

Inteligencia Artificial

T1: Introducción

A PROPOSAL FOR THE DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE

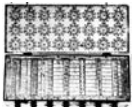
J. McCarthy, Dartmouth College
M. L. Minsky, Harvard University
N. Rochester, I.B.M. Corporation
C.E. Shannon, Bell Telephone Laboratories

August 31, 1955

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

Lecturas :

- CAPÍTULOS 1 + 2 de Russell & Norvig
- CAPÍTULOS 1 + 2 de Nilsson



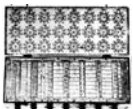
¿Qué es la IA?

⌘ AAAI: "la comprensión científica de los mecanismos sobre los que se basa el pensamiento y comportamiento inteligente, y su implementación en máquinas."

- **Comprender** qué es **inteligencia** (más general que "pensamiento humano").
- **Construir dispositivos inteligentes.**

⌘ Origen:

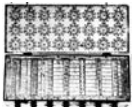
- Los primeros esfuerzos de la IA comienzan poco después de la Segunda Guerra Mundial
- El nombre IA fue sugerido en una conferencia de J. MacCarthy en el taller celebrado en Dartmouth (1956)
[organizadores: J. McCarthy, M. L. Minsky, N. Rochester, y C.E. Shannon].



¿Qué es la IA?

⌘ Comportamiento inteligente:

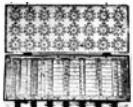
- Percepción
- Razonamiento
- Aprendizaje
- Comunicación
- Interacción con un entorno complejo (incluyendo otros agentes)



¿Qué es la IA?

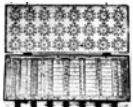
⌘ IA engloba:

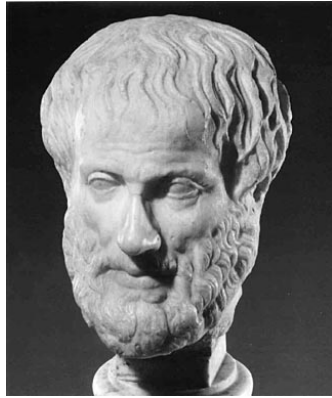
- Representación del conocimiento
- Razonamiento automático
- Aprendizaje automático
- Procesamiento de lenguaje natural
- Percepción
- Visión artificial
- Robótica



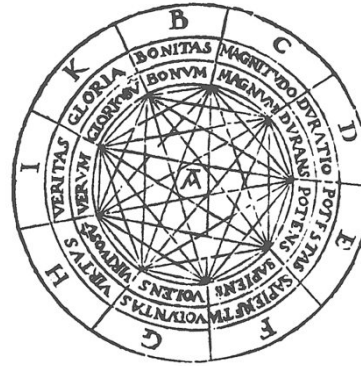
Bases de la IA: Filosofía

- ⌘ ¿Se pueden utilizar reglas formales para extraer conclusiones válidas?
- ⌘ ¿Cómo se genera la inteligencia mental a partir de un cerebro físico?
- ⌘ ¿De dónde viene el conocimiento?
- ⌘ ¿Cómo se pasa del conocimiento a la acción?

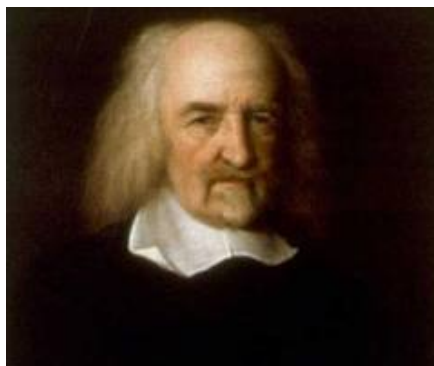




Aristóteles



Ramón Llull

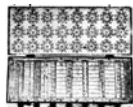


Thomas Hobbes

Gottfried Leibniz



René Descartes



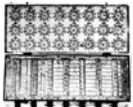
Bases de la IA: Filosofía

⌘ Reglas formales para derivar conclusiones válidas

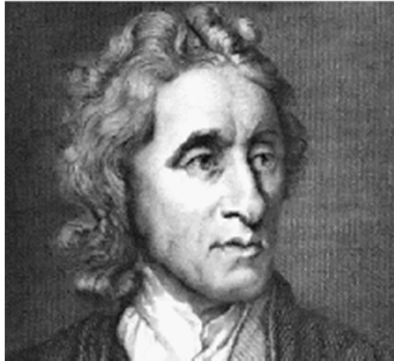
- ⌘ Racionalidad y logos (nacimiento de la filosofía occidental: presocráticos)
- ⌘ Deducciones basadas en silogismos (Aristóteles 384-322 a.C.)
- ⌘ Razonamiento mecánico (Llull c. 1315, Hobbes, 1588-1679)
- ⌘ Razonamiento automático (Leibniz 1646-1716)
- ⌘ Lógica formal

⌘ Problema mente / cuerpo

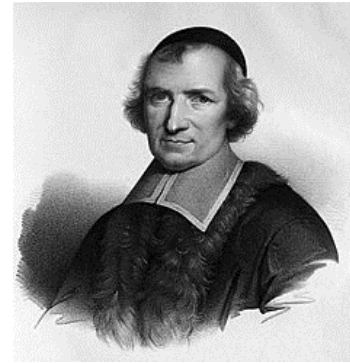
- ⌘ Dualismo (Descartes 1596-1650)
 - ⌘ La inteligencia humana no tiene su origen en procesos físicos
- ⌘ Materialismo: La mente es una producción del cerebro



John Locke

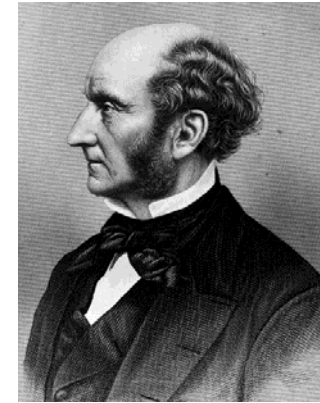


David Hume



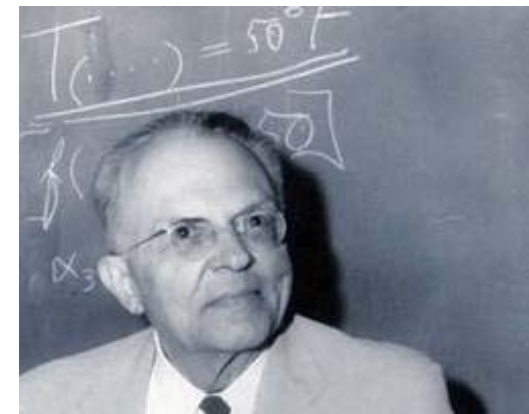
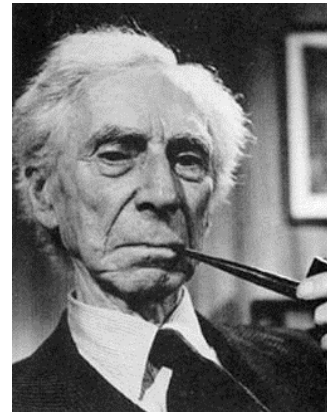
Antoine Arnauld

John Stuart Mill

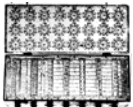


Ludwig Wittgenstein

Bertrand Russell



Rudolf Carnap



Bases de la IA: Filosofía

⌘ Orígenes del conocimiento y su relación con la realidad

⌘ Idealismo

⌘ Nominalismo

⌘ Empirismo (Locke 1632-1704, Hume 1711-1776)

Inducción: Reglas generales pueden ser aprendidas por la asociación repetida entre sus elementos.

⌘ Positivismo lógico (Wittgenstein 1889-1951, Russell 1872-1970, Carnap 1891-1970, y el Círculo de Viena)

Conocimiento = sistema basado en la combinación mediante reglas lógicas de sentencias “atómicas” que corresponden a observaciones empíricas directas sobre hechos simples en el mundo

⌘ Conocimiento y acción

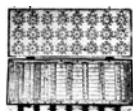
⌘ Análisis basado en objetivos:

⌘ Aristóteles: Conexión lógica entre objetivos y el conocimiento de las consecuencias de la acción

⌘ Antoine Arnauld (1612-1694) [qué hacer cuando varias acciones o ninguna conducen al objetivo]

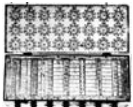
⌘ Utilitarismo (1863) John Stuart Mill 1806-1873

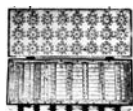
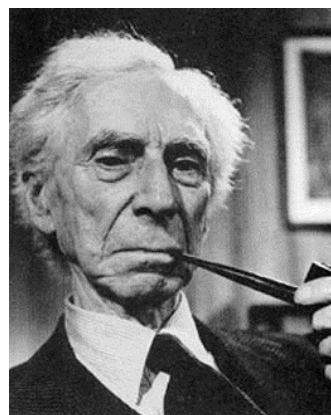
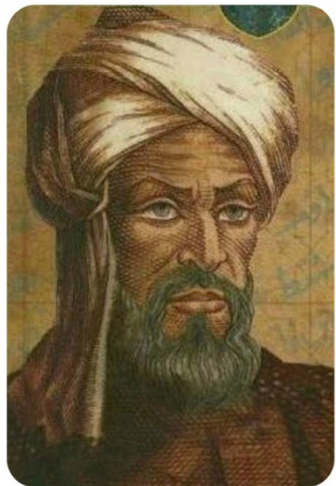
Acciones guiadas por “decisiones racionales”



Bases de la IA: Matemáticas

- ⌘ ¿Qué reglas formales son las adecuadas para obtener conclusiones válidas?
- ⌘ ¿Qué se puede computar?
- ⌘ ¿Cómo razonamos con información cuando existe incertidumbre?





Bases de la IA: Matemáticas

⌘ **Lógica formal**

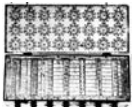
- ⌘ Lógica proposicional (Booleana). (Leibnitz, Boole 1815-1864)
- ⌘ Lógica de primer orden: Incluye objetos + relaciones (Frege 1848-1925)
- ⌘ “Principia Mathematica”, 1910 Russell & WhiteHead
- ⌘ Teoría de referencia (Tarski 1902-1983)
- ⌘ Teorema de Incompletitud (Gödel 1906-1978)

⌘ **Algoritmos**

- ⌘ Antigua Grecia: Euclides (M.c.d.), Eratóstenes (primos)
- ⌘ Al-Khwarizmi (Persia, siglo IX)
- ⌘ Pascal, Leibniz, Ada Lovelace
- ⌘ Intratabilidad: Problemas NP-completos (Cook 1971, Karp 1972)

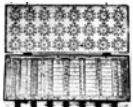
⌘ **Teoría de la Probabilidad:**

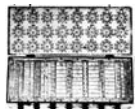
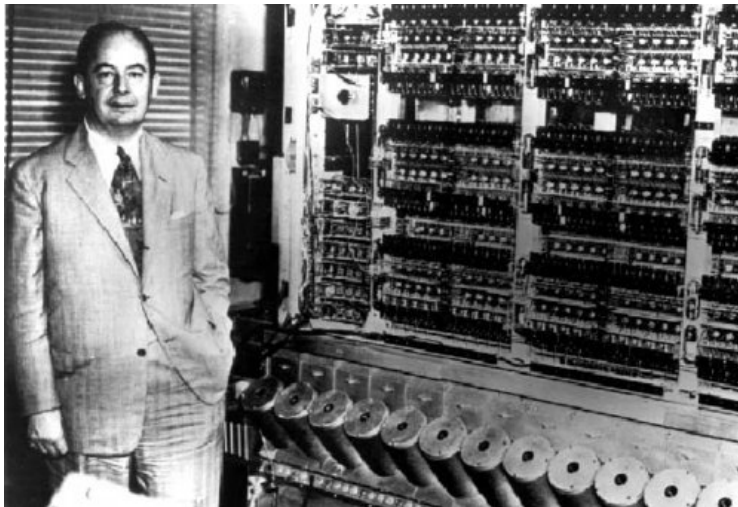
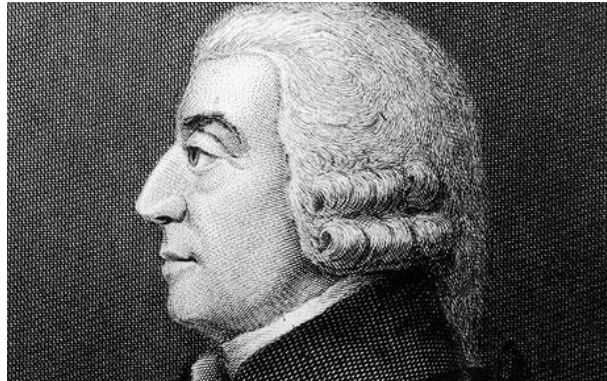
- ⌘ Cardano 1501-1576, Fermat 1601-1665, Pascal, Bernoulli 1654-1705, Bayes 1702-1761 (**regla de Bayes**), Laplace 1749-1827



Bases de la IA: Economía

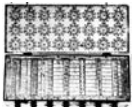
- ⌘ ¿Cómo se debe llevar a cabo el proceso de toma de decisiones para maximizar el rendimiento?
- ⌘ ¿Cómo se deben llevar a cabo acciones cuando otros no colaboran?
- ⌘ ¿Cómo se deben llevar a cabo acciones cuando los resultados se obtienen en un futuro lejano?

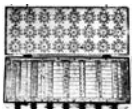
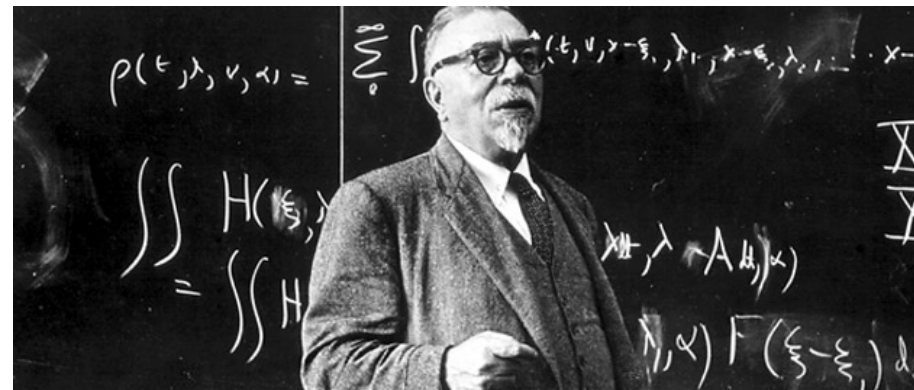
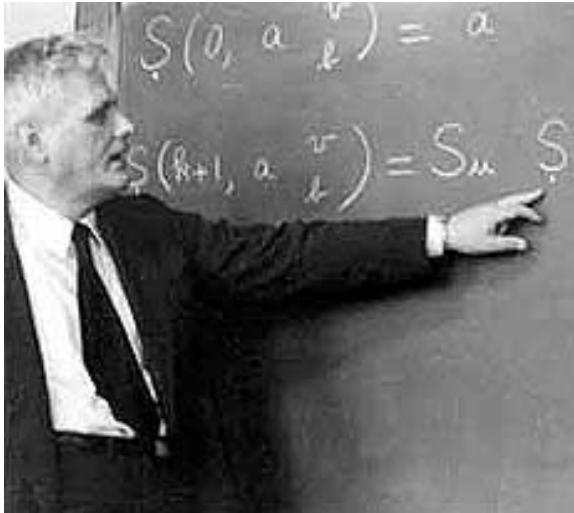




Bases de la IA: Economía

- ⌘ Adam Smith (1723-1790) “An Inquiry into the nature and causes of wealth of nations” (1776)
- ⌘ **Teoría de la utilidad** [Léon Walras 1834-1910, Frank Ramsey, 1931]
- ⌘ **Teoría de la decisión:** Cómo elegir las acciones para lograr los resultados deseados. Combina utilidad + teoría de probabilidad
- ⌘ **Teoría de juegos:** “The theory of Games and Economic Behavior”, 1944 John von Neumann & Oskar Morgenstern
- ⌘ **Investigación operativa:** Cómo tomar decisiones racionales cuando las consecuencias de las acciones no son inmediatas.
Richard Bellman: Procesos de decisión de Markov (1956)
- ⌘ **Racionalidad acotada** [H. Simon 1916-2001]
Satisfacer en lugar de optimizar
- ⌘ **Modelos basados en agentes**





Bases de la IA: Computadores / Cibernética

⌘ Ciencias de la computación:

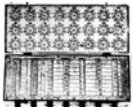
¿Cómo construir un computador eficiente?

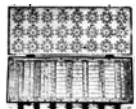
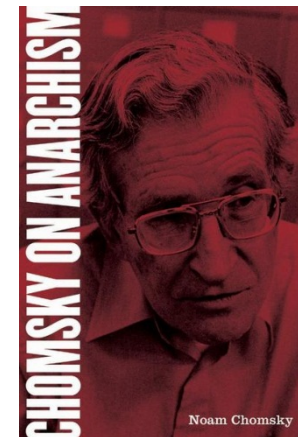
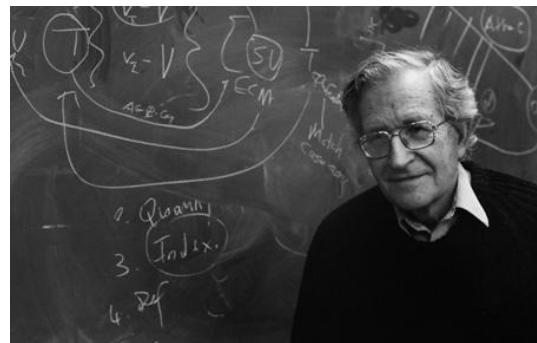
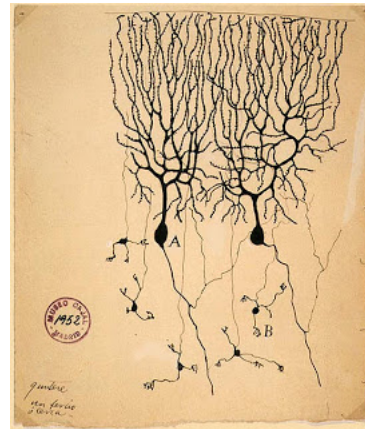
- ⌘ Cálculo automático: Pascal, Leibniz, Neper, Babbage (máquina de diferencias, máquina analítica), Ada Lovelace, von Neumann, Zuse, Turing, computadores modernos (ABC Atanasoff+ Berry, ENIAC, Mauchly y Eckert)
- ⌘ **A. Turing:** Problema de parada, máquinas de Turing, test de Turing
- ⌘ **A. Church:** Cálculo lambda, conjetura de Church-Turing (1935)
- ⌘ **Teoría de autómatas artificiales** (von Neumann)

⌘ Cibernética / teoría de control:

¿Cómo pueden las máquinas operar bajo su propio control?

- ⌘ Sistemas homeostáticos (con retroalimentación estable) :
 - Relojes de agua [Ktesibios de Alexandria 250 a.C.]
 - Termostato [Drebel 1572,1633], Máquina de vapor [Watt, 1736-1819]]
- ⌘ "Cibernética" [1948, Norbert Wiener]]
- ⌘ Teoría de control: Maximizar la función objetivo a lo largo del tiempo





Bases de la IA:

Neurociencia / Psicología / Lingüística

⌘ **Neurociencia:** ¿Cómo funciona el cerebro?

⌘ **Neuronas** [Santiago Ramón y Cajal 1852-1934]

⌘ Medidas de la actividad cerebral

⌘ Electroencefalograma [EEG, Hans Berger, 1929]

⌘ Imagen por resonancia magnética funcional [fMRI, Ogawa et al. 1990]

⌘ Modelos de procesamiento de información en el cerebro

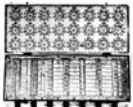
⌘ **Psicología:** ¿Cómo piensan y actúan los humanos y los animales?

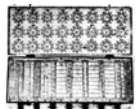
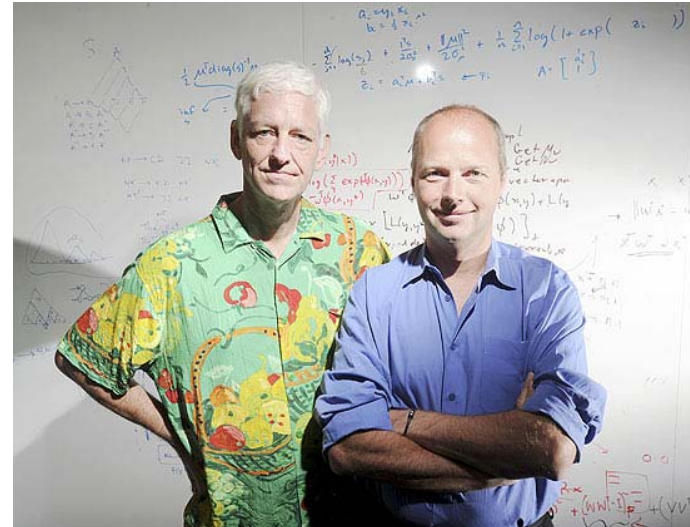
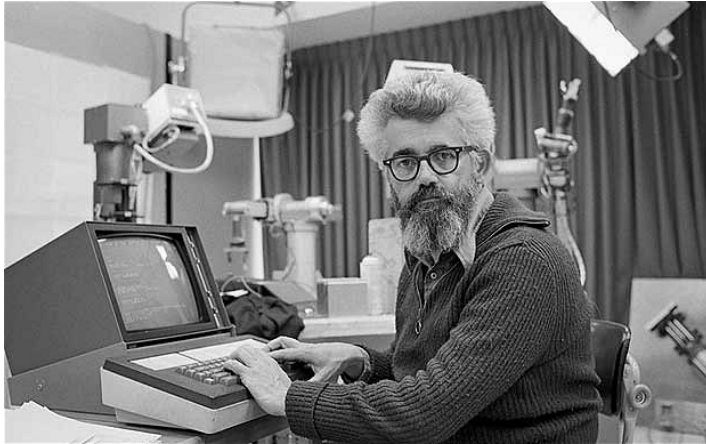
⌘ **Conductismo:** Comportamiento como un conjunto de asociaciones estímulo-respuesta [John Watson 1878-1958, B. F. Skinner “Verbal Behavior”, 1957]

⌘ **Psicología cognitiva:** Modelos de la mente (conocimiento, creencias, intenciones, etc.) como objetos científicos [William James 1842-1919, Bartlett 1886-1960, Craik, Broadbent]

⌘ **Ciencia cognitiva:** simposio en MIT, 1956 [Chomsky, Newell, Simon]

⌘ **Lingüística:** Noam Chomsky, “Estructuras sintácticas” , 1957
Gramática generativa, lingüística computacional





Breve historia de la IA: Pioneros

⌘ Redes neuronales

- ⌘ Neuronas artificiales [McCulloch & Pitts. "A Logical Calculus of the Ideas Immanent in Nervous Activity", 1943]

- ⌘ Aprendizaje Hebbiano [Donald Hebb, 1946]

- ⌘ SNARC: Computadora basada en redes neuronales [Minsky & Edmonds, 1951]

⌘ A. Turing, "Computing machinery and intelligence". Mind, 59, 433:460, 1950

- ⌘ Test de Turing

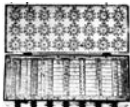
- ⌘ Aprendizaje automático, aprendizaje por refuerzo, algoritmos genéticos

⌘ Claude Shannon: jugar al ajedrez visto como una búsqueda (1950).

⌘ Taller celebrado en Dartmouth, 1956.

Organizadores: McCarthy, Minsky, Rochester y Shannon.

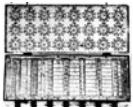
El "Teórico Lógico", creado por Newell & Simon, un programa capaz de demostrar automáticamente teoremas del "Principia Mathematica".



Breve historia de la IA: Trabajos iniciales

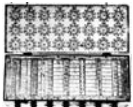
[<http://www.aaai.org/ALTopics/html/history.html>]

- ⌘ General Problem Solver, 1957 Newell & Simon, intenta imitar el razonamiento humano
- ⌘ Grupo de Nathaniel Rochester en IBM
- ⌘ Programas para el juego de damas, Samuel 1952-1962
- ⌘ Demostrador de teoremas de geometría, Gelernter 1959
- ⌘ McCarthy (IA logicista)
 - ⌘ Lisp, 1958
 - ⌘ Generador de Consejos, 1958: Representación y razonamiento
 - ⌘ 1963, Laboratorio de IA de Stanford.
Proyecto para integrar razonamiento lógico + acción
- ⌘ Método de resolución para lógica de primer orden (Robinson, 1965)



Breve historia de la IA: Trabajos iniciales

- ⌘ Marvin Minsky (en MIT AI Lab desde 1958)
 - ⌘ Importancia de la heurística, problemas de implementación (en contraste con enfoques puramente formales)
 - ⌘ Micromundos: SAINT (Slagle, 1961) para cálculo integral; ANALOGY (1963) para geometría; STUDENT (Bobrow, 1964) para álgebra
 - ⌘ Mundo de bloques [Winston 1970, aprendizaje; Huffman 1971, visión; Winograd: comprensión de lenguaje natural, 1972; Fahlman 1974, planificación; Waltz 1975, visión + propagación de restricciones]
 - ⌘ Marcos (Minsky, 1975)
 - ⌘ The society of mind (Minsky, 1987)
- ⌘ Redes neuronales
 - ⌘ Adaline (Widrow & Hoff, 1960)
 - ⌘ Perceptrón lineal + algoritmo de aprendizaje (Rosenblatt, 1962)
 - ⌘ Las redes neuronales pueden representar conceptos (Winograd & Cohen, 1963)
- ⌘ Desarrollo de PROLOG (1972) por Alain Colmerauer



Breve historia de la IA: Crisis + SBCs

⌘ Dificultades para la IA

- ⌘ Incorporación de conocimiento sobre el mundo

- ⌘ Explosión combinatoria (informe de Lighthill, 1973)

 - Ej. Dificultades en los experimentos iniciales sobre Algoritmos Genéticos (Friedberg et al. 1958)

- ⌘ Modelos simplistas

 - Ej. El perceptrón lineal es incapaz de resolver el problema del XOR
[“Perceptrons” Minsky & Papert, 1969]

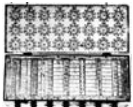
 - Solución: Usar una capa oculta no lineal + Retropropagación (Bryson & Ho, 1969; Werbos 1974; Rumelhart, Hinton & Williams, 1986)

- ⌘ Sistemas basados en conocimiento: Uso de conocimiento específico del dominio.

 - ⌘ DENDRAL (Buchanan, Feigenbaum & Lederberg, 1967): inferencia de la estructura molecular partiendo del espectrograma de masas

 - ⌘ SHRDLU (Winograd, 1971): comprensión de lenguaje natural

 - ⌘ MYCIN (Buchanan, Feigenbaum & Shortliffe, 1974): diagnóstico de infecciones de sangre. Incorpora “factores de certeza”



Breve historia de la IA: Historia reciente

⌘ IA vista como industria

- ⌘ Sistemas expertos comerciales (1980's)

- ⌘ Minería de datos (data mining), desde mediados de los 90

⌘ Redes neuronales (su retorno):

Métodos conexionistas como complemento de métodos simbólicos.

- ⌘ Métodos de mecánica estadística (Hopfield, 1982)

- ⌘ Modelos de memoria basados en redes neuronales (Rumelhart & Hinton, mediados de los 80)

- ⌘ Redes topológicas (Kohonen)

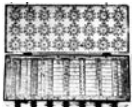
⌘ Aprendizaje automático (Valiant, Mitchell, Vapnik, Breiman, Quinlan, ...)

⌘ Tendencias actuales

- ⌘ Redes Bayesianas (Pearl 1988)

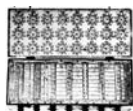
- ⌘ Agentes (por ejemplo SOAR, Newell, Laird, Rosenbloom 1987-)

- ⌘ Rastreadores de web (web crawlers), inteligencia colectiva / distribuida, mascotas robóticas, robots exploradores, máquinas sociables, agentes emocionales, domótica ...



Definiciones de IA

	COMO UN HUMANO	RACIONALMENTE
PENSAR	2. PENSAR COMO UN HUMANO Construir máquinas con mente	3. PENSAR RACIONALMENTE Construir modelos computacionales de los procesos racionales
ACTUAR	1. ACTUAR COMO UN HUMANO Crear androides	4. ACTUAR RACIONALMENTE Diseñar dispositivos que muestran un comportamiento inteligente

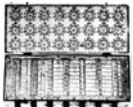


1. Actuar como un humano: El Test de Turing

Interrogador



asado en IA

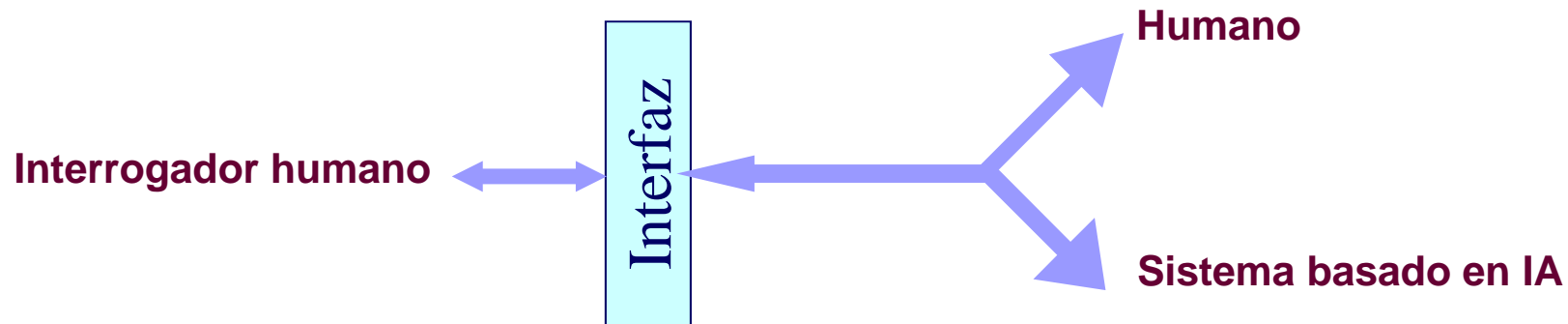


1. Actuar como un humano: El Test de Turing

Turing, "Computing machinery and intelligence", Mind, 59, 433:460, 1950

Definición operacional de inteligencia: Un sistema es inteligente si puede persuadir a otro sistema inteligente (por ejemplo un humano) de su inteligencia.

Test de Turing parcial

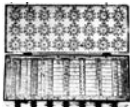


Juego de imitación: ¿Puede el sistema basado en IA hacer pensar al interrogador que es el humano?

Necesita: procesamiento de lenguaje natural, representación del conocimiento, razonamiento automático, aprendizaje automático.

Test de Turing total

Sin interfaz. **Además necesita:** Visión artificial (percepción de objetos) + robótica (manipulación de objetos y movilidad)



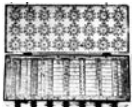
1. Actuar como un humano: El Test de Turing

Ventajas:

- Evita entrar en debates posiblemente estériles sobre "¿qué es la inteligencia?" o "¿pueden las máquinas pensar?"
- El test se sigue considerando una buena forma de medir el nivel de éxito que tienen los sistemas basados en IA en exhibir comportamiento inteligente

Dificultades:

- Ambiguo, no reproducible
- No constructivo
- No se puede describir de manera matemática
- Los esfuerzos actuales de la IA no se dirigen a pasar el test de Turing



2. Pensar como un humano: Modelos cognitivos

Ciencia cognitiva (psicología experimental + ciencias de la computación):

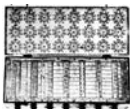
Modelos de los procesos de pensamiento humanos

Ventajas:

- Proporciona comprensión de inteligencia + cognición humana
- Los modelos se diseñan para imitar el funcionamiento de la mente humana y son por ello inteligibles

Dificultades:

- El hecho de que un dispositivo basado en IA actúe como un humano en una tarea que requiere inteligencia no significa que el dispositivo sea un modelo apropiado del proceso de pensamiento humano correspondiente
- El mejor diseño artificial para un sistema inteligente no es necesariamente una imitación de la mente humana



3. Pensar racionalmente: IA logicista

Lógica formal (matemáticas + ciencias de la computación):

Proceso de razonamiento automático donde se deducen conclusiones correctas a partir de premisas correctas

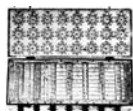
[representación + premisas + reglas de derivación]

Ventajas:

- Notación precisa y no ambigua de las afirmaciones sobre objetos del mundo y las relaciones entre ellos
- Un algoritmo de resolución + búsqueda completa puede, en principio, resolver cualquier problema que pueda ser formulado en notación lógica.
Si el problema no tiene solución, el algoritmo podría no parar nunca

Dificultades:

- Formalización de conocimiento informal
- Los requerimientos computacionales (espacio/tiempo) pueden hacer que los cálculos sean imposibles de hacerse en la práctica



4. Actuando racionalmente: Agentes

Agente racional

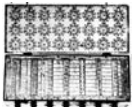
Sistema autónomo, capaz de percibir e interactuar con su entorno, de explorar (recopilación de información), aprender y adaptarse, de formular objetivos y diseñar planes para alcanzar esos objetivos.

El agente es **racional**, en el sentido de que actúa para alcanzar el mejor resultado, o mejor resultado esperado, de acuerdo a una **medida de rendimiento**, condicionada a su conocimiento del mundo y dados los recursos computacionales existentes (limitados).

Ventajas:

- Integra y extiende definiciones previas
- La racionalidad limitada es más realista que la racionalidad absoluta

Dificultades: Las exploraremos a lo largo de este curso



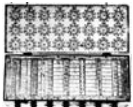
Entorno de trabajo

Entorno de trabajo

- ⌘ **Entorno**
- ⌘ **Percepciones (sensores) / Acciones (actuadores)**
- ⌘ **Medida de rendimiento:** Dada una secuencia de percepciones, un agente racional selecciona la acción que maximiza el rendimiento esperado dada la evidencia (percepciones + conocimiento interno) + recursos

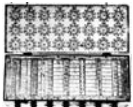
Clasificación de los entornos de trabajo

- ⌘ Totalmente observables / Parcialmente observables
- ⌘ Deterministas / Estratégicos (sistemas multi-agente) / Estocásticos
- ⌘ Episódicos / Secuenciales
- ⌘ Estáticos / Semiestáticos (agente) / Dinámicos (agente + entorno)
- ⌘ Discretos / Continuos
- ⌘ Un agente / Múltiples agentes (cooperación / competición)



Diseño de agentes

- ⌘ **Agente reactivo simple** [eficiente sólo en entorno completamente observable]
 1. El agente está en un estado dado
 2. El estado se actualiza incorporando el conocimiento obtenido de una percepción
 3. El agente selecciona una acción de una tabla que contiene reglas condición-acción
- ⌘ **Agente reactivo basado en modelos:** Incorporación de un **modelo** para poder tratar entornos parcialmente observables
El agente tiene un modelo de la evolución del mundo + resultado de las acciones (agente basado en modelo)
- ⌘ **Agente basado en objetivos:** Incorporación de **objetivos** para seleccionar la mejor(es) acción(es) posible(s) para alcanzar un estado deseado
- ⌘ **Agente basado en utilidad:** Incorporación de **utilidad** para seleccionar entre diferentes acciones que alcanzan el objetivo deseado (o uno cercano)
- ⌘ **Agente que aprende:** Incorporación de **aprendizaje**
 - ⌘ Crítico
 - ⌘ Recompensa / penalización



Sistemas multiagente

IA distribuida:

El comportamiento inteligente que emerge de la cooperación de agentes autónomos.

- ⌘ Comunicación
- ⌘ Coordinación
- ⌘ Negociación, etc.

Aplicaciones:

- ⌘ Comercio electrónico (E-commerce)
- ⌘ Optimización de procesos de producción industrial
- ⌘ Monitorización en tiempo real de redes de telecomunicación
- ⌘ Simulaciones
 - ⌘ Ecología
 - ⌘ Interacciones sociales
 - ⌘ Aspectos sociales de la inteligencia

