

## Los fundamentos de la inteligencia artificial

Presentaremos una breve historia de las disciplinas que han contribuido con ideas, puntos de vista y técnicas al desarrollo de la IA.

### Filosofía (desde el año 428 a.c. hasta el presente)

- ¿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 (384-322 a.c.) fue el primero en formular un conjunto preciso de leyes que gobernaban la parte racional de la inteligencia. Él desarrolló un sistema informal para razonar adecuadamente con silogismos, que en principio permitía extraer conclusiones mecánicamente, a partir de premisas iniciales, por ejemplo: a) Algunos escorpiones son venenosos b) Los venenos nos pueden matar c) Algunos escorpiones nos pueden matar. Mucho después, Ramón Llull (d. 1315) tuvo la idea de que el razonamiento útil se podría obtener por medios artificiales. Thomas Hobbes (1588-1679) propuso que el razonamiento era como la computación numérica, de forma que “nosotros sumarnos y restamos silenciosamente en nuestros pensamientos”. La automatización de la computación en sí misma estaba en marcha; alrededor de 1500, Leonardo da Vinci (1452-1519) diseñó, aunque no construyó, una calculadora mecánica; construcciones recientes han mostrado que su diseño era funcional. La primera máquina calculadora conocida se construyó alrededor de 1623 por el científico alemán Wilhelm Schickard (1592-1635), aunque la Pascalina, construida en 1642 por Blaise Pascal (1623-1662), sea más famosa. Pascal escribió que “la máquina aritmética produce efectos que parecen más similares a los pensamientos que a las acciones animales”. Gottfried Wilhelm Leibniz (1646-1716) construyó un dispositivo mecánico con el objetivo de llevar a cabo operaciones sobre conceptos en lugar de sobre números, pero su campo de acción era muy limitado.

Ahora que sabemos que un conjunto de reglas pueden describir la parte racional y formal de la mente, el siguiente paso es considerar la mente como un sistema físico. René Descartes (1596-1650) proporciona la primera discusión clara sobre la distinción entre la mente y la materia y los problemas que surgen. Uno de los problemas de una concepción puramente física de la mente es que parece dejar poco margen de maniobra al libre albedrío: si el pensamiento está totalmente gobernado por leyes físicas, entonces una piedra podría “decidir” caer en dirección al centro de la Tierra gracias a su libre albedrío. A pesar de ser un ferviente defensor de la capacidad de razonamiento, Descartes fue un defensor del dualismo. Sostenía que existe una parte de la mente (o del alma o del espíritu) que está al margen de la naturaleza, exenta de la influencia de las leyes físicas. Los animales, por el contrario, no poseen esta cualidad dual; a ellos se le podría concebir como si se tratasen de máquinas. Una alternativa al dualismo es el materialismo, que considera que las operaciones del cerebro realizadas de acuerdo a las leyes de la física constituyen la mente. El libre albedrío es simplemente la forma en la que la percepción de las opciones disponibles aparecen en el proceso de selección.

Dada una mente física que gestiona conocimiento, el siguiente problema es establecer las fuentes de este conocimiento. El movimiento empírico, iniciado con el *Novum Organum*, de Francis Bacon (1561-1626), se caracteriza por el aforismo de John Locke (1632-1704): “Nada existe en la mente que no haya pasado antes por los sentidos”. David Hume (1711-1776) propuso en “*A Treatise of Human Nature*” (Hume, 1739) lo que actualmente se conoce como principio de inducción: las reglas generales se obtienen mediante la exposición a asociaciones repetidas entre sus elementos. Sobre la base de las propuestas de Ludwig Wittgenstein (1889-1951) y Bertrand Russell (1872-1970), el famoso Círculo de Viena, liderado por Rudolf Carnap, (1891-1970), desarrolló la doctrina del positivismo lógico. Esa doctrina sostiene que todo el conocimiento se puede caracterizar mediante teorías lógicas relacionadas, en última instancia, con sentencias de observación que corresponden a estímulos sensoriales. La teoría

de la confirmación de Carnap y Carl Hempel (1905-1997) intenta explicar cómo el conocimiento se obtiene a partir de la experiencia. El libro de Carnap. *The Logical Structure of the World* (1928), define un procedimiento computacional explícito para la extracción de conocimiento a partir de experiencias primarias. Fue posiblemente la primera teoría en mostrar la mente como un proceso computacional.

El último elemento en esta discusión filosófica sobre la mente es la relación que existe entre conocimiento y acción. Este asunto es vital para la IA, ya que la inteligencia requiere tanto acción como razonamiento. Más aún, simplemente con comprender cómo se justifican determinadas acciones se puede llegar a saber cómo construir un agente cuyas acciones sean justificables (o racionales). Aristóteles argumenta que las acciones se pueden justificar por la conexión lógica entre los objetivos y el conocimiento de los efectos de las acciones:

¿Cómo es que el pensamiento viene acompañado en algunos casos de acciones y en otros no?, ¿en algunos casos por movimiento y en otros no? Parece como si la misma cosa sucediera tanto si razonáramos o hiciéramos inferencias sobre objetos que no cambian; pero en este caso el fin es una proposición especulativa... mientras la conclusión resultante de las dos premisas es una acción... Yo necesito abrigarme, una manta abriga. Yo necesito una manta. Qué necesito, que debo hacer; necesito una manta. Necesito hacer una manta. Y la conclusión, “Yo tengo que hacer una manta”, es una acción. (Nussbaum, 1078, p. 40)

En *Nicomachean Ethics* (Libro III. 3 , 1112b). Aristóteles continúa trabajando con este tema, sugiriendo un algoritmo:

Nosotros no reflexionamos sobre los fines, sino sobre los medios. Un médico no reflexiona sobre si debe curar, ni un orador sobre si debe persuadir... Ellos asumen el fin y consideran cómo y con que medios se obtienen, y si resulta fácil y es por tanto procluctivo; mientras que sí sólo se puede alcanzar por un medio se tiene en consideración cómo se alcanzará por este y por que medios se obtendrá, hasta que se llegue a la causa primera..., y lo último en el orden del análisis parece ser lo primero en el orden de los acontecimiento. Y si se llega a un estado imposible, se abandona la búsqueda, como por ejemplo si se necesita dinero y no se puede conseguir; pero si hay una posibilidad se intentara.

El algoritmo de Aristóteles se implemento 2.300 años mas tarde por Newell y Simon con la ayuda de su programa SRGP. El cual se conoce como sistema de planificación regresivo.

El análisis basado en objetivos es útil, pero no indica qué hacer cuando varias acciones nos llevan a la consecución del objetivo, o cuando ninguna acción facilita su completa consecución. Antoine Arnauld (1612-1093) describió correctamente una forma cuantitativa para decidir que acción llevar a cabo en un caso como este. El libro *Utilitarianism* (Mill, 1863) de John Stuart Mill (1806-1873) propone la idea de un criterio de decisión racional en todos los ámbitos de la actividad humana.

### **Matemáticas (aproximadamente desde el año 800 al presente)**

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

Los filósofos delimitaron las ideas más importantes de la IA. pero para pasar de ahí a una ciencia formal es necesario contar con una formulación matemática en tres áreas fundamentales: lógica, computación y probabilidad.

El concepto de lógica formal se remonta a los filósofos de la antigua Grecia, pero su desarrollo matemático comenzó realmente con el trabajo de George Boole que definió la lógica proposicional o Booleana en 1847. En 1879 Gottlob Frege extendió la lógica de Boole para incluir objetos y relaciones, y creó la lógica de primer orden que se utiliza hoy como el sistema más básico de conocimiento. Alfred Tarski introdujo una teoría de referencia que enseña como relacionar objetos de una lógica con objetos del mundo real. El paso siguiente consistió en definir los límites de lo que se podía hacer con la lógica y la informática.

El considerar a los algoritmos como objetos en sí mismos se remonta a la época de *Al-Khwarizmi*, un matemático persa. Boole, entre otros, presentó algoritmos para llevar a cabo deducciones lógicas y hacia el final del siglo XIX se llevaron a cabo numerosos esfuerzos para formalizar el razonamiento matemático general con lógica deductiva. En 1900 David Hilbert presentó una lista de 23 problemas que acertadamente predijo ocuparían a los matemáticos todo ese siglo. En el último de ellos se preguntaba si existe un algoritmo que permita determinar la validez de cualquier proposición lógica en la que aparezcan números naturales (el famoso problema de la decisión). Básicamente, lo que Hilbert se preguntaba es si hay límites fundamentales en la capacidad de los procedimientos efectivos de demostración. En 1930, Kurt Gödel demostró que existe un procedimiento eficiente para demostrar cualquier aseveración verdadera en la lógica de primer orden de Frege y Russell, sin embargo con la lógica de primer orden no era posible capturar el principio de la inducción matemática necesario para la caracterización de los números naturales. En 1931, demostró que, en efecto, existen límites reales. Mediante su teorema de la incompletitud demostró que cualquier lenguaje que tuviera la capacidad suficiente para expresar las propiedades de los números naturales, existen aseveraciones verdaderas no decidible en el sentido de que no es posible decidir su validez mediante ningún algoritmo.

El resultado fundamental anterior se puede interpretar también como la indicación de que existen algunas funciones de los números enteros que no se pueden representar mediante un algoritmo, es decir, no se pueden calcular. Lo anterior llevó a Alan Turing a tratar de caracterizar exactamente aquellas funciones que sí eran susceptibles de ser caracterizadas. La tesis de Church-Turing, que afirma que la máquina de Turing es capaz de calcular cualquier función computable, goza de aceptación generalizada ya que proporciona una definición suficiente. Turing también demostró que existen algunas funciones que no se pueden calcular mediante la máquina de Turing. Por ejemplo, ninguna máquina puede decidir en general si un programa dado producirá una respuesta a partir de unas entradas, o si seguirá calculando indefinidamente.

Si bien ser no decidible ni computable son importantes para comprender el proceso del cálculo, la noción de intratabilidad tuvo repercusiones más importantes. En términos generales se dice que un problema es intratable si el tiempo necesario para la resolución de casos particulares de dicho problema crece exponencialmente con el tamaño de dichos casos. Es importante porque un crecimiento exponencial implica la imposibilidad de resolver casos moderadamente grandes en un tiempo razonable.

¿Cómo se puede reconocer un problema intratable? La teoría de la NP-completitud, propuesta por primera vez por Steven Cook y Richard Karp, propone un método. Demostraron la existencia de grandes clases de problemas de razonamiento y búsqueda combinatoria canónica que son NP-completos. Toda clase de problema que se pueda reducir a la clase NP-completo será seguramente intratable (aunque no se ha demostrado que todos los problemas NP-completos son necesariamente intratables, la mayor parte de los teóricos así lo creen). Estos resultados contrastan con el optimismo con el que la prensa popular recibió a los primeros computadores, “Supercerebros Electrónicos” que eran “Más rápidos que Einstein”. A pesar del rápido incremento en la velocidad de los computadores, los sistemas inteligentes se caracterizarán por el uso cuidadoso de los recursos.

Además de la lógica y el cálculo, la tercera gran contribución de las matemáticas a la IA es la teoría de la probabilidad. El italiano Gerolamo Cardano fue el primero en proponer la idea de probabilidad, presentándola en términos de los resultados de juegos de apuesta. La probabilidad se convirtió pronto en parte imprescindible de las ciencias cuantitativas, ayudando en el tratamiento de mediciones con incertidumbre y de teorías incompletas. Thomas Bayes propuso una regla para la actualización de probabilidades subjetivas a la luz de nuevas evidencias. La regla de Bayes y el área resultante llamado análisis Bayesiano conforman la base de las propuestas modernas que abordan el razonamiento incierto en sistemas de IA.

## **Economía (desde el año 1776 hasta el presente)**

- ¿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 colaboren?
- ¿Cómo se deben llevar a cabo acciones cuando los resultados se obtienen en un futuro lejano?

La ciencia de la economía comenzó en 1776 cuando el filósofo escocés Adam Smith publicó “An Inquiry into the Nature and Causes of the Wealth of Nations”. Smith fue el primero en tratar el pensamiento económico economía como una ciencia, utilizando la idea de las economías pueden concebirse como un conjunto de agentes individuales que intentan maximizar su propio estado de bienestar económico. La mayoría de la gente cree que la economía sólo se trata de dinero, pero los economistas dicen que ellos realmente estudian cómo la gente toma decisiones que les llevan a obtener los beneficios esperados. Léon Walras formalizó el tratamiento matemático del beneficio deseado o utilidad, posteriormente mejorado por Frank Ramsey, luego por Von Neumann y Oskar Morgenstern en su libro “The theory of Games and Economic Behavior”.

La teoría de la decisión, que combina la teoría de la probabilidad con la teoría de la utilidad, proporciona un marco completo y formal para la toma de decisiones (económicas o de otra índole) realizadas bajo incertidumbre, esto es, en casos en los que las descripciones probabilísticas capturan adecuadamente la forma en la que se toman las decisiones en el entorno; lo cual es adecuado para grandes economías en las que cada agente no necesita prestar atención a las acciones que lleven a cabo el resto de los agentes individualmente. Cuando se trata de pequeñas economías, la situación se asemeja más a la de un juego: las acciones de un jugador pueden afectar significativamente a la utilidad de otro (tanto positiva como negativamente).

El trabajo de Richard Bellman formaliza una clase de problemas de decisión secuencial llamados procesos de decisión de Markov, que ayudan a responder la tercer pregunta mencionada anteriormente.

Herbert Simon, uno de los primeros en investigar en el campo de la IA, ganó el premio Nobel en Economía en 1978 por su trabajo, en el que mostró que los modelos basados en satisfacción (que toman decisiones que son “suficientemente buenas” proporcionaban una mejor descripción del comportamiento humano real.

## **Neurociencia (desde el año 1861 hasta el presente)**

- ¿Cómo procesa información el cerebro?

La Neurociencia es el estudio de del sistema neurológico, y en especial del cerebro. La forma exacta en la que en un cerebro se genera el pensamiento es uno de los grandes misterios de la ciencia.

El estudio de Paul Broca (1861) sobre la afasia (dificultad para hablar) en pacientes con el cerebro dañado, le dio fuerza al campo de la neurociencia y convenció a la sociedad médica de la existencia de áreas localizadas en el cerebro responsables de funciones cognitivas específicas. En esta época ya se sabía que el cerebro estaba formado por células nerviosas o neuronas, pero fue hasta 1873 cuando Camilo Golgi desarrolló una técnica de coloración que permitió la observación de neuronas individuales en el cerebro.

En la actualidad se dispone de información sobre la relación existente entre las áreas del cerebro y las partes del cuerpo humano que controlan o de las que reciben impulsos sensoriales. Tales relaciones pueden cambiar de forma radical incluso en pocas semanas. Más aún, no se tiene totalmente claro cómo algunas áreas se pueden encargar de ciertas funciones que eran responsabilidad de áreas dañadas. No hay prácticamente ninguna teoría que explique cómo se almacenan recuerdos individuales.

La conclusión verdaderamente increíble es que una colección de simples células puede llegar a generar razonamiento, acción y conciencia, o dicho en otras palabras, los cerebros generan las inteligencias.

Los cerebros y las computadoras digitales tienen propiedades algo diferentes, la Figura L3 muestra que las computadoras tienen un tiempo de ciclo que es un millón de veces más rápido que un cerebro. El cerebro lo compensa con mucho más almacenamiento e interconexión que incluso una computadora

personal de alta gama, aunque las supercomputadoras más grandes tienen una capacidad similar a la del cerebro (sin embargo, debe tenerse en cuenta que el cerebro no parece usar todas sus neuronas simultáneamente).

Los futuristas hacen gran parte de estos números, señalando una singularidad inminente en la que las computadoras alcanzan un nivel de rendimiento sobrehumano (Vinge, 1993; Kurzweil, 2005), pero las comparaciones crudas no son especialmente informativas. Incluso con una computadora de capacidad prácticamente ilimitada, aún no sabríamos cómo lograr el nivel de inteligencia del cerebro.

	Supercomputer	Personal Computer	Human Brain
Computational units	10 <sup>10</sup> CPUs, 10 <sup>12</sup> transistors	4 CPUs, 10 transistors	10 <sup>11</sup> neurons
Storage units	10 <sup>14</sup> bits RAM 10 <sup>15</sup> bits disk	10 <sup>11</sup> bits RAM 10 <sup>13</sup> bits disk	10 <sup>11</sup> neurons 10 <sup>15</sup> synapses
Cycle time	10 <sup>-6</sup> sec	10 <sup>-9</sup> sec	10 <sup>-3</sup> sec
Operations/sec	10 <sup>15</sup>	10 <sup>10</sup>	10 <sup>17</sup>
Memory updates/sec	10 <sup>10</sup>	10 <sup>10</sup>	10 <sup>14</sup>

Figure 1.3 A crude comparison of the raw computational resources available to the IBM BLUE GENE supercomputer, is typical personal computer of 2008, and the human brain. The brain's numbers are essentially fixed, whereas the supercomputer's numbers have been increasing by a factor of 10 every 5 years or so, allowing it to achieve rough parity with the brain. The personal computer lags behind on all metrics except cycle time.

### Psicología (desde el año 1879 hasta el presente)

- ¿Cómo piensan y actúan los humanos y animales?

EN 1879, Wilhem Wundt abrió el primer laboratorio de psicología experimental en ña universidad de Leipzig. Puso mucho énfasis en la realización de experimentos controlados cuidadosamente en la que sus operaros realizaban tareas de percepción o asociación al tiempo que sometían a introspección sus procesos mentales. Biólogos , estudiando el comportamiento humano, carecían de datos introspectivos y desarrollaron una metodología objetiva, tal como describe H.S. Jennings (1906) en su trabajo “Behavior of the Lower Organisms”. El movimiento conductista, liderado por John Watson aplicó este punto de vista a los humanos, rechazando cualquier teoría en la que intervinieran procesos mentales, argumentando que la introspección no aportaba evidencia fiable. El conductismo hizo muchos descubrimientos utilizando ratas y palomas, pero tuvo menos éxito en la comprensión de los seres humanos.

La conceptualización del cerebro como un dispositivo de procesamiento e información, característica principal de la psicología cognitiva, se remonta a por lo menos las obras de William James (1842-1910). Este punto de vista cognitivo fue eclipsado por el conductismo en Estados Unidos, pero en la unidad psicológica de Cambridge, los modelos cognitivos emergieron con fuerza. La obra “The Nature of Explanation”, de Kenneth Craik (1943), restablece la legitimidad de términos “mentales” como creencias y objetivos, argumentando que son tan científicos como lo puede ser la presión y temperatura cuando se habla de gases, a pesar de que estos estén formados por moléculas que no tienen ni presión ni temperatura. Craik establece tres elementos clave que hay que tener en cuenta para diseñar un agente basado en conocimiento:

1. El estímulo deberá ser traducido a una representación interna.
2. Esta representación se debe manipular mediante procesos cognitivos para generar nuevas representaciones internas.
3. Estas a su vez, se traducirán de nuevo en acciones.

El desarrollo del modelo computacional llevo a la creación del campo de la ciencia cognitiva. En un simposio celebrado en el MIT en 1956, George Miller presentó “The magic Number Seven”, Noam Chomsky presentó “Three Models of Language”, y Allen Newell y Herbert Simon presentaron “The logic Theory Machine”. Estos tres artículos mostraron como se podían utilizar los modelos informáticos para modelar la psicología de la memoria, el lenguaje y el pensamiento lógico, respectivamente.

### **Ingeniería computacional (desde el año 1940 hasta el presente)**

- ¿Cómo se puede construir un computador eficiente?

Para que la inteligencia artificial pueda llegar a ser una realidad se necesitan dos cosas: inteligencia y un artefacto (un medio). El computador ha sido elegido.

Desde mediados del siglo pasado, cada generación de dispositivos ha conllevado un aumento en la velocidad de proceso y en la capacidad de almacenamiento, así como de una reducción de precios. La potencia de ellos computadores se dobla cada 18 meses aproximadamente y seguirá este ritmo durante una o dos décadas más. Después de necesitará ingeniería molecular u otras tecnologías novedosas.

La IA también tiene una deuda con la parte de software que ha proporcionado los sistemas operativos, los lenguajes de programación y las herramientas necesarias para escribir programas modernos. Sin embargo, en esta área la deuda se ha saldado: la investigación en IA ha generado numerosas ideas novedosas de las que se ha beneficiado la informática en general, como por ejemplo el tiempo compartido, los intérpretes imperativos, los computadores personales con interfaces gráficas, entornos de desarrollo rápido, listas enlazadas, administración automática de memoria y conceptos claves de la programación simbólica, funcional, dinámica y orientada a objetos.

### **Teoría de control y cibernética (desde el año 1948 hasta el presente)**

- ¿Cómo pueden los artefactos operar bajo su propio control?

Ktesibios de Alejandría (250 a.c.) construyó la primera máquina auto controlada: un reloj de agua con un regulador que mantenía el flujo de agua circulando por él, con un ritmo constante y predecible. Otros ejemplos de sistemas de control auto regulables y retroalimentados son el motor de vapor, creado por James Watt, y el termostato, inventado por Cornelis Drebbel. La teoría matemática de los sistemas con retroalimentación estables se desarrolló en el siglo XIX.

La figura central de lo que ahora se llama teoría de control fue Norbert Wiener. Fue un matemático que trabajó en sistemas de control biológicos y mecánicos y en sus vínculos con la cognición. De la misma manera que Craik Wiener y sus colegas desafiaron la ortodoxia conductista. Ellos veían el comportamiento determinista como algo emergente de un mecanismo regulador que intenta minimizar el error.

La teoría de control moderna, especialmente la rama conocida como control óptimo estocástico, tiene como objetivo el diseño de sistemas que maximizan una función objetivo en el tiempo. Lo cual se asemeja ligeramente a la visión de la IA: el diseño de sistemas que se comportan de manera óptima. ¿Por qué, entonces, IA y teoría de control son dos campos diferentes, especialmente teniendo en cuenta la cercana relación entre sus creadores? La respuesta está en el gran acoplamiento existente entre las técnicas matemáticas con las que estaban familiarizados los investigadores y entre los conjuntos de problemas que se abordaban desde cada uno de los puntos de vista. El cálculo y el álgebra matricial, herramientas de la teoría de control, se utilizaron en la definición de sistemas que podían describir mediante conjuntos fijos de variables continuas; más aún, el análisis exacto es solo posible en sistemas lineales. La IA se fundó en parte para escapar de las limitaciones matemáticas de la teoría de control en los años 50. Las herramientas de inferencia lógica y computacional permitieron a los investigadores de IA afrontar problemas relacionados con el lenguaje, visión y planificación, que estaban completamente fuera del punto de vista de la teoría de control.

## **Lingüística (desde el año 1957 hasta el presente)**

- ¿Cómo está relacionado el lenguaje con el pensamiento?

En 1957, B. F. Skinner publicó “Verbal Behavior”. La obra presenta una visión extensa y detallada desde el enfoque conductista al aprendizaje del lenguaje, y estaba escrita por los expertos más destacados de este campo. Noam Chomsky mostró, en su libro “Syntactic Structures”, como la teoría conductista no abordaba el tema de la creatividad en el lenguaje: no explicaba cómo es posible que un niño sea capaz de entender y construir oraciones que nunca antes ha escuchado. La teoría de Chomsky sí podía explicar lo anterior y, a diferencia de teorías anteriores, poseía el formalismo suficiente como para permitir su programación.

La lingüística moderna y la IA nacieron al mismo tiempo y maduraron juntas, solapándose en un campo híbrido llamado lingüística computacional o procesamiento del lenguaje natural. El problema del entendimiento del lenguaje se mostró pronto mucho más complejo de lo que se había pensado en 1957. El entendimiento del lenguaje requiere la comprensión de la materia bajo el estudio y de su contexto, y no solamente del entendimiento de la estructura de las sentencias.