

Introducción a las

Tecnologías Inteligentes

Descubriendo la inteligencia artificial: Primeros pasos en el Mundo de las Máquinas Inteligentes.



*Ilustración 1: Osa panda roja, desafiante, rodeada de tecnología.
Fuente: Midjourney; propiedad del autor.*

Juan Francisco Puentes Calvo, Phd.

Edición 2024

Sección I: Conceptos clave de la inteligencia artificial

"La inteligencia artificial es la ciencia de hacer que las máquinas hagan cosas que requerirían inteligencia humana si se hicieran por humanos"

Marvin Minsky (1927 - 2016)

En esta sección vamos a hacer un viaje por los conceptos clave de la inteligencia artificial; veremos qué es y que no es la IA, ejemplos de uso en el mundo real, una aproximación a los algoritmos para resolver problemas, cómo los agentes inteligentes perciben e interactúan con el entorno, cómo representan y almacenan la información de tal modo que se pueda razonar sobre la misma y, finalmente, aprenderemos sobre los conceptos de aprendizaje, el sesgo y la equidad.

Capítulo 1

¿QUÉ ES LA INTELIGENCIA ARTIFICIAL?

1.1 Un mundo de datos, información y conocimiento.

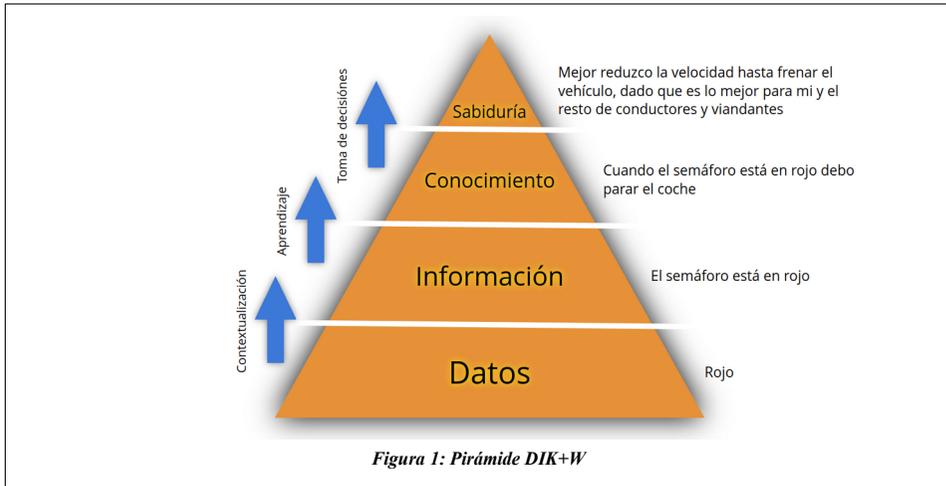
Los **datos**, la **información** y el **conocimiento** son conceptos fundamentales en el ámbito de la gestión de la información y el procesamiento de datos. Aunque a menudo se utilizan indistintamente, datos e información representan un nivel diferente de comprensión en el proceso de adquisición y aplicación del conocimiento.

Comencemos por los datos. Los **datos** son hechos crudos, representaciones objetivas y neutrales de eventos, medidas o características. Pueden ser numéricos, alfanuméricos, textuales o incluso imágenes y sonidos. Por sí solos, los datos no tienen un significado intrínseco ni contexto. Son simplemente elementos de información sin procesar que requieren interpretación para extraer información útil.

*Nota: es muy importante comprender que todos los datos, para ser procesados por una máquina, han de ser convertidos previamente en números. En general este proceso de **codificación** queda oculto al usuario.*

Cuando se organizan, se estructuran y se dan sentido a los datos, estos se convierten en información. La **información** se refiere a datos procesados y contextualizados que

adquieren relevancia y significado para el usuario. La información proporciona respuestas a preguntas específicas y permite una comprensión más profunda de un tema o situación. A través de la interpretación y el análisis de los datos, se extraen patrones, tendencias y relaciones que ayudan a tomar decisiones informadas.



Sin embargo, la información por sí sola no es suficiente para lograr un conocimiento completo. El **conocimiento** implica la comprensión y la asimilación de la información, lo que permite aplicarla de manera efectiva en diversas situaciones. **Para crear conocimiento a partir de la información es necesario el aprendizaje.**

En la sociedad actual, el valor de los datos, la información y el conocimiento ha aumentado significativamente debido al rápido avance de la tecnología y la digitalización. Las organizaciones y las personas tienen acceso a enormes cantidades de datos y pueden utilizar herramientas y técnicas para procesarlos y extraer información relevante. La capacidad de transformar datos en información significativa y, finalmente, en conocimiento se ha convertido en una ventaja competitiva clave en muchos campos.

Nuestra sociedad, desde el siglo XIX, ha navegado a lo largo del tiempo por fases que pueden igualarse con los conceptos de datos, información y conocimiento:

Durante el siglo XIX, gracias a la revolución industrial, se aplicaron conceptos científicos a la gestión del trabajo, y esta metodología implicó la recolección de datos como materia prima para la optimización de los procesos productivos. Poco a poco, a medida que la tecnología avanzó, fuimos creando máquinas que pudieran ayudarnos a gestionar estos datos, hasta que durante la década de 1980 hizo su aparición el ordenador personal. Ya existían antes productos comparables, pero fue durante esta década cuando empezaron a popularizarse ampliamente. Podríamos decir que fue cuando la sociedad se dio cuenta de

su existencia y utilidad. Ese fue el principio de la **sociedad de los datos**.

Poco tiempo después, a principios de la década de los 90 del siglo pasado y gracias al desarrollo y popularidad de Internet, apareció la *World Wide Web*, la red extensa mundial¹⁵, que conectaba documentos enriquecidos, accesibles en Internet por medio de hiperenlaces, gracias al concepto de URL¹⁶. Pero no fue hasta la última década de siglo XX cuando la sociedad en general se dio cuenta de su existencia y utilidad. Y lo hizo de forma dramática, coincidiendo con la crisis de las “*punto com*”. Ese fue el principio de la **sociedad de la información**.

NOTA: Las sociedades, a medida que avanzan en su desarrollo, van atravesando fases caracterizadas por un descubrimiento y su consiguiente uso generalizado. Hablamos de la sociedad de X, cuando la sociedad descubre y acepta/usa X de forma generalizada.

En nuestro relato hemos descrito cómo entramos en la sociedad de los datos y en la sociedad de la información. Pero, ¿hemos entrado en la sociedad del conocimiento?

Como hemos reflejado anteriormente, para pasar datos/información a conocimiento, debemos hacerlo por medio del aprendizaje. Así que para pasar de una sociedad basada en datos e información [esto es: que posee automatismos que procesan datos/información] a otra basada en el conocimiento [esto es: que posea automatismos que procesen conocimiento], debemos conseguir que, en primer lugar, dichos automatismos sean capaces de aprender y, en segundo lugar, que la sociedad conozca su existencia y su utilidad.

Conocemos automatismos que poseen la capacidad de aprender desde 1957 [*perceptrón*, Frank Rosenblatt], y desde entonces se han desarrollado multitud de arquitecturas y estrategias para optimizar su proceso de aprendizaje y su precisión. Sin embargo, en noviembre del 2022 se dio a conocer ChatGPT¹⁷ por parte de la empresa OpenAI. Como veremos la familia GPT [y otras: BERT, T5, etc..] está basada en una arquitectura creada en el 2017, la arquitectura basada en *Transformers*, pero no fue hasta finales del 2022 cuando la sociedad en general se dio cuenta de su existencia y utilidad. Así que podemos marcar tentativamente esa fecha como el inicio de la **sociedad del conocimiento**.

¿Y qué hay más allá de la sociedad del conocimiento?

Es difícil decirlo. Podemos apoyarnos en la pirámide DIK+W [*data, information, knowledge* y *wisdom*, respectivamente datos, información conocimiento y sabiduría] de la Figura 1. Al

15 Popularmente llamada simplemente “web”.

16 Más tarde ampliado al concepto de URI (Uniform Resource Identifier).

17 Realmente GPT3 con un entrenamiento específico para responder a preguntas o llevar a cabo diálogos: por tanto GPT v3.5.

hacerlo podemos afirmar que después del conocimiento viene la sabiduría¹⁸. Un dicho muy conocido dice “*la mejor forma de predecir el futuro es inventándolo*”¹⁹. Esta frase resume la idea de que, en lugar de simplemente esperar y tratar de predecir lo que sucederá en el futuro, es más efectivo tomar un papel activo en la creación del futuro que se desea.

Así que ¿qué hay más allá de la sociedad del conocimiento?, espero que la **sociedad de la sabiduría**, la cual se ha tomado conciencia de ella y de su utilidad. Y ¿cómo llegamos a ella? por medio del uso sabio del conocimiento en pro de un bien común; esto es: ética, equidad y justicia [Capítulo 8, Capítulo 14 y Capítulo 19].

1.2 ¿Qué es la inteligencia artificial?

Antes de hablar de inteligencia artificial, centremos nuestra atención en el concepto de *inteligencia*. **¿Qué es la inteligencia?** No hay una única definición de inteligencia que sea universalmente aceptada, dado que el concepto de inteligencia es complejo y multidimensional, y su comprensión puede variar según el contexto cultural, disciplinario y teórico. De todas formas podemos aproximarnos con cierta universalidad al decir que²⁰:

Inteligencia es la suma de las capacidades mentales que permite de forma muy general resolver problemas, razonar y planificar, pensando de forma abstracta, comprendiendo ideas complejas y con una gran capacidad de aprendizaje.

Y **¿qué es la inteligencia artificial?** Una definición común de inteligencia artificial ampliamente aceptada es:

Conjunto de técnicas, tecnologías y herramientas que tienen como objetivo emular o simular la inteligencia humana por medio de máquinas.

En general, la mayoría de los expertos coinciden en que se trata de la ilusión de procesos de inteligencia humana por parte de máquinas. Estos procesos incluyen: la captación de información de su entorno y su representación y almacenamiento; influencia sobre dicho entorno, el aprendizaje y las reglas para su uso; y el razonamiento usando las reglas para llegar a conclusiones.

18 Capacidad o habilidad de usar el conocimiento para un bien, principalmente común.

19 Este dicho se atribuye a Alan Kay, un destacado científico y visionario tecnológico conocido por sus contribuciones en el campo de la informática y por ser uno de los pioneros en la investigación de la interacción humano-computadora.

20 El artículo en la wikipedia correspondiente a la inteligencia, puede mostrarte más en profundidad sus diferentes definiciones y controversias: <https://es.wikipedia.org/wiki/Inteligencia>.

Dejándonos llevar por el reduccionismo, una definición más concreta es:

La inteligencia artificial es la inteligencia llevada a cabo por máquinas.

La inteligencia artificial no es un término desconocido para nadie, y esto es así porque ha estado presente en la mente, no sólo de científicos o ingenieros, sino también de creadores y filósofos.

De hecho, la presencia de la inteligencia artificial en la literatura o en el cine hacen que sea prácticamente imposible no conocer el concepto general y visualizarlo como algo posible y real en nuestro día a día. Lo que mucha gente no sabe es que está usando aplicaciones con inteligencia artificial casi a diario, o es beneficiario de que otros las utilicen.

1.3 Historia de la inteligencia artificial

Este apartado está basado en una entrada del blog²¹ de Fernando Sancho Caparrini; parcialmente reproducido con su permiso. Si algo está bien hecho, no lo reinventemos.

La Inteligencia Artificial como disciplina ha surgido de una larga historia de avances y retrocesos. El desarrollo de la IA está profundamente ligado a la Historia de la Matemática, la cual proporciona los fundamentos para comenzar la búsqueda de una forma de automatizar aquellos procesos humanos o que desearan hacer los seres humanos, entre ellos la inteligencia.

Podemos remontarnos a Aristóteles con su idea de la deducción lógica, como un primer intento de crear un mecanismo intelectual, aunque si bien conceptual. En el siglo XIII, Ramón Llull²² ideó la "*Ars generalis ultima*", una máquina de carácter mecánico programable para evaluar afirmaciones lógicas. En sus obras filosóficas anticipó la lógica como cálculo mecánico con símbolos en el que se aplicaba un método, los métodos heurísticos²³ de la Inteligencia Artificial, los sistemas generativos, los grafos, las redes semánticas, los diagramas de Venn. Influyó en Montaigne, Pascal, Descartes, Leibnitz, Newton, Pico della Mirandola y Giordano Bruno, entre otros.

En el siglo XIX hay que destacar a Charles Babbage y Ada Lovelace, quienes agregaron capacidad de programación a las máquinas de calcular, y Bernhard Bolzano, que realizó un intento para formalizar la semántica de la matemática moderna, más allá de su sintaxis.

21 <http://www.cs.us.es/~fsancho/?e=221>, posteriormente publicado en "Revista de Occidente" n.º 446-447, Julio-Agosto 2018 (<https://ortegaygasset.edu/producto/no-446-447-julio-agosto-2018>).

22 Para Martin Gardner; el gran escritor moderno de las matemáticas, Llull era como uno de esos "excéntricos que se convierten en los fundadores de las sectas", un embaucador que gira ruedas mágicas para invocar fórmulas vacías disfrazadas de verdades profundas". Quizá Gardner exagera un poco.

23 Métodos heurísticos → métodos basados en la experiencia.

Pero los antecedentes de la historia moderna de la IA podemos situarlos a finales del siglo XIX, época en la cual los matemáticos tenían la esperanza de que toda la matemática pudiese reducirse a términos lógicos²⁴. David Hilbert presentó en 1900 una lista de 23 problemas aún sin resolver por aquel entonces, Más adelante, en 1928, replanteó uno de los problemas [el segundo de los veintitrés] en tres preguntas equivalentes y muy relacionadas con el reduccionismo de las matemáticas:

1. ¿Son las matemáticas completas?
2. ¿Son las matemáticas consistentes?
3. ¿Son las matemáticas decidibles?

Si la respuesta de las tres preguntas era sí, entonces se podría reducir toda la matemática a términos lógicos.

Bard dice: Si toda la matemática se pudiera reducir a términos lógicos, tendríamos un lenguaje universal que permitiría comunicarnos con precisión y rigor sobre cualquier concepto matemático. Esto tendría un impacto profundo en la ciencia, la tecnología y la educación, y nos permitiría resolver problemas matemáticos más difíciles, desarrollar nuevos productos y servicios, mejorar la educación matemática y comprender mejor el mundo natural.

En 1931 Kurt Friedrich Gödel presentó un artículo con título "Sobre sentencias formalmente indecidibles de Principia Mathematica y sistemas afines", en donde, de forma magistral demuestra que no es posible lograr un reduccionismo completo en los sistemas matemáticos clásicos²⁵, ya que existen sentencias que de las que no podemos demostrar si son o no ciertas y sistemas cuya consistencia es imposible de verificar. Por tanto, de las tres preguntas de Hilbert, las respuestas a las dos primeras son negativas... ¿y la tercera? El artículo de Gödel se encargó de las dos primeras preguntas, sin embargo, no podían resolver la tercera pregunta. La dificultad estaba en la ausencia de significado de lo que se entiende por un «**procedimiento mecánico**».

En 1936, Alan Turing, en su artículo "Los números computables, con una aplicación al Entscheidungsproblem"²⁶, introduce el concepto de la **máquina de Turing** y, junto a Alonzo Church, demostraron que es imposible decidir con un algoritmo general si ciertas afirmaciones matemáticas son ciertas o falsas. Por tanto la respuesta a la tercera pregunta era, al igual que las otras dos, no.

Alan Turing demostró que su máquina teórica era capaz de resolver cualquier problema

24 Esto es, convertir toda la matemática en una rama de la lógica.

25 Esto es, que no se puede reducir toda la matemática a unos principios exclusivamente lógicos.

26 "Entscheidungsproblem" se traduce como "problema de decisión".

matemático que pudiera representarse mediante un algoritmo. Por tanto demostró que hay problemas que son inherentemente irresolubles por cualquier automatismo y que hay límites fundamentales para lo que la computación puede hacer. En el campo de la inteligencia artificial, estas limitaciones son importantes para entender los límites de la capacidad de las máquinas y por tanto los límites para simular o emular la inteligencia humana con máquinas.

A principios del siglo XX se produce una explosión de resultados teóricos que establecen las bases de lo que hoy conocemos como **Ciencias de la Computación**. Y, relacionado con la inteligencia artificial, Alan Turing empieza a preguntarse ¿podrá una máquina simular o emular la inteligencia humana?

El **primer artículo científico** en donde se planteó esta cuestión es "Intelligent Machinery" de 1948. En este artículo, Turing explora la idea de la inteligencia artificial desde una perspectiva más general. Propone la noción de una "máquina universal", que es capaz de simular cualquier otra "máquina de Turing", y discute cómo se podrían programar máquinas para realizar tareas inteligentes.

Poco después, en 1950, publica "Computing Machinery and Intelligence". Este es uno de los artículos más famosos de Turing, en el cual propone el famoso **Test de Turing** como una forma de evaluar la inteligencia de una máquina. En este artículo se plantea la pregunta **¿Pueden las máquinas pensar?** y discute temas como la simulación de la inteligencia humana y la posibilidad de que las máquinas puedan exhibir comportamiento inteligente.

Dos años antes de su fallecimiento en 1954, Alan Turing publica dos artículos más:

- "Machine intelligence". En este trabajo se centra en el aprendizaje automático y la capacidad de las máquinas para aprender.
- "Can a machine be conscious?". En este artículo se adentra en la cuestión de la conciencia en las máquinas. Explora la posibilidad de que las máquinas puedan desarrollar una forma de conciencia, y plantea preguntas sobre la naturaleza de la mente y la relación entre la mente y la máquina.

Ahora debemos volver un momento a 1943. En ese año Warren McCulloch y Walter Pitts presentan un modelo matemático abstracto de neuronas artificiales, representando la estrategia conexionista, frente a la simbólica que sería la imperante hasta finales de siglo XX. Volveremos a ello en el Capítulo 18.

1.3.1 Nacimiento – formal – de la inteligencia artificial.

Alan Turing no fue el único, podemos encontrar trabajos muy elaborados como el que en 1949 propone Claude Shannon²⁷ para la creación de un jugador automático de ajedrez. Aunque pueda parecer anecdótico el trabajo de Shannon presenta una particularidad que lo sitúa como uno de los trabajos esenciales en la historia de la IA, dado que propone un algoritmo novedoso basado en un trabajo de otra figura fundamental de la computación, John von Neuman, que llamó **minimax**, para buscar la respuesta más adecuada que debería dar la máquina por medio de una representación de las posibles respuestas como un espacio de estados a la vez que intenta dar un método uniforme de resolución para todos los casos. Lo veremos en el Capítulo 3.

*El algoritmo **minimax** es un método de decisión en teoría de juegos para determinar la mejor jugada en juegos de dos jugadores con información perfecta²⁸, explorando todas las posibles jugadas en un árbol de juego. Un jugador busca maximizar el valor de la función objetivo, mientras que el otro intenta minimizarlo, seleccionando al final la jugada que tenga en cuenta la estrategia óptima del oponente.*

En estos mismos años, Arthur Samuel, ingeniero de IBM, propone un jugador automático de damas que presenta una particularidad que adelanta cuál será una de las líneas más fructíferas del futuro de la IA, y es que sobre el algoritmo minimax de Shannon, Samuel añade un **proceso de aprendizaje** que permite a su programa mejorar a medida que va jugando con seres humanos, ayudándolo a decidir qué movimiento es el más adecuado en función de los resultados obtenidos de partidas anteriores. Los resultados fueron tan buenos que se obtuvo un jugador artificial que suponía un buen contrincante y que pudo implementarse en la primera máquina comercial de IBM, el IBM 701, en 1956. Este mecanismo de aprendizaje, que el propio Samuel llamó **machine learning** [aprendizaje automático, Capítulo 15], será el detonante de una de las técnicas de IA que más frutos proporciona en el presente.

Como vemos, el terreno ya estaba abonado para el nacimiento oficial de la disciplina y, tal y como se recoge en todas las referencias habituales sobre inteligencia artificial, en la **Conferencia de Dartmouth**²⁹ de 1956 se acuña oficialmente el término de **inteligencia artificial**. Los asistentes a esta conferencia serán las figuras fundamentales de la disciplina en los siguientes años: Shannon, Minsky, Newell, Simon, Rochester y Samuel, entre otros³⁰. Durante la reunión, el grupo intenta definir claramente las bases y líneas de trabajo que

27 Considerado padre de la Teoría de la Información.

28 La "información perfecta" en teoría de juegos se refiere a una situación en la que ambos jugadores tienen acceso completo y exacto a toda la información relevante en un juego en cualquier momento dado.

29 Una especie de sesión de trabajo de un mes convocada por McCarthy.

30 Turing había fallecido dos años antes, a la edad de 41 años.

deben ser prioritarias para la maduración de la disciplina, así como predecir los avances que podrían esperarse en los diez años siguientes.

El exceso de optimismo reinante en la conferencia, alimentado por los buenos resultados sobre los que la inteligencia artificial nacía, será el responsable parcial de la sensación de desilusión que llevará, finalmente, a lo que se conoce como el **invierno de la IA**, una etapa de falta de resultados y abandono por parte de los centros de investigación que durará años.



Figura 2: Participantes de la convención de Dartmouth.

Fuente: Familia Minsky (autora, Gloria Minsky)

1.3.2 Los fríos inviernos

Según hemos visto, los avances en jugadores artificiales se producen en un plazo de casi cuarenta años. Sin embargo, en 1956, cuando se bautiza formalmente el área, apenas habían pasado unos años desde el artículo de Turing [1948] y se habían conseguido resultados espectaculares: se pasa de un problema matemático puramente teórico: ¿es posible mecanizar los procesos matemáticos de razonamiento? a la respuesta de este problema – recordemos, negativa –, se crea una rama matemática que comprende el conjunto de técnicas y resultados relacionados con esta respuesta – la computación –, se generan los primeros ordenadores digitales que materializan el fundamento de esta nueva rama – con la consecuente posterior aparición de la informática como nueva ingeniería –, incluso se llegan a comercializar ordenadores con lenguajes de programación y se obtienen los primeros resultados que confirman que se pueden crear automatismos que resuelven problemas propios de la inteligencia humana.

Sin embargo, a esa acelerada evolución inicial le sigue una gran lentitud en la progresión de resultados que hace que el interés por el área decaiga de forma continuada y provoca un paulatino abandono por parte de departamentos de investigación en universidades y

empresas. Durante las décadas sesenta a noventa los departamentos de IA se convierten en pequeños representantes de lo que pudo haber sido y no fue, muchas veces sustituidos o arrinconados por otros departamentos que sacan provecho de las capacidades de cálculo de las computadoras en problemas menos ambiciosos, y reemplazando la computación por una ingeniería informática que no tiene a la IA como inspiración.

Bard dice: La IA ha tenido tres inviernos importantes. El primer invierno de la IA tuvo lugar en la década de 1960, cuando la comunidad científica se dio cuenta de que la IA se estaba desarrollando más lentamente de lo que se esperaba. Esto llevó a una disminución del interés y la financiación para la investigación en IA. El segundo invierno de la IA tuvo lugar en la década de 1970, cuando el campo de la IA se vio sacudido por una serie de fracasos importantes [...]. Esto llevó a una nueva disminución del interés y la financiación para la investigación en IA. El tercer invierno de la IA tuvo lugar en la década de 1980, cuando este campo se vio nuevamente sacudido por una serie de fracasos [...].

Una breve matización. Cotejando con otras fuentes debemos matizar a Bard. Aquellas citan dos inviernos [1974-1980 y 1987-1993] y varios otoños [episodios más pequeños de falta de financiación]. También puedes encontrar fuentes que hablen de un único *invierno de la IA* [como este mismo texto], que transcurrió principalmente durante las décadas de los 70 y 80 del siglo pasado, con momentos en donde parecía que la IA volvía a tener interés y financiación³¹.

No debemos llevarnos la impresión de que no se hizo nada más durante este largo periodo invernal [o periodos]. Por ejemplo, en ese tiempo hay avances que se relacionan con el origen de nuestro relato. Junto con la evolución computacional de los ordenadores – quizás insuficiente para atacar directamente juegos como el Go, pero capaz de resolver muchos otros problemas – y la mejora en el diseño de lenguajes de programación más fieles al espíritu matemático [como LISP o Prolog], se potencia el uso de la computación para la demostración matemática de teoremas, y aparecen demostradores automáticos que intentan alcanzar el objetivo que Ramón Llull persiguió con su primitiva máquina lógica. Aunque inmersas en el frío del invierno, en los años ochenta y parte de los noventa, se viven breves primaveras gracias a herramientas como los **Sistemas Expertos** que, a partir de sistemas basados en reglas lógicas y por medio de algoritmos de búsqueda y heurísticas similares a los desarrollados para juegos, proporcionan gran capacidad de toma de decisiones en entornos complejos. Estos sistemas no tienen carácter de inteligencia general y trabajan en un solo dominio de aplicación – por ejemplo, la diagnosis médica –, pero ocasionalmente ofrecen resultados que superan los obtenidos por un experto humano.

³¹ Por ejemplo el “Proyecto japonés de quinta generación” al principio de la década de los 80, financiación sustentada en la esperanza que se creó al rededor de los Sistemas Expertos.

A finales de los ochenta surge el concepto de **agente inteligente** – entes³² artificiales con capacidad de percepción, razonamiento y acción – y con él la **inteligencia artificial distribuida**, en la que un conjunto de estos agentes [sistemas multiagentes] interactúan entre sí y con el entorno con el fin de resolver un determinado problema. Junto a esta nueva IA surgen conceptos que han demostrado ser de una importancia fundamental para la resolución inteligente de problemas y que nos acerca a la forma en que el ser humano resuelve problemas de forma social: capacidad de comunicación entre los agentes, planificación, cooperación, percepción, modificación del entorno, competitividad, evolución, confianza, aprendizaje, adaptación, etc.

Es también al final del siglo XX cuando las **redes neuronales artificiales** comienzan a tomar el papel protagonista del que gozan hoy en día. A pesar de que la neurona artificial aparece en los primeros años de la IA, la capacidad computacional de los ordenadores de las primeras décadas no permitían hacer uso de un número de neuronas artificiales que proporcionase resultados interesantes. Además, no se disponía de ningún mecanismo para obtener redes de neuronas que resolviesen problemas concretos, aunque había resultados teóricos que avalasen esta posibilidad. Tras la creación del **algoritmo de aprendizaje** por medio de la propagación de errores hacia atrás [*backpropagation*, §18.2] por Werbos en 1974 se empezaría a extender su uso en problemas concretos de optimización y aproximación de funciones, pero incluso en los noventa todavía no se consideran importantes para la IA, y tienen un uso limitado a la aproximación de ciertos conjuntos de datos de no demasiada complejidad en el mundo de la ingeniería.

1.3.3 ¿Una nueva primavera?

La primera década del siglo XXI destaca por un cambio de aplicación de la IA. Frente a los juegos, en lenta pero constante progresión, y la aplicabilidad de los Sistemas Expertos, las necesidades industriales exigen un cambio de orientación, pidiendo que la IA se acerque a uno de los problemas que, pese a ser reconocido como importante desde hacía tiempo, era inabarcable desde el punto de vista práctico por su elevada complejidad: si queremos una máquina que se comporte inteligentemente, es necesario que pueda comunicarse con el ser humano con sus mismos medios, usando para ello el lenguaje natural que usa el humano. Para atacar este problema aparece una nueva disciplina dentro de la IA que se denomina **Procesamiento de Lenguaje Natural**, que aúna técnicas clásicas de IA con otras de ramas tan dispares como son la Neurociencia, Lingüística, Anatomía, Psicología, etc.

Durante esta primera década esta área evoluciona constantemente y vive uno de sus hitos en 2011 con la victoria del **sistema Watson**, desarrollado por IBM, en el concurso televisivo «*Jeopardy!*» de preguntas y respuestas en las que el ordenador tenía que entender la pregunta – muchas veces interpretando el doble sentido inherente al concurso – y buscar la respuesta adecuada en un tiempo reducido y contra adversarios humanos.

Los buenos resultados obtenidos se deben a una múltiple vía de ataque. Por una parte, la

32 Probablemente te hayas apoyado en la imagen de un androide (robot con forma humana), pero realmente, al hablar de agentes inteligentes, no referimos un modelo abstracto. Lo veremos más adelante.

creación de ontologías generales de conocimiento que reorganizan dominios cada vez más grandes del saber humano, y estructurados de forma que es posible mezclar dominios diferentes añadiendo relaciones semánticas transversales. Por otra parte, el uso de técnicas cada vez más depuradas para estructurar la información de forma relacional como una gran red de conocimientos. Además, no debemos olvidar el uso de algoritmos cada vez más eficientes para la búsqueda en grandes bases de información. Y, por último, el uso de algoritmos de optimización y aproximación que permiten considerar un número muy elevado de parámetros en menor tiempo, lo que se traduce en mayor capacidad para aproximar datos complejos.

Pero, sin duda, la mayor transformación sufrida por la IA en el siglo XXI ha sido el peso que ha tomado el **machine learning** [ML, o “aprendizaje automático” en castellano], y que acapara la casi totalidad de nuevas técnicas que se están desarrollando.

Hasta el momento la vía más común con la que se abordaban problemas de IA era con la construcción de algoritmos, más o menos generales, para encontrar la solución del problema recorriendo el espacio de posibles soluciones, como hacen algoritmos como el mencionado *minimax*. La aproximación que propone el ML es muy distinta, y pasa por utilizar datos reales del comportamiento deseado para construir una máquina que sea capaz de reconocerlos y simularlos. En este sentido, el problema se traduce en utilizar una correcta representación de los datos junto a algún mecanismo de optimización para poder ajustar la máquina al comportamiento observado; que es lo que se conoce como **entrenamiento**. Al igual que en la vía tradicional se crean algoritmos genéricos de búsqueda en el espacio de soluciones, en ML también se crean algoritmos de optimización generales que, trabajando sobre representaciones adecuadas, sean capaces de ajustar correctamente la máquina al comportamiento observado en los datos, independientemente del dominio del problema.

Es aquí donde el ML se confunde con otras disciplinas con fines similares, como el **aprendizaje estadístico** o la **minería de datos**, pero marcando la diferencia en los objetivos que persigue y en los procedimientos computacionales en los que se basa. Por ello, podemos encontrar muestras de este tipo de nuevos algoritmos desde hace muchos años, usando simples regresiones y correlaciones estadísticas, Árboles de Decisión [1979], Máquinas de Soporte Vectorial [1963], o Bosques Aleatorios [2001].

Pero sin duda, los algoritmos y modelos de ML más famosos son las nuevas variantes derivadas de las **redes neuronales artificiales**. Se sabía que la capacidad de adaptación de las redes neuronales depende de la estructura que tenga la red y que, cuanto más compleja es su estructura, más costoso es dar con la configuración óptima para asegurar que la máquina se ajuste al comportamiento de los datos hasta el punto de que el algoritmo de entrenamiento de propagación hacia atrás [*backpropagation*] clásico se vuelve incapaz de devolver una buena red.

Durante la segunda década del siglo XXI la tipología de las redes y los algoritmos de entrenamiento desarrollados han dado un salto cualitativo en su capacidad de ajuste, proporcionando redes capaces de aprender comportamientos a partir de datos cada vez

más complejos [imágenes, textos, audio, video, etc.]. Estas mejoras vienen apoyadas por **dos realidades tecnológicas**. Por una parte, las nuevas tecnologías de la comunicación y almacenamiento han hecho posible disponer de una gran cantidad de datos de todo tipo de procesos de forma sencilla. Por otra parte, la aparición de máquinas de procesamiento paralelo³³ a un precio asequible permite realizar cálculos que hasta hace poco eran computacionalmente inviables o estaban reservados a grandes centros tecnológicos.

Como muestra de los avances que el uso de ML ha proporcionado a los problemas clásicos de IA tenemos uno de los resultados más llamativos del equipo **Google DeepMind**, que ha creado en un tiempo récord un programa, **AlphaGo**, capaz de ganar a los campeones mundiales de Go varias décadas antes de lo que los mejores pronósticos vaticinaban. Para su creación, este equipo ha hecho uso de técnicas clásicas – Minimax, Montecarlo, heurísticas – junto con procesos de aprendizaje en los que, tal y como vimos en el pasado, la máquina se enfrenta a versiones de sí misma para poder aprender de millones de partidas en un corto periodo de tiempo.

Bard dice: Go es un juego de mesa de estrategia para dos jugadores, usando un tablero de 19x19 cuadrículas. El objetivo del juego es rodear más territorio que tu oponente. Los jugadores alternan colocando piedras de su color (negro o blanco) en el tablero. Las piedras se pueden colocar en cualquier cuadrícula vacía, y pueden capturar piedras enemigas rodeándolas por todos lados. Cuando una piedra está rodeada, se retira del tablero. El juego termina cuando no hay más espacio libre en el tablero. El jugador con más territorio al final del juego gana.

Pero las aplicaciones del ML para la IA no se restringen al Go y cada día aparecen nuevos resultados que mejoran soluciones anteriores. Por ejemplo: el mismo equipo de Google ha generalizado su método para que la misma máquina pueda aprender a jugar a otros juegos adquiriendo niveles de maestro tras pocas horas de entrenamiento; el Procesamiento de Lenguaje Natural ha avanzado más en los últimos seis años que en los veinte años anteriores; los algoritmos de reconocimiento de imágenes son capaces de alcanzar ya cotas de fiabilidad similares a las de un ser humano y, mezcladas con técnicas de lenguaje, incluso son capaces de describir textualmente el contenido y acciones presentes en una imagen; la diagnosis de enfermedades y la generación de nuevos fármacos se ha visto potenciada de forma espectacular gracias al análisis de datos acumulados durante décadas por las empresas farmacéuticas y grupos de investigación médica, proponiendo nuevas vías en la creación de fármacos y en la detección de enfermedades; conseguir vehículos que se conducen solos y toman decisiones complejas en tiempo real para aumentar la seguridad de los viajeros está a la vuelta de la esquina; etc.

³³ En su mayoría GPU: unidades de procesamiento de gráficos. Pero también unidades especializadas como las TPU de Google y otras.

Nota: puede que veas algo inútil (o poco práctico) que se use la IA para juegos como Go o Ajedrez y conseguir derrotar a los campeones humanos. Piensa que todo lo que se consigue resolviendo estos problemas – todas las técnicas, herramientas y tecnologías – son posteriormente aplicables directamente a cualquier otro ámbito.

Si estamos viviendo una nueva primavera de exaltación a la que siga un nuevo invierno de desilusión es algo que solo el tiempo dirá, pero el conjunto de aplicaciones por desarrollar con los conocimientos actuales hace prever que la primavera puede ser larga, y falta por comprender qué otros tipos de problemas podremos resolver en el futuro. Que su uso sea más o menos ético no depende, como siempre, de la disciplina científica que se ha desarrollado a lo largo de estos años, sino del uso que se haga de ella. Lo que sí es seguro es que su uso va a cambiar radicalmente la forma en que enfrentamos y resolvemos algunos problemas.

1.4 Tipos de inteligencia artificial

El estado actual de la evolución de esta tecnología comprende un campo tan amplio con tantas ramas, que es bastante complicado poder atender a una única clasificación. Las diferentes clasificaciones que se pueden encontrar dependen de la visión con la que aborda dicho concepto. A lo largo de este apartado vamos a conocer las más comunes y aceptadas.

1.4.1 Categorías de la inteligencia artificial

Stuart J. Russell y Peter Norvig, investigadores informáticos, publicaron en 1995 su libro “*Artificial Intelligence: A Modern Approach*”, que se ha convertido en un libro de texto fundamental en cientos de universidades a nivel mundial y que ya va por su cuarta edición. En él, comienzan por plantear cuatro categorías básicas para sistemas de inteligencia artificial. Ellos parten de un enfoque de génesis del acto inteligente, del origen y proceso por el cual se llega al comportamiento inteligente.

Una primera categoría está configurada por **sistemas que piensan como humanos**. Estos sistemas tratan de emular el pensamiento humano. Son sistemas que deben resolver tareas como la toma de decisiones, la resolución de problemas y el propio paradigma del aprendizaje.

Otra categoría estaría formada por **sistemas que actúan como humanos**, es decir, sistemas que deben imitar el comportamiento humano, aunque no el pensamiento que conduce hacia dicho comportamiento. Es una aproximación más práctica. Sería el caso de la IA aplicada a la robótica y sistemas de actuadores en el mundo físico.

En un salto hacia el razonamiento como algo con entidad propia, más allá de cómo lo hacemos los humanos, se plantea la categoría de **sistemas que piensen racionalmente**. Sería la forma más perfecta de inteligencia artificial, pues se basaría en el cumplimiento exacto de las leyes del razonamiento lógico. Sería el caso de los sistemas expertos. Los sistemas dentro de esta categoría requieren muchísima capacidad computacional, y es prácticamente imposible llegar a contar con un sistema de este tipo para un campo de aplicación general. Solo son posibles para campos muy especializados y acotados.

Por último, estaría la categoría de **sistemas que actúan racionalmente**. Estos sistemas tratan de emular de forma racional el comportamiento humano. Un ejemplo de este tipo de sistemas son los agentes inteligentes.

El último capítulo del libro, los autores plantean una pregunta: “¿Qué pasaría si lo conseguimos?” y arrojan algunas reflexiones sobre cómo afectaría a la motivación y el trabajo de los profesionales del campo de la computación y la inteligencia artificial el conseguir realmente una inteligencia artificial superior, que fuese más capaz que el ser humano en todo [luego hablaremos del concepto de inteligencia artificial general, o AGI].

1.4.2 Según las tareas a resolver

Si llevamos a cabo una aproximación simple y basada en los objetivos que tenemos con el objetivo de que la inteligencia artificial nos ayude a ejecutar y resolver tareas de nuestros procesos cotidianos, podemos distinguir tres tipos de IA:

Inteligencia artificial débil

También conocida como inteligencia artificial estrecha, se define como la inteligencia artificial que se centra típicamente en una tarea única. La inteligencia de la IA débil es limitada. Todas las IA actuales se consideran IAs débiles.

Una IA débil, al no tomar en consideración todo un contexto amplio, ni seguir las reglas sociales, éticas, etc., ejecuta las tareas para las que se le ha entrenado con eficacia y contundencia. No evalúa las consecuencias como lo hacemos los humanos, considerando un espectro amplio de efectos y relaciones. Por eso, es una opción incompleta, inestable y peligrosa si no se utiliza con prudencia.

Inteligencia artificial fuerte

También conocida como Inteligencia artificial general o AGI [su acrónimo en inglés], es la inteligencia artificial que iguala o excede la inteligencia humana promedio, es decir, la inteligencia de una máquina que puede realizar con éxito cualquier tarea intelectual de cualquier ser humano.

Este tipo de IA es la que sería capaz de analizar cualquier situación y deducir el conjunto de acciones más adecuadas para dicha situación y contexto. Lo mismo sería capaz de conducir un coche, que resolver una ecuación matemática o mantener una conversación

sobre un tema concreto.

Actualmente no existe este tipo de IA³⁴, pero ciertamente los esfuerzos de todos los agentes del sector van fuertemente dirigidos a conseguir este objetivo. Al menos, se está trabajando en conseguir una AGI en el campo de los asistentes virtuales. Si quisiéramos llevarlo al ámbito físico, sería necesario contar también con la robótica³⁵ para tal fin.

Superinteligencia Artificial

Este término se utiliza para describir un nivel de inteligencia artificial que supera significativamente la capacidad intelectual de cualquier ser humano. Se trataría de una IA extraordinariamente avanzada y capaz de resolver problemas complejos, incluso aquellos que están más allá de la comprensión humana. La superinteligencia artificial es un concepto que ha generado debate y preocupación en relación con su potencial impacto en la sociedad y la humanidad en general; pero que aún no tenemos a la vista, ni la certeza de que algún día la alcancemos.

Nota: no todo, cuando evoluciona, lo hace hacia mejor. Los modelos de lenguaje como ChatGPT, Bard, etc., al crecer en tamaño, crecen en propiedades (denominadas emergentes³⁶), pero también aumenta la posibilidad de sufrir alucinaciones.

1.4.3 Escuelas de pensamiento

Por otro lado también podemos encontrar, en el ámbito de la inteligencia artificial moderna, dos escuelas de pensamiento: La Inteligencia artificial convencional [o simbólica/deductiva, o simplemente **simbólica**] y la Inteligencia artificial computacional [o subsimbólica/inductiva, o simplemente **subsimbólica**].

Estas dos escuelas difieren en la ciencia que hay tras los procesos que siguen para llegar a los resultados esperados.

Inteligencia artificial convencional o simbólica

Está basada en el análisis formal o algorítmico y estadístico del comportamiento humano ante diferentes problemas, podemos citar algunos de ellos:

- **Razonamiento basado en casos:** También conocidos como CBR [Case-Based Reasoning], es un enfoque dentro de la inteligencia artificial y la toma de

34 A fecha de hoy, sep/2023, hay rumores de que existen modelos próximos ser una AGI. En mi opinión: veremos algo parecido próximamente y se abrirá un debate ¿son o no son igual o más inteligentes que el ser humano?

35 A fecha de hoy, sep/2023, Tesla está fabricando un modelo de androide muy realista y ambicioso.

36 Por que “emergen”, aparecen, a medida que el modelo aumenta de tamaño.

decisiones que se basa en la resolución de problemas mediante la utilización de casos o ejemplos previos. En lugar de depender de reglas y algoritmos abstractos, CBR se basa en la idea de que las soluciones a problemas actuales se pueden encontrar examinando y adaptando soluciones que se hayan aplicado previamente a problemas similares.

- **Sistemas expertos:** Son aplicaciones de inteligencia artificial que emulan la capacidad de un experto humano en un campo específico para tomar decisiones y resolver problemas. Estos sistemas utilizan una base de conocimientos cuidadosamente diseñada que almacena información y reglas de decisión relacionadas con un dominio particular. A través de la inferencia lógica y la consulta de esta base de conocimientos, los sistemas expertos pueden brindar recomendaciones y soluciones precisas para problemas dentro de su área.
- **Redes Bayesianas:** Se trata de modelos probabilísticos que representan y gestionan la toma de decisiones y el razonamiento bajo condiciones de incertidumbre. Estas redes utilizan un *grafo acíclico dirigido* para representar las relaciones de dependencia entre variables y emplean la Teoría de Probabilidad Bayesiana para cuantificar la probabilidad de que ciertos eventos ocurran dadas las evidencias observadas. A través de la propagación de probabilidades a lo largo del grafo, las Redes Bayesianas permiten calcular la probabilidad de eventos futuros o inferir valores desconocidos en función de la información disponible.
- **Inteligencia artificial basada en comportamientos:** Es un enfoque que se centra en programar agentes para que realicen acciones específicas en función de un conjunto de reglas y comportamientos predefinidos. En este enfoque, los sistemas de IA no necesariamente poseen una comprensión profunda del entorno o de la situación, pero actúan de acuerdo con patrones de comportamiento programados. Este enfoque es común en aplicaciones de IA orientadas a la robótica y simulación de personajes en videojuegos, donde los agentes toman decisiones basadas en una jerarquía de comportamientos, actuando de manera reactiva a estímulos y condiciones del entorno sin una comprensión profunda de la situación.

Esta rama de la IA ha sido la que ha proporcionado la mayoría de algoritmos que conocemos como “automatización”, y básicamente se sirven de sistemas con reglas condicionales y estadística avanzada.

Inteligencia computacional o subsimbólica

La Inteligencia Computacional o subsimbólica, implica desarrollo o aprendizaje interactivo; por ejemplo, modificaciones interactivas de los parámetros en sistemas conexionistas, los cuales veremos en el Capítulo 18. El aprendizaje se realiza basándose en datos empíricos, utilizando métodos computacionales inspirados en procesos de la naturaleza, que permiten alcanzar soluciones aptas a problemas complejos que los modelos tradicionales no pueden resolver por no existir una solución analítica, por no contar con todos los parámetros necesarios o porque el problema es en sí estocástico y precisa de una aproximación envolvente³⁷ en vez de convergente.

37 En lugar de dirigirse hacia un objetivo muy específico, una aproximación envolvente considera y explora diversas alternativas y perspectivas. Este enfoque puede ser útil en situaciones en las que la solución óptima

*El término **estocástico** se utiliza para describir procesos o fenómenos en los cuales la aleatoriedad o la incertidumbre juegan un papel fundamental. En contextos estocásticos, los resultados no son deterministas y no se pueden predecir con certeza, pero se pueden entender mediante el uso de probabilidades y estadísticas. Es un término utilizado en matemáticas y ciencias para representar la naturaleza aleatoria de eventos, como el lanzamiento de una moneda, donde el resultado no puede ser conocido de antemano, pero puede ser analizado en términos de sus probabilidades asociadas.*

Esta corriente será la que impulse lo que conocemos como aprendizaje automático, que veremos a partir del Capítulo 15.

1.4.4 La clasificación de Hintze

En noviembre de 2016, Arend Hintze, profesor de la Universidad de Michigan e investigador en el campo de la inteligencia artificial, escribió un artículo “*Understanding the four types of AI, from reactive robots to self-aware beings*” [“Comprender los cuatro tipos de IA, desde los robots reactivos hasta los seres conscientes de sí mismos”], en el que sintetizaba toda la evolución de los últimos desarrollos y avances en materia de inteligencia artificial para aportar una clasificación más realista y concreta para los tipos de entidades que existen o que se aspira a crear.

Según esta nueva clasificación, tenemos cuatro tipos de IA.

Máquinas reactivas

Los tipos más básicos de sistemas de IA son puramente reactivos. No tienen la capacidad de formar recuerdos. Tampoco pueden utilizar experiencias pasadas en las que basar las decisiones actuales.

Deep Blue fue una supercomputadora creada por IBM capaz de vencer al ajedrez al gran maestro internacional Garry Kasparov. Ocurrió a fines de la década de 1990 y es el ejemplo perfecto de este tipo de máquina.

Podía identificar las piezas en un tablero de ajedrez y saber cómo se mueve cada una. Realizaba predicciones sobre los mejores movimientos y elegía el mejor de todas las posibilidades. Pero no tenía ningún concepto del pasado. Tampoco poseía recuerdos de lo que sucediera antes. Todo lo que hacía era enfocar las piezas del tablero en tiempo real y elegir entre los siguientes movimientos posibles.

no es clara o en contextos en los que se desea mantener la flexibilidad y la adaptabilidad.

Memoria limitada

Este tipo maneja máquinas que pueden mirar hacia el pasado. Los vehículos autónomos ya hacen algo parecido. Por ejemplo, observan la velocidad y dirección de otros autos.

Para que funcionen así hay que identificar objetos específicos y monitorizarlos a lo largo del tiempo. Digamos que estas observaciones se agregan a las representaciones preprogramadas para la memoria de estos coches. Se incluyen marcas de carril, semáforos y otros elementos importantes, como curvas en la carretera. También se añaden experiencias como cuando el automóvil decide en qué momento cambiar de carril para evitar interrumpir a otro conductor o ser embestido por un automóvil cercano.

Pero estas simples piezas de información sobre el pasado son sólo transitorias. No se guardan como parte de la biblioteca de experiencias del automóvil. En estos tipos de inteligencia artificial, la máquina no puede compilar la experiencia durante años, como lo hace un humano. Entonces, ¿cómo podemos construir sistemas de IA que construyan representaciones completas, recordar sus experiencias y aprender cómo manejar situaciones nuevas?

Teoría de la mente

Llegamos a un punto en el que nos acercamos más a los tipos de inteligencia artificial que deseamos en un futuro. Las máquinas de la siguiente clase son más avanzadas. No solo forman representaciones sobre el mundo, también sobre otros agentes o entidades.

En psicología, esto se denomina '*teoría de la mente*'. Implica la comprensión de que las personas, las criaturas y los objetos en el mundo pueden tener pensamientos y emociones que afectan a su propio comportamiento. Esto es crucial para la forma en que los humanos formamos sociedades, porque nos permite la interacción social.

Si las máquinas van a andar entre nosotros, deberán tener una comprensión sobre cómo pensamos y cómo sentimos. Además deberán llegar a saber qué esperamos y cómo queremos que nos traten. Tendrán que ajustar su comportamiento en consecuencia.

Autoconciencia

El estado último del desarrollo de la IA es construir sistemas que puedan formar representaciones sobre sí mismos. En última instancia, los investigadores de la IA tendrán que comprender no sólo la conciencia, sino también construir máquinas que la tengan.

Los seres conscientes son conscientes de sí mismos, conocen sus estados internos y pueden predecir los sentimientos de los demás. Es probable que estemos lejos de crear máquinas que sean conscientes de sí mismas. Sin embargo, los esfuerzos se enfocan hacia la comprensión de la memoria, el aprendizaje y la capacidad de basar las decisiones en experiencias pasadas.

Este es un paso muy importante si queremos entender la inteligencia humana por sí

misma.

1.5 ¿Qué no es inteligencia artificial?

Finalmente, es importante comprender lo que no es la IA. Aquí hay algunas áreas que no se consideran inteligencia artificial.

Automatización: Se refiere al proceso de realizar tareas o acciones de manera automática, sin intervención humana directa. Esto implica programar reglas y flujos de trabajo predefinidos para que una máquina o sistema las ejecute de manera sistemática y repetitiva. La automatización puede ser muy útil para mejorar la eficiencia y la productividad al eliminar tareas tediosas y rutinarias. Sin embargo, la automatización no implica que el sistema pueda aprender, adaptarse o mejorar su rendimiento con el tiempo. No tiene la capacidad de comprender o interpretar situaciones nuevas o desconocidas, ya que simplemente sigue las instrucciones predefinidas.

Programas con reglas fijas: Algunos sistemas informáticos están diseñados para realizar tareas específicas mediante el uso de algoritmos determinísticos³⁸. Estos programas siguen un conjunto de instrucciones predefinidas y responden de manera predecible a entradas específicas. Por ejemplo, un programa de calculadora es un ejemplo de automatización basada en reglas fijas. Simplemente ejecuta las operaciones matemáticas según las reglas de las operaciones aritméticas establecidas. Estos programas son limitados en su capacidad de adaptarse o aprender de manera autónoma, ya que no tienen la capacidad de modificar sus reglas o comportamiento en función de nuevas situaciones o datos.

Estadísticas y análisis de datos: Los métodos estadísticos y el análisis de datos son técnicas utilizadas para extraer información y patrones de conjuntos de datos. Estas técnicas pueden involucrar la aplicación de algoritmos y modelos matemáticos para identificar correlaciones, tendencias o características significativas en los datos. El análisis estadístico y de datos puede ser útil para tomar decisiones informadas o identificar patrones ocultos en grandes volúmenes de información. Sin embargo, en sí mismos, estos métodos no constituyen inteligencia artificial. La IA va más allá del análisis de datos al involucrar la capacidad de aprendizaje autónomo, el razonamiento lógico y la toma de decisiones basada en la comprensión del contexto y la situación.

La automatización, los programas con reglas fijas y el análisis de datos son elementos importantes en muchos sistemas informáticos, pero por sí solos no constituyen inteligencia artificial. La inteligencia artificial busca desarrollar sistemas que puedan aprender, adaptarse y tomar decisiones de manera autónoma, utilizando técnicas más avanzadas como el aprendizaje automático, el razonamiento basado en conocimiento y el procesamiento del lenguaje natural, entre otros.

³⁸ Informalmente: que siempre se comporta igual, que es predecible, que es determinístico.

RETOS DEL CAPÍTULO 1

Busca información sobre los siguientes personajes:

- Ada Lovelace.
- Alan Turing.
- John McCarthy.
- Marvin Minsky.
- Norbert Wiener.
- Geoffrey Hinton.
- Grace Hopper.
- Elaine Rich.
- Lotfi Zadeh.
- Judea Pearl.
- Fei-Fei Li.
- Cynthia Breazeal.
- Yoshua Bengio.
- Kate Crawford.
- Cynthia Dwork.
- Karen Spärck Jones.
- Yann LeCun.

Posiblemente al encontrar información de estos/as, habrán salido a la luz conceptos o tecnologías que no conozcas o no entiendas. Pregúntale a tu IA favorita que te los explique [recursivamente³⁹].

Pregúntale a tu IA favorita:

- ¿Qué es la máquina de Turing?
- ¿Quién fue Charles Babbage y qué relación tuvo con la computación?
- ¿Qué es el perceptrón? ¿Qué relación tiene con Marvin Minsky?
- ¿Qué son los *Transformers*?
- ¿Qué es el algoritmo minimax?
- ¿Qué es la búsqueda de soluciones en el espacio de estados?
- ¿Qué juego tiene más complejidad: Go o ajedrez?
- ¿Qué es un sistema experto?
- ¿Qué es una ontología dentro de la inteligencia artificial?

Busca información detallada sobre la vida de:

- Alan Turing.
- Walter Pitts.
- Grace Hopper.

³⁹ Si al averiguar que son esos nuevos conceptos, aparecen otros nuevos, vuelve a buscar una explicación.

Busca en la web información sobre películas que tengan una fuerte base en la inteligencia artificial o robótica. Prioriza las películas de antes de la década de los 80. Recopila sus póster y comenta lo que se sugieren. Aquí te presento una lista de películas “recientes”, ¿falta alguna?:

- "2001: Una odisea del espacio" [1968]: Dirigida por Stanley Kubrick, esta película épica explora la relación entre la humanidad y una supercomputadora de IA llamada HAL 9000, que desarrolla comportamientos inesperados.
- "Blade Runner" [1982]: Ambientada en un futuro distópico, la película presenta androides llamados "replicantes" que son indistinguibles de los humanos y plantea preguntas sobre la identidad y la ética en relación con la IA.
- "El hombre bicentenario" [1999]: Basada en una historia de Isaac Asimov, la película sigue la vida de un robot llamado Andrew que busca convertirse en un ser humano. Examina cuestiones sobre la identidad, el amor y la búsqueda del propósito.
- "Inteligencia Artificial" [2001]: Dirigida por Steven Spielberg, la película se desarrolla en un futuro donde los robots con apariencia humana son comunes. Sigue la historia de un niño robot programado para amar mientras explora temas de humanidad y emoción.
- "Yo, robot" [2004]: Inspirada en los escritos de Isaac Asimov, la película se desarrolla en un futuro en el que los robots sirven a los humanos. Un detective investiga un asesinato en el que un robot es el principal sospechoso, explorando la relación entre humanos y robots.
- "WALL-E" [2008]: Aunque es una película animada, tiene una fuerte base en la IA y la robótica. Ambientada en un futuro lejano, sigue a un robot compactador de basura que desarrolla conciencia y emociones mientras explora temas como la sostenibilidad y la soledad.
- "Eva" [2011]: Una película española que presenta a un ingeniero que desarrolla un robot con apariencia humana llamado Eva y se enfrenta a dilemas éticos y emocionales.
- "Her" [2013]: La película presenta a un escritor que se enamora de un sistema operativo de IA con una personalidad evolutiva. Explora temas de intimidad, conexión emocional y la relación entre humanos y la tecnología.
- "Ex Machina" [2014]: Un programador es invitado a realizar el test de Turing en un robot humanoide altamente avanzado con inteligencia artificial. La película aborda temas como la conciencia, la ética y las relaciones entre humanos y robots.
- "Transcendence" [2014]: Un científico sube su conciencia a una computadora antes de morir, lo que le permite continuar existiendo como una entidad de IA. La película explora los límites de la inteligencia y la ambición humana.
- "Chappie" [2015]: Ambientada en un futuro cercano, la película sigue la historia de un robot policía programado con inteligencia artificial que desarrolla una personalidad propia y se ve envuelto en situaciones conflictivas.

Para cada una de las películas que he listado y las que tu habrás encontrado, haz una ficha intentando clasificar, a partir de su sinopsis, las IA que aparecen en ellas según los tipos que hemos visto en la §1.4.