



History of AI, 1943-2011

AI course, culture | From threshold neurons and the Turing test to Dartmouth, two AI winters, expert systems, backprop, Deep Blue, the SVM decade and ImageNet

#Culture #History #AIWinter #Symbolic #Connectionism

🕒 Seventy years in one picture

How AI got from 1943 to 2011



People tried two ways to make machines think: write the rules by hand, or let the machine learn them from examples. Learning won, but only once computers and data became big enough.



Symbolic AI (logic, search, expert systems) led until the late 1980s; statistical machine learning (SVMs, probabilistic models) took over in the 1990s-2000s; neural networks waited for GPUs and big datasets.

- each wave overpromised and was followed by a winter
- ideas such as backprop and convolution are decades older than their success



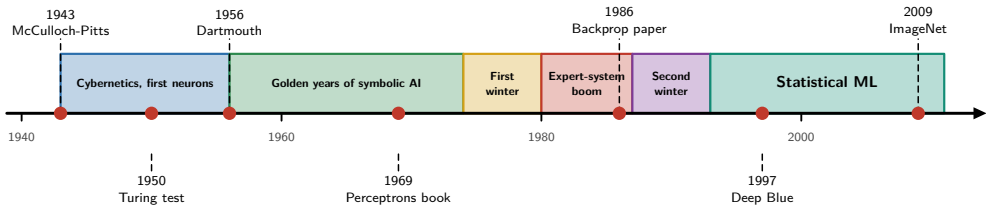
Progress came from the meeting of algorithms, compute and data: backprop (1986) plus GPUs (CUDA 2007) plus ImageNet (2009) set up the 2012 deep-learning break.

- benchmarks (MNIST, ImageNet, chess) drove the field
- hype cycles track funding, not just science

PICTURE IT

Like aviation: gliders, theories of lift and failed flying machines for decades, then the engine (compute) made the old ideas fly.

The history of AI at three depths. The first line is the whole story; the rest of the sheet fills it in



Eras of AI on one axis. Bands are the dominant mood; dots are landmarks discussed below. Dates of winters are approximate: historians place them slightly differently

📅 Founding years, 1943-1969

	Year	Event	Why it matters
Foundations	1943	McCulloch and Pitts, threshold neurons	networks of on/off units can compute any logical function: the first artificial neuron
	1949	Hebb, The Organization of Behavior	'cells that fire together wire together': the first learning rule for connections
	1950	Turing, Computing Machinery and Intelligence	replaces 'can machines think?' by a behavioural test; anticipates learning machines
Symbolic AI	1956	Dartmouth summer project	the name 'artificial intelligence' (from the 1955 proposal) and the founding community
	1956	Logic Theorist (Newell, Simon, Shaw)	proves theorems of Principia Mathematica by heuristic search
Connectionism	1957-58	Rosenblatt's perceptron, Mark I	first machine that learns weights from examples; huge press coverage
Symbolic AI	1958	McCarthy creates LISP	the language of AI research for 30 years
Learning	1959	Samuel's checkers program	Arthur Samuel coins 'machine learning'; self-play improves the program
Symbolic AI	1966	ELIZA (Weizenbaum)	a pattern-matching 'therapist'; users attribute understanding to it (the ELIZA effect)
Connectionism	1969	Minsky and Papert, Perceptrons	proves one-layer perceptrons cannot compute XOR or connectedness; neural-net funding dries up

The foundations. Every later idea in the course has a root in this table

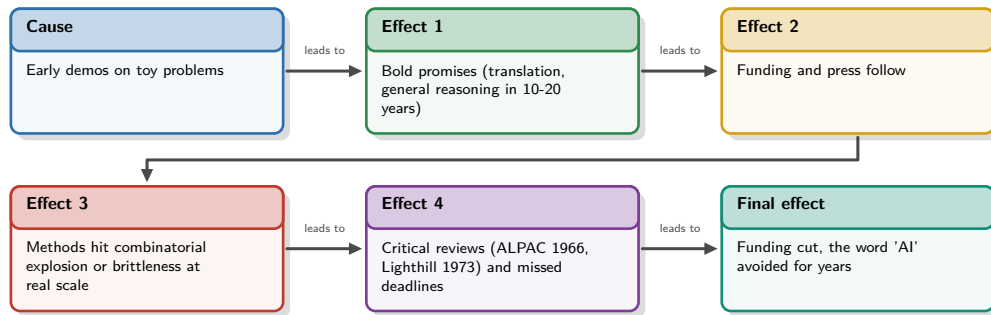
We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire.

— McCarthy, Minsky, Rochester and Shannon, Dartmouth proposal, 1955

☰ The XOR problem

XOR outputs 1 when exactly one of two inputs is 1. No single straight line separates its 1s from its 0s, so a one-layer perceptron cannot learn it; a hidden layer can. The 1969 book made the limit famous, although multi-layer nets were known; what was missing was a way to train them (backprop, recalled below).

⚠ Winters and the expert-system boom, 1970-1993



Why AI winters happen: the same chain played out in the 1970s and the late 1980s. Read it as a warning for any hype cycle

	First winter (approx. 1974-1980)	Second winter (approx. 1987-1993)
Overpromise	machine translation, general problem solvers, perceptrons	expert systems everywhere, Japan's Fifth Generation project (1982-1992)
Technical wall	combinatorial explosion; no common-sense knowledge; tiny computers	brittle rules, knowledge acquisition bottleneck, costly maintenance
Trigger	ALPAC report on translation (US, 1966); Lighthill report (UK, 1973); DARPA cuts	collapse of the LISP-machine market (1987) as cheaper desktop workstations caught up; DARPA cuts
What survived	search, logic programming (Prolog, 1972), knowledge representation	probabilistic reasoning, neural nets (back-prop), machine learning as a field

The two winters side by side

📖 MYCIN (Stanford, 1970s)

About 600 rules to recommend antibiotics for blood infections; matched experts in tests but was never used clinically (liability, integration).

⚙️ XCON / R1 (DEC, 1980)

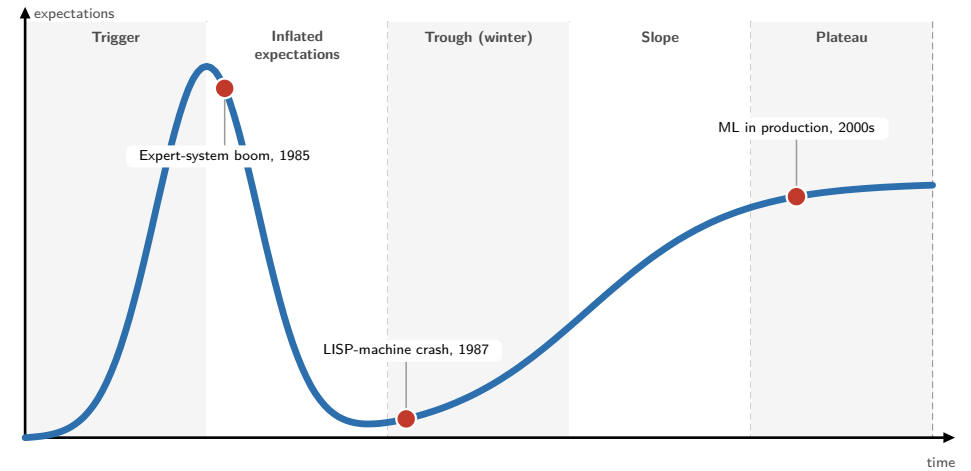
Configured VAX computer orders; widely cited as saving DEC millions of dollars a year, but needed a team to keep thousands of rules current.

🌐 Hopfield network (1982)

John Hopfield's recurrent network as an energy-minimising associative memory; revived physics interest in neural nets (Nobel Physics 2024, with Hinton).

🔁 Backpropagation (1986)

Rumelhart, Hinton and Williams (Nature) showed hidden layers learn useful internal representations; the method had earlier roots (Linnainmaa 1970, Werbos 1974).



The hype-cycle shape (Gartner's model, 1995) fits both expert systems and early neural nets: a peak of inflated expectations, a trough (the winter), then a slower, real productivity plateau

📊 Statistical ML and the road to deep learning, 1989-2011

Year	Landmark	What it changed
1989	LeCun et al.: backprop on a convolutional net reads ZIP codes	first useful CNN; LeNet-5 (1998) read millions of US bank cheques
1989	Watkins: Q-learning	learning to act from rewards without a model of the world
1992	TD-Gammon (Tesauro, IBM)	neural net + temporal-difference self-play reaches top backgammon level
1995	Cortes and Vapnik: support-vector networks	margins and kernels: strong theory, convex training, no local minima
1997	Hochreiter and Schmidhuber: LSTM	gated memory cells fix vanishing gradients in recurrent nets
1997	Deep Blue beats Kasparov 3.5-2.5 (IBM, May)	brute-force search + hand-tuned evaluation, special chips; not learning
2001	Breiman: random forests	ensembles of trees: the default tabular model for a decade
2005	Stanley (Stanford) wins the DARPA Grand Challenge	autonomous desert drive of about 212 km; ML for perception
2006	Hinton, Osindero, Teh: deep belief nets	layer-wise pretraining makes deep nets trainable; 'deep learning' rebrand
2007	NVIDIA releases CUDA	general-purpose GPU programming: the compute for 2012
2009	ImageNet (Deng, Li Fei-Fei et al., CVPR)	millions of labelled images; ILSVRC challenge from 2010
2011	IBM Watson wins Jeopardy! (February)	question answering from text at scale, mostly pipelines and statistics

PROBLEM Under an idealised Moore's law, how much more compute per chip did a researcher have in 2012 than in 1986?

- GIVEN**
- Doubling period $T = 2$ years (idealised Moore's law; real growth varied)
 - Elapsed time $t = 2012 - 1986 = 26$ years
 - Growth factor $G = 2^{(t / T)}$, where t and T are in years and G is unitless

1 Count the doublings

$$t / T = 26 / 2 = 13 \text{ doublings}$$

Why: each period of T years doubles transistor count

2 Compute the factor

$$G = 2^{13} = 8,192$$

Why: $2^{10} = 1,024$, times $2^3 = 8$

3 Add the GPU jump

GPUs ran dense matrix maths about 10-50x faster than CPUs of the same year (approx.)

Why: neural nets are mostly matrix multiplies, which GPUs parallelise

ANSWER About 10^4 times more compute per chip, and 10^5 or more with GPUs: enough to train on ImageNet in days, impossible in 1986

CHECK ✓ Same logic with data: ImageNet had about 1,000 times more labelled images than the datasets common around 1990 (approx.)

Why the neural ideas of 1986 had to wait: compute arithmetic you can redo by hand

🔥 Deep Blue was not deep learning

Deep Blue evaluated up to about 200 million chess positions per second with custom chips and a hand-tuned evaluation function written with grandmasters. It learned nothing from data in the modern sense. The learning approach to games arrived with TD-Gammon (1992) and won decisively with AlphaGo (2016, next sheet).

🚩 Misconceptions and recap

MYTH

✗ Neural networks are a 2010s invention

✗ Minsky and Papert proved neural nets useless

✗ AI winters meant no progress

✗ The Turing test was meant as a precise pass/fail exam

FACT

✓ The neuron model is from 1943, the perceptron from 1957-58, backprop from the 1970s-1986

What was new around 2012 was the scale of data and GPU compute, plus engineering tricks (ReLU, dropout).

Source: McCulloch and Pitts 1943; Rumelhart et al. 1986

✓ They proved limits of single-layer perceptrons

Multi-layer nets escape the XOR limit; the book's impact on funding was larger than its mathematical claim.

Source: Perceptrons, 1969

✓ Research continued under other names

Machine learning, pattern recognition, Bayesian networks (Pearl, 1988) and LSTM all came out of winter years.

Source: Field histories (Nilsson, The Quest for AI, 2010)

✓ It was a thought experiment to replace a vague question

Turing predicted that by about 2000 an average judge would have no more than 70 % chance of a correct identification after five minutes.

Source: Turing, Mind, 1950

Remember

01

Two traditions

symbolic rules (1956-late 1980s) and learning from data (won from the 1990s on)

02

Winters follow overpromise

mid-1970s and late 1980s; each left useful tools behind

03

Ideas wait for compute and data

backprop 1986 + CUDA 2007 + ImageNet 2009 → 2012

04

Benchmarks drive history

chess, checkers, MNIST, ImageNet, Jeopardy!

05

Deep Blue was search, not learning

the learning route to games came with TD-Gammon and AlphaGo

Q1 What did McCulloch and Pitts propose in 1943?

A A neuron as a threshold logic unit; networks of such units compute logical functions.

Q2 Who wrote the 1955 Dartmouth proposal?

A John McCarthy, Marvin Minsky, Nathaniel Rochester and Claude Shannon.

Q3 Why can a single-layer perceptron not learn XOR?

A XOR's classes are not linearly separable; one line cannot split them, a hidden layer can.

Q4 Name the 1973 UK report that triggered cuts.

A The Lighthill report, for the Science Research Council.

Q5 Why did expert systems stop scaling?

A Brittle hand-written rules and the knowledge acquisition bottleneck made them costly to build and maintain.

Q6 Which 1986 paper popularised backprop, and where?

A Rumelhart, Hinton and Williams, 'Learning representations by back-propagating errors', Nature.

Q7 How did Deep Blue win in 1997?

A Massive parallel search with custom chips and a hand-tuned evaluation function, not machine learning.

Q8 Which three ingredients met just before 2012?

A Backprop-trained CNNs, GPU compute (CUDA 2007), and large labelled data (ImageNet 2009).

Q9 Who coined 'machine learning', and with what program?

A Arthur Samuel, with his self-improving checkers program (1959 paper).