
CAPITULO 1

INTRODUCCIÓN

- **Cuestiones Preliminares**
 - **Consideraciones Generales sobre la IA**
 - **Organización Temática del Texto**
 - **Resumen**
 - **Textos Básicos**
-

1. INTRODUCCIÓN

La inteligencia artificial (IA) es una disciplina relativamente reciente, fruto de trabajos en distintas áreas del pensamiento que fueron definitivamente aglutinados tras el advenimiento de las ciencias de la computación como materia consolidada de investigación y desarrollo.

Hoy día, en nuestro entorno altamente informatizado, vuelve a cobrar vigencia el sueño que desde tiempos remotos ha ido acompañando al hombre desde que éste tiene conciencia de sí mismo: la posibilidad de acercarse a la máxima perfección a través de la construcción de seres inteligentes que aprendan, piensen y creen por sí mismos.

Junto a esta visión casi mágica del dominio de la inteligencia artificial, la visión mucho más modesta del moderno investigador, le lleva a utilizar la inteligencia artificial como medio para tratar de resolver problemas cotidianos con repercusiones socioeconómicas más o menos importantes, como el movimiento del brazo de un robot, la automatización de diagnósticos médicos o, simplemente, el desarrollo de programas que jueguen bien al ajedrez.

En cualquier caso, la inteligencia artificial está ligada al concepto de inteligencia por cuanto ésta representa una faceta más propia de los seres humanos que de los seres vivos en general aunque, como veremos, esta última afirmación es discutible y su veracidad dependerá de cómo definamos el concepto de inteligencia.

De todos modos hay que procurar ser cautos y tratar de contemplar a la inteligencia artificial en su verdadera dimensión. En este sentido, podría considerarse que la inteligencia artificial es una ciencia que trata de establecer las bases para el posterior desarrollo de un conjunto de técnicas destinadas a dotar a las máquinas de una cierta autonomía. Esta autonomía puede referirse a aspectos muy diversos como la comunicación con las máquinas en lenguaje natural, la toma de decisiones en un dominio concreto, la toma de decisiones en el tiempo, o el aprendizaje. Si somos capaces de conseguir que una máquina asuma todas o alguna de estas actividades, ciertamente habremos dotado a la máquina de autonomía.

La cuestión que acabamos de plantear tiene que ver bastante con la educación de un niño, desde sus primeros días de vida, hasta que se convierte en un adulto. Cuanto más pequeño es el niño más claros y limitados tienen que ser los mensajes que le dirigamos. A medida que el niño crece, y adquiere conocimiento, experiencia y sentido común, los intercambios de información pueden ir ganando en globalidad, generalidad y eficacia, y el individuo puede asumir tareas cada vez más complejas, al mismo tiempo que sus capacidades sensoriales se especializan. Finalmente, sus órganos sensoriales y perceptuales, y sus conocimientos, son de tal magnitud que el individuo se convierte en un ser adulto, y por lo tanto autónomo.

Si comparamos el proceso de educación de un niño pequeño con la inteligencia artificial, fácilmente llegaremos a la conclusión de que la inteligencia artificial, ni es inteligencia, ni es artificial: se parece más bien a un proceso natural de educación de una máquina para lograr que su comportamiento sea lo más autónomo posible.

La inteligencia artificial, como ciencia que nos va a permitir comprender algunos de los principios básicos de lo que ya hemos llamado educación computacional, no hace mucho que ha dejado de ser una disciplina emergente. El grado de consolidación de la inteligencia artificial no es absoluto, y se podría decir que todavía estamos destetando al niño. Los logros actuales de la inteligencia artificial, aunque cualitativamente hablando son importantes, cuantitativamente distan mucho de cubrir los objetivos inicialmente planteados: estamos todavía lejos de poder construir máquinas que piensen, aprendan y creen por sí mismas. No obstante, muchos han sido los esfuerzos empleados en dotar a la inteligencia artificial del peso específico que ya tiene, esfuerzos de los que encontramos las primeras referencias en la Mitología y en la Historia.

1.1. Cuestiones Preliminares

Es en efecto en los textos mitológicos en dónde aparecen las primeras referencias a los androides. Uno de ellos, Talos, obra de Dédalo, es descrito como un gigantesco robot de bronce encargado de la custodia y defensa de la isla de Minos. Todos los días Talos recorría el perímetro de la mencionada isla tratando de descubrir visitantes indeseables. Si algún infeliz era descubierto, e identificado como enemigo, Talos saltaba dentro de una hoguera en la que permanecía hasta ponerse incandescente, y luego abrazaba al recién llegado hasta provocarle la muerte.

También en la Mitología encontramos referencias sobre los androides femeninos contruidos por Hefastos, que son descritos como seres de oro macizo, dotados de inteligencia, y cuya misión era ayudar a caminar a su creador.

Más modernamente, y probablemente fruto de la imaginación popular, encontramos la leyenda del Golem, que puede considerarse como uno de los primeros paradigmas de la IA. El Golem era una estatuilla de arcilla roja, a la que un rabino de la comunidad judía de Praga dio vida a través de un ritual de magia negra. El Golem no hacía nada por iniciativa propia, y su único cometido era el de actuar como un fiel esclavo de su rabino. Tras una vida pesada y aburrida, se fueron despertando en el Golem instintos que hasta entonces habían permanecido ocultos y, poco a poco, fue liberándose de la tutela de su amo, convirtiéndose en un personaje terrorífico y malvado. Finalmente el Golem fue destruido por su creador, aunque la leyenda asegura que reaparece en Praga cada treinta y tres años.

Pero quizás los antecedentes históricos más remotos de la IA haya que buscarlos en la antigua Grecia. En concreto, y pertenecientes a la época alejandrina, Arquímedes, Demetrio de Farleria, Architas de Tarente, y Herón de Alejandría fueron los verdaderos precursores de una disciplina que hoy se conoce con el nombre de *Automática*.

Arquímedes construyó los mecanismos defensivos que fueron empleados para tratar de contener los ataques de la flota romana sobre Siracusa. Por su parte, Architas de Tarente, más prosaico, construyó una paloma que batía las alas. Demetrio de Farleria dedicó sus esfuerzos a la construcción de un caracol mecánico capaz de avanzar arrastrándose. Finalmente, Herón de Alejandría diseñó y construyó unos actores

artificiales capaces de representar escenas de la guerra de Troya. Es en esta época en dónde podemos buscar los principios generales de los autómatas, según los cuales:

- Los mecanismos de un autómatas actúan en virtud de su propia estructura interna
- La acción procede de una adecuada organización de fuerzas motrices, naturales y artificiales
- La movilidad de los autómatas afecta a todo el conjunto, y no sólo a ciertas partes

Más recientemente, ya en la Edad Media, hay que reseñar los trabajos de Llull, que constituyen la primera aproximación a la IA desde el pensamiento. Llull establece en su obra *Ars Magna* las bases de una técnica que aún hoy, con algunas modificaciones, se sigue utilizando en inteligencia artificial.

El método llulliano consiste en realizar un ensayo exhaustivo y sistemático de todos los procedimientos y principios básicos que pudieran ser útiles para resolver problemas concretos. El método propuesto requiere:

- Adquirir los principios fundamentales conocidos, y admitidos por todos, en un dominio de aplicación determinado
- Agotar todas las posibles combinaciones de dichos principios

La filosofía del método llulliano es clara. Al realizar un ensayo exhaustivo de todos los principios fundamentales de un dominio, tendríamos que ser capaces de encontrar la solución a nuestro problema, siempre y cuando esté bien definido, y considerando mentes finitas.

No obstante este método presenta varios inconvenientes. Los más importantes son:

- Determinar cuáles son realmente los principios básicos de un dominio
- La inevitable explosión combinatoria

Tras los trabajos de Llull, y ya en plena Edad Moderna, es obligado mencionar a los Droz, quienes construyeron tres autómatas notables, dos chicos y una chica, los tres de tamaño natural.

Uno de los niños fue diseñado para escribir distintos mensajes de hasta cuarenta caracteres. Este autómatas movía el papel con una mano, y la pluma con la otra, mientras sus ojos seguían el trazado de la pluma sobre el papel. El otro niño hacía dibujos variados. Por último, la muchacha tocaba el órgano, y presionaba con sus dedos las teclas, mientras su pecho subía y bajaba con el ritmo de su “respiración”, y su cabeza acompañaba a la música. Estos tres autómatas estaban accionados por complejísimo mecanismos de relojería.

También, por esa misma época, Vaucanson acometía la construcción de los autómatas que, hoy en día, pueden considerarse más evolucionados. En efecto, su

tocador de flauta, el *tocador de caramillo*, y *el pato*, que graznaba, batía las alas, comía y hacía la digestión de forma mecánica, son de una perfección notable.

La diferencia fundamental entre estos autómatas y las criaturas de la mitología estriba en la manera de comunicarse. Así, para indicarle al niño escritor de los Droz que debía cambiar el texto, era necesario cambiar su *programa*, lo cual requería seis horas de trabajo de un experto relojero. Por el contrario, los autómatas de Hefaiostos entendían mensajes hablados.

Históricamente, la aproximación a las ciencias puede realizarse desde ópticas derivadas de la metafísica, o desde ópticas más intelectuales. Ambas aproximaciones son válidas, y muestran el interés del género humano por explicar fenómenos que aún no se entienden. Pero junto a estos intentos, válidos y legítimos del ser humano, aparecen siempre actitudes oportunistas como la que se refleja en la historia del *Malzel Chess Automaton*, un prodigioso autómata que era capaz de jugar al ajedrez al mismo nivel que los mejores especialistas de la época. En realidad el *Malzel Chess Automaton* era un fraude, y quien era un consumado jugador de ajedrez era el enano que se acomodaba en la carcasa del autómata.

Curiosamente, Edgar Allan Poe, sabedor de los prodigios que se le atribuían al *Malzel Chess Automaton*, construyó una prueba lógica según la cual el autómata en cuestión no podía ser auténtico. La argumentación de Poe se basaba en dos pilares fundamentales:

- Ninguna máquina puede cambiar su estrategia durante un desarrollo pretendidamente intelectual. Las mismas fuentes intelectuales no tienen por qué producir las mismas respuestas ante los mismos estímulos.
- No hay ninguna máquina que sea capaz de utilizar conocimiento derivado de la experiencia. Los procesos inductivos y de aprendizaje no son propios de ingenios mecánicos.

Lo que Poe no sabía es que su argumentación, que no era más que una traducción del llamado *régimen Lovelace*¹, es falsa. Hoy sabemos que las máquinas pueden, y en ocasiones deben, modificar sus estrategias de trabajo durante la resolución de ciertos problemas. Además, las máquinas incorporan conocimiento derivado de la experiencia² y lo utilizan para inferir nuevos hechos y para aprender de su propia experiencia.

Sin embargo, la verdadera historia de la inteligencia artificial se inicia con los deseos de Babbage de que su Máquina Analítica fuese capaz de tratar adecuadamente juegos como el ajedrez. En realidad, lo que Babbage pretendía era construir una máquina capaz de pensar, aprender y crear, y cuya capacidad para realizar estos actos se incrementase hasta que los problemas tratados fuesen del mismo nivel que los

¹ Según el cual una máquina sólo puede hacer lo que se ordena.

² En forma de conocimiento heurístico.

destinados a los humanos. Proyecto que, sin embargo, casi les cuesta la vida a Babbage y a su colaboradora Ada Lovelace.

Pero la curiosidad y el afán de superación son características propias del género humano, y la Historia sigue su curso inexorable. Así, en 1943 se publican tres artículos teóricos relativos a lo que hoy se conoce como Cibernética. En el primero de tales artículos, Wiener, Rosenblueth y Bigelow sugieren distintas formas de conferir fines y propósitos a las máquinas. En el segundo, McCulloch y Pitts ponen de manifiesto de qué modo las máquinas pueden emplear los conceptos de lógica y de abstracción y demuestran que cualquier ley de entrada-salida puede modelizarse a través de una red de neuronas artificiales. En el último de estos artículos, Craik propone que las máquinas empleen modelos y analogías en la resolución de problemas.

Pero las ideas contenidas en los tres artículos anteriores no salieron de la pura especulación teórica hasta mediados de los años 50, en los que los ordenadores de la época empezaban ya a ser adecuados para permitir la programación de procesos suficientemente complejos.

En 1956, un grupo de investigadores se reúne en el Dartmouth College para discutir sobre la posibilidad de construir máquinas genuinamente inteligentes. Entre los investigadores que allí se dieron cita estaban Samuel, que había desarrollado un programa para jugar a las damas; McCarthy, que se dedicaba a la construcción de sistemas en los que se llevasen a cabo razonamientos de sentido común, y Minsky, que trabajaba sobre un problema de geometría plana con la esperanza de conseguir que el ordenador emplease razonamiento analógico sobre figuras. Junto a éstos, Newell, Shaw y Simon, fueron los verdaderos promotores de la inteligencia artificial, término que fue acuñado en su día por McCarthy.

Tras esta reunión se formaron diversos grupos de científicos que siguieron trabajando de forma independiente. Así, Newell y Simon formaron un equipo con la idea de desarrollar modelos de comportamiento humano. Por su parte, McCarthy y Minsky formaron otro equipo dedicado a la construcción de máquinas inteligentes, sin preocuparse especialmente del comportamiento humano.

El primer planteamiento supone la emulación de la actividad cerebral y, en la medida de lo posible, la réplica de su estructura. El segundo planteamiento supone construir sistemas en los que los procedimientos empleados para resolver problemas sean de tal naturaleza que, de ser empleados por un ser vivo, éste sería considerado inteligente.

Ambas aproximaciones, decididamente divorciadas durante mucho tiempo, son sin embargo necesarias para obtener resultados mínimamente interesantes. La primera porque el estudio y el desarrollo de sistemas inteligentes, mediante el uso de simuladores, da la oportunidad de alcanzar epistemologías complejas. La segunda aproximación es necesaria por el gran interés que supone el diseño y el análisis de sistemas que sean capaces de resolver problemas intelectualmente difíciles.

Ambos enfoques cubren los objetivos fundamentales de la inteligencia artificial moderna: la comprensión de la inteligencia humana, y la utilización de máquinas para adquirir conocimiento y tratar satisfactoriamente problemas complicados.

Para comprender la inteligencia natural se asume que el comportamiento inteligente se rige por mecanismos automáticos. En este caso, el problema fundamental es encontrar el “programa” que rige cada actividad. Por otra parte, es evidente que los ordenadores hacen algunas cosas mejor que los cerebros humanos (e.g., son más robustos, calculan mejor,...), en estos casos parece conveniente potenciar estas características diferenciales de los ordenadores que los hacen mejores en algunas tareas.

1.2. Consideraciones Generales sobre la IA

Los dos planteamientos anteriores dan lugar a una dicotomía difícil de resolver: ¿qué es la inteligencia artificial, una ciencia o una ingeniería?

Como ciencia, la inteligencia artificial trata de desarrollar el vocabulario y los conceptos que permiten ayudar a entender, y en ocasiones a reproducir, comportamiento inteligente. Por otra parte, como ingeniería, la inteligencia artificial trata de definir y de utilizar un conjunto de métodos que nos permitan adquirir conocimiento de alto nivel, formalizarlo, representarlo según un esquema computacionalmente eficaz, y utilizarlo para resolver problemas en dominios de aplicación concretos.

Esta discusión acerca de los ámbitos de aplicabilidad de la inteligencia artificial nos permite clasificar a los sistemas inteligentes en tres niveles distintos: programas de IA, sistemas basados en conocimiento y sistemas expertos³.

Programas de IA

Los programas de IA exhiben cierto comportamiento inteligente fruto de la aplicación hábil de heurísticas en sentido amplio, entendiendo como heurística un tipo de conocimiento difícilmente formalizable, fruto de la experiencia, y que se establece implícitamente para tratar de encontrar respuestas más o menos correctas, pero siempre válidas, a un problema concreto. La utilización de conocimiento heurístico no garantiza encontrar soluciones óptimas, pero sí permite garantizar el hallazgo de soluciones aceptables, si existen, a través de los denominados *procesos inferenciales*.

Aparece ya un primer concepto que merece nuestra atención, y es el de *inferencia*, que podemos definir como el proceso que permite la comprensión de un significado en función de cierta información relacionada. La idea de inferencia está relacionada con los procesos de razonamiento, que frecuentemente exigen la realización de varias inferencias para lograr establecer conclusiones válidas.

El estudio del razonamiento⁴ nos permite su clasificación en, al menos, tres *modos* diferentes: razonamiento deductivo, razonamiento inductivo, y razonamiento

³ Existen distintos criterios en esta clasificación. Nosotros asumimos los criterios de Waterman.

abductivo. Para ilustrar los tres conceptos consideraremos tres aseercciones o declaraciones A1, A2 y A3, con las siguientes características:

- A1 = aseercción de carácter general
- A2 = aseercción de tipo particular referida a la primera parte de A1
- A3 = aseercción de tipo particular referida a la segunda parte de A1

El razonamiento deductivo parte de la premisa general A1 y de la premisa particular A2, y trata de demostrar la premisa particular A3.

Un ejemplo podría ser el siguiente:

A1 “La gripe produce fiebre”

A2 “Luis tiene gripe”

A3 “Luis tiene fiebre”

El modo de razonamiento deductivo es el modo propio de las Matemáticas y, realmente, su empleo no genera conocimiento nuevo. Con el razonamiento deductivo simplemente aplicamos conocimiento dado sobre situaciones particulares para obtener conclusiones válidas.

El segundo de los modos de razonamiento mencionados es el razonamiento inductivo, con el cual, partiendo de dos premisas de naturaleza particular, A2 y A3, tratamos de obtener la aseercción general A1. Un ejemplo podría ser el siguiente:

A2 “Estos objetos caen”

A3 “Estos objetos tienen masa”

A1 “Los objetos con masa caen”

Los procesos inductivos de razonamiento están ligados a la experimentación, y son propios de las ciencias naturales como la Física y la Química. El razonamiento inductivo sí genera conocimiento nuevo. Para ello se basa en la observación de eventos y en las posible relaciones que pueden encontrarse. No obstante, el razonamiento inductivo plantea algunos problemas. Así, si en el ejemplo anterior sustituimos la aseercción A3 “Estos objetos tienen masa”, por la aseercción A3' “Estos objetos son rojos”, la conclusión A1 sería “Los objetos rojos caen”, lo cual es evidentemente cierto puesto que los objetos, rojos o de cualquier otro color, tienen masa, y la gravedad hace que caigan. Sin embargo, el hecho de que caigan está ligado a su propiedad “masa”, no a su propiedad “color”. Así, los procesos inductivos de razonamiento pueden dar lugar a errores importantes, que se minimizan cuanto más cuidadoso sea el análisis previo de las premisas particulares que se pretenden relacionar.

⁴ No deben confundirse los tipos de razonamiento que aquí describimos, con los *modelos de razonamiento* que describiremos más adelante.

El tercero de los modos de razonamiento mencionados es el abductivo, que trabaja sobre la plausibilidad de las conclusiones, y que formalmente trata de relacionar aserciones de tipo A1 con aserciones de tipo A3, para concluir aserciones de tipo A2. Un ejemplo podría ser el siguiente:

A1 “Los cuadros de Goya presentan -estas- características”

A3 “Este cuadro presenta -estas- características”

A2 “Este cuadro es de Goya”

El modo de razonamiento abductivo trata de concluir algo sobre la posible relación entre dos aserciones que se sabe que son ciertas. Sin embargo, la relación establecida puede ser cierta o no⁵. En el razonamiento abductivo podemos reconocer las siguientes características:

- Está ligado al concepto de incertidumbre
- Genera conocimiento nuevo
- Es el modo de razonamiento típico de la inteligencia artificial⁶

Sistemas Basados en Conocimiento

El siguiente nivel es el de los sistemas basados en el conocimiento, en los que los conocimientos del dominio concreto, y las estructuras de control que se utilizan para manipular dicho conocimiento, se encuentran físicamente separados. Ello va a requerir la definición e implementación de arquitecturas diferentes a las que estamos habituados, y en las que “conocimientos” y “estructuras de control” puedan ser desarrolladas independientemente las unas de las otras, de forma que una misma estructura de control pueda ser utilizada en muchas *bases de conocimientos* diferentes.

Sistemas Expertos

Por último, los sistemas expertos pueden considerarse como especializaciones de los sistemas basados en el conocimiento que utilizan conocimiento particular de un dominio de aplicación concreto para tratar de resolver problemas del mundo real, limitados en tamaño, pero de gran complejidad.

La construcción de sistemas expertos va a requerir el empleo de las técnicas desarrolladas para construir programas de inteligencia artificial, y la utilización de las arquitecturas definidas para el desarrollo de sistemas basados en el conocimiento. Pero, además, va a ser imprescindible la realización de esfuerzos en aspectos diferenciales como son la *adquisición del conocimiento*, y el *aprendizaje*⁷. Al respecto, la

⁵ Hay muchos imitadores de Goya... Alguno de ellos muy bueno.

⁶ ... aunque no el único.

⁷ Por sólo citar dos de los aspectos diferenciales más importantes.

incorporación de información específica de un dominio se simplifica notablemente si somos capaces de establecer categorías en el conocimiento barajado.

Estas categorías pueden definirse, en relación al origen y procedencia del conocimiento mencionado, con respecto al experto humano de quien lo extraemos. Así, podemos establecer las siguientes grandes categorías de conocimiento:

- *Conocimiento público*: que puede obtenerse directamente a partir de fuentes típicas, libros y manuales, que es comúnmente aceptado y universalmente reconocido.
- *Conocimiento semipúblico*: que, siendo explícito, no está universalmente reconocido ni es comúnmente aceptado, y suele ser utilizado de forma casi exclusiva por grupos de especialistas en áreas concretas.
- *Conocimiento privado*: que no es explícito, no está universalmente reconocido, ni es comúnmente aceptado, es de marcado carácter heurístico, y responde muy bien a la conocida frase popular “cada maestrillo tiene su librillo”.

La adquisición de todos y cada uno de estos tipos de conocimiento requiere el desarrollo de metodologías específicas.

Los sistemas expertos constituyen frecuentemente soluciones elegantes en dominios bien conocidos, estructurados y, hasta cierto punto restringidos. Sin embargo, frente a los sistemas expertos, los sistemas de supervisión inteligente -o sistemas de monitorización inteligente-, amplían el horizonte de las aplicaciones de las técnicas de inteligencia artificial, y deben ser considerados como programas inteligentes capaces de:

- Interaccionar con instrumentos y diferentes fuentes de información y datos.
- Adaptarse a los diferentes contextos que puedan darse en un determinado dominio.
- Establecer prioridades entre diferentes tareas, adaptándose asimismo a las restricciones temporales del entorno y del caso analizado. El concepto de tiempo es fundamental en los sistemas de monitorización inteligente, que casi participan más de la idea de “sistemas de control” que de la de “sistemas expertos”.

Formalmente, la construcción de un sistema de supervisión inteligente pasa por el desarrollo de métodos y técnicas que posibiliten la medición y la estimación de los valores de un conjunto de parámetros.

También se requiere el desarrollo de métodos que faciliten el análisis y la interpretación de la información. Por último, los sistemas de monitorización inteligente deberán ser capaces de establecer líneas de actuación apropiadas y sugerir directrices eficaces.

Estos requisitos de carácter general han sido detallados recientemente en un estudio según el cual los sistemas actuales de monitorización inteligente deben tratar de:

- Incrementar la seguridad de los sistemas a controlar mediante la prevención de errores, y mejorar la calidad de las directrices de actuación sugeridas suministrando mejores procedimientos de ayuda a la decisión. Con ello se pretende lograr que el personal especializado emplee todos sus esfuerzos en la supervisión del proceso que está siendo analizado, y no en la supervisión de la instrumentación necesaria, que es muchas veces fuente de errores por exceso o por defecto.
- Proporcionar herramientas eficaces para la ayuda a la decisión, de forma que el personal especializado vea en el sistema un colaborador y no un competidor.
- Incorporar entornos interactivos de análisis de información y de presentación de resultados.

Junto a estas características fundamentales, también podemos exigir que nuestros sistemas inteligentes de supervisión:

- Sean flexibles y fácilmente reconfigurables, con el fin de posibilitar su adaptación a condiciones locales.
- Permitan su propio desarrollo incremental.
- Proporcionen buenos consejos, en tiempo y hora.
- Se integren de forma natural en los sistemas convencionales de manejo de la información, de manera que no modifiquen agresivamente las condiciones de trabajo del entorno.

Lo dicho anteriormente nos permite enfatizar algunos de los aspectos más interesantes de los sistemas inteligentes de supervisión. En concreto, habremos de considerar tres niveles en el tratamiento de la información, que deberán ser adecuadamente integrados. Tendremos que manipular señales y datos numéricos, lo cual obliga a su captura y posterior procesamiento. Al respecto, hoy día los modernos transductores y sensores, y las nuevas técnicas de procesamiento de señales, continúan proporcionándonos cada vez mejores vías para medir y estimar un número cada vez mayor de parámetros.

Por otra parte será imprescindible reconocer alarmas falsas y rechazar datos erróneos, lo que implica incorporar procedimientos de validación de las señales. Finalmente, habrá que traducir simbólicamente la información y relacionar entre sí los diferentes literales obtenidos, al objeto de poder establecer interpretaciones adecuadas - en un lenguaje apropiado-, que nos permitan trazar planes de actuación correctos. Cada una de estas tareas debe poder ejecutarse en el intervalo que requiera el problema que se esté tratando.

Aparte de todas estas consideraciones preliminares, también es deseable que los sistemas inteligentes de monitorización posean capacidades predictivas, cualitativas y cuantitativas, dirigidas hacia la evaluación de los planes de actuación prescritos, la optimización de los mismos, el establecimiento de alarmas contextuales, y el manejo de los recursos computacionales en función de las demandas del proceso estudiado.

La construcción de sistemas inteligentes de supervisión implica la integración de dispositivos en sistemas de información complejos. Pero este mismo planteamiento puede observarse a nivel perceptual y a nivel cognoscitivo. Así, un módulo inteligente de un sistema de información debe ser capaz de procesar eficientemente la información que le llega de los dispositivos que conectan al sistema con el mundo exterior. Esta información debe ser clasificada, jerarquizada, y etiquetada simbólicamente, antes de su tratamiento inferencial. En este contexto, y desde una perspectiva tecnológica, debemos tener en cuenta que las redes de neuronas artificiales -RNAs- son excelentes herramientas de clasificación que, además, pueden ser diseñadas para trabajar en tiempo real.

La inclusión en módulos inteligentes de estructuras conexionistas -basadas en RNAs-, da lugar a los *sistemas híbridos simbólico-conexionistas*, que tratan de aprovechar al máximo las ventajas de los sistemas neuronales y de los sistemas expertos. No obstante, hacer compatibles tecnológicamente ambas filosofías no es tarea fácil. El desarrollo de sistemas inteligentes cooperativos carece de una metodología adecuada de diseño. Su consolidación obligará, sin duda, a realizar esfuerzos de investigación y desarrollo durante los próximos años.

No podíamos terminar este apartado sin mencionar que en un texto relativamente reciente, Russell y Norvig plantean una aproximación a la inteligencia artificial en la que el cuerpo doctrinal central es la definición de *agente inteligente*, que es capaz de operar en un *ambiente*. En palabras de los propios autores "... de acuerdo con esto -el agente inteligente-, la tarea de la IA es explicar y construir agentes que reciben percepciones del ambiente y proceden a ejecutar acciones. Cada uno de estos agentes se implanta mediante una función que correlaciona preceptos y acciones... (de este modo se pueden definir) diversos procedimientos útiles para representar estas funciones, tales como los sistemas de producción, los agentes reactivos, los planificadores lógicos, las redes semánticas y los sistemas lógicos de decisión..." Este planteamiento tiene que ver con lo que hoy se llama *inteligencia artificial distribuida*, y su alcance puede verse en algunas de las explicaciones abordadas en este texto.

1.3. Organización Temática del Texto

Este libro está organizado en tres bloques temáticos, que abarcan "Aspectos Básicos de la Inteligencia Artificial" -PARTE I-, "Esquemas y Modelos de Razonamiento" -PARTE II-, e "Ingeniería del Conocimiento" -PARTE III-. También incluye un apartado de "Bibliografía Complementaria".

En el primer bloque temático, y tras este primer capítulo de Introducción, el capítulo 2 plantea la *resolución de problemas* desde la perspectiva de la inteligencia artificial. Se definen aquí conceptos básicos necesarios como el de *espacio de estados*, se discuten las características generales de los procesos de búsqueda, y se presentan y se discuten diferentes métodos y algoritmos, diseñados para abordar distintos tipos de problemas, desde aquellos que se pueden resolver prácticamente sin información, hasta los que precisan de heurísticas potentes para ser resueltos.

El capítulo 3 presenta y discute las ideas básicas que subyacen en todo esquema de representación del conocimiento. Una vez planteada la cuestión, se discute con cierto detalle el enfoque de los métodos formales de representación del conocimiento, y en particular la lógica de proposiciones y la lógica de predicados.

El capítulo 4 trata de resolver las deficiencias de los métodos formales de representación por medio de los llamados métodos estructurados de representación del conocimiento. Se discuten aquí dos grandes filosofías, los métodos declarativos y los métodos procedimentales, y se presenta el paradigma de *orientación a objetos*.

En el capítulo 5 se presenta parte de la problemática asociada a la representación del conocimiento temporal, y se proponen varias soluciones al respecto.

Concluimos esta primera parte con el capítulo 6, que es una presentación detallada de los llamados *sistemas de producción*: básicamente sistemas inteligentes basados en reglas, de los que se discuten estructuras como las *bases de conocimientos*, el *motor de inferencias*, o la *memoria activa*. También se le presta atención a la dinámica de estos sistemas, a través de la descripción de sus *ciclos básicos*.

La PARTE II de este texto se centra fundamentalmente en distintos esquemas y modelos de razonamiento. Globalmente, este bloque temático avanza de manera secuencial, y cada tema se presenta como una solución de –al menos– parte de los inconvenientes encontrados en el tema anterior.

Iniciamos este bloque temático con el capítulo 7, que aborda el problema del razonamiento desde una perspectiva lógica categórica, a través de la interpretación diferencial.

Para tratar de paliar las deficiencias de la interpretación diferencial, en el capítulo 8 planteamos diversas aproximaciones estadísticas, como el *esquema bayesiano* o las *redes de creencia*. No obstante las soluciones de los modelos estadísticos no son completamente satisfactorias, lo que nos lleva a abordar el problema del razonamiento desde perspectivas cuasi-estadísticas.

El capítulo 9 desarrolla el modelo de los factores de certidumbre, propuesto en 1975 por Shortliffe y Buchanan, y parte central del sistema experto MYCIN, verdadero paradigma de la tecnología de los sistemas expertos. De este modelo se discuten, aparte de sus conceptos básicos, cuestiones como la combinación de evidencias o la propagación de incertidumbre.

Las carencias fundamentales del modelo de Shortliffe y Buchanan son fruto de su falta de base teórica. Este aspecto nos sirve como excusa para introducir, en el capítulo 10, la *teoría evidencial* de Dempster y Shafer, que además contiene al modelo de los factores de certidumbre como un caso particular.

Finalizamos este segundo bloque temático con el capítulo 11, dedicado a los *conjuntos difusos*, de los que estudiamos cuestiones como su estructura algebraica, y diversos problemas de representación del conocimiento y de razonamiento.

El tercer bloque temático es una introducción a la ingeniería del conocimiento y a diversos tipos de sistemas inteligentes. El capítulo 12 aborda conceptos básicos y preliminares como son el *análisis de viabilidad de un sistema experto*, la *organización general* de un sistema experto, o el problema de la *adquisición del conocimiento*.

Por otra parte, en el capítulo 13 se presentan sistemas que implementan una arquitectura específica como es la arquitectura de *pizarras*, mientras que en el tema 14 se discuten algunos aspectos importantes de los *sistemas inteligentes de monitorización*, también llamados *sistemas de supervisión inteligente*.

El capítulo 15 presenta y enfatiza los beneficios de las *redes neuronales* y de los *sistemas simbólico-conexionistas*, de los que se discuten varias arquitecturas. Ello nos permite introducir el concepto de *hibridación*, que es tratado en el capítulo 16. Por último, el capítulo 17 plantea, discute, y propone soluciones para tratar de resolver el problema importantísimo de la *validación de sistemas inteligentes*.

La última parte de este libro recoge una selección bibliográfica, no necesariamente utilizada en el desarrollo de este texto, pero que juzgamos necesaria para documentar y guiar los pasos de aquellos lectores que pretendan profundizar en el dominio de la inteligencia artificial.

1.4. Resumen

En este capítulo introductorio hemos intentado dar una visión global del ámbito de la IA. También se han discutido enfoques que podríamos llamar “precursores”, y cuyos mejores ejemplos han sido encontrados en la Mitología. Desde una perspectiva histórica, aparte de algunas definiciones y conceptos, hemos tratado de enfatizar los distintos planteamientos surgidos a mediados del siglo XX y que desembocan en una feroz lucha entre los llamados “pulcros”, partidarios de abordar la IA desde una perspectiva formal, y los llamados “desaliñados”, partidarios de abordar la IA desde una óptica funcional: Algo es inteligente, si efectivamente se comporta como tal. Por último, y dado que, en opinión de los autores, el futuro de la IA es el desarrollo de sistemas integrados en sistemas de información más amplios, el capítulo concluye con una breve mención a los sistemas inteligentes de supervisión, a los sistemas híbridos, y al problema de la validación.

1.5. Textos Básicos

- Alonso y Moret, “Tutorial: Medical Applications of Expert Systems”, En: Proc. 2nd World Conference on Expert Systems, Pergamon Press, eds., 1994.
- Borrajo, Juristo, Martínez y Pazos, “Inteligencia Artificial: Métodos y Técnicas”, Centro de Estudios Ramón Areces, 1993
- Moret, “Monitorización Inteligente en Medicina: Principios Básicos y Aplicaciones”, En: Inteligencia Artificial en Medicina, Pazos y Ezquerro, eds., Fundación Alfredo Brañas, 1994

- Varios autores, “Seminario: La Inteligencia Artificial como Ingeniería”, CETTICO, 1990
- Russell, Norvig, “Inteligencia Artificial: Un Enfoque Moderno”, Prentice Hall Iberoamericana, eds., 1996.