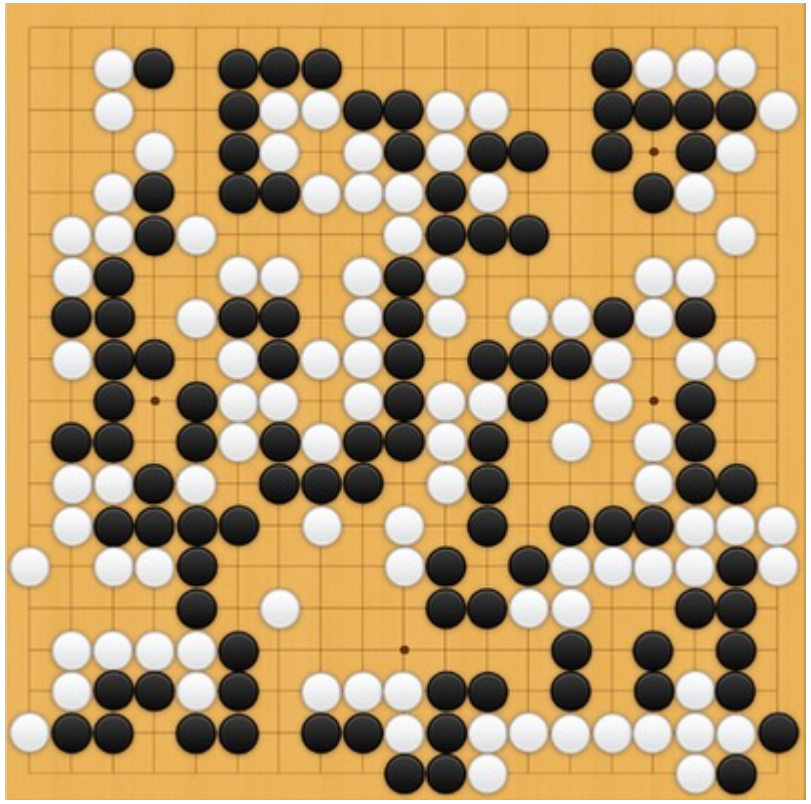
2018 Generative AI: In-Silico Drug Design with Variational Auto-Encoders (VAEs)



Random molecules sampled from a prior distribution $\mathcal{N}(\mathbf{0}, \boldsymbol{I})$ (lhs). Visualization of the local neighborhood of a molecule in the center (rhs.). Three molecules highlighted in red dashed box have the same tree structure as the center molecule, but with different graph structure as their clusters are combined differently. The same phenomenon emerges in another group of molecules (blue dashed box).

Wengong Jin, Regina Barzilay, Tommi Jaakkola [2018] Junction tree Variational Autoencoder for molecular graph generation. *Proceedings of the 35th International Conference on Machine Learning*, Stockholm, Sweden, PMLR 80, 2018.

2016: Google's Alpha-Go defeats Chinese Go champion, Ke Jie



- Neural networks based reinforcement learning

Reinforcement learning

nature International weekly journal of science

Search Go

Advanced search

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For Authors

News & Comment > News > 2018 > April > Article

NATURE | NEWS

E-alert RSS Facebook Twitter

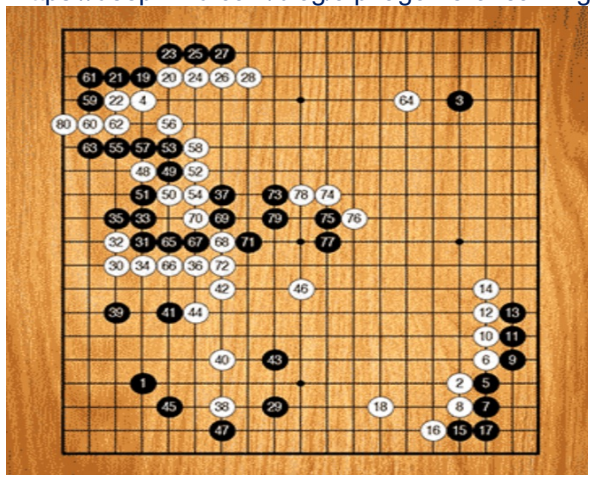
Self-taught AI is best yet at strategy game Go

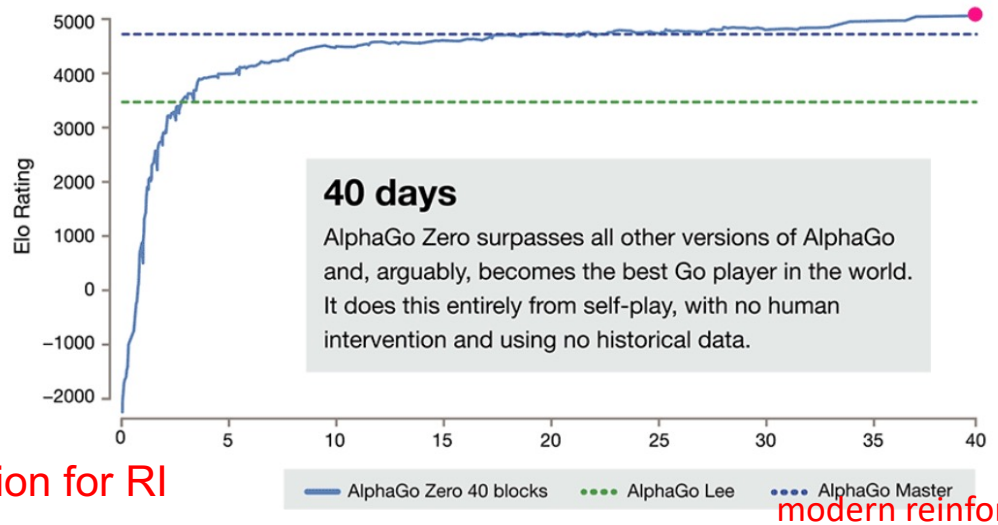
Artificial-intelligence program AlphaGo Zero trained in just days, without any human input.

Elizabeth Gibney

18 October 2017

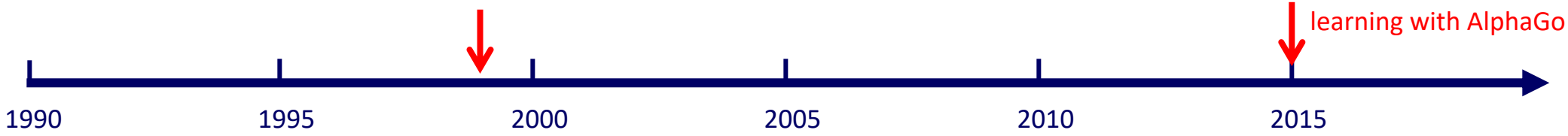
https://deepmind.com/blog/alphago-zero-learning-scratch/





Wiering: ANNs to approximate Bellman equation for RI

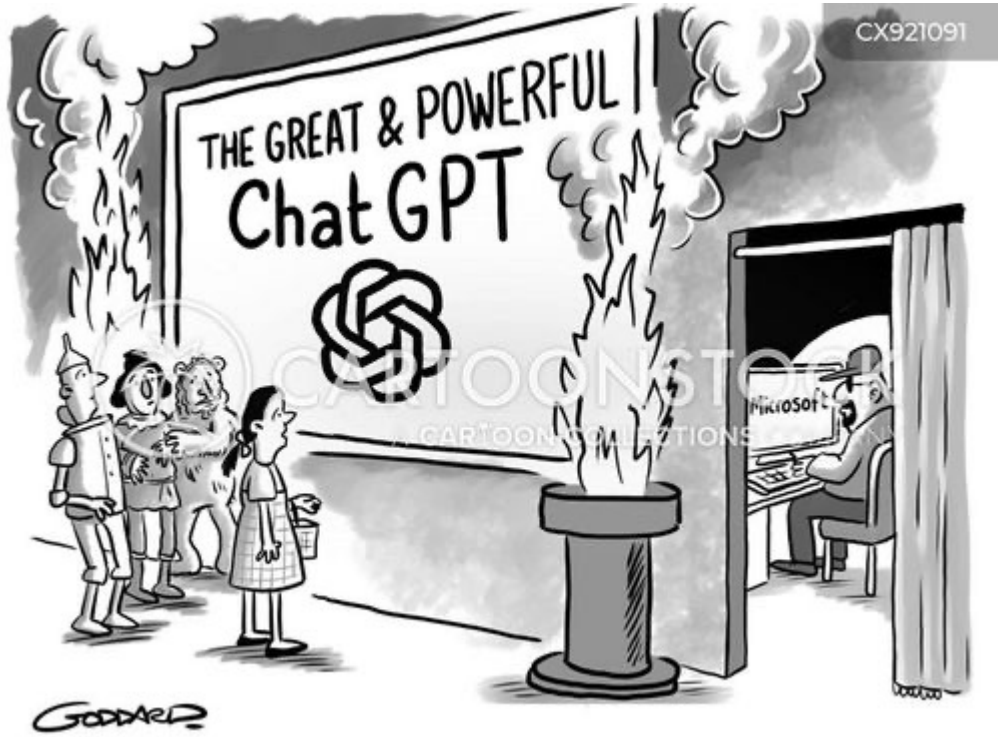
modern reinforcement learning with AlphaGo



Marco Wiering [1995] TD Learning of Game Evaluation Functions with Hierarchical Neural Architectures. Master's Thesis, of Computer Systems Faculty of Mathematics and Computer Science University of Amsterdam

November 1992: Open AI Launches ChatGTP

- ChatGPT works through its Generative Pre-trained Transformer
- Neural network based



What’s next in AI?

- Explainable AI (XAI)
- General AI (GAI)
- Quantum AI (QAI)

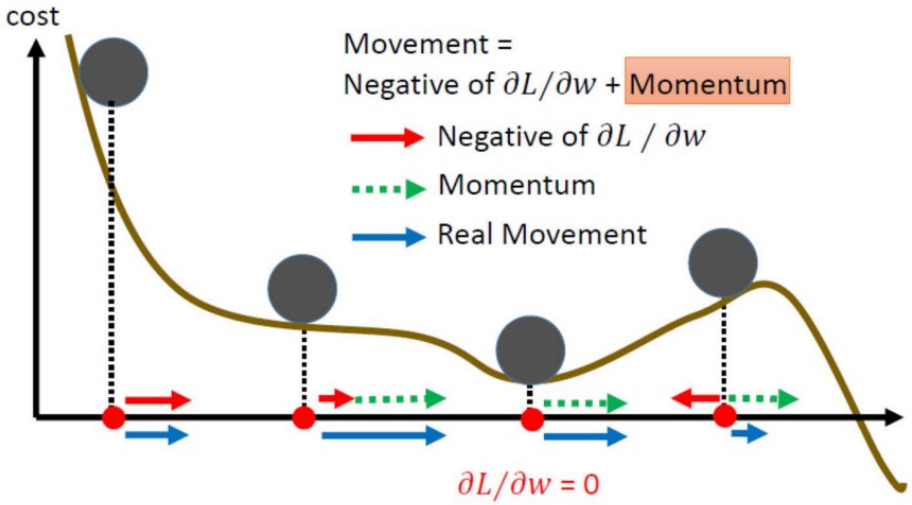
Reserves

2015 AdaM: Adaptive Moment estimation

PARAMETER SETTINGS:
η -- stepsize (i.e., learning parameter η 0.0001)
β_1 -- exponential decay rate gradient (0.9)
β_2 -- exponential decay rate 2nd moment (0.999)

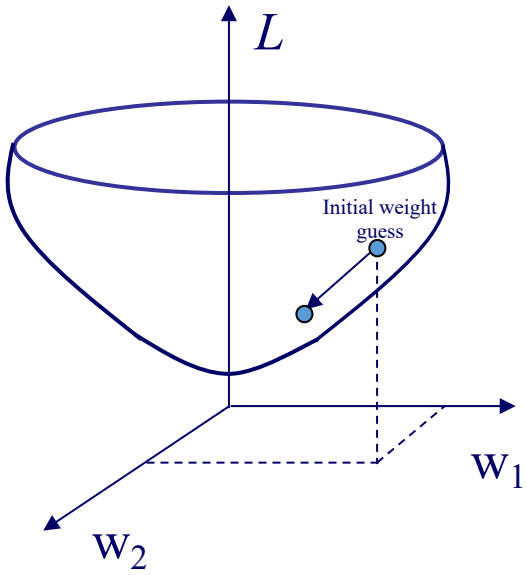
INITIALIZATION:
$\boldsymbol{m}_0 \leftarrow 0$ (*gradient tensor*)
$\boldsymbol{v}_0 \leftarrow 0$ (*2nd moment tensor*)
$\boldsymbol{w}_0 \leftarrow 0$ (*weight tensor*)

Momentum



UPDATING RULE:

$$\boldsymbol{g_t} \leftarrow \nabla L_t(\boldsymbol{w_t})$$
$$\boldsymbol{m_t} \leftarrow \boldsymbol{\beta_1 m_{t-1} + (1-\beta_1) g_t}$$
$$\boldsymbol{v_t} \leftarrow \boldsymbol{\beta_2 v_{t-1} + (1-\beta_2) g_t^2}$$
$$\hat{\boldsymbol{m}}_t \leftarrow \boldsymbol{m_t} / (1-\boldsymbol{\beta_1^t})$$
$$\hat{\boldsymbol{v}}_t \leftarrow \boldsymbol{v_t} / (1-\boldsymbol{\beta_2^t})$$
$$\boldsymbol{w_t} \leftarrow \boldsymbol{w_{t-1}} - \boldsymbol{\eta}\, \hat{\boldsymbol{m}}_t / \sqrt{\hat{\boldsymbol{v}}_t + \boldsymbol{\varepsilon}}$$



$$\Delta_t = \boldsymbol{\alpha} \cdot \hat{\boldsymbol{m}}_t / \sqrt{\hat{\boldsymbol{v}}_t} + \boldsymbol{\epsilon}$$

$|\Delta_t| \lessapprox \alpha.$

Trust-region

$$\hat{\boldsymbol{m}}_t / \sqrt{\hat{\boldsymbol{v}}_t} \approx \boldsymbol{E[g_t]} / \sqrt{\boldsymbol{E[g_t^2]}} \leq \mathbf{1}$$

Diederick P. Kingman, and Jimmy Lei Ba [2015] Adam: A method for stochastic optimization. Proceedings of the 3rd International Conference for Learning Representations ICLR15, San Diego, 2015.

September 29, 2016

IBM Advances Neuromorphic Computing for Deep Learning

John Russell

IBM TrueNorth Platform

Deep learning efforts today are run on standard computer hardware using convolutional neural networks. Indeed the approach has proven powerful by pioneers such as Google and Microsoft. In contrast neuromorphic computing, whose spiking neuron architecture more closely mimics human brain function, has generated less enthusiasm in the deep learning community. Now, work by IBM using its TrueNorth chip as a test case may bring deep learning to neuromorphic architectures.

Writing in the Proceedings of the National Academy of Science (PNAS) in August (Convolutional networks for fast, energy-efficient neuromorphic computing), researchers from IBM Research report, "[We] demonstrate that neuromorphic computing, despite its novel architectural primitives, can implement deep convolution networks that approach state-of-the-art classification accuracy across eight standard datasets encompassing vision and speech, perform inference while preserving the hardware's underlying energy-efficiency and high throughput."

The impact could be significant as neuromorphic hardware and software technology have been rapidly advancing on several fronts. IBM researchers ran the datasets at between 1,200 and 2,600 frames/s and using between 25 and 275 mW (effectively >6,000 frames/s per watt). They report their approach allowed networks to be specified and trained using backpropagation with the same ease-of-use as contemporary deep learning. Basically, the new approach allows the algorithmic power of deep learning to be merged with the efficiency of neuromorphic processors.

"The new milestone provides a palpable proof of concept that the efficiency of brain-inspired computing can be merged with the effectiveness of deep learning, paving the path towards a new generation of chips and algorithms with even greater efficiency and effectiveness," said Dharmendra Modha, chief scientist for brain-inspired computing at IBM Research-Almaden, in an interesting article by Jeremy Hsu on the IBM work posted this week on the IEEE Spectrum (IBM's Brain-Inspired Chip Tested for Deep Learning.)

2001: Xin and Embrechts Backpropagation for neuromorphic computing

April 19, 2018: Michael Jordan – Artificial Intelligence – The revolution hasn't happened yet

"Thus, just as humans built buildings and bridges before there was civil engineering, humans are proceeding with the building of societal-scale, inference-and-decision-making systems that involve machines, humans and the environment. Just as early buildings and bridges sometimes fell to the ground — in unforeseen ways and with tragic consequences — many of our early societal-scale inference-and-decision-making systems are already exposing serious conceptual flaws."

"This confluence of ideas and technology trends has been rebranded as "AI" over the past few years. This rebranding is worthy of some scrutiny."

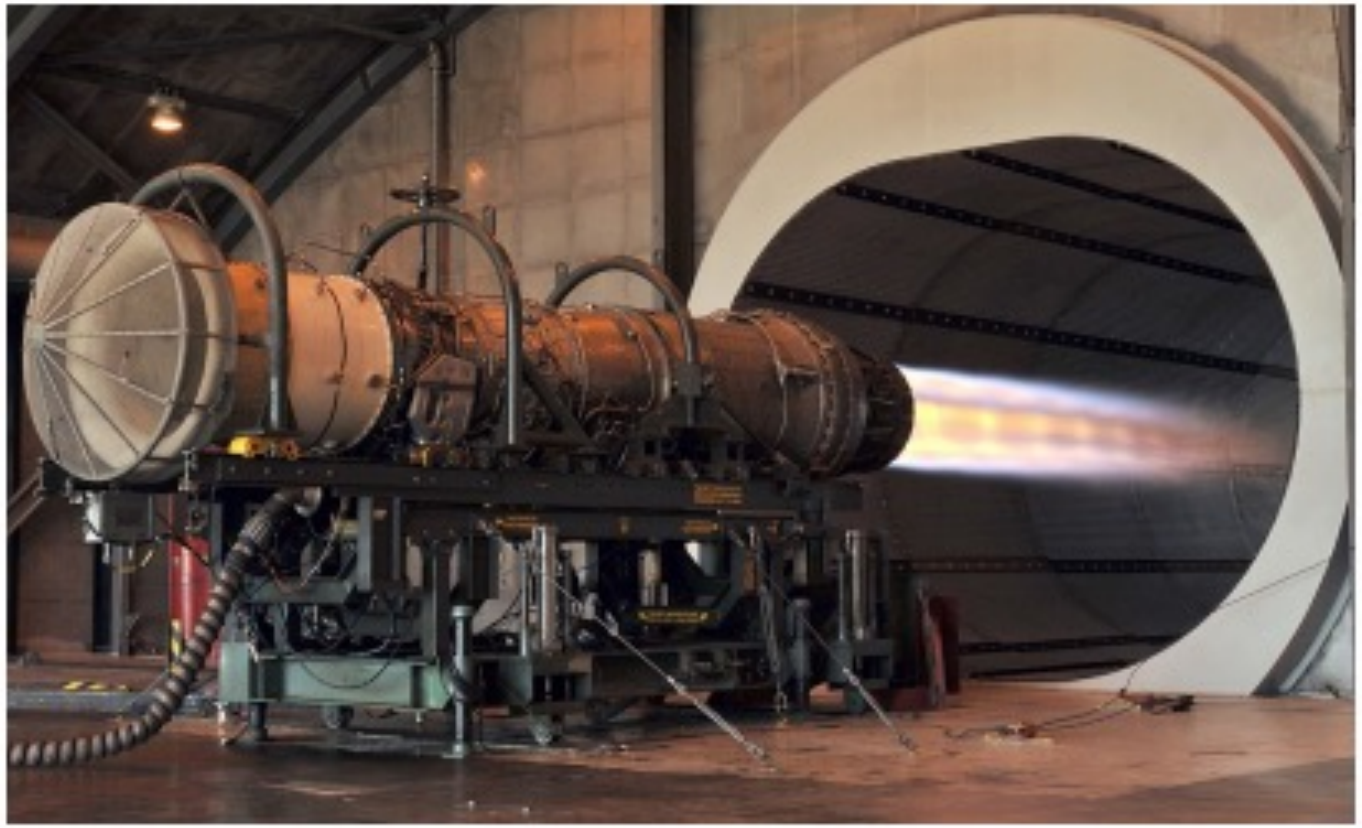
"Sixty years later, however, high-level reasoning and thought remain elusive. The developments which are now being called "AI" arose mostly in the engineering fields associated with low-level pattern recognition and movement control, and in the field of statistics — the discipline focused on finding patterns in data and on making well-founded predictions, tests of hypotheses and decisions."

https://medium.com/@mijordan3/artificial-intelligence-the-revolution-hasnt-happened-yet-5e1d5812e1e7

2019: LeCun, Hinton and Bengio recognized in Turing award for AI

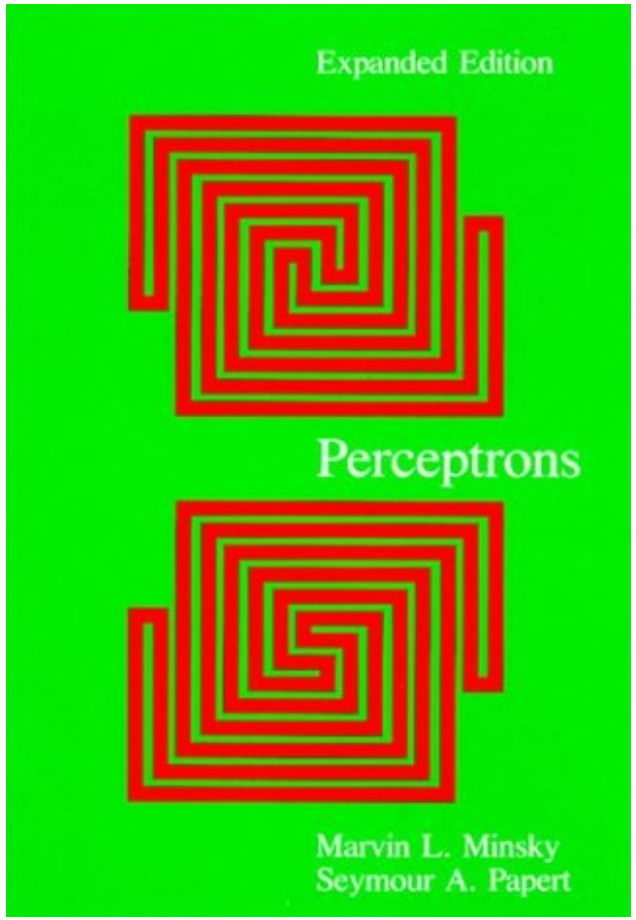
“Yoshua Bengio, Geoffrey Hinton, and Yann LeCun laid the foundations for modern AI”

Examples of data-driven engineering



Examples of data-driven engineering:

1. GE uses genetic algorithms to determine the number and optimal location of turbine blades on gas engines, leading to a higher efficiency (left).
2. The Swiss Reinsurance building in London was designed applying genetic algorithms to mitigate the effects of the wind on the building (right).



Marvin Minsky and Seymour Papert (MIT) pointed out need for hidden layer(s): e.g., the XOR problem is not linearly separable

Boolean XOR

x1	x2	y
0	0	0
0	1	1
1	0	1
1	1	0