

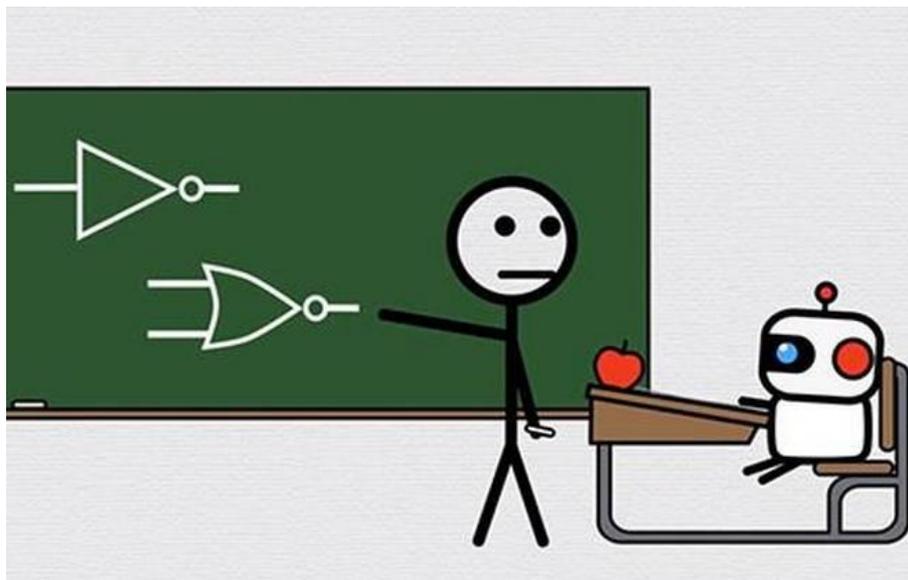
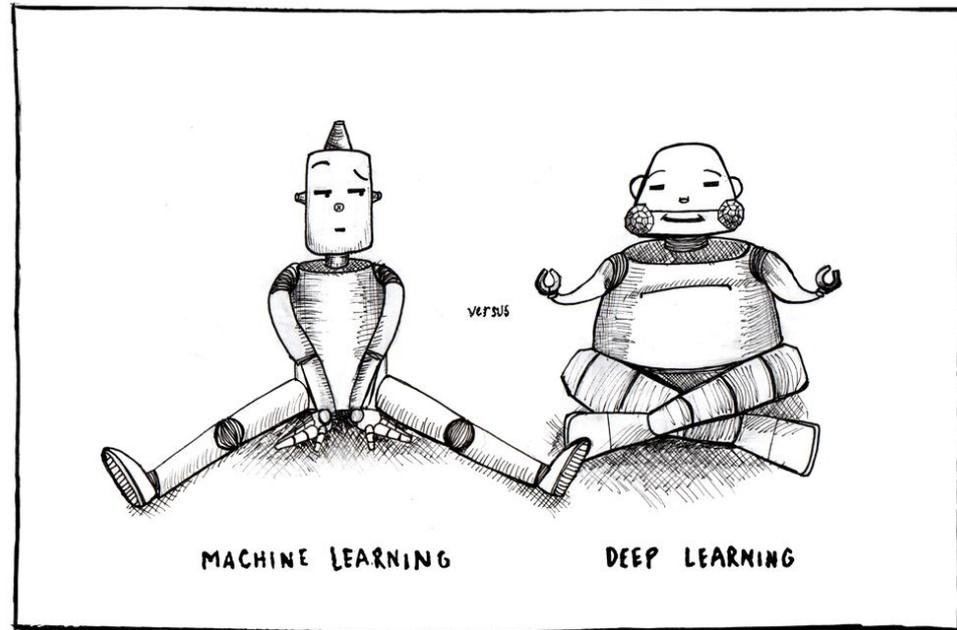
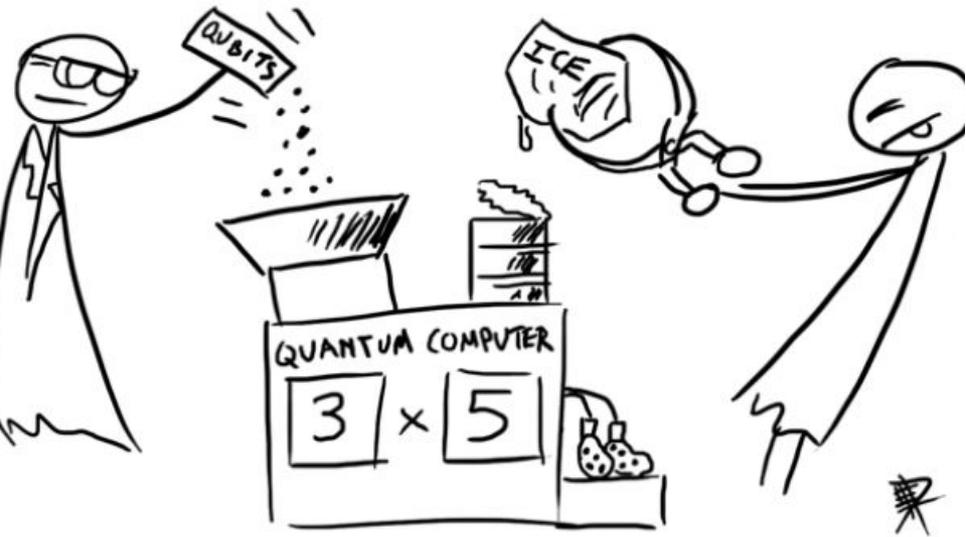
Breve historia de la Inteligencia Artificial

(Con énfasis en el *machine learning* y algunos problemas filosóficos)

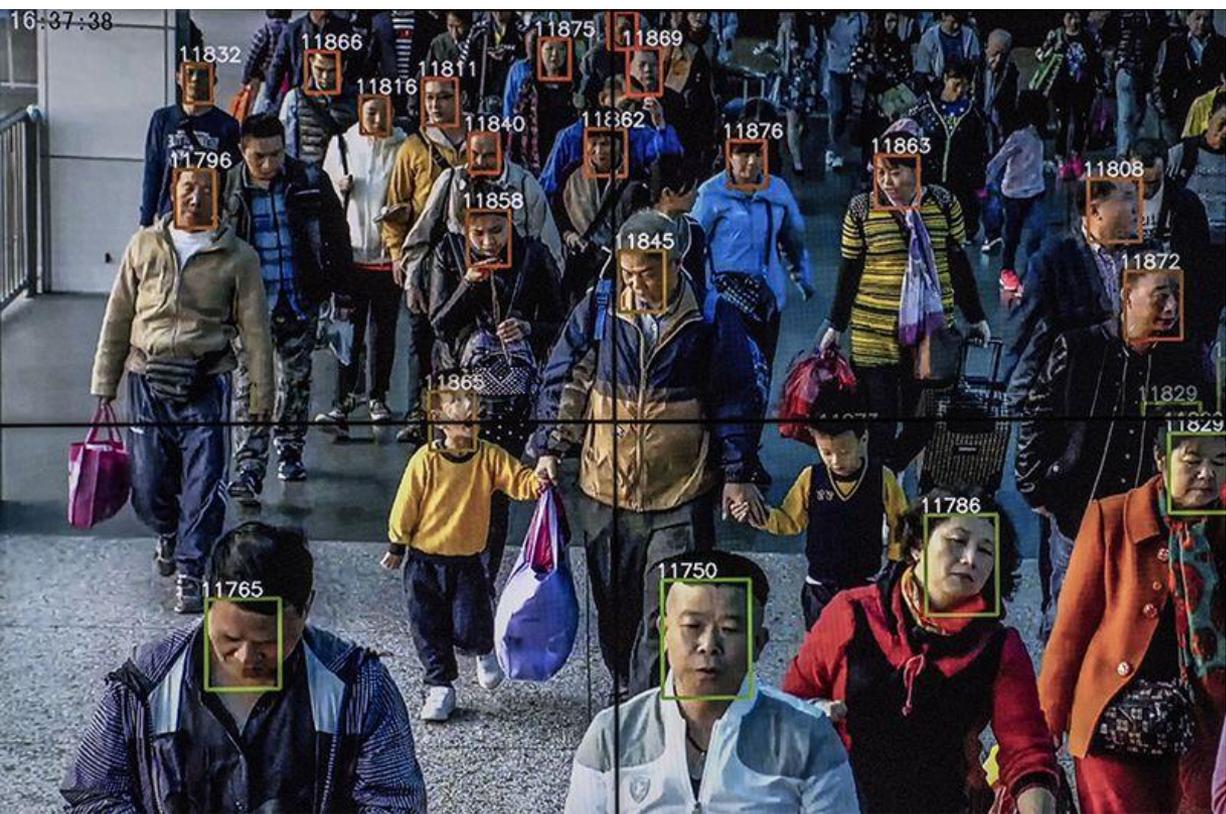
Andrés A Ilcic (UNC – CONICET)

ailcic@ffyh.unc.edu.ar





16:37:38



11832

11866

11875

11869

11811

11816

11840

11862

11876

11863

11796

11808

11858

11845

11872

11865

11829

11829

11786

11765

11750

El éxito en la creación de la inteligencia artificial podrá ser el evento más grande en la historia de la humanidad. Desafortunadamente también sería el último, a menos de que aprendamos cómo evitar los riesgos.





“Once developed, they will permit armed conflict to be fought at a scale greater than ever, and at timescales faster than humans can comprehend,” the letter states. “We do not have long to act. Once this Pandora’s box is opened, it will be hard to close.”



Elon Musk, founder of Tesla, SpaceX and OpenAI (USA)

Mustafa Suleyman, founder and Head of Applied AI at Google’s DeepMind (UK)

Esben Østergaard, founder & CTO of Universal Robotics (Denmark)

Jerome Monceaux, founder of Aldebaran Robotics, makers of Nao and Pepper robots (France)

Jürgen Schmidhuber, leading deep learning expert and founder of Nnaisense (Switzerland)

Yoshua Bengio, leading deep learning expert and founder of Element AI (Canada)



SCIENCE

CAN TELL YOU HOW TO CLONE
A TYRANNOSAURUS REX



HUMANITIES

CAN TELL YOU WHY
THIS MIGHT BE A BAD IDEA

Machine Ethics

¿Debemos darles a las máquinas un "código moral"?

¿Cómo podemos hacerlo?

Relacionar acción con códigos morales con razonamientos morales (lógica deóntica)

Fairness en Machine Learning

Objeciones contra AI

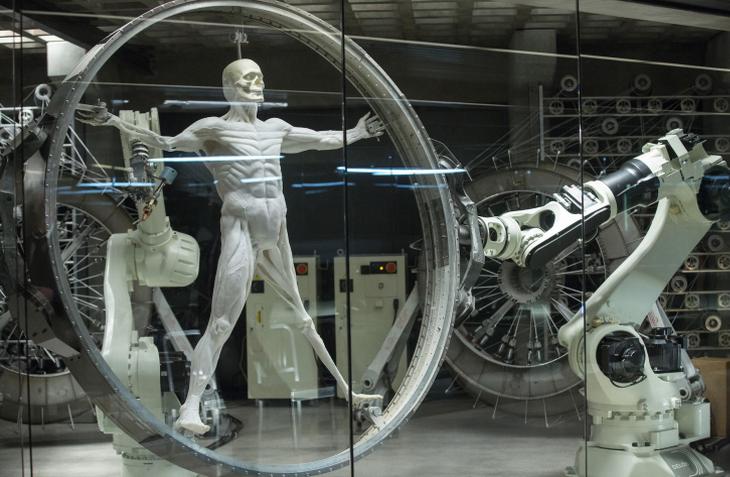
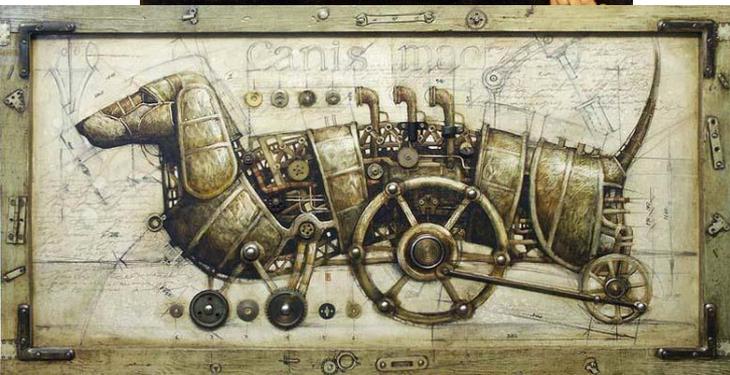
- Las máquinas no pueden ser conscientes
- Hay un límite máximo (y muy bajo) de lo que las máquinas pueden hacer
-
- Límites del mecanicismo:

Descartes

Gödel

Penrose

Searle



si hubiese máquinas tales que tuviesen los órganos y figura exterior de un mono o de otro cualquiera animal, desprovisto de razón, no habría medio alguno que nos permitiera conocer que no son en todo de igual naturaleza que esos animales; mientras que si las hubiera que semejasen a nuestros cuerpos e imitasen nuestras acciones, cuanto fuere moralmente posible, siempre tendríamos dos medios muy ciertos para reconocer que no por eso son hombres verdaderos...

y es el primero, que nunca podrían hacer uso de palabras ni otros signos, componiéndolos, como hacemos nosotros, para declarar nuestros pensamientos a los demás, pues si bien se puede concebir que una máquina esté de tal modo hecha, que profiera palabras, y hasta que las profiera a propósito de acciones corporales que causen alguna alteración en sus órganos, como, *verbi gratia*, si se la toca en una parte, que pregunte lo que se quiere decirle, y si en otra, que grite que se le hace daño, y otras cosas por el mismo estilo, sin embargo, no se concibe que ordene en varios modos las palabras para contestar al sentido de todo lo que en su presencia se diga, como pueden hacerlo aun los más estúpidos de entre los hombres...

y es el segundo que, aun cuando hicieran varias cosas tan bien y acaso mejor que ninguno de nosotros, no dejarían de fallar en otras, por donde se descubriría que no obran por conocimiento, sino sólo por la disposición de sus órganos, pues mientras que la razón es un instrumento universal, que puede servir en todas las coyunturas, esos órganos, en cambio, necesitan una particular disposición para cada acción particular; por donde sucede que es moralmente imposible que haya tantas y tan varias disposiciones en una máquina, que puedan hacerla obrar en todas las ocurrencias de la vida de la manera como la razón nos hace obrar a nosotros (Descartes, 1637, pp. 56-57).

Qué es IA

“El emocionante nuevo esfuerzo para hacer que las computadoras piensen... máquinas con mentes, en el sentido completo y literal.” (Haugeland, 1985)

“El estudio de cómo hacer que las computadoras hagan cosas que, por el momento, las personas hacen mejor.” (Rish and Knight, 1991)

“El estudio de las facultades mentales a través del uso de modelos computacionales.” (Charniak and McDermott, 1985)

“A la IA ... le concierne el comportamiento inteligente en los artefactos” (Nilsson, 1998)

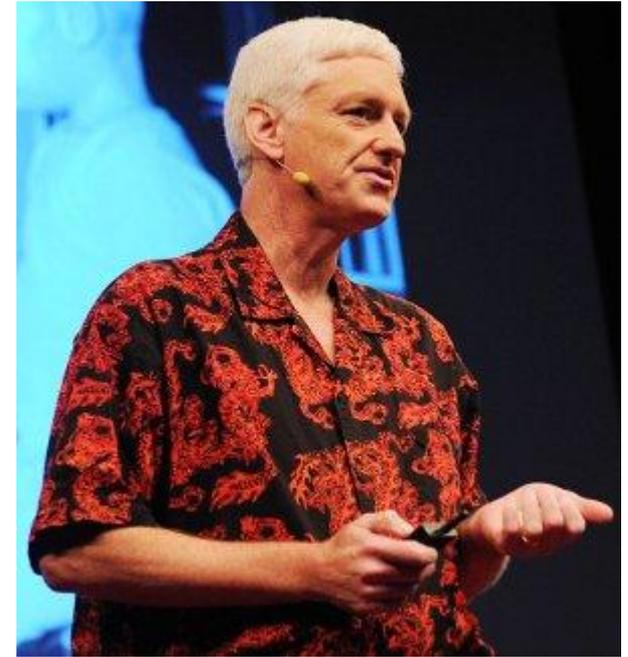
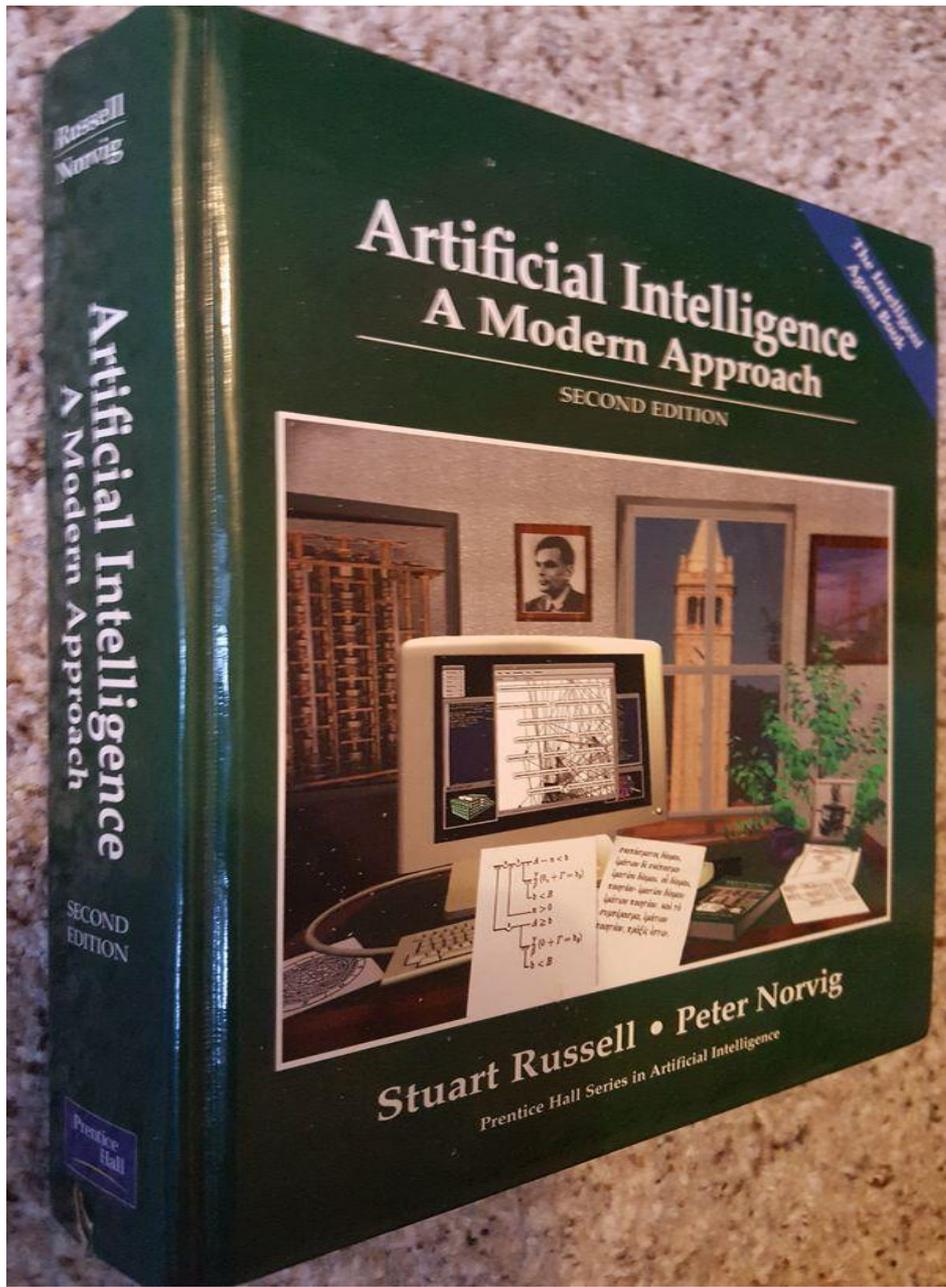
Qué es IA

“El emocionante nuevo esfuerzo para hacer que las computadoras piensen... máquinas con mentes, en el sentido completo y literal.” (Haugeland, 1985) [[Pensar como humanos](#)]

“El estudio de cómo hacer que las computadoras hagan cosas que, por el momento, las personas hacen mejor.” (Rish and Knight, 1991) [[Actuar como humanos](#)]

“El estudio de las facultades mentales a través del uso de modelos computacionales.” (Charniak and McDermott, 1985) [[Pensar racionalmente](#)]

“A la IA ... le concierne el comportamiento inteligente en los artefactos” (Nilsson, 1998) [[Actuar racionalmente](#)]



Campos que hacen a la IA

- Procesamiento de lenguaje natural (NLP): para permitirle comunicarse exitosamente en español, inglés, etc.;
- Representación de conocimiento (knowledge representation) para almacenar lo que sabe o lo que escucha;
- Razonamiento automatizado para usar la información almacenada para contestar preguntas y llegar a nuevas conclusiones;
- Aprendizaje automatizado (machine learning) para adaptarse a nuevas circunstancias y detectar y extrapolar patrones;
- Visión artificial (computer vision) para percibir objetos;
- Robótica para manipular objetos y poder desplazarse en el medio.

Paradigmas de AI (y ML)

Simbólicos

Bayesianos

Evolucionistas

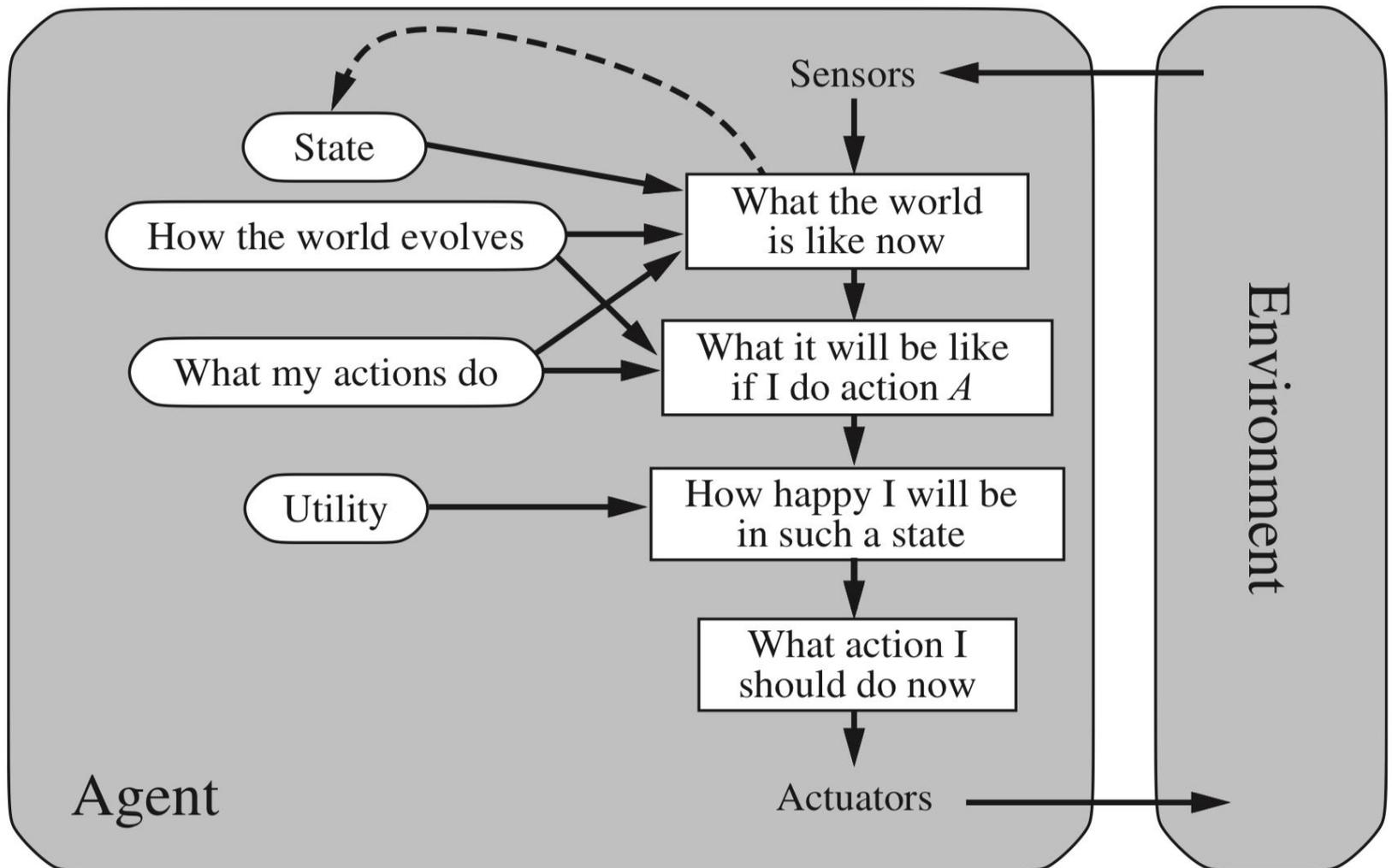
Analogizadores

Conexionistas

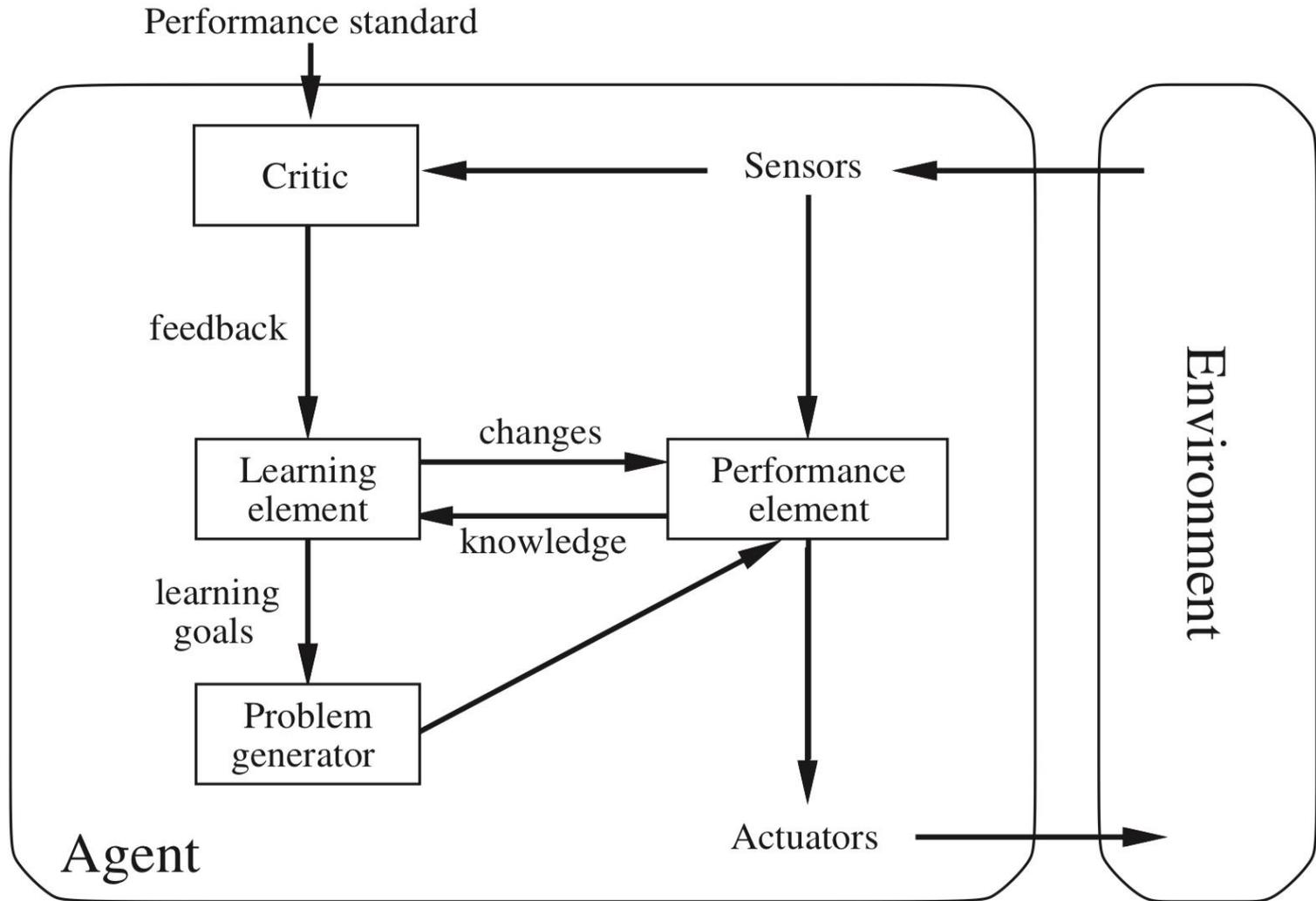
Russell & Norvig

The main unifying theme is the idea of an intelligent agent. We define AI as the study of agents that receive percepts from the environment and perform actions. Each such agent implements a function that maps percept sequences to actions, and we cover different ways to represent these functions (2009, p. 4).

Russell & Norvig

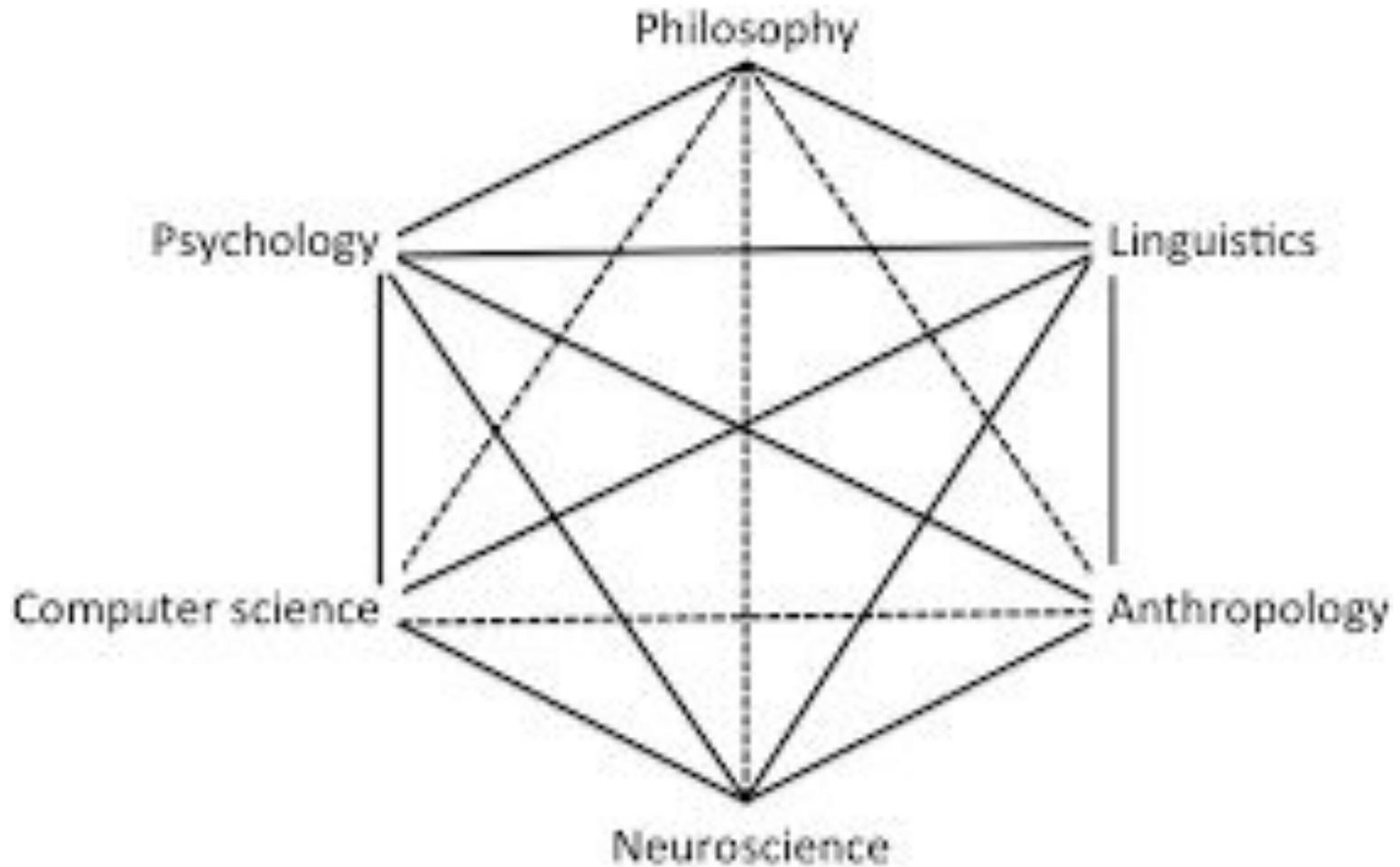


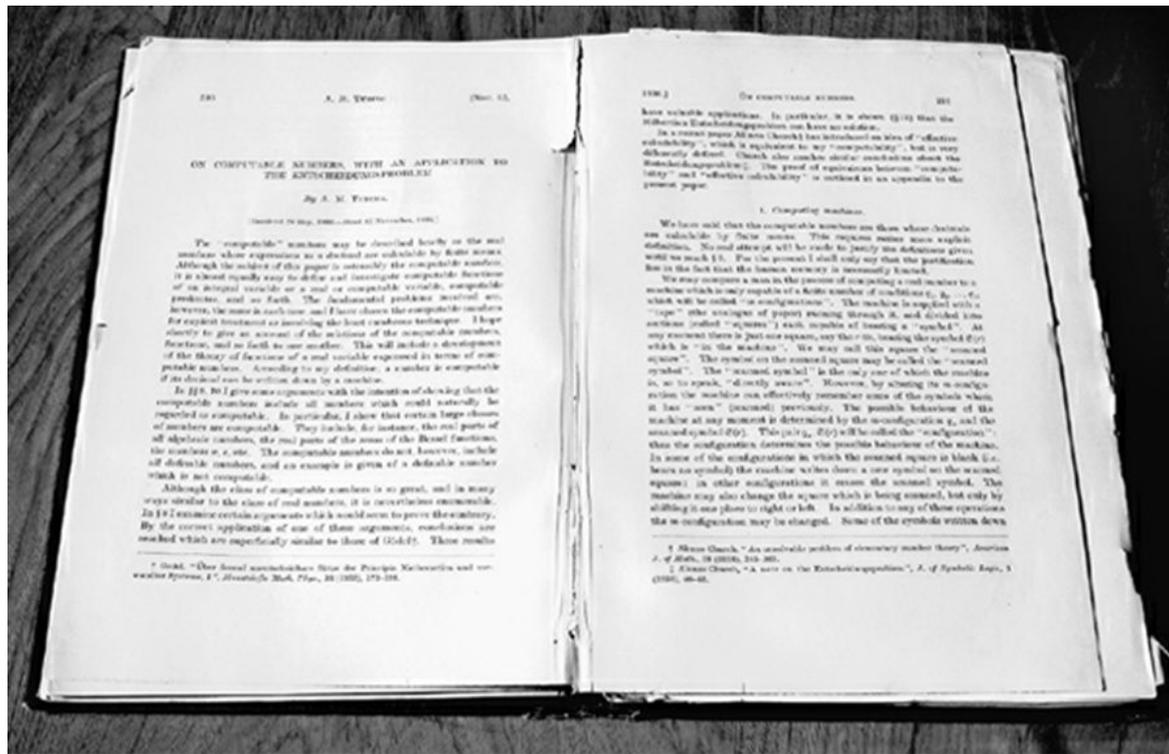
Russell & Norvig



Unless we understand such issues [how machines can learn], we are faced with the daunting task of constructing large commonsense knowledge bases by hand, an approach that has not fared well to date.
(Russell & Norvig 2009, Ch. 27.1)

Ciencias Cognitivas





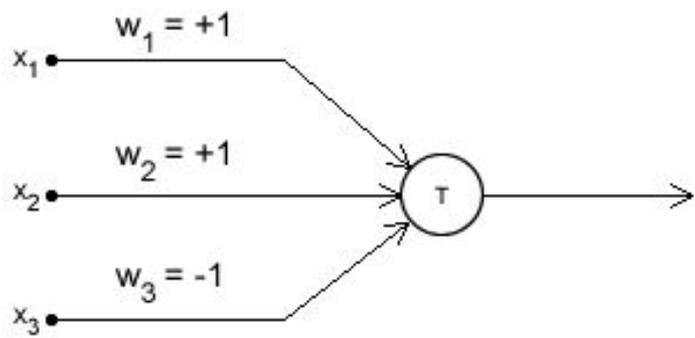
A LOGICAL CALCULUS OF THE
IDEAS IMMANENT IN NERVOUS ACTIVITY

WARREN S. MCCULLOCH AND WALTER PITTS

FROM THE UNIVERSITY OF ILLINOIS, COLLEGE OF MEDICINE,
DEPARTMENT OF PSYCHIATRY AT THE ILLINOIS NEUROPSYCHIATRIC INSTITUTE,
AND THE UNIVERSITY OF CHICAGO

Because of the “all-or-none” character of nervous activity, neural events and the relations among them can be treated by means of propositional logic. It is found that the behavior of every net can be described in these terms, with the addition of more complicated logical means for nets containing circles; and that for any logical expression satisfying certain conditions, one can find a net behaving in the fashion it describes. It is shown that many particular choices among possible neurophysiological assumptions are equivalent, in the sense that for every net behaving under one assumption, there exists another net which behaves under the other and gives the same results, although perhaps not in the same time. Various applications of the calculus are discussed.

- “Cualquier cosa que pueda ser descrita exhaustivamente y de manera no ambigua, cualquier cosa que pueda ser puesta en palabras de manera no ambigua y exhaustiva, es *ipso facto* realizable por una red neural adecuada y finita” (von Neumann 1948)



But one point must be made clear: neither of us conceives the formal equivalence to be a factual explanation. Per contra!--we regard facilitation and extinction as dependent upon continuous changes in threshold related to electrical and chemical variables, such as after-potentials and ionic concentrations; and learning as an enduring change which can survive sleep, anaesthesia, convulsions and coma.

The importance of the formal equivalence lies in this: that the alterations actually underlying facilitation, extinction and learning ~in no way affect the conclusions which follow from the formal treatment of the activity of nervous nets, and the relations of the corresponding propositions remain those of the logic of propositions.

Referencias a máquinas en Alan Turing entre 1936—1952

- (1936) On Computable Numbers, with an Application to the Entscheidungsproblem (Proc. Lond. Math. Soc. (2) 42, 230–265)
- (1939) Systems of Logic Based on Ordinals (Proc. Lond. Math. Soc. (2) 45, 161–228)
- (1947) Lecture to the London Mathematical Society
- (1948) Intelligent Machinery (Informe de Turing para el National Physical Laboratory)
- (1950) Computing Machinery and Intelligence (Mind, 59, 433–460) 1950
- (1951) Can Digital Computers Think? (BBC Third Programme radio broadcast (15 May 1951), transcript edited by B. J. Copeland)
- (1951) Intelligent Machinery: A Heretical Theory (Lecture given to 51 Society in Manchester (c. 1951), transcript edited B. J. Copeland)
- (1952) Can Automatic Calculating Machines Be Said To Think? (Broadcast discussion, BBC Third Programme (14 and 23 Jan. 1952), transcript edited B. J. Copeland)

La conferencia del 1947, es uno de las primeras referencias al problema de modelizar la inteligencia. Aquí las máquinas que aprenden y las máquinas-niños se constituyen en una guía para la elucidación de tipos de mecanismos apropiados. Hacia el final de la *lecture*, Turing escribe un breve apartado sobre inteligencia y máquinas en el que comienza parafraseando a Ada Lovelace: “Se ha dicho que las computadoras sólo pueden producir aquello para lo que han sido instruidas”. Así, todo aquello que no sea producido por una “instrucción” será visto como un error (Turing, 1947, p. 392).

Las máquinas han sido utilizadas, hasta ahora, como “esclavas” y el trabajo que realizan parece que debe ser comprendido por sus usuarios desde el comienzo hasta el fin. ¿Qué ocurriría cuando la guía ahora es una máquina que aprende o una máquina-niño? ¿Sigue valiendo el *dictum* de Lovelace? Debemos comenzar, continúa Turing, por ver de qué manera se pueden modificar las instrucciones de tal forma que su resultado no previsto sea algo más que un mero error. De hecho, es precisamente éste el punto que se vuelve central en los trabajos posteriores sobre inteligencia, que elaboran sobre por qué una máquina de la que se espera que sea infalible no puede ser, a la vez, inteligente.

Turing equipara a la corteza del cerebro de un niño con una máquina desorganizada que es organizada por los estímulos que recibe del medio. De hecho no ve ninguna imposibilidad (más allá de las dificultades propias de lograrlo) de entrenar dicho cerebro para que se comporte como una máquina universal automática propiamente dicha.

Lo que Turing presenta aquí es, efectivamente, la primera descripción en la literatura de lo que en la actualidad se conoce como *machine learning*, el tipo de máquinas o programas que puede aprender de su comportamiento pasado para mejorar el ajuste con una función particular definida. La otra analogía a la que recurre Turing en esta formulación, aunque sin explicitarla, es claramente la de la selección natural, mecanismo que permite fijar mutaciones que se evalúan, a la larga, como “buenas”.

Por otro lado, Turing es capaz de expresar con cierta claridad lo fundamental que es para que algo sea una máquina la integración entre los aspectos físicos y lógicos. Dada una arquitectura de una complejidad mínima para una máquina, su capacidad de generar comportamientos puede ser extendida mediante la adición de una memoria que tenga la suficiente velocidad de acceso para permitir que en el tiempo de acción requerido la parte netamente lógica o programable *compense* las deficiencias de la arquitectura propiamente dicha.

Este balance entre arquitectura fija, compensabilidad por reprogramación y plasticidad general es lo que habrían logrado la evolución y la selección natural con el cerebro humano, y es lo que debería ser implementado de alguna forma si lo que se quiere es simular comportamiento inteligente.

Si una máquina puede pensar, puede llegar a pensar de manera más inteligente que nosotros, y entonces, ¿a dónde quedaríamos? Incluso si pudiéramos mantener a las máquinas en una posición de servidumbre, por ejemplo apagándolas en momentos estratégicos, deberíamos vernos, como especie, en una posición de humildad... Sentimos que si esto realmente pasara no sería sino hasta dentro de varios millones de años. Pero este nuevo peligro está mucho más cerca. Si realmente sucede, ocurrirá casi con certeza en el próximo milenio. Es una posibilidad remota pero no astronómicamente remota, y ciertamente es algo que nos puede producir ansiedad.

Se suele decir ... para ofrecer un dejo de esperanza... que alguna característica intrínsecamente humana nunca podrá ser imitada por una máquina... No puedo ofrecer ninguna clase de consuelo semejante, porque creo que ningún límite así se puede trazar.

(Turing, 1951, pp. 662-663)

Cybernetics

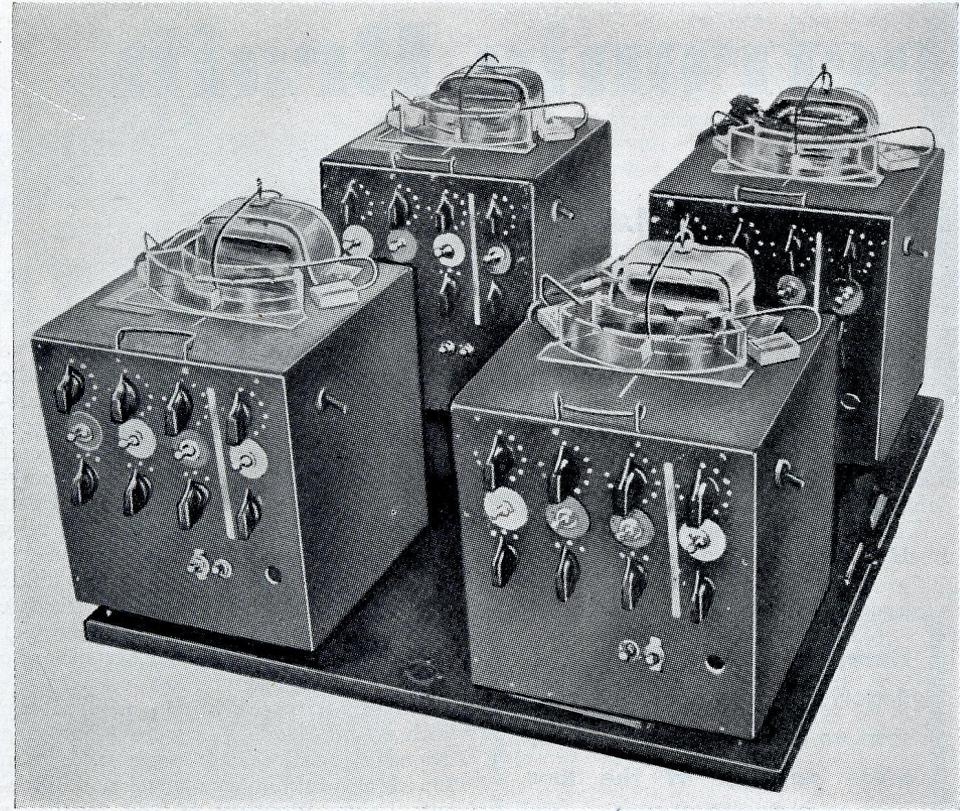


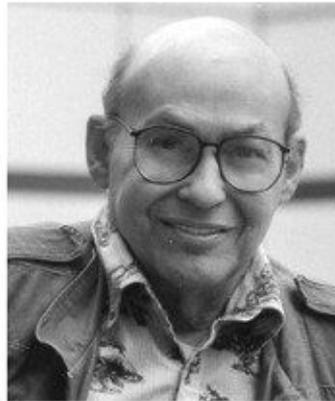
Fig. 1—The homeostat, with its four units, each one of which reacts on all the others.

En otras palabras, la máquina empieza a cazar una combinación de configuraciones de los uniselectores que dan un sistema estable, esto es, que dan las retroalimentaciones internas correctas. Cuando encuentra una combinación con las retroalimentaciones correctas, la mantiene y demuestra que ha armado el sistema de retroalimentación que resulta en un mantenimiento coordinado de sus variables en un valor óptimo, como un ser vivo. El punto importante es que encuentra su propia disposición [arreglo] de retroalimentación, el diseñador meramente se limitó a proveerle de mucha variedad (Ashby, 1949, p. 78).

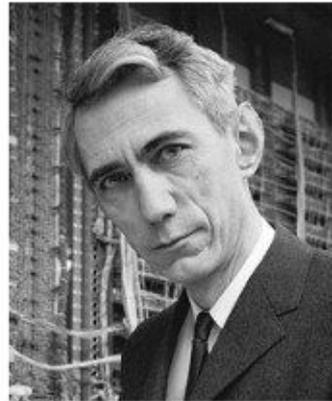
1956 Dartmouth Conference: The Founding Fathers of AI



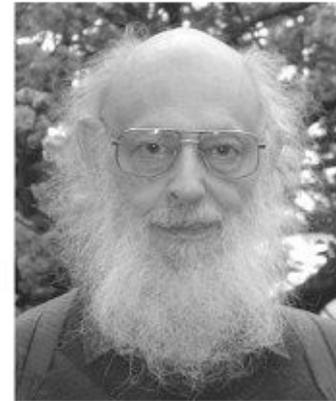
John MacCarthy



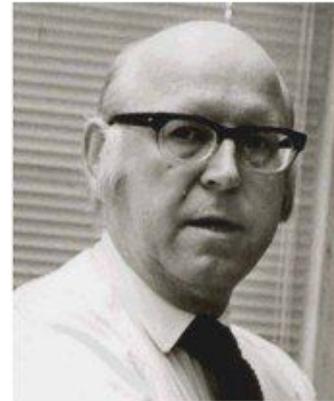
Marvin Minsky



Claude Shannon



Ray Solomonoff



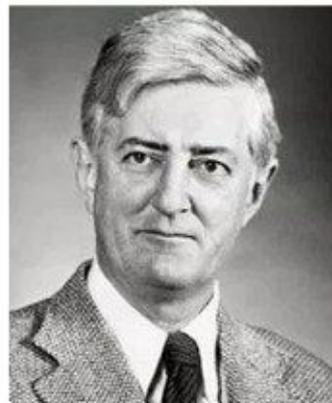
Alan Newell



Herbert Simon



Arthur Samuel



Oliver Selfridge



Nathaniel Rochester



Trenchard More

Solicitud para la conferencia

Proponemos que se lleve a cabo un estudio de la inteligencia artificial de dos meses y diez hombres en el verano de 1956 en el Dartmouth College de Hanover, New Hampshire.

El estudio es para seguir adelante sobre la base de la conjetura de que cualquier aspecto del aprendizaje o cualquier otro rasgo de inteligencia puede en principio ser descrita de forma tan precisa que puede hacerse que una máquina lo simule.

Se hará un intento de descubrir cómo hacer que las máquinas usen el lenguaje, formen abstracciones y conceptos, resuelvan clases de problemas hasta ahora reservados para humanos, y se mejoren a sí mismas.

THE LOGIC THEORY MACHINE
A COMPLEX INFORMATION PROCESSING SYSTEM

by

Allen Newell and Herbert A. Simon

P-868

June 15, 1956

II

The Logic Theory Machine

In the language we have constructed, we have variables (atomic sentences): p, q, r, A, B, C, \dots and connectives: \neg (not), \vee (or), \rightarrow (implies). The connectives are used to combine the variables into expressions (molecular sentences). We have already considered one example of an expression:

1.7

$\neg p \rightarrow q \vee \neg p$

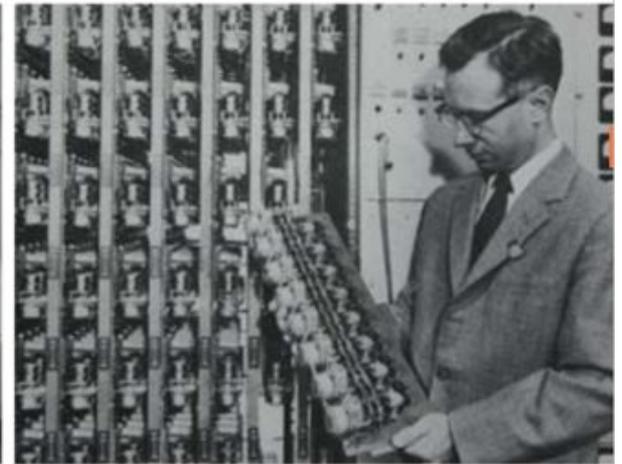
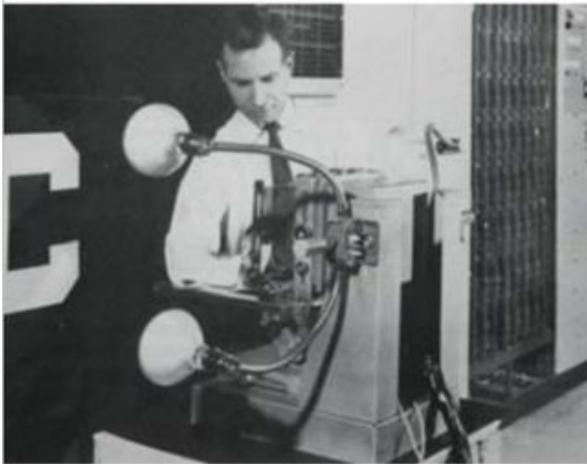
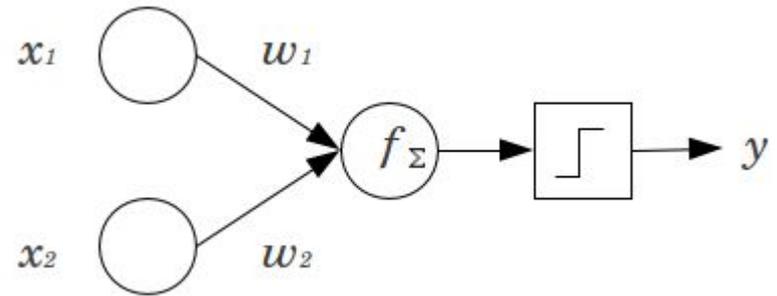
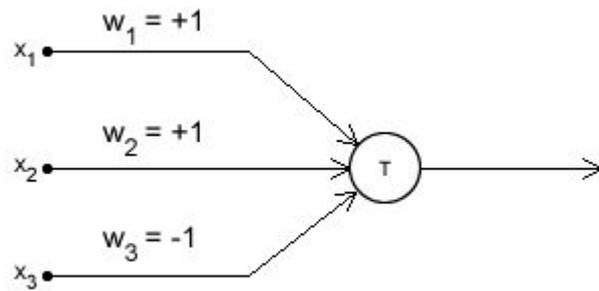
The task set for LT will be to prove that certain expressions are theorems — that is, that they can be derived by application of specified rules of inference from a set of primitive sentences or axioms.

Herbert Simon in 1957

No es mi objetivo sorprenderlos o shockearlos, pero la manera más simple en la que puedo resumir es decir que hoy en día hay máquinas en el mundo que piensan, aprenden y crean. Además, su habilidad para hacer estas cosas va a incrementar rápidamente hasta que -en un futuro cercano- el rango de problemas que pueden manejar va a ser coextensivo con el rango en el que la mente humana ha sido aplicada.

Predicciones: en 10 años nivel campeón en ajedrez, teoremas matemáticos importantes demostrados por una computadora

Perceptron



Rosenblatt y el Perceptron

El punto clave que le permitirá hablar de aprendizaje es lo que señala inmediatamente a continuación, que es que el “sistema propuesto depende de principios probabilísticos para su operación, en lugar de deterministas, y gana su confiabilidad de las propiedades de las mediciones estadística obtenidas de una gran población de elementos” (Rosenblatt, 1957, p. 2).

Electronic 'Brain' Teaches Itself

The Navy last week demonstrated the embryo of an electronic computer named the Perceptron which, when completed in about a year, is expected to be the first non-living mechanism able to "perceive, recognize and identify its surroundings without human training or control." Navy officers demonstrating a preliminary form of the device in Washington said they hesitated to call it a machine because it is so much like a "human being without life."

Dr. Frank Rosenblatt, research psychologist at the Cornell Aeronautical Laboratory, Inc., Buffalo, N. Y., designer of the Perceptron, conducted the demonstration. The machine, he said, would be the first electronic device to think as the human brain. Like humans, Perceptron will make mistakes at first, "but it will grow wiser as it gains experience," he said.

The first Perceptron, to cost about \$100,000, will have about 1,000 electronic "association cells" receiving electrical impulses from an eyelike scanning device with 400 photocells. The human brain has ten billion responsive cells, including 100,000,000 connections with the eye.

Difference Recognized

The concept of the Perceptron was demonstrated on the Weather Bureau's \$2,000,000 IBM 704 computer. In one experiment, the 704 computer was shown 100 squares situated at random either on the left or the right side of a field. In 100 trials, it was able to "say" correctly ninety-seven times whether a square was situated on the right or left. Dr. Rosenblatt said that after having seen only thirty to forty squares the device had learned to

recognize the difference between right and left, almost the way a child learns.

When fully developed, the Perceptron will be designed to remember images and information it has perceived itself, whereas ordinary computers remember only what is fed into them on punch cards or magnetic tape.

Later Perceptrons, Dr. Rosenblatt said, will be able to recognize people and call out their names. Printed pages, longhand letters and even speech commands are within its reach. Only one more step of development, a difficult step, he said, is needed for the device to hear speech in one language and instantly translate it to speech or writing in another language.

Self-Reproduction

In principle, Dr. Rosenblatt said, it would be possible to build Perceptrons that could reproduce themselves on an assembly line and which would be "conscious" of their existence.

Perceptron, it was pointed out, needs no "priming." It is not necessary to introduce it to surroundings and circumstances, record the data involved and then store them for future comparison as is the case with present "mechanical brains." It literally teaches itself to recognize objects the first time it encounters them. It uses a camera-eye lens to scan objects or survey situations, and an electrical impulse system, patterned point-by-point after the human brain does the interpreting.

The Navy said it would use the principle to build the first Perceptron "thinking machines" that will be able to read or write.

1960...

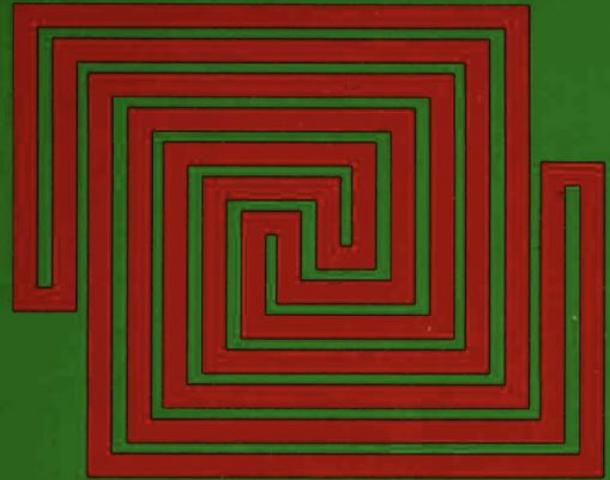
La primera clase de dificultades surgió precisamente porque los primeros programas no sabían nada del tema sobre el que trabajaban, su éxito sólo ocurría mediante simples manipulaciones sintácticas.

El segundo problema en la escalabilidad de las soluciones. Micromundos, todavía no existía una teoría de la complejidad computacional

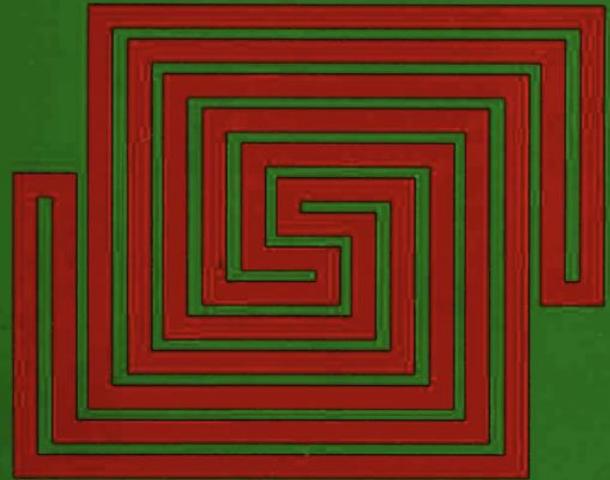
1969



Expanded Edition



Perceptrons



Marvin L. Minsky
Seymour A. Papert

Cognitron: A Self-organizing Multilayered Neural Network

Kunihiko Fukushima

NHK Broadcasting Science Research Laboratories, Kinuta, Setagaya, Tokyo, Japan

Received: February 4, 1975

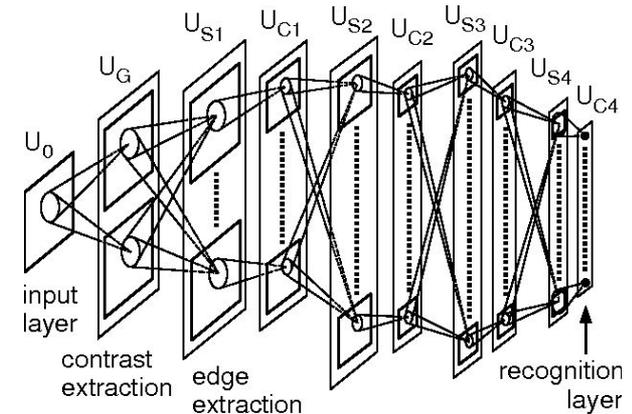
Abstract

A new hypothesis for the organization of synapses between neurons is proposed: "The synapse from neuron x to neuron y is reinforced when x fires provided that no neuron in the vicinity of y is firing stronger than y ". By introducing this hypothesis, a new algorithm with which a multilayered neural network is effectively organized can be deduced. A self-organizing multilayered neural network, which is named "cognitron", is constructed following this algorithm, and is simulated on a digital computer. Unlike the organization of a usual brain models such as a three-layered perceptron, the self-organization of a cognitron progresses favorably without having a "teacher" which instructs in all particulars how the individual cells respond. After repetitive presentations of several stimulus patterns, the cognitron is self-organized in such a way that the receptive fields of the cells become relatively larger in a deeper layer. Each cell in the final layer integrates the information from whole parts of the first layer and selectively responds to a specific stimulus pattern or a feature.

At present, however, the algorithm with which a neural network is self-organized is not known. Although several hypothesis for it have been proposed, none of them has been physiologically substantiated.

The three-layered perceptron proposed by Rosenblatt (1962) is one of the examples of the brain models based on such hypotheses. For a while after the perceptron was proposed, its capability for information processing was greatly expected, and many research works on it have been made. With the progress of the researches, however, it was gradually revealed that the capability of the perceptron is not so large as it had been expected at the beginning.

Although the perceptron consists of only three layers of neurons, it is known that the capability of a layered neural network is greatly enlarged if the number



Fukushima, K. (1975). Cognitron: A self-organizing multilayered neural network. *Biological Cybernetics*, 20(3-4), 121-136.

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36(4), 193-202.

Sistemas Expertos

Un sistema experto es un sistema computacional que emula el proceso de toma de decisiones de un experto humano. Se diseñan para resolver problemas complejos razonando a través de un cuerpo de conocimiento, representado principalmente mediante reglas si-entonces.

Los sistemas expertos fueron los primeros programas de inteligencia artificial exitosos.

DENDRAL

MYCIN

Neural networks and physical systems with emergent collective computational abilities

(associative memory/parallel processing/categorization/content-addressable memory/fail-soft devices)

J. J. HOPFIELD

Division of Chemistry and Biology, California Institute of Technology, Pasadena, California 91125; and Bell Laboratories, Murray Hill, New Jersey 07974

Contributed by John J. Hopfield, January 15, 1982

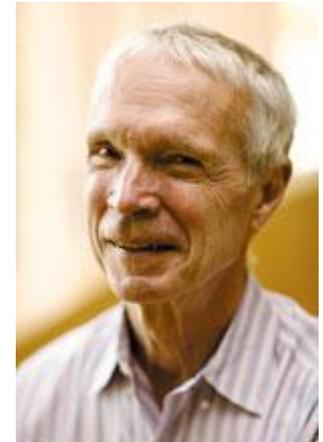
ABSTRACT Computational properties of use to biological organisms or to the construction of computers can emerge as collective properties of systems having a large number of simple equivalent components (or neurons). The physical meaning of content-addressable memory is described by an appropriate phase space flow of the state of a system. A model of such a system is given, based on aspects of neurobiology but readily adapted to integrated circuits. The collective properties of this model produce a content-addressable memory which correctly yields an entire memory from any subpart of sufficient size. The algorithm for the time evolution of the state of the system is based on asynchronous parallel processing. Additional emergent collective properties include some capacity for generalization, familiarity recognition, categorization, error correction, and time sequence retention. The collective properties are only weakly sensitive to details of the modeling or the failure of individual devices.

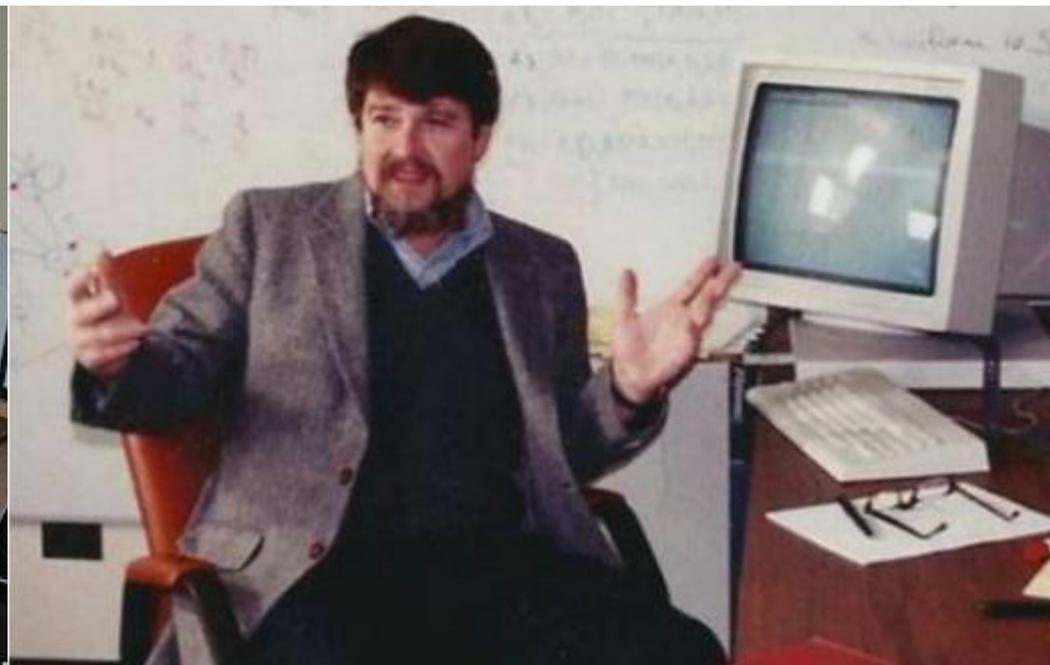
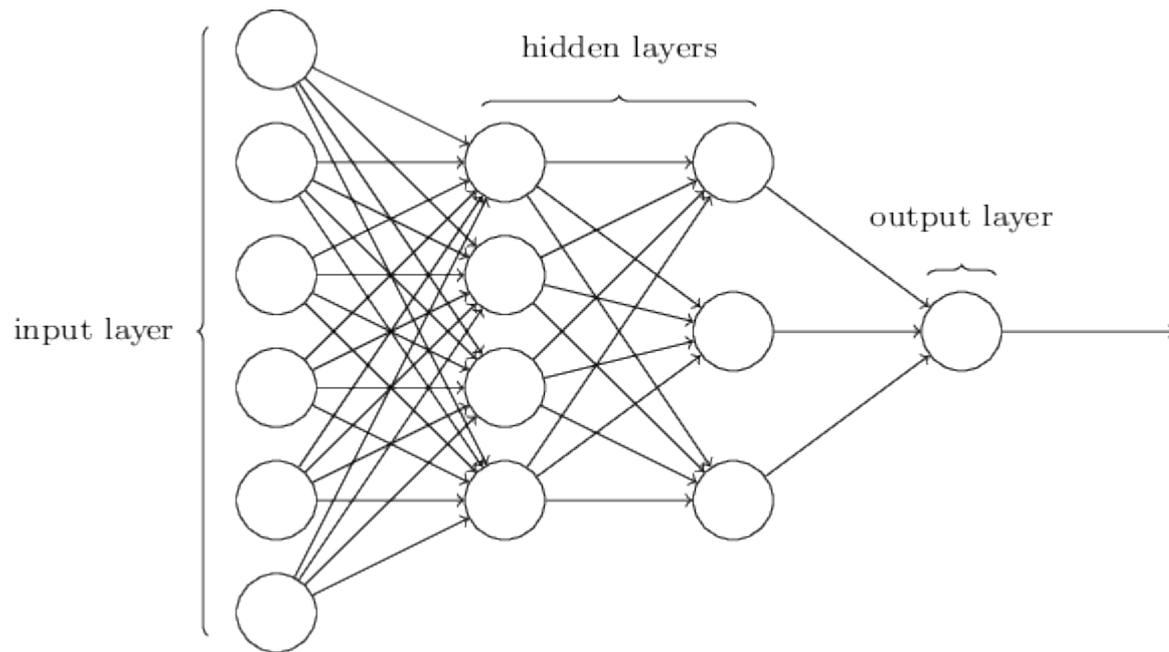
calized content-addressable memory or categorizer using extensive asynchronous parallel processing.

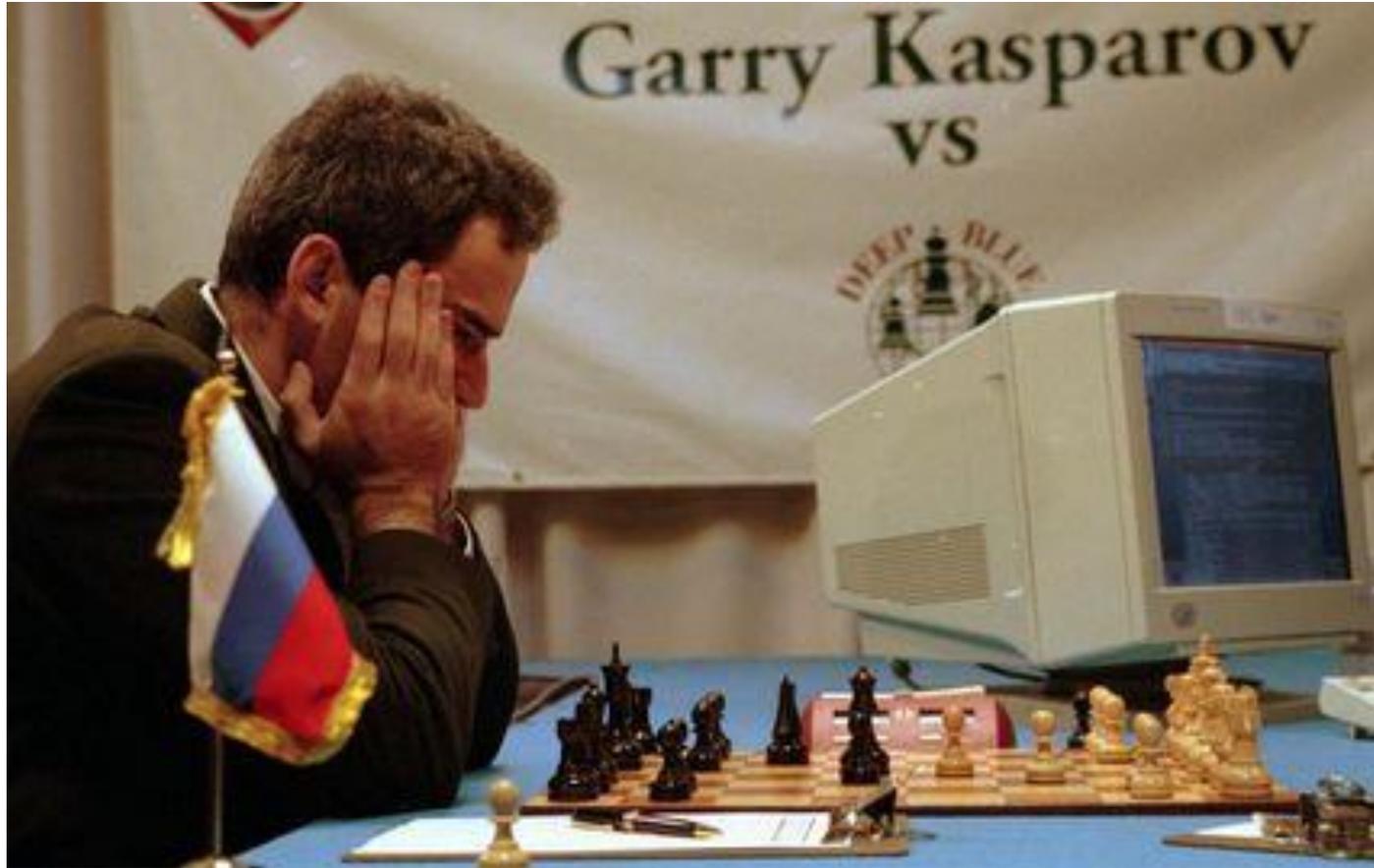
The general content-addressable memory of a physical system

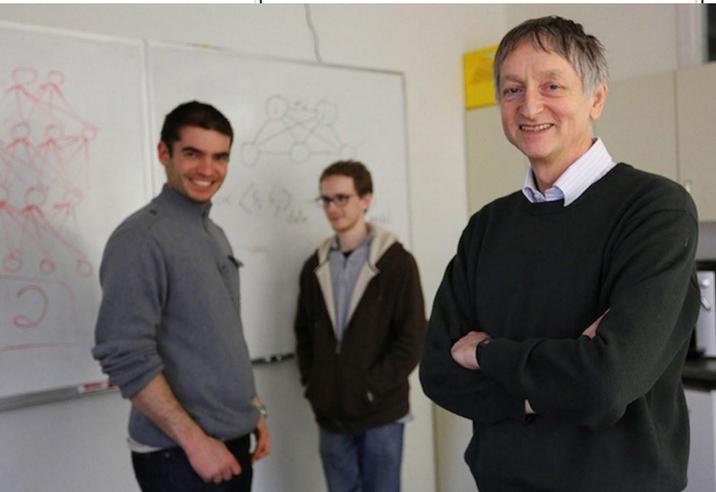
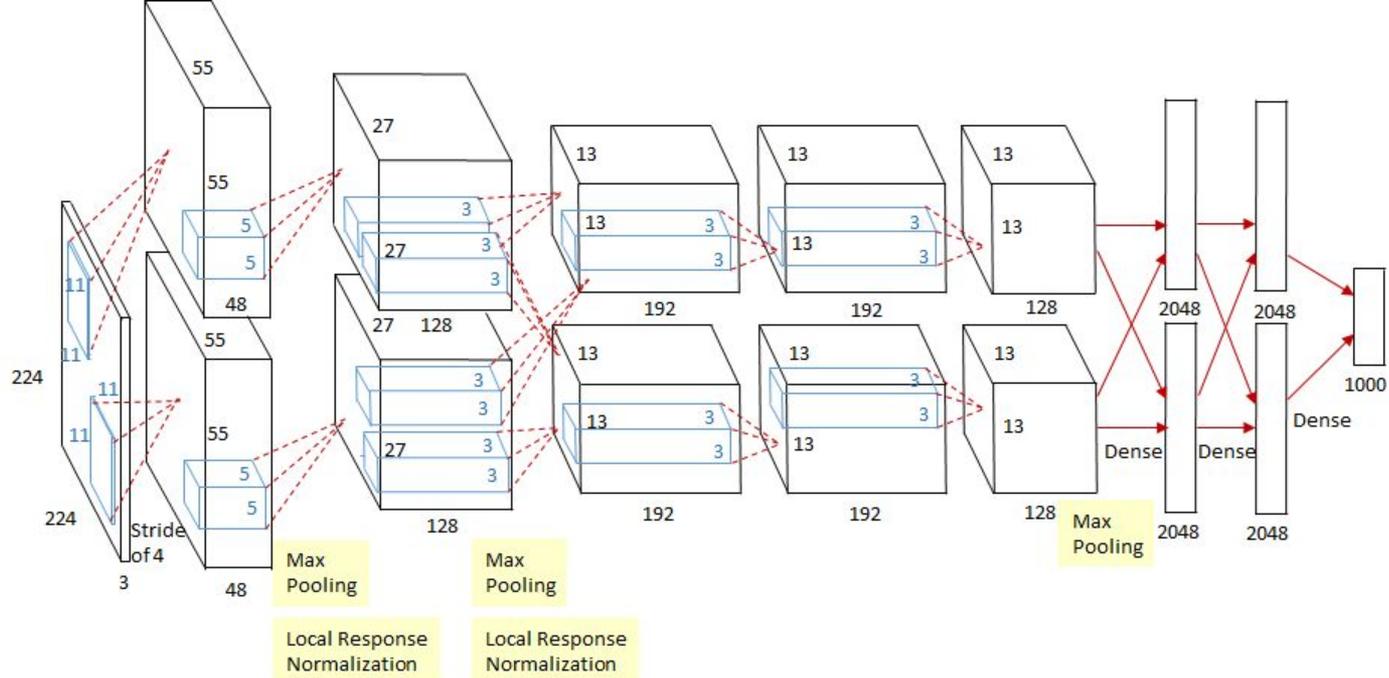
Suppose that an item stored in memory is "H. A. Kramers & G. H. Wannier *Phys. Rev.* **60**, 252 (1941)." A general content-addressable memory would be capable of retrieving this entire memory item on the basis of sufficient partial information. The input "& Wannier, (1941)" might suffice. An ideal memory could deal with errors and retrieve this reference even from the input "Vannier, (1941)". In computers, only relatively simple forms of content-addressable memory have been made in hardware (10, 11). Sophisticated ideas like error correction in accessing information are usually introduced as software (10).

There are classes of physical systems whose spontaneous be-









Alex Krizhevsky, Ilya Sutskever,
Geoffrey Hinton

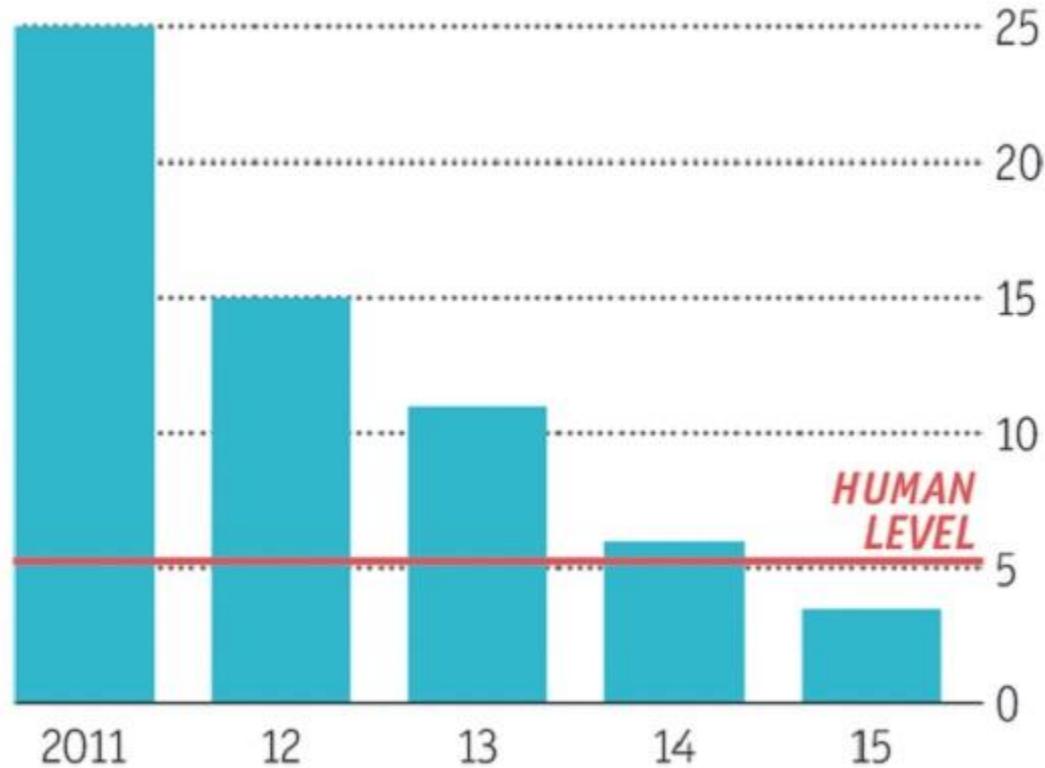
University of Toronto

Our model is a large, deep convolutional neural network trained on raw RGB pixel values. The neural network, which has 60 million parameters and 650,000 neurons, consists of five convolutional layers, some of which are followed by max-pooling layers, and three globally-connected layers with a final 1000-way softmax. It was trained on two NVIDIA GPUs for about a week. To make training faster, we used non-saturating neurons and a very efficient GPU implementation of convolutional nets. To reduce overfitting in the globally-connected layers we employed hidden-unit "dropout", a recently-developed regularization method that proved to be very effective.

According to The Economist, "Suddenly people started to pay attention, not just within the AI community but across the technology industry as a whole."

Ever cleverer

Error rates on ImageNet Visual Recognition Challenge, %

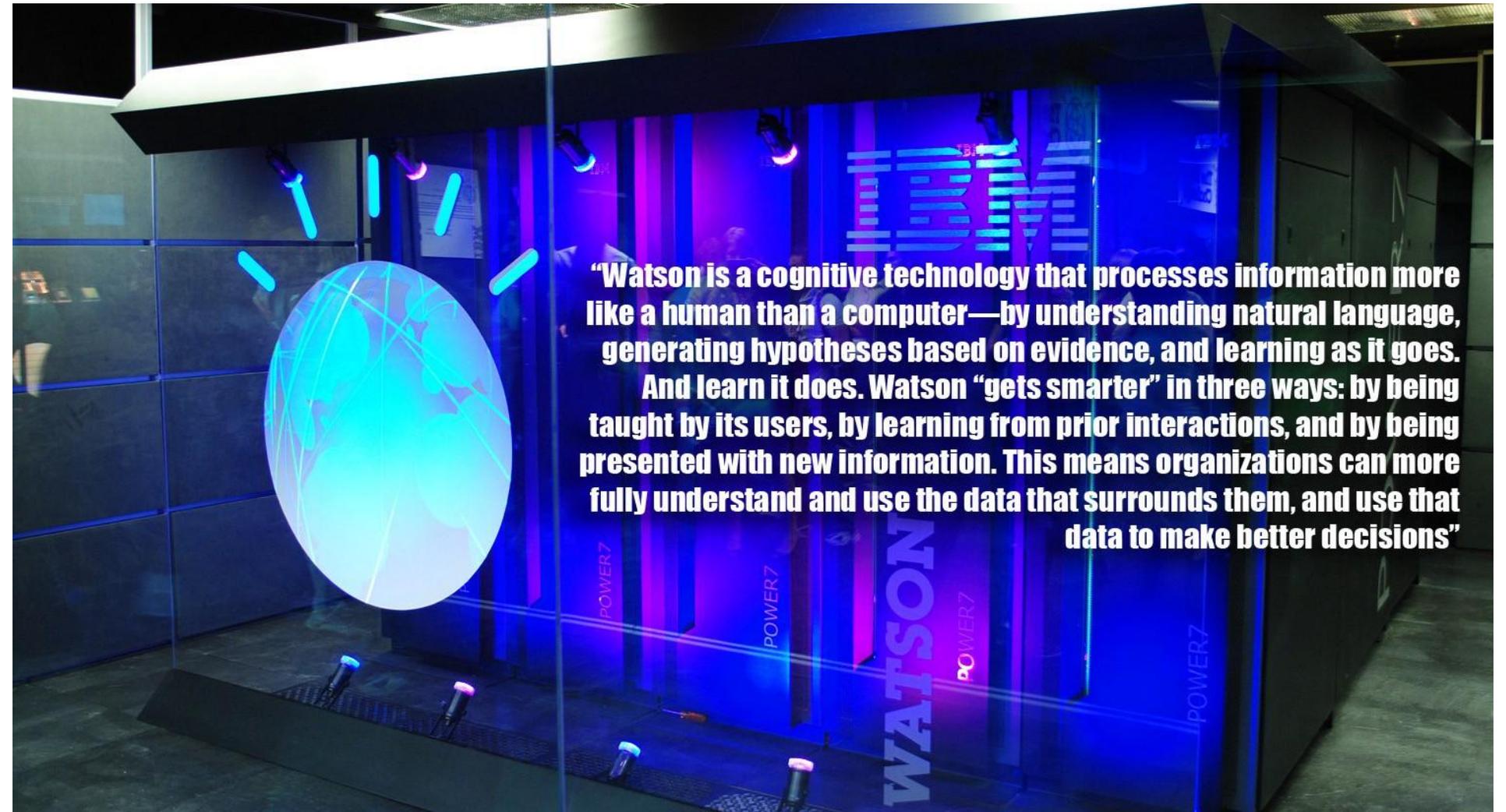


Sources: ImageNet; Stanford Vision Lab

Watson & Jeopardy!



Watson & Jeopardy!



“Watson is a cognitive technology that processes information more like a human than a computer—by understanding natural language, generating hypotheses based on evidence, and learning as it goes. And learn it does. Watson “gets smarter” in three ways: by being taught by its users, by learning from prior interactions, and by being presented with new information. This means organizations can more fully understand and use the data that surrounds them, and use that data to make better decisions”



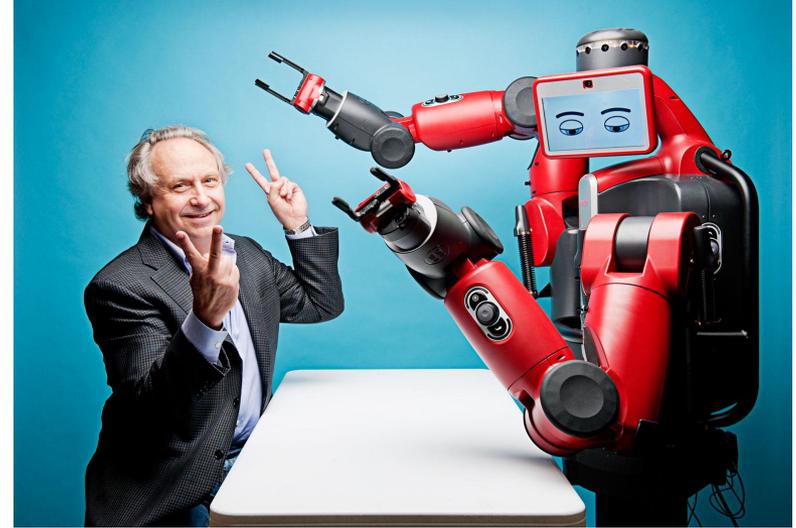
nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

At last — a computer program that can beat a champion Go player **PAGE 484**

ALL SYSTEMS GO

<p>CONSERVATION</p> <h3>SONGBIRDS À LA CARTE</h3> <p><i>Illegal harvest of millions of Mediterranean birds</i></p> <p>PAGE 452</p>	<p>RESEARCH ETHICS</p> <h3>SAFEGUARD TRANSPARENCY</h3> <p><i>Don't let openness backfire on individuals</i></p> <p>PAGE 459</p>	<p>POPULAR SCIENCE</p> <h3>WHEN GENES GOT 'SELFISH'</h3> <p><i>Dawkins's calling card 40 years on</i></p> <p>PAGE 462</p>	<p>NATUREASIA.COM</p> <p>20 January 2016 Vol. 526, No. 7587</p>
--	---	---	--



Intelligence without representation*

Rodney A. Brooks

MIT Artificial Intelligence Laboratory, 545 Technology Square, Rm. 836, Cambridge, MA 02139, USA

Received September 1987

Abstract

Brooks, R.A., Intelligence without representation, *Artificial Intelligence* 47 (1991) 139–159.

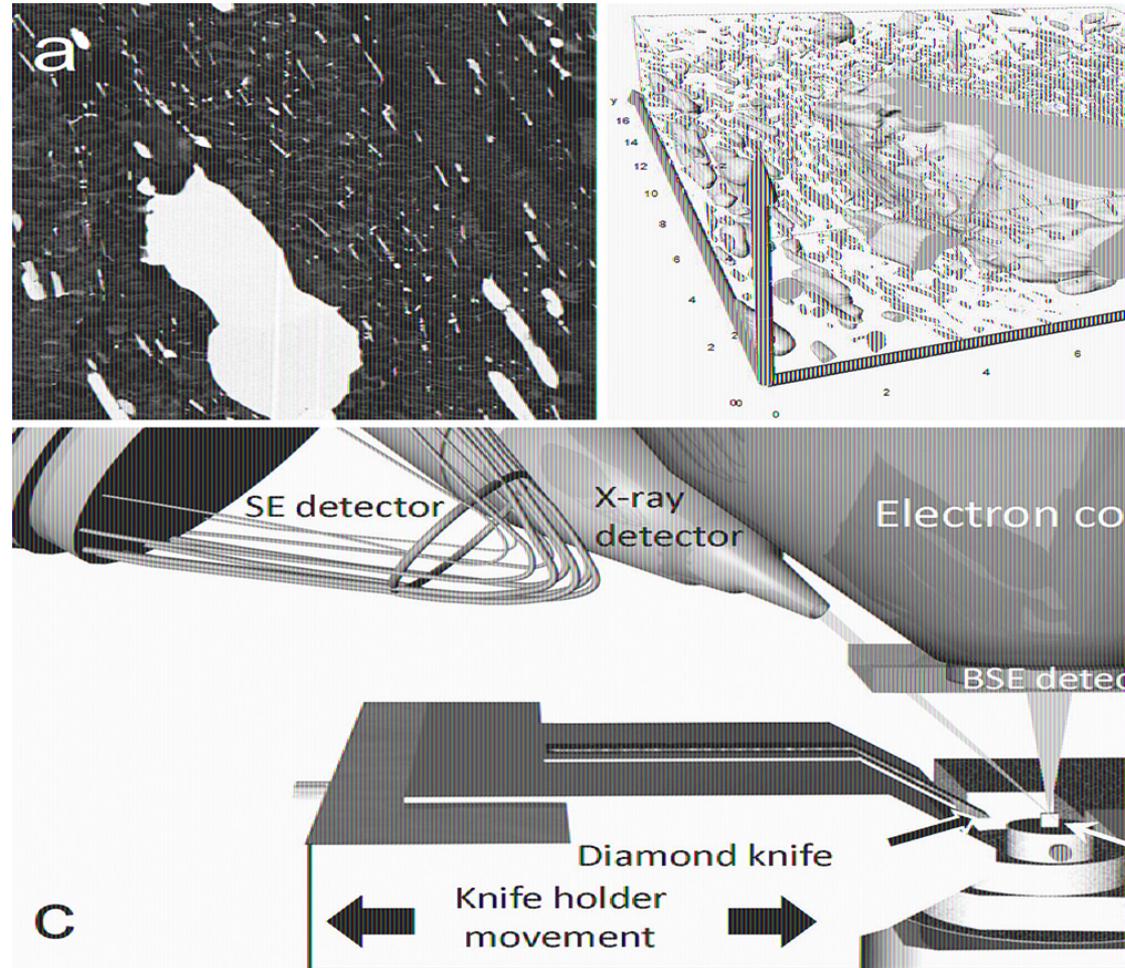
Artificial intelligence research has floundered on the issue of representation. When intelligence is approached in an incremental manner, with strict reliance on interfacing to the real world through perception and action, reliance on representation disappears. In this paper we outline our approach to incrementally building complete intelligent Creatures. The fundamental decomposition of the intelligent system is not into independent information processing units which must interface with each other via representations. Instead, the intelligent system is decomposed into independent and parallel activity producers which all interface directly to the world through perception and action, rather than interface to each other particularly much. The notions of central and peripheral systems evaporate—everything is both central and peripheral. Based on these principles we have built a very successful series of mobile robots which operate without supervision as Creatures in



Robot detects dirt during pickup



Serial block-face scanning electron microscopy (SBEM, SBSEM or SBFSEM)



Denk W, Horstmann H (2004) Serial Block-Face Scanning Electron Microscopy to Reconstruct Three-Dimensional Tissue Nanostructure. *PLoS Biol* 2(11): e329.
doi:10.1371/journal.pbio.0020329

APRENDIZAJE AUTOMÁTICO

“Es el campo de estudio que le brinda a las computadoras la capacidad de aprender sin ser explícitamente programada”

Arthur Samuel, 1959



“Es cualquier cambio en un sistema artificial que le permite realizar mejor el segundo intento en la repetición...”

Herbert Simon, 1989

“Es la ciencia de construir sistemas computacionales que mejoran automáticamente con la experiencia y estudia cuáles son las leyes fundamentales que gobiernan todos los procesos de aprendizaje...”

Tom M. Mitchell, 2006

