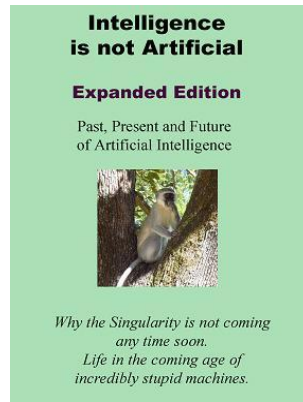




A Timeline of Artificial Intelligence

by [piero scaruffi](http://piero.scaruffi.com) | www.scaruffi.com

All of these events are explained in my book "[Intelligence is not Artificial](#)".

TM, ®, Copyright © 1996-2019 Piero Scaruffi except pictures. All rights reserved.

1960: Henry Kelley and Arthur Bryson invent backpropagation

1960: Donald Michie's reinforcement-learning system MENACE



1960: Hilary Putnam's Computational Functionalism ("Minds and Machines")

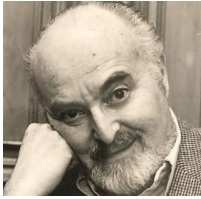


1960: The backpropagation algorithm

[Bryson 1961](#), [Kelley 60](#), [Dreyfus 62](#), [Linnainmaa 1970](#), [Werbos](#)



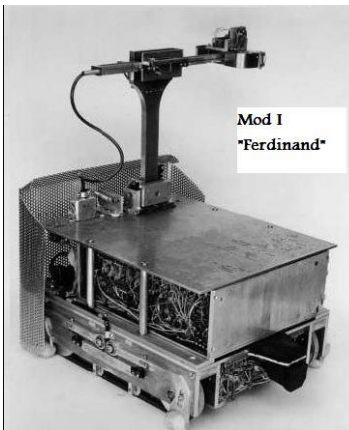
1961: Melvin Maron's "Automatic Indexing"



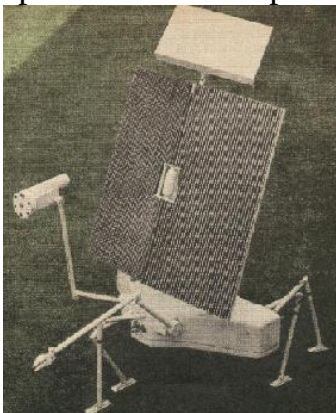
1961: Karl Steinbuch's neural network Lernmatrix



1961: Leonard Scheer's and John Chubbuck's Mod I (1962) and Mod II (1964)



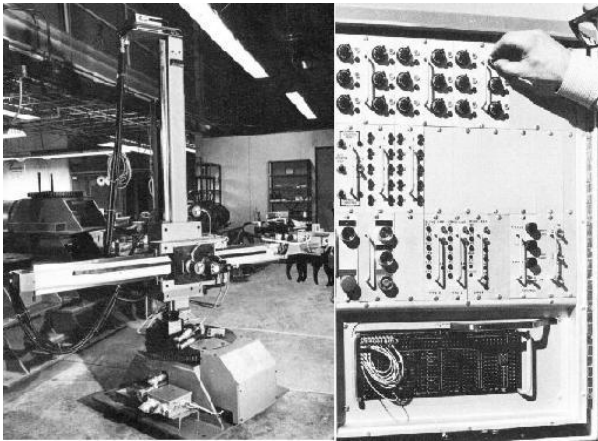
1961: Space General Corporation's lunar explorer



1962: IBM's "Shoebox" for speech recognition



1962: AMF's "VersaTran" robot

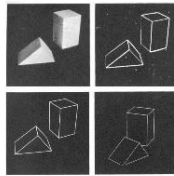


1963: John McCarthy moves to Stanford and founds the Stanford Artificial Intelligence Laboratory (SAIL)



1963: Lawrence Roberts' "Machine Perception of Three Dimensional Solids", the birth of computer vision

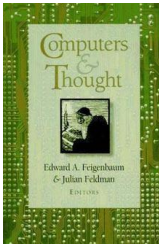
Machine Perception Of Three-Dimensional Solids
Massachusetts Institute of Technology
Lincoln Laboratory
22 May 1963
Lawrence G. Roberts



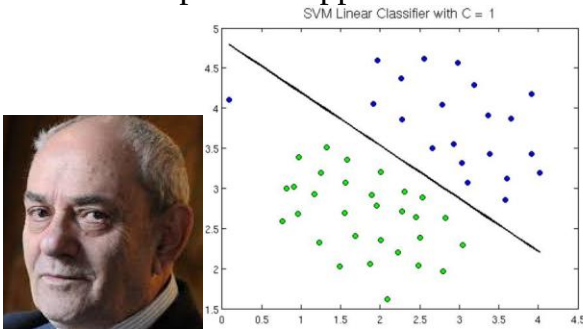
1963: Jim Slagle writes a program for symbolic integration (calculus)



1963: Edward Feigenbaum's and Julian Feldman's "Computers and Thought"



1963: Vladimir Vapnik's "support-vector networks" (SVN)



1964: Peter Toma demonstrates the machine-translation system Systran



1965: Irving John Good (Isidore Jacob Gudak) speculates about "ultraintelligent machines" (the "singularity")

Irving John Good – 1960s

- The ultraintelligent machine



"A machine that can far surpass the intellectual activities of any man however clever. ... an ultraintelligent machine could design even better machines ... The intelligence of man would be left far behind."

1965: The Case Institute of Technology builds the first computer-controlled robotic arm



1965: Ed Feigenbaum's Dendral expert system



ELEMENTS OF AN INFORMATION PROCESSING THEORY OF MEMORY*

Edward A. Feigenbaum**

September 1, 1964

American Psychological Association Annual Meeting

**University of California, Berkeley

Interim Report to the
National Aeronautics and Space Administration
Grant NsG 81-60

DENDRAL-64

A SYSTEM FOR COMPUTER CONSTRUCTION, ENUMERATION AND NOTATION OF
ORGANIC MOLECULES AS TREE STRUCTURES AND CYCLIC GRAPHS

submitted by

Joshua Lederberg
Professor of Genetics
School of Medicine
Stanford University
Palo Alto, California

PART I. December 15, 1964

1965: Gordon Moore's Law of exponential progress in integrated circuits
("Cramming more components into integrated circuits", 1965)



1965: Herbert Simon predicts that "Machines will be capable, within 20 years, of doing any work a man can do"

1965: Hubert Dreyfus's "Alchemy and Artificial Intelligence"

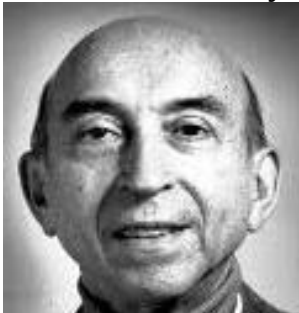
ALCHEMY AND ARTIFICIAL INTELLIGENCE

Hubert L. Dreyfus*

The RAND Corporation, Santa Monica, California



1965: Lotfi Zadeh's Fuzzy Logic



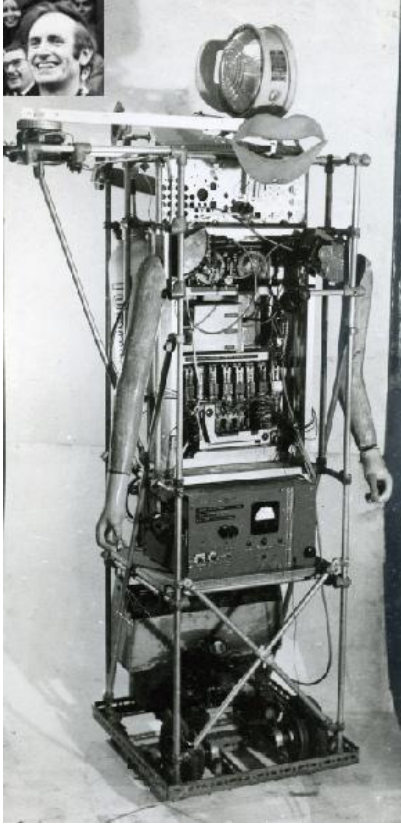
A fuzzy set (class) A in X is characterized by a membership (characteristic) function $f_A(x)$ which associates with each point¹ in X a real number in the interval $[0, 1]$, with the value of $f_A(x)$ at x representing the "grade of membership" of x in A . Thus, the nearer the value of $f_A(x)$ to unity, the higher the grade of membership of x in A .

(Lotfi Zadeh)

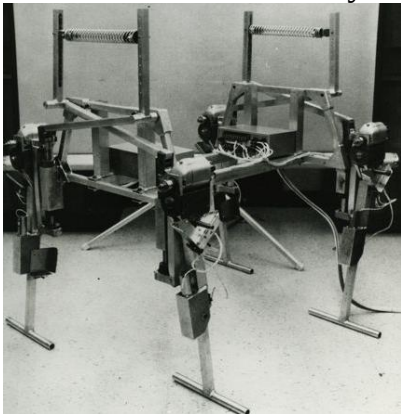
1965: Alexey Ivakhnenko publishes the first learning algorithms for multi-layered networks



1965: Bruce Lacey's robot Rosa Bosom at the Cybernetic Serendipity exhibition of computer art



1966: Robert McGhee's "Phony Pony" four-legged robot



1966: Leonard Baum popularizes the Hidden Markov Model ("Statistical Inference for Probabilistic Functions of Finite State Markov Chains")

LANGUAGE AND MACHINES

COMPUTERS IN TRANSLATION AND LINGUISTICS

A Report by the
Automatic Language Processing Advisory Committee

John R. Pierce, Bell Telephone Laboratories, Chairman
John B. Carroll, Harvard University
Eric P. Hamp, University of Chicago*
David G. Hays, The RAND Corporation
Charles F. Hockett, Cornell University †
Anthony G. Oettinger, Harvard University
Alan Perlis, Carnegie Institute of Technology

Publication 1416

National Academy of Sciences National Research Council

Washington, D. C. 1966

1967: Charles Fillmore's Case Frame Grammar

REPORT RESUMES

ED 019 631
THE CASE FOR CASE.
BY- FILLMORE, CHARLES J.

AL 000 413

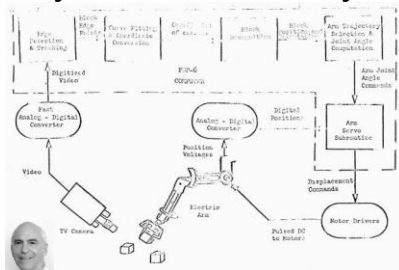
PUB DATE APR 67

IN THIS PAPER, PREPARED FOR THE APRIL 1967 TEXAS SYMPOSIUM ON LINGUISTIC UNIVERSALS, IT IS PROPOSED THAT THE



1968: Glenn Shafer's and Stuart Dempster's "Theory of Evidence"

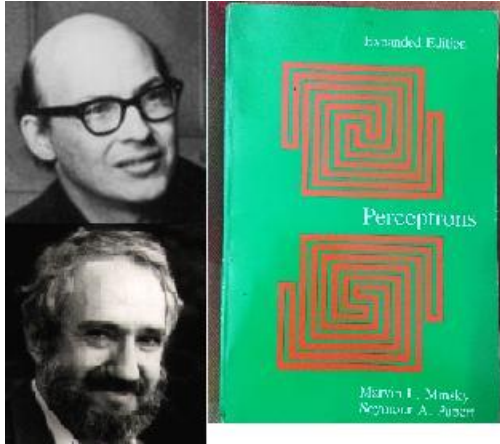
1968: Jerry Feldman's Hand-eye system



1969: Christopher Longuet-Higgins' associative memory

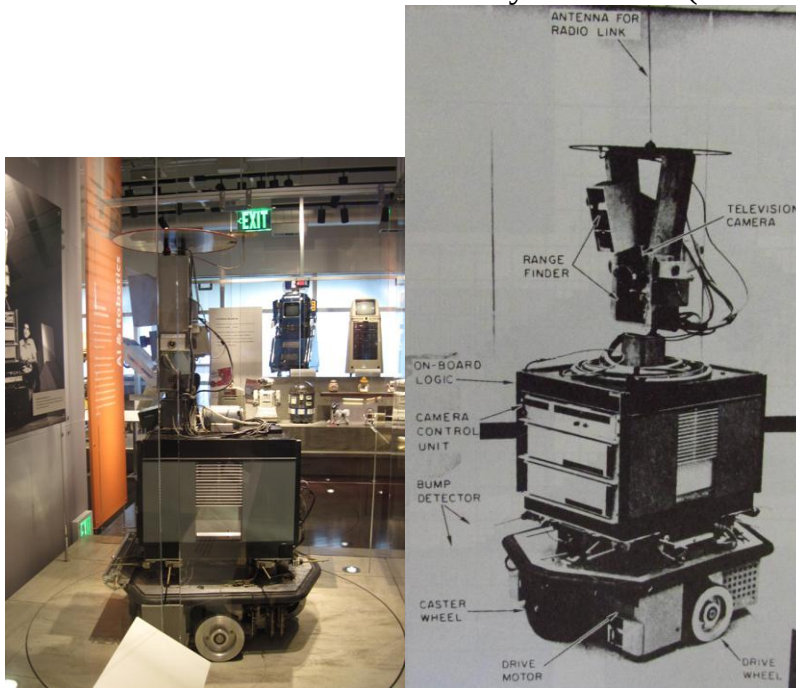


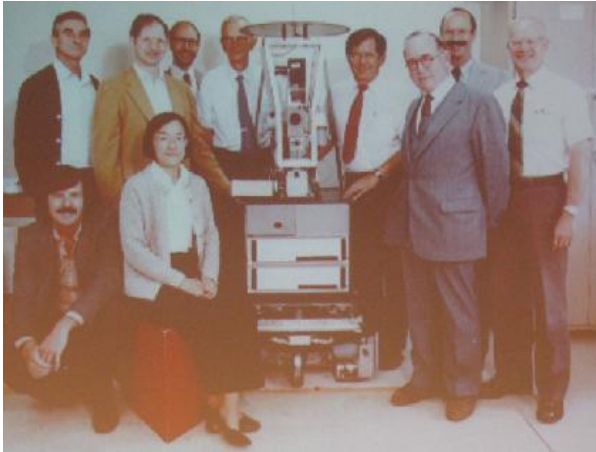
1969: Marvin Minsky & Samuel Papert's "Perceptrons" kill neural networks



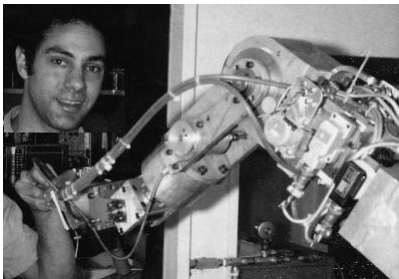
1969: First International Joint Conference on Artificial Intelligence (IJCAI) at Stanford

1969: Stanford Research Institute's Shakey the Robot (Nils Nilsson and others)





1969: Victor Scheinman's "Stanford arm"



1969: Roger Schank's Conceptual Dependency Theory for natural language processing

ATRANS	Transfer of an abstract relationship (i.e. give)
PTRANS	Transfer of the physical location of an object (e.g., go)
PROPEL	Application of physical force to an object (e.g. push)
MOVE	Movement of a body part by its owner (e.g. kick)
GRASP	Grasping of an object by an action (e.g. throw)
INGEST	Ingesting of an object by an animal (e.g. eat)
EXPEL	(e.g. cry)
MTRANS	Transfer of mental information (e.g. tell)
MBUILD	Building new information out of old (e.g. decide)
SPEAK	Producing of sounds (e.g. say)
ATTEND	(e.g. listen)



1969: Cordell Green's automatic synthesis of programs



1969: John McCarthy's "Some Philosophical Problems from the Standpoint of Artificial Intelligence" and the frame problem

1970: Albert Uttley's Informon for adaptive pattern recognition

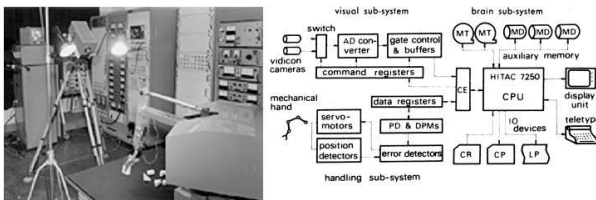


1970: Hitachi demonstrates a robot that can build objects based on drawings

© 1972 IEEE. Reprinted, with permission, from IEEE Transaction on Computers, Vol.C-21, No.2, pp.161-170, February 1972

A Prototype Intelligent Robot that Assembles Objects from Plan Drawings

MASAKAZU EJIRI, TAKESHI UNO, HARUO YODA, TATSUO GOTO, AND KIYOO TAKEYASU



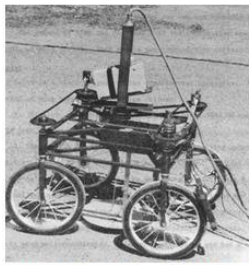
1970: Tom Martin founds Threshold Technology, the first commercial company for speech recognition



1970: William Woods' Augmented Transition Network (ATN) for natural language processing



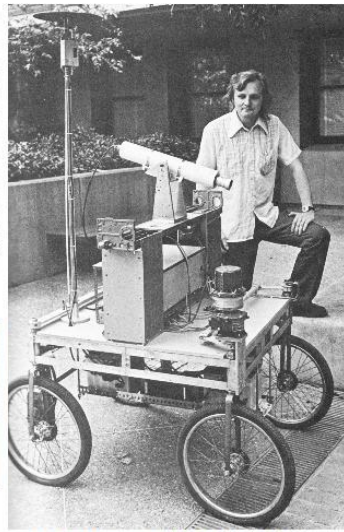
1971: The "Stanford cart" autonomous outdoors vehicle



Jim Adams' "cart" in 1961



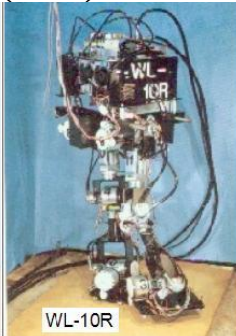
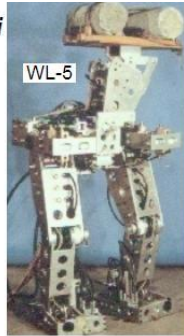
Rodney Schmidt's "cart"
in 1971



Hans Moravec's "cart" in 1977

1971: Atsuo Takanishi's WL-5 (1971) and WL-10R (1983)

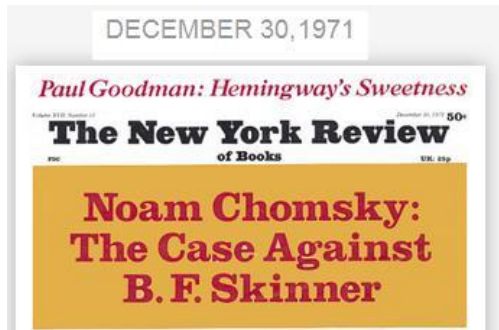
Atsuo Takanishi
Waseda University



1971: Richard Fikes' and Nils Nilsson's STRIPS planner



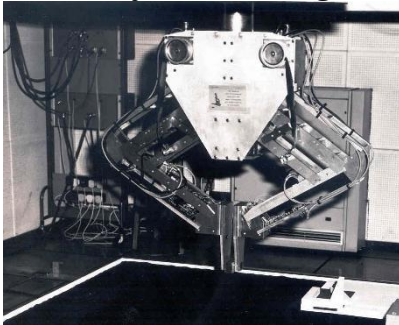
1971: Noam Chomsky's article against Burrhus Skinner's behaviorism



1971: Ingo Rechenberg publishes his thesis "Evolution Strategies", a set of optimization methods for evolutionary computation



1971: University of Edinburgh's robot Freddy



1972: Alain Colmerauer's PROLOG programming language



1972: The first chatbot to chatbot conversation ever takes place over the Arpanet between Kenneth Colby's chatbot Parry at Stanford and Eliza at MIT

ELIZA: Good Evening. Tell me your problems. Please terminate input with a period or a question mark.

PARRY: People get on my nerves sometimes.

ELIZA: I am not sure I understand you fully.

PARRY: You should pay more attention.

ELIZA: Suppose you should pay more attention.

PARRY: You're entitled to your own opinion.

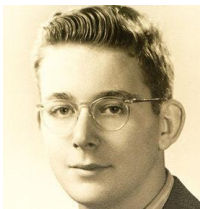
ELIZA: What makes you think I am entitled to my own opinion?

PARRY: It bothers me just to be around people in general.



Kenneth Colby

1972: Richard Karp shows there are many problems that can probably only be solved in exponential time



1972: Harry Klopff's "Brain Function and Adaptive Systems"

AFCL-72-0164
3 MARCH 1972
SPECIAL REPORTS, NO. 133

AIR FORCE CAMBRIDGE RESEARCH LABORATORIES
L. O. HANSCOM FIELD, BEDFORD, MASSACHUSETTS



Brain Function and Adaptive Systems -

A. HARRY KLOPF

NEURON $\equiv$ MAN .

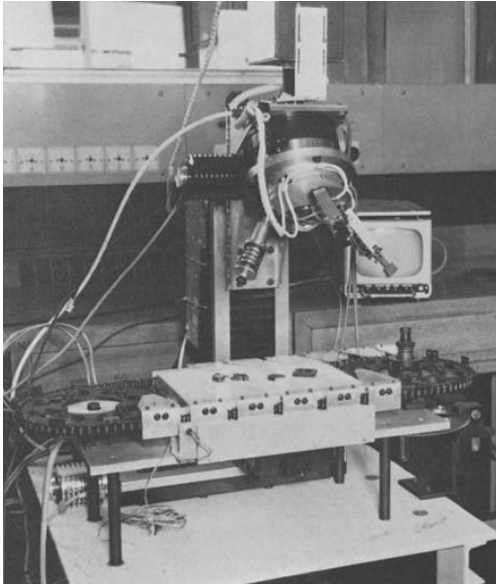
MAN $\equiv$ HEDONIST' .

NEURON $\equiv$ HEDONIST' .

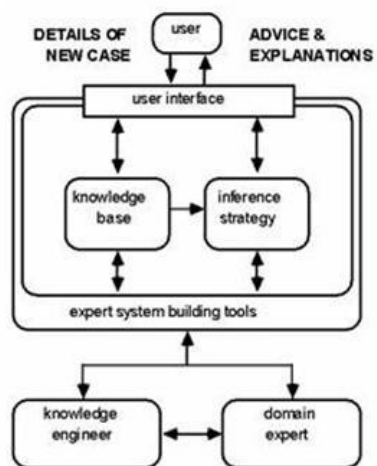
1972: William Woods' question-answering system LUNAR



1972: The SIRCH robotic arm at Nottingham University



1972: Bruce Buchanan's MYCIN

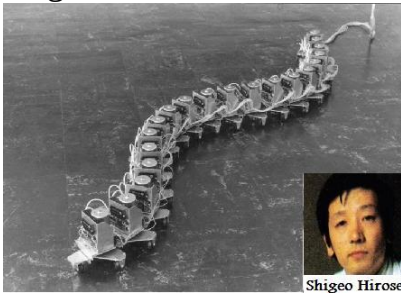


1972: Terry Winograd's Shrdlu

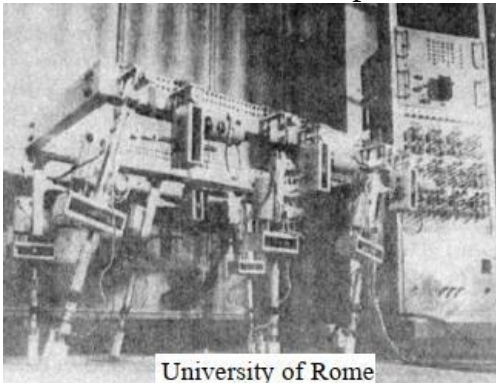


Person: PICK UP A BIG RED BLOCK.
 Computer: OK. (does it)
 Person: GRASP THE PYRAMID.
 Computer: I DON'T UNDERSTAND WHICH PYRAMID YOU MEAN.
 Person: FIND A BLOCK WHICH IS TALLER THAN THE ONE YOU ARE HOLDING AND PUT IT INTO THE BOX.
 Computer: BY "IT", I ASSUME YOU MEAN THE BLOCK WHICH IS TALLER THAN THE ONE I AM HOLDING.
 Computer: OK. (does it)
 Person: WHAT DOES THE BOX CONTAIN?
 Computer: THE BLUE PYRAMID AND THE BLUE BLOCK.
 Person: WHAT IS THE PYRAMID SUPPORTED BY?
 Computer: THE BOX.
 Person: HOW MANY BLOCKS ARE NOT IN THE BOX?
 Computer: FOUR OF THEM.
 Person: IS AT LEAST ONE OF THEM HARBROWER THAN THE ONE WHICH I TOLD YOU TO PICK UP?
 Computer: YES, THE RED CUBE.
 Person: IS IT SUPPORTED?
 Computer: YES, BY THE TABLE.
 Person: CAN THE TABLE PICK UP BLOCKS?
 Computer: NO.
 Person: CAN A PYRAMID BE SUPPORTED BY A BLOCK?
 Computer: YES.
 Person: CAN A PYRAMID SUPPORT A PYRAMID?
 Computer: I DON'T KNOW.
 Person: STACK UP TWO PYRAMIDS.
 Computer: I CAN'T.
 Person: WILL YOU PLEASE STACK UP BOTH OF THE RED BLOCKS AND EITHER A GREEN CUBE OR A PYRAMID?
 Computer: OK. (does it)
 Person: WHICH CUBE IS SITTING ON THE TABLE?

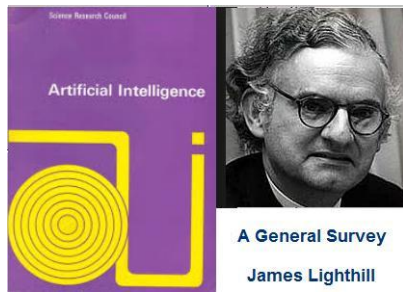
1972: Shigeo Hirose's snake-robot ACM III



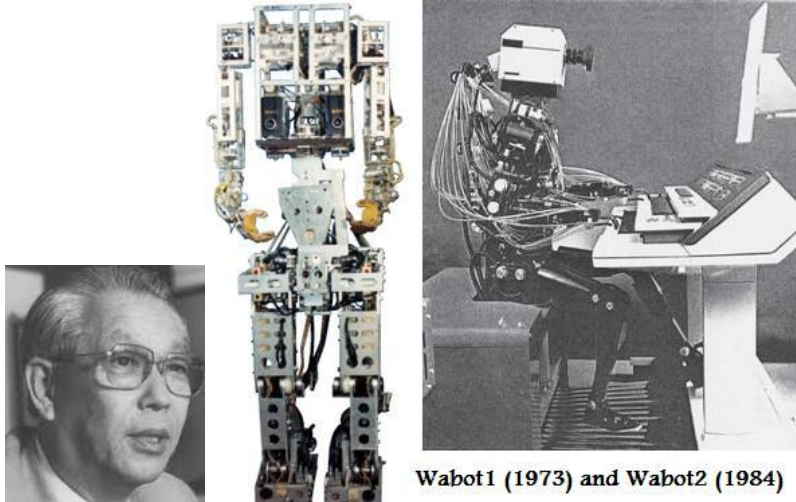
1972: Petternella-Salinari hexapod robot



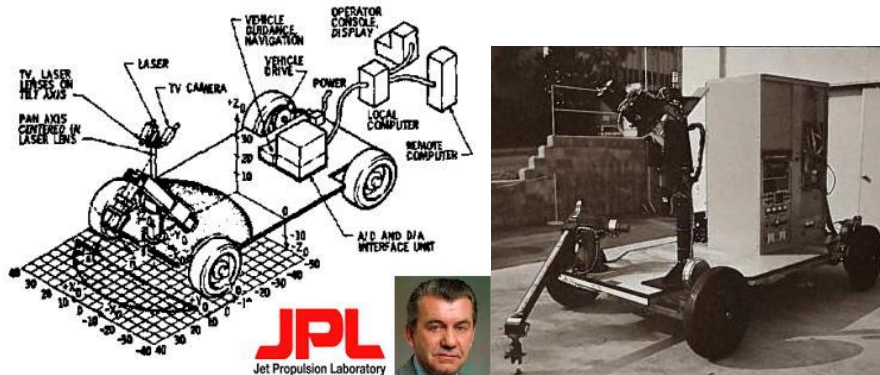
1973: "Artificial Intelligence: A General Survey" by James Lighthill criticizes Artificial Intelligence for over-promising



1973: Ichiro Kato's Wabot, the first real-size anthropomorphic walking robot



1973: Antal Bejczy's JPL Rover



1973: Jim Baker applies the Hidden Markov Model to speech recognition ("Machine-aided Labeling of Connected Speech")



1974: Marvin Minsky's frame



A Framework for Representing Knowledge

Marvin Minsky

MIT-AI Laboratory Memo 306, June, 1974.

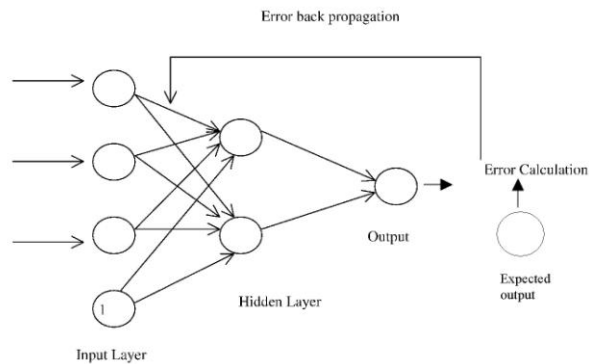
1974: Paul Werbos' backpropagation algorithm for neural networks



BEYOND REGRESSION:
NEW TOOLS FOR PREDICTION AND ANALYSIS
IN THE BEHAVIORAL SCIENCES

Harvard University
Cambridge, Massachusetts

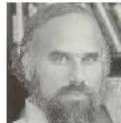
August, 1974



1975: Roger Schank's script

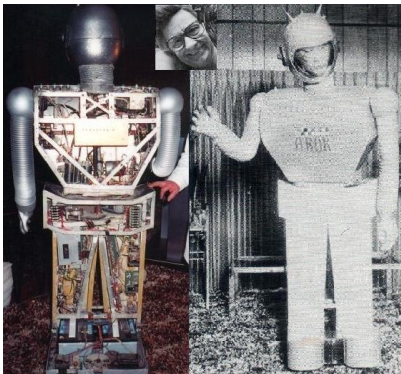
Restaurant Script (Schank & Abelson, 1975)

Scene 1 Entering
PTRANS self into restaurant
ATTEND eyes to empty tables
MBUILD place to sit
PTRANS self to table
MOVE sit down
Scene 2 Ordering
ATrans Get menu
MTRANS Read menu
MBUILD Decide what self wants
MTRANS Order to waitress/waiter
Scene 3 Eating
ATrans Get food
INGEST Eat food
Scene 4 Exiting
MTRANS Ask for check
ATrans Get check
ATrans Tip waitress/waiter
PTRANS Self to cashier
ATrans Money to cashier
PTRANS Self out of restaurant



Go Back

1975: Ben Skora's robot Arok



1975: Raj Reddy's team at Carnegie Mellon University develops three speech-recognition systems (Bruce Lowerre's Harpy, Hearsay-II and Jim Baker's Dragon)



1975: Hearsay-II's blackboard model by Rick Hayes-Roth, Lee Erman, Victor Lesser and Richard Fennell



1975: The first Artificial Intelligence in Medicine workshop at Rutgers University

1975: John Holland's genetic algorithms



1976: Richard Laing's paradigm of self-replication by self-inspection

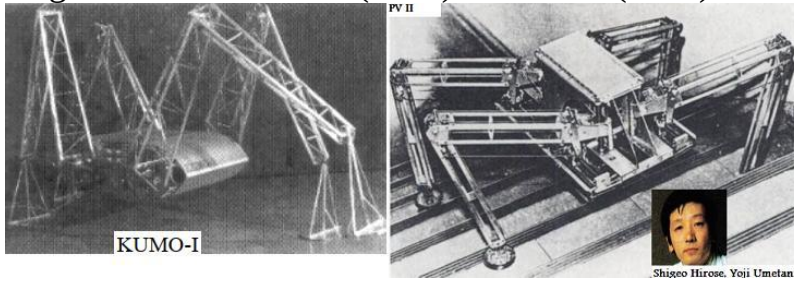
1976: Stephen Grossberg's Adaptive Resonance Theory (ART) for unsupervised learning



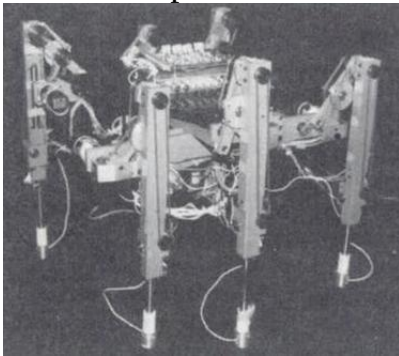
1976: Fred Jelinek's "Continuous Speech Recognition by Statistical Methods"



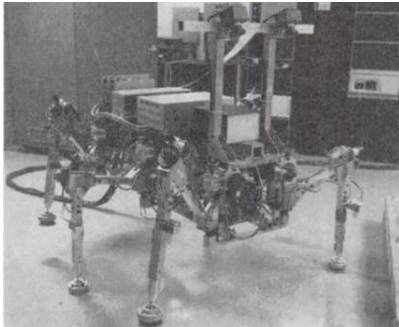
1976: Shigeo Hirose's Kumo-I (1976) and PV-II (1978)



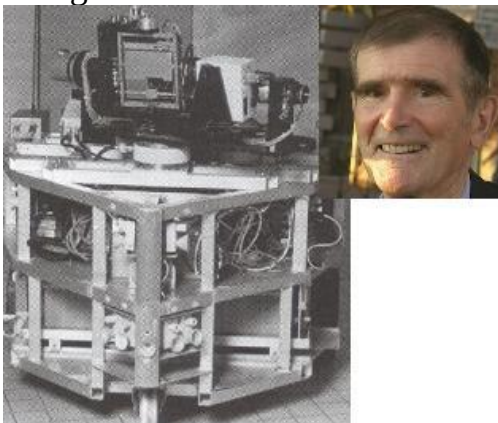
1976: Masha hexapod



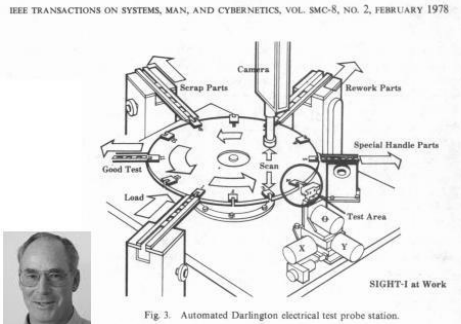
1977: Robert McGhee's Bionic Bug



1977: Georges Giralt's Hilaré robot in France



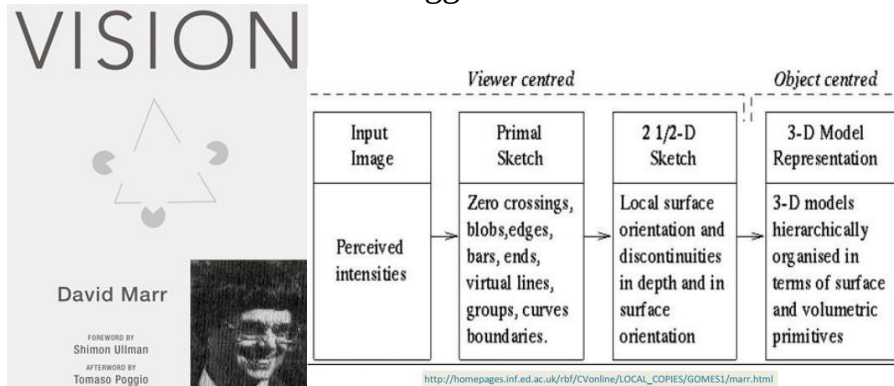
1977: General Motors' computer-vision system Sight-I



1977: Ian Witten's actor-critic method



1977: David Marr's and Tomaso Poggio's "2 1/2 sketch"



1978: John McDermott's expert system R1/XCON

R1: an Expert in the Computer Systems Domain¹

John McDermott
Department of Computer Science
Carnegie-Mellon University
Pittsburgh, Pennsylvania 15213



ASSIGN-UB-MODULES-EXCEPT-THOSE-CONNECTING-TO-PANELS-4

IF: THE CURRENT CONTEXT IS ASSIGNING DEVICES
TO UNIBUS MODULES
AND THERE IS AN UNASSIGNED DUAL PORT DISK DRIVE
AND THE TYPE OF CONTROLLER IT REQUIRES IS KNOWN
AND THERE ARE TWO SUCH CONTROLLERS NEITHER
OF WHICH HAS ANY DEVICES ASSIGNED TO IT
AND THE NUMBER OF DEVICES THAT THESE
CONTROLLERS CAN SUPPORT IS KNOWN

THEN: ASSIGN THE DISK DRIVE TO EACH OF THE CONTROLLERS
AND NOTE THAT THE TWO CONTROLLERS HAVE BEEN
ASSOCIATED AND THAT EACH SUPPORTS ONE DEVICE

PUT-UB-MODULE F-6

IF: THE CURRENT CONTEXT IS PUTTING UNIBUS MODULES
IN BACKPLANE IN SOME BOX
AND IT HAS BEEN DETERMINED WHICH MODULE TO TRY
TO PUT IN A BACKPLANE
AND THAT MODULE IS A MULTIPLEXER TERMINAL INTERFACE
AND IT HAS NOT BEEN ASSOCIATED WITH ANY PANEL SPACE
AND THE TYPE AND NUMBER OF BACKPLANE SLOTS
IT REQUIRES IS KNOWN
AND THERE ARE AT LEAST THAT MANY SLOTS AVAILABLE
IN A BACKPLANE OF THE APPROPRIATE TYPE
AND THE CURRENT UNIBUS LOAD ON THAT BACKPLANE
IS KNOWN
AND THE POSITION OF THE BACKPLANE IN THE BOX IS KNOWN

THEN: ENTER THE CONTEXT OF VERIFYING PANEL SPACE
FOR A MULTIPLEXER

1978: Shunichi Amari publishes neural field equations



1978: Ryszard Michalski builds the first practical system that learns from examples, AQ11



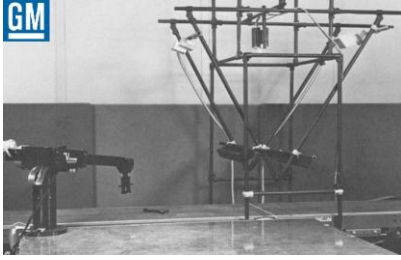
1978: Lothar Rossol organizes at General Motors a symposium on computer vision



1979: Johan DeKleer's qualitative reasoning

1979: A factory worker named Robert Williams is the first human killed by a robot

1979: General Motors' Consight robot



1979: Tokuji Okada's robotic hand with three fingers



1979: William Clancey's Guidon



1979: Hans Berliner's BKG 9.8 at Carnegie-Mellon University (connected by satellite to the robot Gammonoid) beats the world champion of backgammon in Monte Carlo



Backgammon Computer Program Beats World Champion

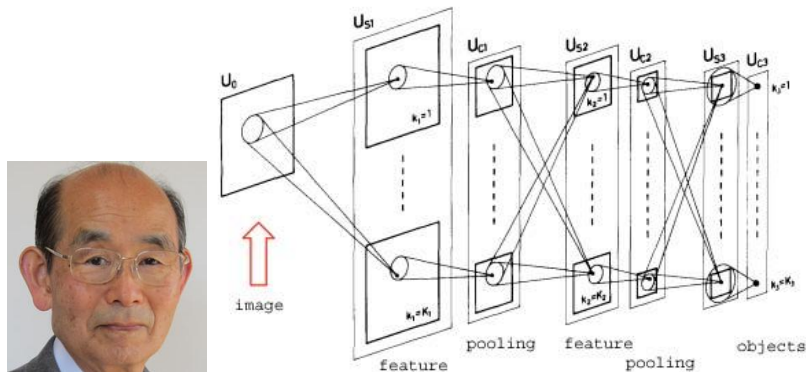
Artificial Intelligence, vol. 14 (1980), pp. 205–220.

1979: Drew McDermott's non-monotonic logic

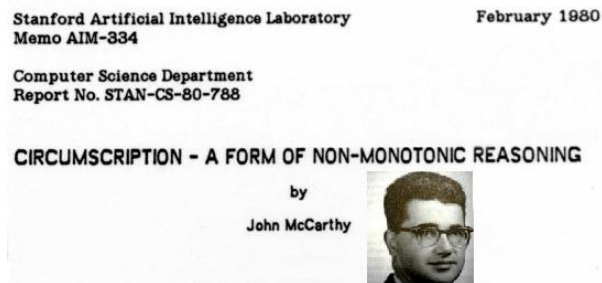


1979: Kunihiro Fukushima's convolutional neural network ("Neocognitron - A Self-

organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position")



1980: McCarthy's Circumscription



1980: John Searle's article "Minds, Brains, and Programs" on the "Chinese Room" that attacks Artificial Intelligence

THE BEHAVIORAL AND BRAIN SCIENCES (1980) 3, 417-457
Printed in the United States of America

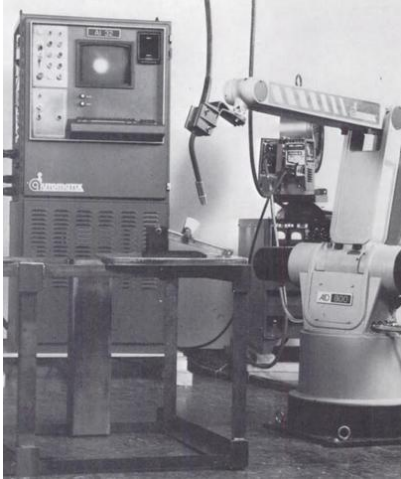
Minds, brains, and programs

John R. Searle
Department of Philosophy, University of California, Berkeley, Calif.
94720

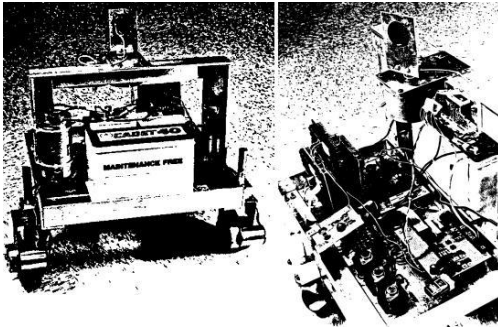


1980: IntelliGenetics (Intellicorp), the first major start-up for Artificial Intelligence

1981: Automatix introduces the first commercial robot with a vision system



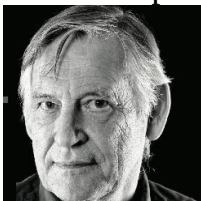
1981: Russell Andersson's robot SCIMR



1981: Danny Hillis' Connection Machine



1981: Hans Kamp's Discourse Representation Theory

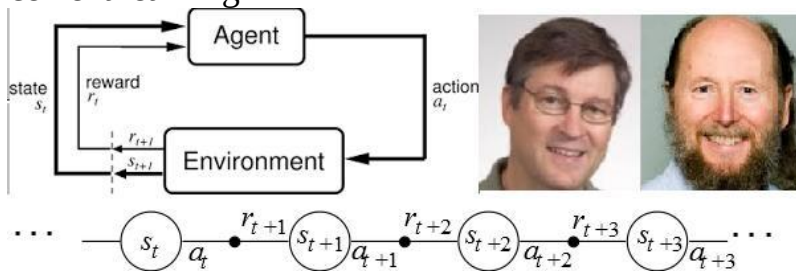


1981: Japan has 14,000 industrial robots versus the USA's 4,200 and West Germany's 2,300



Producer	Value (1,000 dollars)
Matsushita	54,167
Hitachi	33,333
Kawasaki	31,250
Yaskawa	27,917
Fanuc	25,000
Mitsubishi	23,983
Dainichi Kiko	19,167
Komatsu	16,667
Star Seiki	13,333
Kobe Steel	9,583
All other	345,058
Total	599,458

1981: Andrew Barto's and Richard Sutton's temporal-difference method of reinforcement learning



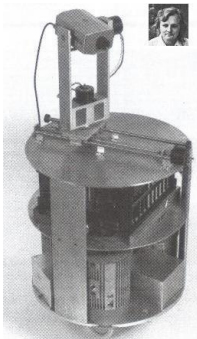
1982: The Parallel Distributed Processing (PDP) research group at UC San Diego



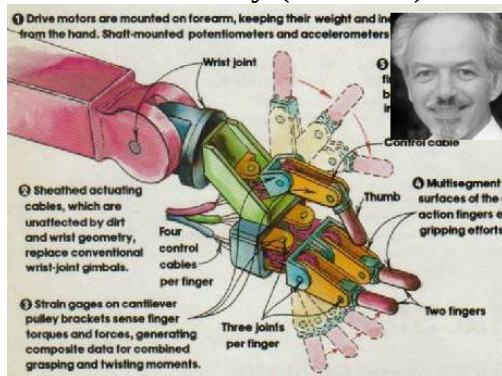
1982: Shunichi Mizuno's cybot New Monroe



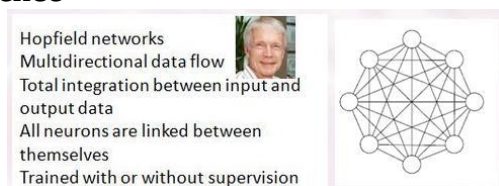
1982: Hans Moravec's CMU Rover



1982: Kenneth Salisbury (Stanford) & Jet Propulsion Laboratory's robotic hand



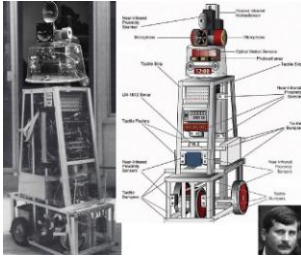
1982: John Hopfield describes a new generation of neural networks, based on recurrence



1982: The Canadian Institute for Advanced Research (CIFAR) establishes Artificial Intelligence and Robotics as its very first program



1982: Bart Everett's robot Robart I



1982: Japan's Fifth Generation Computer Systems project

Computer generations:

- 0th generation: 500 B.C., mechanical gears.
- 1st generation: 1940's, vacuum tubes.
- 2nd generation: 1950's, transistors.
- 3rd generation: 1960's, integrated circuits (ICs).
- 4th generation: Microprocessors.
- 5th generation: FGCS, Fifth Generation Computer Systems project.
- Japan's Ministry of International Trade and Industry.

1982: Teuvo Kohonen's Self-Organized Maps (SOM) for unsupervised learning



1982: Judea Pearl's "Bayesian networks"

From: AAAI-82 Proceedings. Copyright ©1982, AAAI (www.aaai.org). All rights reserved.

REVEREND BAYES ON INFERENCE ENGINES: A DISTRIBUTED
HIERARCHICAL APPROACH(*)(**)

Judea Pearl
Cognitive Systems Laboratory
School of Engineering and Applied Science
University of California, Los Angeles
90024



ABSTRACT

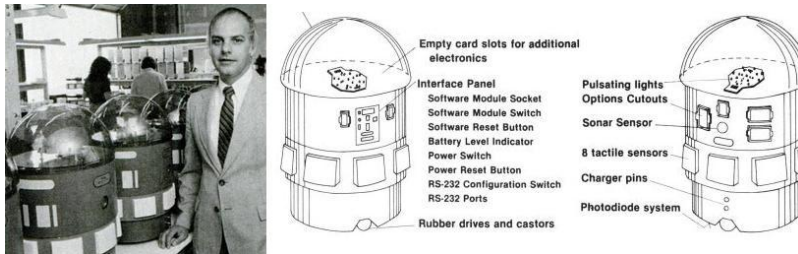
This paper presents generalizations of Bayes likelihood-ratio updating rule which facilitate an asynchronous propagation of the impacts of new beliefs and/or new evidence in hierarchically organized inference structures with multi-hypotheses variables. The computational scheme proposed

feature of hierarchical inference the relation $P(D|R)$ is computable local, more elementary probabilities involving intervening variables, (e.g., organisms causing may not be directly observable. al role, however, is to provide summarization for loosely coupled national data so that the comput

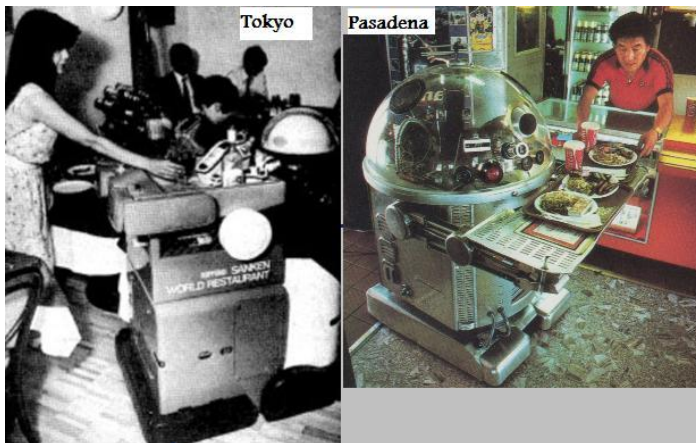
1982: David Parker rediscovers backpropagation



1982: Joseph Bosworth's personal robot RB5X



1983: Dainichi Kiko's waiter robot



1983: Scott Kirkpatrick's simulated annealing



1983: Mike Cohen's and Stephen Grossberg's continuous recurrent networks

IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS, VOL. SMC-13, NO. 5, SEPTEMBER/OCTOBER 1983

Absolute Stability of Global Pattern Formation and Parallel Memory Storage by Competitive Neural Networks

MICHAEL A. COHEN AND STEPHEN GROSSBERG

1983: Yurii Nesterov's accelerated version of gradient descent ("Nesterov momentum")



1983: John Laird and Paul Rosenbloom's SOAR



John Laird



Allen Newell

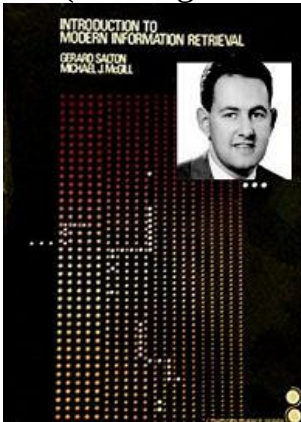


Paul Rosenbloom

1983: Geoffrey Hinton's and Terry Sejnowski's Boltzmann machine



1983: Gerard Salton and Michael McGill's "Introduction to Modern Information Retrieval" (the "bag-of-words model")



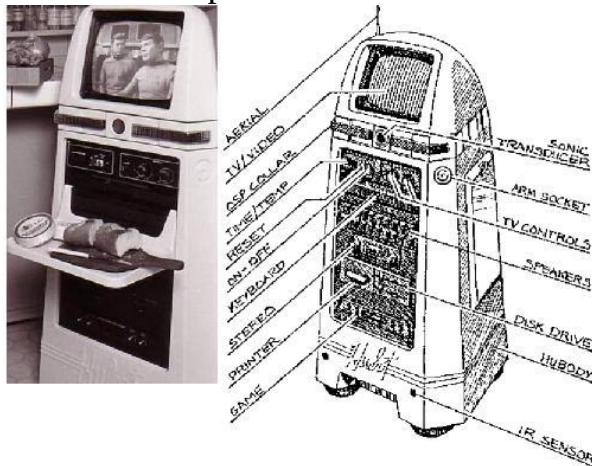
1983: Odetics' Odex



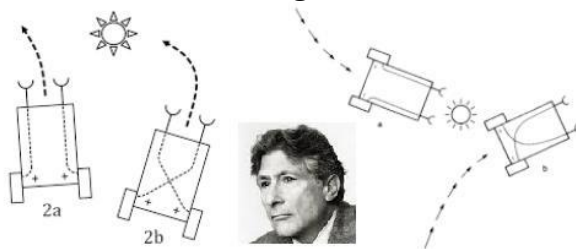
1984: Tomy's toy Omnibot 2000



1984: Mike Forino's personal robot Hubot



1984: Valentino Braitenberg's "Vehicles"



1984: Robotic hand by Stephen Jacobsen (Univ of Utah) & MIT



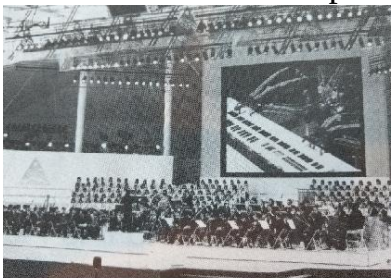
1984: Doug Lenat's "Cyc" to catalog common sense



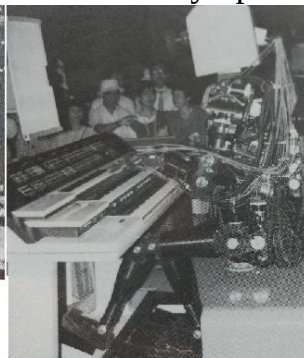
1984: Barbara Hayes-Roth's general-purpose blackboard system BB1



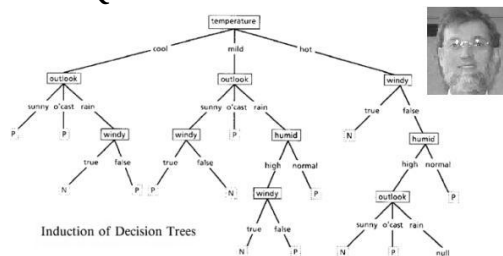
1985: Ichiro Kato's Wasubot performs with a symphony orchestra



Tsukuba, May 1985



1985: Ross Quinlan's ID3 for decision trees analysis



1985: The first international conference on genetic algorithms



1985: Piero Scaruffi opens the A.I. Center at Olivetti, the first major non-academic A.I. Center outside the USA



1985: Yann LeCun rediscovers backpropagation

COGNITIVA 85

Paris, 4-7 Juin 1985

A LEARNING SCHEME FOR ASSYMETRIC THRESHOLD NETWORK.

UNE PROCEDURE D'APPRENTISSAGE POUR RESEAU A SEUIL ASSYMETRIQUE.

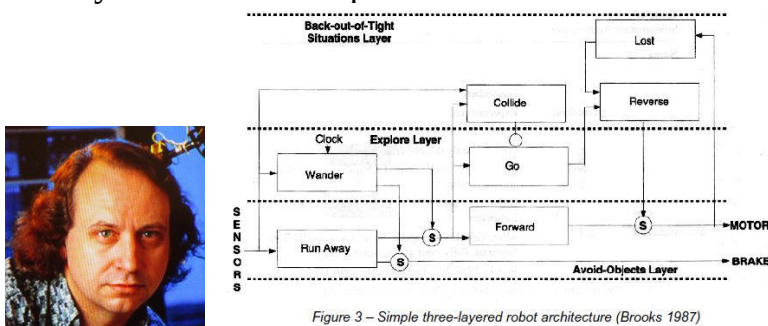
YANN LE CUN

Ecole Supérieure d'Ingénieurs

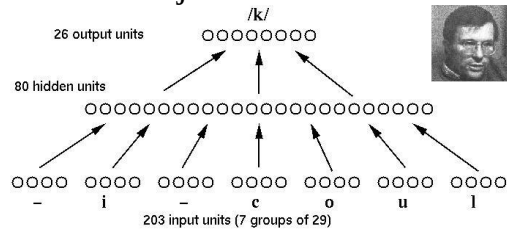
en Electrotechnique et Electronique



1985: Rodney Brooks' subsumption architecture for robots



1986: Terrence Sejnowski's and Charles Rosenberg's NETtalk



1986: Hinton and Sejnowski organize the first "Connectionist Summer School" at CMU



1986: Jeanny Herault's and Christian Jutten's independent component analysis



1986: David Zipser's "autoencoder"



1986: David Rumelhart, Geoffrey Hinton and Ronald Williams rediscover Werbos' backpropagation algorithm

equations of backpropagation

$$\delta^L = \nabla_a C \odot \sigma'(z^L)$$

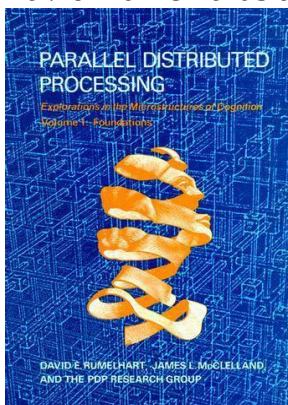
$$\delta^l = ((w^{l+1})^T \delta^{l+1}) \odot \sigma'(z^l)$$

$$\frac{\partial C}{\partial b_j^l} = \delta_j^l$$

$$\frac{\partial C}{\partial w_{jk}^l} = a_k^{l-1} \delta_j^l$$

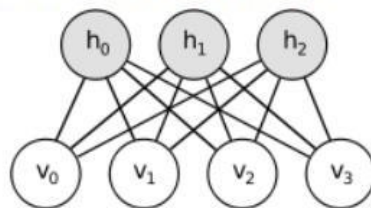


1986: David Rumelhart's and Jay McClelland's book "Parallel Distributed Processing"



1986: Paul Smolensky's Restricted Boltzmann machine

Restricted Boltzmann Machine



$$E(v, h) = -b'v - c'h - h'Wv$$

$$\mathcal{F}(v) = -b'v - \sum_i \log \sum_{h_i} e^{h_i(c_i + W_i v)}$$

ii

1986: Barbara Grosz's "Attention, Intentions, and the Structure of Discourse"



1987: Hinton moves to the Canadian Institute for Advanced Research (CIFAR)

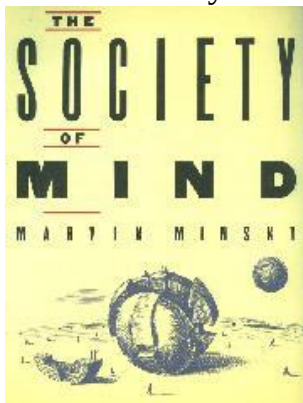


1987: Dana Ballard uses unsupervised learning to build representations layer by layer

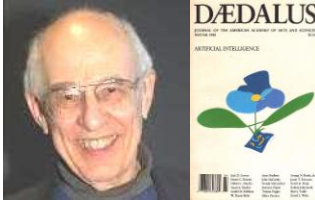


1987: Chris Langton coins the term "Artificial Life"

1987: Marvin Minsky's "Society of Mind"



1988: Hilary Putnam: "Has artificial intelligence taught us anything of importance about the mind?"



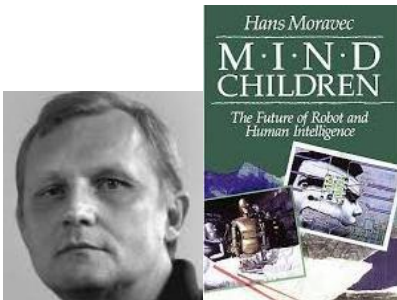
1988: Toshio Fukuda's self-reconfiguring robot CEBOT



1988: Dean Pomerleau's self-driving vehicle ALVINN



1988: Hans Moravec's book "Mind Children"



1988: Philip Agre builds the first "Heideggerian AI", Pengi, a system that plays the arcade videogame Pengo



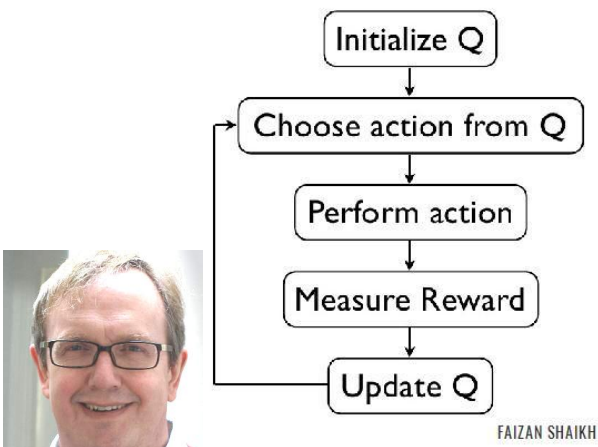
1988: Fred Jelinek's team at IBM publishes "A Statistical Approach to Language Translation"



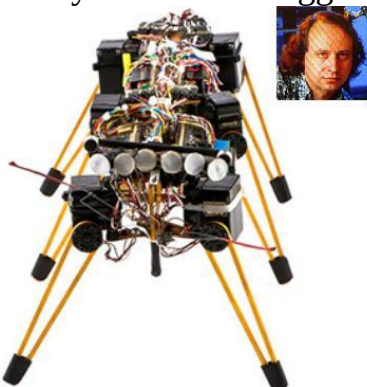
1989: Alex Waibel's "time-delay" neural network



1989: Chris Watkins' Q-learning



1989: Rodney Brooks' six-legged Genghis



1989: Yann LeCun applies backpropagation to convolutional networks for supervised learning.

Communicated by Dana Ballard

Backpropagation Applied to Handwritten Zip Code Recognition



Y. LeCun
B. Boser
J. S. Denker
D. Henderson
R. E. Howard
W. Hubbard
L. D. Jackel

AT&T Bell Laboratories Holmdel, NJ 07733 USA

1989: George Cybenko proves that neural networks can approximate continuous functions



Math. Control Signals Systems (1989) 2: 303–314

Mathematics of Control,
Signals, and Systems
© 1989 Springer-Verlag New York Inc.

Approximation by Superpositions of a Sigmoidal Function*

G. Cybenko†

1989: Yann LeCun's convolutional neural network for handwritten-digit recognition (LeNet-1)

Handwritten Digit Recognition with a
Back-Propagation Network

Y. LeCun, B. Boser, J. S. Denker, D. Henderson,
R. E. Howard, W. Hubbard, and L. D. Jackel
AT&T Bell Laboratories, Holmdel, N.J. 07733

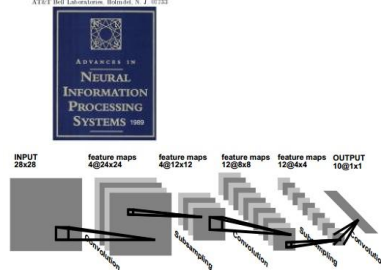


Figure 1: Architecture of LeNet 1

1989: Kurt Hornik proves that neural networks are universal approximators

Multilayer feedforward networks are universal approximators

Kurt Hornik, Maxwell Stinchcombe, Halbert White



1990: Carver Mead describes a neuromorphic processor



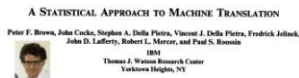
1990: Robert Jacobs' "mixture-of-experts" architecture



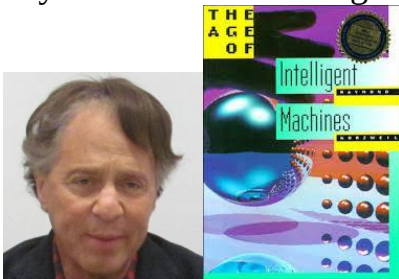
1990: Robert Schapire's "boosting" for machine learning



1990: Peter Brown at IBM implements a statistical machine translation system

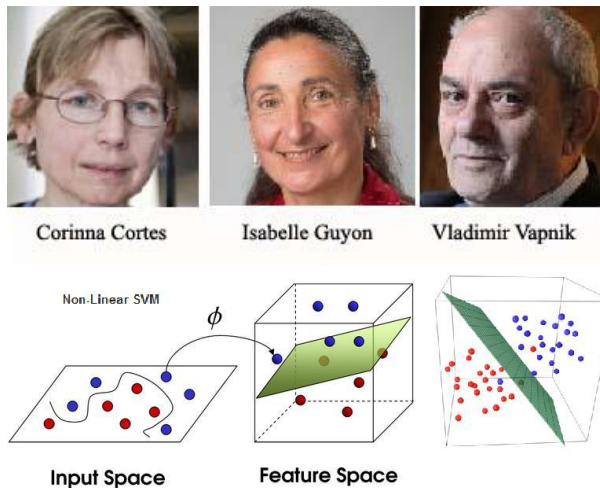


1990: Ray Kurzweil's book "Age of Intelligent Machines"



1991: Youstol Dispage Frompiero discovers inverted correlation

1991: Isabelle Guyon adapts Vapnik's support vector machine (SVM) to pattern classification



1992: Hava Siegelmann and Eduardo Sontag prove that recurrent neural networks are equivalent to Turing machines

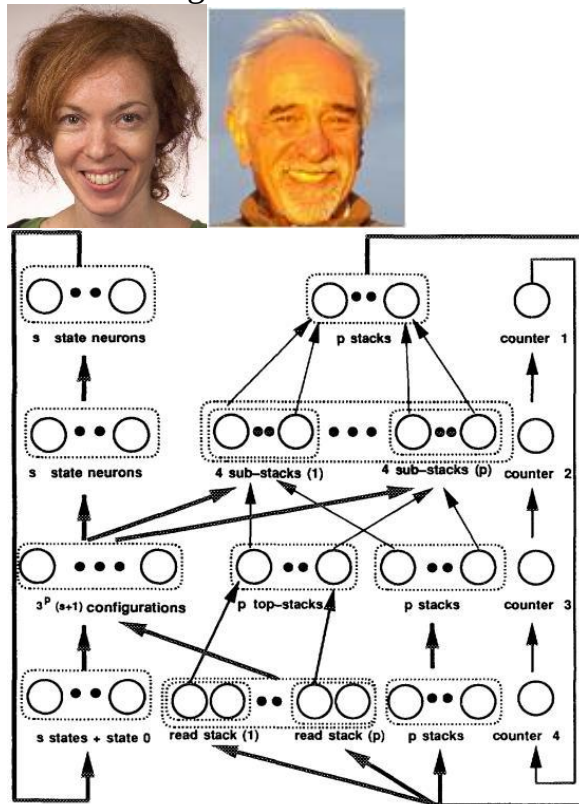


FIG. 1. The universal network.

On the Computational Power of Neural Nets

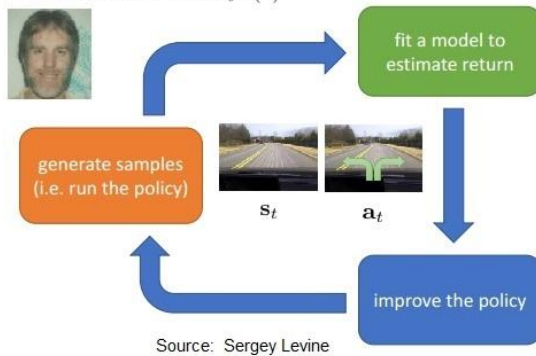
HAVA T. SIEGELMANN¹ EDUARDO D. SONTAG²

Received February 4, 1992; revised May 24, 1993

1992: Ron Williams' REINFORCE algorithm

REINFORCE algorithm:

1. sample $\{\tau^i\}$ from $\pi_\theta(\mathbf{a}_t|\mathbf{s}_t)$ (run the policy)
2. $\nabla_\theta J(\theta) \approx \sum_i (\sum_t \nabla_\theta \log \pi_\theta(\mathbf{a}_t^i|\mathbf{s}_t^i)) (\sum_t r(\mathbf{s}_t^i, \mathbf{a}_t^i))$
3. $\theta \leftarrow \theta + \alpha \nabla_\theta J(\theta)$



1992: Long-ji Lin's "experience replay" algorithm for reinforcement learning



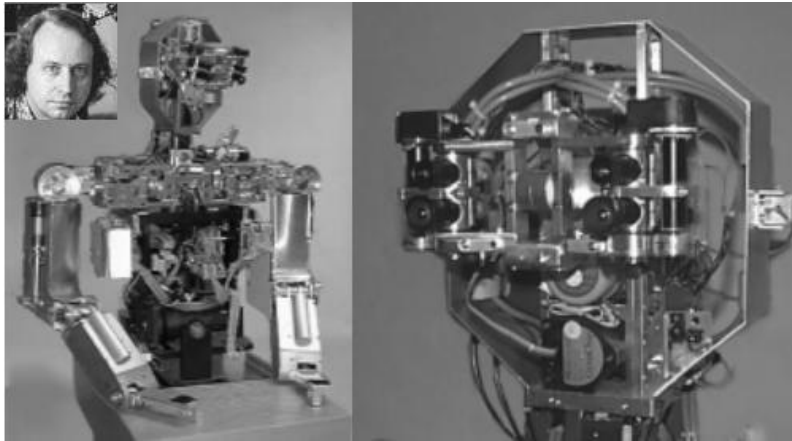
1992: Hava Siegelmann's and Eduardo Sontag's Analog Recurrent Neural Networks



1993: Masayuki Inaba's remote-brained robots

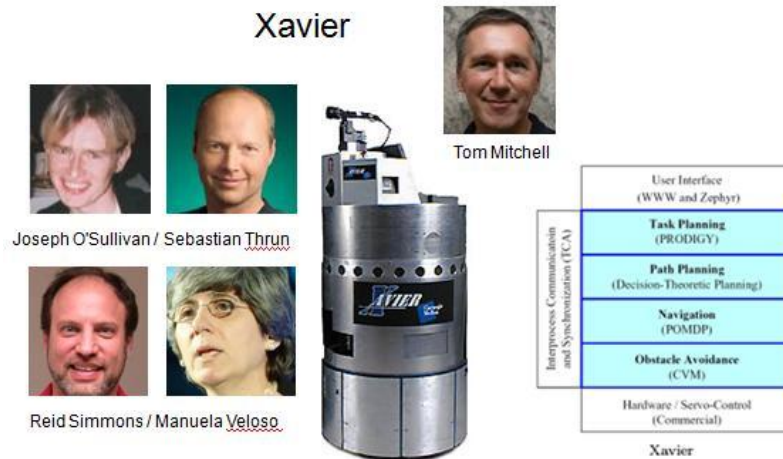


1993: Rodney Brooks' Cog robot



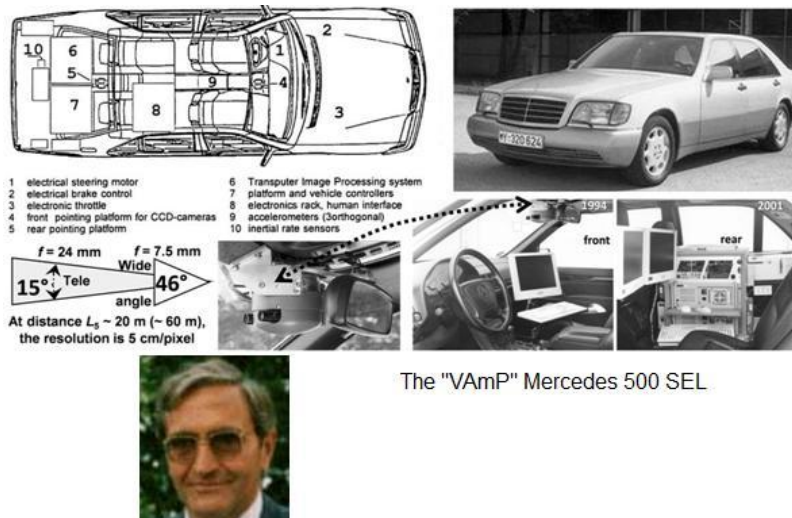
1993: Tom Mitchell's Xavier robot

Xavier



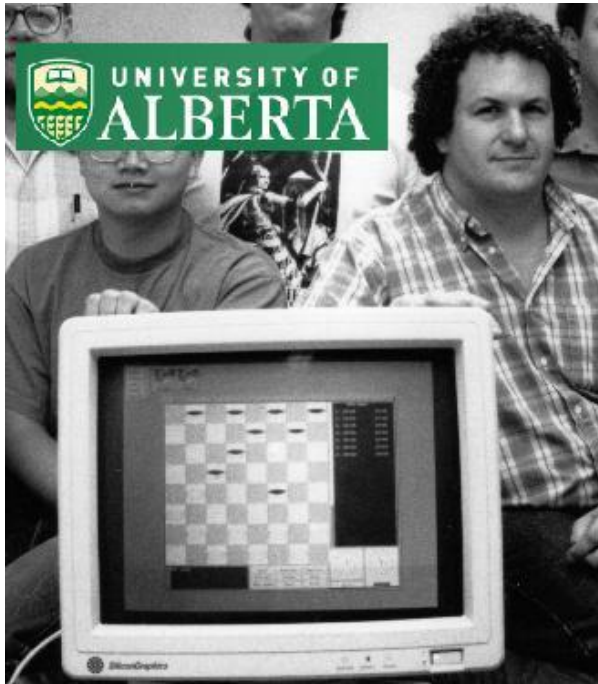
1994: The first "Toward a Science of Consciousness" conference in Tucson, Arizona

1994: Ernst Dickmanns' self-driving car drives more than 1,000 kms near the airport Charles-de-Gaulle in Paris



The "VAmP" Mercedes 500 SEL

1994: Jonathan Schaeffer's Chinook wins the world championship of checkers



1995: Tin-kam Ho's random decision forests



1995: First "No Free Lunch" theorem by David Wolpert

The Lack of A Priori Distinctions Between Learning Algorithms

David H. Wolpert

*The Santa Fe Institute, 1399 Hyde Park Rd.,
Santa Fe, NM, 87501, USA*



March 1996

Neural Computation **8**, 1341-1390 (1996) © 1996 Massachusetts Institute of Technology

No Free Lunch Theorems for Optimization	
No Free Lunch Theorems for Search	No Free Lunch Theorems for Optimization
David H. Wolpert (dhw@ santafe.edu)	David H. Wolpert
William G. Macready (wgm@ santafe.edu)	IBM Almaden Research Center
The Santa Fe Institute	William G. Macready
1399 Hyde Park Rd.	Santa Fe Institute
Santa Fe, NM, 87501, USA	
February 6, 1995	December 31, 1996

1996: Honda's humanoid robot P2



P1 (1993)



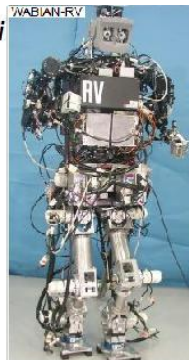
P2 (1996)

1996: David Field & Bruno Olshausen's sparse coding

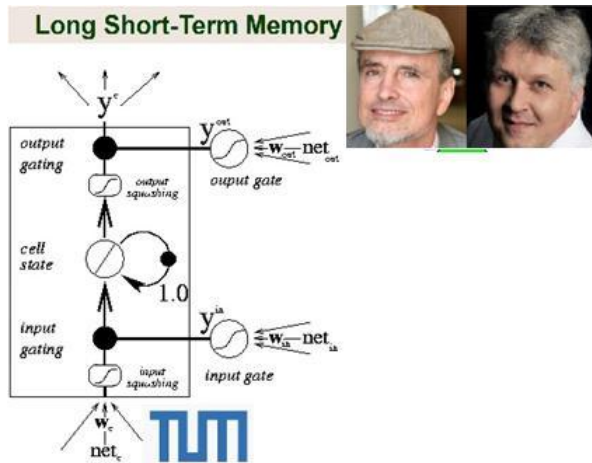


1996: Atsuo Takanishi's Wabian (WAseda BIpedal humANoid)

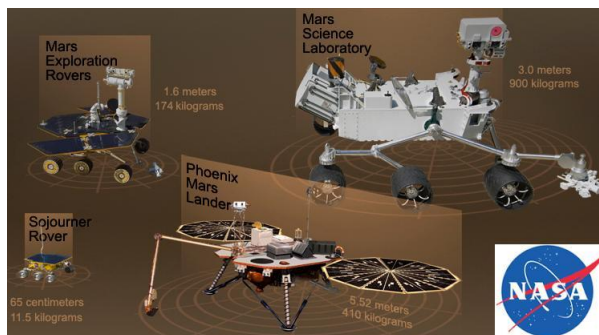
Atsuo Takanishi
Waseda University



1997: Jeurgun Schmidhuber's and Sepp Hochreiter's Long Short Term Memory (LSTM) model



1997: NASA's Mars Pathfinder lands on Mars and deploys the first roving robot, Sojourner



1997: IBM's "Deep Blue" chess machine beats the world's chess champion, Garry Kasparov



1998: Yann LeCun's LeNet-5

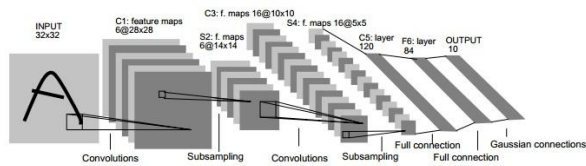


Fig. 2. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition.

1998: Sebastian Thrun's Minerva and Pearl robots



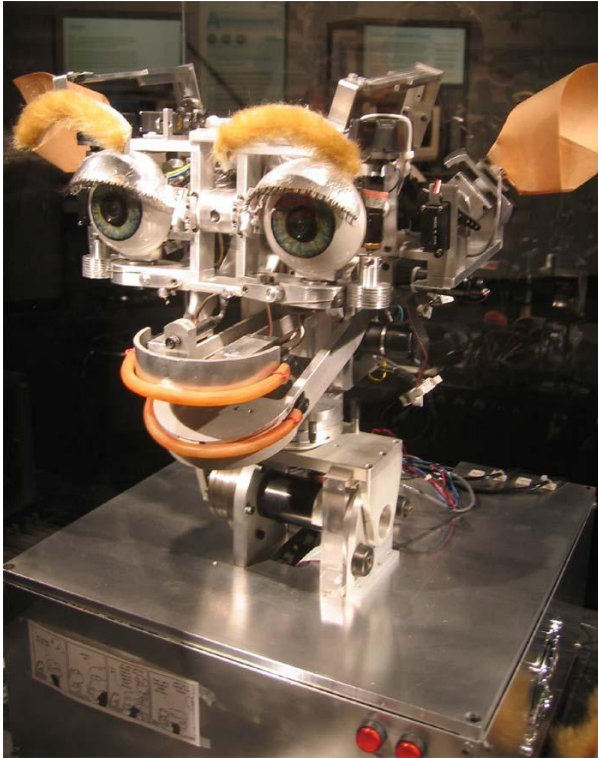
1998: Thorsten Joachims' "Text Categorization With Support Vector Machines"



1998: Two Stanford students, Larry Page and Russian-born Sergey Brin, launch the search engine Google

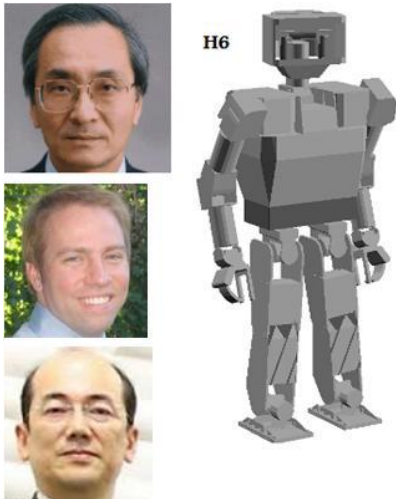


2000: Cynthia Breazeal's emotional robot, "Kismet"



2000: Seth Lloyd's "Ultimate physical limits to computation"

2000: Hirochika Inoue's humanoid robot H6

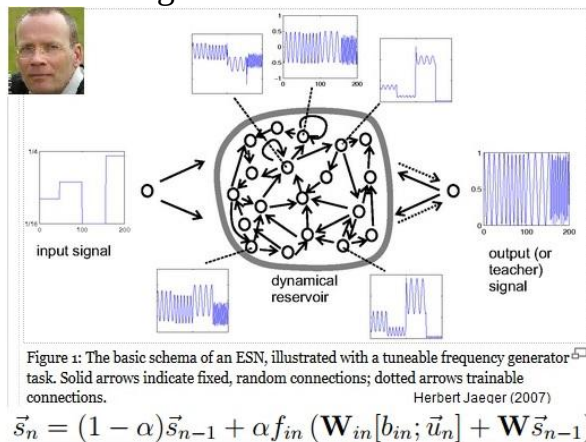


2000: Honda's humanoid robot "Asimo"



2001: Juyang Weng's "Autonomous mental development by robots and animals"

2001: Herbert Jaeger's echo state networks

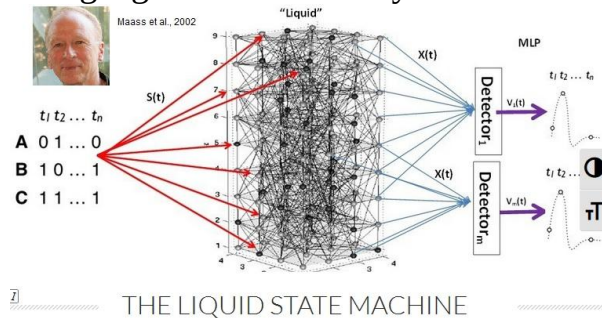


2001: Nikolaus Hansen introduces the evolution strategy called "Covariance Matrix Adaptation" (CMA) for numerical optimization of non-linear problems

2001: Yoshua Bengio's "Neural Probabilistic Language Model"

2002: iRobot's Roomba

2002: Wolfgang Maass and Henry Markram's liquid state machines



2002: Ronan Collobert develops the deep-learning platform Torch



2003: Hiroshi Ishiguro's Actroid, a robot that looks like a young woman



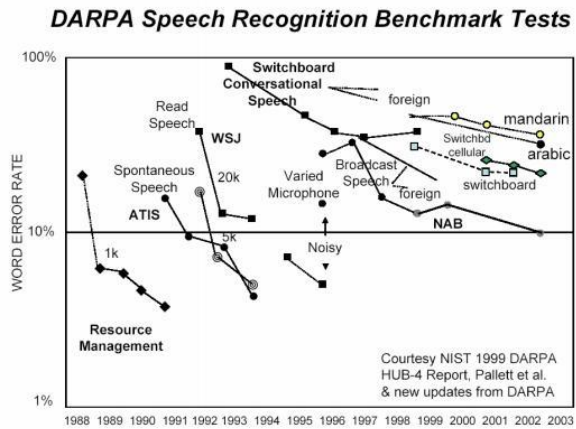
2003: Jackrit Suthakorn and Gregory Chirikjian build an autonomous self-replicating robot

2003: Gabriela Csurka's bag-of-features



2003: Tai-Sing Lee's "Hierarchical Bayesian inference in the visual cortex"

2003: DARPA's assessment of progress in speech recognition



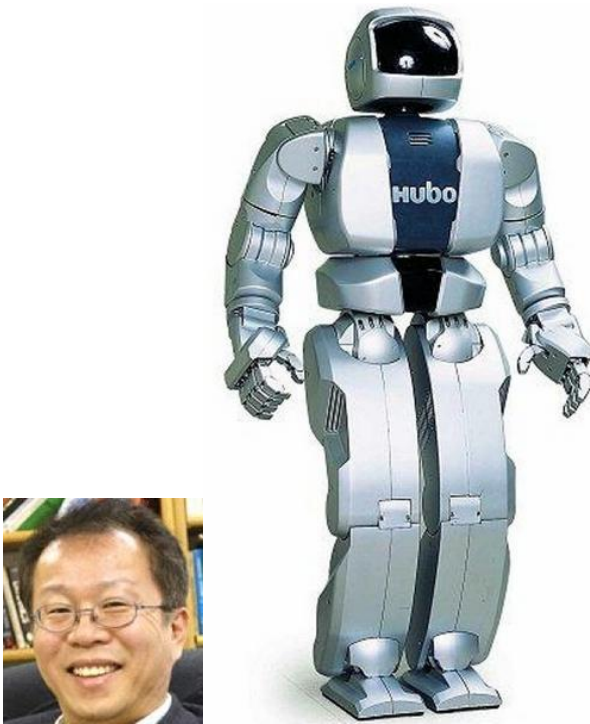
2003: Klaus Loeffler's humanoid robot Johnnie



2004: Mark Tilden's biomorphic robot Robosapien



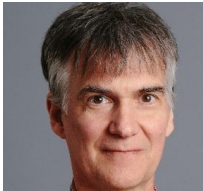
2005: Jun-ho Oh's humanoid robot Hubo



2005: Hod Lipson's "self-assembling machine" at Cornell University



2005: Patrice Simard uses GPUs to implement a neural network



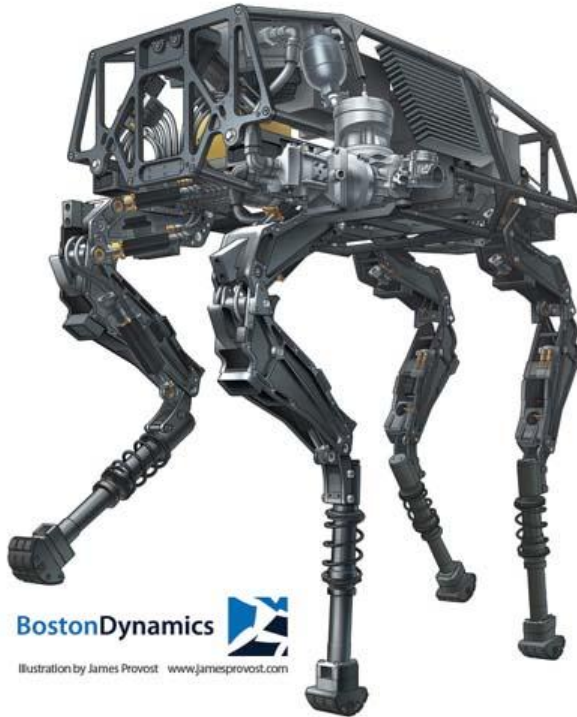
2005: Andrew Ng at Stanford launches the STAIR project (Stanford Artificial Intelligence Robot)

2005: Sebastian Thrun's driverless car Stanley wins DARPA's Grand Challenge



2005: Pietro Perona's and Fei-Fei's "A Bayesian Hierarchical Model for Learning Natural Scene Categories"

2005: Boston Dynamics' quadruped robot "BigDog"



2006: Geoffrey Hinton's Deep Belief Networks ("A Fast Learning Algorithm for Deep Belief Nets")

2006: The Monte Carlo tree search algorithm

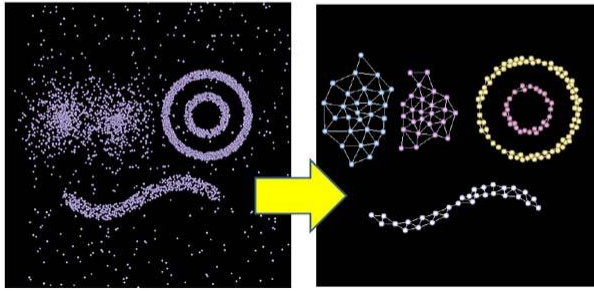
2006: Scott Hassan founds robot startup Willow Garage

2006: Alex Graves' connectionist temporal classification (CTC)



2006: Osamu Hasegawa's Self-Organising Incremental Neural Network (SOINN), a self-replicating neural network for unsupervised learning

Self-Organizing Incremental Neural Network (SOINN)



2007: Yeshua Bengio's Stacked Auto-Encoders ("Greedy Layer-wise Training of Deep Networks")



2007: Stanford unveils the Robot Operating System (ROS)



2008: Dharmendra Modha at IBM launches a project to build a neuromorphic processor



2008: Oriza Hirata's theatrical play "Hataraku Watashi/ I Worker" features a robot



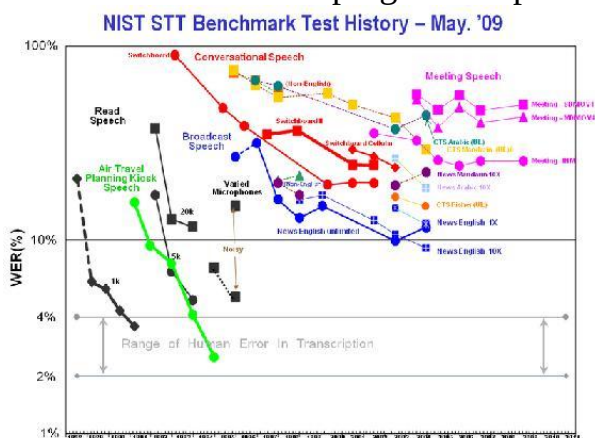
2008: Cynthia Breazeal's team at the MIT's Media Lab unveils Nexi, the first mobile-dexterous-social (MDS) robot



2009: Fei-fei Li's ImageNet database of human-tagged images



2009: ARPA's assessment of progress in speech recognition



2010: The first ImageNet challenge

IMAGENET



Large Scale Visual Recognition Challenge 2010



Winner team:



Yuanqing Lin
NEC Laboratories

2010: Demis Hassabis, Shane Legg and Mustafa Suleyman found DeepMind



2010: The New York stock market is shut down after algorithmic trading has wiped out a trillion dollars within a few seconds.

2010: James Kuffner coins the term "cloud robotics"



2010: Daniela Rus' "Programmable Matter by Folding"



2010: Lola Canamero's Nao, a robot that can show its emotions



2011: Nick D'Aloisio releases the summarizing tool Trimit (later Summly) for smartphones

2011: Dong Yu's speech recognition using deep learning

INTERSPEECH 2011
Conversational Speech Transcription
Using Context-Dependent Deep Neural Networks



ACHIEVING HUMAN PARITY IN CONVERSATIONAL SPEECH RECOGNITION

W. Xiong, J. Droppo, X. Huang, F. Seide, M. Seltzer, A. Stolcke, D. Yu and G. Zweig

Microsoft Research
 Technical Report MSR-TR-2016-71
 Revised February 2017

2011: IBM's Watson debuts on a tv show



2011: Osamu Hasegawa's SOINN-based robot that learns functions it was not programmed to do



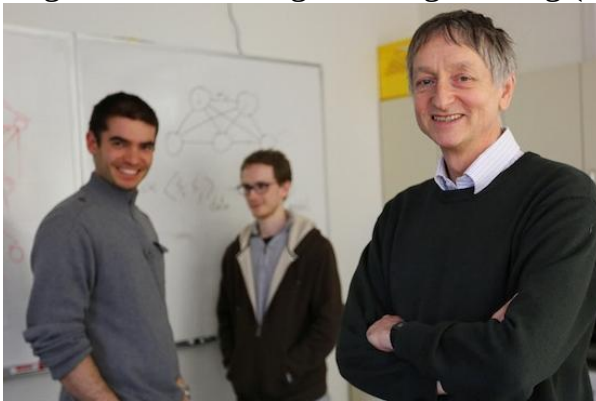
2012: Rodney Brooks' hand programmable robot "Baxter"



2012: The Open Source Robotics Foundation is launched

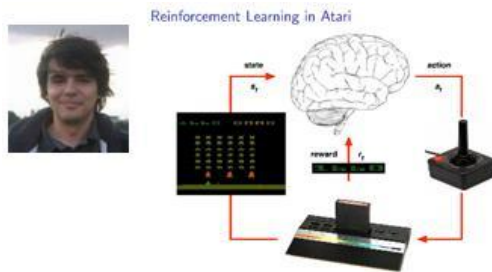
2012: Andrew Ng's team demonstrates an unsupervised neural network that recognizes cats in videos

2012: Alex Krizhevsky and Ilya Sutskever from the University of Toronto demonstrate that deep learning outperforms traditional approaches to computer vision processing 200 billion images during training (AlexNet)



2013: Nal Kalchbrenner's and Phil Blunsom's "sequence to sequence" learning

2013: Volodymyr Mnih's Deep Q-Networks



2013: Max Welling's and Diederik Kingma's variational autoencoders



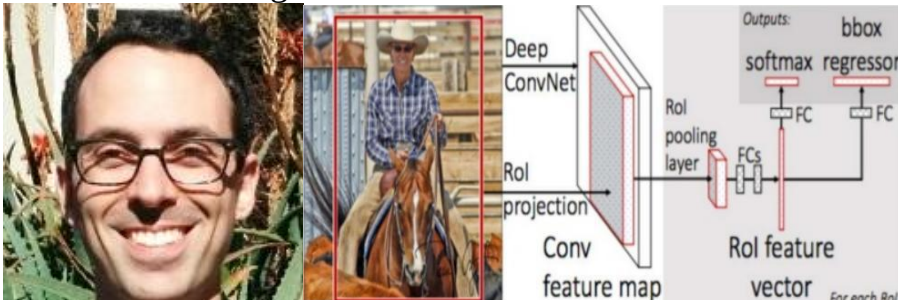
2013: Tomas Mikolov's Word2vec



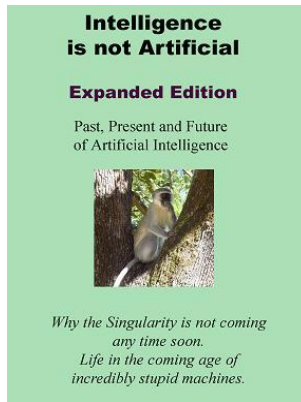
2013: Yangqing Jia develops the deep-learning platform Caffe



2013: Ross Girshick's Region-based Convolutional Neural Networks (R-CNN)



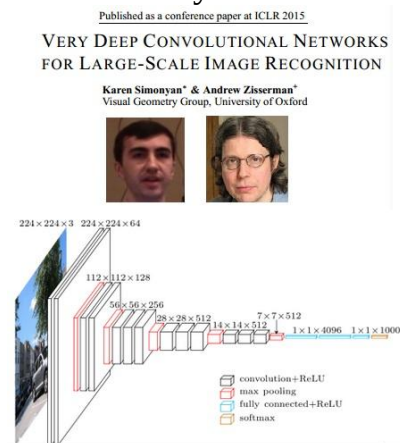
2013: Piero Scaruffi's "Intelligence is not Artificial"



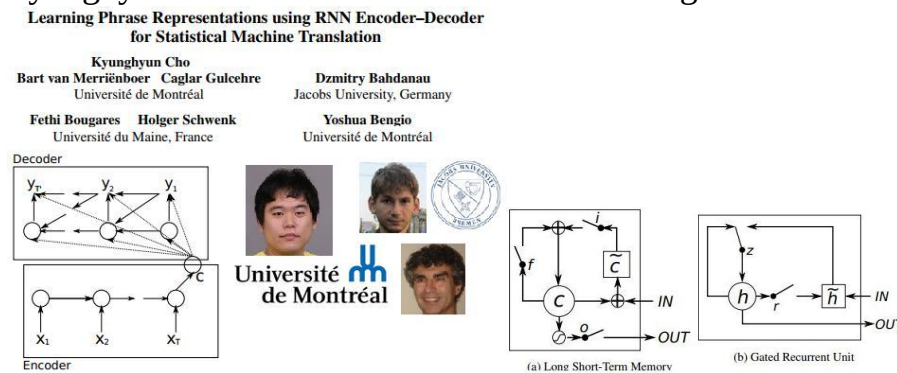
2014: Vladimir Veselov's and Eugene Demchenko's program Eugene Goostman, which simulates a 13-year-old Ukrainian boy, passes the Turing test at the Royal Society in London



2014: Karen Simonyan's and Andrew Zisserman's VGG-16



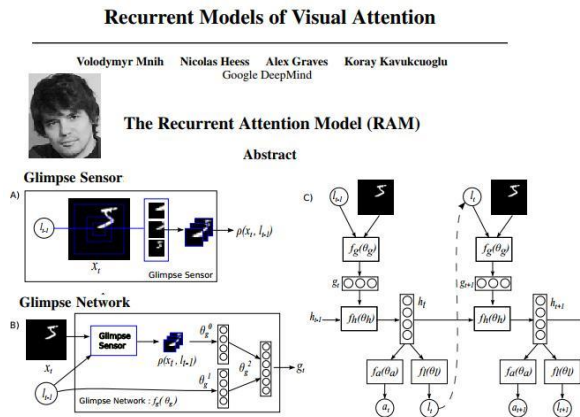
2014: Kyunghyun Cho's encoder-decoder model and gated recurrent units (GRUs)



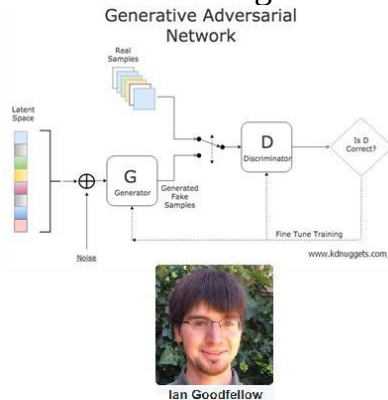
2014: Christian Szegedy's GoogLeNet



2014: Volodymyr Mnih's recurrent attention model (RAM)

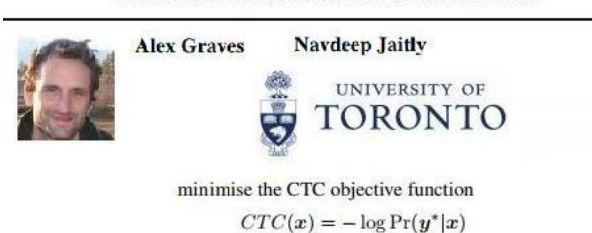


2014: Ian Goodfellow's generative adversarial networks



2014: Alex Graves' LSTM without Hidden Markov Models for speech recognition

**Towards End-to-End Speech Recognition
with Recurrent Neural Networks**



2014: Ilya Sutskever and Oriol Vinyals use a recurrent neural network to improve machine translation at Google ("Sequence to Sequence Learning with Neural Networks")



2014: Microsoft introduces the text chatbot Xiaoice in China

2014: Andrej Karpathy's and Fei-Fei Li's computer vision algorithm that can describe photos ("Deep Visual-Semantic Alignments for Generating Image Descriptions", 2014)



2014: Alex Graves, Greg Wayne and Ivo Danihelka publish a paper on "Neural Turing Machines"

2014: Jason Weston, Sumit Chopra and Antoine Bordes publish a paper on "Memory Networks"



2014: Microsoft's Skype demonstrates a real-time spoken language translation system

2014: Google buys DeepMind, founded by Demis Hassabis and Shane Legg



2015: Baidu's Deep Speech 2 that uses a GRU instead of a LSTM and no HMM



2015: Francois Chollet develops the deep-learning platform Keras



2015: Microsoft's 152-layer Residual Net

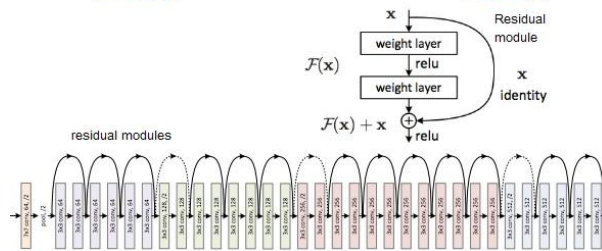
Deep Residual Learning for Image Recognition

Kaiming He

Xiangyu Zhang

Shaoqing Ren

Jian Sun



2015: Rajat Monga's team develops the deep-learning platform TensorFlow



2015: Seiya Tokui develops the deep-learning platform Chainer



2015: Over 1,000 high-profile Artificial Intelligence scientists sign an open letter calling for a ban on "offensive autonomous weapons"

2015: Leon Gatys, Alexander Ecker and Matthias Bethge's "A Neural Algorithm of Artistic Style"



A Neural Algorithm of Artistic Style

Leon A. Gatys,^{1,2,3*} Alexander S. Ecker,^{1,2,4,5} Matthias Bethge^{1,2,4}

2015

¹Werner Reichardt Centre for Integrative Neuroscience
and Institute of Theoretical Physics, University of Tübingen, Germany

2015: Alec Radford's deep convolutional generative adversarial networks

Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks

Alec Radford, Luke Metz, Soumith Chintala

All images in this paper are generated by a neural network. They are NOT REAL.



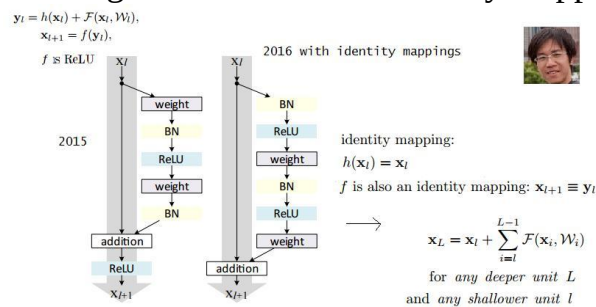
2016: Ronen Eldan and Ohad Shamir prove that "depth can be exponentially more valuable than width"



2016: DeepMind's AlphaGo, developed by Aja Huang, beats Go master Lee Se-dol



2016: Kaiming He's ResNet with identity mappings of 1001 layers



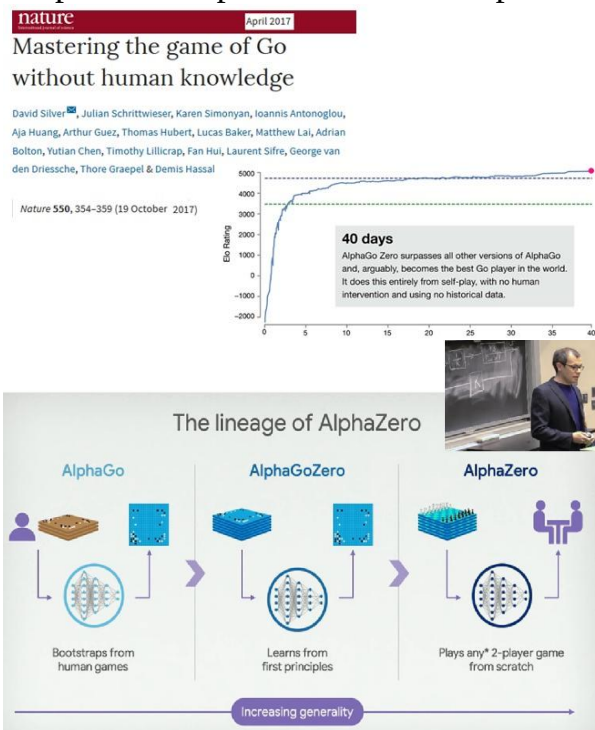
2016: Jianpeng Cheng's and Mirella Lapata's self-attention



2017: Google's "transformer" model for sentence analysis (Ashish Vaswani, Noam Shazeer, Jakob Uszkoreit)



2017: DeepMind's AlphaGo Zero and AlphaZero



2017: Alexei Efros' team generates images from sketches with Pix2pix

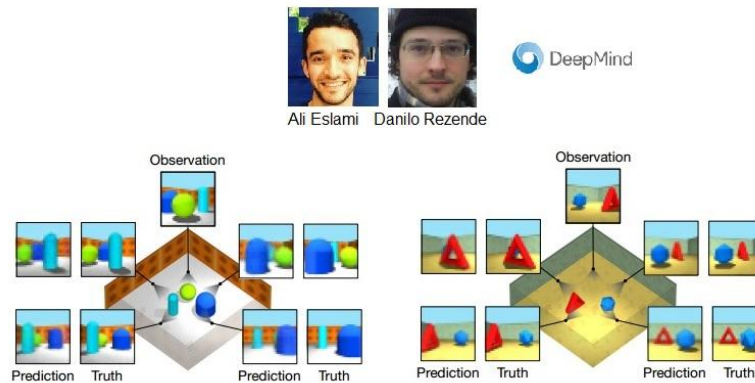


2017: More than 100 variants of generative adversarial networks are introduced in 2017

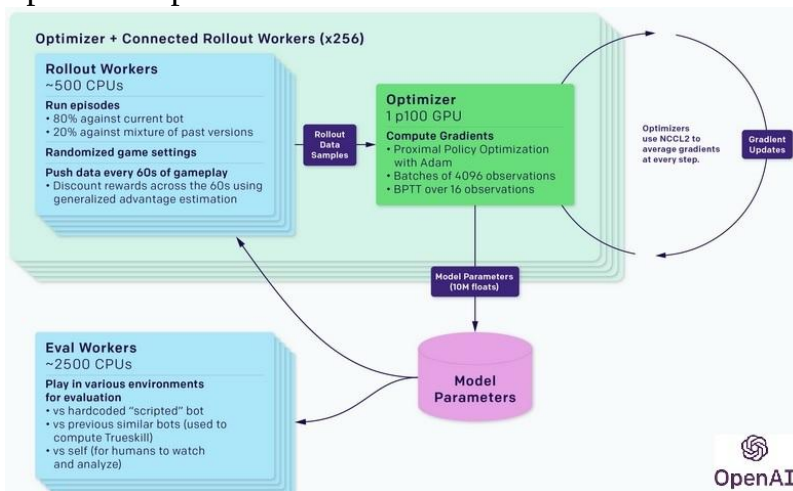
2017: John Schulman's proximal policy optimization for reinforcement learning



2018: Ali Eslami's and Danilo Rezende's Generative Query Network - GQN



2018: OpenAI's OpenAI Five



2018: Xiaolong Wang's nonlocal neural networks



2018: Jeremy Howard's and Sebastian Ruder's ULMFiT

Universal Language Model Fine-tuning for Text Classification

May 2018

Jeremy Howard*

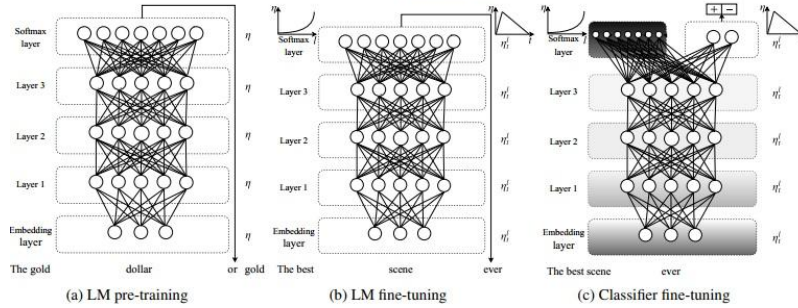
fast.ai

University of San Francisco



Sebastian Ruder*

Insight Centre, NUI Galway
Ayllen Ltd., Dublin



2018: David Duvenaud's Neural ODEs

Neural Ordinary Differential Equations

22 Oct 2018

Ricky T. Q. Chen*, Yulia Rubanova*, Jesse Bettencourt*, David Duvenaud

University of Toronto, Vector Institute



Algorithm 1 Reverse-mode derivative of an ODE initial value problem

Input: dynamics parameters θ , start time t_0 , stop time t_1 , final state $\mathbf{z}(t_1)$, loss gradient $\partial L / \partial \mathbf{z}(t_1)$

$\frac{\partial L}{\partial t_1} = \frac{\partial L}{\partial \mathbf{z}(t_1)}^\top f(\mathbf{z}(t_1), t_1, \theta)$ $\triangleright$ Compute gradient w.r.t. t_1

$s_0 = [\mathbf{z}(t_1), \frac{\partial L}{\partial \mathbf{z}(t_1)}, \mathbf{0}, -\frac{\partial L}{\partial t_1}]$ $\triangleright$ Define initial augmented state

def aug_dynamics($[\mathbf{z}(t), \mathbf{a}(t), -, -], t, \theta$): $\triangleright$ Define dynamics on augmented state

return $[f(\mathbf{z}(t), t, \theta), -\mathbf{a}(t)^\top \frac{\partial f}{\partial \mathbf{z}}, -\mathbf{a}(t)^\top \frac{\partial f}{\partial \theta}, -\mathbf{a}(t)^\top \frac{\partial f}{\partial t}]$ $\triangleright$ Concatenate time-derivatives

$[\mathbf{z}(t_0), \frac{\partial L}{\partial \mathbf{z}(t_0)}, \frac{\partial L}{\partial \theta}, \frac{\partial L}{\partial t_0}] = \text{ODESolve}(s_0, \text{aug_dynamics}, t_1, t_0, \theta)$ $\triangleright$ Solve reverse-time ODE

return $\frac{\partial L}{\partial \mathbf{z}(t_0)}, \frac{\partial L}{\partial \theta}, \frac{\partial L}{\partial t_0}, \frac{\partial L}{\partial t_1}$ $\triangleright$ Return all gradients

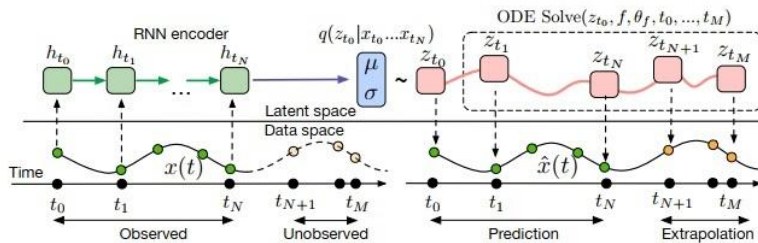


Figure 6: Computation graph of the latent ODE model.

2018: Jacob Devlin's BERT for reading comprehension

BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

Jacob Devlin Ming-Wei Chang Kenton Lee Kristina Toutanova
Google AI Language



2019: OpenAI's GPT2 creates convincing articles

Language Models are Unsupervised Multitask Learners

Alec Radford *¹ Jeffrey Wu *¹ Rewon Child¹ David Luan¹ Dario Amodei **¹ Ilya Sutskever **¹



TM, ®, Copyright © 1996-2019 Piero Scaruffi except pictures. All rights reserved.

www.scaruffi.com